UNSW
THE UNIVERSITY OF NEW SOUTH WALES

Cohesive Subgraph Detection

Clique Model

Never Stand Still | Faculty of Engineering | Computer Science and Engineering

Xuemin Lin

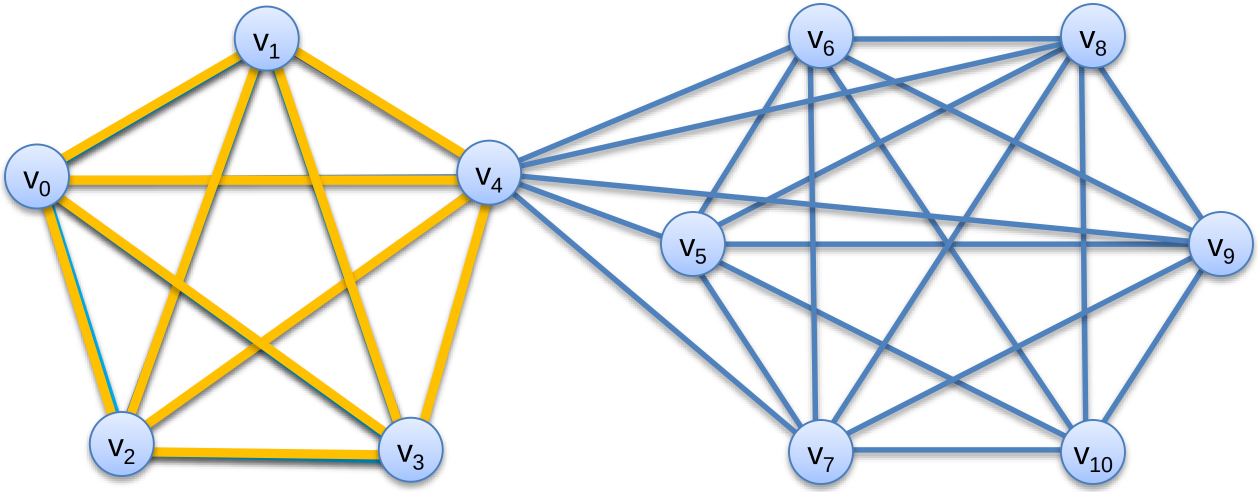
The University of New South Wales, Australia

Outline

- Clique Model
- K-Core vs K-Truss
- K-Edge Connected
- K-Vertex Connected

Clique

- Given a graph G, a clique is a set of nodes such that for any pair of them have an edge
- A clique is called maximal clique if there exist no other bigger cliques that contain it



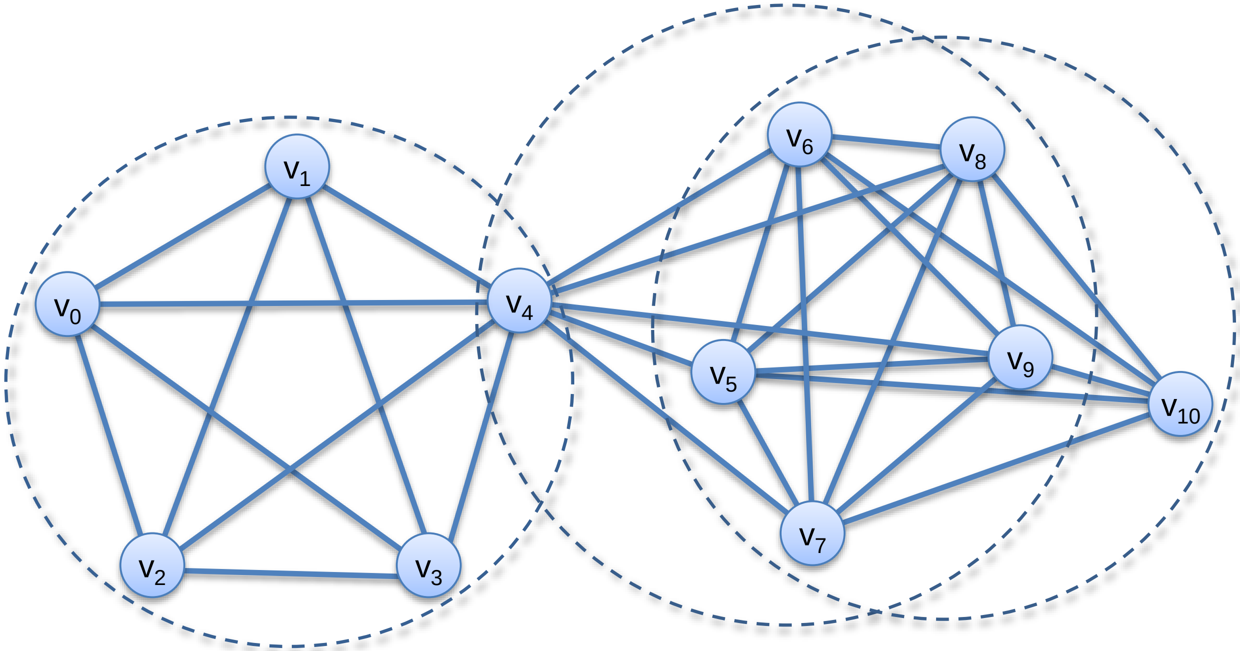
Application

- Community detection
- Gene expression and motif discovery
- Anomaly detection
- Stock market data visualization
- Signal transmission analysis
- ……

UNSW THE UNIVERSITY OF NEW SOUTH WALES

Maximal Clique Enumeration

- Given a graph G, find all the maximal cliques in G.
  - ➢ NP-Hard Problem



In-Memory MCE

- Bron-Kerbosch Algorithm
  - First Practical In-Memory MCE Algorithm
    - In-Memory means all the input and auxiliary data structure can be loaded in main memory during the computation
    - C. Bron et al., “Algorithm 457: finding all cliques of an undirected graph", *Communications of the ACM*, **16** (9): 575–577,1973
  - Based on a recursive backtracking paradigm

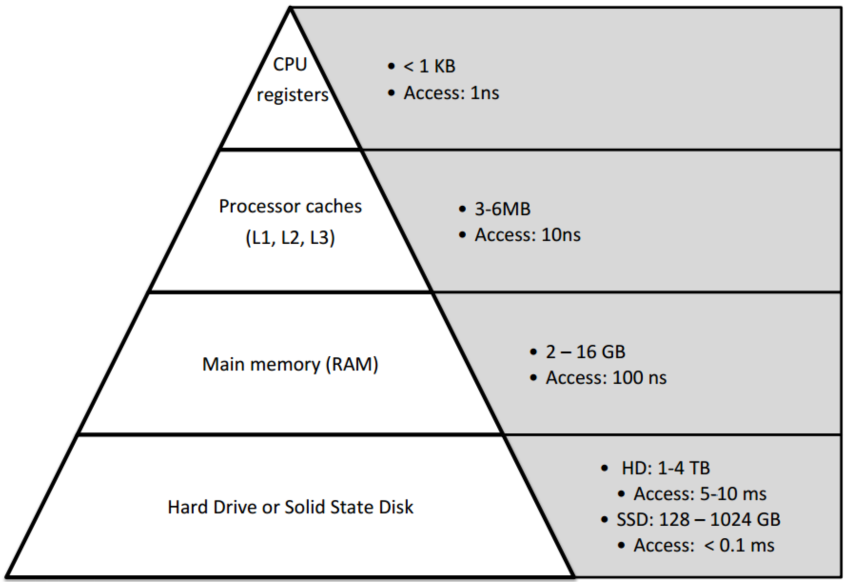
I/O Efficient MCE

- Why I/O Efficient?
  - Real graph is massive
    - Facebook contains 1.32 billion nodes and 140 billion edges
    - EU-2015 (sub-domain of web graph) contains 1.07 billion nodes and 91 billion edges
  - Memory is fast but small while disk is slow but large

UNSW THE UNIVERSITY OF NEW SOUTH WALES

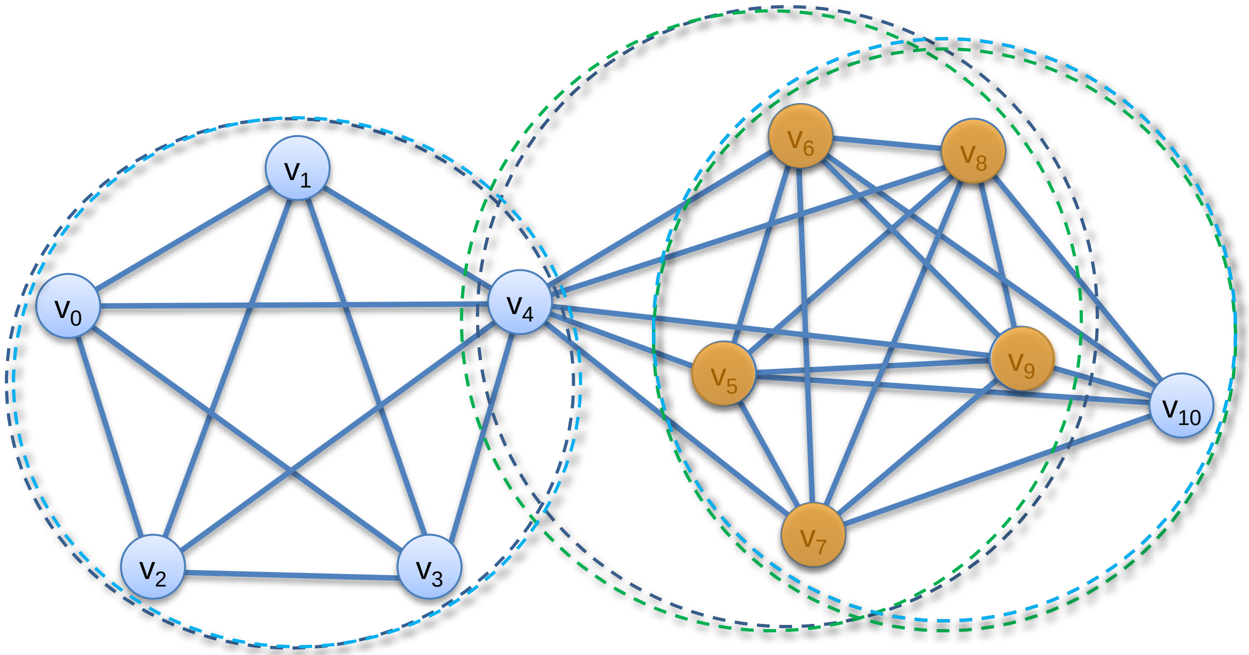
I/O Efficient MCE

- Why I/O Efficient?
  - Memory Hierarchy



Diversified Top-K Clique Search

- Traditional models vs diversified top-k clique



Diversified Top-K Clique Search

- Our Solution
  - ➢ treat it as an online k coverage problem
  - ➢ store k maximal cliques in memory
  - ➢ update these k candidate maximal cliques while enumerating cliques
    - replace small existing cliques with big new cliques

UNSW
THE UNIVERSITY OF NEW SOUTH WALES

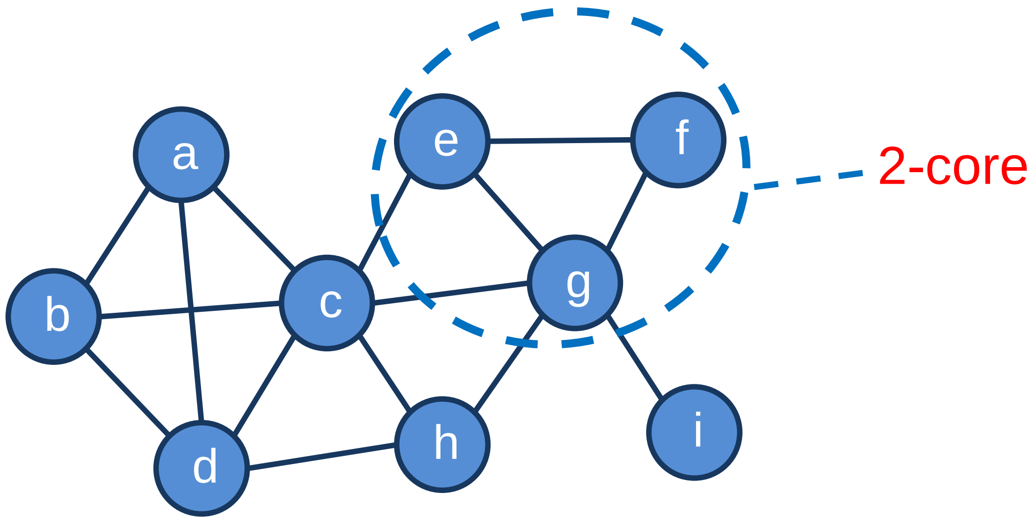
Diversified Top-K Clique Search

- PNP-Index
  - An naïve implementation for candidate set maintenance needs $O(|\mathcal{A}| * k * |C_{max}|)$
  - With the help of PNP-Index, our algorithm can only take $O(\sum_{c \in \mathcal{A}} |C|)$ time

UNSW THE UNIVERSITY OF NEW SOUTH WALES

k-Core

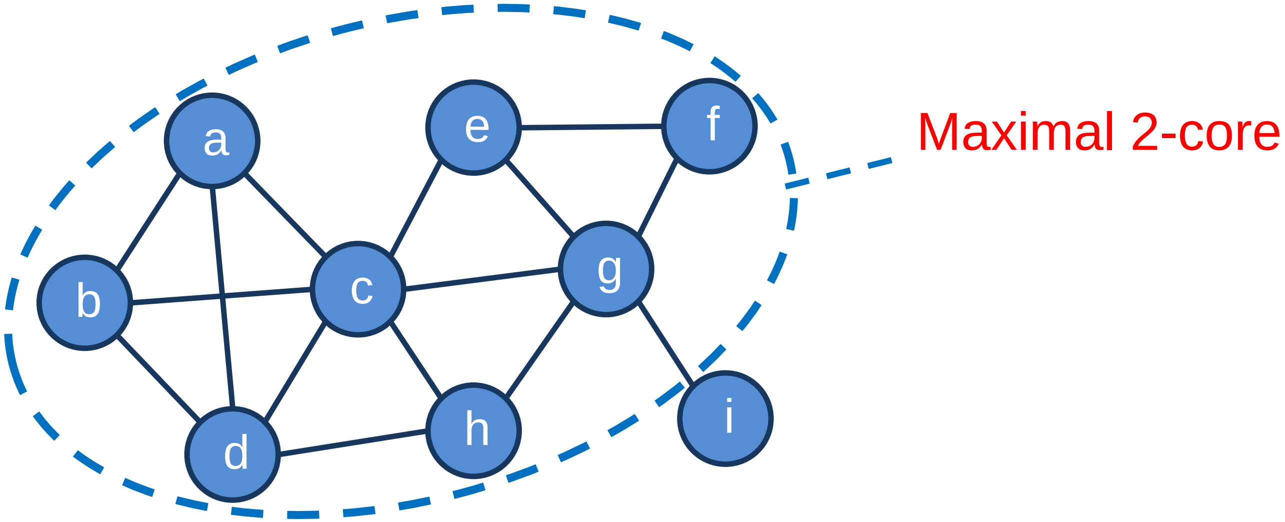
- Given a graph *G*, the *k*-core of *G* is a subgraph where each node has at least *k* neighbors (i.e., *k* adjacent nodes, or a degree of *k*).



S. B. Seidman. Network structure and minimum degree. Social networks, 5(3):269–287, 1983.

Maximal *k*-Core

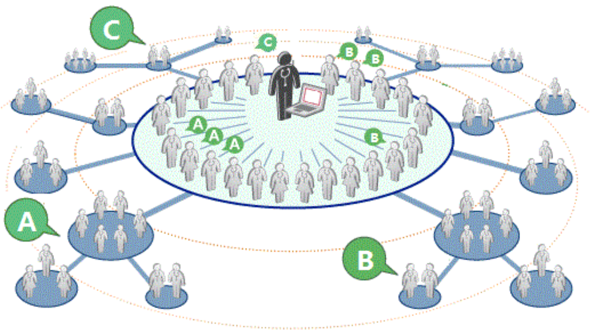
- A *k*-core *C* is called maximal if any supergraph of *C* is not a *k*-core (i.e., no another *k*-core which contains *C*).

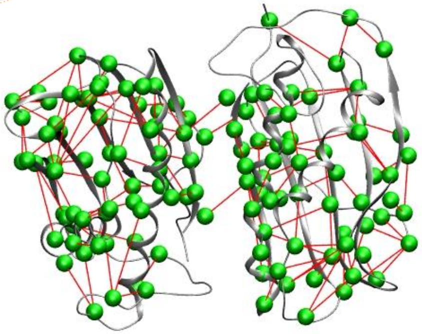


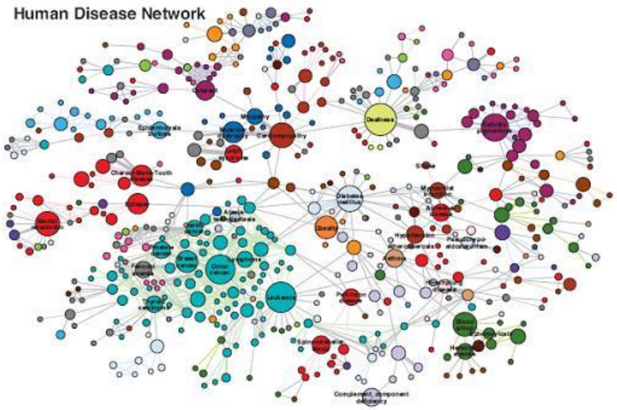
S. B. Seidman. Network structure and minimum degree. Social networks, 5(3):269–287, 1983.

Applications

- Community detection
- Social contagion
- User engagement
- Event detection
- Network analysis and visualization
- Influence study
- Graph clustering
- Protein function prediction
- Human Cerebral Cortex
- ……





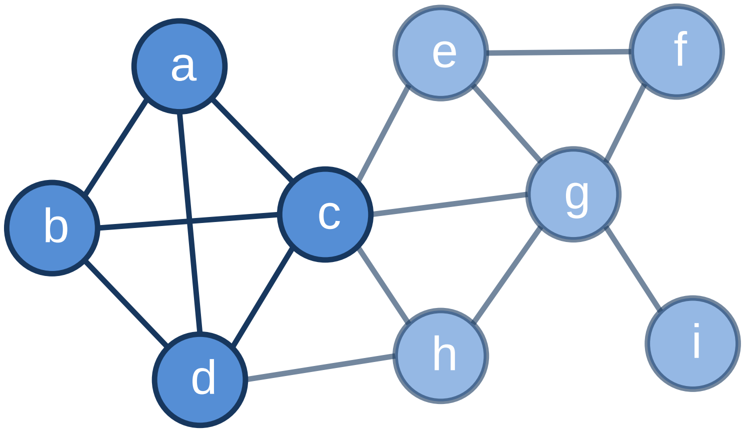


UNSW
THE UNIVERSITY OF NEW SOUTH WALES

Compute Maximal *k*-Core

- Given a graph *G*, the maximal *k*-core of *G* can be computed by recursively deleting every node and its adjacent edges if its degree is less than *k*.

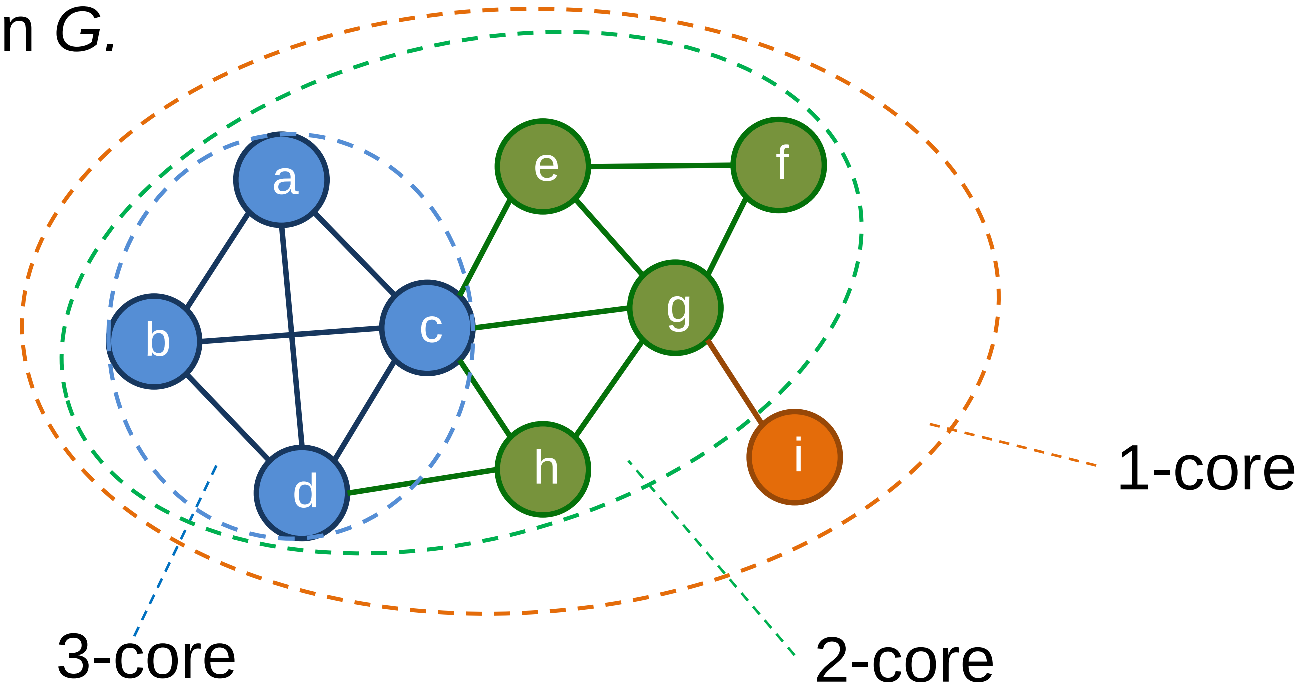
Maximal 3-core



S. B. Seidman. Network structure and minimum degree. Social networks, 5(3):269–287, 1983.

k-Core Decomposition

- **Core number** of a node v: the largest value of *k* such that there is a *k*-core containing *v.*
- Core decomposition: compute the **core number** of each node in *G.*



k-Truss

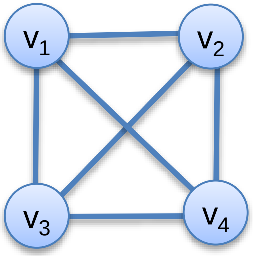
- Given a graph *G*, the *k*-truss of *G* is a subgraph where edge is at least involved in (k-2) triangles.

- k-truss is an enhancement of k-core; each vertex of k-truss has a degree at least k-1.

S. B. Seidman. Network structure and minimum degree. Social networks, 5(3):269–287, 1983.

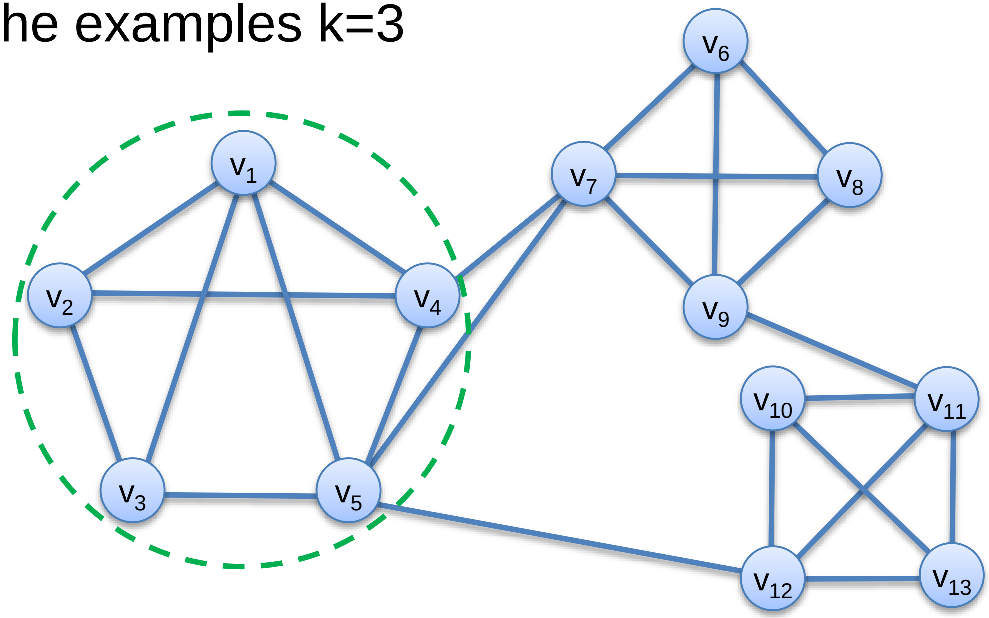
k-edge Connectivity

- A graph is k-edge connected if it is still connected after removing any set of (k-1) edges from it



k-edge Connected Component

- A k-edge connected component (k-ECC) of a graph G is a maximal subgraph g of G such that g is k-edge connected
  - All the examples k=3



UNSW THE UNIVERSITY OF NEW SOUTH WALES

Application

- Community detection
- Social behaviour mining
- Graph visualization
- Steiner Component Search
- Hierarchy Study in Networks
- ……

k-Vertex Connectivity

- A graph is k-vertex connected if it is still connected after removing any set of (k-1) vertex from it