Graphs to analyze medieval social networks

toward advanced applied graph theory tools

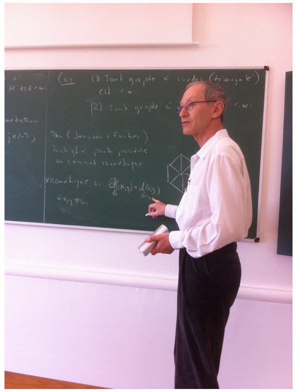
Bertrand Jouve (CNRS, Toulouse/Lyon)
Etienne Fieux (IMT, Toulouse)
Florent Hautefeuille (TRACES, Toulouse)
Romain Boulet (Lyon)
Ted Gragson (Athens University)

1. Introduction

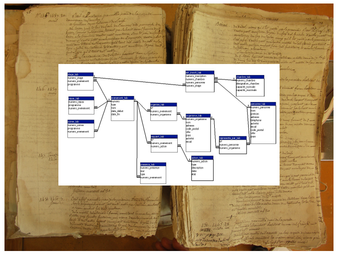
✓ General context

- ❑ Across a large corpus, better understand the organization of the peasant world of XIIIe-XVIe centuries by identifying and analyzing networks of social relations and their dynamics.
- ❑ Illustrate how graph theory may help in the study of real world networks providing a complementary point of view (to historical research or statistical approach)

★ *Historical research*



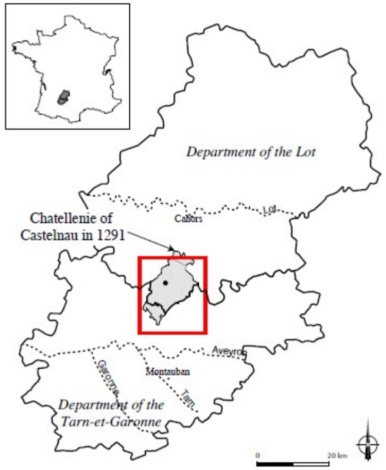
★ *Mathematical models*

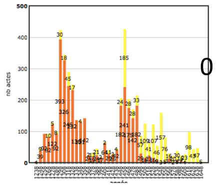


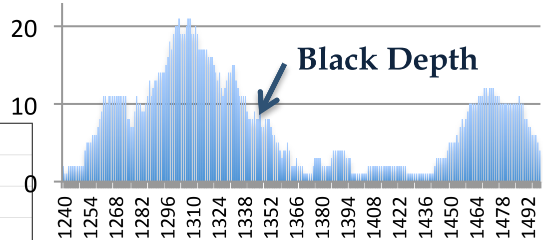
★ *From large corpus to databases*

✓ The database

- ❑ A territory of about 150 km^2;
- ❑ A rich corpus of 3356 legal document primarily agrarian contracts;







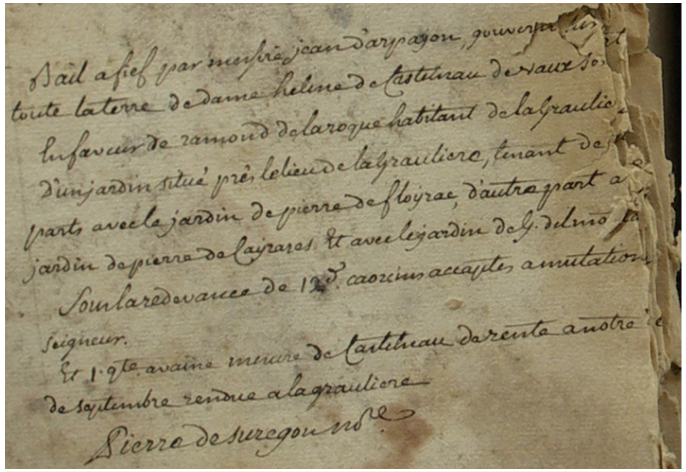
- ❑ Two periods: 1240-1340 and 1440-1520
- ❑ 4191 and 2895 mentions of individuals

- ❑ Different roles of the individuals in any one transaction :

  1) **initiator**: they requested the transaction recorded in the document,

  2) **participant**: they are party to a transaction initiated by another individual

- ❑ Our networks don't have millions of nodes but enough to be studied with tools of network science



Bail à fief par messire **Jean d'Arpajon**, gouverneur de toute la terre de Dame **Hélène de Castelnau de Vaux**, en faveur de **Ramond de Laroque**, habitant de la Graulière, d'un jardin situé près le lieu de la Graulière, tenant de deux parts avec le jardin de **Pierre de Floyrac**, d'autre part à jardin de **Pierre de Cayrases** et avec le jardin de **G del Moli**. Sous la redevance de 12 d cahorcien d'acapte à mutation de seigneur et une quarte avoine mesure de Castelnau de rente à notre dame de septembre rendue à la Graulière. **Pierre de Suregone** notaire.

✓ construction of the social networks

- The nodes are the individuals mentioned in the corpus
- Two nodes are linked by an edge if the corresponding individuals are initiators in at least one same transaction (contractual relationship)

➢ Such a network has a high number of small connected components

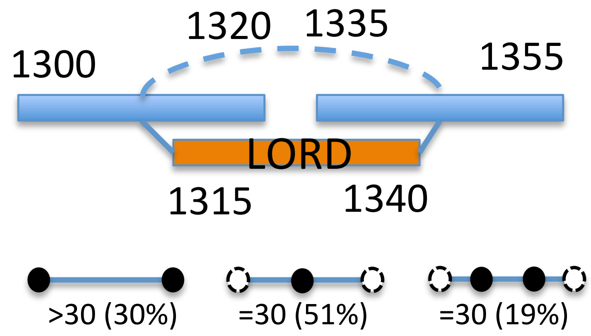
➢ The relationships are only contractual

- We also consider “mimetic relationships”:
  - Two individuals who are geographically proximate to each other had prior relations between themselves (-> the “owners” of neighboring land parcels)
  - Two individuals who depend on the same lord or notary had a social relationship (up to 20 co-lords within a same parish)

➢ The “mimetic” network is a sort of connected substrate with which the contract network develops.

Two nodes are linked by an edge if the corresponding individuals are linked by at least one contractual or mimetic relationship

✓ Three difficulties



- ❑ "the ripple effect": The nodes are the individuals mentioned in the corpus
  - ➢ model of time activity (30 years)
- ❑ "the seignorial effect": three major seignorial families (Ratier de Castelnau, Laperarede, and Roquefeuil) are omnipresent and mask other segments of the society
  - ➢ These three families are deleted from the network and ignored in construction of the "mimetic" ties
- ❑ "the homonymy effect": the most common type is when a son and his father have the same first name.
  - ➢ when we could not resolve name-ambiguity by using other attributes (e.g., geographical location, personal network, *etc.*), we preserved both individuals

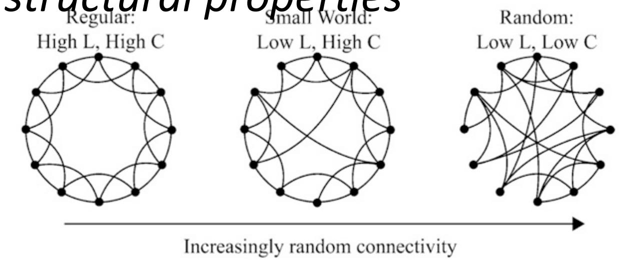
Following these principles, we constructed two social networks and the largest connected component is kept in each:

- **1240-1340: n=2462 / m=51891**
- **1440-1520: n=1786 / m=80546**

2. Network analysis

✓ Network science: brief overview

An important discovery is that large scale real world networks [social networks, real neural networks, ecosystems networks, …] share numerous common structural properties

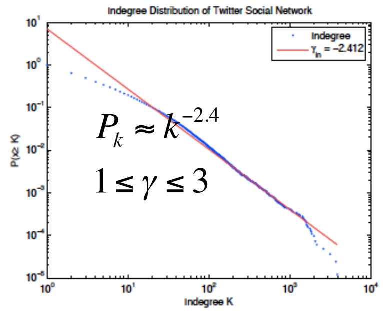


[Watts & Strogatz (1998)]

- ❑ Small-world pattern

it is possible to go from one vertex to any other passing through a very small number of intermediate vertices AND two neighbors of a same third vertex have a high probability to be neighbors.

Low average path length (L) and high clustering (C)



[Ghosh *et al.* (2012)]

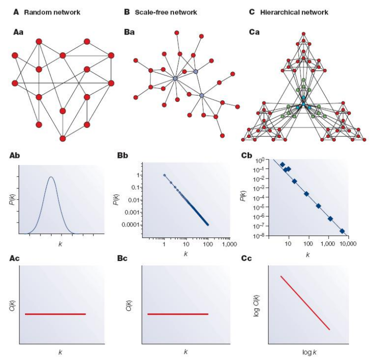
- ❑ Scale-free property

Statistical distributions (degree, betweenness, …) are heterogeneous and follow a power law : many vertices have just a few connections while a few hubs have a high number of connections.

- ❑ Hierarchical organization and “rich club”

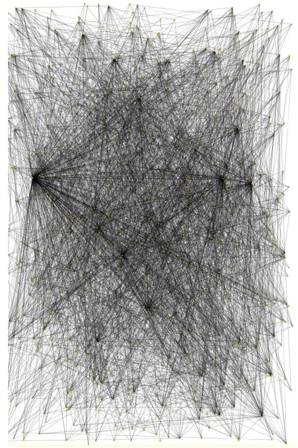
Communities of individuals are generally identified by subgraphs that have a high density of connections and the “rich club” phenomena is that high degree nodes (hubs) are very well connected to each other.

$$P_{C(k)} \approx k^{-\beta}$$

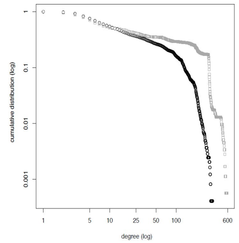


[Barabasi & Oltvai (2004)]

2.a. Network analysis: around "communities"



- The degree of a vertex is the number of neighbors it has within the graph;
- The betweenness (centrality) of a vertex is the number of shortest paths going through the vertex
- Both networks are small-world and scale-free (TPL)

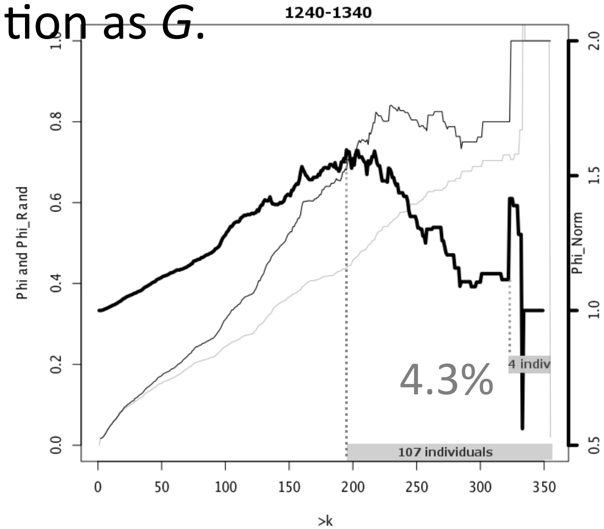


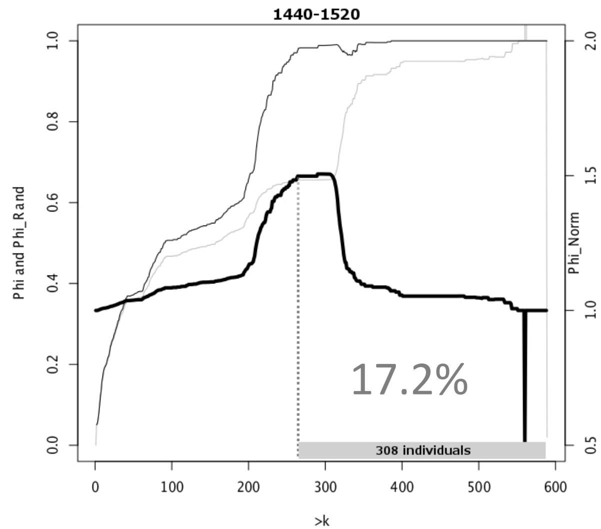
✓ The rich club

- The "rich-club" phenomenon is a usual concept in social network analysis that refers to the tendency of high degree nodes of a graph *G* to be extremely connected among themselves relative to the connections in a random graph with the same degree distribution as *G*.

$$\Phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k}-1)}$$

$$\Phi_{norm}(k) = \frac{\Phi(k)}{\Phi_{rand}(k)}$$



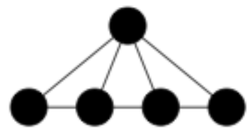


2.a. Network analysis: around “communities”

✓ The perfect communities

- The concept of “community” is often central because it allows a schematization of the network and therefore an approach to another granulometry scale. If a “’community” is generally a set of individuals highly interconnected, other definitions may correspond to well studied mathematical concepts :
  - Two vertices of a same community are linked -> complete subgraphs
  - Two individuals of a same community have similar neighbors -> twins [CJ 2008, BIJS 2014]
  - Two individuals of a same community have similar neighbors outside the community -> interval [Schmerl & Trotter 1993, Boudabbous & Ille 2009]
  - Two individuals of a same community have the same graph distances within the community and the entire graph -> isometric nodes [Anstee & Farber 1988]
- An interval of a graph G is a subset of vertices of G such that:

$$\forall i, j \in I, \forall x \in V(G) - I, i \sim x \iff j \sim x.$$



Example of graph with a 4-interval but without 2-interval

Definition

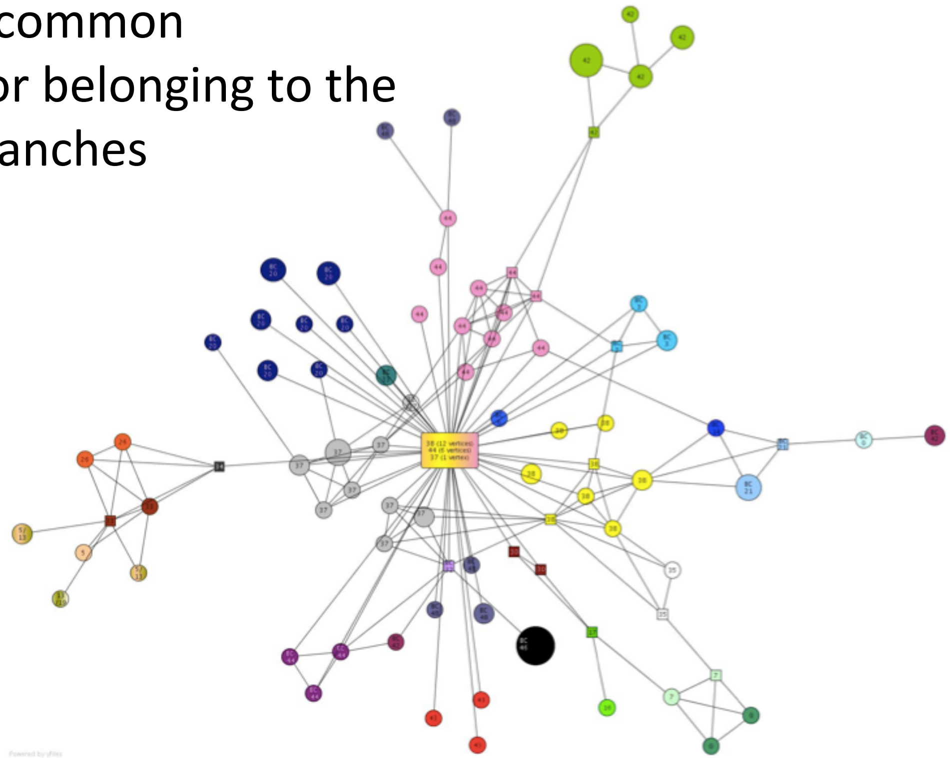
A perfect community of a graph G is a complete subgraph of G which is an interval of G.

2.a. Network analysis: around "communities"

✓ The perfect communities and the rich club in the XIV°s / XVI°s.

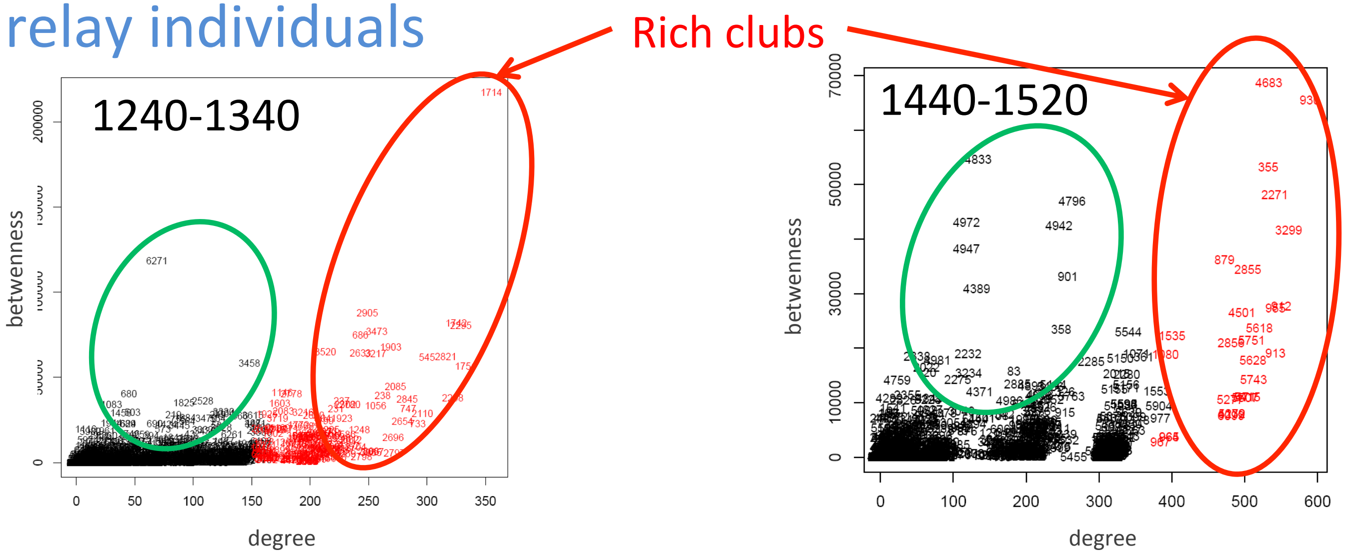
- A non trivial perfect community is a set of twins, that is individuals who have the same relationships.
- We are interested in perfect communities which are not trivial

- A perfect community brings together individuals having a common geographical origin or belonging to the same family main branches



2.a. Network analysis: around “communities”

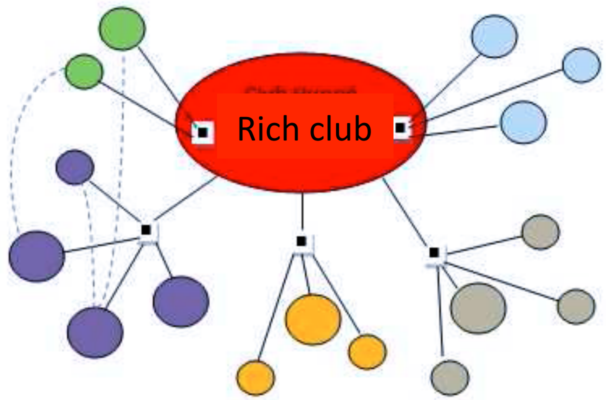
✓ The relay individuals



- We are interested in nodes which are not necessary in the rich club but have a high betwenness centrality
- These “relay individuals” are the links connecting perfect communities with the rich club and part of a hierarchization of the “rich club / relay individuals / communities” network.

➢ The records reveal that “relay individuals” often have characteristics close to those of the rich-club individuals but they are very marked in terms of their involvement in land-related transactions. It is certain that some of them were not peasants but rather members of an upper echelon of well-off craftsmen and small merchants with local clout.

2.a. Network analysis: around “communities”



✓ conclusions

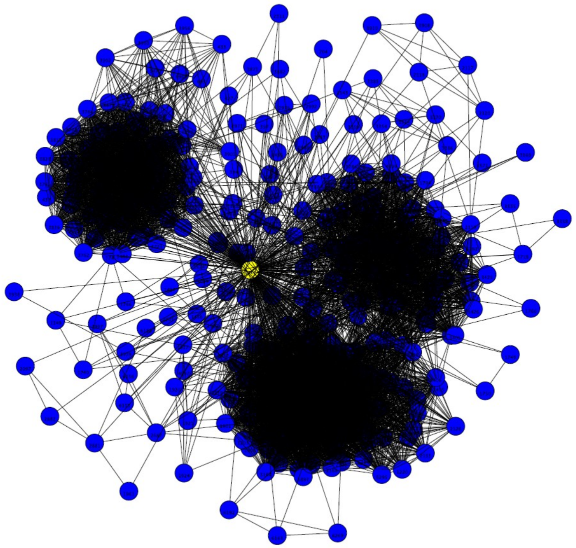
- A method to highlight hierarchical organizations
- A schematic organization which traverses the crises
- A rich club which increases form 4.3% to 17.2% of the population,
- If the “elite” is somewhat represented by the individuals of the “rich club”, we have a dilution when crossing the war
- An individual is not only determined by the social class of his family but also by his social individual relationships
- Spatial proximities and kinship ties are not sufficient to explain the organization [VJRH 2012]

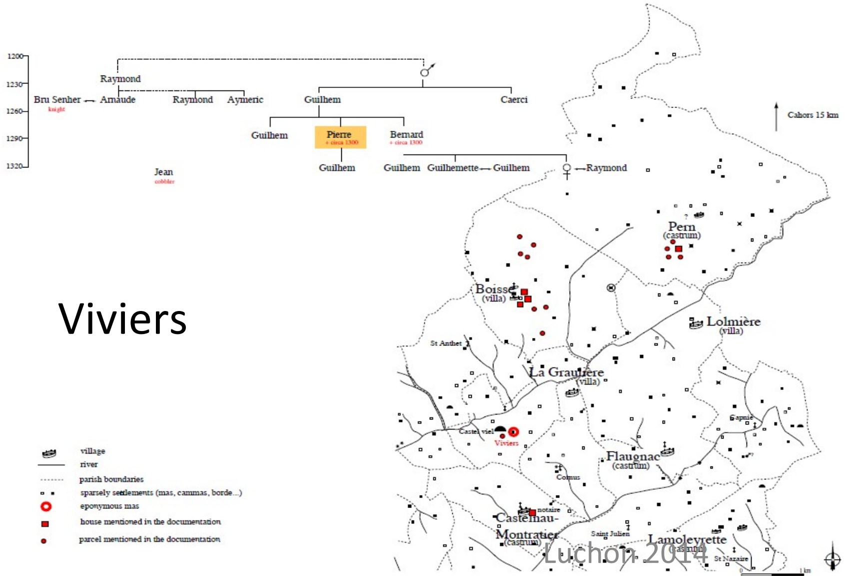
2.a. Network analysis: around "communities"

✓ Perspectives in that direction

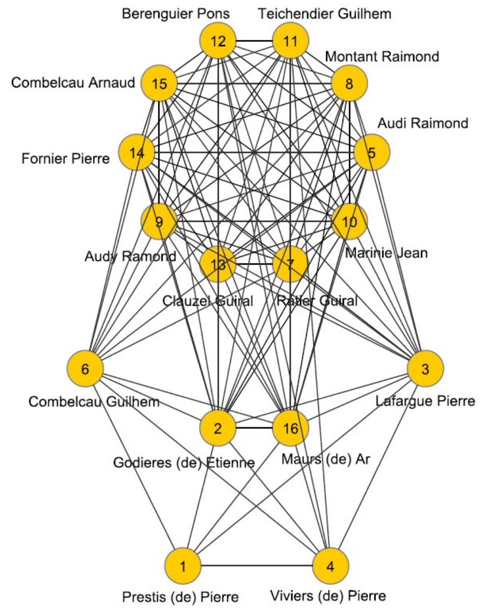
- ❑ Deeper study of the rich-club and relay individuals

Prestis





Viviers

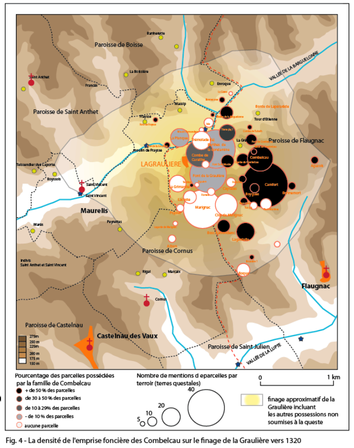


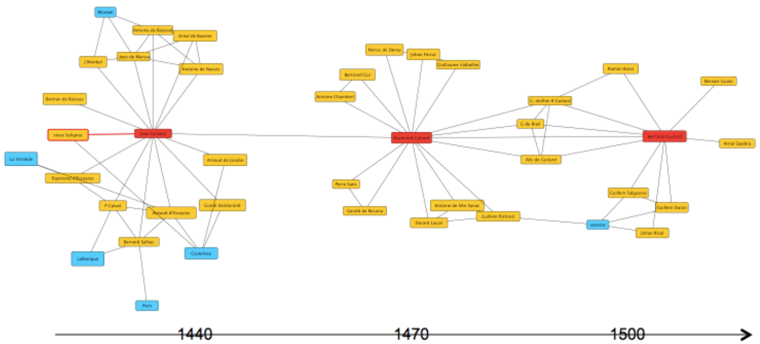
Top 16 of the rich-club

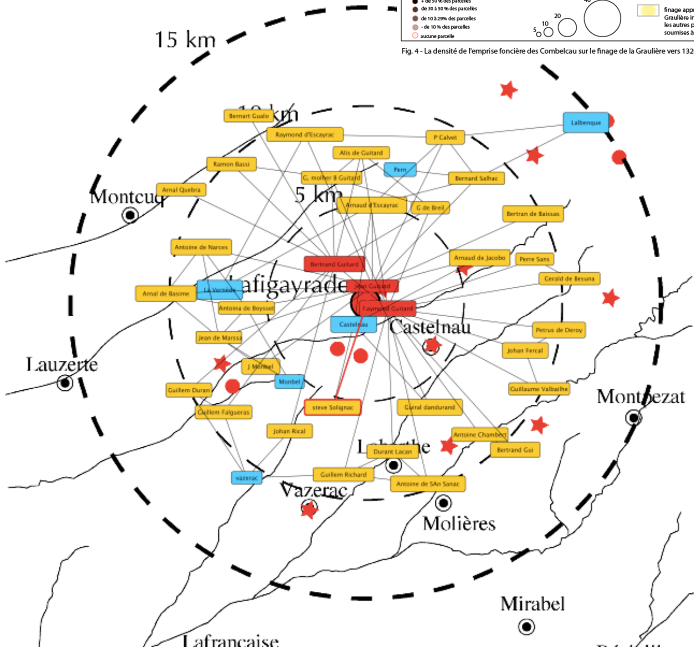
2.a. Network analysis: around “communities”

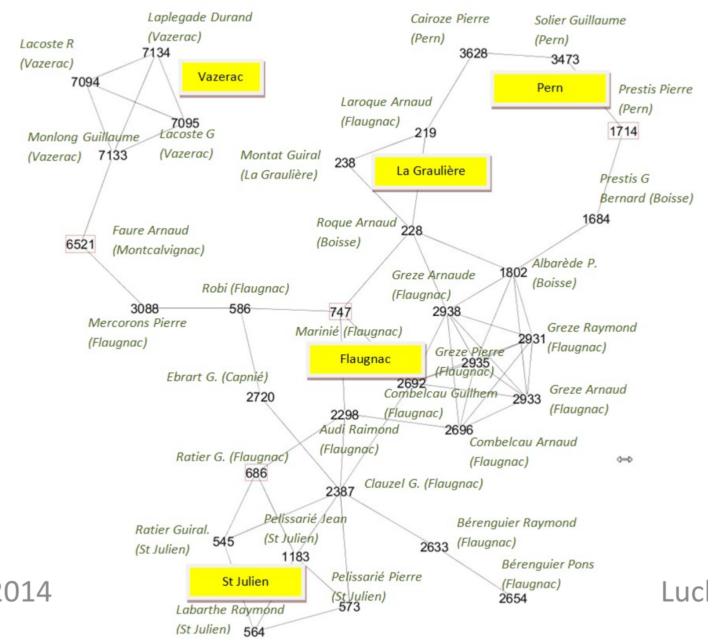
✓ Perspectives

❑ Dynamics of the network and land distributions



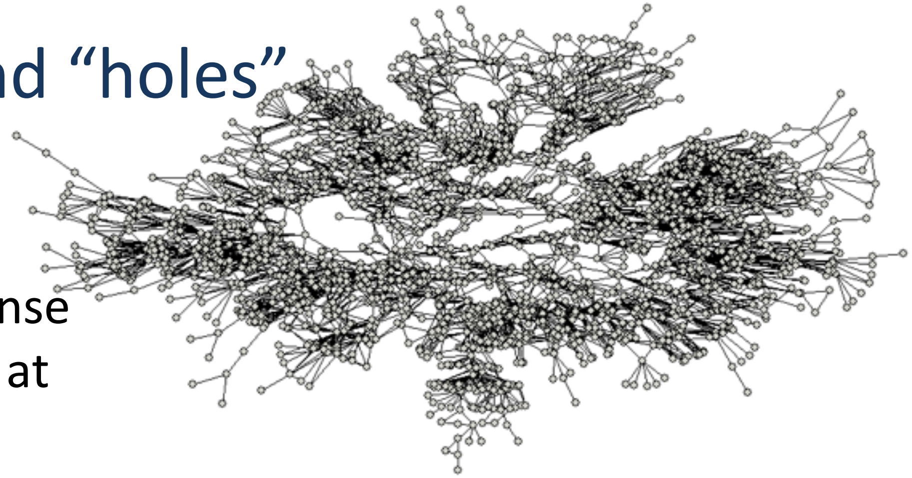






2.a. Network analysis: around "holes"

or how peeling a graph …



- Although most of the tools focus on dense parts of a network, we propose to look at non-dense parts : "holes"

- Sociological concept -> "structural holes" of R.S. Burt (1982) which are places with a low density of links and that an individual must hold to increase his influence ?
  - Topology is the "natural" branch of mathematics for dealing with the form of objects
  - It exists a common link between "graphs" and "topology" by the way of simplicial complex.

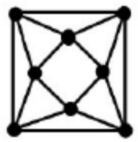
- A simplicial complex *K*, with vertex set *V*, is a collection of non empty sets of *V* (the simplices) *s.t.:*

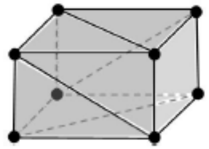
$$V = \bigcup_{\sigma \in K} \sigma \text{ and if } (\sigma \in K, x \in \sigma) \text{ then } \sigma - \{x\} \in K$$

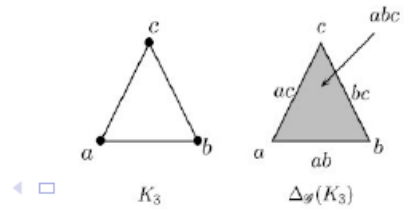
2.a. Network analysis: around "holes"

- *Δ(G)* denote the simplicial complex whose k-simplices are the complete subgraphs with k vertices.

  (flag complexes)



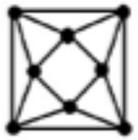


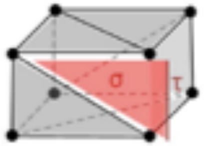


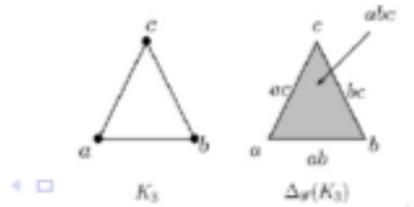
2.a. Network analysis: around “holes”

- *Δ(G)* denote the simplicial complex whose k-simplices are the complete subgraphs with k vertices.

  (flag complexes)







- An elementary reduction in *Δ(G)* is the suppression of a pair of simplices *(σ,τ)* s.t. τ is a proper maximal face of σ and τ is not the face of another simplex.
- Two flag complexes have the same homotopy type if we can go from one to the other by a finite succession of elementary reductions or increases (formalization of a continuous deformation)

 is not homotopic to  (you have to tear the ball)

> ➢ How to define such a notion of homotopy on graphs ?

2.a. Network analysis: around “holes”

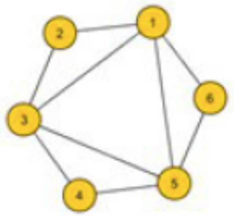
Definition [Quillot, 1983 ; Nowakowski and Winkler, 1983]

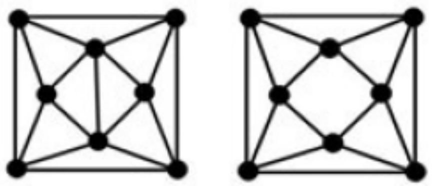
- A vertex i is dismantable if it exists $j \neq i$ such that $V_G[i] \subseteq V_G[j]$.
- A graph G is dismantable if it exists an order $1, 2, \cdots, n$ of its vertices such that i is dismantable in $G - \{1, 2, \cdots, i-1\}$.

$N[1] = \{1,2,3,5,6\}$

$N[2] = \{1,2,3\}$

$N[2] \subset N[1]$

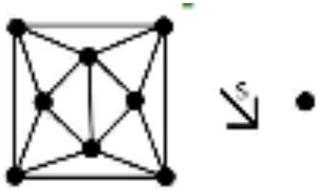




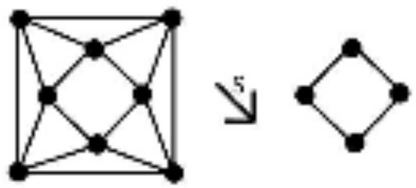
No dismantlable vertex

Definition [2008]

- A vertex i is s-dismantable if $V_G(i)$ is dismantable.
- A graph G is s-dismantable if it exists an order $1, 2, \cdots, n$ of its vertices such that i is dismantable in $G - \{1, 2, \cdots, i-1\}$.



and



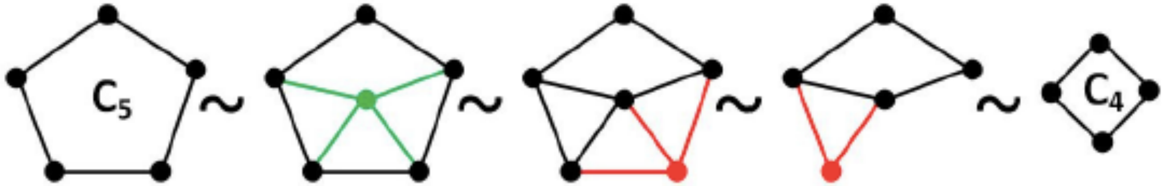
2.a. Network analysis: around "holes"

Definition [2010]

We say that G and H have the same homotopy type iff it exists

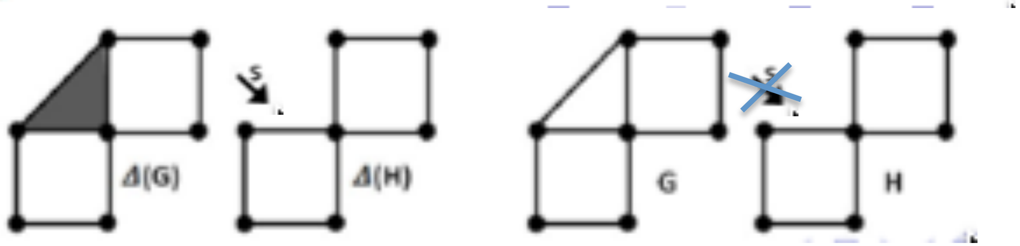
$G = J_1, J_2, \cdots, J_{k-1}, J_k = H$ s.t. $G = J_1 \overset{s}{\to} J_2 \overset{s}{\to} \ldots \overset{s}{\to} J_{k-1} \overset{s}{\to} J_k = H$

$\overset{s}{\to}$ represents the addition or suppression of a s-dismantlable vertex.



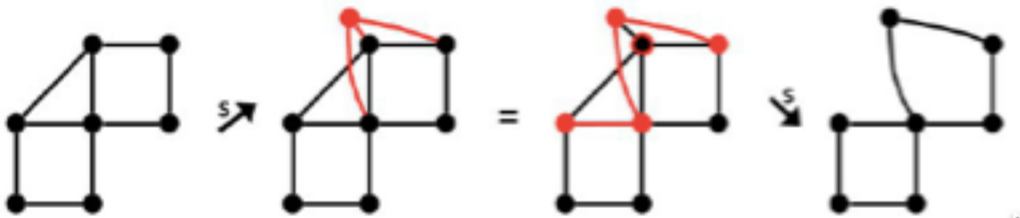
Résultat [2008, 2010]

$G \searrow_s H \Rightarrow \Delta(G) \searrow_s \Delta(H)$.



Résultat [BFJ 2010]

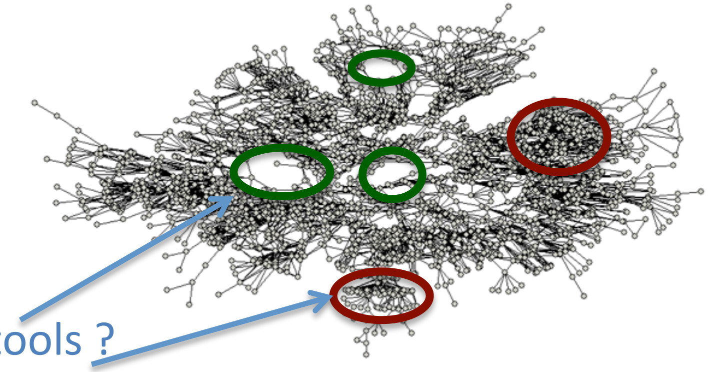
$[G]_s = [H]_s \Leftrightarrow [\Delta(G)]_s = [\Delta(H)]_s$



The application to our dataset is still … in progress ☺

3. Conclusions

- New advanced tools in the domain of applied graph theory
- A close cooperation with the specialists of the application domain (historians)
- A lot of questions both in the field of history and mathematics



Thank you

Questions :

① It is known that if G is vertex transitif (ie. the neighborhoods of the vertices are all isomorphic) and dismantlable then G is a complete graph. Is it true if dismantlable is replaced by s-dismantlable ?

② What are the tools for studying a network constructed from a single database but which contains a lot of small connected components ? (this is the case for a lot of social networks when the rules for defining the links are very restrictive)