DirDense: A Tool for Mining Dense Subgraphs from a Big Directed Graph

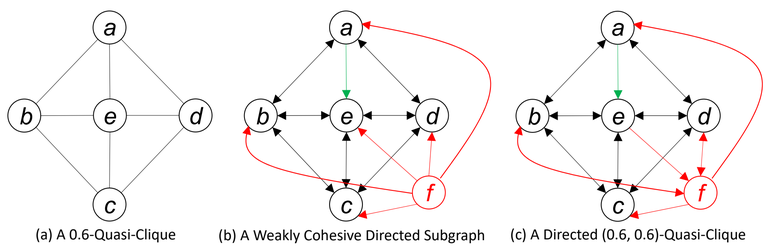
Jalal Khalil[2], Ahmad Akhlaque[1], Da Yan[1], Lyuheng Yuan[1],
Saugat Adhikari[1], Jiao Han[1], Yang Zhou[3], Zhe Jiang[4]

[1]Indiana University Bloomington, [2]St. Cloud State University,
[3]Auburn University, [4]University of Florida

The Problem

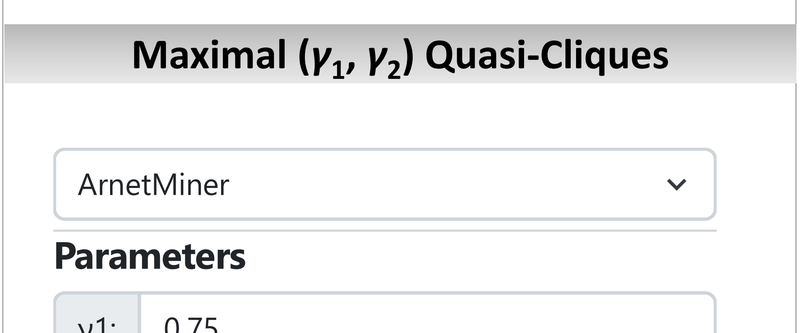
Given a big graph G, find all dense directed structures (aka. subgraphs) in G:

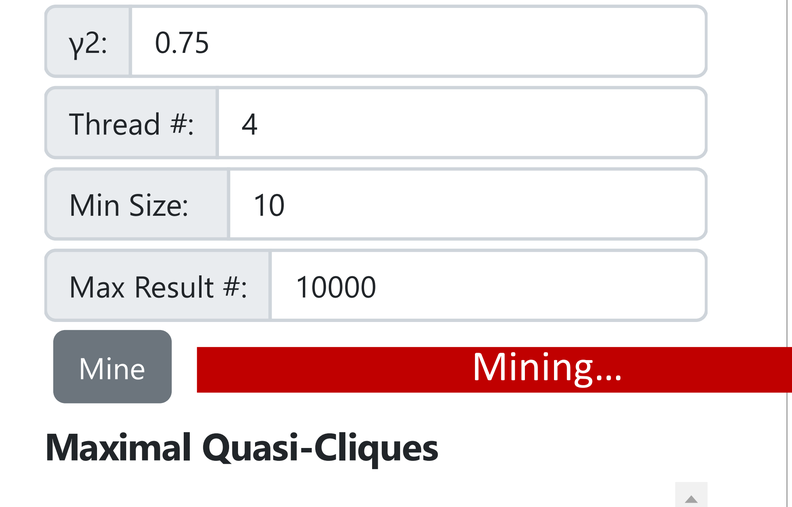
- Maximal (γ_1, γ_2)-quasi-cliques,
- Maximal (k_1, k_2)-plexes
- Directed densest subgraph



(a) A 0.6-Quasi-Clique (b) A Weakly Cohesive Directed Subgraph (c) A Directed (0.6, 0.6)-Quasi-Clique

System Architecture





Maximal (γ_1, γ_2) Quasi-Cliques

ArnetMiner

Parameters

γ1: 0.75

γ2: 0.75

Thread #: 4

Min Size: 10

Max Result #: 10000

Mine

Mining...

Maximal Quasi-Cliques

Quasi-Clique 1, Size: 11

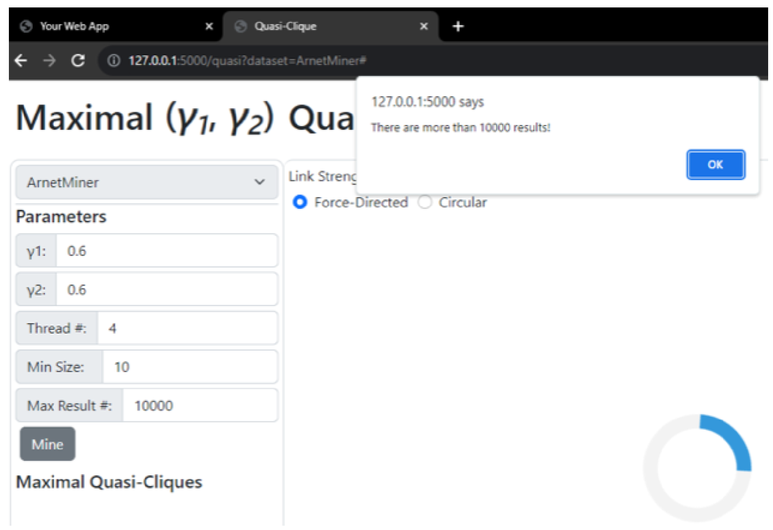
Quasi-Clique 2, Size: 10

Quasi-Clique 3, Size: 10

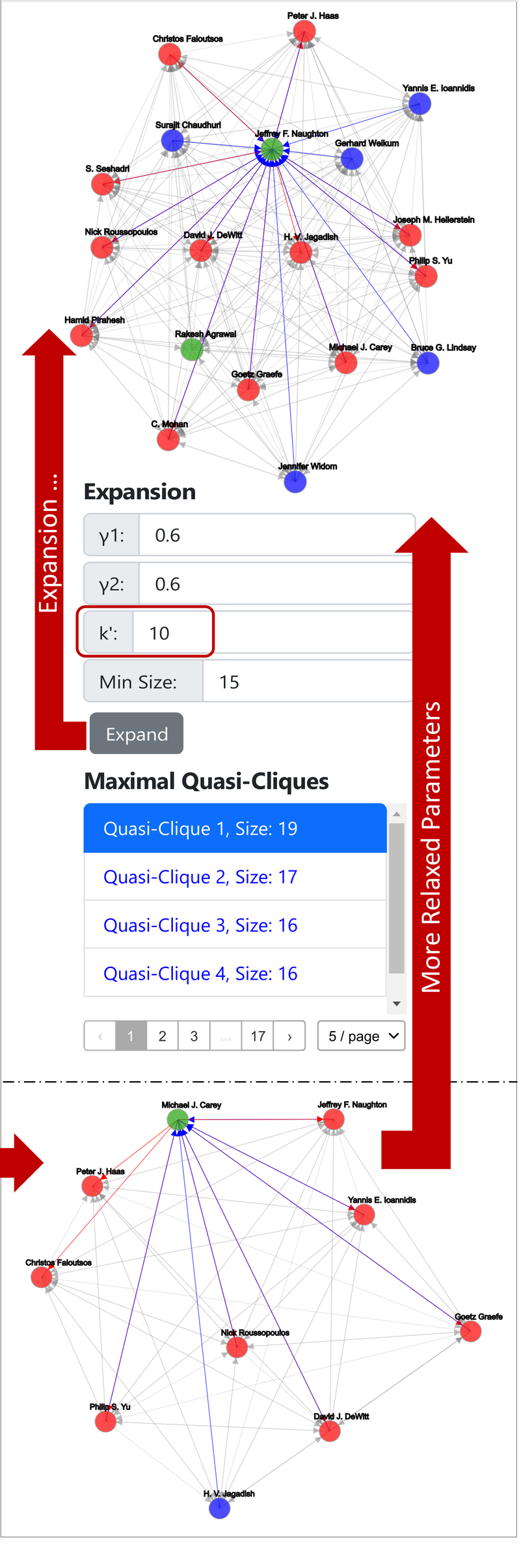
Quasi-Clique 4, Size: 10

1 2 3 4 5 › 5 / page

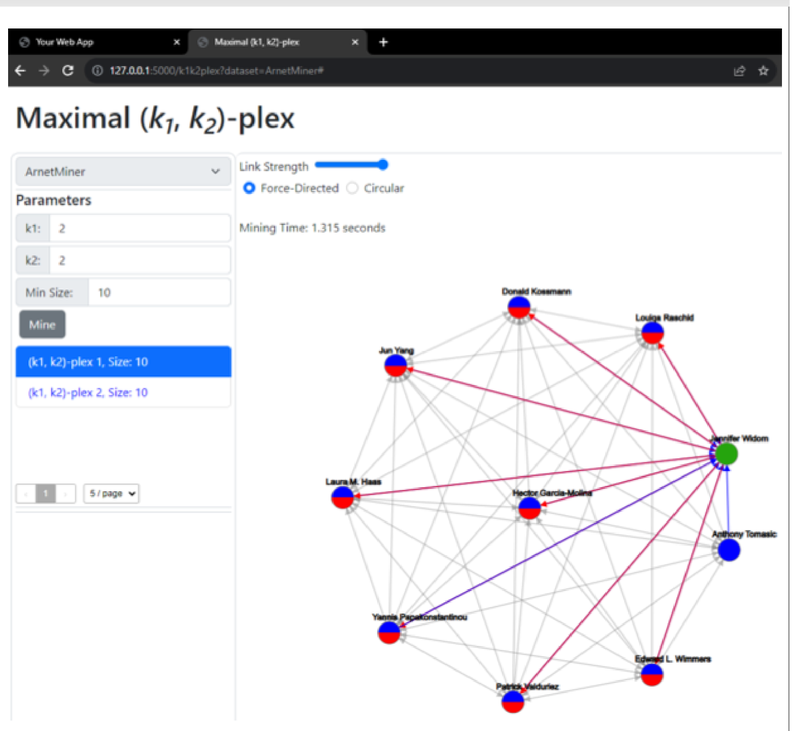
Prompt When Params Are Not Selective

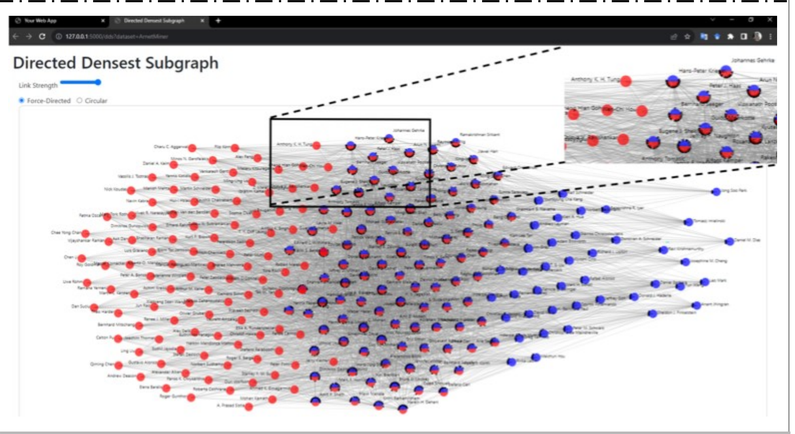


Task Cascading



Other Dense Directed Structures





Backend with T-thinker

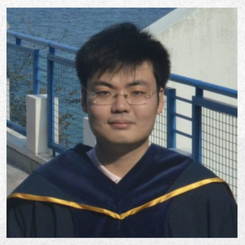
Think-Like-A-Task (TLAT) Model: Divide and Conquer

CCC Computing Community Consortium Catalyst

Catalyzing the computing research community and enabling the pursuit of innovative, high-impact research.

ABOUT VISIONING LEADERSHIP DEVELOPMENT TASK FORCES RESOURCES BLOG PODCAST

The following Great Innovative Idea is from Da Yan, tenure-track assistant professor in the Department of Computer Sciences at the University of Alabama at Birmingham (UAB). Yan presented his poster, T-thinker: A Task-Centric Framework to Revolutionize Big Data Systems Research, at the Computing Community Consortium (CCC) Early Career Researcher Symposium, August 1-2, 2018.



The Idea

Big Data frameworks such as Apache Hadoop and Apache Spark are becoming increasingly popular due to their emphasis on ease of programming, but they are dominantly designed for data-intensive iterative computations, and there lacks an efficient solution to compute-intensive Big Data analytics. Based on my insight that compute-intensive problems are often solved by divide and conquer (e.g., a recursive algorithm), a general task-centric framework, called T-thinker, is developed for compute-intensive Big Data problems. The framework effectively utilizes the CPU cores in a cluster by properly dividing a problem over a big dataset into tasks over smaller subsets of the dataset, and by overlapping CPU processing with network communication (e.g., for requesting a subset of dataset).

References & Related Work

- Jalal Khalil et al., **DirDense: A Tool for Mining Dense Subgraphs from a Big Directed Graph.** CIKM'24.
- Da Yan et al., **Systems for Scalable Graph Analytics and Machine Learning: Trends and Methods (Tutorial).** CIKM'24.
- Jalal Khalil et al., **FSM-Explorer: An Interactive Tool for Frequent Subgraph Pattern Mining from a Big Graph.** ICDE'24 (Best Demo Runner-up).
- Lyuheng Yuan et al., **Faster Depth-First Subgraph Matching on GPUs.** ICDE'24.
- Lyuheng Yuan et al., **T-FSM: A Task-Based System for Massively Parallel Frequent Subgraph Pattern Mining from a Big Graph.** SIGMOD'23.
- Guimu Guo et al., **Maximal Directed Quasi-Clique Mining.** ICDE'22.
- Jalal Khalil et al., **Parallel Mining of Large Maximal Quasi-Cliques.** VLDBJ'22.
- Da Yan et al., **G-thinker: A General Distributed Framework for Finding Qualified Subgraphs in a Big Graph with Load Balancing.** VLDBJ'22.
- Da Yan et al., **PrefixFPM: A Parallel Framework for General-Purpose Mining of Frequent and Closed Patterns.** VLDBJ'22.
- Guimu Guo et al., **Scalable Mining of Maximal Quasi-Cliques: An Algorithm-System Codesign Approach.** PVLDB'21
- Da Yan et al., **G-thinker: A Distributed Framework for Mining Subgraphs in a Big Graph.** ICDE'20.
- Da Yan et al., **T-thinker: A Task-Centric Distributed Framework For Compute-Intensive Divide-and-Conquer Algorithms.** PPoPP'19.

Acknowledgement & Contact

Supported by DOE ECRP Award 000025228, NSF OIA-2229394, South Big Data Innovative Hub 2022 S.E.E.D.S. Award

Contact: jalal.khalil@stcloudstate.edu (Jalal Khalil),
yanda@iu.edu (Da Yan)