Data driven motif discovery across large connectomes

J Matelsky, M Robinette, P Rivlin, E Reilly, E Johnson

Johns Hopkins Applied Physics Laboratory

Identifying statistically significant motifs is crucial for understanding the fundamental building blocks of neural circuitry.

We present a scalable method that discovers both directed and undirected motifs without assuming prior structure.

To keep computation tractable, we implement a greedy progressive-refinement strategy. This means we can perform efficient motif discovery even in very large connectomes.

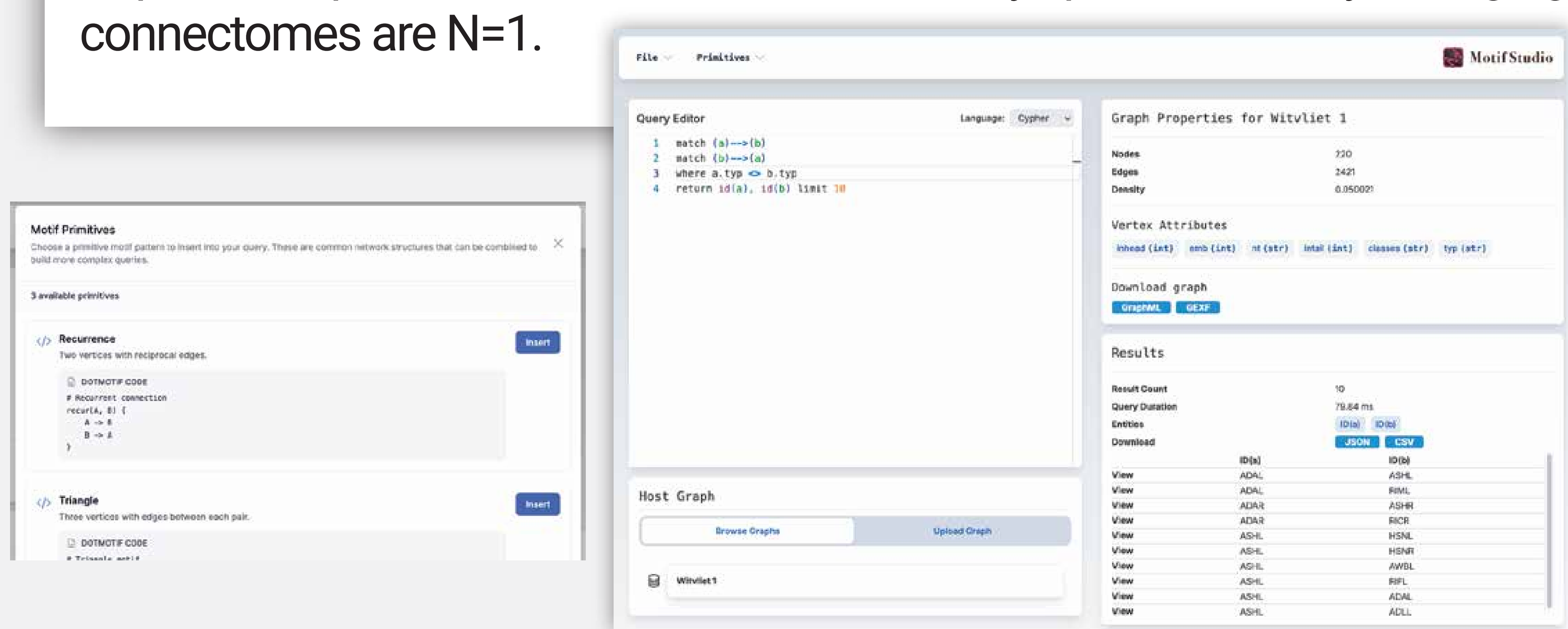
We show results for *C. elegans* and validate findings in the ellipsoid body of *Drosophila*.

Background

Neural circuits contain recurring **network motifs** — small, overrepresented connectivity patterns. Identifying these **statistically significant** motifs helps uncover organizing principles of neural computation.

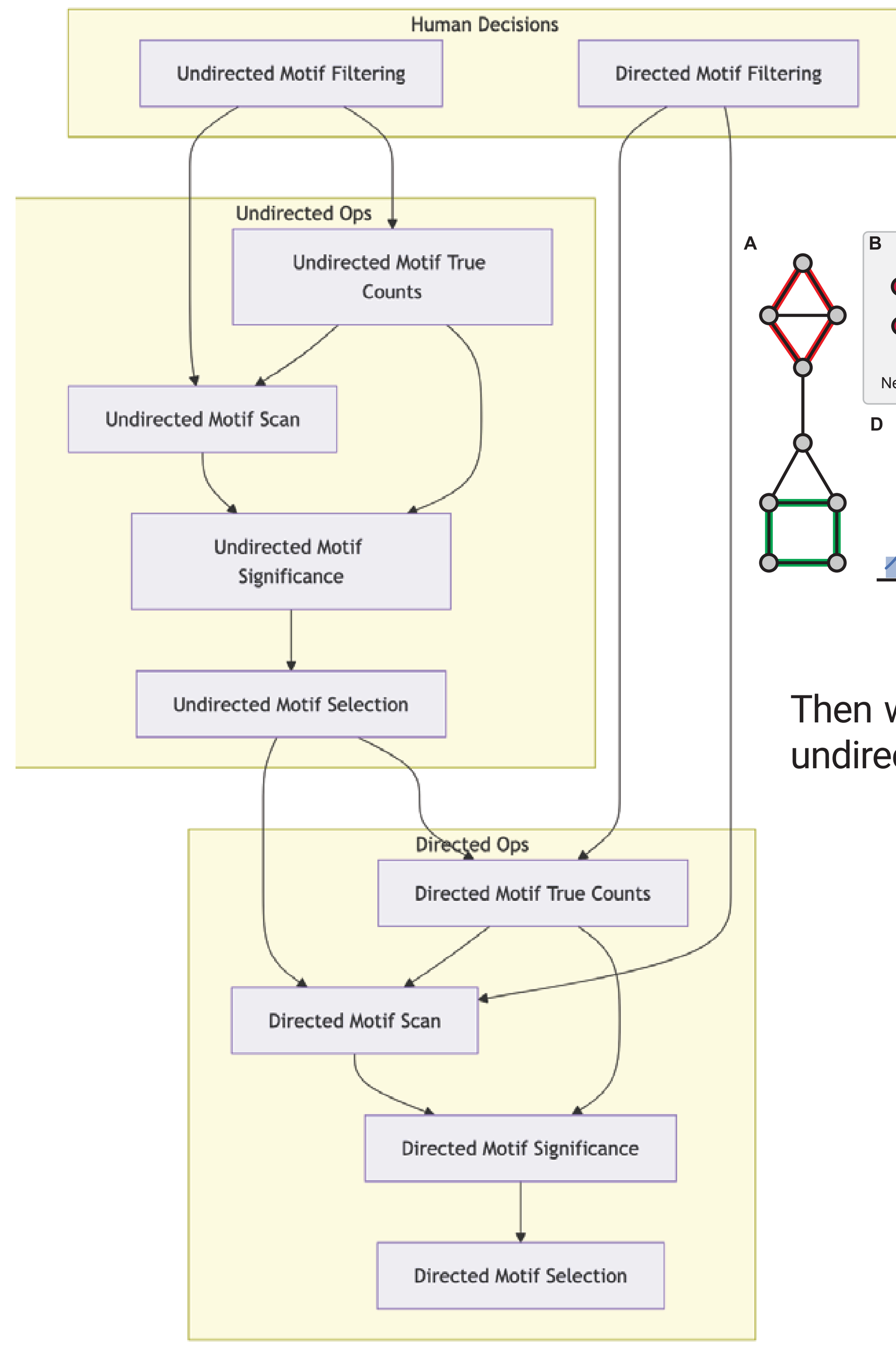
Our prior work has included tools like DotMotif (a query language for searching for known motifs) and pm2m (a toolkit for testing the statistical significance of identified subgraphs).

These and other traditional methods often assume a known structure or require multiple connectomes. This is not always possible; many emerging connectomes are N=1.

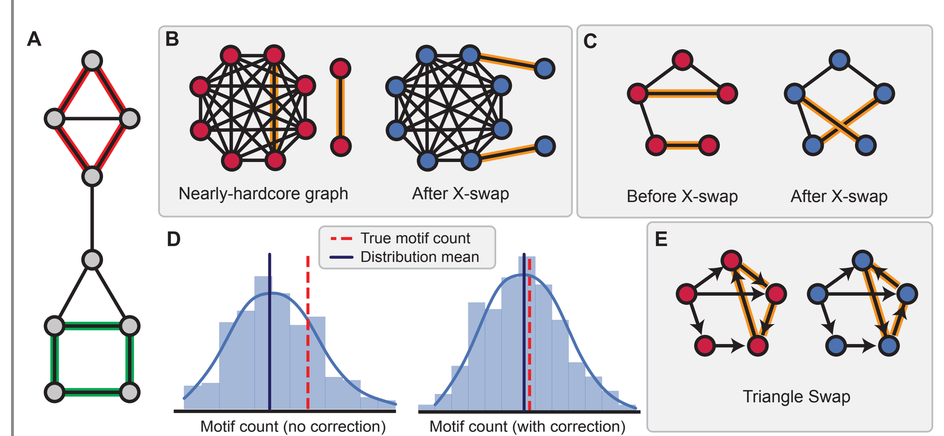


Methods: Motif Search

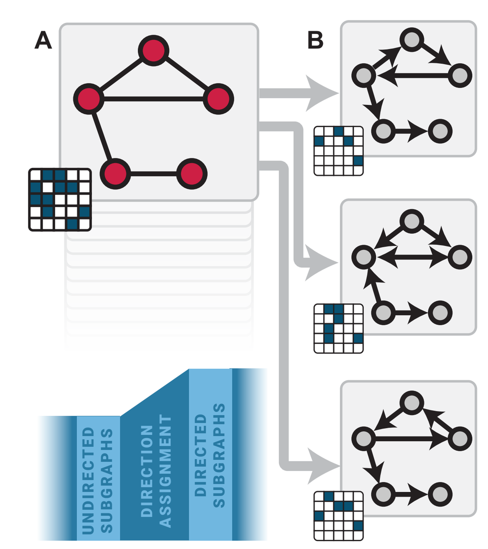
To test statistical significance we need a null model. The brain is not randomly distributed, but nor do we know its generating model.



We first find significant undirected motifs, using a random graph generator to serve as a null model:



Then we assign direction using the significant undirected motifs as an edge mask template:



Methods: Random Graph Models

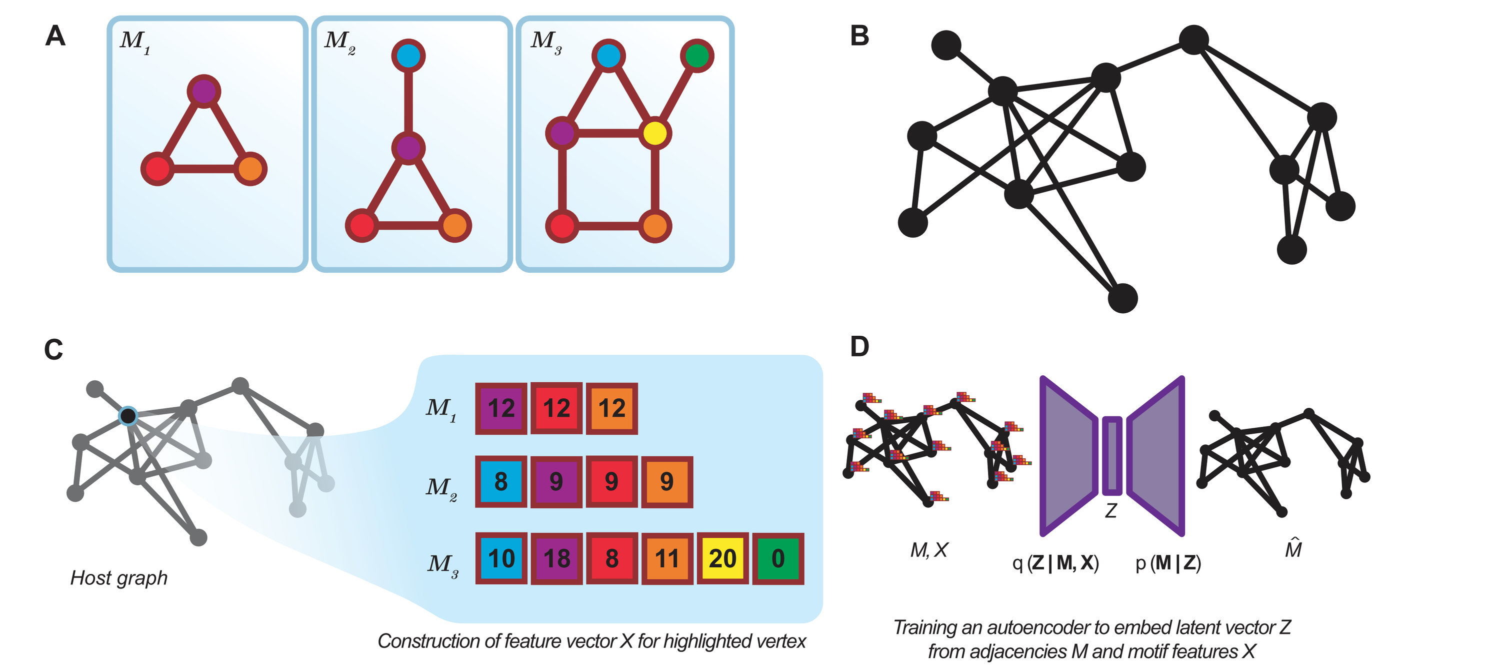
To test statistical significance we need a null model. The brain is not randomly distributed, but nor do we know its generating model.

X-swap randomization maintains degrees but is non-ergodic on directed graphs. A "triangle swap" fixes this problem (**C & E** below)

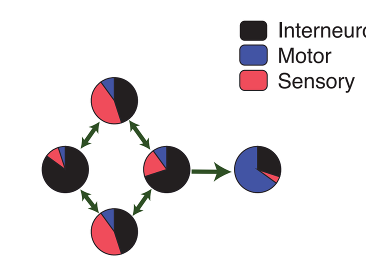
Graph mobility:

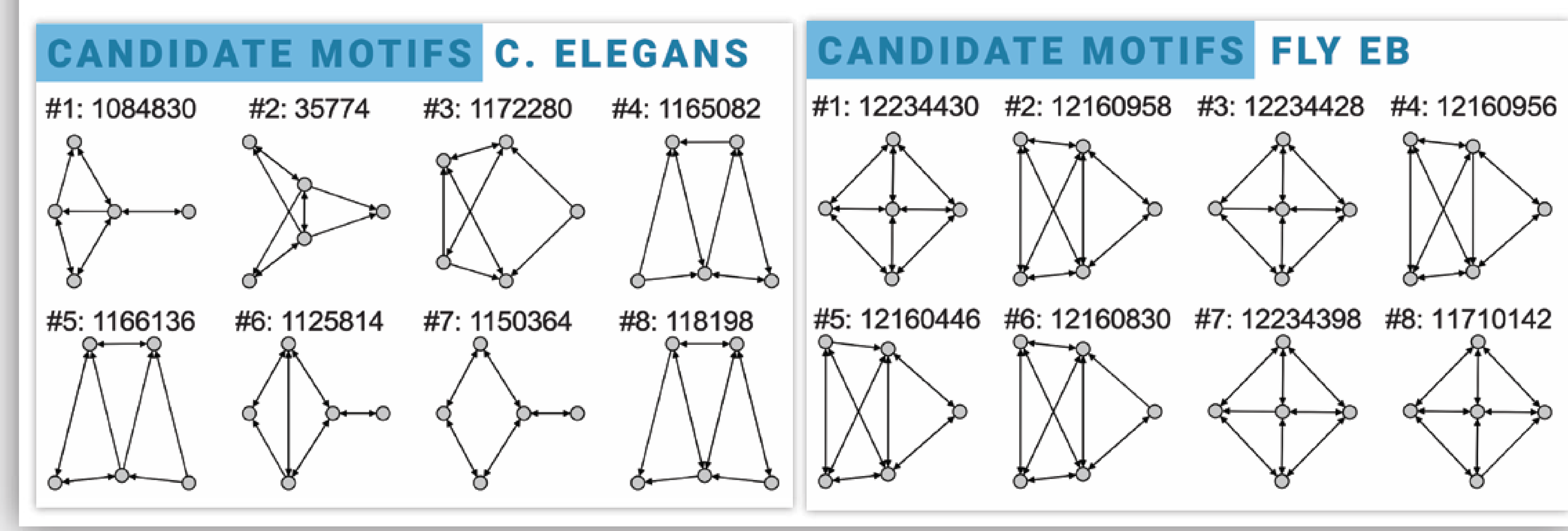
$$n(\mathbf{c}) = \tfrac{1}{4}\left(\textstyle\sum_i k_i\right)^2 + \tfrac{1}{4}\textstyle\sum_i k_i - \tfrac{1}{2}\textstyle\sum_i k_i^2 - \tfrac{1}{2}\textstyle\sum_{ij} k_i \mathbf{c}_{ij} k_j + \tfrac{1}{4}\mathrm{Tr}(\mathbf{c}^4) + \tfrac{1}{2}\mathrm{Tr}(\mathbf{c}^3)$$

Results: Fingerprinting

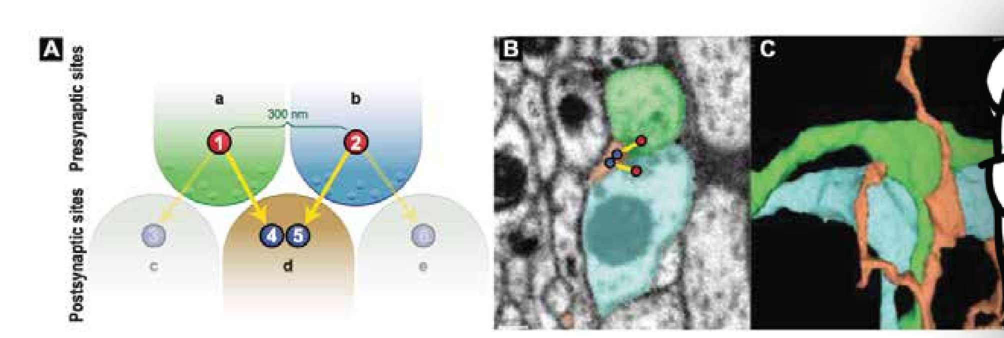


Results: Motif Identification





Results: Convergence



We can also use this technique to identif polyadic convergent synapses in the

Here we show several instances of