From Mind to Machine: *Neuronal Circuits, Learning Algorithms,* and Beyond

“Tony” Runzhe Yang

09/26/2023

Thesis committee

H. Sebastian Seung
(advisor ♥)
Evnin Professor of Neuroscience.
Professor of Computer Science.

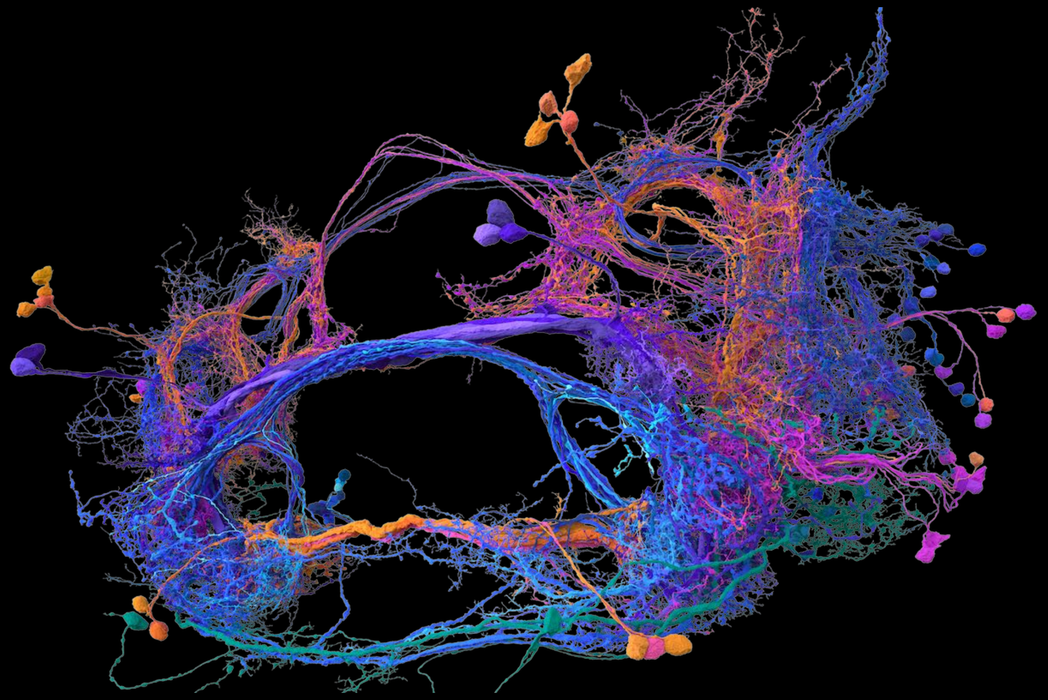
Karthik Narasimhan
(advisor ♥)
Assistant Professor, Computer Science.
Co-director, Princeton NLP.

Mala Murthy
Karol and Marnie Marcin ’96 Professor of Neuroscience. Director, Princeton Neuroscience Institute.

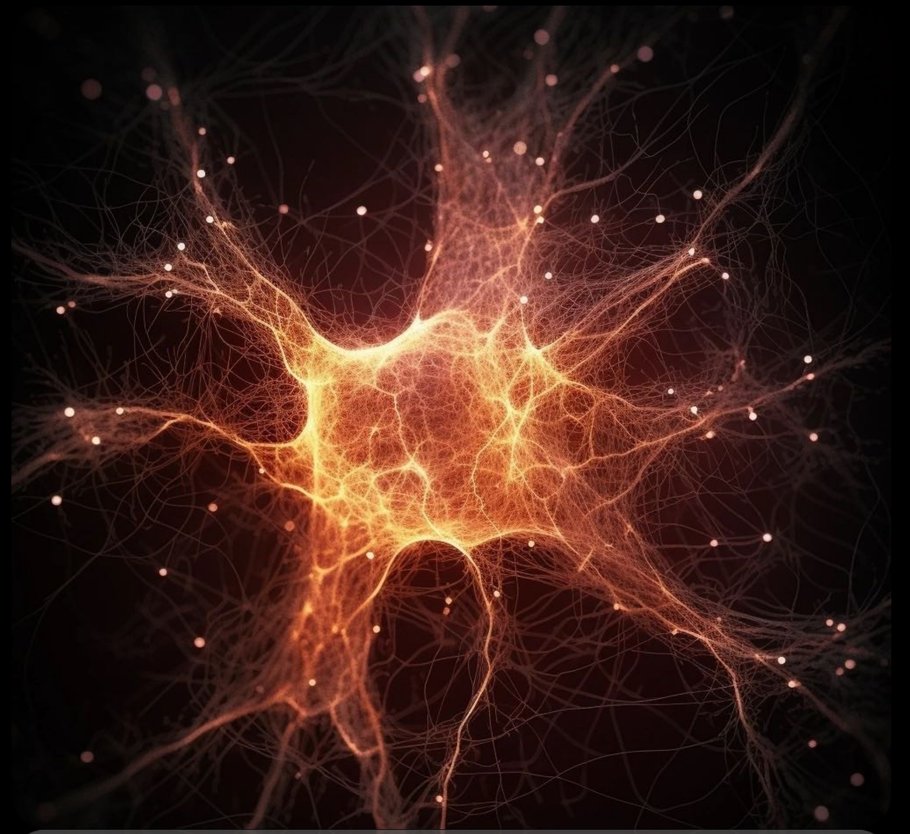
Ryan P. Adams
Professor of Computer Science.
Associate Chair, Princeton Computer Science.

Dmitri ‘Mitya’ Chklovskii
Group Leader, Neural Circuits and Algorithms, CCN, Flatiron Institute.
Research Associate Professor, NYU.

New frontiers during my PhD: from brain reconstruction to generative AI

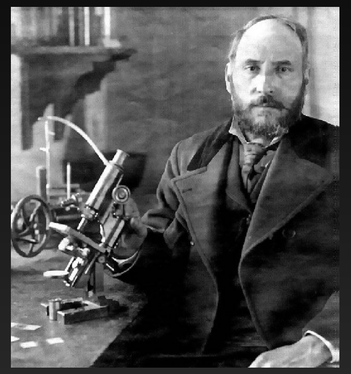


3D Reconstruction of Neuronal Circuits



Prompt: "A photograph of neuron-like stardust"

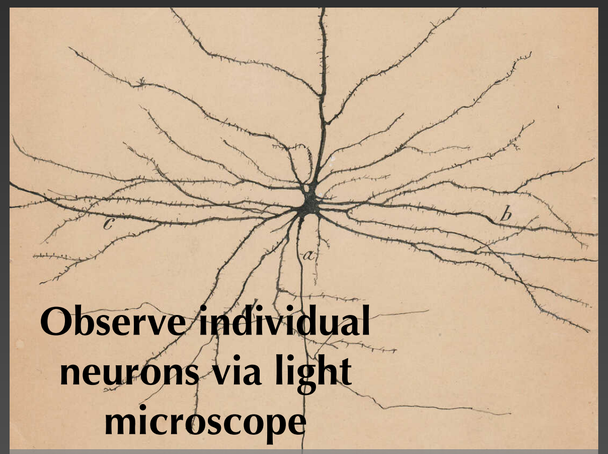
Part I: the organization of biological neural networks



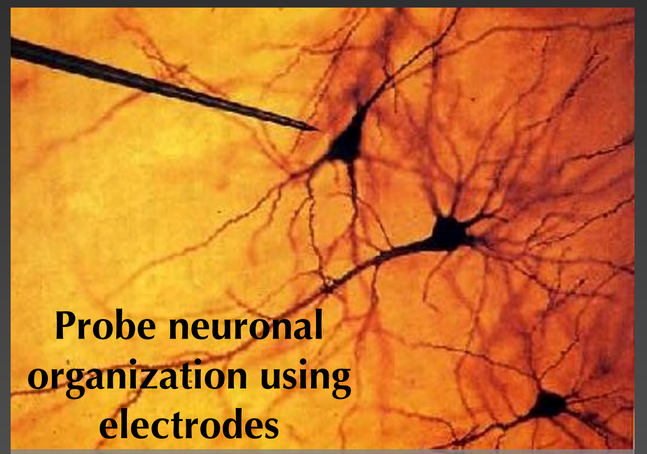
Ramón y Cajal
(1906 Nobel Prize)

Hubel & Wiesel
(1981 Nobel Prize)

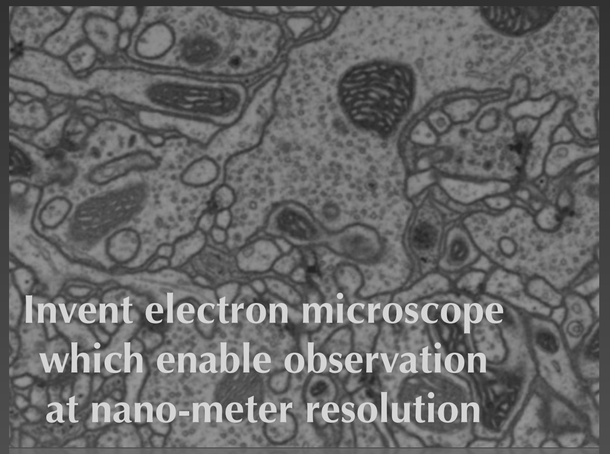
Ernst Ruska
(1986 Nobel Prize)



Influence to AI:
McCulloch-Pitts Neurons,
Perceptron…

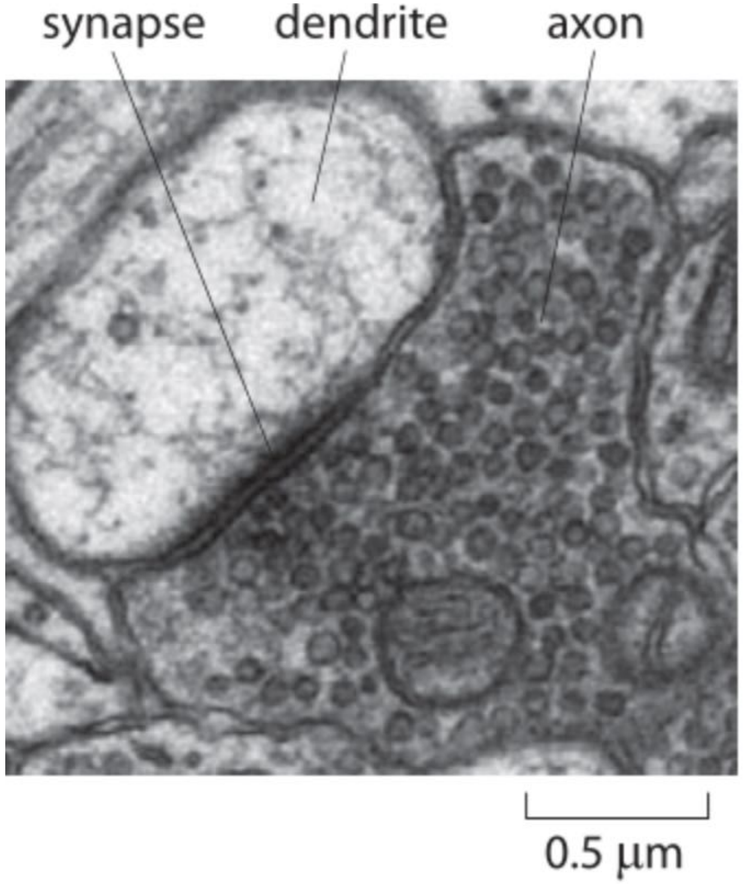


Influence to AI:
Cognitron, Convolutional
Neural Networks…

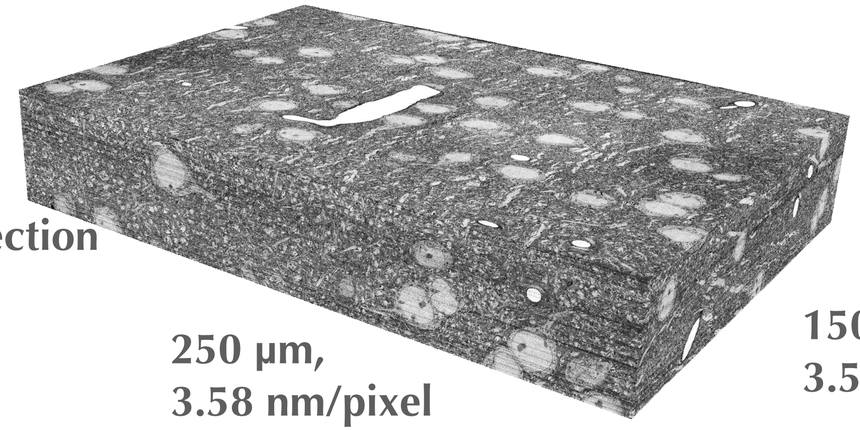


Influence to AI &
Neuroscience:
EM Connectomes …

Electron microscopy (EM): making synapse identification possible



E.g., Serial Section of Mouse V1 L2/3



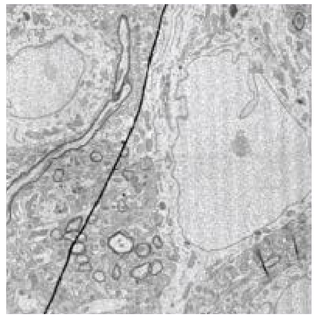
40 µm,
40 nm/section

250 µm,
3.58 nm/pixel

150 µm,
3.58 nm/pixel

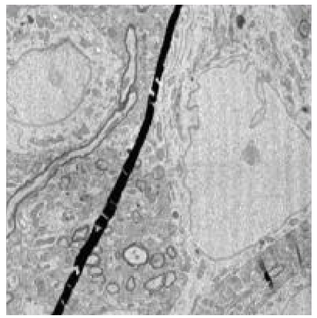
https://www.microns-explorer.org
Layer 2/3 Dataset

Computational pipeline: reconstructing biological neural networks with AI



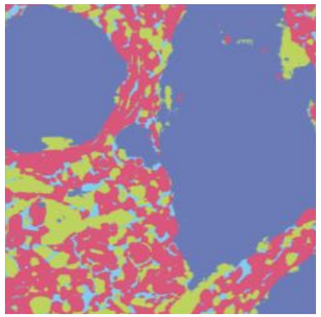
Raw EM Image

alignment



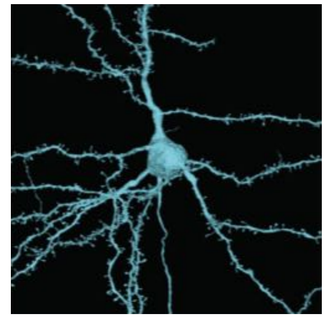
Aligned EM Image

semantic labelling



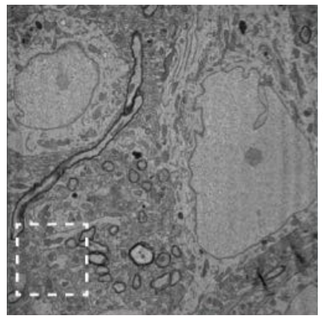
Semantic Segmentation

agglomeration



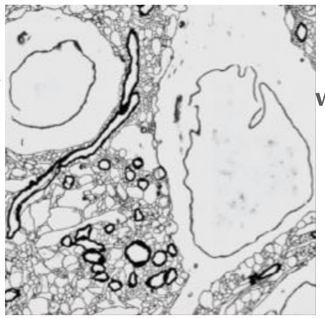
Meshes

meshing



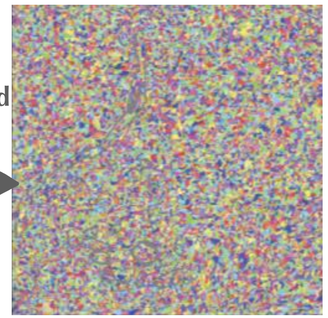
EM Image - 3D

boundary detection

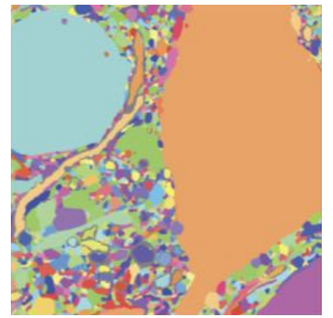


Boundary Map

watershed variants



Oversegmentation

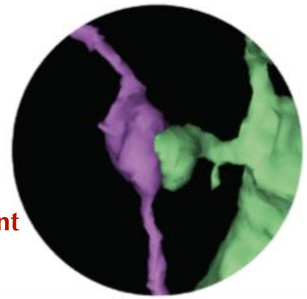


Refined Segmentation

synapse detection

Synapse Candidates

partner assignment



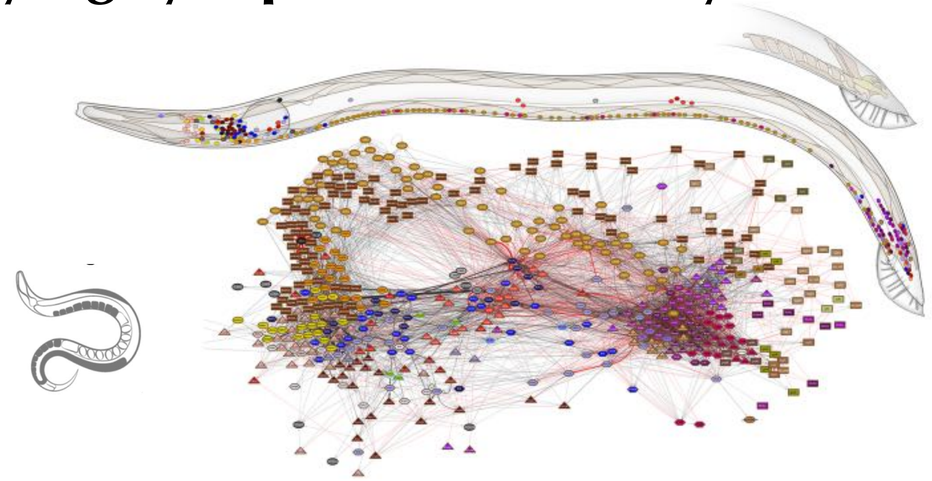
Synaptic Partners

Inferred Connectivity (Neural Networks)

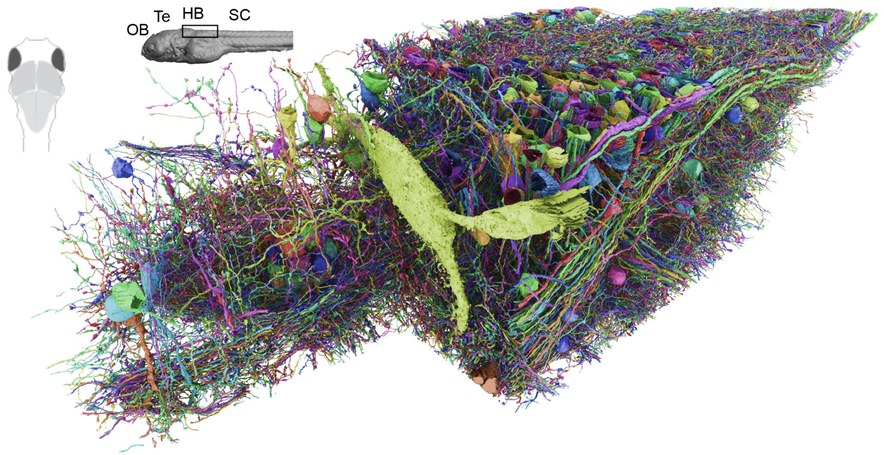
"→": Convolutional Nets + Supervised Learning

Figures modified from *Petascale neural circuit reconstruction: automated methods*. Macrina *et al.*, 2021

Studying synaptic connectivity at scale is possible



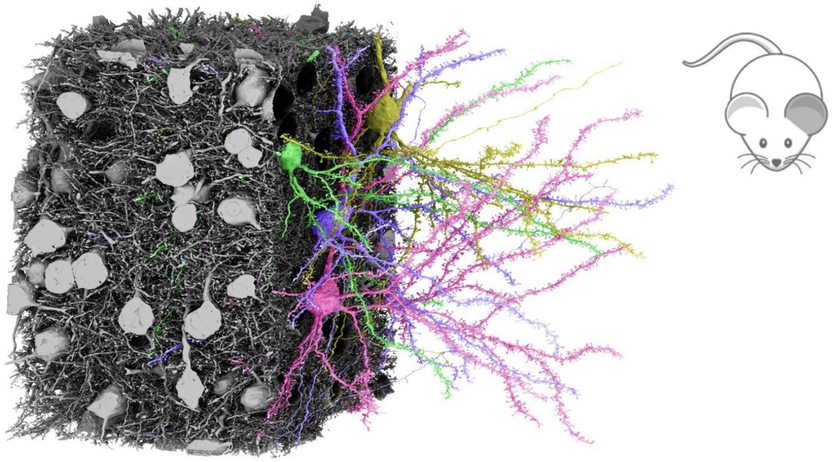
***C. elegans* whole connectome (Cook et al., 2019)**
385 neurons (male), 4,887 chemical connections



Larval zebrafish hindbrain (Vishwanathan et al., 2019)
~3000 neurons, ~45k chemical connections

...

Adult *Drosophila melanogaster*
whole-brain connectome (Dorkenwald et al., 2023)
120K+ neurons, 30M+ chemical synapses

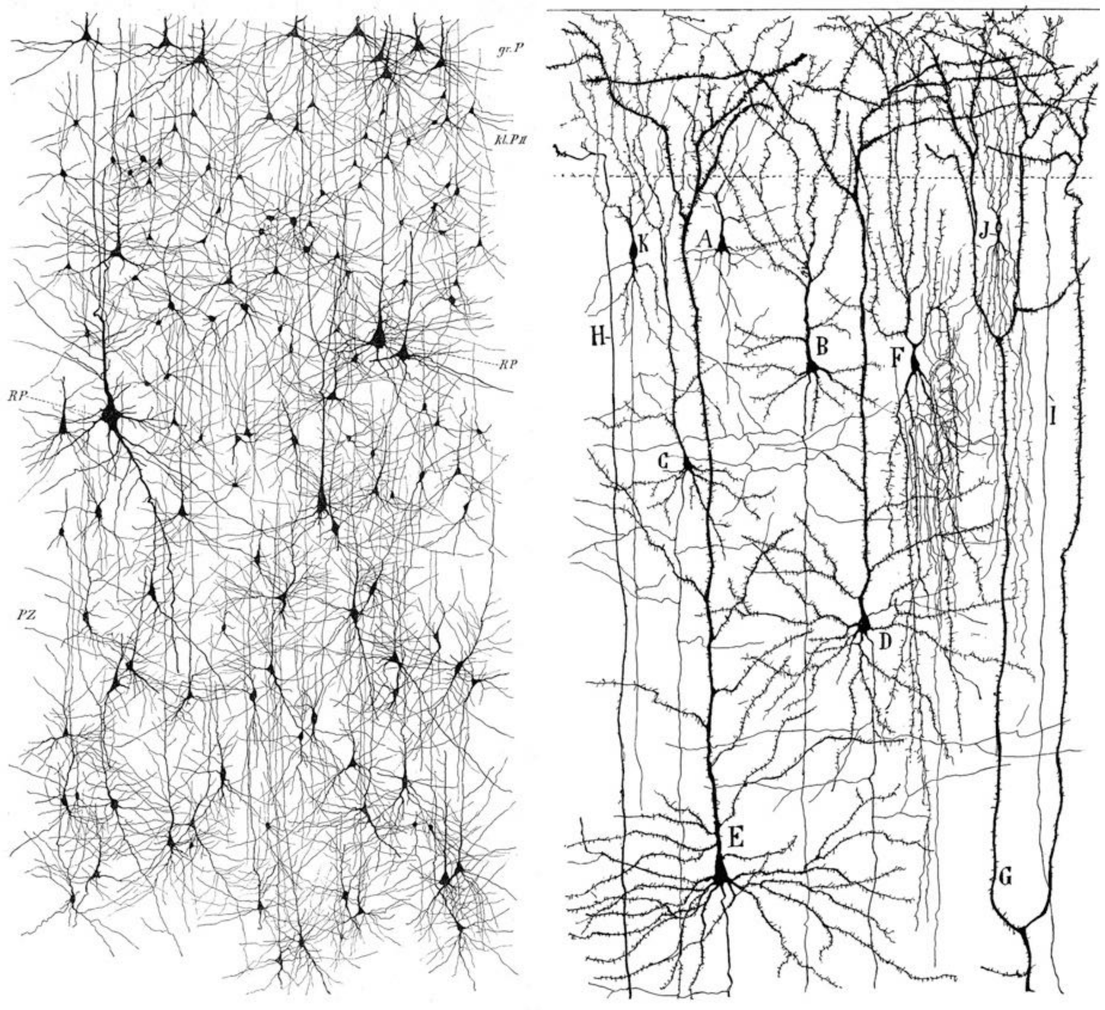


1mm³ mouse cortex (MICrONs Consortium et al., 2021)
~200,000 cells, ~ 523M synapses

...

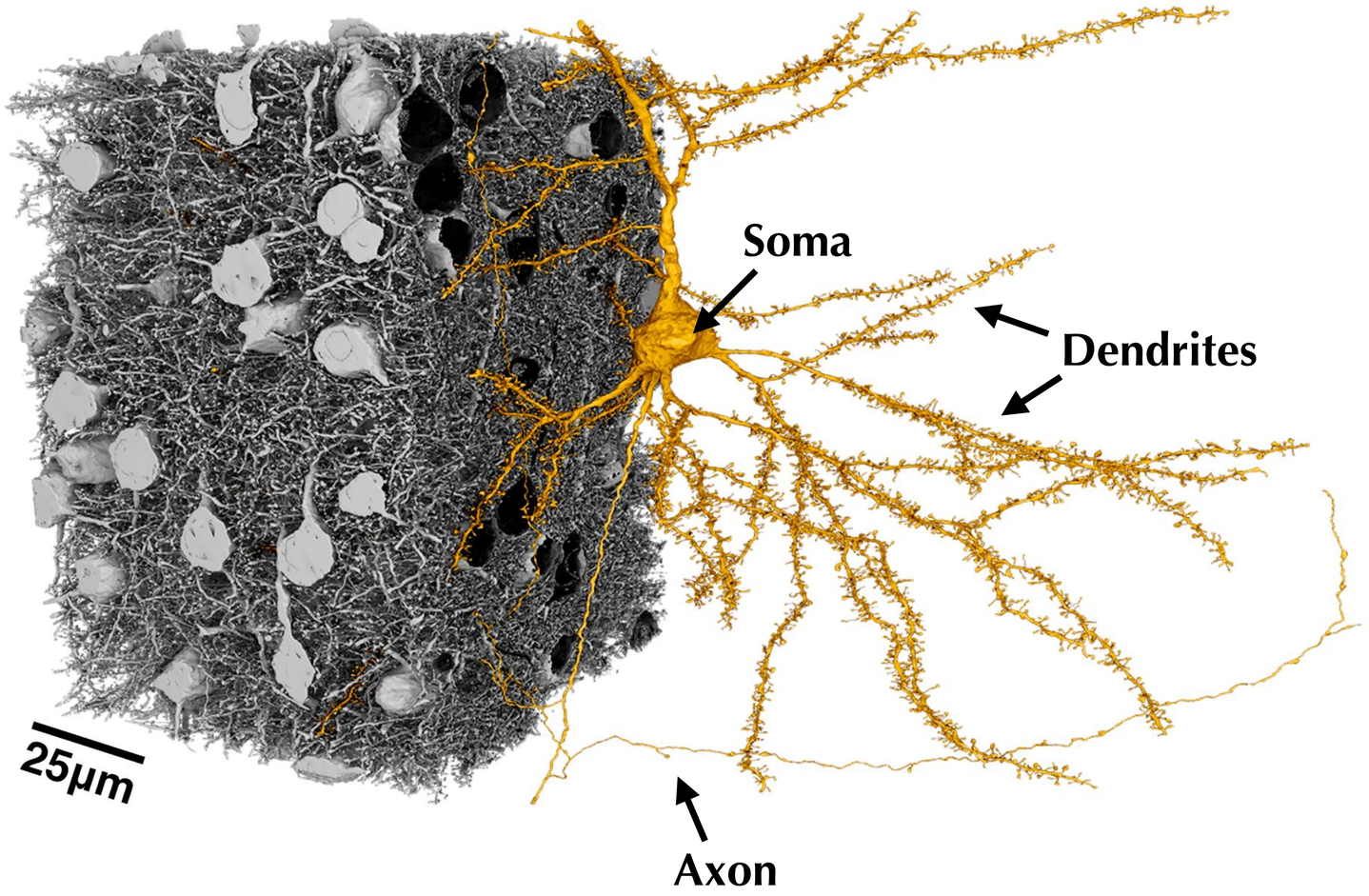
(Naive) Question 1:
Are biological neural networks wired randomly?

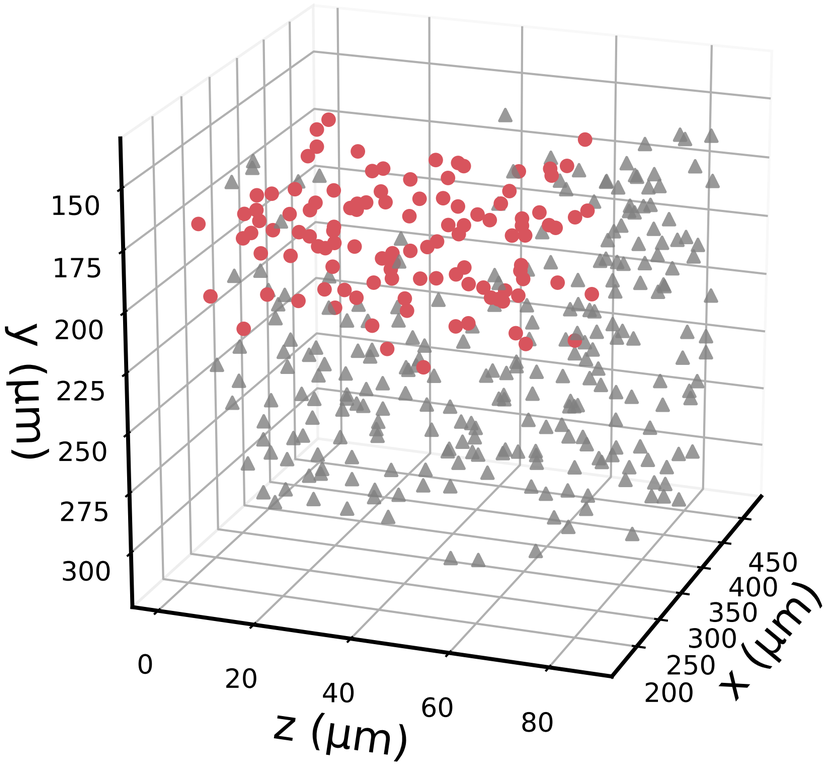
Are cortical connections random?



Drawings of the human cerebral cortex (Golgi method),
from von Kölliker (1893) and Cajal (1899).

A network of pyramidal cells (PyCs) in layer 2/3 of the mouse visual cortex





111 out of total 363 PyCs with ≥100µm axons

659 synaptic connections among them

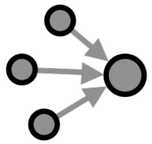
The degree distribution of the PyC subgraph is “wider” than expected

Erdős-Rényi Model (ER)

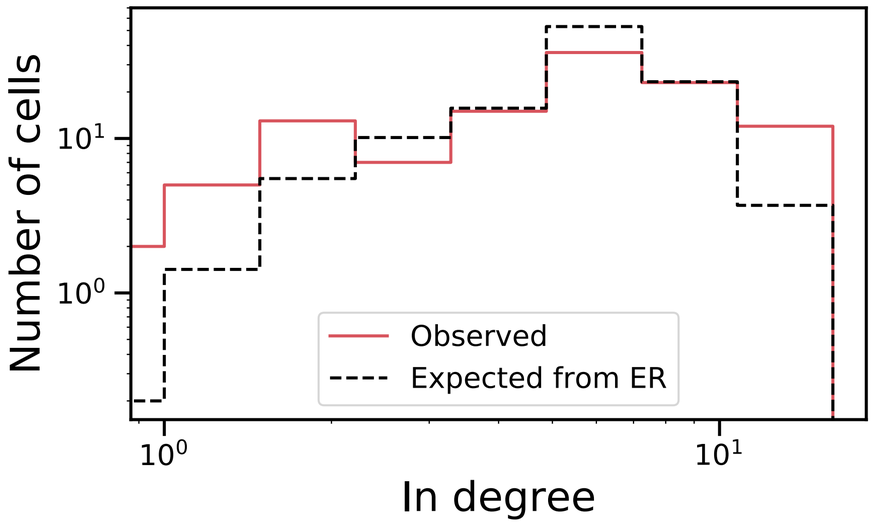
$\Pr[(u,v) \in E] = m/(n(n-1))$

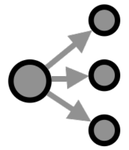
Connection Probability ≈ 0.054

- Directed edges are drawn independently with the same probability.
- The expected number of edges matches the observation.

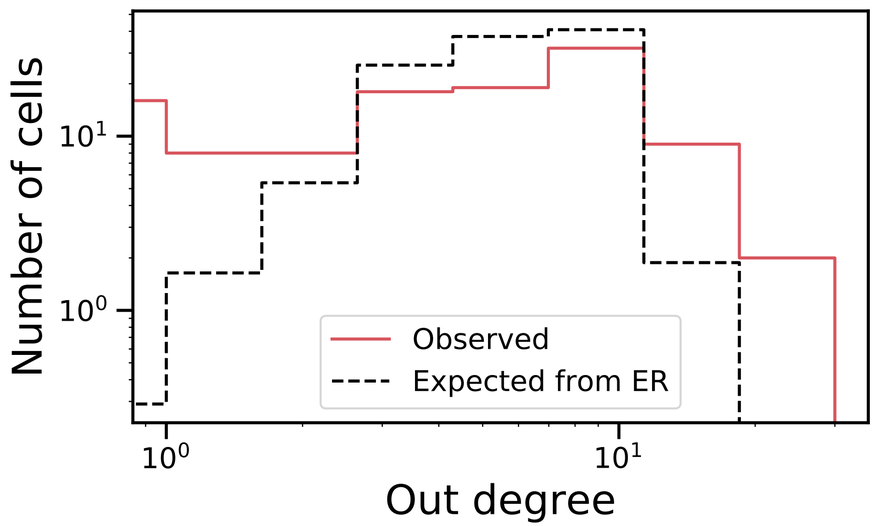


In degree =
# of presynaptic PyCs

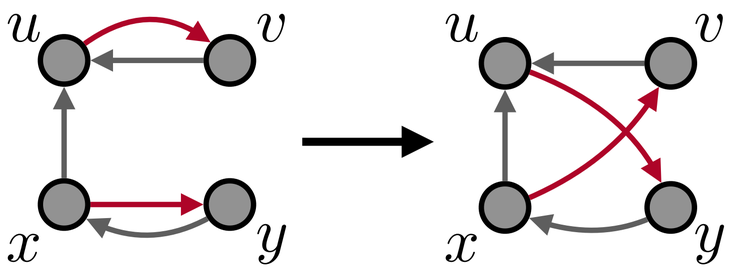




Out degree =
# of postsynaptic PyCs



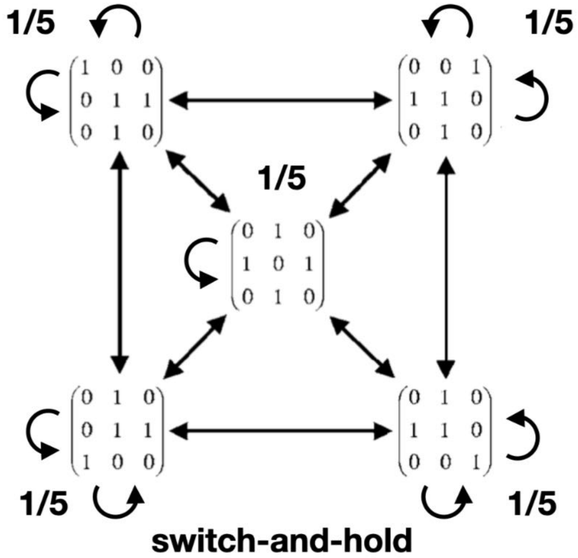
An alternative null-model: hold the degree sequence during random rewiring



Configuration model (CFG)

degree-preserving rewiring

$$(u,v),(x,y) \rightsquigarrow (u,y),(x,v)$$



- The “simplest” random model preserving in- and out-degree sequences.
- A switch-and-hold algorithm samples graphs with same degree sequences uniformly.

Reciprocally connected cells are overrepresented in the PyC subgraph

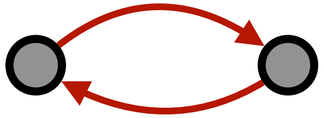
5,475 pairs

Not connected.

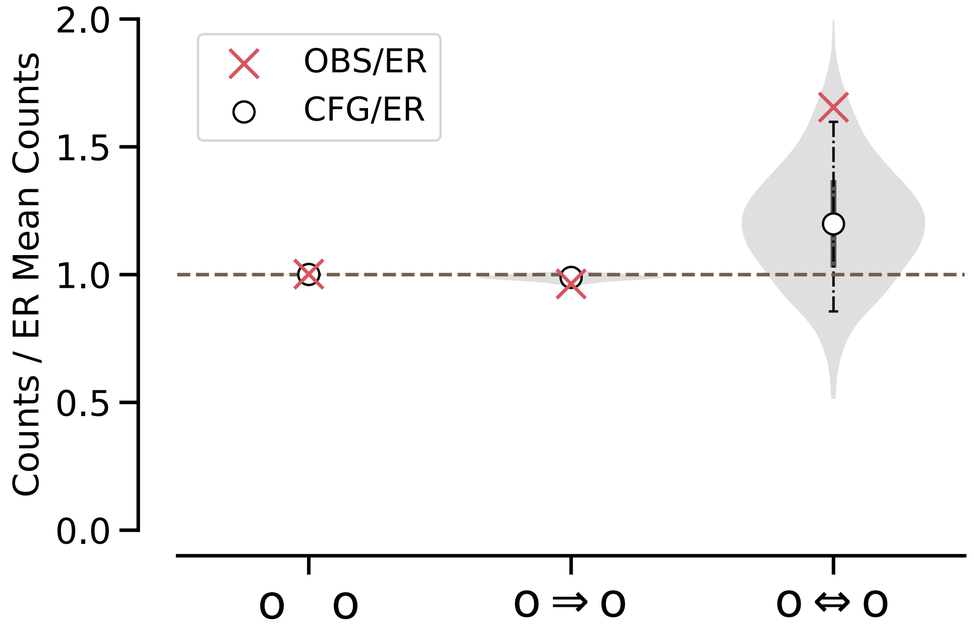
601 pairs

Uni-directionally connected.

29 pairs

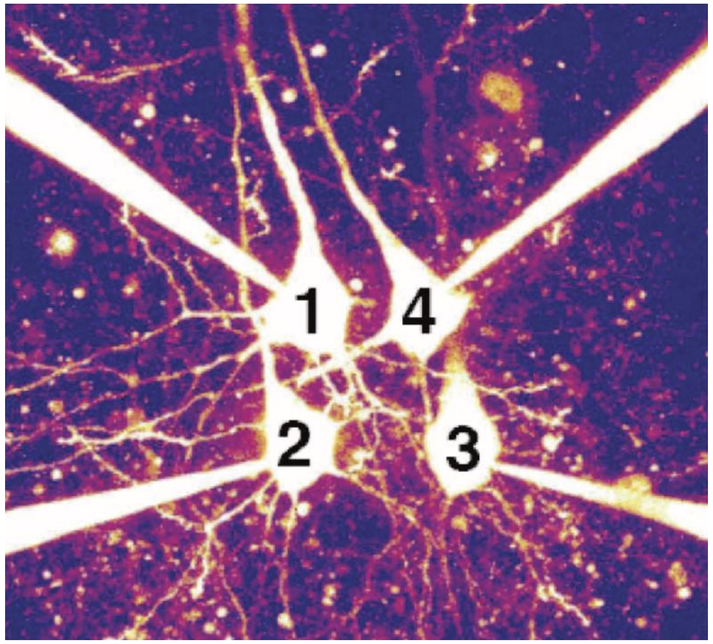


Bi-directionally connected.

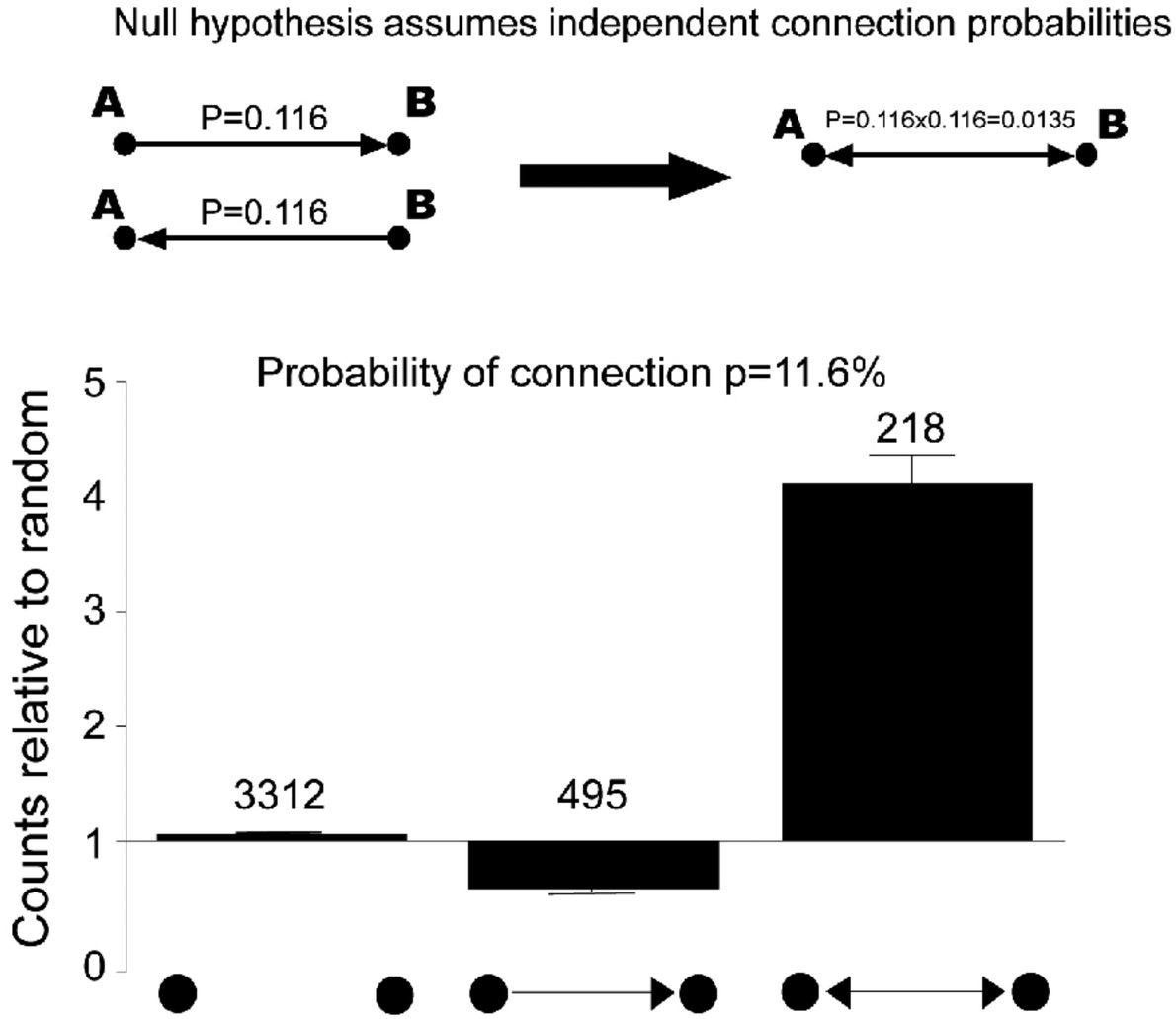


The overrepresentation of bi-directional connections is significant with respect to both ER and CFG.

Reciprocal cortical connections are overrepresented, Song *et al.*, 2005

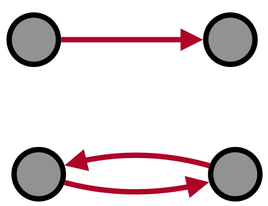


Probe connectivity using quadruple whole-cell recordings



Figures modified from *Highly Nonrandom Features of Synaptic Connectivity in Local Cortical Circuits*. Song *et al.*, 2021

Generalized ER: hold the expected number of bi-directional connections



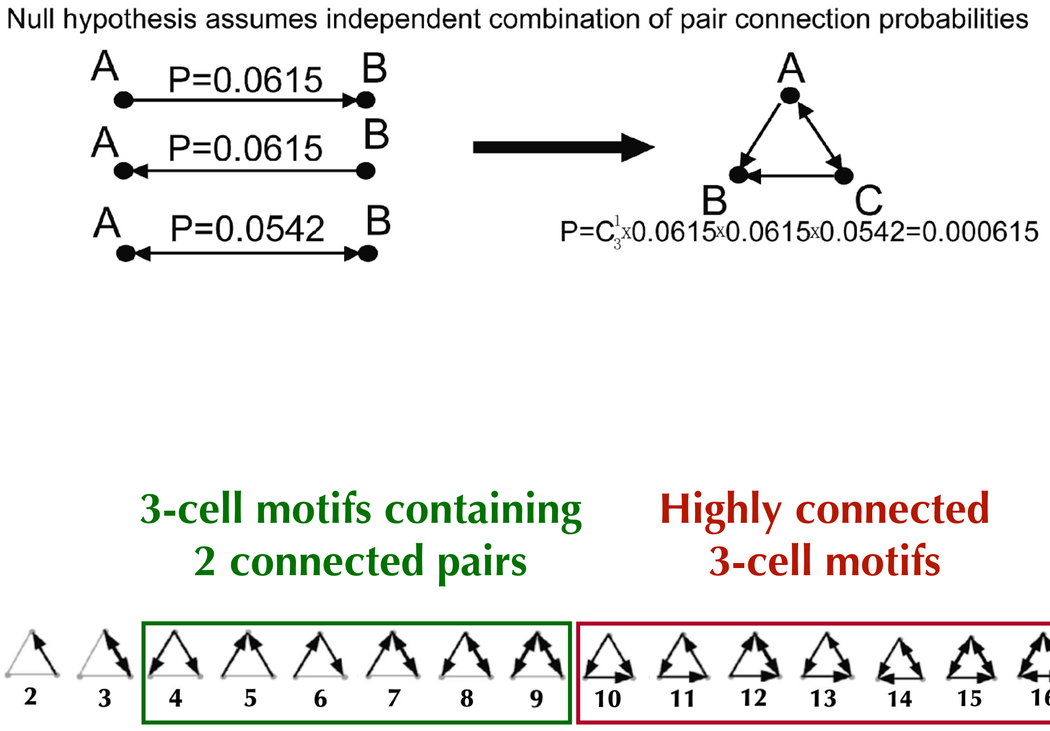
$$p_{uni} = \frac{\#\text{uni-edge}}{n(n-1)}$$

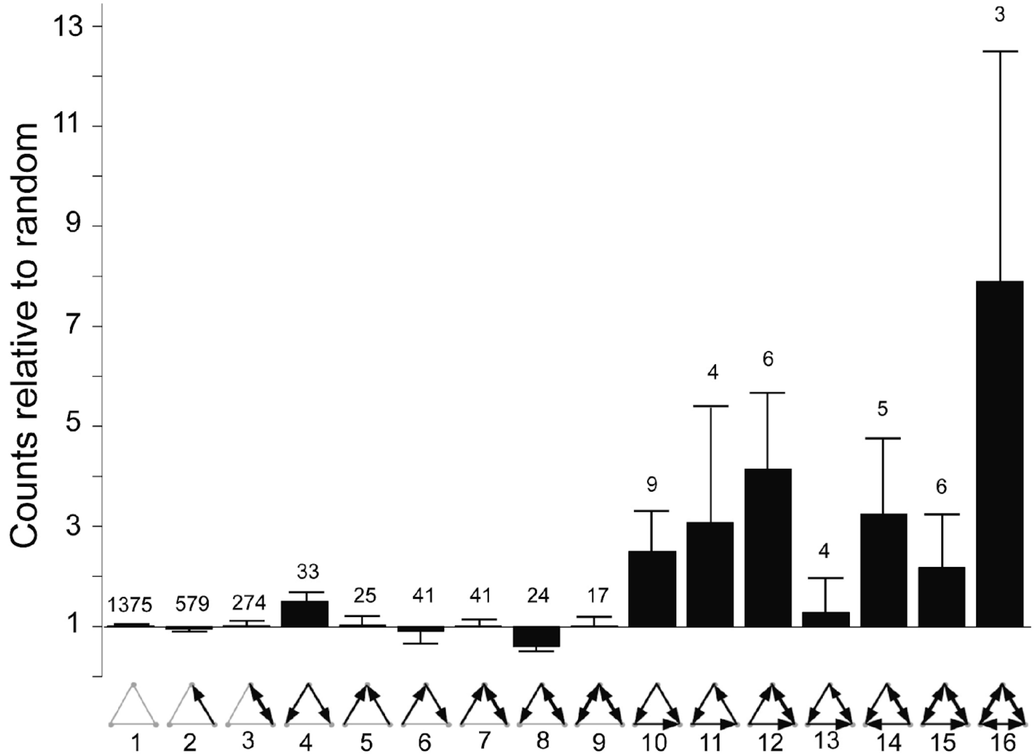
$$p_{bi} = \frac{\#\text{bi-edge}}{n(n-1)/2}$$

Generalized Erdős-Rényi model (gER)
preserving stats of 2-cell motifs

- Edges are drawn independently at random for all pairs of nodes (e.g., A and B):
  - A→B w/ probability p_{uni}
  - A←B w/ probability p_{uni}
  - A↔B w/ probability p_{bi}
  - A B w/ probability$(1 - 2p_{uni} - p_{bi})$

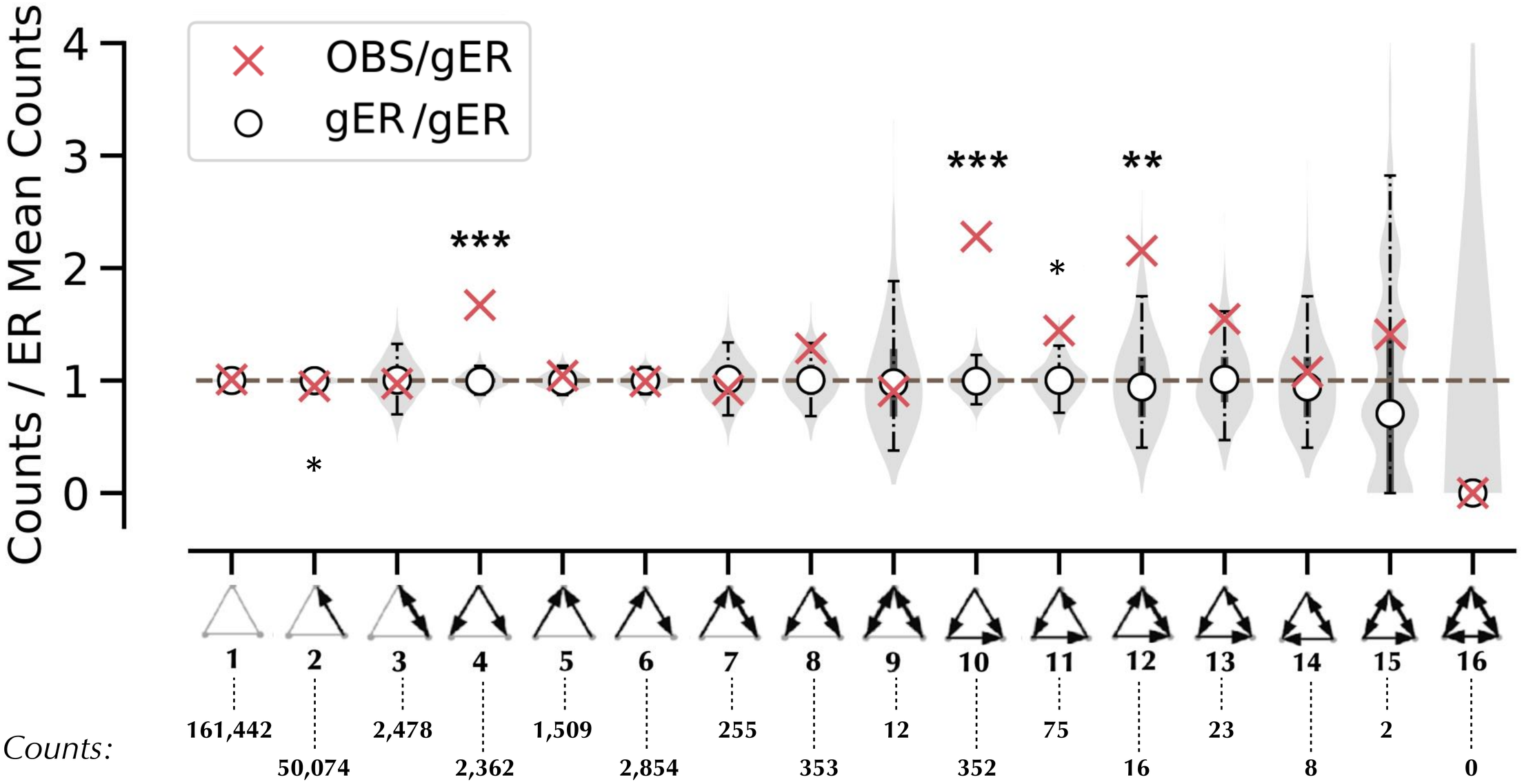
Song *et al.*, 2005 reports overrepresentation of highly connected 3-cell motifs



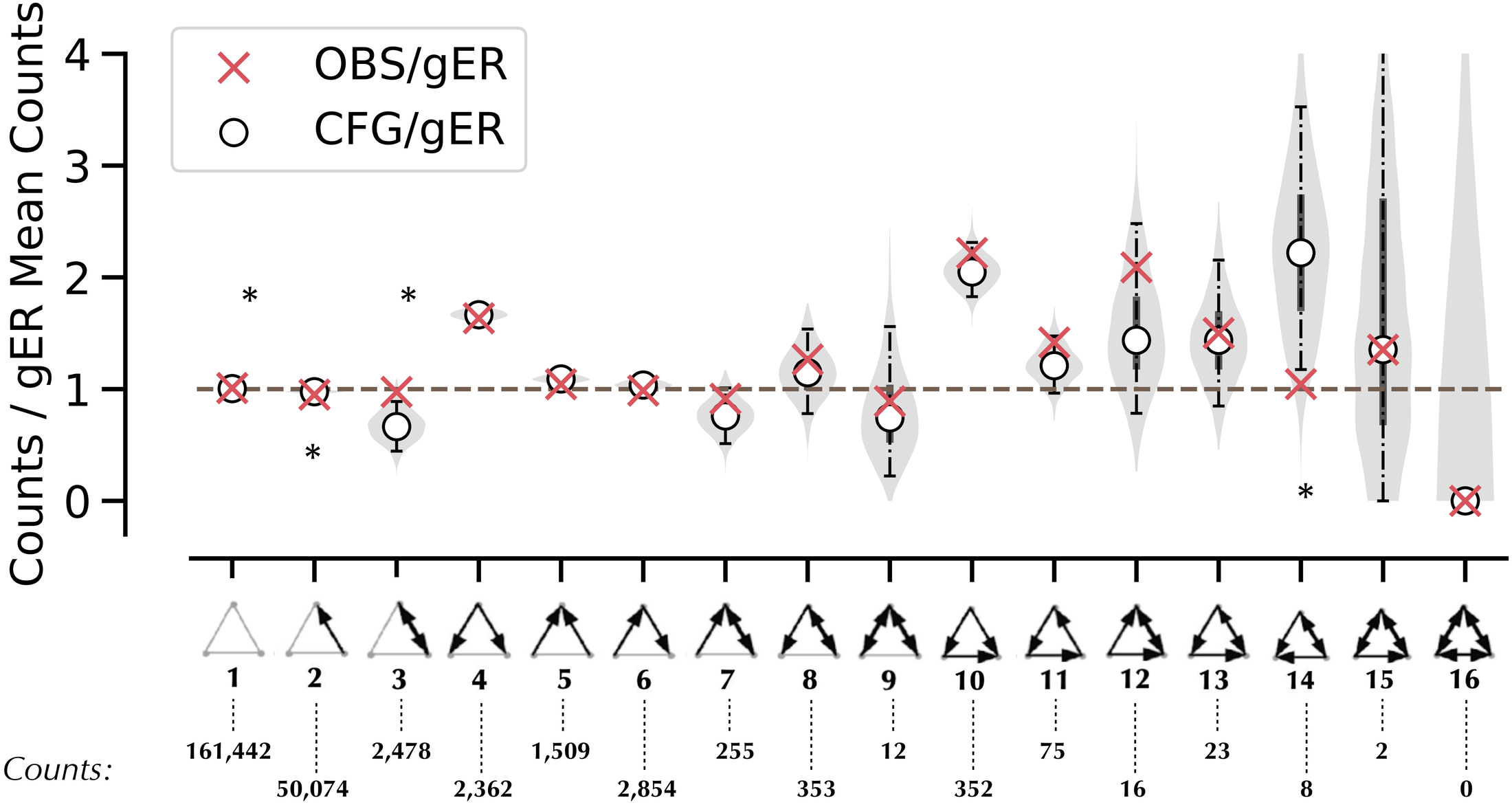


Figures modified from *Highly Nonrandom Features of Synaptic Connectivity in Local Cortical Circuits*. Song *et al.*, 2021

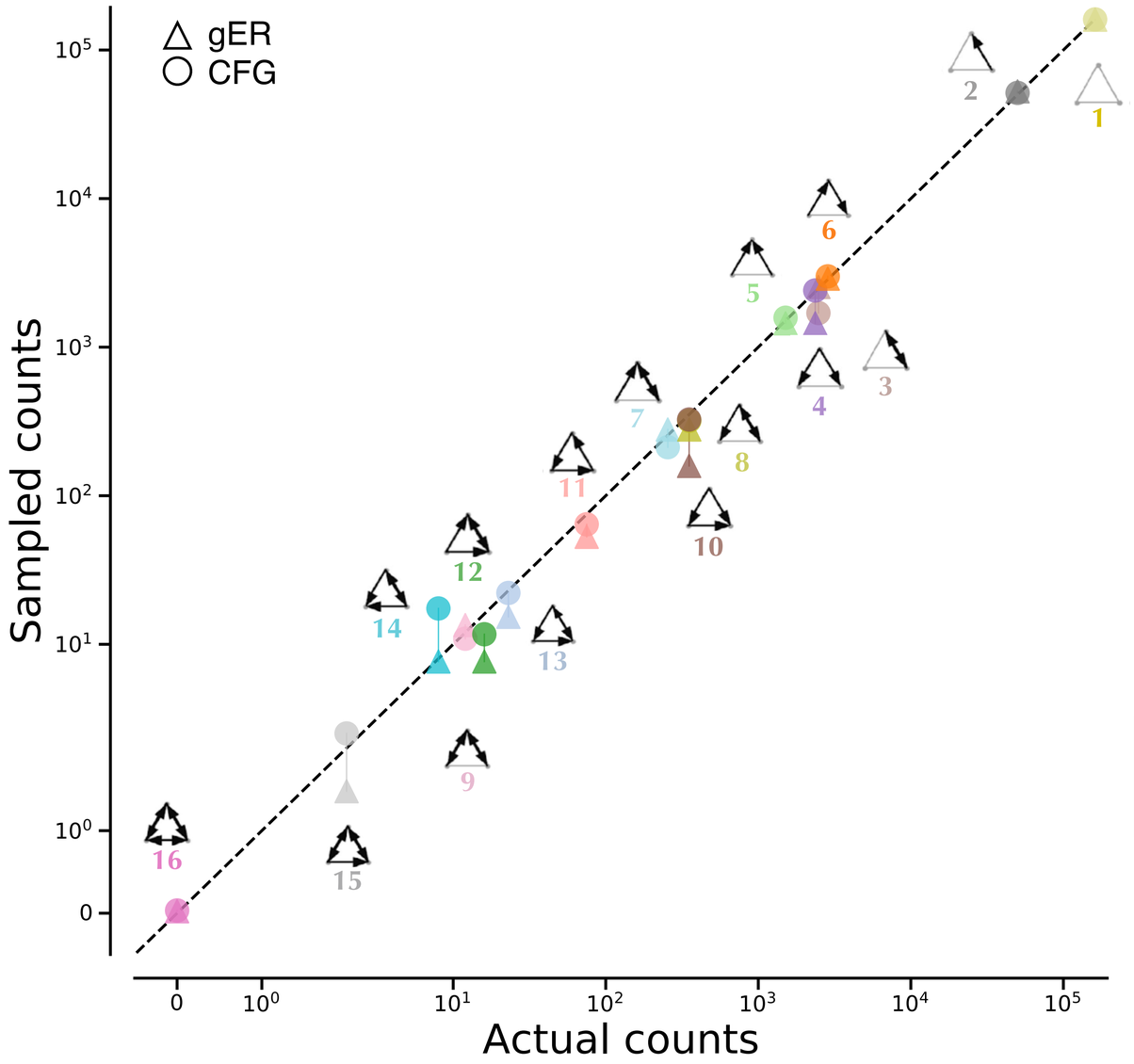
Highly connected 3-cell motifs are highly overrepresented compared w/ gER

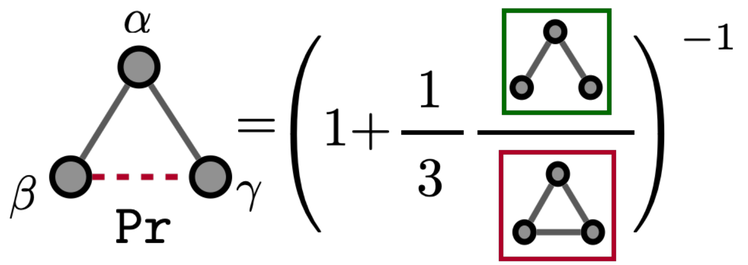


However, most overrepresentation was gone when comparing with CFG



CFG predicts frequencies better than the gER for most motifs





Clustering coefficient

$$C = \mathtt{Pr}[\beta \sim \gamma | \alpha \sim \beta \wedge \alpha \sim \gamma]$$

CFG produces a clustering coefficient close to the observed.

OBS	ER	Generalized ER	CFG
0.161	0.102	0.101	0.150

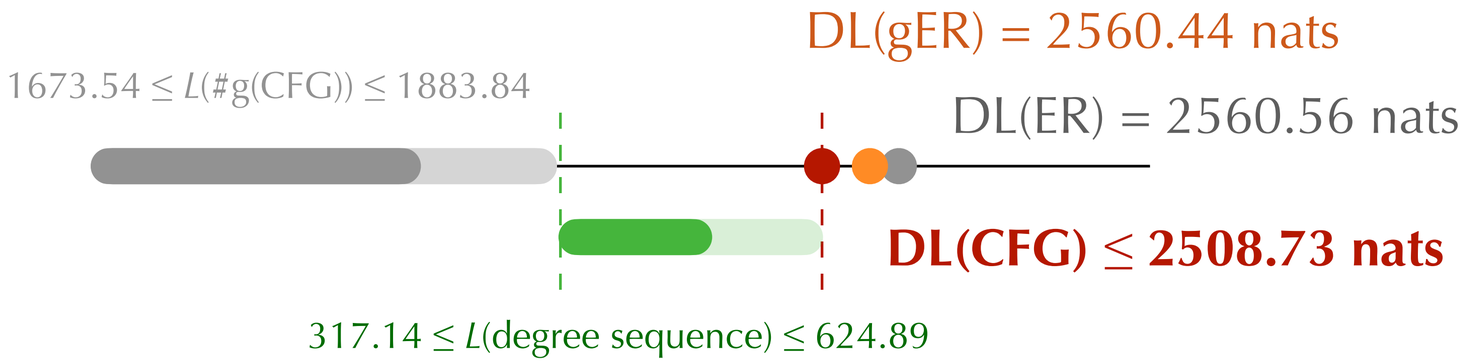
CFG does not simply “overfit” data

Minimal description length (MDL) principle:

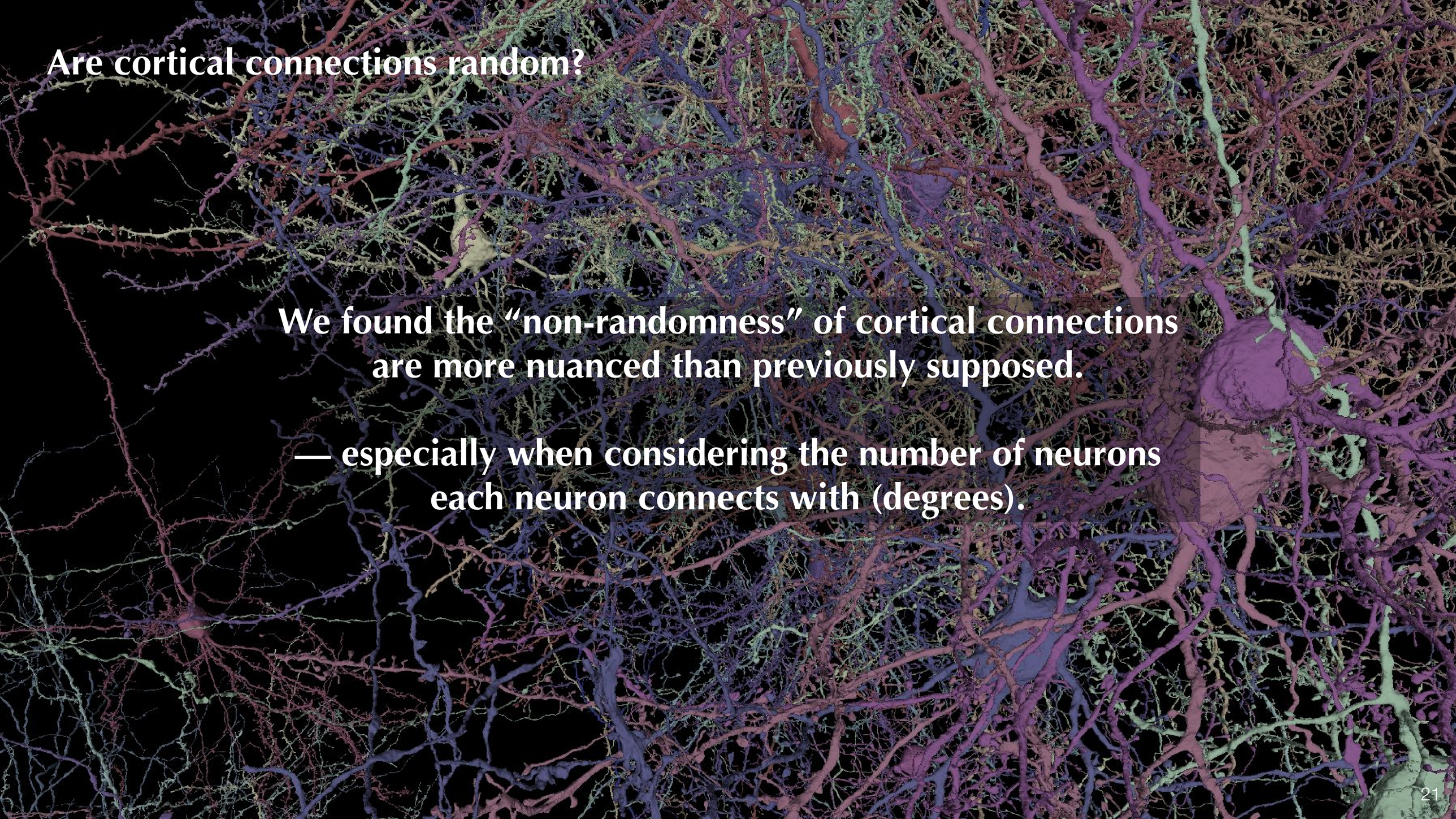
$$L(D) = \min_{H \in \mathcal{H}} \{L(D|H) + L(H)\}$$

# of nats required to specify the data D given the model H

# of nats required to encode the model H



Are cortical connections random?

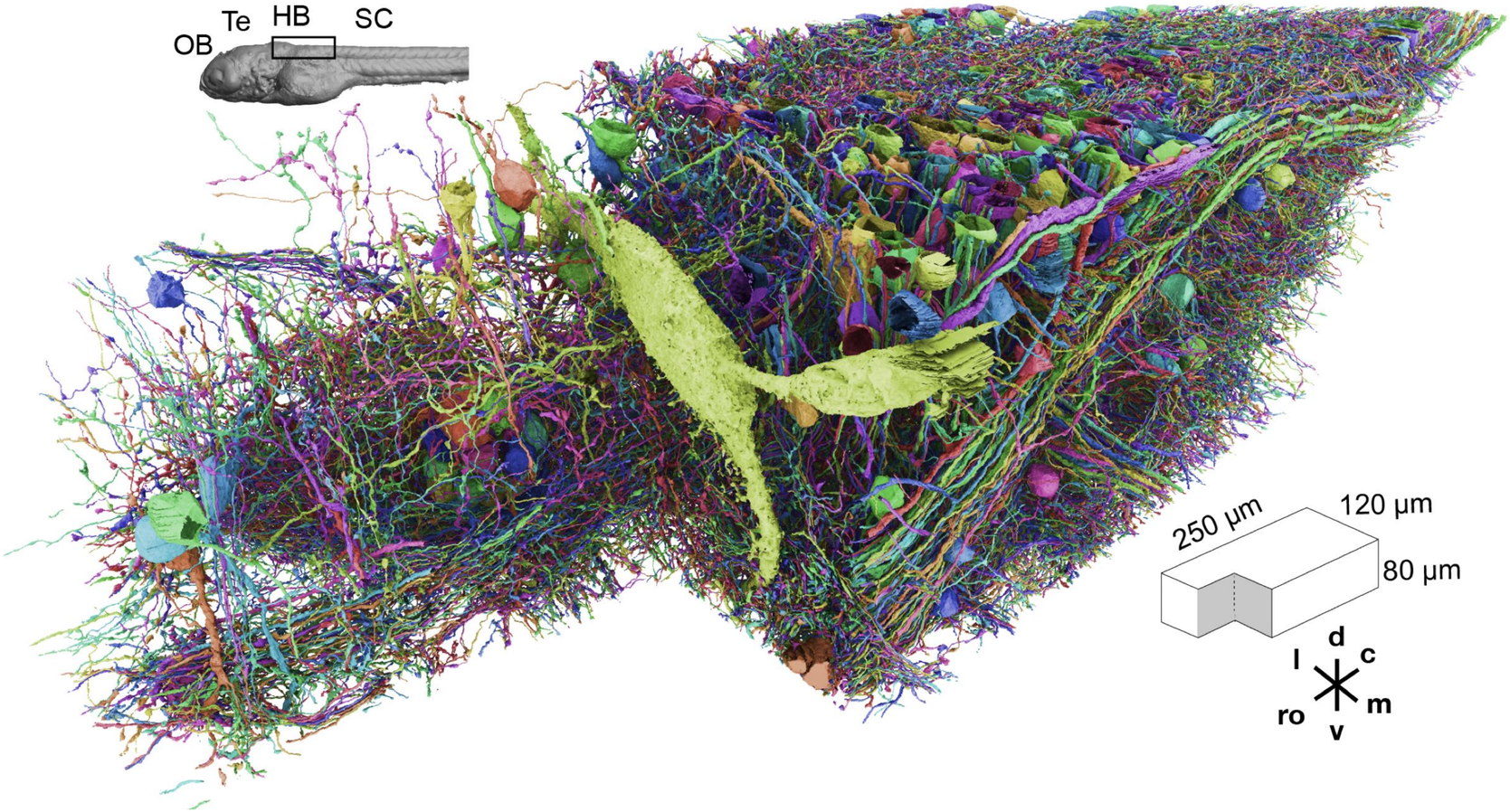


We found the “non-randomness” of cortical connections are more nuanced than previously supposed.

— especially when considering the number of neurons each neuron connects with (degrees).

Question 2:
Are there any examples of interesting/puzzling wiring patterns in the brain?

Connectivity motifs in a larval zebrafish hindbrain



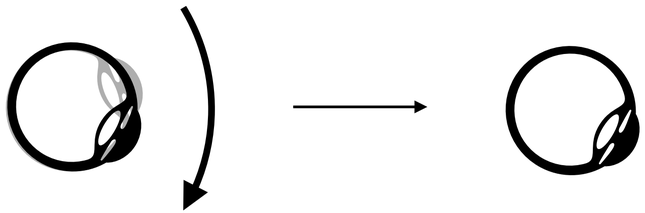
2,883 neurons

44,969 connections

75,195 synapses (within graph)

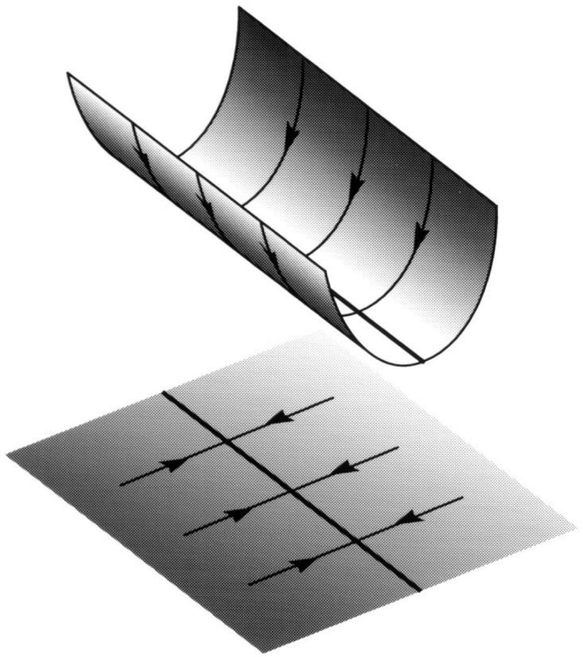
most (~2344) neurons are in the "periphery"

Why interesting? Test the classical idea of low dimensional attractors

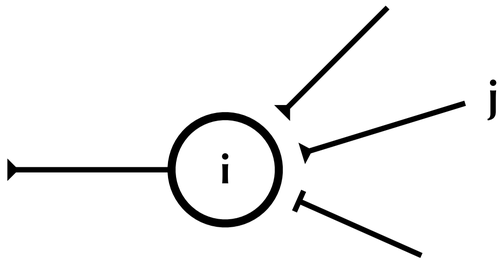


saccade
~100 msec

fixation
1 sec~min



Seung 1996



postsynaptic firing rate

synaptic weights

vestibular inputs, etc.

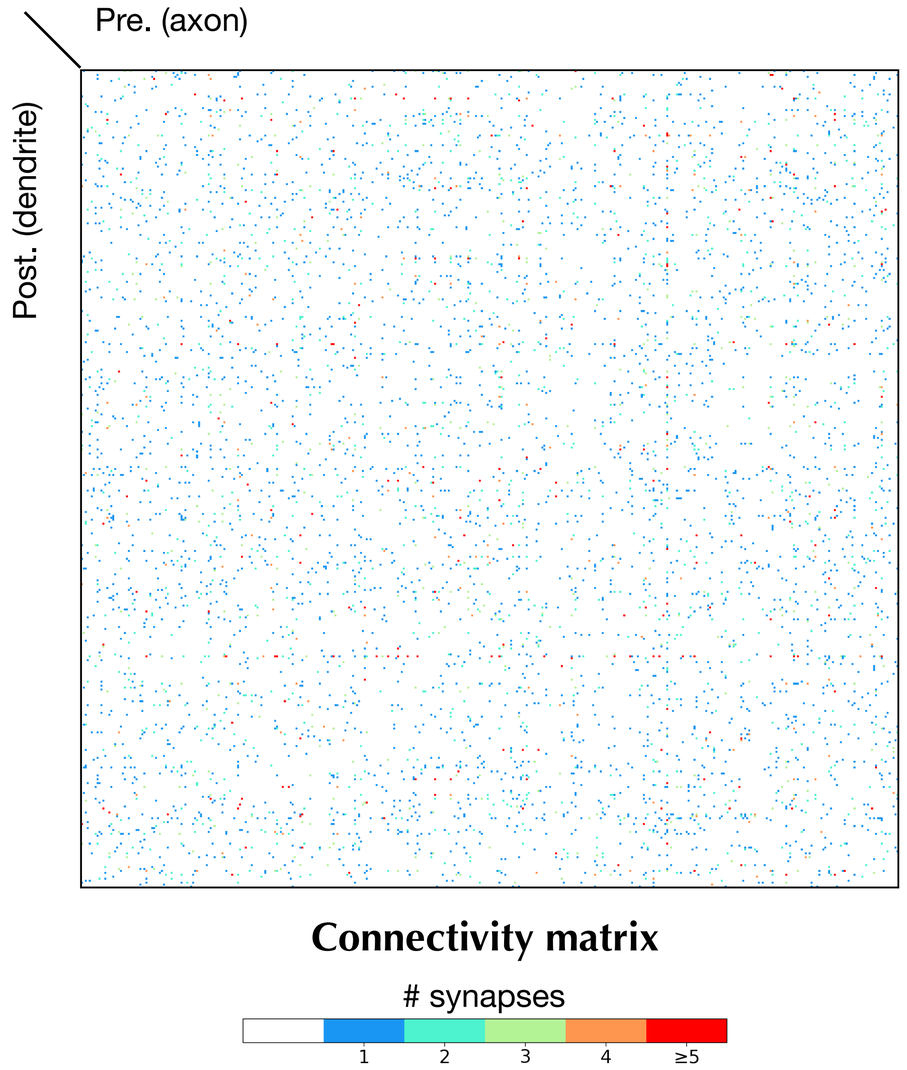
$$v_i = \sum_{j=1}^{N} W_{ij} v_j + f_i$$

presynaptic firing rate

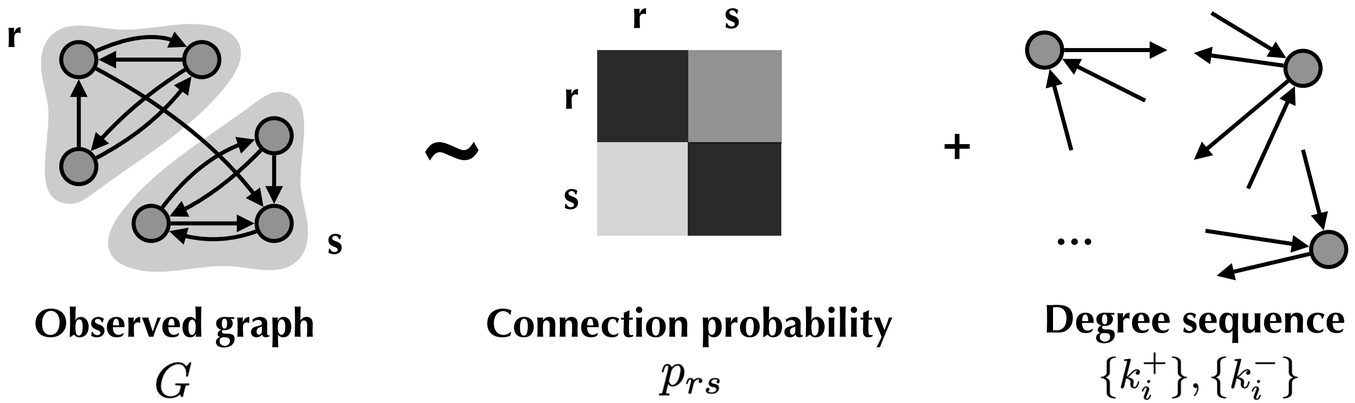
Theory: **synaptic connections *W* must be "recurrent"** to maintain the persistent activity.

Are there recurrent motifs in the circuit?

Raw connectivity matrix of a “center” subgraph of 419 less truncated neurons



Degree-corrected stochastic block modeling (DC-SBM)



Observed graph G

Connection probability p_{rs}

Degree sequence $\{k_i^+\}, \{k_i^-\}$

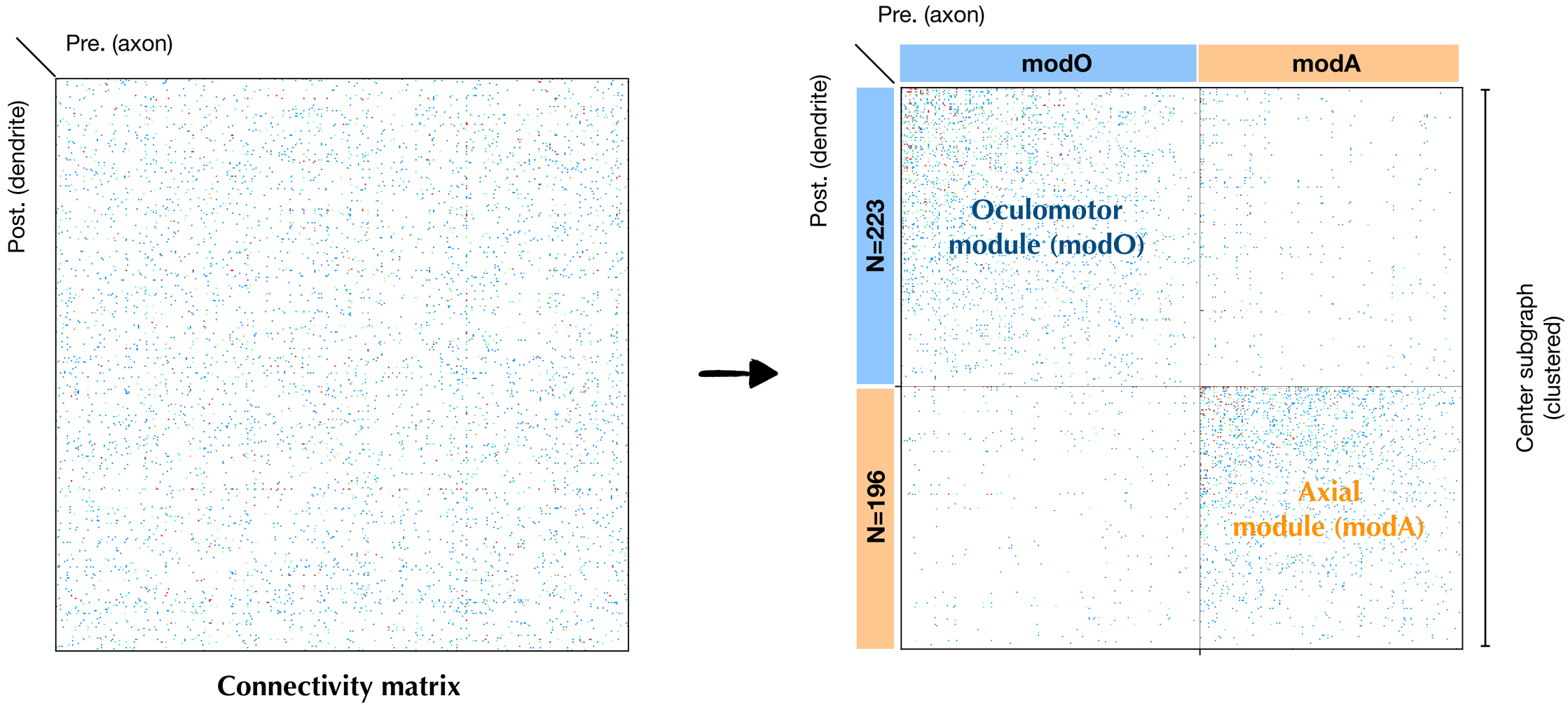
- Connection probability between two neuron is determined by the their block membership, and degrees.

Maximizing likelihood = Minimizing microcanonical entropy (Bianconi 2009):

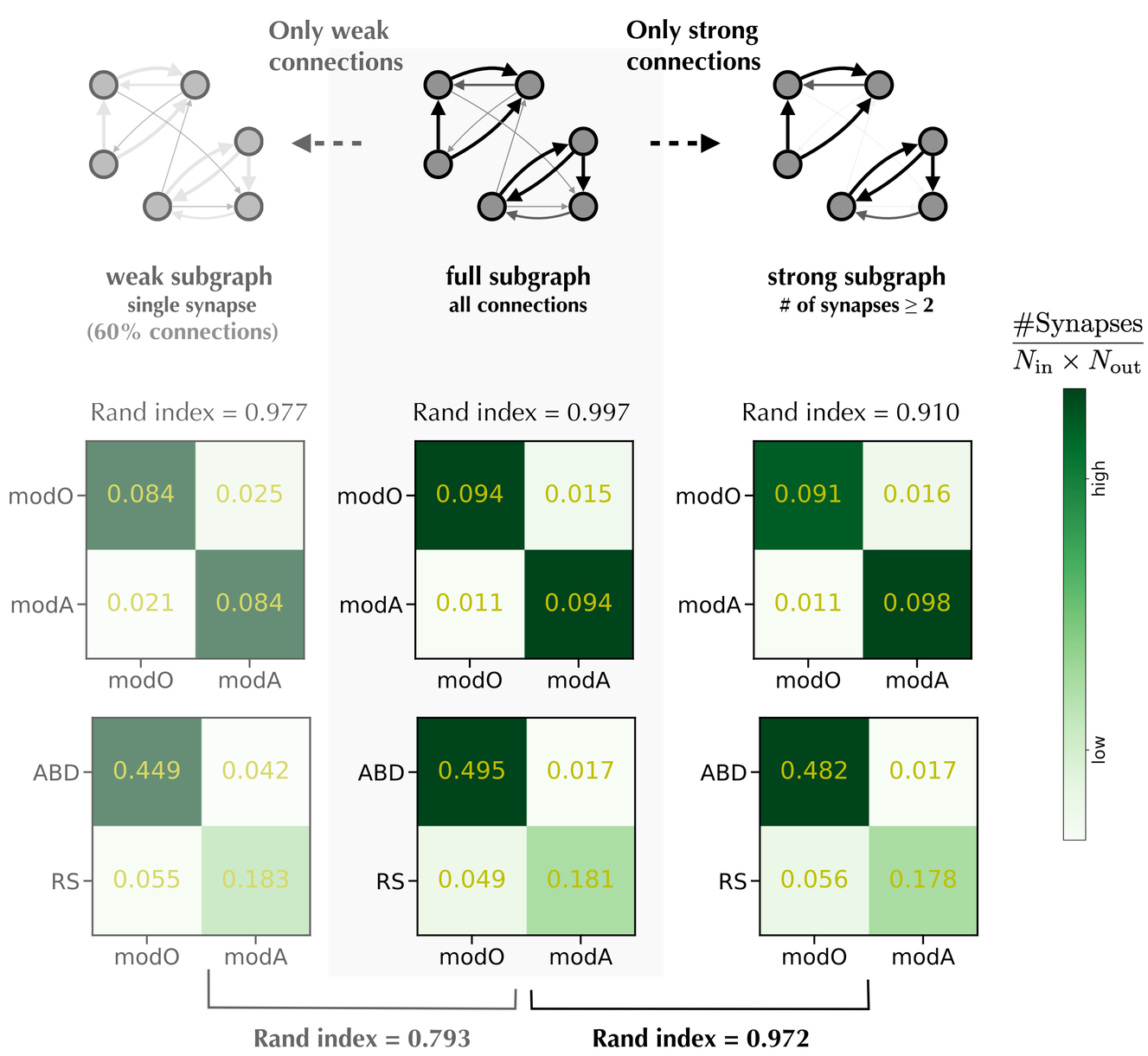
$$S \simeq -M - \sum_i \ln(k_i^+!) - \sum_i \ln(k_i^-!) - \sum_{rs} e_{rs} \ln\left(\frac{e_{rs}}{\sum_s e_{rs} \sum_r e_{rs}}\right)$$

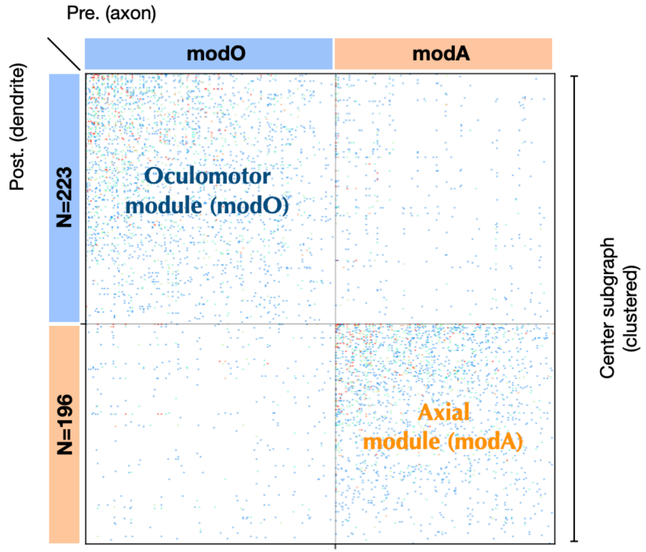
- M: total number of edges
- k_i^+: in-degree of neuron *i*
- k_i^-: out-degree of neuron *i*
- e_{rs}: number of edges between block *r* and block *s*

DC-SBM reveals a “modular” structure of the center subgraph



The modular structure is robust in both weak and strong subgraphs





- The neurons in oculomotor module (modO) send more synapses to abducens neurons (ABD), which controls eye movement.
- The neurons in axial module (modA) send more synapses to large RS neurons, which send outputs to motor neurons that controls body movement.

Reciprocally connected cells are overrepresented in the zebrafish modO

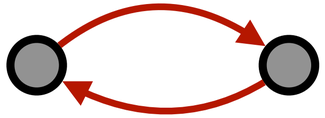
22,139 pairs

Not connected.

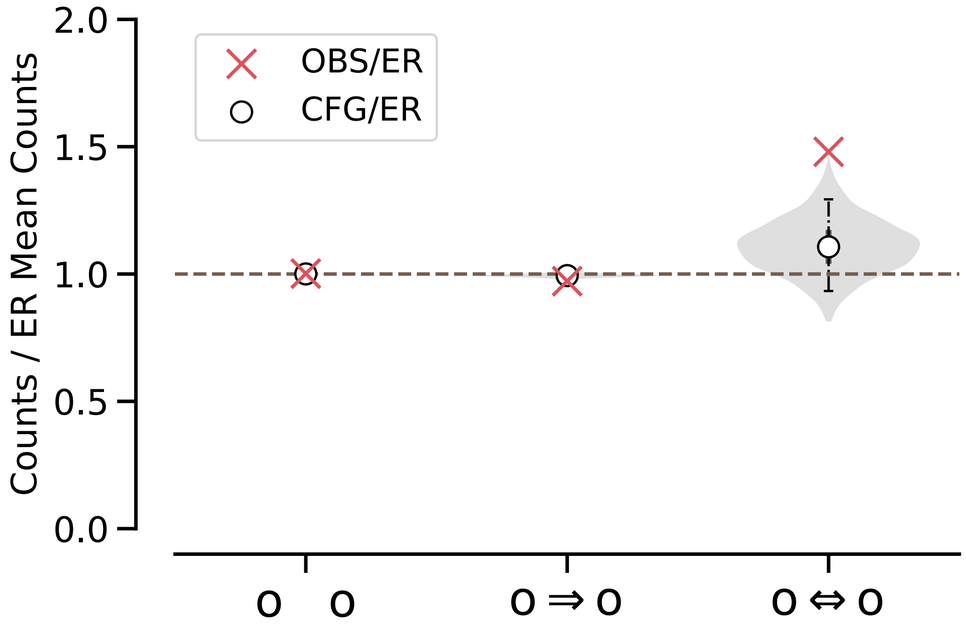
2,503 pairs

Uni-directionally connected.

111 pairs

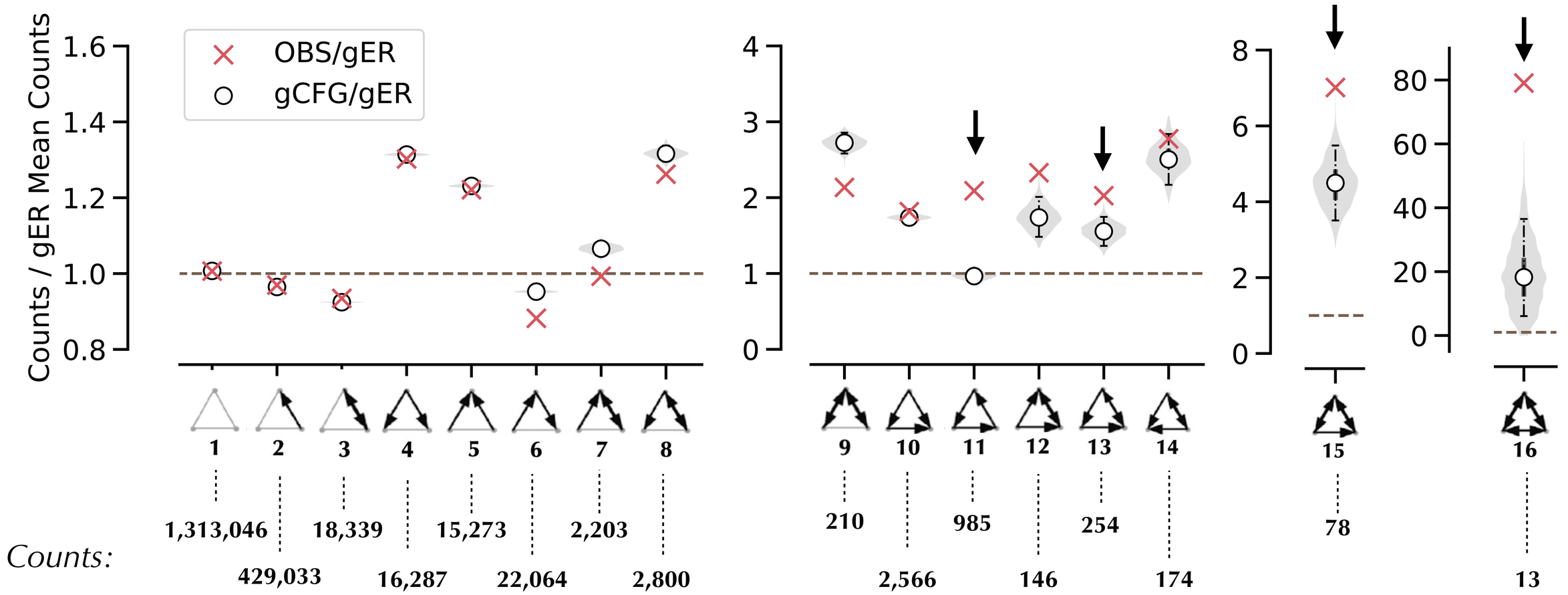


Bi-directionally connected.

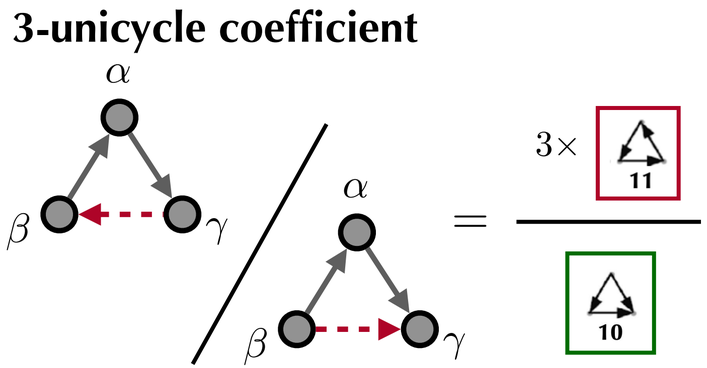


The overrepresentation of bi-directional connections is significant with respect to both ER and CFG.

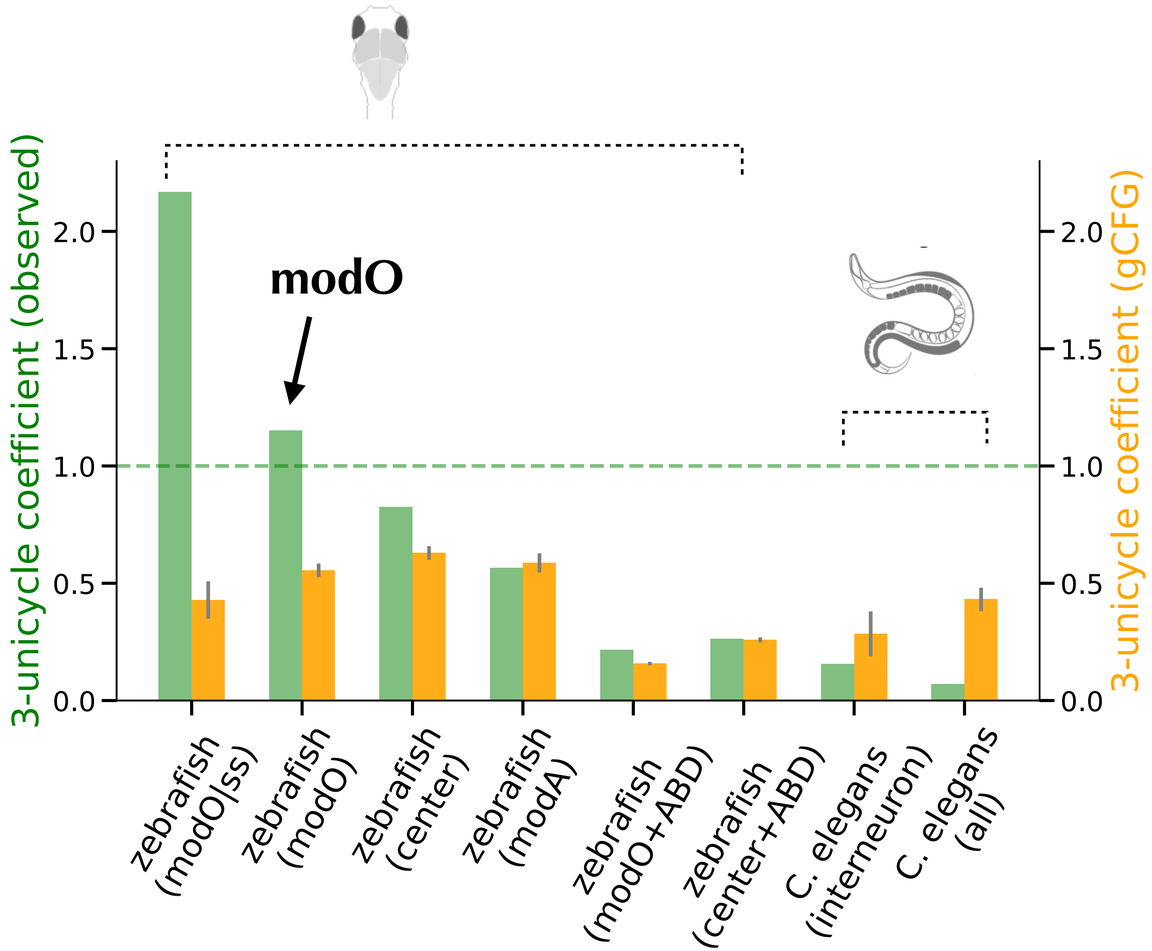
3-cell motifs in the modO deviate significantly from generalized CFG



Zebrafish modO is more recurrent than modA, and *C. elegans*

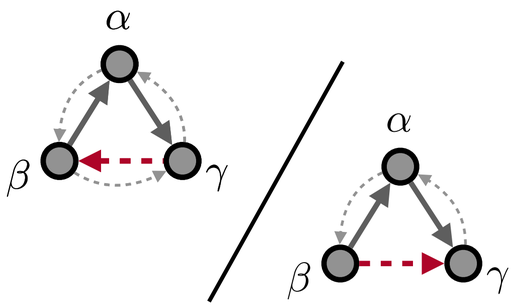


$$U_3 = \frac{\mathbf{Pr}[\gamma \overrightarrow{\not\leftarrow} \beta | \beta \overrightarrow{\not\leftarrow} \alpha \wedge \alpha \overrightarrow{\not\leftarrow} \gamma]}{\mathbf{Pr}[\beta \overrightarrow{\not\leftarrow} \gamma | \beta \overrightarrow{\not\leftarrow} \alpha \wedge \alpha \overrightarrow{\not\leftarrow} \gamma]}$$



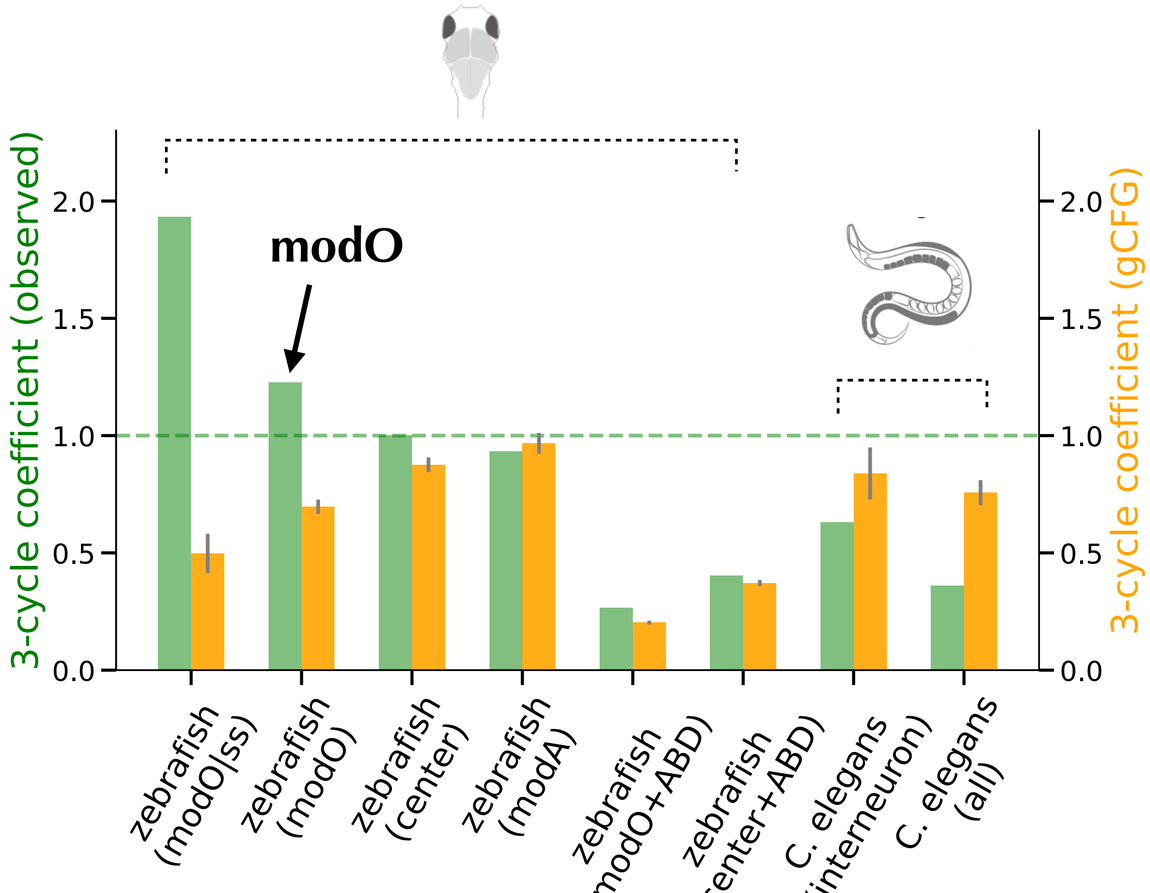
Zebrafish modO is more recurrent than modA, and *C. elegans*

3-cycle coefficient

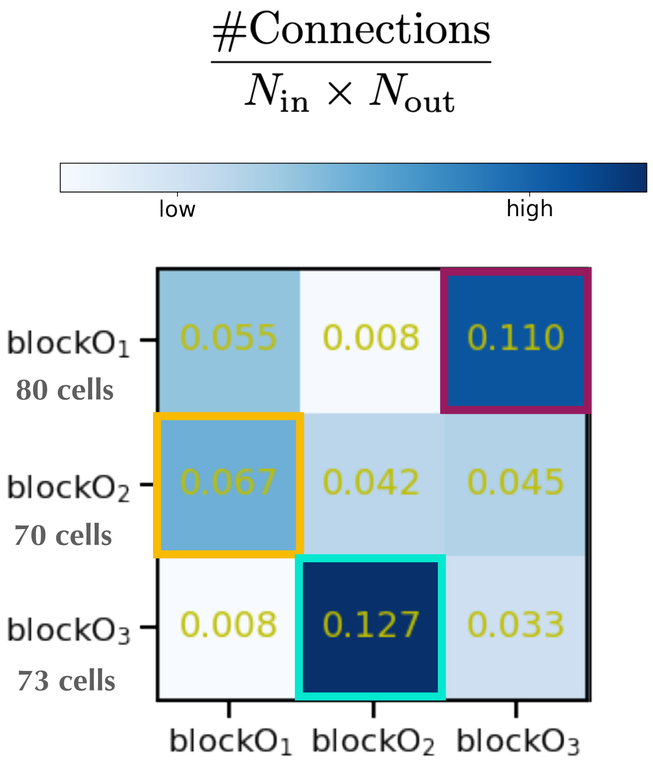


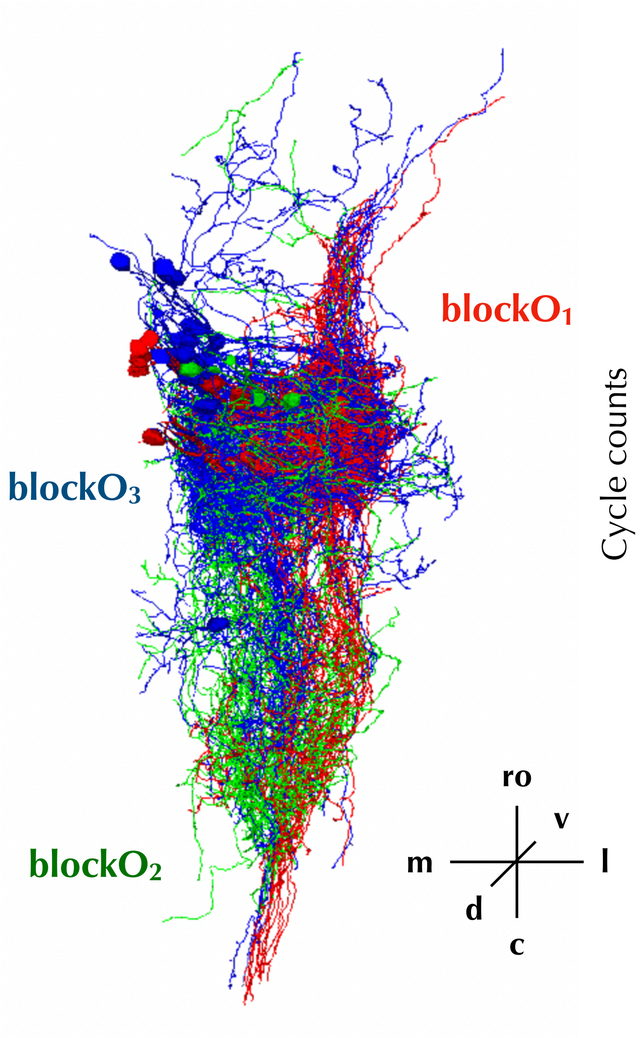
$$C_3 = \frac{\mathbf{Pr}[\gamma \to \beta | \beta \to \alpha \wedge \alpha \to \gamma]}{\mathbf{Pr}[\beta \overrightarrow{\not\leftarrow} \gamma | \beta \to \alpha \wedge \alpha \to \gamma]}$$

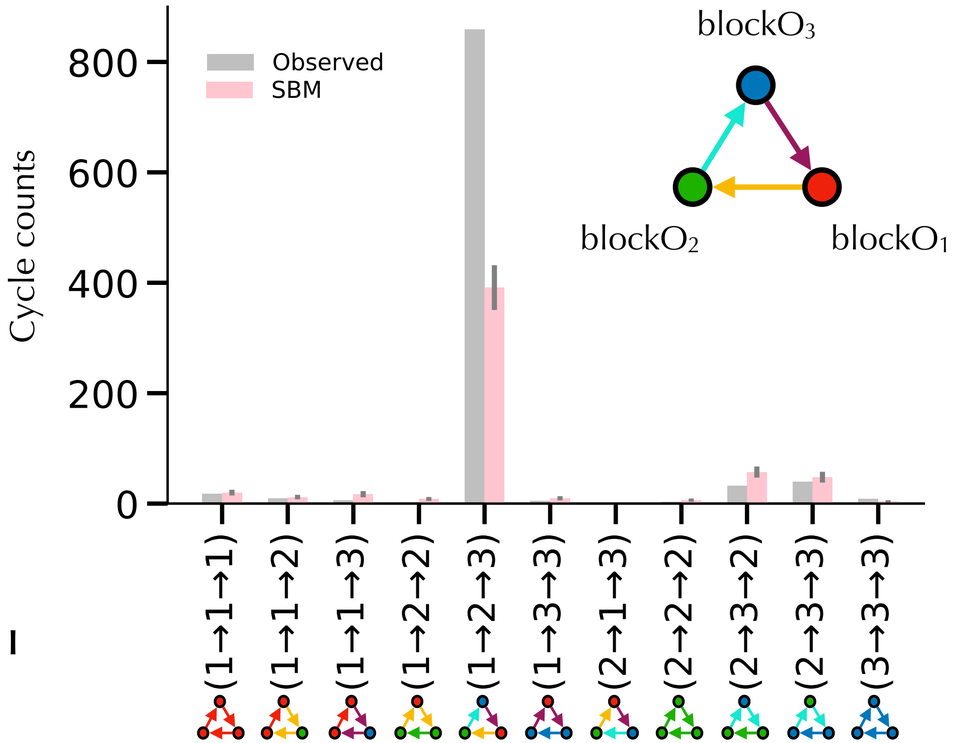
$$= \frac{3\times \boxed{11} \; \boxed{13} \; \boxed{15} \; +6\times \boxed{16}}{\boxed{10} \; +2\times \boxed{12} \; \boxed{14} \; + \; \boxed{15}}$$



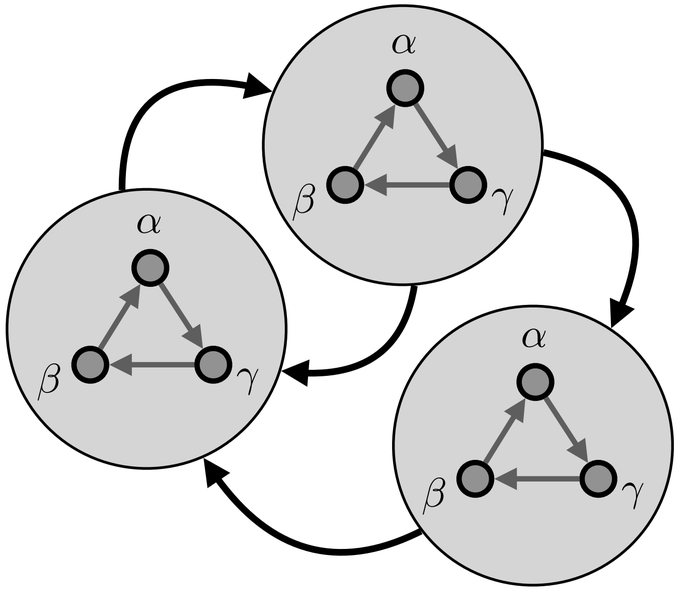
DC-SBM reveals a global cyclic structure of modO, aligns with cellular cycles







Potential function of the precise cellular 3-cycles?



Each unit has a longer effective cellular time constant

synaptic weights

$$v_i = \sum_{j=1}^{N} W_{ij} v_j + f_i$$

Embedding 3-cycle units in the larger network lowers the need for tight tuning globally. (Koulakov *et al*. 2002)

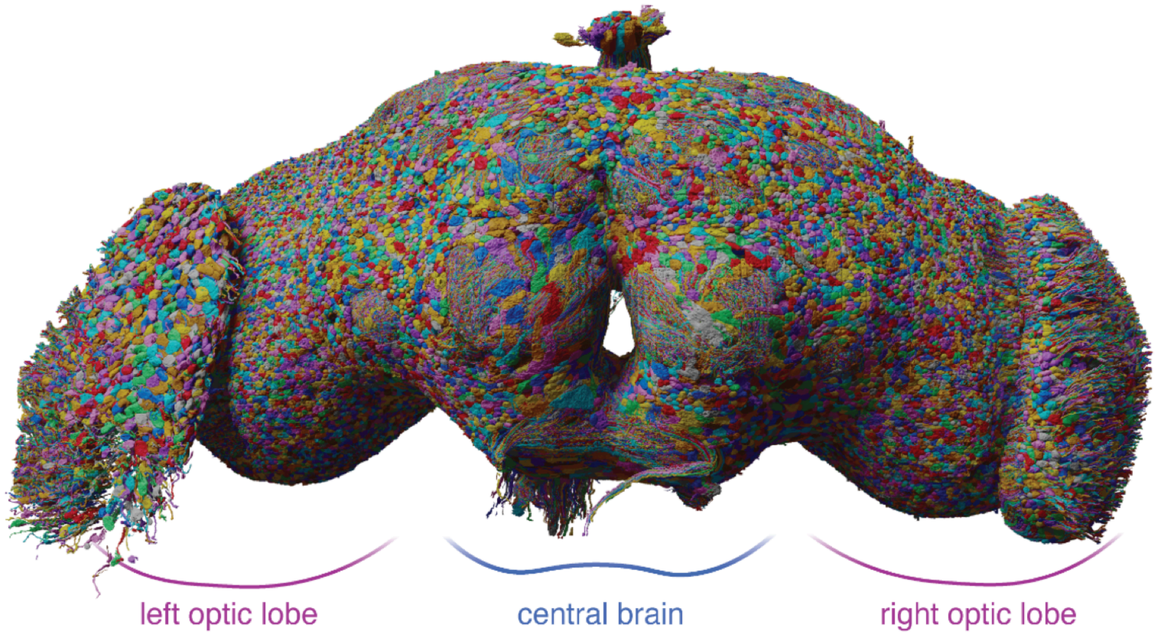
Cyclic structure with cellular precision in a zebrafish oculomotor module

For the first time, for any neuronal wiring diagram reconstructed by EM, we found three-cycles are highly overrepresented.

The cyclic structure could be relevant for recurrent network models of temporal integration by the oculomotor system.

Question 3:
Are there any wiring patterns that scale?

Network statistics of the whole-brain connectome of *Drosophila*



127K+ neurons

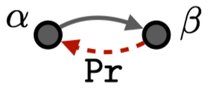
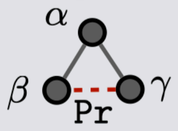
2M+ connections (w/ 5+ syn.)

50M+ chemical synapses

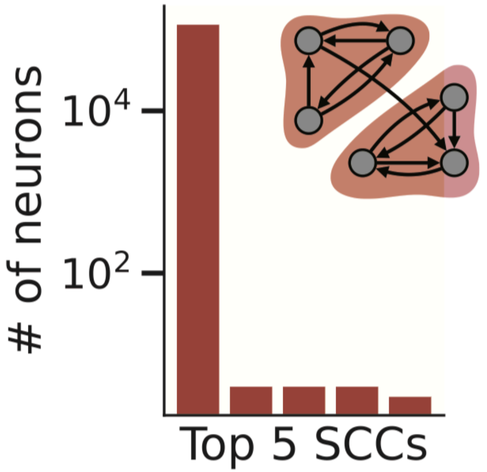
> 81% cells are typed or labelled

"The first neuronal wiring diagram of a whole adult brain"

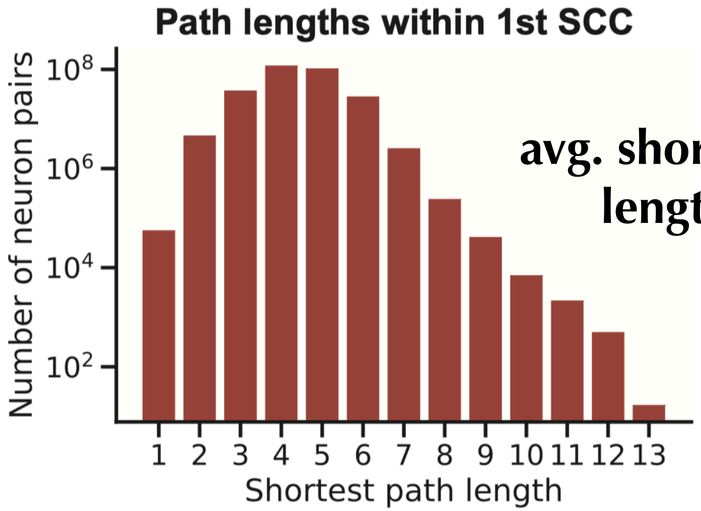
Connection probabilities compared across animals

Neuronal wiring diagrams	**Fruit fly** *Drosophila melanogaster* (Dorkenwald *et al.*, 2023) 	**Nematode** *Hermaphrodite C. elegans* (Cook *et al.*, 2019) 	**Nematode** *Male C. elegans* (Cook *et al.*, 2019) 	**Larval zebrafish (sub-vol.)** *Danio rerio (hindbrain)* (Yang *et al.*, 2023) 	**Mouse (sub-vol.)** *Mus musculus (V1 L2/3)* (Turner *et al.*, 2022) 
Network size 	**127,978 neurons** **2,613,129 connections**	**302 neurons** **3,242 connections**	**364 neurons** **3,467 connections**	**419 neurons** **5,605 connections**	**111 neurons** **659 connections**
Avg. connection strength 	**12.61 synapses** 5 ~ 2358	**3.15 synapses** 1 ~ 36	**3.59 synapses** 1 ~ 63	**1.69 synapses** 1 ~ 21	**1.14 synapses** 1 ~ 5
Connection probability 	**0.000160** x1	**0.0356** x222 denser than fly	**0.0262** x164 denser than fly	**0.0320** x200 denser than fly	**0.0540** x360 denser than fly
Connection reciprocity 	**0.138** x858 than ER x43.8 than CFG	**0.372** x10.4 than ER x5.03 than CFG	**0.386** x14.7 than ER x6.02 than CFG	**0.113** x3.53 than ER x2.64 than CFG	**0.088** x1.63 than ER x1.33 than CFG
Clustering coefficient 	**0.0463** x144 than ER x7.57 than CFG	**0.284** x4.06 than ER x1.86 than CFG	**0.331** x6.39 than ER x2.40 than CFG	**0.182** x2.89 than ER x1.90 than CFG	**0.159** x1.51 than ER x1.06 than CFG

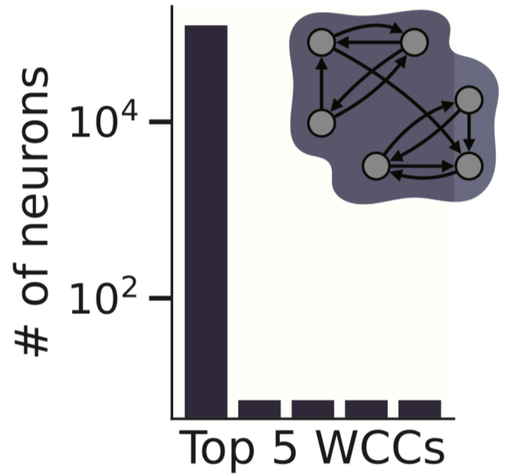
Fly brain is sparse, but highly interconnected



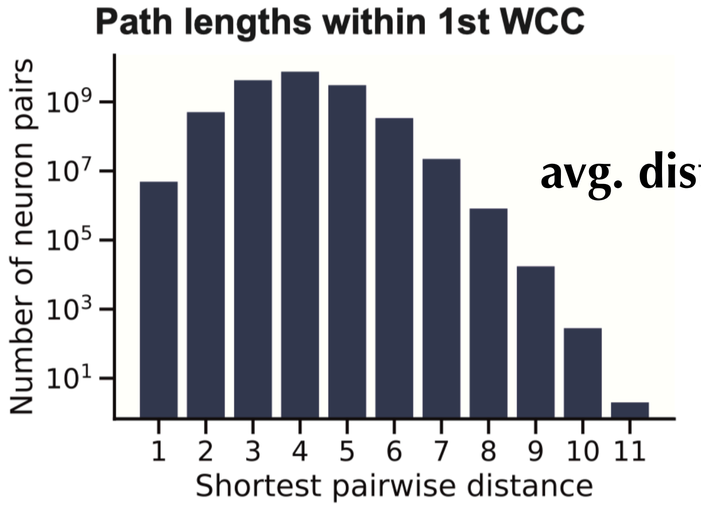
93.3% of neurons are in one giant strongly connected component (SCC)



avg. shortest directed path length is 4.42 hops

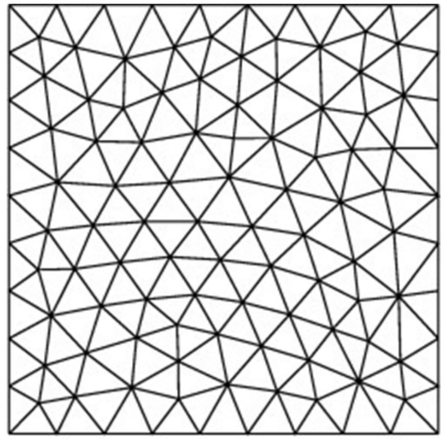


98.8% of neurons are in one giant weakly connected component (SCC)

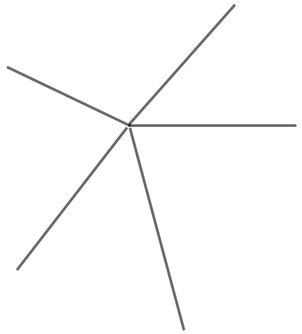


avg. distance is 3.91 hops

Fly brain is a “small world”



highly clustered,
long path length



short path length,
not clustered

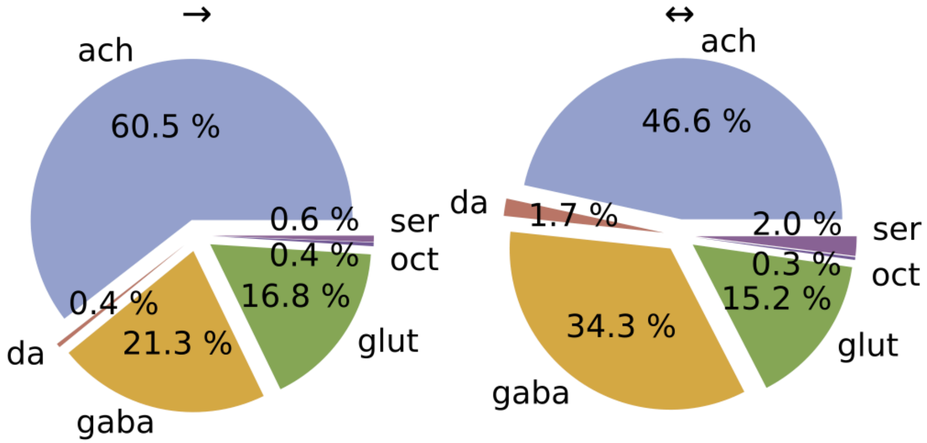
$$S^{\Delta} = \frac{C^{\Delta}_{\text{obs}} / C^{\Delta}_{\text{rand}}}{\ell_{\text{obs}} / \ell_{\text{rand}}} = 141$$

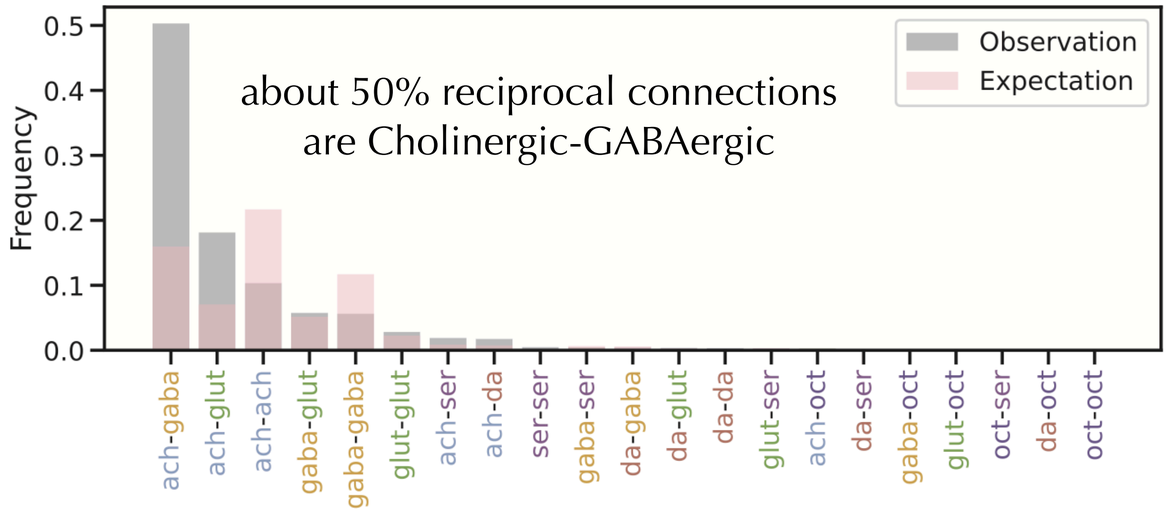
highly effective global communication as the internet!

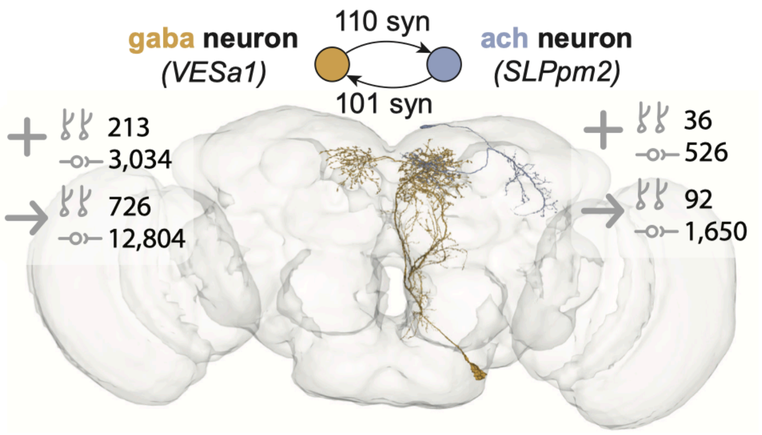
class	#	network	n	m	$\langle k \rangle$	ξ	L	C^{Δ}	S^{Δ}
Social	1	Dolphins†	62	159	5.13	0.084	3.36	0.31	2.8
	2	film actors	449913	25516482	113.43	2.5×10^{-4}	3.48	0.2	627
	3	company directors	7673	55392	14.44	0.002	4.6	0.59	228
	4	math coauthorship	253339	496489	3.92	1.6×10^{-5}	7.57	0.15	11666
	5	physics coauthorship	52909	245300	9.27	1.8×10^{-4}	6.19	0.45	2026
	6	biology coauthorship	1520251	11803064	15.53	1×10^{-5}	4.92	0.088	9089
...									
Technological	18	Internet	10697	31992	5.98	5.6×10^{-4}	3.31	0.035	98.09
	19	power grid	4941	6594	2.67	5.4×10^{-4}	18.99	0.1	84.45
	20	train routes	587	19603	66.79	0.114	2.16	-	-
	21	software packages	1439	1723	1.2	0.0017	2.42	0.07	1403
	22	software classes	1377	2213	1.61	0.0023	1.51	0.033	285.26
	23	electronic circuits	24097	53248	4.42	1.8×10^{-4}	11.05	0.01	33.5
	24	peer-to-peer network	880	1296	2.95	0.0034	4.28	0.012	5.26
Biological	25	metabolic network	765	3686	9.65	0.0126	2.56	0.09	8.18
	26	yeast protein interactions	2115	2240	0.001	2.12	6.8	0.072	107.85
	27	marine food web	135	598	4.43	0.0661	2.05	0.16	7.84
	28	freshwater food web	92	997	10.84	0.2382	1.9	0.2	1.7
	29	C.Elegans†	277	1918	13.85	0.05	2.64	0.2	3.21
	30	Macaque cortex†	95	1522	32.04	0.34	1.78	0.7	1.53
...									

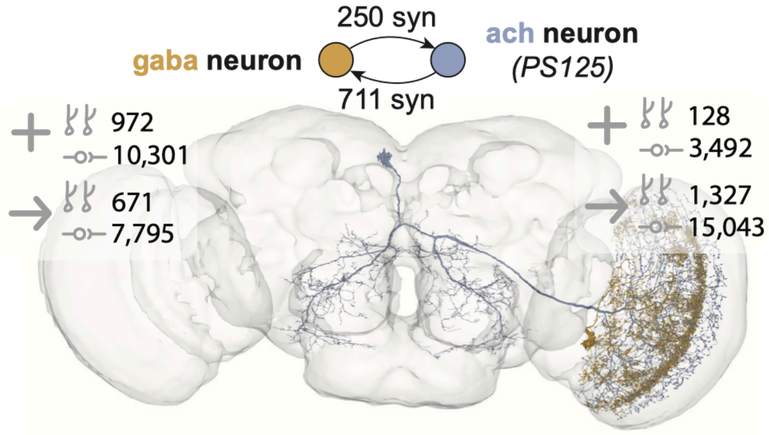
Table modified from *Humphries and Gurney, 2008*

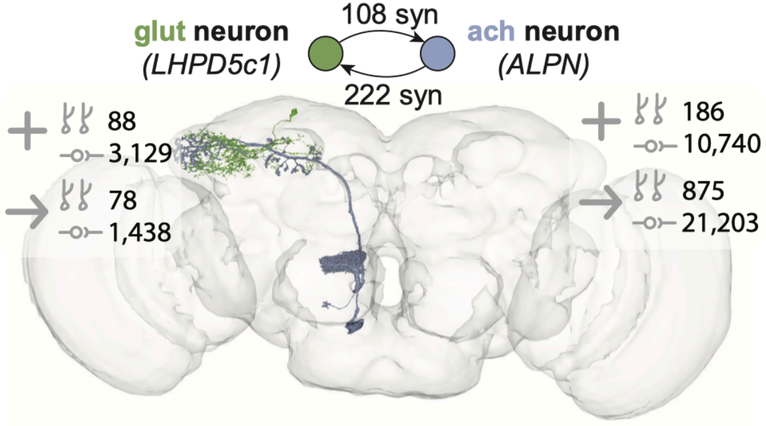
2-cell motifs in the fly brain: most reciprocal connections are excitatory-inhibitory



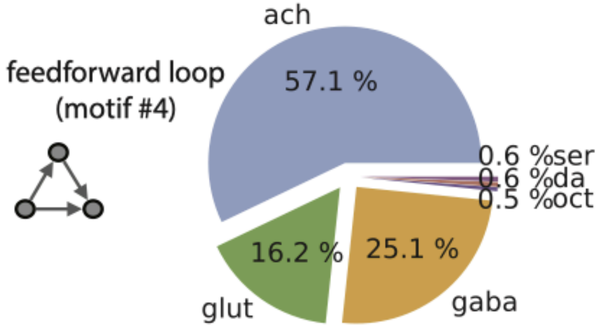


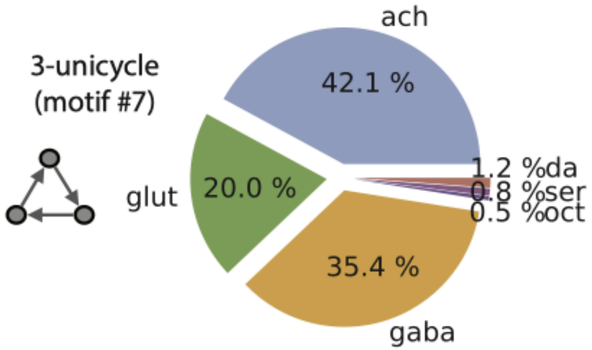


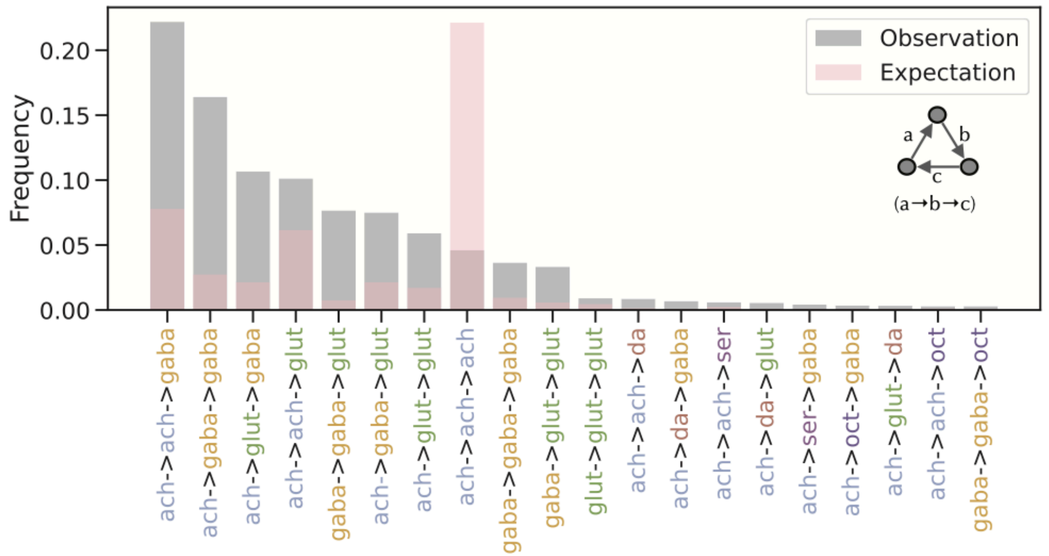


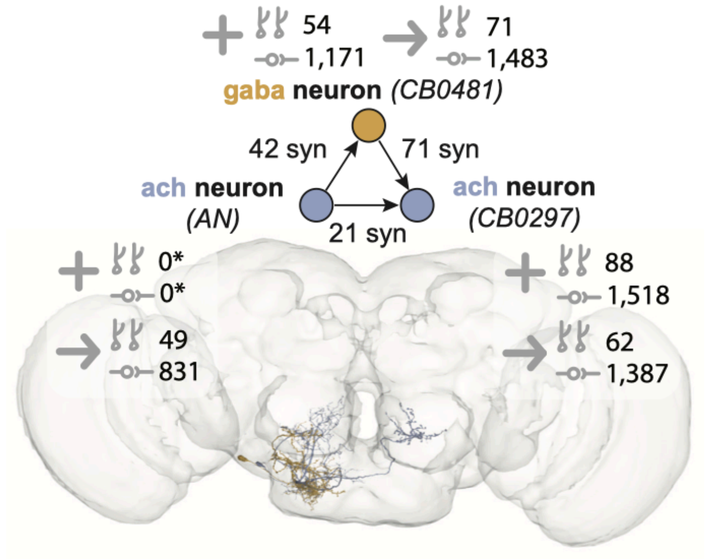


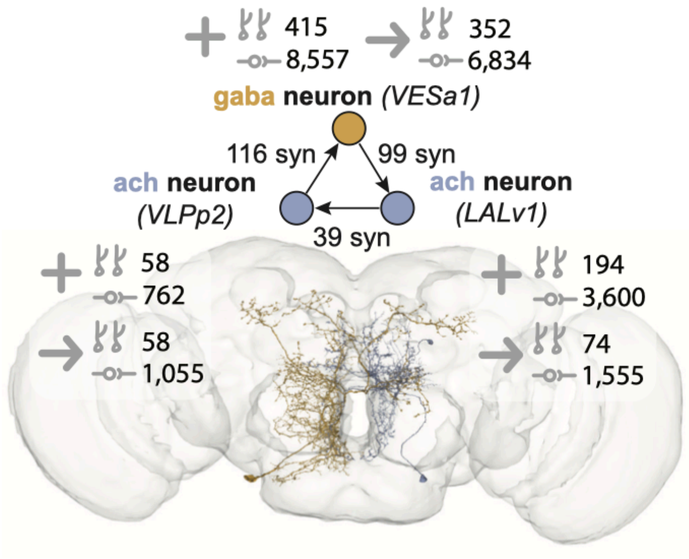
3-cell motifs in the fly brain: most feedback cycles involve inhibitory connections

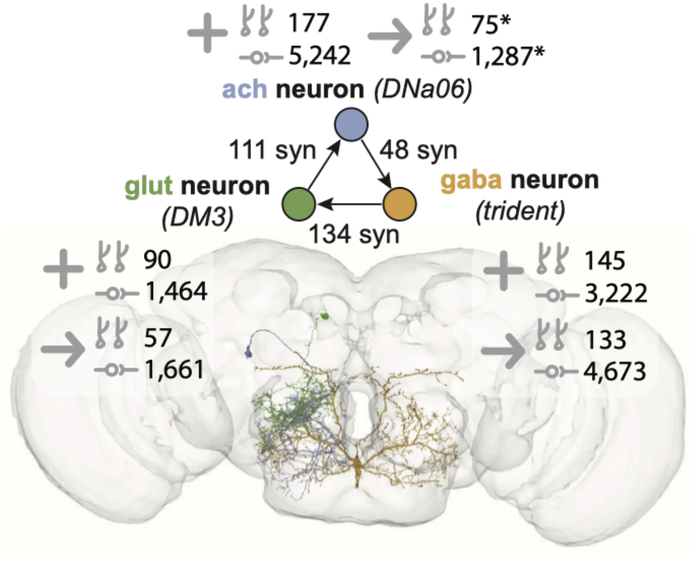




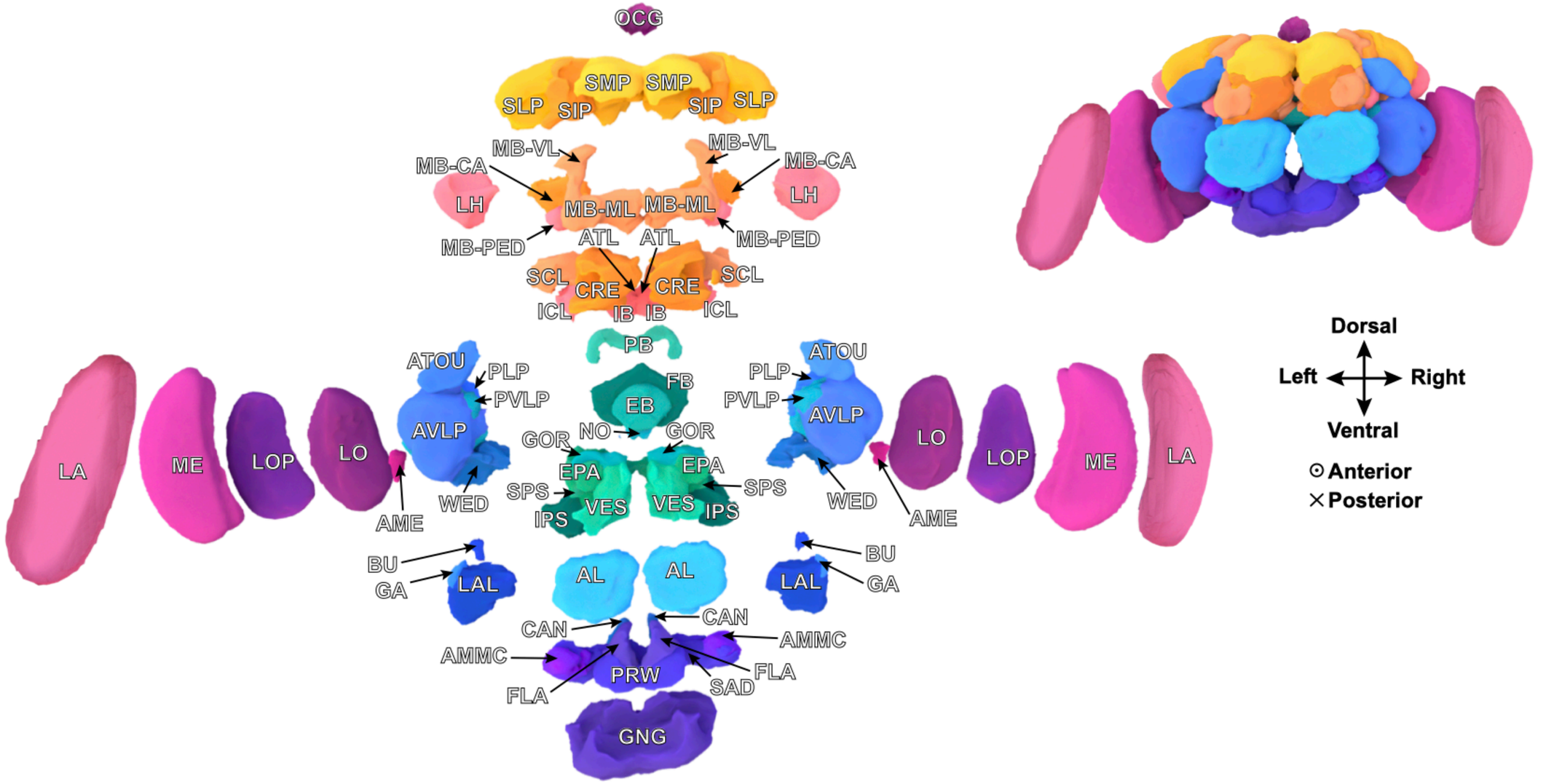




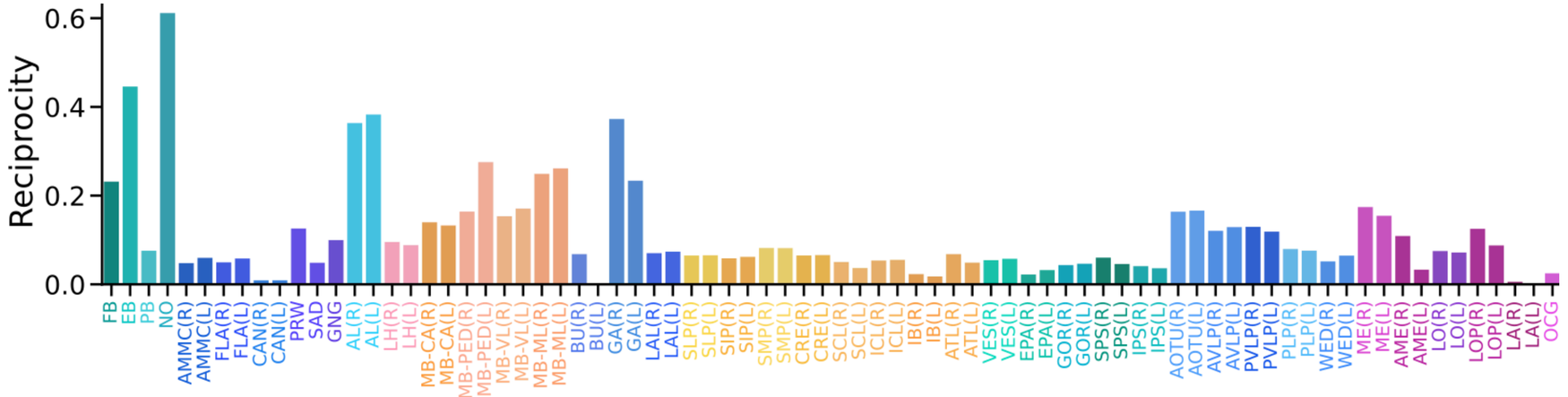


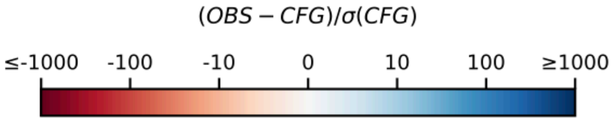


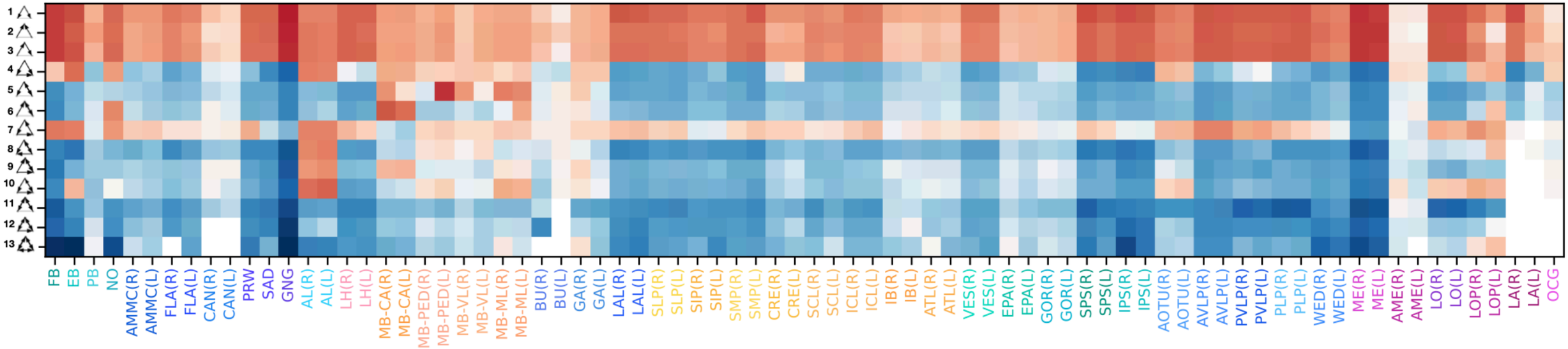
Connectivity difference across 78 anatomically defined brain regions



Connectivity difference across 78 anatomically defined brain regions

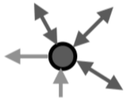
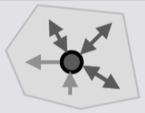
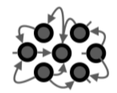
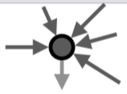
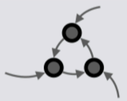
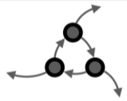




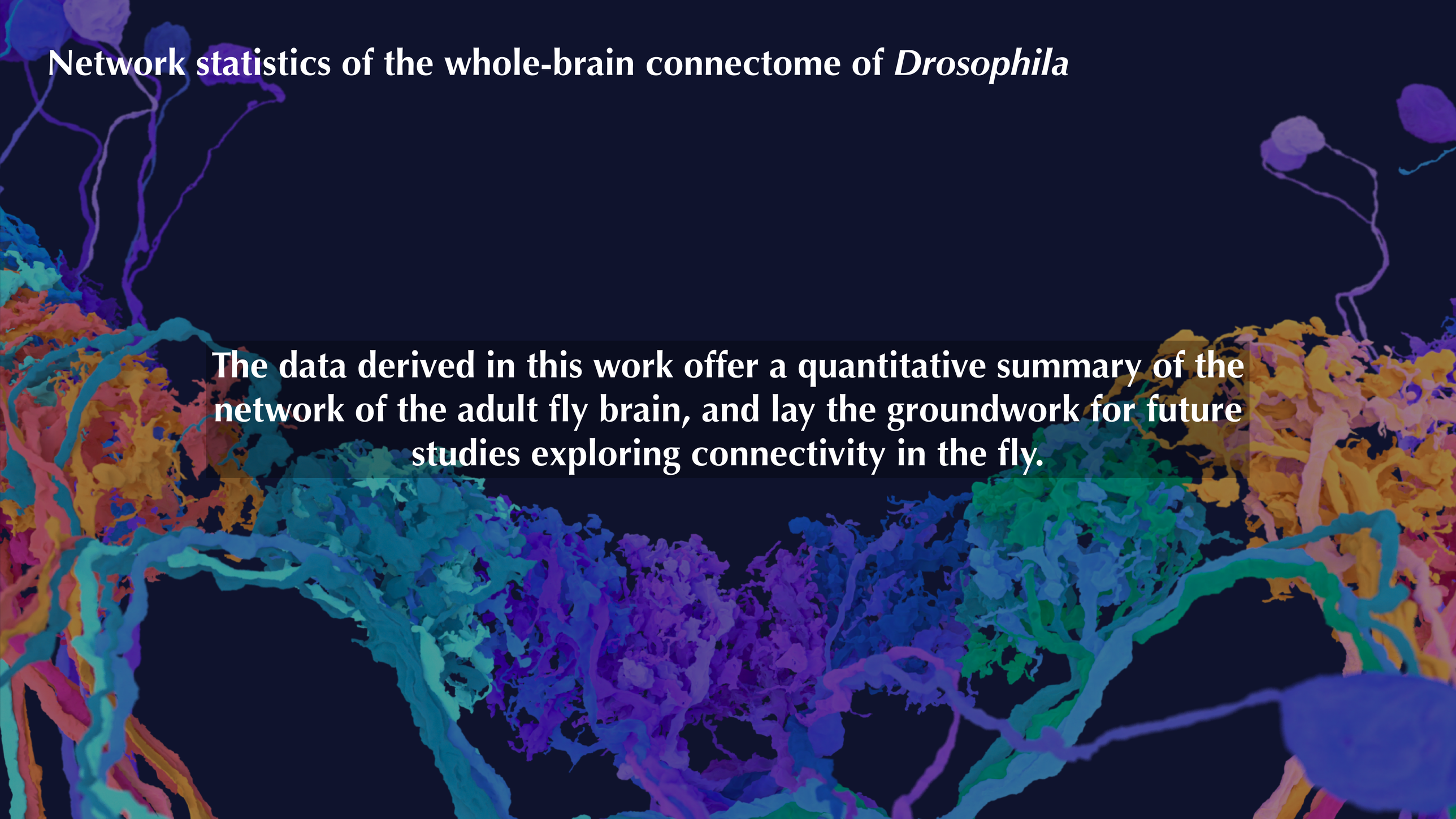


Available network statistics on codex (https://codex.flywire.ai/)

	Computed network statistics	
Connected components	strongly connected components	Figure 1d
	weak connected components	Figure 1e
Path length analysis	directed shortest path lengths	Figure 1d
	undirected shortest path lengths	Figure 1e
Percolation analysis	vertex percolation	Figures 1f, g; S1f
	edge percolation	Figure S1a
Rich-club analysis	total-degree rich club	Figure 1h
	in-degree rich club	Figure S1g
	out-degree rich club	Figure S1g
Small-word analysis	clustering coefficient	Table 2
	small-wordness	Equation 1
2-neuron motifs	reciprocity	Table 2; Figures 5c; S5c
	connection strength	Figures 2a, d; 5f, S6a
	neurotransmitter types	Figures 2c, e, f; 5d, e; S5
3-neuron motifs	motif frequencies	Figures 3a; 6a, d; S7a
	motif strength	Figures 3b; 6c; S7b
	neurotransmitter types	Figures 3c, d, e
Large-scale connectivity	degree distribution	Figure 1c
	cell categories	Figure 4
Spectral analysis	forward random walk	Figure S1d
	reversed random walk	Figure S1e
Neuropil subgraphs	internal/external connection weights	Figure S4
	2-neuron motifs	Figures 5, S5, S6
	3-neuron motifs	Figures 6, S7

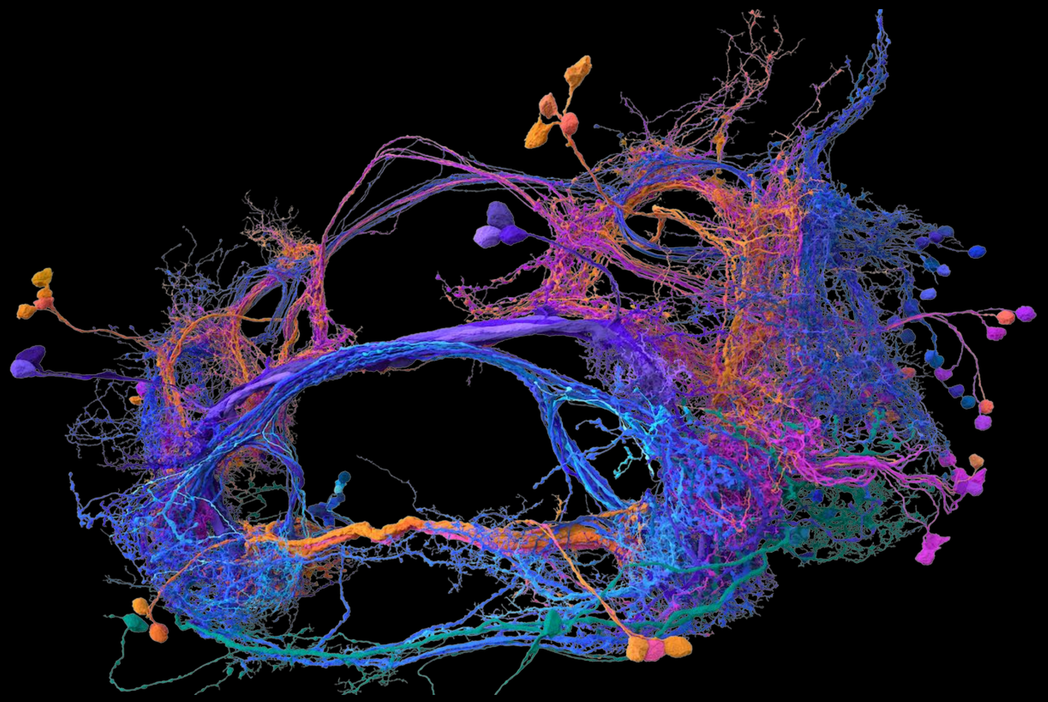
	Neuron lists available on Codex		# of neurons
2-neuron motifs	reciprocal connection participants		77,607
3-neuron motifs	feedforward loop participants		113,978
	3-unicycle participants		66,835
N-neuron motifs	highly reciprocal neurons		2,183
	neuropil-specific highly reciprocal neurons (NSRNs)		704
Rich-club analysis	rich-club neurons		40,218
	broadcasters		676
	integrators		638
Spectral analysis	attractors		3,469
	repellers		3,469

Network statistics of the whole-brain connectome of *Drosophila*



The data derived in this work offer a quantitative summary of the network of the adult fly brain, and lay the groundwork for future studies exploring connectivity in the fly.

New frontiers during my PhD: from brain reconstruction to generative AI



3D Reconstruction of Neuronal Circuits

Prompt: "A photograph of neuron-like stardust"

Part II: deep learning research - trends during my PhD

before large models

Network architectures?
CNNs: VGG, ResNet, …
RNNs: LSTM, GRU, …
Attentions, Transformers,…

Objective functions?
autoregression,
contrastive learning,
negative sampling,
+ regularizations…

Transformers

pre-training + fine-tuning

in-context learning

huge models >100B params

massive data

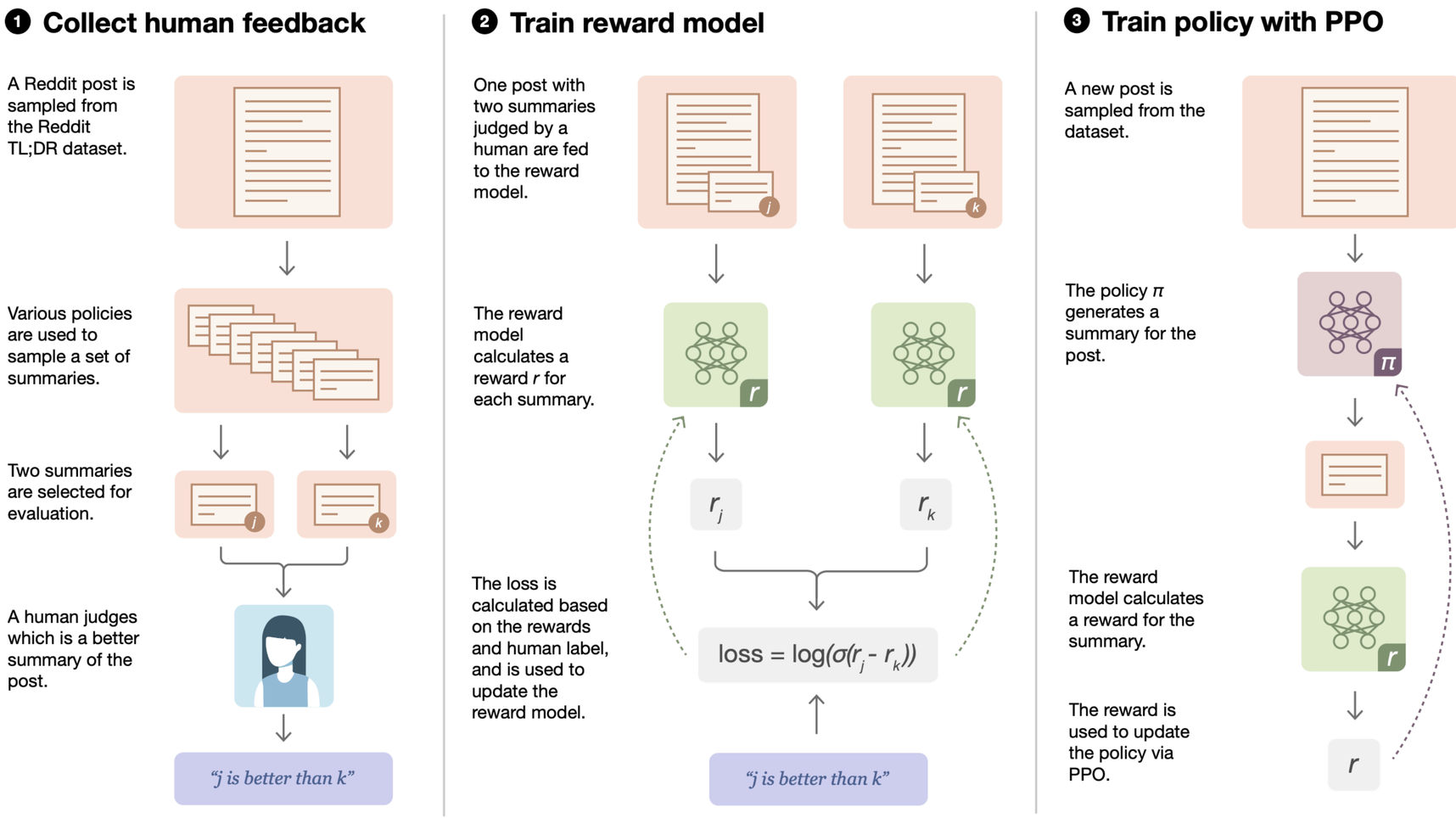
after large models

How to make large language models more efficient?
Multiplexing, LoRA, etc.

How to better use LLMs / language agents?
Prompt engineering, CoT, etc.

How to ensure AI safety?
RL with human feedback, consistency model

Reward in reinforcement learning: a single scalar value

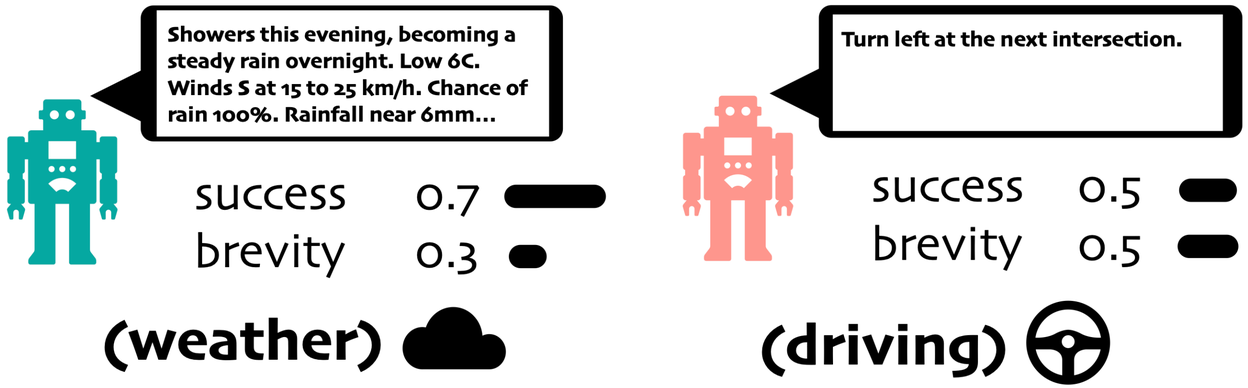


- **In RLHF for ChatGPT:** A ***linear order*** among model outputs is assumed when training the reward model.
- What if two outputs have their own advantages?
  E.g., being concise vs being informative

Figures modified from *Learning to summarize from human feedback*. Stiennon *et al.*, 2021

Algorithm 1:
Reinforcement learning with *multiple objectives*

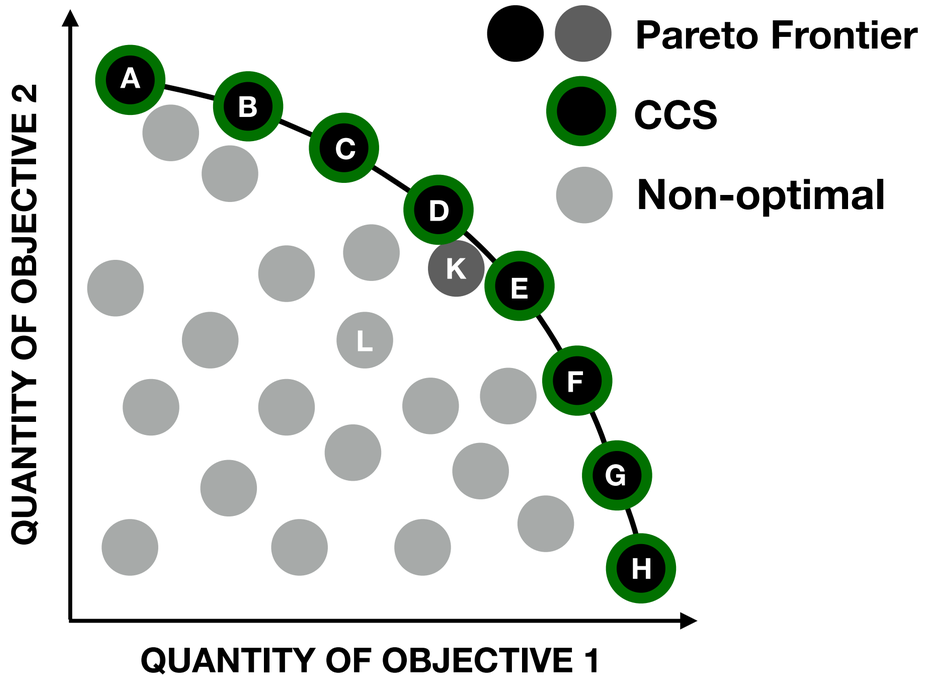
Potential solution: *multi-objective* reinforcement learning (NeurIPS '19)



E.g., in **task-oriented dialogue systems,** users may expect either **briefer** or **more informative** dialogue.

Key Idea:

1. **maintain the entire convex coverage set of Pareto Frontier using a single value network.**
2. **adapt to user's preference with few-shot interactions, without fine-tuning.**



Linear preferences: $f_\omega(r) = \omega^\intercal r$

Potential solution: *multi-objective* reinforcement learning

for *episode* $= 1, \ldots, M$ **do**

Sample a linear preference $\boldsymbol{\omega} \sim \mathcal{D}_\omega$.

for $t = 0, \ldots, N$ **do**

Exploration

Observe state s_t.

Sample an action ϵ-greedily:

$$a_t = \begin{cases} \text{random action in } \mathcal{A}, & \text{w.p. } \epsilon; \\ \max_{a \in \mathcal{A}} \boldsymbol{\omega}^\intercal \boldsymbol{Q}(s_t, a, \boldsymbol{\omega}; \theta), & \text{w.p } 1 - \epsilon. \end{cases}$$

Receive a vectorized reward $\boldsymbol{r}_t$ and observe s_{t+1}.

Store transition $(s_t, a_t, \boldsymbol{r}_t, s_{t+1})$ in $\mathcal{D}_\tau$.

if *update* **then**

Envelope Q-update

Sample N_τ transitions $(s_j, a_j, \boldsymbol{r}_j, s_{j+1}) \sim \mathcal{D}_\tau$.

Sample N_ω preferences $W = \{\boldsymbol{\omega}_i \sim \mathcal{D}_\omega\}$.

Compute $y_{ij} = (\mathcal{T}\boldsymbol{Q})_{ij} =$

$$\begin{cases} \boldsymbol{r}_j, & \text{for terminal } s_{j+1}; \\ \boldsymbol{r}_j + \gamma \arg_Q \max\limits_{\substack{a \in \mathcal{A}, \\ \boldsymbol{\omega}' \in W}} \boldsymbol{\omega}_i^\intercal \boldsymbol{Q}(s_{j+1}, a, \boldsymbol{\omega}'; \theta), & \text{o.w.} \end{cases}$$

for all $1 \le i \le N_\omega$ and $1 \le j \le N_\tau$.

Update Q_θ by descending its stochastic gradient according to equations 1. and 2.:

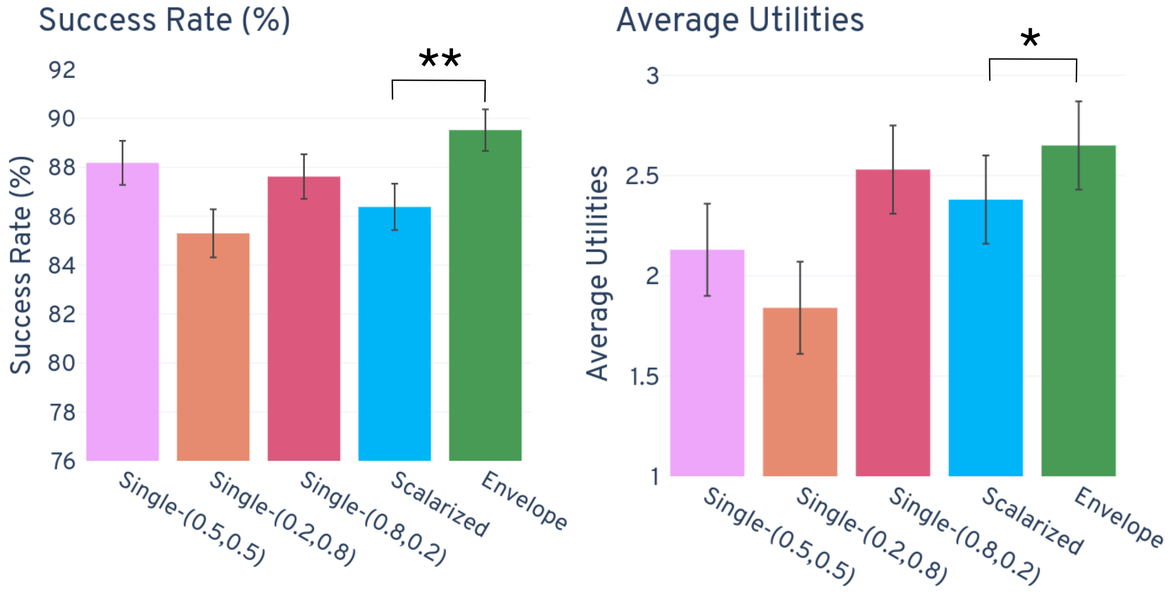
$$\nabla_\theta L(\theta) = (1-\lambda) \cdot \nabla_\theta L^{\mathrm{A}}(\theta) + \lambda \cdot \nabla_\theta L^{\mathrm{B}}(\theta).$$

Increase λ along the path p_λ.

Preference elicitation after learning

$$\arg\max_{\mu_1, \ldots, \mu_m} \mathbb{E}_{\boldsymbol{\omega} \sim \mathcal{D}_\omega^m} \left[\mathbb{E}_{\tau \sim (\mathcal{P}, \Pi_\mathcal{L}(\boldsymbol{\omega}))} \left[\sum_{t=0}^{\infty} \gamma^t r_t(s_t, a_t) \right] \right]$$

Dialog objectives: brevity / success



Algorithm 2:
Theory of mind for dialog generation

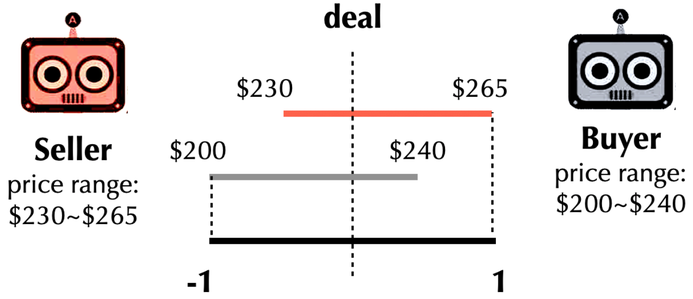
Can AI dialog systems learn and adapt to different user personality type?

Description of an item for sale on Craigslist

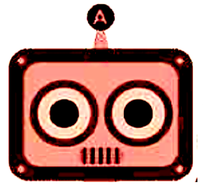
GoPro Hero4 Black + Battery BacPac

- HERO4 Black Camera
- Standard Housing 131',
- Rechargeable Battery,
- Flat Adhesive Mount,
- 3-Way Pivot Arm

Price: $265



reward is zero-sum + no deal penalty (-0.5)

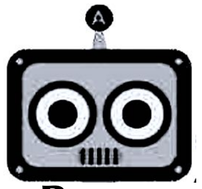


Seller

($230~$265)

Are you interested in this GoPro? Honestly, I barely used it and decided to sell it. I'm selling it for $265.

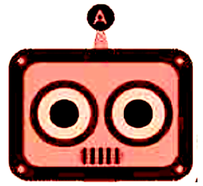
I am definitely interested in the GoPro I'm on somewhat of a budget. Would you be willing to drop the price a tad, maybe $200?



Buyer

($200~$240)

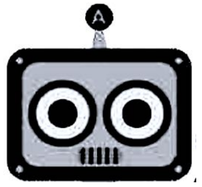
...



Seller

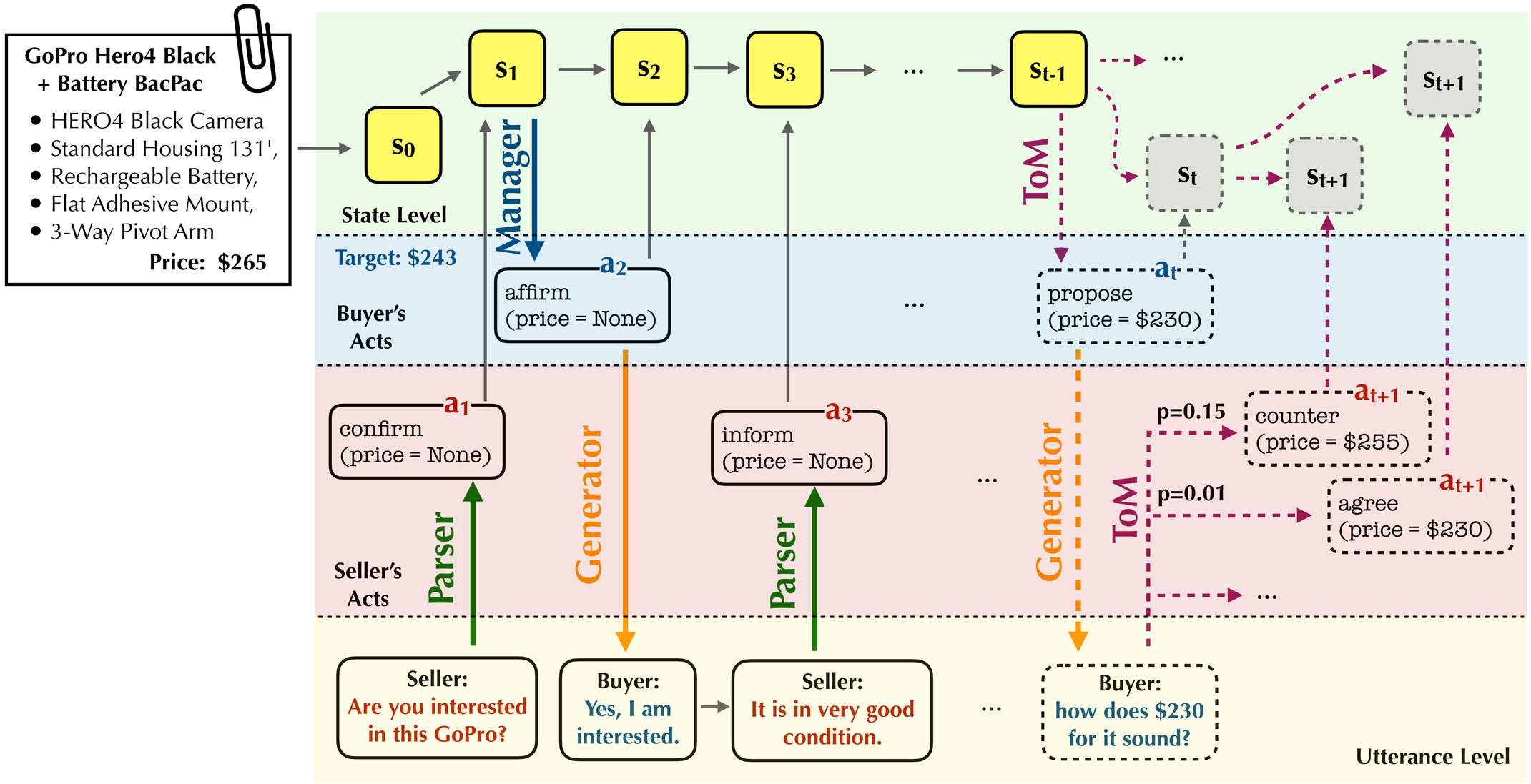
OFFER $235 (Click the Button)

ACCEPT (Click the Button)



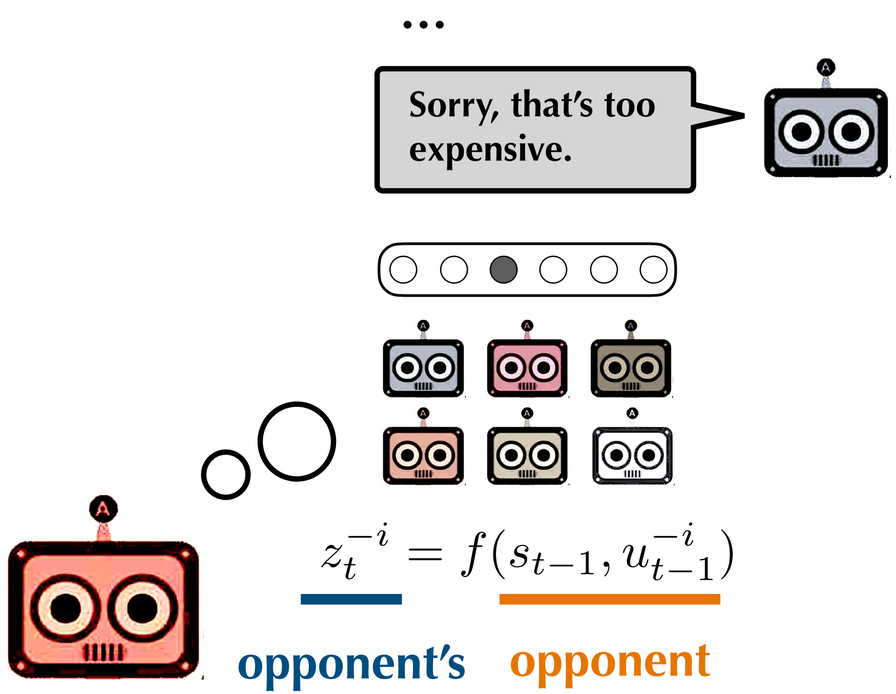
Buyer

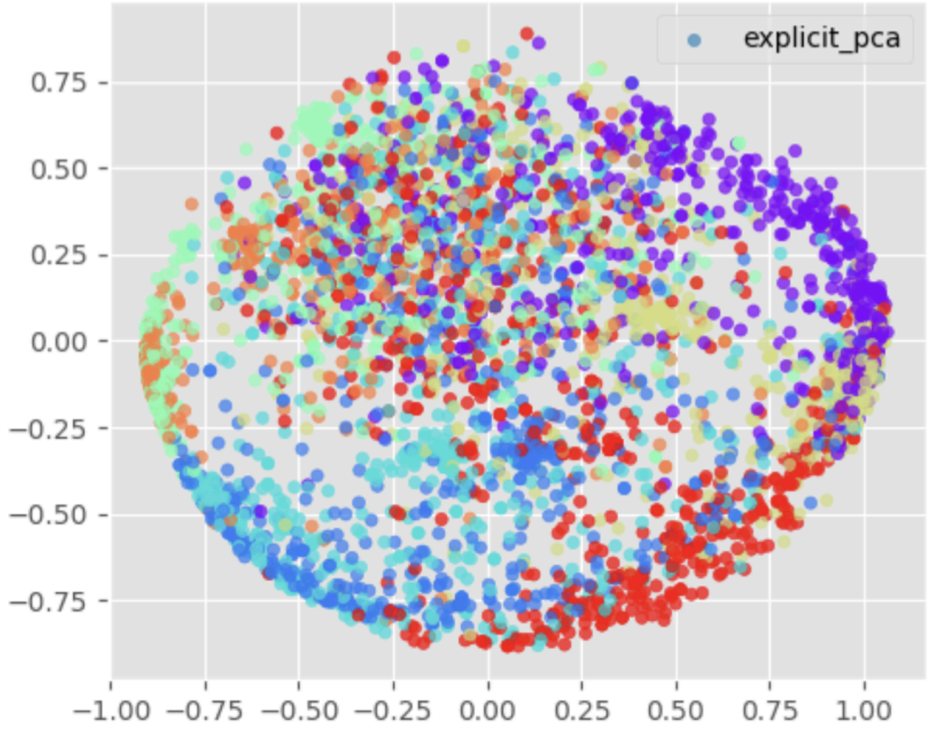
Potential solution: Theory-of-Mind model enables one-step looking ahead (ACL '21)



Our solution: explicitly predict user populations from on utterance and state transition

1st-order ToM model: $T(s_{t+1}|z_{t-1}^{-i}, s_t, u_t^i)$





PCA visualization of latent variables in the explicit (left). Colors indicate different opponent populations.

Improvement of agreement rate against different population

ToM policy:

$$\pi_{\texttt{ToM}}(a_t^i|s_{t-1}, z_{t-1}^{-i}) \propto \exp\left\{\frac{1}{\beta}\sum_{u_t^i}\underbrace{G(u_t^i|s_t, z_{t-1}^{-i})}_{\text{Generator}}\sum_{s_{t+1}}\underbrace{T(s_{t+1}|z_{t-1}^{-i}, s_t, u_t^i)}_{1^{\text{st}}\text{-order ToM}}\underbrace{V(s_{t+1})}_{\text{Value Fn.}}\right\}$$

contains a_t

Method	Cooperative Opponents (id=5)				Competitive Opponents (id=6)				Mixed Population (id=0∼6)				
	Ag ↑	Ut ↑	Fa ↑	Len ↓	Ag ↑	Ut ↑	Fa ↑	Len ↓	Ag ↑	Ut ↑	Fa ↑	Len ↓	Re ↑
SL+rule	0.006	0.006	0.00	**10.51**	0.005	0.005	0.00	**10.64**	0.009	0.008	0.00	**10.59**	−0.48
RL	0.57	0.57	0.00	15.48	0.42	0.18	0.32	16.10	0.47	0.38	0.00	15.79	0.00
ToM (implicit)	0.76	0.72	0.00	13.14	0.34	0.20	0.45	14.26	0.48	0.44	0.00	13.87	0.15
ToM (explicit)	**0.88**	**0.78**	**0.03**	11.34	**0.44**	**0.24**	**0.55**	12.10	**0.56**	**0.47**	**0.10**	11.74	**0.16**

Table 3: Agreement rate (Ag), agent utility (Ut), deal fairness (Fa), dialog length (Len), and reward (Re) for dialog managers (SL, RL, implicit and explicit ToM, with the best $\beta = 0.05$) playing against cooperative, competitive, and mixed populations. Small negative fairness scores are truncated as zeros in the table.

Algorithm 3:
LLMs + role playing / coding

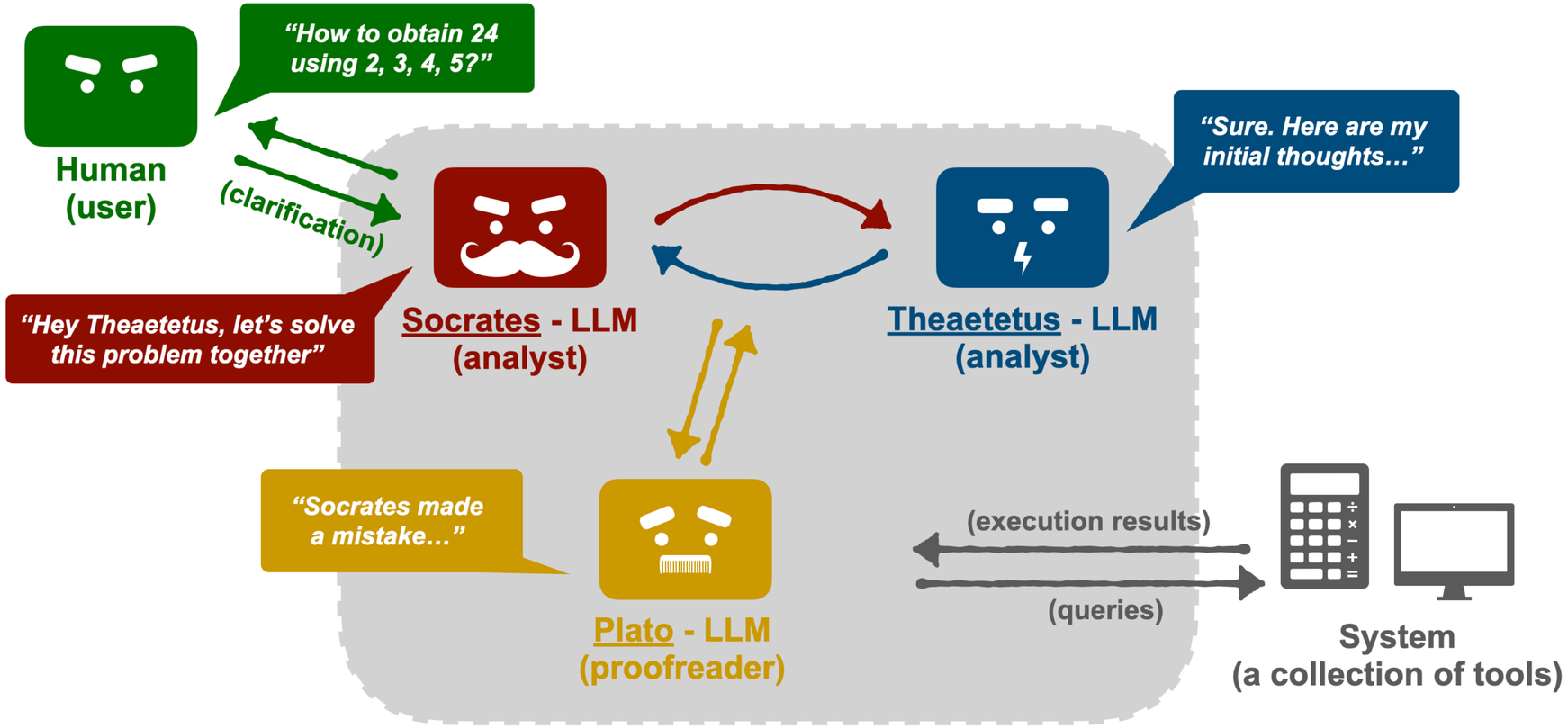
The magic of prompt design for LLMs: "Take a deep breath?"

Table 1: Top instructions with the highest GSM8K zero-shot test accuracies from prompt optimization with different optimizer LLMs. All results use the pre-trained `PaLM 2-L` as the scorer.

Source	Instruction	Acc
Baselines		
(Kojima et al., 2022)	Let's think step by step.	71.8
(Zhou et al., 2022b)	Let's work this out in a step by step way to be sure we have the right answer.	58.8
	(empty string)	34.0
Ours		
`PaLM 2-L-IT`	Take a deep breath and work on this problem step-by-step.	**80.2**
`PaLM 2-L`	Break this down.	79.9
`gpt-3.5-turbo`	A little bit of arithmetic and a logical approach will help us quickly arrive at the solution to this problem.	78.5
`gpt-4`	Let's combine our numerical command and clear thinking to quickly and accurately decipher the answer.	74.5

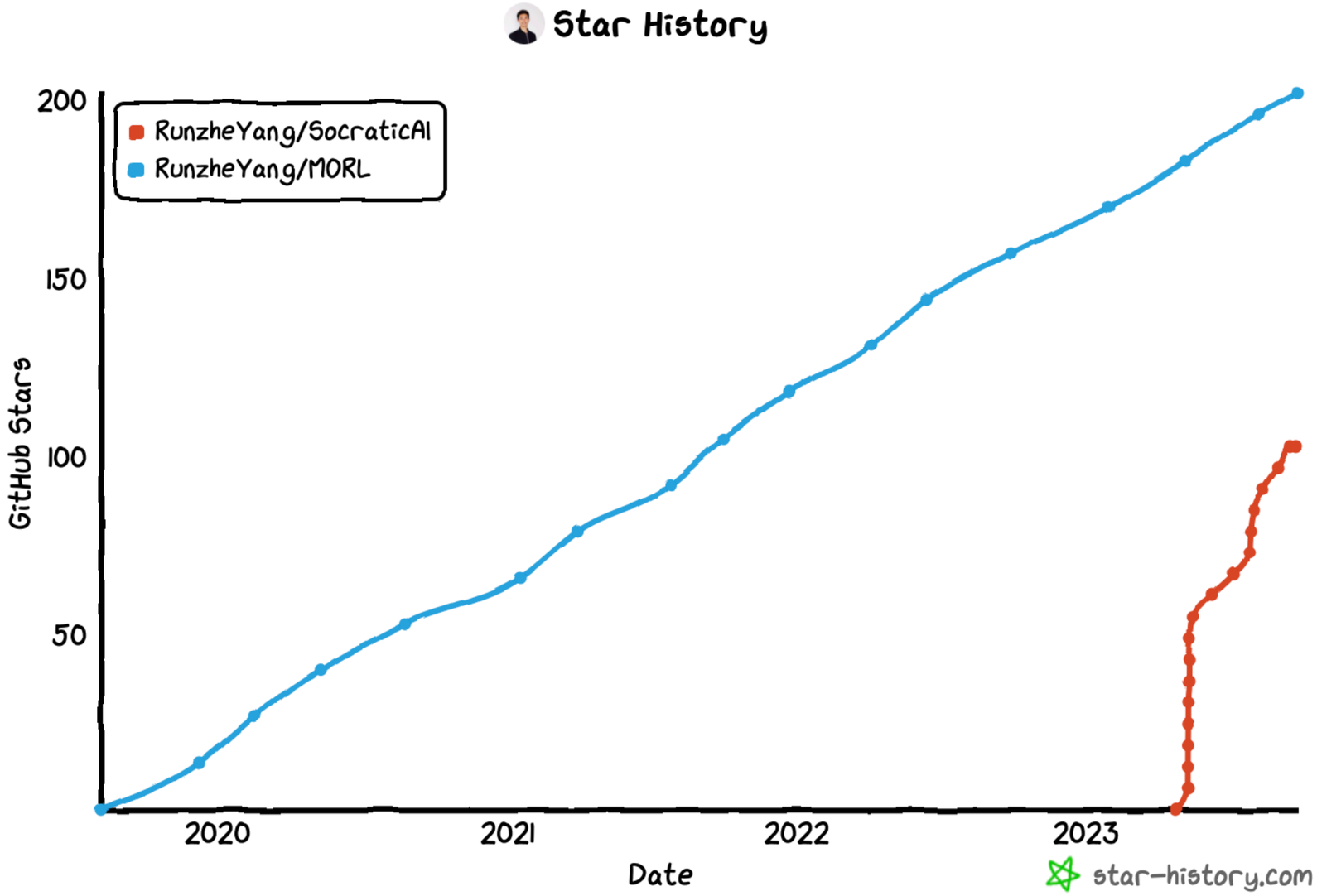
Table modified from *Large Language Models as Optimizers*. Yang *et al.*, 2023

Preliminary exploration: LLMs generate prompts for themselves

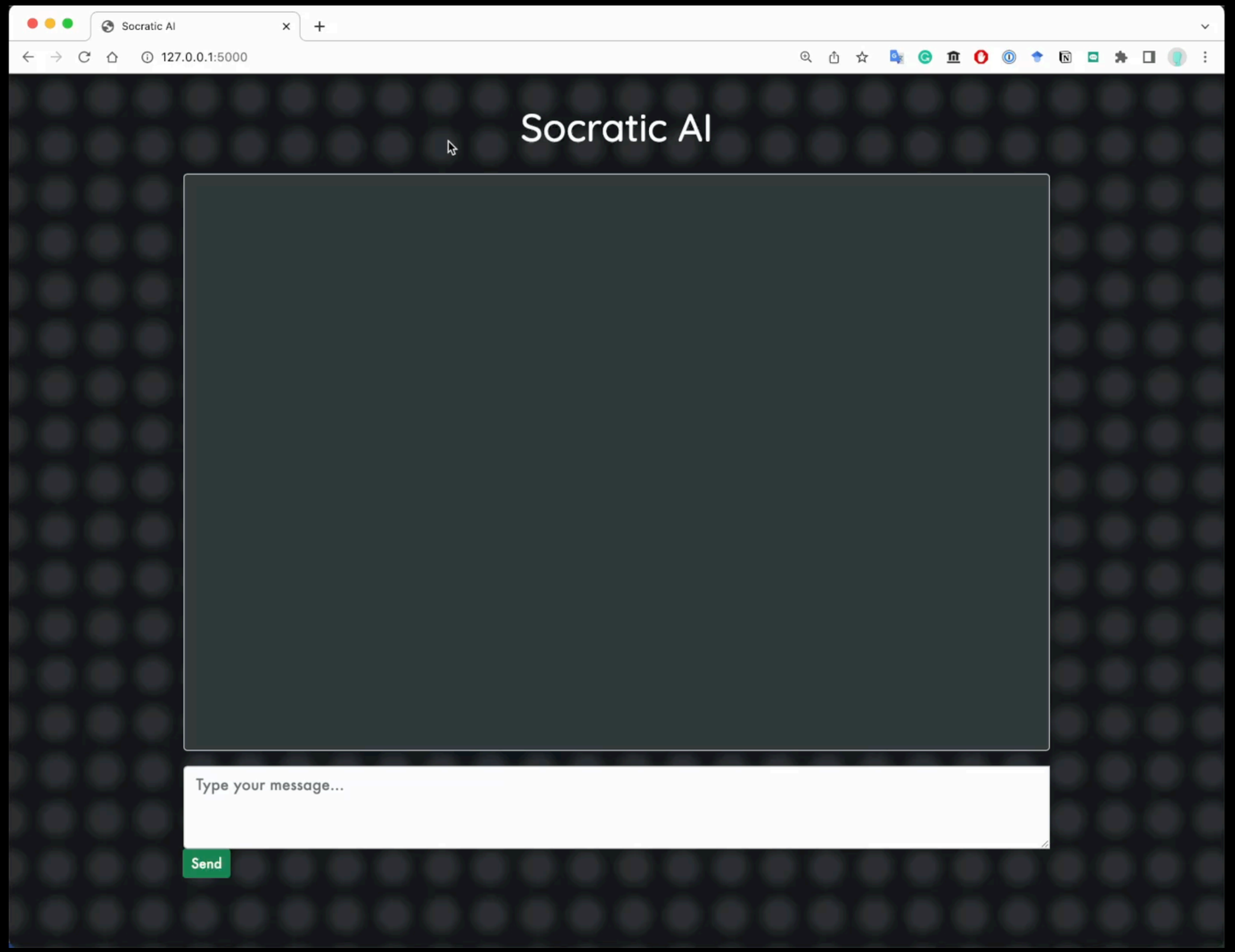


Socratic AI: a framework for collaborative problem-solving with LLMs

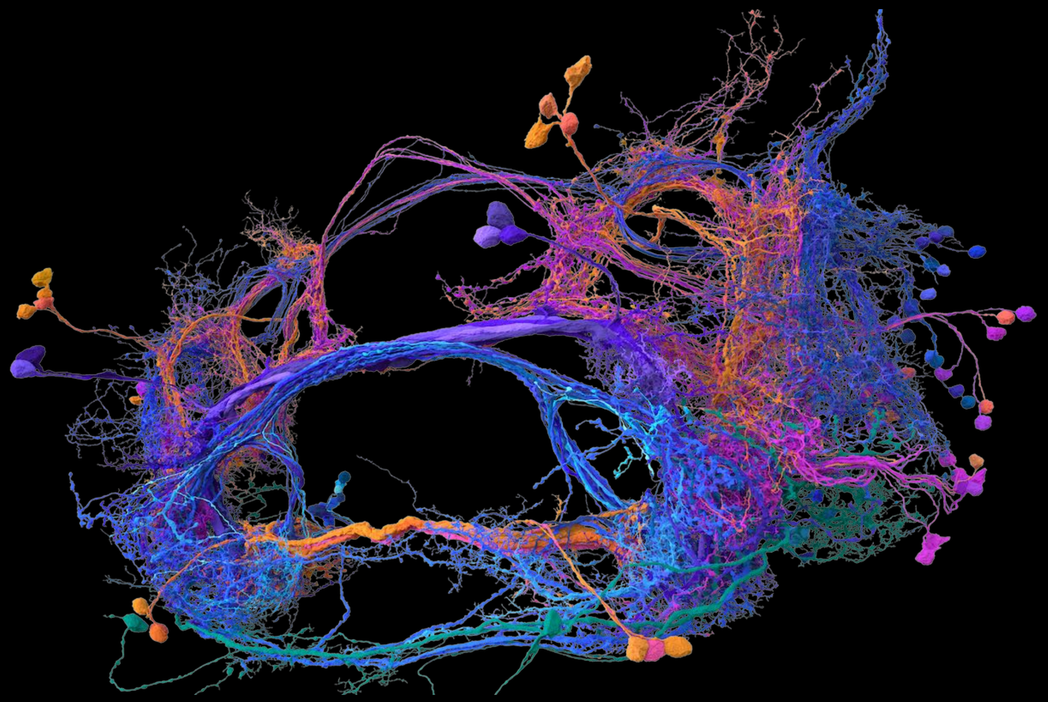
Stars on Github…



Demo: estimate the synapse density in the fly brain



New frontiers during my PhD: from brain reconstruction to generative AI

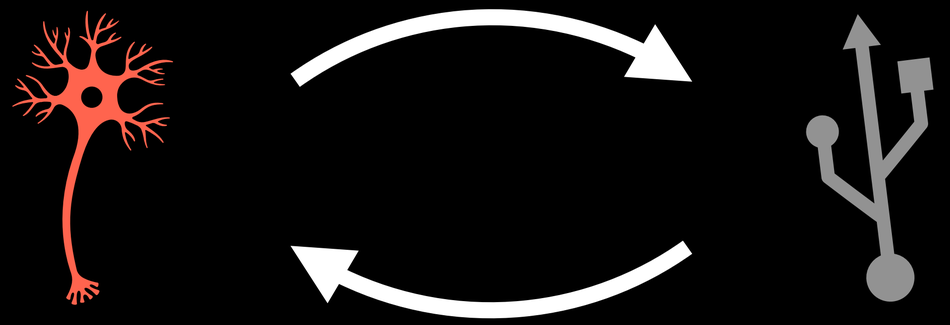


3D Reconstruction of Neuronal Circuits

Prompt: "A photograph of neuron-like stardust"

The future: a virtuous cycle between AI and Neuroscience

Inspire new algorithms



Accelerate discoveries

List of publications during my PhD

Neuroscience

- **Cyclic structure with cellular precision in a vertebrate sensorimotor neural circuit. *Current Biology*, 2023.**
  Runzhe Yang, Ashwin Vishwanathan, Jingpeng Wu, Nico Kemnitz, Dodam Ih, Nicholas Turner, Kisuk Lee, Ignacio Tartavull, William M. Silversmith, Chris S. Jordan, Celia David, Doug Bland, Amy Sterling, Mark S. Goldman, Emre R. F. Aksay, H. Sebastian Seung, and the EyeWirers.

- **Reconstruction of neocortex: Organelles, compartments, cells, circuits, and activity. *Cell*, 2022.**
  Nicholas Turner*, Thomas Macrina*, Alexander Bae*, **Runzhe Yang***, Alyssa Wilson*, Casey Schneider-Mizell*, Kisuk Lee*, Ran Lu*, Jingpeng Wu*, Agnes Bodor*, Adam Bleckert*, Derrick Brittain*, Emmanouil Froudarakis*, Sven Dorkenwald*, Forrest Collman*, Nico Kemnitz* (equal contribution), Dodam Ih, William M. Silversmith, Jonathan Zung, Aleksandar Zlateski, Ignacio Tartavull, Szi-chieh Yu, Sergiy Popovych, Shang Mu, William Wong, Chris S. Jordan, Manuel Castro, JoAnn Buchanan, Daniel J. Bumbarger, Marc Takeno, Russel Torres, Gayathri Mahalingam, Leila Elabbady, Yang Li, Erick Cobos, Pengcheng Zhou, Shelby Suckow, Lynne Becker, Liam Paninski, Franck Polleux, Jacob Reimer, Andreas S. Tolias, R. Clay Reid, Nuno Maçarico da Costa , Sebastian Seung.

- **Network statistics of the whole-brain connectome of *Drosophila*. *bioRxiv 2023. under submission*.**
  Albert Lin*, **Runzhe Yang***, Sven Dorkenwald, Arie Matsliah, Amy Sterling, Philipp Schelgel, Szi-chieh Yu, Claire McKellar, Marta Costa, Kathi Eichler, Alexander Shakeel Bates, Nils Eckstein, Jan Funke, Gregory S.X.E. Jefferis, Mala Murthy

- **Neuronal wiring diagram of an adult brain. *bioRxiv 2023. under submission*.**
  Sven Dorkenwald, Arie Matsliah, Amy R Sterling, Philipp Schlegel, Szi-chieh Yu, Claire E. McKellar, Albert Lin, Marta Costa, Katharina Eichler, Yijie Yin, Will Silversmith, Casey Schneider-Mizell, Chris S. Jordan, Derrick Brittain, Akhilesh Halageri, Kai Kuehner, Oluwaseun Ogedengbe, Ryan Morey, Jay Gager, Krzysztof Kruk, Eric Perlman, Runzhe Yang, David Deutsch, Doug Bland, Marissa Sorek, Ran Lu, Thomas Macrina, Kisuk Lee, J. Alexander Bae,, Shang Mu, Barak Nehoran, Eric Mitchell, Sergiy Popovych, Jingpeng Wu, Zhen Jia, Manuel Castro, Nico Kemnitz, Dodam Ih, Alexander Shakeel Bates, Nils Eckstein, Jan Funke, Forrest Collman, Davi D. Bock, Gregory S.X.E. Jefferis, H. Sebastian Seung*, Mala Murthy*, the FlyWire Consortium.

- **Functional Connectomics Spanning Multiple Areas of Mouse Visual Cortex. *bioRxiv 2021, under submission***
  MICrONS Consortium et al.

- **Modularity and Neural Coding from a Brainstem Synaptic Wiring Diagram. *bioRxiv 2020, under submission***
  Ashwin Vishwanathan, Alexandro D. Ramirez*, Jingpeng Wu*, Alex Sood, **Runzhe Yang**, Nico Kemnitz, Dodam Ih, Nicholas Turner, Kisuk Lee, Ignacio Tartavull, William M. Silversmith, Chris S. Jordan, Celia David, Doug Bland, Mark S. Goldman, Emre R. F. Aksay, H. Sebastian Seung, the EyeWirers.

List of publications during my PhD

Machine Learning

- **Improving Dialog Systems for Negotiation with Personality Modeling.** ***(ACL 2021, Selected Oral)***
  Runzhe Yang*, Jingxiao Chen* and Karthik Narasimhan.

- **A Generalized Algorithm for Multi-Objective Reinforcement Learning and Policy Adaptation.** ***(NeurIPS 2019)***
  Runzhe Yang, Xingyuan Sun and Karthik Narasimhan.

- **DataMUX: Data Multiplexing for Neural Networks.** ***(NeurIPS 2022, Bell Labs 2nd Prize)***
  Vishvak Murahari, Carlos Jimenez, **Runzhe Yang**, Karthik Narasimhan.

- **LLMs are Superior Feedback Providers: Bootstrapping Reasoning for Lie Detection with Self-Generated Feedback.** ***under submission***
  Tanushree Banerjee, Richard Zhu, **Runzhe Yang**, Denis Peskov, Brandon Stewart, Karthik Narasimhan.

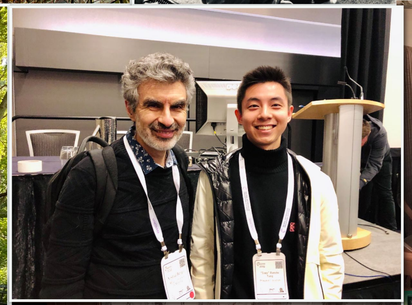
- **COLLIE: Systematic Construction of Constrained Text Generation Tasks.** ***under submission***
  Shunyu Yao*, Howard Chen*, Austin Wang*, **Runzhe Yang***, Karthik Narasimhan.

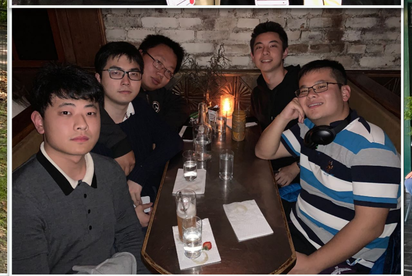
NeuroAI

- **Neuronal Circuits for Robust Online Fixed-point Detection.** ***(COSYNE 2023)***
  Runzhe Yang, David Lipshutz, Tiberiu Tesileanu, Johannes Friedrich, Dmitri Chklovskii.

- **Unsupervised Feature Discovery by Neural Networks with Disynaptic Recurrent Inhibition (NAISys 2020, selected talk)**
  Runzhe Yang, Kyle Luther, H. Sebastian Seung

- **Unsupervised Learning by a "Softened" Correlation Game: Duality and Convergence.** ***(ACSSC 2019)***
  Kyle Luther*, **Runzhe Yang*** and Sebastian Seung.





THANK YOU!! ❤