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

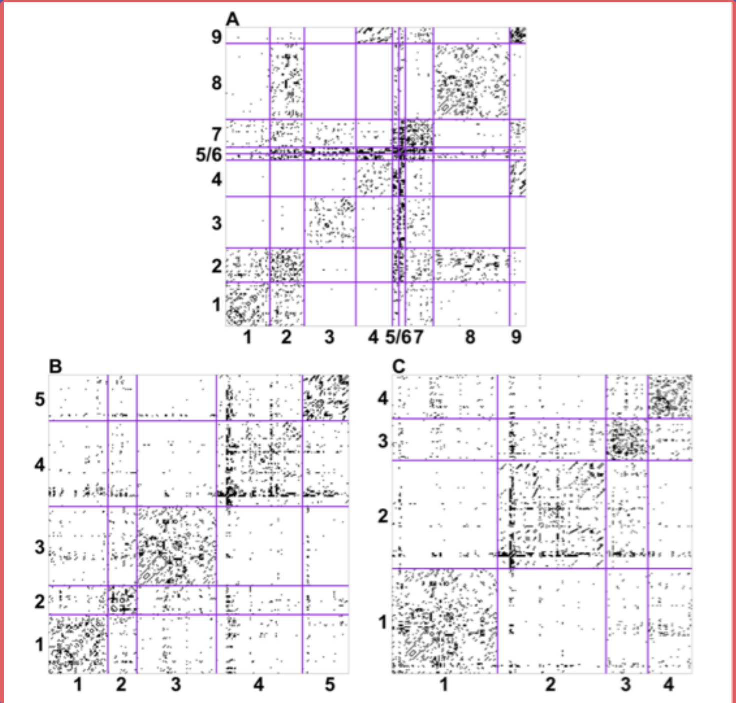
Dragana M. Pavlovic, Petra E. Vertes, Edward T. Bullmore, William R. Schafer, Thomas E. Nichols

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

Overview

Comparison of clustering techniques on *C. elegans* network

- ERMM
- Spectral Clustering
- 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 Clustering

Spectral Algorithm

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

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

ERMM

Adjacency matrix:

$$\boldsymbol{X} = \left((X_{ij})\right)_{1\le i\ne j\le n}$$

Group assignment

$$\boldsymbol{Z_i} = (Z_{i1}, \dots, Z_{iQ}),\ \textstyle\sum_q \boldsymbol{Z_{iq}} = \boldsymbol{1}$$

$\boldsymbol{Z_i} \sim M(1, \boldsymbol{\alpha})$ where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_Q), \sum_{q=1}^{Q} \boldsymbol{\alpha_q} = 1$

Connectivity matrix

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

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

Fast Louvain

Modularity

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

Action

ARI

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

OR

$$ARI = \frac{TP + FP}{\binom{n}{2}}$$

ICC

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

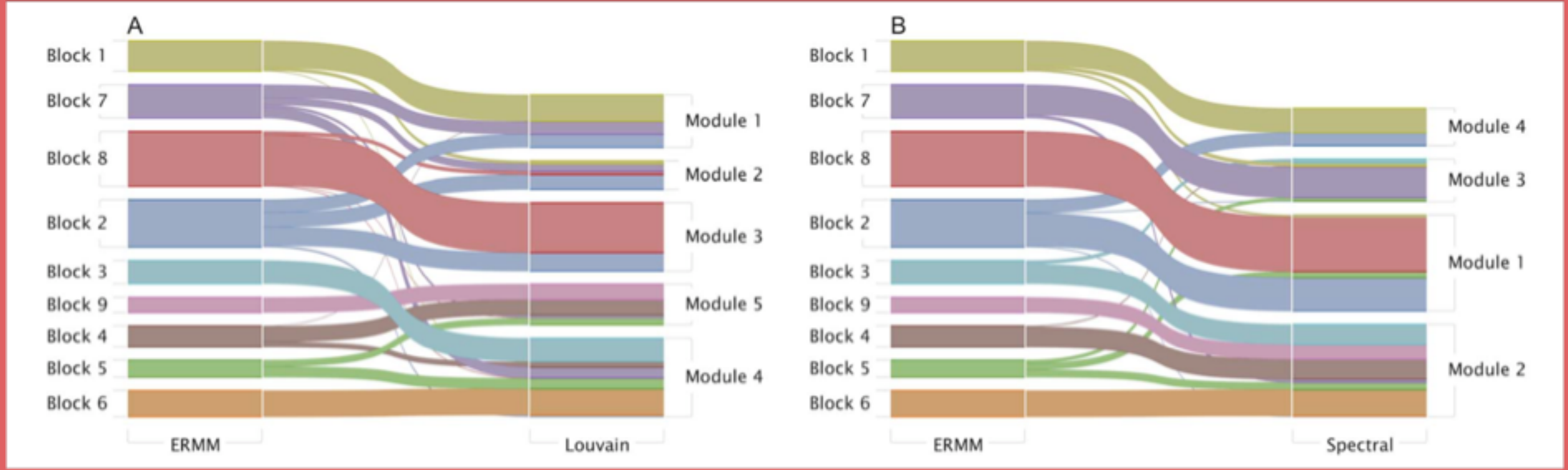
AIC

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

Comparison of performance

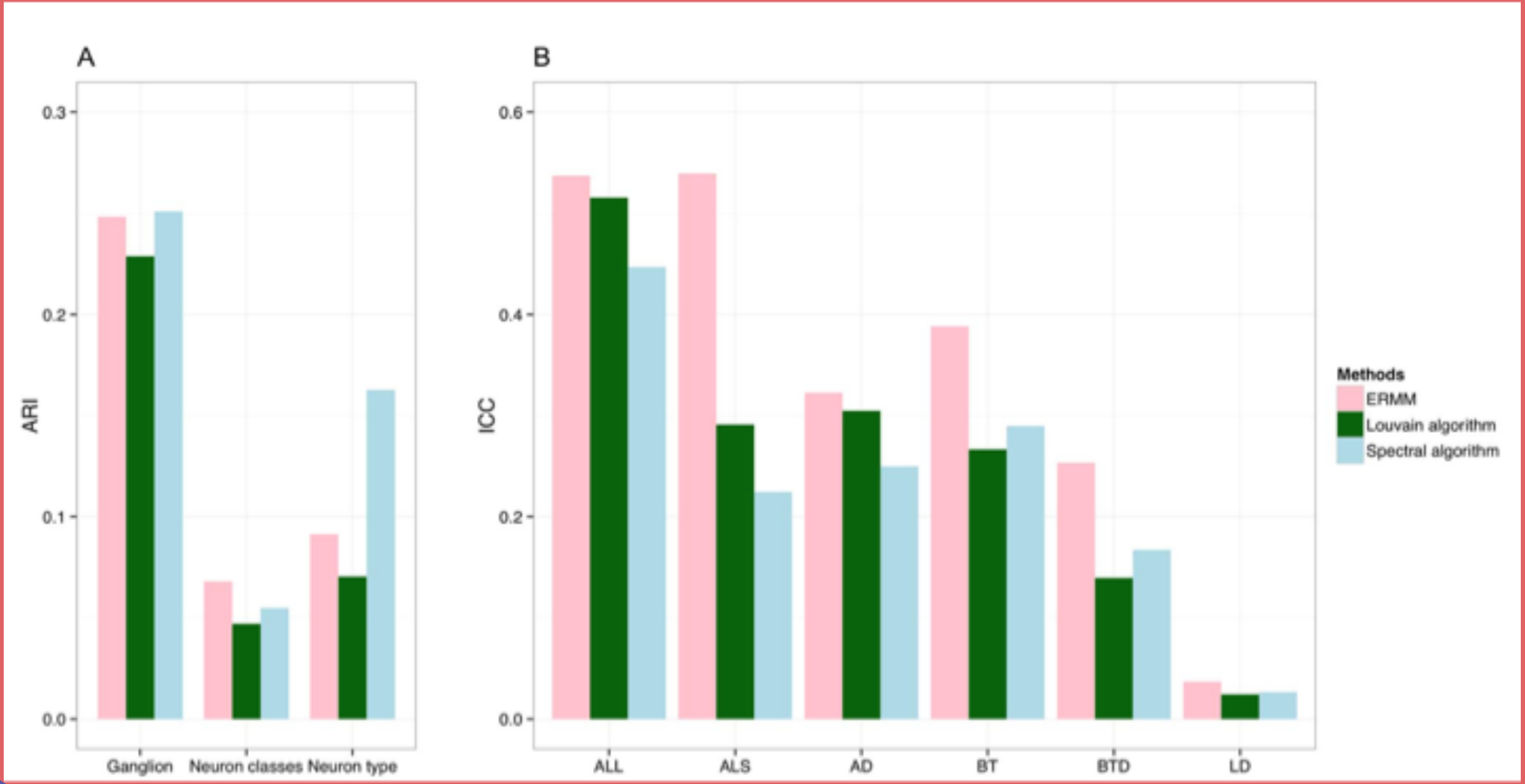
$$AIC(w_h) - AIC_{min}$$

Resolution



A
Block 1
Block 7
Block 8
Block 2
Block 3
Block 9
Block 4
Block 5
Block 6
Module 1
Module 2
Module 3
Module 5
Module 4
ERMM
Louvain
B
Block 1
Block 7
Block 8
Block 2
Block 3
Block 9
Block 4
Block 5
Block 6
Module 4
Module 3
Module 1
Module 2
ERMM
Spectral

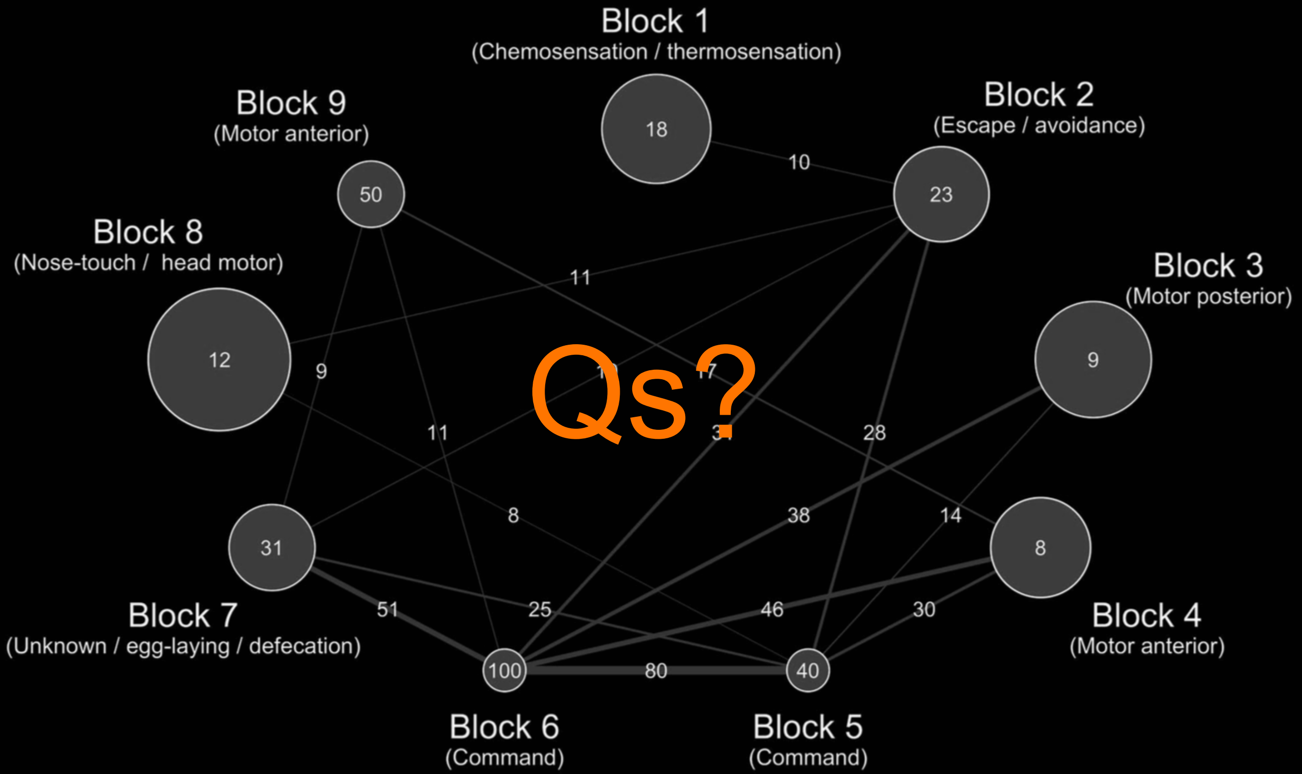
Resolution


A
ARI
0.3
0.2
0.1
0.0
Ganglion
Neuron classes
Neuron type
B
ICC
0.6
0.4
0.2
0.0
ALL
ALS
AD
BT
BTD
LD
Methods
ERMM
Louvain algorithm
Spectral algorithm

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?


Block 1
(Chemosensation / thermosensation)
18
Block 2
(Escape / avoidance)
23
Block 3
(Motor posterior)
9
Block 4
(Motor anterior)
8
Block 5
(Command)
40
Block 6
(Command)
100
Block 7
(Unknown / egg-laying / defecation)
31
Block 8
(Nose-touch / head motor)
12
Block 9
(Motor anterior)
50
10
11
9
11
28
8
38
14
51
25
46
30
80

Qs?

Thank you!