

Modélisation d'expertise scientifique pour la constitution de comités de programme

Soutenance de thèse de doctorat

Hong Diep TRAN

Directeur de thèse : Gilles HUBERT

Co-directeur de thèse : Guillaume CABANAC

Institut de recherche en informatique de Toulouse – IRIT

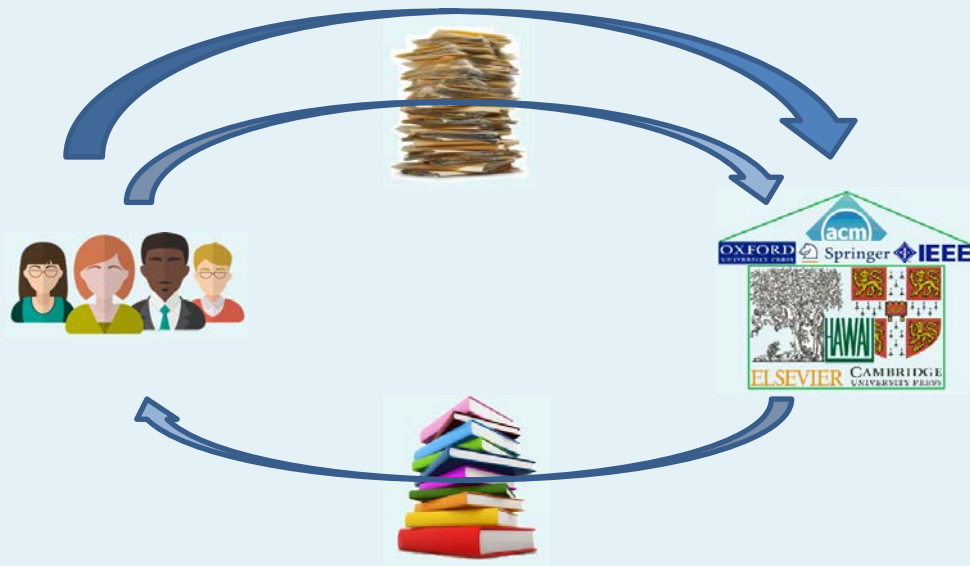
Université Toulouse III Paul Sabatier



Introduction

1. Publier et diffuser les découvertes scientifiques

Publication et pilotage de la recherche dans la communauté scientifique



Introduction

2. Guider les recherches scientifiques

Publication et pilotage de la recherche dans la communauté scientifique



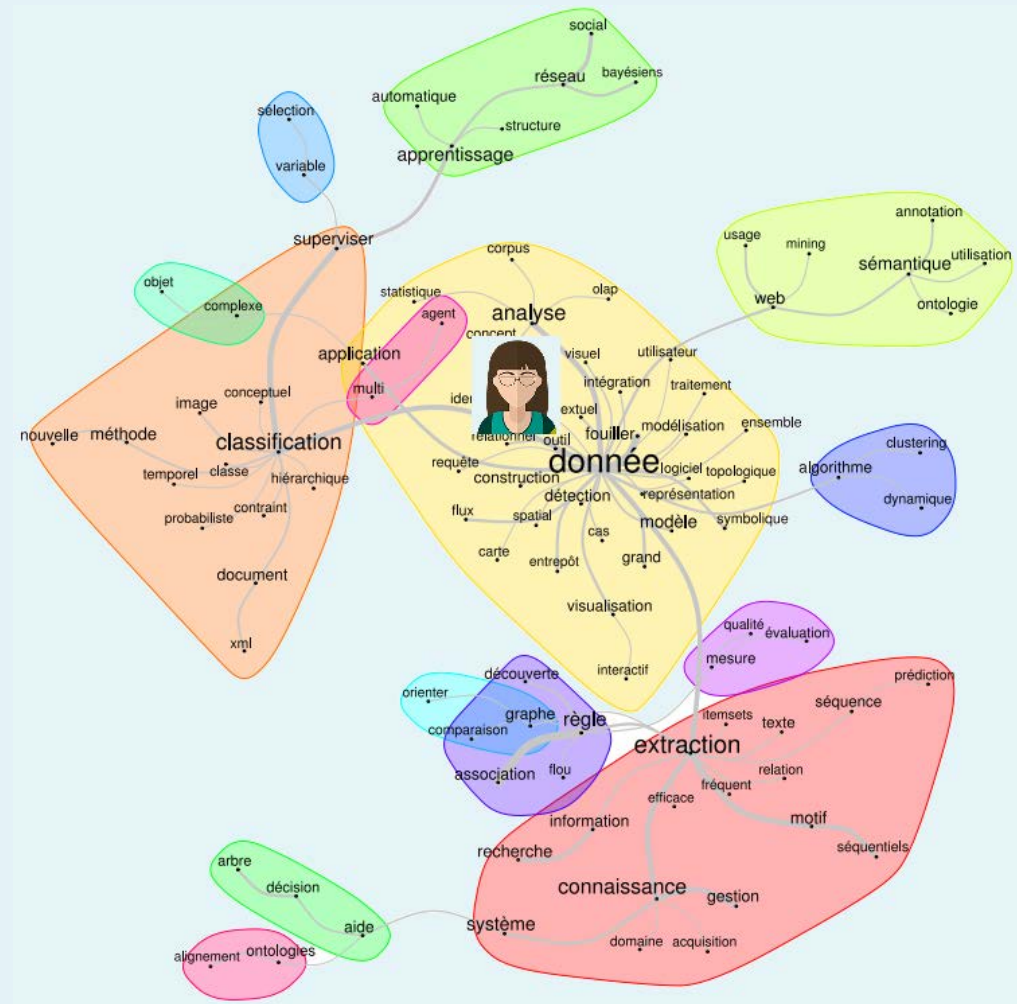
Introduction

3. Orientation des recherches scientifiques

« Positionnement » de la communauté scientifique :

- Sujets au cœur / à la périphérie
- Nouveaux concepts émergents
- Collaborations en science
- Proximité conceptuelle entre les articles et entre les chercheurs
- Travaux les plus plébiscités

➤ Nécessité des études scientométriques



Contexte

- Pression et demande de diffusion des savoirs et des découvertes scientifiques
- Revues et, notamment, conférences sont des vecteurs de publication

Contexte

Évaluation des soumissions de contributions

- Évaluation par les pairs
- Comités d'édition et comités de programme
- Président·e·s des comités

Contexte

Évaluation des soumissions de contributions

Rôle des membres du comité de programme

- Solliciter les chercheurs à soumettre leurs travaux de recherche
- Expertiser les articles soumis et sélectionner ceux qui seront publiés

Contexte

Évaluation des soumissions de contributions

Rôle des membres du comité de programme

Souhaits pour constituer le comité de programme

- Élaborer le CP contenant les membres les plus pertinents
- Renouveler une certaine partie du CP pour les éditions annuelles

Contexte

Évaluation des soumissions de contributions

Rôle des membres du comité de programme

Souhaite pour constituer le comité de programme

Difficultés et risques

- Nécessité de connaître suffisamment les experts dans toutes les thématiques promues par la conférence
- Difficulté d'estimer l'expertise du CP envisagé vis-à-vis de la conférence

Suggestion de membres du comité de programme

- Modélisation de la conférence et du domaine scientifique
- Définition d'indicateurs scientométriques

Évaluation de la proposition

- Évaluation de la pertinence des suggestions de CP

PLAN

- 1 Introduction
- 2 Positionnement
- 3 Modèle d'expertise
- 4 Indicateurs d'expertise de CP
- 5 Suggestion de CP
- 6 Conclusion et Perspectives

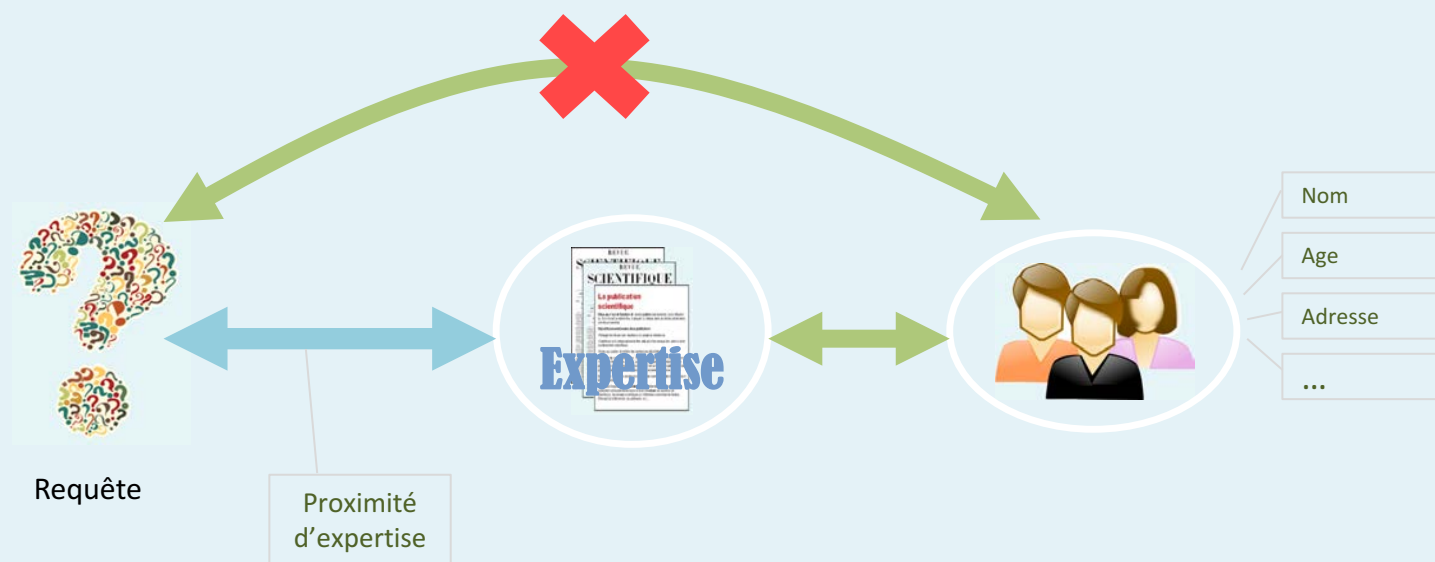
PLAN

- 1 Introduction
- 2 **Positionnement**
- 3 Modèle d'expertise
- 4 Indicateurs d'expertise de CP
- 5 Suggestion de CP
- 6 Conclusion et Perspectives

1. Recherche d'expert et d'expertise

Défis principaux

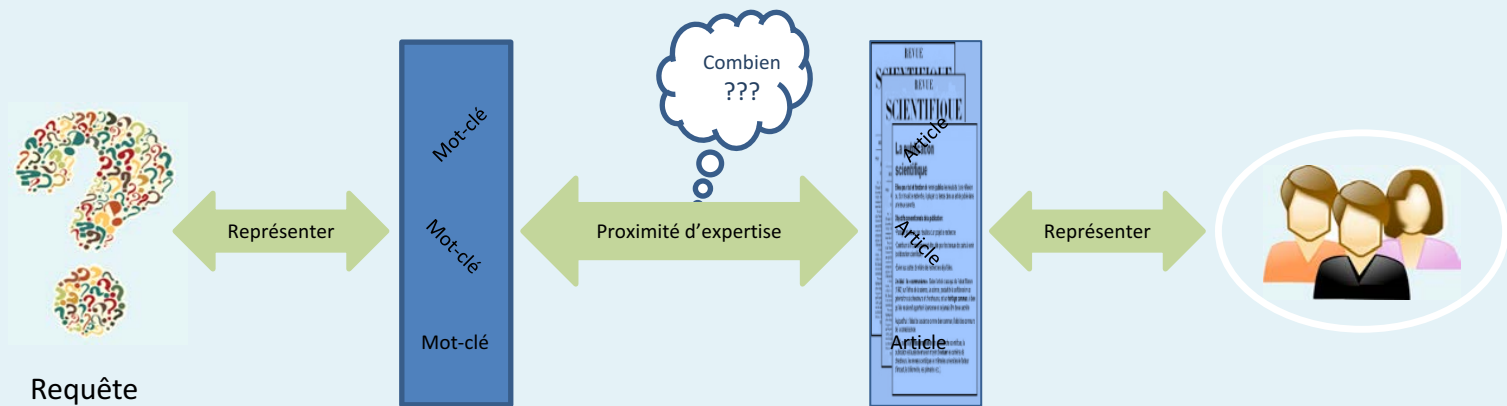
- L'expertise est liée à des « connaissances implicites »
- L'expertise n'est pas définie clairement
- L'expert n'est pas représenté comme une unité d'information qu'on peut rechercher



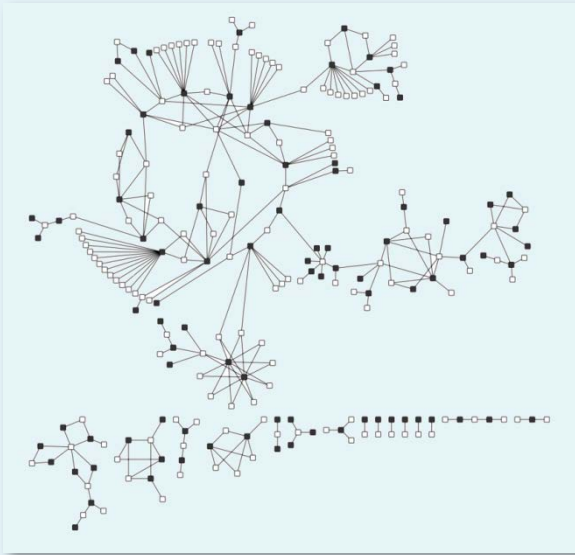
1. Recherche d'expert et d'expertise

Profilier et trouver l'expert

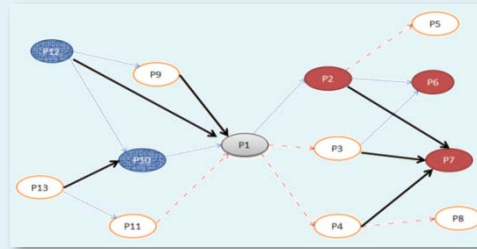
- Modéliser la requête
- Modéliser l'expert candidat
- Associer la requête et l'expert candidat



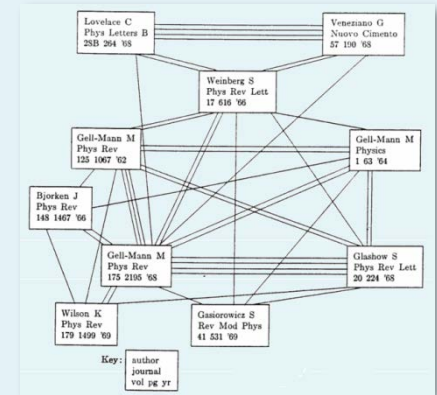
1. Recherche d'expert et d'expertise



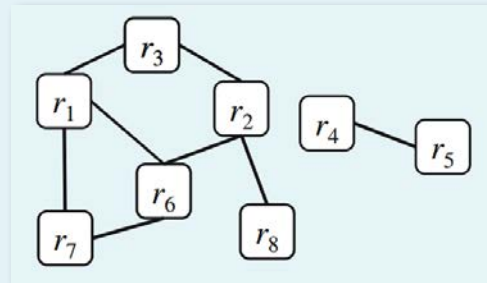
Graphe d'auteurs



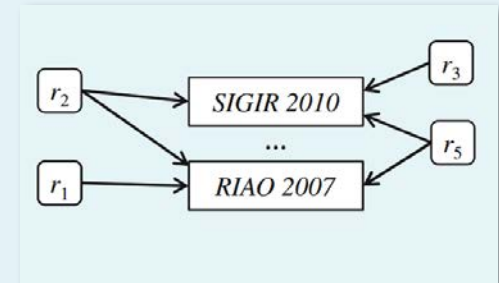
Graphe de citations



Graphe de co-citations



Graphe de co-auteurs

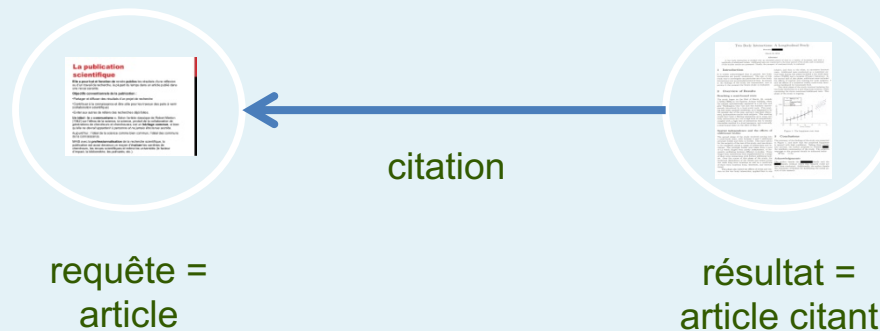


Graphe de lieux de publications

1. Recherche d'expert et d'expertise

Approches orientées graphe

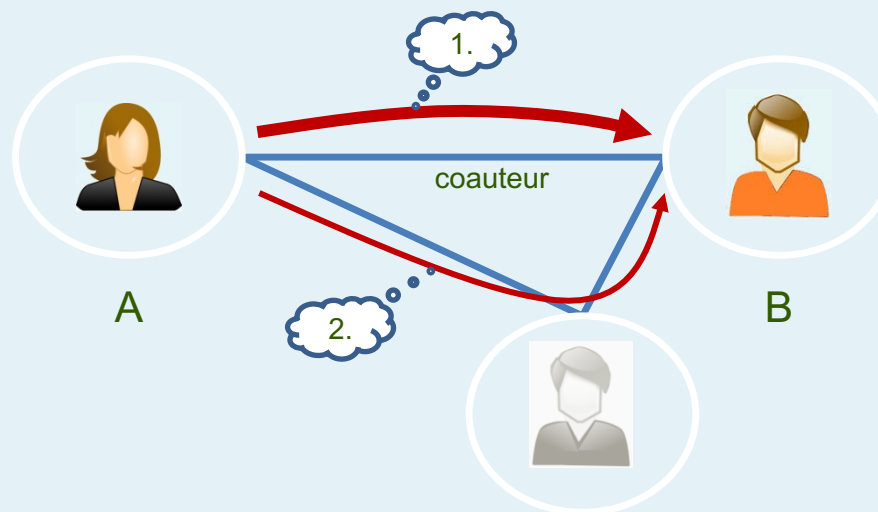
- Objectif : recherche d'expertise
- Principe : relation d'expertise entre les items



1. Recherche d'expert et d'expertise

Approches orientées graphe

- Objectif : recherche d'expertise
- Principe : relation d'expertise entre les items
- Comment quantifier la similarité d'expertise ?
 1. détermination par la relation directe
 2. détermination par le structure du graphe



2. Approche orientée graphe pour le problème de CP

Énonciation du travail

- Requête : l'édition future d'une conférence
- Résultat : liste ordonnée des experts candidats pour constituer le comité de programme



2. Approche orientée graphe pour le problème de CP

Énonciation du travail

Profilage d'expertise

- L'édition donnée est représentée par les éditions précédentes
- Un expert candidat est représenté par ses articles publiés et ses participations au CP



Représentant de la requête

Représentant de l'expert

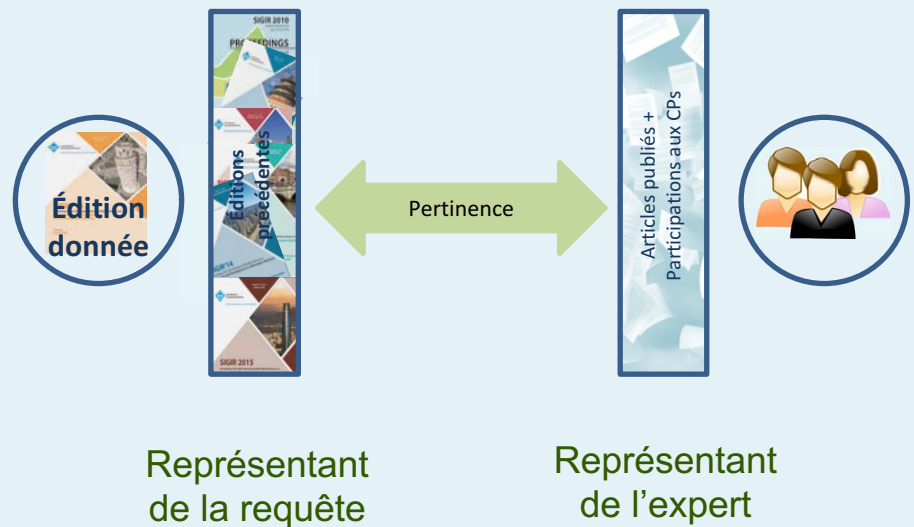
2. Approche orientée graphe pour le problème de CP

Énonciation du travail

Profilage d'expertise

Découverte d'expert

- Graphe hétérogène modélisant l'expertise scientifique multi-facette
- La pertinence de l'expert candidat est estimée par les différentes relations d'expertise avec la conférence donnée



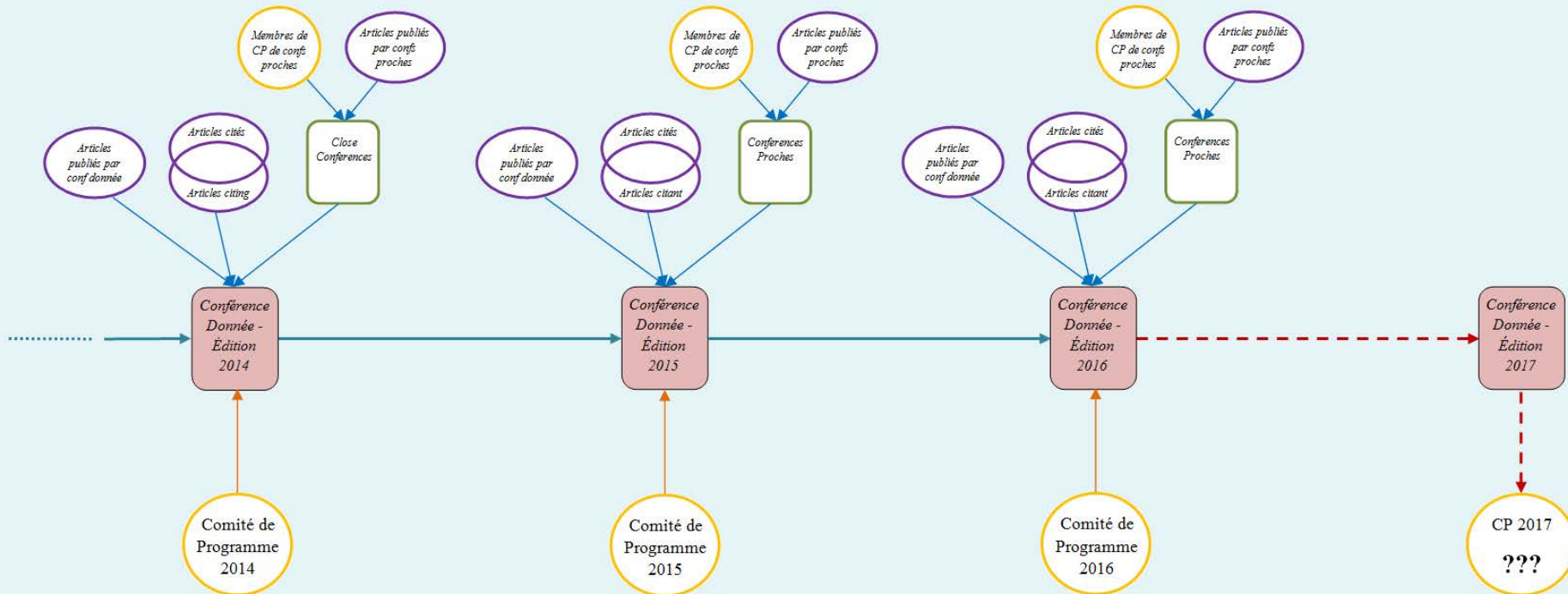
PLAN

- 1 Introduction
- 2 Positionnement
- 3 **Modèle d'expertise**
- 4 Indicateurs d'expertise de CP
- 5 Suggestion de CP
- 6 Conclusion et Perspectives

1. Modélisation

Preuves d'expertise

- Actes de publication dans le passé
- Données disponibles
- Indication de la relation d'expertise



1. Modélisation

Preuves d'expertise

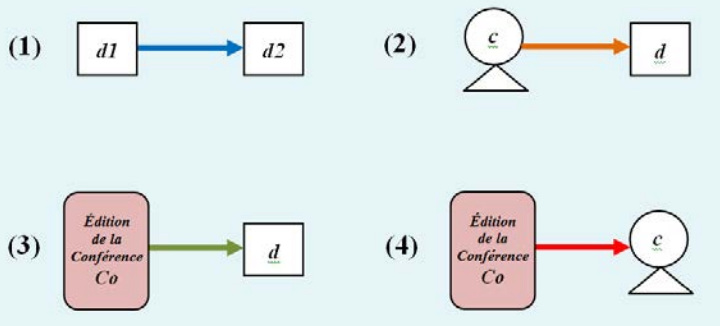
Graphe hétérogène pondéré

● Nœuds :

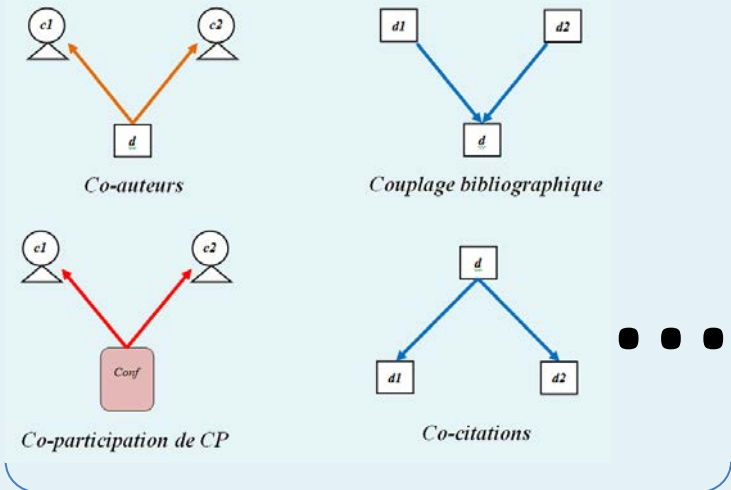
1. Édition de conférence
2. Article
3. Expert

● Arêtes :

1. Lien de citation
2. Lien de signature
3. Lien de publication
4. Lien de participation au CP



Liens principaux : relations directes

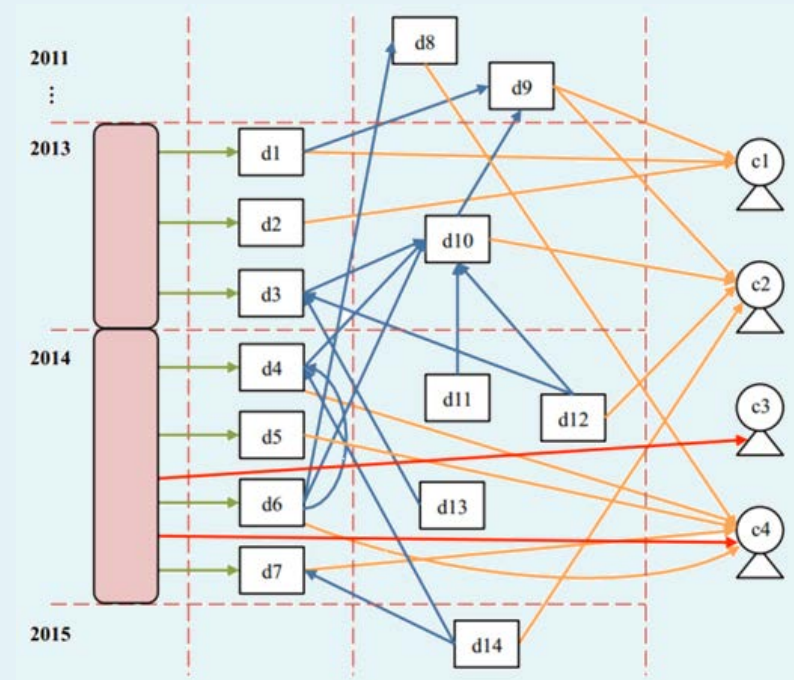
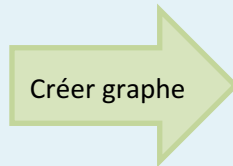
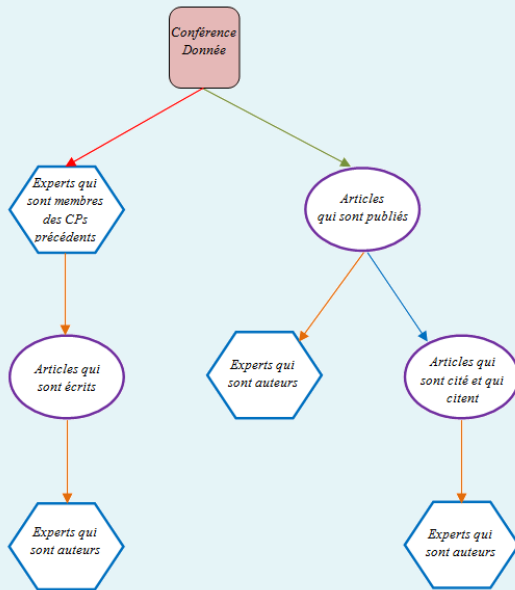


Relations déduites (par exemple)

2. Modèle de conférence

Sphère académique d'une conférence

- Graphe : la conférence donnée est au centre
- Liste des candidats : les relations proches avec la conférence

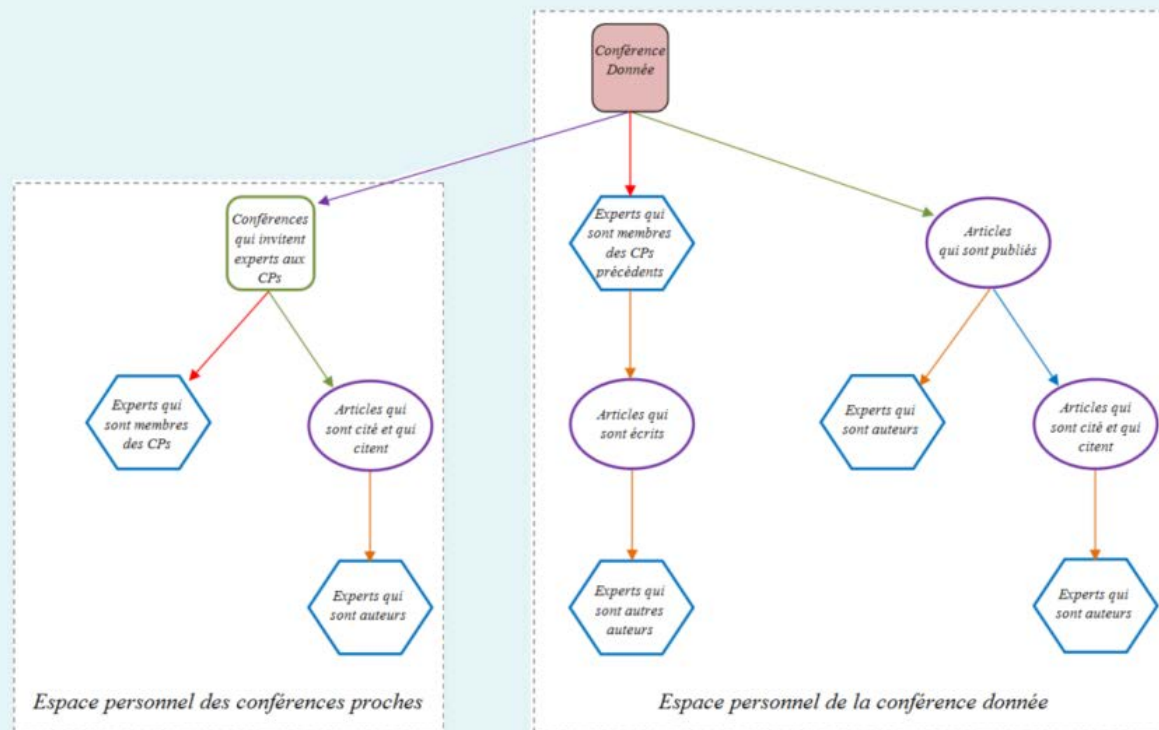


Déterminer les items proches pour la conférence

3. Modèle de domaine scientifique

Sphère académique des conférences

- Les conférences d'un même domaine : conférence donnée et les conférences proches
- Liste des candidats : les relations « proches » avec la conférence donnée et « très proches » avec une conférence proche

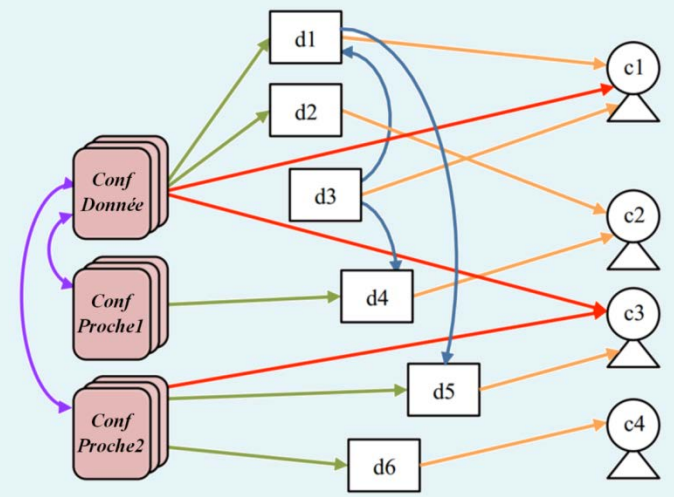
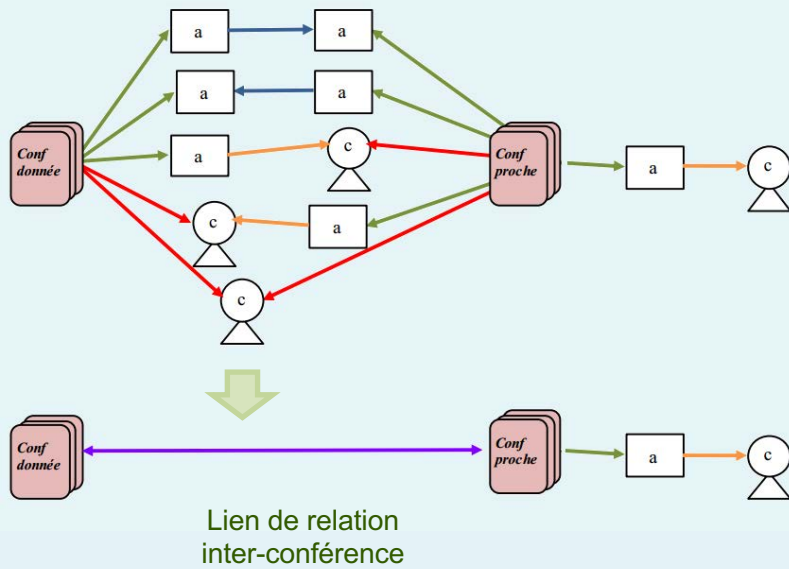


3. Modèle de domaine scientifique

Sphère académique des conférences

Conférences proches

- Les conférences d'un même domaine : conférence donnée et les conférences proches
- Liste des candidats : les relations « proches » avec la conférence donnée et « très proches » avec une conférence proche



PLAN

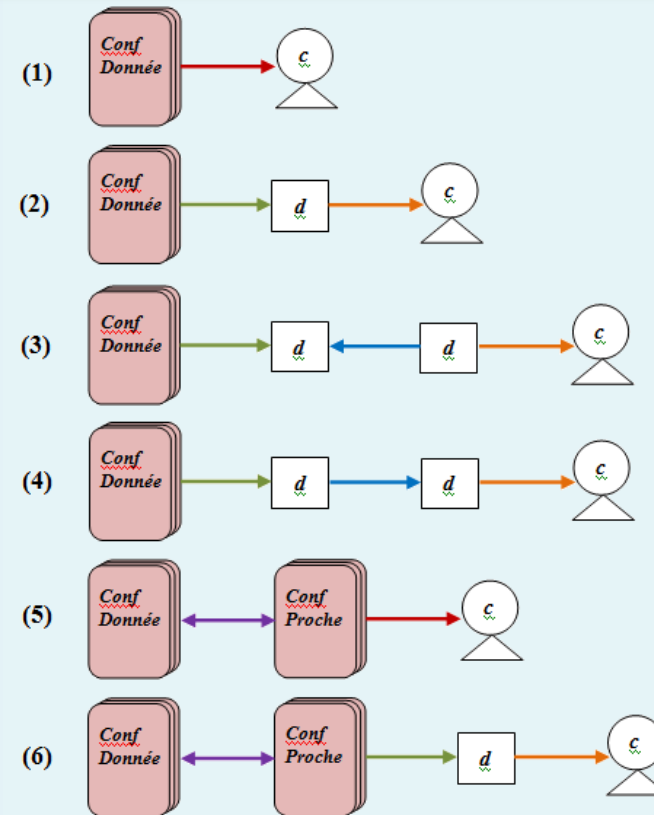
- 1 Introduction
- 2 Positionnement
- 3 Modèle d'expertise
- 4 Indicateurs d'expertise de CP
- 5 Suggestion de CP
- 6 Conclusion et Perspectives

1. Composition des comités de programme

Rôle – facette d'expertise

- 6 rôles de l'expert par rapport à la conférence donnée
- 6 facettes d'expertise correspondantes

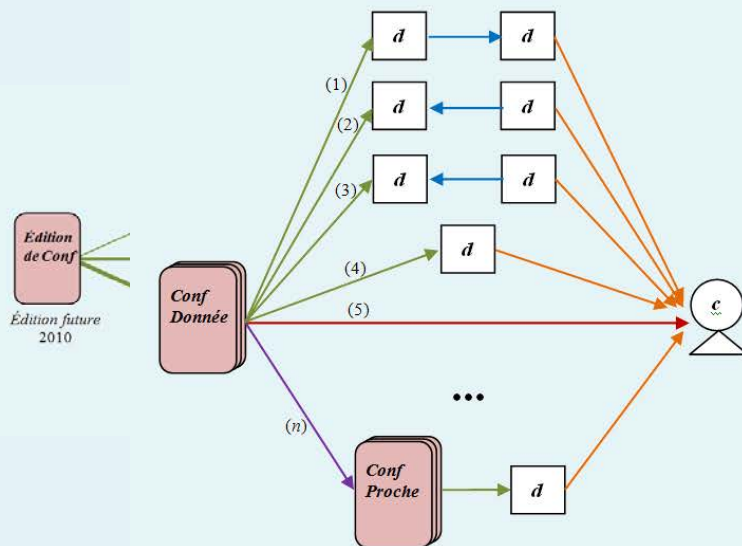
1. Membre interne de CP
2. Auteur interne
3. Auteur externe citant
4. Auteur externe cité
5. Membre externe
6. Auteur externe proche



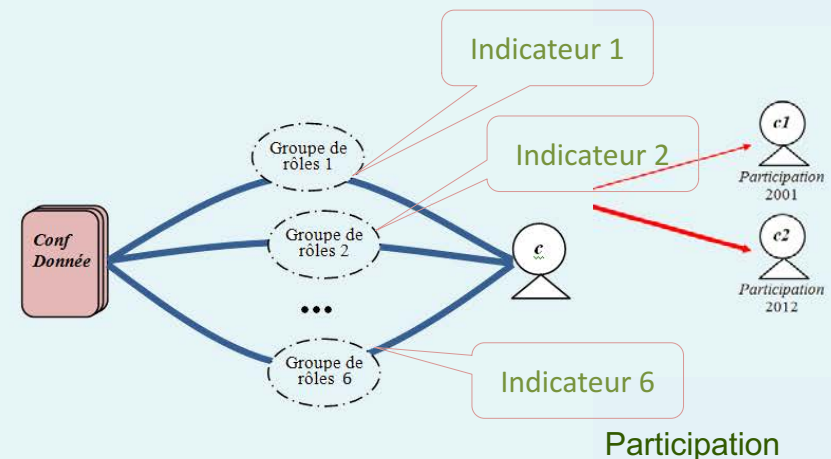
1. Composition des comités de programme

Indicateurs scientométriques des rôles d'un expert

- 6 indicateurs scientométriques correspondant à 6 rôles qui représentent les différentes facettes d'expertise de l'expert
- Quantification de chaque indicateur



Estimation de la force du chemin

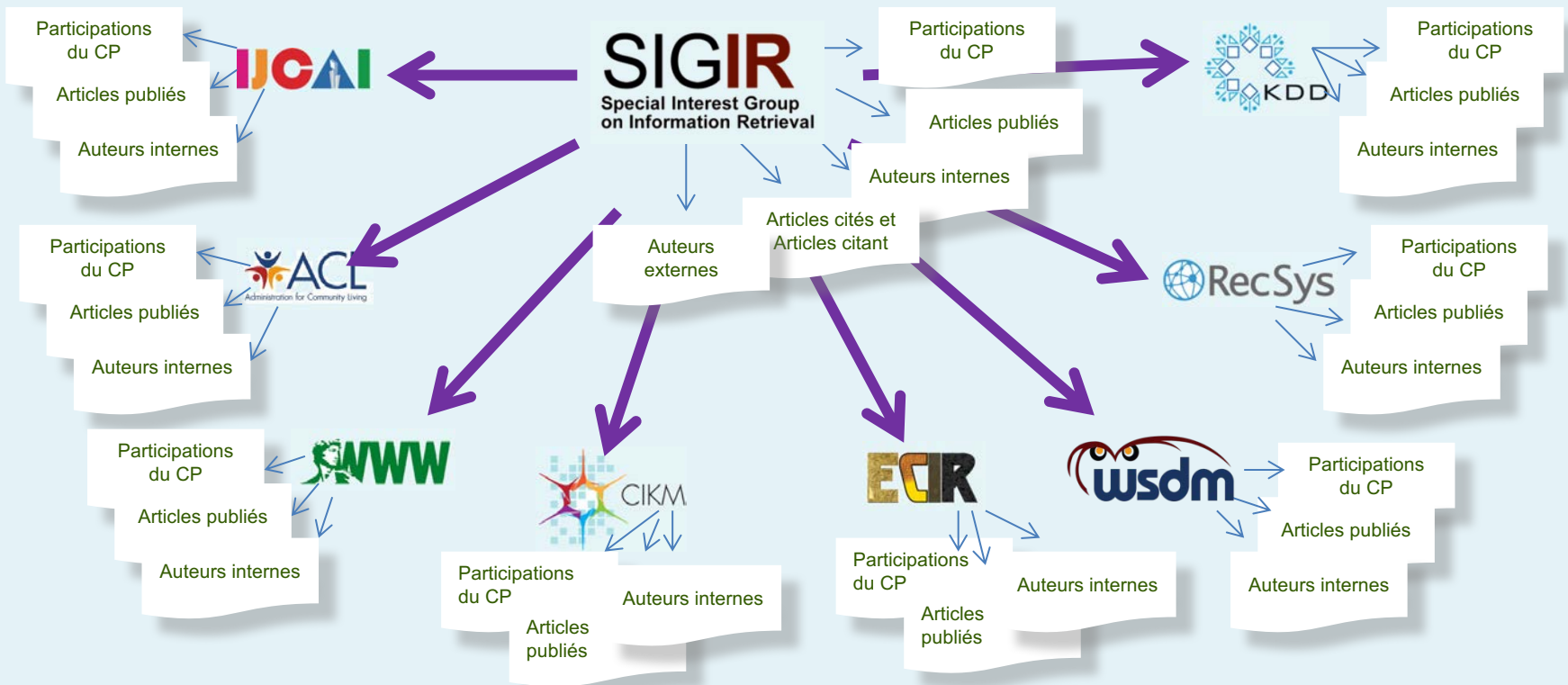


Quantification de l'indicateur

2. Expérimentations

Données

- Conférence : SIGIR et les conférences proches
- Bases bibliographiques : ACM et DBLP



2. Expérimentations

Données

- Conférence : SIGIR et les conférences proches
- Bases bibliographiques : ACM et DBLP

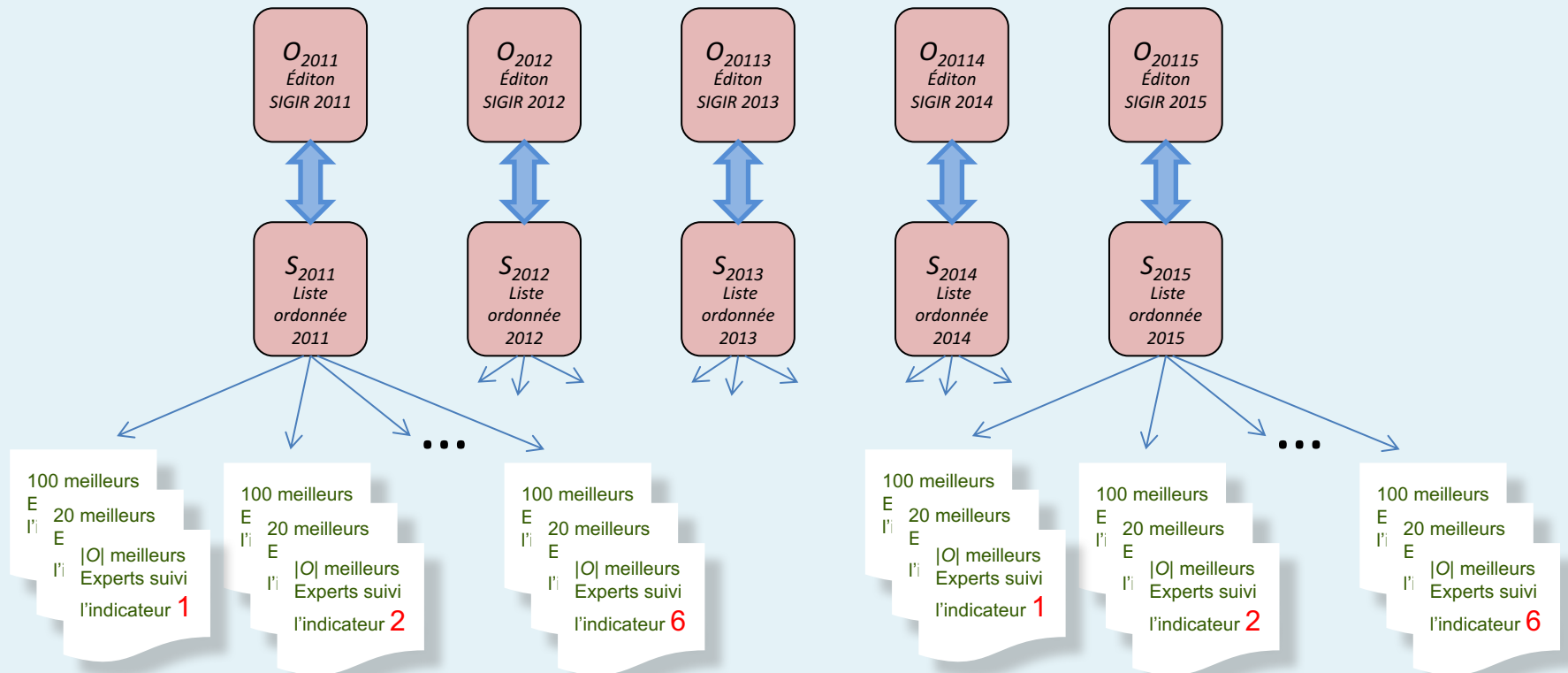
SIGIR 2015 Conference Organization	
Conference Chair:	Ricardo Baeza-Yates, Yahoo Labs, USA
Program Committee Chairs:	Mounia Lalmas, Yahoo Labs, UK Alistair Moffat, University of Melbourne, Australia Berthier Ribeiro-Neto, Google, Brazil, and UFMG, Brazil
Short Papers Chairs:	Maarten de Rijke, University of Amsterdam, The Netherlands Fe-Peng Lim, Singapore Management University, Singapore Ryan White, Microsoft Research, USA
Demonstrations Chairs:	Djoerd Hiemstra, University of Twente, The Netherlands Mirella Moro, Universidade Federal de Minas Gerais, Brazil
Tutorial Chair:	Yoelle Maarek, Yahoo Labs, Israel
Workshop Chairs:	Fernando Diaz, Microsoft Research, USA Diane Kelly, University of North Carolina at Chapel Hill, USA
Doctoral Consortium Chairs:	Shane Culpepper, RMIT University, Australia Brian Davison, Lehigh University, USA
Industry Day Chairs:	Hang Li, Huawei, China Jaime Teevan, Microsoft Research, USA
Best Paper Chair:	Susan Dumais, Microsoft Research, USA
Sponsorship Chairs:	Vanesia Mardock, Microsoft Research, USA Sergei Vassilvitskii, Google, USA
Local Arrangements Chair:	Diego Arroyuelo, UTFSM & Yahoo Labs, Chile
Finance Chair:	Mauricio Marín, Universidad de Santiago de Chile & Yahoo Labs, Chile
Publications Chairs:	Roberto González-Ibáñez, Universidad de Santiago de Chile, Chile Nicolás Hidalgo, Universidad de Santiago de Chile, Chile
Website Chair:	Marcelo Mendoza, UTFSM, Chile
Social Events Chairs:	Erika Rosas, Universidad de Santiago de Chile, Chile Carolina Bonacic, Universidad de Santiago de Chile & Yahoo Labs, Chile

- 9 conférences
- 95 éditions
- 14854 personnes

2. Expérimentations

Protocole

- Construire la liste ordonnée des experts candidats pour chaque indicateur : S
- Confronter à la vérité terrain : le comité de programme officiel O

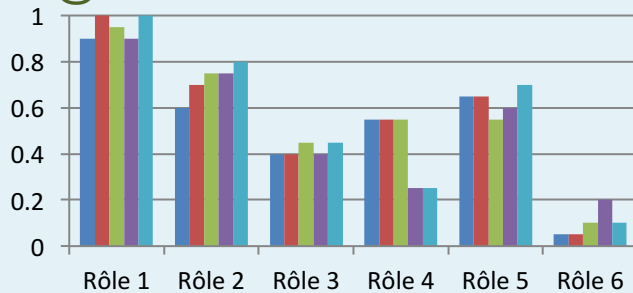


2. Expérimentations

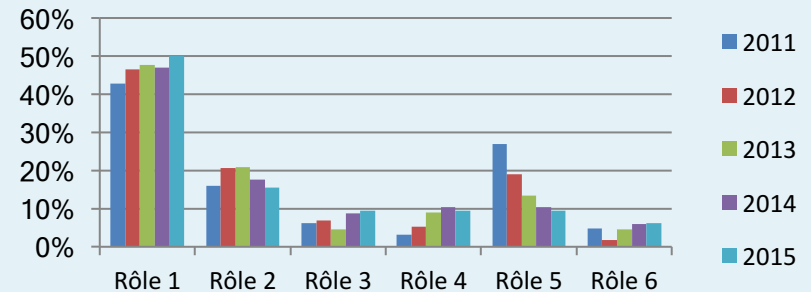
Résultats

- P@20
- P@100
- Proportion d'experts issus d'un seul indicateur et présents dans CP

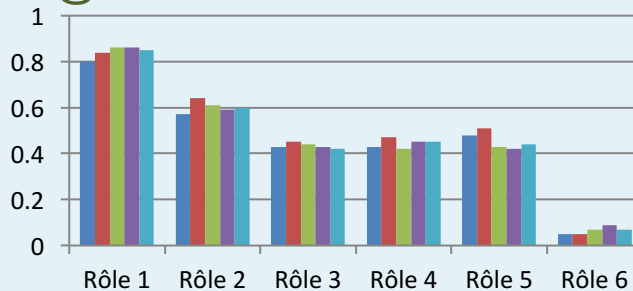
P@20



Proportion d'experts suggérés qui participent au CP en réalité



P@100



Rôle 1 : Membre interne de CP

Rôle 2 : Auteur interne

Rôle 3 : Auteur externe citant

Rôle 4 : Auteur externe cité

Rôle 5 : Membre externe

Rôle 6 : Auteur externe proche

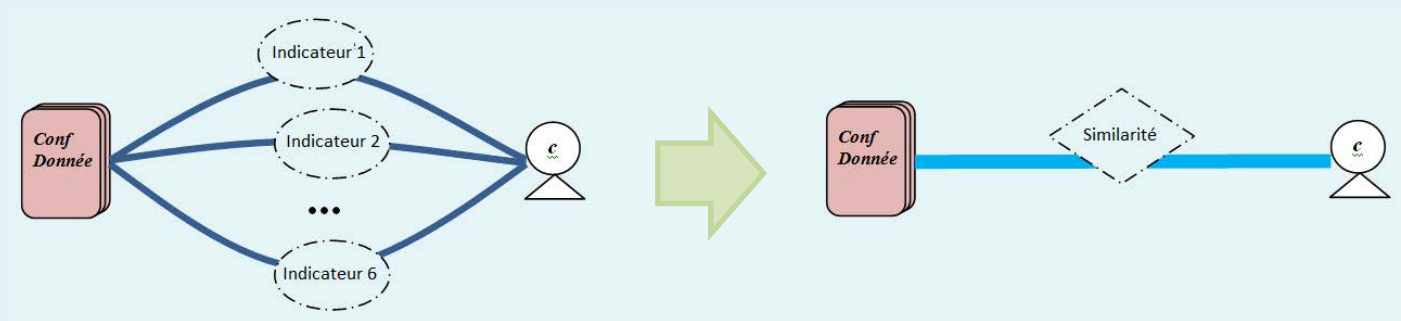
PLAN

- 1 Introduction
- 2 Positionnement
- 3 Modèle d'expertise
- 4 Indicateurs d'expertise de CP
- 5 Suggestion de CP
- 6 Conclusion et Perspectives

1. Similarité entre la conférence et l'expert candidat

Définition de la similarité

- Combinaison des indicateurs
- Paramétrage afin de moduler l'influence de l'indicateur associé

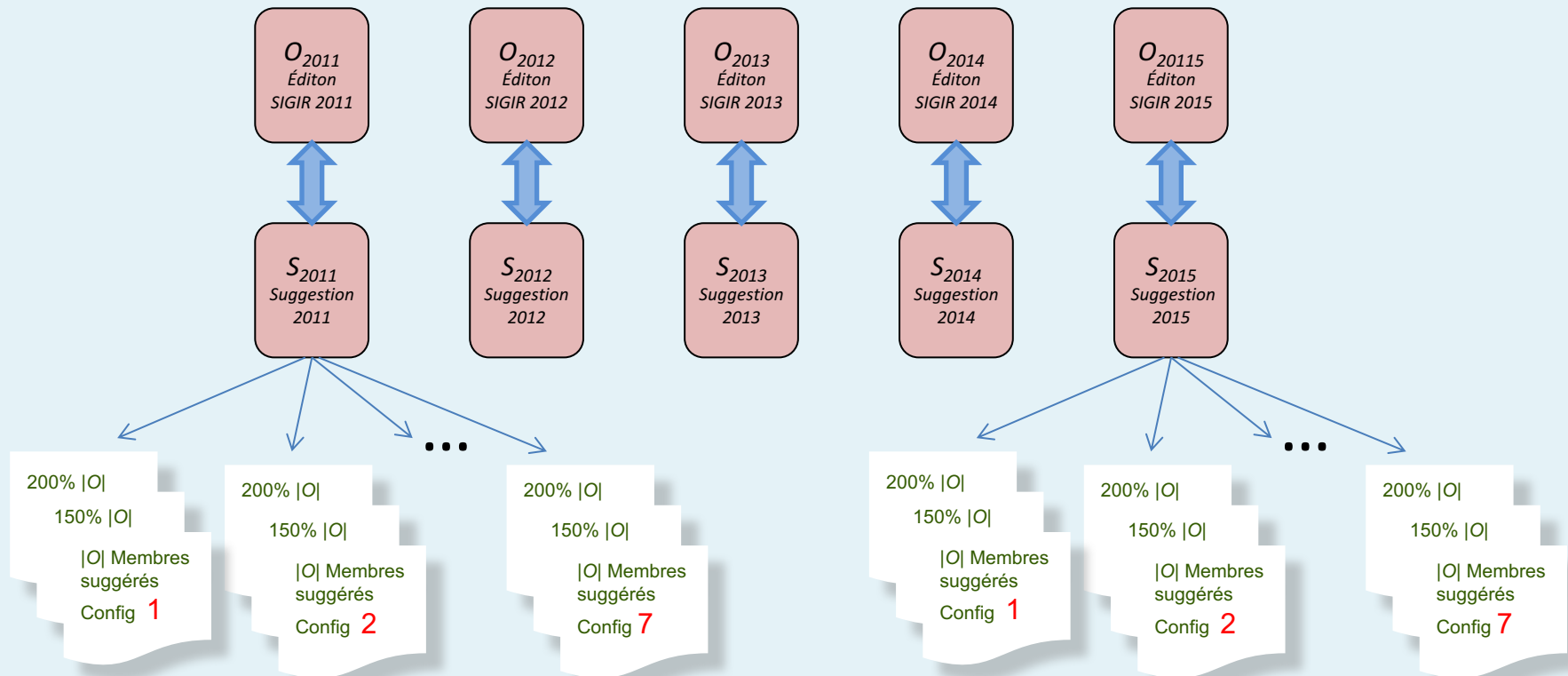


$$Similarité_{C,c,x} = \sum_{i=1}^6 \eta_i \cdot Indicateur_R\acute{o}le_i^{\star}$$

2. Évaluation des suggestions

Protocole commun

- Construire l'ensemble des membre suggérés
- Moduler les 6 paramètres par 7 configurations



2. Évaluation des suggestions

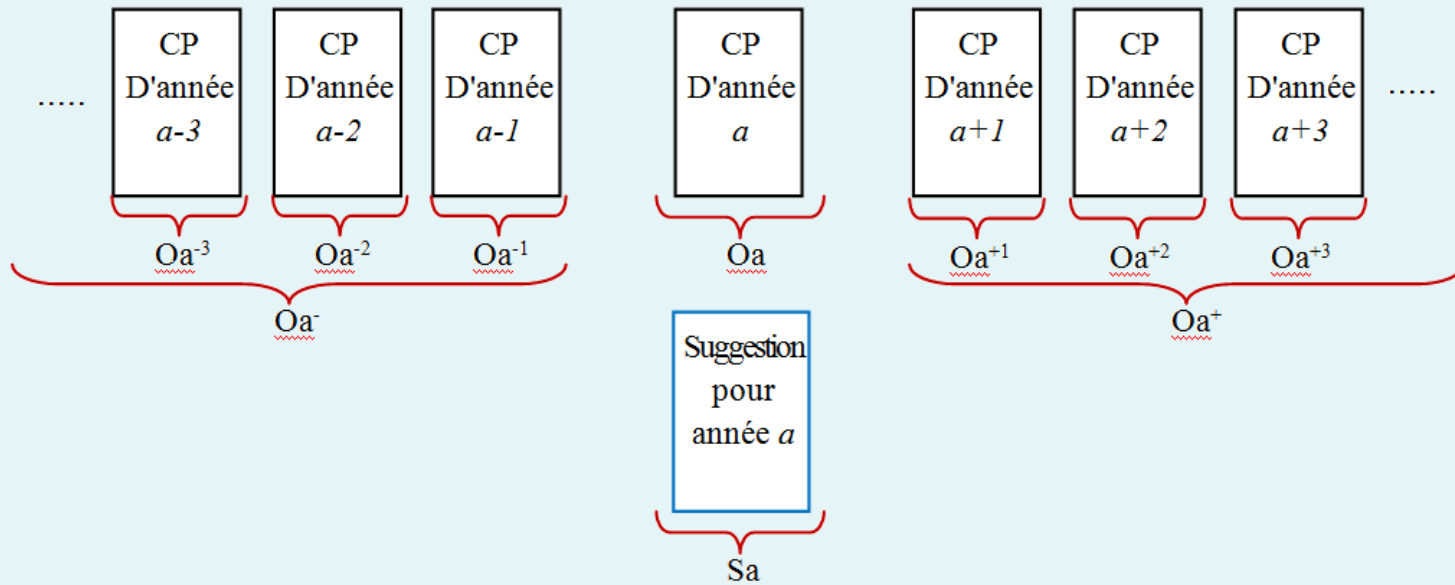


Q1. Dans quelle mesure notre approche réussit-elle à suggérer les membres des CP officiels ?

2. Évaluation des suggestions



Q1. Dans quelle mesure notre approche réussit-elle à suggérer les membres des CP officiels ?



$$C1 = S_a \cap O_a$$

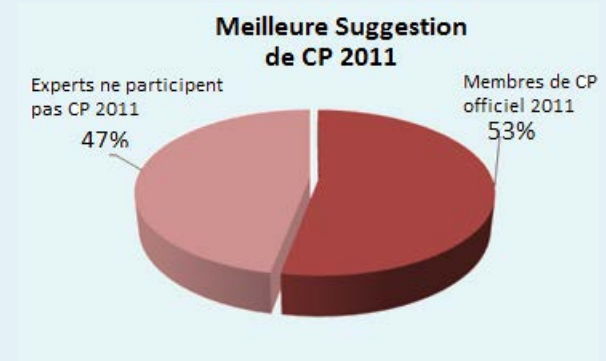
$$C2 = S_a \cap O_a^{-}$$

2. Évaluation des suggestions

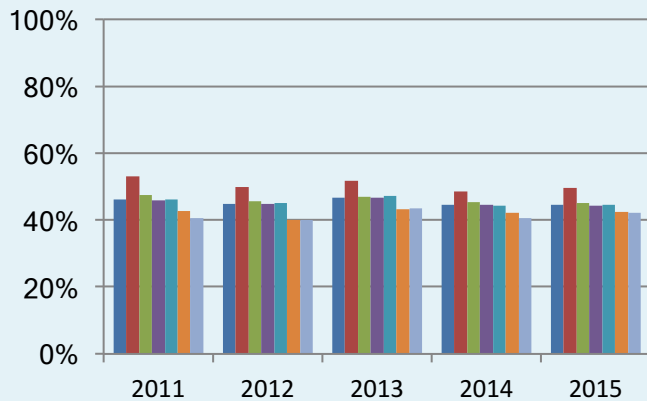


Q1. Dans quelle mesure notre approche réussit-elle à suggérer les membres des CP officiels ?

$$C1 = S_a \cap O_a$$



Précision



- Config 1 Même valeur pour 6 paramètres des indicateurs
- Config 2 L'indicateur sur le rôle de membres interne de PC plus important
- Config 3 L'indicateur sur le rôle d'auteur interne plus important
- Config 4 L'indicateur sur le rôle d'auteur externe citant plus important
- Config 5 L'indicateur sur le rôle d'auteur externe cité plus important
- Config 6 L'indicateur sur le rôle de membre externe de CP plus important
- Config 7 L'indicateur sur le rôle d'auteur externe proche plus important

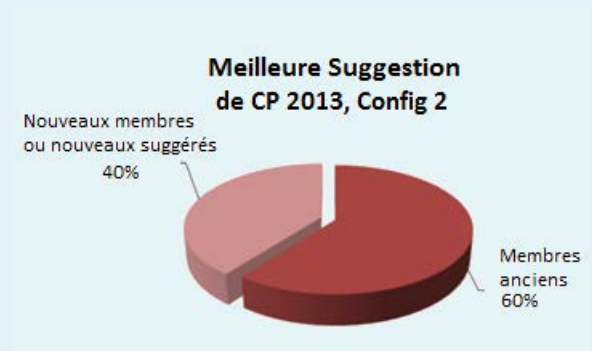
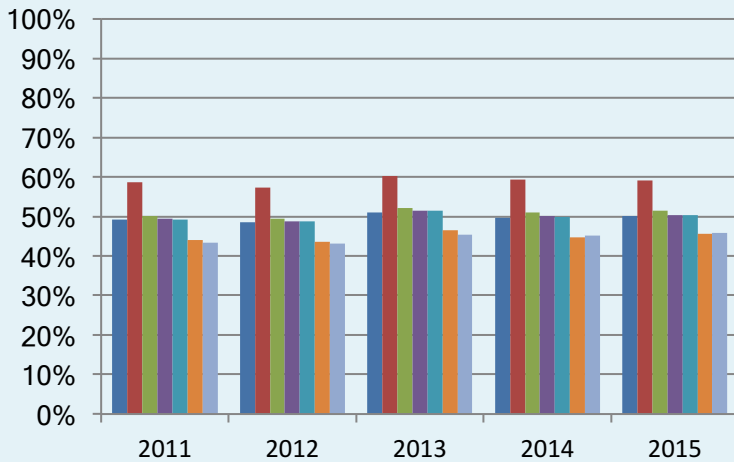
2. Évaluation des suggestions



Q1. Dans quelle mesure notre approche réussit-elle à suggérer les membres des CP officiels ?

$$C2 = S_a \cap O_a^-$$

Précision



- Config 1 : Même valeur pour 6 paramètres des indicateurs
- Config 2 : L'indicateur sur le rôle de membres interne de PC plus important
- Config 3 : L'indicateur sur le rôle d'auteur interne plus important
- Config 4 : L'indicateur sur le rôle d'auteur externe citant plus important
- Config 5 : L'indicateur sur le rôle d'auteur externe cité plus important
- Config 6 : L'indicateur sur le rôle de membre externe de CP plus important
- Config 7 : L'indicateur sur le rôle d'auteur externe proche plus important

2. Évaluation des suggestions

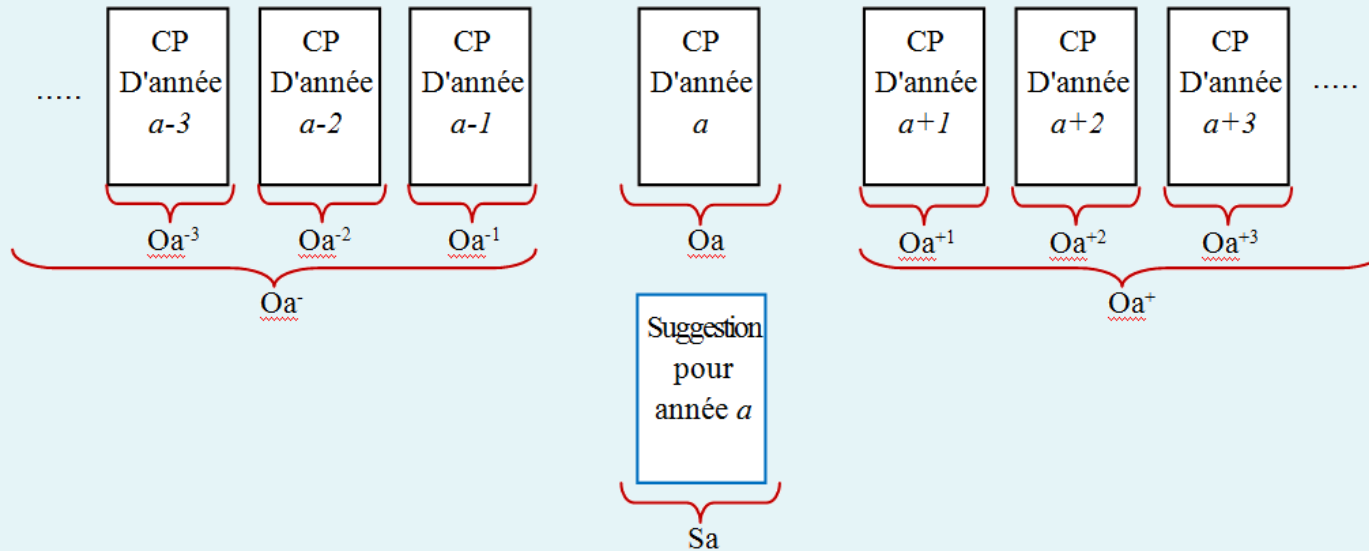
?

Q2. Dans quelle mesure notre approche réussit-elle à suggérer de nouveaux membres pour les prochains CP officiels ?

2. Évaluation des suggestions

?

Q2. Dans quelle mesure notre approche réussit-elle à suggérer de nouveaux membres pour les prochains CP officiels ?



$$C3 = N_{S_a} \cap (O_a \setminus O_a^-)$$

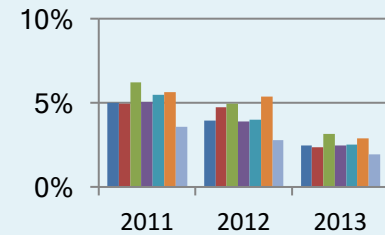
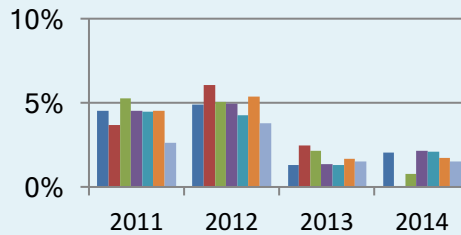
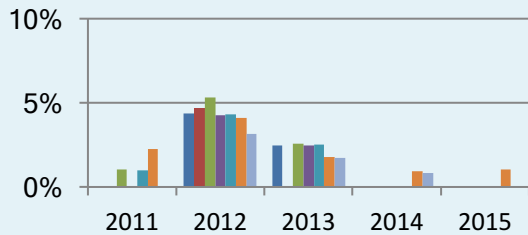
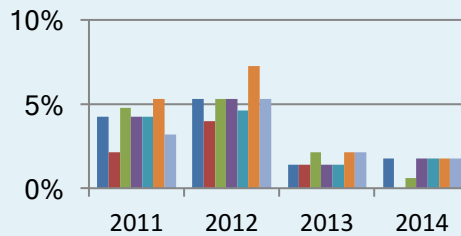
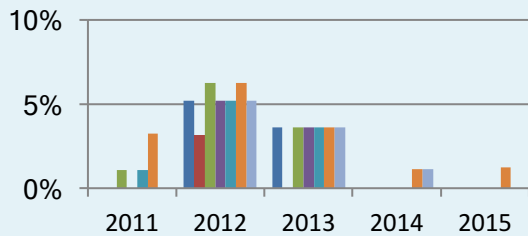
$$C4 = N_{S_a} \cap ((O_a^{+1} \cup O_a) \setminus O_a^-)$$

$$C5 = N_{S_a} \cap ((O_a^{+1} \cup O_a^{+2} \cup O_a) \setminus O_a^-)$$

2. Évaluation des suggestions



Q2. Dans quelle mesure notre approche réussit-elle à suggérer de nouveaux membres pour les prochains CP officiels ?



$$C3 = N_{S_a} \cap (O_a \setminus O_a^-)$$

$$C5 = N_{S_a} \cap ((O_a^{+1} \cup O_a^{+2} \cup O_a) \setminus O_a^-)$$

$$C4 = N_{S_a} \cap ((O_a^{+1} \cup O_a) \setminus O_a^-)$$

PLAN

- 1 Introduction
- 2 Positionnement
- 3 Modèle d'expertise
- 4 Indicateurs d'expertise de CP
- 5 Suggestion de CP
- 6 Conclusion et Perspectives

Conclusion

Étude des CP de conférences scientifiques

- Pilotage de la recherche scientifique
- Scientométrie appliquée au CP pour le constituer et l'évaluer

Modèle d'expertise

- Données disponibles
- Graphe hétérogène pondéré de différentes découvertes scientifiques

Scientométrie

- Analyse de l'expertise scientifique multi-facette par les rôles de l'expert
- Définition des indicateurs scientométriques sous-tendant la participation au CP

Suggestion des membres de CP

- Renouvellement des comités en suggérant des chercheurs pertinents
- Jeu de données coupé en deux ensembles par une année frontière : observation et test

Perspectives

Modélisation multi-variable de la sphère académique

- Intégrer le genre, l'ancienneté et les affiliations des scientifiques
- Tenter d'expliquer la participation au CP par le modèle multi-variables

Étudier complètement la constitution des CP

- Examiner les vrais critères employés par le président de CP pour inviter les chercheurs
- Pourquoi certains chercheurs bien considérés par notre modèle ne participent pas un CP ?

Étudier la participation de certains cas exceptionnels

- Une personne qui participe au CP sans un bon capital scientifique
- Responsable scientifique dans des divisions de R&D du secteur industriel

Intégrer d'autres preuves d'expertise

- Co-signature ou collaboration au sens large : projets de recherche, journées d'études, etc.

MERCI POUR VOTRE ATTENTION