



# Graphes pour les sciences sociales

École d'été du GRETSI - Peyresq

Tiphaine Viard  
i3, Institut Polytechnique de  
Paris

# Qu'est-ce que la sociologie ?

Étude **systématique** de la société et des interactions sociales

*« il n'y a pas de danseur sans la danse,  
Il n'y a pas de danse sans les danseurs »*

Plusieurs niveaux d'analyse :

- **Micro** : individus, petits groupes...
- **Méso** : groupes, petites institutions...
- **Macro** : institutions, états...
- **Global** : qui dépasse une société, un état

Ethnographie

Philosophie

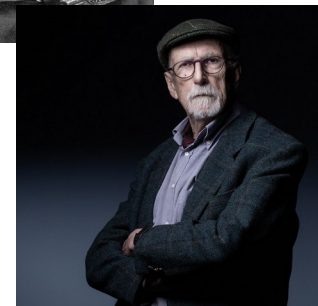
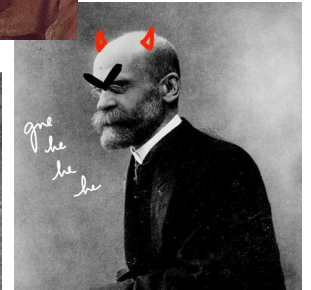
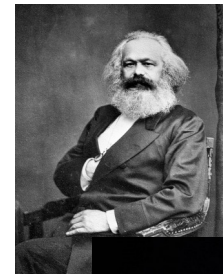
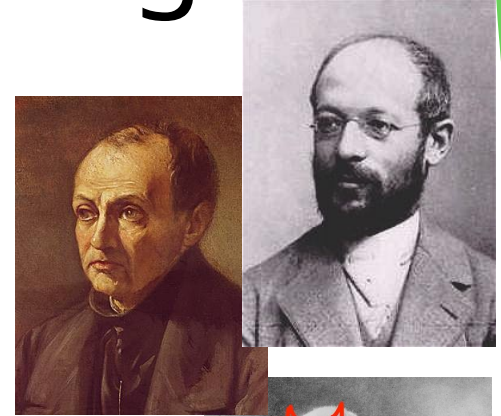
Psychologie

Quelques questions :

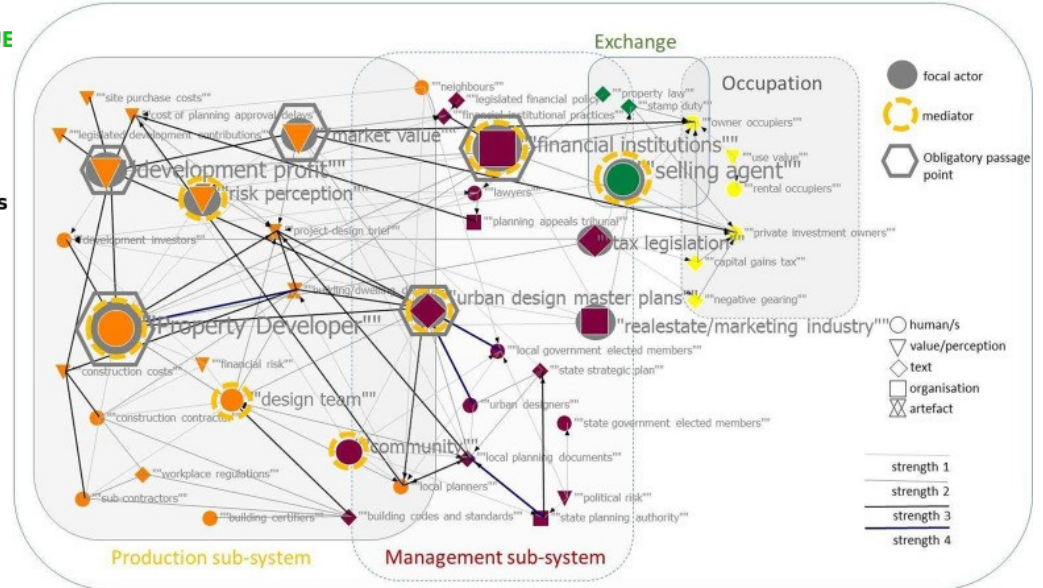
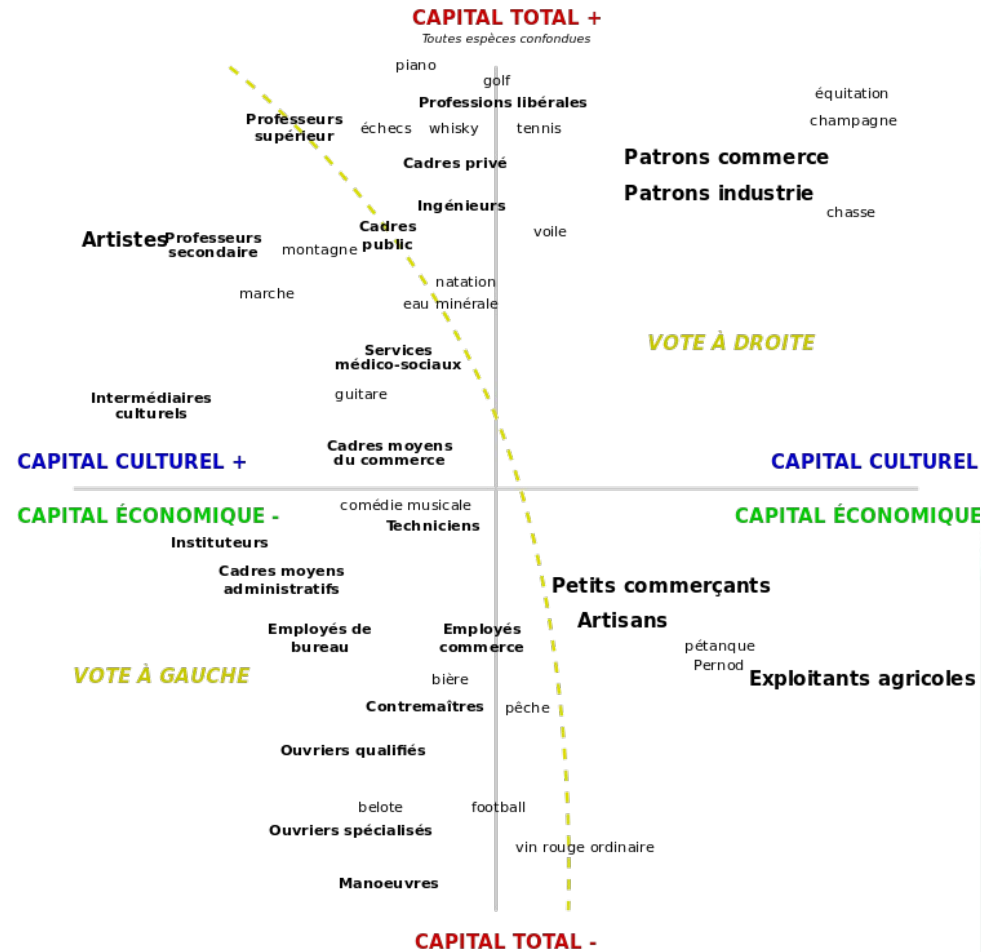
- Comment ces niveaux s'articulent-ils ?
- Peut-on les étudier conjointement ?
- Décrire les effets d'**habitualisation** et d'**institutionnalisation**

# Brève histoire de la sociologie

- Le **positivisme** (A. Comte) : étudier la société comme la physique étudie la nature
  - Passer d'une société *théologique* à une société *métaphysique* puis *positiviste*
- **Fait social** vs **action sociale** (Durkheim, Weber)
  - Approche *holiste* : « toute manière de faire, fixée ou non, susceptible d'exercer sur l'individu une contrainte extérieure; ou bien encore, qui est générale dans l'étendue d'une société donnée tout en ayant une existence propre, indépendante de ses diverses manifestations au niveau individuel »
  - Approche *située* : « les faits humains ne sont pas entièrement déterminés, qu'ils ne sont pas soumis à des lois comme le sont les faits de la nature, et qu'ils ne peuvent être expliqués qu'en prenant en compte le sens donné par les acteurs. »
- La **critique** de la société (Marx)
  - Les forces sociales se structurent autour de rapports de force entre classes dominantes et dominées économiquement ; la sociologie doit mettre en lumière ces mécanismes de domination pour les dépasser
- L'analyse **d'interactions** (Simmel)
  - « Society exists where a number of individuals enter into interaction. »
  - Les structures macro n'ont pas de propriétés propres (*sui generis*), on peut les réduire aux sommes d'interactions individuelles
- Au 20<sup>e</sup> siècle (Bourdieu, Latour...)
  - Dépasser, combiner ces constructions, apporter du quantitatif...



Chaque auteur apporte des grilles d'analyse différentes et complémentaires

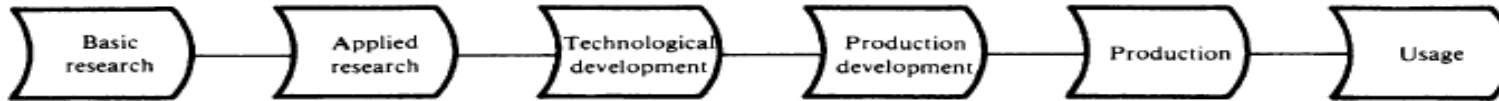


# Social Construction of Technology (SCOT) (1)

- Wiebe Bijker et Trevor Pinch (1984), Bijker et al (1987), Bijker (1995)
- Aborder la problématique « techniques en société »  
à partir des réflexions menées sur la science:  
EPOR → SCOT
- Repoussoirs :



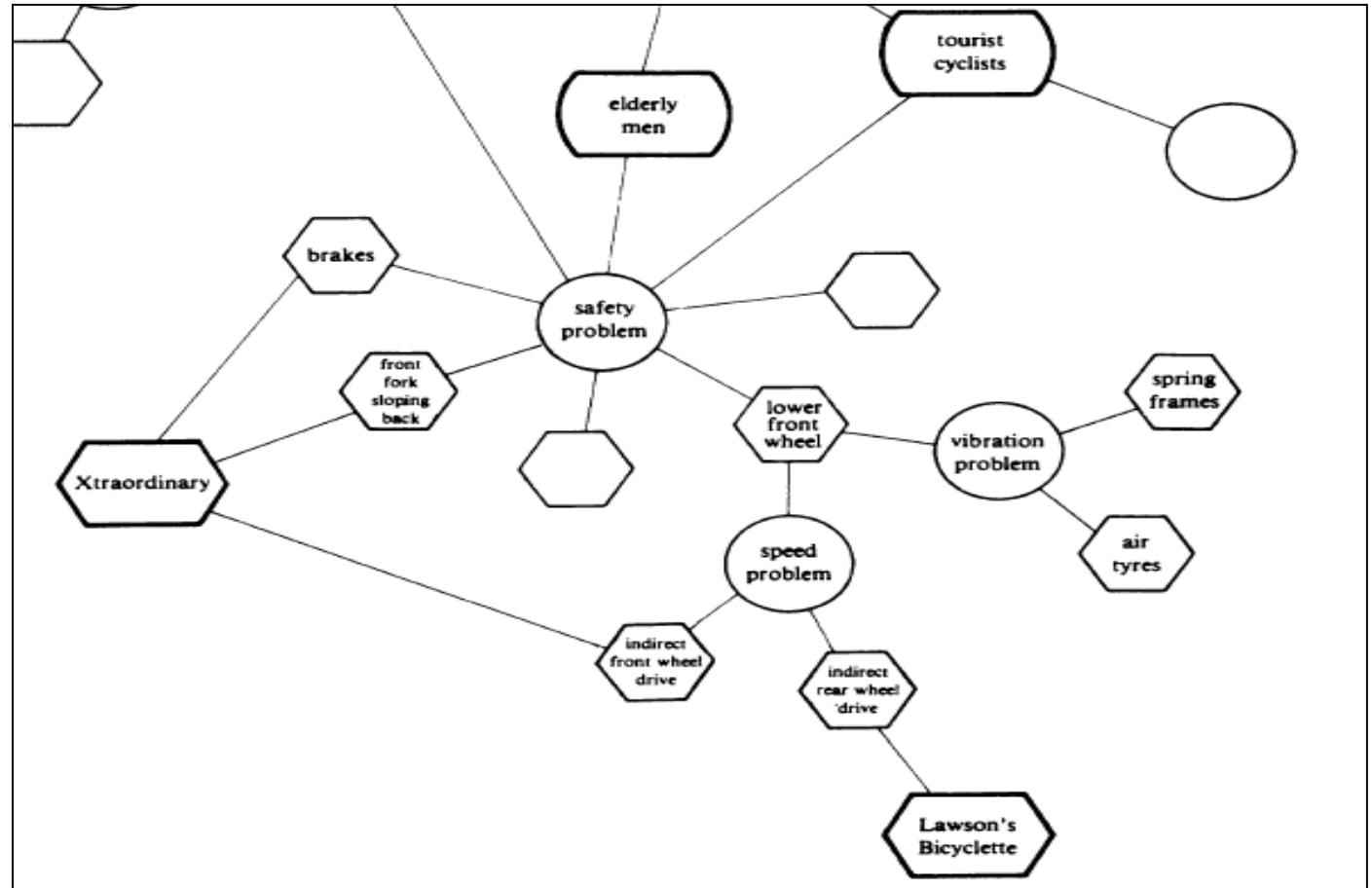
- Les études d'innovation réalisées par les économistes prennent peu en compte le contenu technique et s'appuient sur un « modèle linéaire » irréaliste



- Les études d'histoires sont assez descriptives, ne s'intéressent qu'aux réussites et ramènent leur succès à leurs caractéristiques techniques (alors que celles-ci sont émergentes)
- La sociologie des technologies est émergente et se cherche autour de notions comme les paradigmes, les controverses
- Propositions:
  - étudier la création technique en suivant les opérations de **variation** et de **sélection**
  - Étudier l'intrication entre des artefacts, des groupes sociaux, des problèmes et des solutions techniques
  - « **Relevant social groups** », « **Interpretative flexibility** », « **Closure mechanisms** »

## Social Construction of Technology (SCOT) (2)

- La méthode: retracer les liens entre artefacts, groupes, problèmes, solutions pour suivre le cheminement de la création technique et l'avancée vers une stabilisation éventuelle.



# Social Construction of Technology (SCOT) (3)

- Cas d'étude: l'histoire de la bicyclette
- Non pas 1884, mais une longue période de variations et sélections entre 1879 et 1898



- « **Relevant social groups** »: des hommes sportifs, des femmes qui veulent s'émanciper, des personnes âgées, etc
- « **Interpretative flexibility** »: le vélo à grande roue permet d'aller plus vite (pour les sportif) mais est perçu comme plus dangereux (pour les personnes âgées)
- **Stabilisation** et **clôture**: rhétorique quand la publicité prétend que le vélo à grande roue est sûr; par redéfinition du problème à propos des pneumatiques considérés comme inesthétiques et ridicules.

# La sociologie des techniques dans le courant de l'acteur réseau (1)



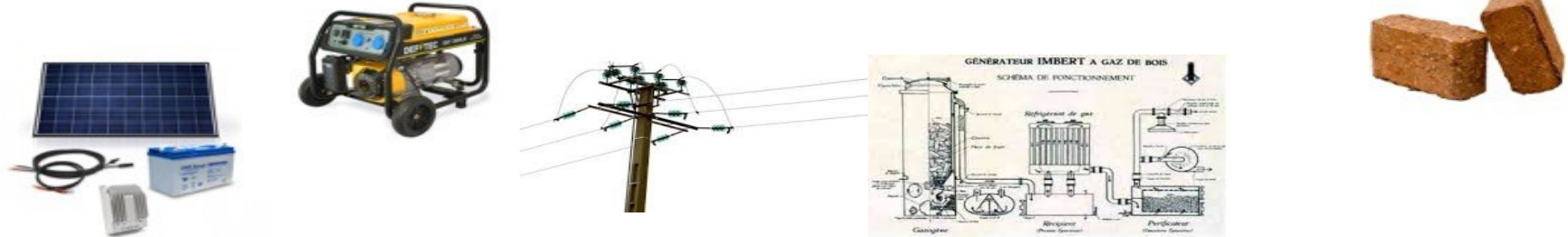
- Comme pour SCOT, un courant qui développe une analyse sur la technique en société en lien avec une réflexion sur l'activité scientifique ...
- ... mais pas une transposition: plutôt des analyses sur la science, l'innovation et la technique qui viennent se projeter dans la perspective commune de l'acteur réseau.
- En fait, presque un retournement de perspective (Akrich, 1987):
  - Les sciences se veulent hors du monde social, le travail des sociologues a consisté à montrer qu'elles y appartiennent bien.
  - À l'inverse, les objets techniques sont hétérogènes « mi-chair mi-poisson »: partiellement distinctifs du monde social et culturel, et partiellement adhérents. Comment les « décrire » ?
- Une méthode: l'étude de l'innovation (ou le transfert de technologie, ou l'archéologie...) pour prendre de la distance par rapport à l'évidence trompeuse des rapports technique / société.
- Au-delà de la vulgate (« le travail de la conception consiste à inscrire des scripts dans les objets techniques »), et de la réduction disciplinaire (« approche sémiotique ») une approche des modalités d'élaboration des relations entre contenus techniques et leur inscription socio-culturelle.

# La sociologie des techniques dans le courant de l'acteur réseau (2)

- 7 idées fortes dans l'analyse des processus d'innovation technique
  - 1. Il n'y a **pas d'extériorité de position** du concepteur par rapport à l'utilisateur: pour comprendre les modalités du processus d'innovation qui les met en relation, il faut examiner le réseau qui assure les processus de représentations (Akrich, 1987; Akrich, Callon, Latour, 1992)
  - 2. La conception ne relève pas d'un processus d'inscription ... mais d'un **partage** entre ce qui est inscrit dans le dispositif, ce qui peut être prescrit à son utilisateur, ce qui peut être pré-inscrit dans l'environnement (Akrich, 1992). Lien avec la notion de configuration (Woolgar)
  - 3. La rencontre entre le projet et son usage est une de-scription, c'est-à-dire un processus **imprévisible et irréductible** qui met à l'épreuve le lien entre l'usager inscrit et l'usager réel (Akrich, 1987)
  - 4. Au modèle linéaire, il faut substituer un modèle itératif et **tourbillonnaire** dont le terme correspond à la stabilisation de l'innovation (Akrich, Callon, Latour, 1992)
  - 5. Ce processus tourbillonnaire assure une **co-construction**: il produit un objet technique, mais aussi des acteurs sociaux aux identités spécifiques qui l'utilisent (# SCOT).
  - 6. Une fois les dispositifs techniques mis en place, ils peuvent opérer comme des objets frontière permettant **d'articuler des régimes techniques et des régimes sociologiques** de l'action (Akrich, 1994)
  - 7. Les utilisateurs peuvent **agir en innovateurs** en s'emparant de dispositifs techniques stabilisés et les transformer pour en inventer de nouveaux usages (Akrich, 1998)

# La sociologie des techniques dans le courant de l'acteur réseau (3)

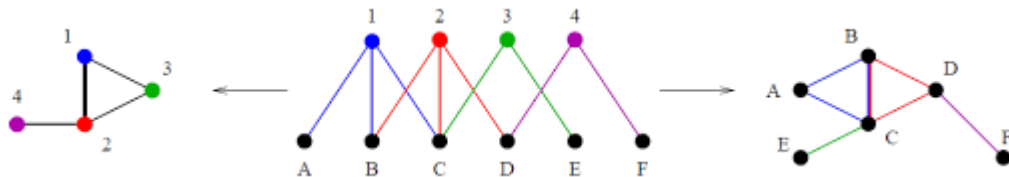
- Une série d'études sur les technologies énergétiques ...



- ... qui informant la réflexion sur la construction de systèmes énergétiques:
  - Une réflexion sur l'inscription socio-politique des réseaux énergétiques ...
  - ... et sur la manière dont ils contribuent à organiser les relations permettant le partage et la distribution de l'action technique dans le collectif
  - Un dialogue possible avec les théories des systèmes techniques (Hughes, 1988)
- Critiques de la sociologie de l'innovation proposée par l'acteur réseau:
  - Elle focalise l'attention sur l'innovateur, dont elle donne une figure héroïque (Chateauraynaud, Dodier)
  - Elle fait l'hypothèse trop forte de la stabilisation des techniques et se rend aveugle au travail de l'utilisateur (Oudshoorn et Pinch, Orlikowski)
  - Elle ne rend pas compte des usages sociaux, mais de l'usabilité (Jouët, Flichy)

# Graphes de corpus

- Étant donné  $D$  documents, on peut construire un graphe de **co-occurrence**
  - Les nœuds sont des termes/mots, et l'on met un lien entre deux termes s'ils apparaissent dans le même document
- Problèmes :
  - Il y a beaucoup de termes différents → n-grammes
  - Tous les termes ne sont pas porteurs de sens → stop words et TF-IDF
  - Cela revient à projeter un réseau biparti, et à former d'immenses cliques → Quelle est la nature du **lien** ? Comment rendre l'information **visible** ?



$$tf(t) \cdot \frac{|D|}{|\{d : t \in d, d \in D\}|}$$

# Analyser un graphe de corpus

- Que veut-on regarder ? Importance de définir les hypothèses en amont
- Nécessité d'effectuer des **aller-retours** constants
- Par ex, on peut s'intéresser aux **disciplines sociales**

*« Les formes de solidarité et d'exclusion, de socialisation et d'apprentissage, de contrôle social et de résolution de conflits, de régulation et d'institutionnalisation. » (Lazega, 2015)*

- Jouer avec les paramètres, et raffiner l'analyse
- La connaissance du **terrain** étudié est indispensable

# Préserver le dialogue quali-quant

- Les sociologues, comme scientifiques, font **partie** de la société étudiée

*La science en action, Latour, 1987*

- Le **social** n'existe pas en dehors de nos interactions **situées** : pour comprendre les faits de société, il faut analyser le **réseau d'acteurs**
- Les graphes permettent de **modéliser** et **comparer** les structures sociales
- La description scientifique ne peut se faire que par **aller-retours** successifs entre le terrain et la modélisation
  - C'est ce qui permet de suivre le processus tourbillonnaire, et d'adapter les formalismes aux réalités du terrain

*Éléments pour une sociologie de la traduction : la domestication des coquilles Saint-Jacques à St-Brieuc, Callon, 1986*

# Analyser les élections présidentielles en France

*Gaumont, Panahi, Chavalarias (2018)*

- But : définir une méthodologie d'extraction de données et d'analyse pour le suivi des **fonctions démocratiques dans les espaces numériques**
- Cas pratique : Twitter, l'élection présidentielle française de 2017 – 60M de tweets, ~2.5M comptes
- Une approche mixte :
  - Catégorisation automatique de contenus politiques ;
  - Catégorisation manuelle de 5000 profils ;
  - Correspondance avec les « temps forts » de l'élection et impact sur le réseau.
- En creux, une analyse des *fake news* et *echo chambers* (voir D. Cardon)
- **Niveau méso** : entre statistiques nationales et questionnaires sociologiques

# Analyser les élections présidentielles en France

- Identifier les **groupes sociaux**

*« Une collection d'êtres en tant qu'ils sont en train de s'imiter entre eux ou en tant que, sans s'imiter actuellement, ils se ressemblent et que leurs traits communs sont des copies anciennes d'un même modèle » (G. Tarde, 1890)*

- $G = (V, E)$ ,  $u \in V$  a Twitter account and  $(u, v, w) \in E$  means  $w$  retweets between  $u$  and  $v$

*$G$  is computed over a time window  $T$*

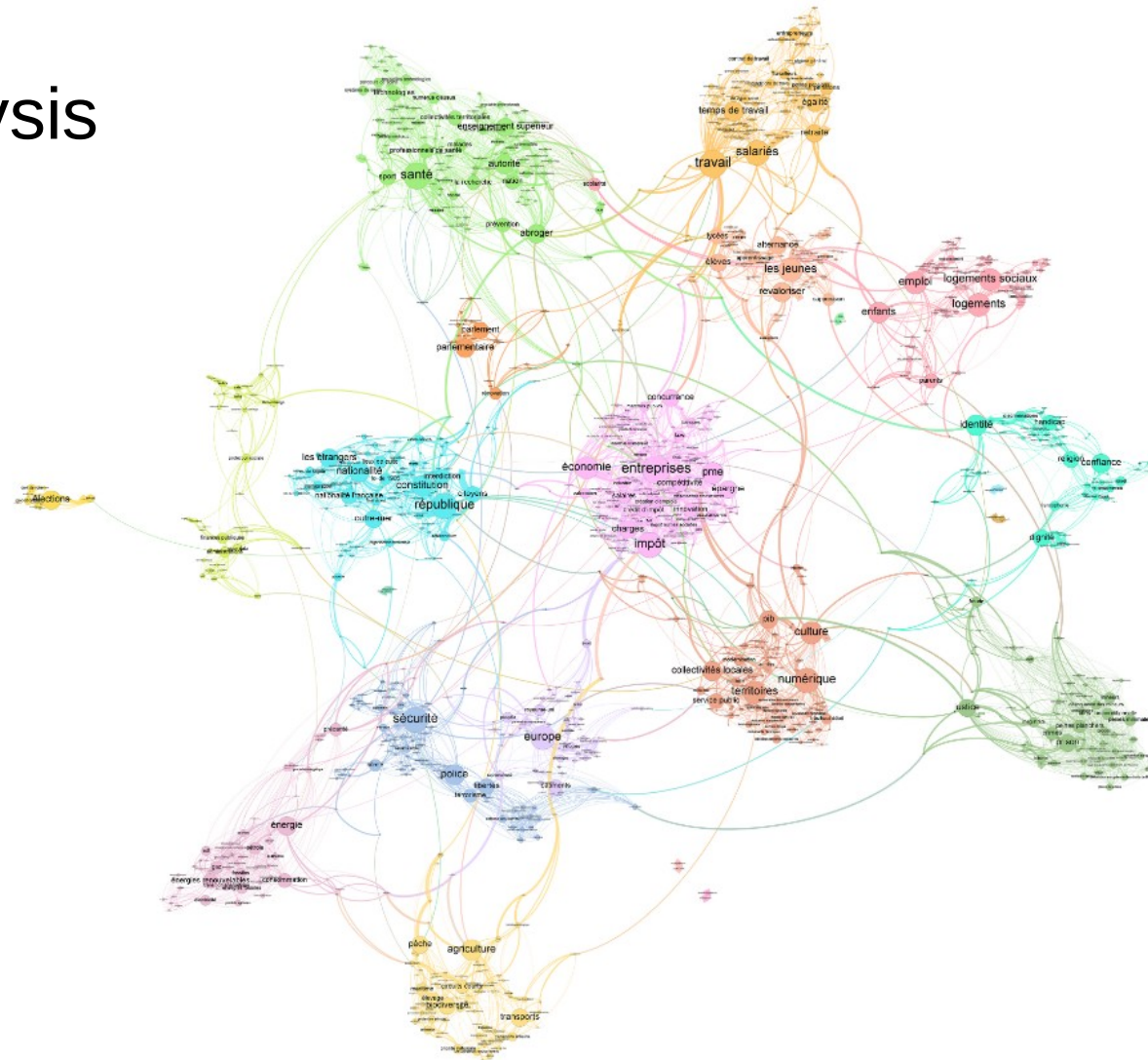
- $w$ -communities = communities on the graph with edges of weight  $> w$

*with Louvain*

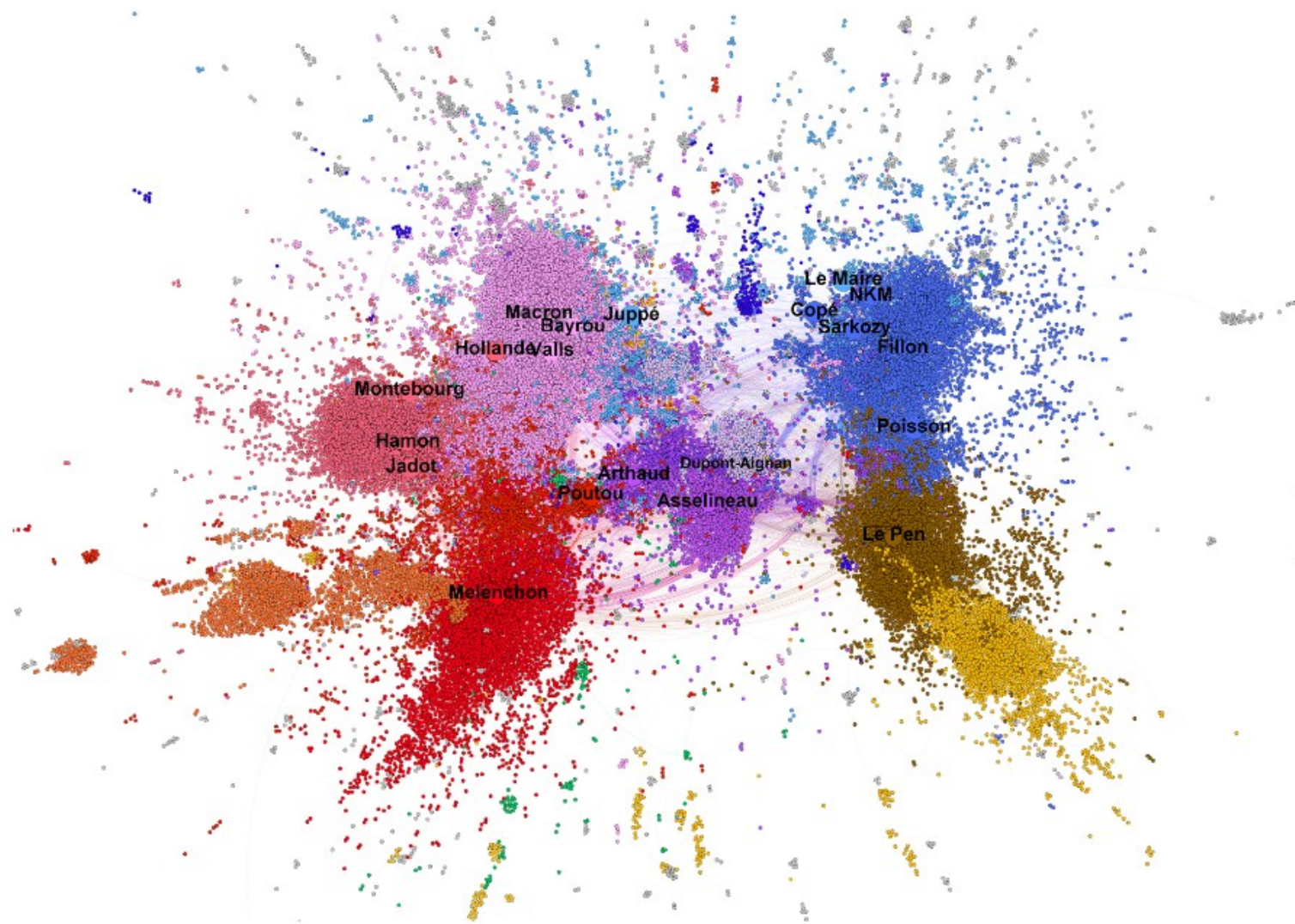
- Time window is set to be 2 weeks with 1 week overlap

$$|T_k| = 14 \text{ days}, |T_{k-1} \cap T_k| = |T_k \cap T_{k+1}| = 7 \text{ days}$$

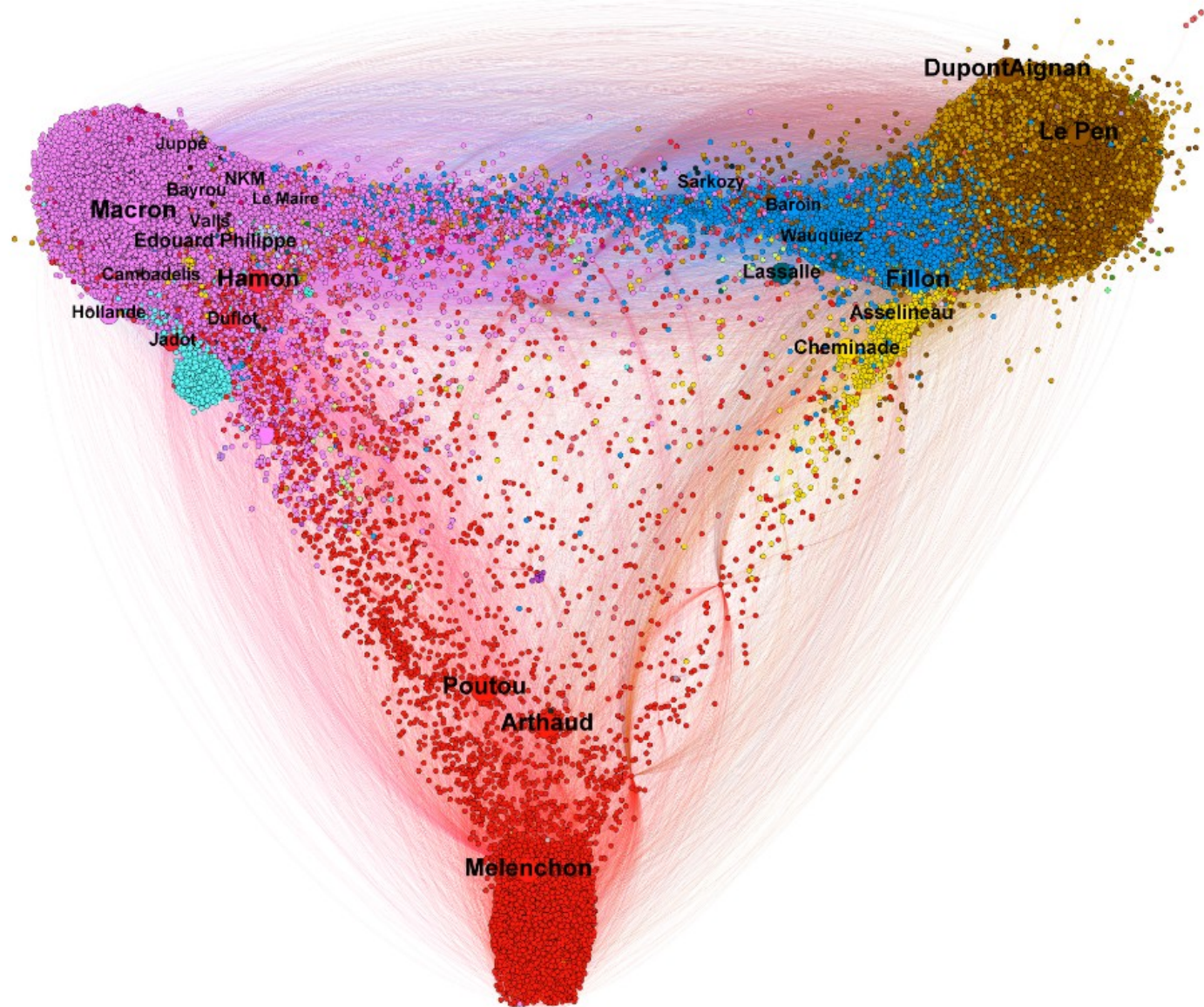
# Topic analysis

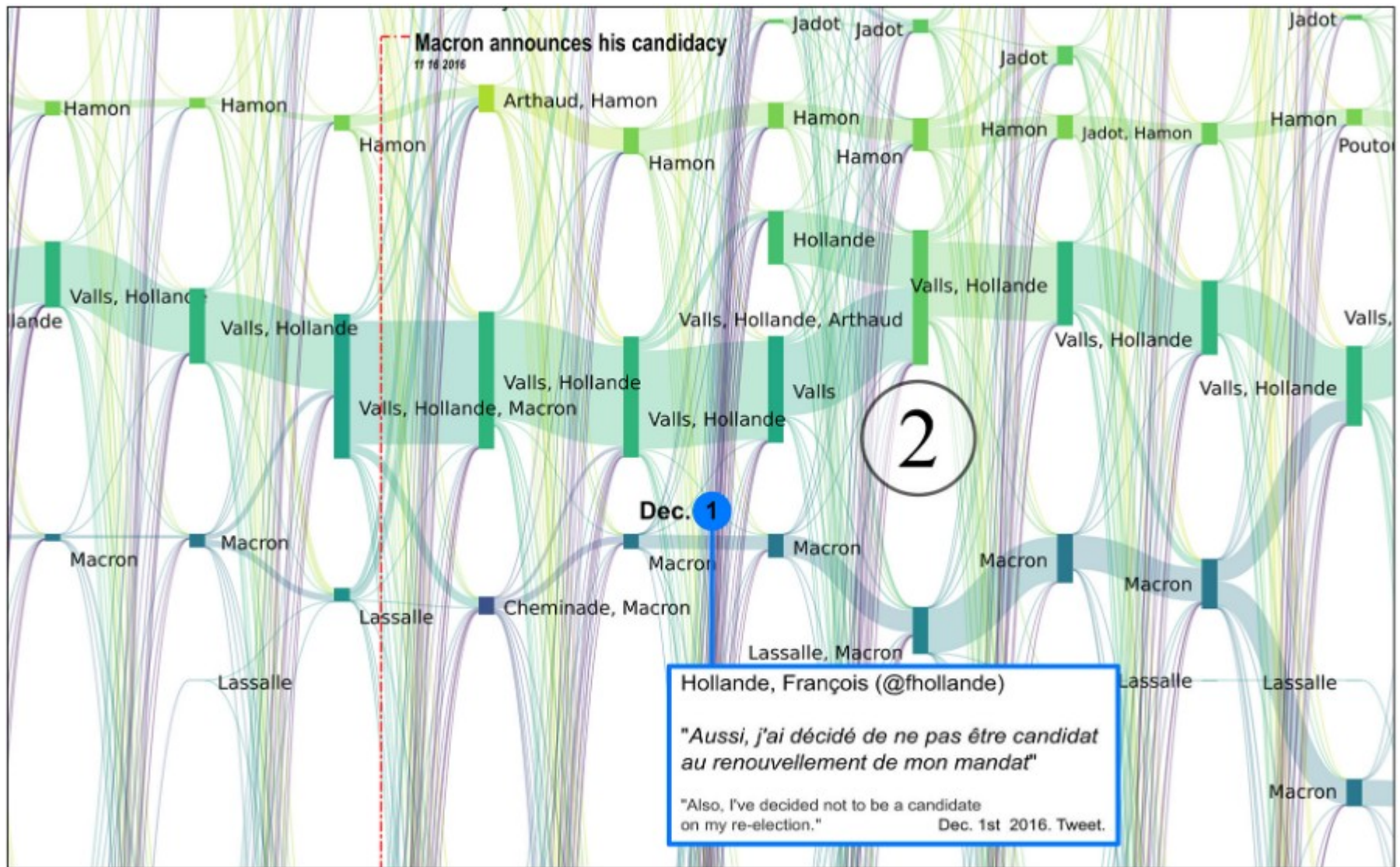


3-communautés  
au premier tour  
(avril 2017)



### 3-communautés Entre deux tours



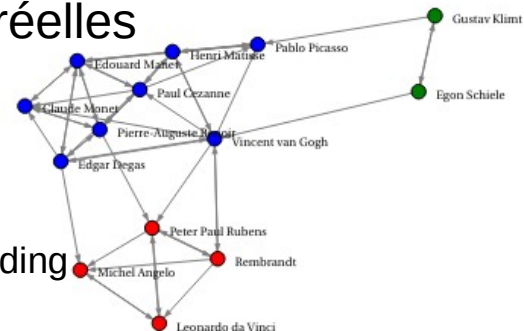


# Grands réseaux pauvres, petits réseaux riches : cesser la dichotomie

- Sociologie néo-structurale (E. Lazega)
  - Étudier les structures d'opportunité et de contrainte, modéliser et comparer les formes de disciplines sociales
- Dynamique (Latapy, Casteigts...)
  - Comment prendre en compte les temporalités ? Quelles garanties théoriques ?
  - Distinguer *graphes dynamiques* et *séquences d'interactions*
- Multicouches
- Attribués
- ...de très grande taille ! (millions de nœuds, milliards de liens)
- Quels liens avec l'apprentissage sur graphe ?
- Prendre en compte la spécificité algorithmique des graphes de données réelles : sparse, degrés hétérogènes, petit monde

# Une librairie : scikit-network

- But : gérer le traitement efficace de grands graphes de données réelles
- Basée sur le format **CSR** implémenté dans scipy + **Cython**
- Implémentation des algos classiques
  - PageRank (personnalisé), Louvain, Katz, Betweenness, Closeness, GNNs, embedding spectral...
- Des jeux de données
  - **Wikivitals**, US politics, Karate Club, Movies...
- Libre et ouvert aux contributions



[github.com/sknetwork-team/scikit-network/](https://github.com/sknetwork-team/scikit-network/)

	scikit-network	NetworkX	iGraph	graph-tool
Louvain	771	✗	1,978	✗
PageRank	48	📀	236	45
HITS	109	📀	80	144
Spectral	534	📀	✗	✗

Table 2: Execution times (in seconds). ✗: Not available. 📀: Memory overflow.

	scikit-network	NetworkX	iGraph	graph-tool
RAM usage	1,222	📀	17,765	10,366

Table 3: Memory usage (in MB). 📀: Memory overflow.

# Compressed Sparse Row

A (not so) sparse matrix (in dense format):

2 0

0 3

The same matrix in coordinates:

(0, 0, 2), (1, 1, 3)

or equivalently, in CSR format:

indices = 0, 1

indptr = 0, 1, 2

data = 2, 3

Dense = 4 values, sparse = 7 values

# Compressed Sparse Row format (2)

A sparse matrix (in dense format) :

$$\begin{bmatrix} 3 & 0 & 5 & 0 & 0 & 0 & 4 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 & 2 & 4 & 0 & 0 \\ 1 & 0 & 2 & 0 & 0 & 3 & 5 & 3 & 0 & 0 \\ 0 & 2 & 0 & 4 & 1 & 5 & 0 & 0 & 1 & 0 \\ 5 & 0 & 3 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Coordinates:

$$(0, 0, 3), (0, 2, 5), (0, 6, 4), \dots$$

The same matrix in CSR (Compressed Sparse Row) format:

$$\text{indices} = 0, 2, 6, 7, 4, 6, 7, 0, \dots$$

$$\text{indptr} = 0, 4, 7, 12, 17, 20$$

$$\text{data} = 3, 5, 4, 4, 4, 2, 4, \dots$$

Dense = 50 values, sparse = 46 values (20 + 6 + 20)

# Merci!

# Un TP : l'analyse des controverses autour des perturbateurs endocriniens

