

Flow-based network analysis of the *Caenorhabditis elegans* connectome - Supplementary Information

Karol A. Bacik,^{1, *} Michael T. Schaub,^{2, 3, †} Mariano Beguerisse-Díaz,^{1, ‡} Yazan N. Billeh,⁴ and Mauricio Barahona^{1, §}

¹*Department of Mathematics, Imperial College London, London SW7 2AZ, United Kingdom*

²*naXys & Department of Mathematics, University of Namur, B-5000 Namur, Belgium*

³*ICTEAM, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium*

⁴*Computation and Neural Systems Program, California Institute of Technology, CA 91125 Pasadena, USA*

(Dated: December 8, 2015)

* karol.bacik13@imperial.ac.uk

† michael.schaub@unamur.be

‡ m.beguerisse@imperial.ac.uk

§ m.barahona@imperial.ac.uk

S1. MULTISCALE COMMUNITY STRUCTURE OF THE *C. ELEGANS* CONNECTOME WITH MARKOV STABILITY

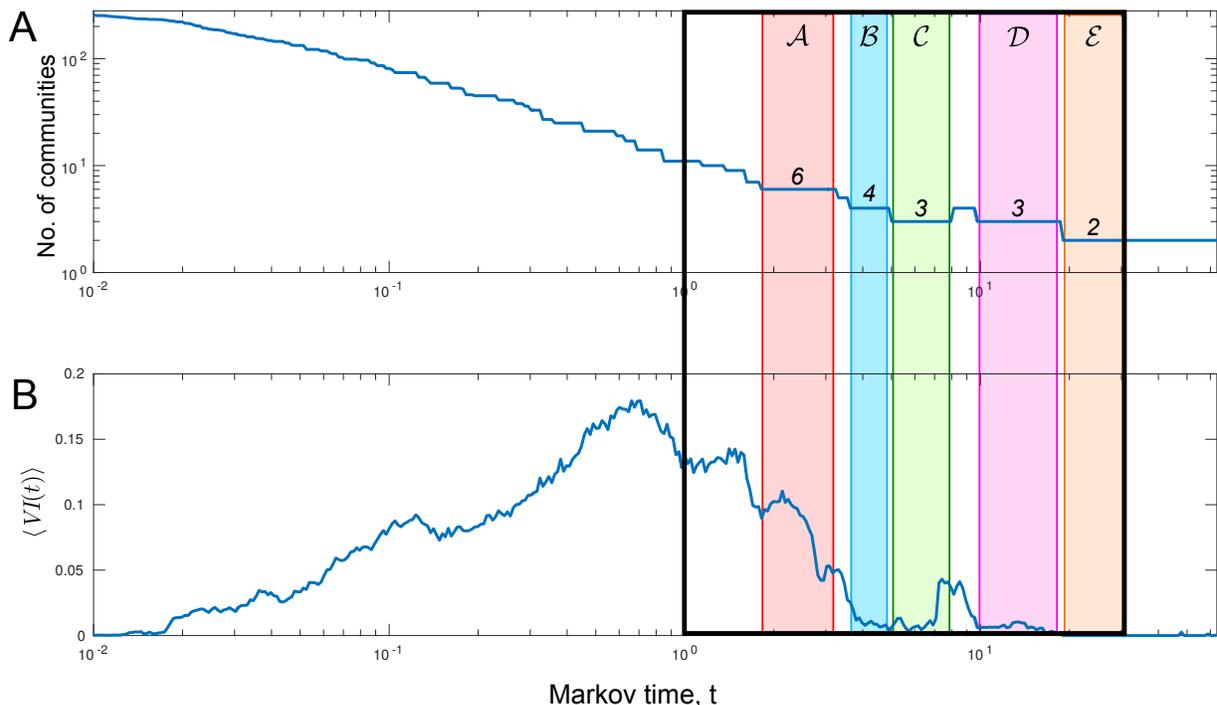


Figure S1. Full analysis with Markov Stability (MS) across all Markov times, from the finest possible partition (every node in its own partition) at small Markov times to the bipartition at large Markov times. The highlighted time interval corresponds to Fig. 1 in the main text (Fig. 1), which focusses on the medium to coarse scales.

Full scan of Markov times: Figure S1 displays the full MS analysis of the *C.elegans* connectome, from the finest to the coarsest scales. The analysis in the main text concentrates on the medium to coarse partitions $\mathcal{A} - \mathcal{E}$.

Quasi-hierarchical character of partitions: Figure S2 shows the conditional entropy of the partitions found by MS. The normalized conditional entropy $\Omega(\mathcal{P}(t')|\mathcal{P}(t))/\log(n) \in [0, 1]$ quantifies how much uncertainty there is about the community assignment in $\mathcal{P}(t')$, given the known partition $\mathcal{P}(t)$. If partition $\mathcal{P}(t')$ can be predicted from $\mathcal{P}(t)$, i.e. if $\mathcal{P}(t')$ is a strictly hierarchical agglomeration of the communities of $\mathcal{P}(t)$, then the conditional entropy will be zero. As seen in Figure S2, $\Omega(\mathcal{P}(t')|\mathcal{P}(t))/\log(n)$ has a strong upper-triangular character, implying that the communities are almost hierarchical.

Comparison with other partitions: The flow-based MS partitions obtained here are distinct from partitions obtained by several other methods. In particular, we have compared against partitions obtained with Modularity, Stochastic Block models, and Infomap.

Modularity has been used to obtain optimised partitions in Refs. [1, 2]. The partition found in Ref. [1] is closest to our 4-way Partition \mathcal{B} ($VI = 0.185$), whereas the partition found in Ref. [2] is closest to our 3-way Partition \mathcal{C} ($VI = 0.186$). Modularity optimisation imposes a particular intrinsic scale (or resolution) to the partition, so that partitions found with modularity are well matched to a particular scale (i.e., a particular Markov time) in the Markov Stability framework, as shown previously [3, 4]. On the other hand, as discussed in the main text, the Markov Stability framework is based on a systematic scanning across Markov times [5] allowing the intrinsic multiscale organisation to become apparent.

The partitions based on stochastic block models [6] and hierarchical Infomap [7] are less similar to the ones found by MS: the partition found by stochastic block models in [6] is closest to our 3-way Partition \mathcal{C} (but with a higher $VI=0.272$), and the partition found by hierarchical Infomap in [7] is closest to our 6-way Partition \mathcal{A} (yet with an even higher $VI=0.282$). These differences in the outcomes are expected due to the contrasting methodological approaches. In particular, Infomap is known to impose a clique-like structure to the modules leading to groupings where strong local density is favoured [8].

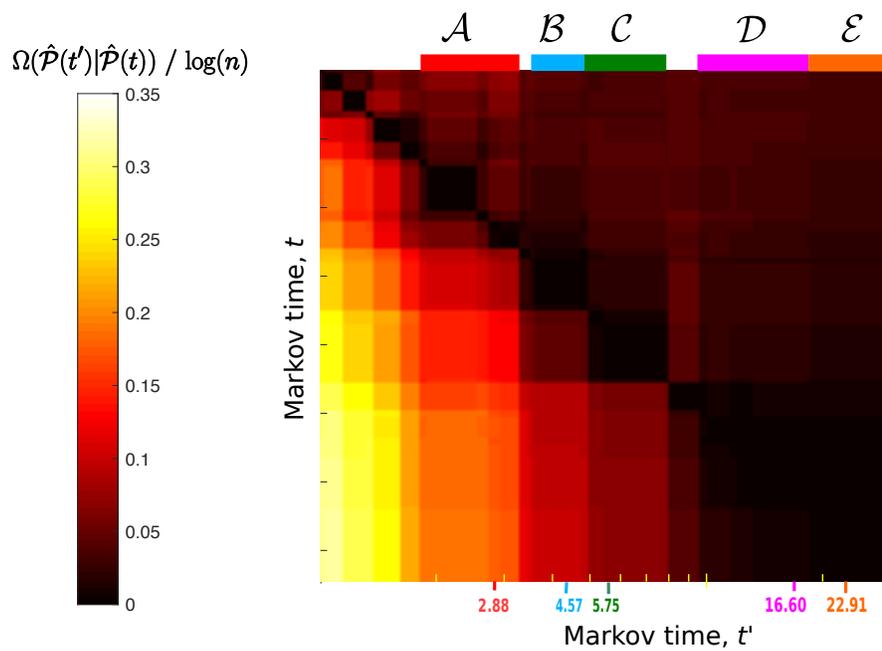


Figure S2. Normalised conditional entropy of the optimised MS partitions $\mathcal{A} - \mathcal{E}$. The observed asymmetry in this measure implies a quasi-hierarchical organisation (see text).

S2. ROLES OF NODES: THE RBS FRAMEWORK

Extracting flow roles in directed networks using RBS: Figure S3 summarises schematically the procedure to obtain flow roles using RBS analysis, as discussed in detail in [9]. First, a similarity network is created by computing a similarity score between each node in the network, based on their incoming and outgoing path-profiles, as described in the main text. Second, a similarity graph is obtained using the RMST method, which subsequently prunes out uninformative links (see Ref. [9] for details). Third, the resulting similarity graph is clustered using MS to obtain relevant groups of nodes that have similar in- and out-flow profiles at all scales.

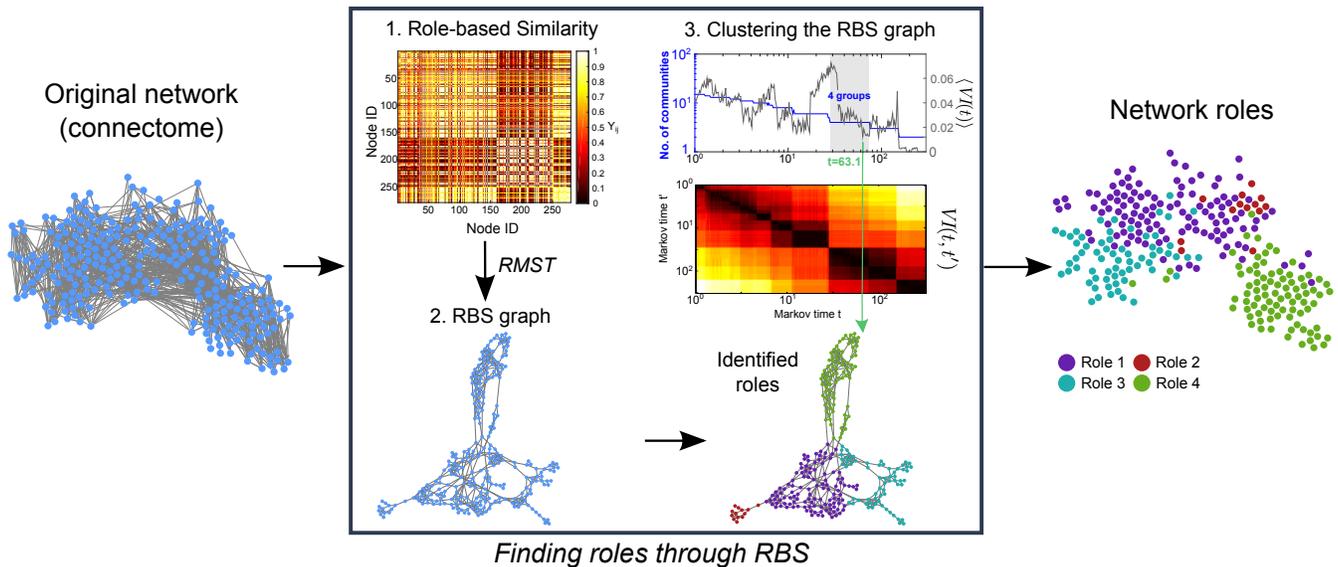


Figure S3. Role extraction with RBS, RMST, and Markov Stability. Left to right: From the original *directed* network of the *C. elegans* connectome we obtain a similarity matrix using the RBS metric. This similarity matrix is transformed into a similarity matrix using the RMST method. When Markov Stability is applied to the similarity graph, we find a robust partition into four communities, corresponding to four *flow roles*. The roles are then shown on the original connectome layout.

Comparison with other analyses of roles: Our *flow roles* are fundamentally different from notions of roles used in social networks based on Structural Equivalence (SE) [10], and Regular Equivalence (RE) [11]. Figure S4 presents the comparison of our RBS roles versus those obtained based on RE by the REGE algorithm [12] (see also the Supplementary Data). Because both RE and SE consider only one-step neighbourhoods and do not incorporate information about the long scales of the network [5], they are less applicable to complex networks such as the *C. elegans* connectome [13]. In particular the roles produced by REGE show undifferentiated PageRank and connectivity profiles.

In Refs. [1, 2], roles were assigned to neurons according to the technique proposed by Guimera et al [14], identifying certain interneurons as relevant hubs between predefined communities. In Ref. [1] command interneurons (e.g. AVA, AVB, AVD, PVC) play the role of global hubs, whereas D-type motor neurons play the role of provincial hubs. These features are in line of our ablation results, where D-type motor neuron ablations alter flows at finer scales and ablation of interneurons modifies flow patterns at larger scales. Chatterjee and Sinha [15] explored the core-periphery structure of the *C. elegans* connectome using a k -core decomposition based on in- and out-degree separately. The k -core of a network is the subgraph with the property that all nodes have (in/out)degree at least k . As expected, motor neurons are overrepresented in the k -cores based on in-degree, and sensory neurons are overrepresented in k -cores based on out-degree. This distinction between neurons with upstream and downstream roles is also an inherent characteristic in the RBS analysis, yet from a different perspective, i.e., based on the global characteristics of a node with respect to the in- and out-flows in the network, rather than based on its local connections.

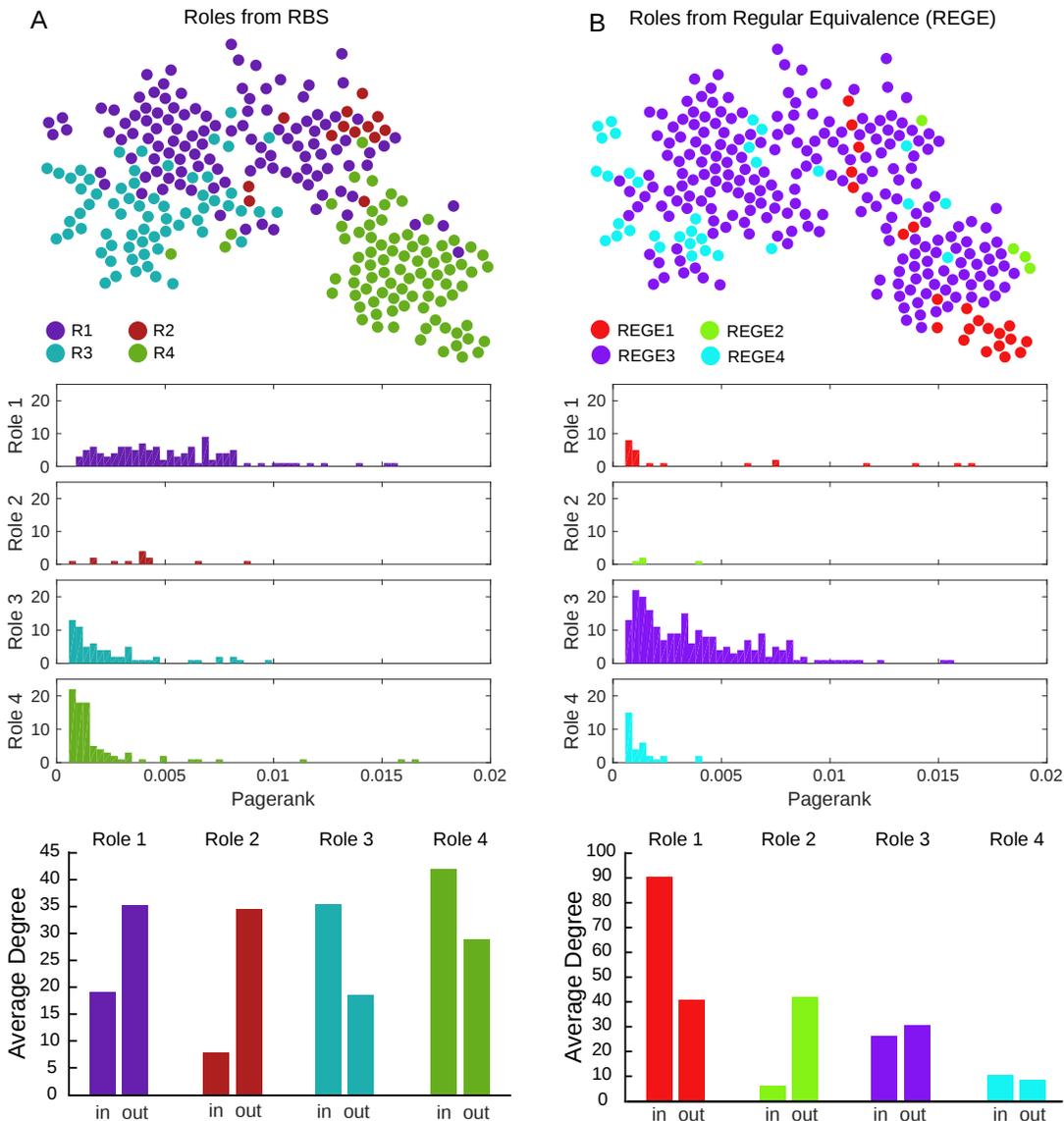


Figure S4. **A:** Roles of the nodes according to RBS with the PageRank distribution for each role and the average in/out degree for each role. **B:** Same for the roles obtained according to Regular Equivalence obtained using the REGE algorithm [12].

S3. STYLISTED SIGNAL PROPAGATION ANALYSIS

Summarising signal propagation: Signal propagation analysis can be analysed from different angles. In Figure S5, we explain concisely our approach to detect pathways of signal flow exemplified on scenario (i1).

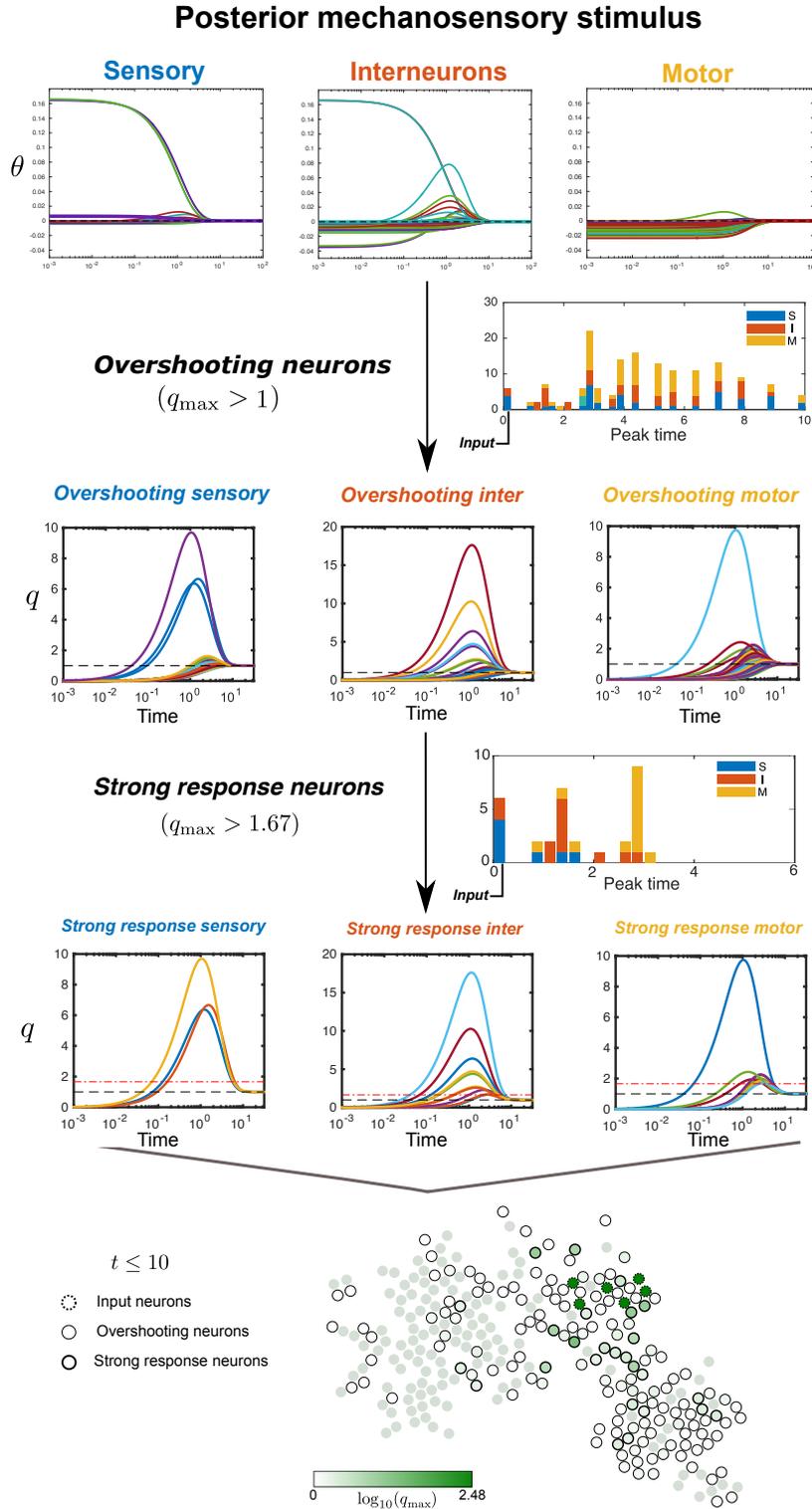


Figure S5. Operational procedure of signal propagation analysis (see text).

For all neurons, we compute the time series $\phi_i(t)$, i.e., the amount of signal present at the node at Markov time t . With increasing time the amount of signal at each node will converge to its stationary value ($\theta_i(t) = \phi(t) - \pi_i \rightarrow 0$). The approach to stationarity can happen in two ways: i) the initially negative $\theta_i(t)$ approaches 0 from below; ii) $\theta_i(t)$ ‘overshoots’, i.e. exceeds its stationary value temporarily before decaying towards its stationary value. We define the relative amount of signal with respect to the stationary value as $q_i(t) = \phi(t)/\pi_i$. We focus on nodes that have overshoot above their stationary value π_i , i.e., those with $q_{\max} := \max_t \phi_i(t)/\pi_i > 1$, and we collect the times at which they reach their peak. To summarise the signal propagation more concisely, we focus on *strong response neurons* that present a large overshoot, i.e., $q_{\max} > 5/3$, and collect the peak times. The peak-time histogram and the particular sequence of strong response neurons is used to characterise the different input-response biological scenarios. Analyses by neuron type and by role groups is also carried out. See Figure S5 for an illustration of these steps applied to the biological scenario (i1), corresponding to the posterior mechanosensory stimulus.

Comparison of signal propagation for the different biological scenarios: Figures S6, S7, and S8 summarise the signal propagation results for the input-response case studies (i2)-(i4), in the same way as done in Figure 7 in the main text.

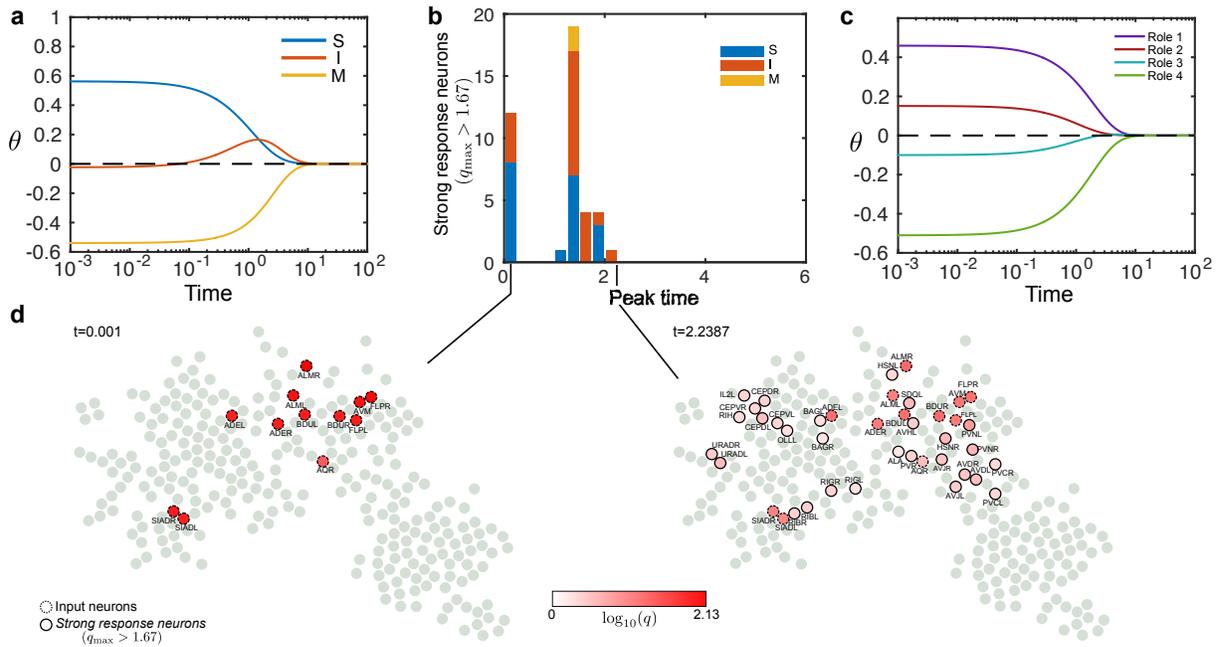


Figure S6. **Signal propagation: anterior mechanosensory stimulus (i2).** Signal propagation evolving from an initial condition localised at the mechanosensory neurons (i2). **(a)** As stationarity is approached ($\theta(t) \rightarrow 0$), the input propagates from sensory to motor neurons through an intermediate stage when interneurons overshoot. **(b)** The propagation seen as a cascade of strong response neurons ($q_{\max} > 1 + 2/3$) with peak times concentrated around two bursts. **(c)** The input (i2), appears localised on R1 and to a lesser extent R2 neurons. The signal diffuses somewhat quicker out of R2 than R1 neurons, but induces not collective overshoot of R3 or R4 neurons. **(d)** Stages of signal propagation in the network showing the strong response neurons that have peaked at each time.

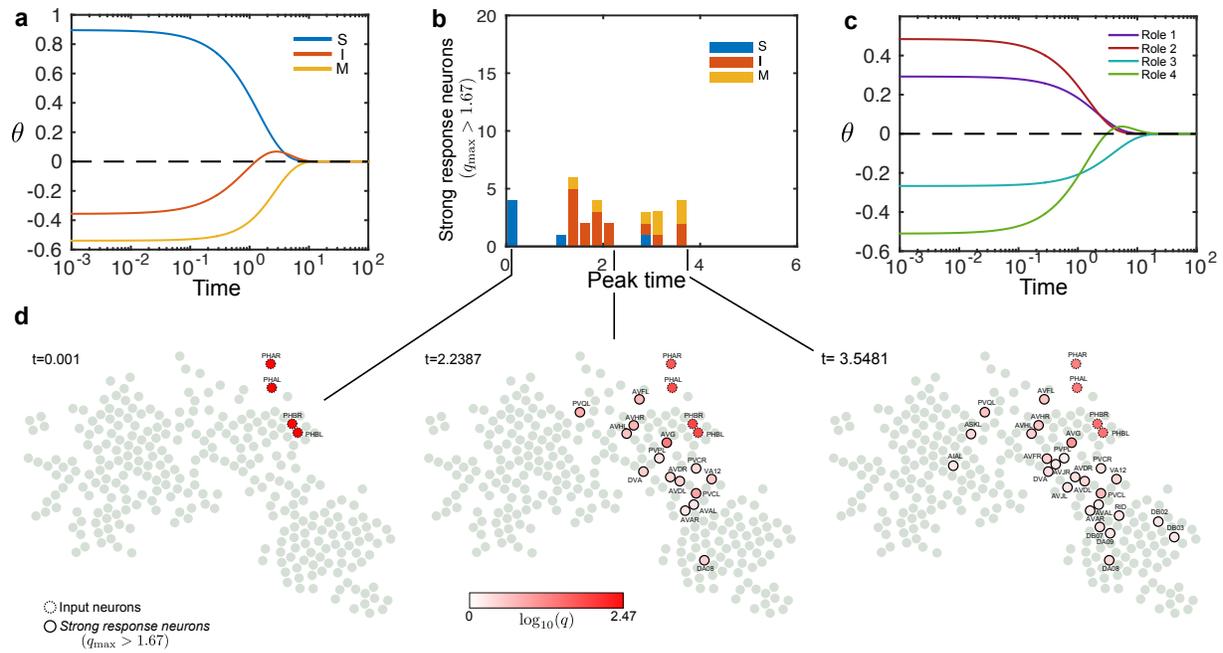


Figure S7. **Signal propagation: posterior chemosensory stimulus (i3).** See Caption Figure S6.

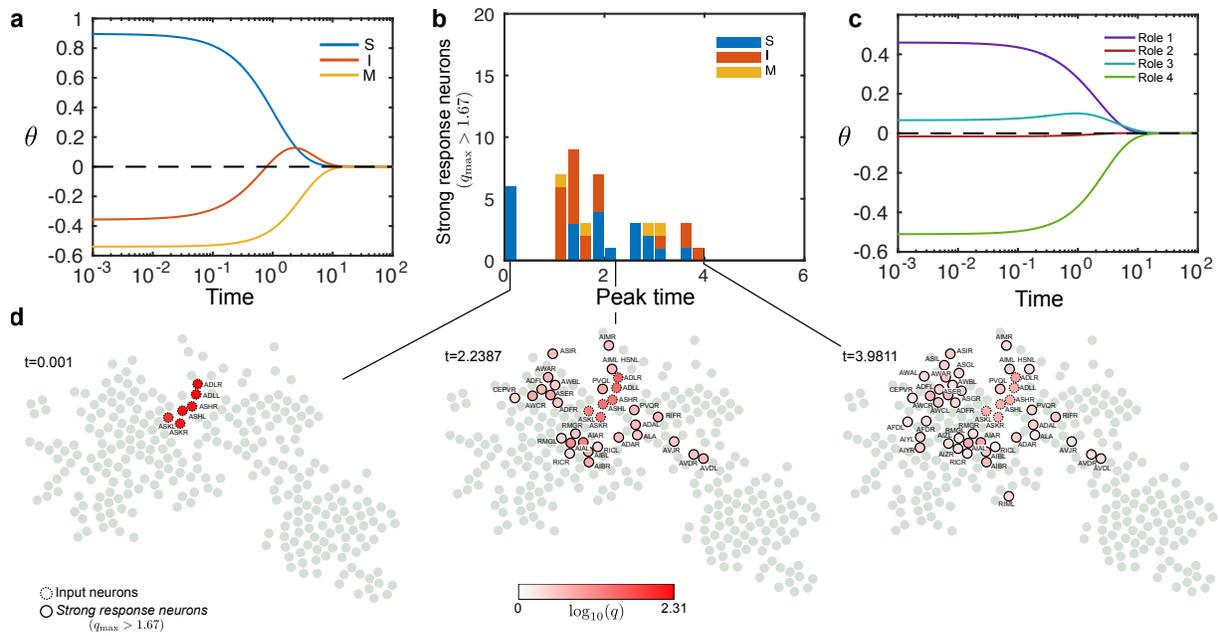


Figure S8. **Signal propagation: anterior chemosensory stimulus (i4).** See Caption Figure S6.

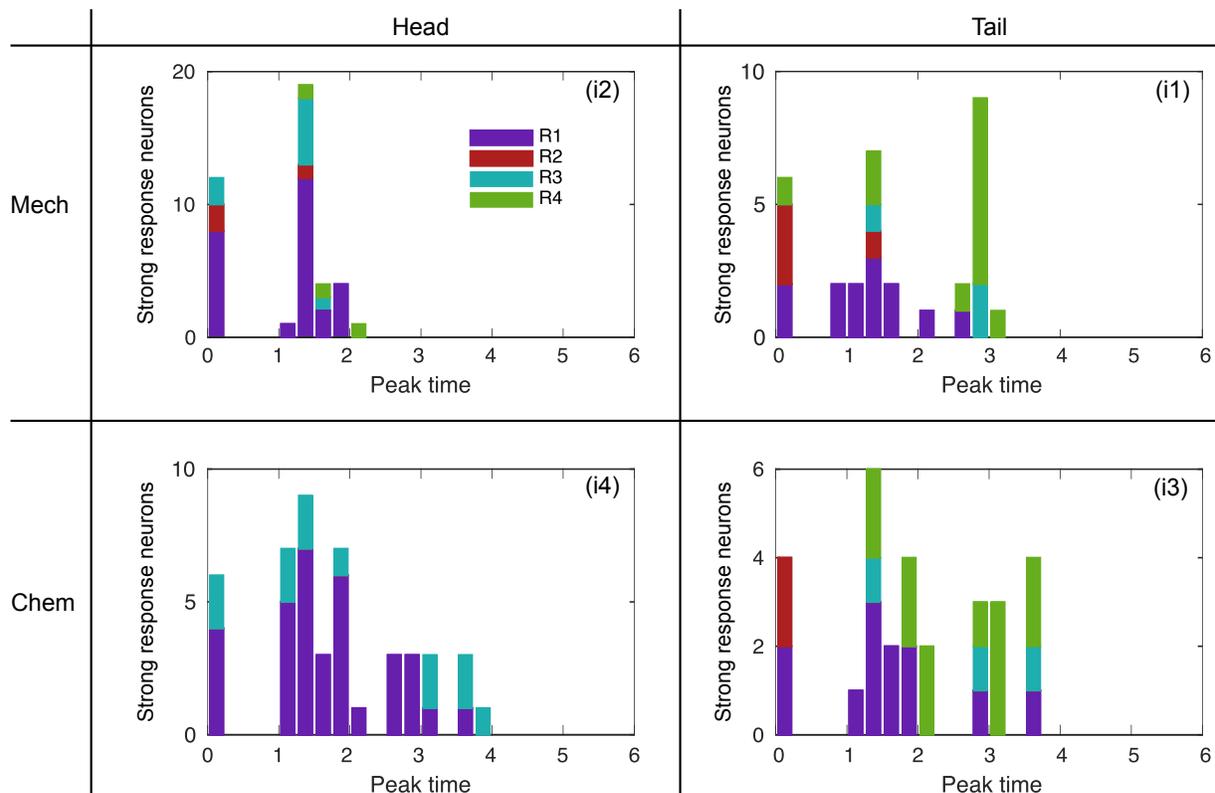


Figure S9. Peak times of strong response neurons summarised by RBS roles for each of the four input scenarios (i1)-(i4).

Figure S9 provides a comparison of the histograms of peak times for the strong response neurons in the four biological scenarios from the perspective of flow roles. Note how the tail inputs induce strong responses on neurons spreading from R2 to R1 and finally R4. On the other hand, the head inputs induce strong responses on neurons heavily based on R1 spreading only to R3.

-
- [1] R. K. Pan, N. Chatterjee, and S. Sinha, PLOS ONE **5** (2010).
 - [2] Y. Sohn, M.-K. Choi, Y.-Y. Ahn, J. Lee, and J. Jeong, PLoS Comput. Biol. **7** (2011).
 - [3] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, Proceedings of the National Academy of Sciences **107**, 12755 (2010), arXiv:0812.1811.
 - [4] R. Lambiotte, J. Delvenne, and M. Barahona, Network Science and Engineering, IEEE Transactions on **1**, 76 (2014), see also arXiv:0812.1770.
 - [5] M. T. Schaub, J.-C. Delvenne, S. N. Yaliraki, and M. Barahona, PLoS ONE **7**, e32210 (2012).
 - [6] D. M. Pavlovic, P. E. Vertes, E. T. Bullmore, W. R. Schafer, and T. E. Nichols, PLoS ONE **9(7)** (2014).
 - [7] D. Edler and M. Rosvall, "The MapEquation software package, available online at <http://www.mapequation.org>."
 - [8] M. T. Schaub, R. Lambiotte, and M. Barahona, Phys. Rev. E **86**, 026112 (2012).
 - [9] M. Beguerisse-Díaz, B. Vangelov, and M. Barahona, in *2013 IEEE Global Conference on Signal and Information Processing (GlobalSIP)* (2013) pp. 937–940.
 - [10] F. Lorrain and H. C. White, The Journal of Mathematical Sociology **1**, 49 (1971).
 - [11] M. G. Everett and S. P. Borgatti, The Journal of Mathematical Sociology **19**, 29 (1994).
 - [12] S. P. Borgatti and M. G. Everett, Social Networks **15**, 361 (1993).
 - [13] M. Beguerisse-Díaz, G. Garduño Hernández, B. Vangelov, S. N. Yaliraki, and M. Barahona, J R Soc Interface **11** (2014), 10.1098/rsif.2014.0940.
 - [14] R. Guimera and L. A. N. Amaral, Nature **433**, 895 (2005).
 - [15] N. Chatterjee and S. Sinha, "Understanding the mind of a worm: hierarchical network structure underlying nervous system function in *C. elegans*," in *Models of Brain and Mind - Physical, Computational and Psychological Approaches*, Progress in Brain Research, Vol. 168 (Elsevier, 2007) pp. 145–153.