BADJI MOKHTAR University – Annaba

Social Network Analysis

Master II, Big Data Management & Analysis (BDMA)/(GADM)

Dr, Mohammed BESNACI

2024/2025

Content

- Introduction
- Reminder on graph theory
- Measures and analysis techniques
  - Clustering Coefficient
  - Distance Measures
  - Robustness Measures
  - Centrality Measures
- Graph models
- Community detection
- Tools
  - NetworkX Python Library
  - GEPHI the analysis graphical tool

Part I

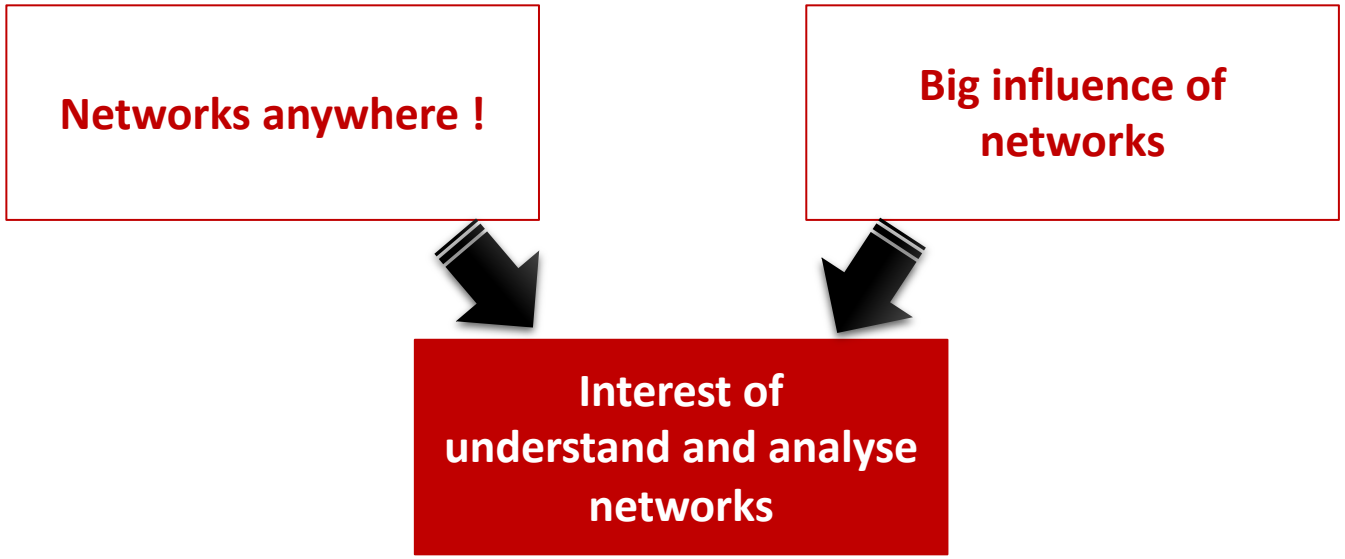
Introduction

Key Social Media Statistics and Facts

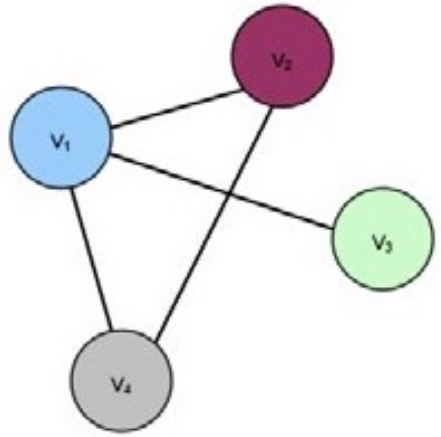
- If **Twitter** users formed a country, it would be the 4th largest in the world.
- There are over 50 billion selfies on **Instagram**.
- **TikTok** played a major role in influencing **Gen Z's** political activism, including mobilizing voters during the 2020 U.S. presidential election
- 90% of businesses are on **LinkedIn**, making it a primary tool for recruitment.
- 87% of consumers trust recommendations from peers over advertisements.
- Over 3 million pieces of content are shared daily on **Facebook**.
- In 2020, 49.03% of the global population (7.77 billion people) were active social media users.

Network anywhere !

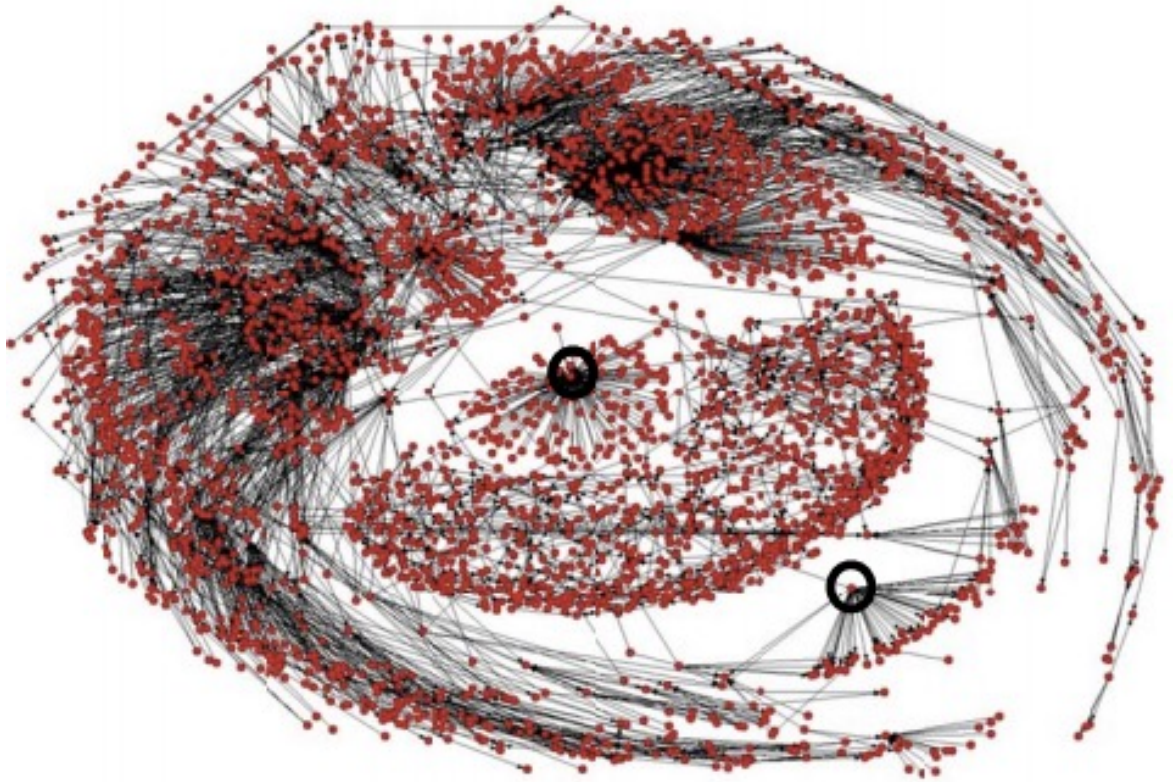
- **Computing**: web pages, routers, P2P, etc.
- **Biology**: proteins, brain neurons, etc.
- **Social sciences**: friendships, collaboration, contacts, marriages, etc.
- **Economics**: financial exchanges, collaboration, etc.
- **Transportation**: air networks, roads, electricity grids, water networks, etc
- **Linguistics**: synonymy, co-occurrence, etc.
- **Science**: articles, co-authors, topics, citations, etc.


Networks anywhere !
Big influence of networks
Interest of understand and analyse networks

Definitions (Network)



- A **network** is a set of nodes connected by edges
- It is a simple (mathematical) abstraction of a complex system.
- **Node**, vertex, actor, ...
- **Edge**, connection, link, relationship, ...
- **Graph** vs **Network**.
- The structure of a network is often non-trivial.

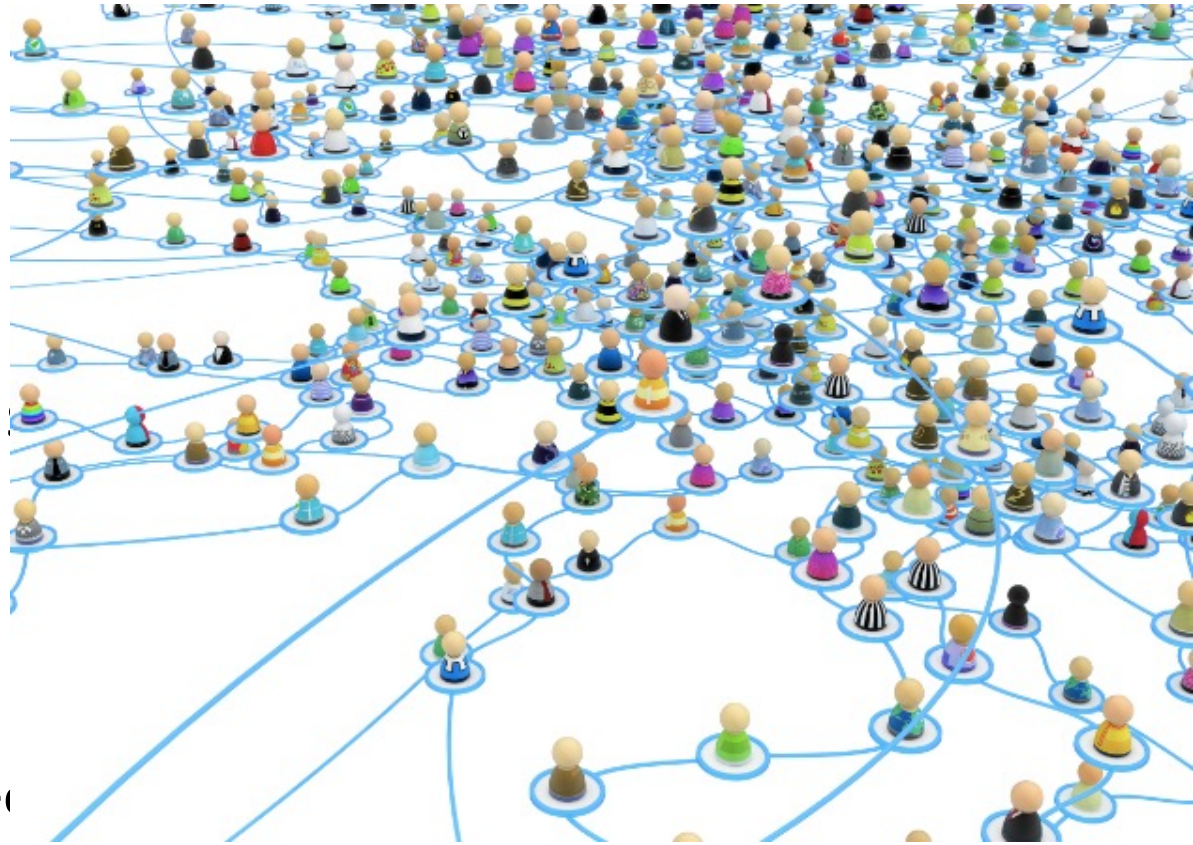


Definitions (Social network)

- A **social network** is a simplified representation consisting of a set of actors connected by social interactions.

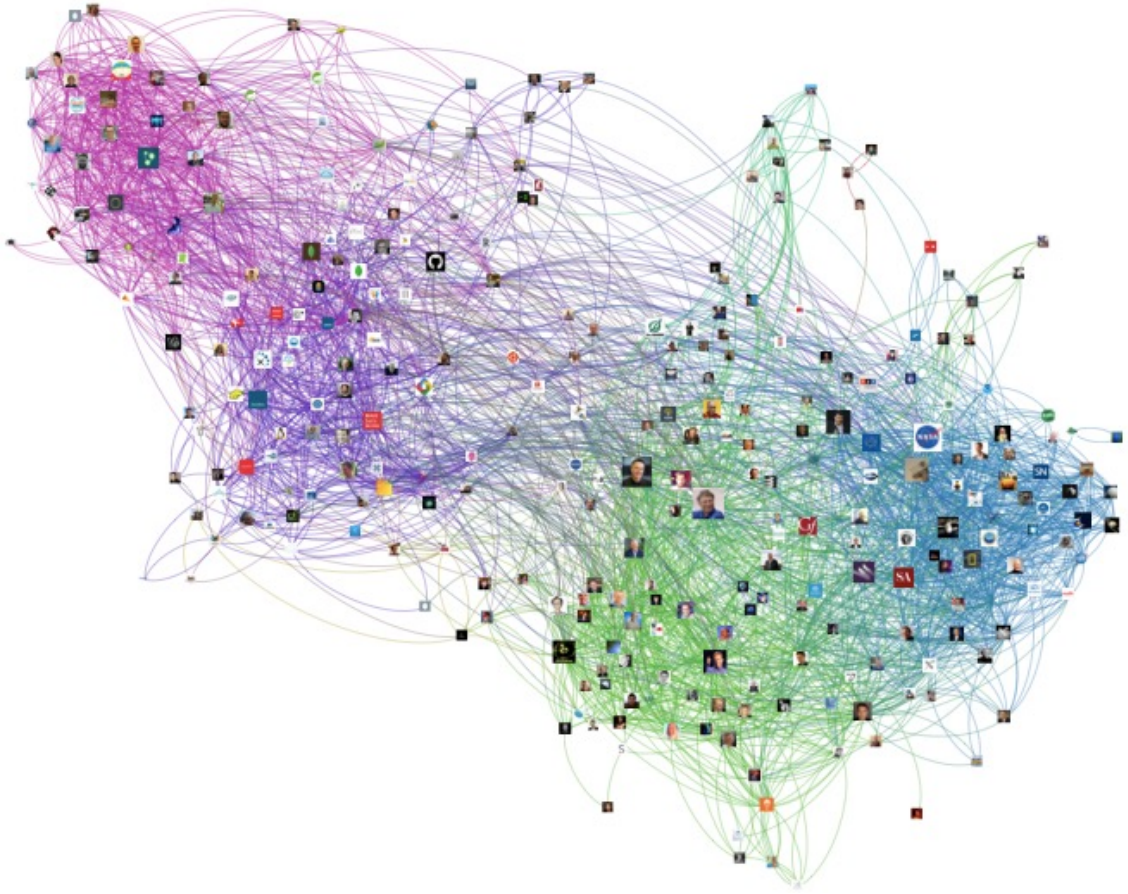
- **Social**, meaning everything related to society, individuals, groups of individuals, communities, and the relationships between individuals...

  - **Actors**: individuals, groups, or organizations.
  - **Social interactions**: familial, romantic, affinity-based, business, work-related... social, virtual worlds.



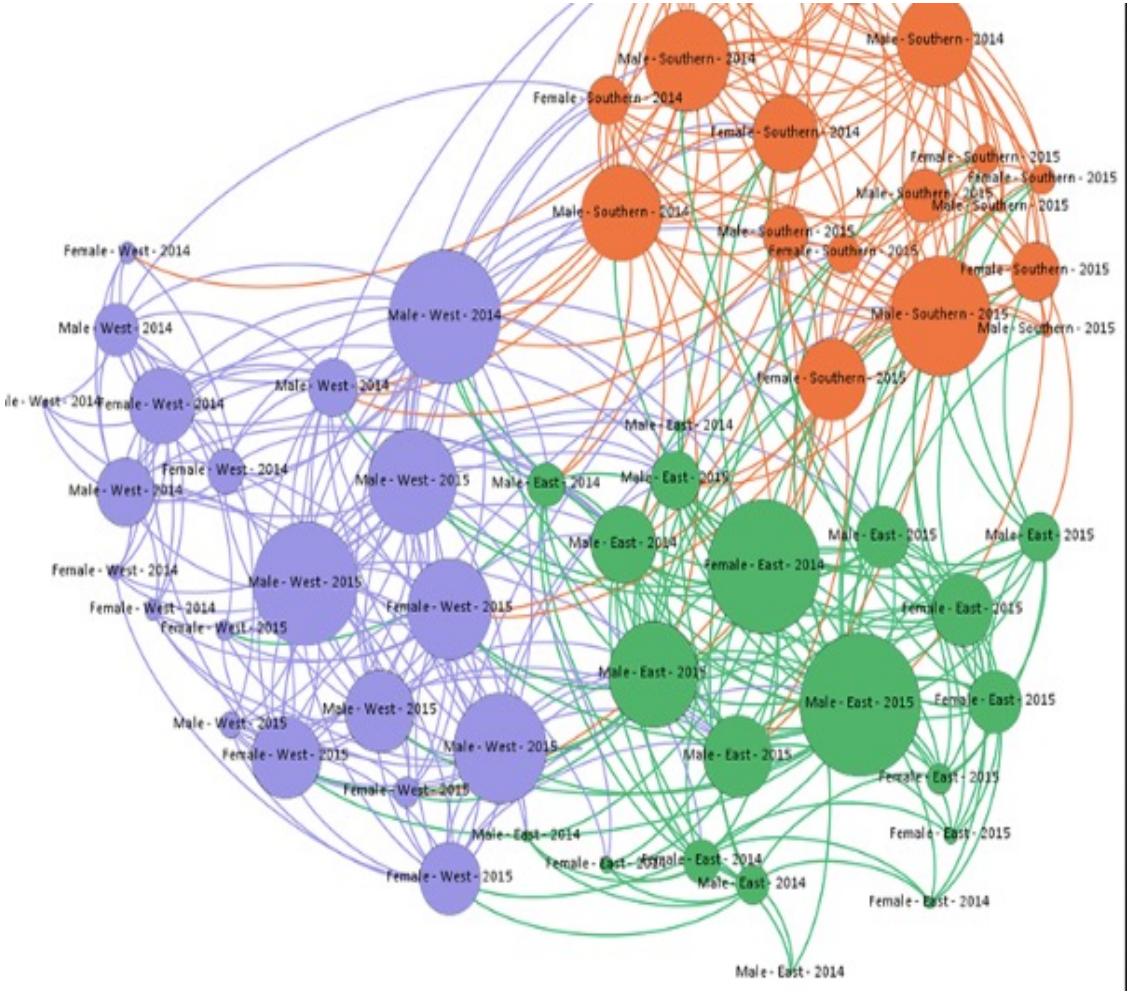
Definitions (Social network)

- Social relationships are formed, depending on the case, based on: origins, interests, needs, viewpoints, professional aspects.
- Classic & new social networks
  - Facebook, Twitter, instagram, Linkedin,
  - supplier/client networks, family networks, and more...



Social Network Analysis (SNA)

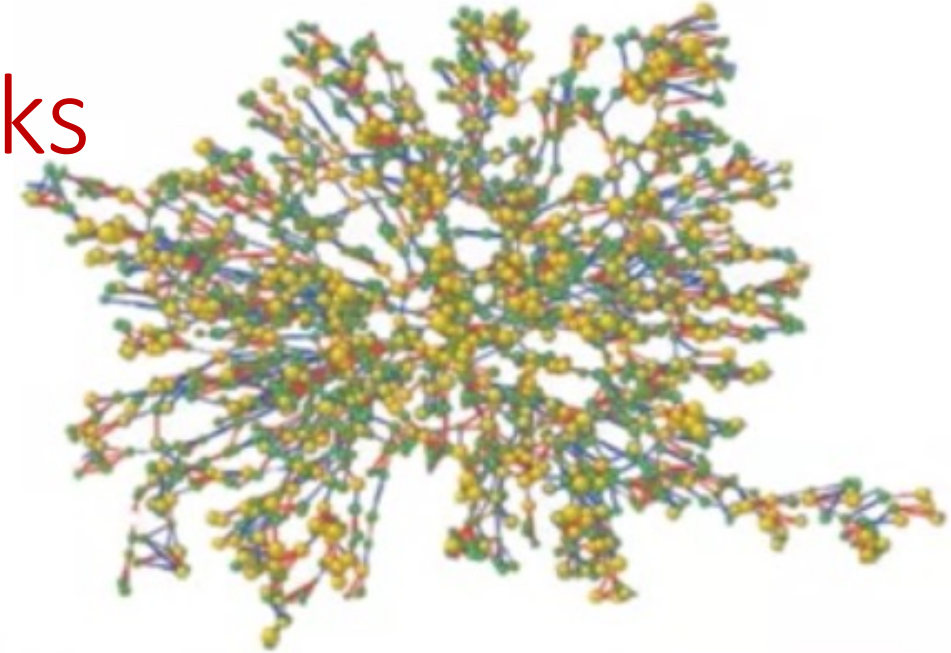
- **SNA** is a field of data analysis that uses networks and **graph theory** to understand social structures.
- It is seen as a toolkit for **visualizing**, **modeling**, and **analyzing** social structures.
- SNA techniques can also be applied to networks beyond the social context.



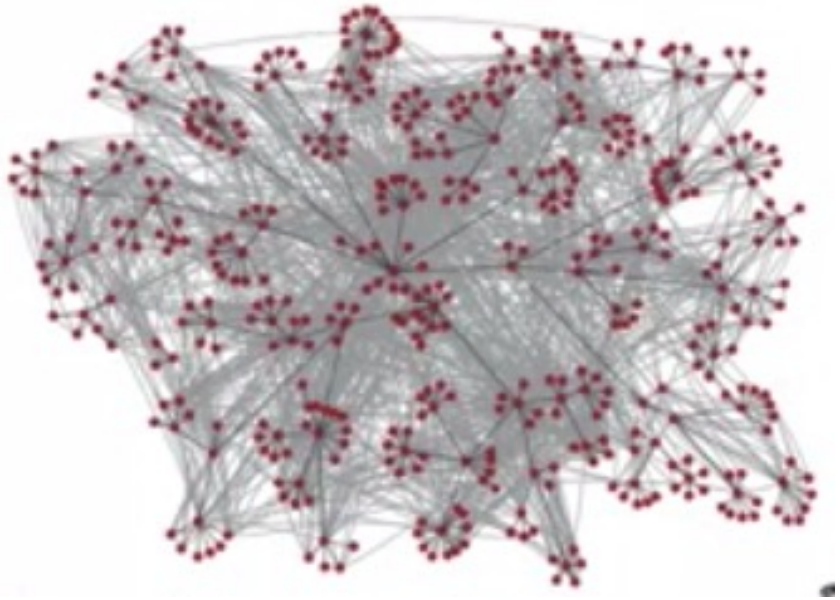
Social Network Analysis (SNA)

- **SNA is** a sociological approach based on the study of graph theory, applied to social networks (conceptualizes social relationships in terms of nodes and links).
- **Multidisciplinary**: computer science, sociology, mathematics, statistics.

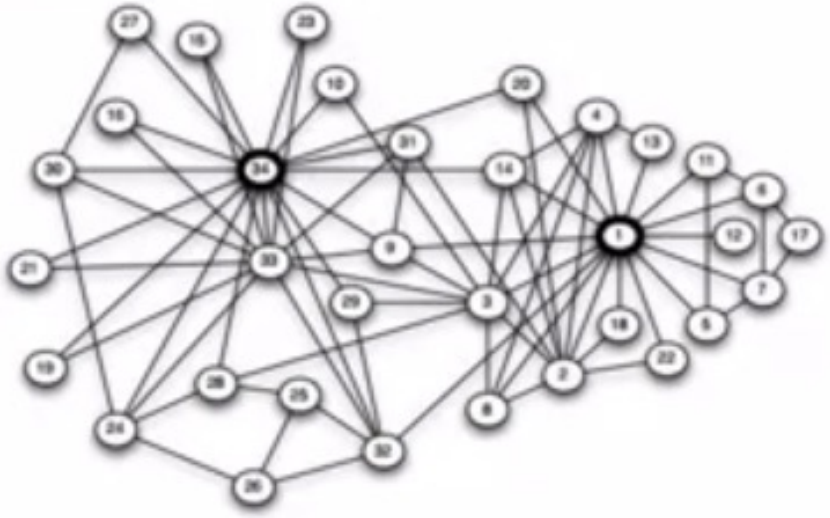
Social Networks



Friendship network, family marital connection among 2,200 people



Email communication network among 436 HP employees

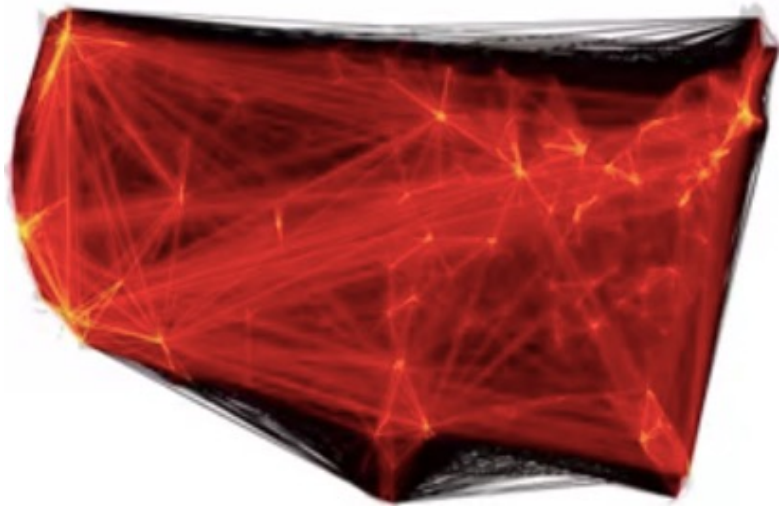


Friendship network in a karate club of 34 people

Transport Networks

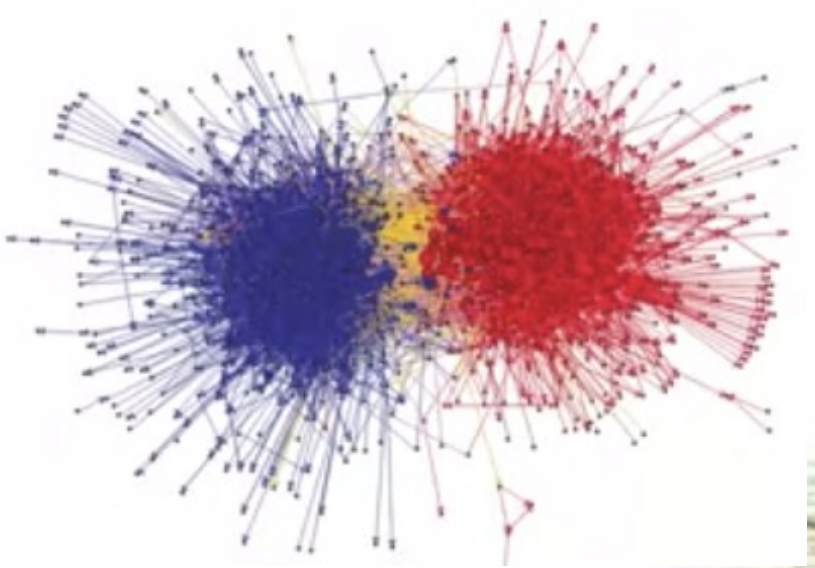
Direct flights network around the world

Bus transportation networks

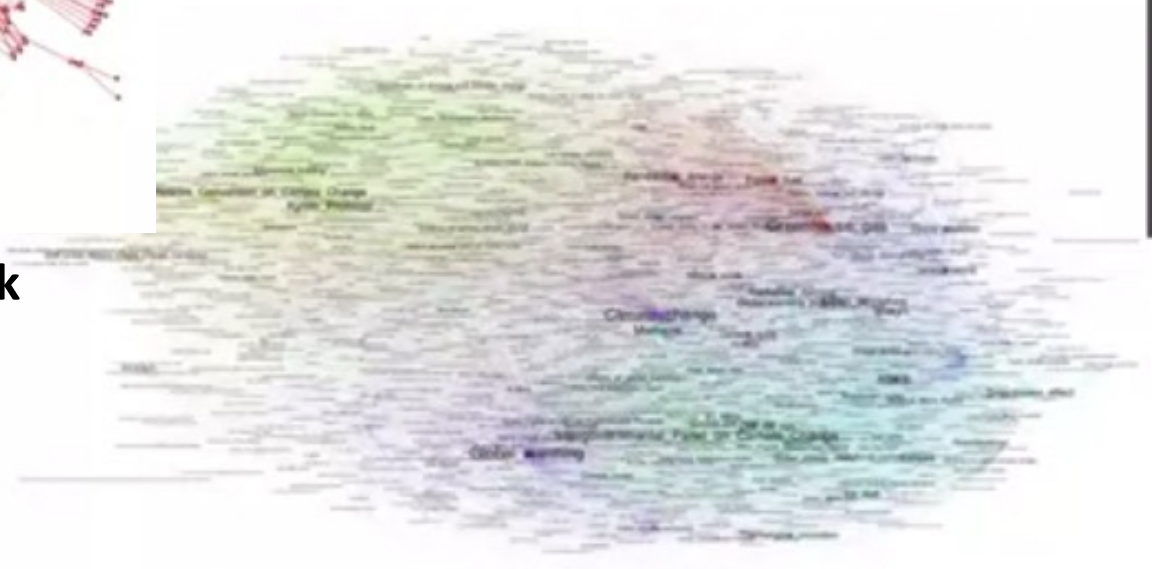


Human mobility network based on the location of one-dollar bills

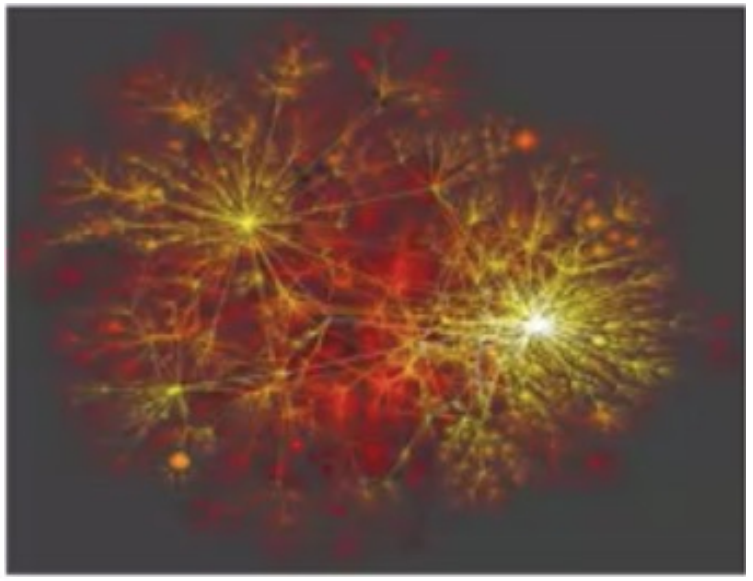
Information Networks



Communication network between left and right political blogs (US)

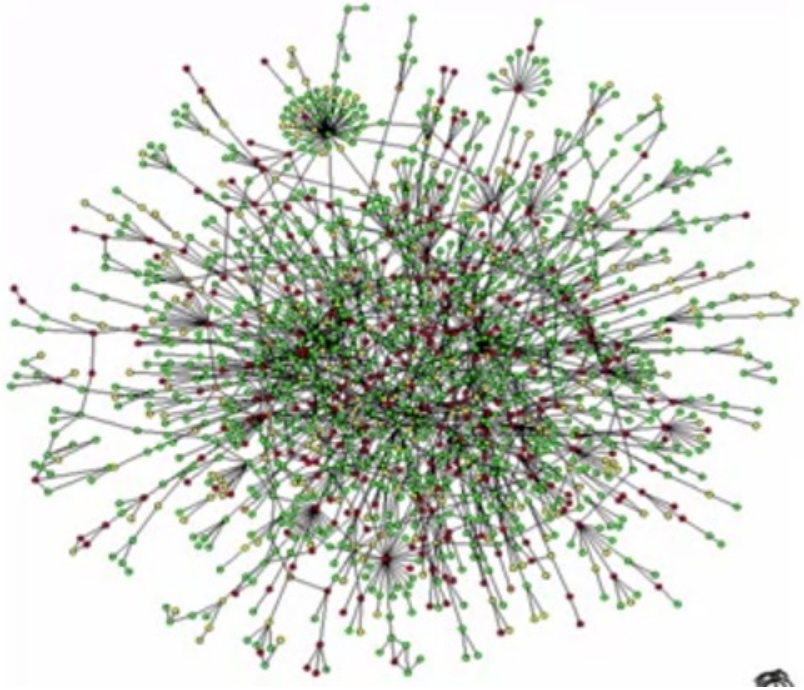


Network of Wikipedia articles on climate change



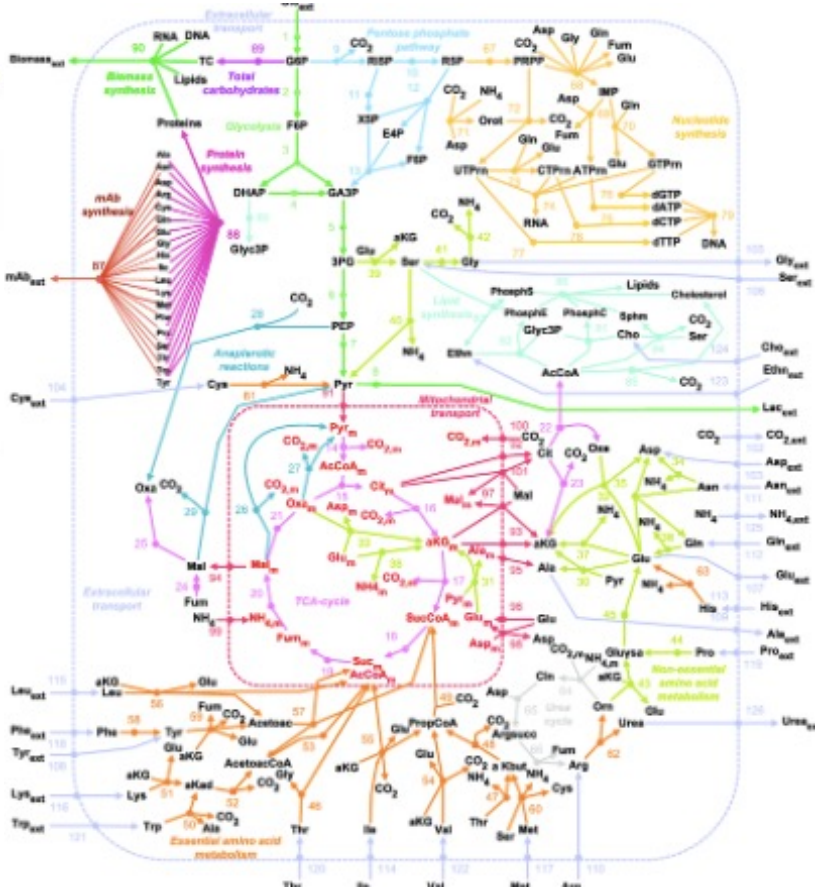
Internet connectivity network

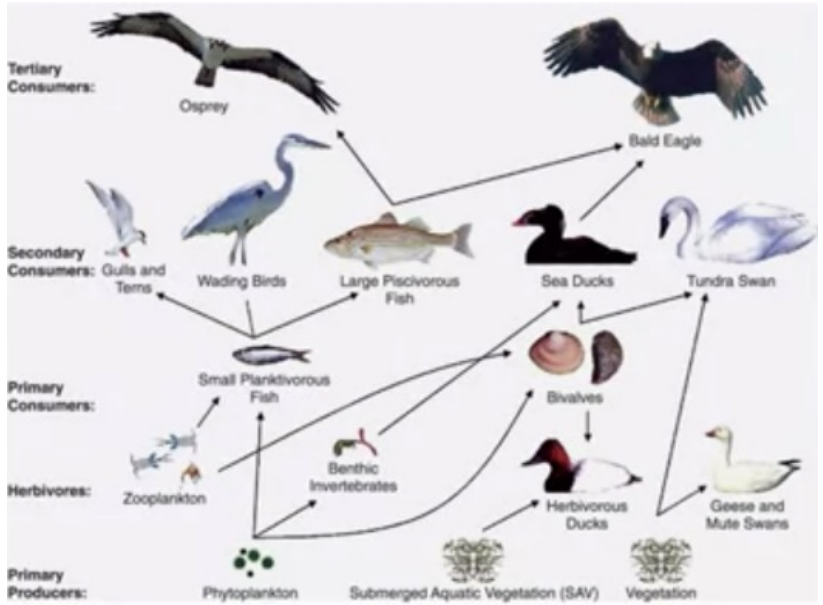
Biological Networks



Protein-protein interaction network

Metabolic network

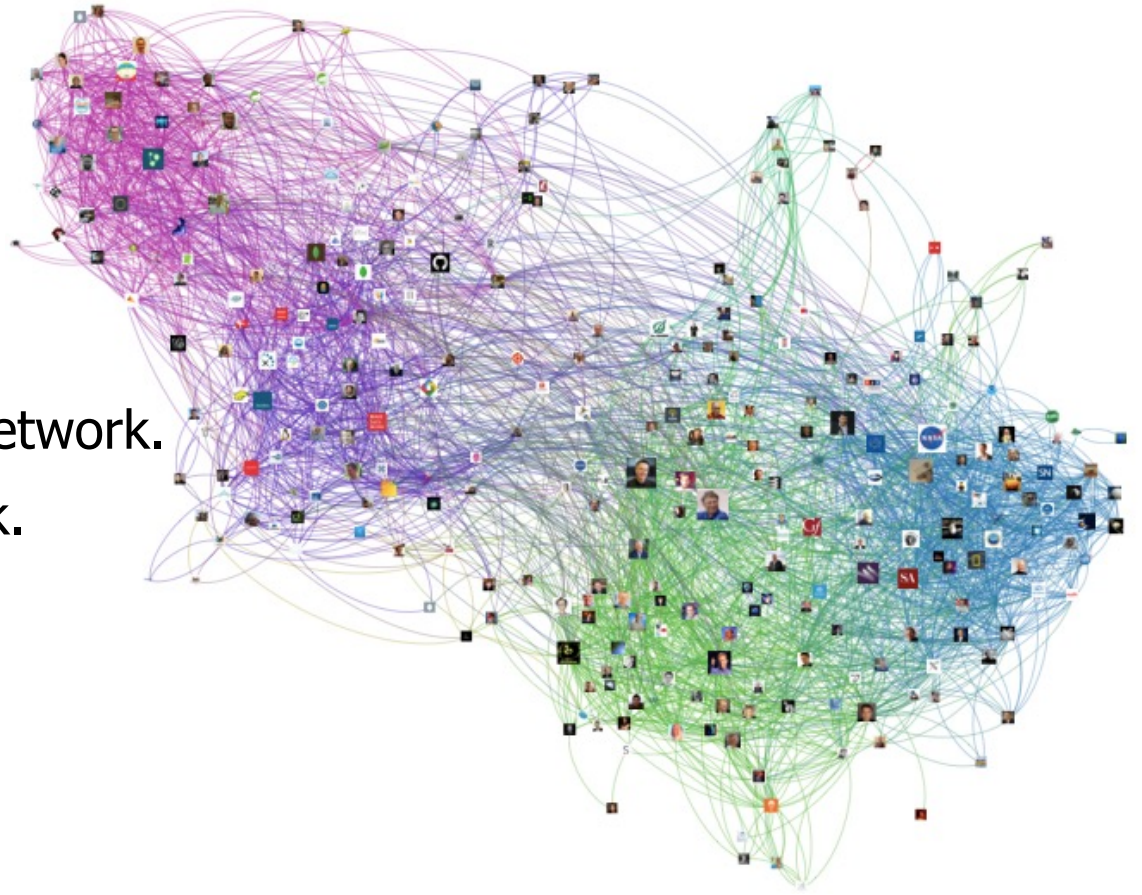




Trophic network of aquatic birds in the Chesapeake Bay

Why SNA ?

- To understand how social interactions influence a network.
- To learn how information flows within a network.
- To understand how structure affects processes (information diffusion, opinion formation, coordination/cooperation, resilience to attacks)
- To characterize the roles of individuals within a network.
- To characterize the communities within a network.
- To characterize the evolution of a network.

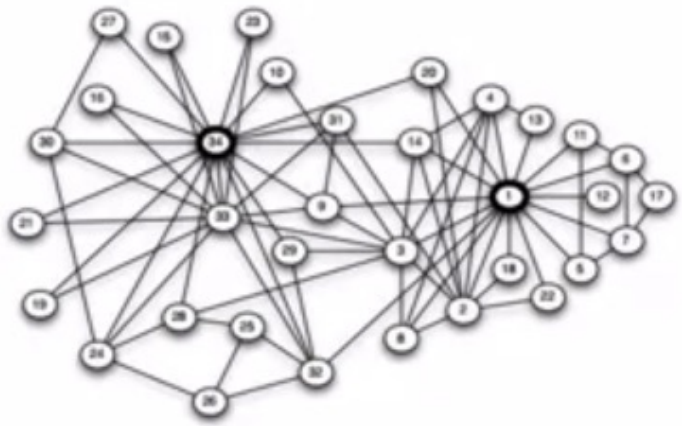


Applications

- Is a rumor likely to spread within this network?
- Who is the most influential person in this organization?

- Is this club likely to split into groups?
- If so, which node will belong to which group?

Email sending



Club Friendship

Applications

- Which airports are most exposed to the risk of virus spread?
- Are certain parts of the world more difficult to reach?

Airports Flights

- Do players exhibit violent behavior?
- Can we determine how addicted a user is to a game?

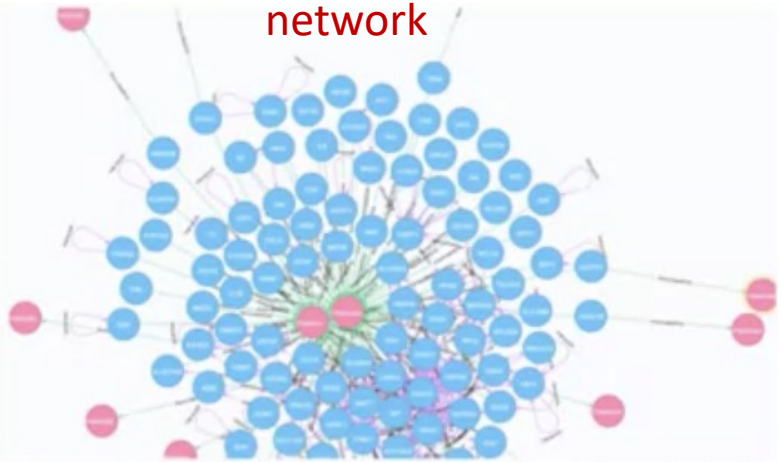
Gamer's tweets

Applications

- Is there a relationship between two diseases by studying the pathway between their genes?
- How do the phenotypes of undiagnosed diseases overlap?

- Are there groups of users?
- What are the most suitable professions for the users?

Biological network



User's Linkedin

Applications (domain)

- Protein interactions and disease spread (health).
- Terrorist networks (politics/security).
- Identifying opinion leaders and "influencers" (politics).
- Assessing network vulnerability (telecommunications).
- Understanding the diffusion of innovations (economics), rumors (sociology), ideas (culture), ...
- Knowledge management (semantic, linguistic networks, ...).

Applications (theme)

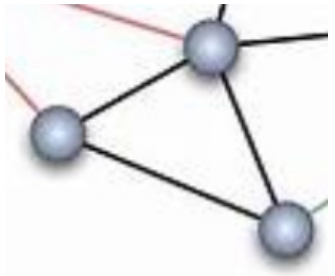
- Identifying new scientific, commercial trends, ... (evolution).
- Analyzing networks of experts, co-authors, co-citations, patents (importance, communities).
- Page importance analysis (PageRank).
- Using propagated trust to combat spam: in emails, in web page ranking.

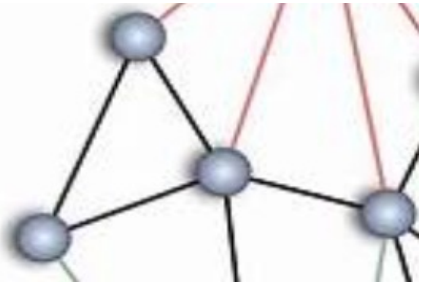
Research lines

- Node classification.
- Link prediction (friend recommendations, product recommendations).
- Prediction of growth and virality.
- Detection of disinformation and anomalies (rumors, propaganda, intrusion attempts, suspicious behavior, ...).

SNA levels (microscopique view)

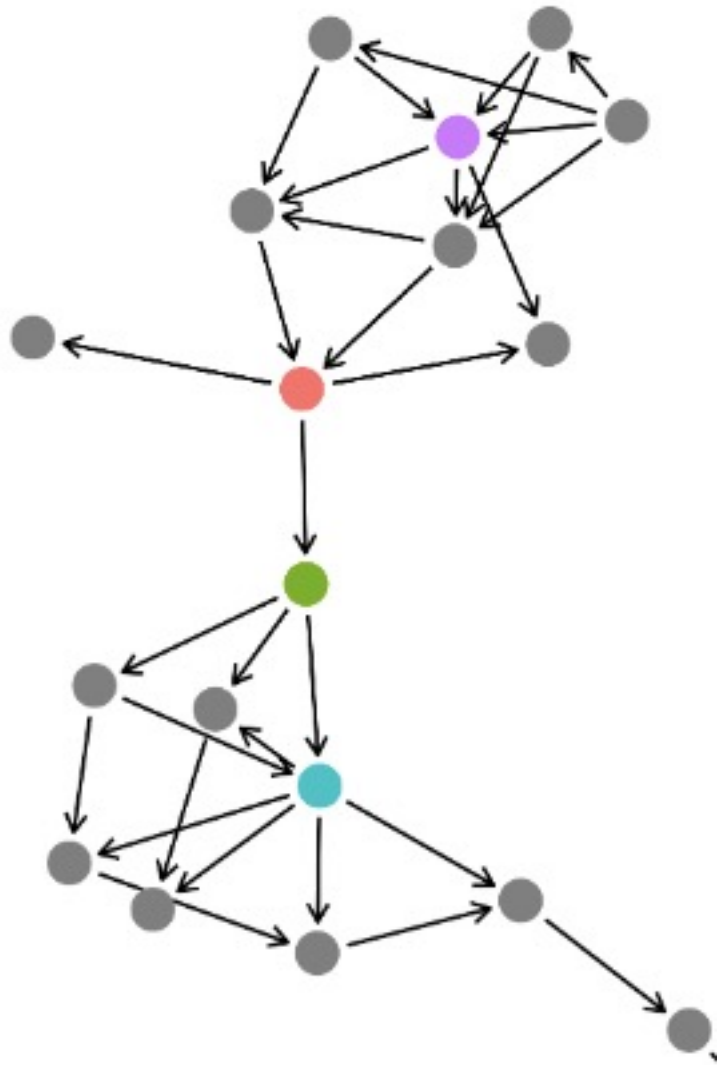
- **Node level**: the most basic, focused on nodes, individual properties (internal variables) and relational properties (degree, centrality, ...).
- **Dyad level**: interaction patterns between 2 nodes, properties of homophily, reciprocity, assortativity, ...
- **Triad level**: interaction patterns between 3 nodes, properties of triadic closure, clustering coefficient, ...
- **Ego-centered level**: interaction patterns between an ego and its alters





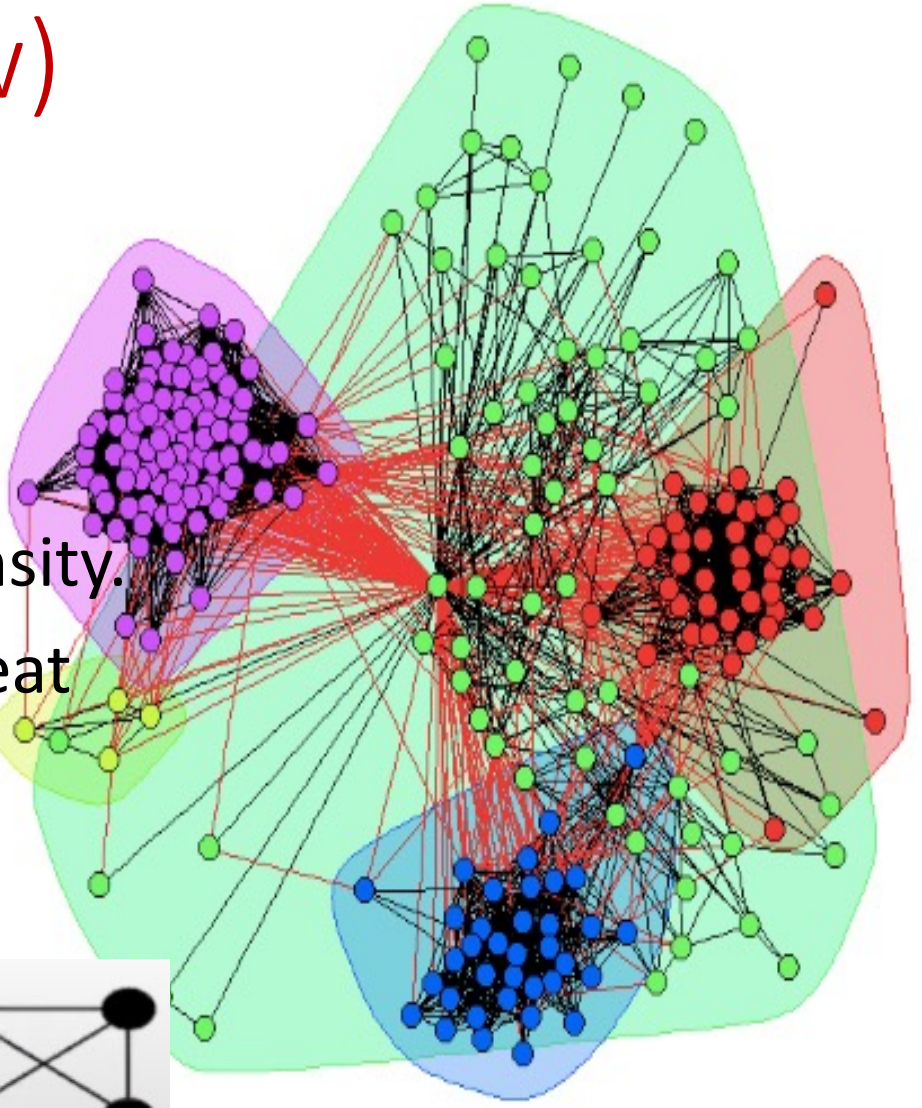
SNA levels (macroscopique view)

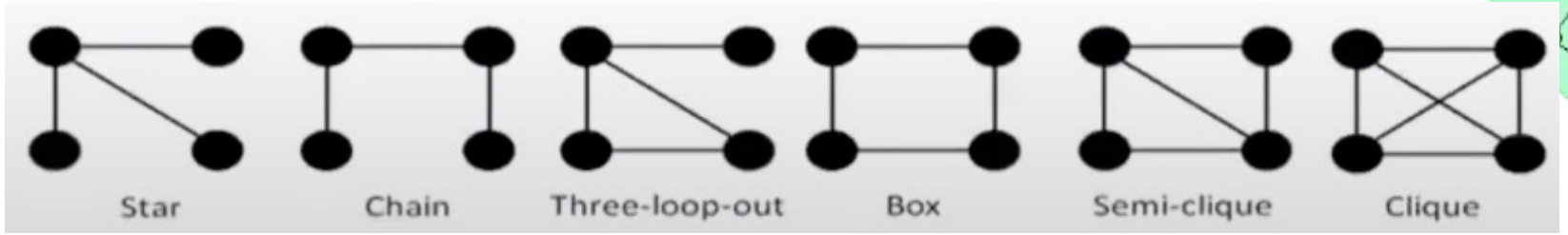
- **Focusing on the network as a whole.**
- Properties of interest:
  - Connectivity, Diameter, Degree distribution
  - Average shortest path length, Link density.
- **Examples:**
  - Very small diameter → star-shaped network.
  - High density → network in the form of cliques.



SNA levels (mesoscopique view)

- **Focusing on parts (subsets) of the network.**
- Communities: formed by frequent interactions between homogeneous nodes.
  - High internal density relative to external density.
- Network motifs: subgraphs that frequently repeat within the network.
  - Used to capture functional properties in the network.





SNA - Synthesis