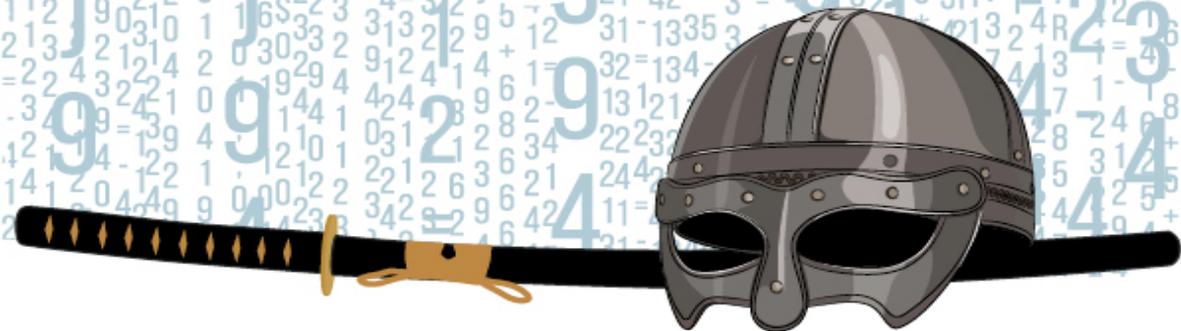


**SAGEO**  
Spatial Analysis and GEomatics 2017

DU 6 AU 9 NOVEMBRE 2017



**Lundi 6 novembre 2017**



# **Les réseaux numériques : un tournant pour l'analyse des relations hommes-milieux ?**

## **Application au Parc national des calanques**

**Siqi FAN, Philippe DEBOUDT, Amel FRAISSE, Eric KERGOSIEN**

<http://renurohm.univ-lille.fr>

**Gérialico**

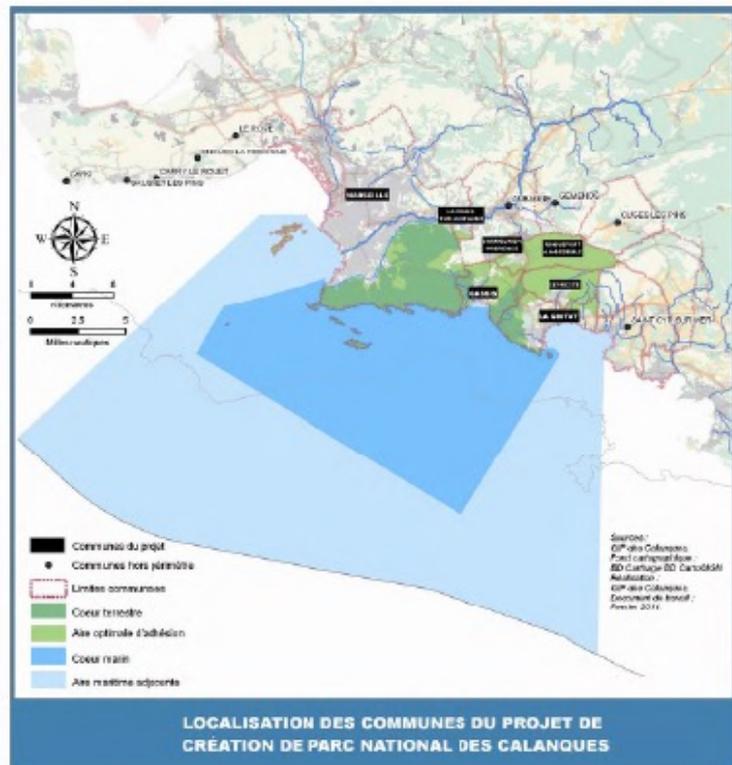


# 2. LE TERRAIN ET RECHERCHES ANTERIEURES SUR LE TERRITOIRE

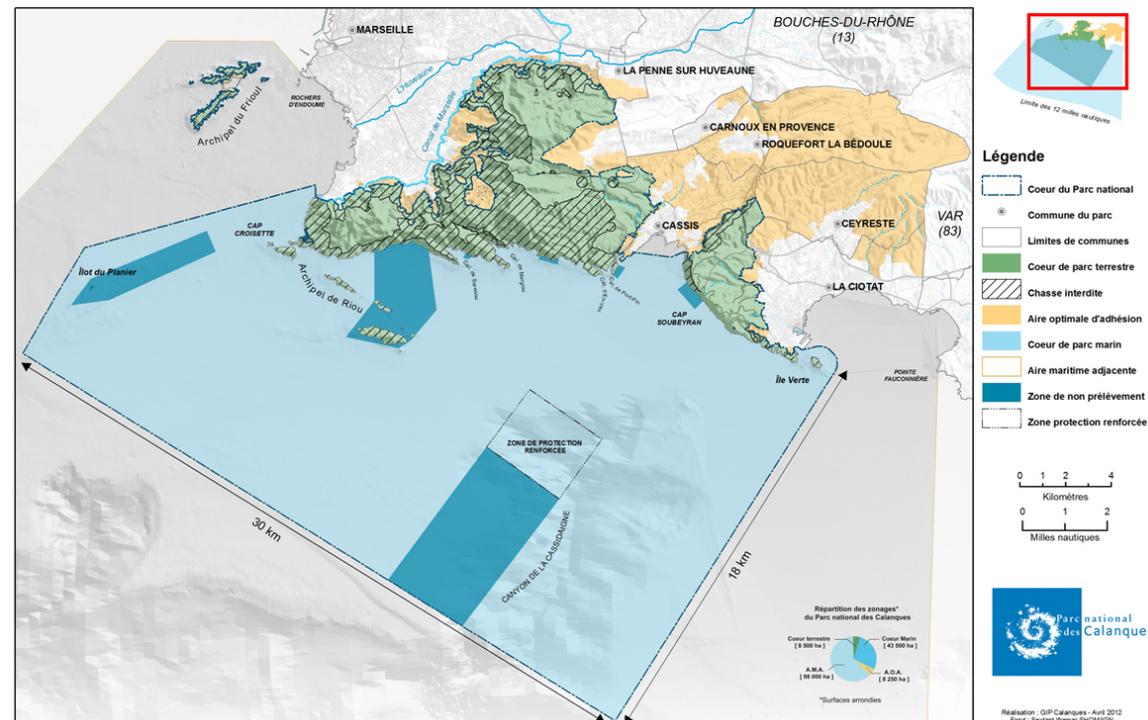
Enjeux maritimes très circonscrits dans la concertation :

*préservation d'espèces emblématiques, prélèvements de ressources et pollutions liées à la navigation ou d'origine terrestre.*

- Questions accès à la mer des usagers des calanques ou celle des usages socio-économiques de la mer (exception de la pêche professionnelle) peu débattues.



Périmètres du Parc national des Calanques



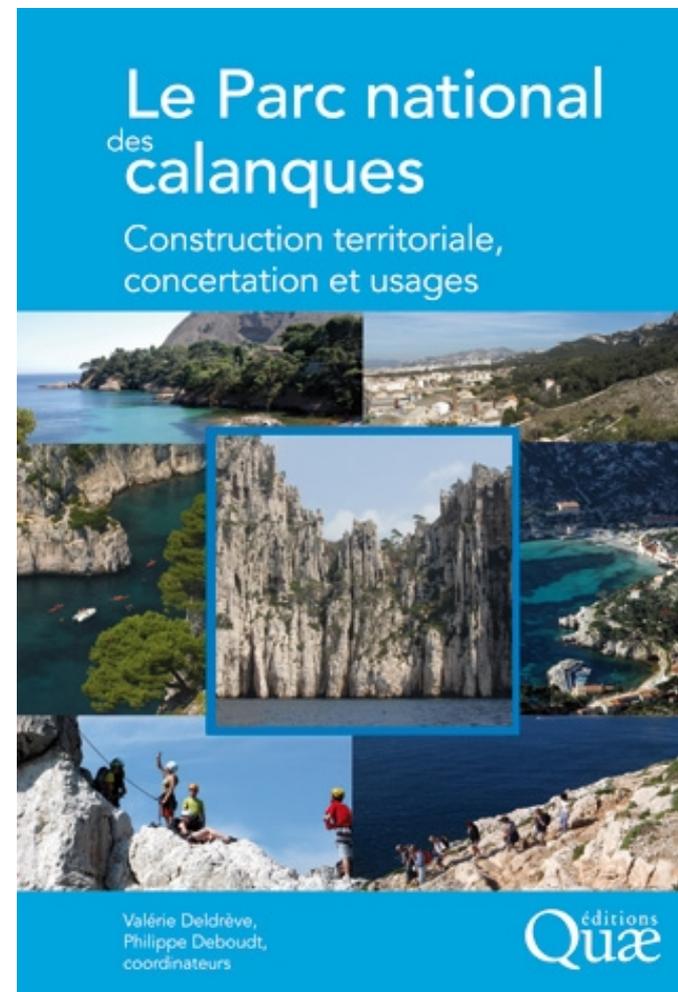
## 2. LE TERRAIN ET RECHERCHES ANTERIEURES SUR LE TERRITOIRE

De 2008 à 2015 :

- Recherches interdisciplinaires en sciences sociales (sociologie, géographie, aménagement-urbanisme) pour analyser le processus de construction territoriale du Parc national des Calanques de Marseille-Cassis-La Ciotat ;
- Concertation mise en œuvre pour réaliser la charte du parc national (Deldrève et Deboudt, 2012).  
→ difficulté d'articuler les enjeux globaux et locaux dans la construction d'un tel projet de territoire qui a principalement bénéficié aux usagers traditionnels et s'est accompagnée d'une exclusion des enjeux urbains et maritimes

**Volonté :**

Volonté de réinterroger ces résultats en proposant une méthodologie mobilisant des corpus de données numériques issues des réseaux sociaux (twitter)



# 1. Hypothèse de départ et objectifs du projet

Objectifs du projet RENRHOM: <http://renurohm.univ-lille.fr>

- Explorer la faisabilité d'une nouvelle forme d'interdisciplinarité entre les sciences de l'information-communication et la géographie de l'environnement ;
- Démarrer la constitution d'une base de données de contenus numériques en lien avec l'OHM ;
- Identifier les contributions de l'analyse des réseaux numériques au rapprochement entre les enjeux écologiques de préservation de la biodiversité et les enjeux sociaux d'accessibilité et de développement des usages dans les espaces de nature.
- Hypothèse : Internet et réseaux sociaux : contribution à une reconfiguration des rapports de pouvoir, prise de parole des « sans voix » (Arsène, 2013)
  - Approche reproductible sur d'autres territoires

## Travaux présentés : Analyse de corpus numériques (Twitter)

- Analyse semi-automatique des contenus des tweets afin d'identifier les entités nommées (organisation, personnes et lieux) et les thématiques abordées dans le texte des tweets ;
- Questions posées :
  - Quelles sont les thématiques et les sujets abordés dans les tweets ?
  - Quels sont les couples d'hashtag les plus utilisés ?
  - Quelles sont les évolutions observées selon différentes temporalités ?
  - Quels sont les acteurs qui s'expriment sur les lieux/sujets en lien avec le Parc national des Calanques ? (*Qui parle de Quoi ?*)
  - Quelles sont les relations entre ces acteurs ? (*Qui parle à Qui ?*)

A terme : Est-ce que les discours, débats, projets contenus dans ces réseaux numériques alimentent, rejoignent ou s'opposent à ceux développés dans l'espace physique ?

## Travaux présentés : Analyse de corpus numériques (Twitter)

### – Problématiques associées :

- Comment identifier le corpus de messages courts Twitter relatif à la thématique des recherches ;
- Comment ensuite identifier automatiquement les informations pertinentes à partir de messages courts non standard (faute, pas de vraies phrases dans les tweets, etc.) ;
- Comment extraire et analyser ces informations : combinaison méthode automatique et analyse experte pour le choix des données, leur traitement et la validation les résultats obtenus

## 4. Données de départ

François GARREAU @GARREAU75 · 7月6日  
#Boues #rouges de #Gardanne : les riverains dénoncent un transfert de #pollution de la mer vers la terre [sco.lt/7V6RNZ](https://sco.lt/7V6RNZ)  
翻译自法文

ChristianFaivre @bleusuper · 6月28日  
Les opposants aux #boues rouges attaquent l'Etat en #justice  
[20minutes.fr/marseille/2095...](https://20minutes.fr/marseille/2095...) via @20minutes  
翻译自法文

VIDEO. Marseille: Les opposants aux boues rouges attaquent l'Et...  
Le tribunal administratif et le conseil d'Etat ont été saisis...  
[20minutes.fr](https://20minutes.fr)

1

2

3

1. ChristianFaivre : nom de l'auteur ; @bleusuper : nom de l'auteur identifié ;
2. #boues / #justice : hashtags (qui permet de trouver ce tweet plus facilement que ce qui n'a rien marqué)
3. personne(s) indiquée(s) qui a lien avec le contenu du tweet

## 4. Données de départ : phase de collecte

2 périodes temporelles choisies pour la collecte des données :

- **2007-2011** : processus de création du Parc national des Calanques et de l'organisation du processus de concertation pour élaborer la charte du Parc national. C'est durant cette période que se sont manifestés des groupes d'acteurs favorables ou opposés à la création d'un Parc national dans le territoire des Calanques ;
- **2012-2017** : année de création du Parc national (en 2012) jusqu'aux années récentes marquées par plusieurs conflits d'usages et notamment celui fortement médiatisé, provoqué par des rejets de boues rouges issues de la production d'alumine par l'usine ALTEO de Gardanne, dans les espaces du cœur maritime du Parc national.

## 4. Données de départ : phase de collecte

Collecte des données :

- **Expertise de géographes spécialistes du territoire d'études** : une liste d'acteurs qui ont participé au processus de création du Parc national des Calanques à Marseille (d'associations, conseils de quartiers, personnalités politiques, et citoyens), filtrage des lieux, évènements, etc.
- **Identification** sur le portail Web des acteurs listés d'une liste de mots clés que nous avons utilisés sous forme de **hashtags** pour collecter les tweets

Hashtags	Description
#ParcNationalCalanques	Parc National des Calanques
#PNCal	Parc National des Calanques
#BouesRouges	Boues Rouges
#Marseille	Marseille
#CreationPNCalanques	Création du Parc National des Calanques
#Alteo	Entreprise
#bouerouge	Boues Rouges
#LittoralSudMarseille	Comité santé
...	....

- **Utilisation API Search de Twitter** pour collecter et filtrer les messages qui contiennent nos hashtags : 40771 tweets collectés.



## 5. Traitement des données : méthodologie

### Etape 2 : Approche statistique

**Hypothèse** : si un mot  $m$  est fortement corrélé à un hashtag  $h$  de notre liste alors ce mot est un candidat pertinent pour notre analyse de relations territoriales.

- Pour chaque hashtags, extraction de l'ensemble des mots qui lui est associé dans le corpus de tweets :

1. Mesure Information mutuelle introduite par (Fano, 1961) afin de mesurer l'association entre un mot  $m$  du corpus et un hashtag  $h$  : pour chaque couple de variables aléatoires  $(m, h)$ , nous mesurons leur degré de dépendance au sens probabiliste. L'information mutuelle est donnée par la formule suivante :

$$IM(h, m) = \log_2 \left( \frac{freq(h, m)}{freq(h) \cdot freq(m)} \right)$$

Avec

- $freq(h, m)$  est le rapport entre le nombre de tweets contenant le mot  $m$  et le hashtag  $h$  ( $|Th, m|$ ) et le nombre total de tweets ( $|T|$ )
- $freq(h)$  est le rapport entre le nombre total de tweets contenant le hashtag  $h$  ( $|Th|$ ) et le nombre total de tweets
- $freq(m)$  est le rapport entre le nombre total de tweets contenant le hashtag  $h$  ( $|Tm|$ ) et le nombre total de tweets

## 5. Traitement des données : méthodologie

### Etape 3 : Visualisation et analyses

Plateforme R utilisée pour ces traitements (version R3.4 du logiciel Rstudio) :

- langage R pour réaliser la fouille de textes
  - Ggplot2 les visualisations des résultats.
  - Mots ordonnés par ordre croissant selon le degré d'association : du plus pertinent au moins pertinent.
  - Suppression des tweets intégrant #marseille et calanques : 12 hashtags sélectionnés après validation experts
  - Filtrage des couples pertinents selon un seuil défini avec les experts : > 2 occurrences du couple dans le corpus traité.
- Corpus sélectionné :
- Sur 2007-2011 : 5 000 tweets ;
  - Sur 2012-aujourd'hui 2 900 tweets.

# 5. Traitement des données : méthodologie

## Etape 3 : Visualisation et analyses

Plateforme R utilisée pour ces traitements (version R3.4 du logiciel Rstudio) :

- langage R pour réaliser la fouille de textes
- Ggplot2 les visualisations des résultats.
- Mots ordonnés par ordre croissant selon le degré d'association : du plus pertinent au moins pertinent.
- Suppression des tweets intégrant #marseille et calanques : 12 hashtags sélectionnés après validation experts
- Filtrage des couples pertinents selon un seuil défini avec les experts : > 2 occurrences du couple dans le corpus traité.

	B	C	D	E	F
	MotRequest	hashtag	IM	Ocurrence	
1	ParcNationalCalanques_	cg13	9.75444836289823		1
2	ParcNationalCalanques_	contratdebaie	9.75444836289823		1
3	ParcNationalCalanques_	enqu	9.75444836289823		1
4	ParcNationalCalanques_	eurom	9.75444836289823		
5	ParcNationalCalanques_	evidence	9.75444836289823		
6	ParcNationalCalanques_	frioul	9.75444836289823		
7	ParcNationalCalanques_	guyteissier	9.75444836289823		
8	ParcNationalCalanques_	marseillegrandest	9.75444836289823		
9	ParcNationalCalanques_	nihous	9.75444836289823		
0	ParcNationalCalanques_	parcnationalcalanc	9.75444836289823		
1	ParcNationalCalanques_	promessesquinengag	9.75444836289823		
2	ParcNationalCalanques_	schpountz	9.75444836289823		
3	ParcNationalCalanques_	sormiou	9.75444836289823		
4	ParcNationalCalanques_	teissier	9.75444836289823		
5	ParcNationalCalanques_	ump	9.75444836289823		
6	ParcNationalCalanques_	unfollowsinonrien	9.75444836289823		
7	PNCAL_	eaart	7.88844566828897		
8	PNCAL_	jamansd	7.88844566828897		
9	PNCAL_	logia	7.88844566828897		

	B	C	D
	MotRequest	hashtag	IM
1	ALTEO_	jobs	0.307293754210503
2	ALTEO_	montreal	0.307293754210502
3	ALTEO_	job	0.307293754210503
4	ALTEO_	canada	0.307293754210503
5	ALTEO_	google	0.307293754210502
6	ParcNationalCalanques_	parcnationalcalanc	9.75444836289823
7	Parc National des Calanques	marseille	5.61052274754212
8	Boues rouges_	hongrie	4.71421241370507
9	ALTEO_	serp	0.307293754210502
0	ALTEO_	seo	0.307293754210503
1	ALTEO_	emploi	0.307293754210503
2	ALTEO_	smm	0.307293754210503
3	ALTEO_	career	0.307293754210503
4	ALTEO_	fresher	0.307293754210503
5	ALTEO_	ecommerce	0.307293754210502
6	ALTEO_	facebook	0.307293754210503

## 5. Traitement des données : premiers résultats

### Etape 3 : Visualisation et analyses

Nuage de hashtags – mots clés thématiques pré-sélectionnés par les experts et mondes lexicaux associés avant 2012

- le cas des couples pour le thème « environnement »

	Avant 2012	Après 2012
<b>Tweets</b>	<b>2064</b>	<b>16683</b>
<b>Couples</b>	<b>345</b>	<b>3133</b>
ALTEO	342	1966
(ALTEO, bouesrouges)	0	382
(ALTEO, pollution)	0	117
(ALTEO, environnement)	0	63
Bouesrouges	111	3180
(Bouesrouges, bouesrouges)	4	484
(Bouesrouges, pollution)	4	101
(Bouesrouges, environment)	4	100
Parc National des Calanques	175	1683
(PNCAL, bouesrouges)	0	67
(PNCAL, pollution)	0	20
(PNCAL, environment)	4	27

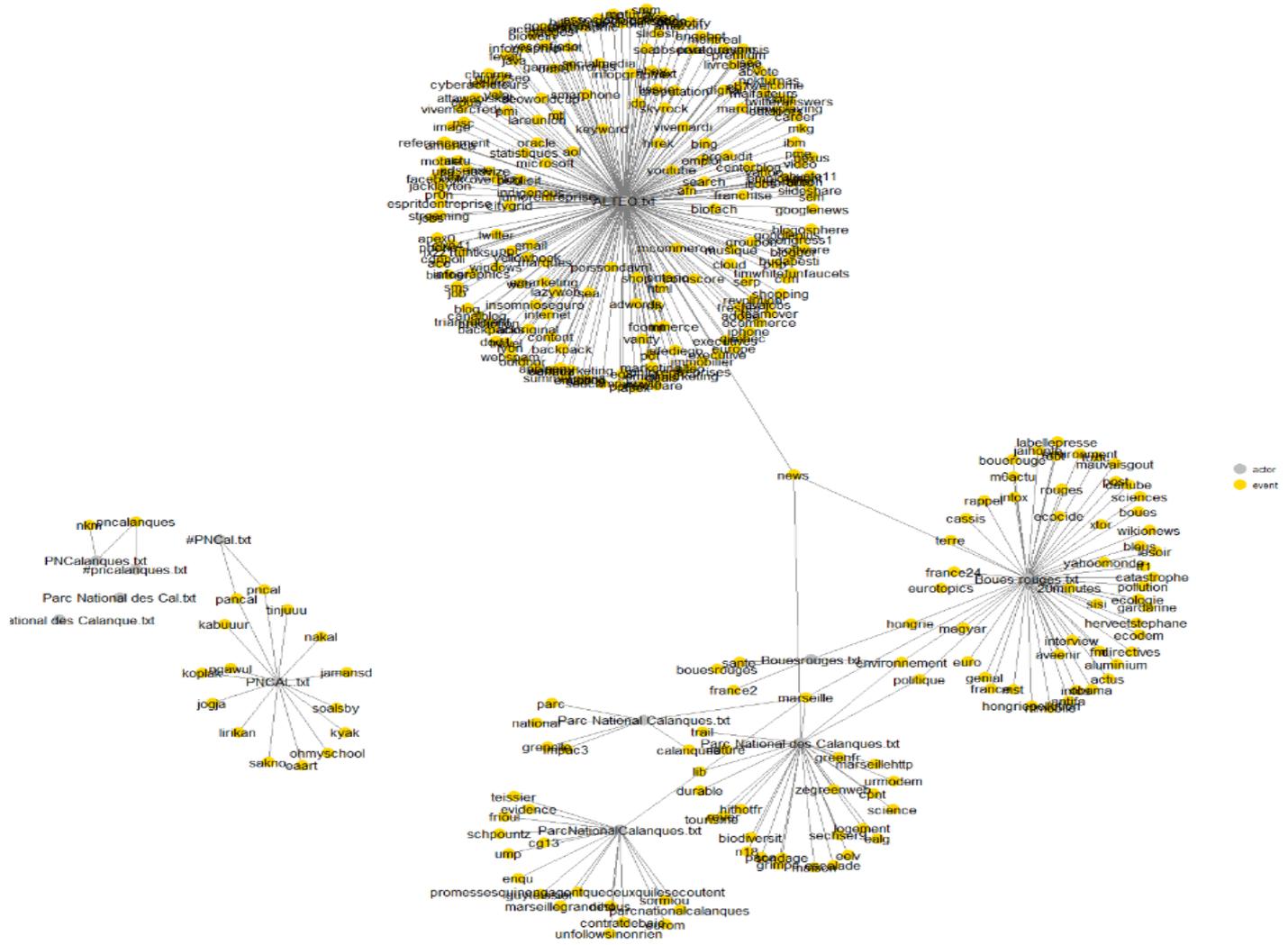
création de  
ALTEO et  
du parc.

# 5. Traitement des données : premiers résultats

## Etape 3 : Visualisation et analyses

Nuage de hashtags – mots clés thématiques pré-sélectionnés par les experts et mondes lexicaux associés avant 2012

- le cas des couples pour le thème « environnement »



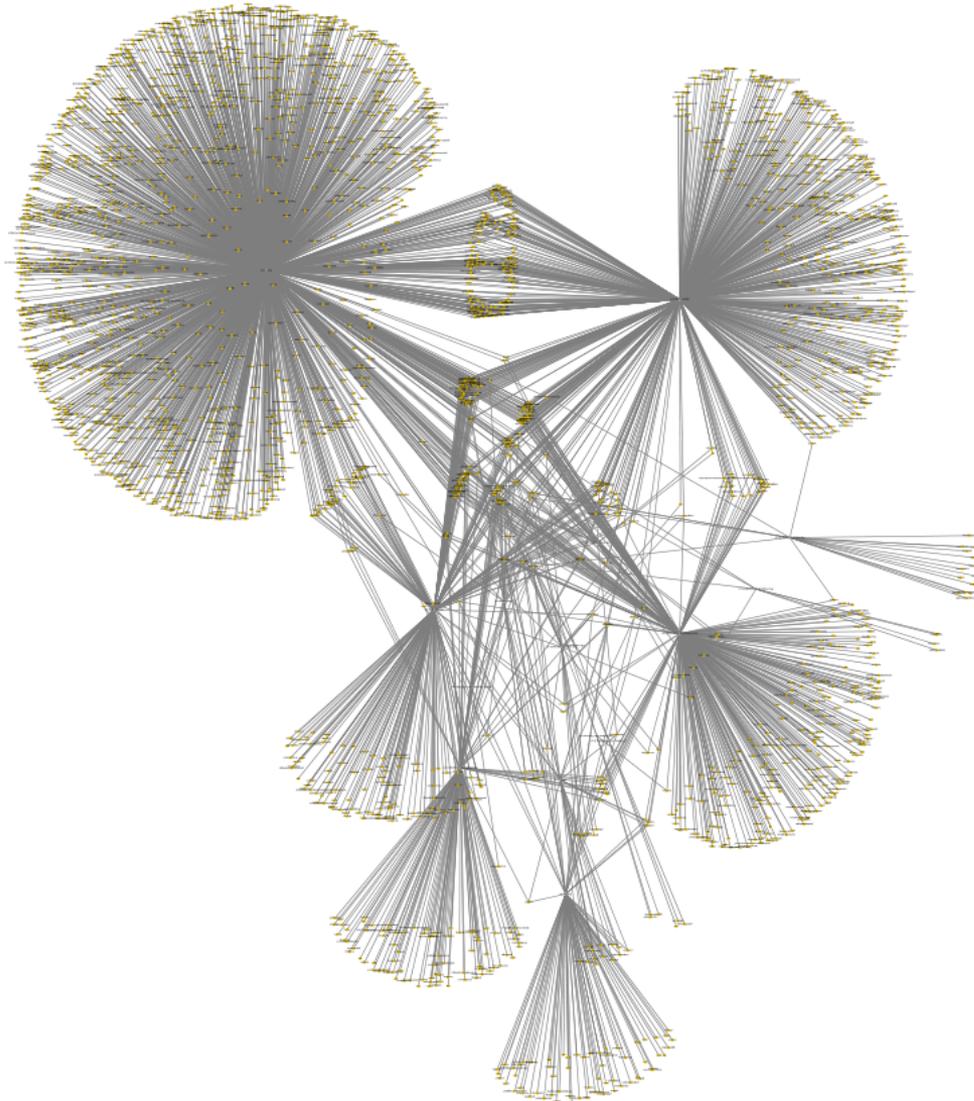


## 5. Traitement des données : premiers résultats

### Etape 3 : Visualisation et analyses

Nuage de hashtags – mots clés thématiques pré-sélectionnés par les experts et mondes lexicaux associés à partir de 2012

- le cas des couples pour le thème « **environnement** »



## 5. Traitement des données : premiers résultats

### Etape 3 : Visualisation et analyses

Nuage de hashtags – mots clés thématiques pré-sélectionnés par les experts

- le cas des couples pour le thème « environnement »

- Evolution des thématiques dans le temps

de 2007 à 2011

Hashtags Experts	Hashtags
Alteo	#jobs, #emploi, #career, etc.
Parc National Calanques	#parcnationalcalanques, #marseille, #nature, #environnement
Boues rouges	#hongrie, #pollution

de 2012 à 2017

Hashtags Experts	Hashtags
Alteo	#bouesrouges, #emploi, #pollution, #calanques, #marseille, #environnement
Parc National Calanques	#calanques, #marseille, #environnement, #ArchipelDeRiou, #parcnationaldescalanques, #bouesrouges, #nature
Boues rouges	#bouesrouges, #calanques, #marseille, #alteo, #pollution



## 6. Premières conclusions

- Une approche quanti – quali intéressante pour le traitement et l'analyse des données provenant des réseaux sociaux
- Un usage des réseaux sociaux de plus en plus important sur la thématique environnement en lien avec le parc national des calanques
  - Des thématiques qui ressortent : #pollution, #nature, #bouesrouges...
  - Une évolution des thématiques dans le temps : forte présence de l'acteur industriel Alteo dans le contenu des tweets

## 6. Perspectives

### Extractions fines des noms de lieux

- Approches de fouille de textes pour préciser :
  - Le marquage des thématiques (Pak *et al.*, 2014)
  - Le marquage des entités spatiales à partir de messages courts (Zenasni *et al.*, 2016 ; 2017)



The image shows a screenshot of three tweets from Twitter. The first tweet is from 'goodplanet @infogoodplanet' dated 21 août 2011, with a profile picture of a globe and the text '#News : Marseille: réouverture de deux plages après une pollution au styrène: MARSEILLE - Deux plages de Marseil... bit.ly/nQoYSp'. The second tweet is from 'Perrin Hugues @HuguesPerrin' dated 7 sept. 2012, with a profile picture of a man and the text 'En réponse à @AlexisPalisson @AlexisPalisson si tu le courage de faire 10 min de voiture , il y a un excellent dans la calanque de Figuerol à la ciotat !'. The third tweet is from 'markstewart @Markstewart000' dated 4 janv. 2016, with a profile picture of a man and the text 'Préfet de la région PACA: Interdire le rejet dans la mer en plein cœur du Parc national des Calanq... change.org/p/pr%C3%A9fet-... via @UKChange'. Each tweet has icons for replies, retweets, likes, and direct messages.

 **goodplanet** @infogoodplanet · 21 août 2011  
#News : Marseille: réouverture de deux plages après une pollution au styrène: **MARSEILLE** - Deux plages de Marseil... [bit.ly/nQoYSp](http://bit.ly/nQoYSp)

 **Perrin Hugues** @HuguesPerrin · 7 sept. 2012  
En réponse à @AlexisPalisson  
[@AlexisPalisson](#) si tu le courage **de** faire 10 min **de** voiture , il y a un excellent dans la **calanque de Figuerol** à la ciotat !

 **markstewart** @Markstewart000 · 4 janv. 2016  
Préfet **de** la région PACA: Interdire le rejet dans la mer en plein cœur du Parc national des **Calanq...** [change.org/p/pr%C3%A9fet-...](http://change.org/p/pr%C3%A9fet-...) via [@UKChange](#)

# 6. Perspectives

## Qui parle sur qui ?

- Avant 2012

B	C	D	E	F
UserID	UserName	Mention	Nombre	
80310555	un_job_de_suite	algebrik	37	
80310555	un_job_de_suite	algebrik2010emploi	17	
19266151	Alexandre Penyauski	agence	8	
19266151	Alexandre Penyauski	alteo	8	
99609168	Alteo	jdnebusiness	6	
1,8E+08	Gengembre Dominique	scoopit	5	
2,74E+08	7AY4	idipodong	5	
4471191	François Cazals	agence	4	
4471191	François Cazals	alteo	4	
4471191	François Cazals	bitchlemagazine	4	
4471191	François Cazals	intempestif	4	

- À partir de 2012

UserID	UserName	Mention
446939787	<U+B808><U+B385><U+C740> <U+BE0C><U+C6>	bot
446939787	<U+B808><U+B385><U+C740> <U+BE0C><U+C6>	alteo
1108658778	l'Encre de Mer	encredemer
481170516	khalidi	royalsegolene
384206154	Didier R	parccalanques
481170516	khalidi	nichelerivasi
874404774	Vialidad Bs. As.	mininfra
874404774	Vialidad Bs. As.	baprovincia
874404774	Vialidad Bs. As.	vialidadba
2936740168	Anarchozy	manuelvalls
1945218524	InsoumisdeBerre13	cnpcf
212778422	SiDD VLAK <f0><U+009F><U+0092><U+0080>	datpiff
874404774	Vialidad Bs. As.	solotransito
1556975869	El interior existe!	rutarota
384206154	Didier R	prefet13
436922925	Dani	royalsegolene
436922925	Dani	didierreault
932234984	PCF 13	cnpcf
7.912099e+17	@antaresdul3	youtube
384206154	Didier R	parccalanque
436922925	Dani	marsactu
104883773	Marsactu	marsactu
267938974	bernard jegou	lemarin
287557798	Andre Dechene	royalsegolene
436922925	Dani	nichelerivasi

- Nécessité d'étudier en détails

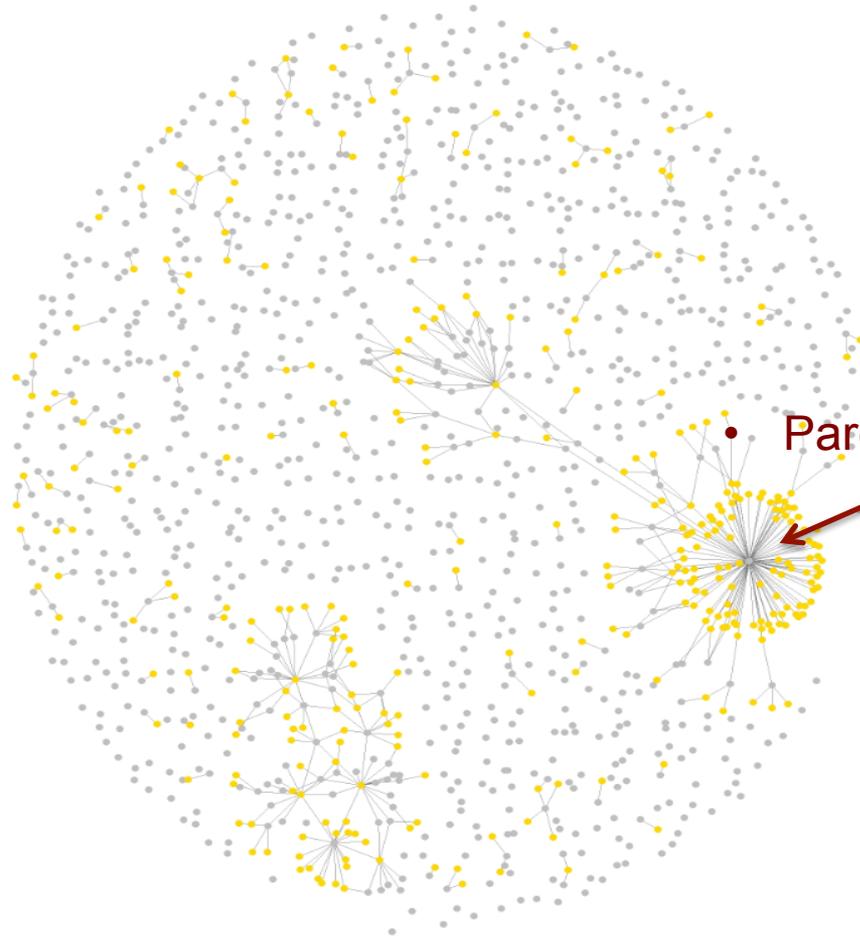
les liens entre acteurs :

- Analyse fine des échanges
- Entretiens sur le terrain (2018)
- Identification des communautés d'acteurs

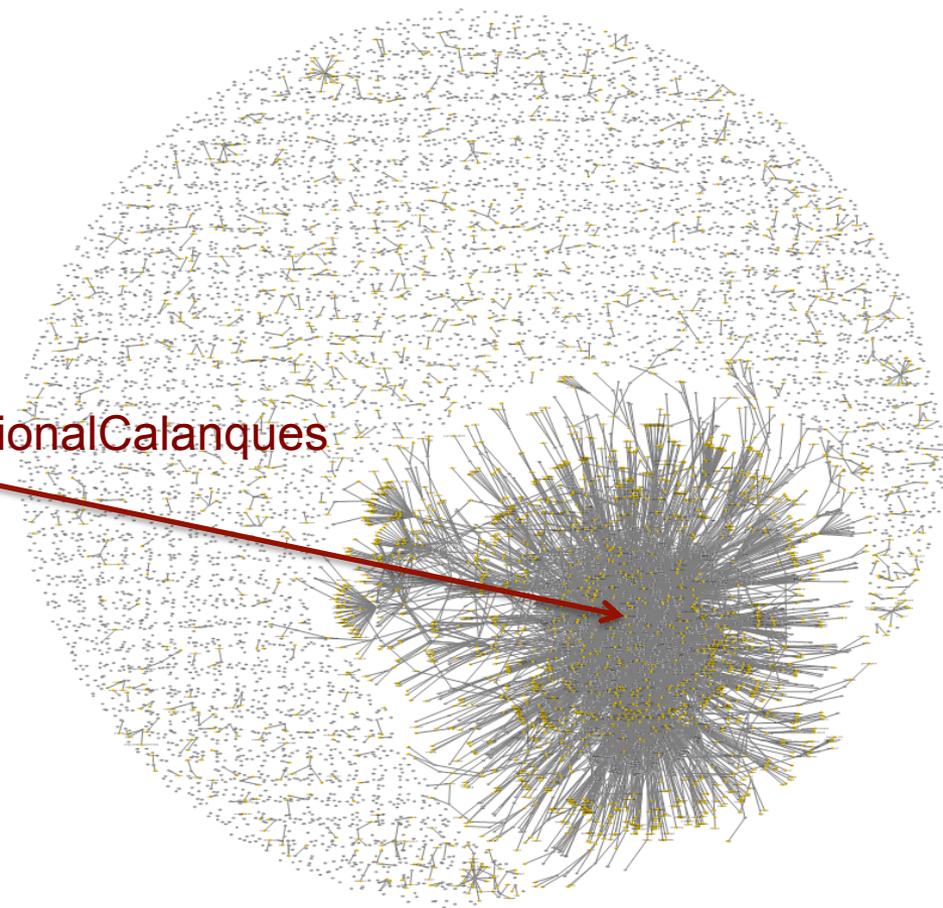
## 6. Perspectives

Qui parle sur quoi ?

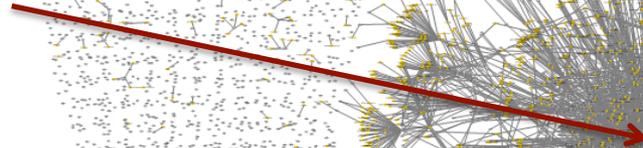
- Avant 2012



- À partir de 2012



• ParcNationalCalanques



- Augmentation de l'usage des thématiques #pollution, #bouesrouges, #environnement dans le temps
- Nécessité de pousser l'étude pour identifier des communautés d'acteurs par thèmes

Contact : [eric.kergosien@univ-lille3.fr](mailto:eric.kergosien@univ-lille3.fr)

- La suite à venir... notamment pour le congrès du CIST :  
<https://cist2018.sciencesconf.org>



### cist2018 - Représenter les territoires

22-23 mars 2018 Rouen (France)



#### DES DÉCOUVERTES

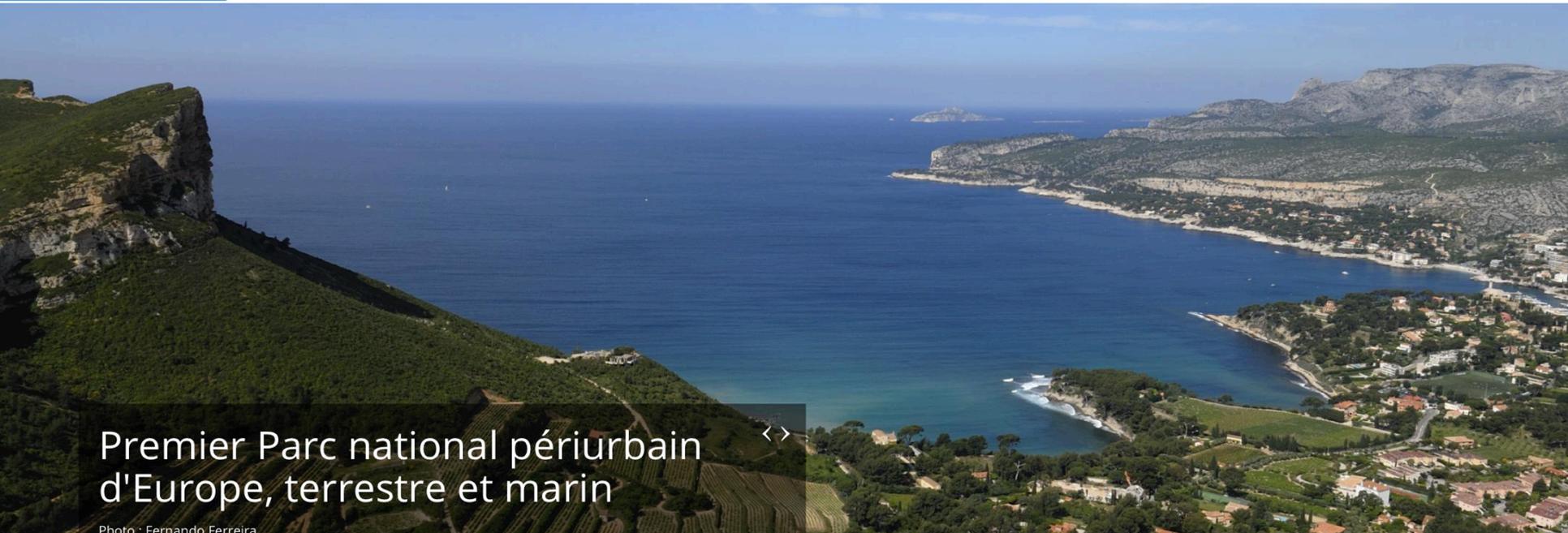
*Faites de rencontres  
& d'émotions*

#### DES CONNAISSANCES

*Au service de la nature  
& des hommes*

#### DES ACTIONS

*Au côté des acteurs  
du Parc national*



Premier Parc national périurbain  
d'Europe, terrestre et marin

## Références bibliographiques :

- Arsène S., 2013, *Vers une recomposition des pouvoirs : Internet et réseaux sociaux*, CERISCOPE Puissance
- Berthelot M.-A., Severo M., Kergosien E., 2016, , Cartographier les acteurs d' un territoire : une pproche appliquée au patrimoine industriel textile du Nord-Pas-de-Calais, In 3ème colloque international du CIST (CIST 2016), pp.6, Grenoble.
- Deboudt P., 2015, L'aménagement du littoral à l'épreuve des inégalités environnementales, *Annales des mines, Responsabilité et Environnement*, n°79, p.83-89.
- Deboudt P., Deldrève V., 2015, Inégalités et concertation « encadrée » : le projet du parc national des calanques, *in* L. Mermet et D. Salles (dir.), *Environnement et transition écologique*, De Boeck éd., coll. Ouvertures Sociologiques, p. 151-166.
- Deldrève V., Deboudt Ph. (dir.), 2012, *Le parc national des calanques : construction territoriale, concertation et usages*, QUAE, 231 p.
- Deboudt Ph. (éd.), 2010, *Inégalités écologiques, territoires littoraux, développement durable*, Presses Universitaires du Septentrion, 409 p.
- Fraisse A., Paroubek P., 2014, *Twitter as a Comparable Corpus to build Multilingual Affective Lexicons*, *in* proceedings of the 7th International Workshop on Building and Using Comparable Corpora at LREC 2014 (BUCC 2014), pages 17-21. May 26-31, 2014., Reykjavik, Iceland.
- Kergosien E., B. Laval, Roche M., Teisseire M., Are opinions expressed in land-use planning documents? In *International Journal of Geographical Information Science*, vol. 28, issue 4, jan. 2014. [Rank A, IF: 1.61 in 2012)
- Pak A. , Paroubek P., Fraisse F., Francopoulo G., 2014, *Normalization of Term Weighting Scheme for Sentiment Analysis. Book Chapter, Human Language technology Challenges for Computer Science and Linguistics. Series: Lecture Notes in Artificial Intelligence*, Springer, Vol. 8387. ISBN 978-3-319-08957-7. Vetulani, Zygmunt, Mariani, Joseph (Eds.). *May 27, 2014.*
- Zenasni S., Kergosien E., Roche M., Teisseire M., 2016, Extracting new Spatial Entities and Relations from Short Messages, In the 8th International ACM Conference on Management of Digita EcoSystems (MEDES'2015), pp. 8, Hendaye (France).