

## Abstract

Recent advances in information science have shown that linked data pervade our society and the natural world around us. Graphs have become increasingly important for representing complicated structures such as Wikipedia, Freebase, and various social networks. However, existing algorithms cannot handle large graphs efficiently, so fast algorithms are needed. We introduce two fast algorithms. They outperform previous algorithms in terms of both speed and quality. Our algorithms are fundamental to many applications. Our algorithms allow many applications to be processed more efficiently and will help to improve the effectiveness of future applications.

- Efficient search algorithm for PageRank [1]

### Overview Find high PageRank nodes from large-scale graphs efficiently and exactly

- Compute lower/upper PageRank scores iteratively

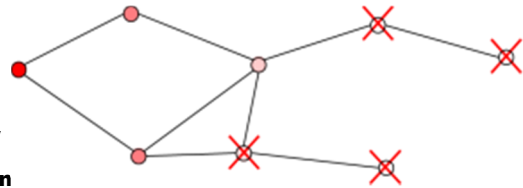
$$s = cWs + (1 - c)e$$

$$= (1 - c)(e + cWe + c^2W^2e + \dots)$$

$S$  : PageRank vector  $C$  : Random walk probability

$W$  : Adjacency matrix  $e$  : Vector of elements of  $1/n$

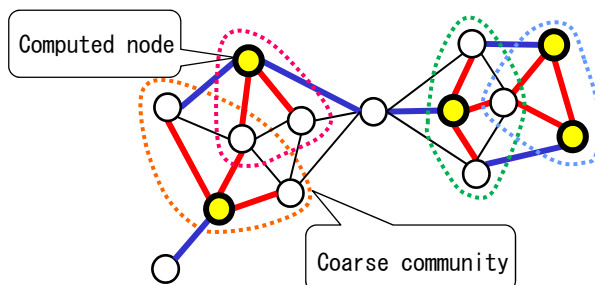
- Prune low PageRank nodes to reduce the size of graphs



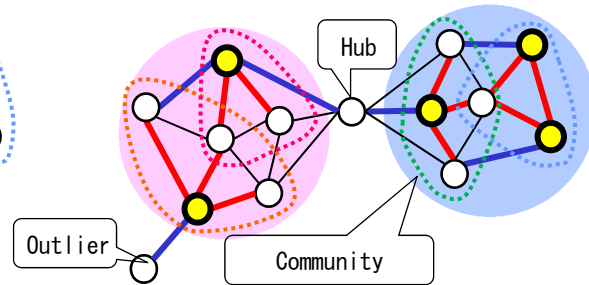
- Fast algorithm for graph clustering [2]

### Overview Efficiently detect communities, hubs and outliers from large-scale graph data

- Find coarse communities by computing 2-hop away nodes



- Refine communities by finding nodes that are belong to multi-communities



## Related work

[1] Y.Fujiwara, M.Nakatsuji, H.Shiokawa, T.Mishima, M.Onizuka, "Fast and exact top-k algorithm for pagerank," In *Proc. the 27th AAAI Conference on Artificial Intelligence (AAAI2013)*, 2013.

[2] 塩川浩昭, 藤原靖宏, 鬼塚真, "構造的類似度に基づくグラフクラスタリングの高速化," 第6回データ工学と情報マネジメントに関するフォーラム(DEIM Forum 2014), D6-2, 2014 (in Japanese).

## Contact

**Makoto Onizuka** Distributed Computing Technology Project, NTT Software Innovation Center  
E-mail : onizuka.makoto{at}lab.ntt.co.jp (Please replace {at} with @)