Finding nodes with the same behavior in large networks

Maguy Trefois Jean-Charles Delvenne

March 27, 2012

Outline

Why networks ?

Outline

Why networks ?

Community detection

Outline

Why networks ?

Community detection

The stability method to detect communities

Outline

Why networks ?

Community detection

The stability method to detect communities

A generalization of the community detection : Weinan E et al method

Outline

Why networks ?

Community detection

The stability method to detect communities

A generalization of the community detection : Weinan E et al method

Future work

Why networks ?

Community detection

The stability method to detect communities

A generalization of the community detection : Weinan E et al method

Future work

A network is another name for "graph".

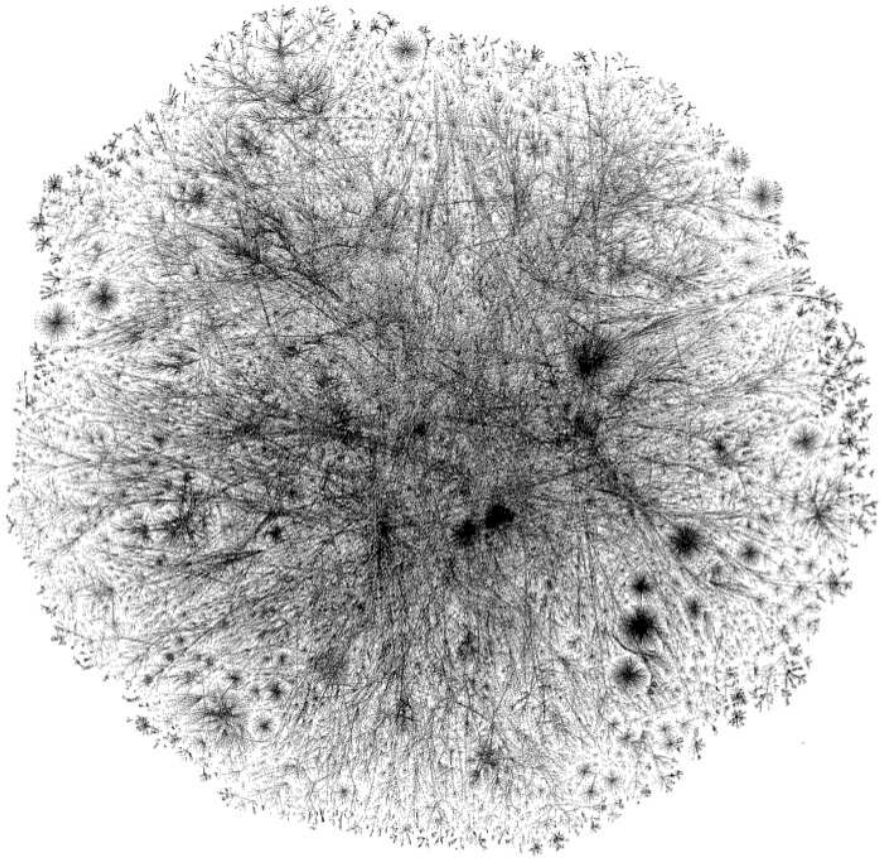
Why networks ?

Networks allow to model systems with interacting agents.

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

Example

- *A technological network : the Internet*



M.E.J. Newman, Networks : an introduction, Oxford University Press, Oxford UK, 2010, page 4.

Why networks ?

Community detection

The stability method to detect communities

A generalization of the community detection : Weinan E et al method

Future work

Community detection :

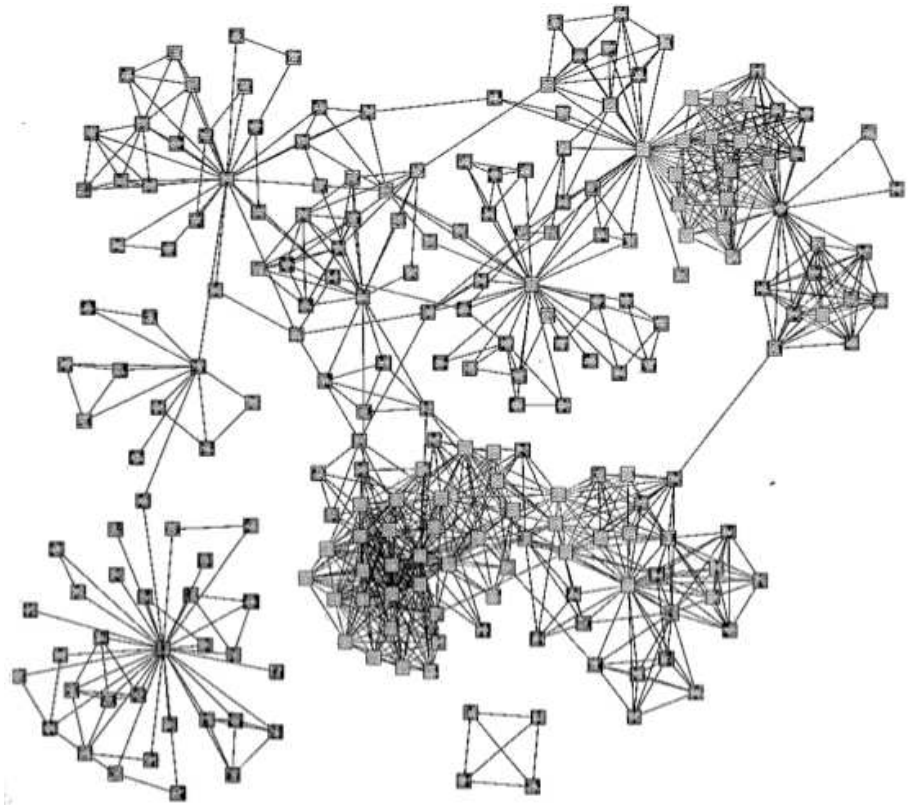
Partitioning the nodes of the network into groups, called communities, with many edges within the communities et few links between them.

Utility :

Revealing the structure and the organisation of the network.

Example

- *Network of coauthorship in a university department*



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

Why networks ?

Community detection

The stability method to detect communities

A generalization of the community detection : Weinan E et al method

Future work

Idea of the stability method :

After defining a dynamic Markov process on the network, we study the behavior of a random walker in order to detect communities.

A discrete-time Markov chain on the network :

Consider a network G.

The adjacency matrix A of the network is a squared matrix defined in this way :

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

The degree $d(i)$ of node i is the number of edges connected to i, that is :

$$d(i) = \sum_{j \in Nodes(G)} A_{ij}.$$

A discrete-time Markov chain on the network is defined in this way :

$$p_{ij} = \frac{A_{ij}}{d(i)},$$

that is, p_{ij} is the probability of going from node i to node j in one step.

A discrete-time Markov chain on the network is defined in this way :

$$p_{ij} = \frac{A_{ij}}{d(i)},$$

that is, p_{ij} is the probability of going from node i to node j in one step.

The stability method ...

Given a Markov time $t \in \mathbb{N}_0$, the spirit of the stability method is to detect a partition for which the probability that a random walker being in some community at time 0 is still in the same community at time t is maximum.

More precisely ...

Let us fix a Markov time $t \in \mathbb{N}_0$.

The stability of a partition $S = \{S_1, ..., S_n\}$ is :

$$r(t, S) := \sum_{i=1}^{n} P_i(t) - P_i(\infty),$$

where $P_i(t)$ is the probability that a random walker starting in community i is still in the same community after t steps.

More precisely ...

Let us fix a Markov time $t \in \mathbb{N}_0$.

The stability of a partition $S = \{S_1, ..., S_n\}$ is :

$$r(t, S) := \sum_{i=1}^{n} P_i(t) - P_i(\infty),$$

where $P_i(t)$ is the probability that a random walker starting in community i is still in the same community after t steps.

The stability method :

Given a Markov time t, we look for a partition with the largest stability.

Why networks ?

Community detection

The stability method to detect communities

A generalization of the community detection : Weinan E et al method

Future work

Idea of Weinan E et al method :

Partitioning the network such that two nodes in a same group have almost the same behavior.

Idea of Weinan E et al method :

Partitioning the network such that two nodes in a same group have almost the same behavior.

What does it exactly mean ?

Let G be a network and $S = \{S_1, ..., S_n\}$ be a partition of G.

Two nodes i, j in a same group are said to have the same behavior if for any group S_l of the partition, the fraction of the neighbors of node i in S_l is equal to the fraction of the neighbors of node j in S_l.

Let G be a network and $S = \{S_1, ..., S_n\}$ be a partition of G.

Two nodes i, j in a same group are said to have the same behavior if for any group S_l of the partition, the fraction of the neighbors of node i in S_l is equal to the fraction of the neighbors of node j in S_l.

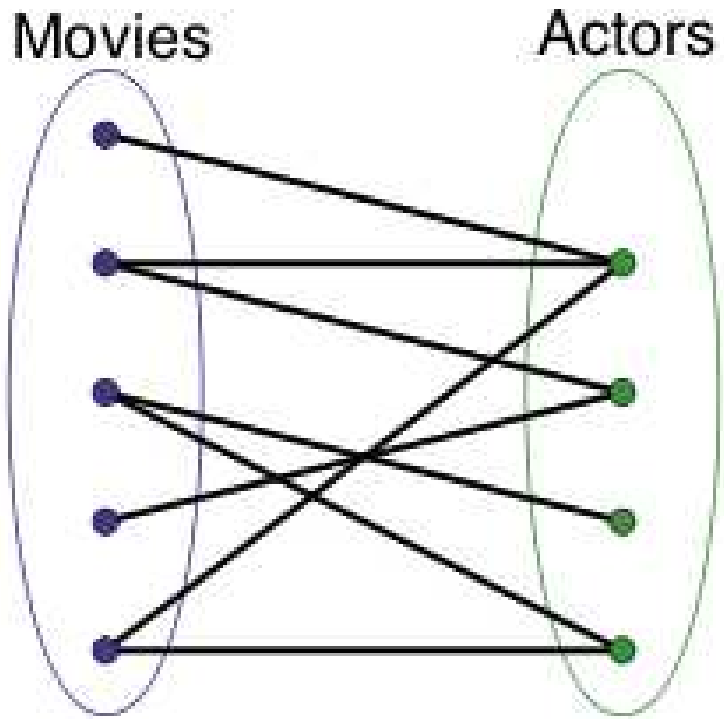
Goal :

Finding a partition of the network such that all the nodes in a same group have almost the same behavior.

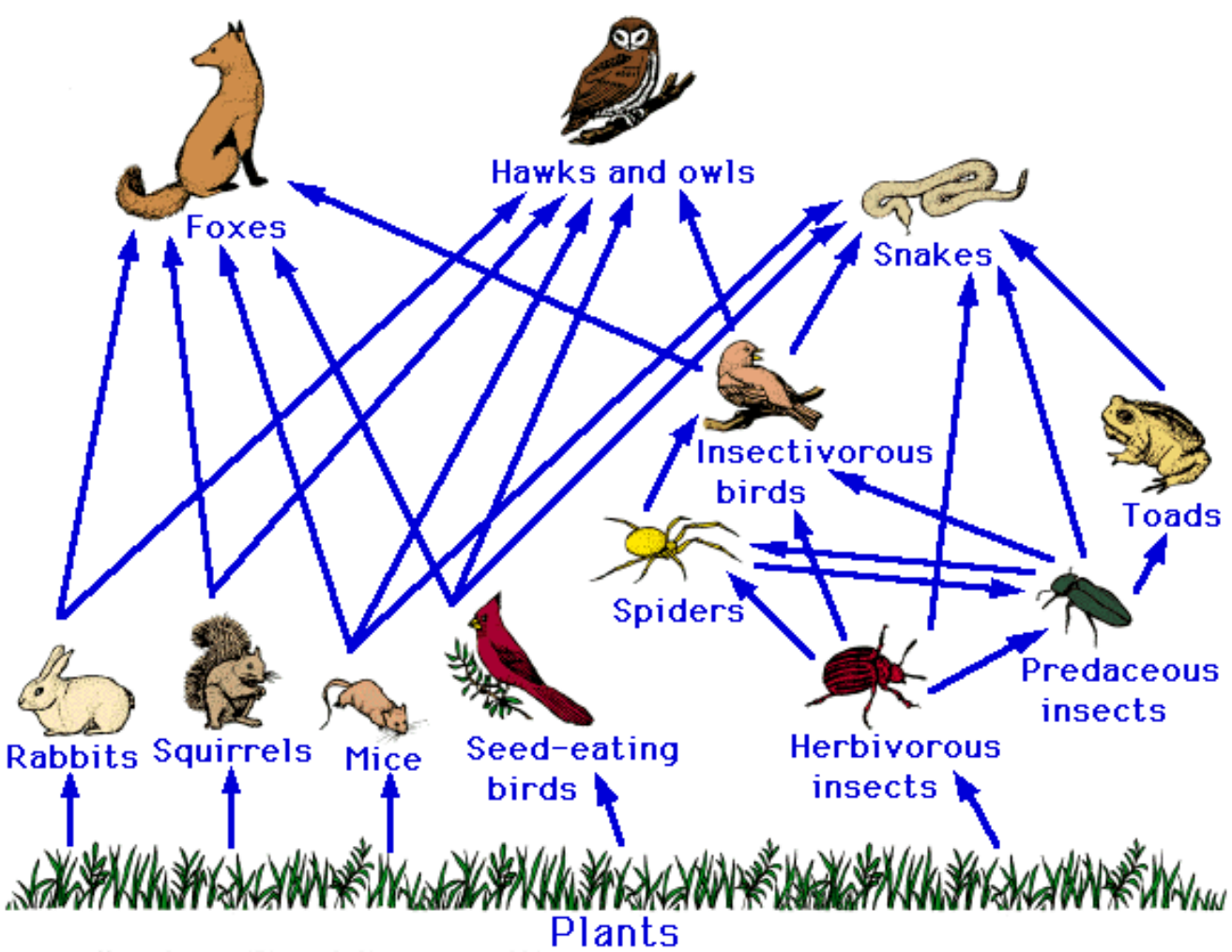
Motivation of this method :

Examples

- *A movie-actor network*



- A food web network



Source: Champaign Unit 4 Schools © 2008; http://www.champaignschools.org

In the community detection problem

Several communities are detected but only one behavior : two nodes in a same community mainly communicate with nodes in the same community and very little with the nodes in the other communities.

In Weinan E et al method :

Several different behaviors can be detected. For example, the nodes in group i only communicate with nodes in group j, but the nodes in group j could have 50% of their contacts in group k and 50% in group l.

So, Weinan E et al method is a generalization of the community detection problem.

Why networks ?

Community detection

The stability method to detect communities

A generalization of the community detection : Weinan E et al method

Future work

- Community detection in complex networks has been extensively studied since 2004.

- Community detection in complex networks has been extensively studied since 2004.
- A generalization was suggested by Weinan E et al (2007) : detection of groups of nodes with a similar behavior.

- Community detection in complex networks has been extensively studied since 2004.
- A generalization was suggested by Weinan E et al (2007) : detection of groups of nodes with a similar behavior.
- Disadvantage of Weinan et al method : they have to fix the number of groups of nodes to detect.

- Community detection in complex networks has been extensively studied since 2004.
- A generalization was suggested by Weinan E et al (2007) : detection of groups of nodes with a similar behavior.
- Disadvantage of Weinan et al method : they have to fix the number of groups of nodes to detect.
- Weinan et al method is similar (Markov chain) to the stability method.

- Community detection in complex networks has been extensively studied since 2004.
- A generalization was suggested by Weinan E et al (2007) : detection of groups of nodes with a similar behavior.
- Disadvantage of Weinan et al method : they have to fix the number of groups of nodes to detect.
- Weinan et al method is similar (Markov chain) to the stability method.
- Advantage of the stability method : the relevant number of communities is given by the method.

- Community detection in complex networks has been extensively studied since 2004.
- A generalization was suggested by Weinan E et al (2007) : detection of groups of nodes with a similar behavior.
- Disadvantage of Weinan et al method : they have to fix the number of groups of nodes to detect.
- Weinan et al method is similar (Markov chain) to the stability method.
- Advantage of the stability method : the relevant number of communities is given by the method.
- Future work : merging the advantages of the stability and Weinan E et al methods.

Thank you for your attention !