L11. Inquiry 2.0

Olivier Godechot
Sciences Po – Campus Reims
Inquiries 4 Sociology

What changes for sociological inquiry?

- Object: New types of social sphere
  - New forms of communication
  - New forms of interaction
  - Reshaping, encounters, communities, circulation of information
  - Leave digital traces
- Method : New tools for research
  - Type of data
    - Already "coded"… Text, images, and variables
    - Big data. Size
    - Digital revolution
  - Type of method
    - All other methods
    - + Artificial intelligence

Online: the sampling problem in an age of access

- Important sphere of social life.
  - Cf. #giletsjaunes
- Presence and activity online greatly unequal
- Gigantic but biased data

Inequality and heterogeneity of online life

Internet: inflexion or transformation of forms of sociability

- Not the first disruptive technological changing remote personal communication
  - 16th century: royal mail
  - 19th century: telegraph
  - End of 19th century and early 20th century: telephone
  - Other abandoned inventions (pneumatics, minitel, etc.)
- Decrease in transportation costs counts also
- Mass communication devices
  - Gutenberg's galaxy. Books and printed press (18th)
  - Early 20th century: Radio
  - Mid 20th: television
  - Online ➔ From top down to peer-to-peer media

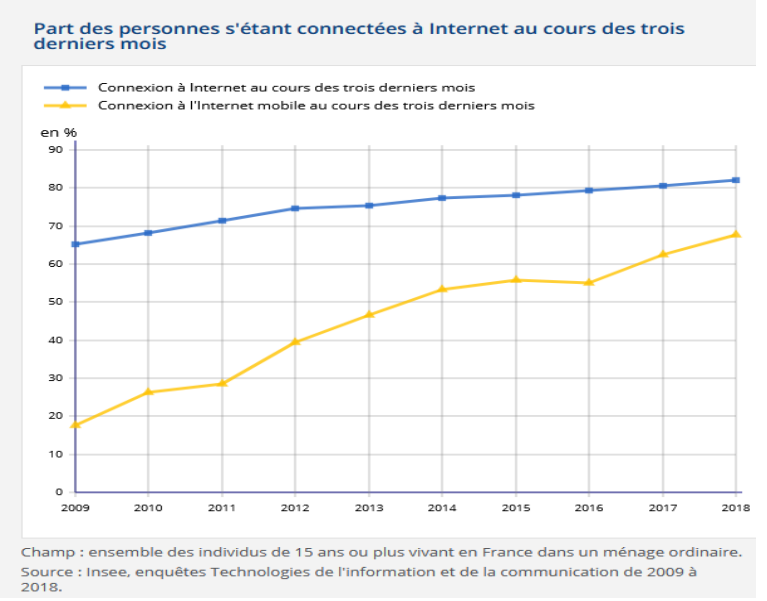
Rapid change with adoption and abandon

- ARPANET (1972)
  - Beginning of electronic communication: big firms, universities, emails
- Beginning 1990: internet
  - Fixed web sites.
  - Visiophone conversation
- Beginning of year 2000: Web 2.0
  - Websites become easy to create modify without coding knowledge (CMS)
  - Blogs (Declined somehow)
  - Social networks: Facebook 2004, Twitter
  - Multiple online activity: housing, furnishing, curing, gaming, dating, knowledge (wikipedia) etc.
- 2007: First iphone. 2009. Samsung Galaxy.. 2010s: Generalization of smartphone.
- Rise & fall of platforms: Altavista, ebazar, MySpace, Second Life, MSN, Google Plus, (Facebook, Twitter ???)…

Internet : fast and contrasted diffusion

- Berret Pierre, 2008, « Diffusion et utilisation des TIC en France et en Europe » , *Culture Chiffres*.
- Donnat Olivier, 2007, « Pratiques culturelles et usages d'internet », *Culture Etudes*.
- Gombault Vincent, 2011,« Deux ménages sur trois disposent d'internet chez eux », *Insee Première*
- Legleye, Stéphane, and Annaïck Rolland. "Une personne sur six n'utilise pas Internet, plus d'un usager sur trois manque de compétences numériques de base." (2019).

- France: Internet at home
  - 2000: 12% (Gombault 2011)
  - 2006: 41% (Berret 2008)
  - 2007: 49% (Berret 2008)
  - 2008: 54% (Gombault 2011)



1 Équipement, usage d'Internet et capacité numérique

en %

	Pas d'équipement Internet	Non-usage d'Internet dans l'année	Au moins une incapacité*	Illectronisme
Ensemble	**12,0**	**15,2**	**47,3**	**16,5**
15-29 ans	2,3	2,8	19,2	3,0
30-44 ans	3,6	2,8	33,6	3,2
45-59 ans	7,0	7,8	47,8	9,3
60-74 ans	15,4	24,1	66,2	26,7
75 ans ou plus	53,2	64,2	90,0	67,2
Femmes	13,1	16,2	49,0	17,4
Hommes	10,9	14,0	45,4	15,5
Études supérieures	2,5	3,0	18,3	3,5
Bac ou équivalent	3,9	4,5	33,2	5,1
CAP, BEP ou BEPC	9,7	14,0	56,0	15,5
Aucun diplôme ou CEP	34,1	41,0	84,2	43,9
En emploi	3,4	3,7	35,4	4,4
Étudiant, apprenti	2,1	2,4	12,9	2,4
Chômeur	10,6	8,1	42,4	8,8
Retraité	28,6	38,1	74,6	40,8
Autre inactif	17,9	22,9	69,0	25,8
5e quintile de niveau de vie (ménages aisés)	3,5	6,5	28,2	7,2
1er quintile (ménages modestes)	15,7	18,2	57,7	19,8
Couple avec au moins un enfant	1,8	3,0	30,2	3,1
Personne seule	29,5	30,3	64,8	32,4
Famille monoparentale	3,5	2,4	38,2	3,2
Couple sans enfant	14,0	20,1	55,4	22,0
Autre type de ménage (ménage complexe)	4,7	8,2	37,7	9,5
Agglomération de Paris	8,2	11,9	40,9	13,5
Unité urbaine de 10 000 à 1 999 999 habitants	12,2	14,7	45,8	16,0
Commune rurale et unité urbaine de moins de 10 000 habitants	13,2	16,7	50,7	18,1
Métropole	11,8	14,9	46,9	16,3
DOM hors Mayotte	19,4	24,4	60,1	26,3

* Information, communication, logiciel ou résolution de problème.
Lecture : en 2019, 2,3 % des 15-29 ans n'ont pas d'équipement Internet à domicile.
Champ : individus de 15 ans ou plus, France hors Mayotte.
Source : TIC ménages 2019.

A practice socially differentiated (Legleye Rolland, 2019)

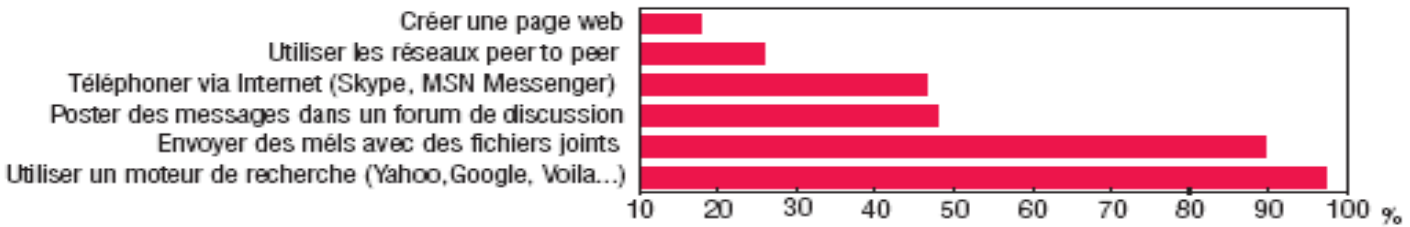
Digital knowledge matters

❷ Les raisons pour ne pas avoir Internet

	Nombre (milliers)		Raisons[1] (%)	
	2008	2010	2008	2010
A un accès dans d'autres endroits	1 765	960	15,3	9,7
Ne veut pas d'Internet	4 365	1 075	37,7	10,8
Internet est inutile	5 705	2 910	49,2	29,3
Le coût de l'ordinateur est trop élevé	4 325	3 010	37,3	30,3
Le coût de l'abonnement est trop élevé	3 900	2 920	33,7	29,3
Pas assez compétent	5 025	4 365	43,4	43,9
Par souci de sécurité ou de protection de la vie privée	1 490	1 395	12,9	14,0
Se déclare inapte	895	590	7,7	5,9
Autre raison	2 070	2 665	17,8	26,8

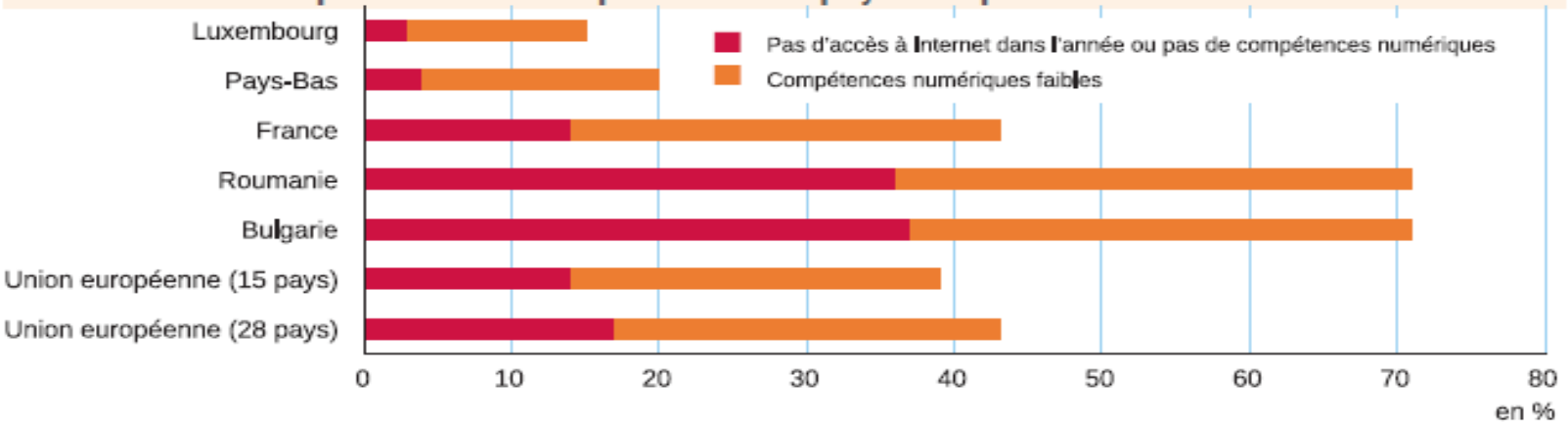
1. Plusieurs réponses possibles ; la somme peut donc être supérieure à 100 %.
Lecture : 43,9 % des ménages n'ayant pas Internet en 2010 ne l'ont pas car ses membres ne sont pas compétents.
Champ : ménages n'ayant pas Internet à domicile (44 % en 2008 et 35,6 % en 2010).
Source : Insee, enquête Technologies de l'information et de la communication d'avril 2010.

① Compétences Internet en 2010



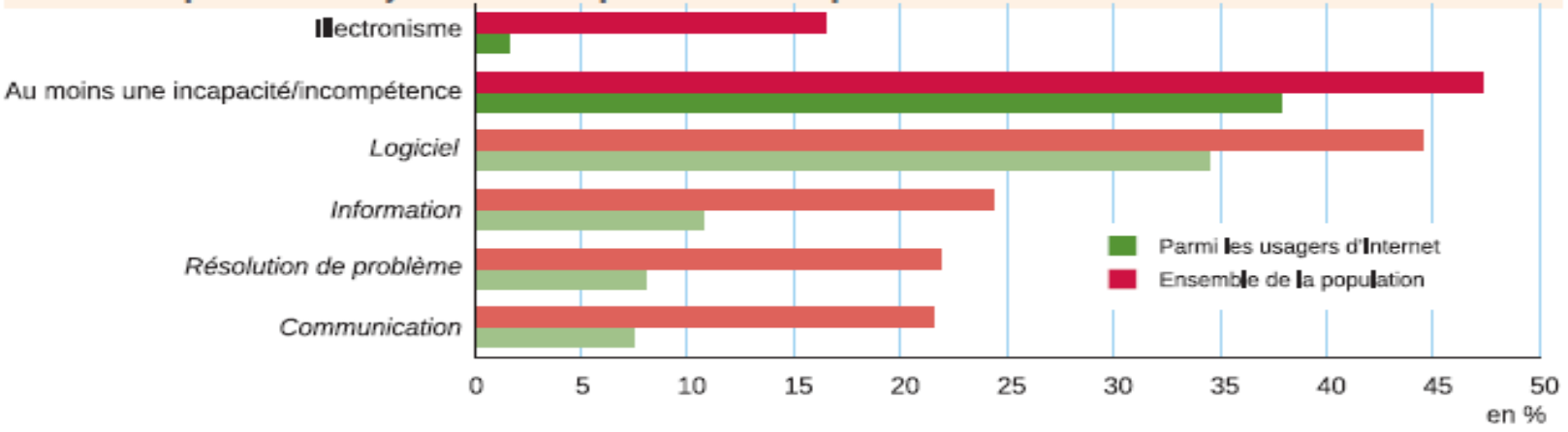
Lecture : 90 % des internautes savent envoyer un mél avec des fichiers joints.
Champ : personnes de 15 ans ou plus ayant utilisé Internet au cours des 3 derniers mois.
Source : Insee, enquête Technologies de l'information et de la communication, avril 2010.

4 Niveaux de compétences numériques dans les pays européens



Lecture : en 2017, au Luxembourg, 3 % de la population n'a pas accédé à Internet pendant l'année ou n'a pas de compétences numériques, et 12 % a des compétences numériques faibles.

3 Part de personnes ayant une incapacité numérique ou en situation d'illectronisme



Lecture : en 2019, 24 % de la population n'a pas su obtenir des informations sur Internet (11 % parmi les usagers d'Internet).
Champ : individus de 15 ans ou plus, France hors Mayotte.
Source : TIC ménages 2019.

In 2010: Emailing and searching for information as the basis of online activity

❸ Les utilisations d'Internet au cours des trois derniers mois

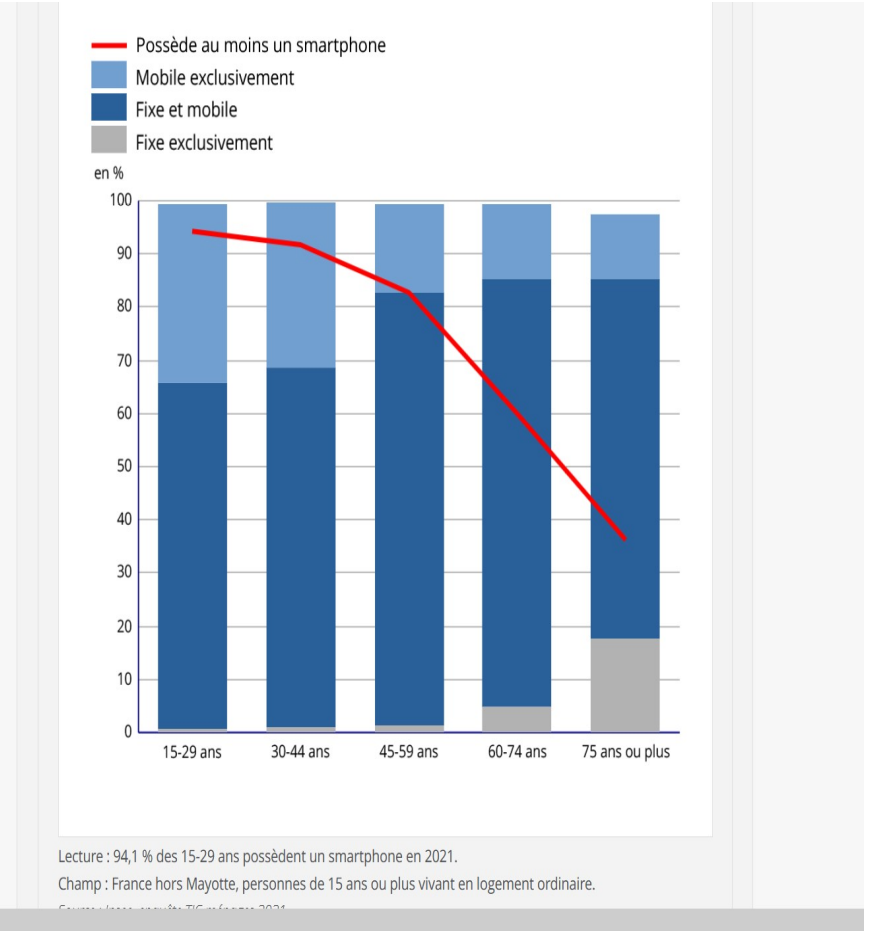
en %

	Messagerie électronique	Recherche d'informations sur des biens et services	Compte bancaire	Apprendre, enrichir ses connaissances	Vacances	Informations sur la santé	Conversations, forums de discussion	Vendre aux enchères	Lire ou télécharger journaux ou magazines	Télécharger des logiciels (autres que jeux)	Rechercher un emploi
Ensemble	**91,1**	**82,6**	**65,6**	**62,1**	**51,8**	**45,9**	**41,1**	**29,2**	**27,0**	**25,2**	**20,1**
Hommes	91,8	82,6	67,3	65,2	52,9	39,2	41,8	32,2	31,8	32,9	20,7
Femmes	90,5	82,6	63,9	59,0	50,8	52,3	40,5	26,2	22,3	17,6	19,6
15-29 ans	93,9	84,4	59,3	68,8	51,0	47,6	77,4	32,4	28,7	40,1	32,5
30-44 ans	90,8	85,4	72,7	60,3	60,1	46,2	34,1	35,0	27,0	23,1	23,0
45-59 ans	90,3	81,9	67,5	58,7	50,6	42,9	25,6	25,0	24,5	17,7	14,1
60-74 ans	89,8	76,5	62,4	58,2	40,2	48,3	15,4	20,2	29,7	15,2	0,5
75 ans et plus	73,3	54,9	34,0	65,4	18,5	42,8	4,7	3,5	22,1	7,6	1,2

Lecture : 51,8 % des internautes ont au cours des 3 derniers mois organisé leurs vacances sur Internet ; cette proportion est de 51 % chez les moins de 30 ans.
Champ : personnes de 15 ans ou plus ayant utilisé Internet au cours des 3 derniers mois.
Source : Insee, enquête Technologies de l'information et de la communication d'avril 2010.

2010's The smartphonization of online activity

- Messenging / Sharing / Liking
- Share of videos → emotional
- Fake truth diffusion
- Populism (?)
- Cf. Smartphone in France 2021

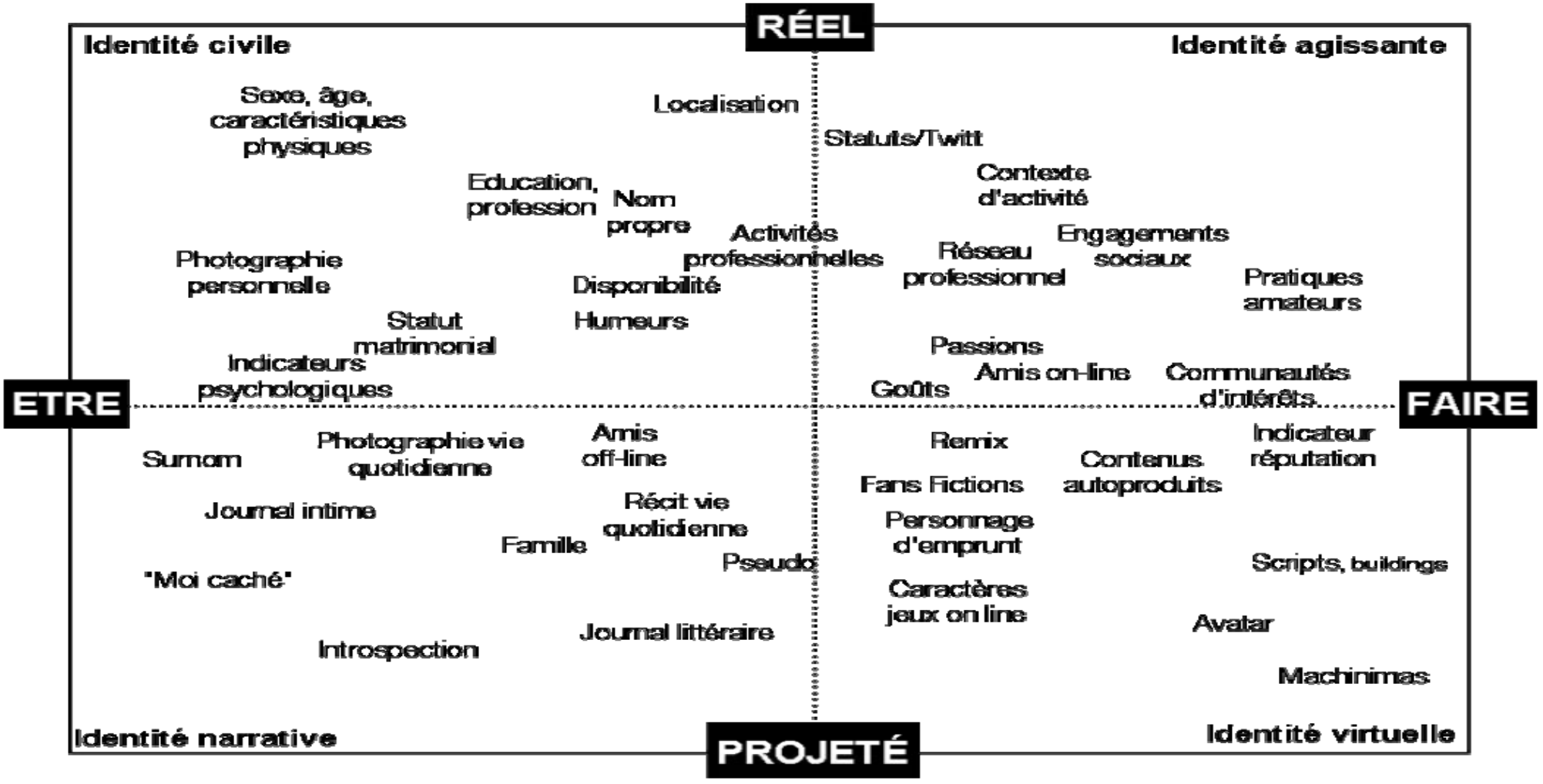


Importance and diversity of the type of formats

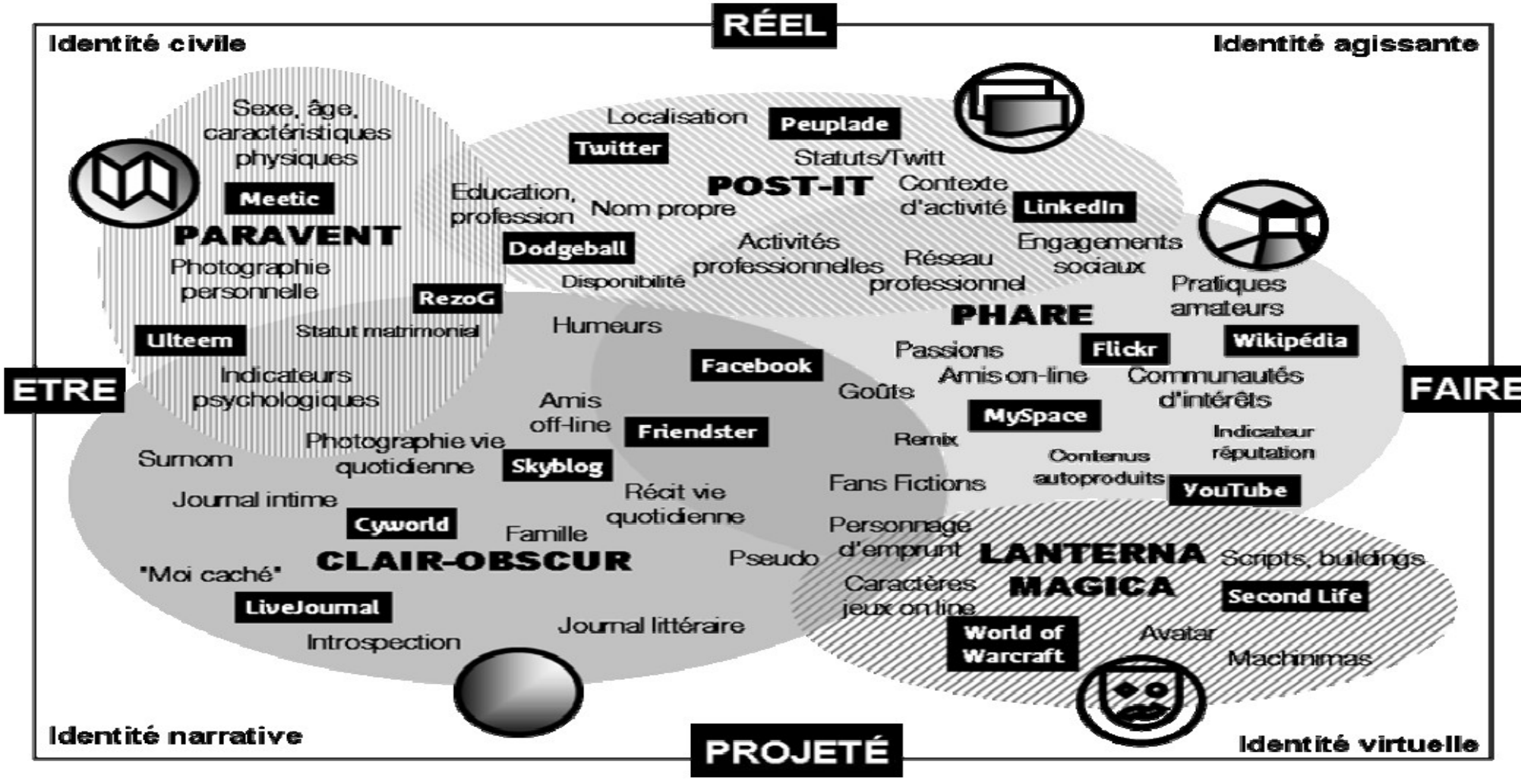
Cardon, Dominique. 2008. "Le design de la visibilité", *Réseaux*, 152 (6): 93-137.

- Web 2.0 qualitative change enabling non-experts to produce/share information on internet
- Panorama of Web 2.0 platforms
- Centering on the question of the production of identities

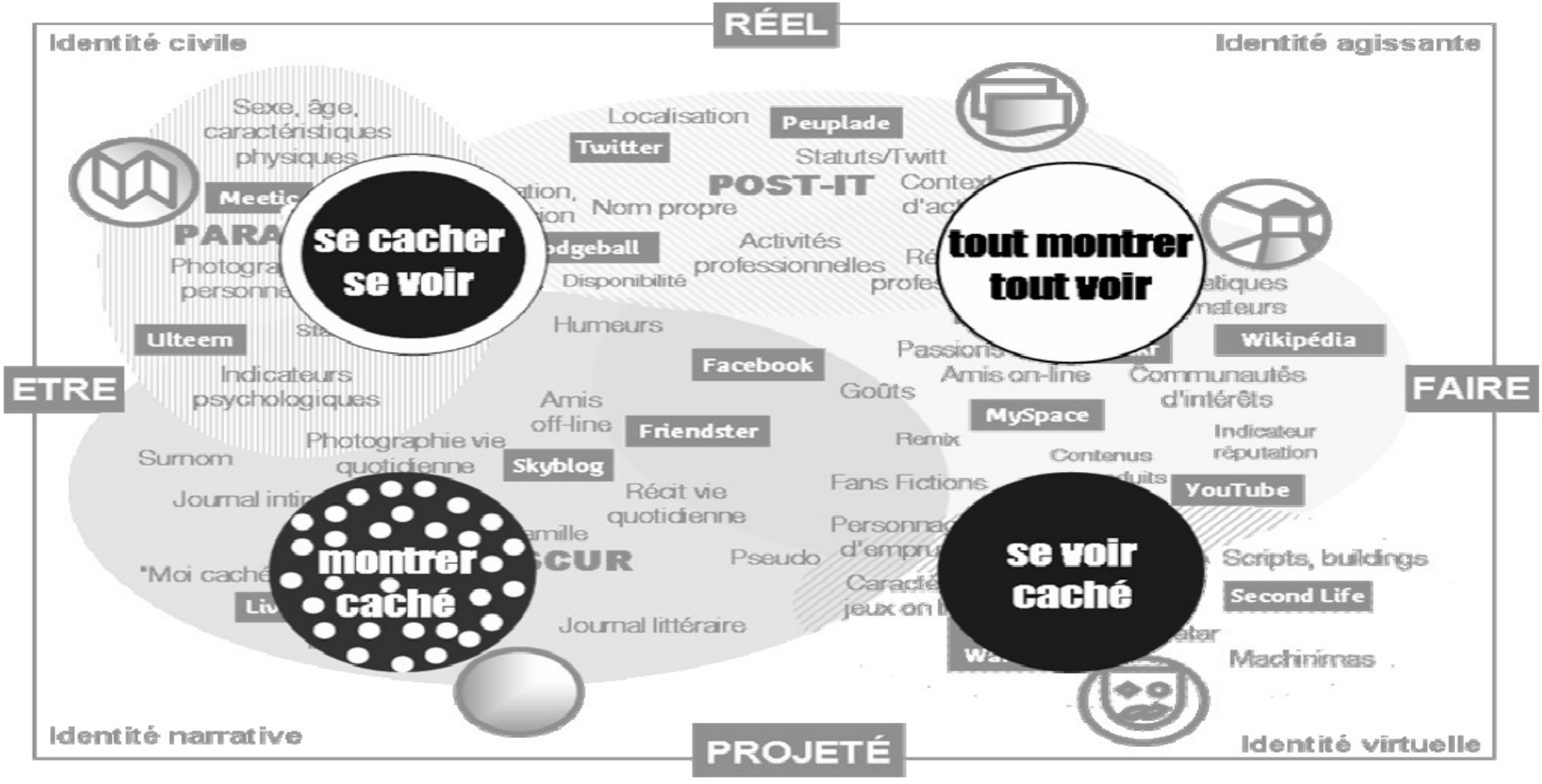
Carte 1. Cartographie des traits identitaires projetés vers les plateformes du web 2.0



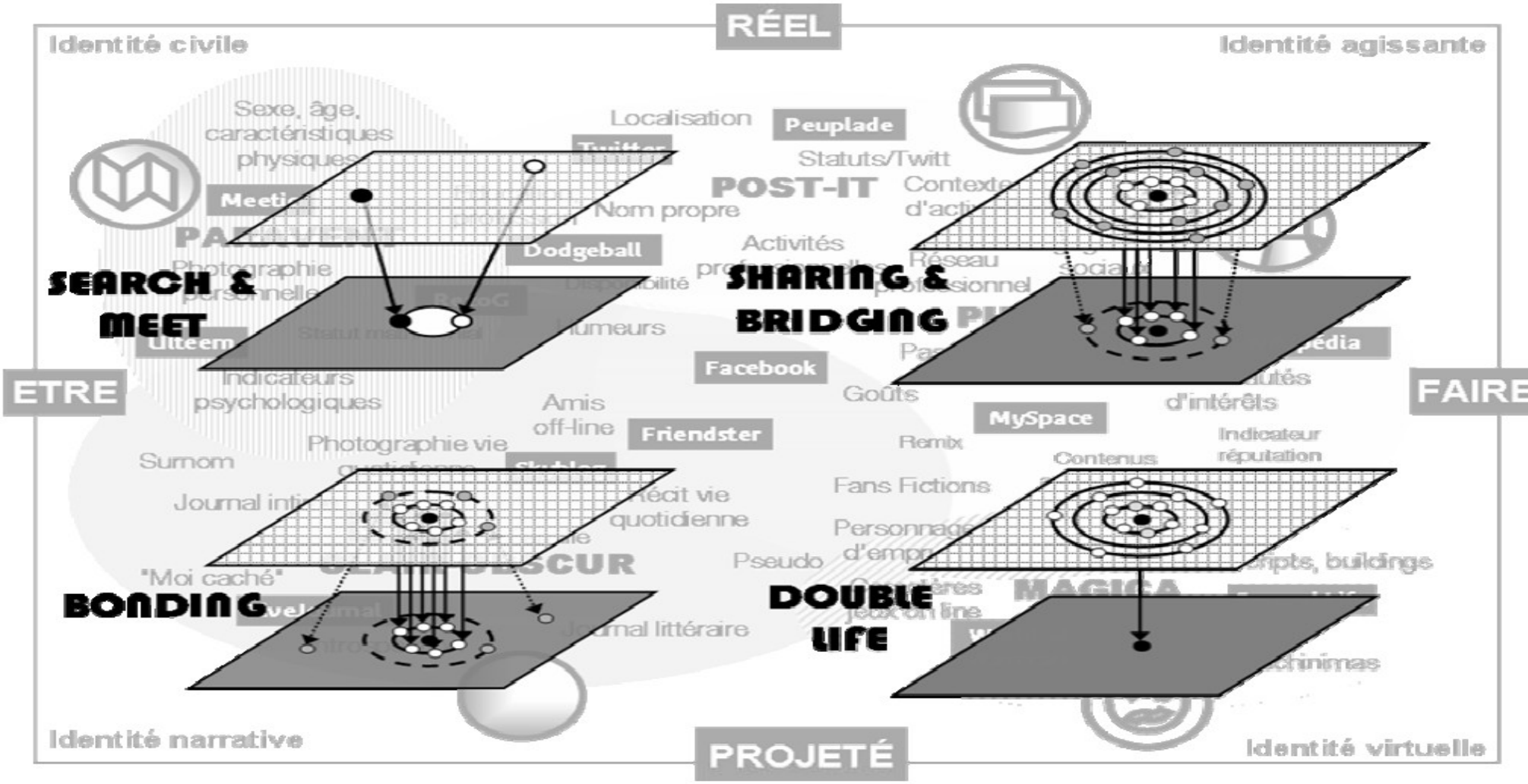
Carte 2. Typologie de la visibilité de l'identité sur les plateformes du web 2.0



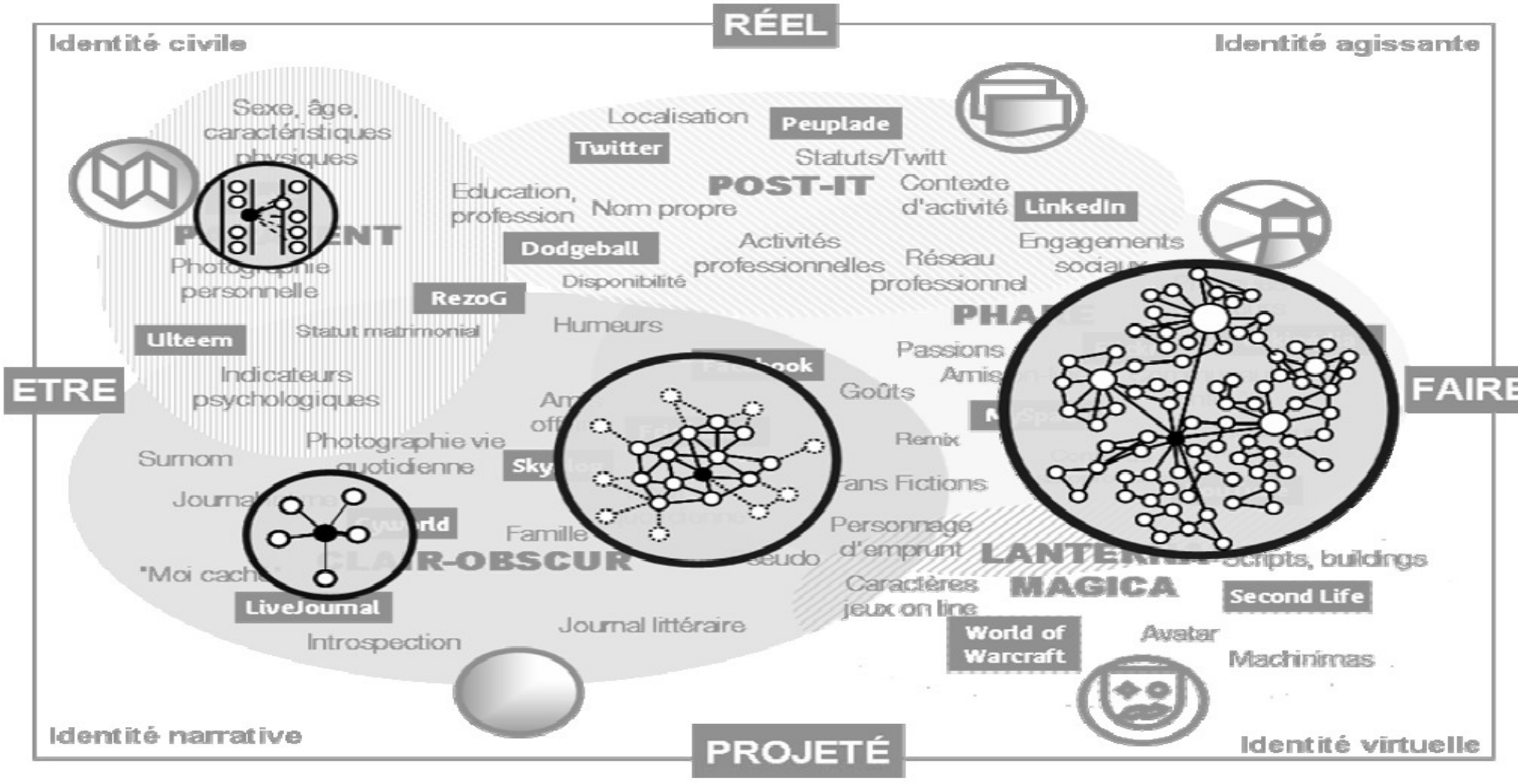
Carte 3. Visibilité/invisibilité

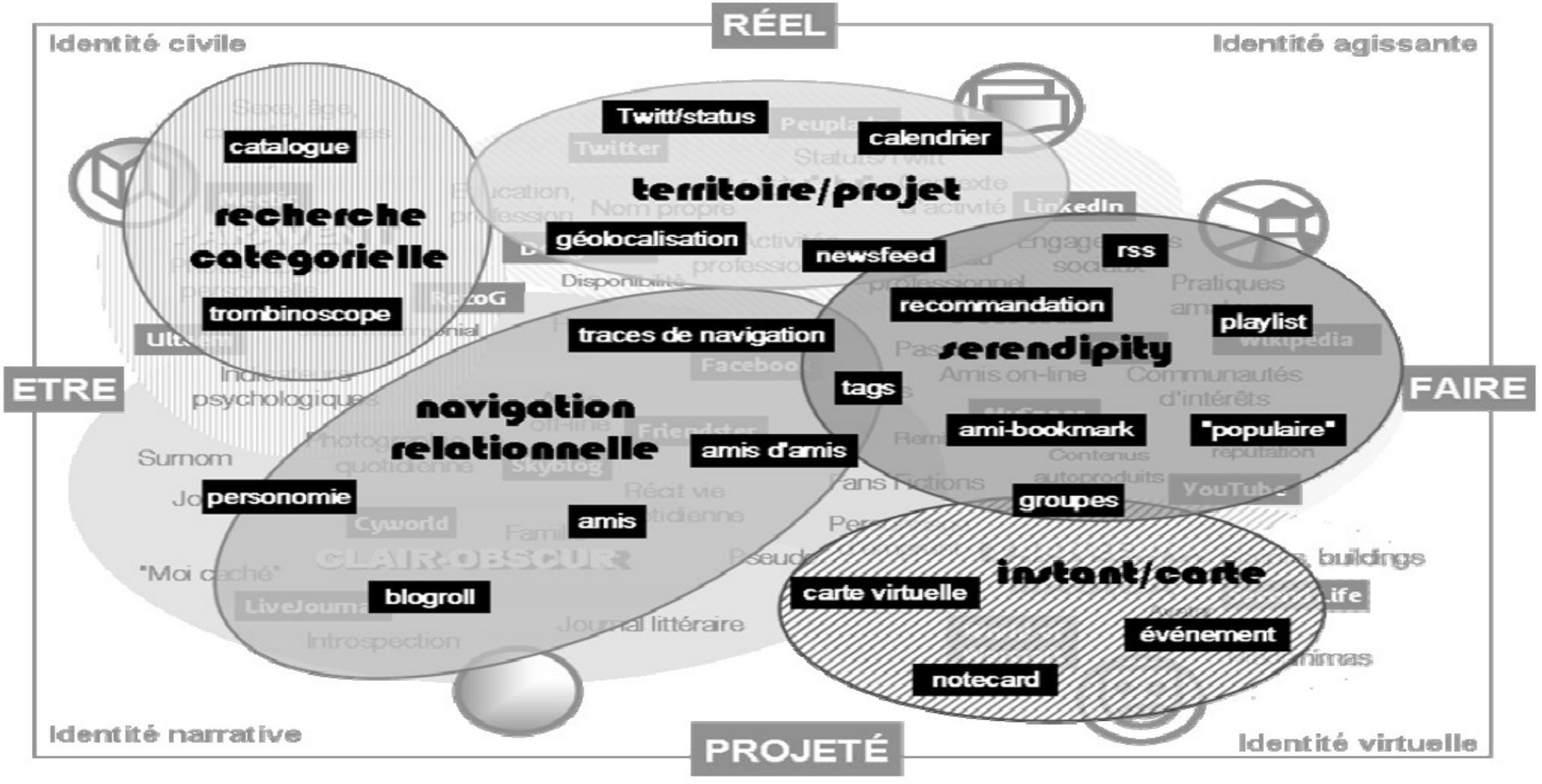


Carte 4. Monde réel/monde numérique



Carte 5. Taille des réseaux




RÉEL
Identité civile
Identité agissante
catalogue
recherche categorielle
trombinoscope
Twitt/status
calendrier
territoire/projet
géolocalisation
newsfeed
rss
recommandation
playlist
serendipity
tags
ami-bookmark
"populaire"
groupes
traces de navigation
navigation relationnelle
amis d'amis
personomie
amis
blogroll
instant/carte
carte virtuelle
événement
notecard
ETRE
FAIRE
PROJETÉ
Identité narrative
Identité virtuelle

Elison, Steinfield, Lampe, 2007, « The benefits of Facebook friends », *Journal of computer-mediated communication*

- Survey on 800 students at Michigan State University in April 2006
- Three forms of social capital: bonding, bridging and

Table 3 Summary statistics for Facebook use for prior contacts and meeting new people

Individual Items and Scales[1]	Mean	S.D.
Off to Online: Use Facebook to connect with offline contacts (Cronbach's alpha = 0.70)	**3.64**	**0.79**
I have used Facebook to check out someone I met socially	3.99	1.05
I use Facebook to learn more about other people in my classes	3.26	1.20
I use Facebook to learn more about other people living near me	2.86	1.22
I use Facebook to keep in touch with my old friends	4.42	0.86
On to Offline: I use Facebook to meet new people (single item measure)	**1.97**	**1.03**

Note: [1]Individual items ranged from 1 = strongly disagree to 5 = strongly agree, scales constructed by taking mean of items.

Facebook and social capital

- Independent measure of social capital and studies the correlation between Facebook use and intensity of social capital.
- A benefit of using Facebook for the 3 forms of social capital
- This benefit is greater for individuals with a low self-estime.

Conducting 2.0 inquiries

Collecting: Web scrapping

- Possibility to collect great amount of data at a relatively minimum cost
- How?
  - Personal ex: SuperMacro → Research on EHESS
  - Example of web scrapping: Package xml R.
  - Depends on the structure of the web site (API: Application Programming Interface)
  - A little bit (or enormous bit) of coding

Type of inquiry: qualitative …

- Internet contains text and meaning➔ qualitative approach

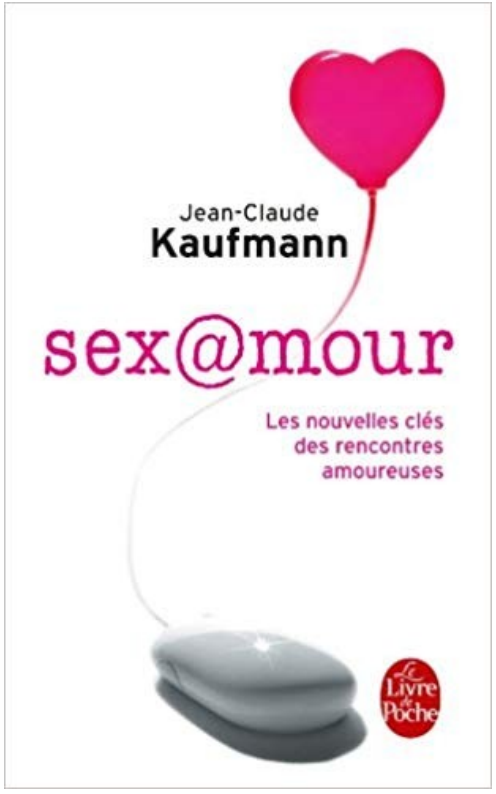
Rosental, Claude. 2003. "Certifying knowledge: the sociology of a logical theorem in artificial intelligence." *American sociological review*: 623-644.

- Studies the reception by a community of mathematician of Elkan's theorem on fuzzy logic, and its equivalence with classical logic
- Online forum : comp.ai.fuzzy
- Shows the progressive stabilization of the interpretation of a controversial theorem

This doesn't make sense to me. I cannot imagine a scenario in which I was uncertain about the truth of (B or not(B)). (H. Lucke, message 813)

You're still thinking Boolean. . . . With fuzzy logic (if B or not B) has much more meaning than in Boolean (it's meaningless in Boolean). (Y. Tanaka, message 814)

Limits of a qualitative approach: people expressing themselves on internet are special

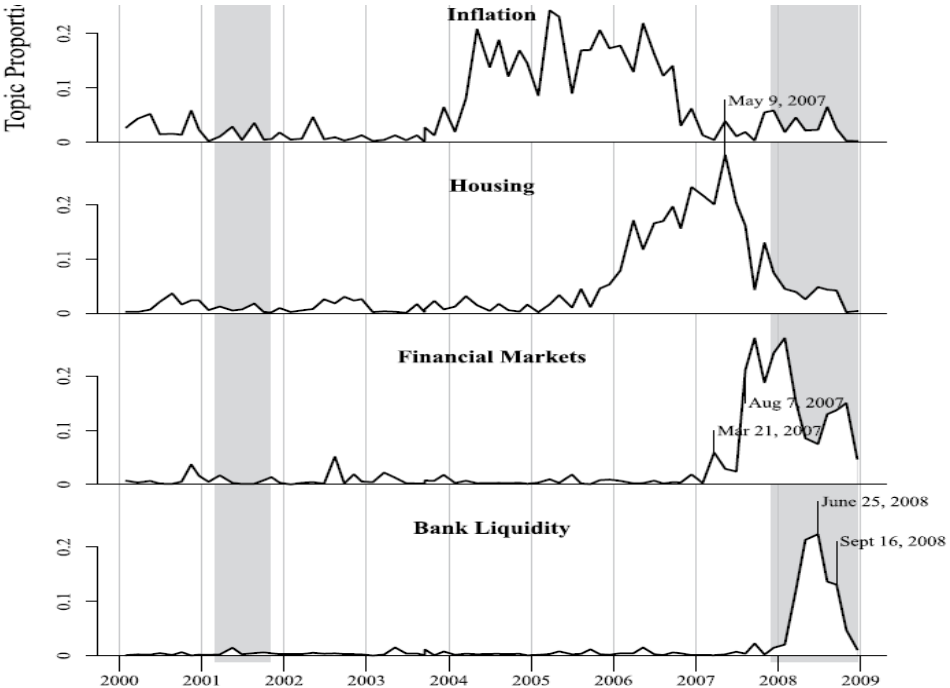


- Kaufman – Dating online
- Rely only on forums and on blogs
- Especially on one blog
  - Very special type of user of who turns dating experience into material for exhibitionist "autofiction"

Quantifying of qualitative data

- Text mining
  - Pure statistical approach
    - Classical text mining
  - Natural Language Processing (NLP)
    - Topic modeling
    - Artificial intelligence techniques help to define topics
- Example
  - Fligstein, Brundage, and Schultz. 2017. "Seeing like the Fed" *American Sociological Review* 82(5): 879-909.
    - Topic modeling on the minute of the FED. Collected online.

      ➔ slow to react to Global Financial Crisis. Over obsessed with inflation



1b. Topic Proportions over Time; Framing Topics

Mostly quantitative data

- Virtue of 2.0. Size of the data downloadable
- Datasets online
  - Ex. IMDB
- Traces of social networks
- Example from students
  - Huang, Mulun. 2018. *Automne - Hiver 2017-18 : une enquête sociologique du marché de la mode,*
  - DHUOT Raphaël, 2013, *Prix et temporalité sur une place marchande en ligne*

Promise and limits of artificial intelligence

- Typical artificial intelligence technique
  - 1 Million of pictures tagged “cat”
  - Give to an algorithm. (Random forest)
  - Learn to recognize cats of 10 billion pictures
- Transposition in social science
  - Train an algorithm to code on a subsample
  - Let the algorithm code it on a larger sample

Limits

- Limit 1.
  - Subsample for training needs to be very large
  - Initial coding cost
  - Solution: use already existing coding algorithm (positive/negative text)
- Limit 2.
  - Prediction based on A.I. is not necessarily better than prediction based on classical techniques (Boalert, Ollion, 2018)
- Limit 3.
  - A.I. is about predicting and not about explaining.
  - Social Sciences is about explaining

Examples

Using the web as a source for studying social questions.

Ex. 1. Facebook friends: an inquiry

Wimmer, Andreas, and Kevin Lewis. 2010. "Beyond and below racial homophily: ERG models of a friendship network documented on Facebook." *American Journal of Sociology* 116(2): 583-642

Research question

- How can we explain racial homogeneity?
  - Racial homophily could be due to ethnic homophily
- Is it due to racial homophily?
  - Or to other factors
    - Availability
    - Propinquity
    - Shared interests
    - Balancing (network effects)

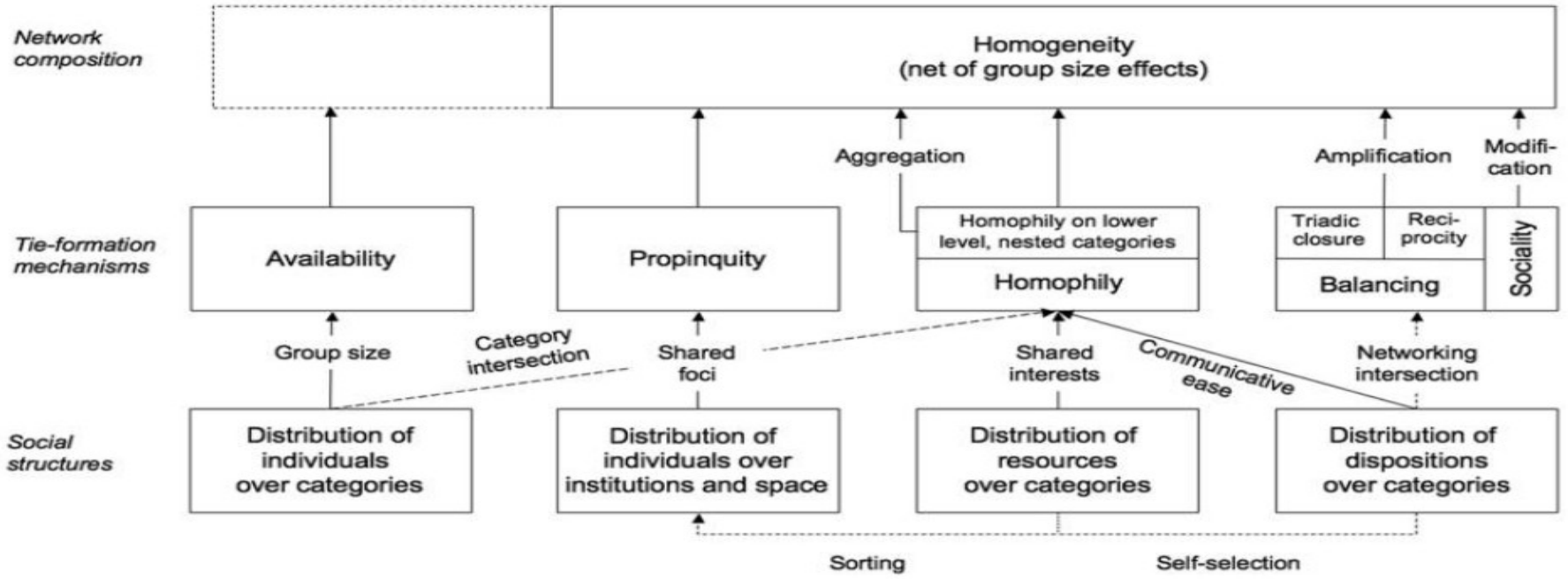


FIG. 1.—Social structures, tie formation mechanisms, and network composition. Indirect effects are represented with dashed lines.

Method

- Population: 1640 first year students in a university.
  - 97% have a Facebook account.
  - Links only within the university.
- Friendship links: tagged friends on a Facebook picture.
  - Stronger tie than Facebook tie
  - Caveat: 45% of Facebook users post pictures of their friends.
- Statistical method: ERGM Exponential Random Graph model

Variables

- Gender
- Cultural tastes
- Social origin
  - Elite/Non elite high school
- Ethnic origin through several proxies
  - Facebook profile picture
  - Surname
  - Facebook ethnic groups

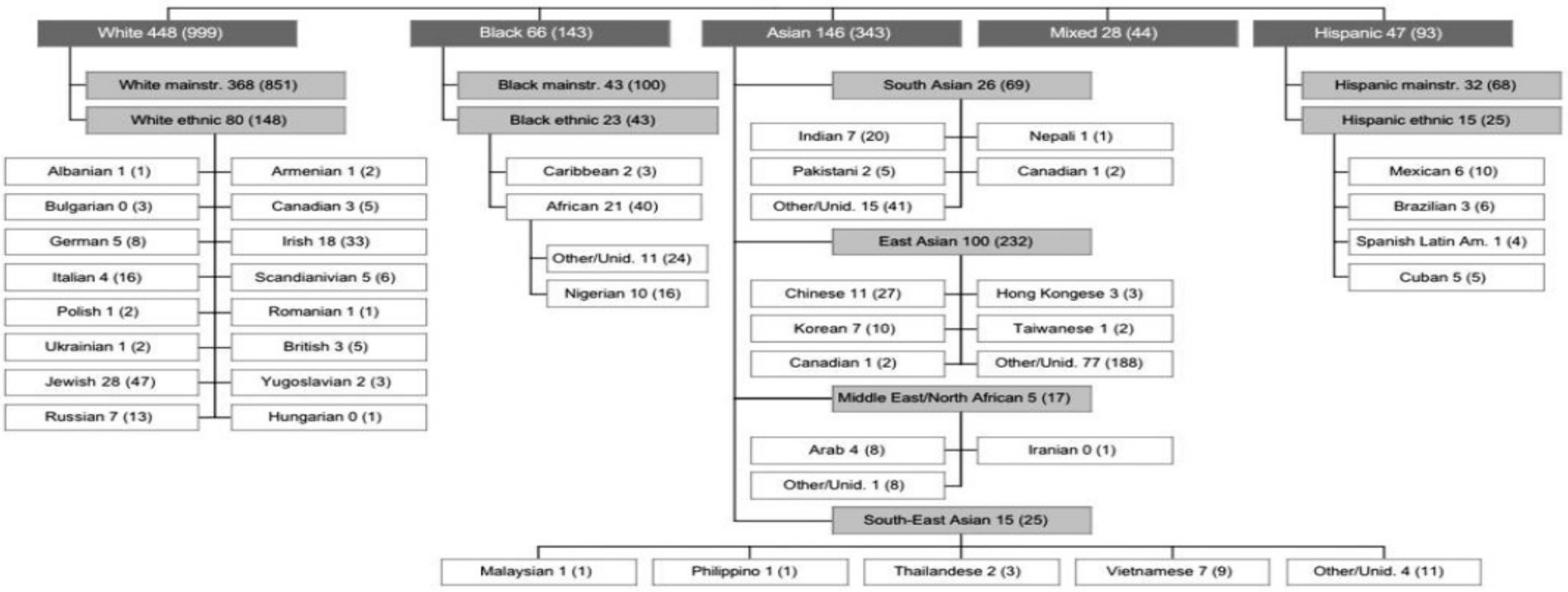


FIG. 2.—A multilevel, nested scheme of ethnoracial classification. Dark grey boxes contain racial categories, light grey boxes ethnic categories, and white boxes "microethnic " (or country/countries-of-origin) terms. Numbers immediately following category names refer to the quantity of individuals in that category who posted pictures ($N = 736$); those in parentheses refer to the composition of the entire population ($N = 1{,}640$). Seventeen students whose race/ethnicity could not be determined and one self-identified Native American are omitted from this diagram.

Models

- Model 1: Racial homophily
- Model 2: Racial homophily + Micro-ethnoracial homophily
- Model 3: Racial homophily + Reciprocity and triadic closure
- Model 4: Racial homophily + racialized sociality for minorities.
- Model 5: Racial homophily + intersection of racial categories and individual characteristics
- Model 6: combination of all the significant mechanisms of model 1-5

TABLE 1
DECOMPOSING RACIAL HOMOGENEITY

TERMS	MODEL 1	2	3	4	5
Edges	−4.82***	−4.82***	−5.96***	−4.91***	−4.85***
	(.02)	(.02)	(.02)	(.03)	(.02)
Racial homophily:					
Whites	.37***	.29***	.25***	.46***	.37***
	(.03)	(.04)	(.03)	(.04)	(.03)
Blacks	2.11***	1.97***	1.14***	2.41***	2.04***
	(.07)	(.10)	(.06)	(.09)	(.07)
Asians	1.01***	.50***	.73***	.96***	.98***
	(.05)	(.09)	(.03)	(.06)	(.05)
Mixed	.85**	.85**	.16	.38	.83**
	(.27)	(.27)	(.64)	(.28)	(.27)
Hispanics	1.50***	1.51***	1.07***	1.32***	1.48***
	(.12)	(.18)	(.09)	(.13)	(.12)
Ethnic homophily:					
Mainstream whites		.10*			

A COMPREHENSIVE MODEL OF TIE FORMATION

	MODEL 6	
	Coefficient	SE
Edges	−4.59***	.03
Racial homophily:		
Whites	.22***	.04
Blacks	1.02***	.05
Asians	.27*	.12
Hispanics	.79***	.21
Ethnic and microethnic homophily:		

Findings

- Various effects (propinquity, shared preferences, balance, etc.) affect differently different groups
- Ethnic homophily can create racial homophily through aggregation ex Asians. Contra-example: Hispanics
- Balance effects (reciprocity and triadic closure) amplify racial homophily

Interpretation

- Explanation of racial or ethnic homophily
  - Experience of discrimination reinforces feeling of solidarity. Ex: proximity African-Americans and Caribbeans. Similar perception by others.
  - Shared language and cultural disposition.
    - ex: Hispanics vs Asians not the same language
  - Closure strategy of white students towards minority

Limits of Wimmer & Lewis (& Lin & Lundquist)

- Limit of looking for a "pure" net ethnic/racial preference/homophily
- What is race? Bundle of many elements
  - Some elements could be a subcomponent of the notion of "race"
    - Phenomenal characteristics - Shape of the face
  - Or highly correlated to it
- Is it possible to disentangle race from Religion (Jews), from culture (music), language…

Ex.2. Dating online

- Lin, Ken-Hou, and Jennifer Lundquist. 2013. "Mate selection in cyberspace: The intersection of race, gender, and education." *American Journal of Sociology* 119 (1): 183-215.
- Studying mating.
- Race barrier.
- Disentangling propinquity and preferences

Data

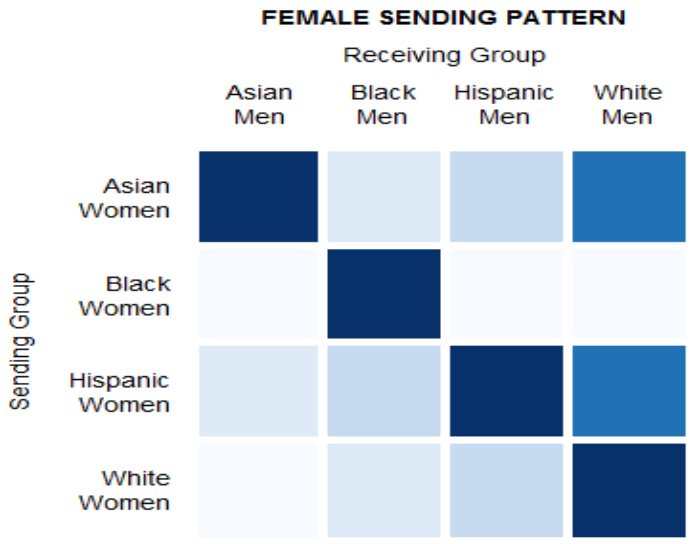
- Largest dating online website
- Selection: Active users in 20 Largest metropolitan areas in the US
- 529 000 straight men, 405 000 straight women
- Ethnicity self declared

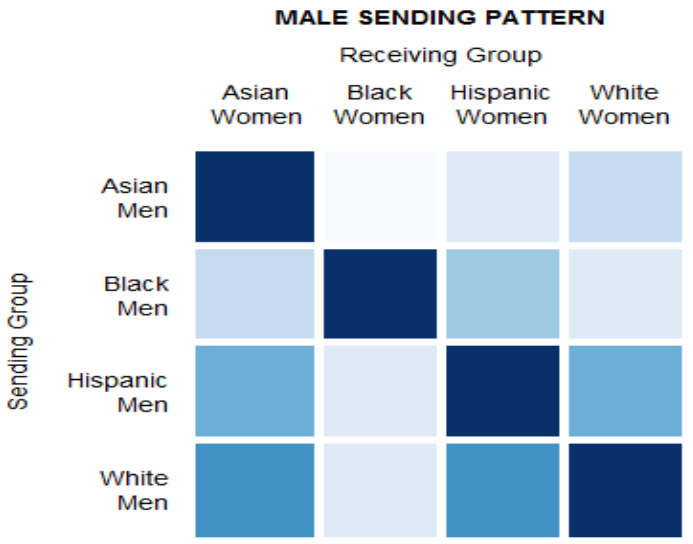
Race	Men (%)	Women (%)
Asian	3.28	3.69
Middle Eastern	.36	.26
Black	3.35	3.86
Native American	.23	.26
Indian	.81	.47
Pacific Islander	.25	.24
Hispanic	4.97	4.66
White	52.05	53.67
Other	1.47	1.48
Undeclared	25.24	24.36
Multiracial	7.99	7.05
N	528,800	405,021

Probability of messages

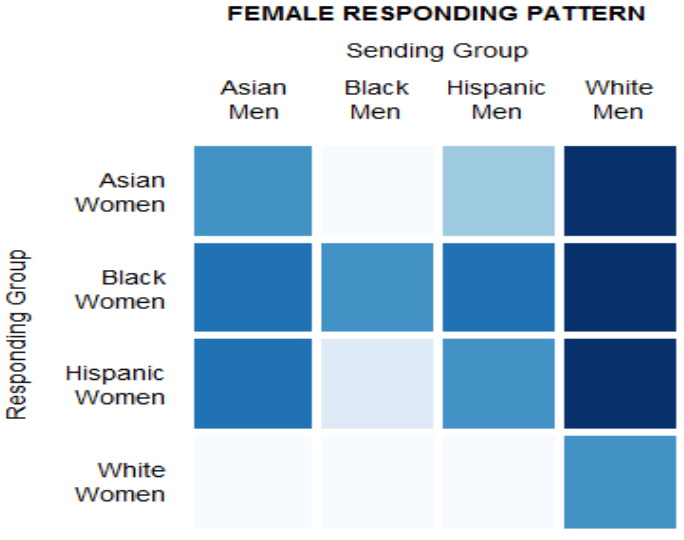
- Sending a message
  - Reconstitution of a counterfactual group  people to which message is not sent
  - All potential matches in the age tranche and within the same metropolitan area.
- Responding
  - Yes/No, conditional to a first message

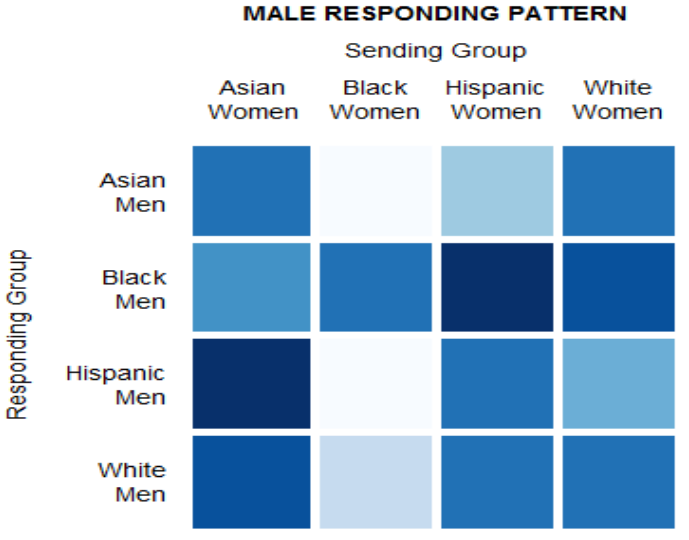
Sending patter governed by racial homophily





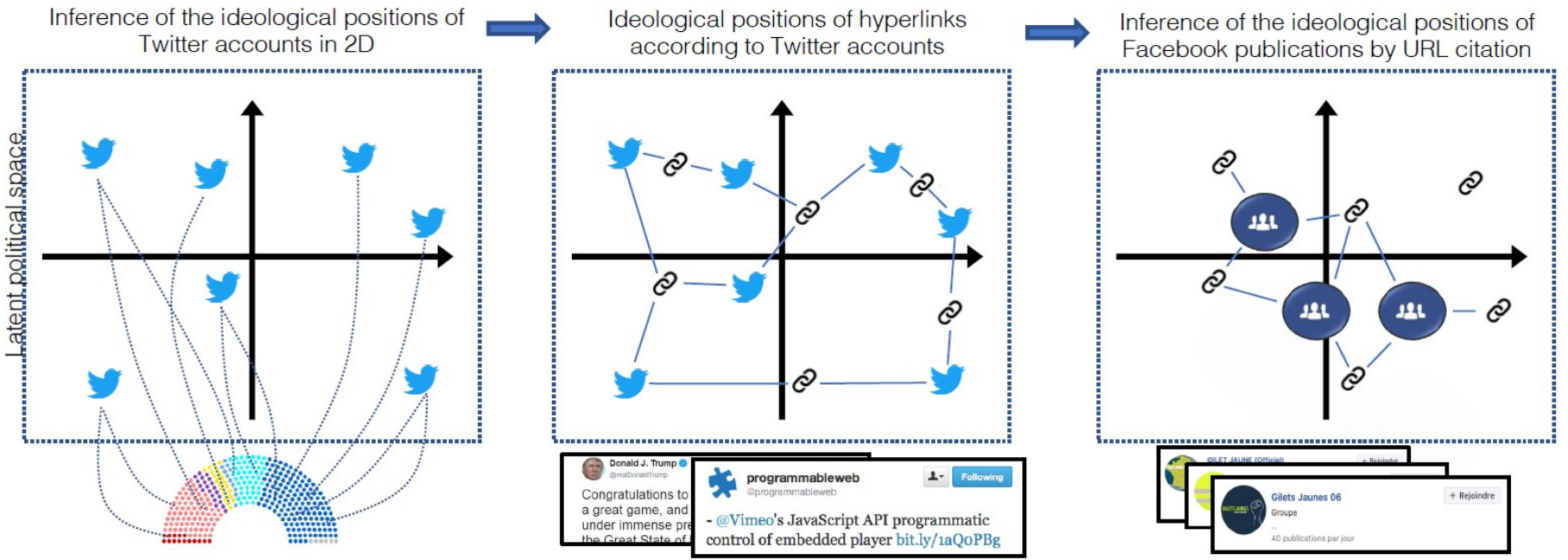
Answering pattern governed by racial hierarchy





Ex. 3 Ideological color of the yellow vest

- Cointet, Jean-Philippe, et al. 2021. "What colours are the yellow vests? an ideological scaling of facebook groups." Statistique et Société 9 (1-2): 79-107.
- Three levels
  - Ideological space of French Mps on twitter
    - 883 Twitter accounts held by French MPs(545) and senators (338).
    - Who follow who ?
  - Informational space of French Mps
    - URL sharing by MPS
  - Informational space of Yellow Vest Facebook groups
    - 900 facebook groups


Inference of the ideological positions of Twitter accounts in 2D
Ideological positions of hyperlinks according to Twitter accounts
Inference of the ideological positions of Facebook publications by URL citation
Latent political space
Donald J. Trump
@realDonaldTrump
Congratulations to
a great game, and
under immense pre
the Great State of
programmableweb
@programmableweb
Following
- @Vimeo's JavaScript API programmatic control of embedded player bit.ly/1aQoPBg
Gilets Jaunes 06
Groupe
40 publications par jour
+ Rejoindre

MPs versus Yellow Vest

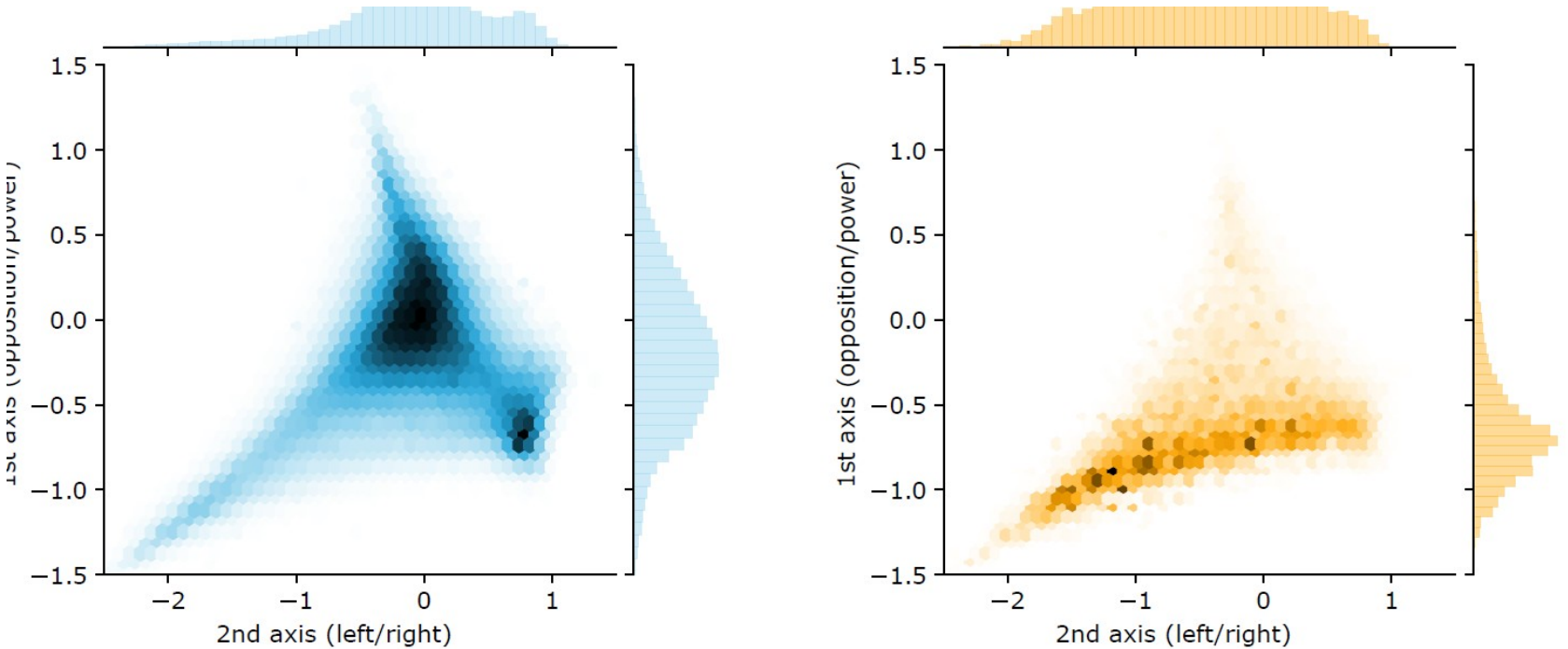


FIGURE 10 – Distribution of the ideological space occupied by URLs shared on Twitter (40 000 random accounts) [left], and URLs shared on Yellow Vests Facebook groups [right]

Shifting left?