SOCIAL NETWORK ANALYTICS

Finding Communities & Overlapped Communities by Clique Percolation Method (CPM)

Prakash C O

Department of Computer Science and Engineering

SOCIAL NETWORK ANALYTICS

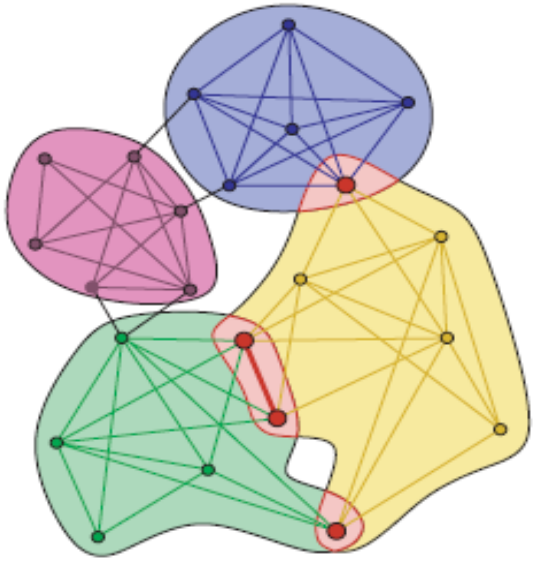
Finding Communities & Overlapped Communities by Clique Percolation Method (CPM)

Prakash C O

Department of Computer Science and Engineering

Finding Overlapped Communities by CPM

- Community membership could overlap: **a node could be part of more than 1 community**.
- A **community** -- also called a **cluster** or **module** -- in a network is a group of nodes more densely connected to each other than to nodes outside the group.
- **In real networks clusters/Communities often overlap.**



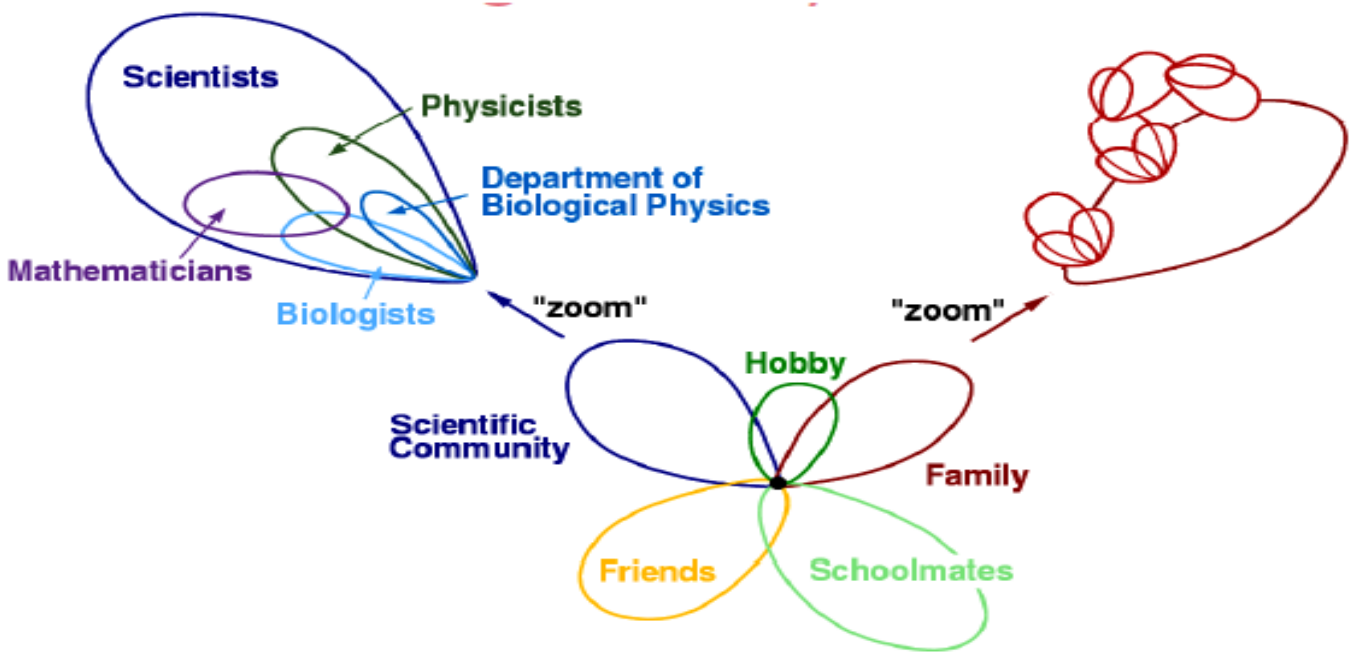
a) An example of overlapping k-clique-communities at k = 4. The yellow community overlaps with the blue one in a single node, whereas it shares 3 nodes and a link with the green one. These overlapping regions are emphasized in red.

Notice that any k-clique (complete subgraph of size k) can be reached only from the k-cliques of the same community through a series of adjacent k-cliques. Two k-cliques are adjacent if they share k − 1 nodes.

Finding Overlapped Communities by CPM

- Nodes can belong to more than 1 social circle!

b) The black dot in the middle represents either of the authors of this Letter, with several of his communities around. Zooming into the scientific community demonstrates the nested and overlapping structure of the communities, while depicting the cascades of communities starting from some members exemplifies the interwoven structure of the network of communities.



Finding Overlapped Communities by CPM

- **Clique** in a graph is **a subset of nodes where each pair of node is connected through an edge**, that is a **complete sub-graph.**
- The notation, **k-clique indicates size of the clique i.e., the clique consist of k nodes** e.g. a 3-clique indicate a complete sub-graphs having 3 nodes.

- **Clique: Complete graph/sub-graph**
- **k-clique: Complete graph/sub-graph with k vertices**

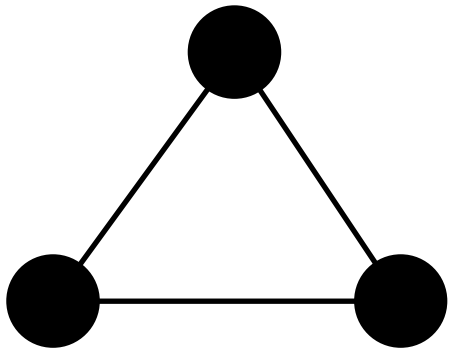
Finding Overlapped Communities by CPM

What is Clique Percolation Method(CPM)?

- It's a method to find **overlapping** communities.
- The clique percolation method (CPM) was proposed by Palla et al.
- **Based on concept:**
  - Internal edges of community likely to form cliques.
  - Intercommunity edges unlikely to form cliques.

Finding Overlapped Communities by CPM

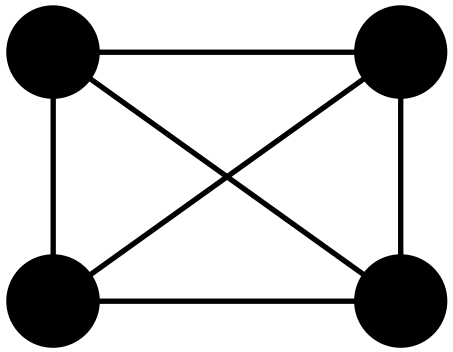
- **3-clique**: Complete graph with 3 vertices



3-clique

Finding Overlapped Communities by CPM

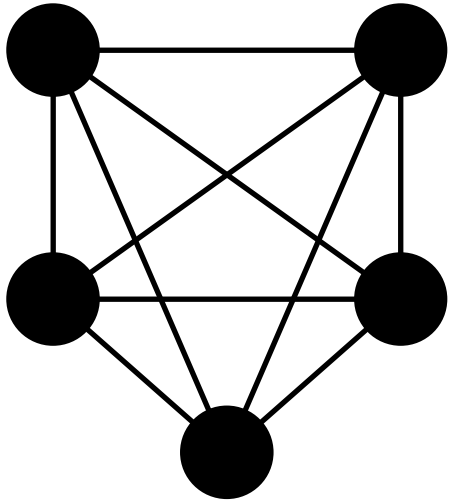
- **4-clique**: Complete graph with 4 vertices



4-clique

Finding Overlapped Communities by CPM

- **5-clique**: Complete graph with 5 vertices



5-clique

Finding Overlapped Communities by CPM

- **Adjacent k-cliques**

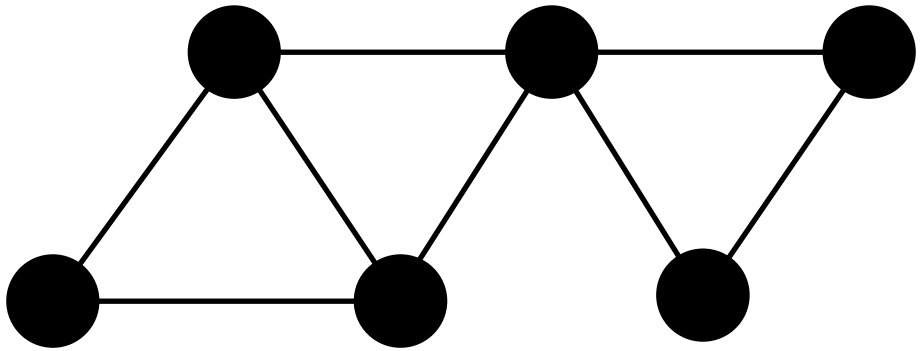
  Two k-cliques are adjacent when they share **k-1** nodes

Finding Overlapped Communities by CPM

- **Adjacent k-cliques**

  Two k-cliques are adjacent when they share **k-1** nodes

k = 3

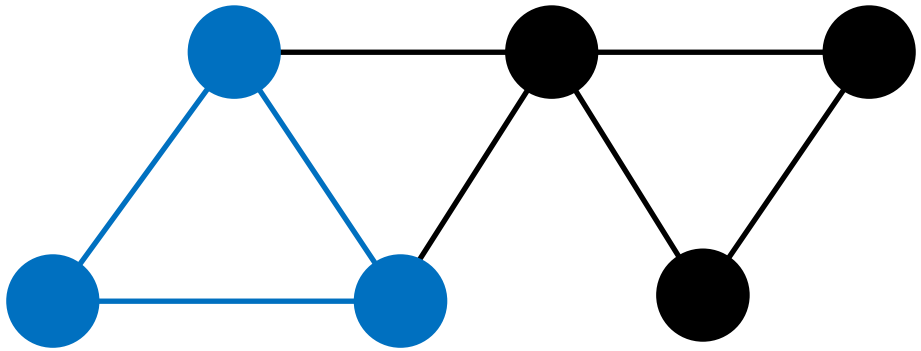


Finding Overlapped Communities by CPM

- **Adjacent k-cliques**

  Two k-cliques are adjacent when they share **k-1** nodes

k = 3



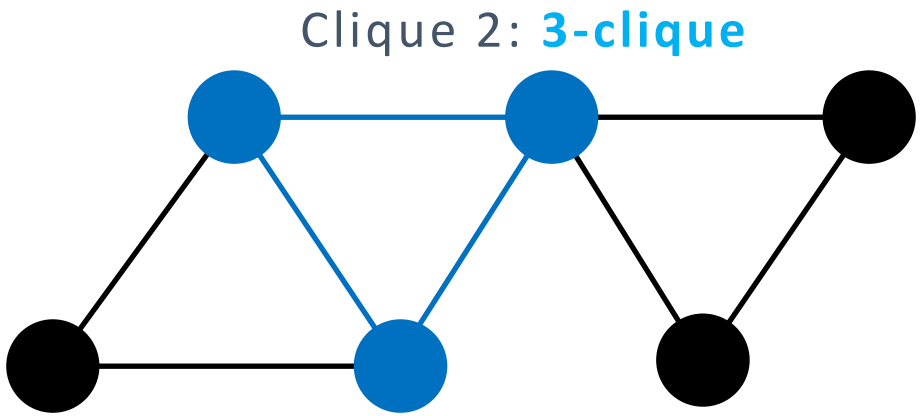
Clique 1: **3-clique**

Finding Overlapped Communities by CPM

- **Adjacent k-cliques**

  Two k-cliques are adjacent when they share **k-1** nodes

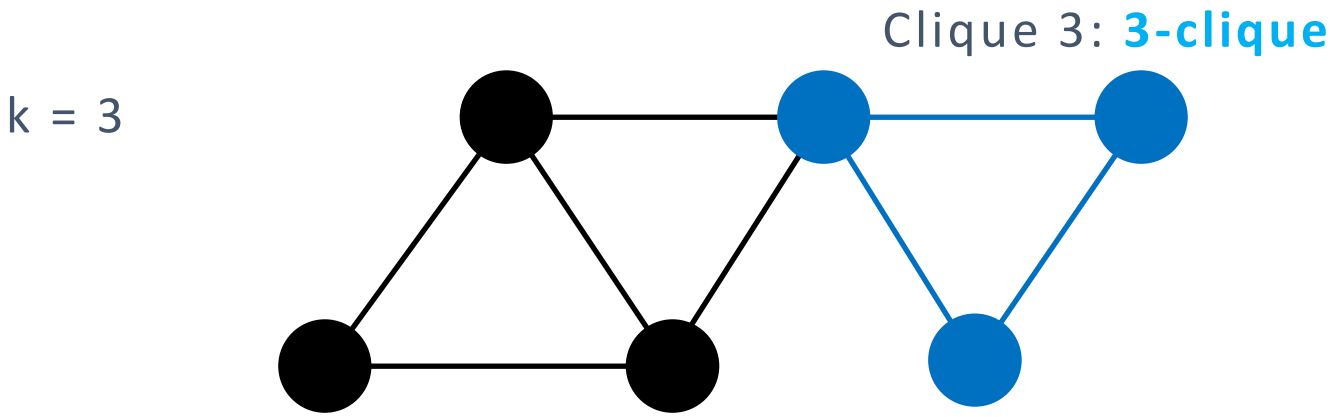
k = 3



Finding Overlapped Communities by CPM

- **Adjacent k-cliques**

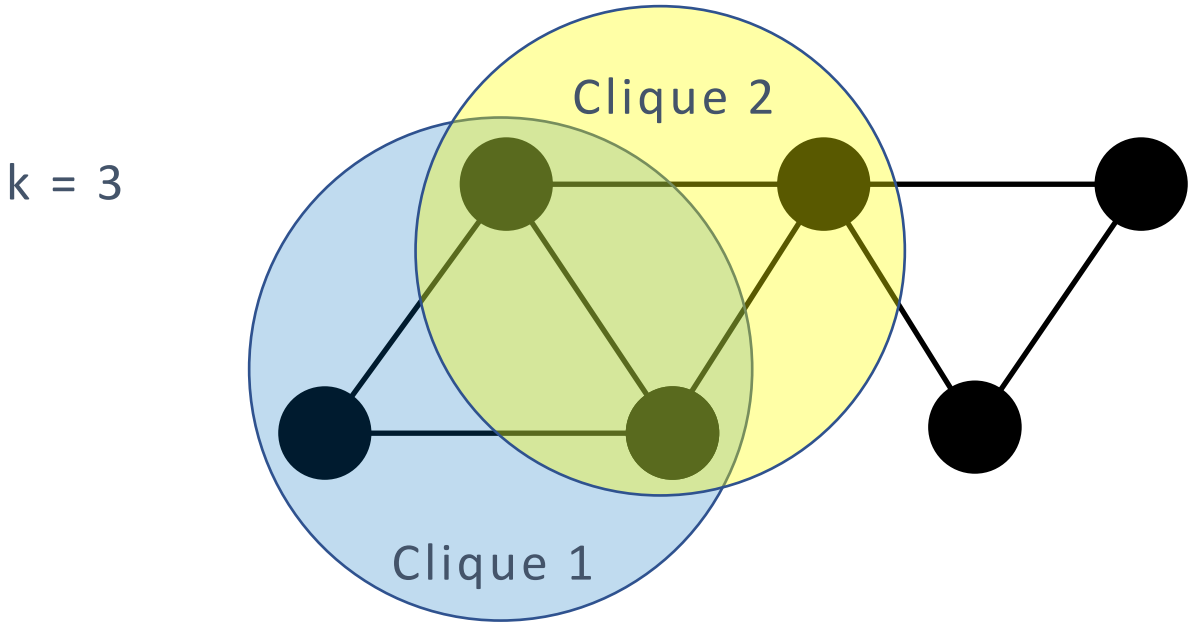
  Two k-cliques are adjacent when they share **k-1** nodes



Finding Overlapped Communities by CPM

➢ Adjacent k-cliques

Two k-cliques are adjacent when they share **k-1** nodes



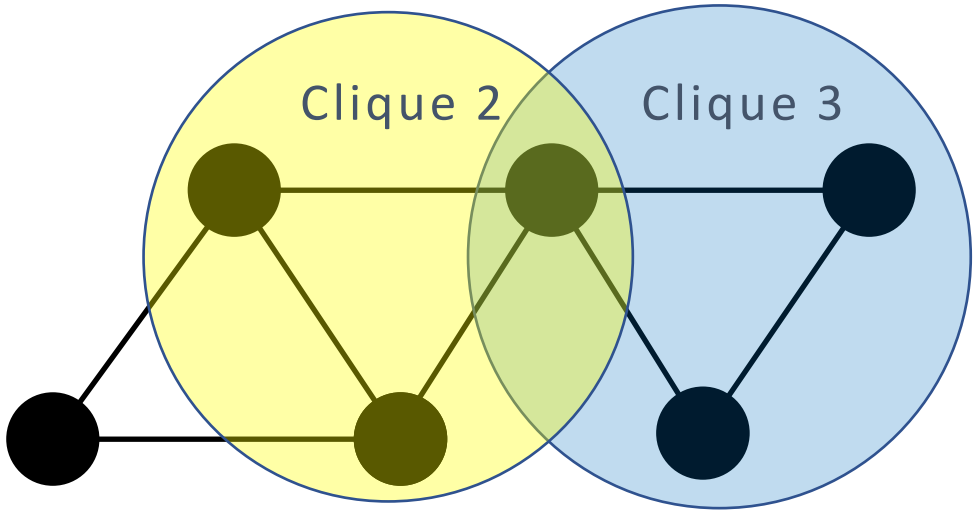
Clique-1 and Clique-2 are adjacent because they share k-1 nodes

Finding Overlapped Communities by CPM

➢Adjacent k-cliques

Two k-cliques are adjacent when they share **k-1** nodes

k = 3



Clique-2 and Clique-3 are not adjacent because they are not sharing k-1 nodes

Finding Overlapped Communities by CPM

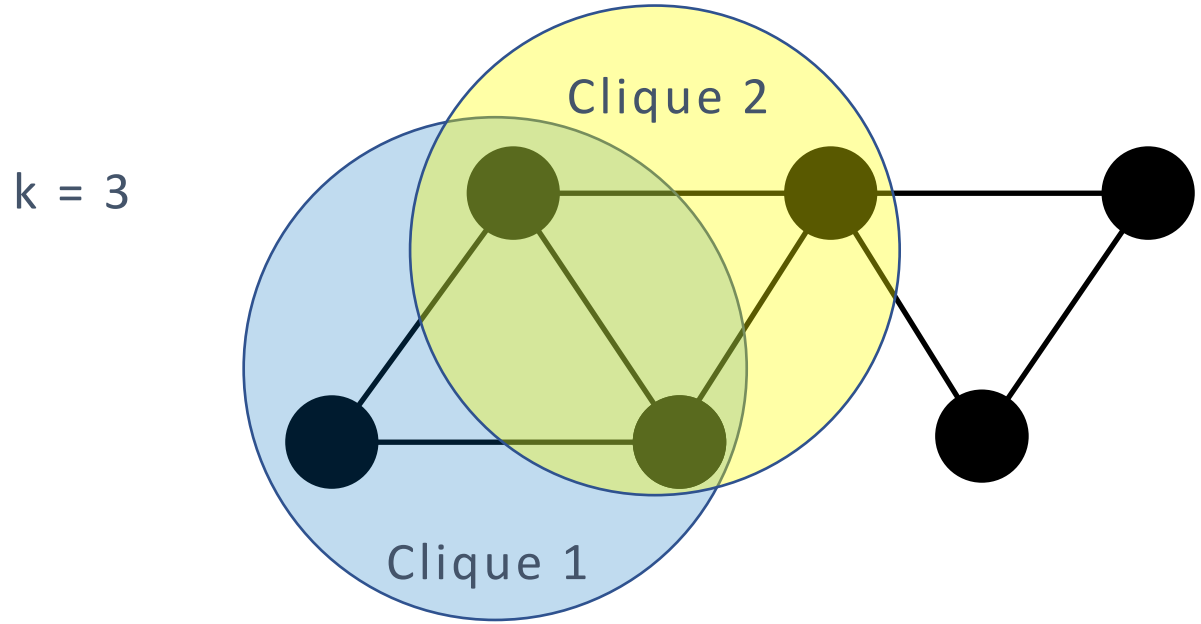
k-Clique Community

Union of all k-cliques that can be reached from each other through a series of adjacent k-cliques.

Finding Overlapped Communities by CPM

k-Clique Community

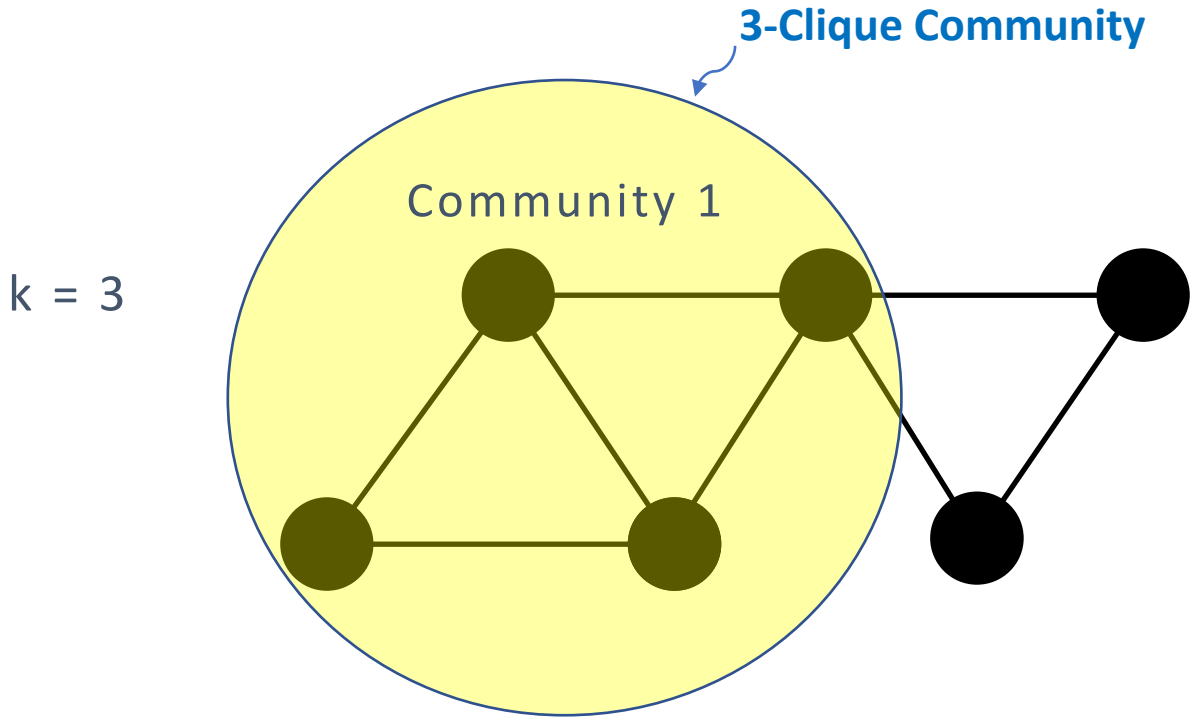
Union of all k-cliques that can be reached from each other through a series of adjacent k-cliques.



Finding Overlapped Communities by CPM

k-Clique Community

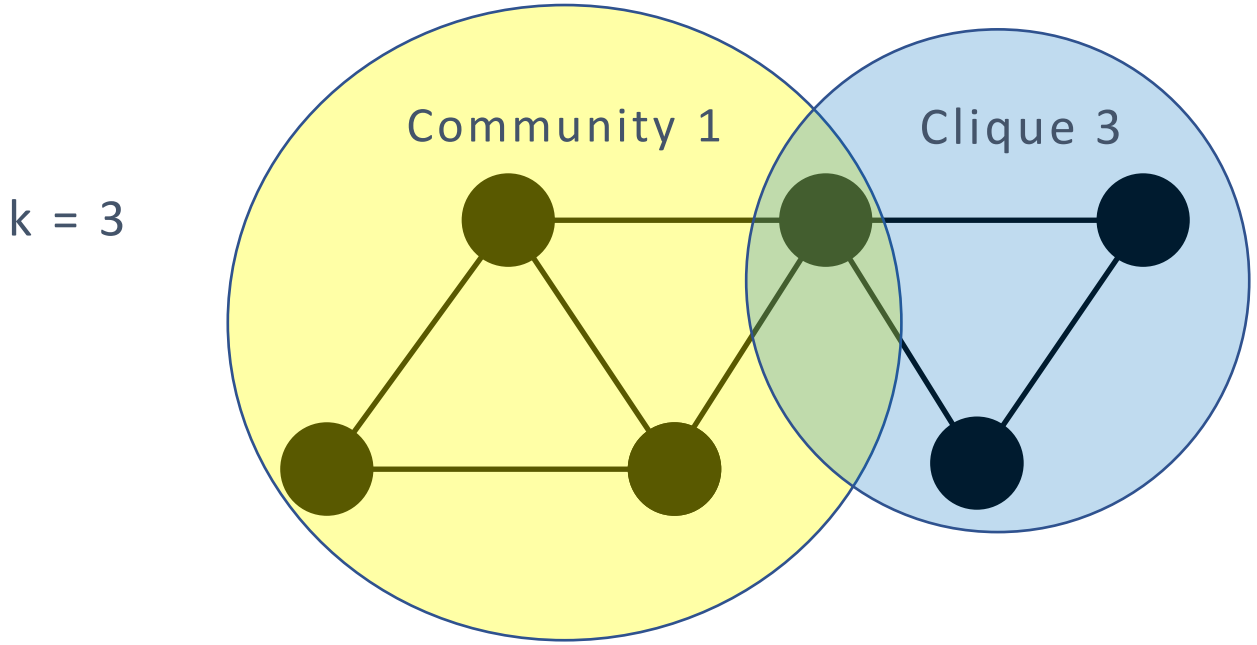
Union of all k-cliques that can be reached from each other through a series of adjacent k-cliques.



Finding Overlapped Communities by CPM

k-Clique Community

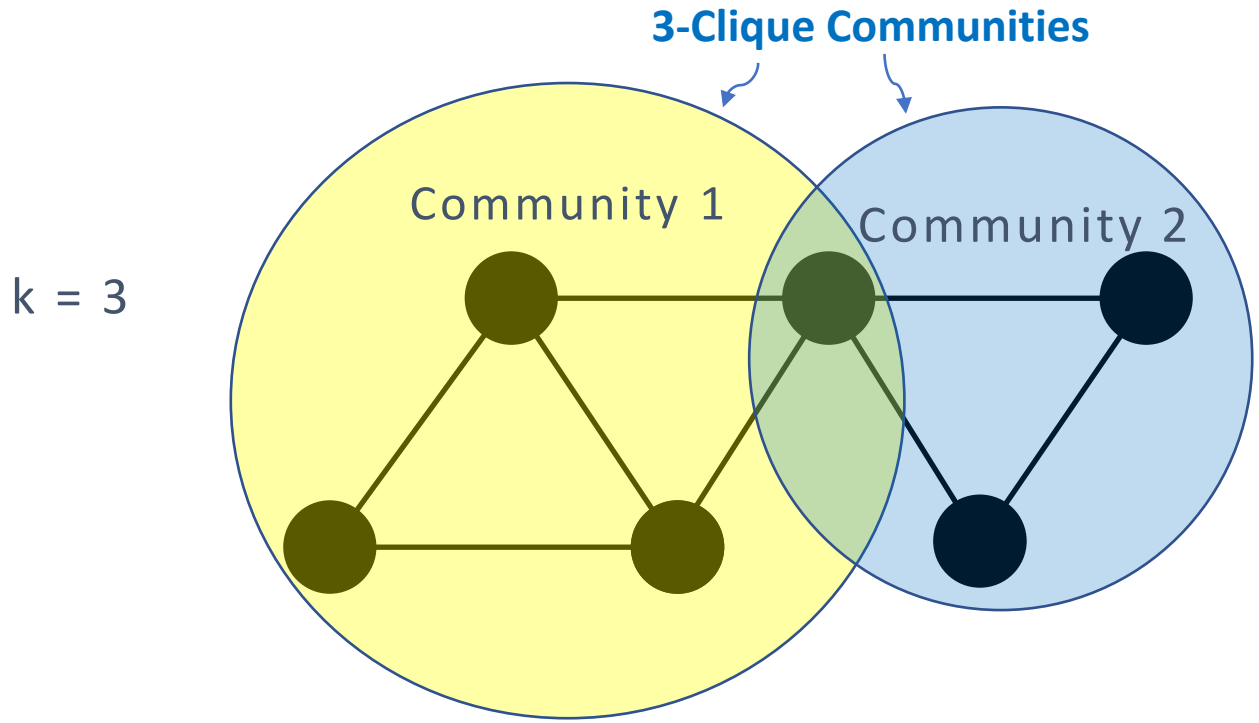
Union of all k-cliques that can be reached from each other through a series of adjacent k-cliques.



Finding Overlapped Communities by CPM

k-Clique Community

Union of all k-cliques that can be reached from each other through a series of adjacent k-cliques.



Community-1 and **Community-2 are overlapped**

Finding Overlapped Communities by CPM

Algorithm:

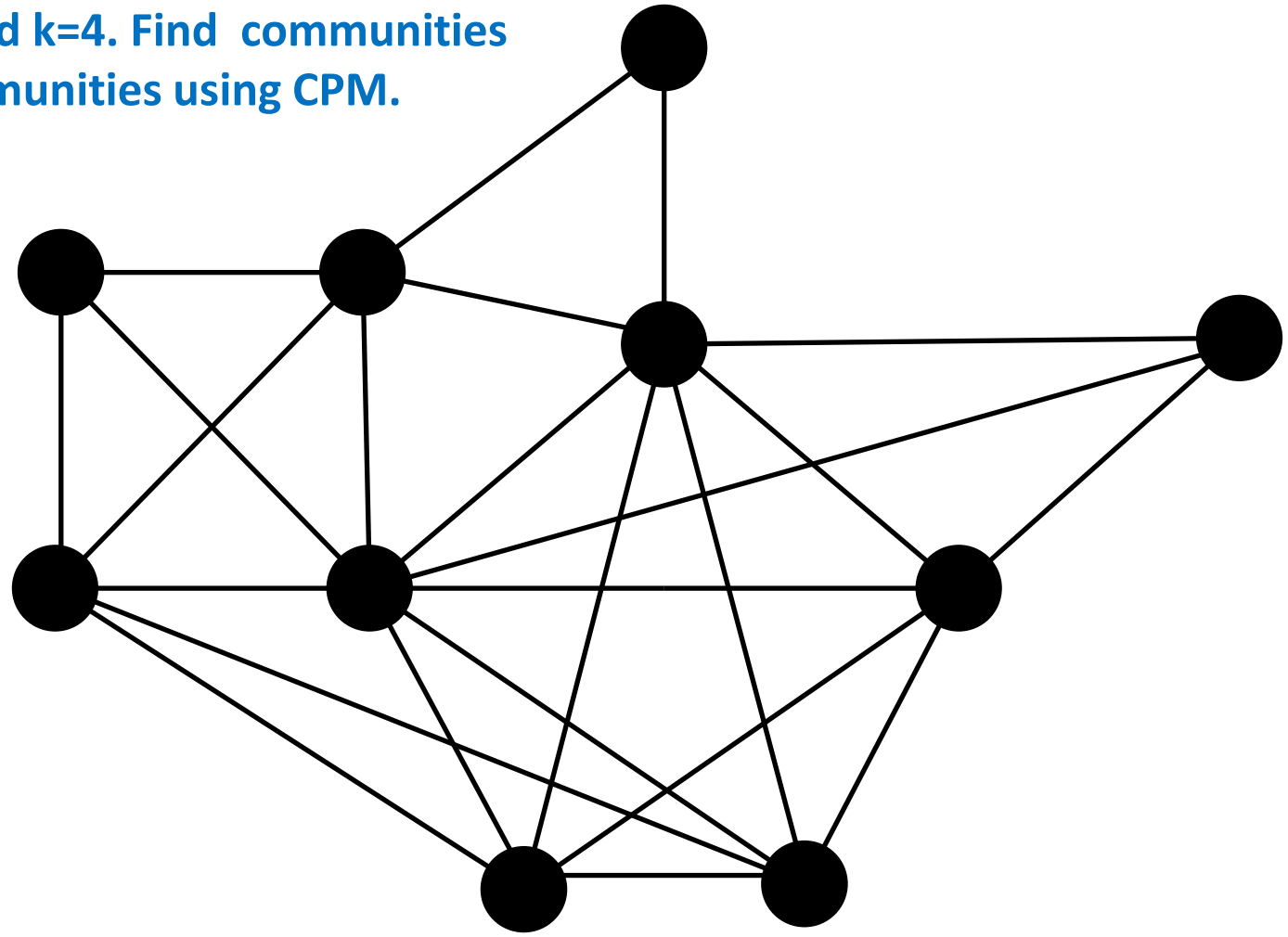
Input: Given graph G, and k (k: find communities of size k or more)

1. **Locate maximal cliques**

2. **Convert from cliques to k-clique communities**

Finding Overlapped Communities by CPM

- **Problem:** Given G and k=4. Find communities and overlapped communities using CPM.



Note: k indicates finding communities of size k or more.

Finding Overlapped Communities by CPM

Algorithm

1. Locate maximal cliques
2. Convert from cliques to k-clique communities

Finding Overlapped Communities by CPM

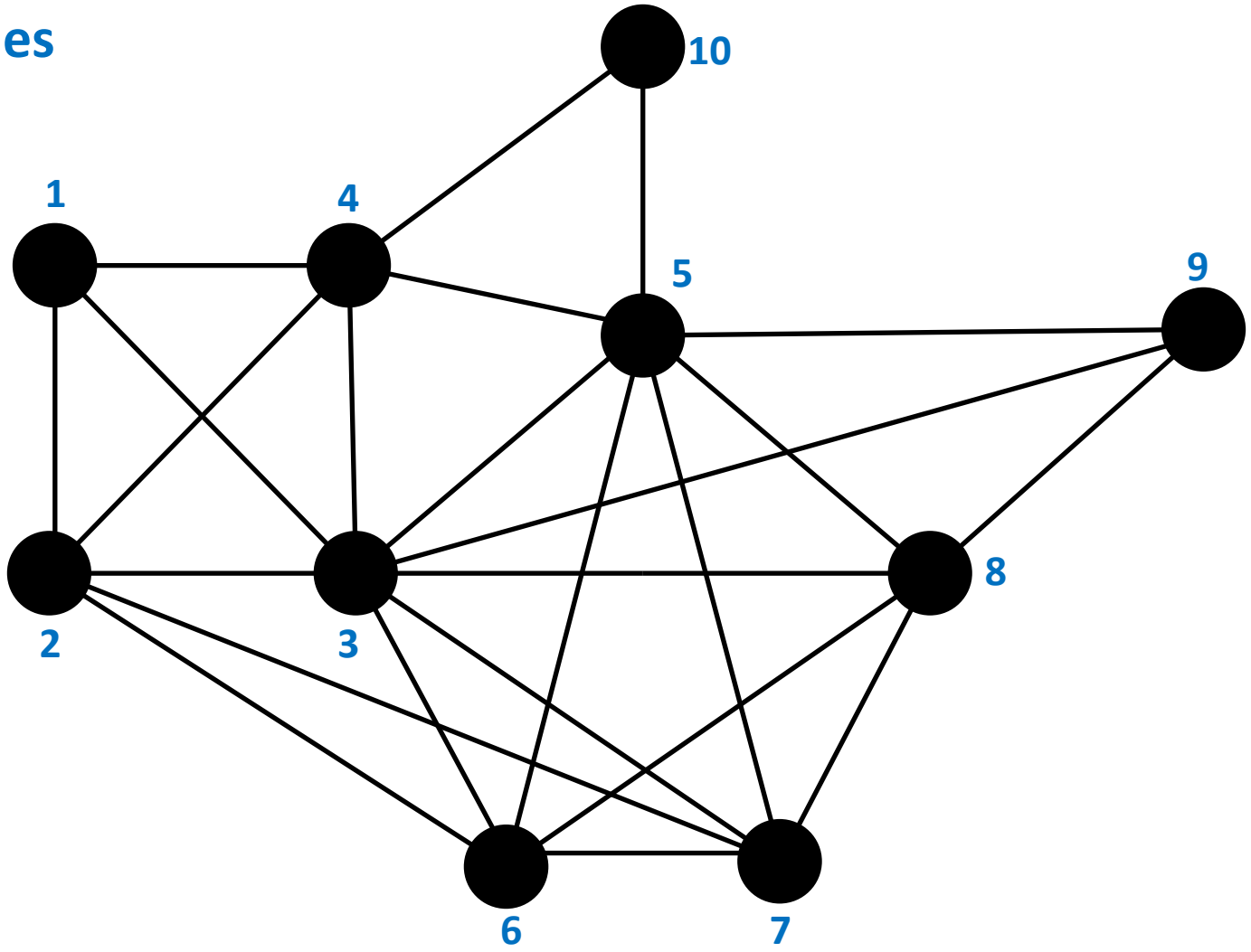
1. Locate Maximal Cliques

a) Largest possible clique size can be determined from degrees of vertices

b) Starting from this size, find all cliques, then reduce size by 1 and repeat

Finding Overlapped Communities by CPM

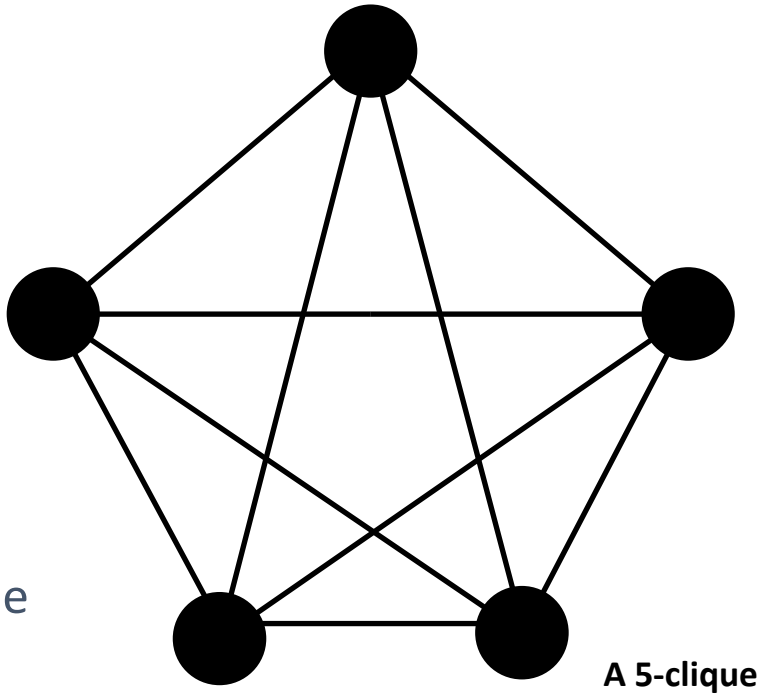
Locate maximal cliques



Finding Overlapped Communities by CPM

Locate maximal cliques

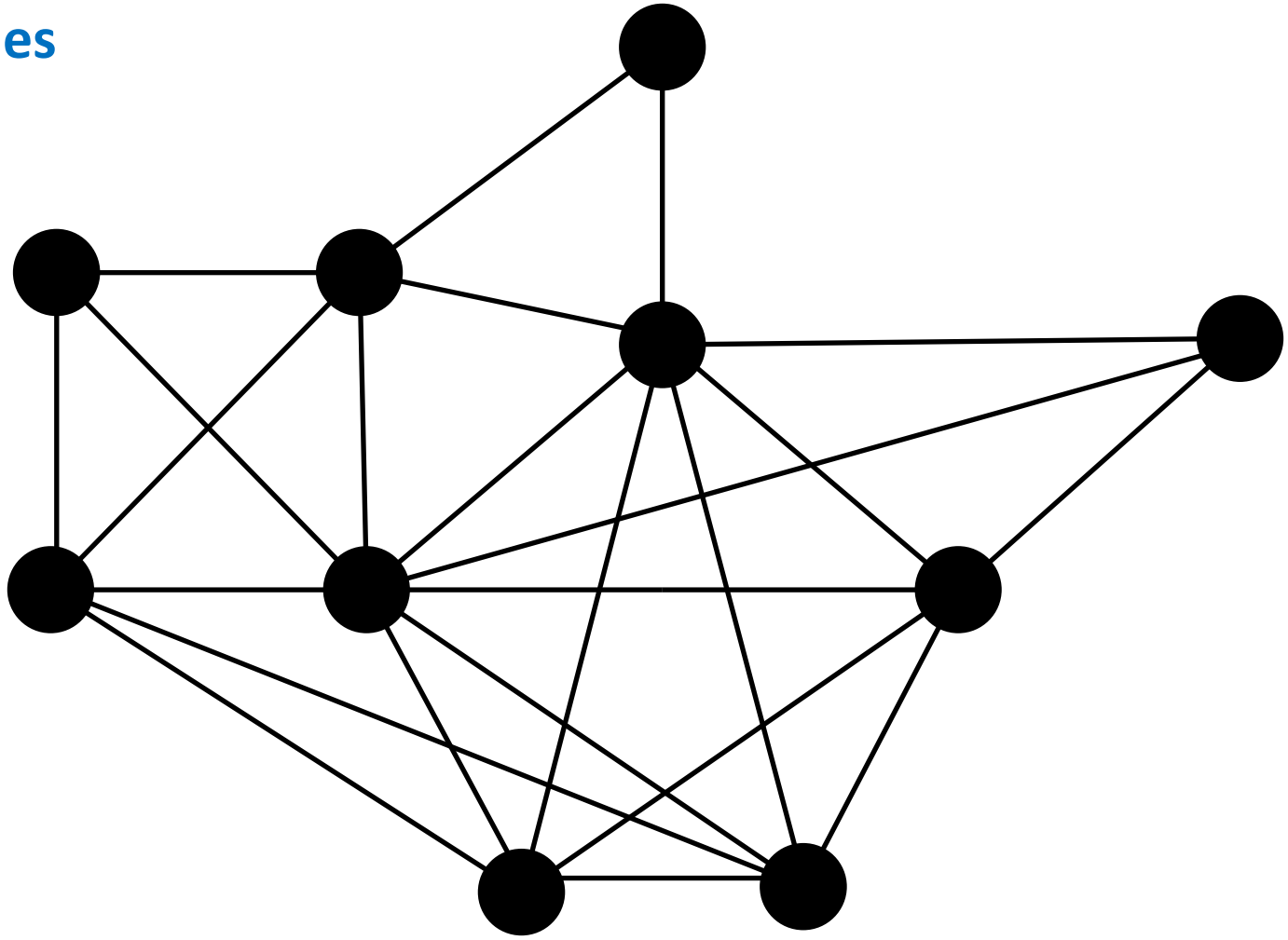
Clique 1: 5-clique



A 5-clique contains five 4-cliques

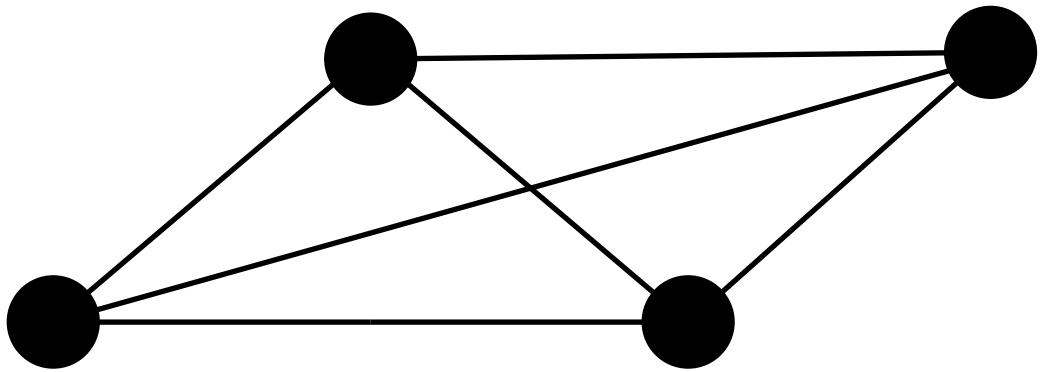
Finding Overlapped Communities by CPM

Locate maximal cliques



Finding Overlapped Communities by CPM

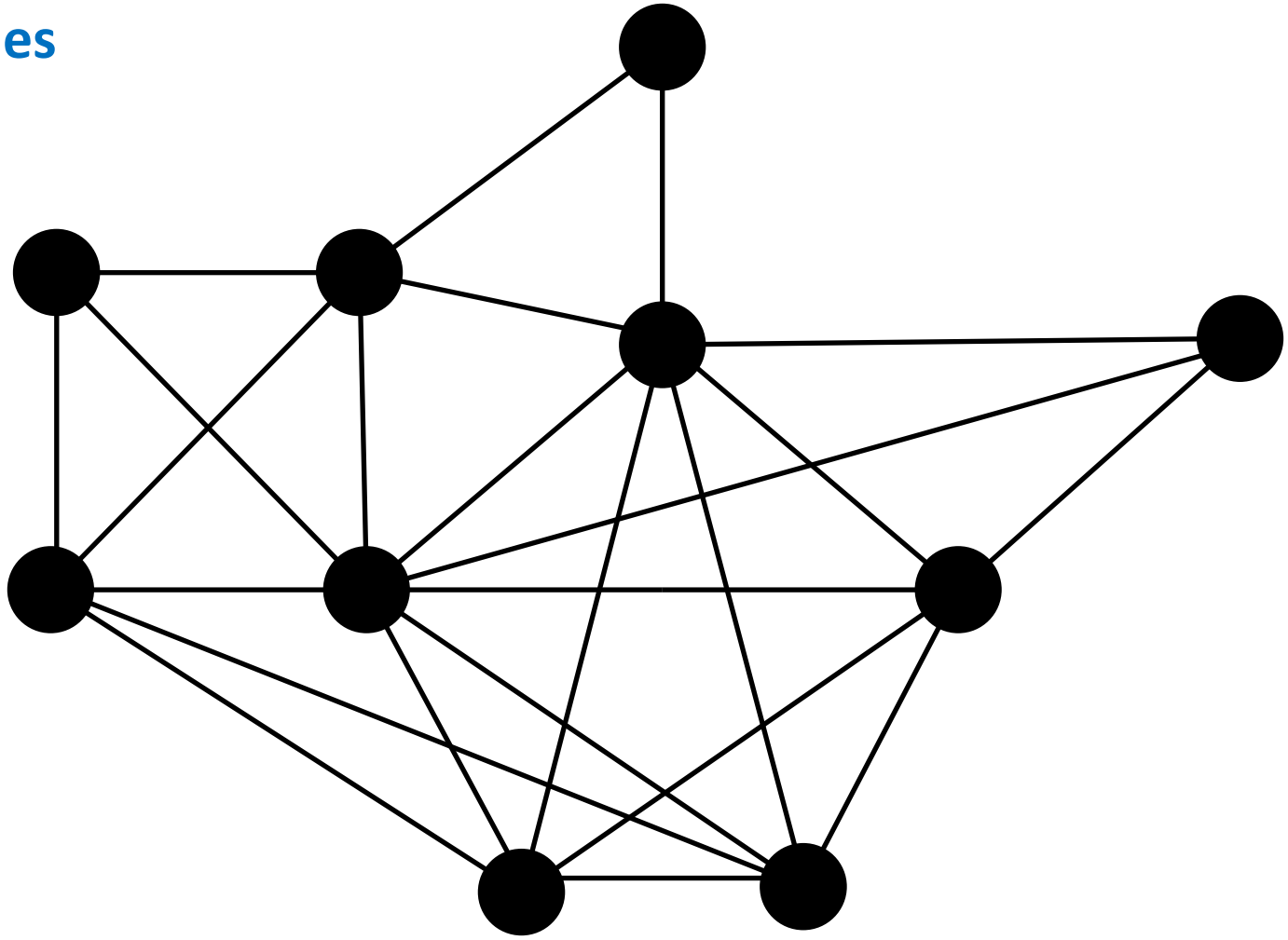
Locate maximal cliques



Clique 2: 4-clique

Finding Overlapped Communities by CPM

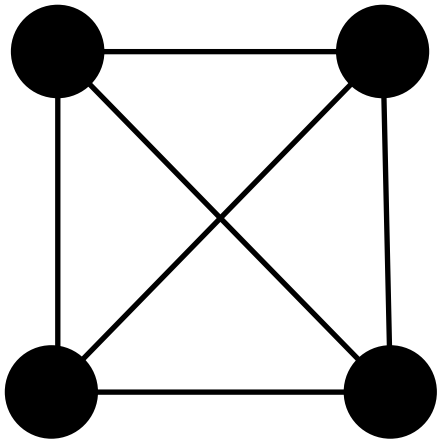
Locate maximal cliques



Finding Overlapped Communities by CPM

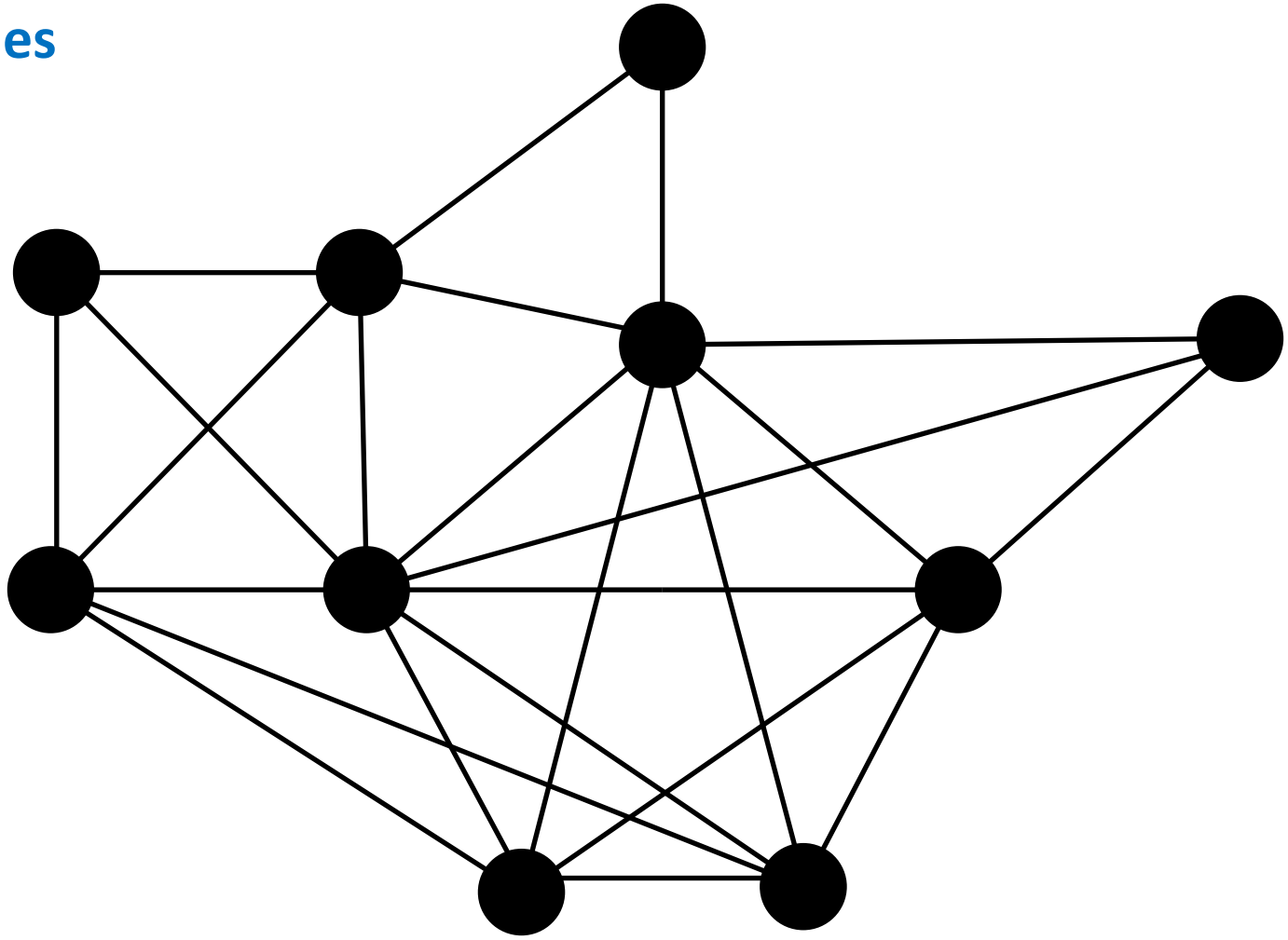
Locate maximal cliques

Clique 3: 4-clique



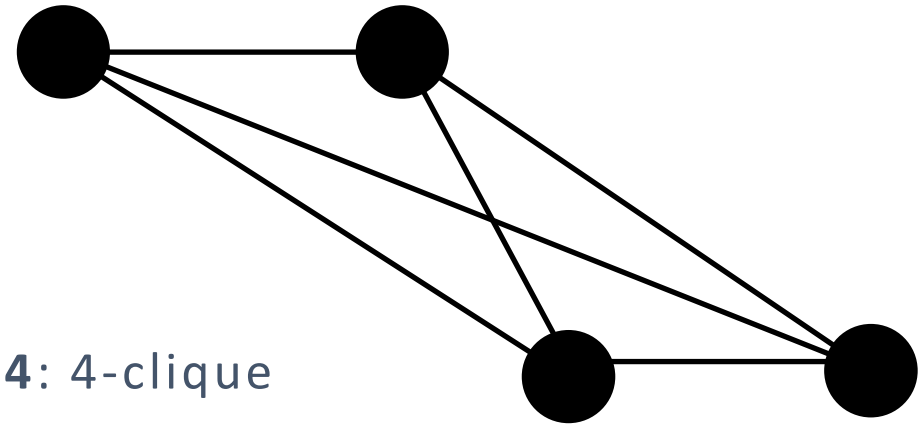
Finding Overlapped Communities by CPM

Locate maximal cliques



Finding Overlapped Communities by CPM

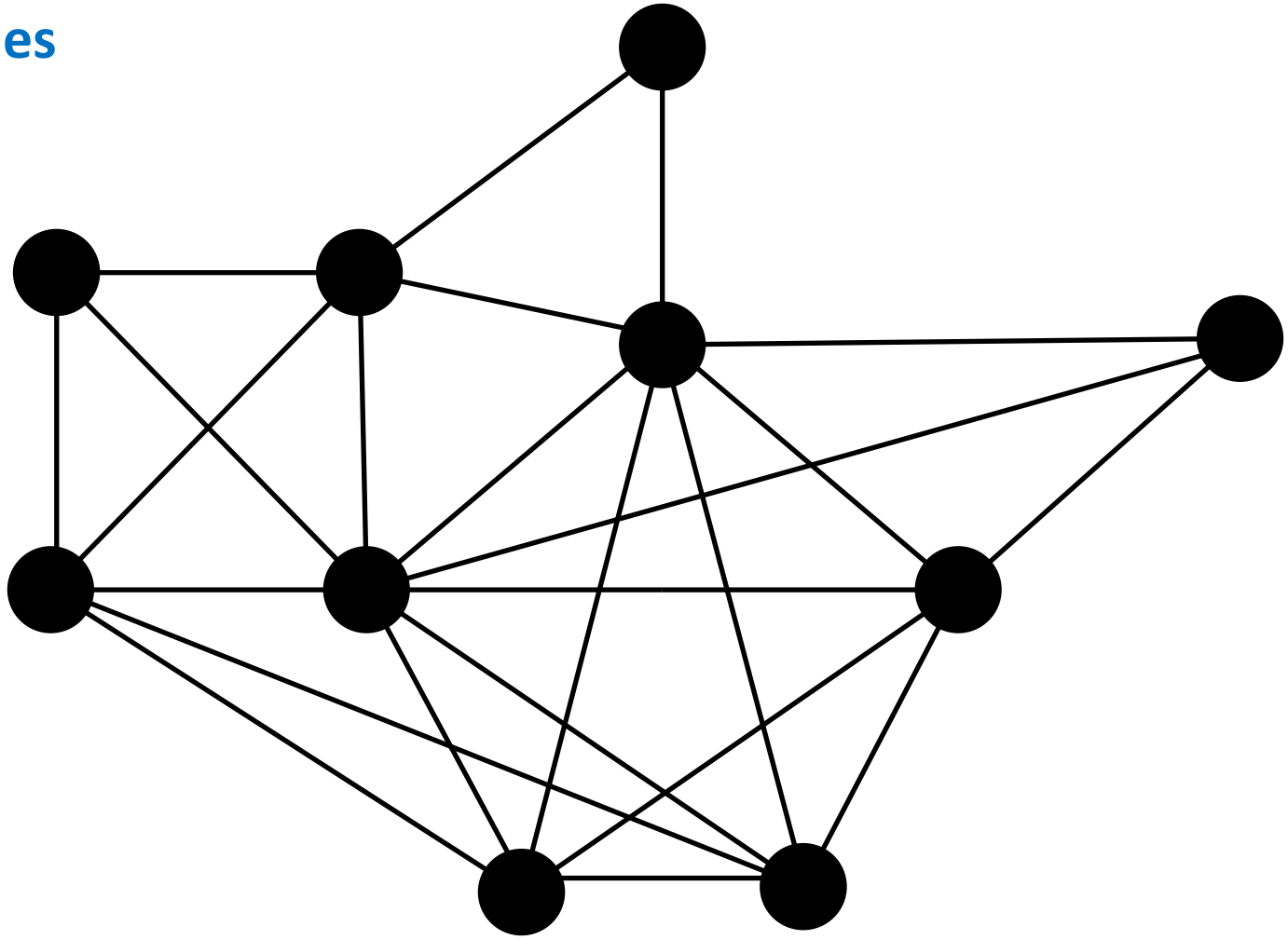
Locate maximal cliques



Clique 4: 4-clique

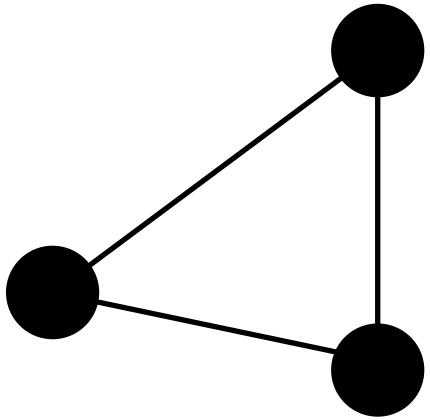
Finding Overlapped Communities by CPM

Locate maximal cliques



Finding Overlapped Communities by CPM

Locate maximal cliques

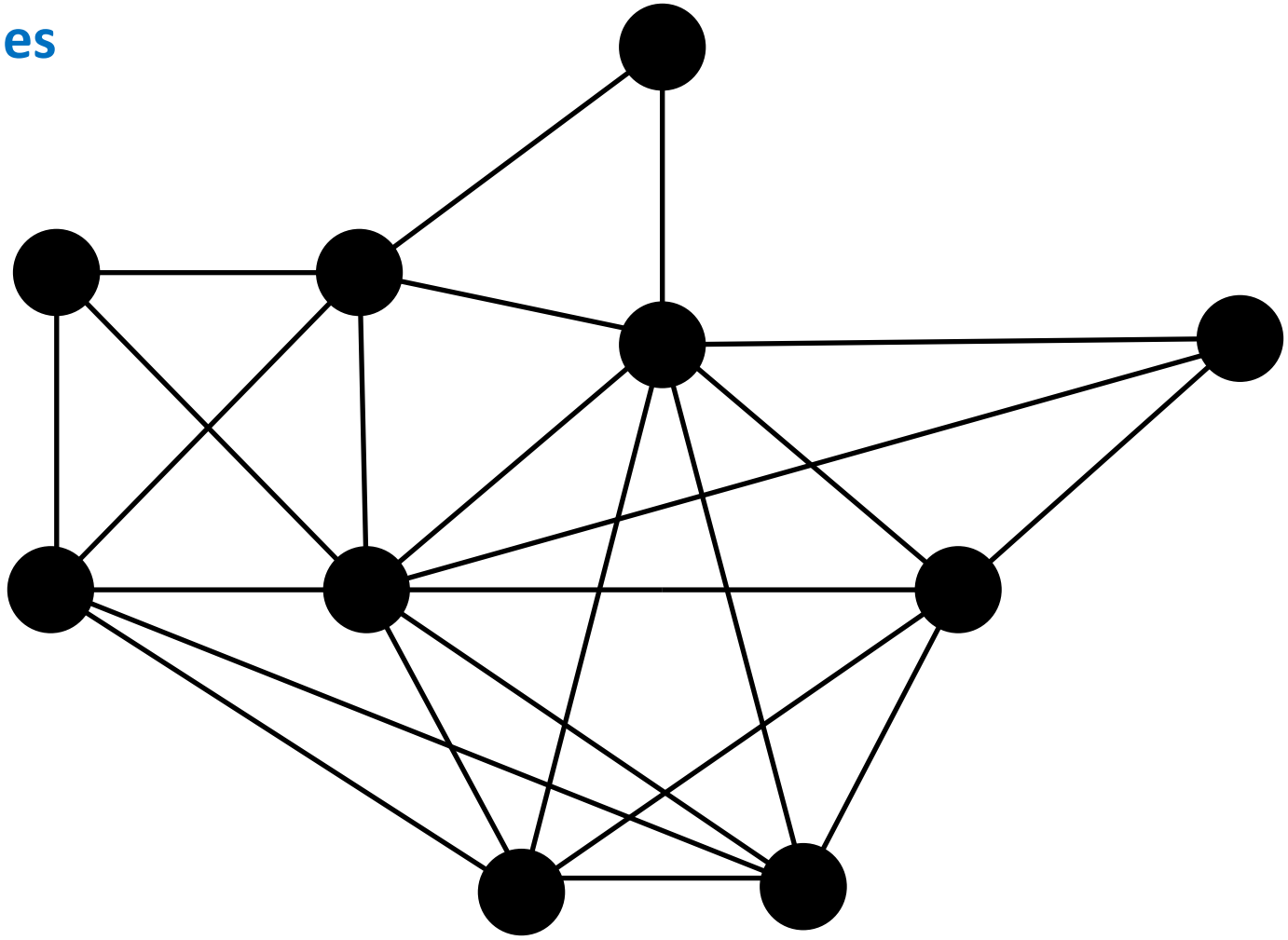


Clique 5: 3-clique

SOCIAL NETWORK ANALYTICS

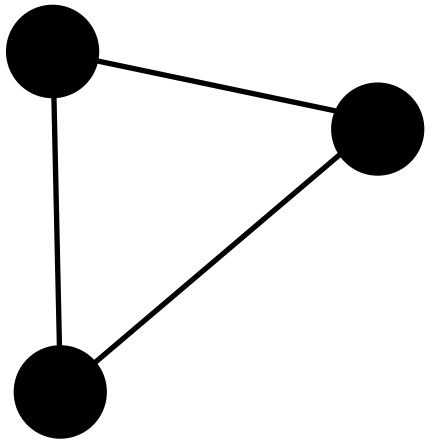
Finding Overlapped Communities by CPM

Locate maximal cliques



Finding Overlapped Communities by CPM

Locate maximal cliques



Clique 6: 3-clique

Finding Overlapped Communities by CPM

Algorithm

1. Locate maximal cliques
2. Convert from cliques to k-clique communities

Finding Overlapped Communities by CPM

2. Convert from cliques to k-clique communities

a) Create a square matrix $CC_{n \times n}$, where n is the number of cliques found.

Each cell CC[*i*, *j*] contains number of nodes shared by cliques C_i and C_j.

b) If CC[i,i] is less than k, then delete.

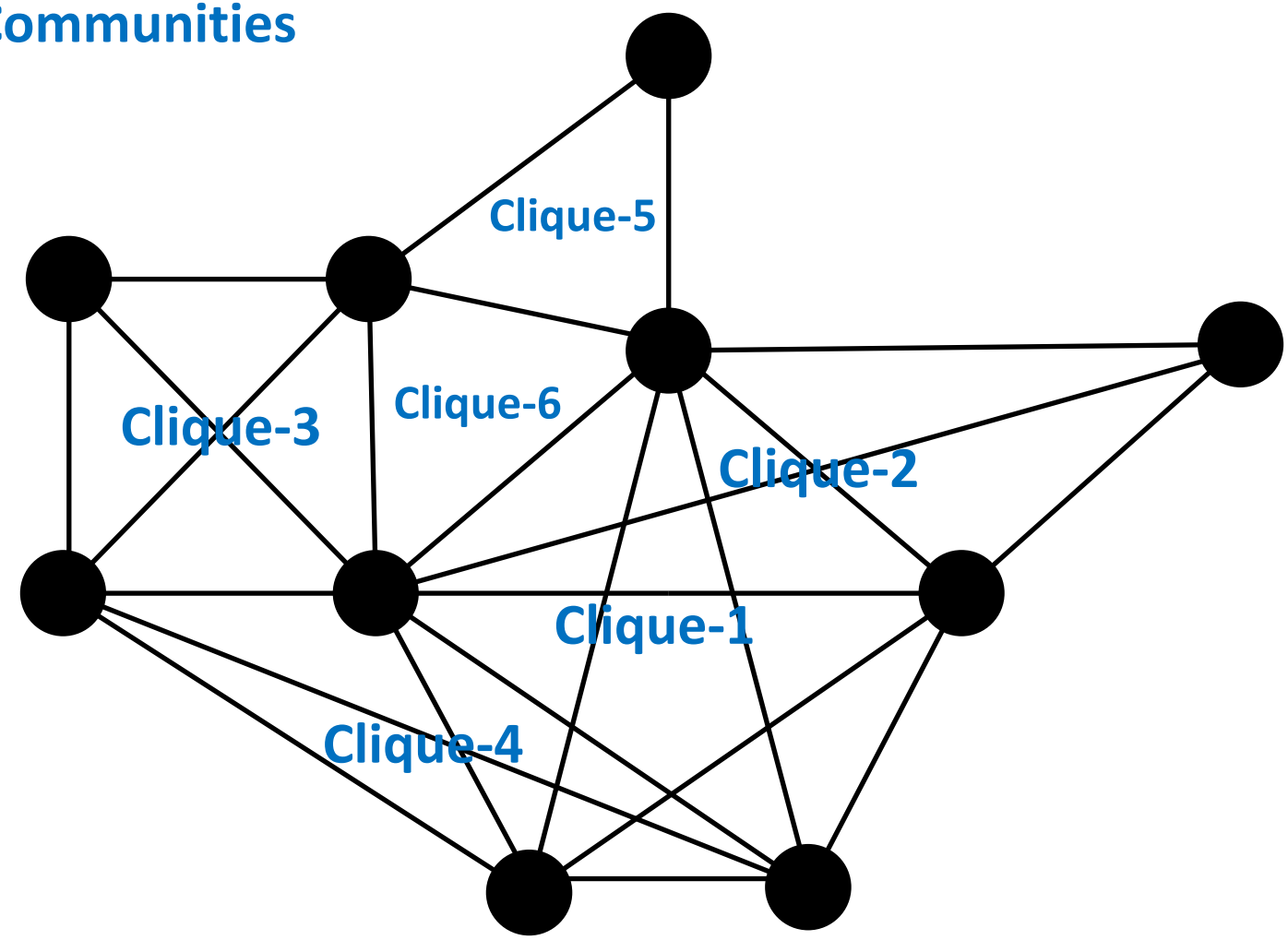
c) If CC[i,j] is less than k-1, then delete.

d) Change all non-zeros to 1.

e) Combine adjacent cliques to form a community.

Finding Overlapped Communities by CPM

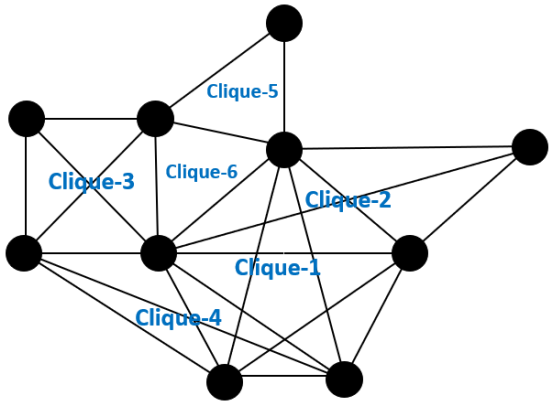
Cliques to k-Clique Communities



Finding Overlapped Communities by CPM

Cliques to k-Clique Communities

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	5					
C_2		4				
C_3			4			
C_4				4		
C_5					3	
C_6						3

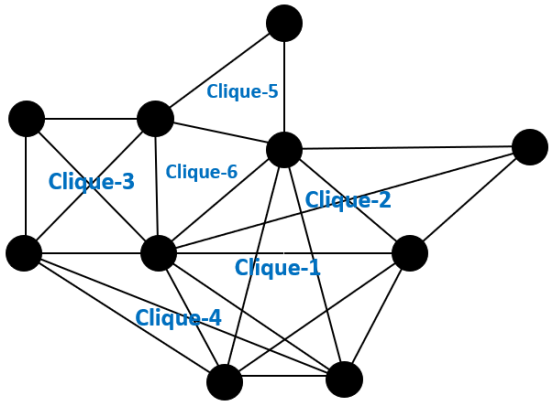


Step-a: Construct a matrix $CC_{n \times n}$,where n represents the total number of cliques. Each cell CC[*i*, *j*] contains number of nodes shared by cliques C_i and C_j.

Finding Overlapped Communities by CPM

Cliques to k-Clique Communities

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	5	3	1	3	1	2
C_2	3	4	1	1	1	2
C_3	1	1	4	2	1	2
C_4	3	1	2	4	0	1
C_5	1	1	1	0	3	2
C_6	2	2	2	1	2	3



Step-a: Construct a matrix $CC_{n \times n}$,where n represents the total number of cliques. Each cell CC[*i*, *j*] contains number of nodes shared by cliques C_i and C_j.

Finding Overlapped Communities by CPM

Cliques to k-Clique Communities

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	5	3	1	3	1	2
C_2	3	4	1	1	1	2
C_3	1	1	4	2	1	2
C_4	3	1	2	4	0	1
C_5	1	1	1	0	3	2
C_6	2	2	2	1	2	3

Symmetric Matrix

Finding Overlapped Communities by CPM

Cliques to k-Clique Communities

Given k=4

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	5	3	1	3	1	2
C_2	3	4	1	1	1	2
C_3	1	1	4	2	1	2
C_4	3	1	2	4	0	1
C_5	1	1	1	0	3	2
C_6	2	2	2	1	2	3

Note: Given k in problem statement represents Community size (i.e., find communities of size k or more)

Finding Overlapped Communities by CPM

Cliques to k-Clique Communities

k=4

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	5	3	1	3	1	2
C_2	3	4	1	1	1	2
C_3	1	1	4	2	1	2
C_4	3	1	2	4	0	1
C_5	1	1	1	0	3	2
C_6	2	2	2	1	2	3

Step-b: If CC[i,i] is less than k, then delete

Finding Overlapped Communities by CPM

Cliques to k-Clique Communities

k=4

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	5	3	1	3	1	2
C_2	3	4	1	1	1	2
C_3	1	1	4	2	1	2
C_4	3	1	2	4	0	1
C_5	1	1	1	0	0	2
C_6	2	2	2	1	2	0

Finding Overlapped Communities by CPM

Cliques to k-Clique Communities

k=4

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	5	3	1	3	1	2
C_2	3	4	1	1	1	2
C_3	1	1	4	2	1	2
C_4	3	1	2	4	0	1
C_5	1	1	1	0	0	2
C_6	2	2	2	1	2	0

Step-c: If CC[i,j] is less than k-1, then delete

Finding Overlapped Communities by CPM

Cliques to k-Clique Communities

k=4

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	5	3	0	3	0	0
C_2	3	4	0	0	0	0
C_3	0	0	4	0	0	0
C_4	3	0	0	4	0	0
C_5	0	0	0	0	0	0
C_6	0	0	0	0	0	0

Finding Overlapped Communities by CPM

Cliques to k-Clique Communities

k=4

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	5	3	0	3	0	0
C_2	3	4	0	0	0	0
C_3	0	0	4	0	0	0
C_4	3	0	0	4	0	0
C_5	0	0	0	0	0	0
C_6	0	0	0	0	0	0

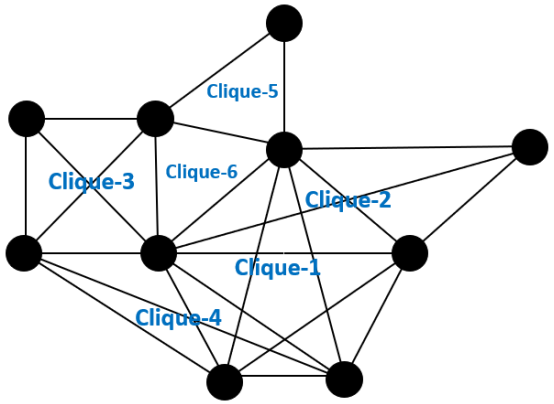
Step-d: Change all non-zeros to 1

Finding Overlapped Communities by CPM

Cliques to k-Clique Communities

k=4

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	1	1	0	1	0	0
C_2	1	1	0	0	0	0
C_3	0	0	1	0	0	0
C_4	1	0	0	1	0	0
C_5	0	0	0	0	0	0
C_6	0	0	0	0	0	0



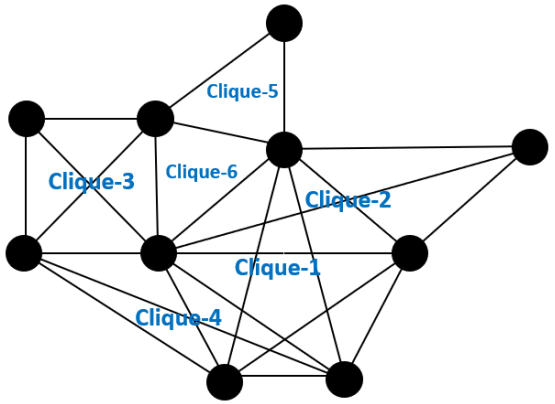
Step-e: Combine adjacent cliques to form a community.

Finding Overlapped Communities by CPM

Cliques to k-Clique Communities

Clique-clique overlap matrix

	C_1	C_2	C_3	C_4	C_5	C_6
C_1	1	1	0	1	0	0
C_2	1	1	0	0	0	0
C_3	0	0	1	0	0	0
C_4	1	0	0	1	0	0
C_5	0	0	0	0	0	0
C_6	0	0	0	0	0	0



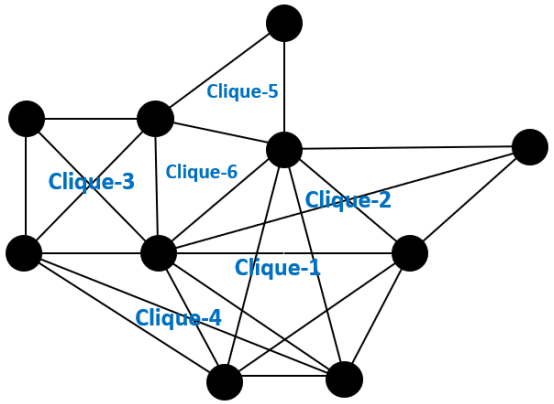
Row 1 indicates clique-1 is adjacent to itself, clique-2 and clique-4.
Row 2 indicates clique-2 is adjacent to clique-1 and itself.
Row 3 indicates clique-3 is adjacent to itself only.
Row 4 indicates clique-4 is adjacent to clique-1 and itself.

Finding Overlapped Communities by CPM

Cliques to k-Clique Communities

Clique-clique overlap matrix

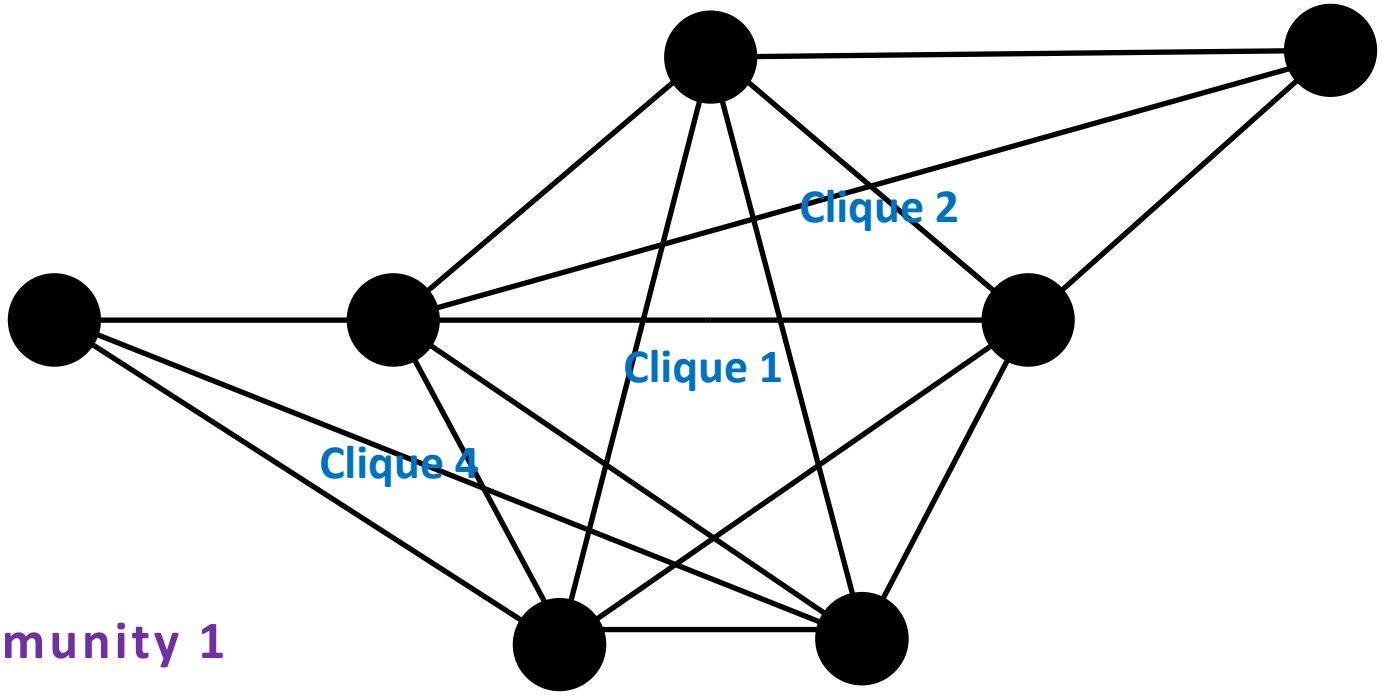
	C_1	C_2	C_3	C_4	C_5	C_6
C_1	1	1	0	1	0	0
C_2	1	1	0	0	0	0
C_3	0	0	1	0	0	0
C_4	1	0	0	1	0	0
C_5	0	0	0	0	0	0
C_6	0	0	0	0	0	0



Adjacent Cliques 1,2 and 4 form community-1 and **Clique 3 alone form community-2**

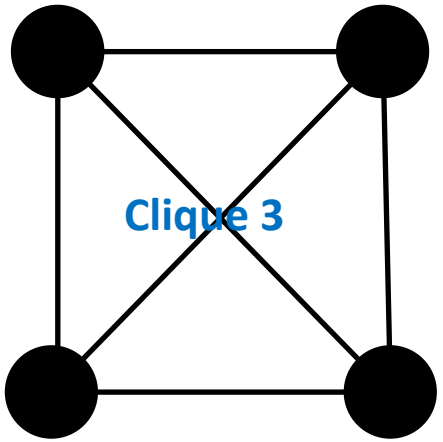
Finding Overlapped Communities by CPM

Cliques to k-Clique Communities



Finding Overlapped Communities by CPM

Cliques to k-Clique Communities

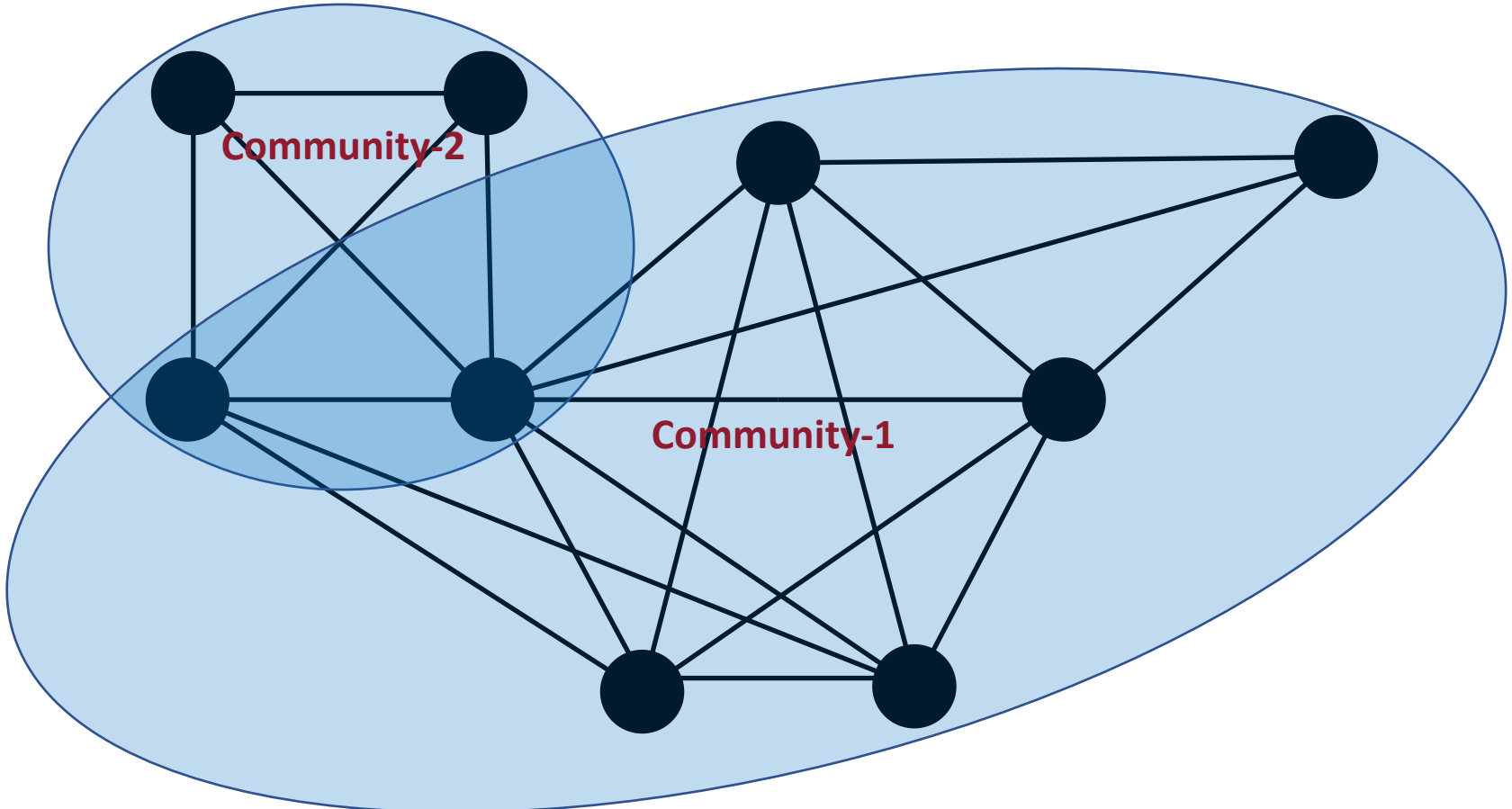


Community 2

Finding Overlapped Communities by CPM

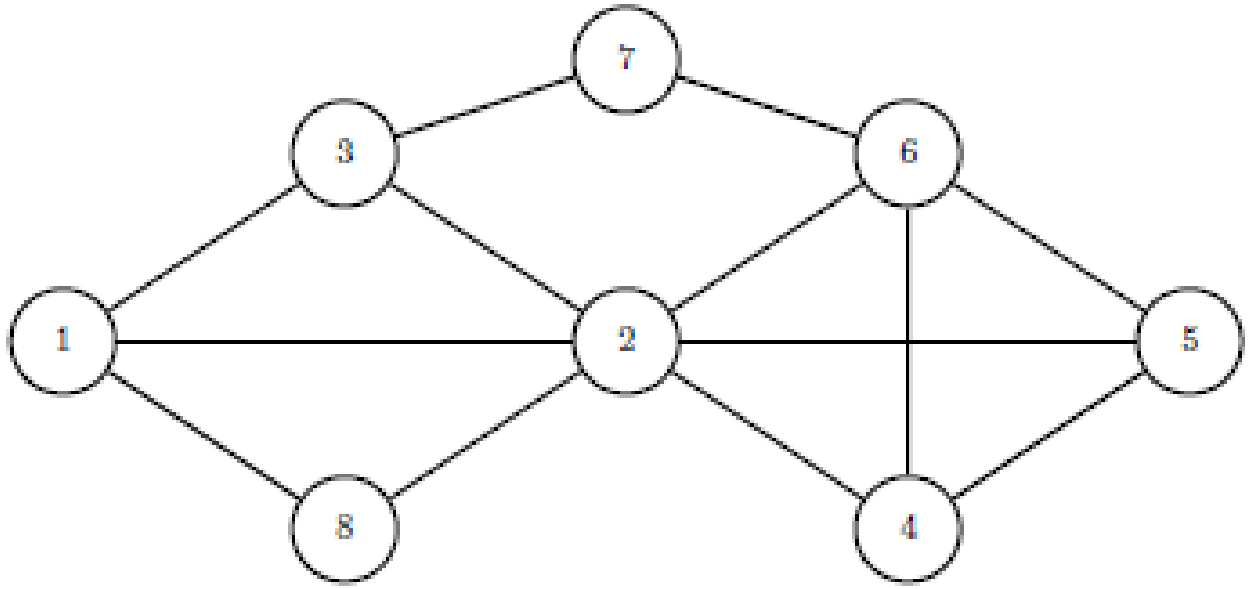
Cliques to k-Clique Communities

Identified communities Community-1 and Community-2 are overlapped.



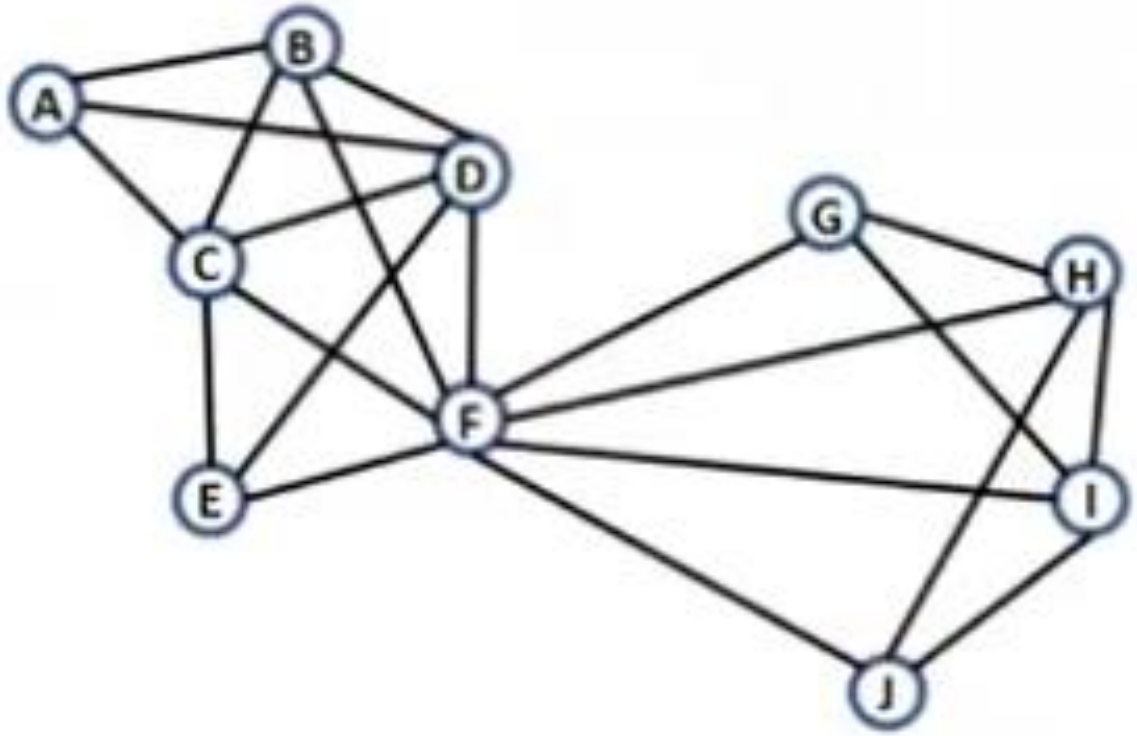
Finding Overlapped Communities by CPM

- **Exercise 1:** Find communities and overlapped communities of size greater than or equal to 3 using Clique Percolation Method (CPM). Show matrix computation.



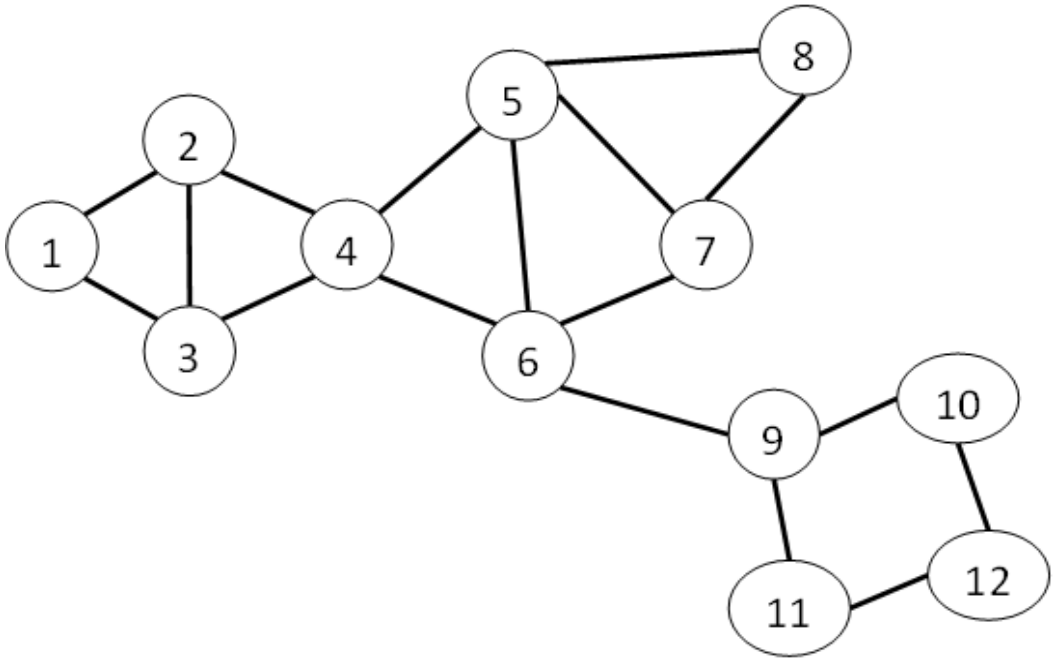
Finding Overlapped Communities by CPM

- **Exercise 2:** Given G and k=3. Find communities and overlapped communities using CPM. Show matrix computation.



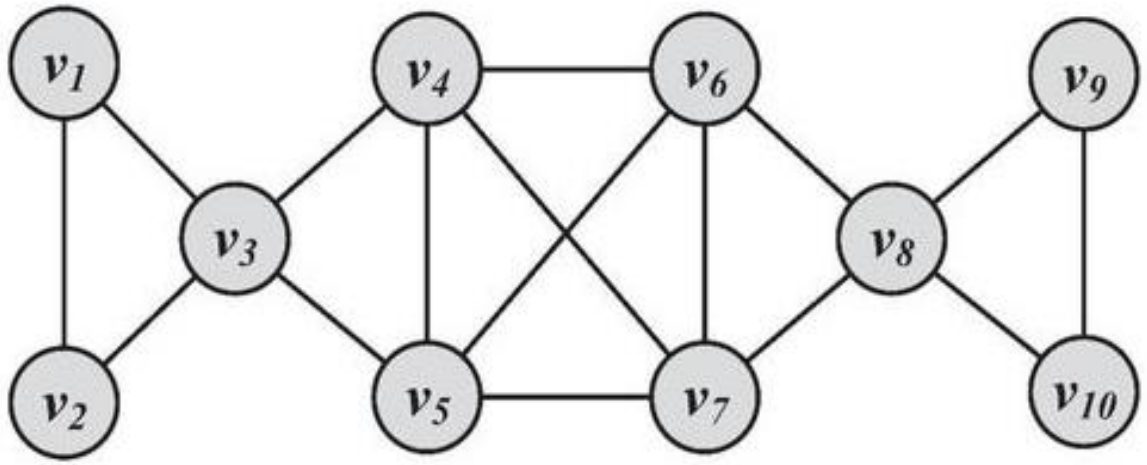
Finding Overlapped Communities by CPM

➢ **Exercise 3:** Given G and k=3. Find communities and overlapped communities using CPM. Show matrix computation.



Finding Overlapped Communities by CPM

➢ **Exercise 4:** Given G and k=3. Find communities and overlapped communities using CPM. Show matrix computation.



Finding Overlapped Communities by CPM

CFinder: (http://www.cfinder.org/)

- **CFinder** is a free software for finding and visualizing overlapping dense groups of nodes in networks, based on the Clique Percolation Method (CPM) of Palla et. al., *Nature* **435**, 814-818 (2005).
- CFinder was recently applied to the quantitative description of the evolution of social groups: Palla et. al.
- CFinder offers a fast and efficient method for clustering data represented by large graphs, such as genetic or social networks and microarray data.
- CFinder is also very efficient for locating the cliques of large sparse graphs.

- **CFinder Manual:** **http://hal.elte.hu/cfinder/wiki/?n=Main.Manual**

CONCLUSION

- Widely used algorithm for detecting overlapping communities
- However:
  - Fail to give meaningful covers for graph with few cliques
  - With too many cliques, might give a trivial community structure
  - Left out vertices?
  - Subgraphs containing many cliques == community?
  - What value of k to choose to give a meaningful structure?

Finding Overlapped Communities by CPM

References

- Palla et al. – Uncovering the overlapping community structure of complex networks in nature and society
- Santo Fortunato - Community detection in graphs
- CPM PPT of Eugene Lim

THANK YOU

Prakash C O

Department of Computer Science and Engineering

coprakasha@pes.edu

+91 98 8059 1946

Finding Overlapped Communities by CPM

Algorithm works as follows:

1. All cliques are found for different values of *k*.
2. A square matrix , where n is the number of cliques found, is created. Each cell [*i*, *j*] contains number of nodes shared by cliques *i* and *j*.
3. All cliques of size equal or greater than *k* are selected and between cliques of the same size connections are found in order to create a *k*-clique chain.