

Effective and Efficient Conductance-Based Community Search at Billion Scale

Presented by: Ian Chen

University of Illinois Urbana-Champaign

August 15, 2026

Community Search

- Discovering communities in social networks
- Query-dependent formulation [2]
 - ▶ Organizing an event
 - ▶ Product recommendations
 - ▶ Link prediction
- Formally, given $Q \subset V$ connected query, find $H \subset V$ s.t.
 1. $Q \subset H, G[H]$ connected
 2. $f(G[H])$ maximized among all feasible H

Goodness Functions

- core (truss) number
- local modularity
- conductance based
 - ▶ Spectral methods fit in this group
 - ▶ This is what Lin et al. [1] focus on
- Previous conductance based methods have not enforced connectivity constraint

PPRCS Algorithm

1. Input is a single query vertex q
2. Rank the nodes by their *influence* on q (personalized page-rank)
 - 2.1 $\hat{\pi}(v)$ = Probability that a certain rank walk will reach v from q
3. Compute communities by threshold nodes by this influence
4. Reject all thresholds that lead to disconnected clusters

SCCS Algorithm

1. Find candidate subgraph via a distance restricted BFS
2. Find a seed graph (maximal clique)
3. Repeat greedy expansion and greedy pruning (i.e. optimize local node moves for changes in conductance)

Experimental Design

- Empirical networks with ground truth (largest Friendster, 1.8B edges); LFR networks
- Compete against other *local* community search algorithms; optimizing
 - ▶ coreness, truss
 - ▶ Modularity
 - ▶ Random walk
- Measure runtime, and size + accuracy of their community

Results

- They have the fastest method (on Friendster)
- SCCS closely matches the cluster size distribution
- Requires an objective that has internal + external connectivity, and method cannot be random walk based to match size

Takeaways

- Optimization function matters: cohesion vs modularity vs conductance
- Local search methods can scale independent of the size of the network

My Questions

- I don't think they describe how they selected the query nodes.
- I don't think they described what accuracy means if their query is in multiple ground-truth communities.
- Does it only make sense to try and find very small communities?
- What are other ways to evaluate community search quality? This paper mainly matches on size, but ground-truth communities in empirical networks are suspect.
- This paper suggests that diffusion based methods find large communities (same as what we've seen in Infomap). Why?

- [1] Longlong Lin et al. “Effective and Efficient Conductance-Based Community Search at Billion Scale”. In: *IEEE Transactions on Big Data* 11.6 (Dec. 2025), pp. 3170–3184. ISSN: 2372-2096. DOI: 10.1109/tbdata.2025.3588028. URL: <http://dx.doi.org/10.1109/TBDATA.2025.3588028>.
- [2] Mauro Sozio and Aristides Gionis. “The community-search problem and how to plan a successful cocktail party”. In: *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. KDD '10. ACM, July 2010, pp. 939–948. DOI: 10.1145/1835804.1835923.