Efficient Estimation of Pairwise Effective Resistance

Renchi Yang and Jing Tang

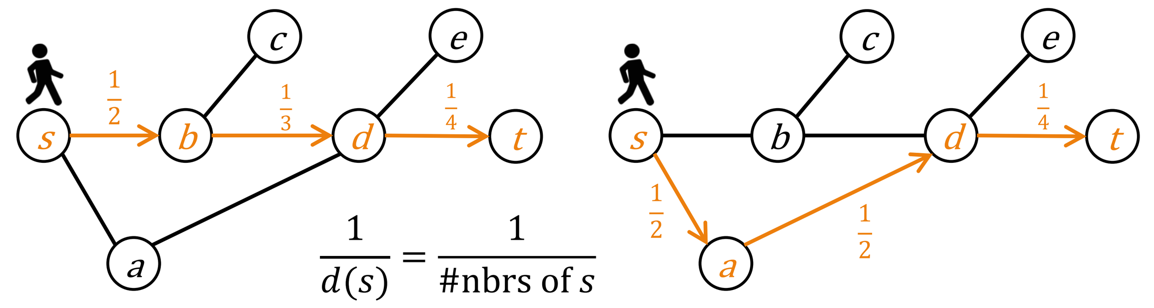
Hong Kong Baptist University and Hong Kong University of Science and Technology (Guangzhou)

香港浸會大學 HONG KONG BAPTIST UNIVERSITY

THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY (GUANGZHOU)

What is Effective Resistance (ER)?

- ***i*-th hop Random Walk Probability**



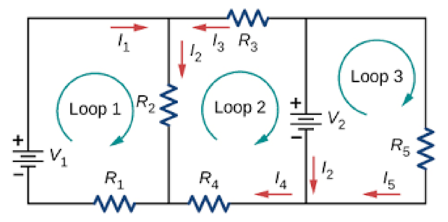
$p^i(s,t)$ = Pr[A random walk starting from s visits t at the i-th hop]

- **Effective Resistance**

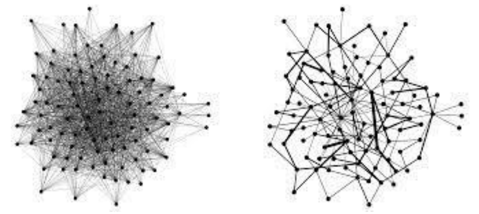
$$ER(s,t) = \sum_{i=0}^{+\infty} \frac{p^i(s,s)}{d(s)} + \frac{p^i(t,t)}{d(t)} - \frac{p^i(s,t)}{d(t)} - \frac{p^i(t,s)}{d(s)}$$

The ER of node pair (s,t) is a sum of the random walk probabilities of all possible numbers of hops between s and t. It describes the *dissimilarity* of nodes s and t.

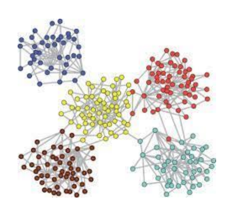
- **Applications of ER**
  - **Electrical Circuit & Power Network Analysis**



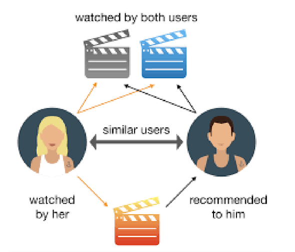
- **Graph Sparsifiers**



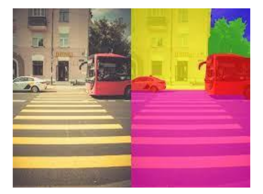
- **Graph Clustering**



- **Recommendation Systems**

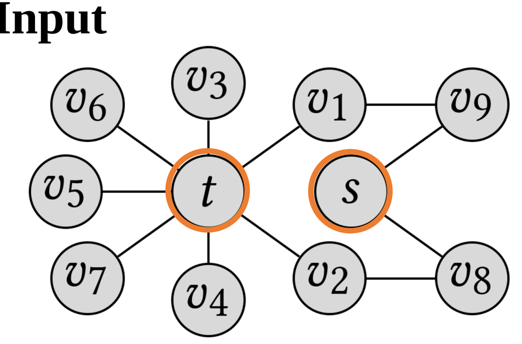


- **Image Segmentation**



Problem Definition

- **ϵ-Approximate Pairwise ER Query**
  - **Input**



An undirected & unweighted graph G

Two nodes s and t

Error threshold ϵ

- **Output**

$$|ER(s,t) - ER'(s,t)| \le \epsilon$$

Exact ER — Approximate ER

Limitations of Existing Work

- **Existing works need**
  - **either expensive matrix operations**
  - **or a huge number of long random walks**
- **The SOTA is a Monte-Carlo Approach**
  - **Maximum Random Walk Length**

$$\ell = \left\lceil \frac{\log\left(\frac{4}{\epsilon(1-\lambda)}\right)}{\log(1/\lambda)} \right\rceil$$ **Too Large!**

λ is the 2nd largest eigenvalue of G's transition matrix, which is usually large (> 0.9) and thus leads to a large ℓ

- **Total #Random-Walks Needed (failure prob.=1%)**

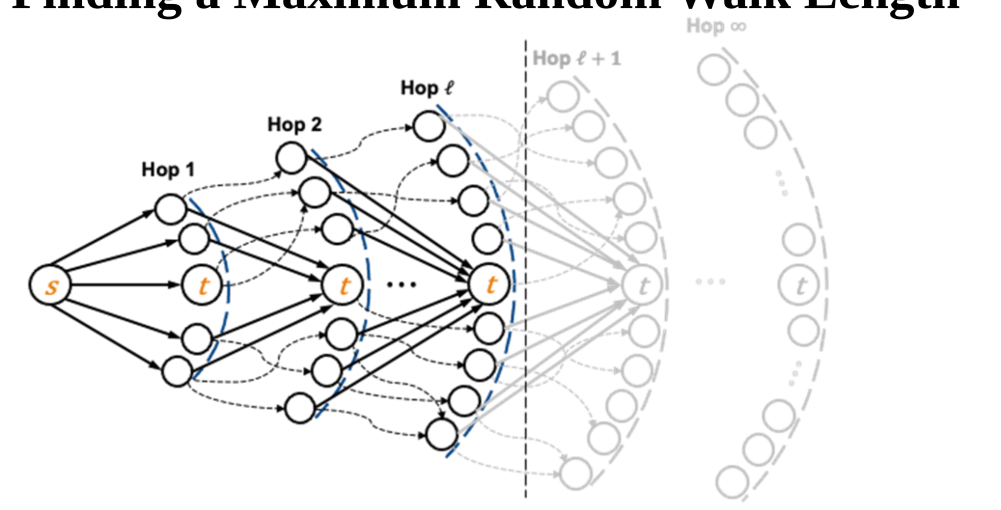
$$r = \frac{40}{\epsilon^2} \cdot \ell^2 \ln(800\ell)$$ **Too Large!**

- **On Facebook graph (#nodes=4k, #edges=88k)**

	$\epsilon = 0.5$	$\epsilon = 0.1$	$\epsilon = 0.05$
ℓ	42	57	64
r	2.94×10^6	139×10^6	710×10^6

Reducing Random Walk Length

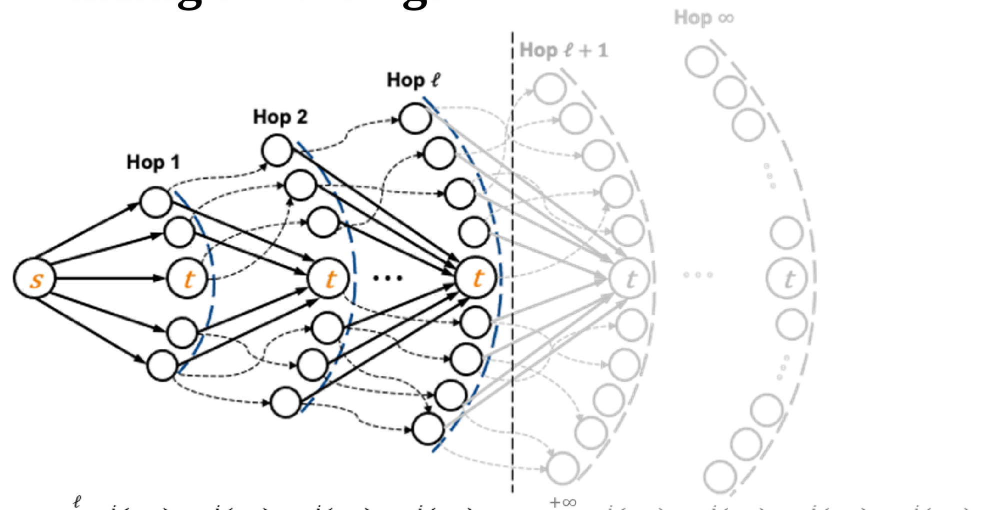
- **Finding a Maximum Random Walk Length**



$$ER(s,t) = \sum_{i=0}^{\ell} \frac{p^i(s,s)}{d(s)} + \frac{p^i(t,t)}{d(t)} - \frac{p^i(s,t)}{d(t)} - \frac{p^i(t,s)}{d(s)} + \sum_{i=\ell+1}^{+\infty} \frac{p^i(s,s)}{d(s)} + \frac{p^i(t,t)}{d(t)} - \frac{p^i(s,t)}{d(t)} - \frac{p^i(t,s)}{d(s)}$$

Estimate this with at most an $\epsilon/2$ error — $\approx \epsilon/2$

- **Utilizing Node Degrees**



$$ER(s,t) = \sum_{i=0}^{\ell} \frac{p^i(s,s)}{d(s)} + \frac{p^i(t,t)}{d(t)} - \frac{p^i(s,t)}{d(t)} - \frac{p^i(t,s)}{d(s)} + \sum_{i=\ell+1}^{+\infty} \frac{p^i(s,s)}{d(s)} + \frac{p^i(t,t)}{d(t)} - \frac{p^i(s,t)}{d(t)} - \frac{p^i(t,s)}{d(s)}$$

Nodes with large degrees have smaller maximum random walk lengths

- **Our Maximum Random Walk Length**

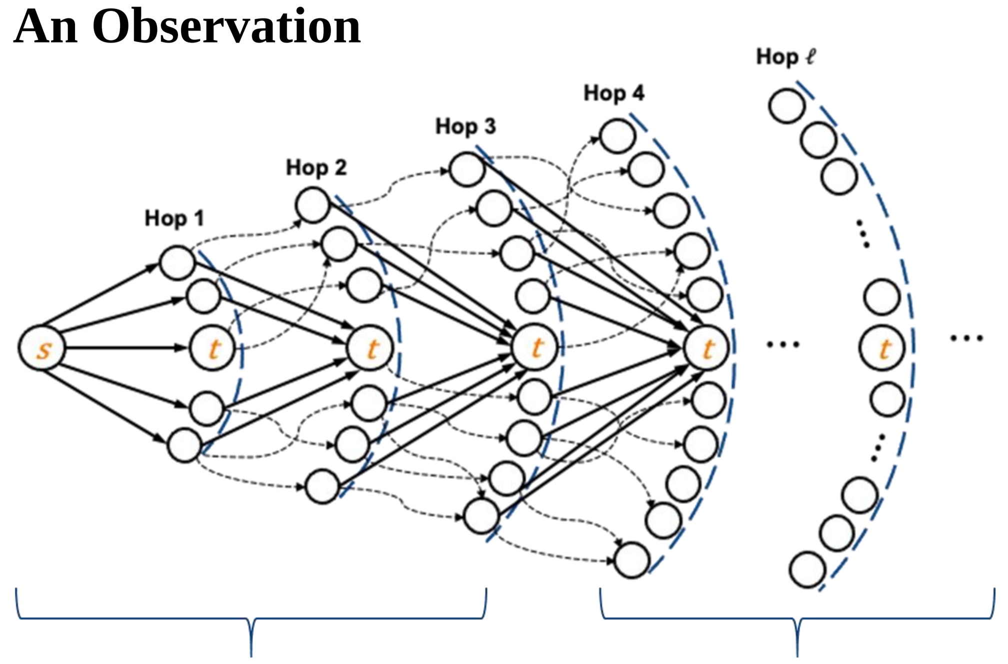
$$\ell = \left\lceil \frac{\log\left(\frac{\frac{2}{d(s)} + \frac{2}{d(t)}}{\epsilon(1-\lambda)}\right)}{\log(1/\lambda)} - 1 \right\rceil$$

- **On Facebook graph (#nodes=4k, #edges=88k)**

	$\epsilon = 0.5$	$\epsilon = 0.1$	$\epsilon = 0.05$
ℓ	42	57	64
Our ℓ	6	21	28
r	2.94×10^6	139×10^6	710×10^6
Our r	48.8×10^3	17.2×10^6	126×10^6

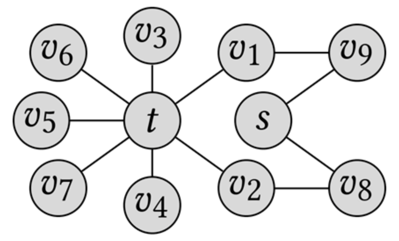
Pruning Random Walks

- **An Observation**



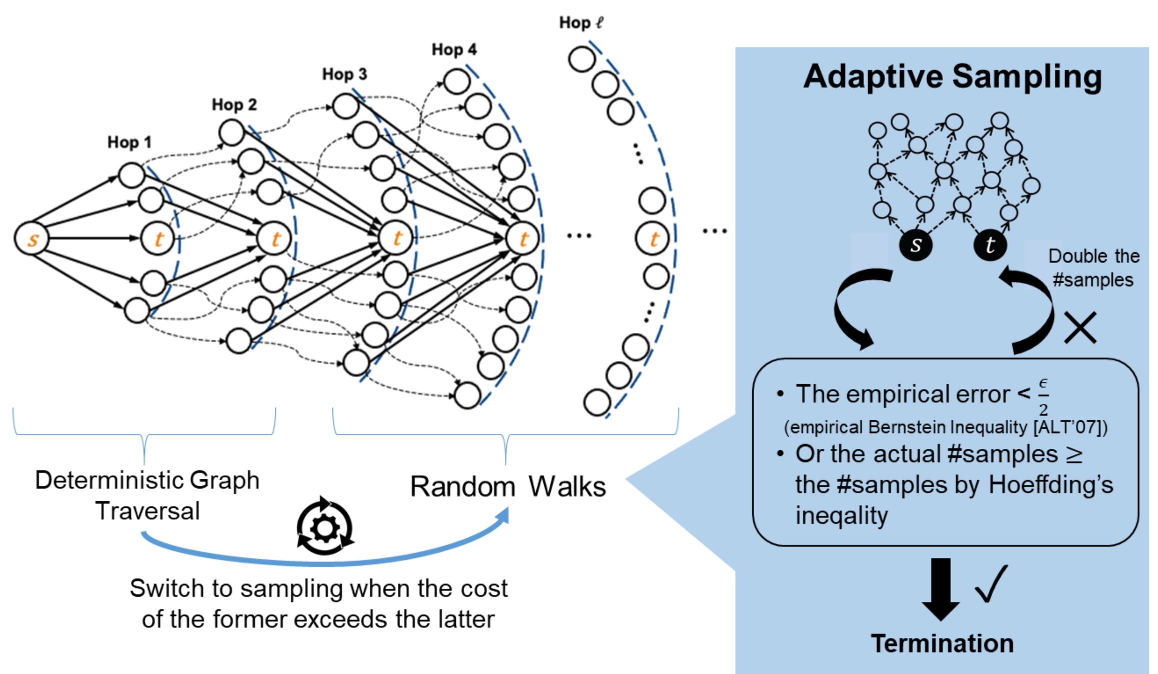
- **#Path & #Random-Walks from Node s to Node t**

Hop	1	2	3	4	5	6	7	8
#paths	9	13	61	97	439	723	**3231**	**5387**
#random walks	**31**	**122**	**275**	**488**	**762**	**1097**	1493	1949



- Deterministic search is better when #hops is small
- Sampling is better when #hops is large

- **Combining Deterministic Graph Traversal and Random Walks in an Adaptive Manner**



Experimental Setup

Dataset	#nodes	#edges	Avg. degree
Facebook	4.04K	88.2K	43.69
DBLP	317.1K	1.05M	6.62
YouTube	1.13M	2.99M	5.27
Orkut	3.07M	117.2M	76.28
LiveJournal	4M	34.68M	17.35
Friendster	65.6M	1.81B	55.06

- Our Methods (AMC (only random walks), GEER (deterministic graph traversal + random walks)) against 5 competitors: EXACT, TP/TPC (KDD'21), RP (STOC'08), and SMM (our baseline)
- Linux machine with an Intel Xeon(R) Gold 6240@2.60GHz 32-core processor and 377GB of RAM

Experimental Results