# Community detection in complex networks

Maguy Trefois    Jean-Charles Delvenne

January 2012

# Outline

Network : definition

# Outline

Network : definition

Why networks ?

# Outline

Network : definition

Why networks ?

Community detection

# Outline

Network : definition

Why networks ?

Community detection

  Some basic notions

# Outline

Network : definition

Why networks ?

Community detection

- Some basic notions
- Spectral partitioning

# Outline

Network : definition

Why networks ?

Community detection
- Some basic notions
- Spectral partitioning
- Modularity maximization

# Outline

Network : definition

Why networks ?

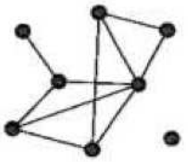
Community detection

- Some basic notions
- Spectral partitioning
- Modularity maximization

Detecting groups of vertices with the same behavior ...

A network (or graph) is a set of points joined by lines.

Example



A point is called a vertex.
A line is called an edge.

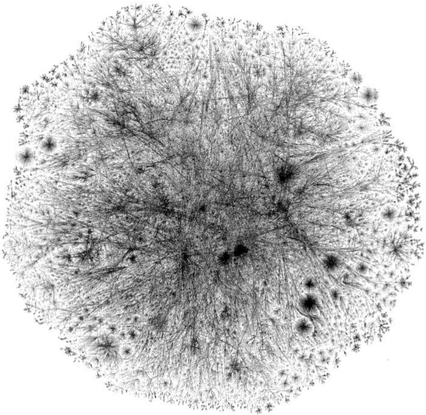
**Why networks ?**

Networks allow to model systems with interacting agents.

The structure of the network is fundamental for the understanding of the underlying system.

# Examples

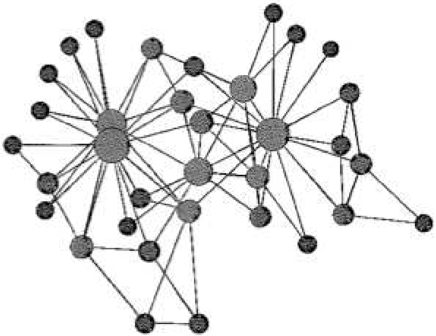
- *Technological networks*
  - *Internet*



*M.E.J. Newman, Networks : an introduction, Oxford University Press, Oxford UK, 2010, page 5.*

  - *Telephone networks*
  - *Transportation networks*

- Social networks
    - Friendship network between members of a club



M.E.J. Newman, *Networks : an introduction*, Oxford University Press, Oxford UK, 2010, page 6.

And so many others ...

**Community detection :**

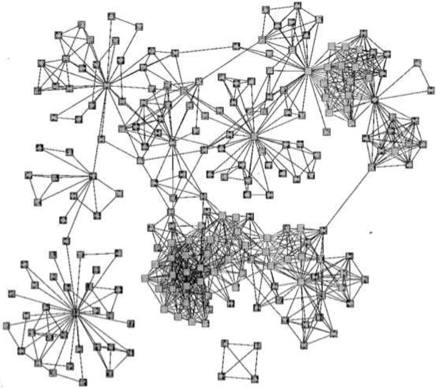
Partitioning the vertices of the network into groups, called communities, with many edges within the communities et few links between them.

**Utility :**

Revealing the structure and the organisation of the network.

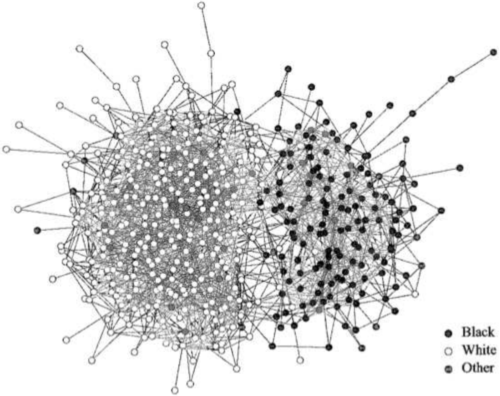
## Examples

- *Network of coauthorship in a university department*



*M.E.J. Newman, Networks : an introduction, Oxford University Press, Oxford UK, 2010, page 355.*

- Friendship network at a US high school




M.E.J. Newman, *Networks : an introduction*, Oxford University Press, Oxford UK, 2010, page 221.

Suppose that the network has $N$ vertices numbered from 1 to $N$.

The network can be mathematically represented thanks to its adjacency matrix $A \in \mathbb{R}^{N \times N}$ :

$$A_{ij} = \begin{cases} 1 & \text{if vertices } i \text{ and } j \text{ are connected} \\ 0 & \text{otherwise} \end{cases}$$

Suppose that the network has $N$ vertices numbered from 1 to $N$.

The network can be mathematically represented thanks to its adjacency matrix $A \in \mathbb{R}^{N \times N}$ :

$$A_{ij} = \begin{cases} 1 & \text{if vertices } i \text{ and } j \text{ are connected} \\ 0 & \text{otherwise} \end{cases}$$

The degree $k_i$ of vertex $i$ is the number of edges connected to it, that is :

$$k_i = \sum_{j=1}^{N} A_{ij}$$

The structure of the network can also be mathematically represented thanks to the Laplacian, strongly related to the adjacency matrix.

Denote $D$ the diagonal matrix with the degrees of the vertices on its diagonal, that is :

$$D = \begin{pmatrix} k_1 & 0 & 0 & \dots & 0 \\ 0 & k_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & k_N \end{pmatrix}$$

The Laplacian of the network is then the matrix : $L := D - A$, where $A$ is the adjacency matrix of the network.

**Goal** : partitioning the network into two communities $C_1$ and $C_2$.

The cut function counts the number of edges between communities $C_1$ and $C_2$, that is :

$$R = \sum_{i \in C_1, j \in C_2} A_{ij}.$$

**Goal** : partitioning the network into two communities $C_1$ and $C_2$.

The cut function counts the number of edges between communities $C_1$ and $C_2$, that is :

$$R = \sum_{i \in C_1, j \in C_2} A_{ij}.$$

$\Rightarrow$ We have to find a partition which minimizes $R$.

**Spectral partitioning method of Fiedler :**

Define the following vector $s \in \mathbb{R}^{N \times 1}$ :

$$s_i = \begin{cases} +1 & \text{if vertex } i \in C_1 \\ -1 & \text{if vertex } i \in C_2 \end{cases}$$

Then,

$$R = \frac{1}{4} s^T L s,$$

where $L$ is the Laplacian of the network to partition.

Denote $\lambda_1 \leq \lambda_2 \leq ... \leq \lambda_N$ the spectrum of $L$.

Let $v_2$ be the eigenvector related to eigenvalue $\lambda_2$.

The vector $s$ given by Fiedler's method is :

$$s_i = \begin{cases} +1 & \text{if } [v_2]_i > 0 \\ -1 & \text{if } [v_2]_i < 0 \end{cases}$$

If one entry $i$ of $v_2$ is zero, then $s_i$ can take equivalently value $+1$ or -1.

## Configuration model :

In this model of random graph :

- the number $N$ of vertices is fixed
- the degree sequence $[k_1, ..., k_N]$ is given.

As a consequence, the number $m$ of edges is fixed. Indeed,

$$m = \frac{1}{2} \sum_{i=1}^{N} k_i.$$

Let us place randomly the edges in the graph.

The expected number of edges between vertices $i$ and $j$ is :

$$\frac{k_i k_j}{2m}.$$

The modularity function compares the fraction of edges between two vertices in a same communtity and the expected fraction of edges (given the degree sequence), that is :

$$Q = \frac{1}{2m} \sum_{i,j} \left( A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j),$$

where :

- $m$ is the number of edges in the network
- $c_i$ is the community index of vertex $i$
- $\delta$ is the Kronecker symbol.

$\Rightarrow$ We have to find a partition which maximizes $Q$.

There exist several methods. The fastest one is the Louvain method.

Network : definition

Why networks ?

Community detection

Some basic notions

Spectral partitioning

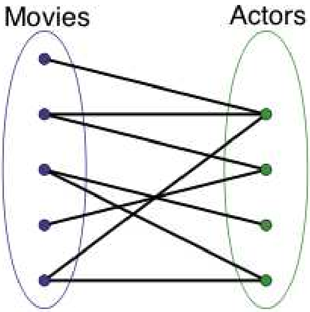
Modularity maximization

Detecting groups of vertices with the same behavior ...

We would like to detect groups of vertices with the same behavior or of the same type.

## Example

*A Movie-Actor network.*



**Actual work :**

1. How to define the behavior of a vertex ?
2. Developing algorithms to detect such groups of vertices.

Thank you for your attention !