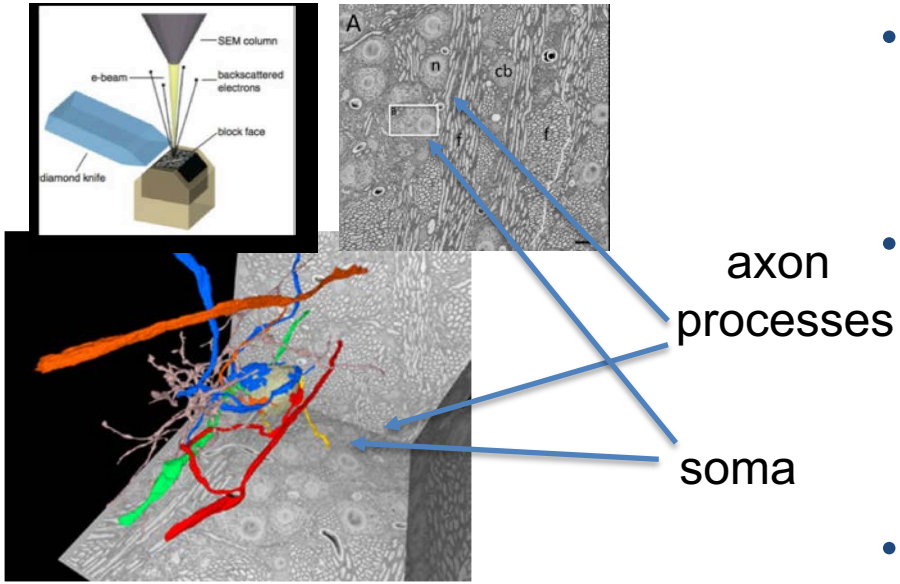
DETECTING HIERARCHICAL COMMUNITY STRUCTURES IN THE MOUSE RETINAL CONNECTOME

Joint work with Scott Payne, CQ Zhang (MATH)
George Spirou (Neuroscience)
Michael Morehead (CS)

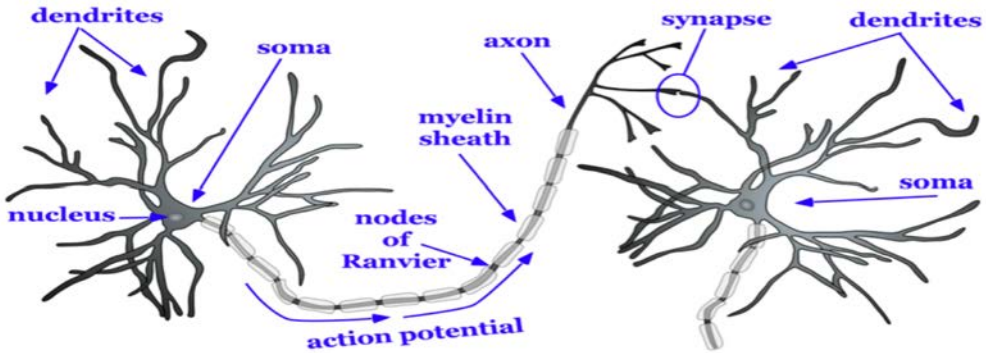
Eddie Fuller
Department of Mathematics
West Virginia University

FROM SBEM IMAGERY TO GRAPHS

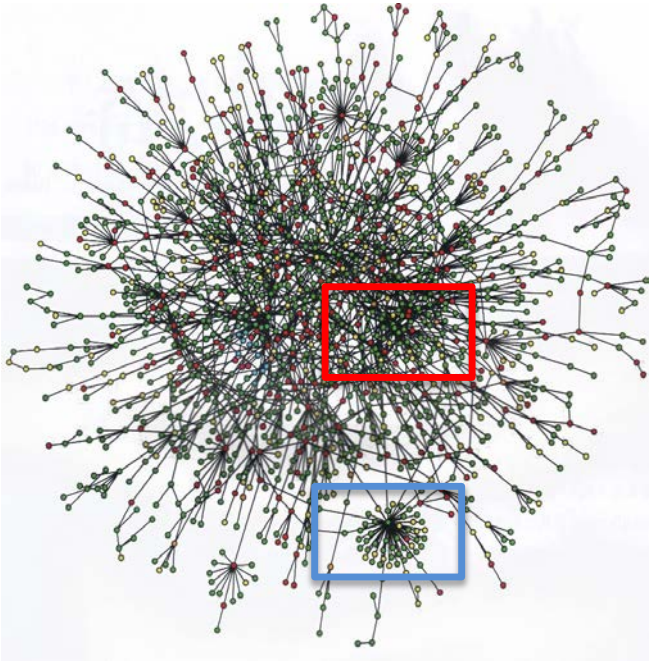


- Serial Blockface Electron Microscopy generates ~terabytes of data (and →petabytes soon)
- Volumes must be coded (neuron, glial cell, blood vessels) and neurons isolated
  - Cell morphology/structure (amacrine, bipolar, ganglion)
- Synaptic connections identified and used to build a network of neurons

CONNECTOMICS - NEURON STRUCTURE AND CONNECTIONS



NETWORK ANALYTICS



Graphs can model data well but

- Are they faithful representations?
- What do we consider ‘important’ structures?
  - Are there important vertices or groups of vertices?
- In the context of connectomics, what are the relationships to biological structures?
  - Can these analytic processes provide useful insights?

NETWORK PROPERTIES

Centrality

- Degree
  - Number of edges incident at a node
- Closeness
  - Closeness is based on the inverse of the path length of each node to every other node in the network.
- Betweenness
  - Counts the number of shortest paths between *i* and *k* that node *j* resides on.

Matrix based methods

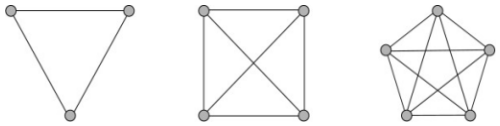
- Analysis of adjacency, weight, Laplacian matrices, spectral methods

STRUCTURE WITHIN NETWORKS

Structure *may* also be found in ‘communities’ or ‘clusters’ of nodes

One definition of a community/cluster: Clique

- Vertices are all adjacent to each other
  - Strict definition,
  - finding all or maximal clique NP-complete problem
- Density = 1
- *Other names: n-clique, n-clan, n-club, k-plex*



Density of a subgraph C with vertex set V(C) and edge set E(C)

$$d(C) = \frac{|E(C)|}{\frac{|V(C)|(|V(C)|-1)}{2}}$$

- Number of complete “triangles

Idea: Search for ‘almost’ cliques as well as relations between them

- *Use density or weighted density*

$$d_w(C) = \frac{\sum_{e \in E(C)} w(e)}{\frac{|V(C)|(|V(C)|-1)}{2}}$$

CLIQUE/COMMUNITY DETECTION

For our work we employ a modified version of the ***Quasi-Clique Merger*** method (QCM)

Ou, Y., & Zhang, C. (2007). A new multimembership clustering method. *Journal of Industrial and Management Optimization*, *3*(4), 619.

- This algorithm constructs 'quasi-cliques' by adding edges to existing collections of nodes and edges (a subgraph) and comparing the density of the new collection to that of the old
- QCM
  - Examines graph structure at multiple resolutions
  - Builds a tree with community cuts at multiple heights
    - Agglomerative
  - Not proper tree as nodes may lie in multiple branches

Modified method in the current work builds on this using similarity measures in combination with graph theory to detect dense intra-node connectivity as well as inter-community connectivity (e.g. bi-partite structures)

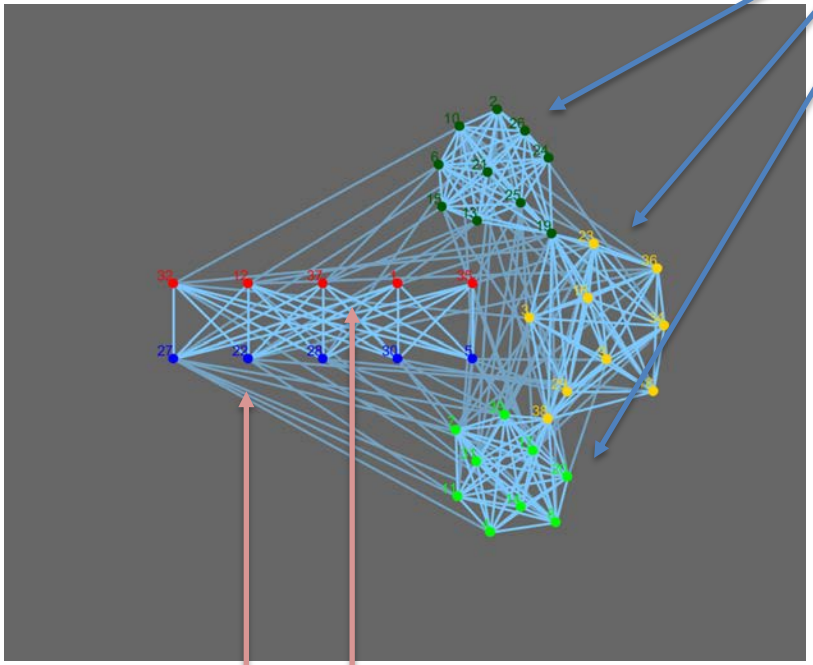
Qi, X., Tang, W., Wu, Y., Guo, G., Fuller, E., & Zhang, C. Q. (2014). Optimal local community detection in social networks based on density drop of subgraphs. *Pattern Recognition Letters*, *36*, 46-53.

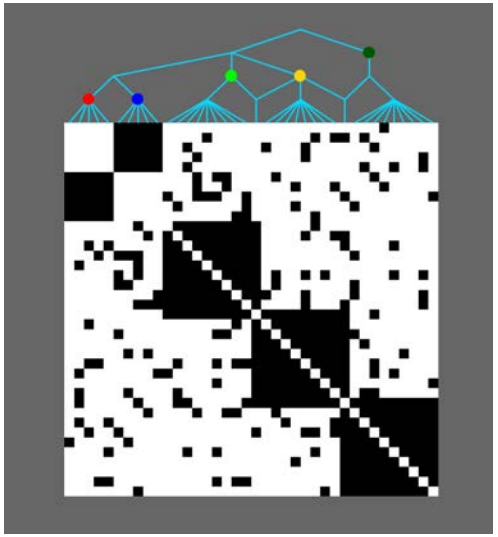
→

Payne, S., Fuller, E., Zhang, CQ. *QCMp*: An Unsupervised Network Community Detection Method. *In preparation.*

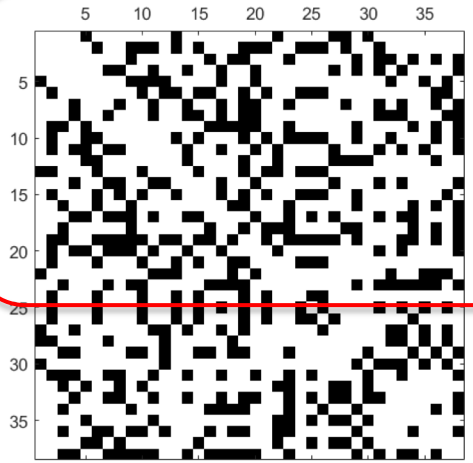
QCM

Similar connectivity within community



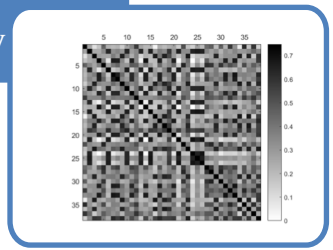


Similar connectivity to other regions



Adjacency/weight information

Map to similarity



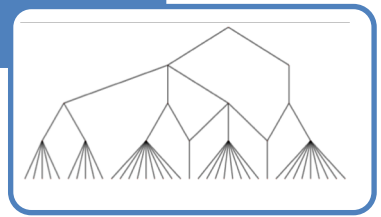
Ex: Network Properties

$$y_t = -(I - W^{+\mathrm{T}})y$$

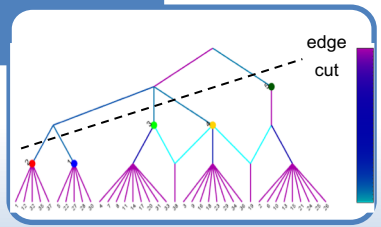
$$\Phi(t) = e^{-(I-W^{+\mathrm{T}})t} = e^{-t}\sum_{k=0}^{\infty}\frac{t^k}{k!}(W^{+\mathrm{T}})^k$$

Cosine similarity of rows

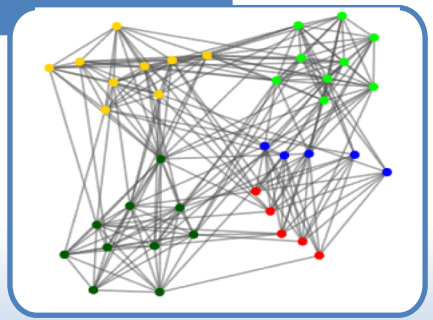
QCM-FPZ algorithm



Detect Communities

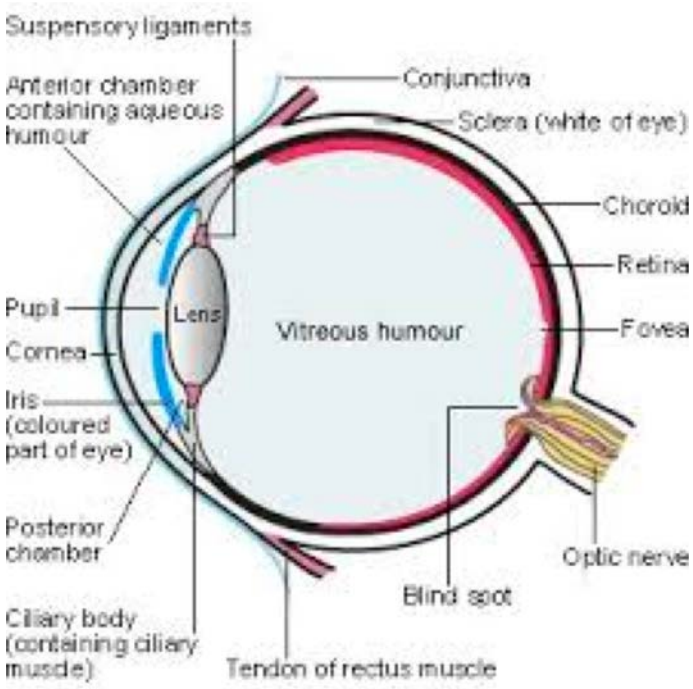


Gravitate to communities

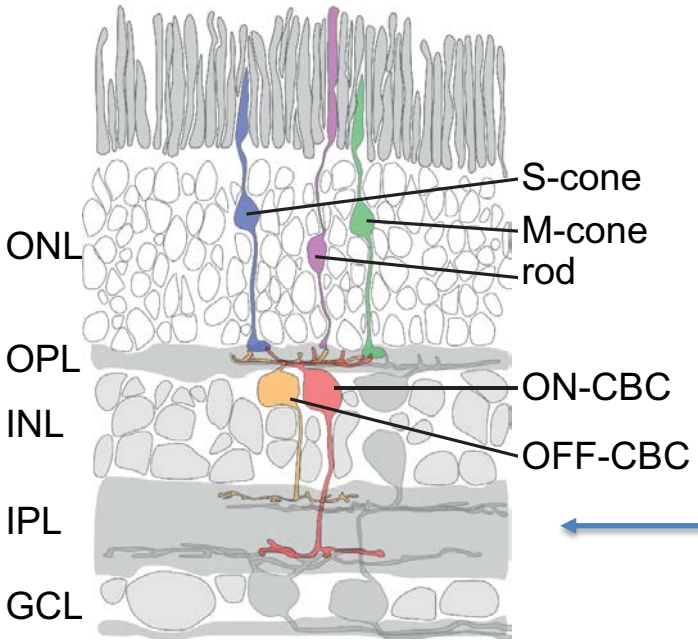


MAMMAL SENSORY NETWORKS

- The optical network
- Retinal neurons on the back of the eye collect signals from rod and cone neurons and feed to ganglion cell ‘collectors’



THE MOUSE RETINA AND IPL



*adapted from Behrens et al 2016

- Retina composed of multiple neuron types
  - Rods/cones react to light
  - Signal along bipolar cells to ganglion cells
  - Ganglion cells 'integrate' signals from multple columnar units
  - Amacrine cells in between connect units and mostly inhibit signaling
- Inner Plexiform Layer
  - Lies between rod/cone layer (back of eye) and ganglion cell layer (center of eye)
  - 'mesh' of dendrites and axon processes from
    - Bipolar cells connected to rods/cones on one end and ganglion cells on the other
    - Amacrine cell networks interacting with these circuits

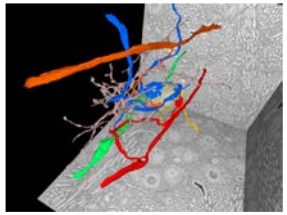
NETWORKS FROM THE IPL

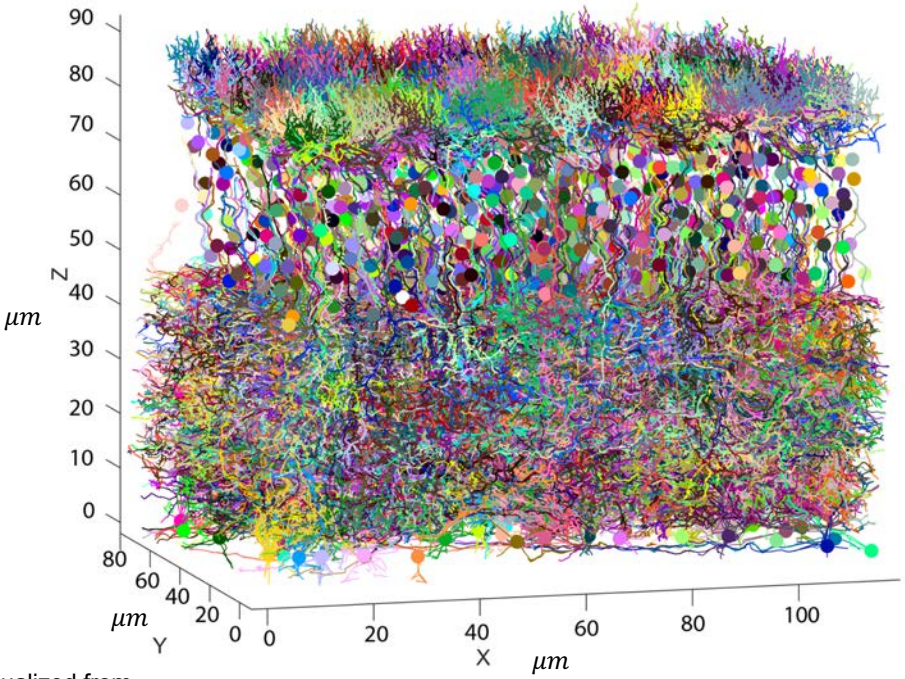
SBEM dataset e2006 published by Helmstaedter et al. (2013)

- Volume from inner plexiform layer of mouse retina
- 1123 neurons isolated, cell type identified by structural examination
  - Ganglion, Amacrine, Cone Bipolar, Rod Bipolar Cells
  - Mueller Glia
- 1076 connected cells
- Total volume $114\mu m$ x $80\mu m$ x $132\mu m$

For our work we use data for all connected cells

- 1076 nodes
- 89,991 weighted edges (sum of areas between unique node pairs)
- Tools developed in Matlab for visualization, data analysis





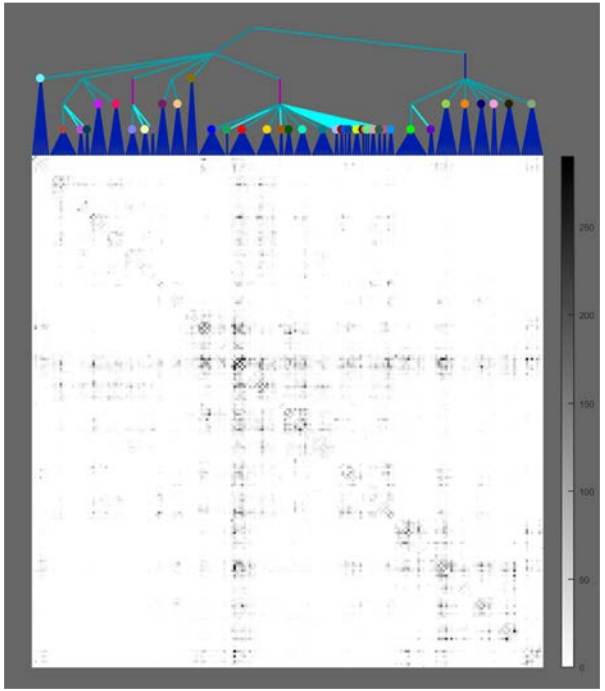
Data visualized from
Helmstaedter, M., Briggman, K. L., Turaga, S. C., Jain, V., Seung, H. S., & Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, *500*(7461), 168.

QCM ANALYSIS OF THE IPL CONNECTOME

We applied the modified version of QCM to 1076x1076 weighted IPL connectivity data by

- Computing similarity matrix from network information (heat kernel, cosine similarity of nodes)
- Apply QCM to similarity matrix derived from this
  - Developed 203 communities
  - Per Community
    - Mean 5.3 cells, median 4 cells per
    - Range 2 to 21 neurons (non-glial)

Result is a 203x203 intermodular connectivity matrix

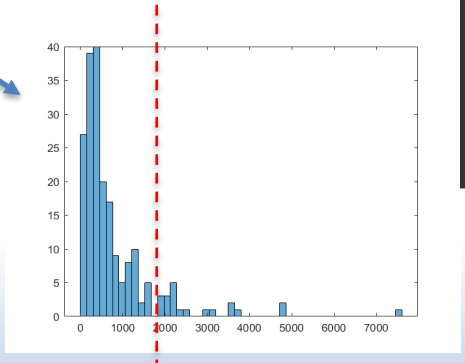


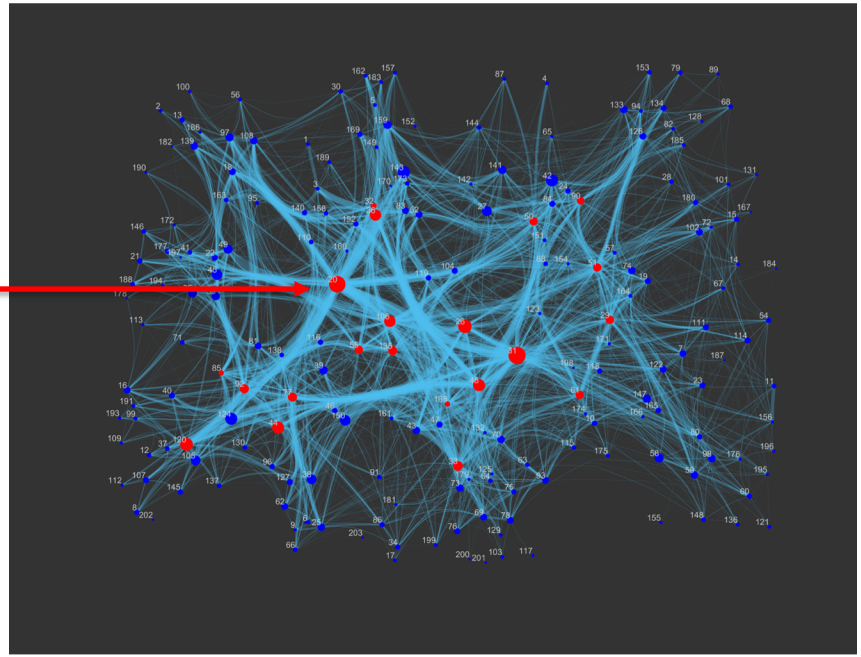
BIOLOGICAL INSIGHTS FROM ANALYSIS

Connected communities identified by QCM reveal low level modular structure as well as hierarchical structure related to wide field connectivity cells calls starburst amacrine cells and collector ganglion cell networks

Red nodes (21 of 203)

- denote most highly connected communities, ‘network hubs’
- 64.7% of intermodular connectivity (summed contact area)





*visualization using divided edge bundling

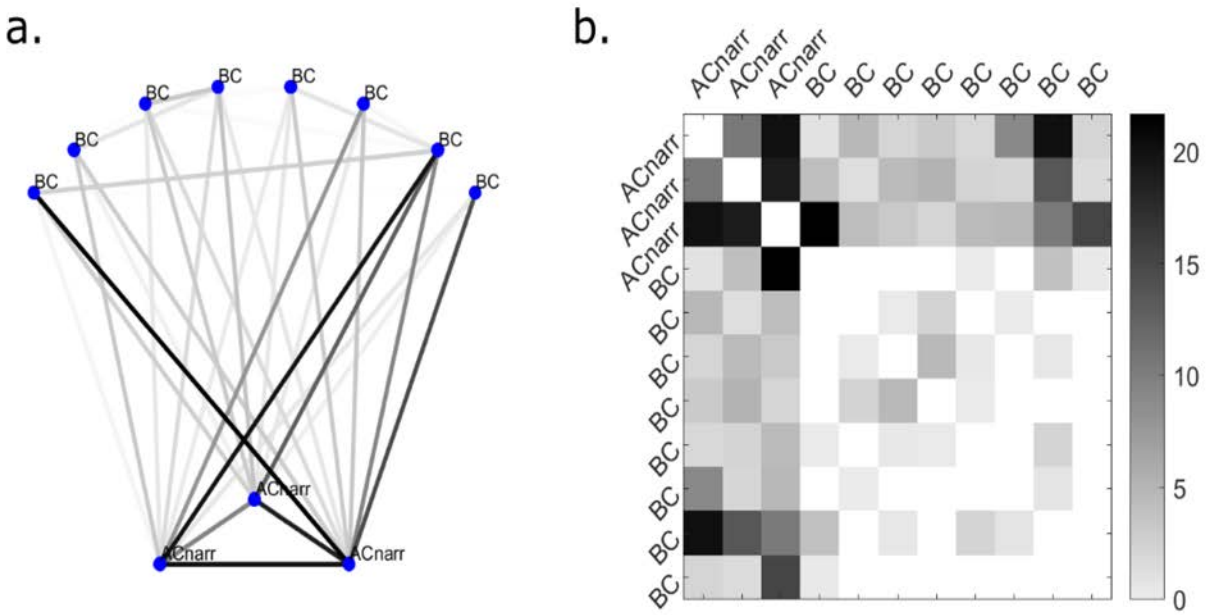
Selassie, D., Heller, B., & Heer, J. (2011). Divided edge bundling for directional network data. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2354-2363.

WEST VIRGINIA UNIVERSITY
Department of Mathematics

BIOLOGICAL INSIGHTS FROM ANALYSIS

Other modules:

Groups of bipolar cells synapsing onto amacrine narrow cells



BIOLOGICAL INSIGHTS FROM ANALYSIS

Select all communities with amacrine narrow cells

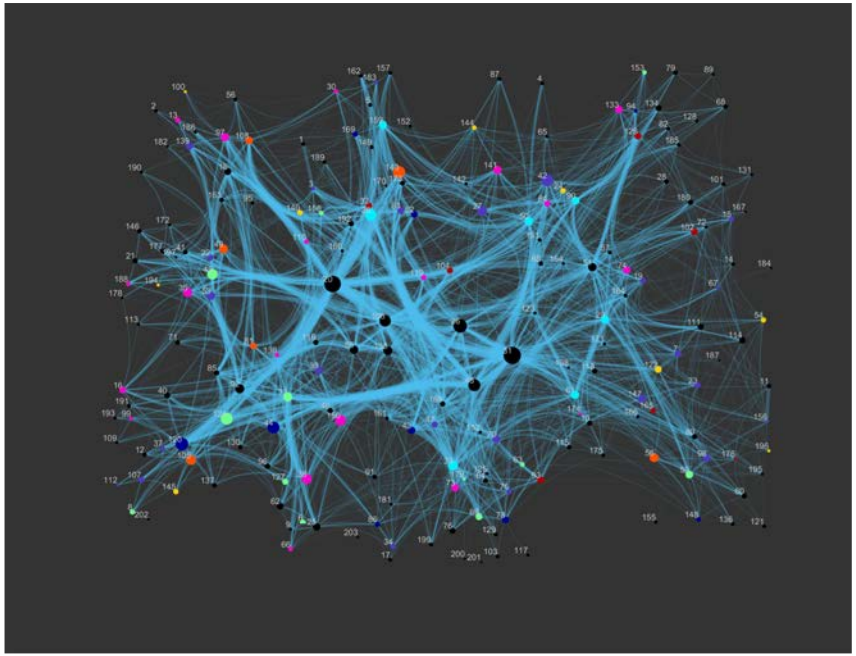
- Present in 93 of the 203

Black = no amacrine narrow cells

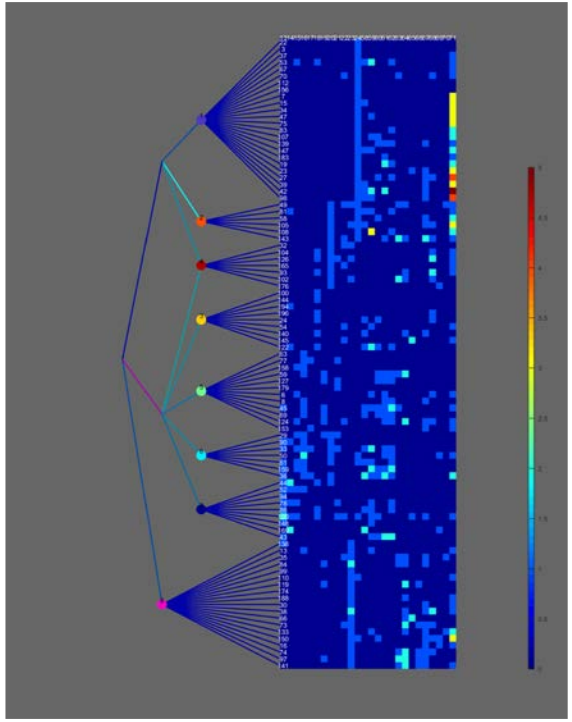
Perform second QCM process on the 93 using number of each type of amacrine narrow

- Define vector for each community
- Compute similarity, run QCM, cut

Found second order collection of 8 groups of communities



BIOLOGICAL INSIGHTS FROM ANALYSIS



Modules with amacrine narrow aligned by subtype as shown

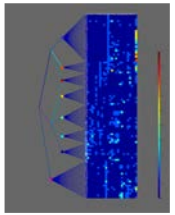
- Vertical columns = subtypes as in (Helmstaedter et al, 2013)
- Each row one of 93

Eight communities identified by QCM from top to bottom

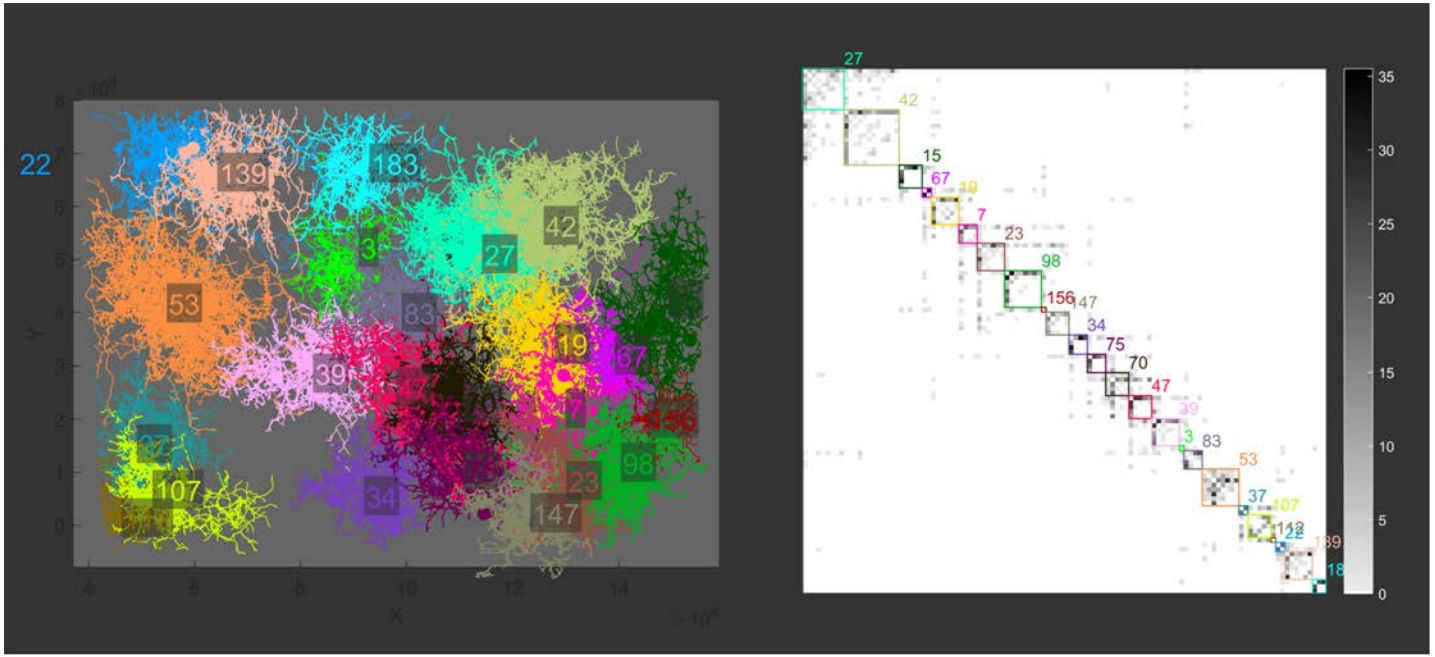
Group 1 (purple)

- Each module has exactly one type 24 (AII) cell.
- 3-5 cells in total
- Expected composition of a module is one type 24, one or two rod bipolar, one or two OFF-cone bipolar.

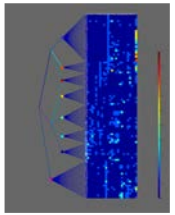
BIOLOGICAL INSIGHTS FROM ANALYSIS

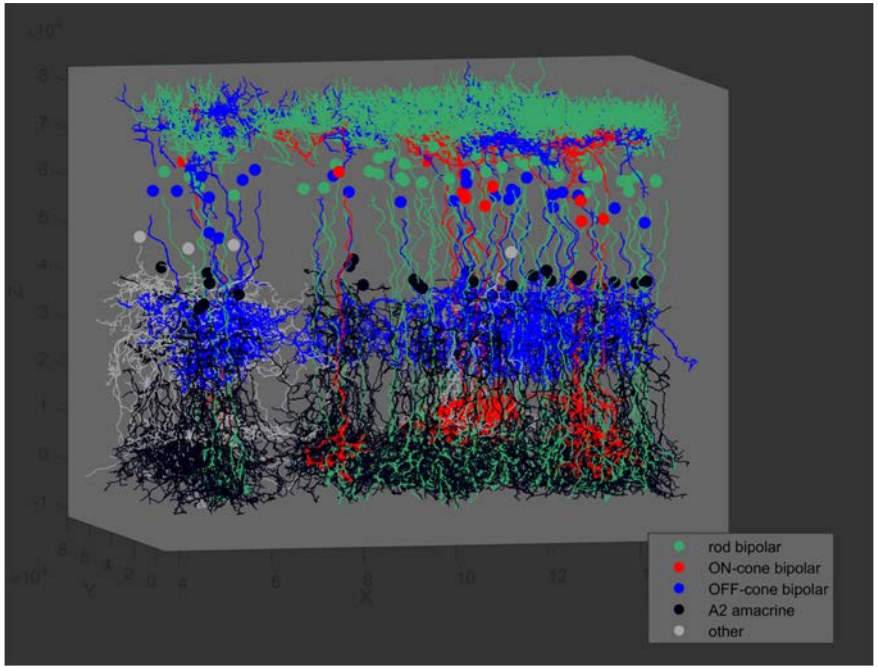


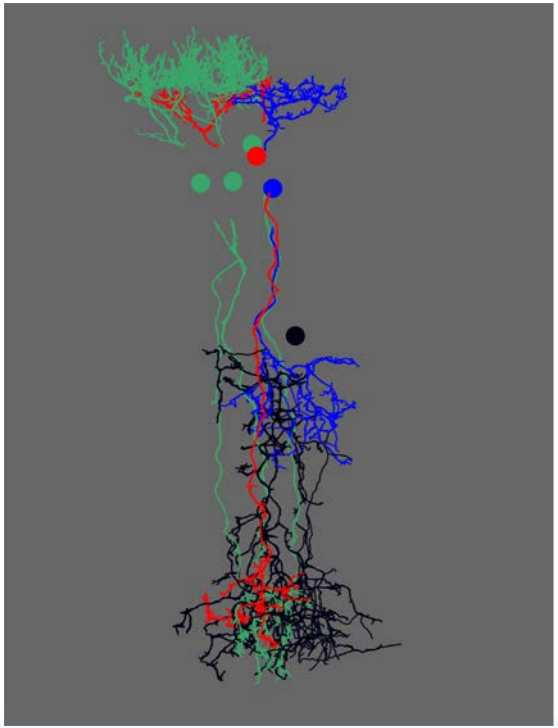
Group 1 amacrine narrow communities



BIOLOGICAL INSIGHTS FROM ANALYSIS







Group 1 amacrine narrow side view

community 39

PROBLEMS AND CHALLENGES

- Ground truth?
  - No ‘best’ clustering algorithm
- Scaling
  - 1000 neurons to 10k, 100k, to ???
  - Fly: 250k with 10^7 connections
  - Mouse: 70,000,000 with 10^12 connections
  - Human: ~10^11 with 10^15 connections
- HHMI/Janelia fly connectome
  - Porting to neo4j
  - Recoding for connectivity + metadata (arbor, cell type, strength…)

QUESTIONS?

- Thanks
- ef@math.wvu.edu or ejfuller@gmail.com