

# NETWORK ANALYTICS

## Community Detection

**“A picture speaks a thousand words”**

# Kristina Khvatova

Software Developer  
Softec S.p.A.

Specialist in Computer Science and  
Applied Mathematics  
Saint-Petersburg State University

Master in Computer Science and  
Data Analysis  
Milano-Bicocca University



[kristina.a.khvatova@gmail.com](mailto:kristina.a.khvatova@gmail.com)

<https://www.linkedin.com/in/kristina-khvatova-a529b2114/>

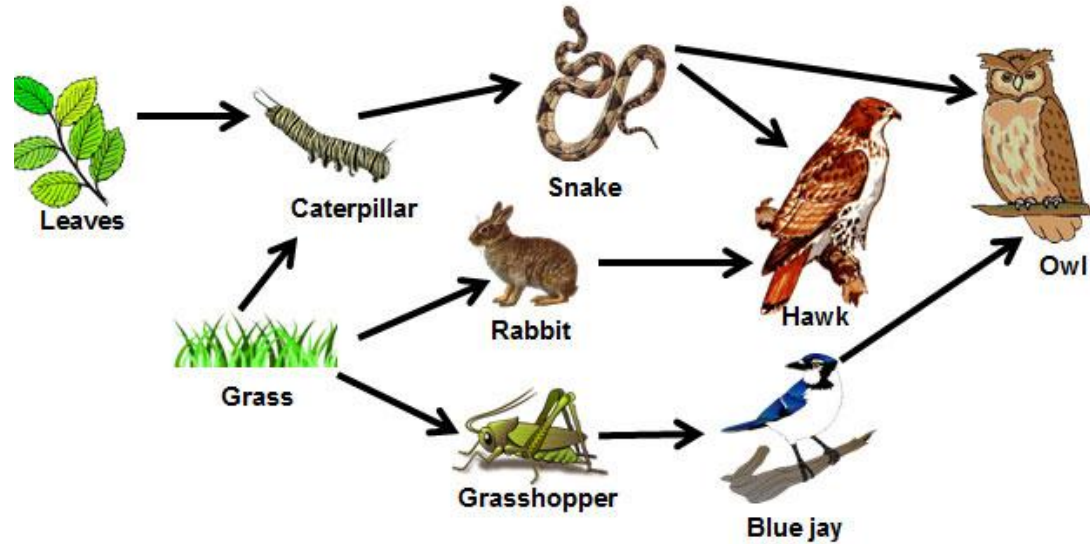
# Network

a large system consisting of many similar parts that are connected together to allow movement or communication between or along the parts, or between the parts and a control centre



# Examples

Animal's network



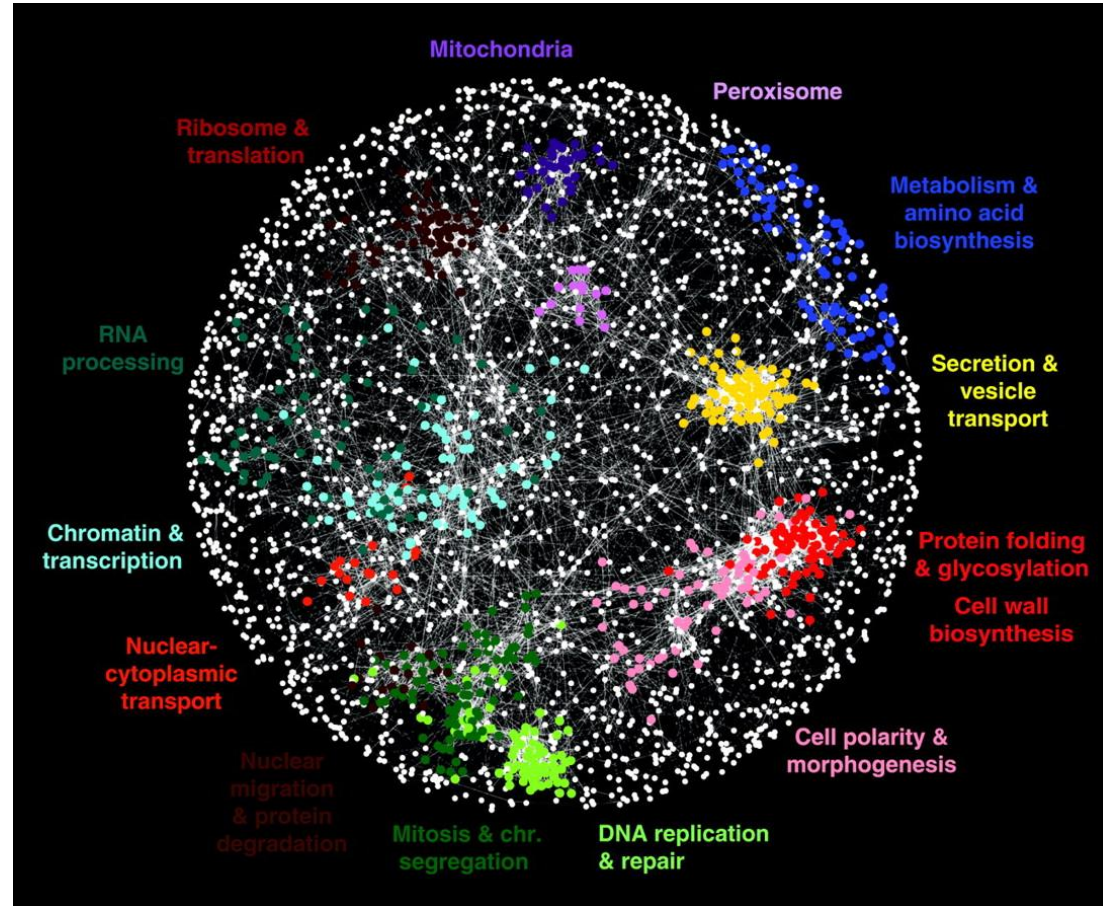
# Examples

Transport(Airplanes) network



# Examples

## Genetic network



# Common features

1. Nodes
2. Links

**Nodes:** Entities that are related to each other  
e.g., people, animals, towns...

- **Links:** Connections between nodes
- may be real and fixed,
- real and dynamic,
- abstract



# Network properties

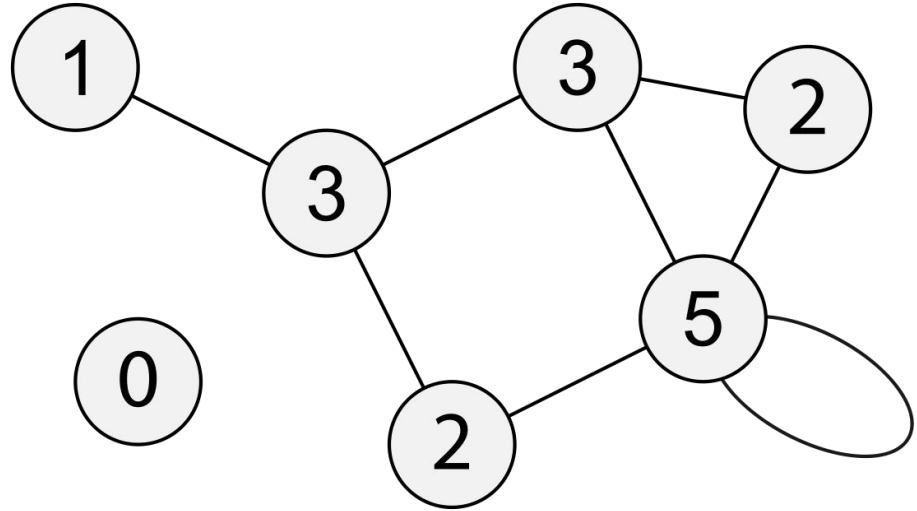
There is a set of quantitative measures (statistics) to describe and compare networks

- Degree
- Clustering
- Centrality:
  - Degree
  - Betweenness
  - Eigenvector



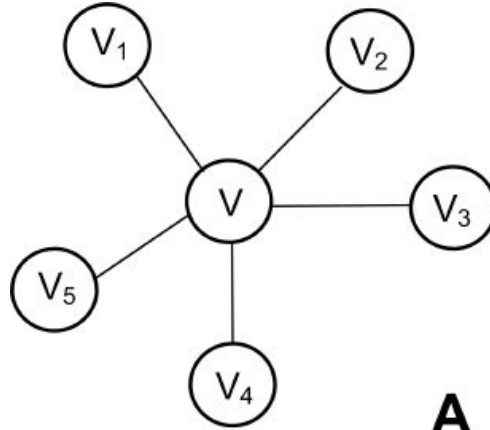
# Node degree

The **degree of a node** is the number of links connected to the node

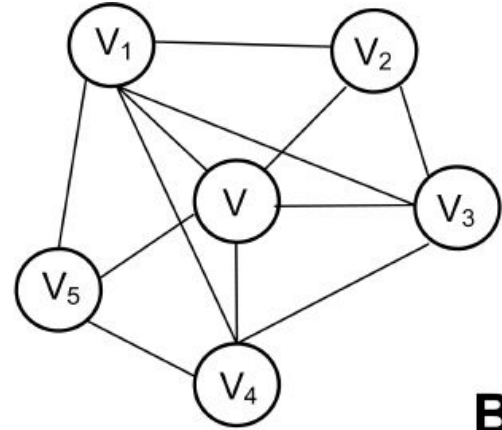


# Clustering coefficient

**Clustering coefficient** is a measure of the degree to which nodes in a graph tend to cluster together



**A**

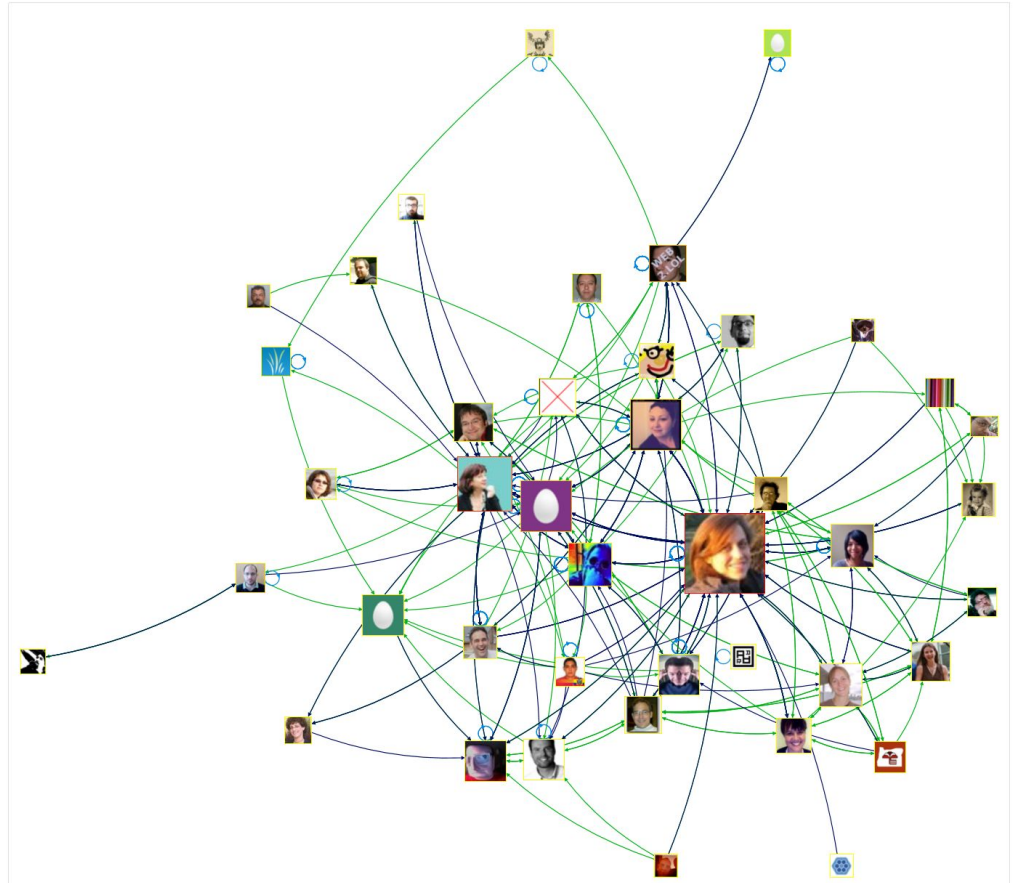


**B**

# Centrality

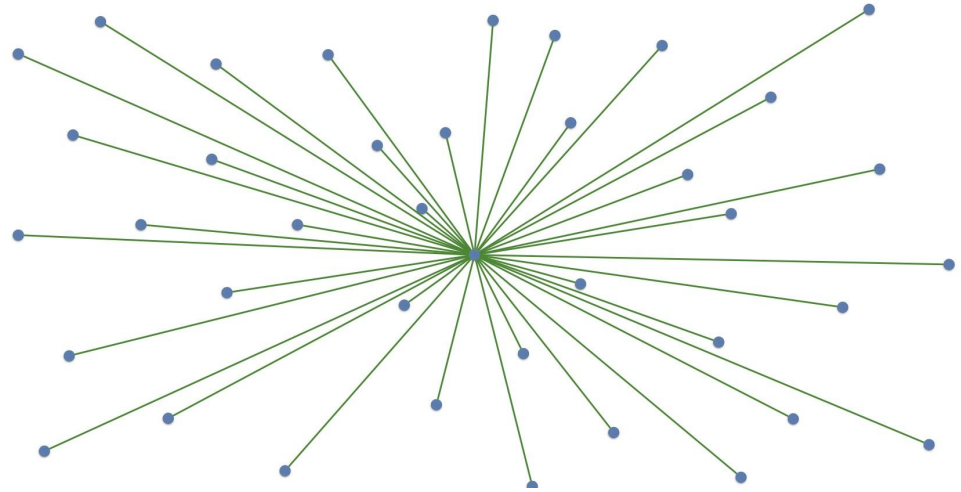
Centrality is the metric that indicates the most “important” node in the network.

Who is the most central?



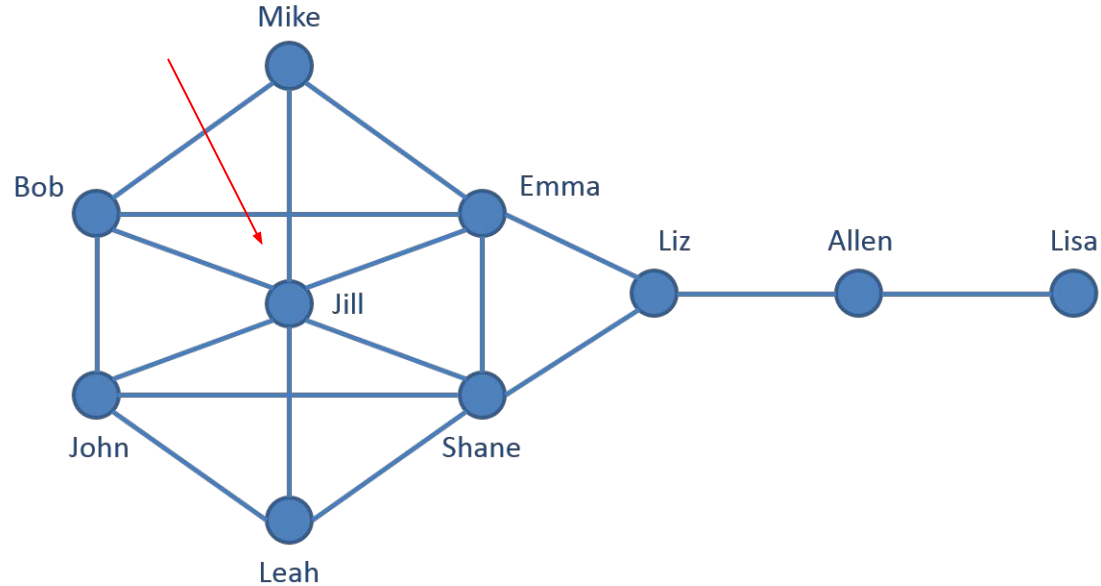
# Centrality as a degree

Higher node degree - higher node centrality



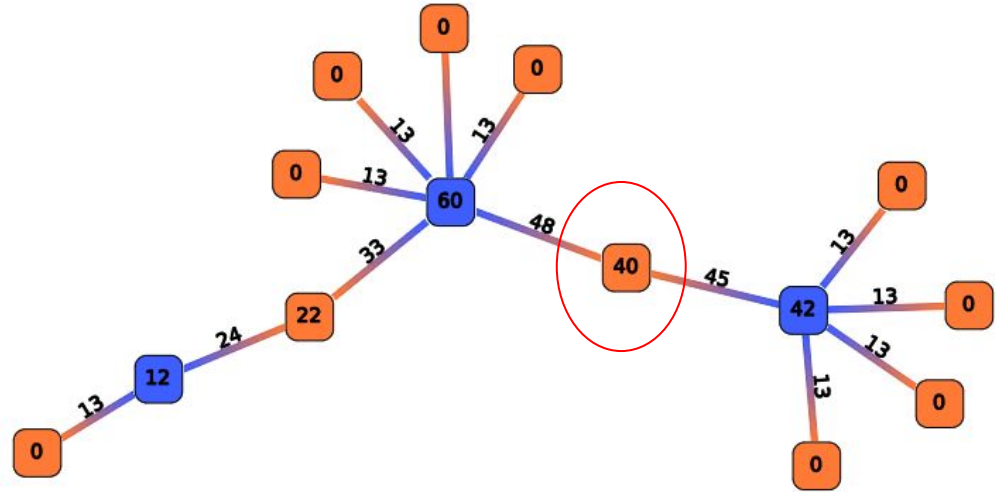
# Centrality as a degree

Higher node degree - higher node centrality



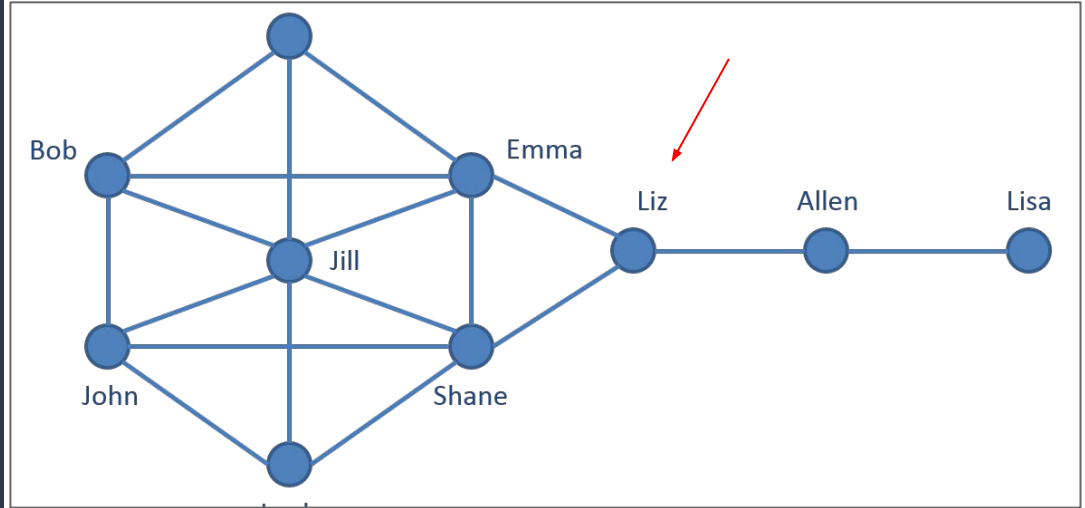
# Centrality as a betweenness

Shorter the way from one node  
to other - higher node  
centrality



# Centrality as a betweenness

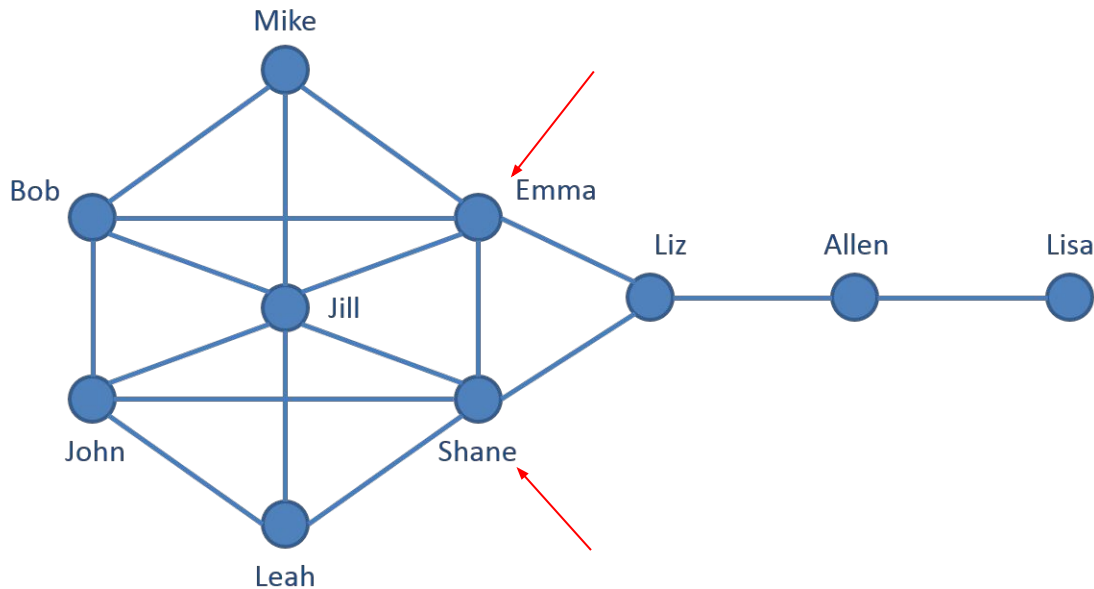
Shorter the way from one node  
to other - higher node  
centrality





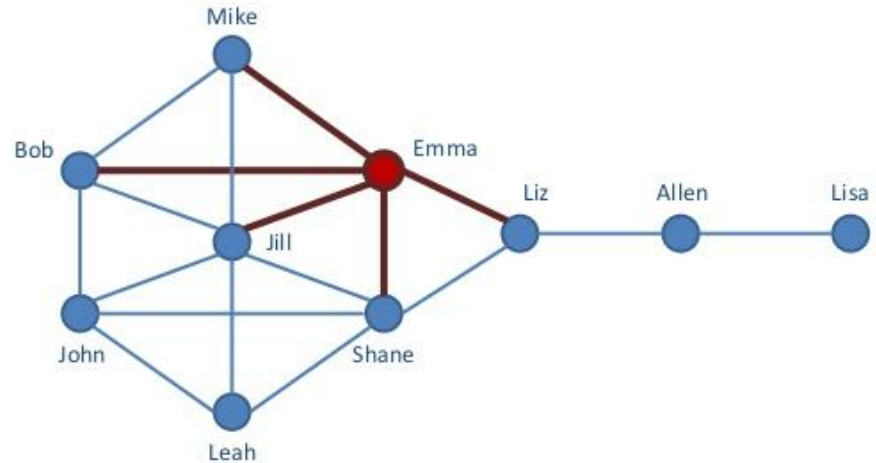
# Centrality as an eigenvector centrality

A node is important if it is linked to by other important nodes.



# Centrality as an eigenvector centrality

The more central a node is, the closer it is to all other nodes



# Communities

subsets of nodes among  
which there are relatively  
strong, direct, intense,  
frequent connection



# Community detection

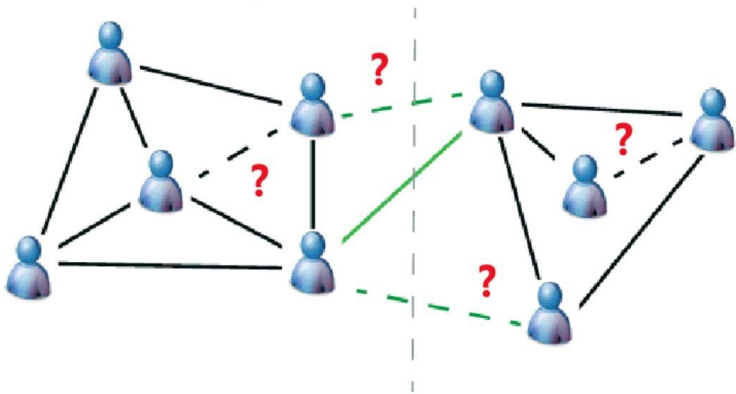
One of the most important tasks when studying networks is that of identifying *communities*.

Identifying network communities allows to:

- discover functionally related objects
- study interactions between groups
- infer missing node values
- predict unobserved connections

# Why community detection?

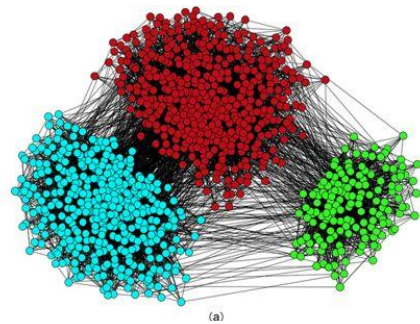
1. Analyzing communities helps better understand nodes



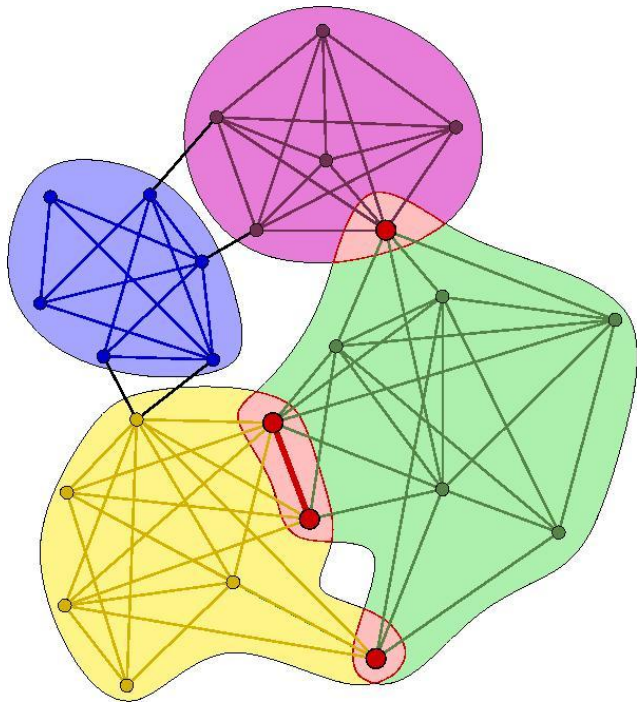
2. Detect missing connections or find new one



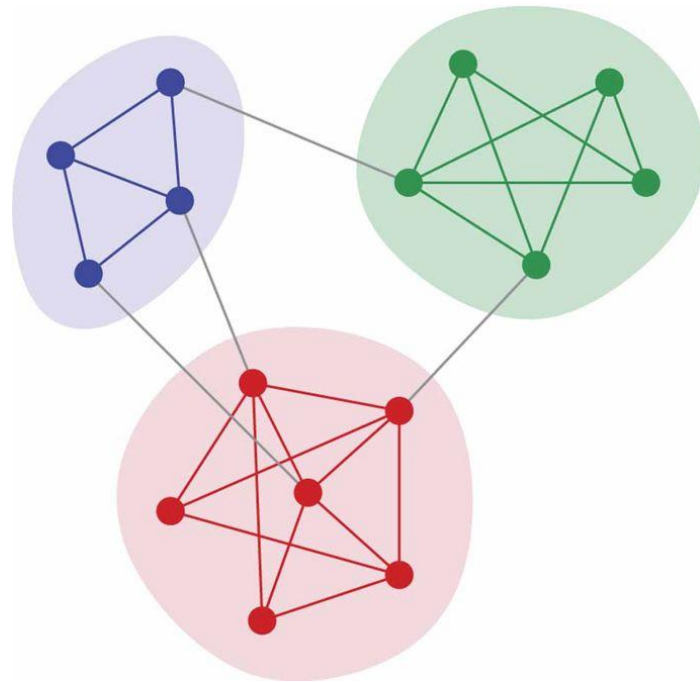
3. To provide a global view of network and how are situated communities



# Overlapping vs disjoint communities



**overlapping**



**disjoint**

# Network community detection

1. **Detect communities.**
2. **Find the most “important” nodes in each community and in the whole network.**





# Laboratory

## Madrid train bombing network

Jose A. Rodriguez of the University of Barcelona created a network of the individuals involved in the bombing of commuter trains in Madrid on March 11, 2004.

Rodriguez used press accounts in the two major Spanish daily newspapers (El Pais and El Mundo) to reconstruct the terrorist network. The names included were of those people suspected of having participated and their relatives.

# Laboratory

## Data challenges

1. Provide network analytics for the entire network, find the terrorists with the highest connection
2. Implement network community detection
3. Analyse two communities and try to find leaders there.

# Results

	Entire network	First community	Second community
Max degree	Jamal Zougam Mohamed Chaoui Imad Eddin Barakat	Basel Ghayoun	Imad Eddin Barakat Jamal Zougam
Highest betweenness	Jamal Zougam Mohamed Chaoui OM. Othman Abu Qutada	Basel Ghayoun	Imad Eddin Barakat
Highest centrality	Jamal Zougam Mohamed Chaoui	Mohamed Chaoui Basel Ghayoun Rafa Zuher	Jamal Zougam Imad Eddin Barakat

# Real facts

Clues from the thirteenth bomb allowed the police to arrest the first alleged perpetrators, three Moroccans - **Jamal Zougam**, **Mohamed Chaoui** and Mohamed Bekkaliand.

Three of the eight main suspects - Emilio Trashorras, **Jamal Zougam** and Othman el-Gnaoui - received sentences of nearly 40,000 years each.

Four other lead defendants - Youssef Belhadj, Hassan el Haski, Abdulmajid Bouchar and **Rafa Zouhier** - were acquitted of murder and trafficking in weapons.

According to Spanish court documents, Spanish telephone intercepts showed Mr. Zougam in contact with the accused leader of that cell, **Imad Eddin Barakat Yarkas**.

**Basel Ghayoun** is among the 18 people charged in the bombing case.