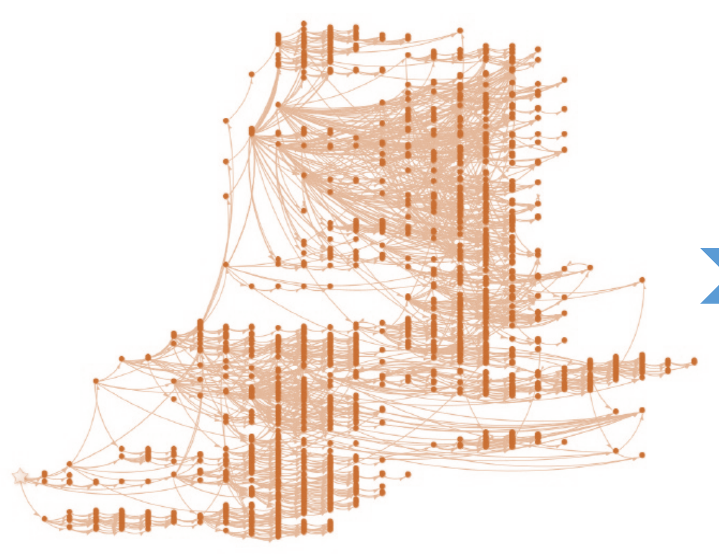
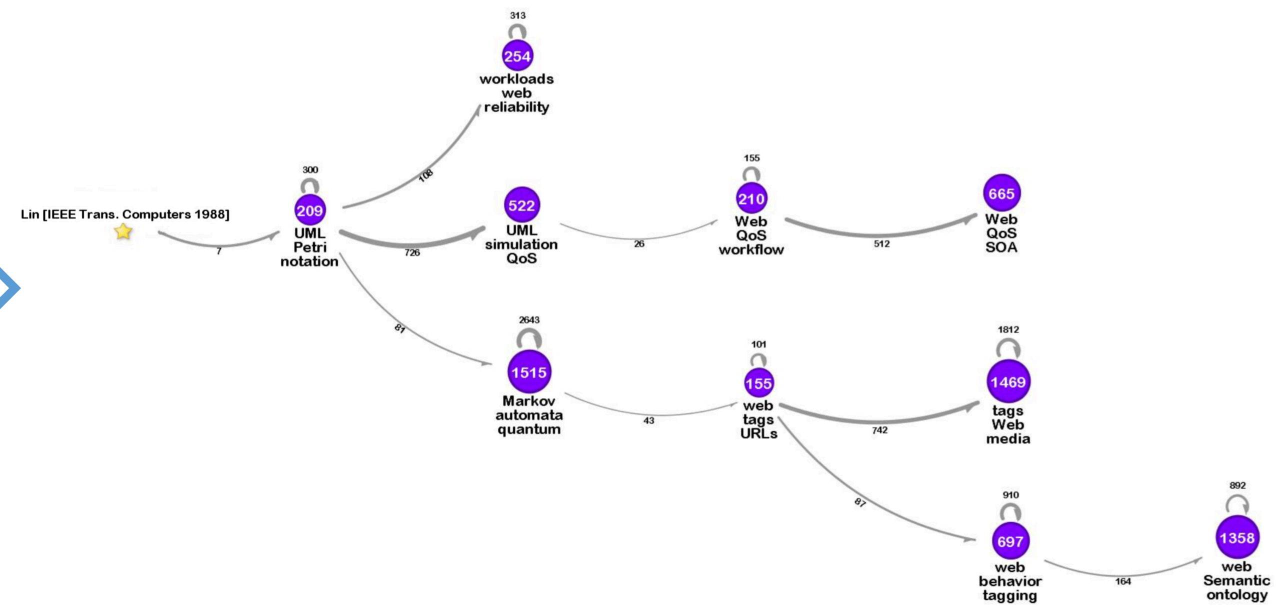
TOPIC: TOward Perfect InfluenCe Graph Summarization

Lei Shi, Sibai Sun, Yuan Xuan, Yue Su, Hanghang Tong, Shuai Ma, Yang Chen

Influence Graph Summarization



Raw Citation Graph



Citation Graph after Summarization

IGS Objective: $\max \sum_{s=1}^{k^2} r(\xi_s)$ which is equal to $\sum_{s=1}^{k^2} \frac{\sum_{v_i \in \pi_{S(\xi_s)}, v_j \in \pi_{T(\xi_s)}} a_{ij}}{\sqrt{|\pi_{S(\xi_s)}||\pi_{T(\xi_s)}|}}$

Motivating Problems on IGS

- Optimality
- Scalability
- Flexibility

New Perspective on IGS Optimality

$$\max \sum_{s=1}^{k^2} r(\xi_s) = \sum_{s=1}^{k^2} \frac{\sum_{v_i \in \pi_{S(\xi_s)}, v_j \in \pi_{T(\xi_s)}} a_{ij}}{\sqrt{|\pi_{S(\xi_s)}||\pi_{T(\xi_s)}|}} = \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{a_{ij}}{\sqrt{|\pi_{C(v_i)}||\pi_{C(v_j)}|}}$$

⇩ (as a quadratic form)

$$\max \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j = x^T A x = x^T A^* x \quad \text{where } x = (|\pi_{C(v_1)}|^{-\frac{1}{2}}, \cdots, |\pi_{C(v_n)}|^{-\frac{1}{2}})^T$$

⇩ (by min-max theorem)

$$\max \frac{x^T A^* x}{x^T x} = \lambda_1 \implies x^T A^* x = k \cdot \lambda_1 \implies x = \sqrt{k} q_1$$

⇩ (optimizing the largest component)

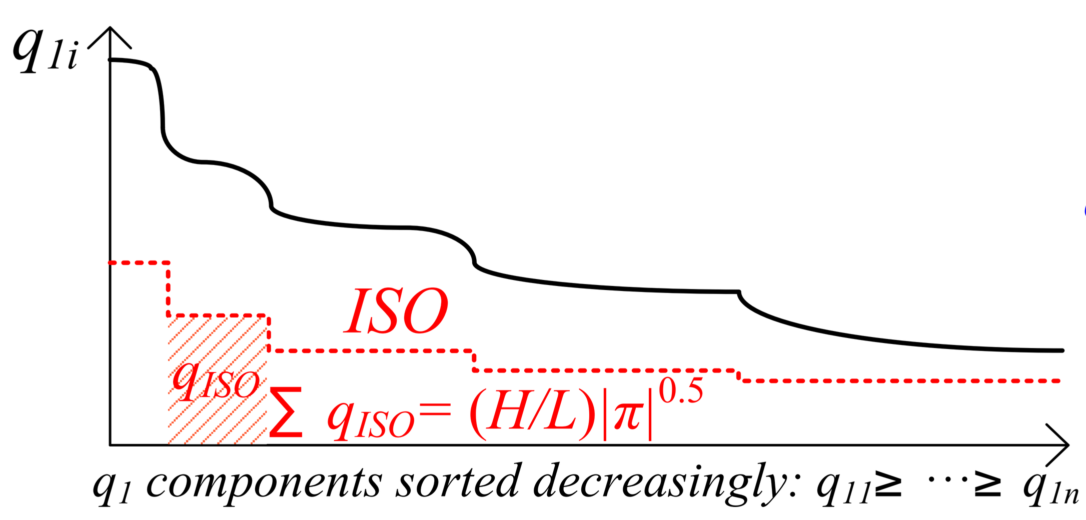
$$\max z_1 = q_1^T x = \sum_{j=1}^{k} |\pi_j|^{-\frac{1}{2}} \sum_{C(v_i)=j} q_{1i}$$

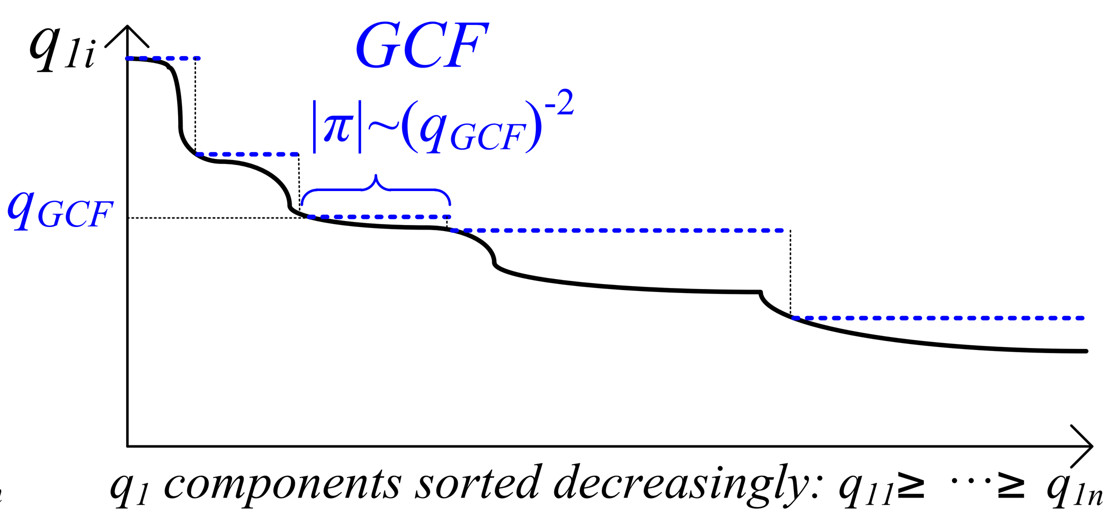
components sorted decreasingly: $q_{11} \geqslant \cdots \geqslant q_{1n}$

$$q_1 = \left[\underbrace{q_{11}}_{\pi_1} \; \underbrace{q_{12}}_{\pi_2} \; \underbrace{\cdots\cdots}_{\cdots\cdots} \; \underbrace{\cdots\cdots \; q_{1n}}_{\pi_k} \right]$$

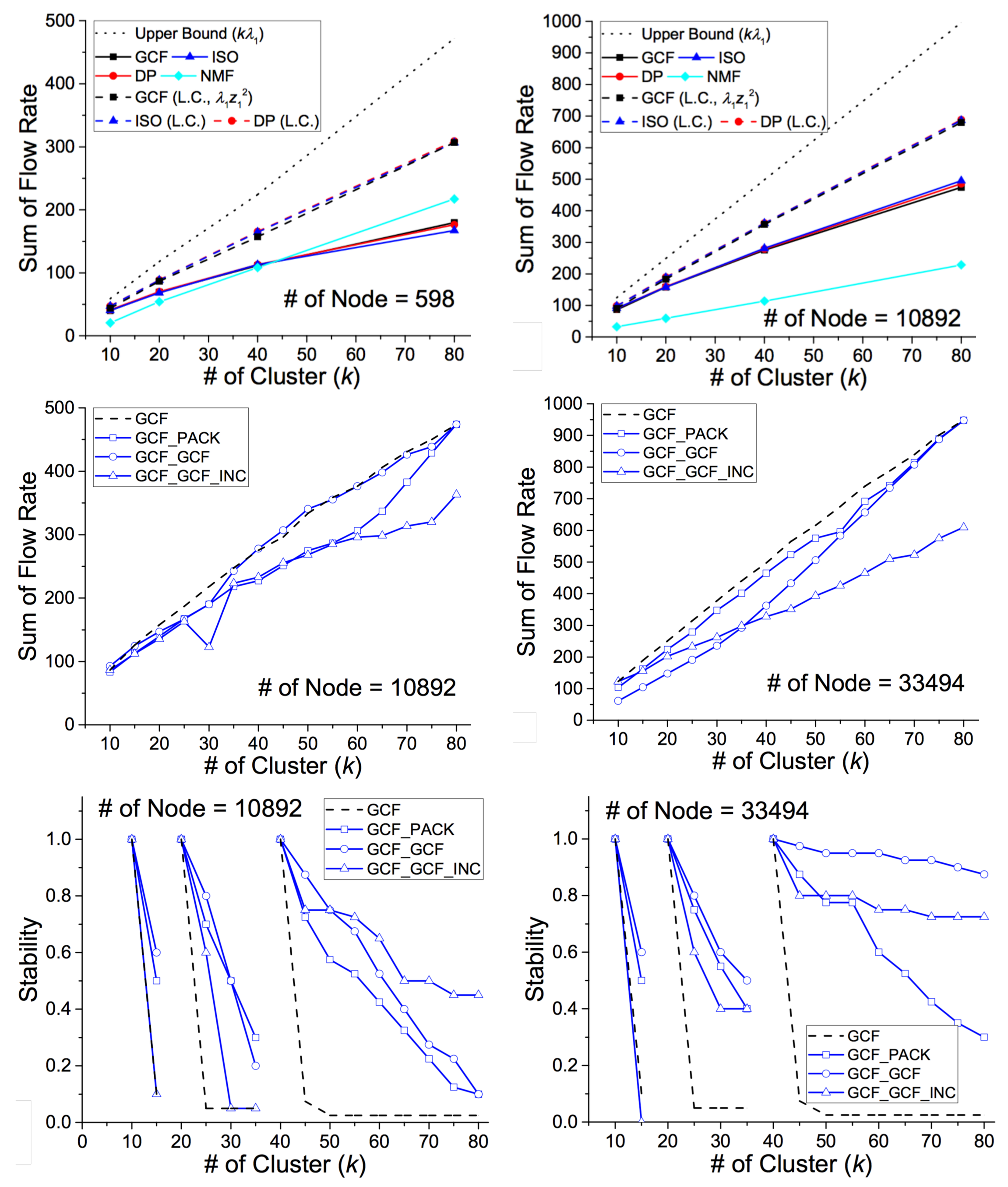
cluster size increasing: $|\pi_1| \leqslant \cdots \leqslant |\pi_k|$

Static Algorithms

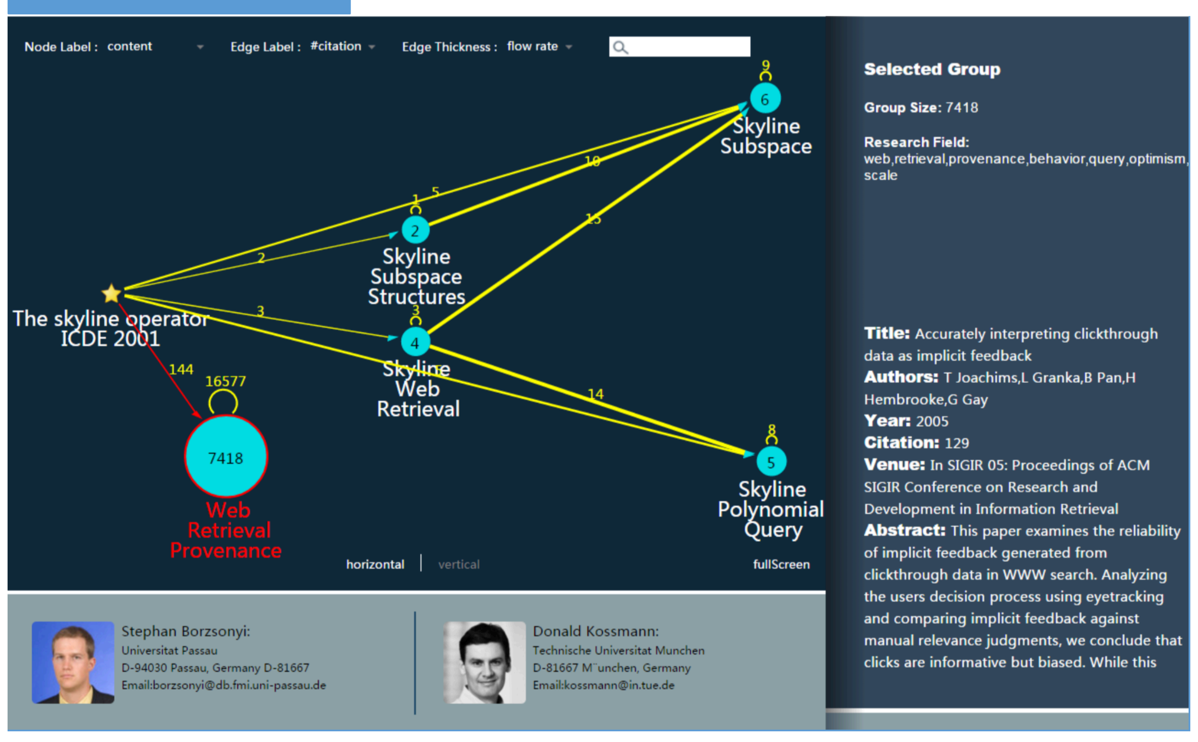




Evaluation



System



Dynamic Algorithm