Human Memory Search as Initial-Visit Emitting Random Walk

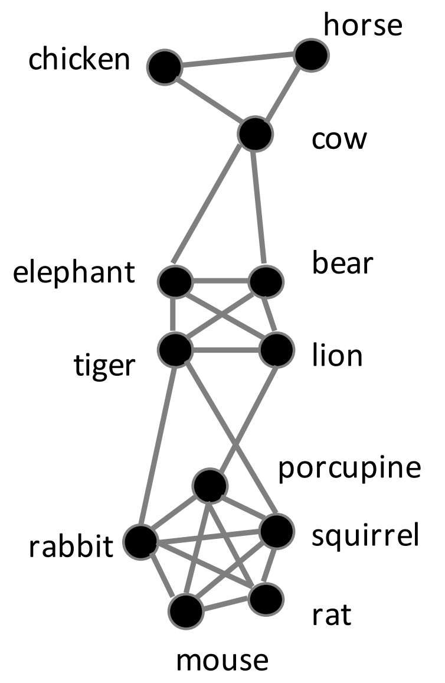
Kwang-Sung Jun* (kjun@discovery.wisc.edu), Xiaojin Zhu†, Timothy Rogers‡, Zhuoran Yang$^\xi$, Ming Yuan$^\zeta$

*Wisconsin Institute for Discovery, Department of †Computer Sciences, ‡Psychology, $^\zeta$Statistics, University of Wisconsin-Madison, $^\xi$Department of Mathematical Sciences, Tsinghua University

INitial-VIsiT Emitting (INVITE) Random Walk

KEY: Output the state only when visiting it for the *first time*

- *n* states, $\boldsymbol{\pi}$: initial distribution, **P**: Markov chain (row stochastic)



$x_1 \sim \boldsymbol{\pi}$

$x_{t+1} \sim \mathbf{P}(. \mid x_t)$

x_1 = cow
x_2 = horse
x_3 = chicken
x_4 = cow
x_5 = bear
x_6 = lion
x_7 = bear
x_8 = tiger

a	output
a_1	cow
a_2	horse
a_3	chicken
a_4	bear
a_5	lion
a_6	tiger
	…

↑ "Censored List"

- The random walk runs *indefinitely*.
- A censored list is not Markovian anymore.
- **GOOD NEWS**: captures important human behavior in a cognitive task (see below) [Abbott12]
- **BAD NEWS**: Parameter Estimation is HARD!

Main Contribution

1. First ***tractable*** method for the maximum likelihood estimate (MLE) of INVITE
2. ***Consistency*** of the INVITE MLE

Verbal Fluency: A Human Memory Search Task

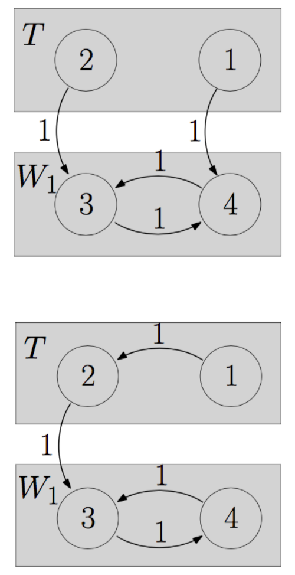
TASK: List examples of *animals* in 60 seconds *without* repetition

- different categories possible: e.g., vehicles
- A "generative" task where participants must remember past productions, inhibit these, and focus on the task.
- Importance
  1. **Clinical application**: different neurological syndromes have different patterns in lists (e.g. repeats more, less/irrelevant items)
     ⟹ important diagnostic information
  2. **Study of human memory search**: responses are runs of semantically related items.
     ⟹ reveal structure in semantic representation
- Our focus is on the second application, so *repeats are ignored*, but can be allowed by a reduction (see future work).

Order	Item
1	cow
2	horse
3	chicken
4	bear
5	lion
6	tiger
7	porcupine
8	rat
9	mouse
10	duck
11	goose
12	…

Censored Lists Generated by INVITE

- A censored list is a permutation of n items or a prefix of it.
- Does it produce every permutation? Or every prefix?



Transient state: a state that has nonzero probability of not coming back to itself in finite time

Recurrent state: a state that is not transient

A is **closed** if $i \in A$ and $j \notin A$ implies a walk from i cannot reach j.

B is **irreducible** if for all $i, j \in B$, a walk from i can reach j.

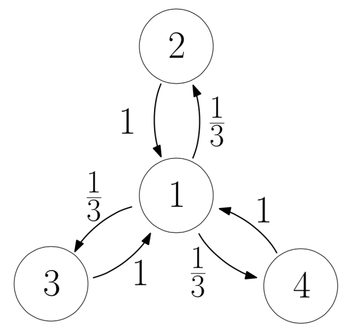
- **Thm 1.** [Durrett12] A finite set of states S can be uniquely decomposed as $S = T \cup W_1 \cup \ldots \cup W_K$, where T is the set of transient states (possibly empty) W_k is a nonempty closed irreducible set of recurrent states.
- **Thm 2.** Given P, let $S = T \cup W_1 \cup \ldots \cup W_K$ be the decomposition by Thm 1. A censored list a generated by INVITE on $\mathbf{P}$ has zero or more transient states (in T), followed by all states in one and only one closed irreducible set (W_k for some k)

Computing INVITE Likelihood

- **Why hard?** A naïve way requires an infinite sum.
  - $\boldsymbol{x} = (x_1, x_2, \ldots)$: uncensored, **hidden**
  - $\boldsymbol{a} = (a_1, \ldots, a_M)$: censored, **observed**

$$\mathbb{P}(\mathbf{a}; \boldsymbol{\pi}, \mathbf{P}) = \sum_{\mathbf{x} \text{ produces } \mathbf{a}} \mathbb{P}(\mathbf{x}; \boldsymbol{\pi}, \mathbf{P})$$

  - E.g., $\boldsymbol{a} = (1, 2, 3, 4)$



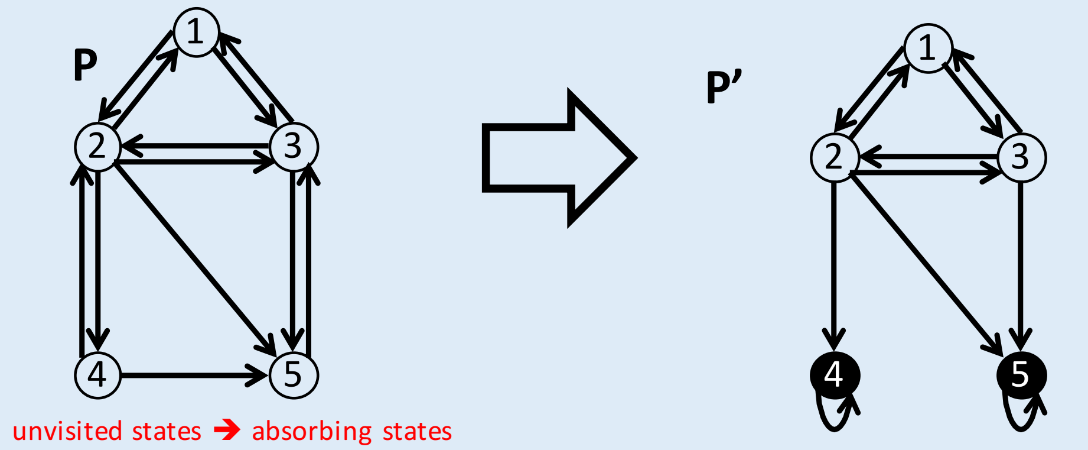
- Key Quantity: $\mathbb{P}(\mathbf{a}; \boldsymbol{\pi}, \mathbf{P}) = \pi_{a_1} \prod_{k=1}^{M-1} \boxed{\mathbb{P}(a_{k+1} \mid a_{1:k}; \mathbf{P})}$
- **Solution**: turn P into an ***absorbing* random walk**!
  - Specifically, turn *unvisited* states into *absorbing* states
  - **Q**: from visited state to visited state
  - **R**: from visited state to unvisited state

$$\mathbf{P} = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{A} & \mathbf{B} \end{pmatrix} \Rightarrow \mathbf{P}' = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

- **Thm 3.** [Doyle84] Let $\mathbf{B} = (\mathbf{I} - \mathbf{Q})^{-1}\mathbf{R}$. $\mathbf{B}_{ik}$ is the probability of a chain starting from i being absorbed by k.
- **Cor 1.** Assume $\mathbf{P}$ is arranged as above. Then,

$$\mathbb{P}(a_{k+1} \mid a_{1:k}, \mathbf{P}) = \begin{cases} [(\mathbf{I} - \mathbf{Q})^{-1}\mathbf{R}]_{k1} & \textit{if } (\mathbf{I} - \mathbf{Q})^{-1} \textit{ exists} \\ 0 & \textit{otherwise} \end{cases}$$

* E.g., $\boldsymbol{a} = (1,2,3,4,5)$. Compute $\Pr(a_4 = 4 \mid a_1 = 1, a_2 = 2, a_3 = 3, \mathbf{P})$.



unvisited states → absorbing states

$$\mathbf{P} = \begin{array}{c|c|c} & 1,2,3 & 4,5 \\ \hline 1,2,3 & \mathbf{Q} & \mathbf{R} \\ \hline 4,5 & \mathbf{A} & \mathbf{B} \end{array} \qquad \mathbf{P}' = \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{pmatrix}$$

$\Pr(a_4 = 4 \mid a_1=1, a_2=2, a_3=3, \mathbf{P}) = ((\mathbf{I} - \mathbf{Q})^{-1}\mathbf{R})_{34}$ [Doyle84]

Consistency of INIVITE MLE

- $\boldsymbol{a}^{(1)}, \boldsymbol{a}^{(2)}, \ldots \sim \text{INVITE}(\boldsymbol{\pi}^*, \mathbf{P}^*)$
- $(\boldsymbol{\pi}^{(m)}, \mathbf{P}^{(m)}) = \arg\max_{(\boldsymbol{\pi},\mathbf{P})} \Sigma_{i \le m} \log \Pr(\boldsymbol{a}^{(i)}; \boldsymbol{\pi}, \mathbf{P})$
- Question: Does $\{ (\boldsymbol{\pi}^{(m)}, \mathbf{P}^{(m)}) \}$ converges to $(\boldsymbol{\pi}^*, \mathbf{P}^*)$?
- A necessary condition: **identifiability**
  - **Thm 4.** INVITE is not identifiable (adjusting self-transitions does not change the model).
  - **Thm 5.** INVITE with $\boldsymbol{\pi}^* > 0$ (element-wise) and *without* self-transitions in $\mathbf{P}^*$ is identifiable.
- **Challenge**: a common strategy is to show *uniform convergence* of the log likelihood, which is not true in INVITE MLE.
- **Solution**: show the **local uniform convergence**
  - uniformly convergent in an intersection of a max-norm ball and a subspace of "equivalent chain decomposition" around the true parameter $(\boldsymbol{\pi}^*, \mathbf{P}^*)$.
- **Thm 6.** If $\boldsymbol{\pi}^* > 0$ (element-wise), INVITE MLE is **consistent.**

Parameter Estimation: Regularized MLE

- Data: $D_m = \left\{ \left(a_1^{(1)}, \ldots, a_{M_1}^{(1)}\right), \ldots, \left(a_1^{(m)}, \ldots, a_{M_m}^{(m)}\right) \right\}$
- Relaxation: assume that the underlying random walk terminates after finite number of steps.
- Initial distribution $\boldsymbol{\pi}$: MAP estimation (easy)
- Transition Matrix **P** is constrained (nonnegative, sum to 1)
  - Easier: unconstrained parameterization
  - $\beta_{ii} := -\infty$ to *disallow* self-transitions.

$$P_{ij} = \frac{e^{\beta_{ij}}}{\sum_{j'=1}^{n} e^{\beta_{ij'}}}$$

- Optimization problem:

$$\min_{\boldsymbol{\beta}} \quad -\sum_{i=1}^{m} \sum_{k=1}^{M_i-1} \log \mathbb{P}(a_{k+1}^{(i)} \mid a_{1:k}^{(i)}; \boldsymbol{\beta}) + \frac{1}{2} C_{\boldsymbol{\beta}} \sum_{i \neq j} \beta_{ij}^2$$

- We run LBFGS.
- For larger dataset, we run *averaged* stochastic gradient descent.

Experiment: Toy

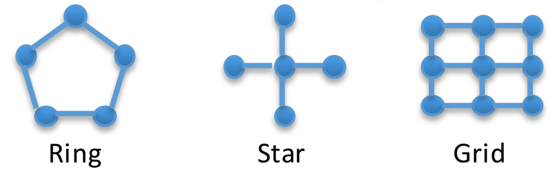
- Goal: (1) confirm the consistency result (2) compare with baselines
- Baselines
  - Naïve Random Walk (RW)

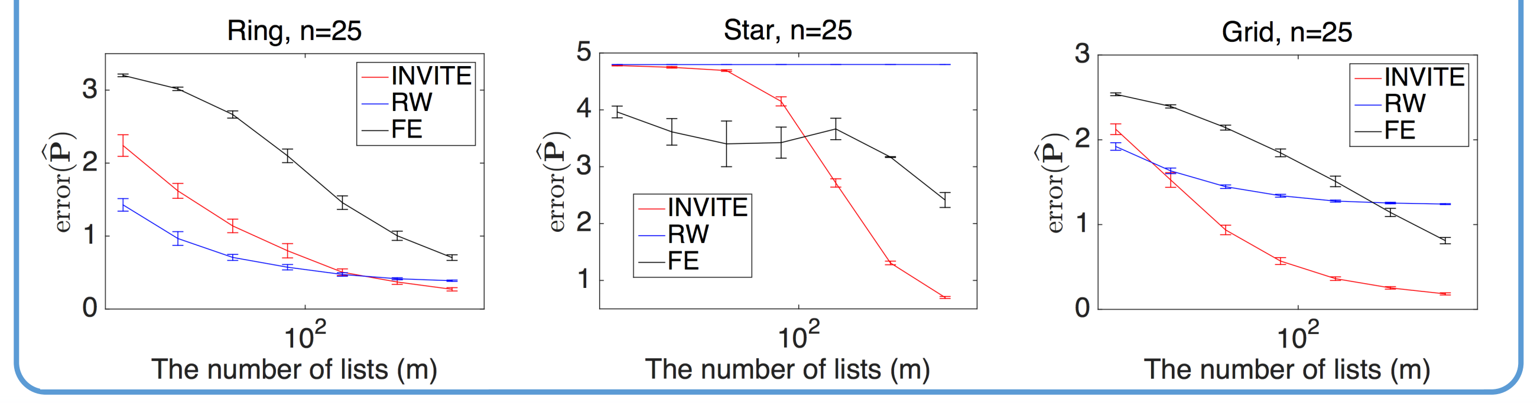
$$\widehat{P}_{rc}^{(RW)} \propto \left(\sum_{i=1}^{m} \sum_{j=1}^{M_i-1} \mathbb{1}_{(a_j^{(i)}=r) \wedge (a_{j+1}^{(i)}=c)} \right) + C_{RW}$$

  - FirstEdge (FE) [Abrahao13]

$$\widehat{P}_{rc}^{(FE)} \propto \left(\sum_{i=1}^{m} \mathbb{1}_{(a_1^{(i)}=r) \wedge (a_2^{(i)}=c)} \right) + C_{FE}$$

- Define $\mathbf{P}^*$ as the uniform transition matrix on toy undirected graphs (# of nodes = 25)
- Use INVITE($\mathbf{P}^*$) to generate censored lists
- Evaluation: $\text{error}(\widehat{\mathbf{P}}) = \sqrt{\sum_{i,j} (\widehat{\mathbf{P}}_{ij} - \mathbf{P}_{ij}^*)^2}$





Experiment: Verbal Fluency

- Data

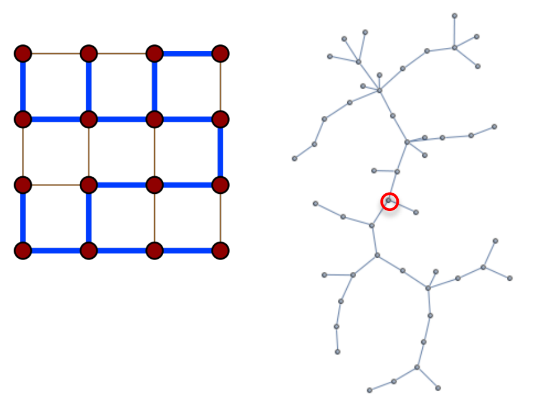
Name	n	m	Length			
			Min.	Max.	Mean	Median
Animal	274	4710	2	36	18.72	19
Food	452	4622	1	47	20.73	21

- Result: negative log likelihood on holdout set (smaller is better)

	Model	Test set mean neg. loglik.
Animal	INVITE	**60.18 (±1.75)**
	RW	69.16 (±2.00)
	FE	72.12 (±2.17)
Food	INVITE	**83.62 (±2.32)**
	RW	94.54 (±2.75)
	FE	100.27 (±2.96)

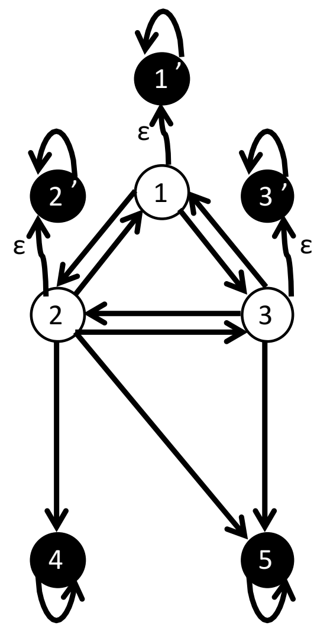
Related Work

- Random spanning tree algorithm [Broder89]
- Self-avoiding Random Walk [Flory53]
- Cascade Model [Gomez-Rodriguez10] for disease / information spread



Future Work

- **Consistency rate**: how fast does it converge to the true parameter?
- **Structured estimation of P**: Given a cluster structure, form a stochastic block matrix P. Or, learn the cluster structure and the parameters at the same time.
  ⟹ reduces the number of parameters
- **Allow repeats**: create "dongle twin" that is an absorbing state.
  ⟹ interpolates between "no repeat" and "repeats as in the standard random walk"
- **Incorporate timing information**



References

[Abbott12] J. T. Abbott, J. L. Austerweil, and T. L. Griffiths, "Human memory search as a random walk in a semantic network," in *NIPS*, 2012, pp. 3050–3058.

[Abrahao13] B. D. Abrahao, F. Chierichetti, R. Kleinberg, and A. Panconesi, "Trace complexity of network inference." *CoRR*, vol. abs/1308.2954, 2013.

[Broder89] A. Z. Broder, "Generating random spanning trees," in *FOCS*. IEEE Computer Society, 1989, pp. 442–447.

[Doyle84] P. G. Doyle and J. L. Snell, *Random Walks and Electric Networks*. Washington, DC: Mathe- matical Association of America, 1984.

[Durrett12] R. Durrett, *Essentials of stochastic processes*, 2nd ed., ser. Springer texts in statistics. New York: Springer, 2012.

[Flory53] P. Flory, *Principles of polymer chemistry*. Cornell University Press, 1953.

[Gomez-Rodriguez10] M. Gomez Rodriguez, J. Leskovec, and A. Krause, "Inferring networks of diffusion and influ- ence," Max-Planck-Gesellschaft. New York, NY, USA: ACM Press, July 2010, pp. 1019- 1028.