NETWORK ARCHITECTURES SUPPORTING LEARNABILITY

SIAM - DS
MAY 23, 2019

Danielle S. Bassett

University of Pennsylvania
Department of Bioengineering
Department of Physics & Astronomy
Department of Neurology
Department of Psychiatry
Department of Electrical & Systems Engineering

"The life that is here proposed to depict was a life singularly devoid of incident. It was the career of a lonely, secluded, fastidious, and affectionate man; it was a life not rich in results, not fruitful in example. It is the history of a few great friendships, much quiet benevolence, tender loyalty, wistful enjoyment. The tangible results are a single small volume of imperishable quality, some accomplished translations of not great literary importance, a little piece of delicate prose-writing, and many beautiful letters."

I. Knowledge is a network

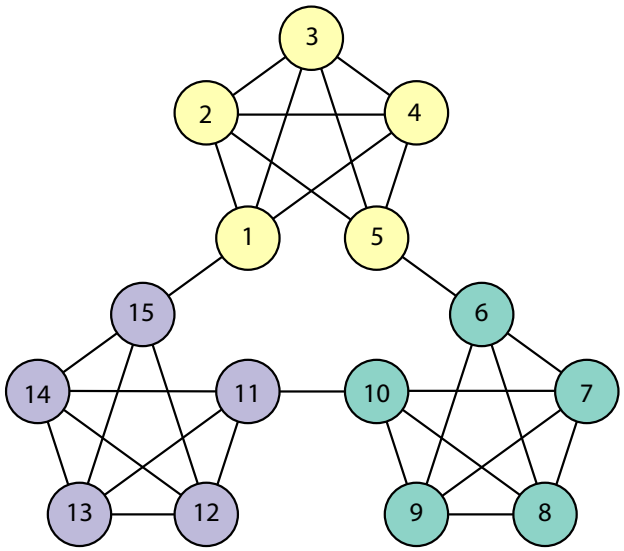
From Henri Poincare's 1905 *Science and Hypothesis*:

"The aim of science is not things themselves, as the dogmatists in their simplicity imagine, but the relations among things; outside these relations there is no reality knowable."

From Dewey's 1916 *Democracy and Education* (NY: Simon & Brown, 2011):

"…[K]nowledge is a perception of those connections of an object which determine its applicability in a given situation. [...] Thus, we get at a new event indirectly instead of immediately - by invention, ingenuity, resourcefulness. An ideally perfect knowledge would represent such a network of interconnections that any past experience would offer a point of advantage from which to get at the problem presented in a new experience" (185).

II. Knowledge is a network learned by example



Suppose I must translate 15 ideas to a class.

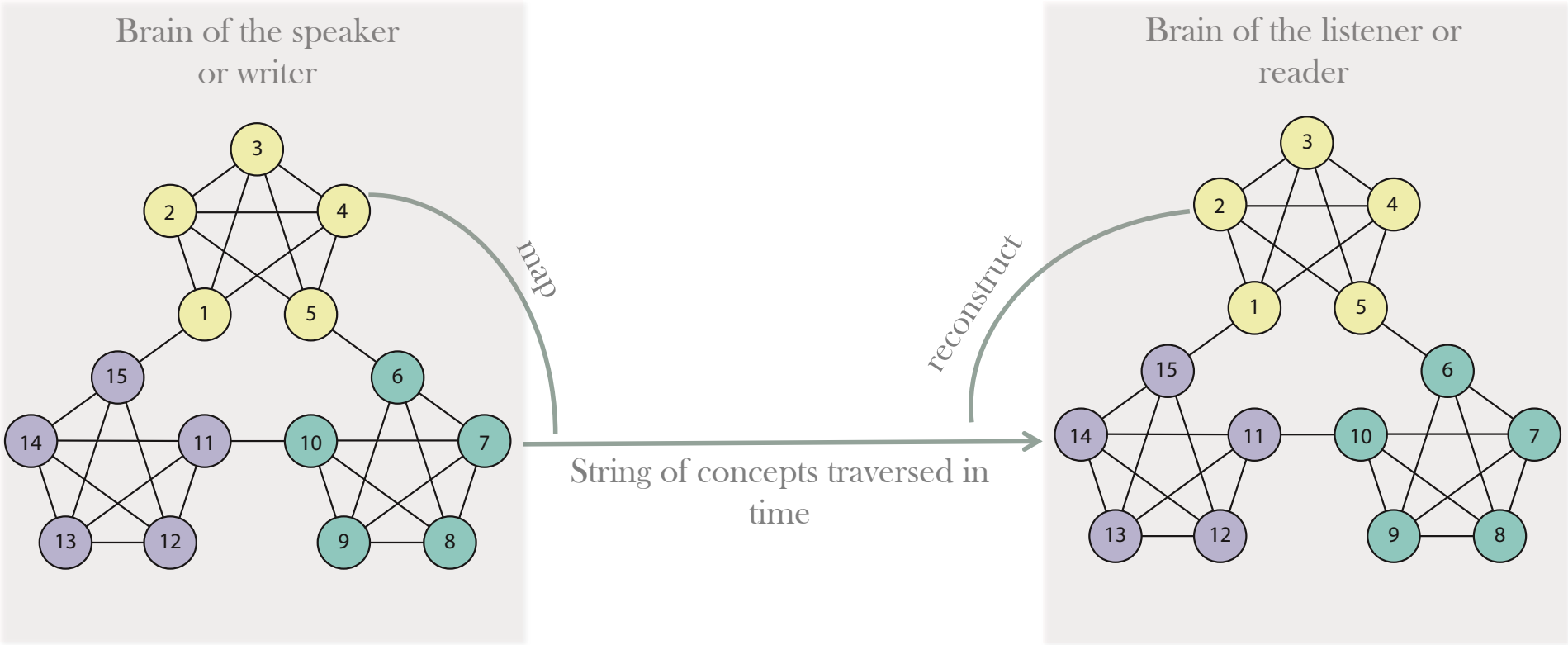
Those 15 ideas are related to one another in a heterogeneous manner, making a network.

I must translate that information linearly; time is one-dimensional and uni-directional.

How should I create this information time series in a way that maximizes learning?

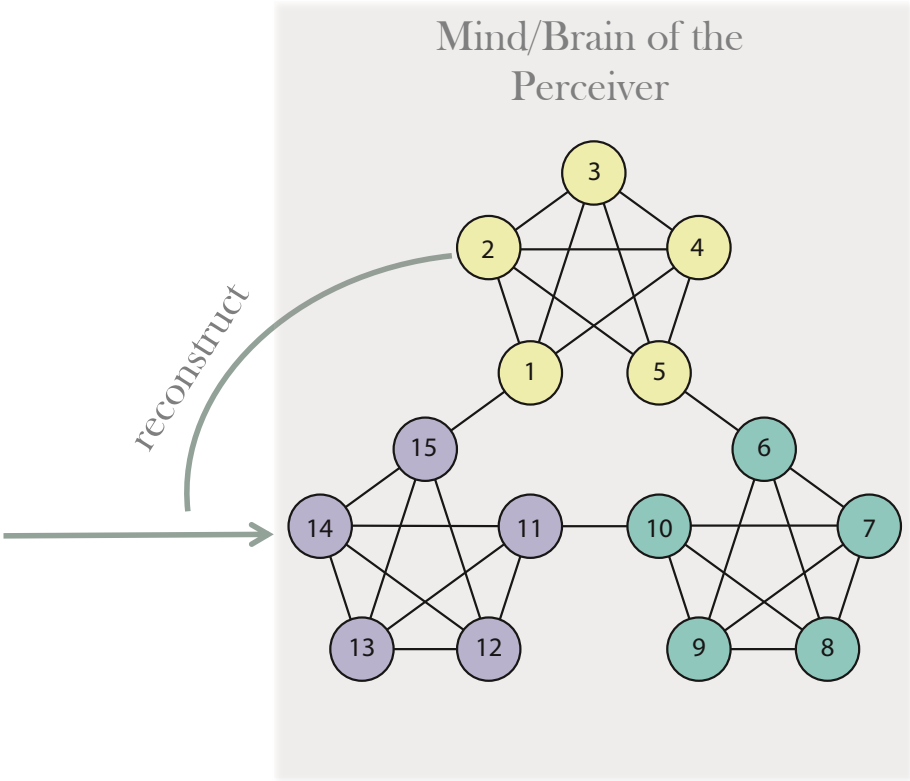
A "good walk" minimizes reconstruction error

and maximizes perception of the network's topology



The problem of inferring the patterns of pairwise dependencies from incoming streams of data allows us to:

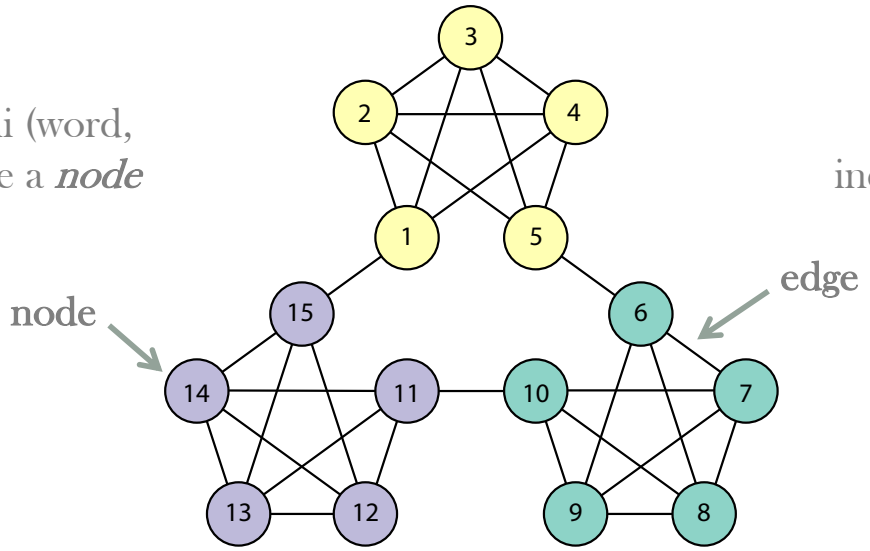
Learn language
Segment visual events
Parse tonal groupings
Parse spatial scenes
Infer social networks
Perceive distinct concepts



Can we measure perception of network topology

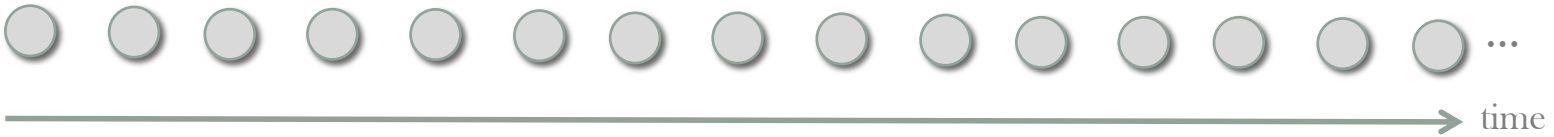
in a continuous stream of stimuli?

Let each specific stimuli (word, image, or movement) be a ***node in a graph***.



Let each ***edge in the graph*** indicate an allowable transition between nodes.

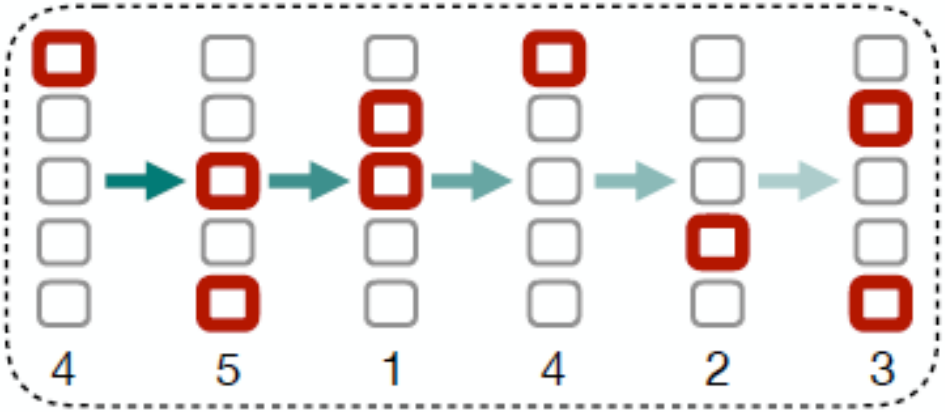
Construct a sequence of stimuli by a random walk on the graph.



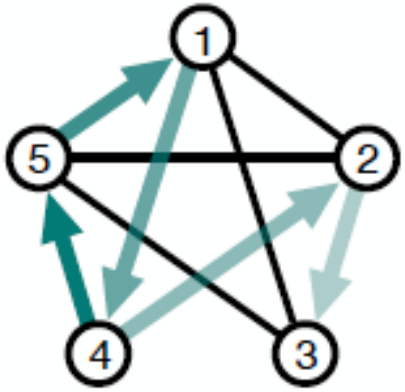
At each stimuli, require the participant to perform a task, so that their time-to-react can be used as a measure of how well that edge in the graph was learned.

Example experimental setup

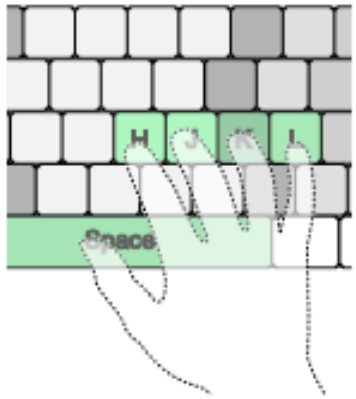
1. Motor: *Kahn et al. 2018 Nature Human Behavior*



Stream of Stimuli



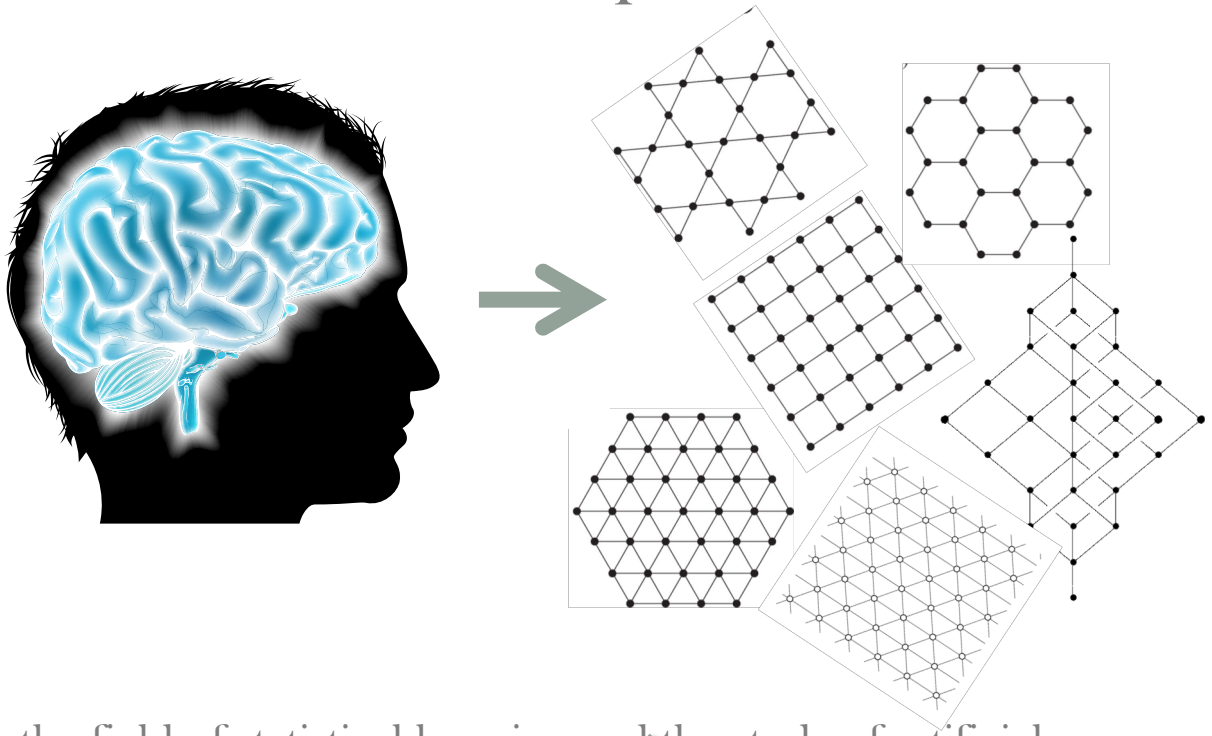
Transition graph



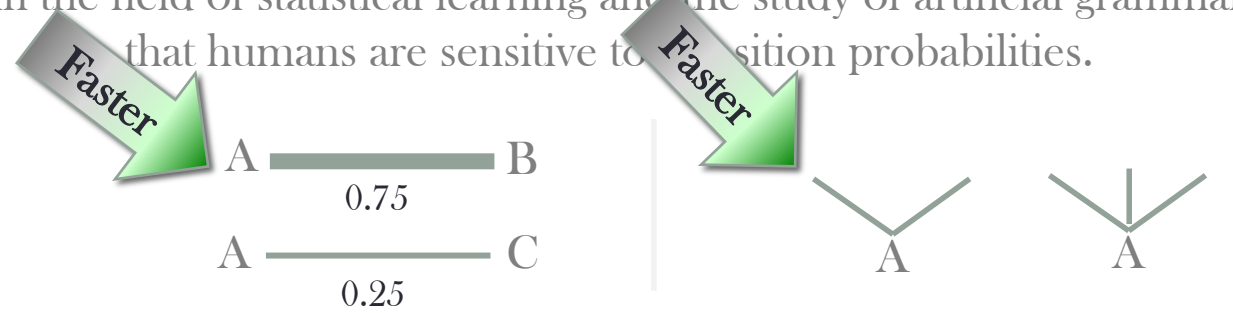
Hand Placement

2. Visual: Karuza et al. 2017 *Scientific Reports*
3. Social: Tompson et al. 2018 *Journal of Experimental Psychology; Learning Memory & Cognition*

What do we know about this problem?



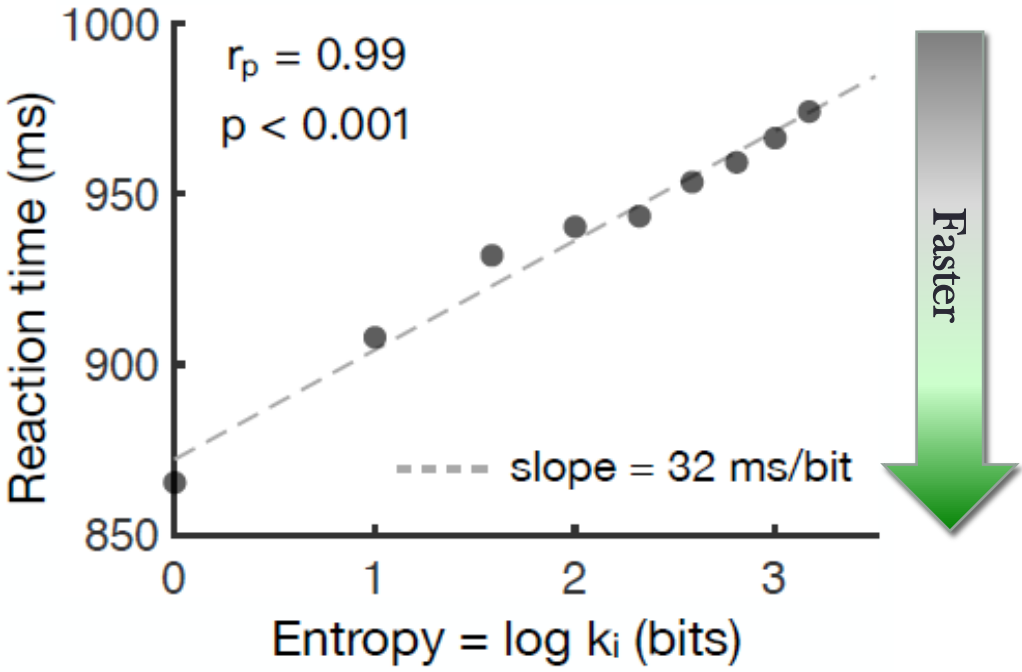
From work in the field of statistical learning and the study of artificial grammars, we know that humans are sensitive to transition probabilities.



Human reactions depend on entropy

Human reactions should vary with the entropy of a transition from one stimulus *i* to another stimulus *j*.

For random walks on a simple undirected network, the entropy of the ($i \rightarrow j$) transition equals log k_j, where k_j is the degree of node *j*.

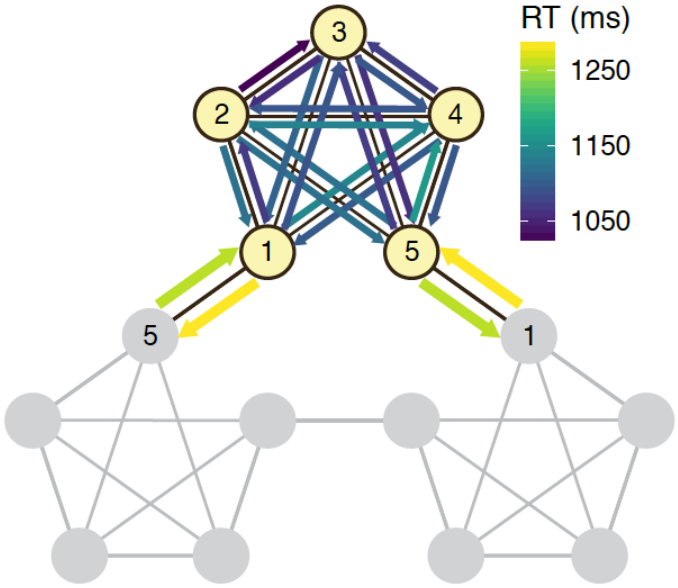


In this experiment, humans appear to process 1 bit of information in 32 ms.

Human reactions depend on MORE THAN entropy

In a k-4 regular graph, traversed by a random walk, the entropy of all transitions is the same.

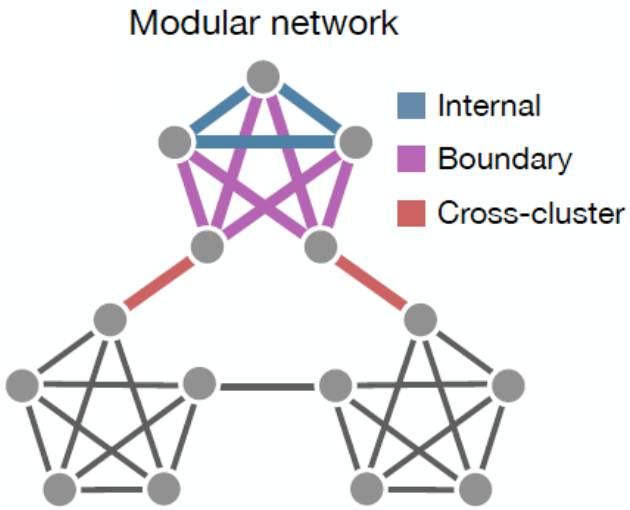
Measure Reaction Time (RT)
Per Transition

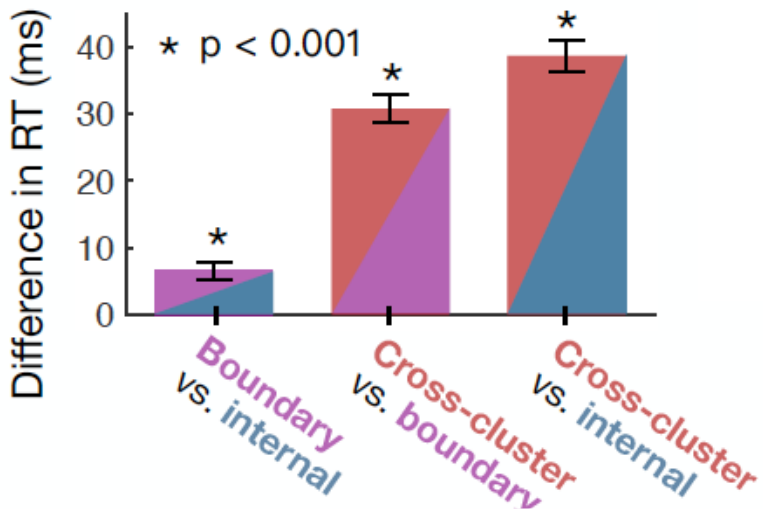


Striking slowing at cluster boundaries
indicating graph learning

In fact, human reaction times are sensitive to hierarchy …

A modular network, which by symmetry, only contains three types of transitions; each type produces reaction times that are distinct from the other two.

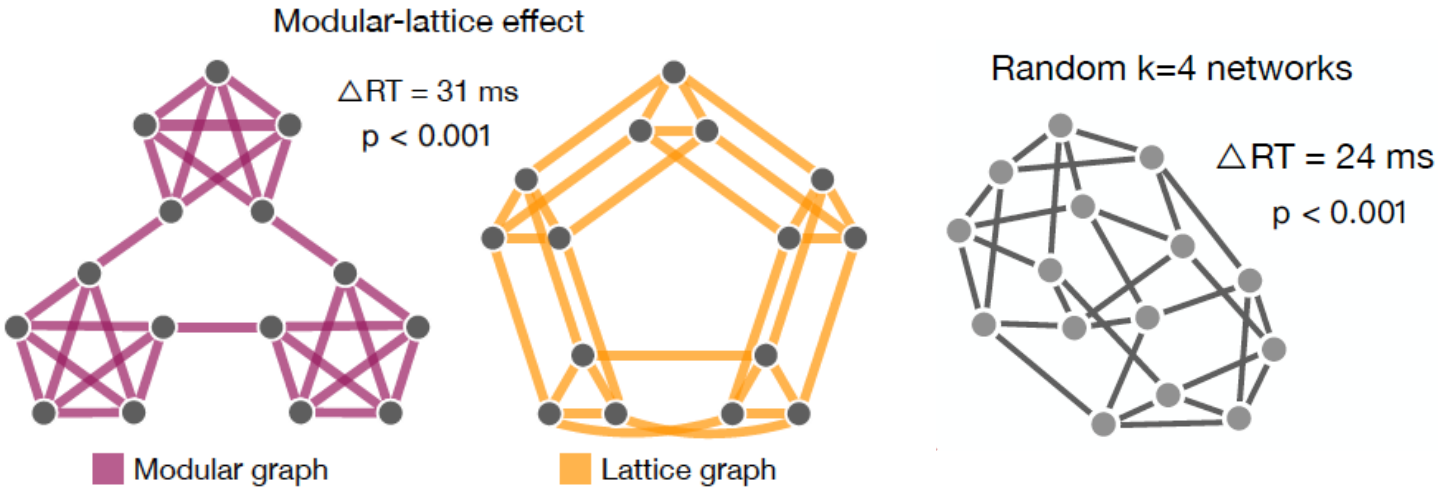




Transitions between or at the boundaries of modules generate more surprise than transitions deep within a module.

… And to network topology.

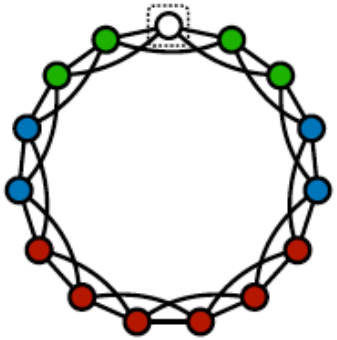
Compared to lattice and random graphs with equal entropy, reactions in the modular graph are significantly faster overall, indicating a decreased in perceived information.



Together, these results reveal that humans process information beyond entropy in a manner that depends systematically on network topology.

Perturbation experiment: effects of network violations

- Ring graph:

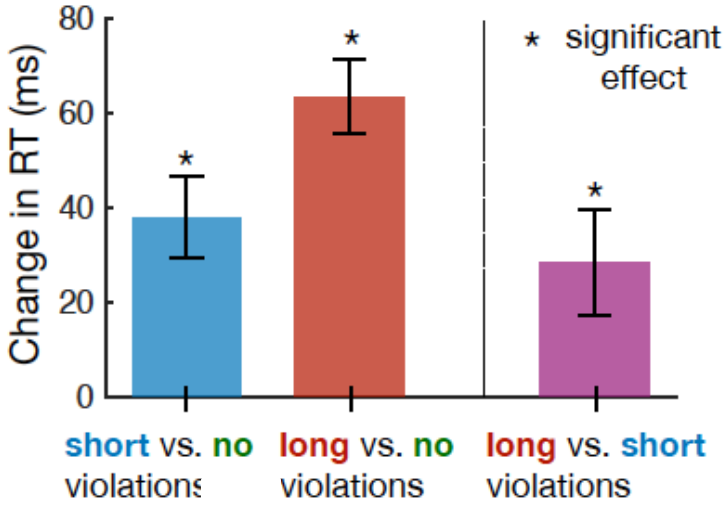


Topological distance:

- 0 (current node)
- 1 (no violation)
- 2 (short violation)
- 3,4 (long violation)

N= 99

- Empirical results:



Humans are more surprised by stimuli from farther away on the ring than closer, indicating their implicit perception of the network topology.

What are people thinking?

Creating expectations of transitions

We build expectations about a network structure with a counts matrix n_{ij}

Free energy principle: brain minimizes errors & computational complexity.

Probability of recalling $x_{t-\Delta t}$ rather than x_t is $Q(\Delta t)$.

Creating expectations of transitions

We build expectations about a network structure with a counts matrix n_{ij}

Free energy principle: brain minimizes errors & computational complexity.

Probability of recalling $x_{t-\Delta t}$ rather than x_t is $Q(\Delta t)$.

The error of a candidate probability distribution is $E(Q) = \sum_{\Delta t} Q(\Delta t)\Delta t$

Complexity of the error distribution is the entropy $-S(Q) = \sum_{\Delta t} Q(\Delta t) \log Q(\Delta t)$

Creating expectations of transitions

We build expectations about a network structure with a counts matrix n_{ij}

Free energy principle: brain minimizes errors & computational complexity.

Probability of recalling $x_{t-\Delta t}$ rather than x_t is $Q(\Delta t)$.

The error of a candidate probability distribution is $E(Q) = \sum_{\Delta t} Q(\Delta t)\Delta t$

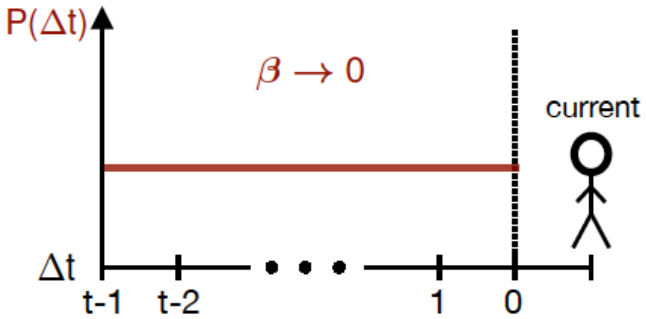
Complexity of the error distribution is the entropy $-S(Q) = \sum_{\Delta t} Q(\Delta t) \log Q(\Delta t)$

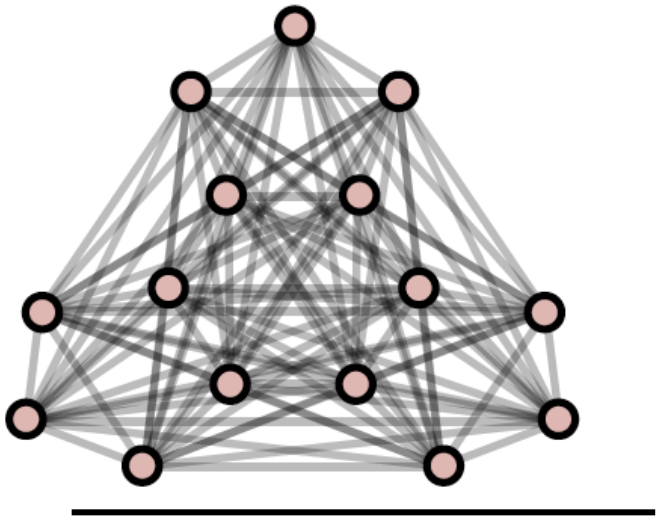
Total cost of the distribution is its free energy: $F(Q) = \beta E(Q) - S(Q)$

Distribution that minimizes the free energy is the Boltzmann distribution $P(\Delta t) = \frac{1}{Z} e^{-\beta \Delta t}$

From a poor memory arises biases in learning

No memory;
Minimizes mental resources

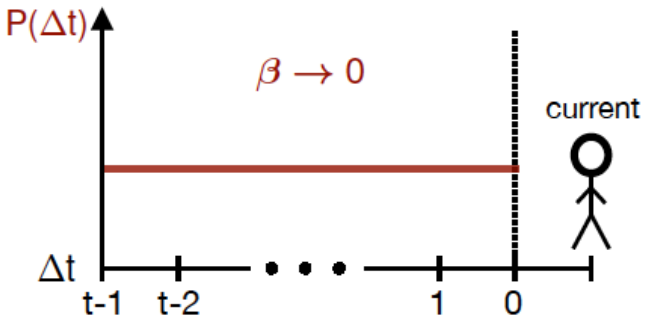




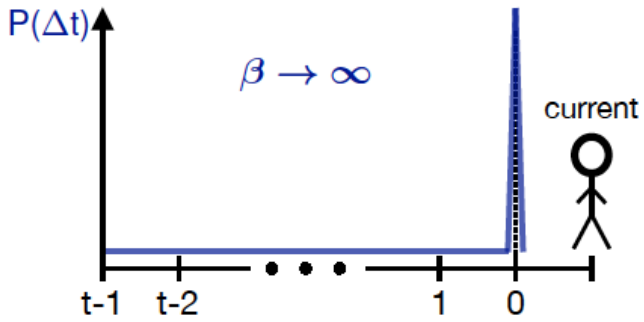
Lynn et al. 2018, arXiv:1805.12491

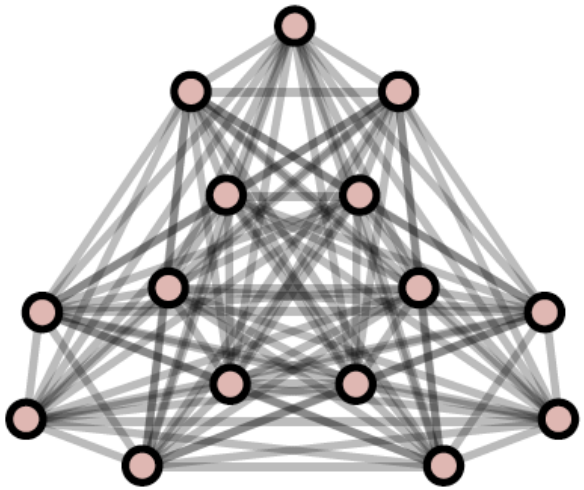
From a poor memory arises biases in learning

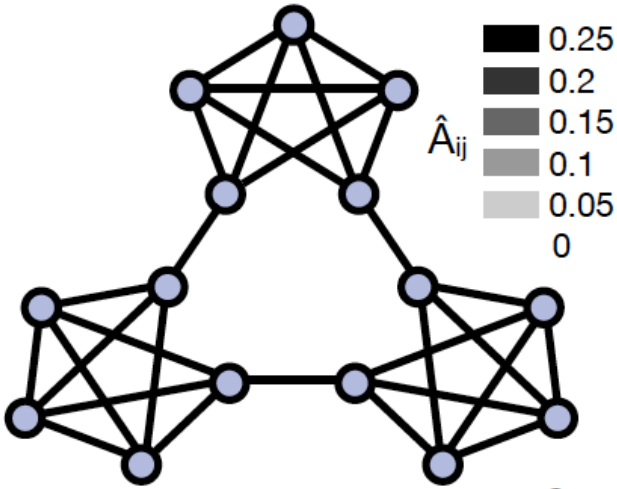
No memory;
Minimizes mental resources



Perfect memory;
Maximizes mental resources







β

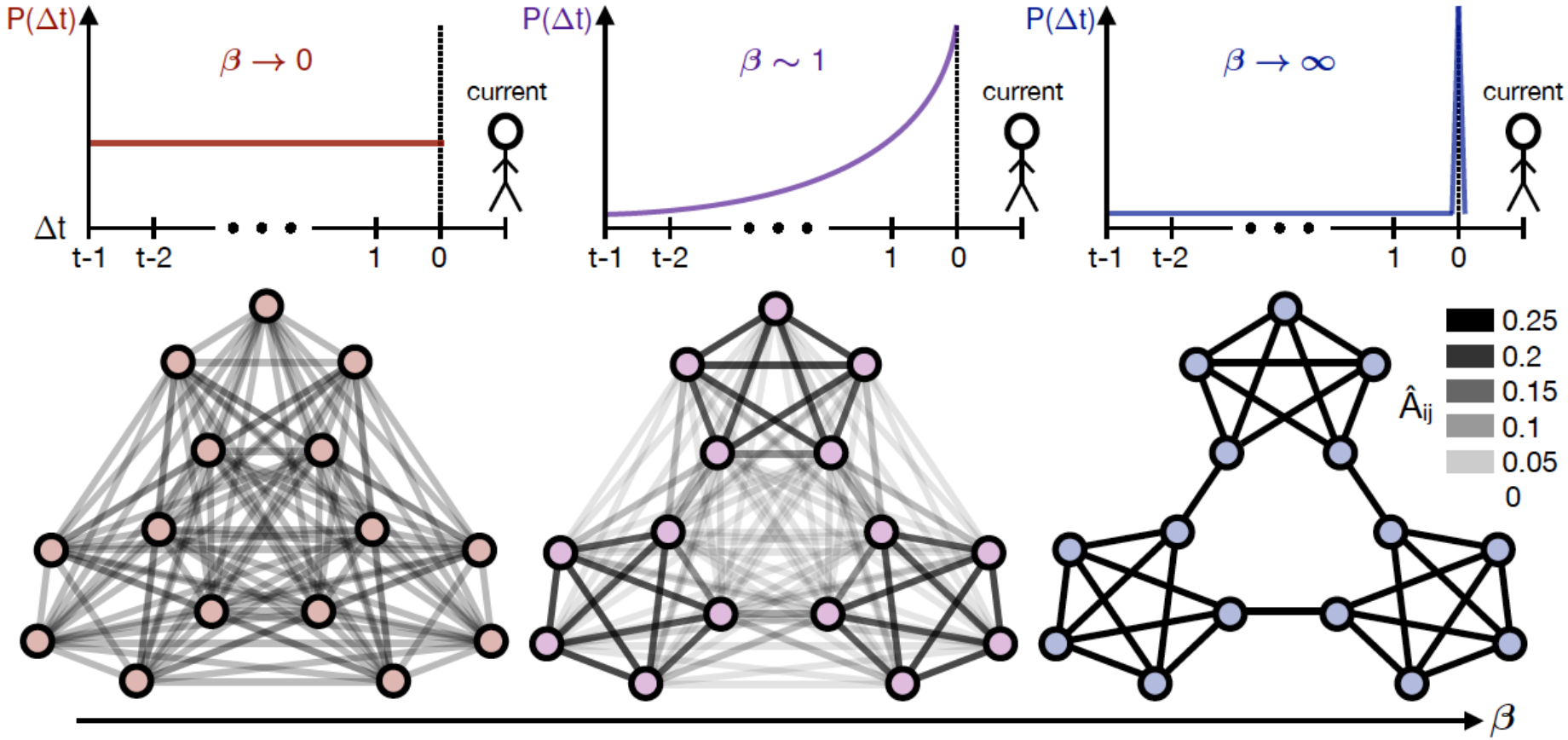
Lynn et al. 2018, arXiv:1805.12491

From a poor memory arises biases in learning

No memory;
Minimizes mental resources

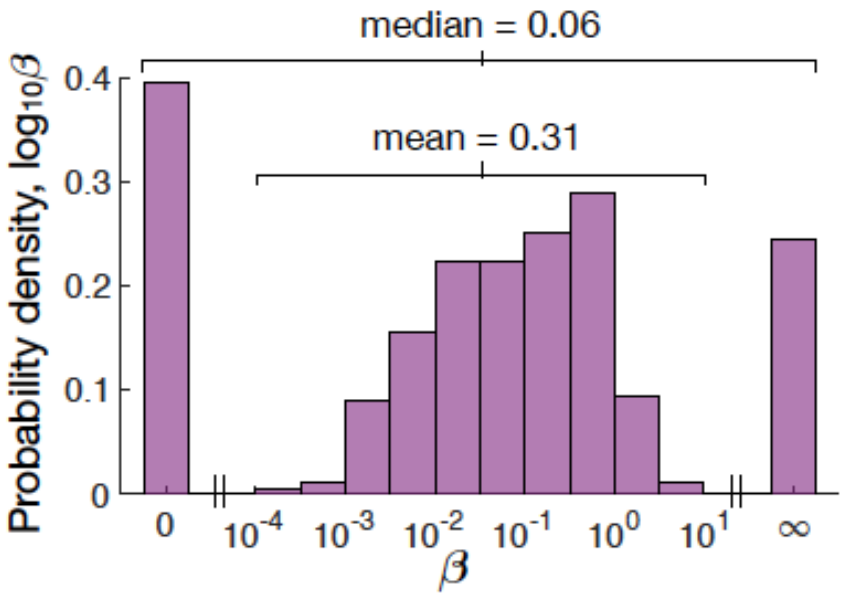
Poor memory

Perfect memory;
Maximizes mental resources



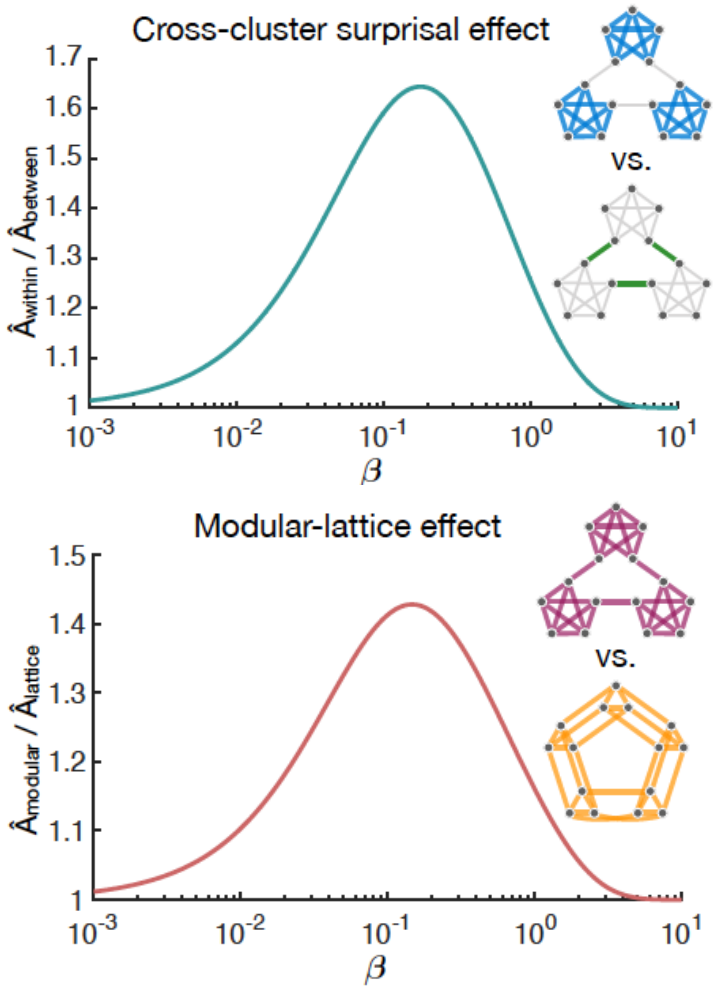
Measuring inverse temperature from RT data

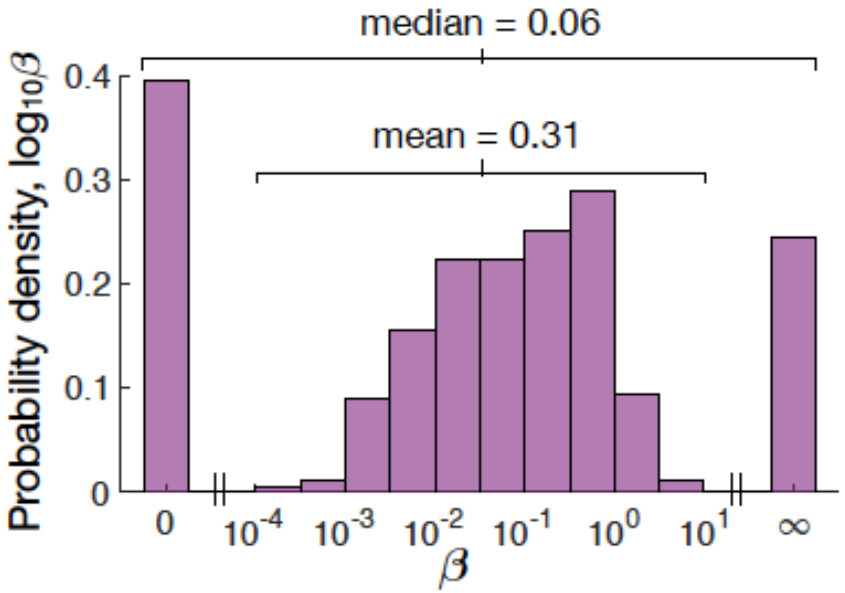
Fitting the model to the graph learning data from 358 participants.



N= 71; 243; 44

Measuring inverse temperature from RT data



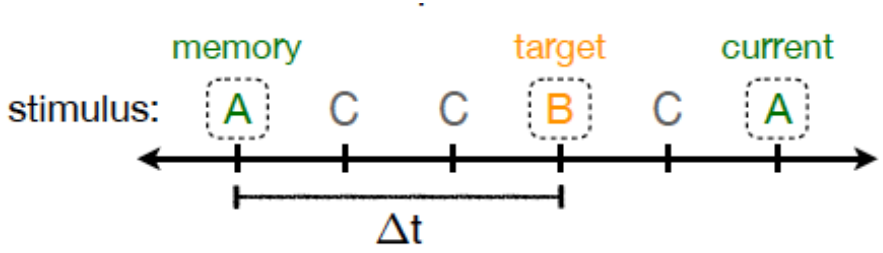


N= 71; 243; 44

Measuring the memory distribution in humans

In the field of psychology, the accuracy of a person's memory is often tested using what is called an "n-back task".

A 2-back task

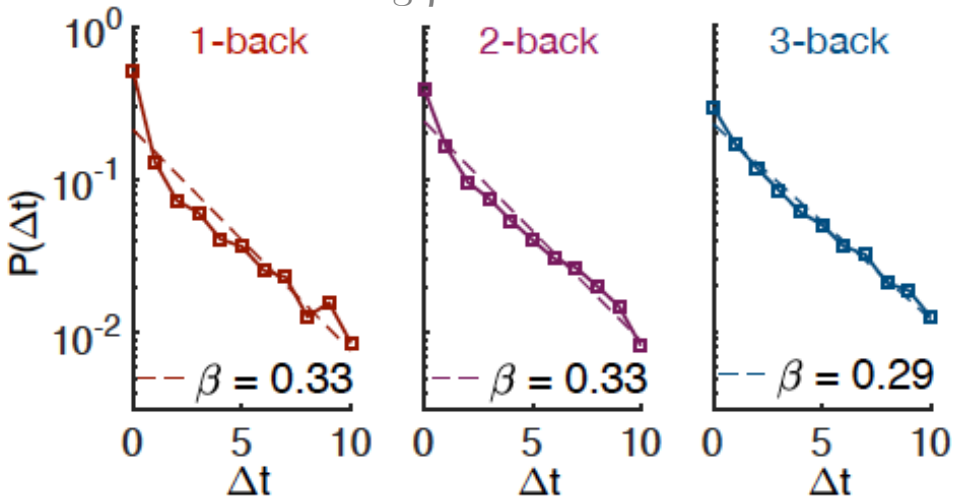


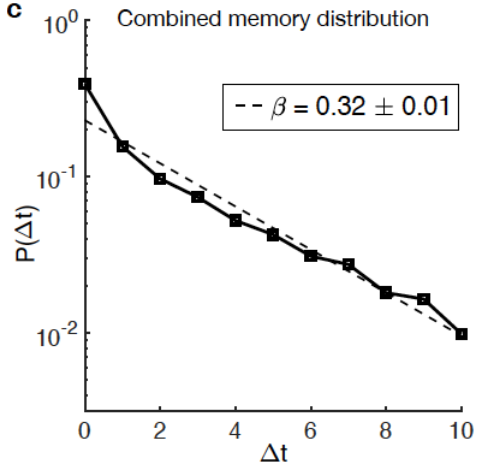
Goal is to determine whether the current letter is the same as the letter two before.

You are currently at A, and you must answer the question: "Did you see A 2x ago?"
You say "yes", but you are incorrect; "B" is 2x ago.
You say "yes" because you confuse 2x ago with what 5x ago.
5x minus 2x is your Δt.

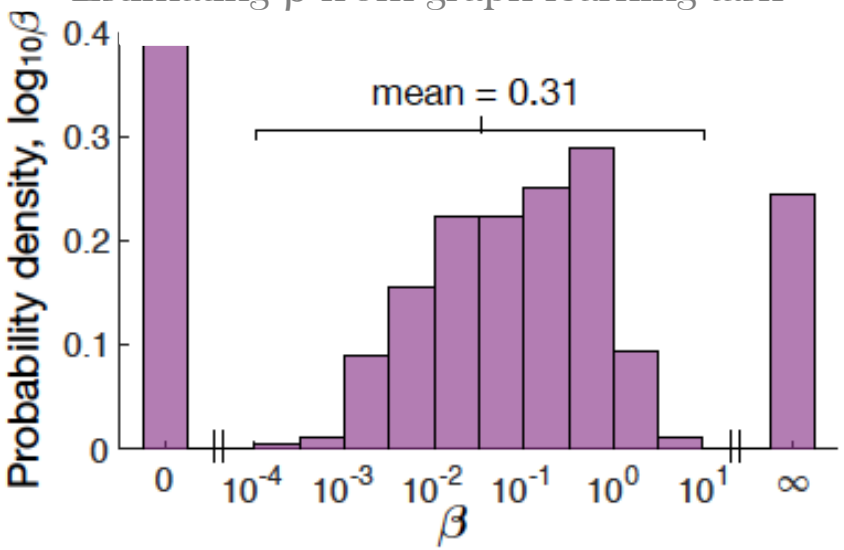
Measuring the memory distribution in humans

Estimating β from the 2-back task





Estimating β from graph learning task



The memory distribution we posit is the theory is consistent with the memory distribution we observe in experiment.

Information Processing in Real Networks

Perceived information

For intuition's sake, let's redefine the transition probability matrix as P, and state the perceived information as:

$$\underbrace{\left\langle -\log \hat{P}_{ij} \right\rangle_P}_{S(P,\hat{P})} = \underbrace{\left\langle -\log P_{ij} \right\rangle_P}_{S(P)} + \underbrace{\left\langle -\log \frac{\hat{P}_{ij}}{P_{ij}} \right\rangle_P}_{D_{\mathrm{KL}}(P||\hat{P})}.$$

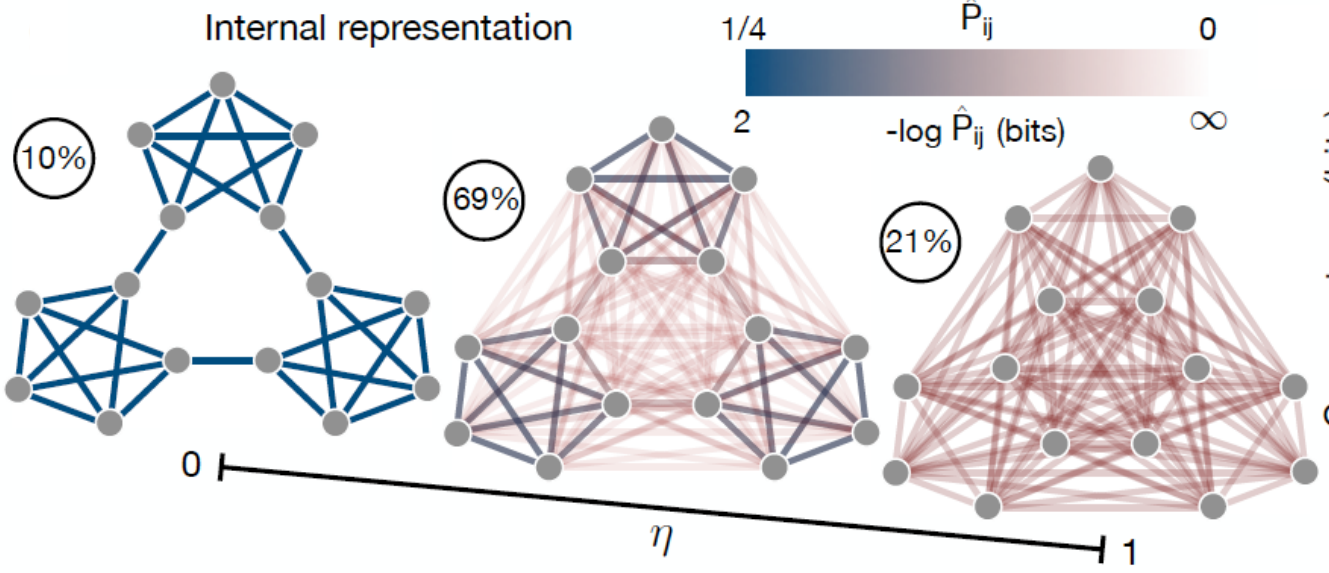
Cross entropy (perceived information)

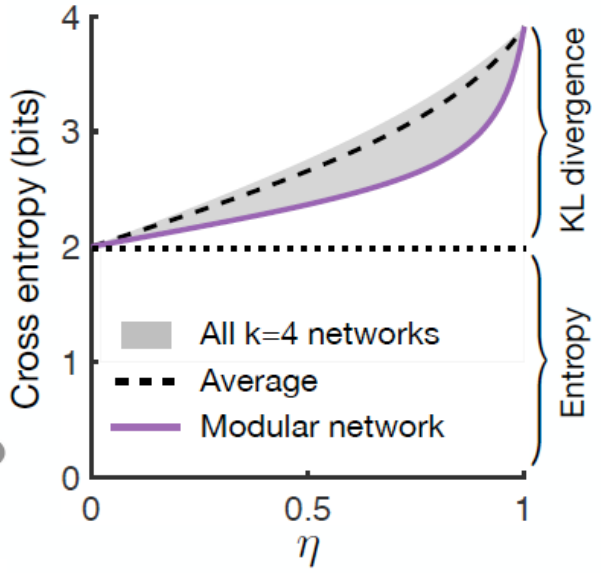
Entropy rate (rate of information production)

KL Divergence (inefficiency of observer's representation)

Perceived information depends on topology

Networks with identical entropy differ in cross entropy; modular having markedly low cross entropy, predicting the observed quickening of subjects' reactions.





$$\hat{P} = C \sum_{t=0}^{\infty} g(t) P^{t+1} \quad \text{where} \quad g(t) \geq 0$$

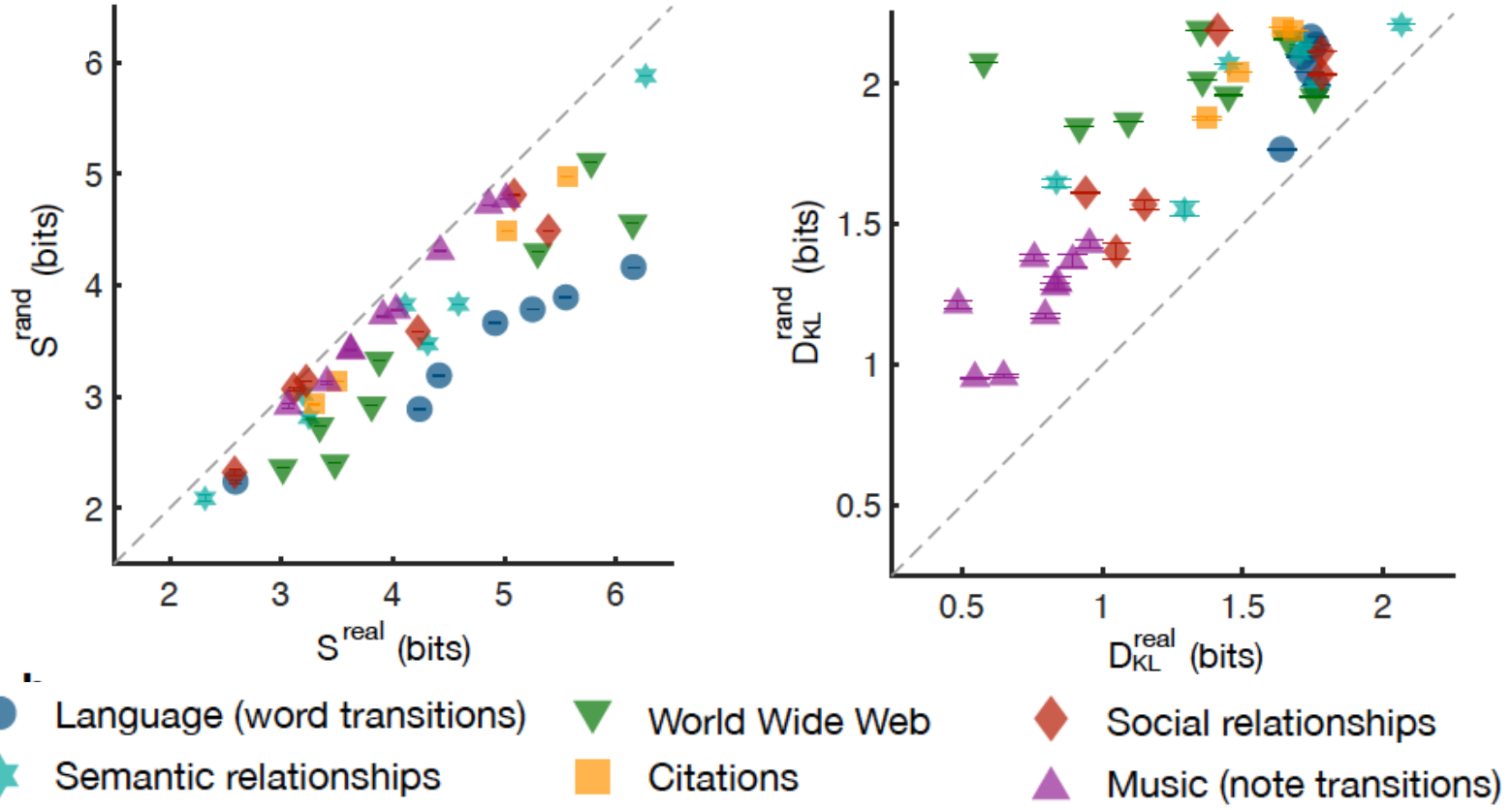
and $g(t) = \eta^t$ where $\eta \in (0, 1)$

Information processing in real networks

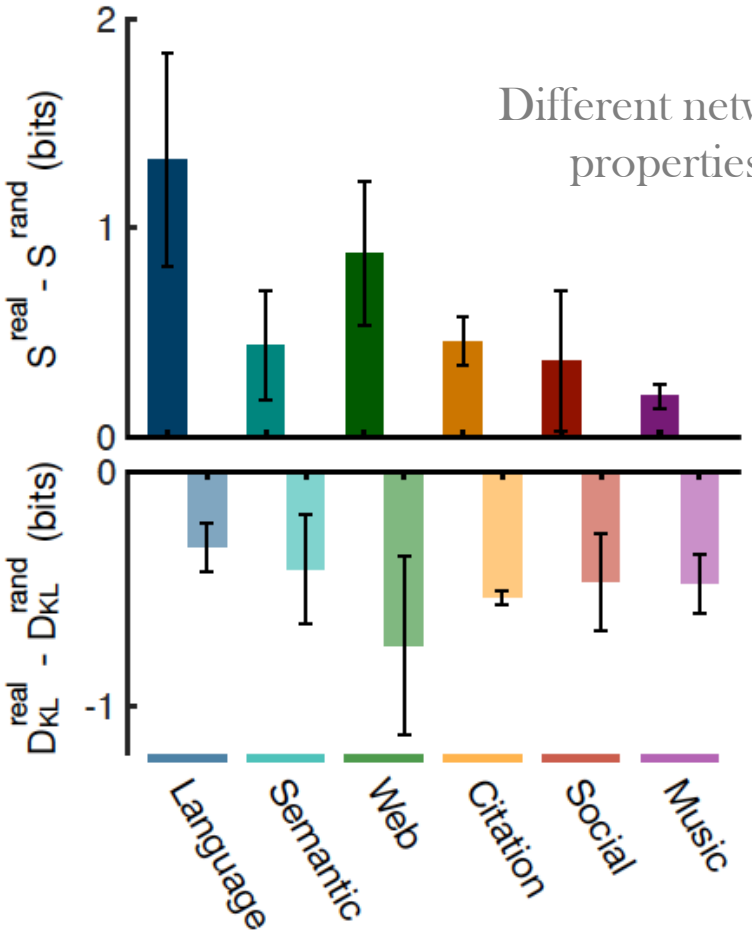
Type / Name	N	E
Language (word transitions)		
Shakespeare	11,234	97,892
Homer	3,556	23,608
Plato	2,271	9,796
Jane Austen	1,994	12,120
William Blake	370	781
Miguel de Cervantes	6,090	43,682
Walt Whitman	4,791	16,526
Semantic relationships		
Bible	1,707	9,059
Les Miserables	77	254
Edinburgh Thesaurus	7,754	226,518
Roget Thesaurus	904	3,447
Glossary terms	60	114
FOLDOC	13,274	90,736
ODLIS	1,802	12,378
World Wide Web		
Google internal	12,354	142,296
Education	2,622	6,065
EPA	2,232	6,876
Indochina	9,638	45,886
2004 Election blogs	793	13,484
Political blogs	643	2,280
Spam	3,796	36,404
Webbase	6,843	16,374
Citations		
arXiv Hep-Ph	12,711	139,500
arXiv Hep-Th	7,464	115,932
Cora	3,991	16,621
DBLP	240	858
Social relationships		
Facebook	13,130	75,562
arXiv Astr-Ph	17,903	196,972
Adolescent health	2,155	8,970
Highschool	67	267
Jazz	198	2,742
Karate club	34	78
Music (note transitions)		
Thriller – Michael Jackson	67	446
Hard Day's Night – Beatles	41	212
Bohemian Rhapsody – Queen	71	961
Africa – Toto	39	163
Sonata No 11 – Mozart	55	354
Sonata No 23 – Beethoven	69	900
Nocturne Op 9-2 – Chopin	59	303
Clavier Fugue 13 – Bach	40	143
Ballade No 1 – Brahms	69	670

Real networks display high entropy and low KL divergence

Since the networks chosen have evolved or were designed to communicate with humans, one might expect them to produce large amounts of information (high entropy) without inducing additional processing costs (low KL divergence).



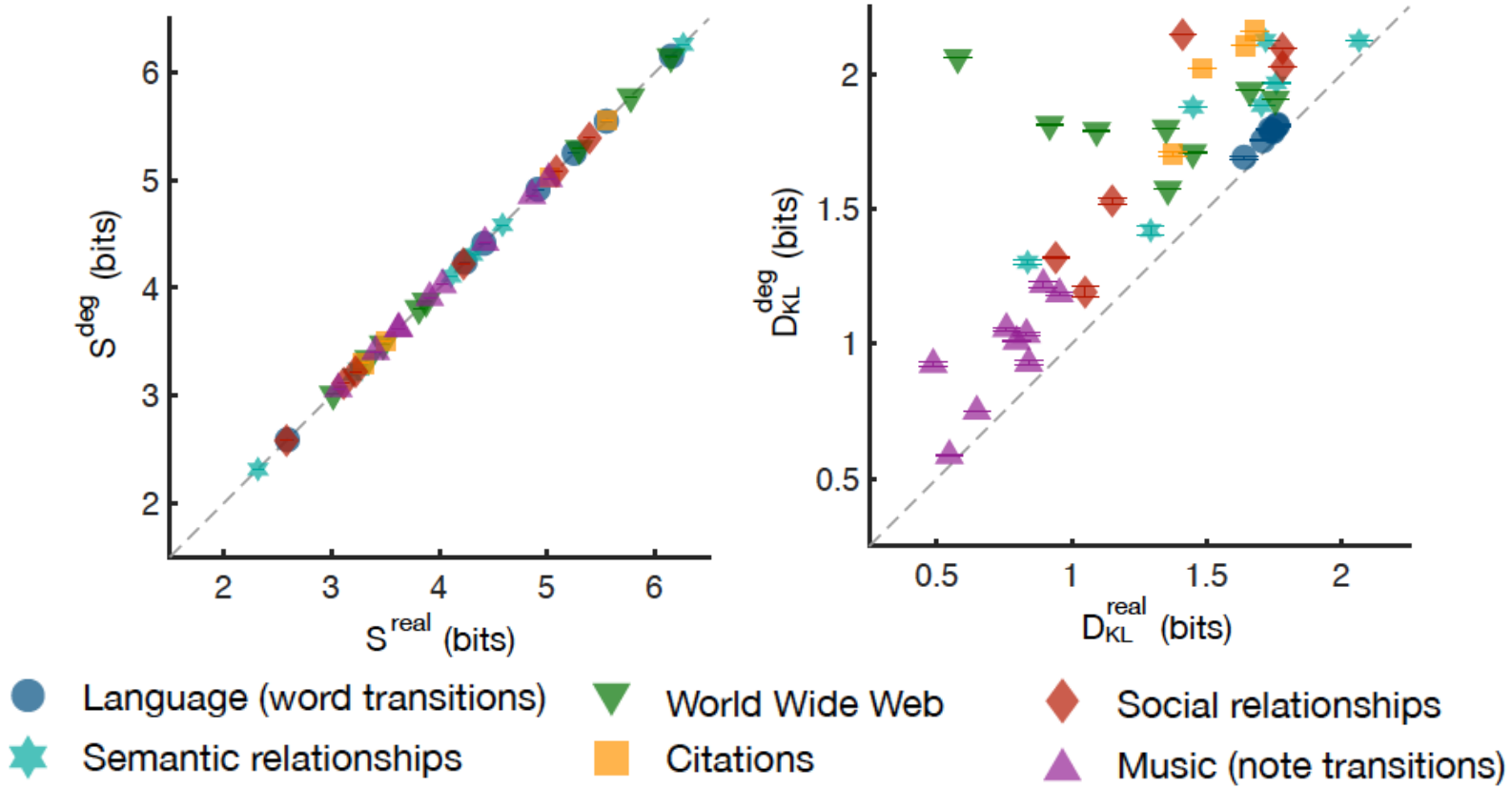
Classes of real networks differ in perceived information



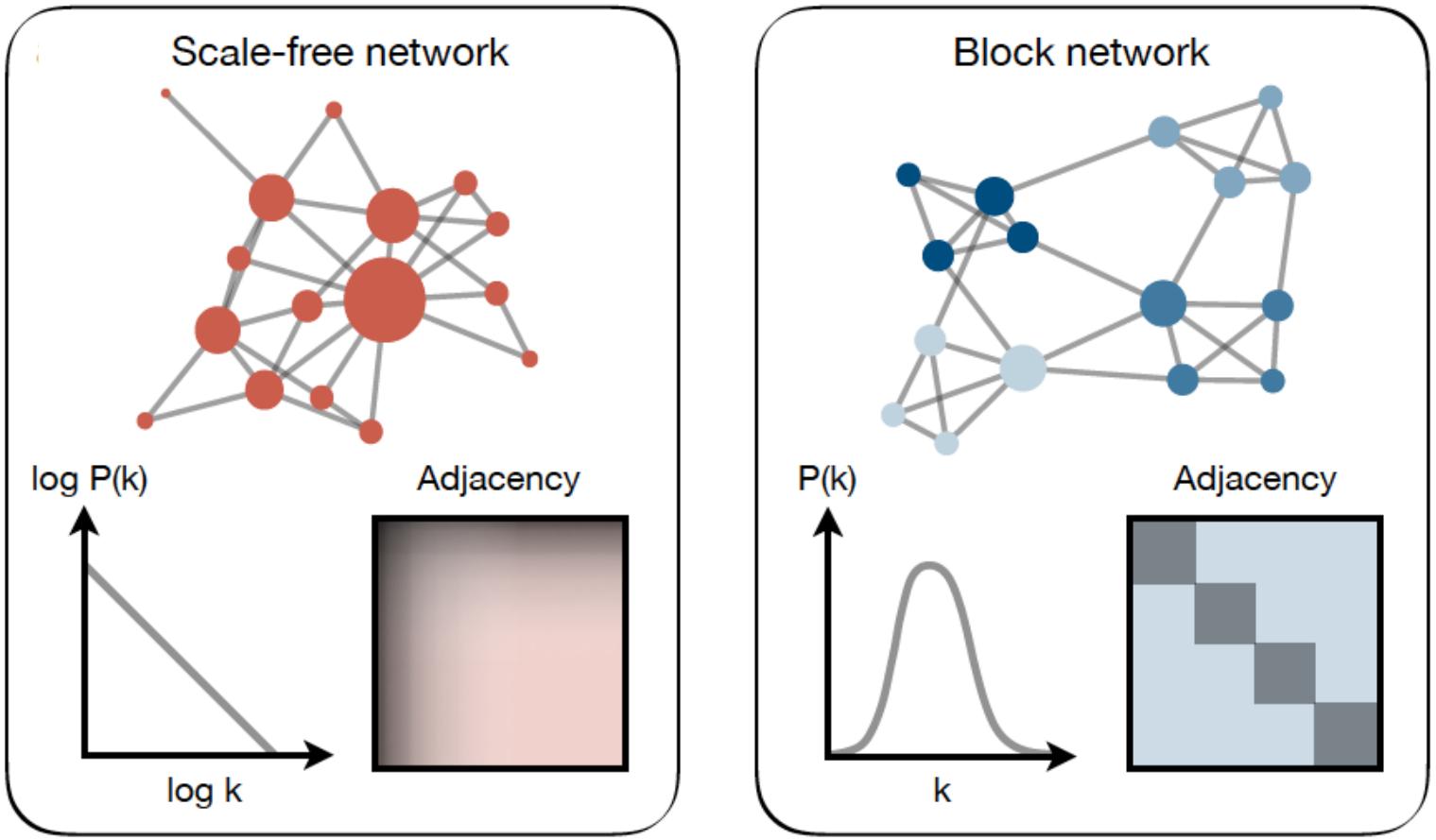
Different network types exhibit these properties to varying degrees.

KL divergence is lower than in degree-preserving null models

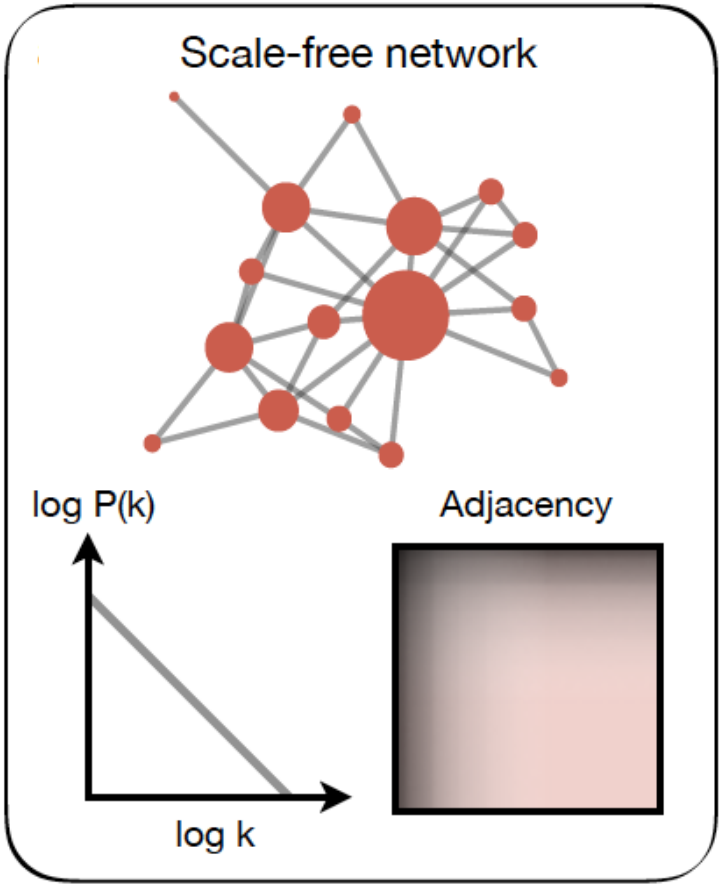
Compared to entropy-preserving null models, real world networks are closer to human expectations, revealing that the KL divergence is driven by the networks' higher-order topology.



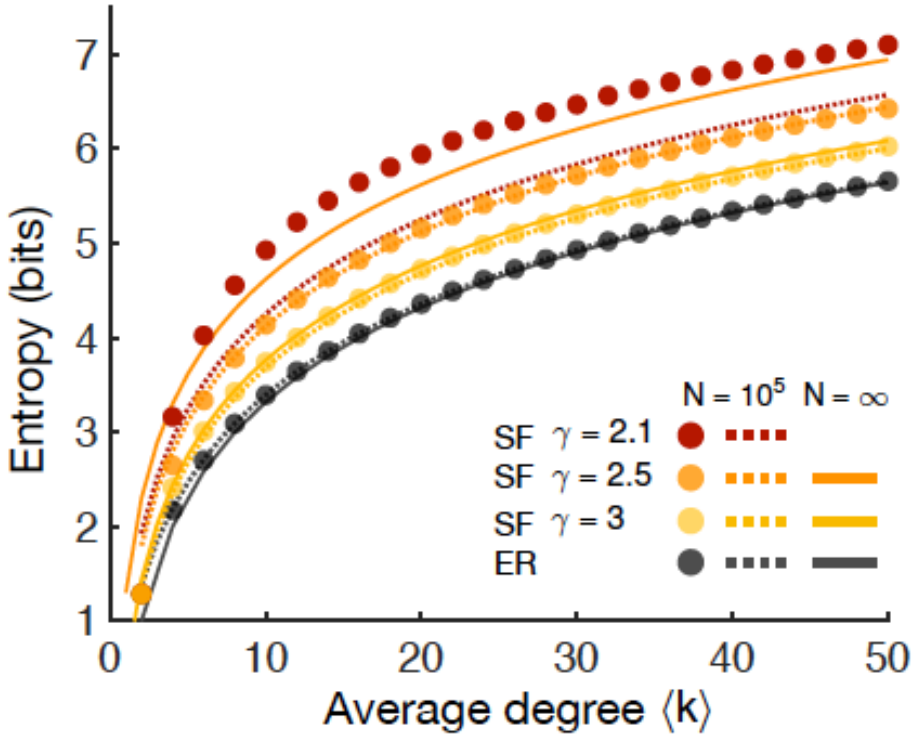
What structural features give rise to these properties?



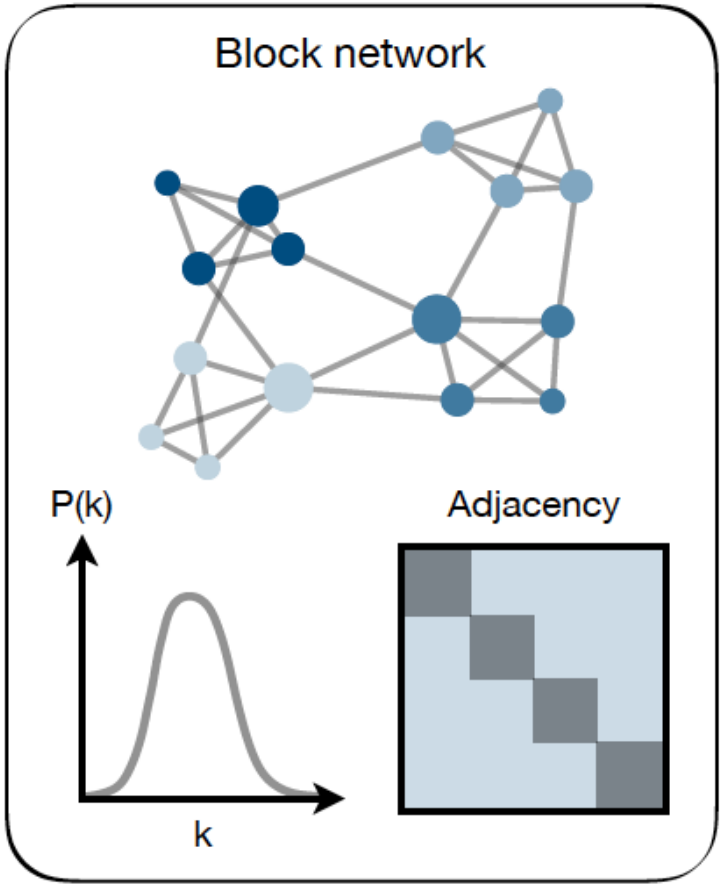
Tuning entropy by broadening the degree distribution



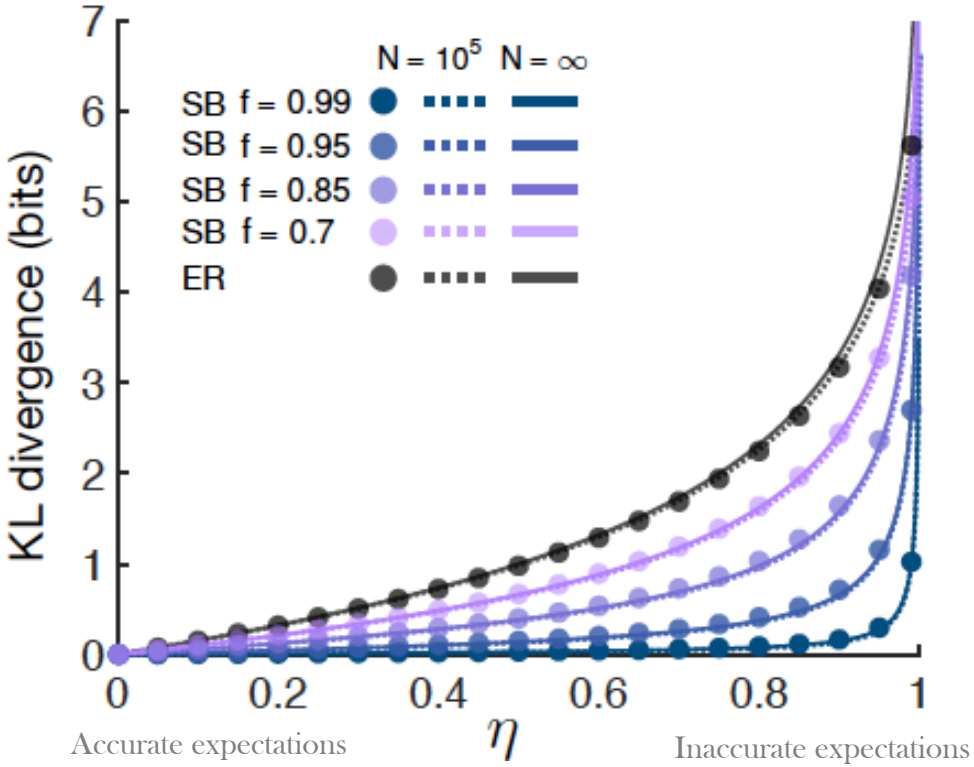
Even after controlling for the density of a network, the entropy is larger for networks with more drastic variation in node degree.



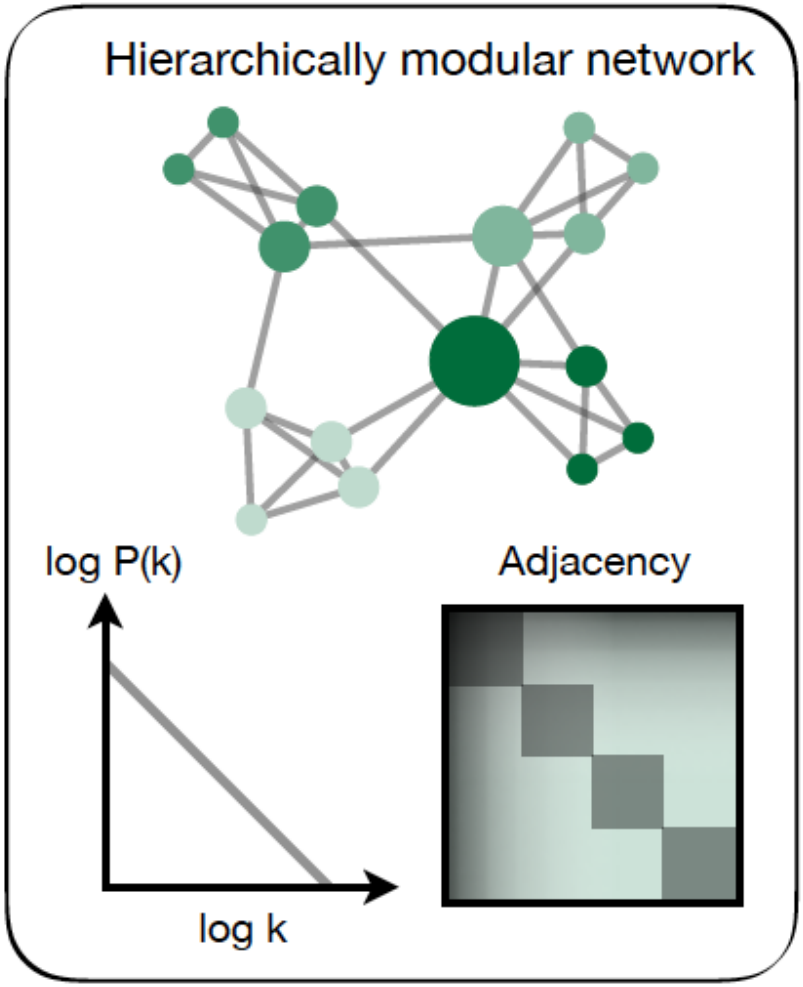
Tuning KL divergence by increasing modularity

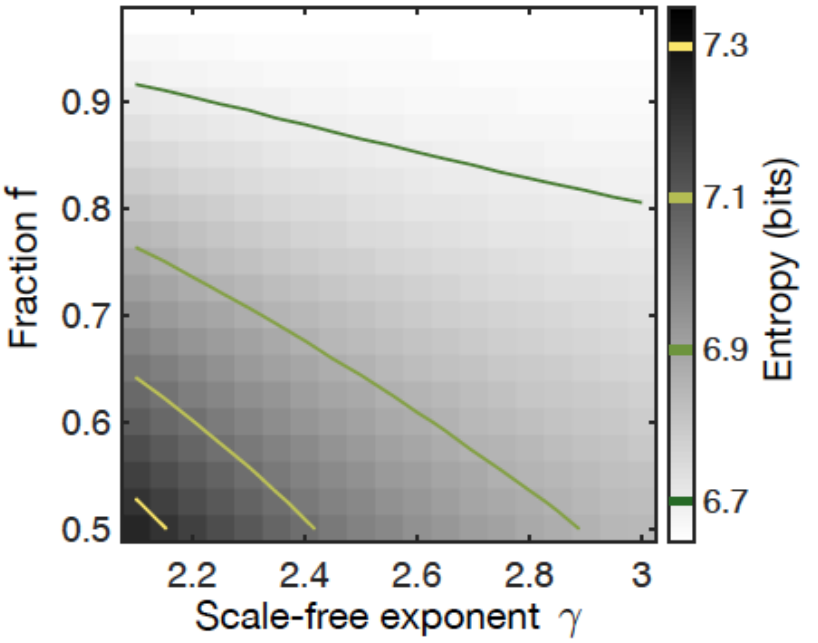


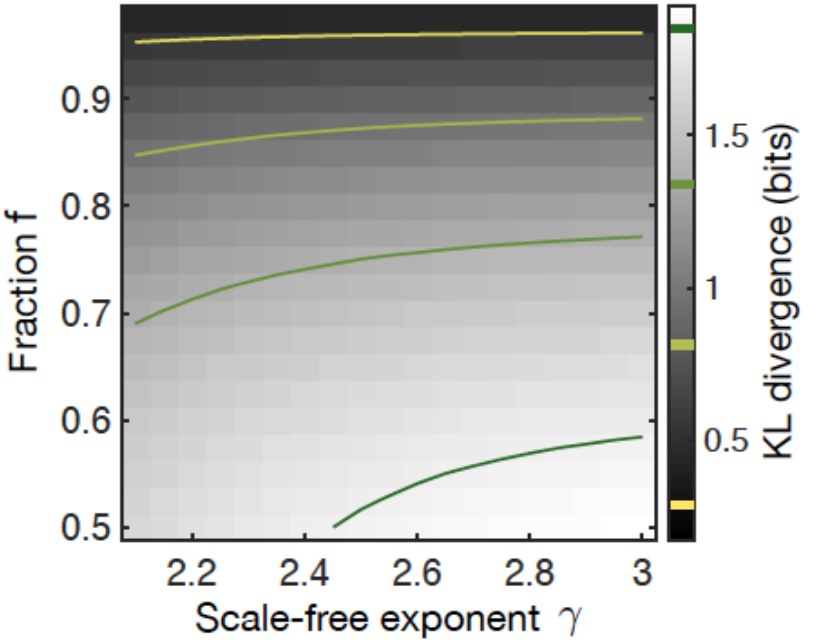
After fixing the density of a network and the accuracy of an observer's representation, modular organization can help to decrease the amount of information that a human is required to process.



A reasonable solution to the problem of maximizing entropy and minimizing KL divergence







Summary

- Humans can learn the architecture of networks underlying a continuous stream of information.

Summary

- Humans can learn the architecture of networks underlying a continuous stream of information.

- Humans display expectations that diverge from the true network, indicating that they are not performing simple maximum likelihood computations.

Summary

- Humans can learn the architecture of networks underlying a continuous stream of information.
- Humans display expectations that diverge from the true network, indicating that they are not performing simple maximum likelihood computations.
- Free energy principle that the brain minimizes errors & computational complexity produces a memory distribution (Boltzmann) that is consistent with empirical measurements from traditional psychology tests.

Summary

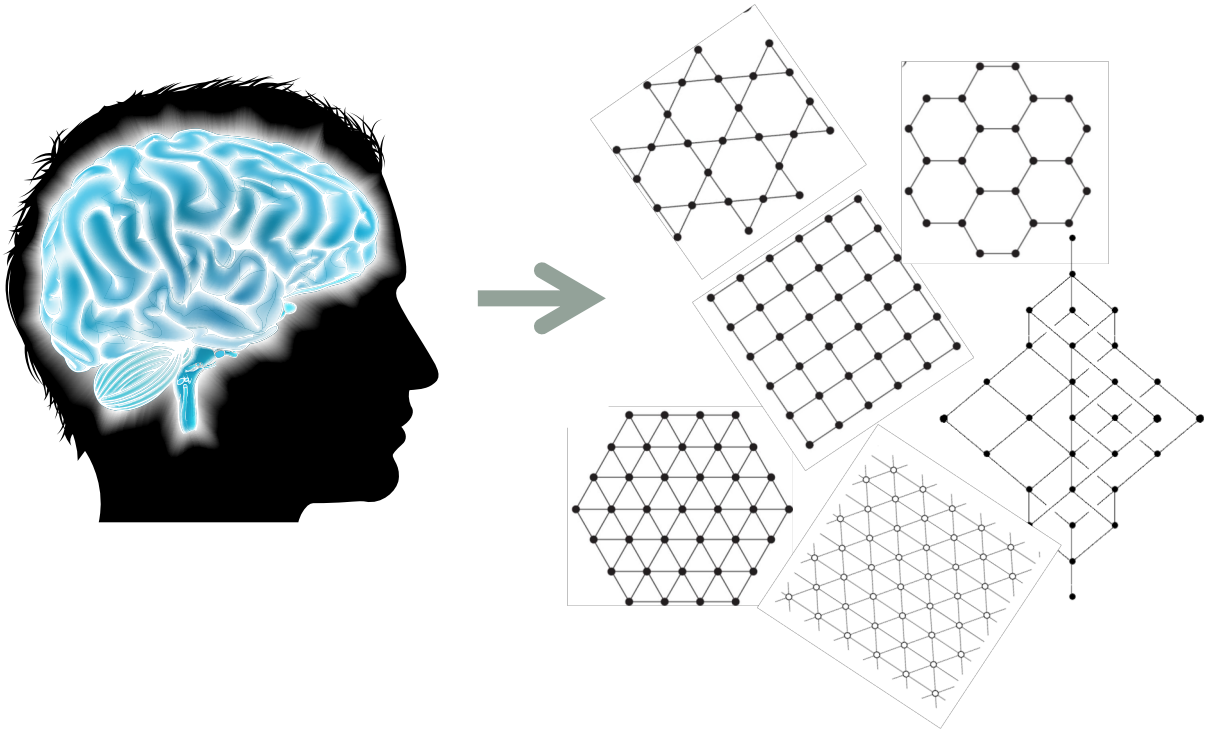
- Humans can learn the architecture of networks underlying a continuous stream of information.
- Humans display expectations that diverge from the true network, indicating that they are not performing simple maximum likelihood computations.
- Free energy principle that the brain minimizes errors & computational complexity produces a memory distribution (Boltzmann) that is consistent with empirical measurements from traditional psychology tests.
- Real networks are organized to communicate large amounts of information (high entropy), and to do so efficiently (low KL divergence from human expectations).

Summary

- Humans can learn the architecture of networks underlying a continuous stream of information.
- Humans display expectations that diverge from the true network, indicating that they are not performing simple maximum likelihood computations.
- Free energy principle that the brain minimizes errors & computational complexity produces a memory distribution (Boltzmann) that is consistent with empirical measurements from traditional psychology tests.
- Real networks are organized to communicate large amounts of information (high entropy), and to do so efficiently (low KL divergence from human expectations).
- Efficient communication is supported by heterogeneous and modular structure, hinting at a general explanation for the hierarchical organization observed in many communication systems.

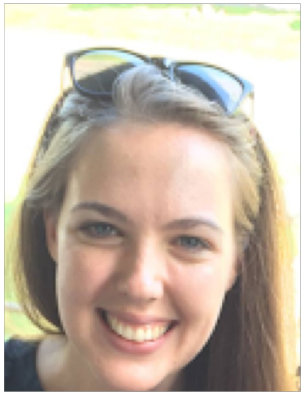
Open Questions

- Does the optimally learnable graph have a topology that is common in well-written papers? (Chai et al. 2019 JCN) Or in well-written textbooks? (Christianson et al. In Prep)
- Can we move beyond random walks to incorporate non-Markovian dynamics?
- Can humans learn more than graphs behind a time series? Can we learn chaotic attractors from time series and how? (Lu arXiv:1807.05214)



Acknowledgments

TEAM: Abigail Poteshman, Andrew Murphy, Ann Sizemore Blevins, Ari Kahn, Arun Mahadevan, Cedric Xia, Christopher Lynn, Dale Zhou, David Lydon-Staley, Eli Cornblath, Erin Teich, Graham Baum, Harang Ju, Jason Kim, Jeni Stiso, Karol Szymula, Keith Wiley, Lia Papadopolous, Lindsay Smith, Lorenzo Caciagli, Mark Choi, Maxwell Bertolero, Pranav Reddy, Urs Braun, Ursula Tooley, Xiaosong He, Zhixin Lu.

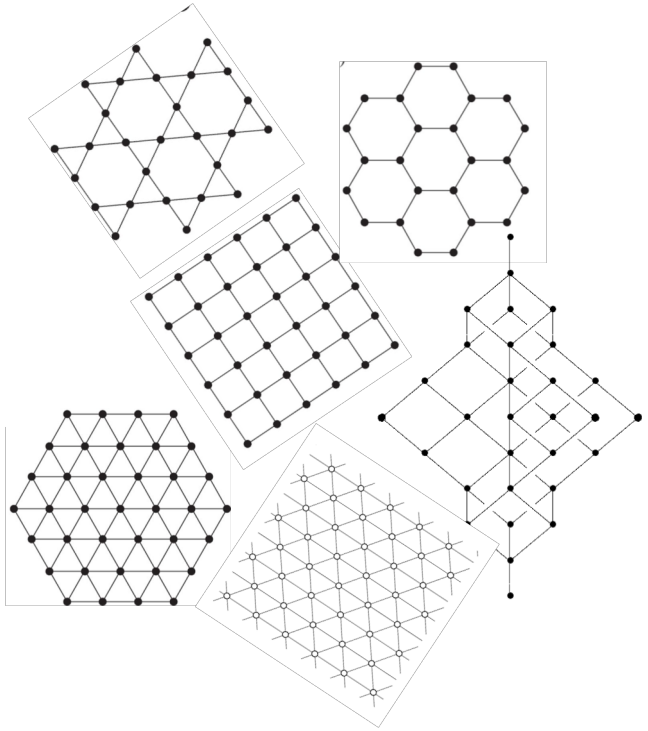


Dr. Lizz Karuza,
Now Asst. Prof of Psychology at PSU

Ari Kahn,
Graduate Student in Neuroscience

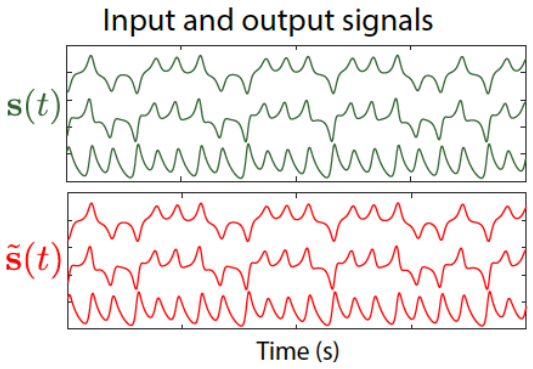
Chris Lynn,
Graduate Student in Physics

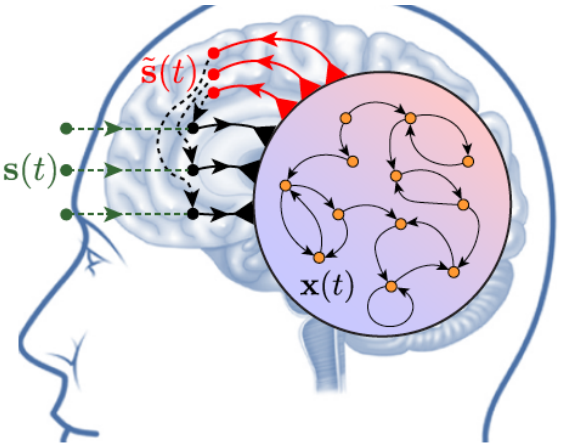
Thanks!

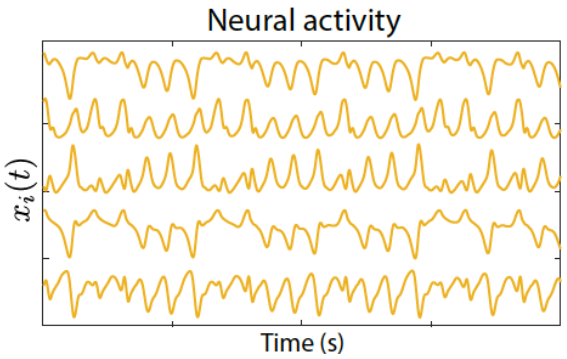


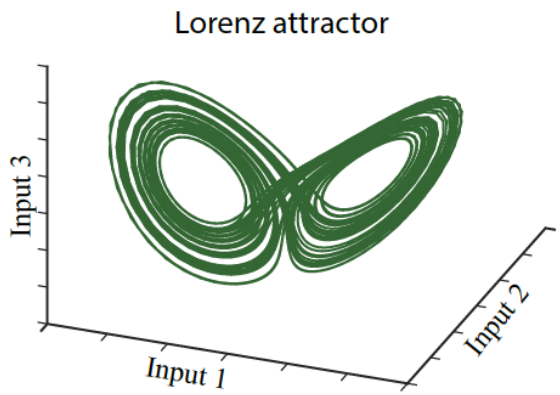
Dynamical systems framework for implicit rule learning

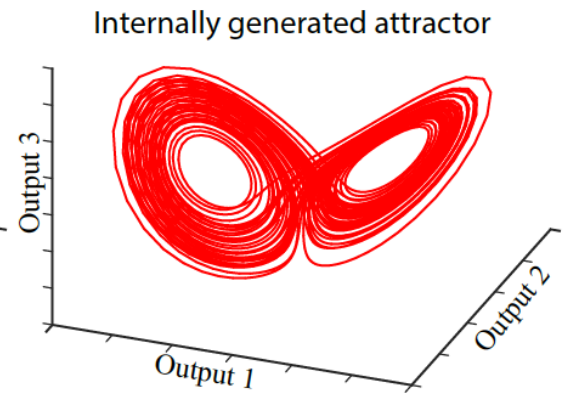
Goals: Acknowledge complex dynamics of real stimuli; couple with neural mechanisms

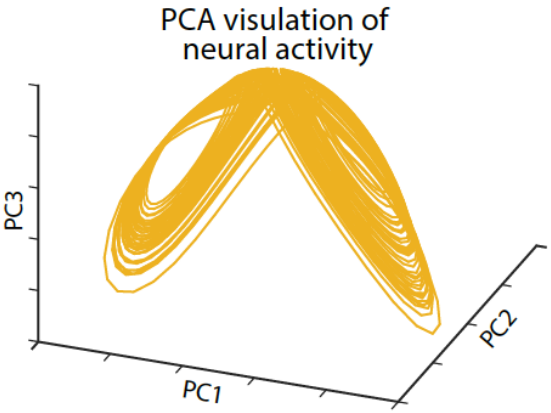








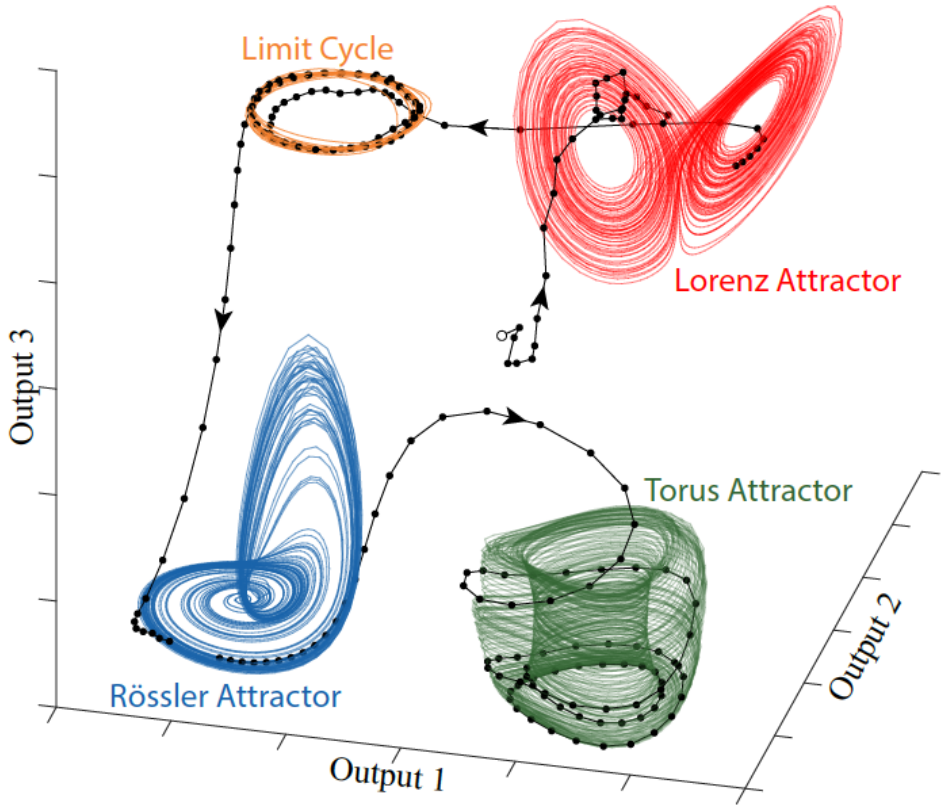


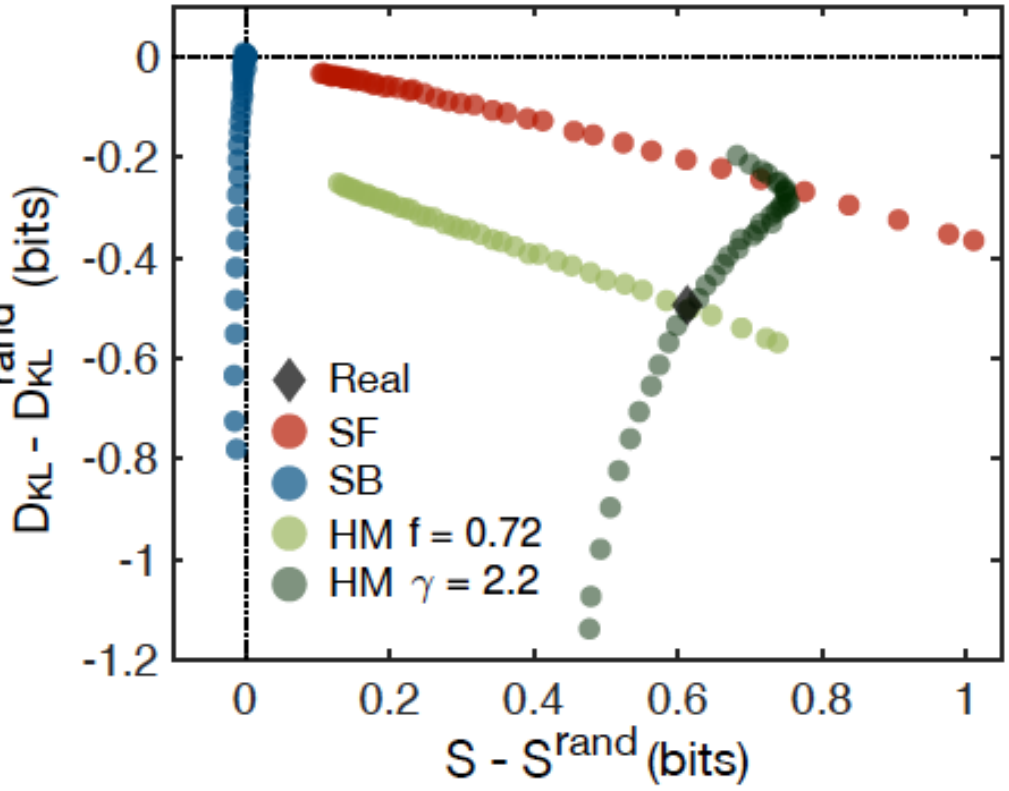


Implicit rule learning of chaotic attractors

Our approach allows us to demonstrate and theoretically explain the emergence of five distinct phenomena reminiscent of functions observed in natural neural systems:

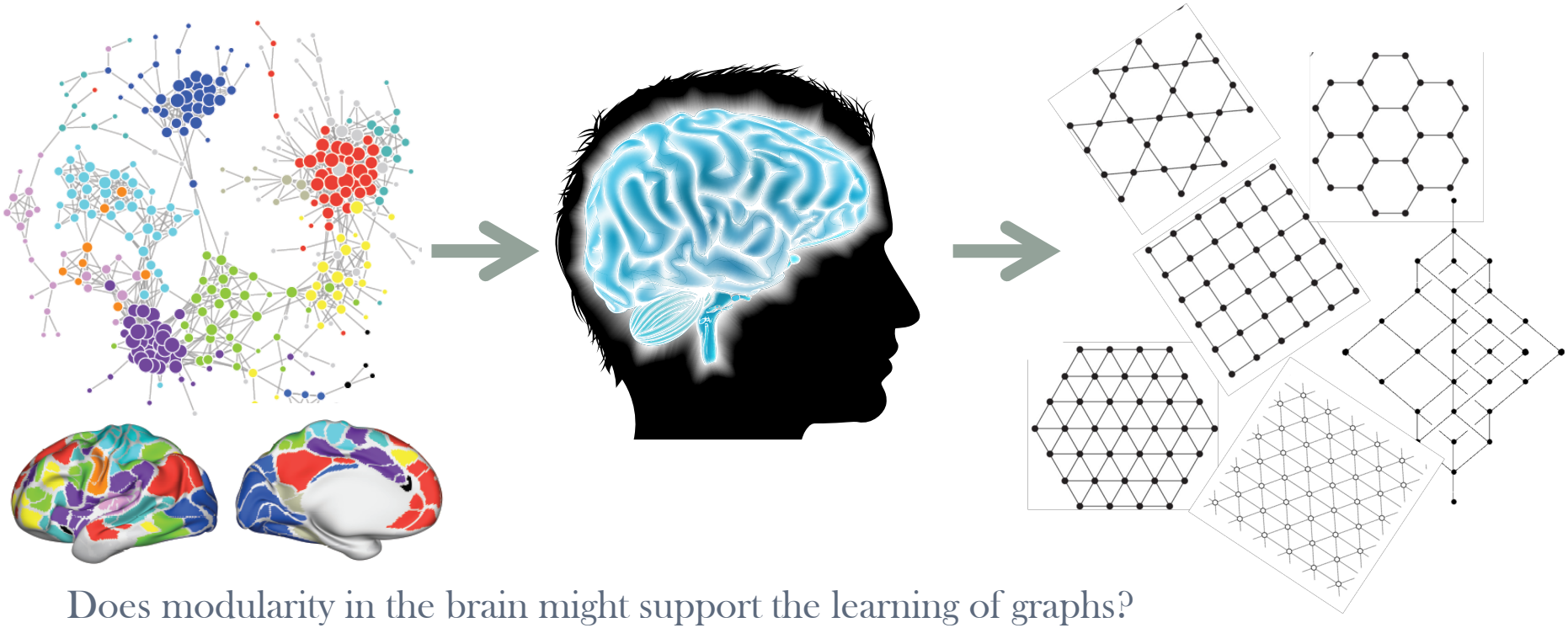
(i) learning the dynamical rule underlying a chaotic attractor purely from time series,
(ii) generating new streams of stimuli from a chaotic attractor,
(iii) switching stream generation among multiple learned chaotic attractors, either spontaneously or in response to external perturbations,
(iv) inferring missing data from sparse observations of the chaotic attractor, and
(v) deciphering superimposed input from different chaotic attractors.




0
-0.2
-0.4
-0.6
-0.8
-1
-1.2
$D_{KL} - D_{KL}^{rand}$ (bits)
0
0.2
0.4
0.6
0.8
1
$S - S^{rand}$ (bits)
Real
SF
SB
HM f = 0.72
HM γ = 2.2

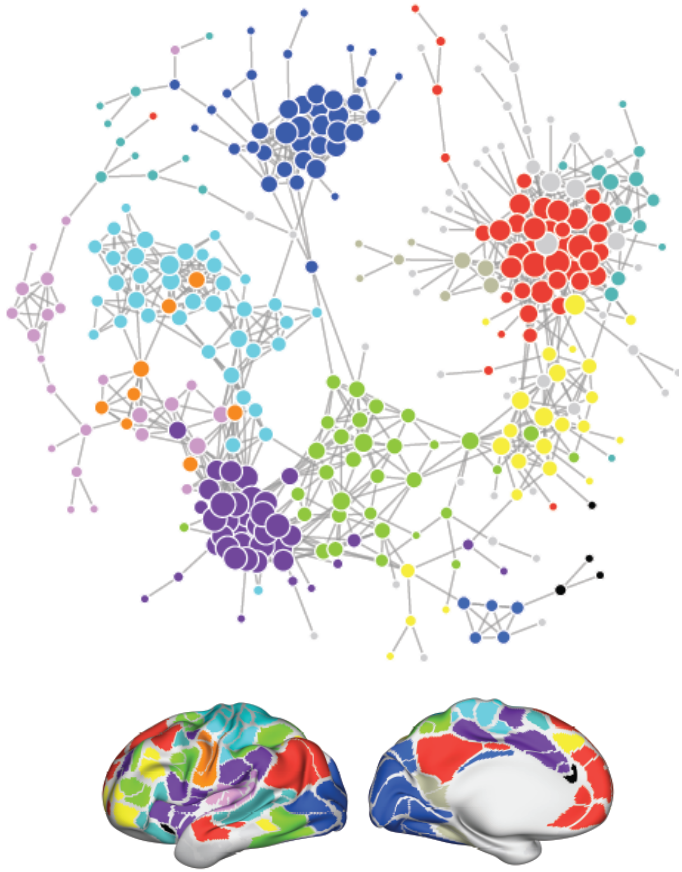
Brain network processes supporting learning

"Mind thinks itself because it shares the nature of the object of thought; for it becomes an object of thought in coming into contact with and thinking its objects, so that mind and object of thought are the same." Aristotle, Metaphysics, Book XII, 7, 1072 b 20



Does modularity in the brain might support the learning of graphs?
Might differences in modularity explain differences in the ability to learn?

Theoretical & Computational Challenges



Challenge: Parsimoniously representing and describing complex connectivity patterns.

- Network models
- Bassett, Zurn, Gold (2018) *Nat Rev Neuro*

Challenge: Detecting modular structure in network models of brain connectivity.

- Modularity maximization (NP Hard)
- Meunier et al. (2009) *NeuroImage*

$$Q = \sum_{ij} [A_{ij} - P_{ij}] \delta(\sigma_i \sigma_j)$$

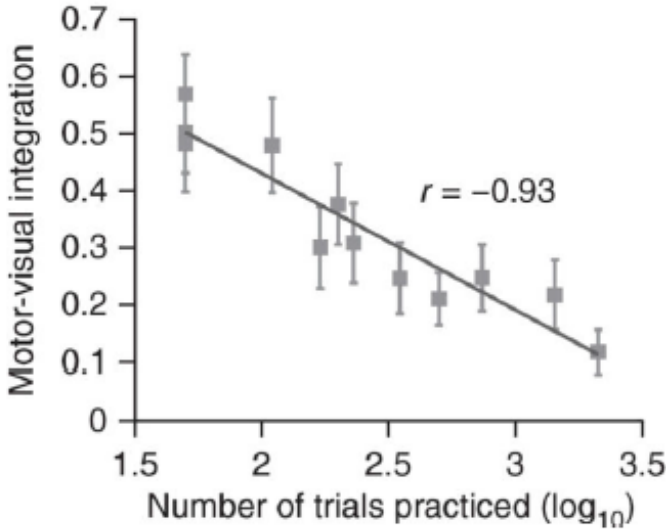
Challenge: Detecting evolving modules.

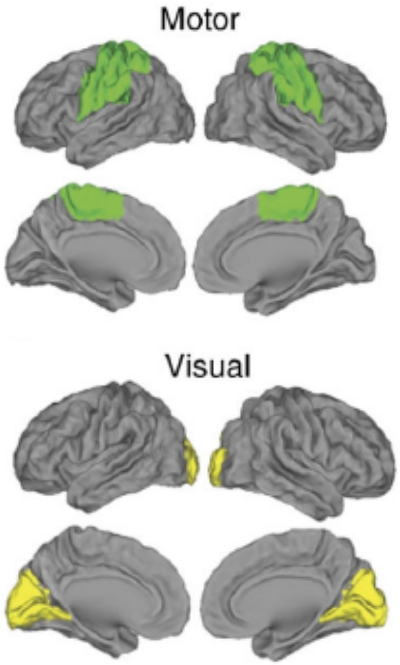
- Multilayer modularity maximization
- Mucha et al. (2010) *Science*

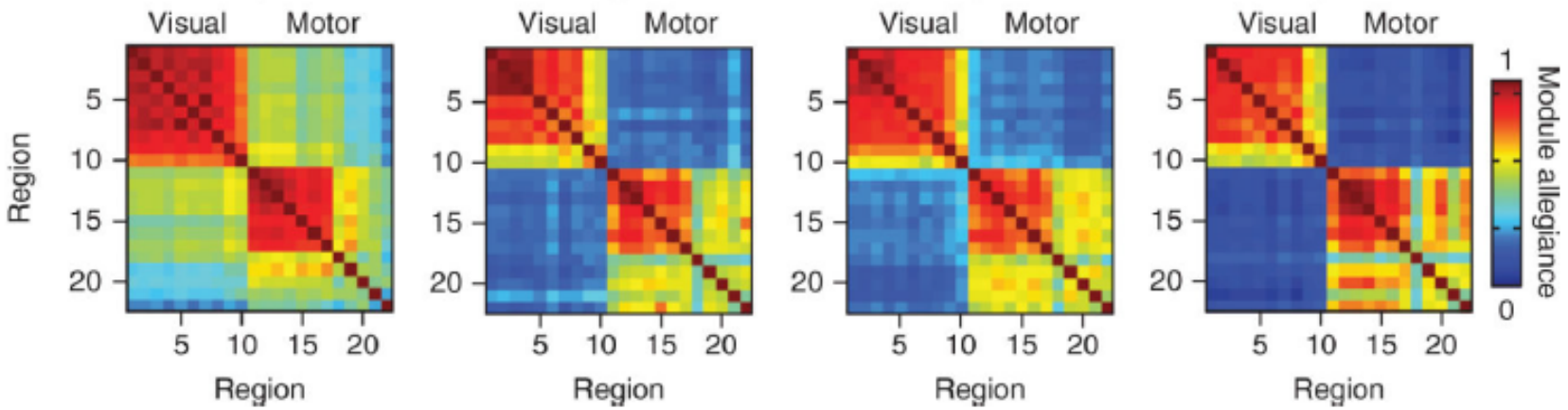
$$Q_{multi} = \frac{1}{2\mu} \sum_{ijls} [(\mathcal{A}_{ijl} - \gamma_l \mathcal{P}_{ijl}) \delta_{lm} + \omega_{jlm} \delta_{ij}] \delta(c_{il}, c_{jm})$$

Module Autonomy in Sequence Learning

The coherence between motor and visual modules decreased markedly with training, suggesting a growing autonomy.

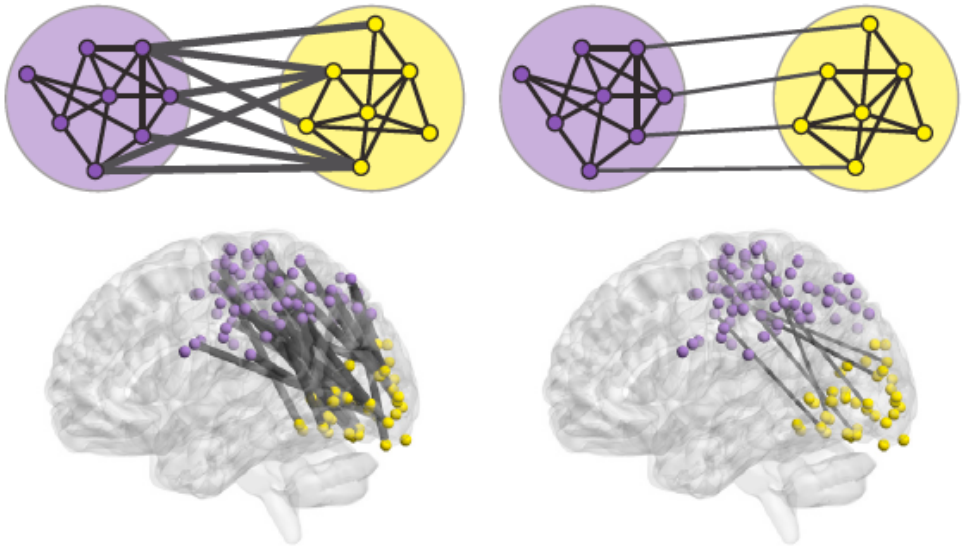






General relevance of modularity for learning

Hypothesis: Networks that can flexibly adapt are those with greater modularity.



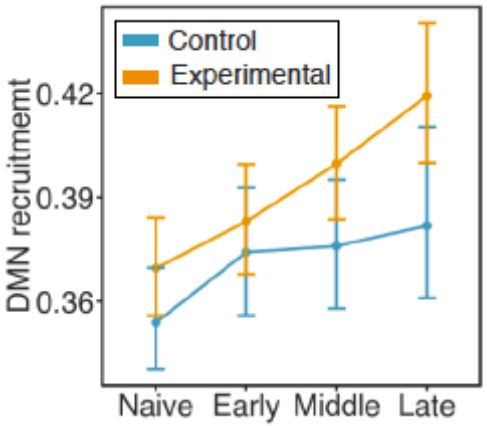
Constrained modules; slow learner

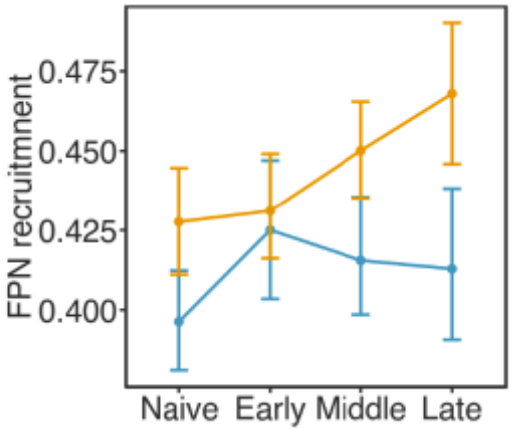
Flexible modules; Fast learner

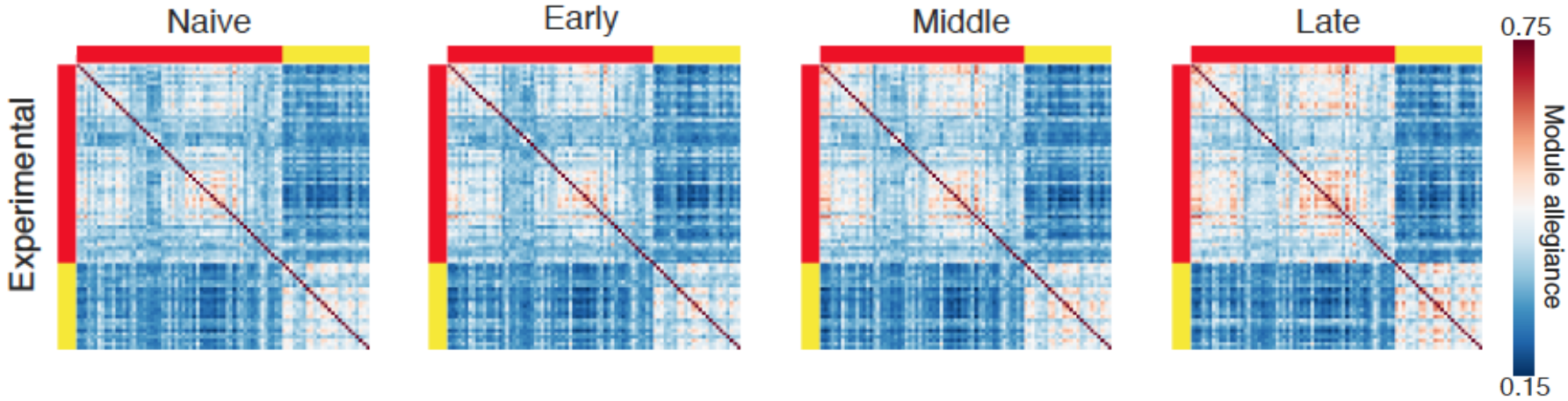
1. Flexible modules support swifter learning over 3 days; *Bassett et al. 2011 PNAS*
2. Swift learning is associated with flexible segregation of modules over 6 weeks; *Bassett et al. 2015 Nature Neuroscience*
3. Segregation of modules at rest predicts learning 6 weeks in the future; *Mattar et al. 2018 NeuroImage*

Module strengthening with dual n-back training

Large-scale collections of brain regions (modules) change in their coherent activity with training, providing coarse-grained markers of function.

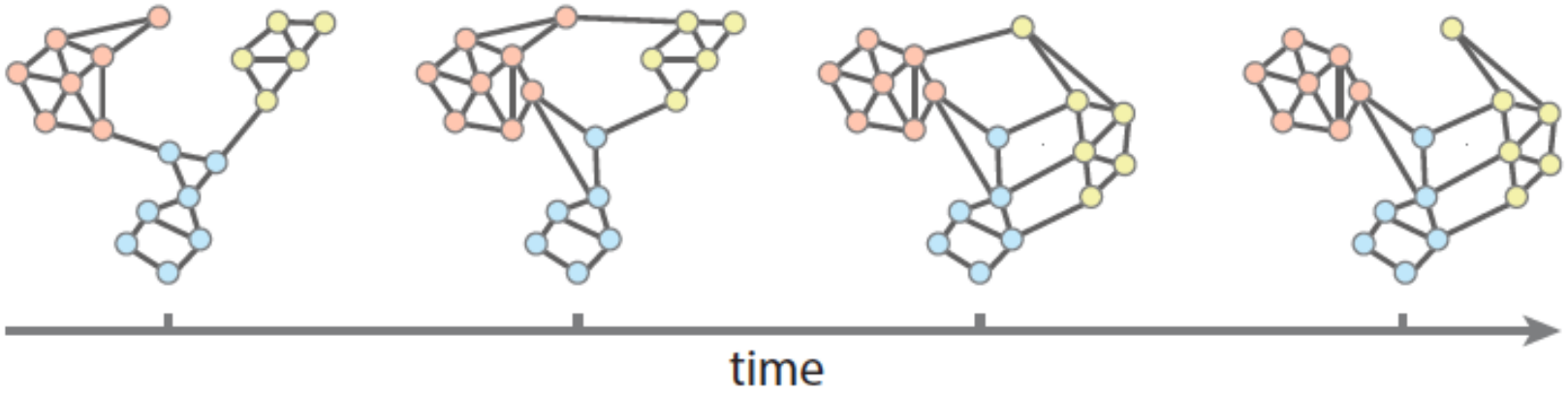






Finc et al. (2019) *In Preparation*

Flexible modularity supports learning (& executive function)



Flexibility in network modules is predicts individual differences in:

Visuo-motor learning (Bassett et al. 2011 *PNAS*)
Cognitive flexibility (Braun et al. 2015 *PNAS*)
Working memory (Braun et al. 2015 *PNAS*)
(Shine et al. 2016 *Neuron*)
Learning rate (Gerraty et al., 2018, *J Neurosci*)
Future learning (Mattar et al. 2018 *NeuroImage*)
Planning & reasoning (Pedersen et al. 2018 PNAS)

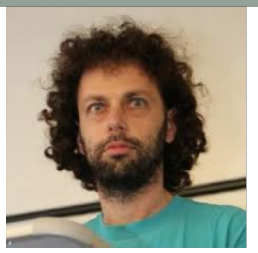
Medication (Braun et al. 2016, *PNAS*)
Positive mood (Betzel et al. 2017 *Sci Rep*)
Amount of sleep (Pedersen et al. 2018 PNAS)

Searching for design rules

Why do some types of learning induce changes in system strength and others induce changes in inter-system connectivity? Could we create a parameterization of task families that would allow us to manipulate these two phenotypes smoothly and continuously?

What induces network reconfiguration? Who is able to respond to training with greater network reconfiguration and why? What constraints determine what sorts of reconfiguration are easier or harder than others? How much energy does it take to induce a network reconfiguration?


Urs Braun


Mason Porter


Marcelo Mattar


Peter Mucha


Rick F. Betzel


Scott Grafton


Raphael Gerraty


Daphna Shohamy

Karuza et al. 2016 *Trends in Cognitive Science*

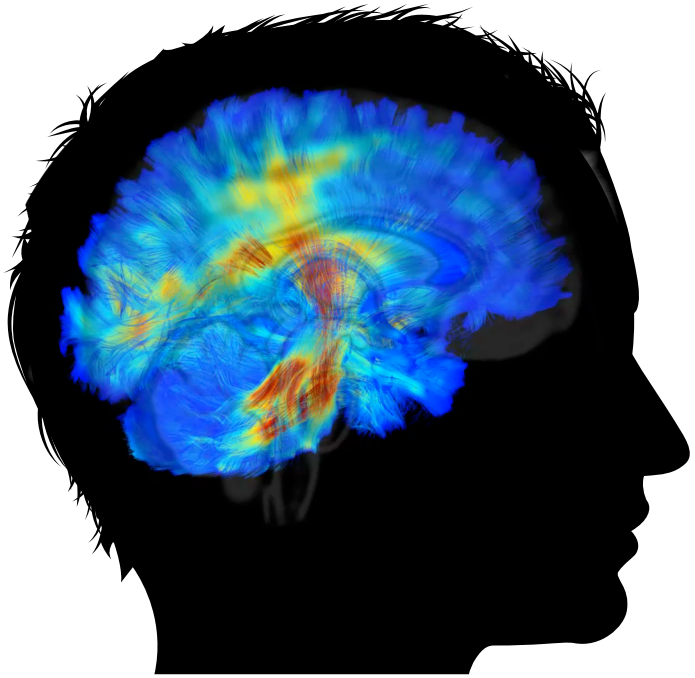
Danielle S. Bassett

"Now if there was a becoming of every changeable thing, it follows that before the motion in question another change must have taken place in which that which is capable of being changed or of causing change had its becoming."

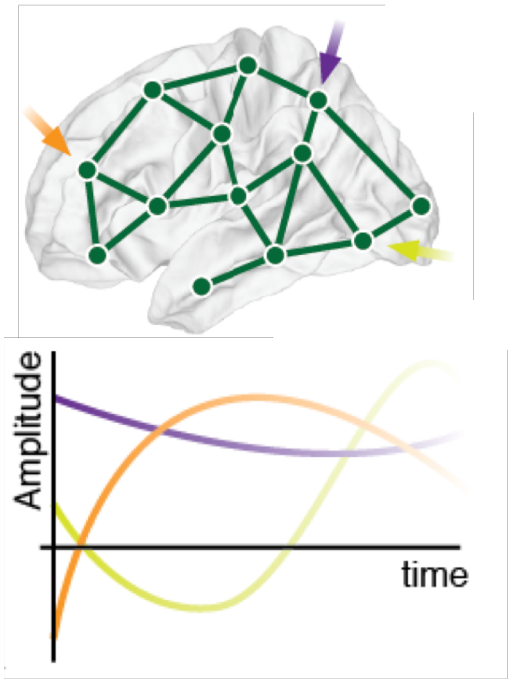
Aristotle, *Physics* VIII.I, 251a9

Constraining Nature of Network Architecture

Can build a theoretical model from data that predicts the changing, the becoming, and the causing of change? Or … how the brain's activity can be altered by a perturbative signal?



What we have: A network of structural links empirically measured by neuroimaging.



What we seek: A theory for how a change in activity in one region affects activity in other regions.

Formalizing the Problem of Network Control

- Neural processes can be approximated by linearized generalizations of nonlinear models of cortical circuit activity (Galan 2008; Honey et al. 2009).
- We consider a noise-free linear discrete-time and time-invariant network model:

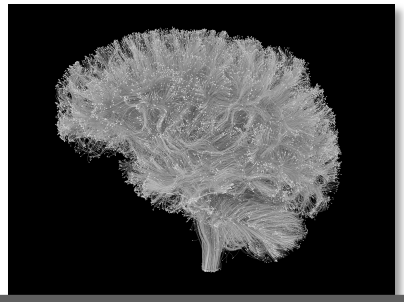
$$x(t+1) = Ax(t) + B_K u_K(t)$$

Number of regions being controlled

State of brain regions over time

Weighted adjacency matrix

Control energy



Gu et al. (2015) *Nature Communications*; Tang et al. (2017) *Nature Communications*

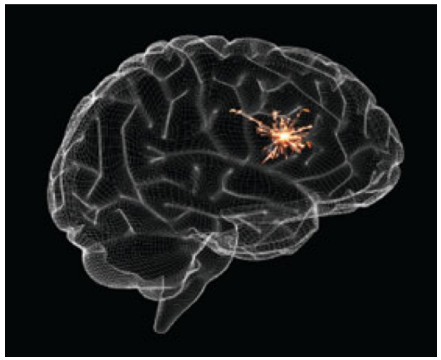
Is the brain theoretically controllable?

How controllable the network is can be estimated using the smallest eigenvalues of the T-steps controllability Gramian:

$$W_{\mathcal{K},T} = \sum_{\tau=0}^{T-1} A^{\tau} B_{\mathcal{K}} B_{\mathcal{K}}^{\mathsf{T}} (A^{\mathsf{T}})^{\tau}$$

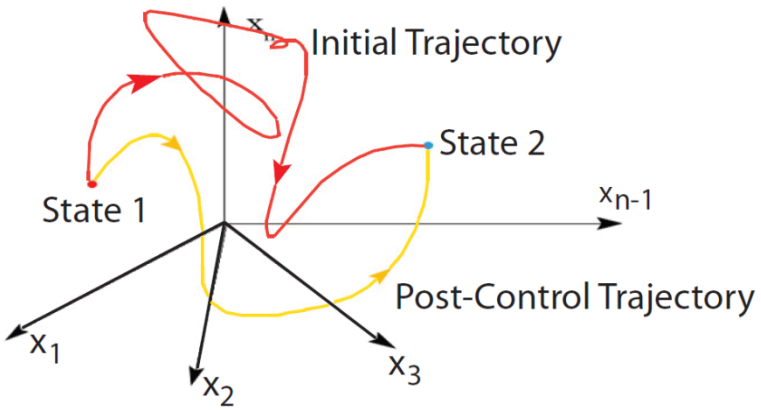
For brain networks, this value was small: 2.5 x 10^(-23)

- Practically extremely hard to control



Types of driver nodes

- Which regions of the brain are most efficient or most difficult to control?

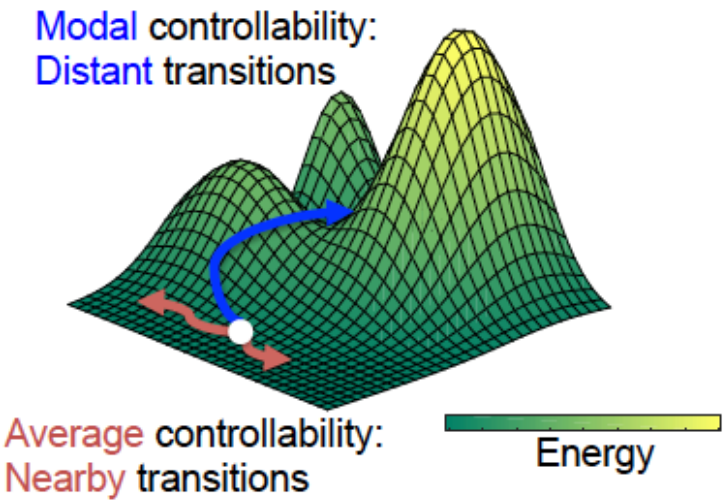


A couple control strategies:

1. **Average Controllability:** Steer to many easily reachable states

2. **Modal Controllability:** Steer to few difficult to reach states

Average and modal control



$$x(t+1) = Ax(t) + B_{\mathcal{K}} u_{\mathcal{K}}(t)$$

$$W_{\mathcal{K},T} = \sum_{\tau=0}^{T-1} A^{\tau} B_{\mathcal{K}} B_{\mathcal{K}}^{\mathsf{T}} (A^{\mathsf{T}})^{\tau}$$

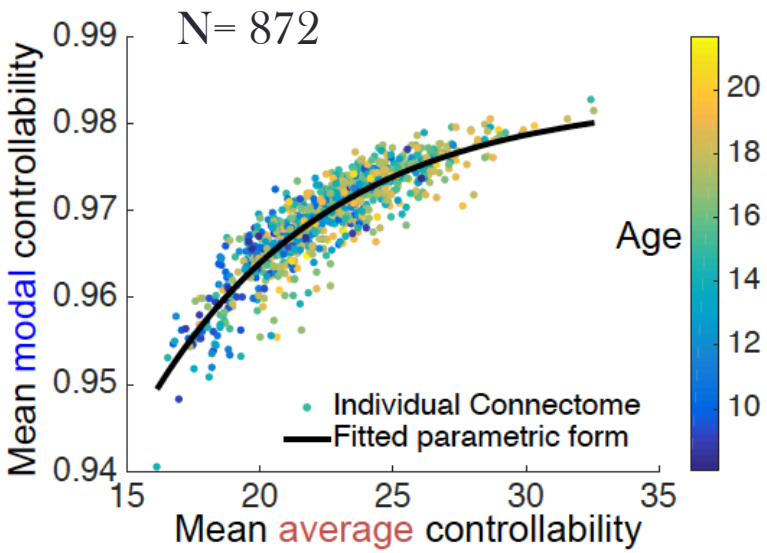
Average: Trace(W_K^{-1}))

Modal: Let v_j be the j^{th} eigenvector of A with eigenvalue λ_j. Then if v_{ij} is small, then the j^{th} mode is poorly controllable from node i. Define $\phi_i = \sum_{j=1}^{N} (1 - \lambda_j^2(A)) v_{ij}^2$ as a scaled measure of controllability of all N modes from region i.)

Gu et al. (2015) *Nature Communications*, Betzel et al. (2016) *Scientific Reports*

Network control as a model for cognitive control

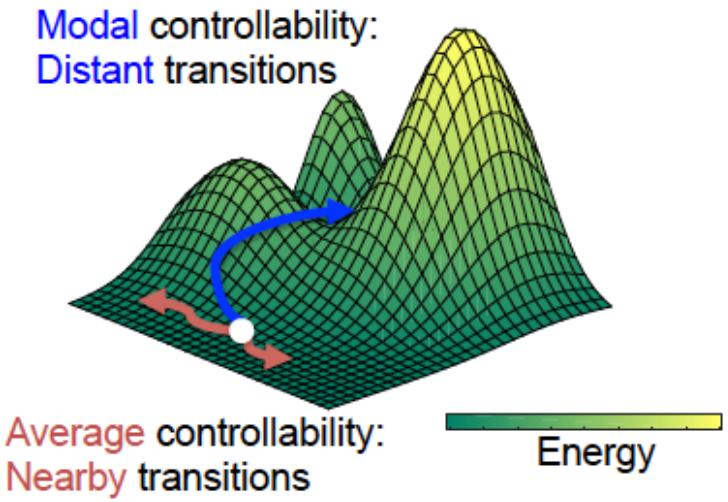
- Different brain regions have more or less average/modal controllability, indicating differential capacity to alter whole-brain dynamics
  (Gu et al. 2015 *Nature Communications*)
- The capacity of brain regions to exert control grows as children develop
  (Tang et al. 2017 *Nature Communications*)
- Network controllability is correlated with impulsivity, a measure of executive function
  (Cornblath et al. 2018 NeuroImage)
- The energy required for control decreases with age, in concert with increasing executive function
  (Ciu et al. 2018 In Revision)

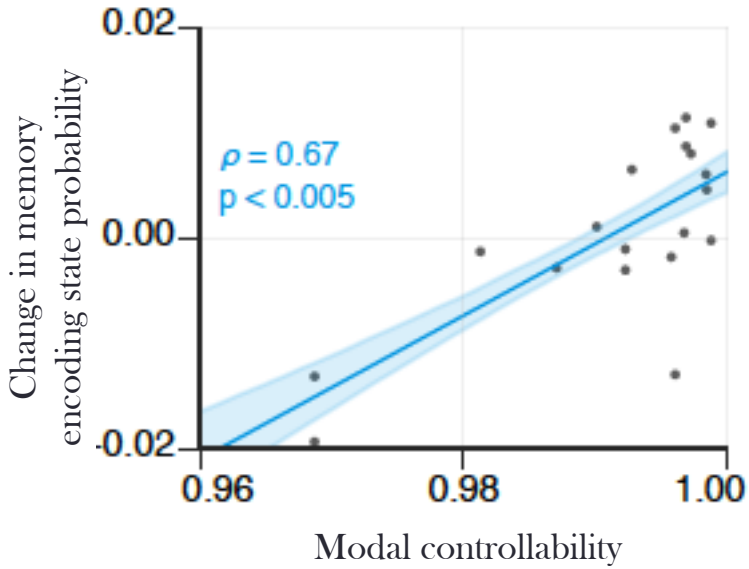


Together, these results suggest that our theory is a useful marker of how the brain enacts control to change network function.

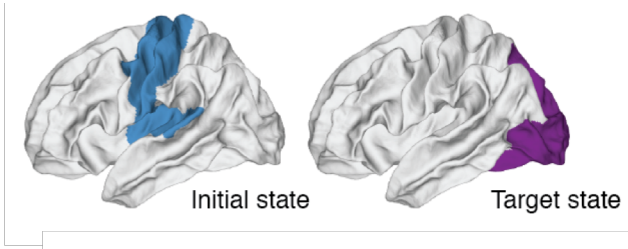
Extending to exogenous control (stimulation)

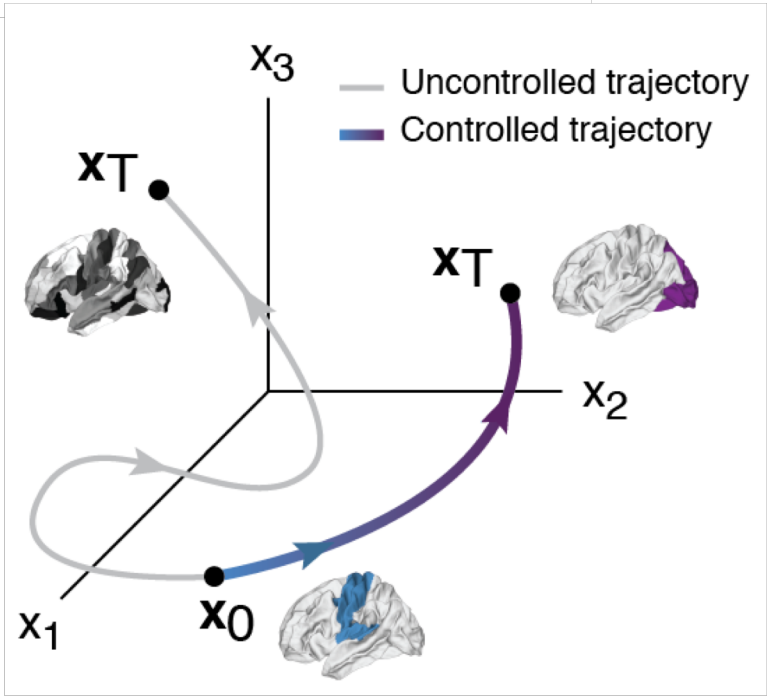
Preliminary work suggests that stimulation to modal controllers pushes the brain into better memory encoding states.





Precise control of specific state transitions





What we want

- Finite time, Finite energy,
- Multi-point control
- Initial state, Target state

Define model of network dynamics.

$$x(t+1) = Ax(t) + B_{\mathcal{K}} u_{\mathcal{K}}(t)$$

Define a cost function penalizes energy and distance of x(t) from the target state.

$$\min_{\mathbf{u}} \int_0^T (\mathbf{x}_T - \mathbf{x})^T (\mathbf{x}_T - \mathbf{x}) + \rho \mathbf{u}_{\mathcal{K}}^T \mathbf{u}_{\mathcal{K}}$$

Betzel et al. (2016) *Scientific Reports*; Gu et al. (2017) *Neuroimage*
Stiso et al. https://doi.org/10.1101/313304 (In Revision)

Open questions

What is it about certain network topologies that makes them easier or harder to control? (Kim et al. 2018 *Nature Physics*) Does the answer to this question help us to understand time scales of control, such as transient versus persistent control, which may be altered in certain patient groups? (Tang et al. 2019, Submitted)

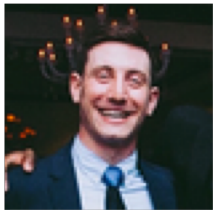
How does control of brain state transitions relate to network reconfiguration exactly? How might brain states relate to neural representations, housing information about the world or our model of the world? What rules constrain the evolution of neural representations during learning? (Tang et al. (2019) *Nature Neuroscience*, In Press)

Jason Z. Kim

Ankit Khambhati

Shi Gu

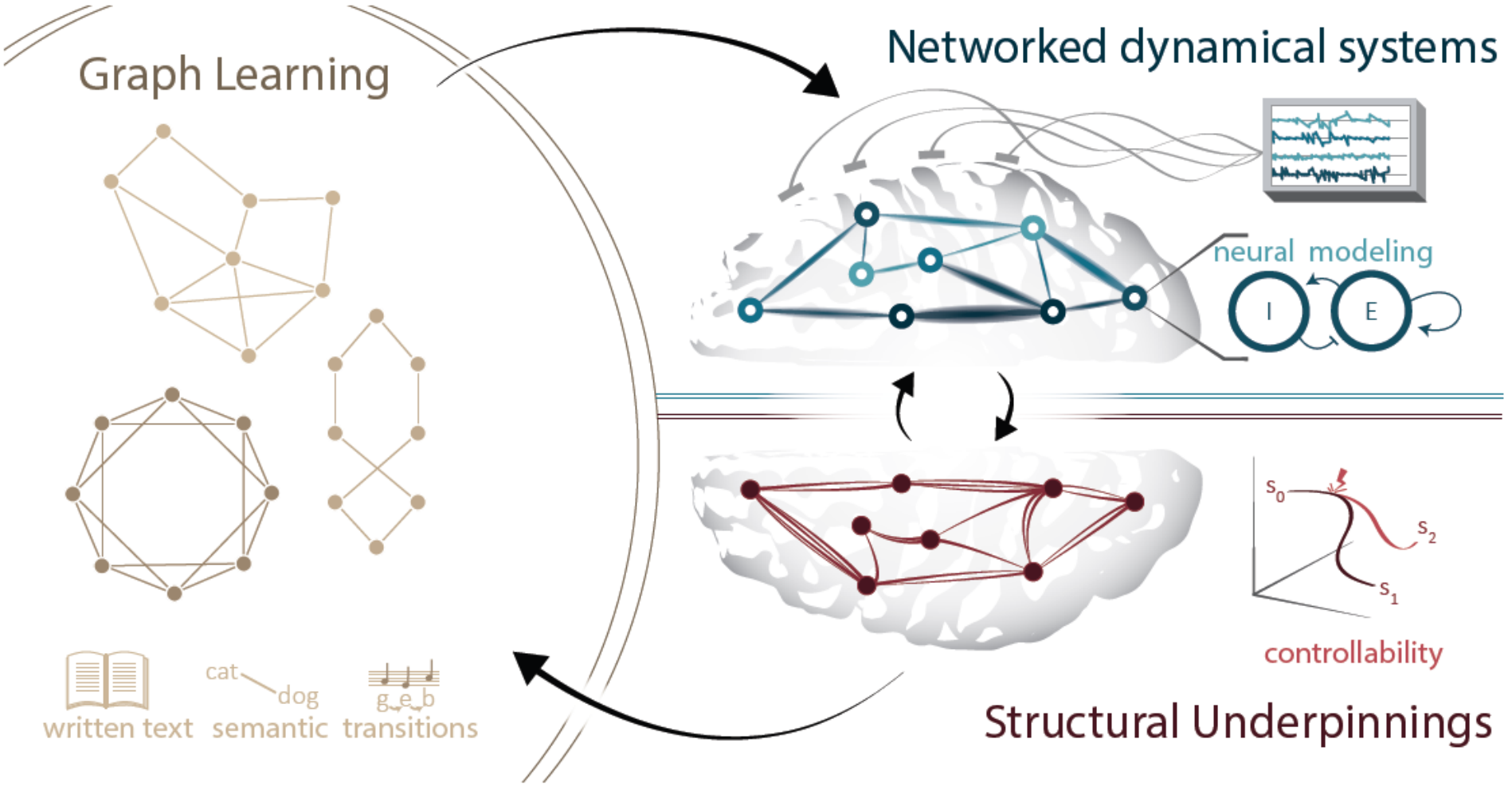
Rick F. Betzel



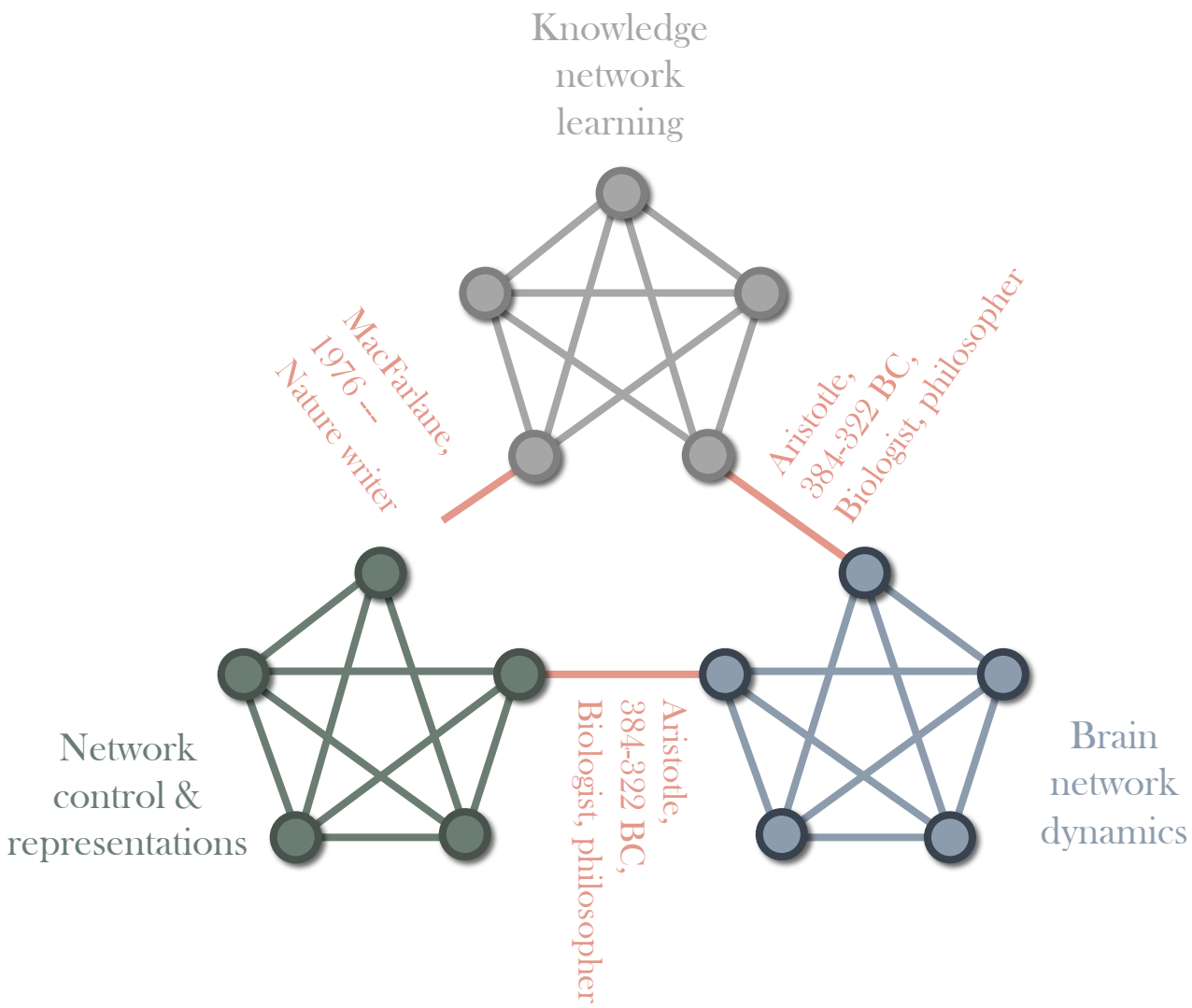
Eli Cornblath

Jeni Stiso

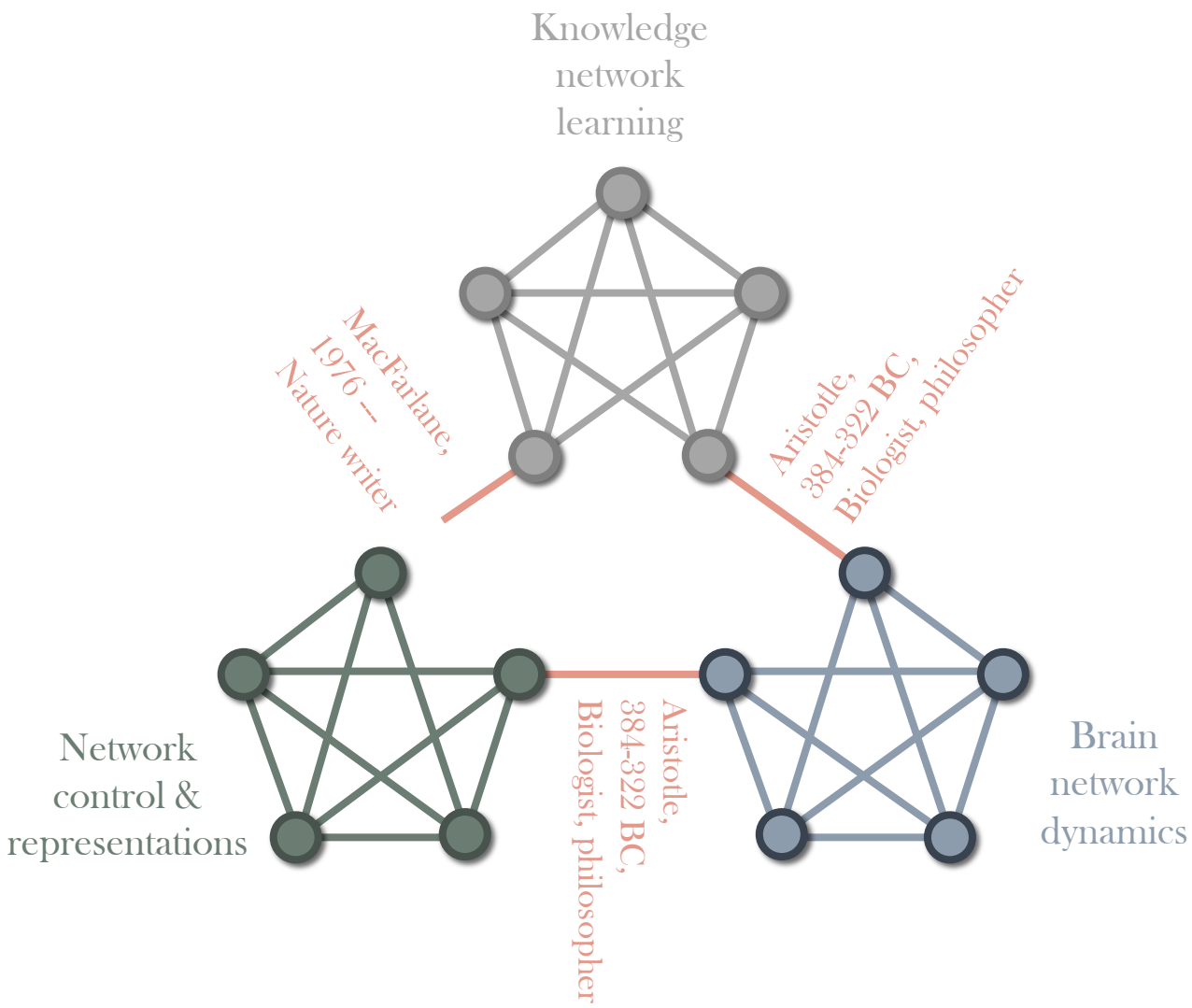
Summary & Future Directions

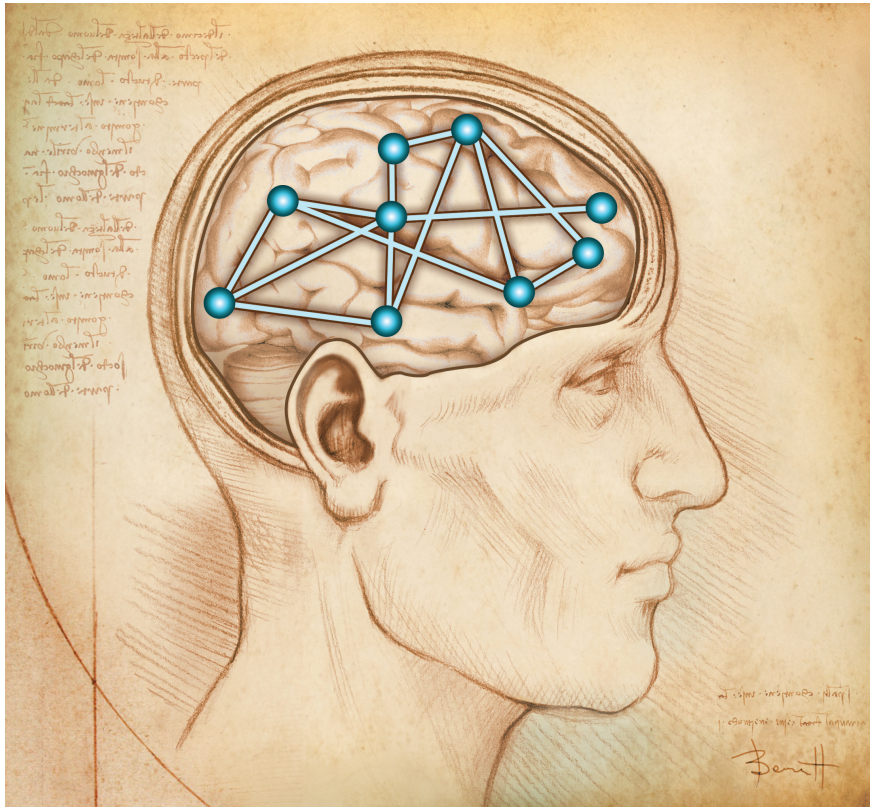


Networks thinking themselves



Networks thinking themselves

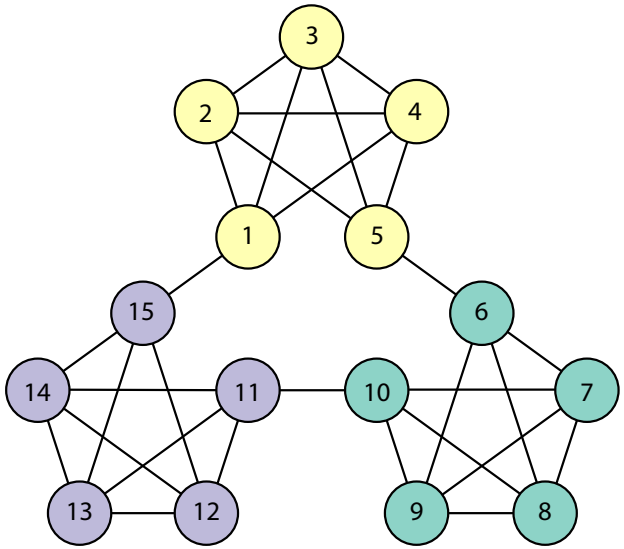




What do we know about this problem?

From work in the field of statistical learning and the study of artificial grammars, we know that humans are sensitive to transition probabilities.

Then what would we predict about the graph below?



Because every edge has a transition probability of 0.25, human expectations should be equivalent across all transitions, and thus so should human reaction times.