

# Communities

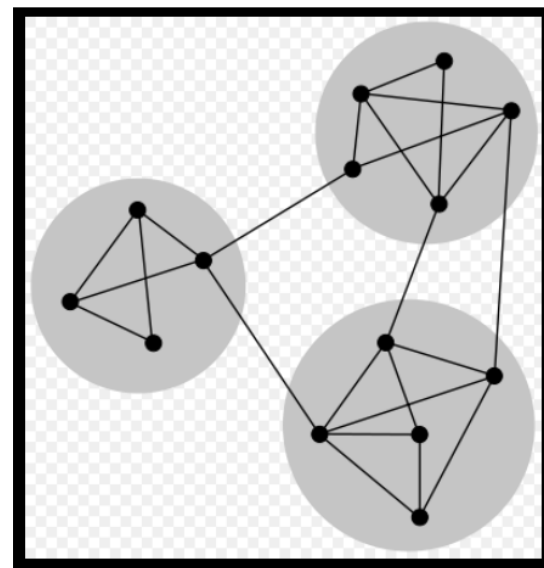
Ifeoma Adaji

# Learning outcomes

- Describe communities in networks with examples
- Explain what makes a community
- Describe how communities are formed

# Communities

- A community is a group of nodes that have a higher likelihood of connecting to each other than to other nodes from other communities
- A community is the gathering of nodes into groups such that there is a higher density of edges within groups than between them
- A community is a group of nodes in a network that are more densely connected internally than with the rest of the network
- Network divides naturally into groups of nodes with dense connections internally and sparser connections between groups
- Overlapping is allowed
- Examples in social networks:
  - Employees of a company
  - Circle of friends
  - Groups of individuals who pursue same hobbies
  - Students in same school
  - People living in same neighbourhood
- Communities are important
  - Play a role in how diseases spread, how misinformation spreads, etc.

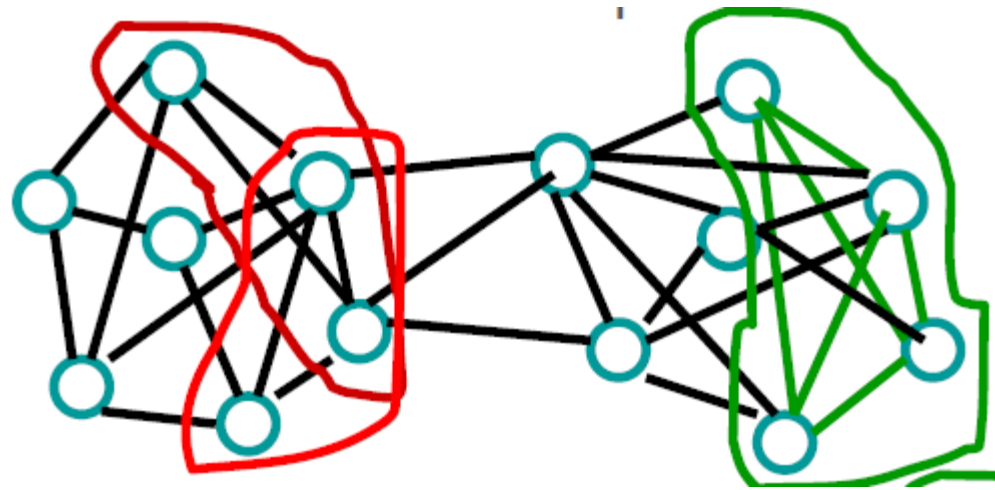


# What makes a community

- Mutuality of ties
  - everybody in the group knows everybody else (cliques)
- Frequency of ties among members
  - everybody in the group has links to at least  $k$  others in the group ( $k$ -cores)
- Closeness or reachability of subgroup members
  - individuals are separated by at most  $n$  hops ( $n$ -cliques)

# Cliques

- Cliques are formed when every member of the group has links to every other member
- Cliques can overlap

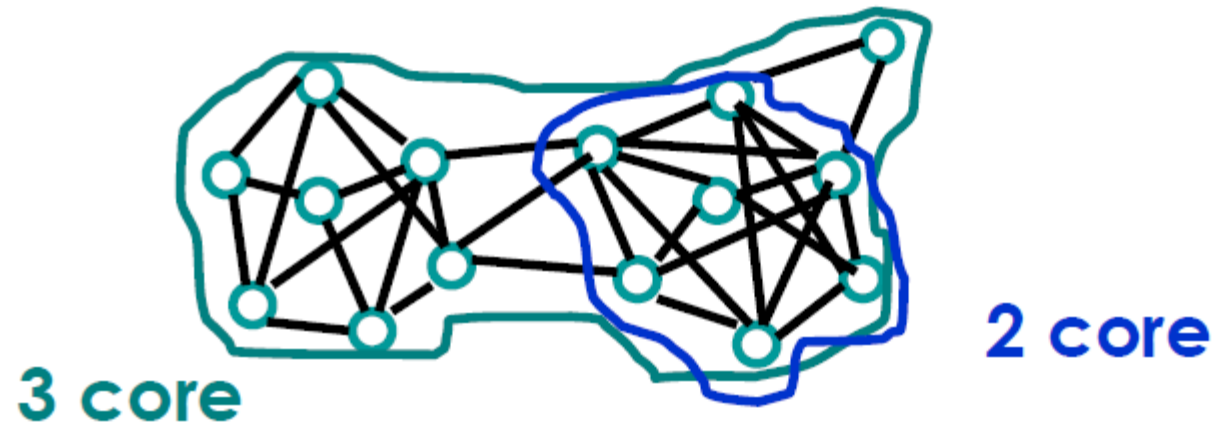
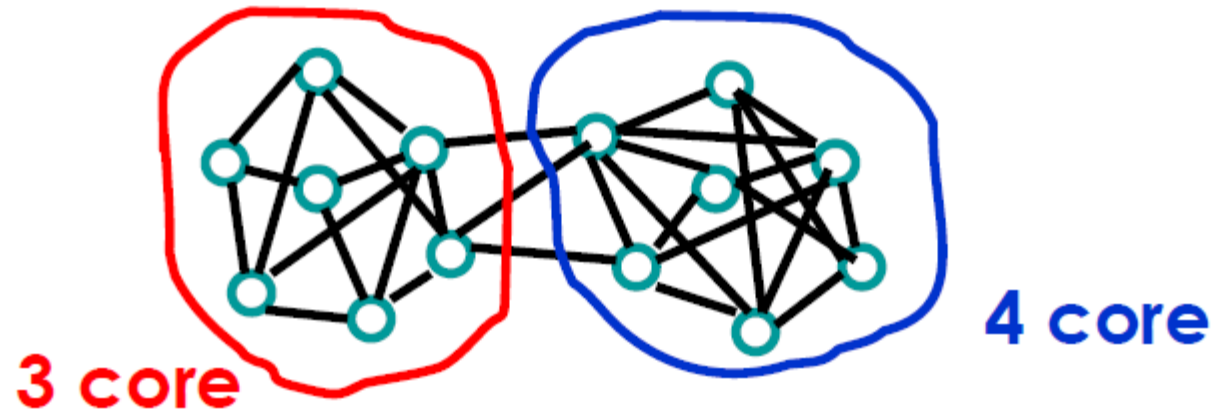


overlapping cliques of size 3

clique of size 4

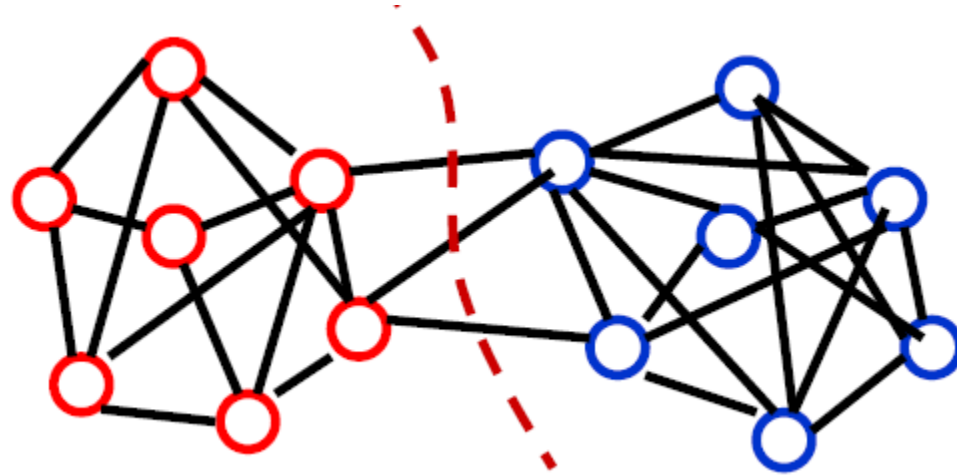
# k-cores

- Similar to cliques but is less stringent
- Each node within a group is connected to at least  $k$  other nodes in the group



# Community finding

- Social and other networks have a natural community structure
- We want to discover this structure rather than impose a certain size of community or fix the number of communities



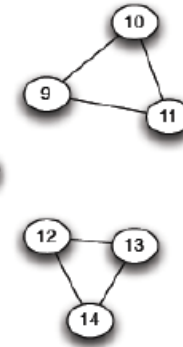
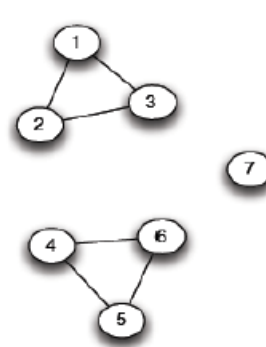
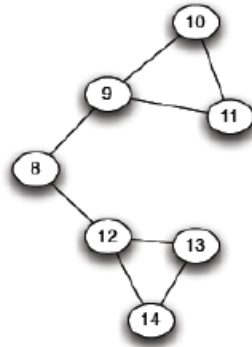
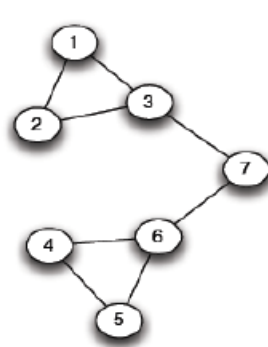
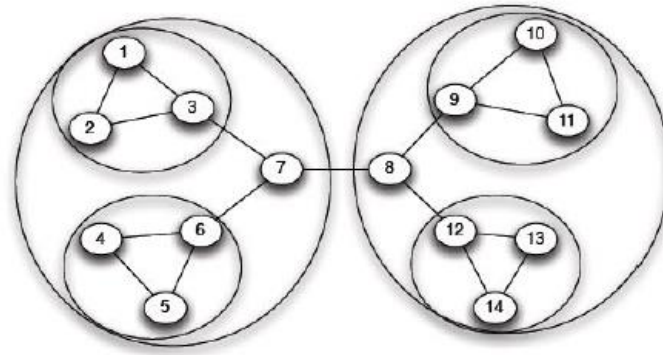
# Betweenness clustering

- Algorithm
  - compute the betweenness of all edges
  - while (betweenness of any edge  $>$  threshold):
    - remove edge with highest betweenness
    - recalculate betweenness
- Betweenness needs to be recalculated at each step
  - removal of an edge can impact the betweenness of another edge
  - very expensive, slow
  - does not scale to more than a few hundred nodes, even with the fastest algorithms



# Betweenness clustering

- Successively remove edges of highest betweenness (the bridges, or local bridges), breaking up the network into separate components



(a) Step 1

(b) Step 2

# Betweenness clustering

- Girvan and Newman
  - Uses edge betweenness as a metric to identify the boundaries of communities.
  - The algorithm makes heavy demands on computational resources
  - Slow

# Other algorithms

- Clauset-Newman-Moore (CNM)
  - Uses modularity
  - faster than the algorithm of Girvan and Newman
- Wakita-Tsurumi
  - Uses consolidation control

# Summary

- Communities are group of nodes that have a higher likelihood of connecting to each other than to other nodes from other communities
- Examples of communities in a network include employees of a company, circle of friends, groups of individuals who pursue same hobbies, students in same school, people living in same neighbourhood
- Communities can be formed through mutuality of ties, frequency of ties among members, closeness or reachability of subgroup members
- Several *community finding* algorithms exist including betweenness clustering.

# References

- *Network science* by Barabási, Albert-László. Cambridge University Press, 2016.
- *Linked. How everything is connected to everything else and what it means for business, science and everyday life* by Barabási, Albert-László, 2002.
- Lectures notes in Social Computing. Julita Vassileva.