Quantifying Human Priors over Abstract Relational Structures

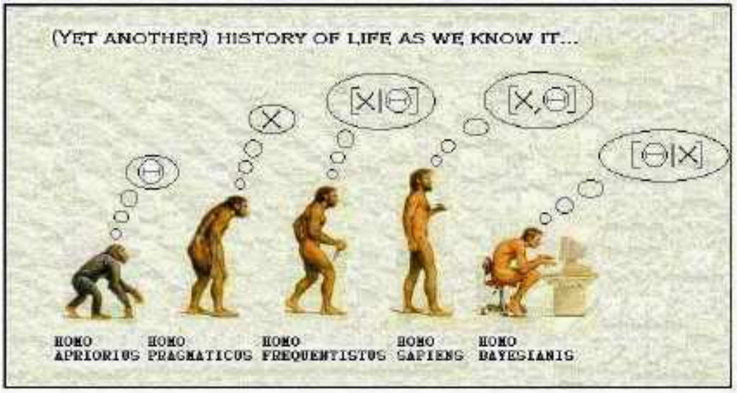
Gecia Bravo-Hermsdorff

Ph.D. Defense Presentation

February 9, 2020

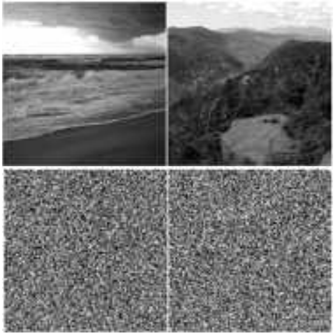
Organisms approximate Bayes' rule

Scarcity and chance are essential facts of life.
$\Longrightarrow$ Organisms need to effectively handle with uncertainty.

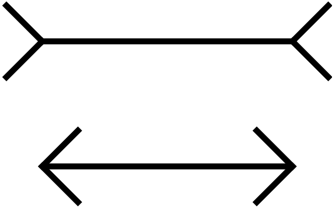


Correct expression: $p(\Theta = h|X = d) = \frac{p(X{=}d|\Theta{=}h)p(\Theta{=}h)}{\sum_{h\in\Theta} p(X{=}d|\Theta{=}h)p(\Theta{=}h)}$

Priors color our perception



(*Graham & Field, 2009*)



(*Howe & Purves, 2004*)

…might there be similar “illusions” with respect to our priors over the structure of tasks?

Hofstadter’s Law: It always takes longer than you expect, even when you take into account Hofstadter’s Law.

Priors cannot be both exhaustive and efficient

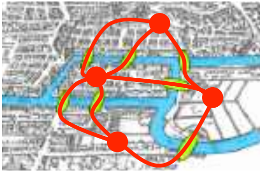
Priors over the structure of tasks
are necessary for learning and generalization.

Formulating a tractable problem: from arbitrary tasks to navigation and social graphs

Navigation

Kaliningrad, Russia

⟶

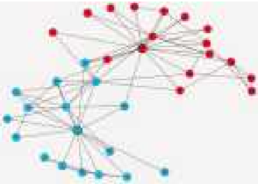


The seven bridges of Königsberg

Social

facebook

⟶



Zachary karate club

Outline

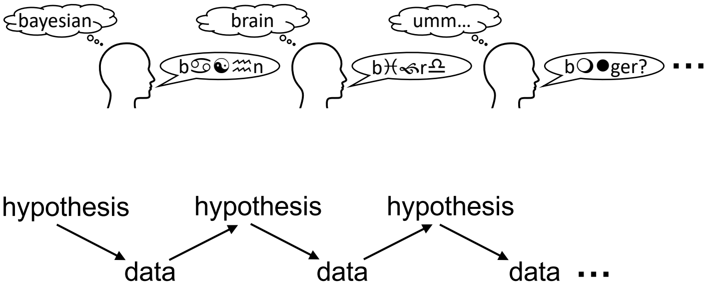
1. MCMCP:
   Experimental framework for obtaining priors over graphs

2. Introducing Graph Cumulants:
   Principled Framework
   for Quantifying the Structure of Priors over Graphs

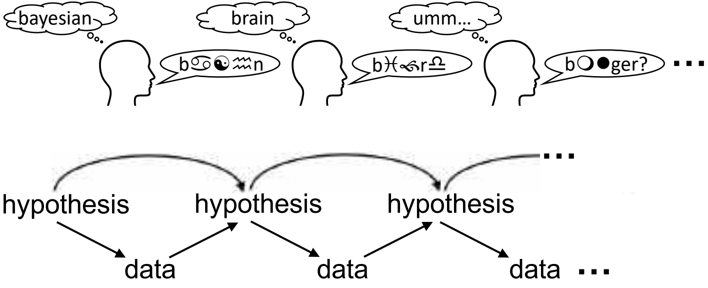
3. The Structure of Human Priors over Navigation and Social Tasks

MCMCP: Experimental framework for obtaining priors over graphs

MCMCP: a Bayesian model of the “telephone game”



MCMCP: a Bayesian model of the “telephone game”



(Kalish & Griffiths, 2007)

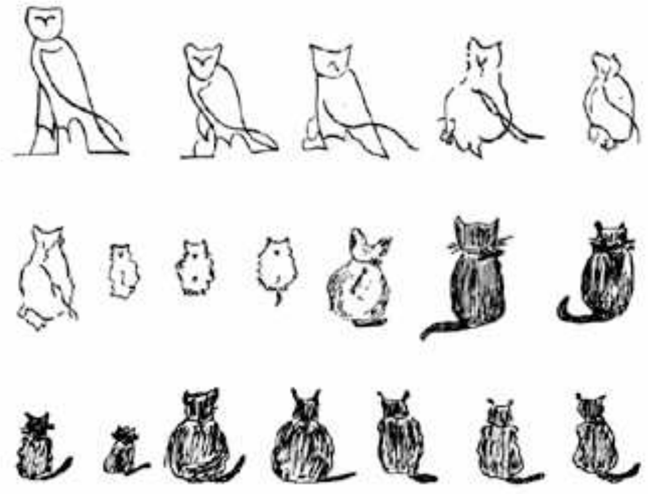
MC over hypotheses:

$$T_{ij} = \sum_{x} p(h_{t+1} = i | d_{t+1} = x) p(d_{t+1} = x | h_t = j)$$

Under certain assumptions...

this process converges to people’s prior.

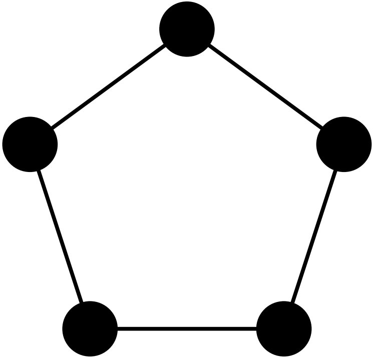
A historical example



(Bartlett, 1932)

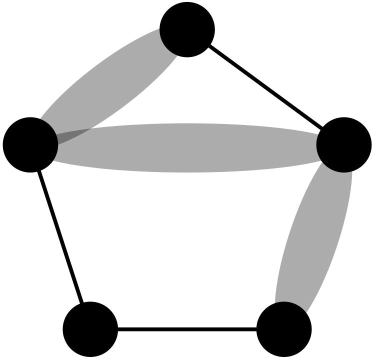
Algorithm for MCMCP experiments over graphs

1. Initialize "task graph"



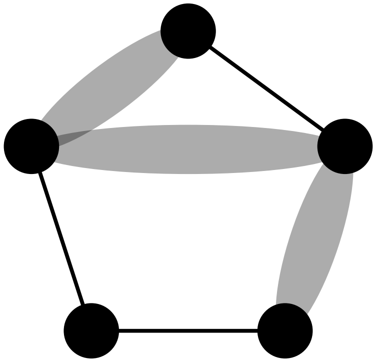
Algorithm for MCMCP experiments over graphs

1. Initialize “task graph”
2. Randomly obscure s pairwise relations



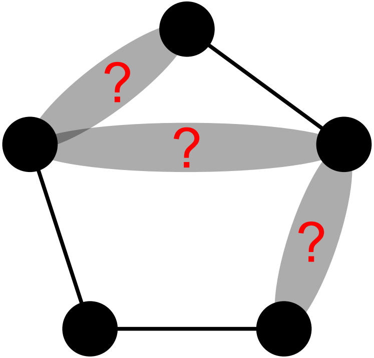
Algorithm for MCMCP experiments over graphs

1. Initialize "task graph"
2. Randomly obscure s pairwise relations
3. Show n^{th} subject the remaining relations



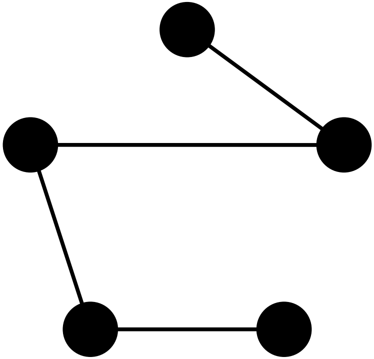
Algorithm for MCMCP experiments over graphs

1. Initialize "task graph"
2. Randomly obscure s pairwise relations
3. Show n^{th} subject the remaining relations
4. Query n^{th} subject about the obscured relations



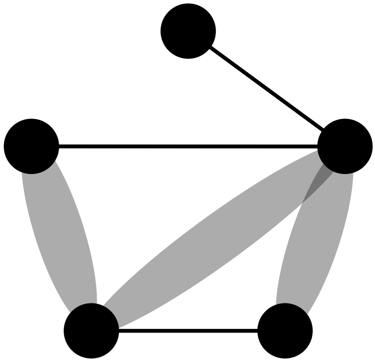
Algorithm for MCMCP experiments over graphs

1. Initialize “task graph”
2. Randomly obscure s pairwise relations
3. Show n^{th} subject the remaining relations
4. Query n^{th} subject about the obscured relations
5. Update task graph, and $n = n + 1$



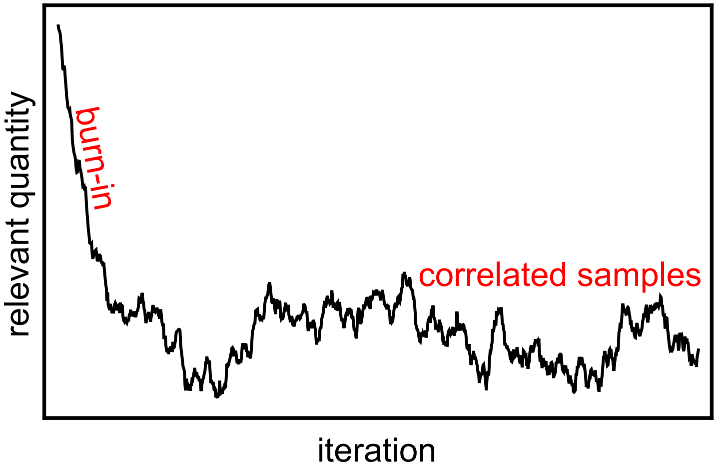
Algorithm for MCMCP experiments over graphs

1. Initialize “task graph”
2. Randomly obscure s pairwise relations
3. Show n^{th} subject the remaining relations
4. Query n^{th} subject about the obscured relations
5. Update task graph, and $n = n + 1$
6. Repeat steps $2 - 5$

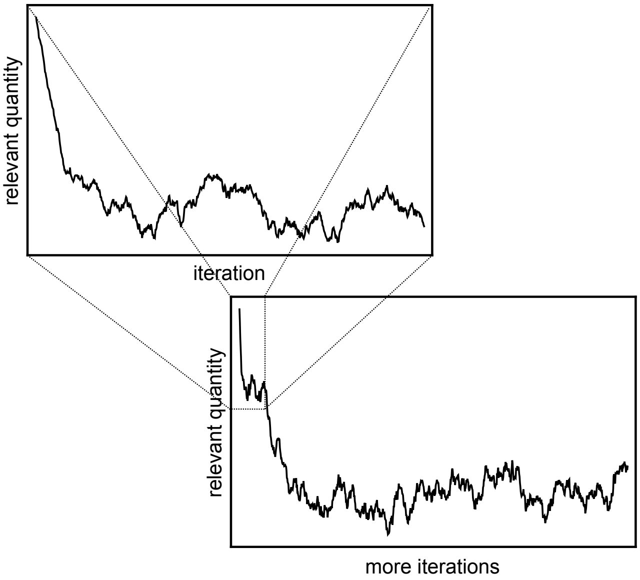


Standard MCMC “wastes” lots of data

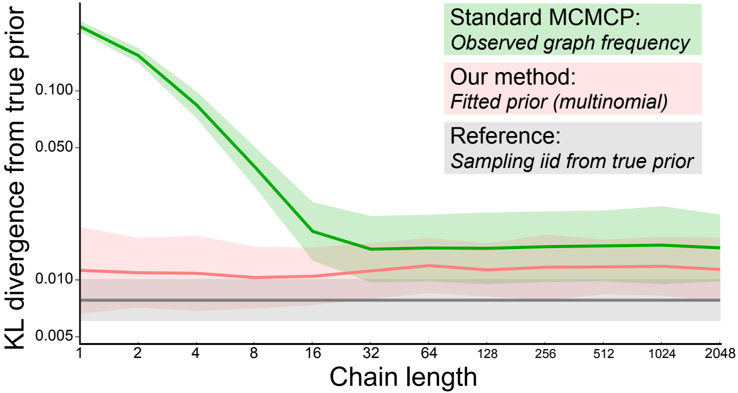
Two main sources of inefficiency:



Estimating convergence is tricky



Priors can be more accurately recovered by fitting the MCMCP model to the aggregated data



Parameterization and interpretation issues

1. How to parametrize distributions over graphs?

nodes:	**unique graphs:**
3	4
4	11
5	34
6	156
7	1044
8	12346
9	274668
10	12005168
11	1018997864
12	165091172592
13	50502031367952
14	29054155657235488
15	31426485969804308768

Parameterization and interpretation issues

1. How to parametrize distributions over graphs?
2. How to summarize these distributions?

nodes:	unique graphs:
3	4
4	11
5	34
6	156
7	1044
8	12346
9	274668
10	12005168
11	1018997864
12	165091172592
13	50502031367952
14	29054155657235488
15	31426485969804308768

Outline

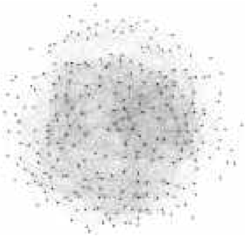
1. MCMCP:
   Experimental framework for obtaining priors over graphs

2. Introducing Graph Cumulants:
   Principled Framework
   for Quantifying the Structure of Priors over Graphs

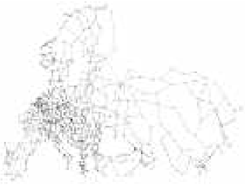
3. The Structure of Human Priors over Navigation and Social Tasks

Introducing Graph Cumulants: Principled Framework for Quantifying the Structure of Priors over Graphs

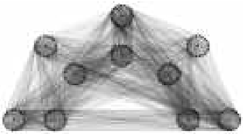
The pluripotent language of graph theory



Biological



Navigational



Social

Recurring themes:

- sparsity
- power law degree distribution
- clustering

A standardized hierarchy of network statistics

Network network null models often aim to replicate these common properties, BUT:

- different models can replicate similar sets of these properties,
- and these properties are often intertwined.

A standardized hierarchy of network statistics

Network network null models often aim to replicate these common properties, BUT:

- different models can replicate similar sets of these properties,
- and these properties are often intertwined.

A standardized hierarchy of descriptive network statistics and associated network null models is lacking.

A standardized hierarchy of network statistics

Network network null models often aim to
replicate these common properties, BUT:

- different models can replicate similar sets of these properties,
- and these properties are often intertwined.

A standardized hierarchy of descriptive network statistics
and associated network null models is lacking.

Vector-valued random variables
enjoy such a description in terms of their *cumulants*.

A standardized hierarchy of network statistics

Network network null models often aim to
replicate these common properties, BUT:

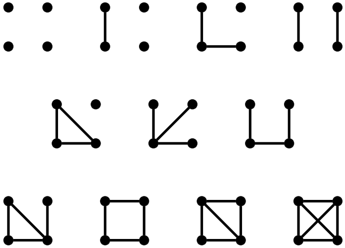
- different models can replicate similar sets of these properties,
- and these properties are often intertwined.

A standardized hierarchy of descriptive network statistics
and associated network null models is lacking.

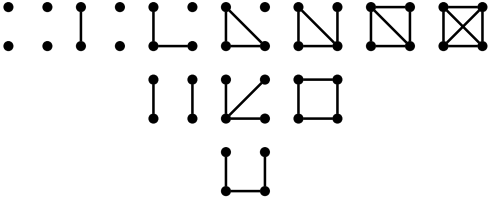
Vector-valued random variables
enjoy such a description in terms of their *cumulants*.

We derived the analogue of cumulants for networks.

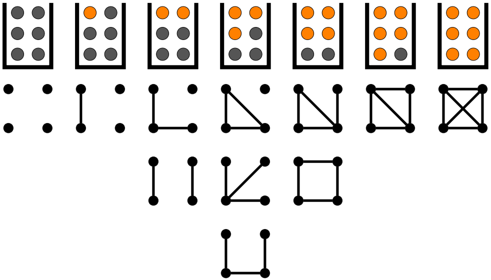
A simpler, non-graphical example



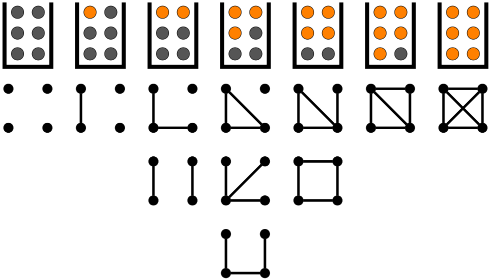
A simpler, non-graphical example

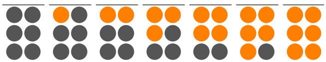


A simpler, non-graphical example

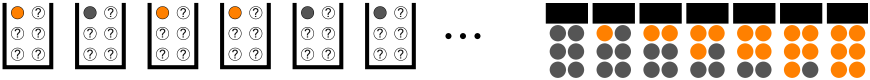


A simpler, non-graphical example

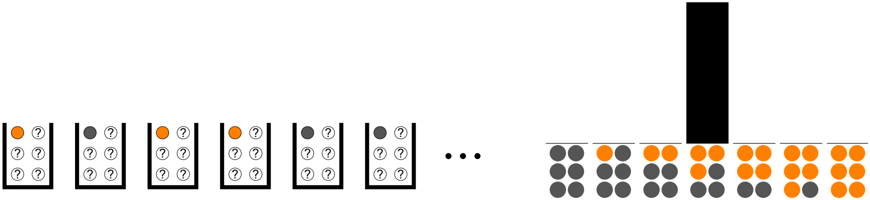




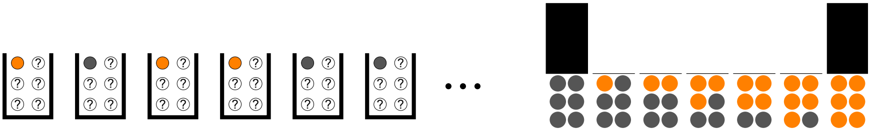
A simpler, non-graphical example



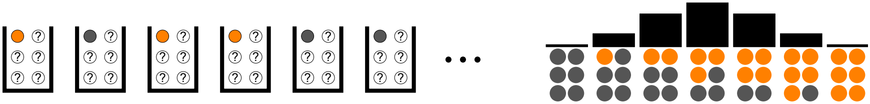
A simpler, non-graphical example



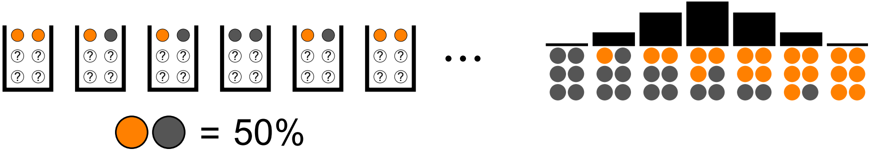
A simpler, non-graphical example



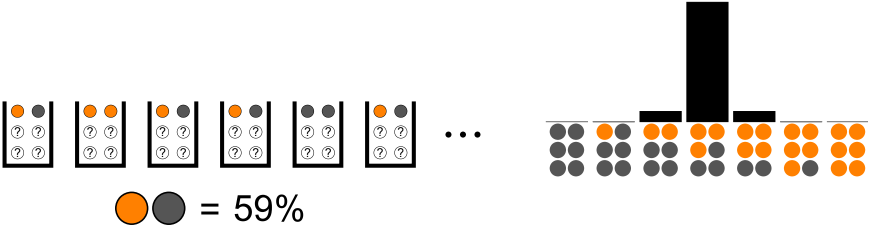
A simpler, non-graphical example



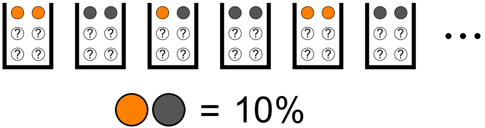
A simpler, non-graphical example

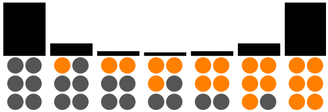


A simpler, non-graphical example

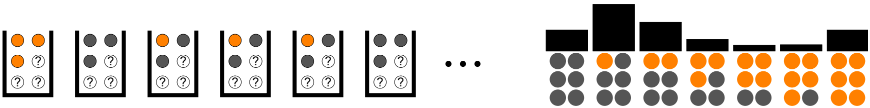


A simpler, non-graphical example



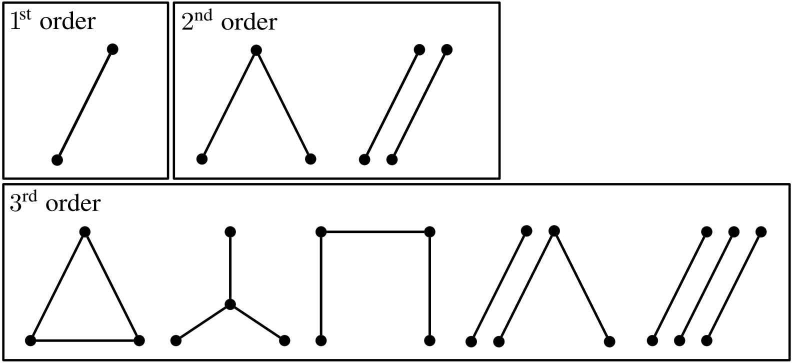


A simpler, non-graphical example



Graph Moments

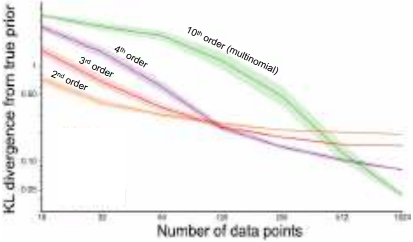
When sampling multiple *relations*,
one must consider the graphical structure.



A natural hierarchical low-dimensional parameterization of distribution over graphs

In the limited data regime, it allows for:

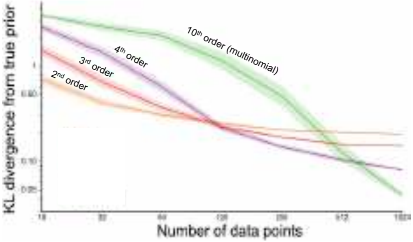
- more accurate recovery of the prior (*in silico data*)

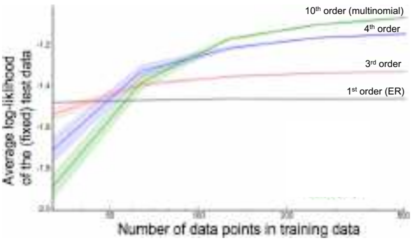


A natural hierarchical low-dimensional parameterization of distribution over graphs

In the limited data regime, it allows for:

- more accurate recovery of the prior (*in silico data*)
- better generalization (*in human data*)





Cumulants as a function of moments

$$\mu_i = i^{\text{th}} \text{ order moment}, \quad \kappa_i = i^{\text{th}} \text{ order cumulant}$$

$$\mu_1 = \langle x \rangle, \quad \mu_2 = \langle x^2 \rangle, \quad \mu_3 = \langle x^3 \rangle, \quad \ldots$$

$$\text{mean: } \kappa_1 = \mu_1$$

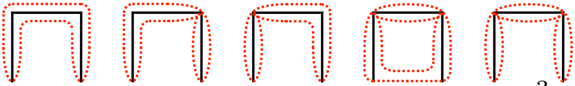
$$\text{variance: } \kappa_2 = \mu_2 - \mu_1^2 \qquad (\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2)$$

$$\text{skew: } \kappa_3 = \mu_3 - 3\mu_2\mu_1 + 2\mu_1^3$$

$$\text{kurtosis: } \kappa_4 = \mu_4 - 4\mu_3\mu_1 - 3\mu_2^2 + 12\mu_2\mu_1^2 - 6\mu_1^4$$

Graph Cumulants

We exploit the combinatorial relationship
between moments and cumulants
to obtain *graph cumulants*.

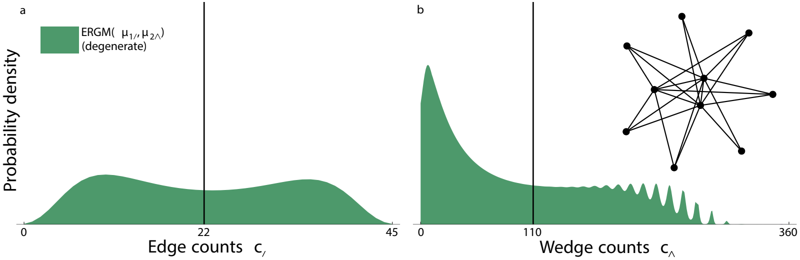


$$\mu_{3\sqcap} = \kappa_{3\sqcap} + \kappa_{2\wedge}\kappa_{1/} + \kappa_{2\wedge}\kappa_{1/} + \kappa_{2/\!/}\kappa_{1/} + \kappa_{1/}^{3}$$

This framework provides a principled method
for quantifying the importance of substructures
in networks of different sizes or edge densities.

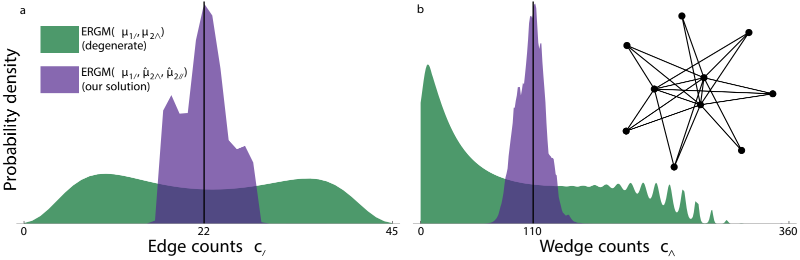
A solution to the degeneracy problem

Many commonly used maximum entropy models of networks yield to bimodal distributions,
even when fitting to a single network observation.



A solution to the degeneracy problem

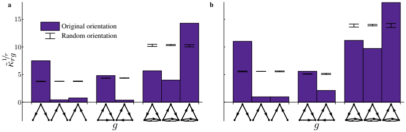
Many commonly used maximum entropy models of networks yield to bimodal distributions,
even when fitting to a single network observation.



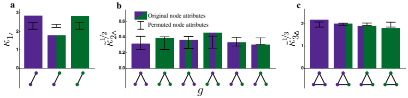
Our hierarchical parameterization solves this problem.

Our framework encompasses networks with additional properties

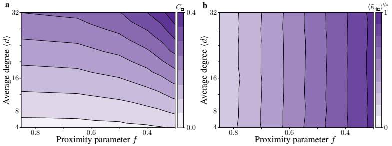
Directed edges:



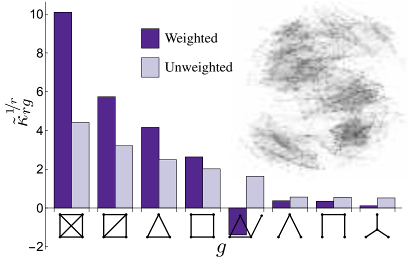
Node attributes:



Bipartite networks:



Weighted edges:



Summary of graph cumulants

- Constructed a hierarchy of summary statistics of networks by considering correlations between an increasing number of relations.
- Provides a principled hierarchical parameterization of distributions over networks.
- Provides interpretable measures of the propensity for arbitrary network substructures.
- Solves the “degeneracy problem” of exponential random graph models.
- Naturally extends to networks with additional properties.

The Structure of Human Priors over Navigation and Social Tasks

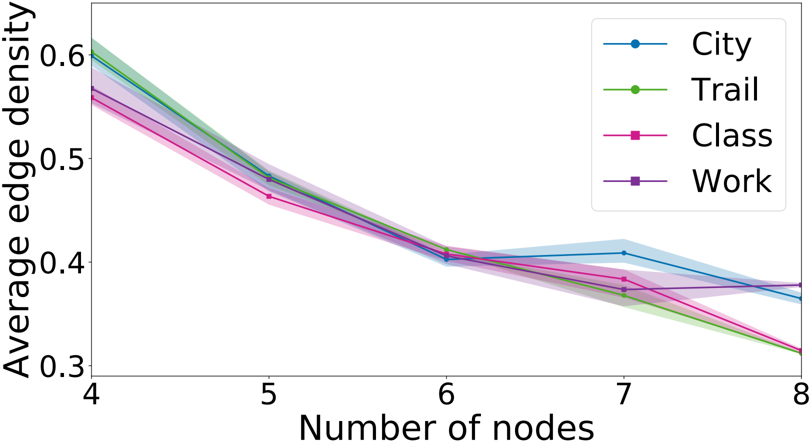
Online experiments

Experimental cover stories:

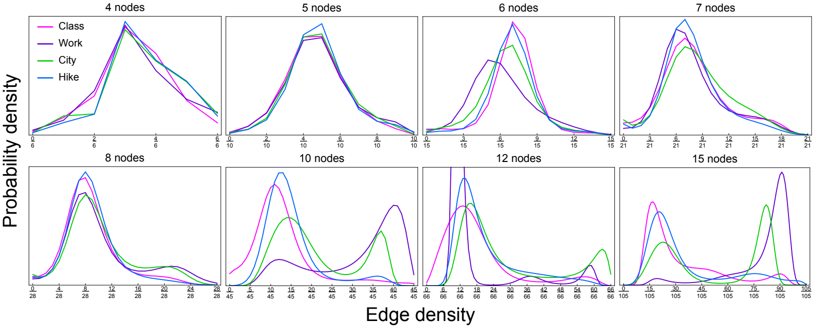
- Social:
    - Friendships in a classroom
    - Friendships in a workplace
- Navigation:
    - Trails in a nature park
    - Neighborhoods in a city

Cover stories were over 4, 5, 6, 7, 8, 10, 12, and 15 nodes.

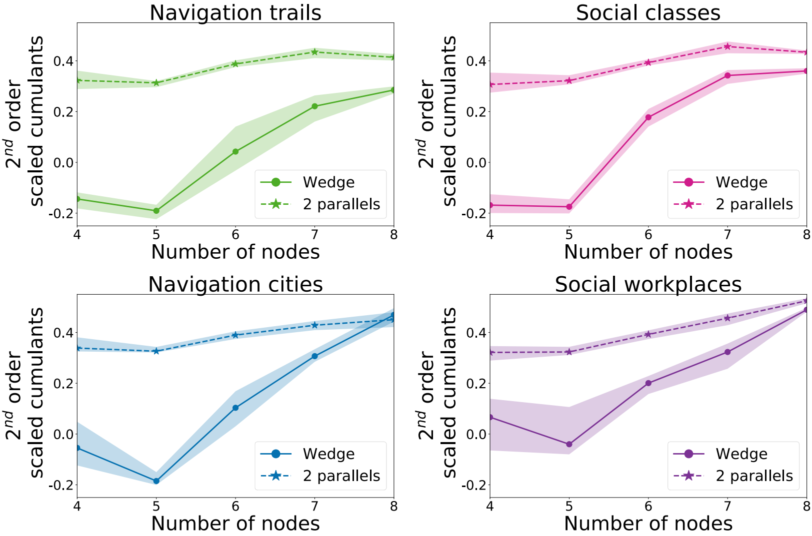
Priors favor sparsity



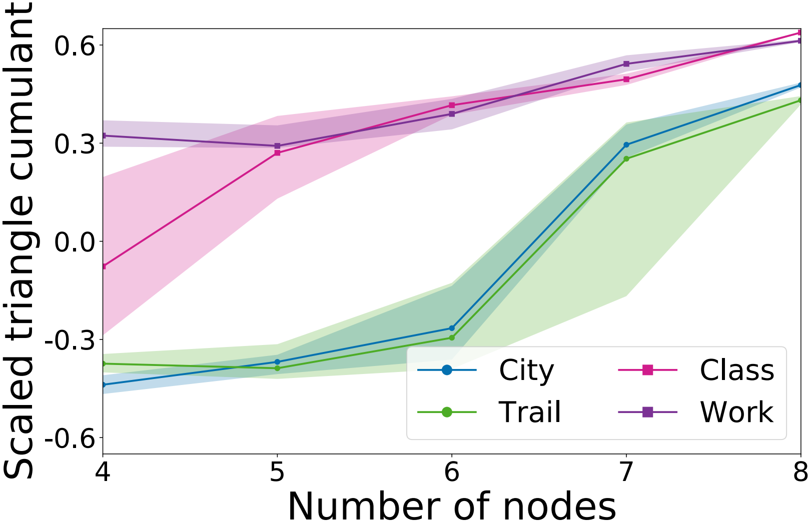
A regime change in the preferred density of edges



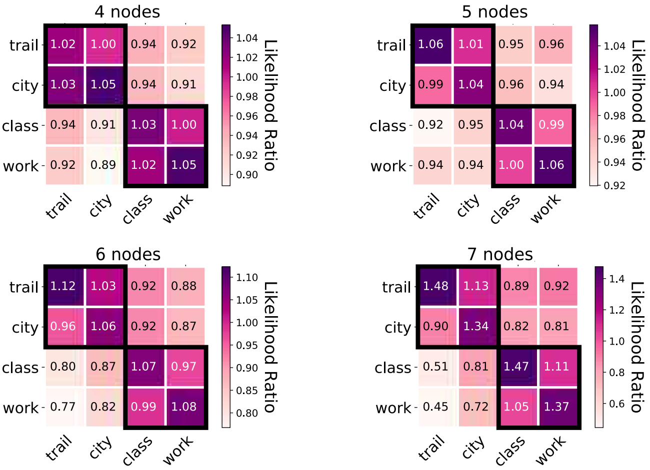
Priors favor more “egalitarian” configurations



Priors over social interactions favor triangles

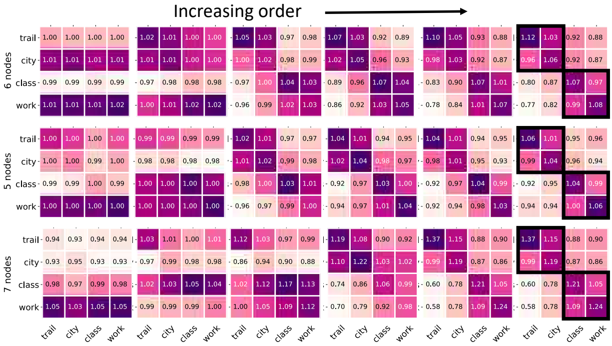


Priors have non-trivial domain-dependent graphical structure



Priors have non-trivial domain-dependent graphical structure

Priors have non-trivial domain-dependent graphical structure



Summary

- Developed a framework for quantifying priors over graphs
- Applied it to human priors over social and navigational tasks
- Resulting priors have nontrivial graphical structure
- Proposed a novel parameterization of distributions over graphs and associated summary statistics