Density Equivalence based Index(DEQ) for Efficient Community Search

Esra Akbas and Daniel Maliro

Department of Computer Science, Georgia State University

Problem Statement

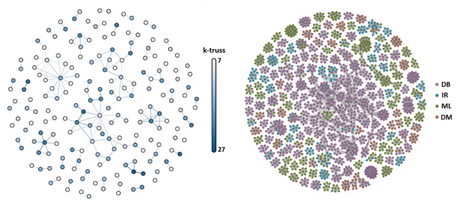
•Community search problem:

- Finding densely connected subgraphs within a large graph, specifically targeting communities containing a set of given query vertices, enabling personalized community discovery.

•Challenges and imitations of Other models:

- Expensive to search on large graph
- Different density measures, e.g. k-clique, k-core
  - Each one encode different information and results in different communities
- Current search models including indexes are for specific measures, not generalizable

Diagram: Summarized Graph and Communities of the DBLP Dataset



(a) The summarized graph **(b)** All k-truss communities

Proposed Model

•DENSITY EQUIVALENCE BASED INDEX(DEQ)

- Designed to support community search with optimal performance,
- Supports different cohesiveness measures including k-truss, k-core and -kedge
- Partitions graph edges into mutually exclusive equivalence classes,
  - encoding cohesive communities within supernodes while
  - maintaining their connectivity across communities via super-edges.
- O(|E|) time and O(|E|) space

Models

•Density Measures

- **k-truss:** the largest subset of a graph where each edge is contained in at least *k*−2 triangles within that subset.
- **k-core:** the largest subgraph where each vertex has a degree of at least k.
- **k-edge-connected (k-ECC) :** A graph is considered connected under the k-ECC model if it remains connected even after removing fewer than k edges.

Diagram: A representation of a k-truss model

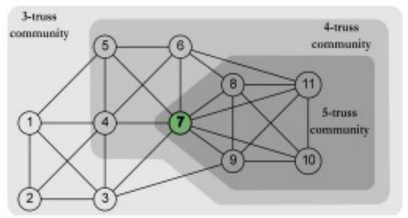
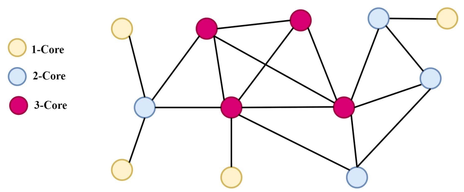


Diagram: A representation of a k-core model



Experiments

•Query time with respect to Graph Size:

- Query time for community search on subgraphs' index with different sizes, given specific query sets across all generated subgraphs..

•Query Time with respect to Query Set :

- Query time for community search on large graphs varying query set densities
  - by adjusting the k value to manipulate density

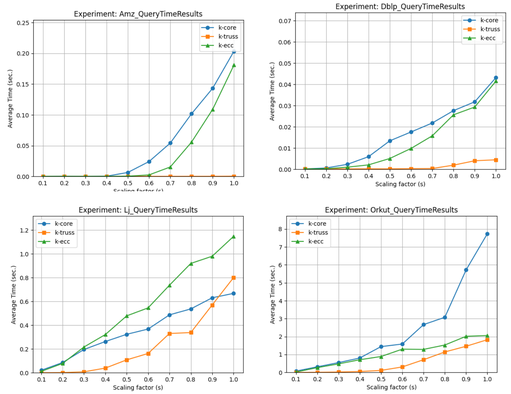
Datasets

Table: Vertex count and Edge count of undirected graphs

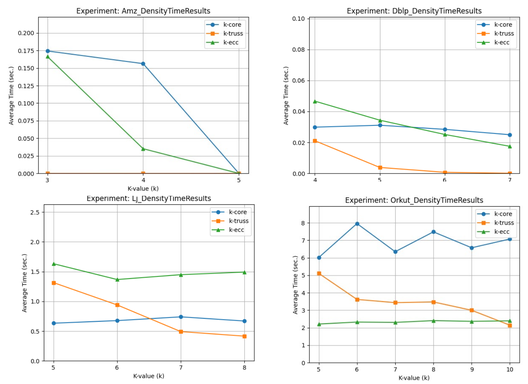
Network	|V|	|E|
Amazon	335K	926K
DBLP	317K	1M
LiveJournal	4M	35M
Orkut	3.1M	117M

Results

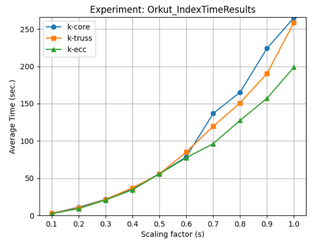
Tables: Performance on Querying 100 sets in Subgraphs of the datasets based on scaled subgraph



Tables: Performance on Querying 100 sets in Subgraphs of the datasets based on query density/k-value



Tables: Index Time for Dataset Orkut Subgraphs



Conclusion and Future Work

•DEQ indexing support efficient community search on large graphs

- Without repetitive access to original graph

•As the future work, we will compare our results with baseline models.

REFERENCES

1. Esra Akbas and Peixiang Zhao. 2017. Truss-based Community Search: a Truss-equivalence Based Indexing Approach
2. Tianyang Xu, Zhao Lu, and Yuanyuan Zhu. Efficient Triangle-Connected Truss Community Search In Dynamic Graphs. PVLDB, 16(3): 519 - 531, 2022.
3. E. Akbas, "Index Based Efficient Algorithms For Closest Community Search," 2019 IEEE International Conference on Big Data (Big Data)

Contact/Additional Info