

# Community Detection

Peter Carragher



Carnegie Mellon University



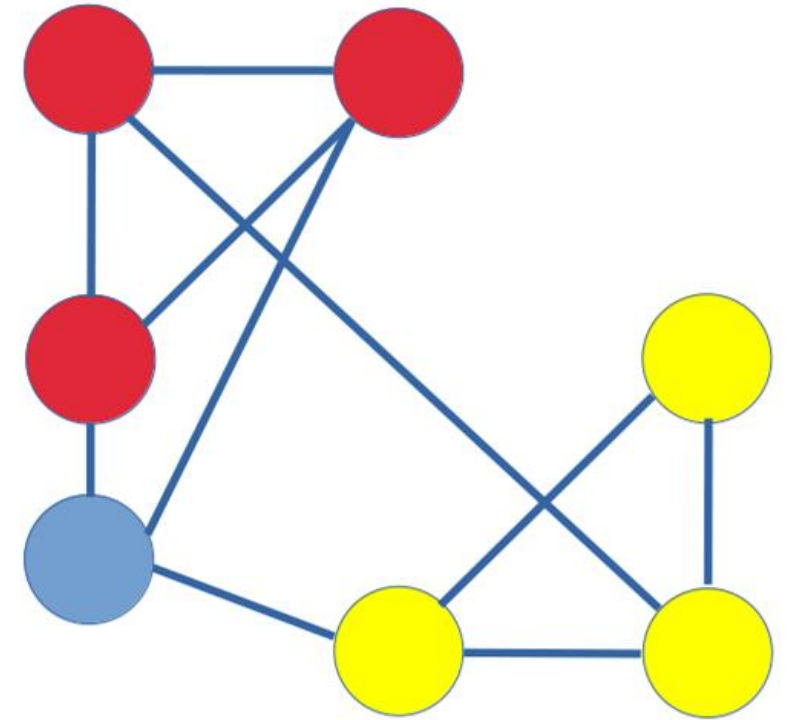
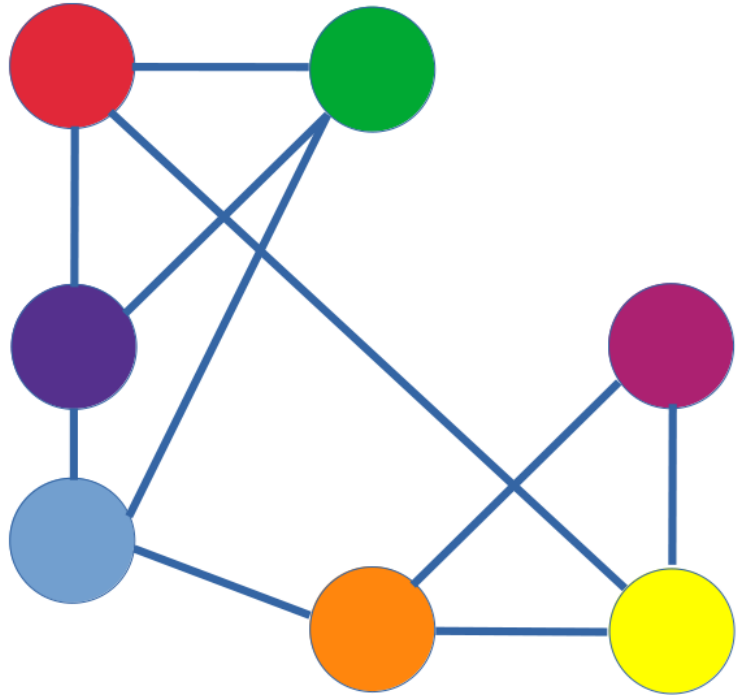
# Learning goals

This session will introduce measures to characterize groups of entities (e.g., users, hashtags, URLs) and the information space as a whole.

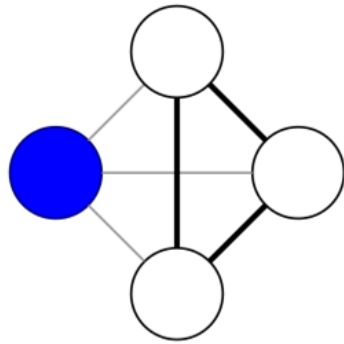
These methods allow us to address questions like:

- ❑ How are communities structured? How fragmented are they?
- ❑ What are the dominant narratives and information sources? How do they relate to each other?
- ❑ To what extent is information spreading hierarchically?
- ❑ How does the network, groups, and narratives evolve over time?

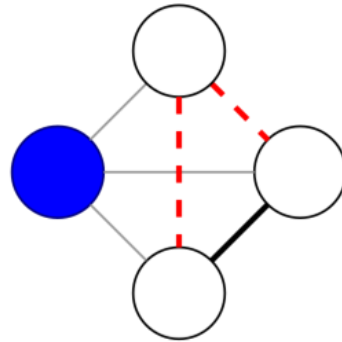
# Community Detection



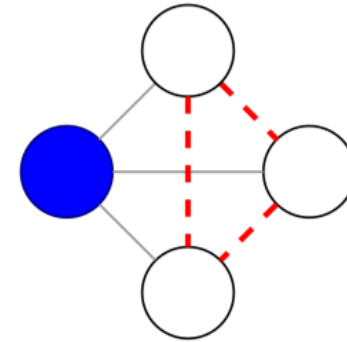
# Clustering Coefficient



$$c = 1$$



$$c = 1/3$$



$$c = 0$$

# Modularity: what makes a good group?

$$Q = \frac{1}{(2m)} \sum_{vw} \left[ A_{vw} - \frac{k_v k_w}{(2m)} \right] \delta(c_v, c_w)$$

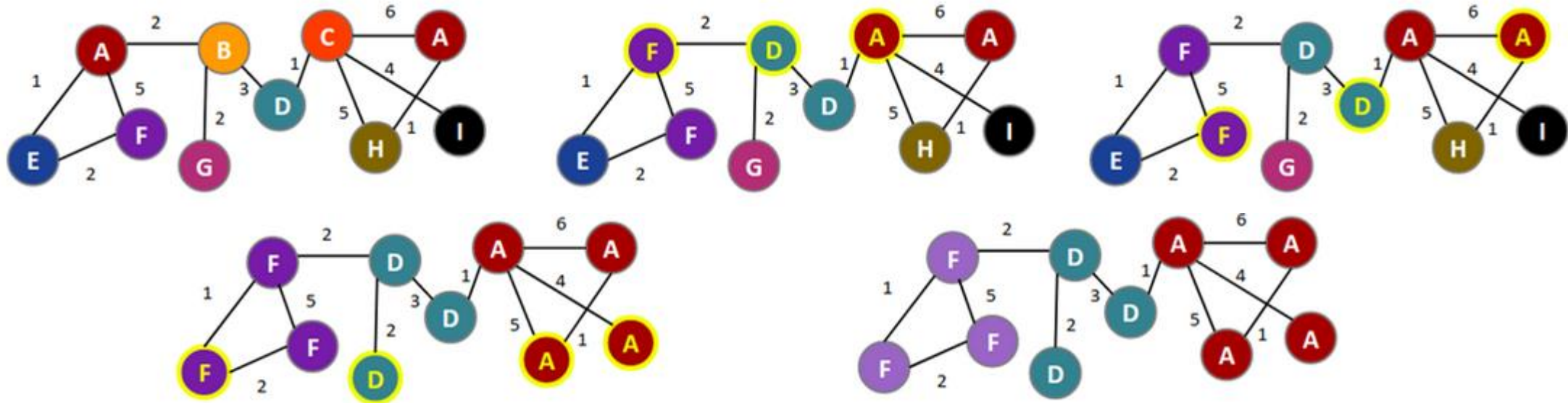
# edges connecting u & v

E[A<sub>vw</sub>] given degree

Are v and w in the same community?

# Louvain and Modularity Maximization

- Iteratively choose a community to merge nodes into such that we increase (maximize) modularity



# Louvain and Modularity Maximization

- However, Louvain (modularity maximization) finds communities in fully random graphs
  - ❑ It cannot be used to find statistically significant community structures in empirical networks.
- It is unable to detect small communities, as it suffers from a resolution limit
- Leiden addresses these issues
- Girvan-Newman: iteratively remove edges with high betweenness
- Other methods introduce soft clustering (nodes can be in multiple groups)

# CONCOR (CONvergence of iterated CORrelations)

