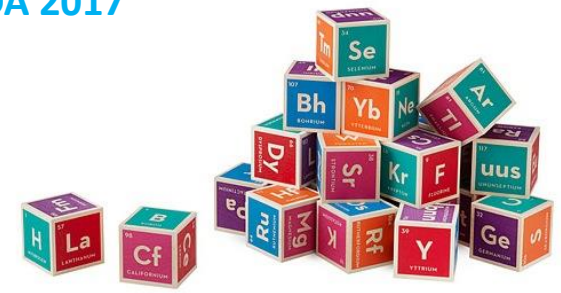
ERIC
ENTREPÔTS, REPRÉSENTATION & INGÉNIERIE des CONNAISSANCES

ADOC
Entrepôts et analyse de documents
3-5 mai 2017 – Lyon
En conjonction avec EDA 2017

CNRS MaDICS



Cécile Favre*
Wararat Jakawat**
Sabine Loudcher*

* Université Lyon 2, ERIC
** Prince of Songkhla University, Thailand

Jeux de cubes pour les graphes … Ou comment des graphes enrichis par des cubes (GreC) peuvent contribuer à l'analyse de données textuelles ?

Lyon 1
UNIVERSITÉ LUMIÈRE LYON 2 UNIVERSITÉ DE LYON
UNIVERSITÉ DE LYON

Données textuelles

- Deux caractéristiques notables :
  - différents axes d'analyse peuvent être pertinents pour résumer l'information
  - possibilité de représenter en réseau(x) certaines informations

- Exemples :
  - Messages Twitter :
    - Nombre de messages peut être analysé en fonction de leur auteur, de leur temporalité, du sujet traité, etc.
    - Réseau des Twittos où les arêtes représentent par exemple les liens d'abonnement.
  - Données de la recherche, cas de la bibliométrie et de l'analyse de la production scientifique
    - Analyser cette production au travers du nombre de papiers publiés selon différents axes que seraient leurs auteurs, l'année de publication, les thématiques, etc.
    - Intéressant d'analyser le graphe des auteurs pour s'intéresser au réseau de collaborations (co-publications), ou au réseau de citations.

Objectifs

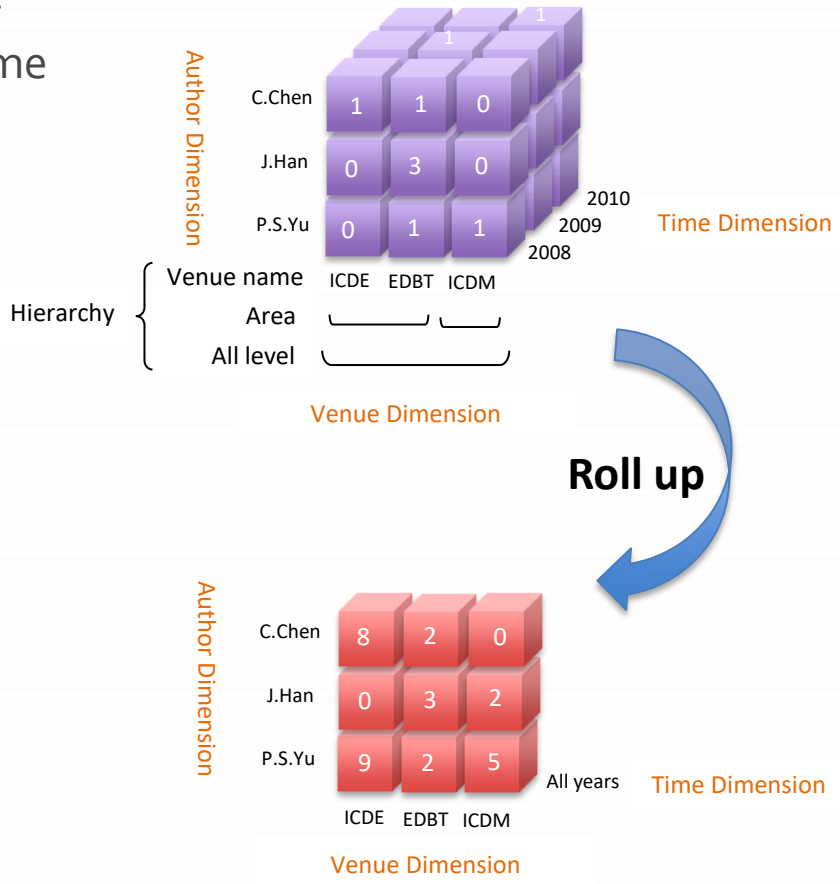
- Traduction de ces deux caractéristiques en dynamique d'analyse
  - Analyse multidimensionnelle de type OLAP (OnLine Analytical Processing) liée aux entrepôts de données avec présentation sous forme de cube de données
  - Approche d'analyse/de visualisation de graphes
- Enjeu de tirer parti de ces deux aspects :
  - Emergence du « Graph OLAP » (Chen *et al.*, 2008)
- Objectifs de la présentation :
  - Présenter les grands principes de notre alternative du Graph OLAP avec GreC (Graphes enrichis par des cubes)
  - Focus sur la mise en œuvre par rapport aux données textuelles

Plan

- Préambules sur l'analyse
  - OLAP
  - Réseaux d'information
- Graph OLAP « classique »
- GreC
  - Principe
  - Positionnement vis-à-vis des données textuelles
- Conclusion et perspectives

OLAP classique

- OLAP (Online Analytical Processing)
  - Analyse de données multidimensionnelle
  - Modèle multidimensionnel avec paradigme du cube de données
  - Sujet d'analyse étudié (fait) au travers de mesure(s)
  - Axes d'analyse : dimensions pouvant comporter des niveaux hiérarchiques
  - Processus d'agrégation des mesures
  - Opérateurs OLAP de navigation



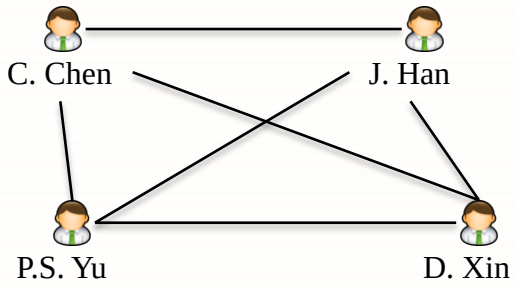
Réseaux d'information

- Caractéristiques
  - Grand nombre d'objets inter-agissant et de types multiples
  - Extraction à partir du web, de bases en ligne, du texte…
  - Des indicateurs spécifiques pour les caractériser

- Deux types de réseaux :

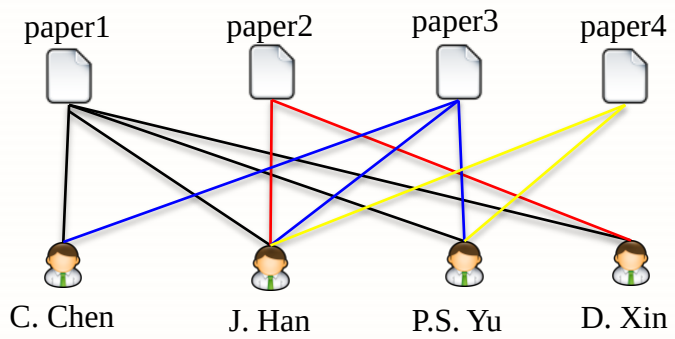
Réseau homogène

- **Types d'objet et de lien uniques**



Réseau hétérogène

- **Types d'objet et de lien multiples**

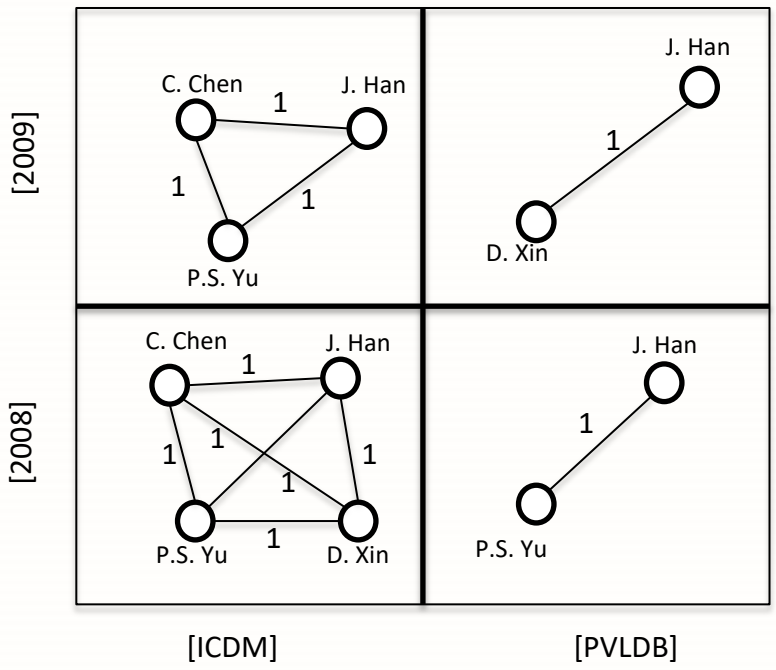


Graph OLAP

Graph OLAP

Graph OLAP "classique"

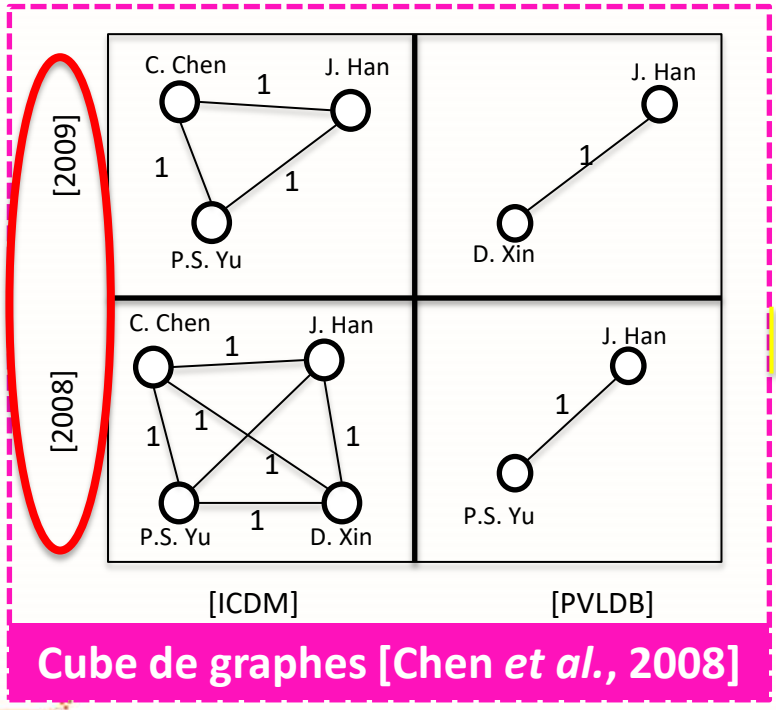
- Cube de graphes (snapshots)
- Chaque cellule du cube : un graphe en tant que mesure
- Navigation au sein du cube
- Deux types de dimension
  - Dimension informationnelle
  - Dimension topologique



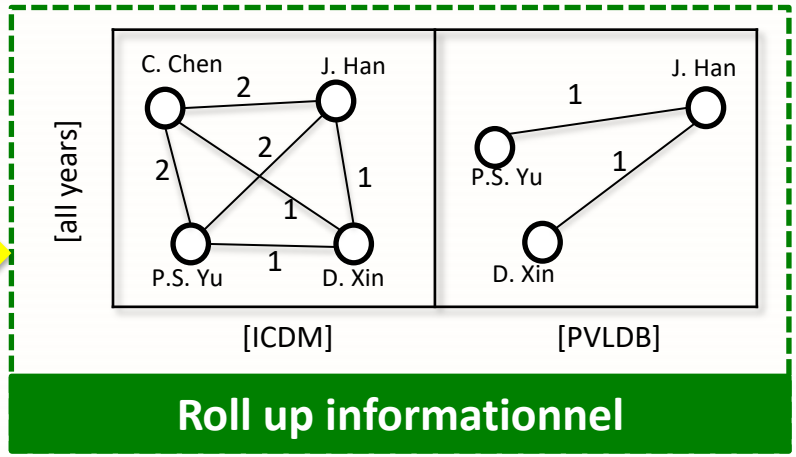
Cube de graphes [Chen *et al.*, 2008]

Graph OLAP "classique"

- Dimension informationnelle :
  - Dimension du cube
  - Organisation des snapshots selon la définition des cellules du cube
  - Opération de navigation informationnelle (par rapport au cube) : le type des nœuds ne change pas



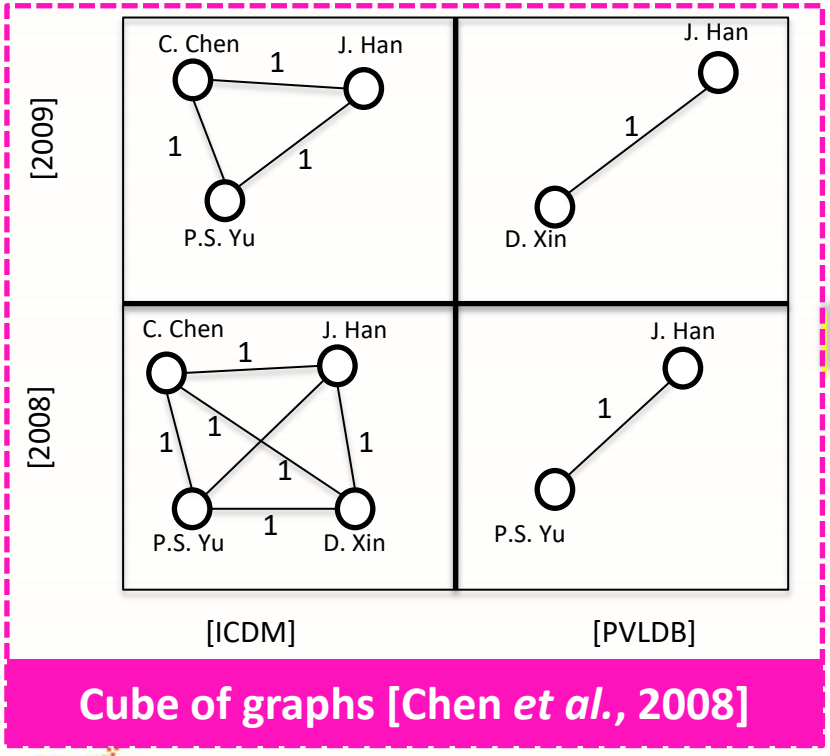
Cube de graphes [Chen *et al.*, 2008]



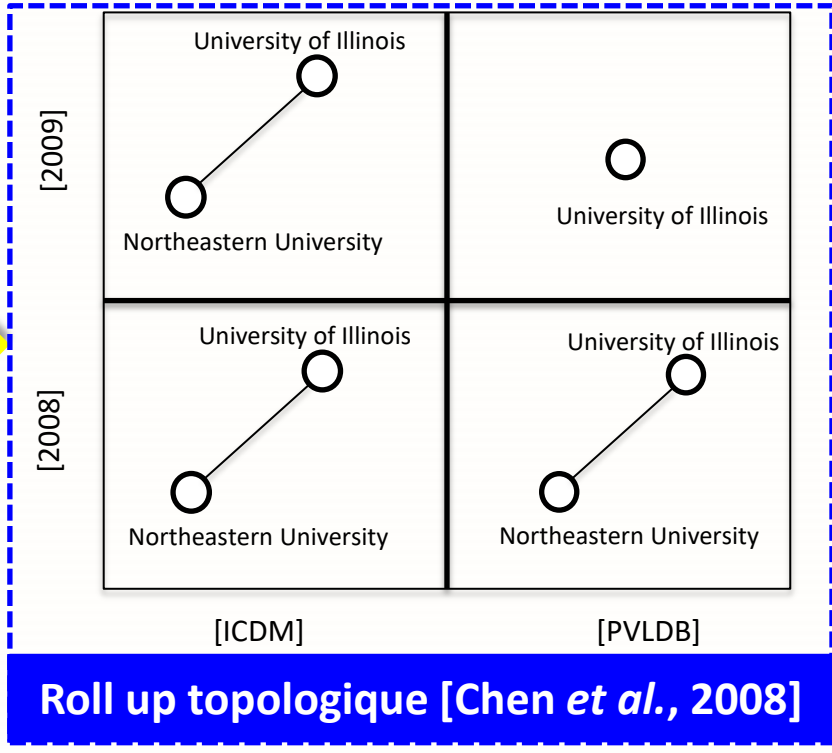
Roll up informationnel

Graph OLAP "classique"

- Dimension topologique :
  - Dimension du graphe
  - Définition de la granularité des graphes
  - Opération de navigation topologique (par rapport aux graphes) : le type des nœuds dans les cellules du cube change en fonction de cette dimension topologique

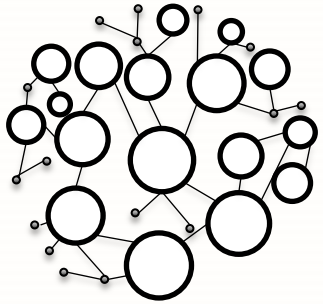


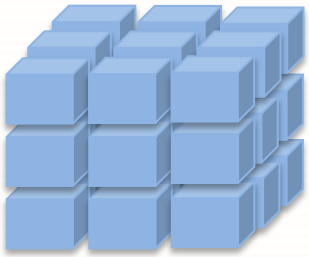
Cube of graphs [Chen *et al.*, 2008]



Roll up topologique [Chen *et al.*, 2008]

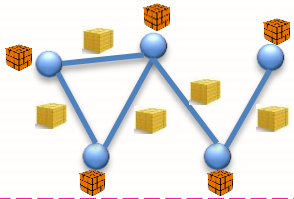
Graph OLAP





Cube de graphes

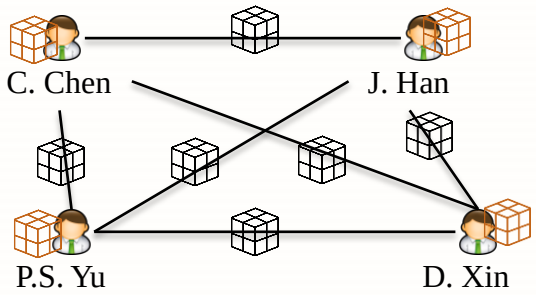
Graphes enrichis par les Cubes



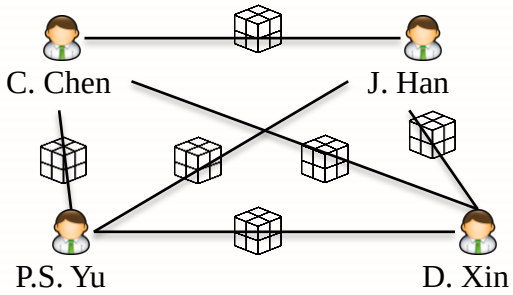
GreC

- Principes de GreC (Graphes enrichis par les Cubes)
  - Enrichir les graphes par des cubes
  - Nœuds et/ou arêtes valués par des cubes en fonction du sujet d'analyse et des types de mesures

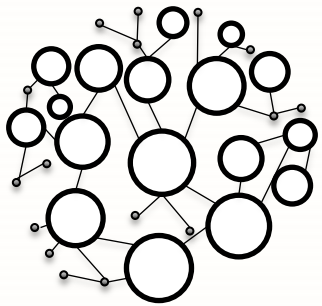
Production

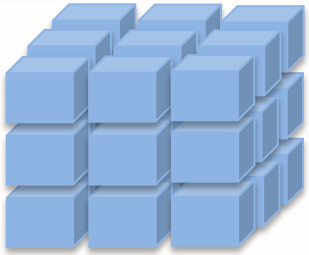


Co-publication



GreC



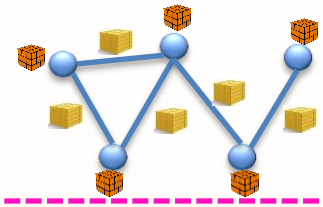


Cube de graphes

Mesures : graphes

Dimensions : informationnelles & topologiques

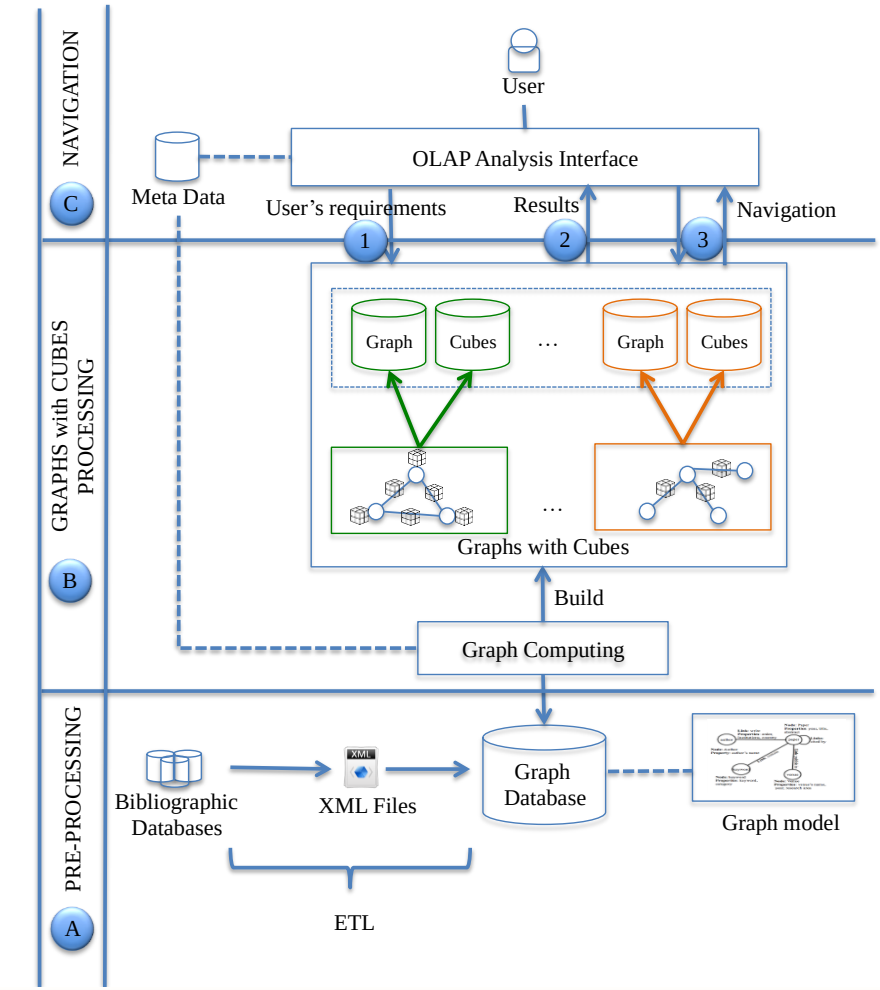
Graphes enrichis par les Cubes



Mesures numériques

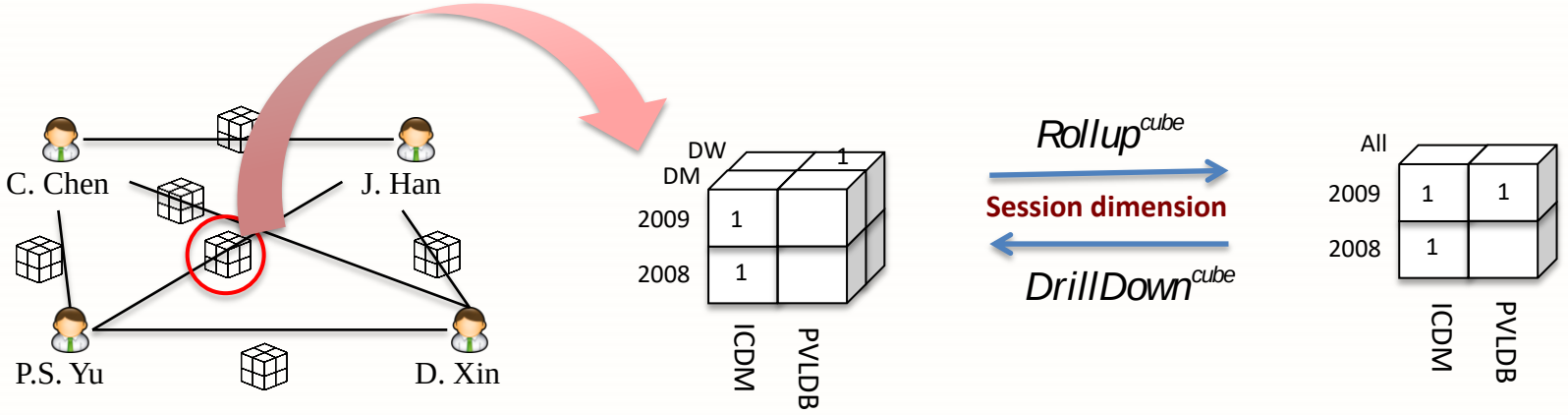
- Indicateurs “simples”
- Mesures basées sur les graphes

Processus


NAVIGATION
C
User
OLAP Analysis Interface
Meta Data
User's requirements
Results
Navigation
1
2
3
GRAPHS with CUBES PROCESSING
B
Graph
Cubes
…
Graph
Cubes
Graphs with Cubes
Build
Graph Computing
PRE-PROCESSING
A
Bibliographic Databases
XML
XML Files
Graph Database
Graph model
ETL

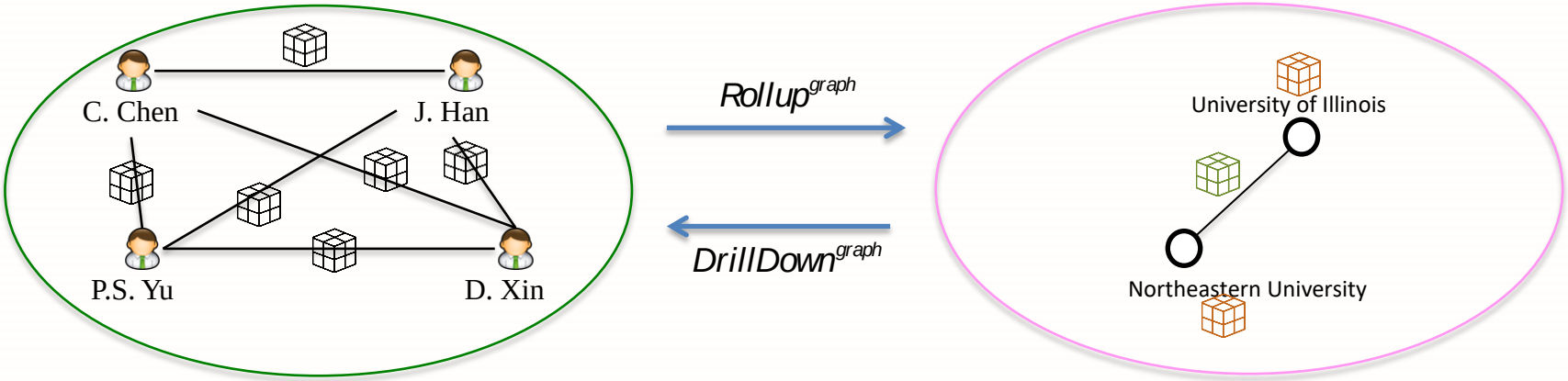
Opérations informationnelles sur les cubes du graphe

- **Roll up** : forage vers le haut selon la dimension spécifiée
- **Drill down**: forage vers le bas selon la dimension spécifiée



Opérations topologiques sur le graphe

- Rollup^{graph} : génération du réseau à un niveau plus général
- DrillDown^{graph} génération d'un réseau à un niveau plus détaillé



Positionnement données textuelles

- Plusieurs étapes de manipulation des données textuelles
  - Construction de réseaux
  - Construction des cubes

- Apport du text-mining et du text-OLAP
  - Construction de réseaux d'auteurs : co-publications, citations…
    - Désambiguisation
    - Complémentarité des bases
  - Enrichissement du contexte d'analyse

  Exemple : détection de thématiques (hiérarchisées) pour nourrir une dimension d'analyse (et sa hiérarchie)
  - Construction de mesures et fonctions d'agrégation basées sur le texte
  - …

Conclusion

- Présentation de l'approche GreC
  - Une approche combinant graphe et cubes avec adaptation des opérations
  - Un bref aperçu du principe
- Pertinence par rapport aux données textuelles

  -> aspects réseau et multi-dimensionnalité au sens de l'OLAP
- Apport des contributions du Text-Mining et du Text-OLAP pour aller plus loin…

Perspectives

Généricité

- Application effective à d'autres domaines, d'autres données textuelles que les données bibliographiques

Nouvelles possibilités d'analyse

- Nouvelles mesures : mesures textuelles / autres mesures basées sur les graphes
- Analyser les liens entre plus de 2 noeuds (modélisation hypergraphes)

Extraction d'informations

Usage du Text-Mining/Text-OLAP

ERIC
ENTREPÔTS, REPRÉSENTATION & INGÉNIERIE des CONNAISSANCES

ADOC
Entrepôts et analyse de documents
3-5 mai 2017 – Lyon
En conjonction avec EDA 2017

CNRS MaDICS

Merci de votre attention

Cécile Favre*
Wararat Jakawat**
Sabine Loudcher*

* Université Lyon 2, ERIC
** Prince of Songkhla University, Thailand

Jeux de cubes pour les graphes …
Ou comment des graphes
enrichis par des cubes (GreC)
peuvent contribuer à
l'analyse de données textuelles ?

Université Lyon 1
UNIVERSITÉ LUMIÈRE LYON 2 UNIVERSITÉ DE LYON
UNIVERSITÉ DE LYON