

TIMES: Temporal Information Maximally Extracted from Structures



Ananth
Gramma
(Purdue)

Abram
Magner
(UIUC)

Wojciech
Szpankowski
(Purdue)

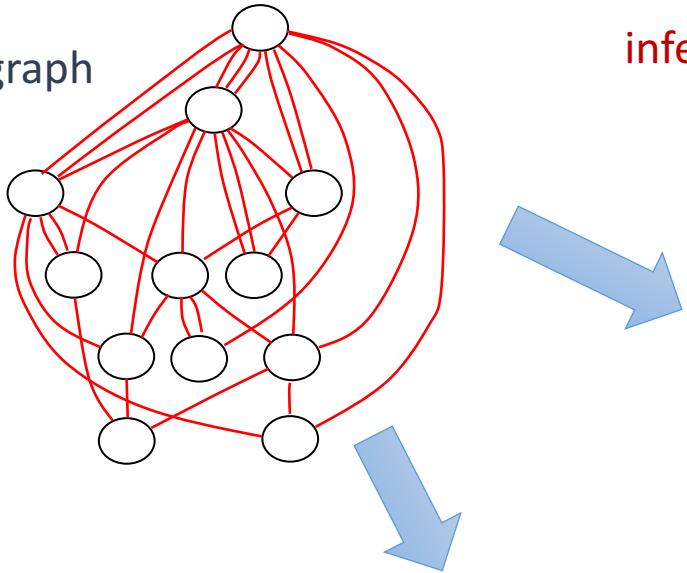
Jithin K. Sreedharan
Purdue University



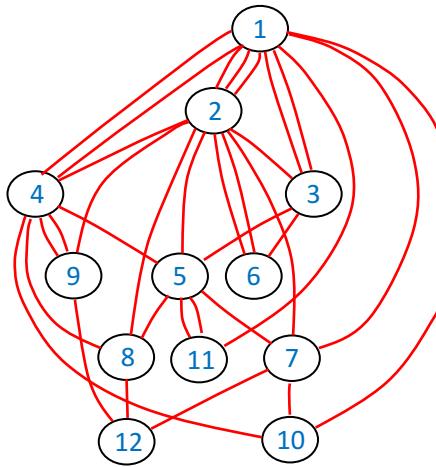
**Center for
Science of Information**
NSF Science and Technology Center

The Problem: Recovery of Node Arrival Order

Dynamic graph

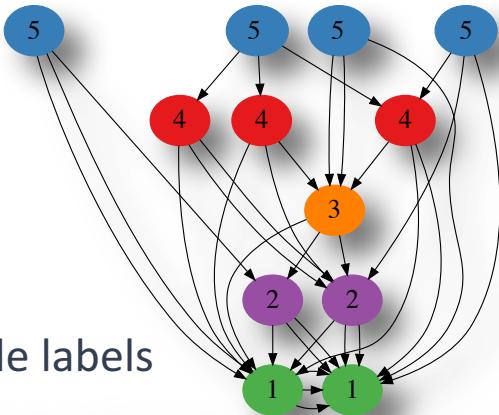


Reverse engineer the dynamic process to infer network trajectories



Graph with node label as arrival order

Graph with partial node labels



Where is arrival order information useful?

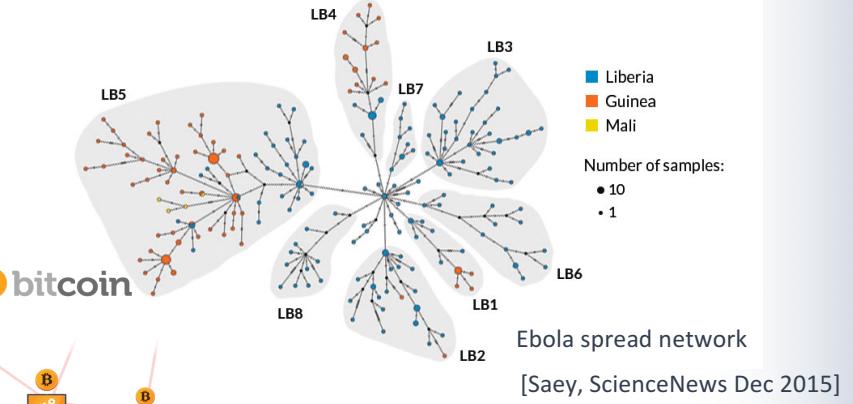
Social networks:

online spam spreading or rumor propagation



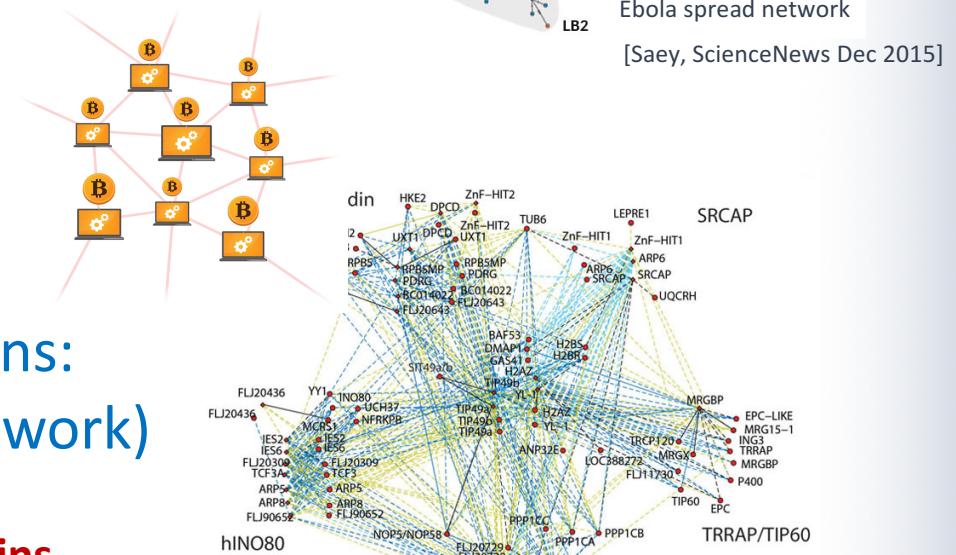
Spread of infectious diseases:

origin and initial carriers



Financial transaction networks:

flow of capital



Network of biochemical reactions:
(protein-protein interaction network)

Study of the phylogenetic tree

Cancer proteins tend to be ancient proteins

[Srivastava et al., Nature 2010]

Prior Works

[Bubeck, Devroye, and Lugosi 2016] For preferential and uniform attachment trees, finds the set that contains root node w.h.p.

[Frieze and Pegden 2017] For preferential attachment graphs, locates the oldest node by a random walk process. Assumes arrival info is known when a node is sampled.

[Shah and Zaman, 2011] Oldest node in Susceptible-Infected epidemic model

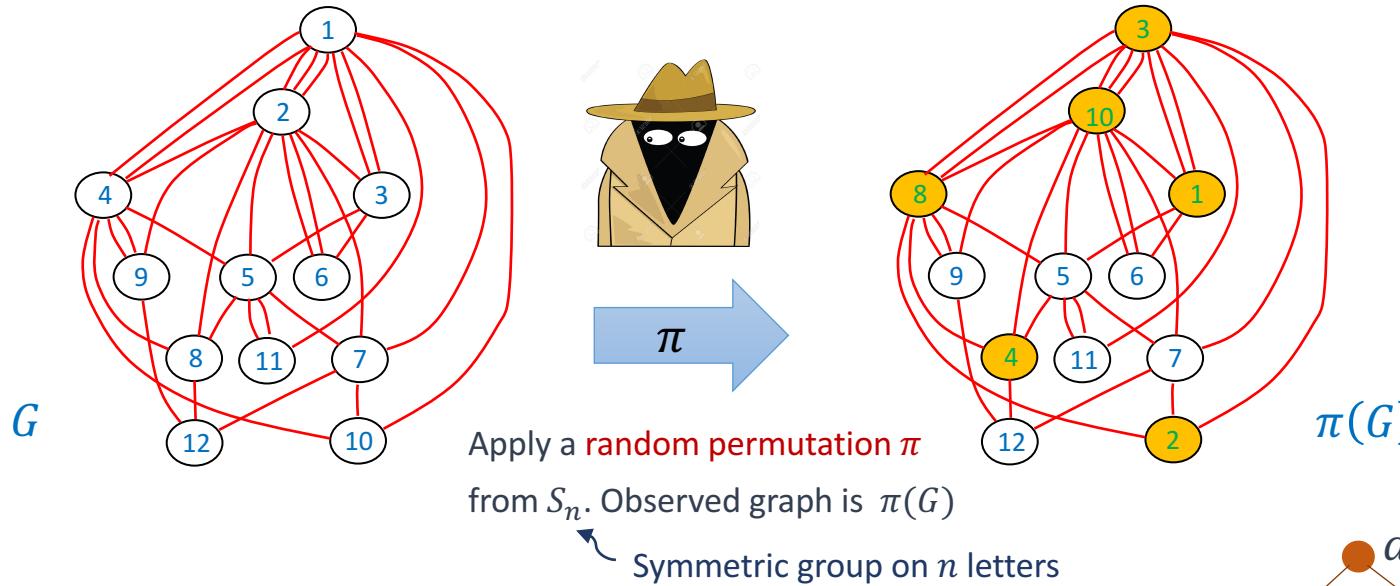
[Zhu and Ying, 2016] Oldest node in Susceptible-Infected-Recovered epidemic model

Several works on "Counting linear extension of partial order sets"

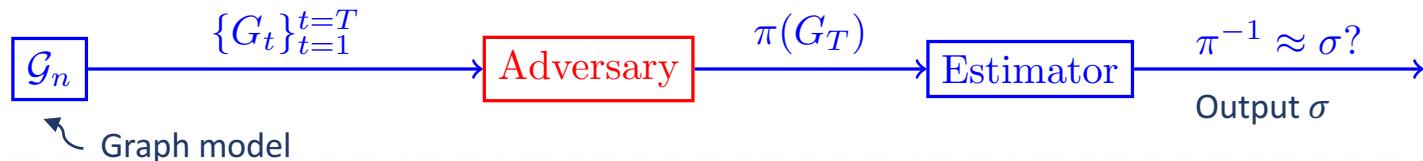
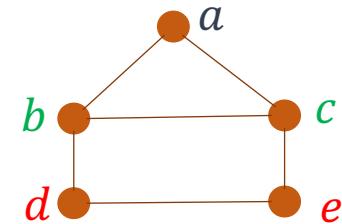
Formulation

Graph G : set of nodes $[n] = \{1, \dots, n\}$

Youngest node label is n ; Oldest node label is 1



Graph model and symmetries play a critical role



Erdős–Rényi and Preferential Attachment models

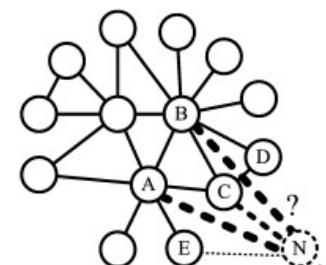
Erdős–Rényi model $G(n, p)$

Each pair of nodes receives an edge independently, with probability p

Preferential Attachment model $\mathcal{PA}(n, m)$

- At $t = 1$, G_1 with single vertex (called 1) is created with m self loops
- At $t > 1$, vertex t joins and makes m connections to existing nodes in G_{t-1} .
- Each of the m connection choice is independent and satisfies

$$\Pr[t \text{ connects to } k | G_{t-1}] = \frac{\deg_{t-1}(k)}{2m(t-1)}$$



Bad News: Inapproximability results

Total Order: Exact and Approximate Recovery of π^{-1}

No algorithm can solve the problem with vanishingly small probability of error in the case of preferential attachment and Erdős–Rényi graphs

Minimax risk with worst adversary and best estimator
 $= 1 - o(1)$

Maximum Likelihood Estimation for Preferential Attachment Graph

$$\mathcal{C}_{\text{ML}}(H) = \arg \max_{\sigma \in S_n} \Pr[\pi^{-1} = \sigma | \pi(G) = H]$$

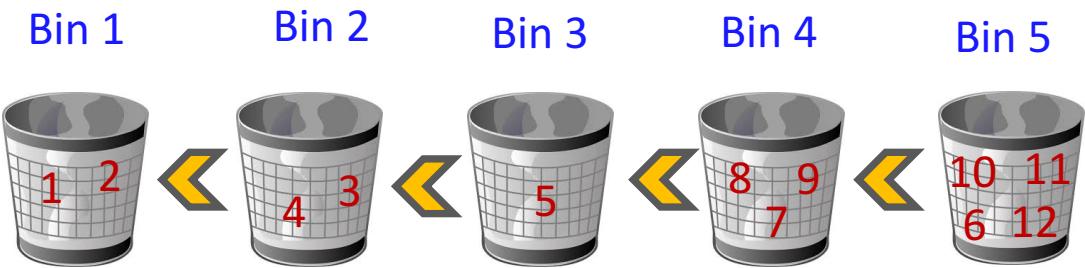
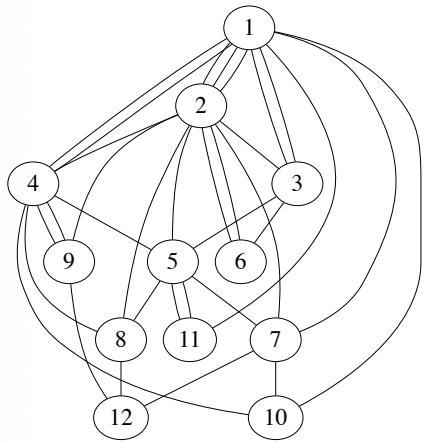
MLE gives a large number of equiprobable solutions,

$$|\mathcal{C}_{\text{ML}}| = e^{n \log n - O(n \log \log n)}$$

Can we do better?

- Formulation
- Infeasibility results
- Optimization formulation
- A simple approximation algorithm for Preferential Attachment graph
- Experimental results: Facebook Wall post and human brain networks

Partial Orders and Binning



Look for partial orders instead of total orders

Estimator:

Set of all labeled graphs with n vertices → Set of all partially ordered set

How to Measure Efficiency of an Estimator?

Recall

How much are we able to recover from a partial order?

$$\frac{\# \text{ correct pairs in the partial order}}{\binom{n}{2}} \quad \rho(\sigma) = \mathbb{E} \left[\frac{1}{\binom{n}{2}} |\{u, v \in [n] : u <_{\sigma} v, \pi^{-1}(u) < \pi^{-1}(v)\}| \right].$$

Precision

How good are the guessed pairs, irrespective of its number?

$$\frac{\# \text{ correct pairs in the partial order}}{\# \text{ pairs in the partial order}} \quad \theta(\sigma) = \mathbb{E} \left[\frac{|\{u, v \in [n] : u <_{\sigma} v, \pi^{-1}(u) < \pi^{-1}(v)\}|}{|\{(u, v) : u <_{\sigma} v\}|} \right]$$

Interpret $\pi^{-1}(v)$ as
original label of node v
of given graph $\pi(G)$

Density

How good are the guessed pairs, irrespective of its number?

$$\frac{\# \text{ pairs in partial ordeer}}{\binom{n}{2}}$$

Constrained Optimization Problem

Different approach: phrase as an integer program.

\max Precision
Set of partial orders
subject to Density $\geq \varepsilon$

$$x_{u,v} : \mathbf{1}\{u <_\phi v\} \text{ for } u, v \in [n]$$
$$p_{u,v}(H) := \Pr[\pi^{-1}(u) < \pi^{-1}(v) | \pi(G) = H]$$

For observed graph $H = \pi(G)$ and $\varepsilon > 0$,

\max Precision \Leftrightarrow

$$\max_{\phi(H)} \frac{\sum_{1 \leq u < v \leq n} p_{u,v}(H) x_{u,v}}{\sum_{1 \leq u \neq v \leq n} x_{u,v}}$$

Subjected to

1. Antisymmetry: $x_{u,v} + x_{v,u} \leq 1$.
2. Transitivity: $x_{u,w} \geq x_{u,v} + x_{v,w} - 1$ for all $u, v, w \in [n]$.
3. Minimum density: $\sum_{1 \leq u \neq v \leq n} x_{u,v} \geq \epsilon \binom{n}{2}$.
4. Domain restriction: $x_{u,v} \in \{0, 1\}$ for all $u, v \in [n]$.

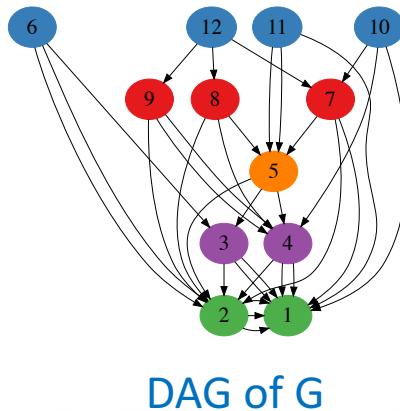
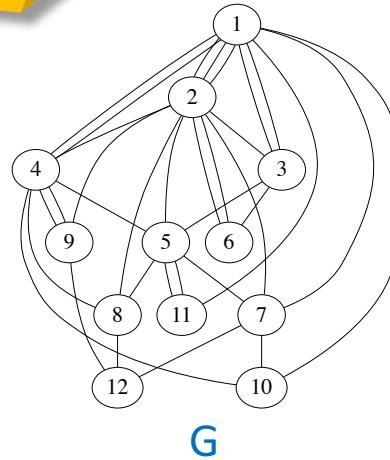
Recollect $\theta(\sigma) = \mathbb{E} \left[\frac{|\{u, v \in [n] : u <_\sigma v, \pi^{-1}(u) < \pi^{-1}(v)\}|}{|\{(u, v) : u <_\sigma v\}|} \right]$

$\Gamma(H)$: permutations σ , such that $\sigma(H)$ has positive probability under the random graph distribution

Estimating $p_{u,v}(H)$

$$= \Pr[\pi^{-1}(u) < \pi^{-1}(v) | \pi(G) = H] = \frac{|\{\sigma : \sigma^{-1} \in \Gamma(H), \sigma^{-1}(u) < \sigma^{-1}(v)\}|}{|\Gamma(H)|}.$$

Focus of Preferential Attachment graphs from now on.

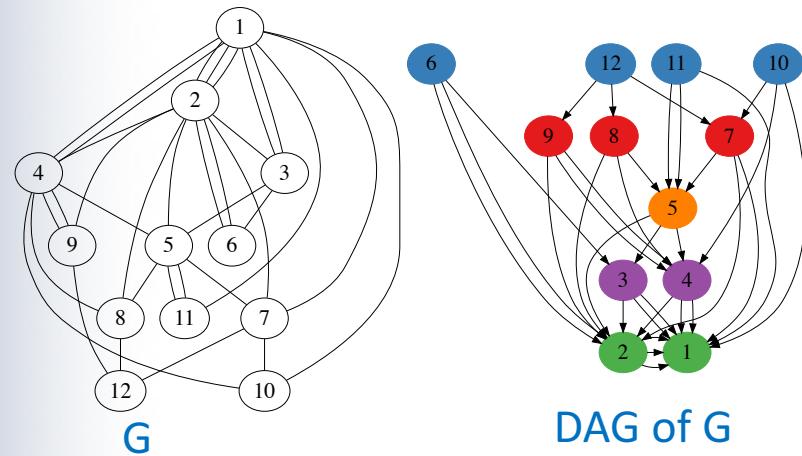


Estimating $p_{u,v}(H)$ is equivalent to counting linear extensions of a partial order

- #P-complete in general!
[Karzanov & Khachiyan,
Brightwell & Winkler]
- Approximate counting in polynomial time (Markov chain algorithm).

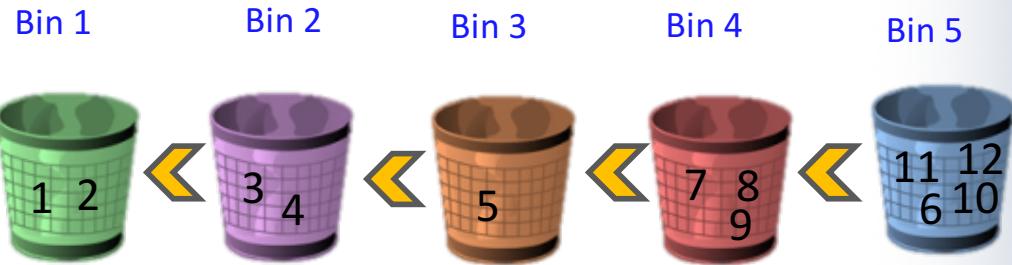
With the best known technique, $p_{u,v}(H)$ can be estimated in $O(n^5 \log^3(n))$ w.h.p.

Approximating via Peeling algorithm



Exact Recovery of the DAG via Peeling

Recursive peeling of minimum degree nodes.
Pealed nodes at each step forms a bin



The Peeling algorithm recovers all the ordered pairs that hold with probability 1

Maximum-Density Precision-1 Estimator

Outputs only probability-1 pairs given by the DAG

Peeling Estimator

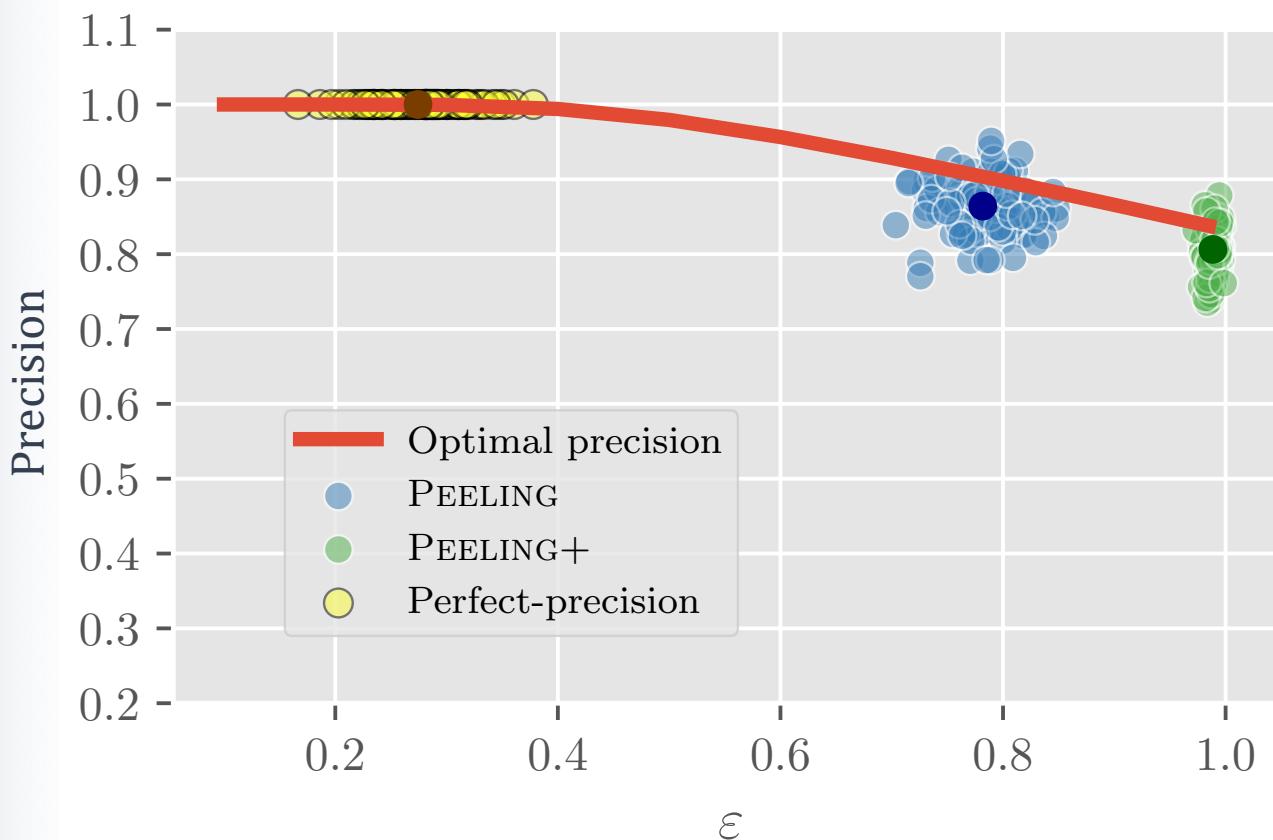
DAG recovered + pairs among non-edge nodes in distinct bins

Peeling+ Estimator

Peeling Estimator + pairs among nodes in same bins

The time complexity of the estimators is dominated by the DAG construction, and is $O(n \log n)$.

For theoretical results, see the paper



Upper bound on optimal precision for density $\geq \varepsilon$

Experiments: Synthetic Graphs

How **robust** is the algorithm?

No. of nodes, $n = 5000$

Result of Peeling estimator

	Technique	Precision	Recall	Density
Uniform Attachment model	$\mathcal{PA}(n, m = 25)$	0.958	0.936	0.977
	$\mathcal{PA}(n, M), M \sim \text{unif}\{5, 50\}$	0.691	0.683	0.988
	$\mathcal{UA}(n, m = 25)$	0.977	0.967	0.99
	$\mathcal{UA}(n, M), M \sim \text{unif}\{5, 50\}$	0.827	0.823	0.995
	Cooper-Frieze (Web graph) model	0.828	0.822	0.993

$\mathcal{PA} + \mathcal{UA} + \text{addition of edges between existing nodes}$

Experiments: Real-World Networks

Result of Peeling estimator

Dataset	# Nodes	# Edges	Genre	Precision	Recall	Density
ArXiv High Energy Physics	7.46K	116K	Citation	0.708	0.681	0.961
Simple English Wikipedia	100K	1.62M	Hyperlink	0.624	0.548	0.878
DBLP CS bibliography	1.13M	5.02M	Coauthorship	0.785	0.728	0.927
Facebook Wall post	43.9K	271K	Social	0.698	0.657	0.941
SMS network	30.2K	447K	Social	0.669	0.610	0.912

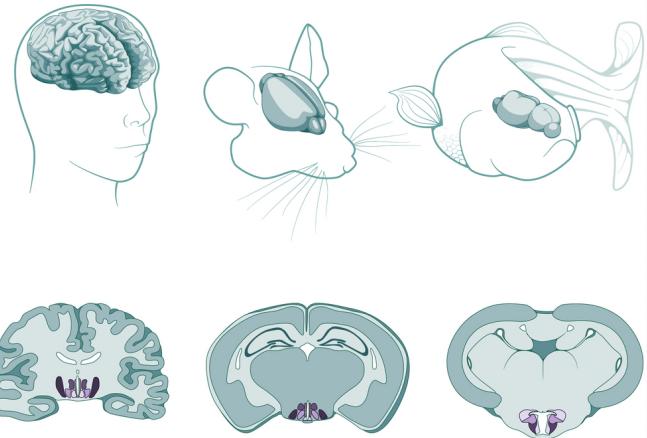
Facebook users in New Orleans region.

Edge of the form (u, v, t) : user u posted on user v 's wall at time t .

Experiments: Brain Networks

Aim: Recover the evolutionary order among the important regions inside the brain.

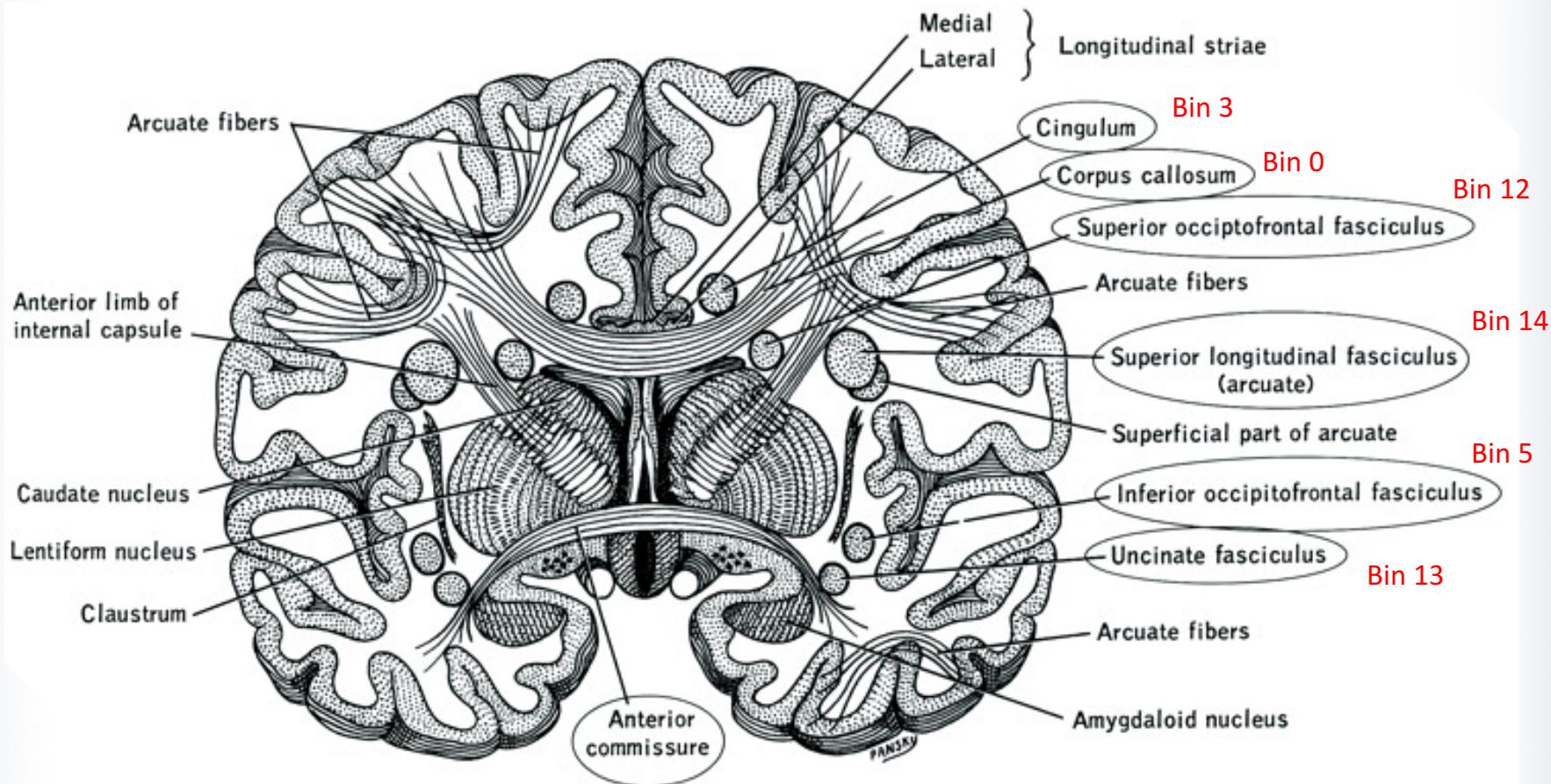
Conjecture: There exists high correlation between these orderings of species evolved from the same genetic parent.



Human Brain Network formation: Cambridge Buckner Dataset

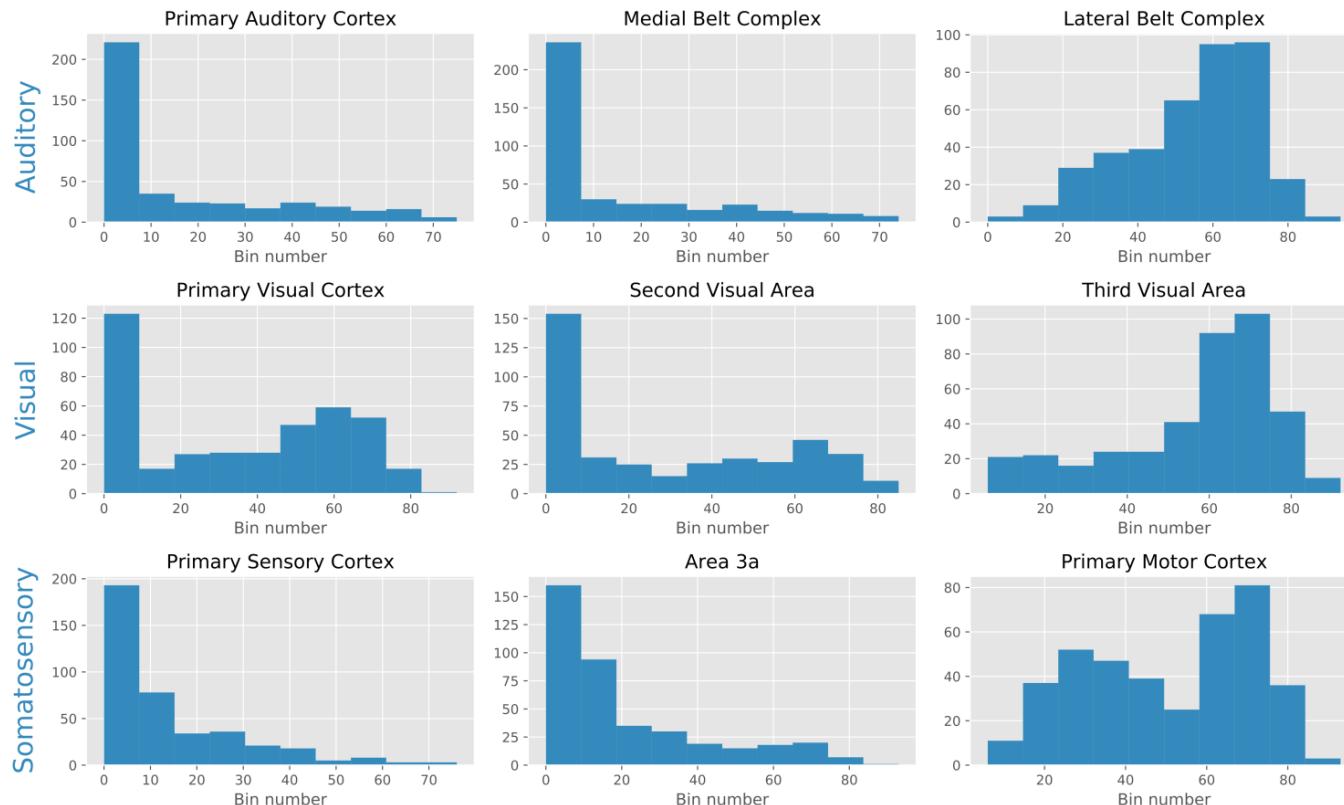
- The network has 46 nodes, each of which represents a region in the brain
- Data: fMRI resting state images
- Correlation matrix with each element as the Gaussianized version of the Pearson correlation coefficient.
- Formation of binary adjacency matrix: find a threshold for the correlation matrix values, above which entries get replaced by 1, and below it by 0.
- The threshold is taken as the largest quantile of elements in the correlation matrix which just makes the graph formed out of it connected.

Peeling estimator on Cambridge Buckner Dataset



Human Connectome Data: Histogram of Prominent Regions in Cortex

Data: Resting state fMRI focusing on the cortex area from 400 healthy young adults. Each network has 300 brain regions.



Conclusions

- Formulated the node arrival order problem for temporal networks
- Fundamental limits derived for inference of node arrival order
- Shown infeasibility of recovering total order
- Recovery of partial orders shows promising results
- Integer programming framework for general random graph models
- Solved the case of standard preferential attachment (PA) model!
- Experimental results reveals that our algorithms admit perturbations in PA model
- Results on Facebook and human brain networks

More details at cs.purdue.edu/homes/jithinks/
(Code and data will be uploaded in a week)

Thank You!