Efficient Approximation Algorithms for Spanning Centrality

Shiqi Zhang, Renchi Yang, Jing Tang, Xiaokui Xiao, Bo Tang

NUS National University of Singapore | 香港浸會大學 HONG KONG BAPTIST UNIVERSITY | THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY | SUSTech Southern University of Science and Technology

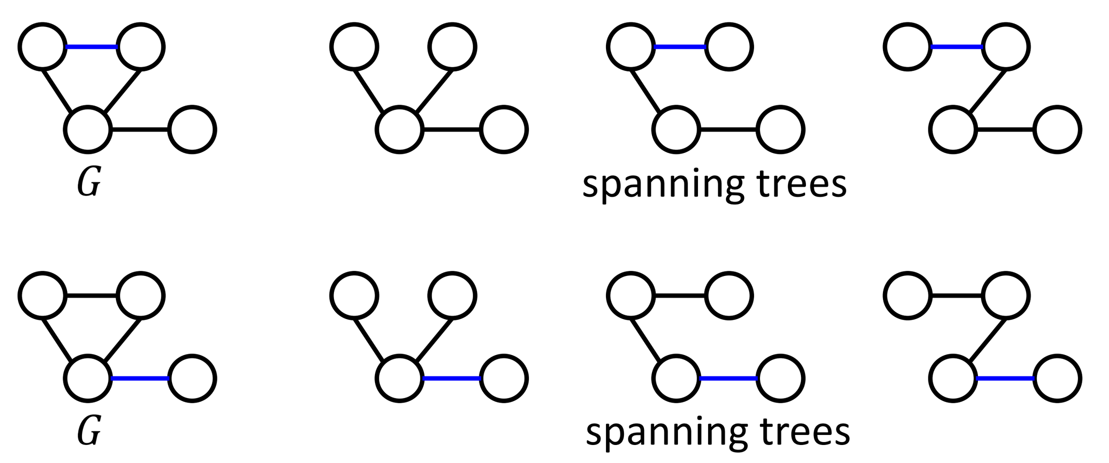
Problem Definition

Input:

an undirected and connected graph G

Spanning centrality $s(e_{i,j})$ of an edge $e_{i,j}$:

the fraction of spanning trees of G that contains $e_{i,j}$



A higher SC $s(e_{i,j}) \rightarrow e_{i,j}$ is more crucial for connectedness

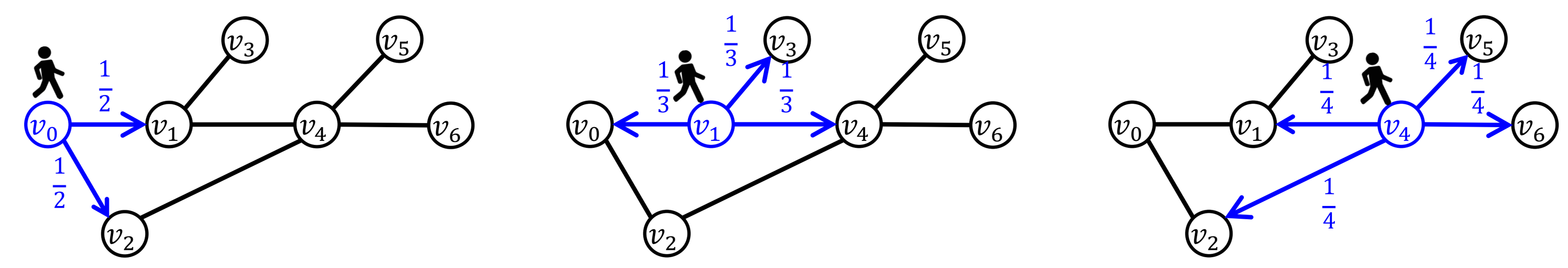
ϵ-approximate All Edge Spanning Centrality (AESC):

the estimated SC $\hat{s}(e_{i,j})$ for every edge $e_{i,j}$ satisfying

$$|s(e_{i,j}) - \hat{s}(e_{i,j})| \le \epsilon$$

State of the Art

Simple random walk from node v_0:



$d(v_i)$ = #neighbors of v_i

$p_\ell(v_i, v_j)$ = Pr[A simple random walk from v_i visits v_j at the ℓ-th step]

SC in a view of simple random walk [Peng et al. KDD'21]:

$$s(e_{i,j}) = \sum_{\ell=0}^{+\infty} \frac{p_\ell(v_i,v_i)}{d(v_i)} + \frac{p_\ell(v_j,v_j)}{d(v_j)} - \frac{p_\ell(v_i,v_j)}{d(v_j)} - \frac{p_\ell(v_j,v_i)}{d(v_i)}$$

↓ Shorten by a derived length threshold τ — TOO LARGE

$$s_\tau(e_{i,j}) = \sum_{\ell=0}^{\tau} \frac{p_\ell(v_i,v_i)}{d(v_i)} + \frac{p_\ell(v_j,v_j)}{d(v_j)} - \frac{p_\ell(v_i,v_j)}{d(v_j)} - \frac{p_\ell(v_j,v_i)}{d(v_i)}$$

↓ Estimated by random walk samplings — TOO MANY

$$\hat{s}(e_{i,j}) = \hat{s}_\tau(e_{i,j})$$

Our Proposal

Separate the random walk interpretation of SC into three parts:

Switch when the cost of step 2 exceeds step 3

Step 1: improved length threshold $\tau_{i,j}$ personalized to each $e_{i,j}$

Step 2: Deterministic graph traversal in a reverse manner for the first $\tilde{\tau}$ steps

Step 3: Random walk samplings for the rest $\tau - \tilde{\tau}$ steps

$$\sum_{\ell=\tau+1}^{+\infty} \frac{p_\ell(v_i,v_i)}{d(v_i)} + \frac{p_\ell(v_j,v_j)}{d(v_j)} - \frac{p_\ell(v_i,v_j)}{d(v_j)} - \frac{p_\ell(v_j,v_i)}{d(v_i)} + \sum_{\ell=0}^{\tilde{\tau}} \frac{p_\ell(v_i,v_i)}{d(v_i)} + \frac{p_\ell(v_j,v_j)}{d(v_j)} - \frac{p_\ell(v_i,v_j)}{d(v_j)} - \frac{p_\ell(v_j,v_i)}{d(v_i)} + \sum_{\ell=\tilde{\tau}+1}^{\tau} \frac{p_\ell(v_i,v_i)}{d(v_i)} + \frac{p_\ell(v_j,v_j)}{d(v_j)} - \frac{p_\ell(v_i,v_j)}{d(v_j)} - \frac{p_\ell(v_j,v_i)}{d(v_i)}$$

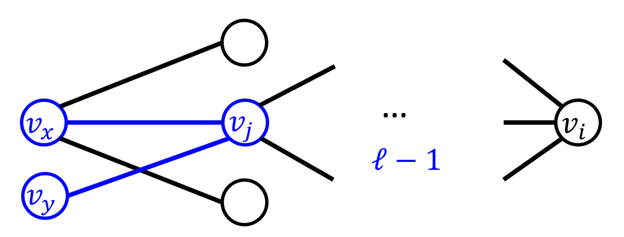
Step 1:

- endpoints with larger degrees ↓ smaller SC values and smaller τ ↓ utilize the degree information of two endpoints
- decomposed by graph spectral property ↓ utilize eigenvectors and eigenvalues of G

Step 2:

$$\sum_{\ell=0}^{\tilde{\tau}} \left(\frac{p_\ell(v_i,v_i)}{d(v_i)} - \frac{p_\ell(v_j,v_i)}{d(v_i)}\right) + \left(\frac{p_\ell(v_j,v_j)}{d(v_j)} - \frac{p_\ell(v_i,v_j)}{d(v_j)}\right)$$

rely on $p_\ell(v_*, v_i)$ ↓ deterministic traversal reversely from v_i



$$p_\ell(v_x, v_i) \mathrel{+}= \frac{p_{\ell-1}(v_j,v_i)}{d(v_x)} = \frac{p_{\ell-1}(v_j,v_i)}{3}$$

$$p_\ell(v_y, v_i) \mathrel{+}= \frac{p_{\ell-1}(v_j,v_i)}{d(v_y)} = p_{\ell-1}(v_j,v_i)$$

Step 3:

$$\sum_{v_x} \frac{p_{\tilde{\tau}}(v_x,v_i)}{d(v_i)} \left(\sum_{\ell=1}^{\tau-\tilde{\tau}} p_\ell(v_i,v_x) - p_\ell(v_j,v_x)\right)$$

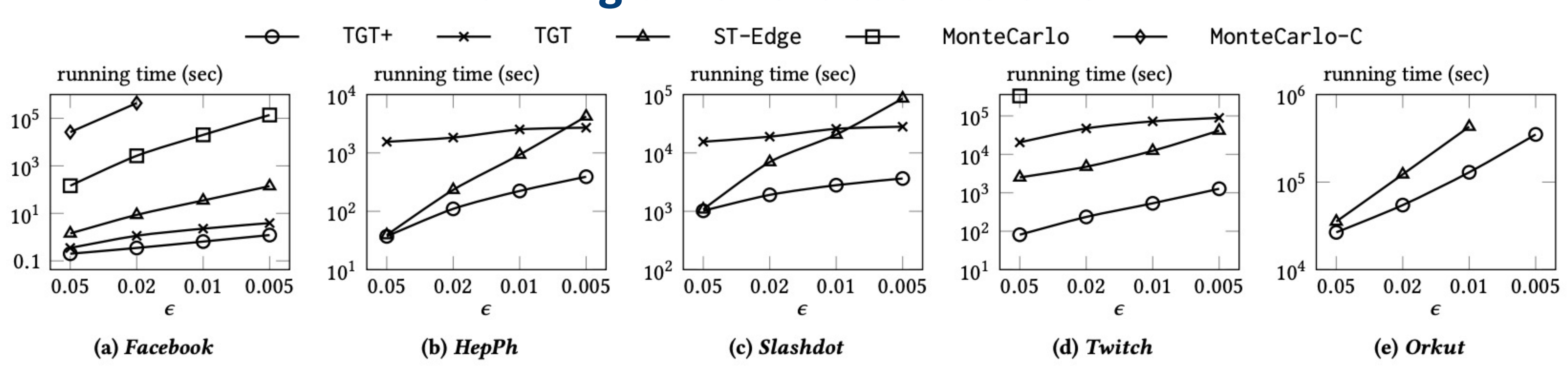
- all $p_{\tilde{\tau}}(v_*, v_i)$ are known in step 2
- estimate by generating random walks from v_i and v_j

Experiments

ϵ-approximate AESC solutions:

- spanning tree sampling: ST-Edge
- random walk sampling with our τ:
  - MonteCarlo, MonteCarlo-C
- our proposal:
  - TGT (steps 1-2), TGT+ (steps 1-3)

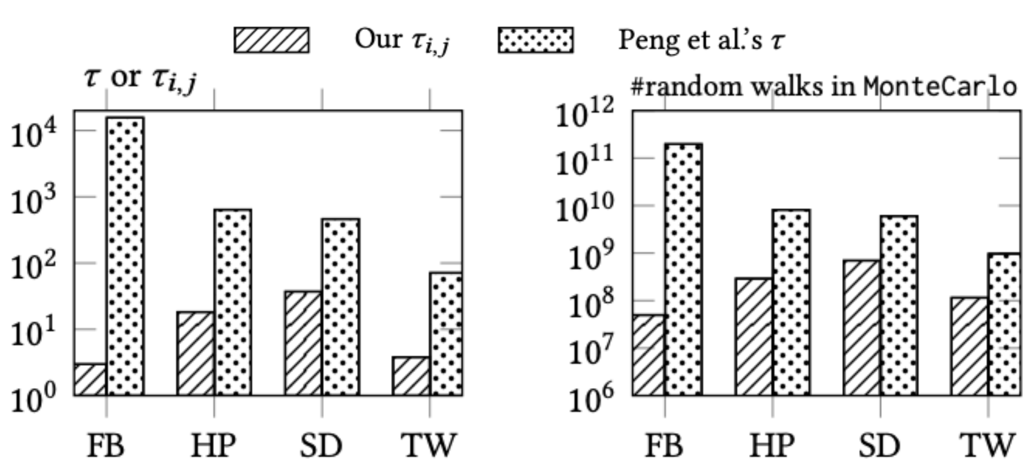
Running time vs. absolute error



Datasets

Name	#nodes	#edges
Facebook [30]	4,039	88,234
HepPh [20]	34,401	420,784
Slashdot [22]	77,360	469,180
Twitch [35]	168,114	6,797,557
Orkut [50]	3,072,441	117,185,082

τ comparison



Acknowledgement, technical report, code repository, etc.

This work is partially supported by the National Natural Science Foundation of China (Grant No. U22B2060), by Shenzhen Fundamental Research Program (Grant No. 20220815112848002), by Guangzhou Municipal Science and Technology Bureau (Grant No. 2023A03J0667), and by HKUST(GZ) and HKBU under Start-up Grants. For the technical report and code, please scan the QR code and visit Shiqi's homepage.