

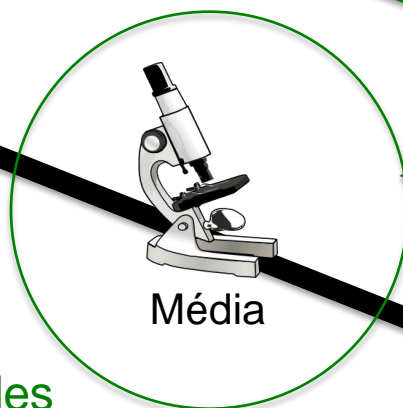
Organisation, agrégation et visualisation d'informations médiatiques

R. Lamarche-Perrin, Y. Demazeau, J.-M. Vincent
Laboratoire d'Informatique de Grenoble

Colloque « Fonder les sciences du territoire »
Paris 23-25 novembre 2011

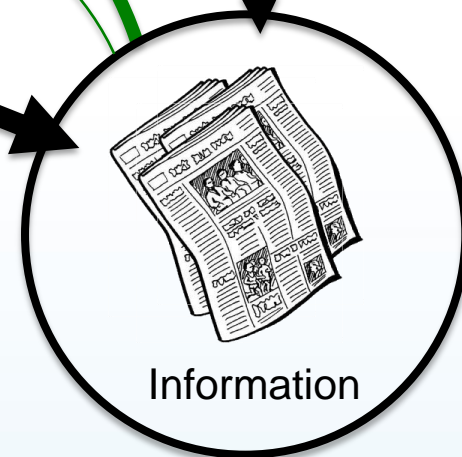
Objectif de l'analyse

Explication des dynamiques sociales

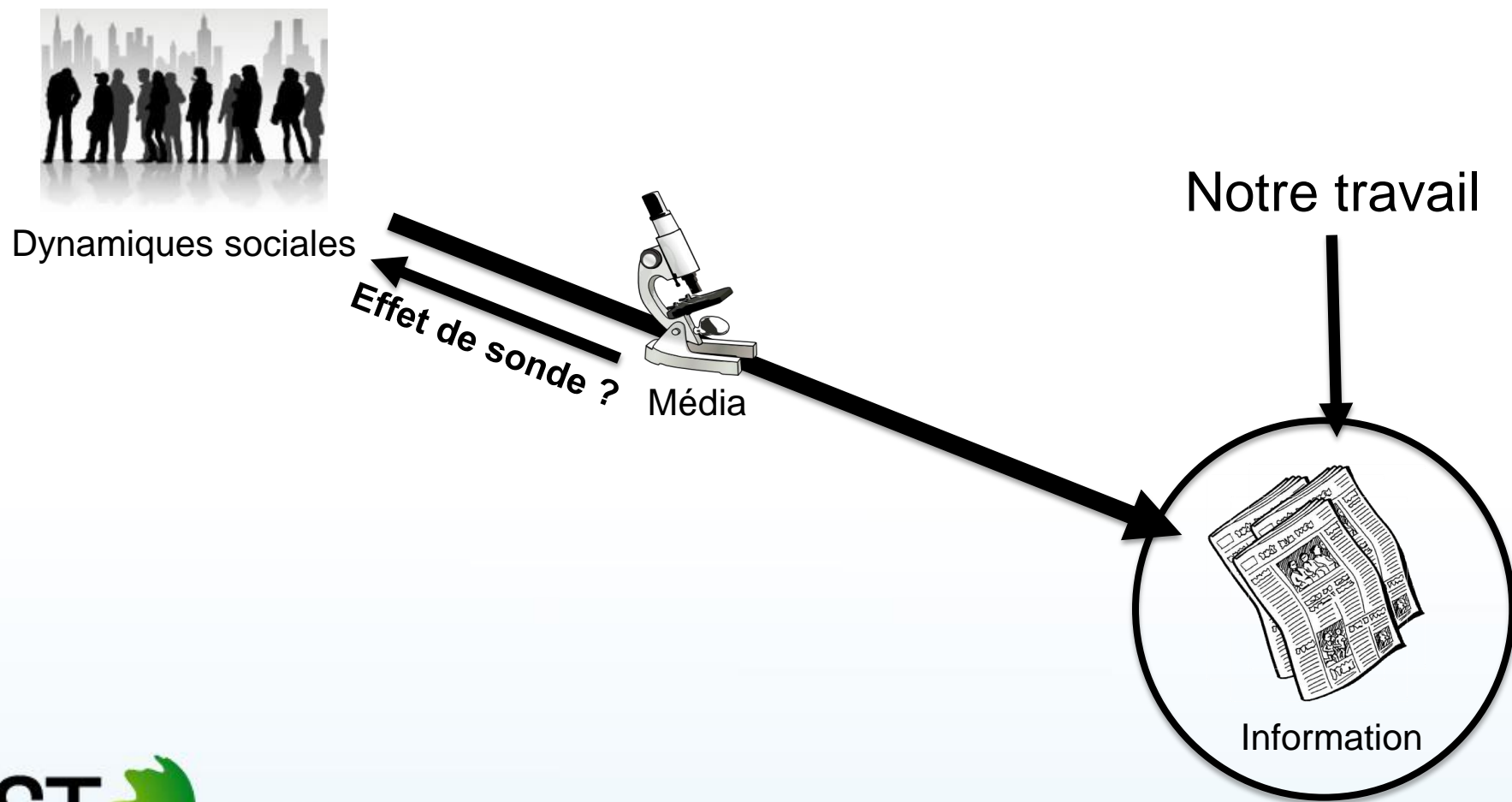


Évaluation des instruments d'observation

Notre travail



Objectif de l'analyse



Difficultés



1. Données hétérogènes, non-structurées

2. Grandes quantités d'information

– **68 flux RSS pendant 206 jours → 280 815 items**

- The Gardian : **100 items/jours**
- Les 10 plus actifs : **62 items/jours**

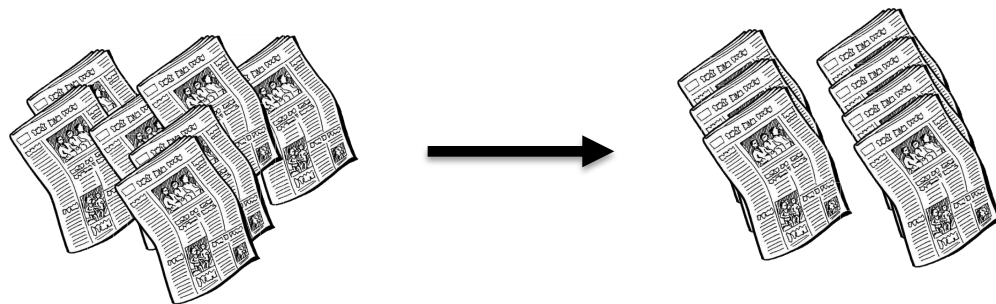
– Exemple : analyse d'un phénomène sur **5 ans**
selon **10 points de vue** → **1 130 000 articles !**

→ **Objectif** : outils pour l'analyse de très grandes quantités de données en « temps réel »

Traitement de l'information

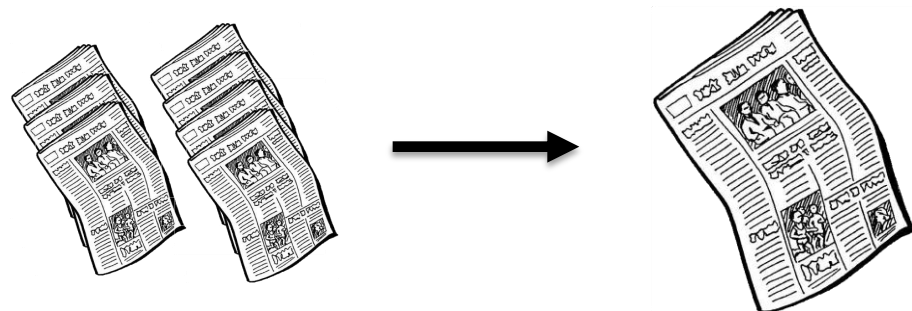
1. Organisation

Quel modèle pour représenter les données ?



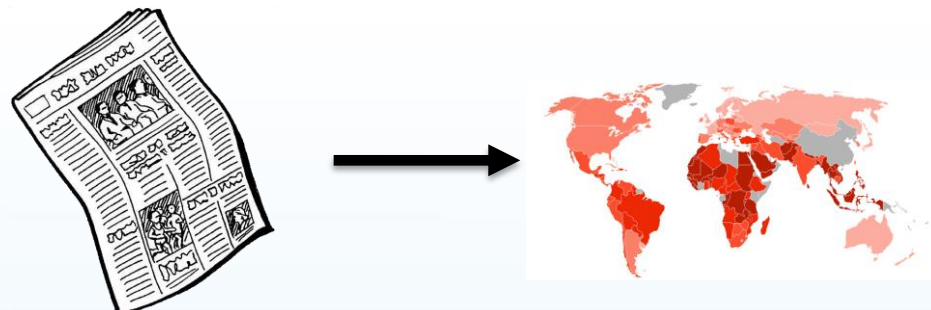
2. Agrégation

Comment engendrer des abstraction ?

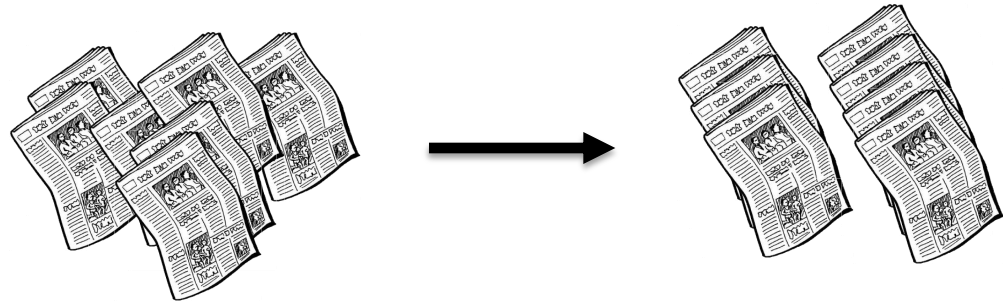


3. Visualisation

Comment afficher les données agrégées à l'utilisateur ?



ORGANISATION DE L'INFORMATION



Les trois dimensions de l'information médiatique

Dimensions		Références indirectes		Références directes
		Absolues	Relatives	
Espace	Où ?	« France » « Allemagne »	« Nos voisins »	Lieu de publication
Temps	Quand ?	« 12 octobre 2006 »	« En mai dernier »	Date de publication
Thèmes	Quoi ?	« Accords diplomatiques »	« Selon l'article X »	Ligne éditoriale

- **Espaces multidimensionnels**

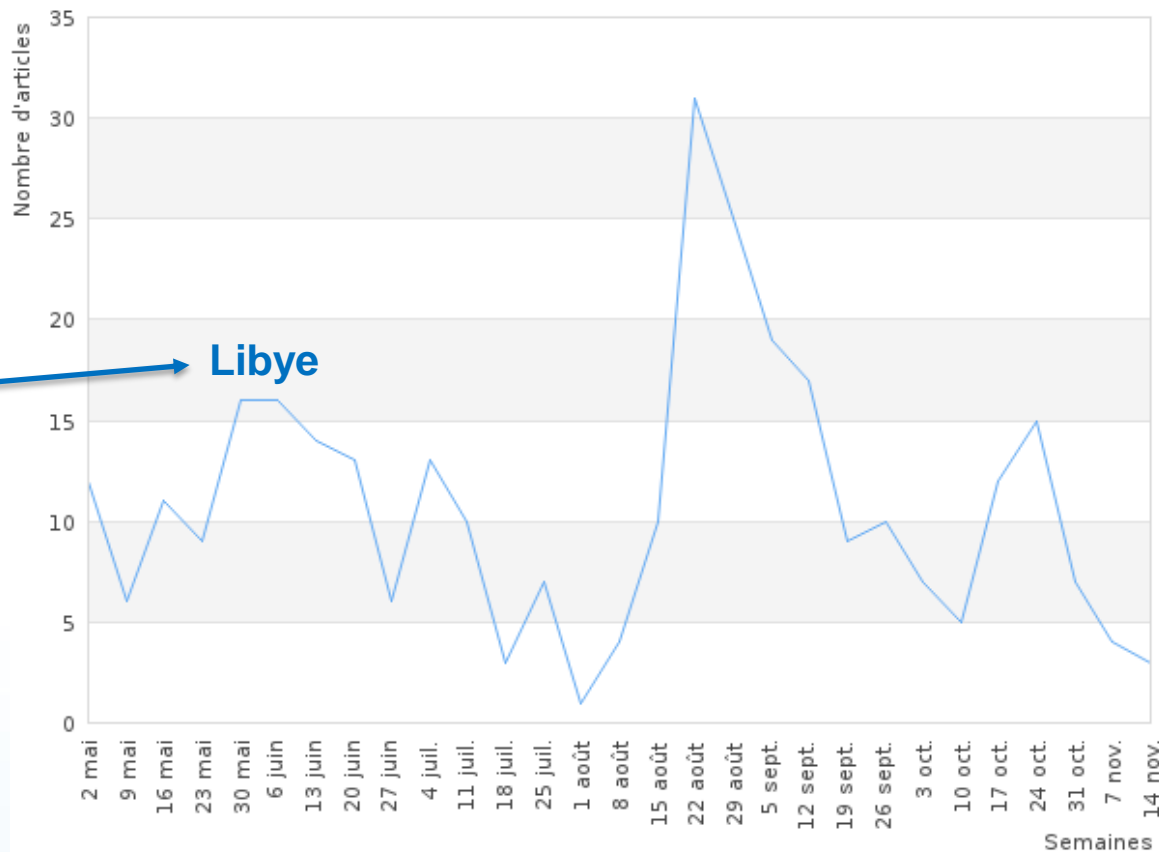
« Causalité » médiatique → Comment ?

Source : Le Monde international,
3520 articles du 02 mai au 20 nov.

Espace × Temps

Variation du poids médiatique des territoires

	USA	Libye	Syrie	France	Israël	...	Total
2 mai	25	12	11	10	4	...	142
9 mai	14	6	12	12	5	...	108
16 mai	20	11	12	6	9	...	142
23 mai	15	9	6	13	5	...	120
30 mai	10	16	17	9	4	...	137
6 juin	14	16	16	9	4	...	114
13 juin	15	14	17	9	6	...	119
20 juin	17	13	12	12	7	...	123
27 juin	7	6	7	20	2	...	103
4 juill.	12	13	8	10	6	...	129
11 juill.	21	10	10	14	3	...	107
18 juill.	7	3	8	4	5	...	61
25 juill.	16	7	6	13	4	...	128
1 août	21	1	9	7	4	...	88
8 août	16	4	10	11	4	...	112
15 août	12	10	7	11	9	...	108
22 août	13	31	7	5	3	...	106
...
Total	423	308	260	248	153	...	3520



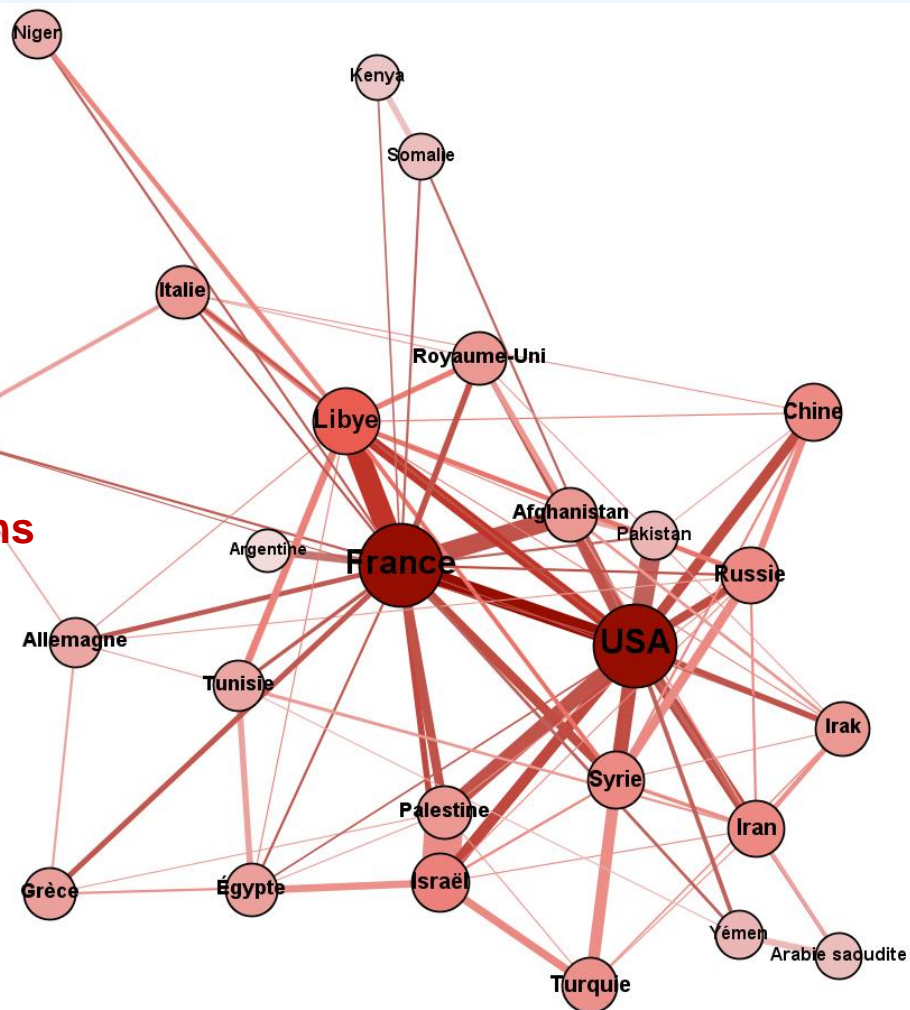
Espace²

Poids des relations médiatiques
entre les territoires

Source : Le Monde international,
3520 articles du 02 mai au 20 nov.

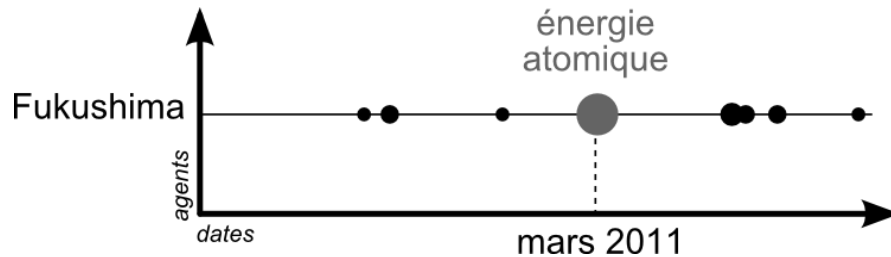
	USA	France	Libye	Israël	Syrie	Palestine	Afghanistan	...	Total
USA	×	25	19	18	24	24	24	...	423
France	25	×	36	10	15	15	30	...	248
Libye	19	36	×	0	7	0	2	...	308
Israël	18	10	0	×	4	62	0	...	153
Syrie	24	15	7	4	×	1	0	...	260
Palestine	24	15	0	62	1	×	0	...	126
Afghanistan	24	30	2	0	0	0	×	...	131
...
Total	423	248	308	153	260	126	131	...	3520

Relations

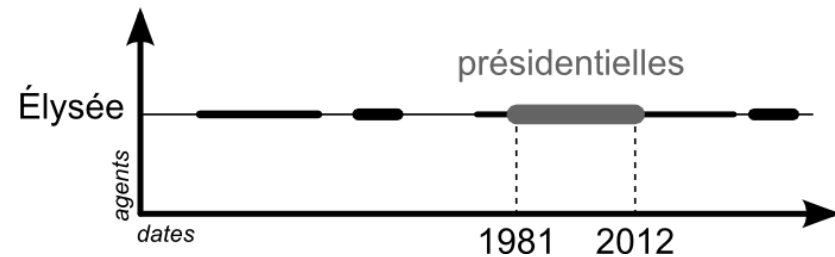


Algèbre sur les dimensions

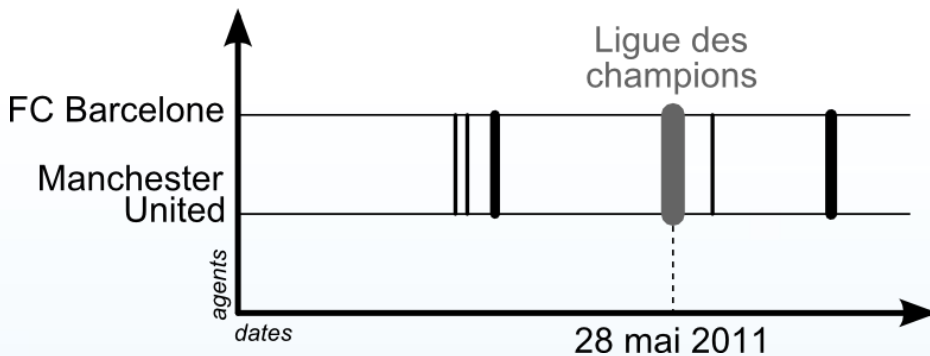
Espace × Temps × Thème
Faits médiatiques



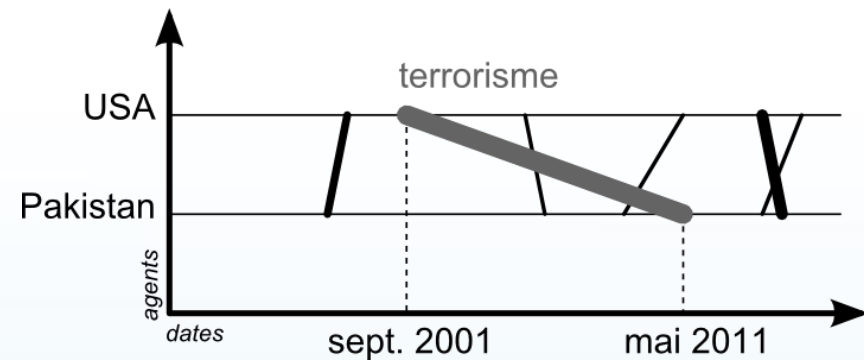
Espace × Temps² × Thème
Interactions temporelles



Espace² × Temps × Thème
Interactions spatiales



Espace² × Temps² × Thème
Interactions spatio-temporelles

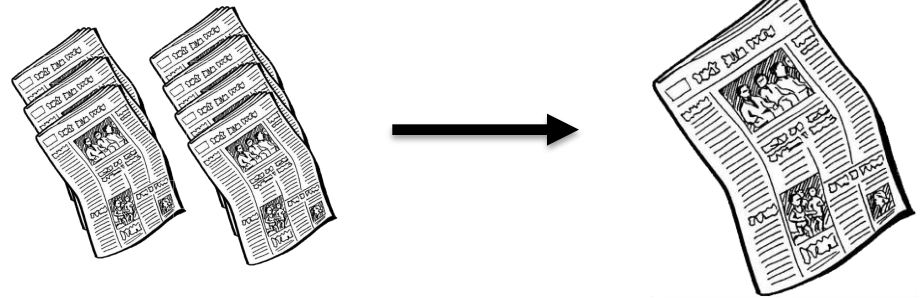


Problématiques

- Quelles structures multidimensionnelles permettent d'expliquer un phénomène donné ou de découvrir de nouvelles dynamiques ?
- Quelles sémantiques donner aux structures multidimensionnelles ?

AGRÉGATION DE L'INFORMATION

(Travail en cours)



Objectif de l'agrégation

- Objectif

Améliorer l'intelligibilité de l'information

→ Réduire la complexité des structures

→ Engendrer des abstractions utiles à l'analyse

- Contraintes

- Contrôler la perte d'information

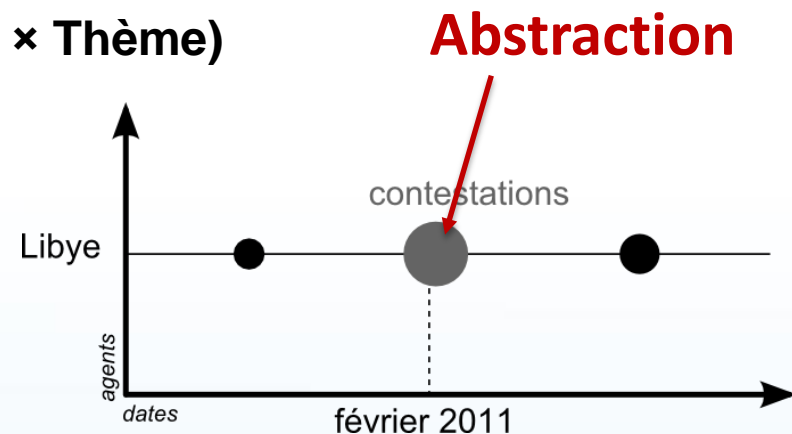
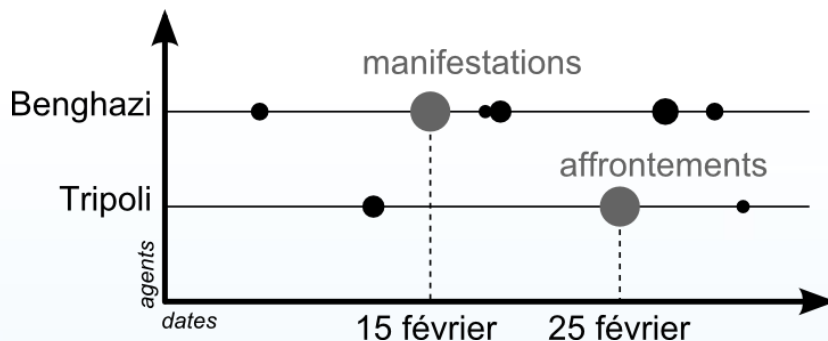
- Conserver les relations de causalité

Agrégation *a priori*

Micro → **Macro**

Espace	Communes	→	Pays
Temps	Jours	→	Mois
Thèmes	Mots-clés	→	Catégories

Faits médiatiques (Espace × Temps × Thème)



Agrégation *a posteriori*

- Agréger les entités qui ont « le même comportement médiatique », *i.e.* qui sont **fortement corrélées**

		Dimensions à agréger		
		Espace	Temps	Thèmes
Mode de corrélation	Espace	Relations spatiales similaires	Espaces co-cités similaires	Espaces co-cités similaires
	Temps	Dates co-citées similaires	Relations temporelles similaires	Dates co-citées similaires
	Thèmes	Thèmes co-cités similaires	Thèmes co-cités similaires	Relations thématiques similaires

Agrégation *a posteriori*

- Agréger les entités qui ont « le médiateur », *i.e.* qui sont formées par un médiateur et un médian

Aggréger les entités qui ont « le même médiateur », *i.e.* qui sont fortement corrélées

Mode de corrélation

		Dim 1	Dim 2	Dim 3	Dim 4	Dim 5	Dim 6	Dim 7	Dim 8	Dim 9	Dim 10
Espace	Espace	Relations spatiales similaires	Espace	Espace	Espace	Espace	Espace	Espace	Espace	Espace	Espace
	Temps	Dates co-citées similaires									
	Thèmes	Thèmes co-cités similaires									
		Thèmes co-cités similaires									

Espace × Espace

	USA	France	Libye	Israël	Syrie	Palestine	Afghanistan	...	Total
USA	×	25	19	18	24	24	24	...	423
France	25	×	36	10	15	15	30	...	248
Libye	19	36	×	0	7	0	2	...	308
Israël	18	10	0	×	4	62	0	...	153
Syrie	24	15	7	4	×	1	0	...	260
Palestine	24	15	0	62	1	×	0	...	126
Afghanistan	24	30	2	0	0	0	×	...	131
...
Total	423	248	308	153	260	126	131	...	3520

similaires

similaires

Relations thématiques similaires

Agrégation a

Espace × Temps

- Agréger les entités qui ont « l'impact médiatique », i.e. qui sont **for**

Mode de corrélation	Dimension	
	Espace	Temps
	Espace	Temps
	Thèmes	Thèmes
	Relations spatiales similaires	Dates co-citées similaires
		Thèmes co-cités similaires

	USA	Libye	Syrie	France	Israël	...	Total
2 mai	25	12	11	10	4	...	142
9 mai	14	6	12	12	5	...	108
16 mai	20	11	12	6	9	...	142
23 mai	15	9	6	13	5	...	120
30 mai	10	16	17	9	4	...	137
6 juin	14	16	16	9	4	...	114
13 juin	15	14	17	9	6	...	119
20 juin	17	13	12	12	7	...	123
27 juin	7	6	7	20	2	...	103
4 juill.	12	13	8	10	6	...	129
11 juill.	21	10	10	14	3	...	107
18 juill.	7	3	8	4	5	...	61
25 juill.	16	7	6	13	4	...	128
1 août	21	1	9	7	4	...	88
8 août	16	4	10	11	4	...	112
15 août	12	10	7	11	9	...	108
22 août	13	31	7	5	3	...	106
...
Total	423	308	260	248	153	...	3520

Agrégation *a posteriori*

- Agréger les entités qui ont « le même comportement médiatique », *i.e.* qui sont **fortement corrélées**

		Dimensions à agréger		
		Espace	Temps	Thèmes
Mode de corrélation	Espace	Relations spatiales similaires	Espaces co-cités similaires	Espaces co-cités similaires
	Temps	Dates co-citées similaires	Relations temporelles similaires	Dates co-citées similaires
	Thèmes	Thèmes co-cités similaires	Thèmes co-cités similaires	Relations thématiques similaires

Problématiques

- Quelles sémantiques associer aux différents modes de corrélation ?
- Quels modes de corrélation privilégier pour chacune des trois dimensions ?

Merci pour votre attention

Refresh Reset

[cleanDB](#)
[exemple1](#)
[exemple2](#)
[exemple3](#)
[fillStructures](#)
[getGephiGraph](#)
[getItems](#)
[getLineGraph](#)
[getOccurrences](#)
[getPlotGraph](#)
[getStructures1](#)
[getStructures2](#)
[getStructures3](#)
[getStructures4](#)
[getStructures5](#)
[printItems](#)
[printOccurrences](#)

SCRIPT fillStructures
Description du script.

Run!

SCRIPT getItems
Chargement des articles : 0% / 10% / 20% / 30% / 40% / 50% / 60% / 70% / 80% / 90% / 100%
Base de données initialisée : 1 média et 3554 articles
Time : 3958 ms

SCRIPT getOccurrences
Recherche des agents : 0% / 10% / 20% / 30% / 40% / 50% / 60% / 70% / 80% / 90% / 100%
Dictionnaire : 200 agents
Agents trouvés : 4048 occurrences (140 distinctes)
Recherche des dates : 0% / 10% / 20% / 30% / 40% / 50% / 60% / 70% / 80% / 90% / 100%
Dates trouvées : 3554 occurrences (205 distinctes)
Recherche des thèmes : 0% / 10% / 20% / 30% / 40% / 50% / 60% / 70% / 80% / 90% / 100%
Dictionnaire : 54 thèmes
Thèmes trouvés : 4179 occurrences (52 distinctes)
Recherche des médias : 0% / 10% / 20% / 30% / 40% / 50% / 60% / 70% / 80% / 90% / 100%
Médias trouvés : 3554 occurrences (1 distinctes)
Time : 60707 ms

SCRIPT getLineGraph
Chargement des dimensions : 0% / 10% / 20% / 30% / 40% / 50% / 60% / 70% / 80% / 90% / 100%
Dimensions chargées : 2
Chargement des structures : 0% / 10% / 20% / 30% / 40% / 50% / 60% / 70% / 80% / 90% / 100%
Structures chargées : 1
Time : 1266 ms

SCRIPT printOccurrences
Décompte des occurrences : 0% / 10% / 20% / 30% / 40% / 50% / 60% / 70% / 80% / 90% / 100%
Calcul des distributions : 0% / 10% / 20% / 30% / 40% / 50% / 60% / 70% / 80% / 90% / 100%
Time : 7344 ms

SCRIPT exemple1
Suppression des dimensions
Enregistrement des dimensions : 0% / 10% / 20% / 30% / 40% / 50% / 60% / 70% / 80% / 90% / 100%
Dimensions enregistrées : 2
Complétion des dimensions : 0% / 10% / 20% / 30% / 40% / 50% / 60% / 70% / 80% / 90% / 100%
Suppression des structures
Enregistrement des structures : 0% / 10% / 20% / 30% / 40% / 50% / 60% / 70% / 80% / 90% / 100%
Structures enregistrées : 1

RELATIONS ENTRE PAYS 2818 éléments (962 distincts) AGENT x AGENT


RELATIONS ENTRE PAYS (Retour)

	USA	France	Israël	Libye	Syrie	Palestine	Afghanistan	Russie	Iran	Chine	...	Total
USA	0	27	22	20	26	26		14	23	20	...	470
France	27	0	12	36	17	16		31	4	0	...	270
Israël	22	12	0	0	4	67		0	2	7	...	172
Libye	20	36	0	0	9	0		2	9	2	...	318
Syrie	26	17	4	9	0	1		0	16	6	...	290
Palestine	26	16	67	0	1	0		0	2	0	...	135
Afghanistan	26	31	0	2	0	0		0	0	3	...	140
Russie	14	4	2	9	16	2		0	0	5	...	65
Iran	23	0	7	2	6	0		3	5	0	...	83
Chine	20	0	0	2	4	0		0	13	2	...	138
...
Total	470	270	172	318	290	135	140	65	83	138	...	3554

distribution4.png (Image PNG, 800x400 pixels) - Mozilla Firefox

distribution2.png (Image PNG, 800x400 pixels) - Mozilla Firefox

Nombre d'occurrences de theme dans les textes du Monde International

CIST 

Robin.Lamarche-Perrin@imag.fr

		Espace		Temps		Thèmes
		Direct	Indirect	Direct	Indirect	
Espace	Direct	Espaces co-cités similaires	Espaces médiatisés similaires	Espaces co-cités similaires	Espaces médiatisés similaires	Espaces co-cités similaires
	Indirect	Répartitions médiatiques similaires		Répartitions médiatiques similaires		Répartitions médiatiques similaires
Temps	Direct	Dates co-citées similaires	Dates médiatisées similaires	Dates co-citées similaires	Dates médiatisées similaires	Dates co-citées similaires
	Indirect	Variations médiatiques similaires		Variations médiatiques similaires		Variations médiatiques similaires
Thèmes		Thèmes co-cités similaires	Thèmes médiatisés similaires	Thèmes co-cités similaires	Thèmes médiatisés similaires	Thèmes co-cités similaires