

ABSTRACT

Recent advances in neuroscience and bioimaging have enabled scientists to explore brain structure at the level of individual nanoscale synaptic connections. This measure of connectivity, when represented as **a graph with neurons as nodes and synapses as directed edges**, can be large and complex, presenting significant barriers to searching for structure and testing neural-circuit hypotheses. **We leverage graph database and analysis libraries to provide an easy-to-use grammar suitable for rapidly constructing queries and searching for subgraph isomorphisms**, or “motifs.” This abstracts many of the computer science and graph theory challenges associated with nanoscale connectomics and allows neuroscientists to quickly achieve reproducible findings.

Our approach, dubbed “DotMotif,” simplifies identification of basic as well as complex neural connectivity structures relevant to the systems-neuroscience community. We contextualize these results and demonstrate efficiently that the found motifs are unlikely to have occurred by chance; we also demonstrate the versatility of our approach through the exploration of errorful graphs. All of **our tools are released open source** to empower other scientists to use and extend these methods.

BACKGROUND

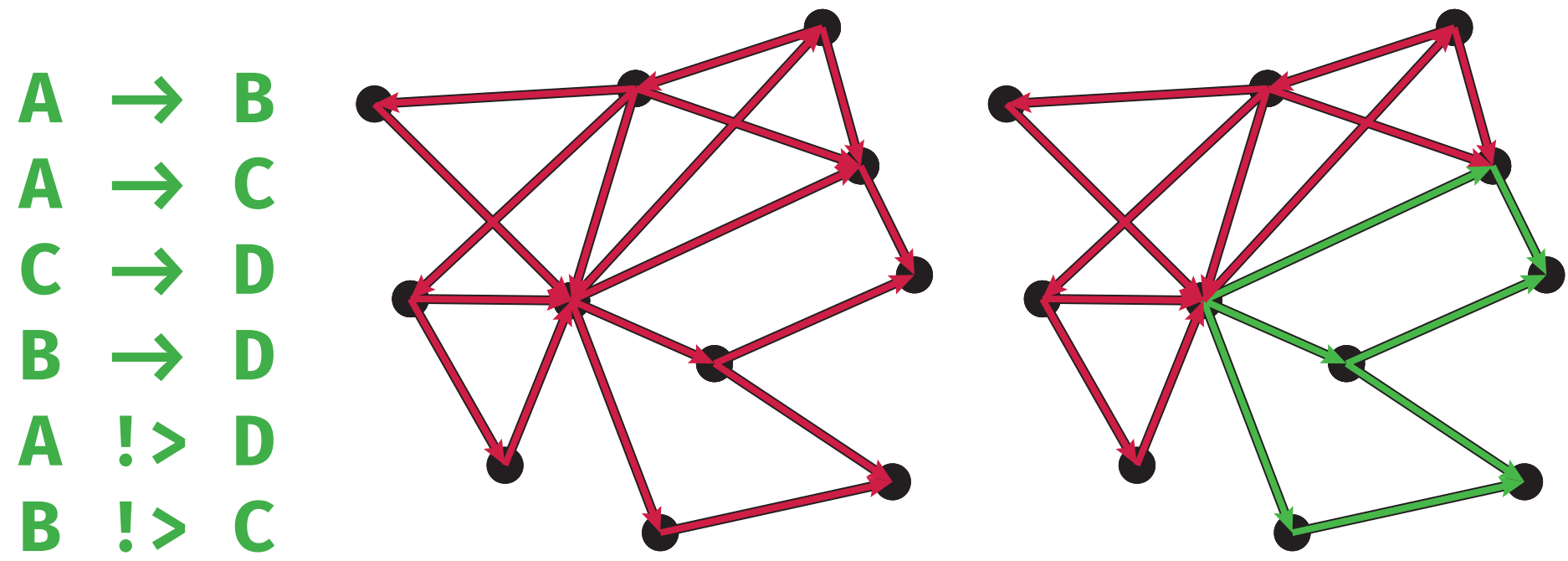
Nanoscale connectomes:

- Neurons are nodes
- Synapses are directed edges

This abstraction enables use of specialized graph analysis software.

Subgraph isomorphism identification: hypothetical small graph is searched for and identified in a larger complete graph.

One common and speedy implementation of this task is VF2 (*Cordella 2004*).



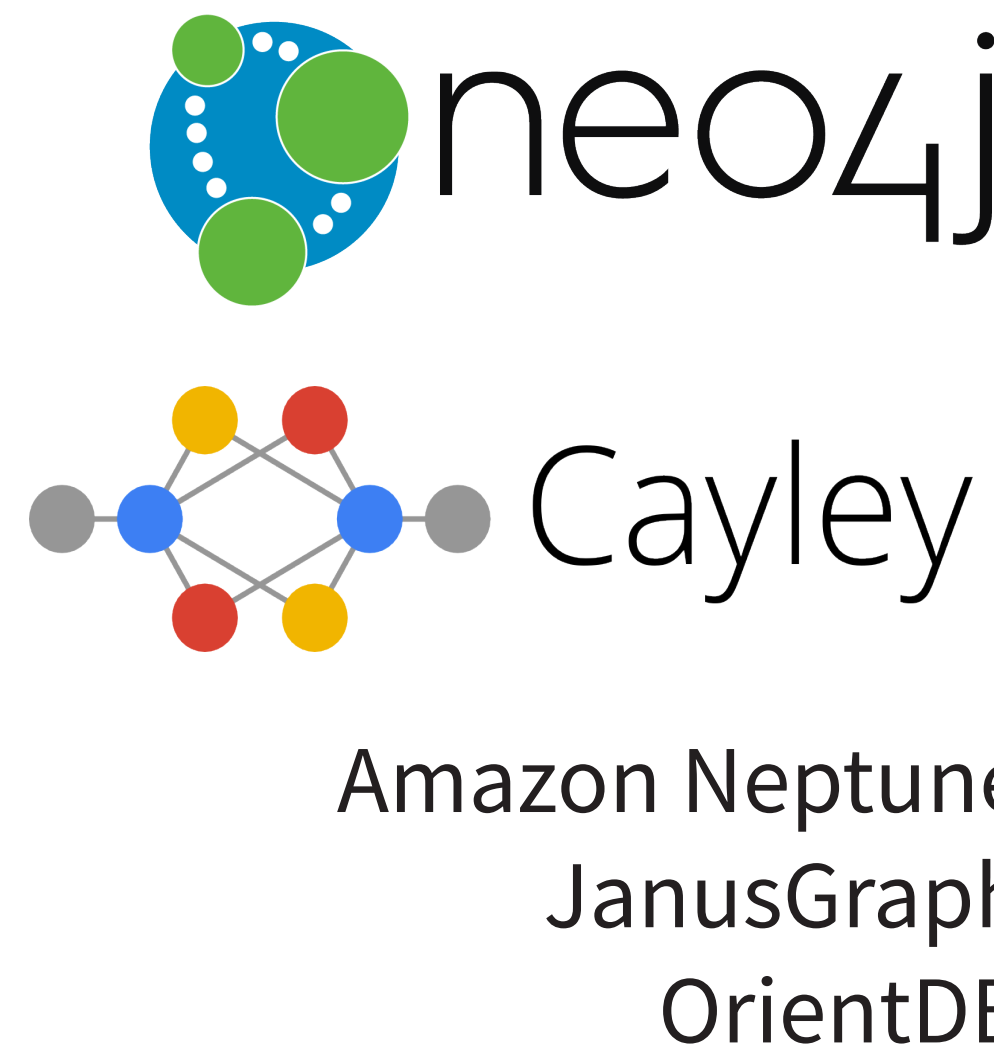
Subgraph isomorphism task:

- Common to industry domains such as e-commerce or search engines
- Graph datastore technologies have been developed to handle this task on very large (billion-node) graphs

Identifying brain motifs in large brain graphs can be rephrased as a subgraph isomorphism task: Graph databases can then be used to efficiently perform connectomics-flavored queries.

Weaknesses of these tools:

- Complex query syntax
- No error-checking prior to execution
- Nontrivial data ingest/export
- Requires tool familiarity/expertise



METHODS

DOTMOTIF DOMAIN-SPECIFIC LANGUAGE

We propose a graph query syntax, the “motif” format. This syntax can be used as an alternative to CSV or hand-curated query design to formulate questions which are then compiled into an intermediate format.

EDGES
A→B
B!>A
C→D
D↔E
E!!F

EDGE ATTRIBUTES
A→B [type=Glu,x>450]

NODE ATTRIBUTES
A.radius > 12

Our Python implementation includes syntax highlighting for major code editors.

MACROS

```
inhibit_feedback(A, B, C) {  
  A→B  
  B→C  
  C→A  
  C.type='inhibitory'  
}  
inhibit_feedback(x, y, z)
```

DOTMOTIF/CYPHER COMPARISON

```
# One-way edge:  
unigraph(A, B) {  
  A→B  
  B!>A  
}  
# One-way triangle:  
unitri(A, B, C) {  
  unigraph(A, B)  
  unigraph(B, C)  
  unigraph(C, A)  
}
```

```
unitri(x, y, z)  
unitri(z, d, e)  
unitri(e, f, g)
```

VALIDATION & EXECUTION

Common mistakes are easily caught at parse/compile time using the **dotmotif.validators** submodule. Validators are used to check both for query formation errors as well as biologically implausible motif components:

IMPOSSIBLE QUERY

A → B
A !> B

BIOLOGICALLY IMPLAUSIBLE

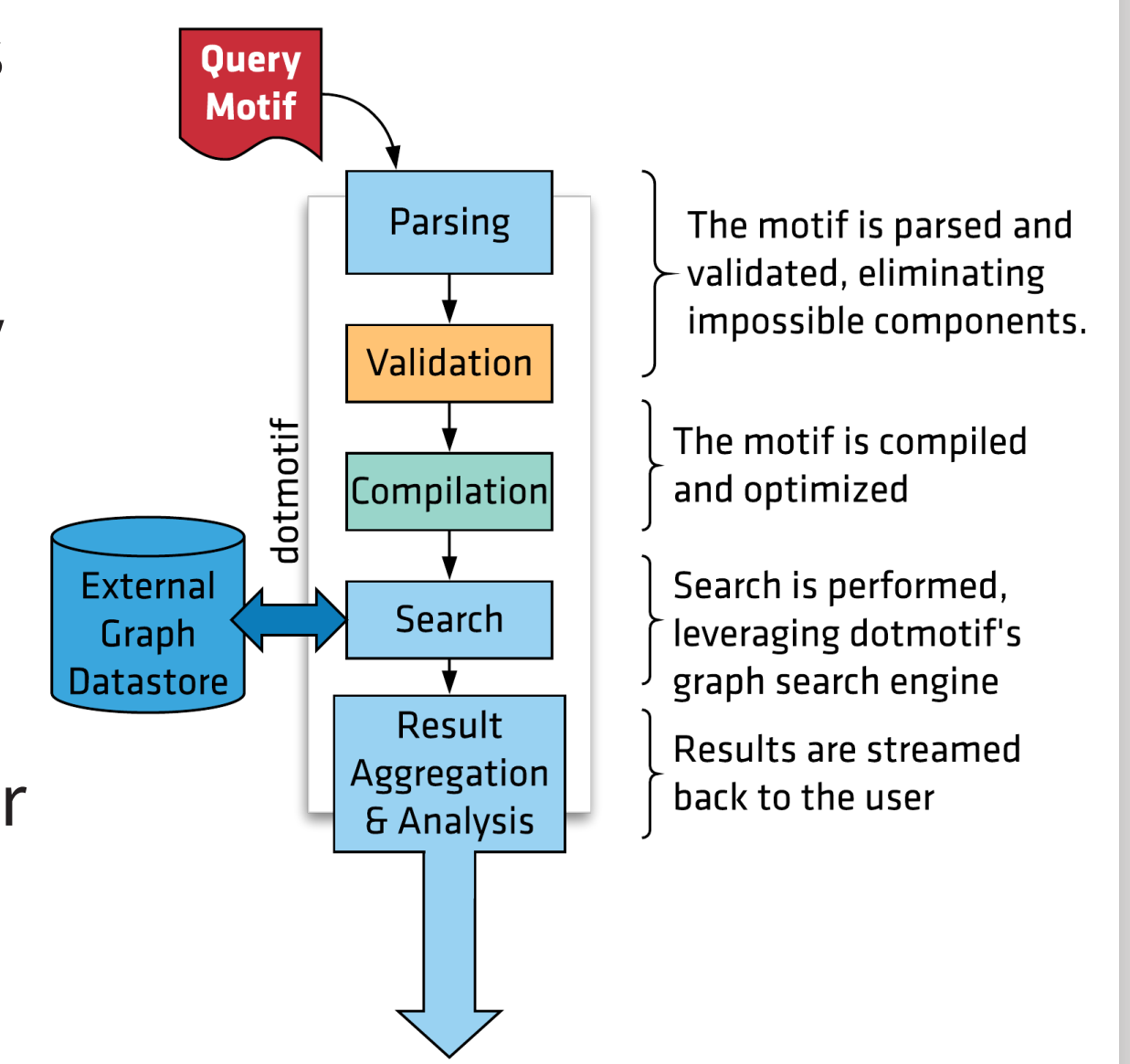
A → B [type=inhibitory]
A → B [type=excitatory]

After validation, the query is compiled into an intermediate format, which can then be executed on a large search graph using a variety of graph datastore technologies, such as networkx, Neo4j, Cayley, Neptune, etc.

INFRASTRUCTURE & INTERPRETATION

DotMotif provisions graph datastores and handles data ingest/export to common data standards:

- Reduces query syntax complexity
- Enables query compilation and optimization
- Standardizes in-memory graph queries for use with arbitrary DB
- Provisions graph database Docker containers where needed
- Performs data ingest
- Sanitizes and standardizes outputs for ease of use



RESULTS

TAKEMURA ET AL (eLIFE 2017) CASE STUDY

This paper identified the circuitry responsible for “ON” motion response in the fruit fly visual system. The circuit spans multiple brain regions and involves multiple cell types whose organizations and inputs onto motion-sensing cells in the lubula.

MOTIF CONSTRUCTION

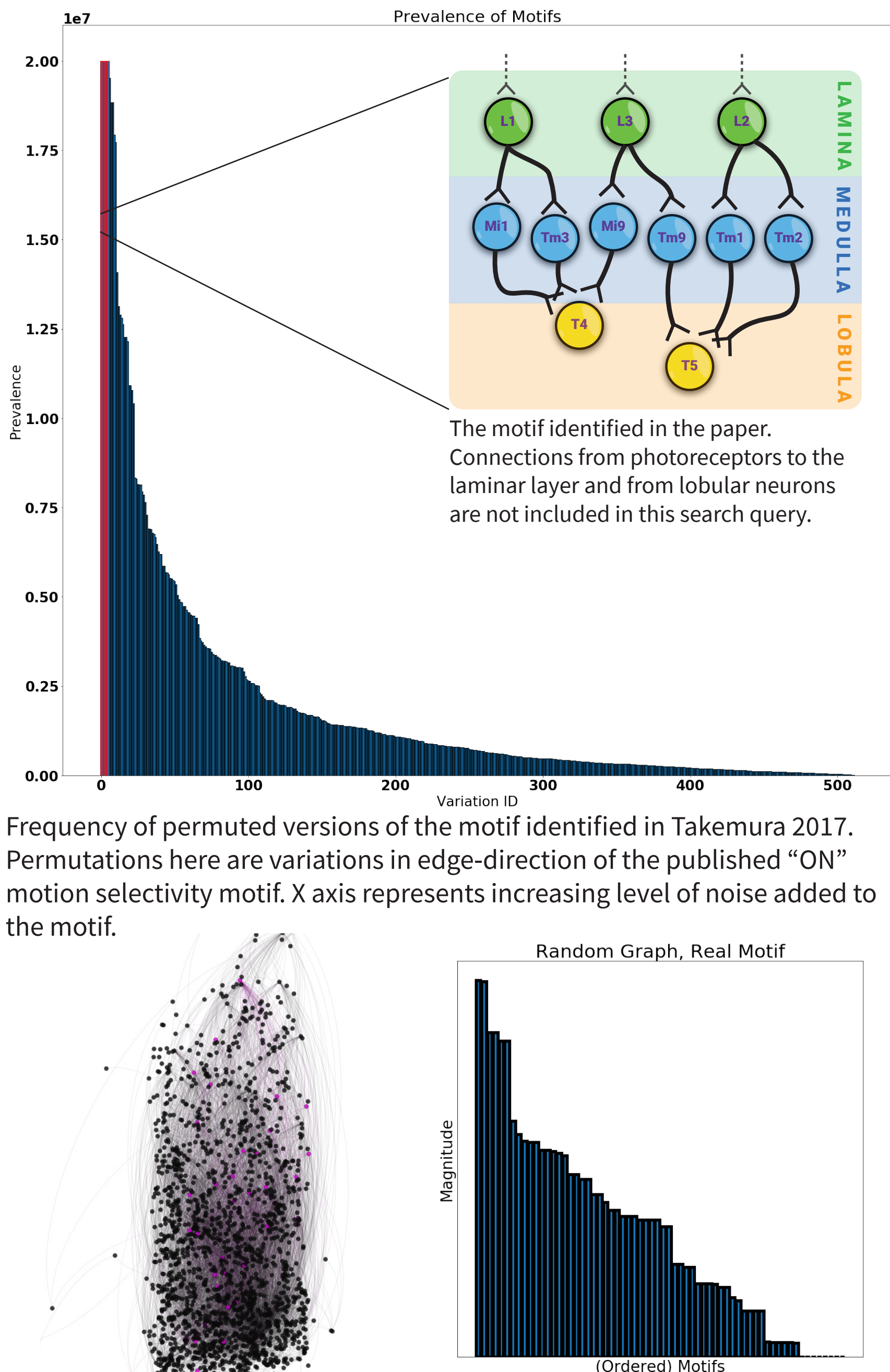
Original motif design from reported circuit:

L1 → Mi1
L1 → Tm3
L3 → Mi9
Mi1 → T4
Tm3 → T4
Mi9 → T4
L3 → Tm9
Tm9 → T5
T4 → Tm1
L2 → Tm2
Tm1 → T5
Tm2 → T5

CONDITIONS
Can add conditions to improve query specificity when such data are available
T5.name contains “T5”
T4.z > 900
...

EXCLUDED CONNECTIONS
Can specify connections to *exclude*:
L1 !> T4
L3 !> T4
...

After construction, can render, execute, or store in .motif file (along with constraints and executor parameters) to share with others or package with publication.



Left: Locations of the motif in the full Takemura 2017 graph (purple). Right: Frequency of permuted versions of the motif identified in Erdős-Rényi random graph of the same size and edge density. This peak is qualitatively less pronounced and the distribution has a shorter tail.

PUBLICATION & REPRODUCIBILITY

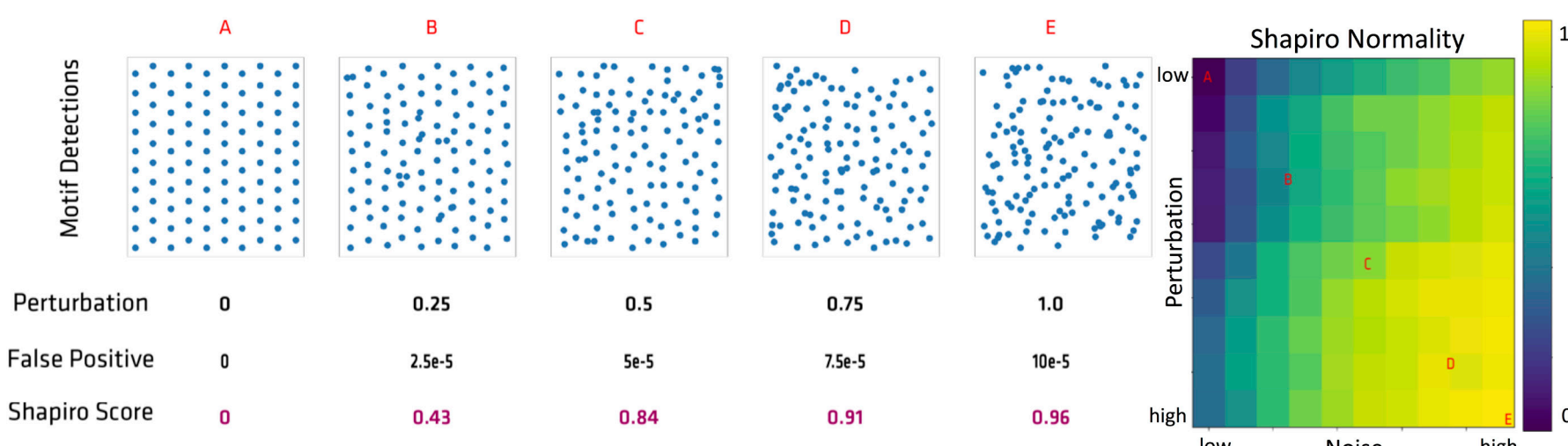
.motif file format includes:

- Motif graph structure, as well as constraints, metadata, and comments
- Parameters used to construct the search query
- Version and release information for the codebase used to generate the file
- Provenance metadata (where available) including date of creation, user/author name
- Checksum for validation

Immutable file can then be shared or published.

SPATIAL MOTIF DISTRIBUTION

DotMotif includes a suite of periodicity measures for analyzing outputs of motif searches. Below, synthetic simulation data are perturbed and tested for periodicity using a simple Shapiro normality test. These measures may be useful for some scientific questions that expect repeated substructure.



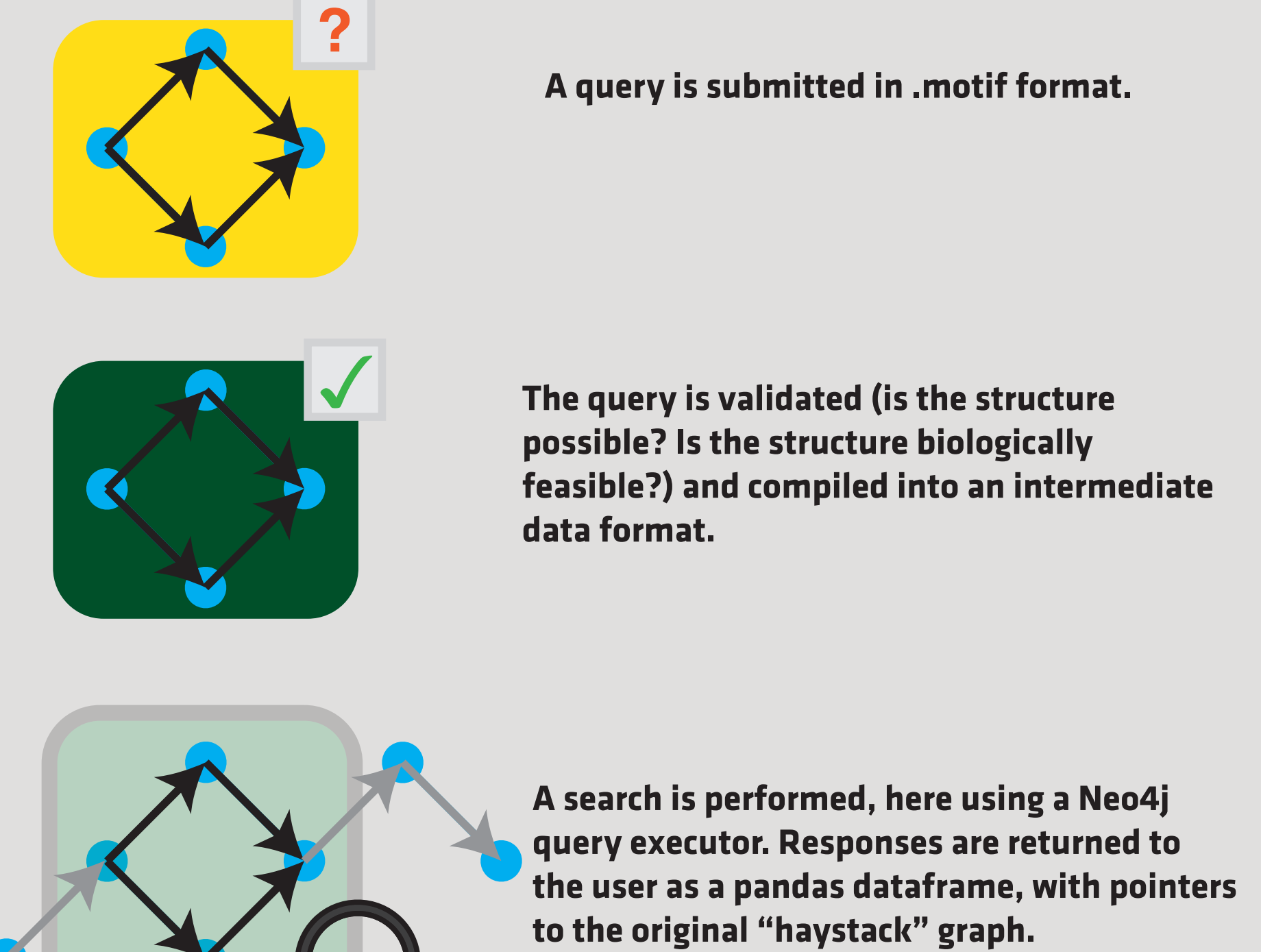
STATISTICAL SIGNIFICANCE TEST

Once motifs are identified, researchers may wish to investigate local structure directly or to determine if motifs occur more likely than chance. Some possible directions for testing significance include:

- Permutation tests
- Comparison with theoretical results
- **Your idea here?**

TOOL COMPATIBILITY

DotMotif imports and exports to major graph framework formats in Python as well as on disk (graphml, CSV, etc), and simplifies graph ingest to graph databases like Cayley or Neo4j. **DotMotif is compatible with the Janelia NeuPrint data format, and the databases can be used interchangeably.**



WORKS CITED & ATTRIBUTIONS

Borst, A., & Helmstaedter, M. (2015). Common circuit design in fly and mammalian motion vision. *Nature Neuroscience*, 18(8), 1067–1076. <https://doi.org/10.1038/nrn.4050>

Brugnera, G., & Beck, H. (2018). The Circuit Motif as a Conceptual Tool for Multilevel Neuroscience. *Trends in Neurosciences*, 41(3), 128–136. <https://doi.org/10.1016/j.tins.2018.03.002>

Burns, R., Reilly, E., Bales, A., Roncal, W. G., Falls, B., Chandrasekhar, V., ... , Llaney, K. (2019). A Community-Developed Open-Source Computational Ecosystem for Big Neuro Data. *ArXiv Preprint ArXiv:1904.02835*.

Connectomics. (n.d.). Retrieved January 12, 2015, from <https://neurodata.io/project/connectomics/>

Cordella, Luigi. "A Subgraph Isomorphism Algorithm for Reaching Large Graphs." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 10, pp. 1387–1392, Oct. 2004.

Docker. (n.d.). Retrieved January 13, 2019, from <https://www.docker.com/index.html>

Helmstaedter, M., Briggman, K. L., Turaga, S. C., Jahn, V., Seung, H. S., & Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7501), 169–174. <https://doi.org/10.1038/nature12346>

Kistnas, D., Hader, R., Pryor, D., Gion, T., Manavalan, P., Matelsky, J., ... , D'Angelo, D. (2017). The block object storage service (bosdb): A cloud-native approach for petascale neuroscience discovery. *BioRxiv*, 217740.

Reilly, E. T., Carrozzini, J. S., Roncal, W. G., Helman, D. M., Weste, B. A., Chevillet, M. A., & Ross, M. J. (2017). Neural Reconstruction Integrity: A metric for assessing the connectivity of reconstructed neural networks. *ArXiv:1702.02684 [q-bio]*. Retrieved from <http://arxiv.org/abs/1702.02684>

Takemura, S., Blanton, A., Liu, Z., Heme, A., Viskochil, S., Rhee, P. K., ... , Chikara, D. B. (2013). A neural motion detection circuit suggested by Drosophila connectomics. *Nature*, 500(7461), 179–181. <https://doi.org/10.1038/nature12450>

Takemura, S., Heme, A., Chikara, D. B., Schiffer, L. K., Rubin, G. M., & Helman, D. B. (2017). The comprehensive connectome of a neural substrate for 'ON' motion detection in Drosophila. *eLife*, 6, e24394. <https://doi.org/10.7554/eLife.24394>

Wolp, B., Chikara, D., & Bloomfield, S. A. (2009). Tracer Coupling Patterns of the Ganglion Cell Subtypes in the Mouse Retina. *The Journal of Comparative Neurology*, 513(5), 464–487. <https://doi.org/10.1002/jcomp.21912>