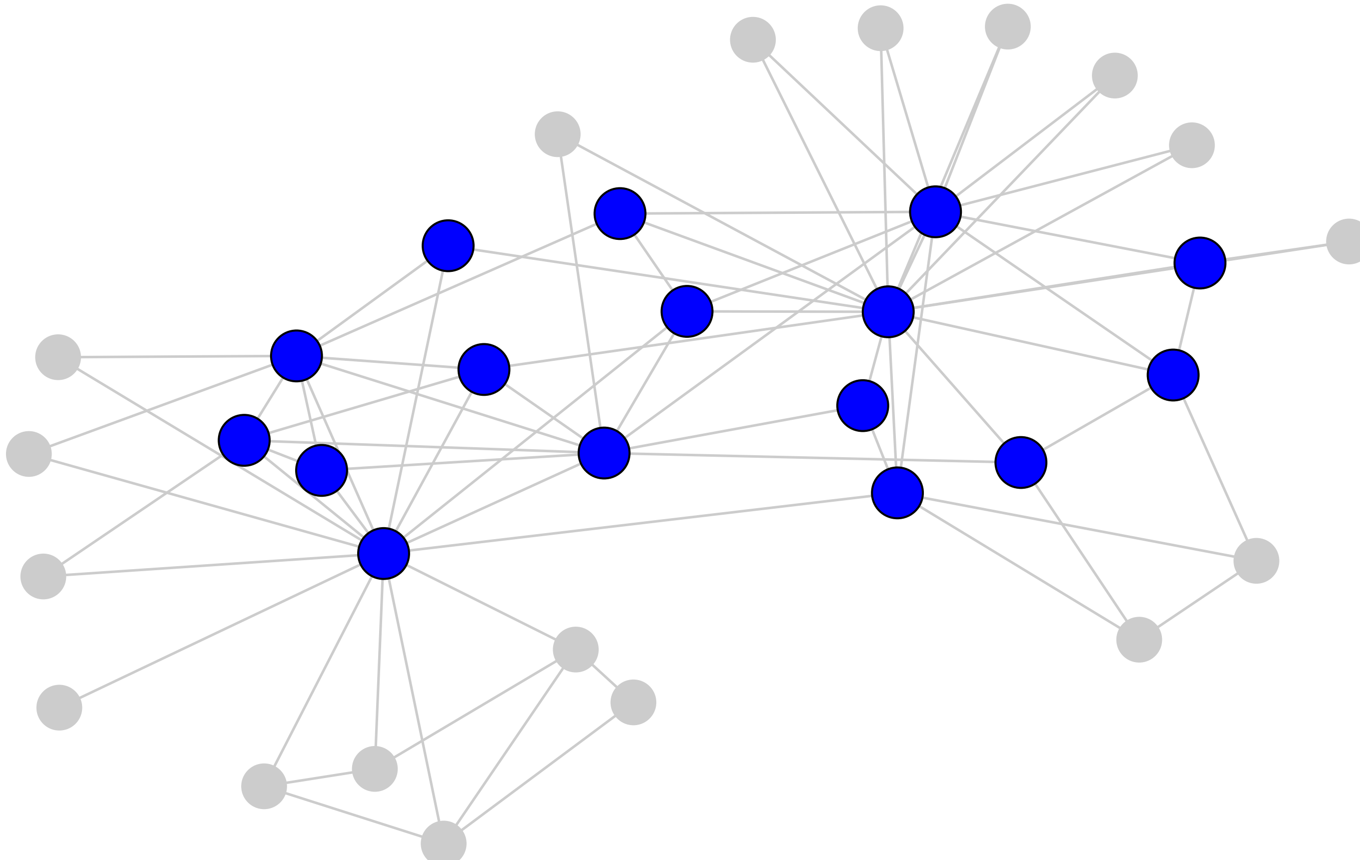


Dense Subgraph Discovery

Sept 20, 2022

Advanced Graph Algorithms

Finding a dense subgraph is a fundamental graph mining primitive

Applications

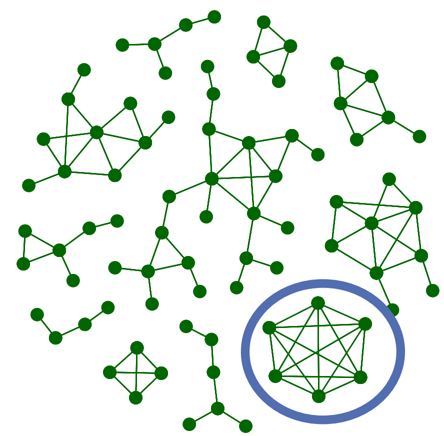
- detecting DNA motifs
- finding modules in gene co-expression data
- identifying trending topics in social media

⋮

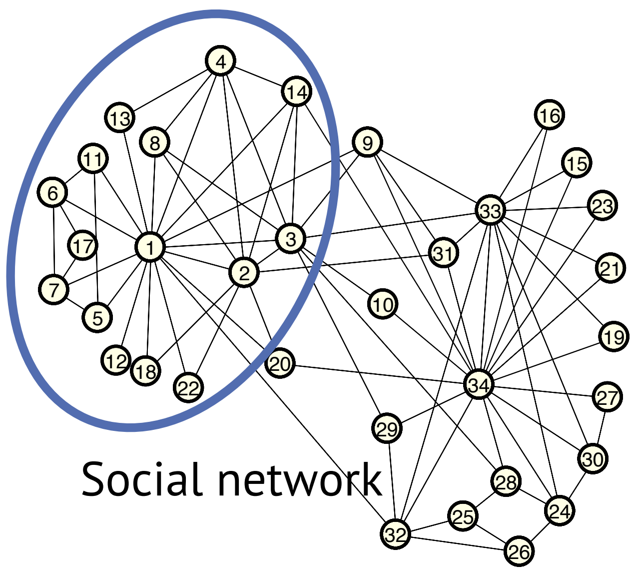
Objectives

- maximum clique
- densest subgraph
- k-core
- k-truss
- clique
- quasi-clique
- nucleus decompositions

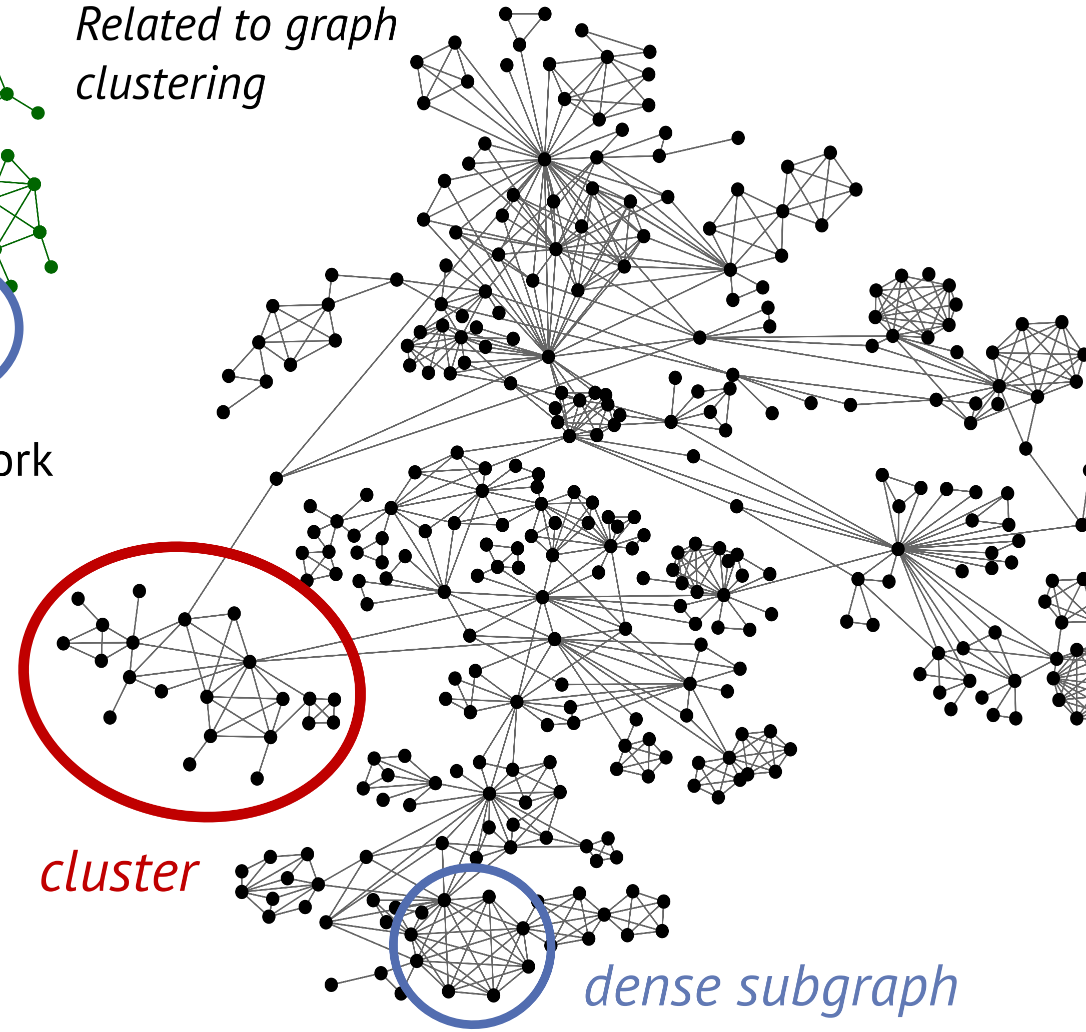
⋮



Biological network

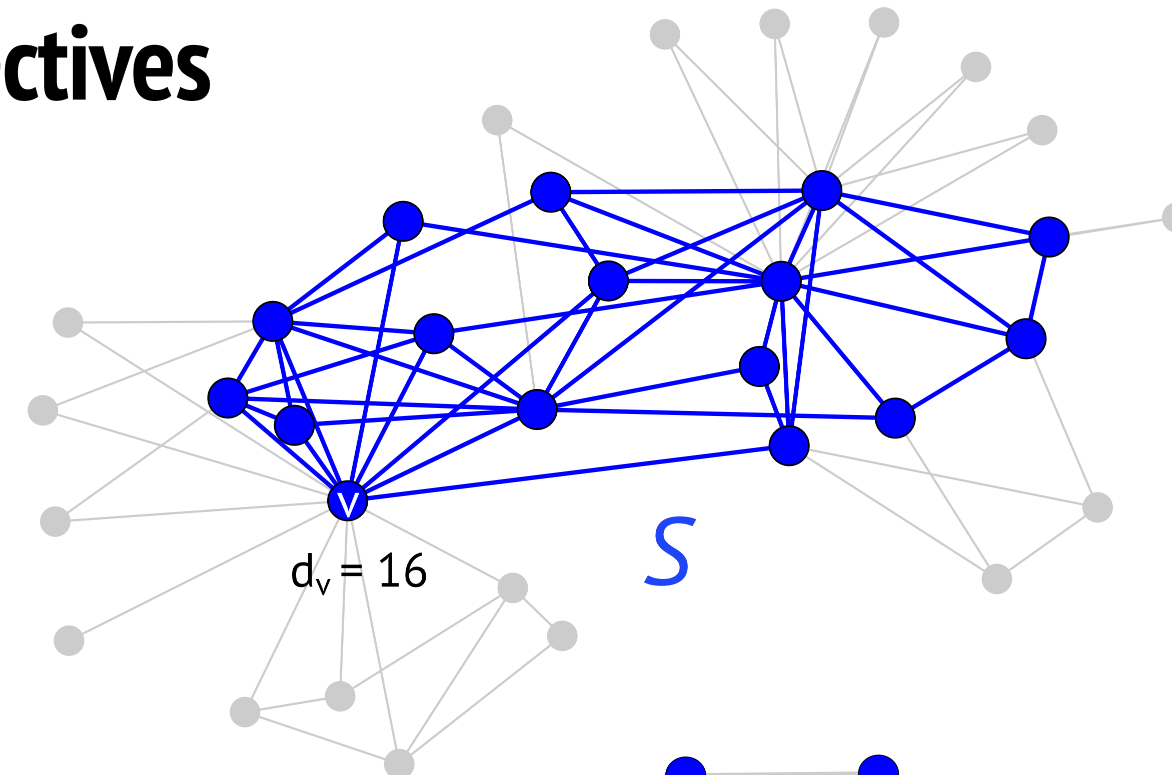


Social network



Collaboration network

The densest subgraph problem is one of the most common objectives

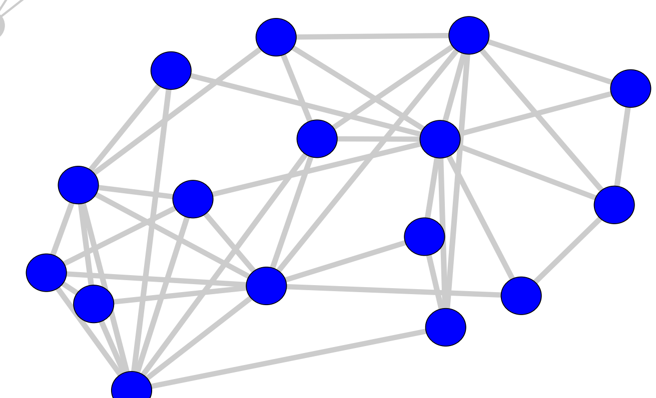


Densest Subgraph Problem
Maximize the ratio between number of nodes and number of edges.

number of edges between S nodes

$$\max_{S \subseteq V} \frac{|E(S)|}{|S|} = \frac{1}{2} \frac{\sum_{v \in S} d_v(S)}{|S|}$$

Equivalently: max average degree
Find subgraph S with maximum average (induced) degree

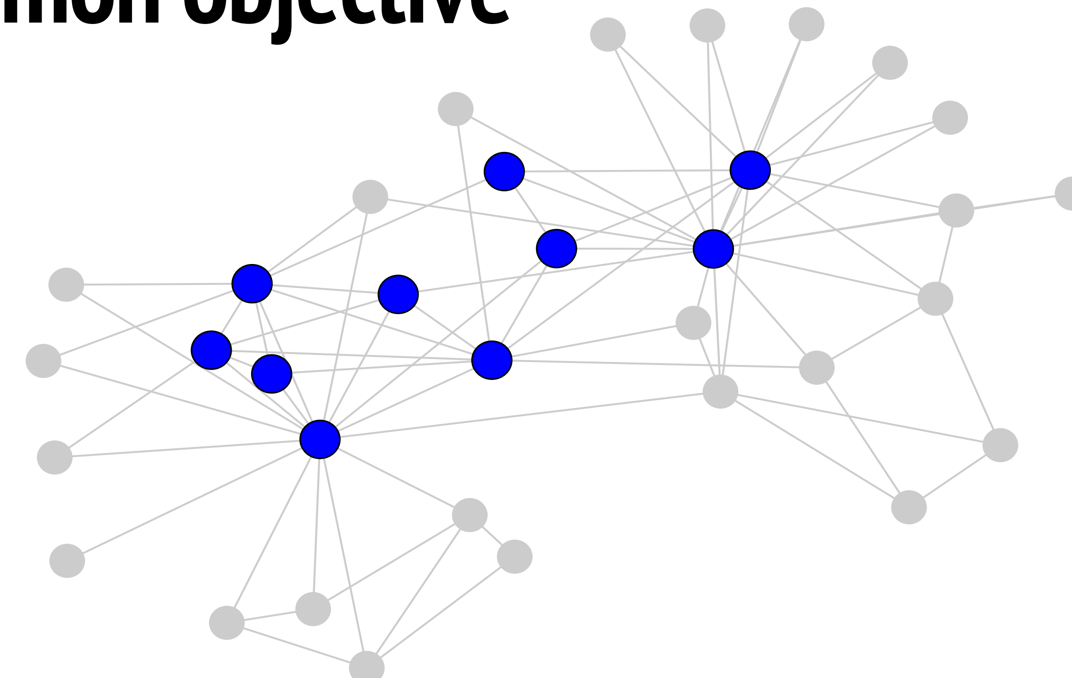


$d_v(S)$ = 8

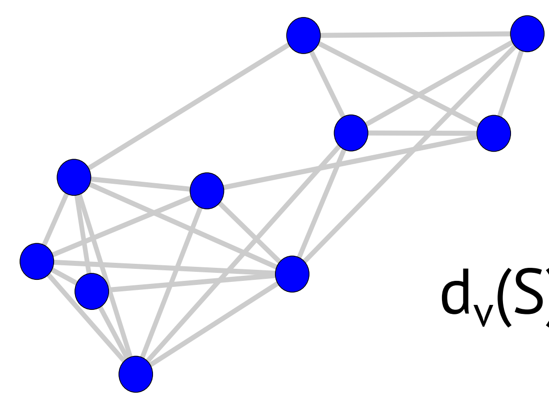
$d_v(S)$ = degree of v in induced subgraph

The k-core is another common objective

k-core
Maximal subgraph where all nodes have induced degree at least k



degeneracy
maximum k such that k-core is nonempty



$d_v(S) \geqslant 4$

We'll cover a few techniques for solving these problems

Exact algorithms based on minimum s-t cuts

Algorithms based on "peeling"

Exact algorithms based on linear programming

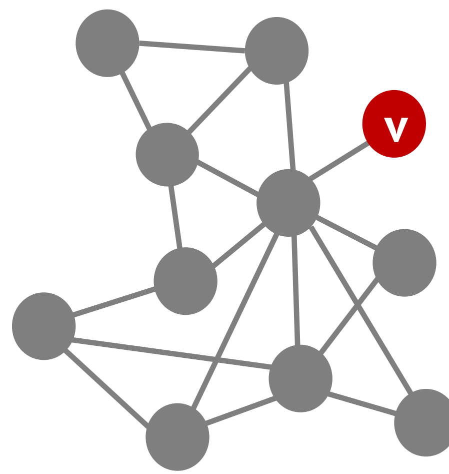
Peeling algorithms provide a fast way to approximate these objectives

GreedyPeel: repeatedly remove the minimum degree node from the graph.

Want small changes to numerator.

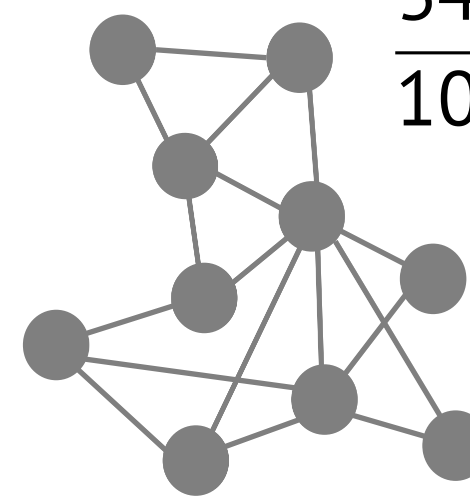
$$\frac{\sum_{v \in S} d_v(S)}{|S|}$$

Removing any node changes |S| the same.



$$\frac{36}{11} = 3.27$$

|S| = 11



$$\frac{34}{10} = 3.4$$

Guarantees.

One of the subgraphs will be a 1/2-approx for densest subgraph [Charikar 2000].

One of the subgraphs returned will be the k-core, with k = degeneracy [Matula & Beck 1983].