"Regards Croisés" sur l' Influenza aviaire
15-19 / 12 / 2008 • Montpellier • France
Rencontres scientifiques autour de deux projets de recherche :
Scientific meeting around two research projects:
GRIPAVI (CIRAD, MAEE) & ARDIGRIP (AIRD)

# Introduction to the Graph Theory in the context of Social Network Analysis and Disease Spread

**Raphaël Duboz**

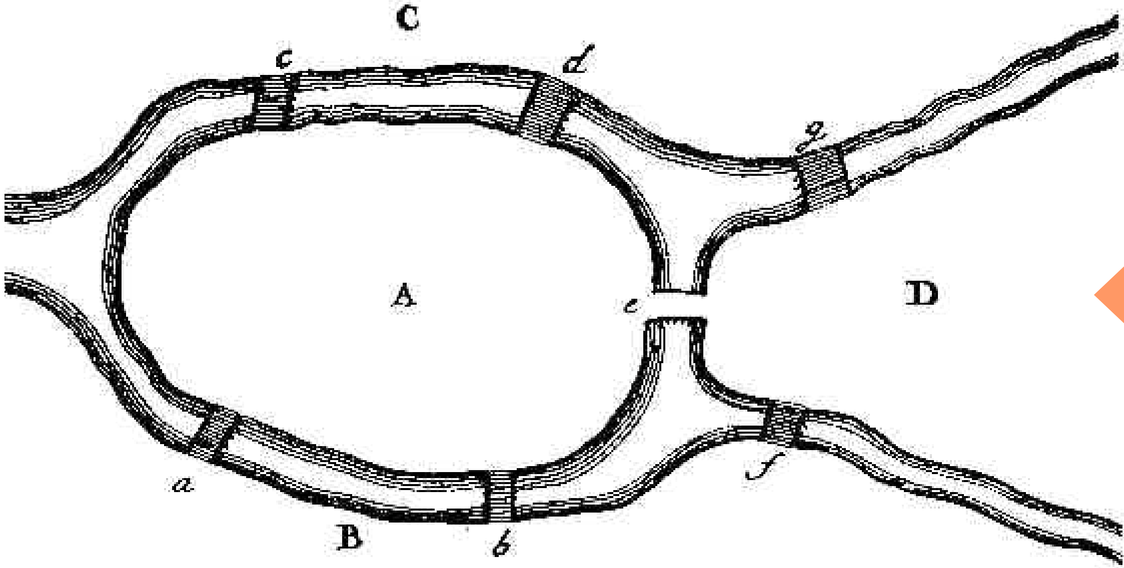
CIRAD – UPR Agirs



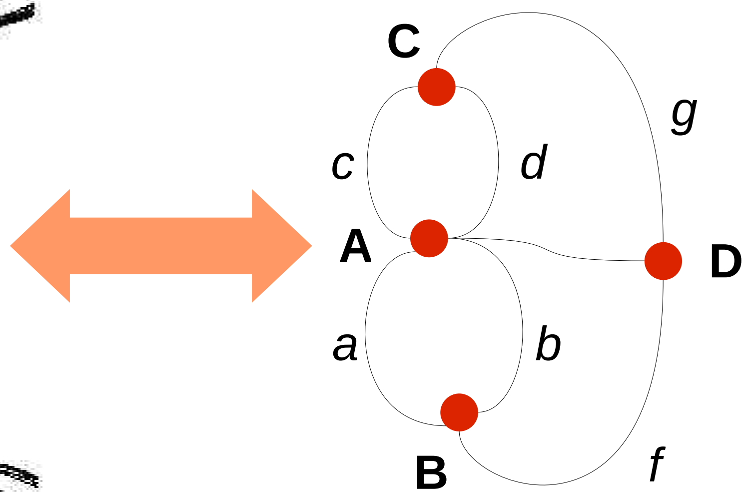
1736 with the communication of Euler (1707-1783)
The bridges of Königsberg

Starting from A, B, C or D, how to go back to the same point by crossing every bridges only once?

**The map**




**The graph**



# Outline

Definitions

Some indices (measures related to graph)

Graph and process models



# Definitions

- A **Graph** is in concern with relational data
  It is a finite set of entities called **vertices**
  and a multiset of relationships among those entities, called **edges**

  We focus on dyadic relationships

  **edges** consist of (**possibly ordered**) two element multisets on the **set of vertices**

- The **elements of an edge** are referred to as its **endpoints**

In the **ordered case**:
first element is **the tail (or sender)**,
the second is **the head (or receiver)**

An edge whose endpoints are identical is called a loop

- The combination of an edge set, **E**, with vertex set **V** is said to be a **graph**

# Definitions

- A graph ***G*** is a triplet $G=(V,E,\phi)$

***V*** the set of vertices ***E*** the set of edges $\phi : E \rightarrow V \times V$

$\phi$ is the incidence function:
in dyadic graphs, every edges is associated with a couple of vertices

- In the following, we denote a graph with $G=(V,E)$
- $|V|$ Is the graph order

- ***E*** is symmetric $(x,y) \in E \Leftrightarrow (y,x) \in E \quad | \quad x \in V \wedge y \in V$

If *a = (x,y)* is an edge, we say:

*x* and *y* are the endpoints of *a*
*a* is incident in *x* and *y*
*y* is a successor or a prior of *x* (directed graph)

# Definitions

- A graph defines a topological space which is the set ***V*** together with the set ***E***, whose elements are subsets of ***V***, such that:

$$\emptyset \in V$$
$$E \in V$$

If $P_j \in V \; \forall \; j \in J \;\; \rightarrow \;\; \cup_{j \in J} P_j \in V$

If $P \in V \wedge Q \in V \;\; \rightarrow \;\; P \cap Q \in V$

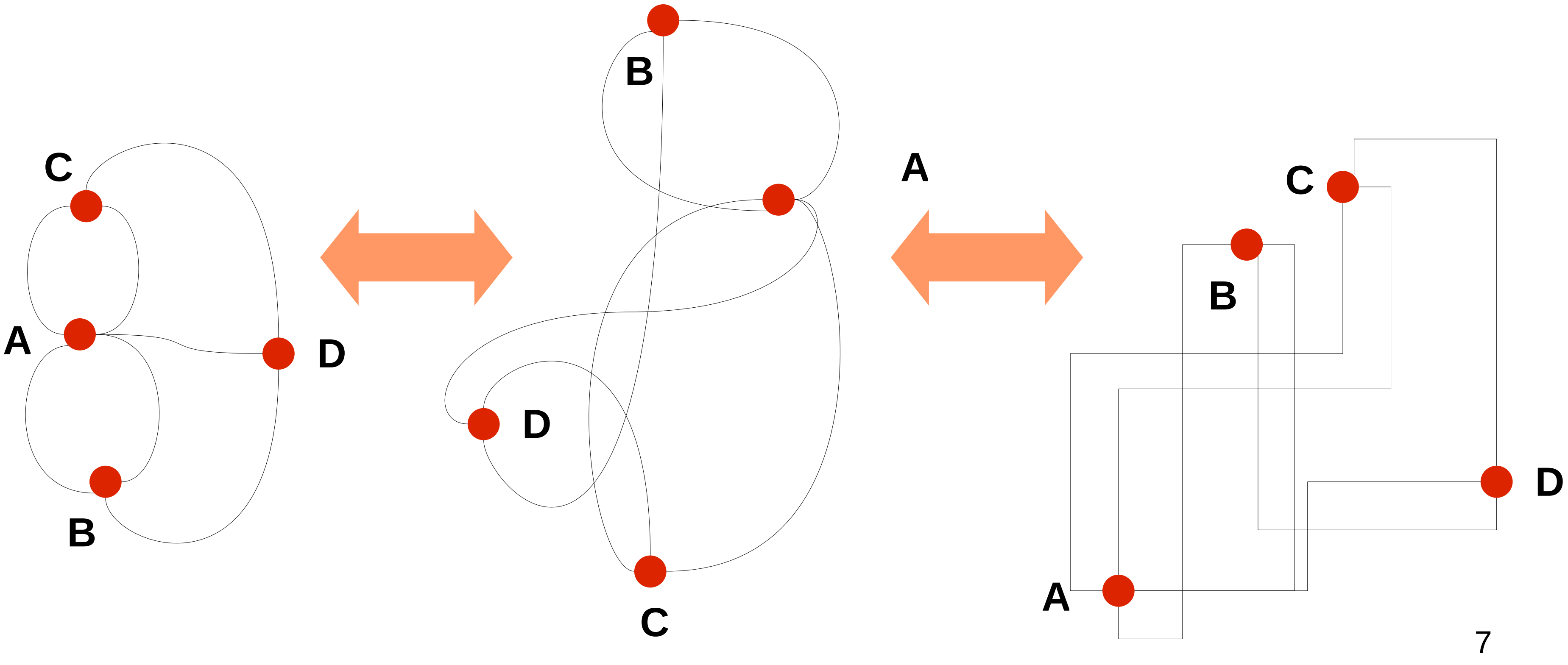
The set ***V*** is called a topology on ***E***

- The topology of a graph is then defined by $\phi : E \rightarrow V \times V$

$\phi$ instantiate the topology on ***E***

# Definitions

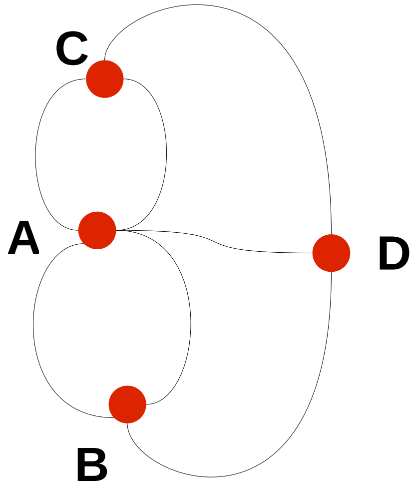
The topology of a graph is not defined by its graphical representation



# Definitions

## Undirected graph

- If the elements of E are unordered multisets, G is said to be an undirected graph
- the set of vertices tied (or adjacent) to vertex v is called the neighborhood of v (denoted $N_v$)



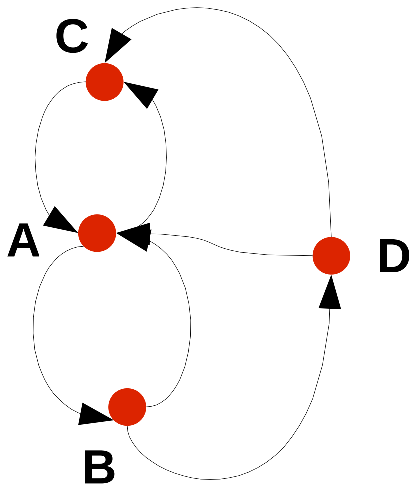

$$V = \begin{Bmatrix} A \\ B \\ C \\ D \end{Bmatrix} \qquad E_{ug} = \begin{Bmatrix} (A,B) \\ (A,B) \\ (A,C) \\ (A,C) \\ (A,D) \\ (B,D) \\ (C,D) \end{Bmatrix}$$

- A graph (undirected or otherwise) is simple if it has no loops and if there exists no edge having multiplicity greater than one

# Definitions

## Directed graph (or digraph)

- If edges are ordered multisets G is said to be a directed graph (or digraph)
- the in-neighborhood or $N^{-}(v)$: the set of vertices sending edges to v
  the out-neighborhood, or $N^{+}(v)$: the set of vertices receiving edges from v




$$V = \begin{Bmatrix} A \\ B \\ C \\ D \end{Bmatrix}$$

$$E_{dg} = \begin{Bmatrix} (A,B) \\ (A,C) \\ (B,D) \\ (B,A) \\ (C,A) \\ (D,A) \\ (D,C) \end{Bmatrix}$$



# Definitions

## Adjacency matrix

- Relational data can be stored in various formats

  UCINET, Pajeck, graphviz, ...

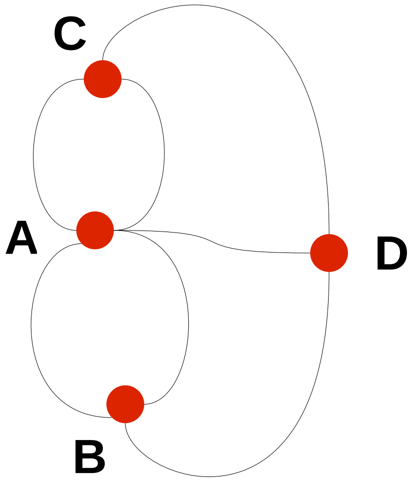
- For computations, graphs are generally formalized by an adjacency matrices

- Adjacency matrix definition

- a square matrix **M** where elements are defined such that $\mathbf{M}_{ij}$ is the value of the (i, j) edge in the corresponding graph

- by convention, $\mathbf{M}_{ij}$ is a dichotomous indicator variable ((0, 1) most of the time)

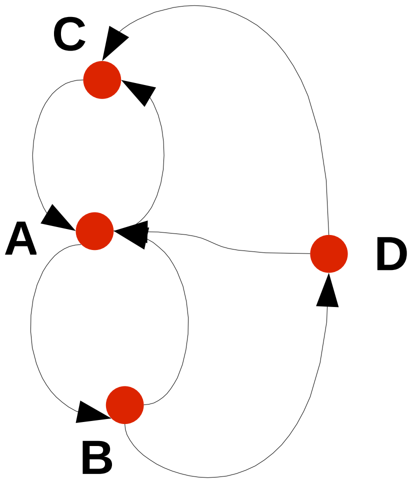
# Adjacency matrix (M)




$$V=\begin{Bmatrix}A\\B\\C\\D\end{Bmatrix} \qquad E_{ug}=\begin{Bmatrix}\{A,B\}\\\{A,B\}\\\{A,C\}\\\{A,C\}\\\{A,D\}\\\{B,D\}\\\{C,D\}\end{Bmatrix}$$

**M**

| | A | B | C | D |
|---|---|---|---|---|
| A | 0 | 2 | 2 | 1 |
| B | | 0 | 0 | 1 |
| C | | | 0 | 1 |
| D | | | | 0 |




$$V=\begin{Bmatrix}A\\B\\C\\D\end{Bmatrix} \qquad E_{dg}=\begin{Bmatrix}(A,B)\\(A,C)\\(B,D)\\(B,A)\\(C,A)\\(D,A)\\(D,C)\end{Bmatrix}$$

**M**

| | A | B | C | D |
|---|---|---|---|---|
| A | 0 | 1 | 1 | 0 |
| B | 1 | 0 | 0 | 1 |
| C | 1 | 0 | 0 | 0 |
| D | 1 | 0 | 1 | 0 |

# Definitions

## Graph or network?

- A graph is the abstract object
- A network is a concrete set of relational data

## The interpretation of a network

Which vertices (called nodes, individuals, locations,etc...)
play a major role in the propagation of the information, energy, disease, ...

→ **Indices at the vertex level**

How the global structure of the network influences the propagation...

→ **Indices at the graph level**

# Indices

## Vertex level indices

### Centrality indices

intuitively reflect some sense in which a vertex occupies a prominent or "central" position within a graph

### Degree

$N_i = \{v_j\} \mid e_{ij} \in E \wedge e_{ji} \in E$ the neighborhood of vertex i

*Undirected graph*

$$c_d(v, G) = |N_v|$$

*Directed graph*

$$c_{d^-}(v, G) = |N_v^-|$$
$$c_{d^+}(v, G) = |N_v^+|$$

Total or Freeman degree

$$c_{d^t}(v, G) = c_{d^-}(v, G) + c_{d^+}(v, G)$$

# Indices

## Vertex level indices

### Betweenness

Betweenness measures the extent to which a given vertex lies on non-redundant geodesics between third parties.

$$c_b(v,G)=\sum_{(v'v'')\subset V\setminus v}\frac{g'(v',v,v'',G)}{g(v',v'',G)}$$

$g(v',v'',G)$ the number of *(v, v')* geodesics in *G*

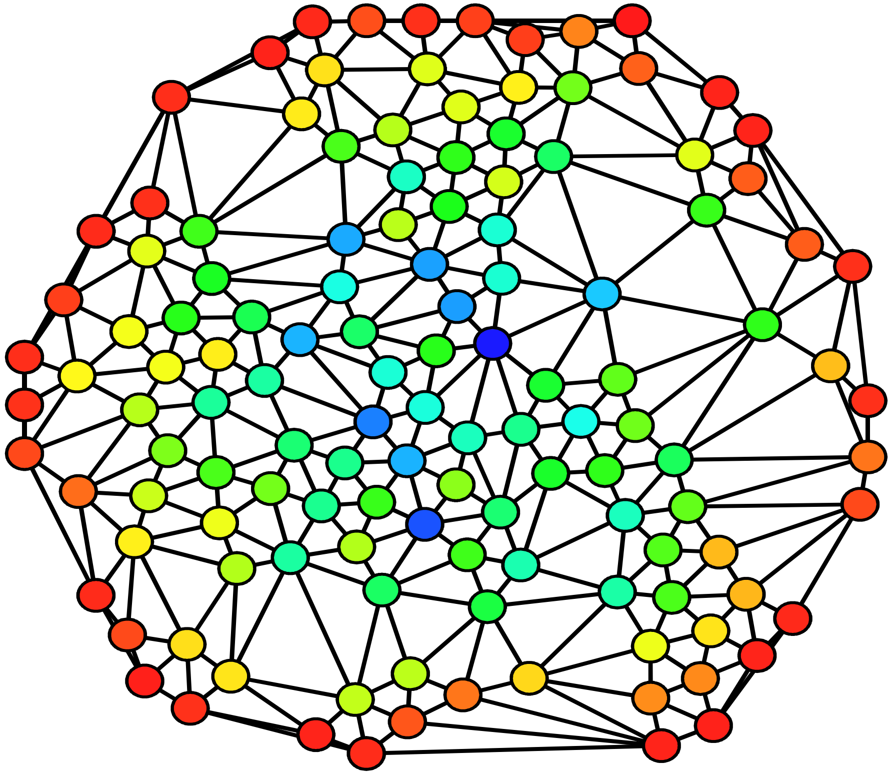
$g(v,v',v'',G)$ the number of *(v, v'')* geodesics in *G* containing *v'*

with $\frac{g'(v',v,v'',G)}{g(v',v'',G)}=0$ where $g(v',v'',G)=0$

# Indices

## Vertex level indices

### Betweenness



Node betweenness from red to dark blue



# Indices

## Vertex level indices

### Clustering Coefficient *C*

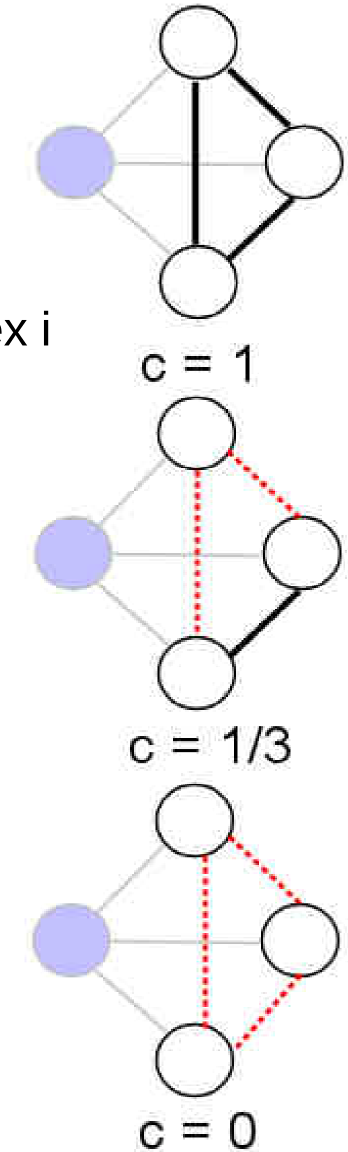
Let $N_i = \{v_j\} \mid e_{ij} \in E \wedge e_{ji} \in E$ be the neighborhood of vertex i

then the clustering coefficient for vertex *i* in a digraph *G* is:

$$C_i = \frac{|\{(j,k)\}|}{c_d(i,G)(c_d(i,G)-1)} \quad \mid \quad v_j, v_k \in N_i, e_{jk} \in E$$

and for a undirected graph *G*:

$$C_i = \frac{2|\{(j,k)\}|}{c_d(i,G)(c_d(i,G)-1)} \quad \mid \quad v_j, v_k \in N_i, e_{jk} \in E$$





# Indices

## Graph level indices

Graph density $$D(G)=\frac{|E|}{|V|\times(|V|-1)}$$

cardinality of E divided by the max cardinality of the graph

Centralization $$C(G)=\sum_{i=1}^{|V|}\left[\left(\max_{v\in V} c(v,G)\right)-c(v_i,G)\right]$$

Following Freeman (1979) and considering a centrality measure *c*

The difference between max and mean centrality scores, scaled by the number of vertices

Graph global clustering coefficient $$\overline{c}=\frac{1}{|V|}\sum_{i=1}^{|V|}C_i$$

the average of the clustering coefficient for each vertex

# Indices

## Graph level indices

**Graph efficiency** (Krackhardt (1994))

Let $G = G_1 \cup G_2 \ldots \cup G_n$ be a digraph with weak components $G_n$

The cardinalities of $G_n$ vertex sets are: $|V(G)| = N \quad |V(G_i)| = N_i \; \forall i \in \{1,..,n\}$

Then the Krackhardt efficiency of G is given by

$$\Psi = \frac{|E(G)| - \sum_{i=1}^{n} (N_i - 1)}{\sum_{i=1}^{n} \left[ N_i (N_i - 1) - (N_i - 1) \right]}$$

$\Psi$ can be interpreted as 1 minus the proportion of possible "extra" edges (above those needed to weakly connect the existing components) actually present in the graph

A graph which an efficiency of 1 has precisely as many edges as are needed to connect its components; as additional edges are added, efficiency gradually falls towards 0

# Connectivity

- Connectivity refers to a range of properties relating to the ability of one vertex to reach another via traversal of edges

- A graph is said "connected" iff all its vertices are reachable (in the weak form)

- The geodesic distance (path length) from vertices *i* to *j* is the number of edges to cross starting from *i* to reach *j* (it can exist several paths from *i* to *j*)



# Connectivity

Fararo and Sunshine's (1964) structure statistics

$G=(V,E)$ with *d(i,j)* the geodesic distance from vertices *i* to *j* in *G*

The series $S_{0,} S_{1,} \ldots , S_{|V|-1}$ are the the "structure statistics" of *G*

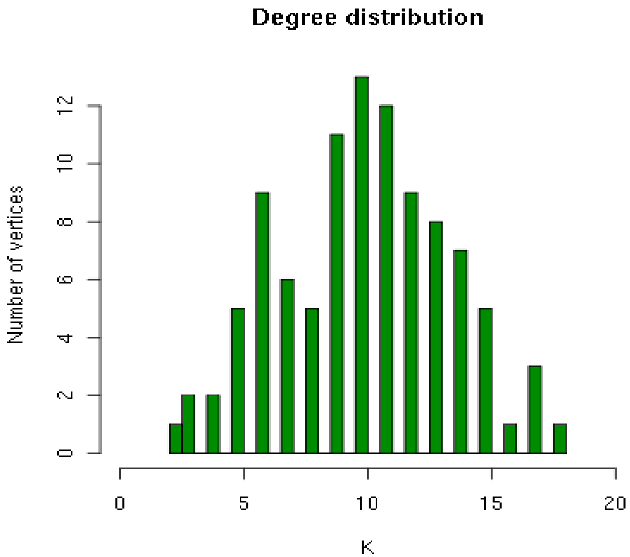
where $S_i = N^{-2} \sum_{j=1}^{|V|} \sum_{k=1}^{|V|} I(d(j,k) \leq i)$ and *I* the standard indicator function

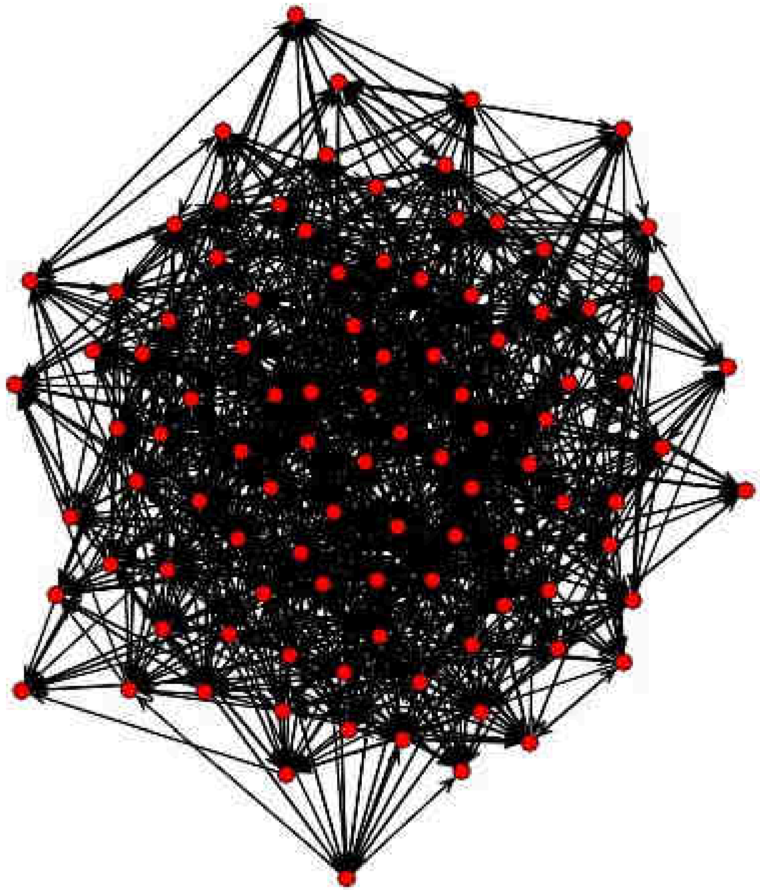
→ the expected fraction of *G* which lies within distance *i* of a randomly chosen vertex

A parsimonious description of global connectivity


# Graph classification

Uniform random graph:
with a poisson distribution of degrees

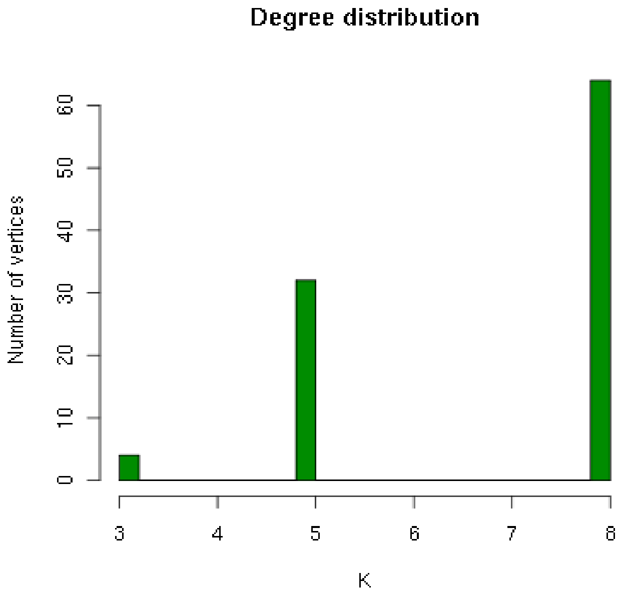


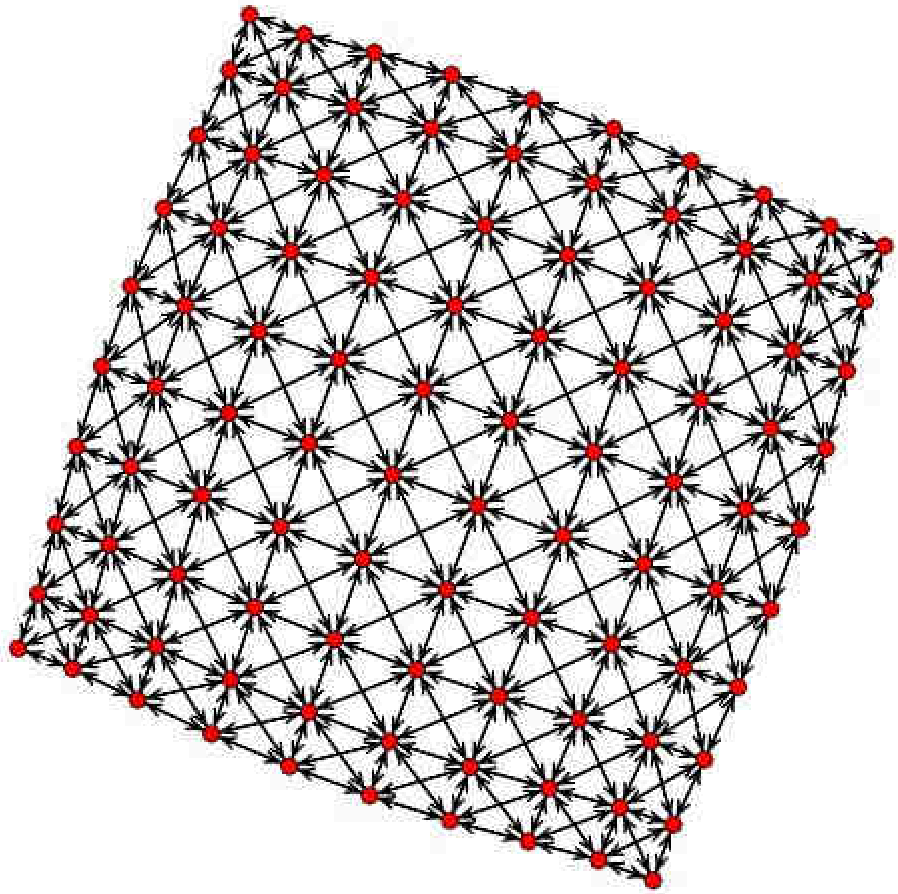





# Graph classification

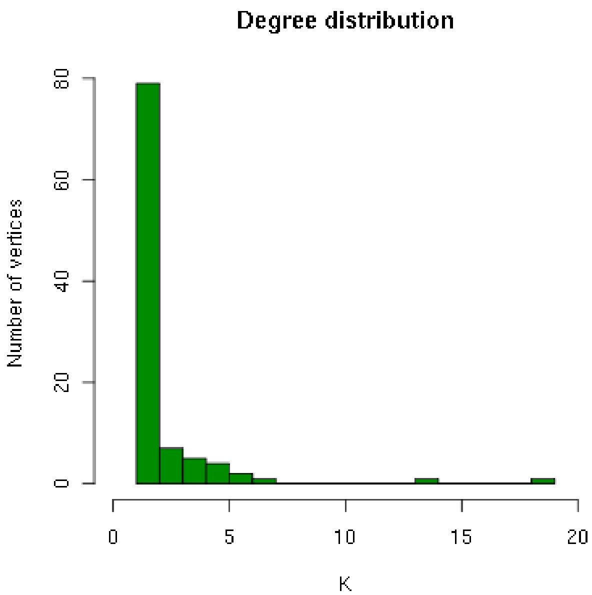
Regular lattice

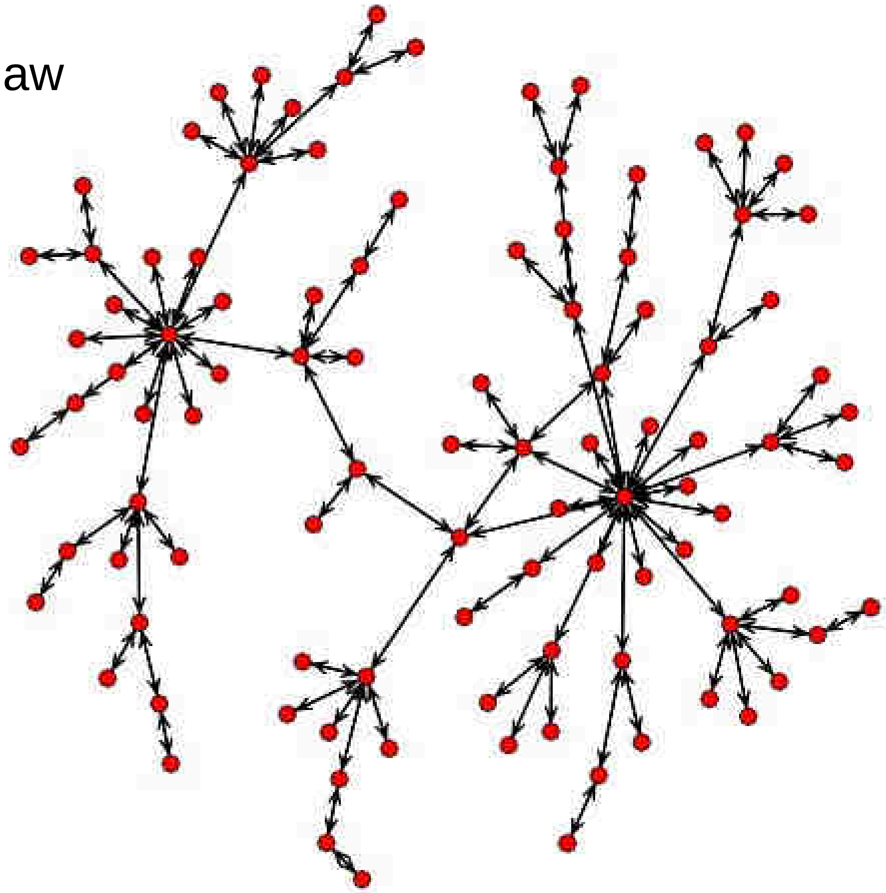


# Graph classification

Scale free
degree distribution following a power law





# Graph and process models

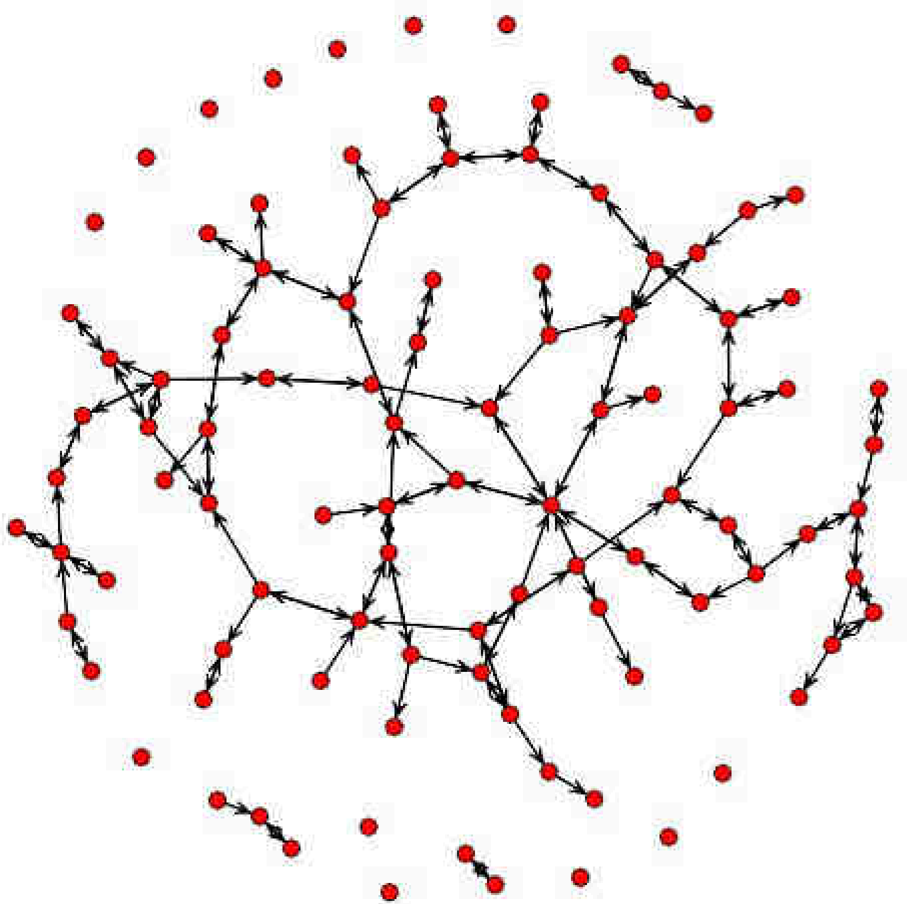
- It is possible to add state variables to vertices and edges
  - Diffusion models, percolation models
    - In the SNA field: Rumor diffusion models
- In the following, we consider a simple epidemiological model
  - 3 states for each vertex: S, I ,R
  - $|V| = 100$
  - Probability for a vertex to infect a neighbor is *p=0.1*
  - Infectiousness period = 10 time steps

"Regards Croisés" sur l' **Influenza aviaire**

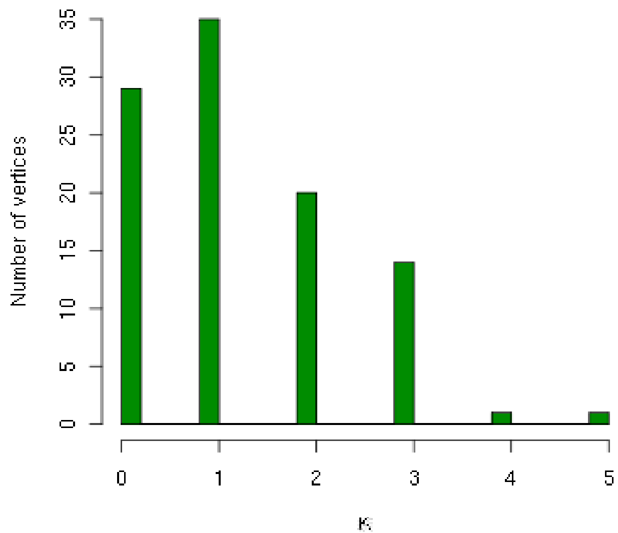
*15-19 / 12 / 2008 • Montpellier • France*

Rencontres scientifiques autour de deux projets de recherche :
Scientific meeting around two research projects:
**GRIPAVI (CIRAD, MAEE) & ARDIGRIP (AIRD)**

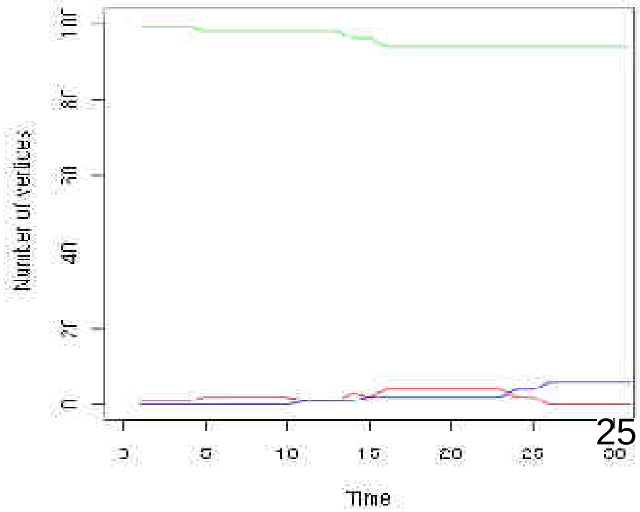
**Random graph (p = 0.01)**



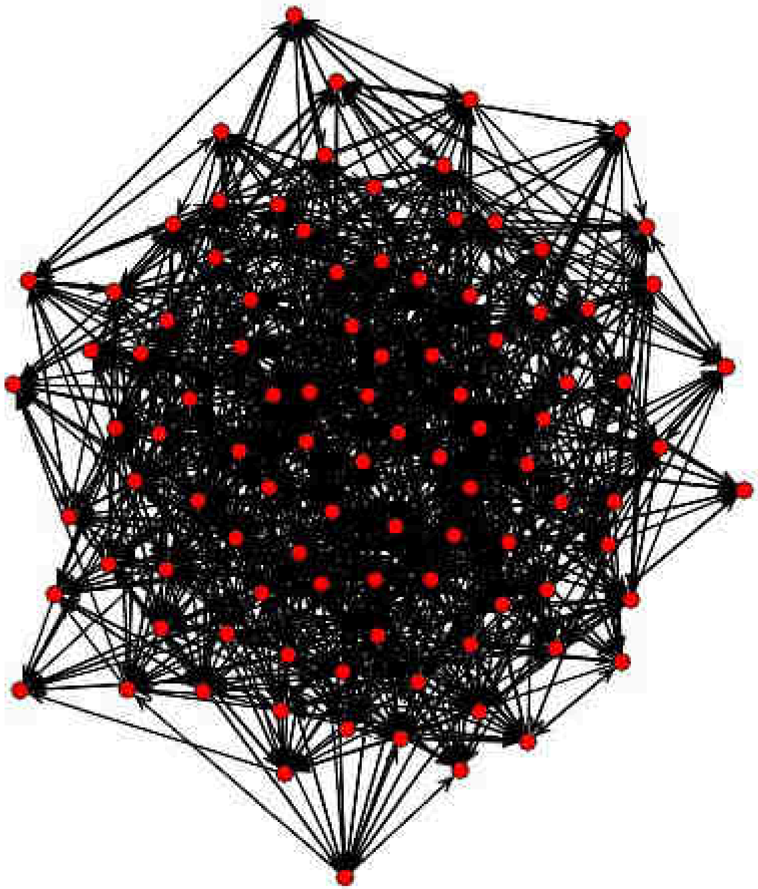
**Degree distribution**



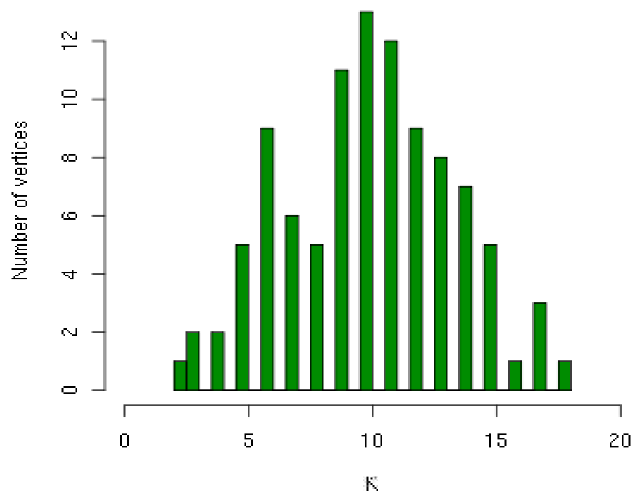

**Epidemic dynamic**



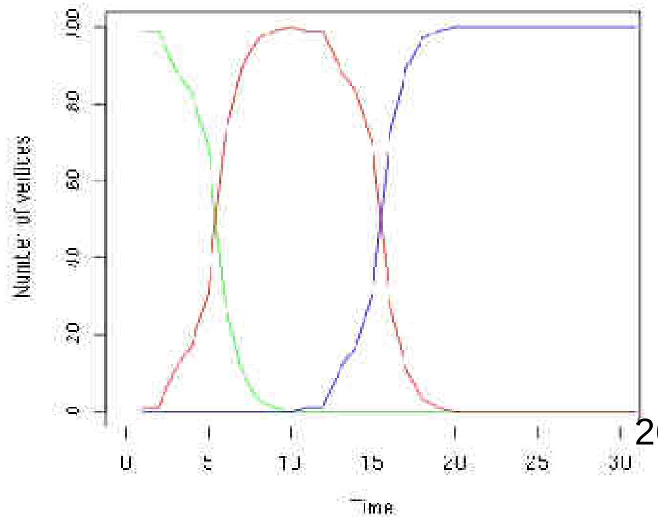
Random graph (p = 0.1)



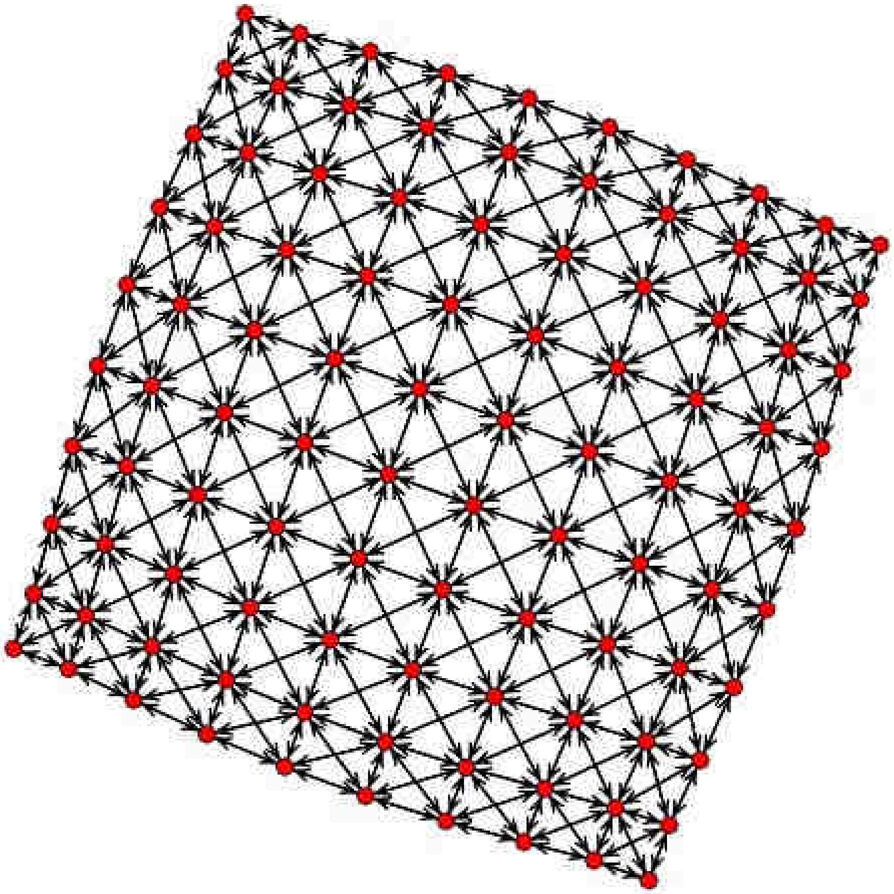
Degree distribution



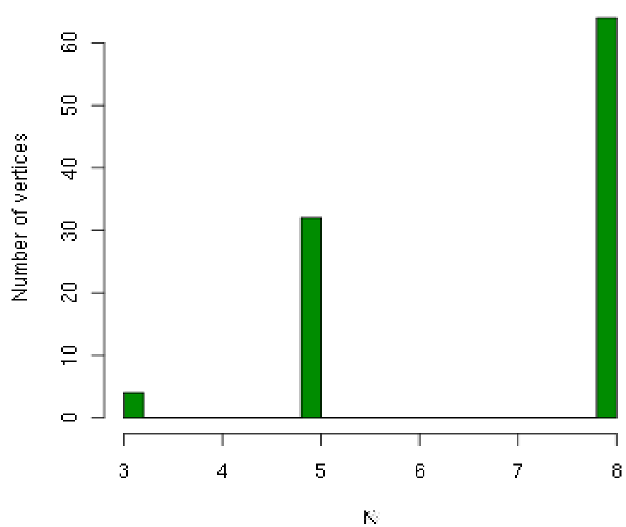

Epidemic dynamic



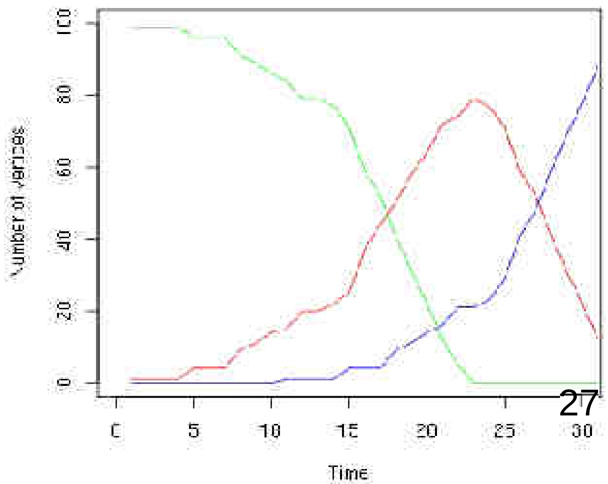
Regular lattice



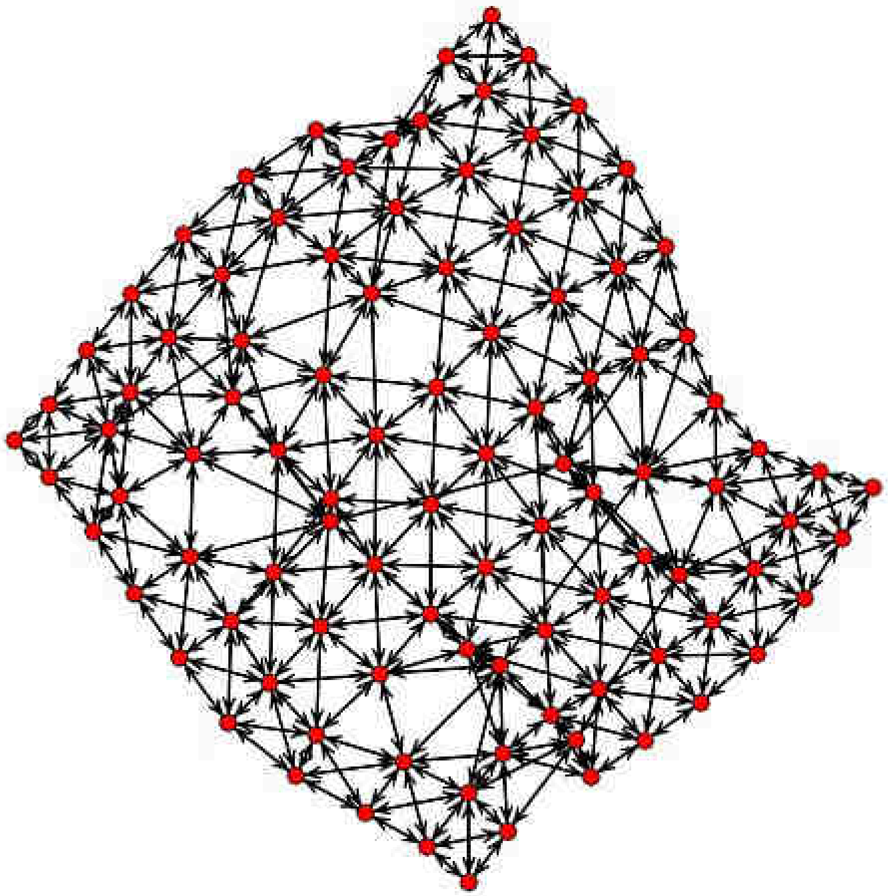
Degree distribution



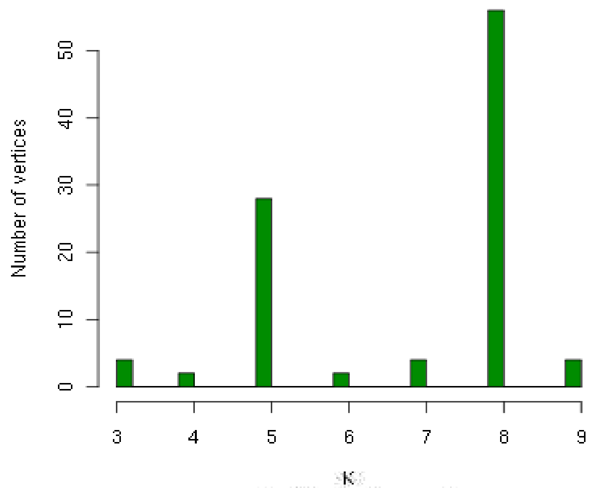

Epidemic dynamic



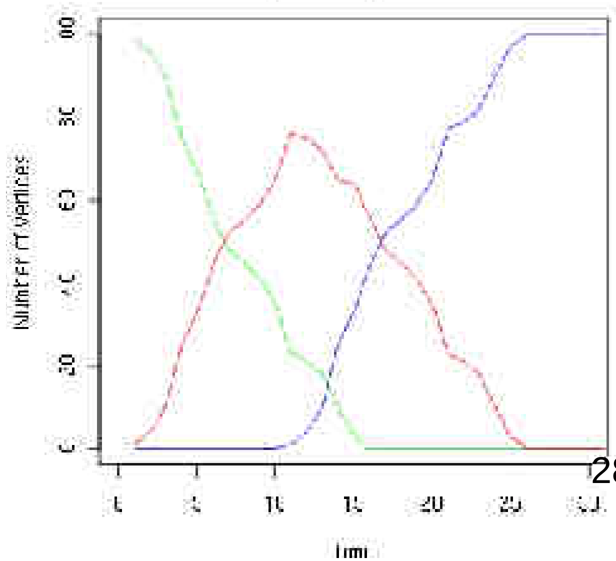
**Rewired lattice (p = 0.01)**



**Degree distribution**

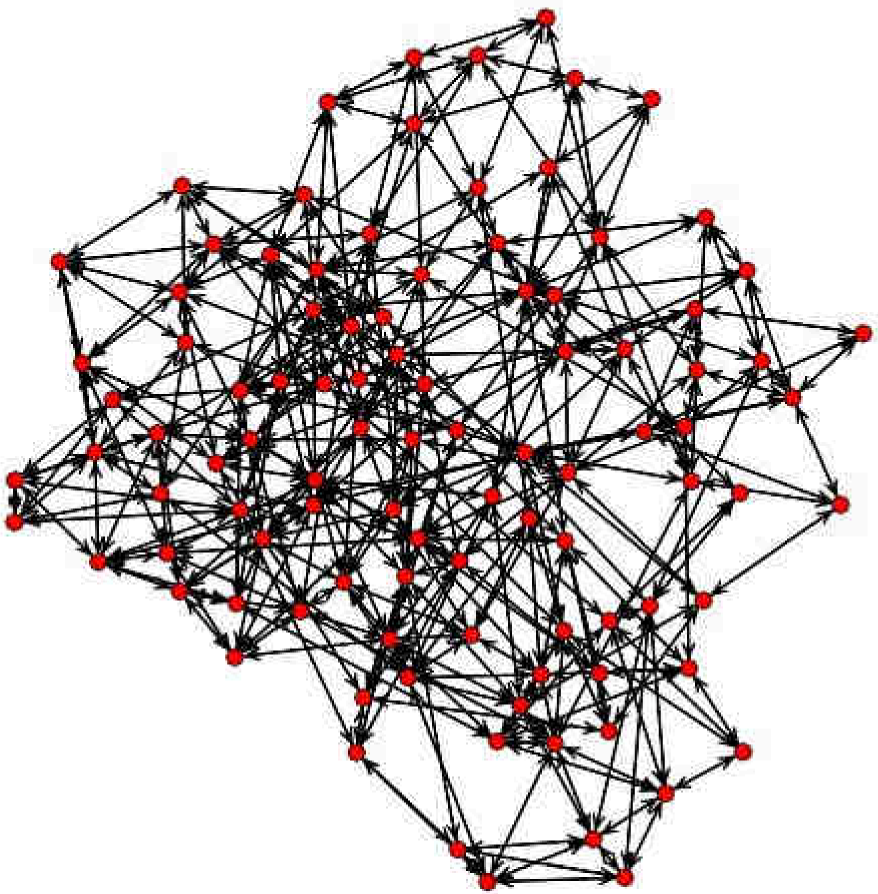


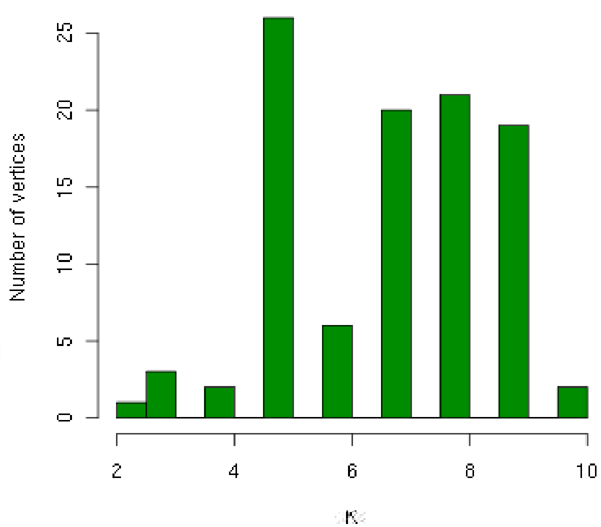

**Epidemic dynamic**

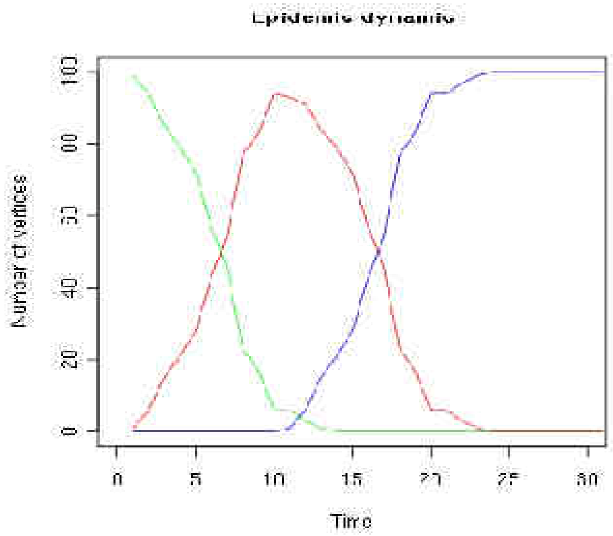




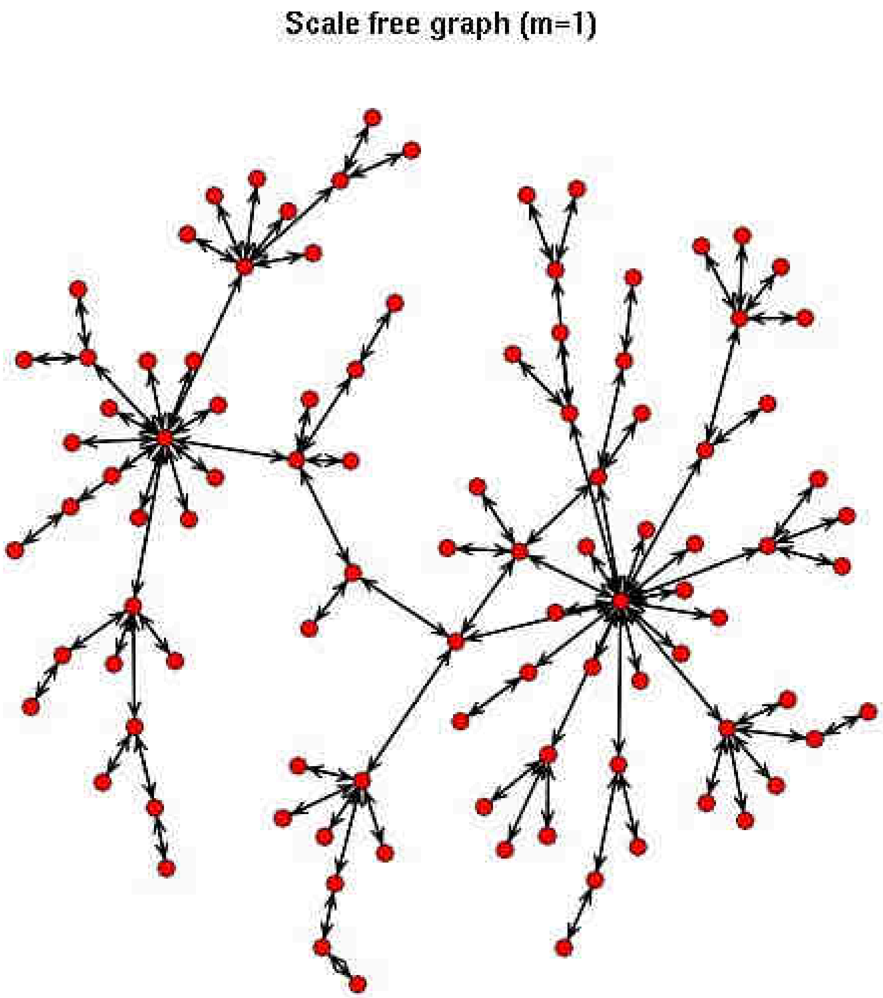
Rewired lattice (p = 0.1)



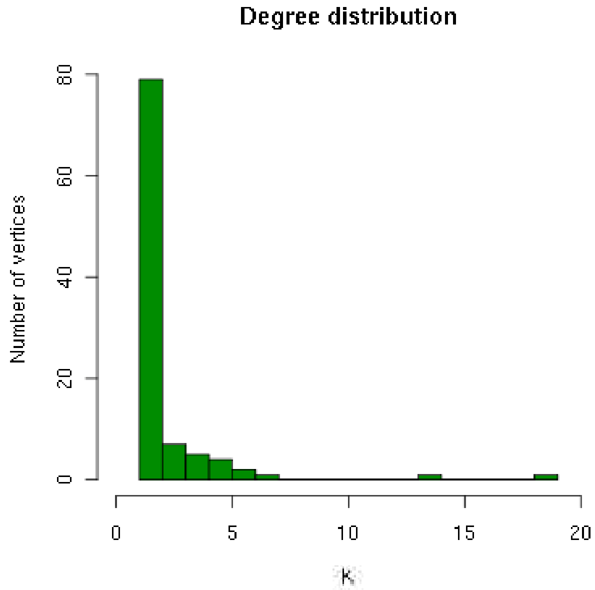


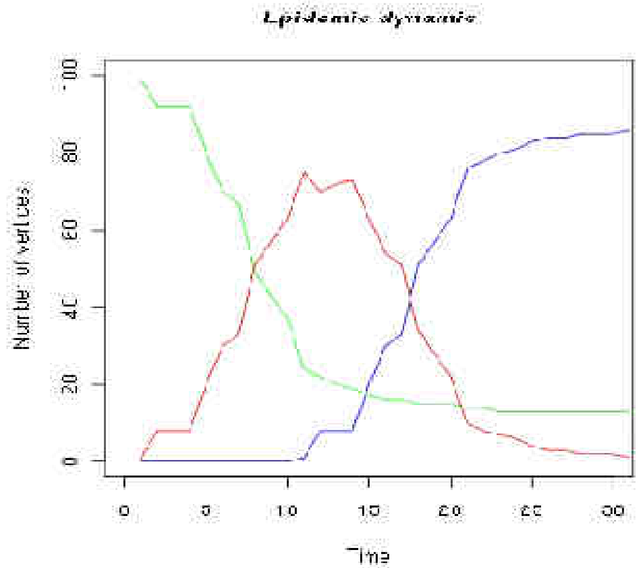

Small world effect



Scale free graph (m=1)

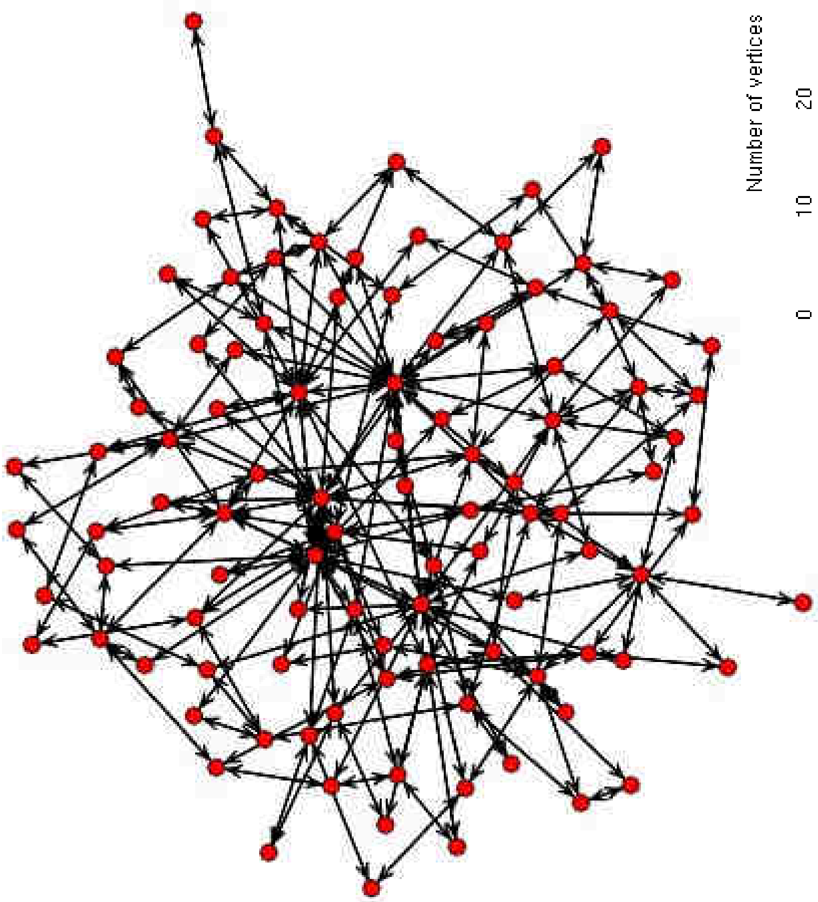



Degree distribution
Number of vertices
0
20
40
60
80
0
5
10
15
20
k

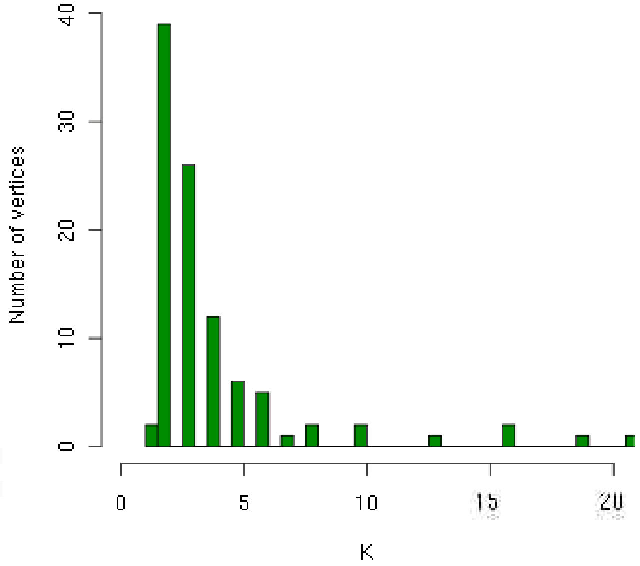



Epidemic dynamic
Number of vertices
Time

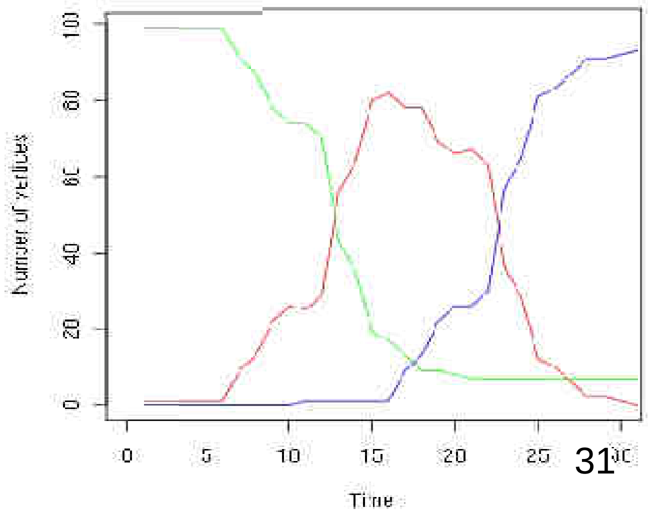
Scale free graph (m = 2)



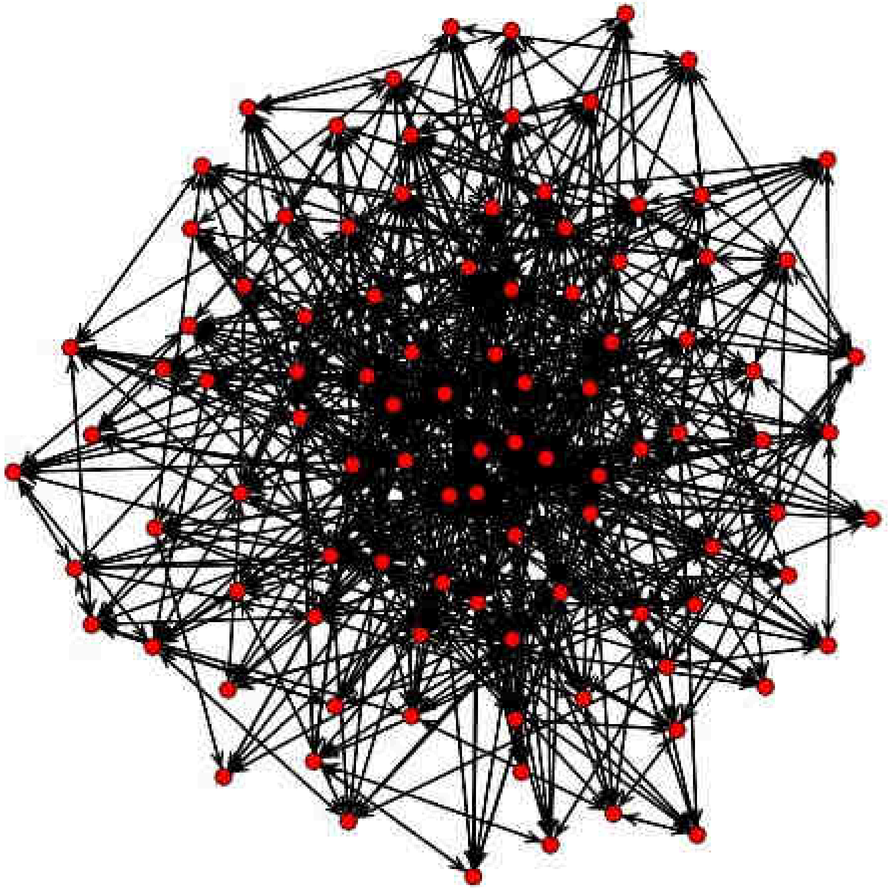
Degree distribution



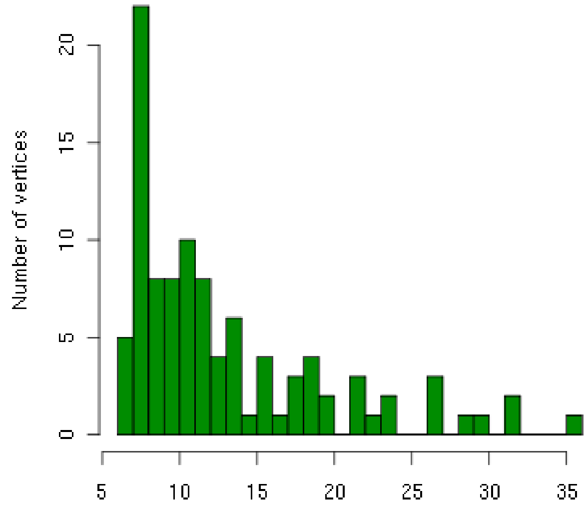

Epidemic dynamic



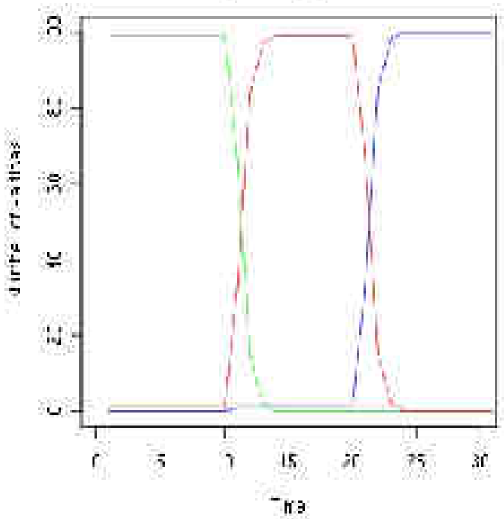
Scale free graph (m = 8)



Degree distribution




Epidemic dynamic



# What have we learned?

- Graph theory is mainly a set of definition related to relational data
- Two families of indices
- Graph structure has a strong influence on the propagation



# Other topics to consider for SNA and disease spread (not seen here)

- Numerous indices
- How to combine indices for analysis
- Subgraph statistics
  - Cluster analysis
  - Subgraph comparison
- Position and role analysis
  - Blockmodelling
  - Prestige measures
- Graph inference
  - Working with sparse or partially known graphs



# References

A. Barabasi and R. Albert. **Emergence of Scaling in Random Networks**. *Science* 286, 509-512, 1999

M.D.F. Shirley and S.P. Rushton. **The impact of network topology on disease spread**
Ecological Complexity, 2, 287-299, 2005

M. Pautasso and M.J. Jeger. **Epidemic threshold and network structure: The interplay of probability of transmission and of persistence in small-size directed networks.** Ecological Complexity, 5, 1-8, 2008

C.T. Butts. **Social Network Analysis with sna**. Journal of Statistical Software, 24, 6, 1-51, 2008
*free at http//www.jstatsoft.org/*