

Stochastic Blockmodeling of the Models and Core of the *Caenorhabditis elegans* Connectome

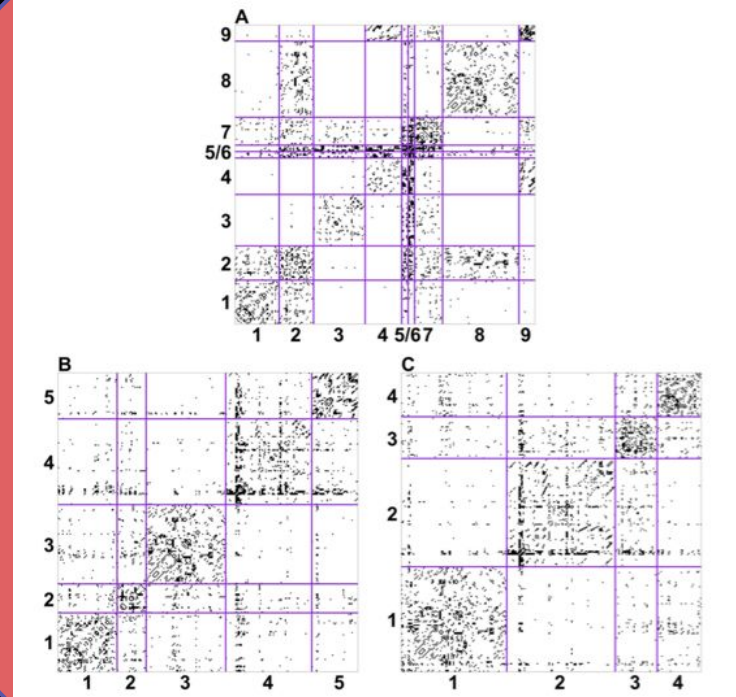
Dragana M. Pavlovic, Petra E. Vertes, Edward T. Bullmore, William R. Schafer, Thomas E. Nichols

As presented by Kristin Gunnarsdottir and Greg Kiar on March 24th, 2015

Summary

Comparison of clustering techniques on *C. elegans* network

- ERMM
- Spectral Algorithm
- Fast Louvain Algorithm



Opportunity



“... network analysis has great potential for addressing some of the key questions in neuroscience...”

Available and untested algorithms for biological network clustering



Challenge

Only one complete nervous system mapped to the cellular scale (even analysis of *C. elegans*, 300 nodes, > 2000 edges, is non-trivial)

Similarity metric?

Exhaustive search (for deterministic solutions) is astronomically large

Action

Spectral

Spectral Algorithm

$$\Delta f_{mod} = \frac{1}{4m} \sum_{i,j \in g} [D_{ij} - \delta_{ij} \sum_{i,j \in g} D_{ik}] s_i s_j = \frac{1}{4m} \mathbf{s}^T \mathbf{D}^{(g)} \mathbf{s}$$

$$D_{ij} = A_{ij} - \frac{\rho(V_i)\rho(V_j)}{2m}$$

ERMM

Adjacency matrix:

$$\mathbf{X} = \left((X_{ij}) \right)_{1 \leq i \neq j \leq n}$$

Group assignment

$$\mathbf{Z}_i = (Z_{i1}, \dots, Z_{iQ}), \sum_q \mathbf{Z}_{iq} = \mathbf{1}$$

$$\mathbf{Z}_i \sim M(1, \alpha) \text{ where } \alpha = (\alpha_1, \dots, \alpha_Q), \sum_{q=1}^Q \alpha_q = 1$$

Connectivity matrix

$$\pi = \left((\pi_{ql}) \right)_{1 \leq q, l \leq Q}, \text{ i. e. } X_{ij} | Z_{iq} = 1, Z_{jl} = 1 \sim \text{Bernoulli}(\pi_{ql})$$

$$ICL(M_Q) = \max_{\psi} \log [\mathcal{L}(\mathbf{x}, \hat{\mathbf{z}} | \mathcal{M}_Q; \psi)] - \frac{1}{2} \frac{Q(Q+1)}{2} \log \left[\frac{n(n-1)}{2} \right] - \frac{Q-1}{2} \log[n]$$

Fast Louvain

Modularity

$$f_{mod} = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{\rho(V_i)\rho(V_j)}{2m} \right) \delta(c_i, c_j)$$

Action

ARI

$$ARI = \frac{RI - E(RI)}{Max(RI) - E(RI)}$$

Where,

$$RI = \frac{TP + TN}{\binom{n}{2}}$$

ICC

$$ICC = \frac{Var(a_q)}{Var(Y_{qi})} = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_\epsilon^2}$$

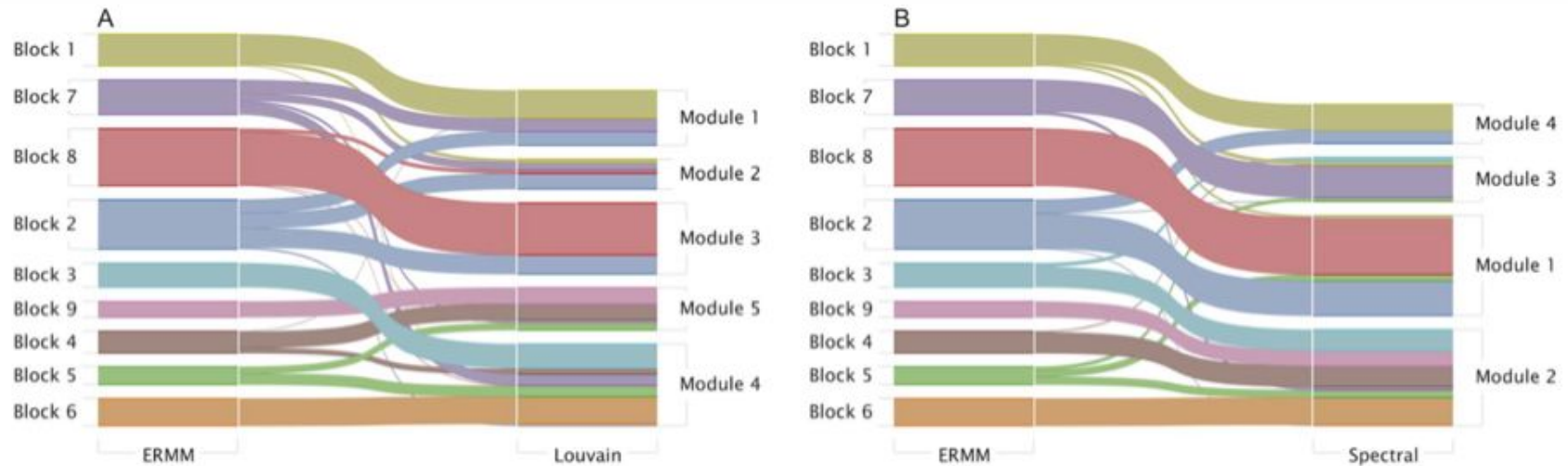
AIC

$$AIC(w_h) = -2 \log[\mathcal{L}(w_h; y)] + 2p$$

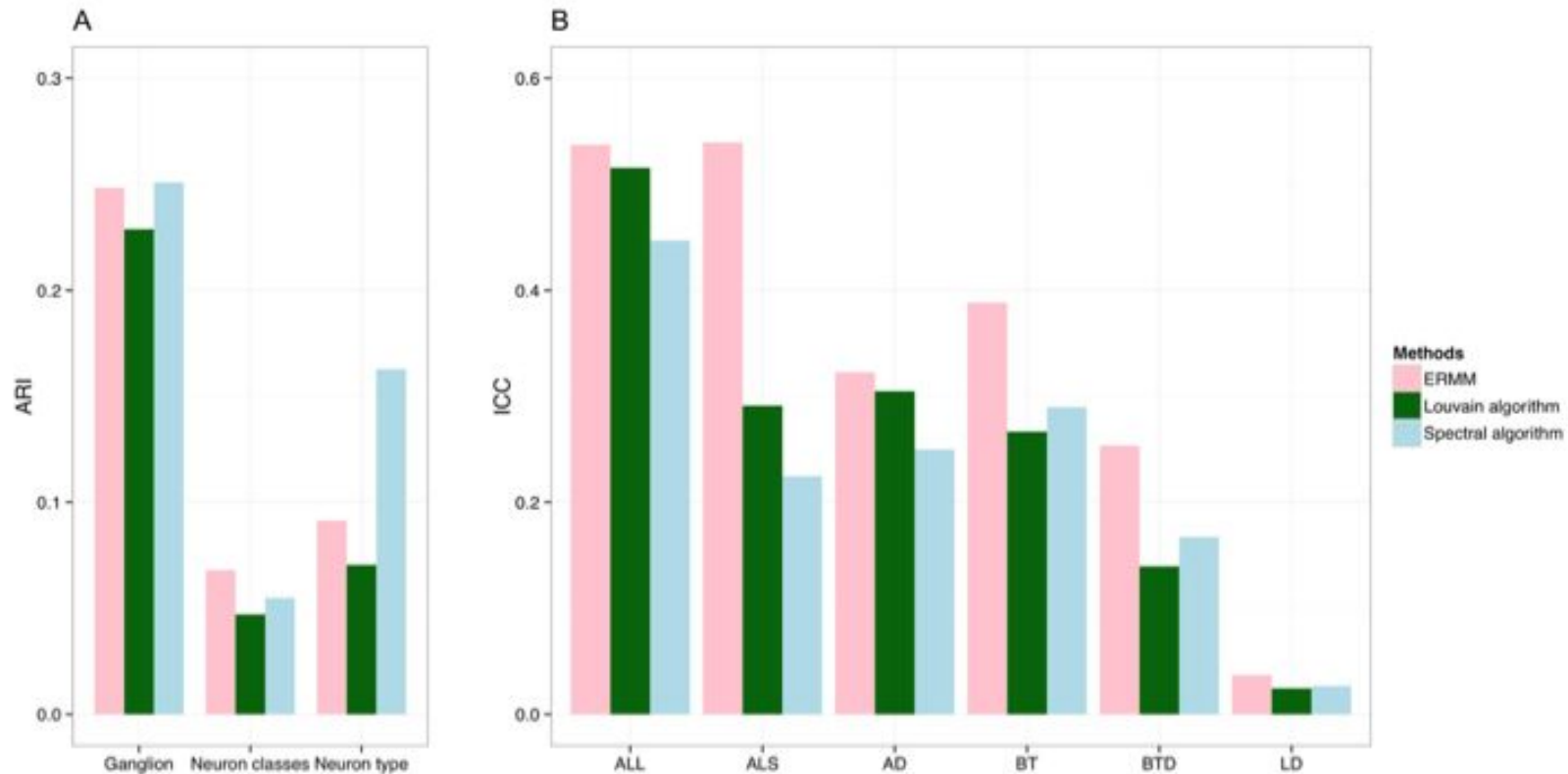
Comparison of performance

$$AIC(w_h) - AIC_{min}$$

Resolution



Resolution



Future work (aka YOUR work)

How do we feel about the similarity metrics used to compare clustering methods? Can you think of a better way to evaluate performance?

How about the clustering algorithms themselves? Is there another method you think might work?

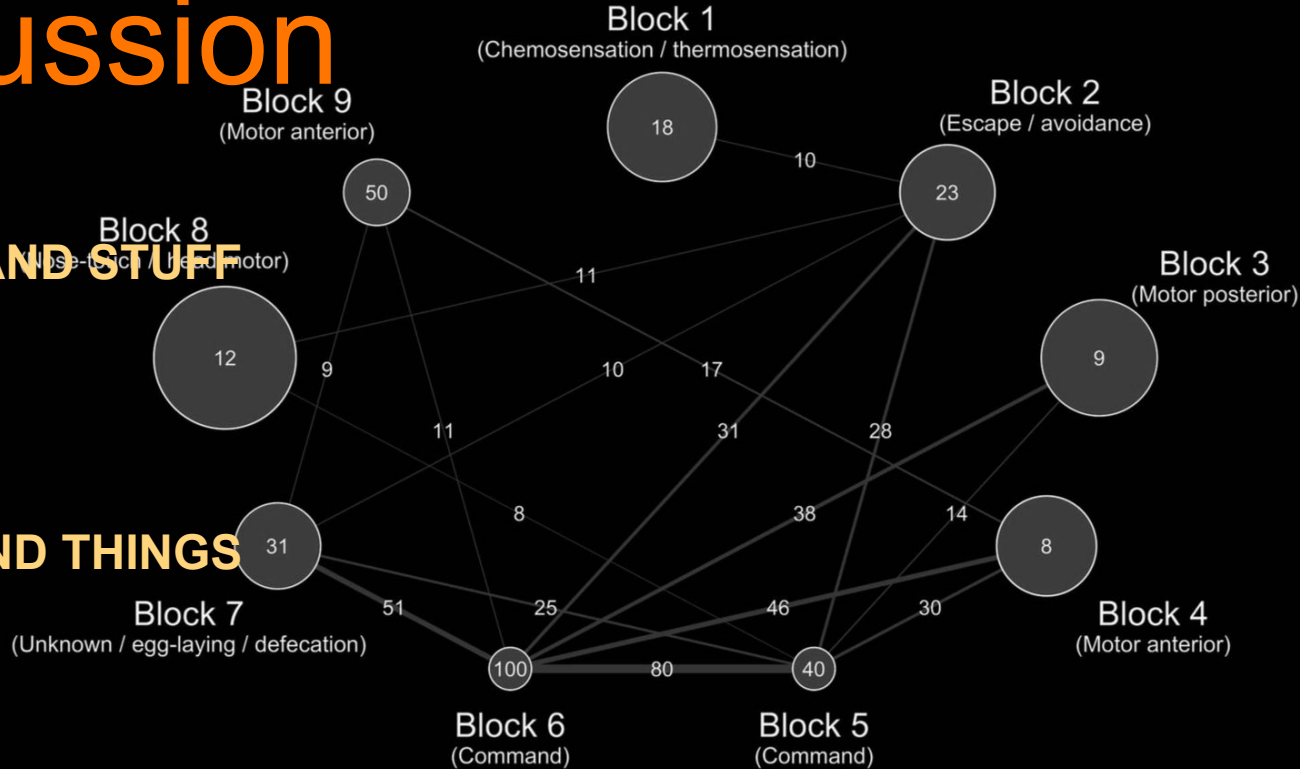
Discussion

Pros

THINGS AND STUFF

Cons

STUFF AND THINGS



Thank you!