Effective and Efficient PageRank-based Positioning for Graph Visualization

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

National University of Singapore

香港浸會大學 HONG KONG BAPTIST UNIVERSITY

SUSTech Southern University of Science and Technology

Graph Visualization

Input: a graph G with n nodes and m edges

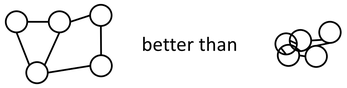
Output: a 2D position matrix $\boldsymbol{X}$

Drawing:

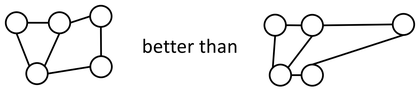
- Position each node v_i at its coordinate $X[i]$
- Link two endpoints of each edge with a straight segment

Aesthetic criteria :

- Evaluate the readability of $\boldsymbol{X}$
- Node distribution (ND):
  - Measure the distribution evenness of the nodes on the screen



- Uniform Length Coefficient Variance (ULCV):
  - Measure the length skewness of edge segments on the screen

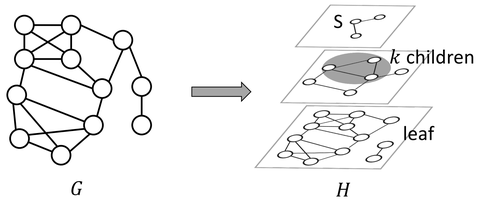


PPRviz Framework

Idea: interactively show the partial view of G level by level

Offline: supergraph hierarchy construction

- Build a hierarchy H for G by Louvain with balanced size

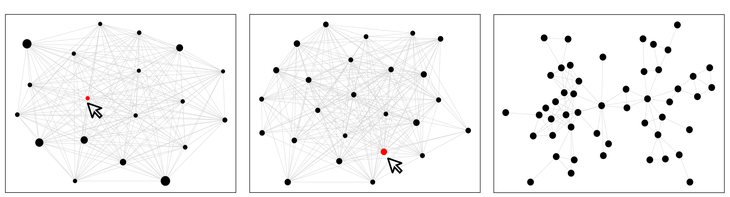


Online #1: node distance computation

- **Propose a new distance measure PDist**
- **Compute PDist matrix Δ for children in S by our Tau-Push**

Online #2: node position embedding

- Compute $\boldsymbol{X}$ by $\boldsymbol{\Delta}$
- Make node pair's Euclidean distance resemble its PDist



PDist for Leaf Nodes

Personalized PageRank (PPR)

- Input: a source v_i, a target v_j, a stopping probability α
- Random walk with restart (RWR) from v_i:
  - At each step, stops with probability α at the current node,
  - With $1-\alpha$ probability randomly jumps to one of neighbors
- PPR from v_i to v_j:

$$\pi(v_i, v_j) = \mathbb{P}[\text{RWR from } v_i \text{ stops at } v_j]$$

PDist between any nodes v_i, v_j:

- Degree-normalized PPR (DPPR):

$$\pi_d(v_i, v_j) = \pi(v_i, v_j) \cdot d(v_i)$$

- Convert DPPR to a distance:

$$1 - \log\left(\pi_d(v_i, v_j) + \pi_d(v_j, v_i)\right)$$

- Pros:
  - Preserve high-order information
  - Guarantee visualization quality in terms of ND and ULCV

Tau-Push for Leaf Nodes

Step #1: Tau value computation

- Compute τ_j for each v_j: $\tau_j = \frac{1}{m} \cdot \sum_i \pi_d(v_i, v_j)$
- Identify v_j with τ_j larger than a pre-defined τ

Step #2: Forward Push

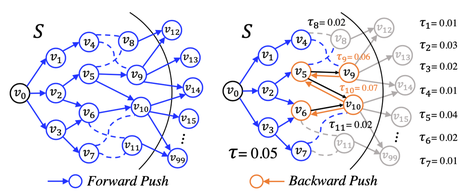
- A deterministic version of RWR
- Estimate $\pi_d(v_i, v_j)$ from v_i for v_j with $\tau_j < \tau$

Step #3: Backward Push

- A reversed deterministic version of RWR
- Estimate $\pi_d(v_i, v_j)$ from v_j for v_j with $\tau_j \geq \tau$

τ_1	0.01	τ_7	0.01
τ_2	0.03	τ_8	0.02
τ_3	0.02	τ_9	0.06
τ_4	0.01	τ_{10}	0.07
τ_5	0.04	τ_{11}	0.02
τ_6	0.02	...	...

$\tau = 0.05$



Index: τ_j in step #1 and $\hat{\pi}_d(v_i, v_j)$ in step #3

Result: For any $v_i, v_j \in S$ and $v_i \neq v_j$:

$$\left|\pi_d(v_i, v_j) - \hat{\pi}_d(v_i, v_j)\right| \leq \epsilon \cdot \delta$$

Experiments

Datasets and competitors

- 12 real-world graphs from different fields
- 11 single-level and 2 multi-level competitors

Dataset ($K = 10^3, M = 10^6, B = 10^9$)

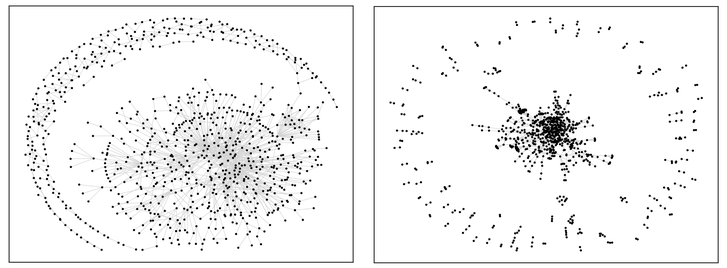
Dataset	n	m	Description
TwEgo	23	52	Ego network
FbEgo	52	146	Ego network
Wiki-ii	186	632	Authorship network
Physician	241	1.8K	Social network
FilmTrust	874	2.6K	User trust network
SciNet	1.5K	5.4K	Collaboration network
Amazon	334.9K	1.9M	Product network
Youtube	1.1M	6.0M	Social network
Orkut	3.1M	234.4M	Social network
DBLP	5.4M	17.2M	Collaboration network
It-2004	41.3M	2.3B	Crawled network
Twitter	41.7M	3.0B	Social network

ULCV: the smaller the better

	TwEgo	*FbEgo*	*Wiki-ii*	*Physician*	*FilmTrust*	*SciNet*
PPRviz	**0.22**	**0.39**	**0.35**	**0.45**	**0.48**	**0.34**
FR	0.35	*0.42*	*0.41*	0.53	*0.54*	0.77
LinLog	0.57	0.67	1.09	0.90	1.99	4.70
ForceAtlas	0.37	0.49	0.64	0.55	0.96	1.52
CMDS	0.40	0.46	0.62	0.80	1.05	1.74
PMDS	*0.23*	0.45	0.78	*0.47*	0.69	*0.74*
GFactor	0.45	0.91	0.62	0.95	0.64	0.86
SDNE	1.96	0.94	0.94	1.67	1.31	1.72
LapEig	1.15	0.98	1.04	1.02	1.70	1.26
LLE	0.46	0.77	1.27	0.77	0.87	-
Node2vec	0.80	0.96	0.86	1.41	0.89	1.32
SimRank	0.84	0.75	0.53	0.53	1.78	1.98

Effectiveness

- PPRviz outperforms all competitors in terms of aesthetic criteria and visualization results



Visualization results of PPRviz (left) and the best competitor FR (right) on FilmTrust

Efficiency

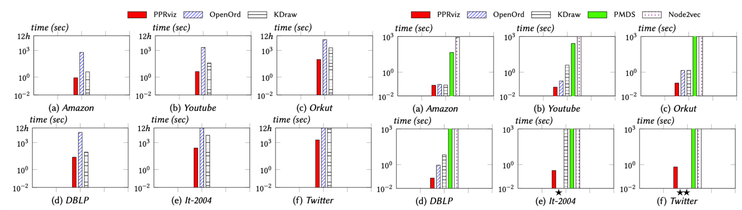
- PPRviz outperforms all competitors in terms of preprocessing time and response time

Preprocessing time:

- compute H and index of Tau-Push in PPRviz
- compute H in multi-level method

Response time:

- visualize S in PPRviz and multi-level methods
- visualize G in single-level methods



Preprocessing time (left) and response time (right)

Acknowledgement, technical report, code repository, etc.

This work was partially supported by A*STAR, Singapore (Grant No. A19E3b0099), by Guangdong Basic and Applied Basic Research Foundation (Grant No. 2021A1515110067), and by Shenzhen Fundamental Research Program (Grant No. 20220815112848002) and a research gift from Huawei. For the technical report and code, please visit Shiqi's homepage: https://www.comp.nus.edu.sg/~s-zhang/.