Graph-based inferences *in a Semantic Web* Server for the Cartography of Competencies in a Telecom Valley

Fabien Gandon, Olivier Corby, Alain Giboin,
Nicolas Gronnier, Cecile Guigard
http://www-sop.inria.fr/acacia/

Context of this work: KmP

- Knowledge management solution at scale of Telecom Valley Sophia Antipolis (industrial & scientific park)
  - http://www-sop.inria.fr/acacia/soft/kmp.html
- Competency management for Telecom Valley to:
  1. provide an up-to-date snapshot of the park;
  2. facilitate partnerships between firms of the park;
  3. facilitate collaboration firms ↔ research institutes.
- Public semantic web portal & annotation of firms
  - common language to describe and compare needs and resources of firms;
  - steering committee = 11 firms working on ontologies: Amadeus, Philips Semiconductors, France Telecom, HP, IBM, Atos Origin, Transiciel, Elan IT, Qwam System Cross Systems

Some uses of semantics

- Corese conceptual graph ↔ RDF/S search engine
  - http://www-sop.inria.fr/acacia/soft/corese/
- Views of the Telecom Valley on-the-fly:
  - Combine models of the economists and data from users
  - Rules to enrich contributions & bridge different viewpoints
  - Semantic query constructors to find partners, build consortiums, extract indicators, build statistics, sort and group results, etc. [Corby et al.]
- Ontology-based metrics based on graph structure SW
  - Approximate query: semantic query relaxation [Corby et al.]
  - Interface inferences: visualization and navigation
  - Cartography: clustering algorithms to produce SVG views of similar competences in the Telecom Valley

Bonjour Admin

Se Décrire

liste des organismes

liste des entreprises

Explorer Telecom Valley

Rechercher un Partenaire

Liens directs

- Chaîne de valeur
- L'ontologie

Autres liens

- Plan du site
- Contacts
- Présentation
- Forum
- Déconnexion

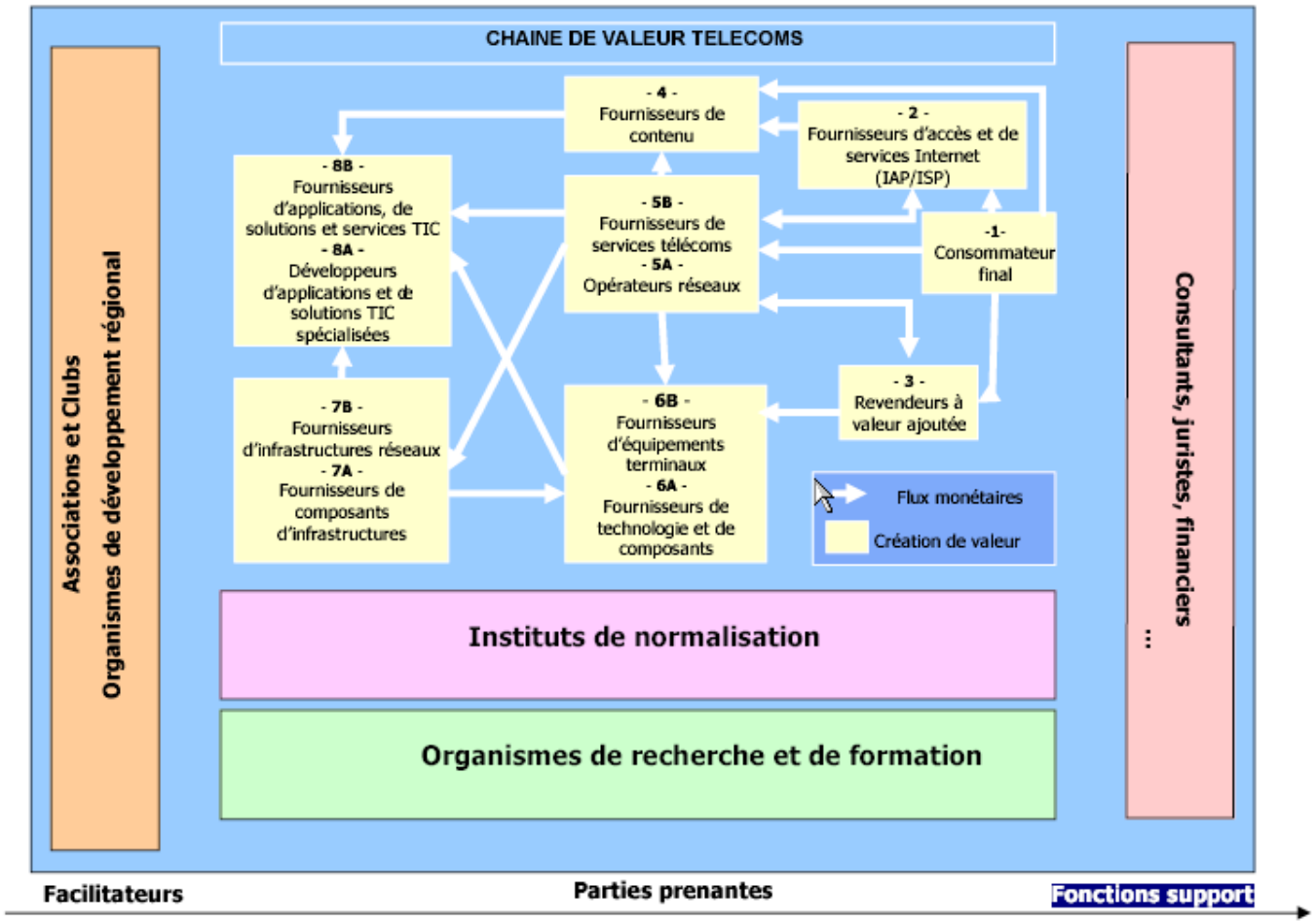
Bienvenue sur KmP "Knowledge Management Platform"

Plugin SVG installé, Version détectée : v.3

User loggé !

Découvrez les compétences de la communauté Télécom Valley de Sophia Antipolis (Alpes-Maritimes FRANCE)

Le cluster Télécoms sophipolitain



Bonjour Admin

Se Décrire
liste des organismes
liste des entreprises
Explorer Telecom Valley
•Naviguer
•Clusters et Pôles
•Connaitre les facilitateurs
•Requête type
Rechercher un Partenaire
Liens directs
- Chaîne de valeur
- L'ontologie

Autres liens
- Plan du site
- Contacts
- Présentation
- Forum
- Déconnexion

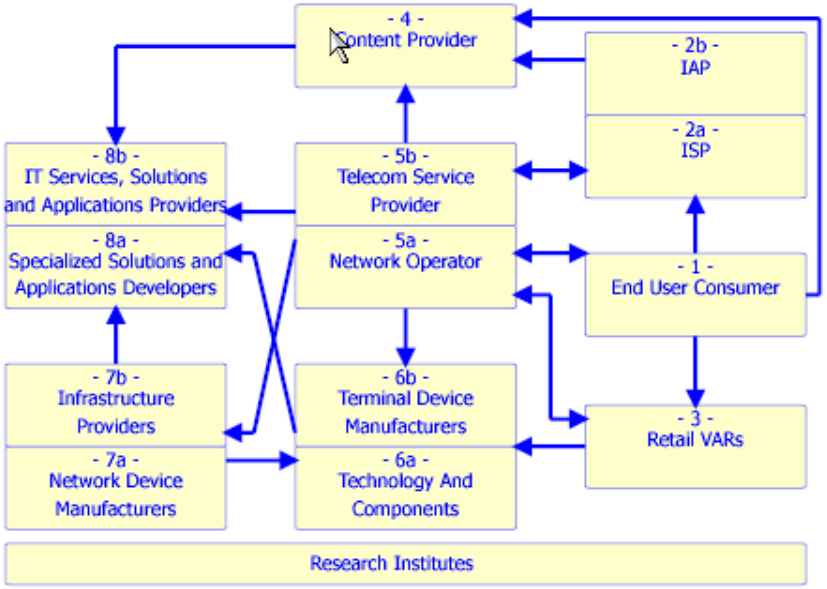
Retour

Quelle est la **répartition des acteurs dans le cluster Telecom** ?

Acteurs : entreprises et/ou équipes de recherche membres de Telecom Valley (T.V.), cartographiées dans KmP

- **11 entreprises** cartographiées dans la base **représentant un effectif de 3220 personne(s)**
- **26 équipes de recherche** cartographiées dans la base **représentant un effectif de 228 personne(s)**

soit un **total au sein de la Telecom Valley : 3448 personne(s)**



Acteur(s) sur ce segment :

Infos sur ce segment :

Bonjour Admin

Se Décrire
liste des organismes
liste des entreprises
Explorer Telecom Valley
•Naviguer
•Clusters et Pôles
•Connaitre les facilitateurs
•Requête type
Rechercher un Partenaire
Liens directs
- Chaîne de valeur
- L'ontologie

Autres liens
- Plan du site
- Contacts
- Présentation
- Forum
- Déconnexion

Quels sont les **échanges dans le cluster Telecom** ?

Retour

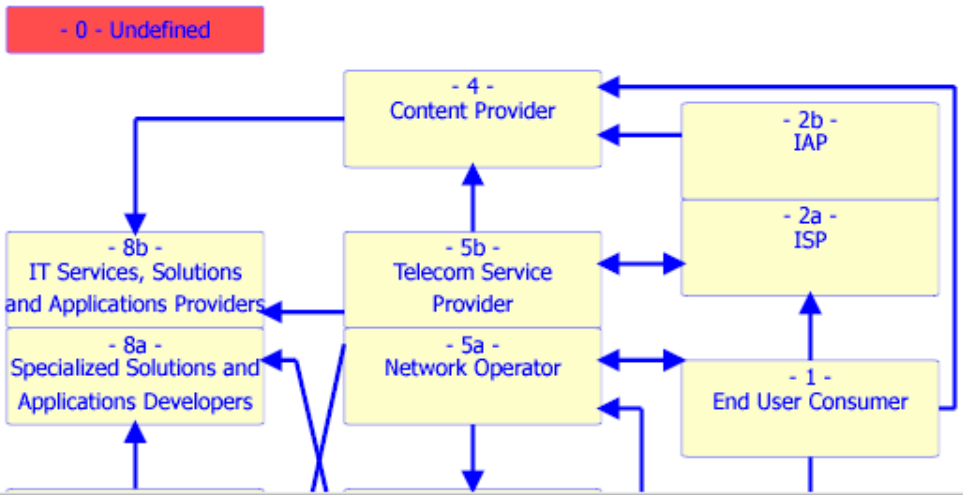
Lecture du tableau :

- une ligne: les "routes" partant d'un segment (NB: le segment 9 correspond aux accords avec les de recherche)
- une colonne: les "routes" arrivant à un segment (NB: le segment 9 correspond aux accords avec équipes de recherche)
- une valeur: le nombre de compétences pour une "route", cliquez sur une valeur pour voir les typ coopérations

	4	1	7b	2b	2a	8b	7a	5a	9	3	8a	8	5b	6b
7b			3								3			
8b	14	1	7	5	6	19	7	7	3	6	19	12	18	11
9						1		1	26			1		
8a									1		3	1		
6a						4			1			16		
5b									1		1			

Cluster Telecom

Répartition des échang



Bonjour Admin

Se Décrire
liste des organismes
liste des entreprises
Explorer Telecom Valley
Rechercher un Partenaire
Liens directs
- Chaîne de valeur
- L'ontologie

Autres liens
- Plan du site
- Contacts
- Présentation
- Forum
- Déconnexion

Compétence 1 | Compétence 2 | Partenaire | Class+ | Chercher

Définition de la PREMIERE compétence recherchée

Action — Realiser / Developper
Delivrable — Choisir
Système d'offre — Systeme d Information

Bénéficiaire

Beneficiaires hors CDV — ☐ Hors de la CDV
Beneficiaires sur CDV — Choisir

Ressource(s) mobilisee(s)

Manageriales
Technologiques
Modèles et théories

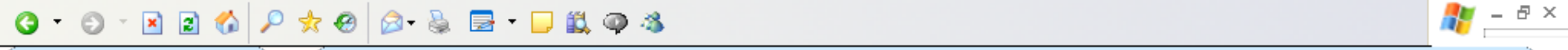
Degré stratégique

Degré Stratégique — Choisir

Type de coopération envisagé

Cooperations en RD — Choisir
Cooperations Industrielles

Modalité (Cas de coop Indus seulement)
- ☐ Forfait
- ☐ Assitance technique
- ☐ Tierce Maintenance Applicative
- ☐ Autre(s)



Bonjour Admin

Se Décrire
liste des organismes
liste des entreprises
Explorer Telecom Valley
Rechercher un Partenaire
Liens directs
- Chaîne de valeur
- L'ontologie

Autres liens
- Plan du site
- Contacts
- Présentation
- Forum
- Déconnexion

Compétence 1 | Compétence 2 | Partenaire | Class+ | Chercher

Définition de la PREMIERE compétence recherchée

Action: Concevoir
Delivrable: Choisir
Système d'offre: Radio-frequences

Bénéficiaire

Beneficiaires hors CDV: ☐ Hors de la CDV
Beneficiaires sur CDV: Choisir

Ressource(s) mobilisee(s)

Manageriales
Technologiques: Wifi
Modèles et theories

Degré stratégique

Degré Stratégique: Choisir

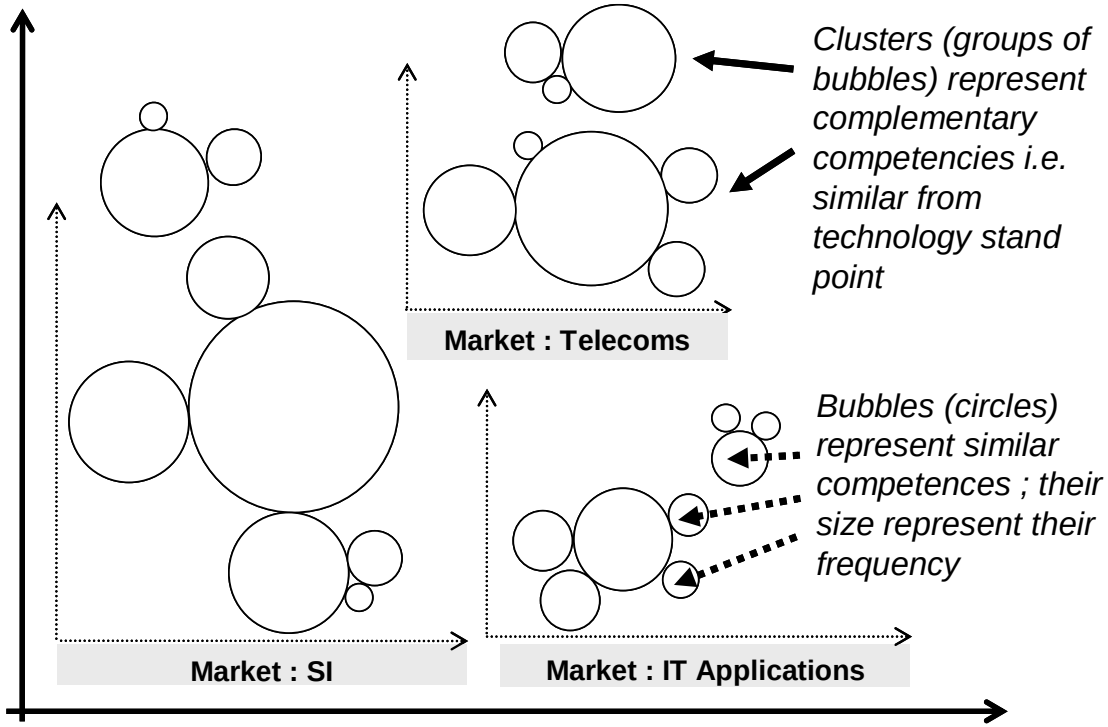
Type de coopération envisagé

Cooperations en RD: Choisir
Cooperations Industrielles

Modalité (Cas de coop Indus seulement)
- ☐ Forfait
- ☐ Assitance technique
- ☐ Tierce Maintenance Applicative
- ☐ Autre(s)

Clustering competencies

- Dynamic up-to-date cartography of competencies in technological park and grouped in clusters.
  - built from visual representations the users are used to;
  - they came up with a "readable representation of the clusters of competencies in the technological pole":



1: Readability implies an expressivity

- Motivating scenario: requirement on the ontology
- Models from management & economics [Lazaric & Thomas]
- Four facets to describe a competence
  - E.g. "designing microchips for the 3G mobile market using GSM, GPRS and UMTS"
  - an action (design);
  - a deliverable (microchip);
  - a market (3G mobile technology);
  - a set of resources (GSM, GPRS, UMTS)

Similarity of competence ≡ similar actions and resources

- Each facet is formalized in a module of the ontology. ~1400 concept types and 90 relation types

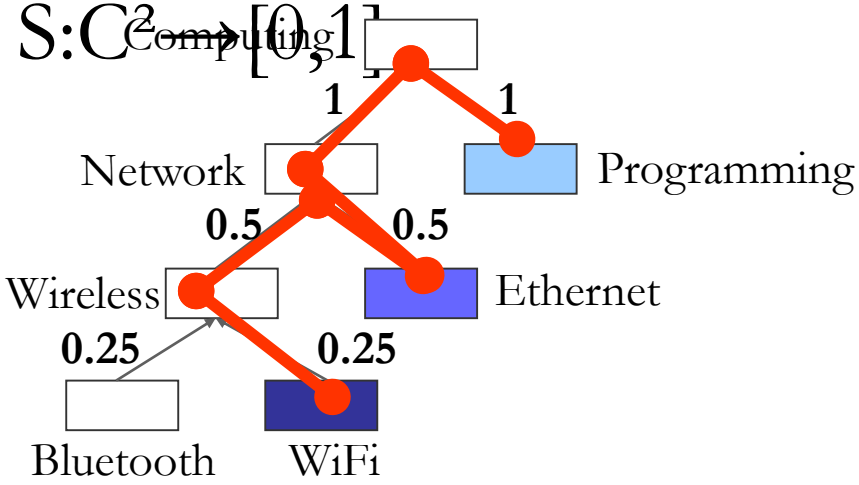
2: Readability implies inferences

- Monothetic clustering algorithm [Jain, Murty & Flynn]
  1. chose a market to limit analysis: sub-types are considered, ancestors/siblings are discarded. → closure of subtype;
  2. group competences by *similarity* of resources (cluster);
  3. in each cluster, group competences by *similarity* of action (bubbles).
- Evaluating conceptual relatedness [Collins & Loftus] [Rada *et al.*] [Resnik][Jiang & Conrath][Wu &Palmer]
  - early works on humans' semantic memory [Quillian]
  - many forms of relatedness e.g. complementarity, similarity
  - here similarities based on shared definitional features
  - concept type hierarchy supporting similarity reasoning
  - apply spreading algorithm to simulate distance

Example in conceptual graphs [Sowa]

- Non binary projection based on a similarity

$S: C^2 \rightarrow [0,1]$



Pattern of W3C ~ approach 4

Programming → Ethernet = 3
Ethernet → Wifi = 3

Programming → Ethernet = 2,5
Ethernet → Wifi = 1,25

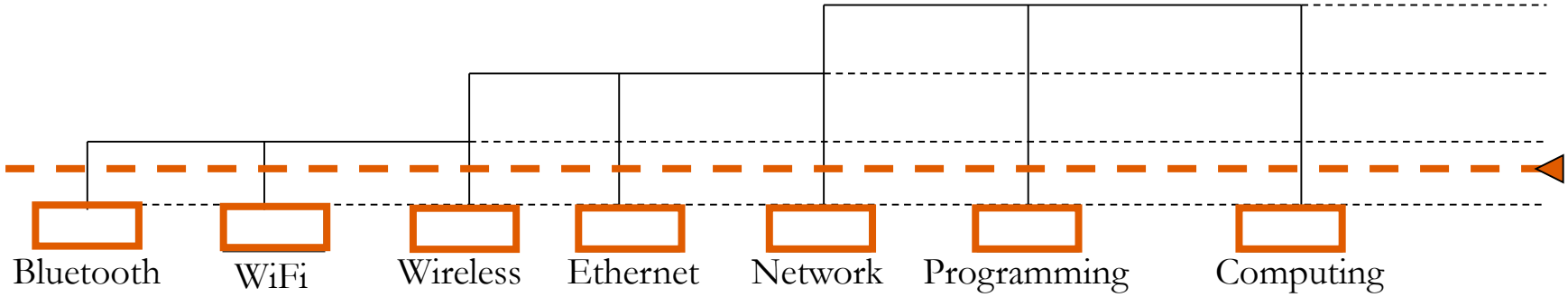
- When the targeted module of hierarchy is a tree:

$$dist(t, father(t)) = \left[\frac{1}{2}\right]^{depth(father(t))}$$

$$dist(t_1, t_2) = \frac{1}{2^{depth(LCST(t_1,t_2))-2}} - \frac{1}{2^{depth(t_1)-1}} - \frac{1}{2^{depth(t_2)-1}}$$

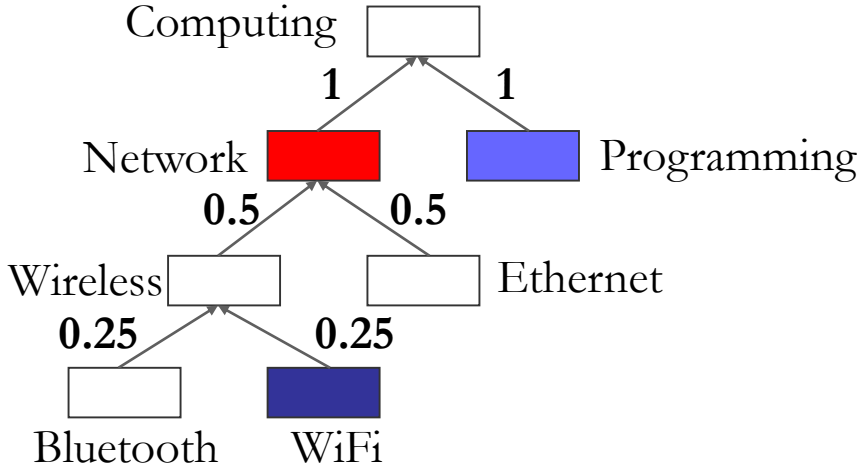
From distance to clusters

- Cluster competencies and control level of detail
  - a dendrogram provides a clustering with chosen granularity



  - a dendrogram relies on an ultrametric

$dist(t1,t2) \leq max(dist(t1,t'), dist(t2,t'))$ *for any* t'



t_1=Wifi t_2= Programming

$dist(t_1,t_2)=2,75$

t'= Network

$dist(t_1,t')=0,75$ and $dist(t_2,t')=2$

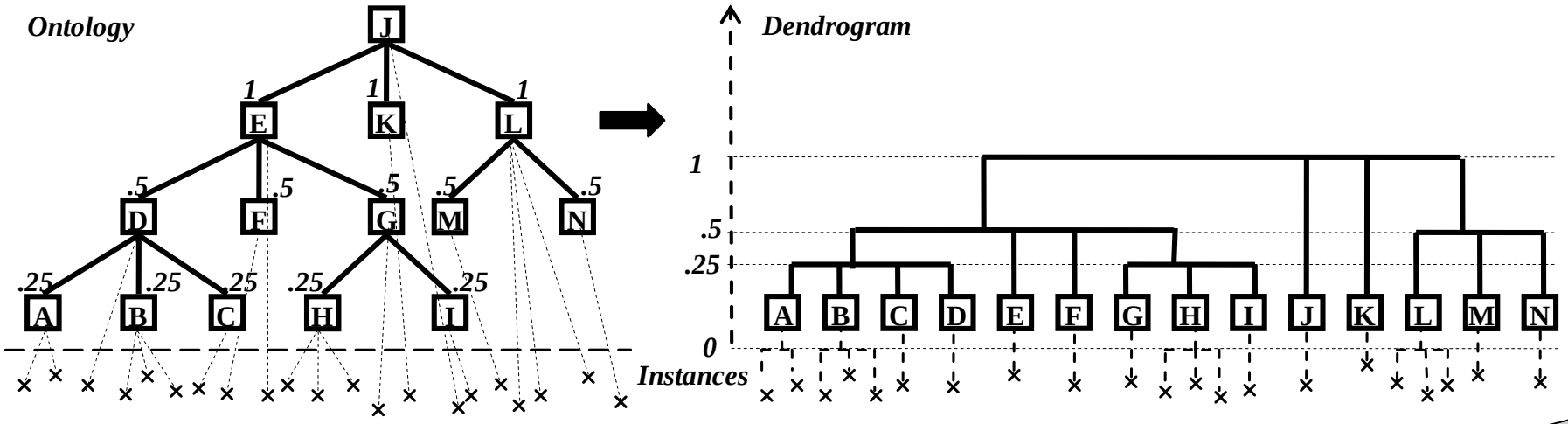
$dist(t_1,t_2) > max(dist(t_1,t'), dist(t_2,t'))$

Transform hierarchy into dendrogram

- First transformation, maximal distance to the least common supertype:

$$dist_{MH}(t_1,t_2) = \max_{\forall t<LCST(t_1,t_2)}(dist(t,ST(t)) = \left[\frac{1}{2}\right]^{depth(LCST(t_1,t_2))}$$

- But no added value compared to the direct use of the ontology depth to control the level of detail

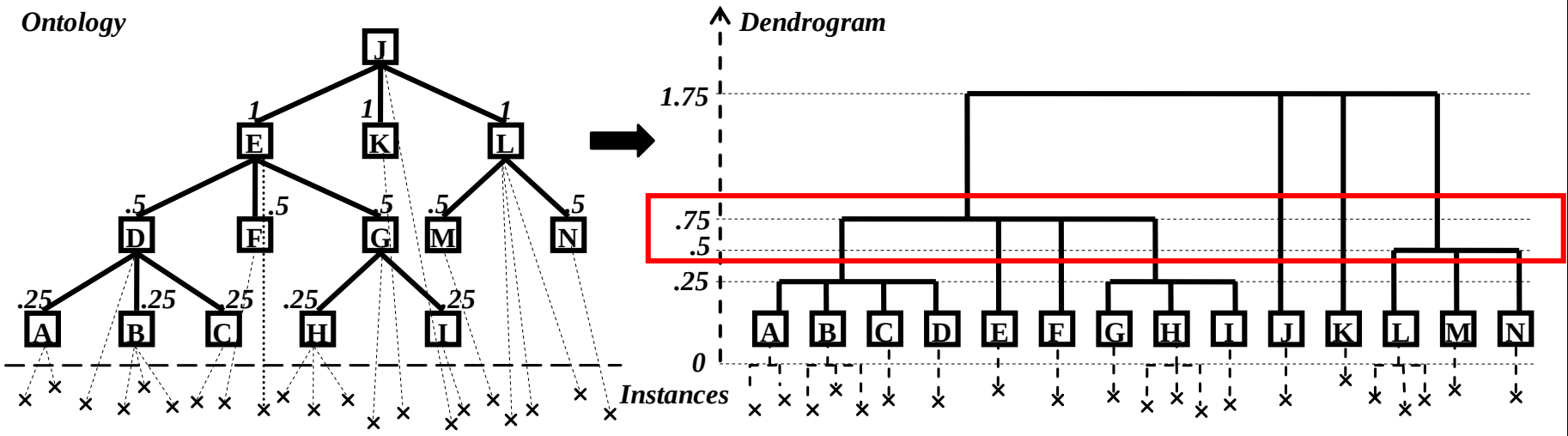


Refine transformation

- Take into account depth of hierarchy below least common supertype: a measure of the loss of details

$$dist_{CH}(t_1, t_2) = \max_{\forall st \leq LCST(t_1,t_2)} \left(dist\left(st, LCST(t_1, t_2)\right)\right) \text{ when } t_1 \neq t_2$$

$$dist_{CH}(t_1, t_2) = 0 \text{ when } t_1 = t_2$$



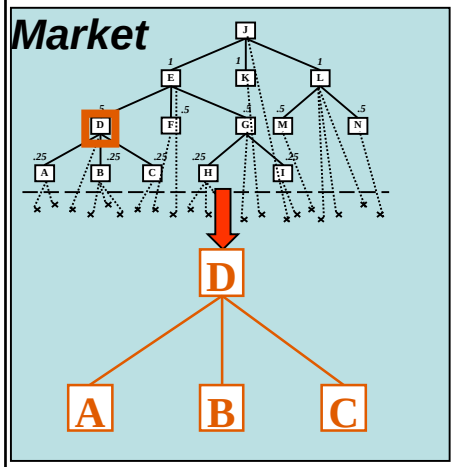
Characteristics of this distance

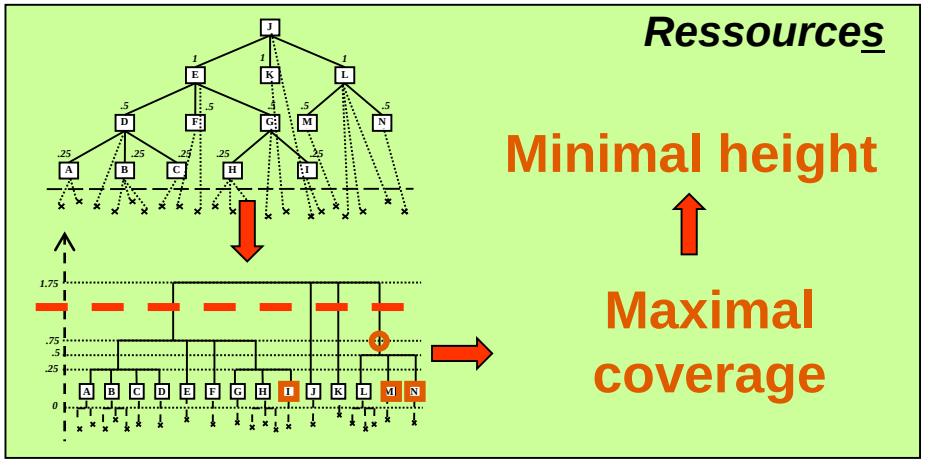
- $dist_{CH}$ is an ultrametric (see paper)
  - $dist_{CH}(t,t) = 0$
  - $dist_{CH}(t_1,t_2) = dist_{CH}(t_2,t_1)$
  - $dist_{CH}(t_1,t_2) = 0 \Rightarrow t_1 = t$
  - $\forall t'\ dist_{CH}(t_1,t_2) \leq max(dist_{CH}(t_1,t'), dist_{CH}(t_2,t'))$
- Thus $dist_{CH}$ generates more levels than $dist_{MH}$ and the proportionally by the square of maximal depth
- One bound per depth *i.e.* $dist_{Max}(d+1) < dist_{Min}(d)$
  - the clustering respects the ontology hierarchy
  - children of a shallow class will be grouped first
- A name can be given to every cluster Cl

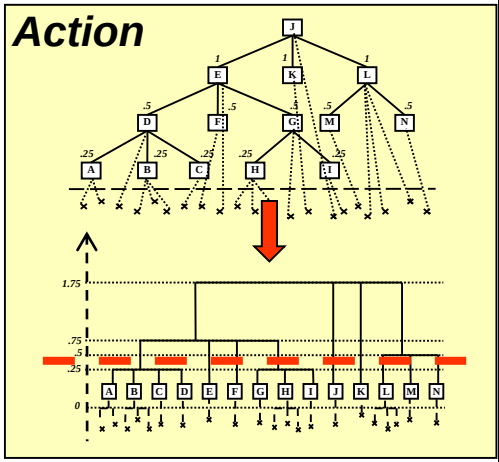
$Name(Cl) = Name\ (LCST(\{t\ ;\ type\ t \in Cl\}))$

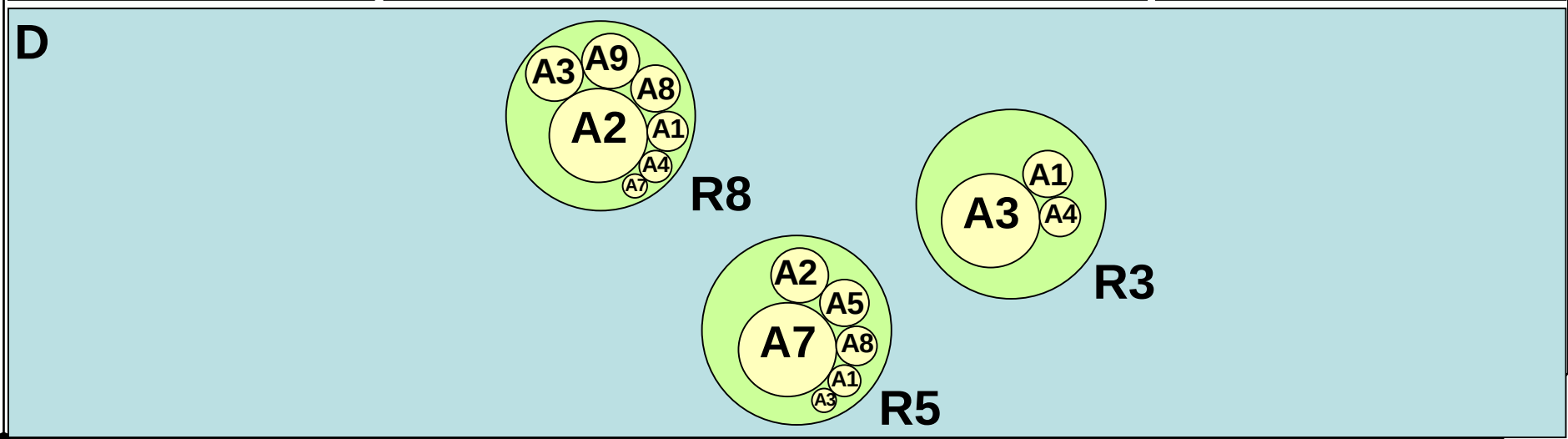
Ready for cartography

- 2 dendrograms from 2 ontologies (resources, actions) to control the vocabulary used for descriptions



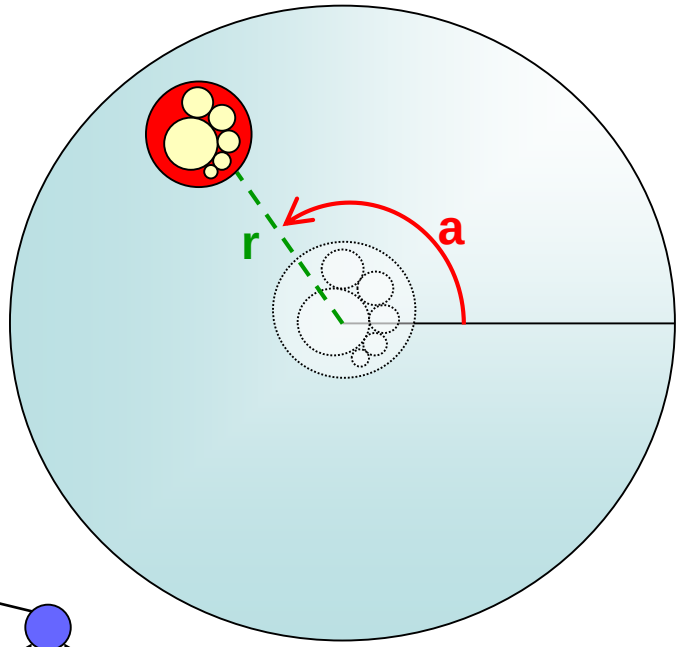


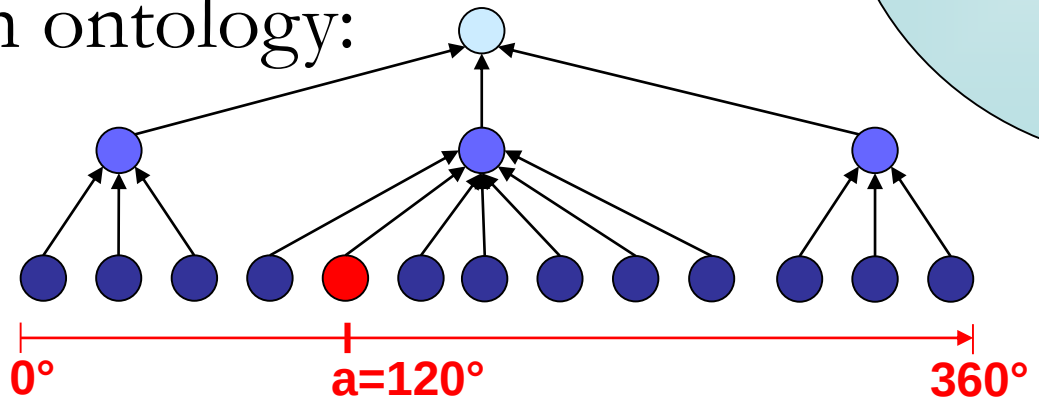




Display: ontology & angular distance

- The size of the cluster → radius
- The type of resources used by competences in cluster → angle derived from place of resource classes in ontology:





  - top-down favours structure of upper modules
  - bottom-up favours the detailed branches
  - here: top-down division to divide equally between domains

Technologie

Bonjour Admin

Se Décrire

liste des organismes

liste des entreprises

Explorer Telecom Valley

- Naviguer
- Clusters et Pôles
- Connaitre les facilitateurs
- Requête type

Rechercher un Partenaire

Liens directs

- Chaîne de valeur
- L'ontologie

Autres liens

- Plan du site
- Contacts
- Présentation
- Forum
- Déconnexion

Identifier le cluster du système d'offre

+Environnement Telecom

Soumettre la requête

Guide de lecture et légende

Grappe (racine technologique)
Aucune

Bulle
Aucune

Système de ressources
Aucune

Acteur(s)
Aucune

Vous pouvez choisir des niveaux de précision

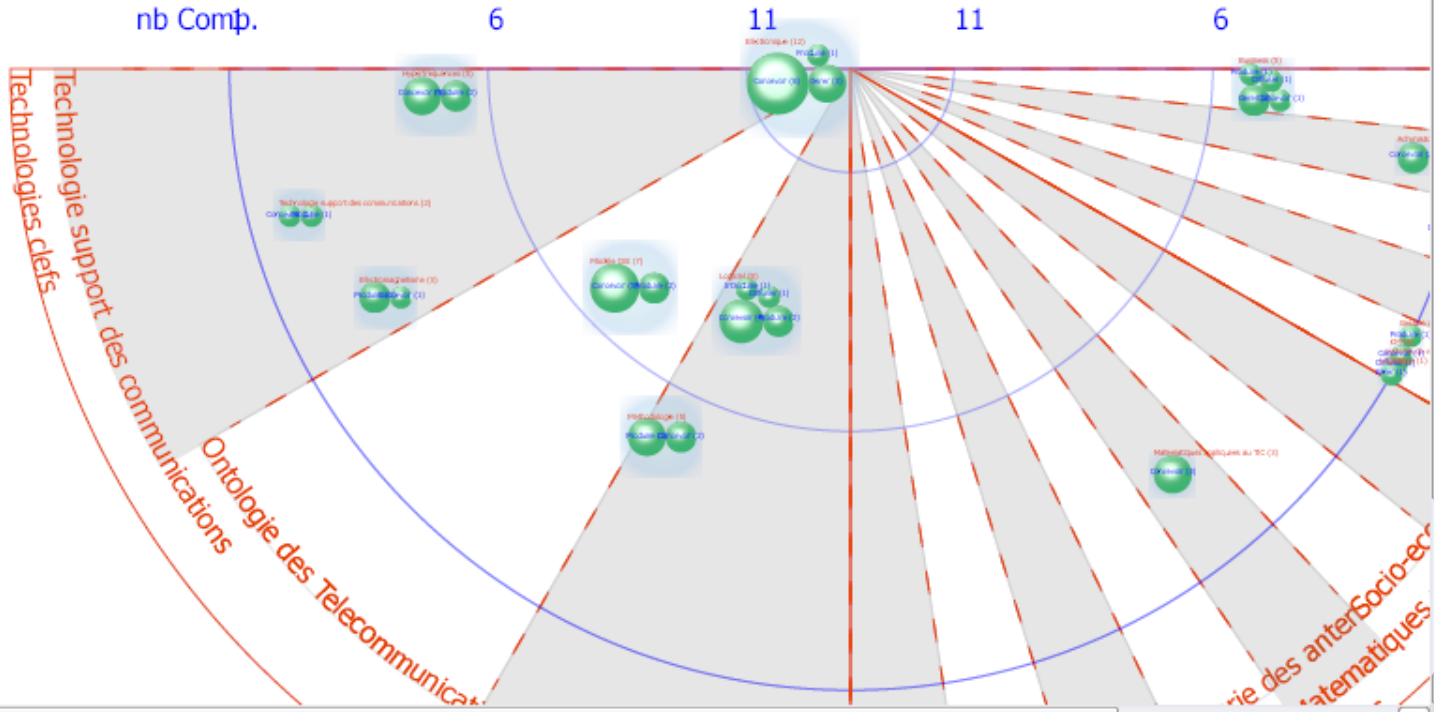
Niveau de détail des Action(s) Faible

Niveau de détail des Ressources(s) Faible

Limiter la recherche

aux entreprises aux équi

Le cluster Environnement Telecom



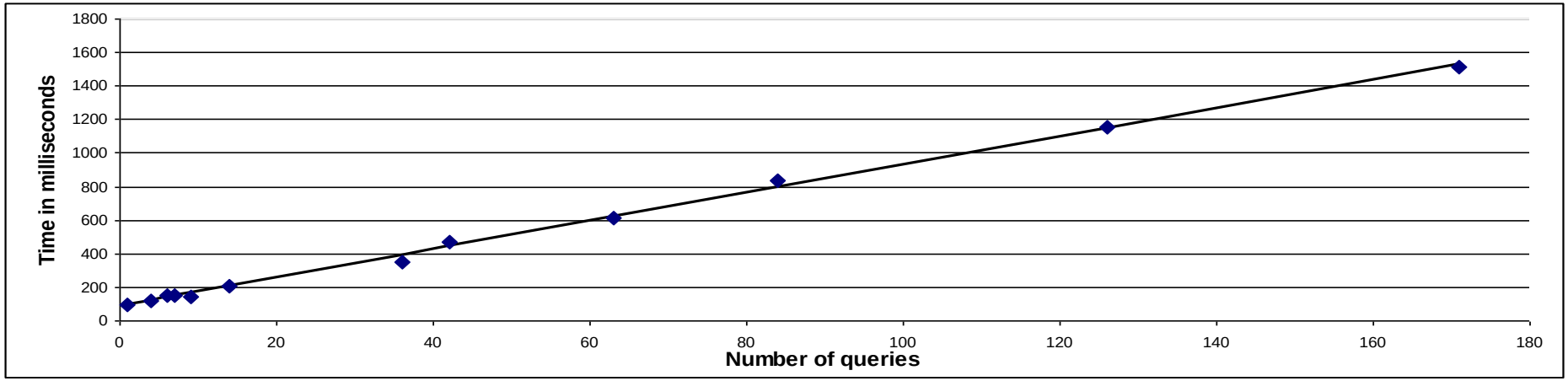
Evaluation (1/2)

- Usage & ergonomics
  - simple widgets & real-time display for levels of detail
  - cluttered centre while it contains important information
  - customize graphical elements (colors, angles, labels)
  - ability to post-process the results
  - ability to capture the dynamics (history, previsions)
  - provide predicted time to compute a clustering
- Complexity & response time (n, m number resources, actions)
  - Parsing the schema to build the tree: $O(n+m)$
  - Initializing depth and angular distribution: $O(n+m)$
  - Sorting the dendrogram: $O(n \times log(n) + m \times log(m))$
  - n' and m' levels of detail in dendrograms of resources and actions : $n' \times m'$ queries for clustering

    $1 \leq n' \leq 596$ $1 \leq m' \leq 118$ $1 \leq n' \times m' \leq 70\,328$

Evaluation (2/2)

- Pentium 4 M / 1.7GHz / 512 Mb Windows XP Pro
  - average minimum time (1 query) is 86 milliseconds.
  - average maximum time (70 328 queries) is 11 minutes.
  - average typical time (981 queries) is 9 seconds.



- linear regression ~ $y = x \times 8.42 + 89.24$
  - rapid way to foresee the duration of a clustering
  - level above which it is better to rely on reporting in batch mode rather than real-time calculation; typically above 15 seconds of response time *i.e.* above $n' \times m' = 2000$

Conclusion

- KmP: industrialization phase
  - from 20 pilot companies to 70 user companies
  - more and more public research centers
  - taken over by a start-up
- Graph structure of semantic web formalisms and inferences in retrieval, clustering, visualization
  - proved the characteristics of the metrics and inferences
  - inferences: domain-independent and reusable
  - other inferences on these graph structures
- Experiments to evaluate and compare simulated metrics with the ones humans naturally use.

- **Thank you.**

Fabien.Gandon
@sophia.inria.fr
http://www-sop.inria.fr/acacia/

Acacia INRIA SOPHIA ANTIPOLIS W3C MEMBER