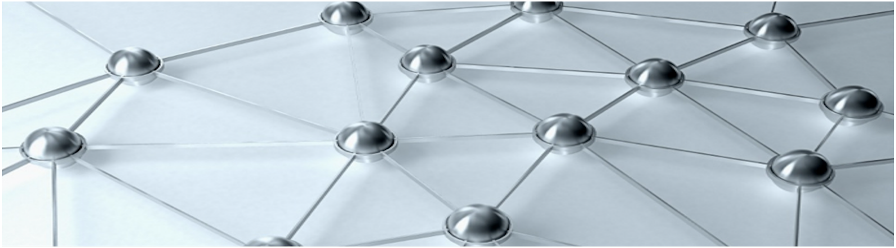


Dense Subgraph Discovery Meets Strong Triadic Closure

Chamalee Wickrama Arachchi, Iiro Kumpulainen, Nikolaj Tatti

University of Helsinki, Finland

KDD'24

Dense subgraph discovery

- Finding dense subgraphs has many applications in diverse domains such as community detection, biological system analysis, and anomaly detection.
- Definition of density :

$$d(S) = \frac{|E(S)|}{|V(S)|}$$

- Densest subgraph problem : Find a subset of vertices which maximizes the density.
- Exact[1] and greedy[2] algorithms.

---

[1] Andrew V Goldberg. Finding a maximum density subgraph. 1984.

[2] Moses Charikar. Greedy approximation algorithms for finding dense components in a graph. In International workshop on approximation algorithms for combinatorial optimization, pages 84–95. Springer, 2000.

STC-constrained dense subgraph discovery problem

Input:

- A graph.
- Weight parameter : λ.

Output:

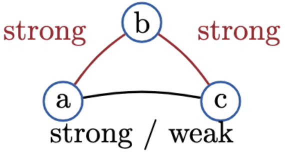
- A subgraph.
- A labeling of each edge as either strong or weak within the subgraph.

Such that:

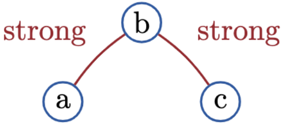
- Our density-like score $\frac{|S|+\lambda|W|}{|U|}$ is maximized.
- Stong edges should obey strong triadic closure (which we define next).

Strong triadic closure (STC)

- If (a, b) and (b, c) are strongly connected, there should be at least a weak connection between a and c.



✓



✗

Our problem : STC-DEN

- Given a graph $G = (V, E)$ and a weight parameter λ,
  find a subset of vertices $U \subseteq V$ and
  split $E(U)$ into two sets: set of strong edges S and weak edges W
  such that the STC property is satisfied in $(U, E(U))$ and
  $\frac{|S|+\lambda|W|}{|U|}$ is maximized.

- Our problem is **NP**-hard.

- The case of $\lambda = 0$ reduces to the MAX-CLIQUE problem.

- The case of $\lambda = 1$ reduces to the densest subgraph problem which is polynomial time solvable.

Auxiliary problem : $\textsc{stc-den}(\alpha)$

- Given a graph $G = (V, E)$, a weight parameter λ, and a number α,
  find a subset of vertices $U \subseteq V$ and
  split $E(U)$ into two sets: set of strong edges S and weak edges W
  such that the STC property is satisfied in $(U, E(U))$ and
  $|S| + \lambda|W| - \alpha|U|$ is maximized.

ILP formulation : $\textsc{stc-den}(\alpha)$

$$
\begin{array}{lrr}
\textsc{maximize} & \sum x_{ij} + \lambda \sum z_{ij} - \alpha \sum y_i & \\
\textsc{subject to} & x_{ij} + z_{ij} \leq y_i & ij \in E \\
 & x_{ij} + z_{ij} \leq y_j & ij \in E \\
 & x_{ij} + x_{jk} \leq y_j & (i,j,k) \in Z \\
 & x_{ij}, z_{ij} \in \{0,1\} & ij \in E \\
 & y_i \in \{0,1\} & i \in V
\end{array}
$$

- α : Guess parameter.
- y_i : Whether the node $i \in U$.
- x_{ij} : Whether the edge (i, j) is strong or not.
- z_{ij} : Whether the edge (i, j) is weak or not.

- $\textsc{stc-den}(\alpha)$ yields the same solution as $\textsc{stc-den}$ for the largest α with a non-empty solution.

LP relaxation

- Relax the integer program formulation.
- Construct a collection of sets $\mathcal{S} = \{S_1, S_2, \ldots, S_n\}$ where $S_j = \{i : y_i^* \geq y_j^*\}$.
- Order the edges based on z_e^* and label edges as strong if the STC property is not violated.
- Out of all subgraphs, pick the subgraph and labeling that maximizes our score.

Algorithms based on label, find the densest subgraph, and relabel

Algorithm 1: $\text{STC-Cut}(G, \lambda)$ and $\text{STC-Peel}(G, \lambda)$.

$L \leftarrow$ find an approximate labeling [3];
$H \leftarrow$ the weighted graph by setting 1 to strong edges and λ to weak edges of G;
$U \leftarrow$the densest subgraph (exact [1] or approximate [2]);
$L' \leftarrow$ find an approximate labeling [3];
return subgraph U and its labeling L';

[1] Andrew V Goldberg. Finding a maximum density subgraph. 1984.

[2] Moses Charikar. Greedy approximation algorithms for finding dense components in a graph. In International workshop on approximation algorithms for combinatorial optimization, pages 84–95. Springer, 2000.

[3] Sintos, S., Tsaparas, P. (2014, August). Using strong triadic closure to characterize ties in social networks. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1466-1475).

Greedy algorithm

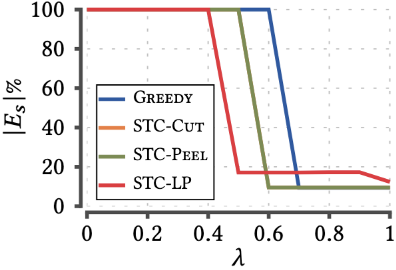
Algorithm 2: Greedy algorithm.

$U \leftarrow V$;
while *there are nodes* **do**
    $L \leftarrow$ find an approximate labeling [1] ;
    $u \leftarrow \arg\min_{v \in U} \deg_\lambda(v, U, L, \lambda)$;
    $U \leftarrow U \setminus \{u\}$;
return best tested U and its labeling L;

[1] Sintos, S., Tsaparas, P. (2014, August). Using strong triadic closure to characterize ties in social networks. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 1466-1475).

Experiments with Synthetic Datasets

- $|E_s|\%$: The percentages of *strong* edges in discovered subgraph.



Experiments with Real-world Datasets

- STC-ILP always give the highest score while all other algorithms produce approximately equal scores.
- When STC-ILP is not usable, STC-LP produce the maximum score in general.
- For high λs all of them produce less deviated scores when compared to lower λs.
- In general, $|E_s|\%$ monotonically decreases as λ increases.
- Greedy, STC-LP, and STC-Cut run significantly slower than STC-Peel.

Conclusion

- Introduced a novel dense subgraph discovery problem that takes into account the strength of ties within the subgraph.
- Our goal was to maximize a density-like measure defined as the sum of the number of strong edges and the number of weak edges weighted by a weight parameter, divided by the number of nodes within the subgraph.
- Proved that our problem is **NP**-hard.
- Designed an exact algorithm based on integer linear programming.
- Designed linear programming, a greedy heuristic algorithm, and two other straightforward algorithms based on the algorithms for the densest subgraph discovery which run in polynomial time.
- Showed experimentally that the algorithms could find the ground truth using synthetic dataset.
- Showed experimentally that our proposed algorithms discovered the subgraphs reasonably fast in practice.

Thank you for your attention!!!