

# Confrontation à la perception humaine de mesures de similarité entre membres d'un réseau social académique

## Enrichissement de la thématique par l'aspect social

Anais Lefeuvre — Guillaume Cabanac

Université de Toulouse, IRIT UMR 5505 CNRS  
118 route de Narbonne, F-31062 Toulouse cedex 9  
anais.lef@gmail.com, cabanac@irit.fr

---

*RÉSUMÉ.* Des mesures de similarité entre chercheurs ont été définies selon les dimensions thématique (analyse du contenu des articles) et bibliométrique (analyse du graphe des citations entre articles). Les rencontres et échanges entre chercheurs, relevant d'un aspect social, n'ont jusqu'alors pas été considérés. Aussi, notre contribution est triple. 1) Nous introduisons la dimension sociale avec notamment le concept de « co-conférenciers ». 2) Nous définissons un protocole expérimental pour confronter la perception humaine des chercheurs aux similarités calculées. 3) Nous rapportons les conclusions d'une expérimentation réalisée avec la participation de 71 chercheurs. Elle montre l'apport significatif de la dimension sociale pour simuler la perception des chercheurs (+8,49%). La prise en compte de notre proposition pourra notamment améliorer la qualité des résultats des systèmes de recommandation d'articles de recherche.

*ABSTRACT.* Similarity measures among researchers were defined according to topical and bibliometric dimensions (i.e., by analyzing publication contents and the associated citation graph). Meetings, encounters, and informal exchanges between researchers—which are related to a social aspect—were not considered to date. In this respect, our contribution is threefold. 1) We introduce the social dimension with notably the « co-lecturer » concept. 2) We design a protocol to check researchers' perception against computed similarities. 3) We report the findings of an experiment with 71 participants. It shows the significant improvement of the simulation of researchers' perception due to the social dimension (+8.49%). As a result, the integration of our proposal into paper recommender systems will help improving their effectiveness.

*MOTS-CLÉS :* Réseau social académique, recommandation, perception humaine, expérimentation  
*KEYWORDS:* Academia, social network, recommendation, human perception, measurement

---

## 1. Introduction

Quelle que soit sa discipline, tout chercheur réalise en continu une veille bibliographique lui permettant de constituer l'état de l'art de son domaine et de le maintenir à jour. Cette activité complexe est habituellement manuelle. Lorsqu'elle est assistée par ordinateur, des systèmes de recommandation identifient pour l'utilisateur les ressources (publications ou chercheurs) qui pourraient l'intéresser. Les modèles sous-jacents reposent sur des fonctions de similarité, notamment entre chercheurs. Ces dernières sont principalement basées sur un appariement entre les termes des publications. Considérer cette seule dimension paraît être réducteur comparé à l'ensemble des paramètres cognitivement intégrés par les chercheurs, notamment en fonction d'aspects sociaux, temporels, etc. Il nous semble donc nécessaire 1) d'évaluer la qualité des similarités thématiques, 2) d'en proposer des extensions, 3) et de vérifier que ces dernières améliorent la pertinence des mesures pour tendre vers une simulation plus proche de la réalité cognitive du chercheur dans son activité de veille bibliographique.

Cet article est structuré comme suit. Nous présentons le contexte de la veille bibliographique assistée par ordinateur et en détaillons les problématiques dans la section 2. Nous introduisons dans la section 3 la dimension sociale, considérée en soutien à la dimension thématique pour mesurer la similarité entre chercheurs. Dans la section 4, nous définissons un protocole d'expérimentation pour évaluer la mesure de similarité thématique, puis le gain obtenu par l'intégration de la dimension sociale. Nous rapportons les analyses réalisées suite à la participation de 71 chercheurs bénévoles. Enfin, nous concluons cet article en section 5 avant d'en ébaucher les perspectives.

## 2. Contexte de la veille bibliographique assistée par ordinateur

Dans le contexte de la recherche scientifique, nous nous intéressons aux activités des chercheurs, en particulier à la tâche d'état de l'art qui consiste à identifier, lire, comprendre et synthétiser des publications (articles de recherche). Ce travail est critique car il permet de positionner ses recherches tout en évitant de réinventer la roue. Pour aider les chercheurs dans ce travail de veille scientifique, il existe des systèmes de recommandation d'articles de recherche. Ces systèmes proposent aux chercheurs des ressources (articles ou autres chercheurs) susceptibles de les intéresser.

### 2.1. *Recommandation d'articles de recherche*

La littérature présente trois types d'approches concernant la recommandation (Montaner *et al.*, 2003). Nous les déclinons dans le cadre de la veille scientifique :

1) l'approche *collaborative* est basée sur une appréciation des ressources (notion de vote) par les chercheurs pour émettre les recommandations. Les chercheurs ayant émis des appréciations similaires sont alors rapprochés, permettant ensuite de proposer les articles jugés pertinents. Dans le cas particulier de la bibliométrie, le graphe

des citations est utilisé pour extraire ces votes. Notons que le contenu textuel des publications n'est pas considéré ;

2) l'approche orientée *contenu* est basée sur l'analyse textuelle des publications. Pour réaliser cela, ces approches mettent en œuvre un processus classique en recherche d'information : l'indexation (Manning *et al.*, 2008, ch. 2). Pour un document donné, ce processus comprend deux phases : extraction des termes significatifs, puis pondération de ces derniers en fonction de leur spécificité. La similarité entre deux documents quelconques est réalisée au travers d'une fonction d'appariement qui dépend du modèle choisi, par exemple cosinus pour le modèle vectoriel (Salton *et al.*, 1975). On peut généraliser ce fonctionnement au calcul de similarités entre chercheurs en modélisant un chercheur comme un méga-document qui concatène l'ensemble de ses publications (Klas *et al.*, 2000) ;

3) l'approche *hybride* combine les deux approches précédentes en considérant aussi bien les jugements des individus que le contenu des publications. C'est le cas par exemple du système COBRAS (Karoui *et al.*, 2006) qui exploite les jugements de plusieurs chercheurs et le contenu des notices bibliographiques (titres, mots-clés, etc.).

Pour plus d'informations, nous référons le lecteur à (McNee *et al.*, 2006) qui détaille ces différentes approches. Plus récemment, la problématique de la recommandation d'articles de recherche continue à faire l'objet de travaux, comme l'attestent notamment (Ben Jabeur *et al.*, 2010; Yang *et al.*, 2010). L'analyse de ces travaux nous a permis d'identifier des limites qui sont présentées dans la section suivante.

## **2.2. Problèmes liés aux dimensions prises en compte**

L'approche collaborative et l'approche orientée contenu, présentées en section 2.1, souffrent de limites que nous identifions ci-dessous :

1) une des limites de l'approche collaborative concerne la dimension bibliométrique. L'analyse purement comptable des citations pour représenter la qualité d'un article est biaisée, puisque ces dernières sont employées par un auteur pour diverses raisons. En effet, par l'emploi d'une citation, l'auteur véhicule une polarité perceptuelle : il peut l'employer négativement (resp. positivement) pour critiquer (resp. louer) l'article associé. Cependant, elles compteront de la même façon. Par ailleurs, la signification des citations peut varier en fonction de leur position dans un document. Une citation placée en introduction ou dans un protocole d'expérimentation n'a pas la même portée et est plus ou moins généraliste. Enfin, il existe des abus qui consistent à intégrer des citations de complaisance, voire des auto-citations non pertinentes. De tels phénomènes, bien que malhonnêtes car visant à augmenter des indicateurs tels que le *h*-index, ne sont pas détectés. À partir de cette base biaisée exploitant le graphe de citations, des approches telles que (Cazella *et al.*, 2005; Ben Jabeur *et al.*, 2010) déterminent la réputation des chercheurs, tentant ainsi d'évaluer leur « prestige ». La justification d'une recommandation peut alors davantage résulter d'un prestige élevé que d'une pertinence thématique avérée. Or, pour la tâche d'état de l'art, le critère le plus important pour un chercheur nous paraît être l'adéquation thématique. Le résultat

d'une recommandation déviée par le prestige ne serait donc pas pertinente. De plus, les chercheurs réputés sont forcément visibles, rendant leurs travaux visibles également. Par contre, on pourrait davantage bénéficier de l'identification d'auteurs peu réputés ou débutants, qui proposent des idées innovantes dans notre thématique spécifique. Ils seraient pourtant masqués par la présence des plus réputés. En conclusion, la démarche actuelle introduisant le prestige relève du cercle vicieux : elle ne favorise ni l'émergence de nouvelles personnes, ni de nouvelles idées car les plus visibles deviennent de plus en plus visibles ;

2) l'approche orientée contenu souffre également de problèmes. Premièrement, la sémantique des termes n'est pas considérée. Par exemple, « taxonomie » et « taxinomie » sont considérés différents. Deuxièmement, les termes sont considérés indépendamment les uns des autres. Par exemple, l'expression « recherche d'information » est indexée par deux termes « recherche » et « information », qui, pris séparément, n'expriment pas le même concept. Troisièmement, les documents rédigés en différentes langues ne seront pas comparables (« retrieval » et « interrogation »).

En plus d'être limitées, ces approches ne renseignent que sur les *réalisations* des chercheurs, sans considérer leur relations sociales. Pour contrebalancer ces limites, nous proposons d'introduire la notion d'*environnement social* dans la section suivante. En effet, la réduction d'un chercheur à sa seule production n'est qu'une vision partielle de ce dernier, car il interagit avec d'autres individus, notamment en phase de rédaction et à l'occasion de la participation à des conférences.

### **3. Définition et intégration de l'aspect social pour la recommandation**

Comme indiqué dans la section précédente, de nombreuses approches actuelles requièrent l'analyse du contenu des publications pour en extraire les termes ainsi que les liens de citation. Elles développent des modèles riches et intéressants sur le plan théorique. Toutefois, leur mise en pratique se heurte à des barrières légales rendant leur mise en œuvre difficile. En effet, les auteurs d'un article publié chez un éditeur (ACM, Springer, etc.) cèdent leurs droits d'auteur. L'éditeur diffuse alors les articles moyennant un abonnement payant. Par conséquent, une approche de recommandation basée sur le contenu intégral des articles nécessite l'abonnement à l'ensemble des éditeurs du domaine étudié, représentant à la fois un coût important et un traitement spécifique à chacun pour récupérer les publications. Pour s'affranchir de ces restrictions, la proposition que nous développons dans cet article se base sur des données publiques et libres de droits. En effet, nous exploitons uniquement les méta-données des publications (titres, auteurs, nom de conférence ou de journal, etc.) sans nécessiter l'acquisition de droits d'exploitation pour le contenu intégral des articles.

L'originalité de notre approche consiste à intégrer et valoriser, dans la recommandation, l'aspect social du travail de recherche. On peut observer cet aspect social, pas tant au travers d'articles pris séparément, mais plutôt en considérant les interactions entre individus. Il nous semble donc plus judicieux de recommander des chercheurs (en donnant ensuite accès à leur production scientifique) que des articles de chercheurs

distincts. En outre, pour un chercheur  $c$  donné, notre approche répond à un besoin récurrent : identifier les chercheurs partageant ses thématiques sur le long terme (au lieu d'un intérêt ponctuel contenu dans un seul article, par exemple) et accessibles, c'est-à-dire proches dans son environnement social. Cette recommandation permettrait au chercheur  $c$  d'identifier des collaborateurs potentiels pour la réalisation de projets, par exemple. Ces derniers seraient pertinents au regard de la thématique, comme au regard de leur distance sociale. Pour compenser les limites de la dimension thématique exposées dans la section 2.2, nous proposons d'exploiter cette dimension et de la compléter par la dimension sociale, décomposée en trois facettes représentées par les fonctions suivantes, où  $C = \{c_1, \dots, c_n\}$  est l'ensemble des chercheurs :

1) l'*inverse du degré de séparation* entre deux chercheurs. Milgram (1967) désigne par degré de séparation le nombre minimal d'intermédiaires nécessaires pour relier deux personnes données. Expérimentalement, Milgram (1967) a montré que le degré de séparation médian entre deux personnes aux USA était de cinq, d'où le concept de « petit monde » introduit pour désigner ce phénomène. Dans notre contexte de travail, nous calculons cette mesure à partir du graphe des coauteurs, où les nœuds représentent les chercheurs et les arcs représentent les relations de co-écriture. Le degré de séparation  $d : C^2 \rightarrow \mathbb{N}_+$  correspond alors à l'inverse de la longueur du chemin le plus court entre deux chercheurs donnés. Sur la base bibliographique DBLP<sup>1</sup> recensant des publications en informatique, l'étude d'Elmacioglu *et al.* (2005) a montré qu'il y a six intermédiaires en moyenne entre deux chercheurs. Cette fonction capture un aspect social en valorisant la proximité entre personnes ;

2) la *force de la connectivité* entre deux chercheurs. Cette notion représentée par la fonction  $f : C^2 \rightarrow \mathbb{N}_+$  est également basée sur le graphe des coauteurs. Elle restitue le nombre de chemins distincts reliant deux chercheurs par l'intermédiaire d'un minimum de personnes (ce nombre correspondant à leur degré de séparation). Cette fonction capture un aspect social en valorisant les situations où une personne a l'opportunité d'en joindre une autre en passant par des intermédiaires différents ;

3) les *conférences communes* entre deux chercheurs, où l'on considère à la fois le nom de la conférence et son édition : « Marami 2010 », par exemple. Cette notion représentée par la fonction  $c : C^2 \rightarrow \mathbb{N}_+$  est basée sur le graphe des co-conférenciers dans lequel deux chercheurs sont liés s'ils ont publié à une même édition de conférence. Cette fonction capture un aspect social dans la mesure où les chances que deux chercheurs se soient rencontrés sont d'autant plus élevées qu'ils ont publié dans les mêmes éditions de conférences et y ont éventuellement participé.

Par ailleurs, nous définissons la fonction thématique  $t : C^2 \rightarrow \mathbb{R}_+$  basée sur une approche classique en RI : le modèle vectoriel et la mesure du cosinus (cf. section 2.1). Par la suite, nous calculons pour un chercheur  $c_i$  donné la liste  $T$  des  $n - 1$  chercheurs les plus proches de lui, soit  $T = \langle (c_i^1, s_i^1), \dots, (c_i^{n-1}, s_i^{n-1}) \rangle$ , où  $s_i^j = t(c_i, c_j)$  représente la similarité thématique entre  $c_i$  et  $c_j$ . De la liste  $T$ , représentée en figure 1(c), on exclut le chercheur  $c_i$  lui-même, ainsi que ses coauteurs (car il ne serait pas inté-

1. <http://www.informatik.uni-trier.de/~ley/db>

ressé par une recommandation le concernant ou concernant ses coauteurs, dont on fait l'hypothèse qu'il connaît les travaux). De façon similaire, on crée les listes D, F et C à l'aide des fonctions  $d$ ,  $f$  et  $c$  précédemment définies.

Nous avons formulé l'hypothèse que l'intégration d'aspects sociaux à la dimension thématique améliore sa pertinence. Pour réaliser cette intégration, nous utilisons le combinateur CombMNZ (Fox *et al.*, 1993) normalisé, issu de travaux de RI. Il combine plusieurs ( $m$ ) listes de résultats issues de moteurs de recherche (couples document-similarité) en une seule liste, comprenant tous les documents distincts associés à une agrégation des similarités initiales. Pour CombMNZ, la similarité calculée  $S$  agrège les  $m$  valeurs de similarité  $s_1, \dots, s_m$  selon  $S = M \cdot \sum_{i=1}^m s_i$ , où  $M$  correspond au nombre de similarités  $s_i$  non nulles. Notons que la supériorité du combinateur CombMNZ par rapport à d'autres combinateurs a été montrée par Lee (1997) sur des collections de test de TREC. Appliqué à notre contexte, il combine les listes obtenues (couples chercheur-similarité) en une seule liste. Concrètement, nous calculons la liste thématico-sociale TS d'un chercheur en deux étapes. La première consiste en la combinaison des listes D, F et C pour obtenir la liste sociale S. Dans cette liste figurent des chercheurs qui ne sont liés que par l'aspect social et pas par la thématique. Aussi, nous les retirons de cette liste, l'objectif de notre approche étant d'améliorer la veille bibliographique. La seconde étape combine S avec T pour obtenir TS, représentée en figure 1(a). Notons que ces deux combinaisons ne font pas d'hypothèse sur l'importance relative des critères en entrée de la combinaison (par exemple, la dimension sociale a le même poids que la dimension thématique). C'est une première approche sans *a priori* que nous évaluons dans la section suivante.

#### **4. Confrontation des similarités entre chercheurs à la perception humaine**

Dans cette section, nous confrontons les similarités thématique et thématico-sociale entre chercheurs à la perception humaine. Notre hypothèse est la suivante : la combinaison des similarités thématique et sociale simule mieux la perception humaine que la similarité thématique seule. Nous présenterons tout d'abord le protocole d'expérimentation, puis sa mise en œuvre et les résultats obtenus.

##### **4.1. Protocole d'expérimentation**

Préalablement à l'évaluation, on calcule pour un chercheur donné les listes TS et T illustrées dans la figure 1. La tâche d'un participant consiste alors en l'évaluation de la liste T, après qu'elle ait été ordonnée aléatoirement pour ne pas influencer son jugement. Pour recueillir la perception de sa proximité avec les  $n$  chercheurs, il note chacun d'entre-eux sur une échelle de zéro à six. La valeur de cette note répond à la question : « Pour progresser dans votre recherche, avec qui faudrait-il discuter ? ». Remarquons que la note zéro signifie que le participant n'est pas du tout intéressé par le chercheur proposé : la recommandation de ce dernier ne serait pas pertinente. Par ailleurs, la recommandation serait d'autant plus pertinente que la note est élevée. Ces notes constituent la liste PH (perception humaine) représentée dans la figure 1(b).

Noms	Sim		Noms	Sim		Noms	Sim
Alan Turing	0,54	← ? →	Donald E. Knuth	6	← ? →	Bill Gates	0,75
Donald E. Knuth	0,48		Alan Turing	4		Alan Turing	0,71
⋮	⋮		⋮	⋮		⋮	⋮
Bill Gates	0,10		Bill Gates	0		Donald E. Knuth	0,24
(a) Liste TS			(b) Liste PH			(c) Liste T	

**Figure 1.** Listes TS, PH et T considérées dans le protocole d’expérimentation.

Pour analyser ces données, nous établissons un parallèle avec l’évaluation en Recherche d’Information (RI). Dans ce domaine, les systèmes de RI (moteurs de recherche) restituent un résultat (ensemble de documents) en réponse à une requête d’un utilisateur, traduisant son besoin en information. Pour évaluer la qualité de ces systèmes, la méthodologie Cranfield (Buckley *et al.*, 2005) est employée, notamment dans le cadre de TREC<sup>2</sup>. Pour cela, les résultats du système  $s$  pour la requête  $r$  sont comparés aux jugements de pertinence  $j$  effectués par des humains. Cette comparaison est réalisée à l’aide d’une mesure de pertinence  $m(s, r, j) \in [0; 1]$ . Dans le cas où  $m(s_1, r, j) > m(s_2, r, j)$ , on peut alors conclure que  $s_1$  obtient de meilleurs résultats que  $s_2$ . Plusieurs mesures  $m$  ont été proposées, notamment *ap*, *Average Precision*, pour des jugements de pertinence  $j$  booléens (le document est pertinent ou non) et *ndcg* (Järvelin *et al.*, 2002), *Normalized Discounted Cumulative Gain*, pour des  $j$  graduels (le document est plus ou moins pertinent).

Au même titre qu’un système de RI  $s$  restitue des documents pour une requête  $r$  donnée, un système de veille bibliographique restitue des recommandations pour un chercheur donné. Par conséquent, on assimile la liste T aux résultats d’un système  $s_1$  et la liste TS aux résultats d’un système  $s_2$ . De plus, un chercheur  $c$  participant est assimilé à  $r$  ; les jugements de pertinence  $j$  correspondent à sa perception humaine PH. Enfin, pour évaluer  $m(T, c, PH)$  et  $m(TS, c, PH)$ , nous retenons la mesure  $m = ndcg$  car la perception humaine PH est graduelle (entre zéro et six). La mesure de la qualité des listes T et TS peut alors être calculée pour chaque chercheur  $c_i$ . Celle-ci peut être très variable d’un chercheur à l’autre. Le calcul de la moyenne des  $n$  mesures *ndcg* de pertinence individuelles permet de comparer les deux systèmes  $s_1$  et  $s_2$  globalement. Cet indicateur est d’autant plus robuste lorsque le nombre  $n$  de participants est élevé et lorsqu’ils sont variés. Dans le cadre de TREC, Buckley *et al.* (2000) ont montré que cet indicateur est robuste à partir de  $n = 25$  valeurs, bien que  $n = 50$  soit préférable. Le gain résultant de l’intégration de l’aspect social est donc mesuré par  $(\sum_{i=1}^n ndcg(TS, c_i, PH) / \sum_{i=1}^n ndcg(T, c_i, PH)) - 1$ . Pour tester la robustesse de ce gain, nous vérifions que la différence entre les deux séries testées est statistiquement significative. On calcule alors la valeur de significativité  $p$  avec le test  $t$  de Student pairé (la différence est calculée entre les paires de valeurs  $ndcg(T, c_i, PH)$  et  $ndcg(TS, c_i, PH)$ ) et bilatéral. Bien que nécessitant théoriquement une distribution normale des données,

2. Text REtrieval Conference, cf. <http://trec.nist.gov>

Hull (1993) précise que ce test est en pratique robuste aux violations de cette condition. Par ailleurs, Sanderson *et al.* (2005) montrent que ce test est bien plus fiable que d'autres, tel que le test de Wilcoxon. Concrètement, lorsque  $p < \alpha$  avec  $\alpha = 0,05$  la différence entre les deux échantillons testés est qualifiée de statistiquement significative (Hull, 1993). Plus la valeur  $p$  est petite, plus la différence est significative.

#### **4.2. Mise en œuvre du protocole d'expérimentation**

Pour évaluer la qualité des mesures de similarité entre chercheurs au travers du protocole décrit, il est nécessaire de se baser sur une source de données bibliographique. Nous avons retenu DBLP (Ley, 2002) qui est focalisée sur les sciences informatiques. Ce choix est motivé par trois raisons. Premièrement, au 11 mai 2010, DBLP indexait 1 392 143 publications en différentes langues (dont l'anglais et le français), réalisées par 811 787 chercheurs distincts. Deuxièmement, ces données sont mises à disposition sous la forme d'un fichier XML de 713 Mo associé à une DTD. Troisièmement, DBLP a déjà été utilisée dans le cadre de travaux de recherche antérieurs (Elmacioglu *et al.*, 2005; Biryukov, 2008).

Notre but étant de valider les mesures de similarité par rapport à la perception humaine, il s'avère nécessaire d'impliquer les chercheurs en tant que participants à l'expérimentation. Dans la lignée des approches de *crowdsourcing* (Alonso *et al.*, 2008), afin d'obtenir un panel de nombreux jugements variés, nous adoptons une démarche basée sur une expérimentation en ligne (Reips, 2007). En juillet 2010, nous avons contacté individuellement 90 personnes ayant au moins deux publications dans DBLP. Cet échantillon de chercheurs a été constitué à partir de nos listes de contact sur la plateforme sociale LinkedIn<sup>3</sup>. Le courrier électronique envoyé contenait une description du travail à réaliser dont la durée approximative est de cinq minutes, ainsi qu'un hyperlien vers la plateforme d'expérimentation<sup>4</sup>.

#### **4.3. Analyse des données recueillies suite à l'expérimentation**

##### **4.3.1. Caractéristiques de l'échantillon des 71 participants**

Suite à notre courrier électronique, 74 chercheurs ont commencé l'expérimentation, représentant un taux de réponse de 82 %. Au final, 71 personnes appelées « participants » dans la suite de l'article l'ont terminée. Le taux d'abandon est donc de 4 %. En comparaison avec le taux d'abandon de 45 % mesuré par Reips (2007) lors d'expérimentations en ligne non rémunérées, l'abandon que nous avons observé est faible. Ceci résulte certainement de deux spécificités de notre expérimentation : la durée nécessaire pour la réaliser était courte (en moyenne cinq minutes) et nous avons sollicité des accointances qui étaient davantage susceptibles de nous aider que des inconnus.

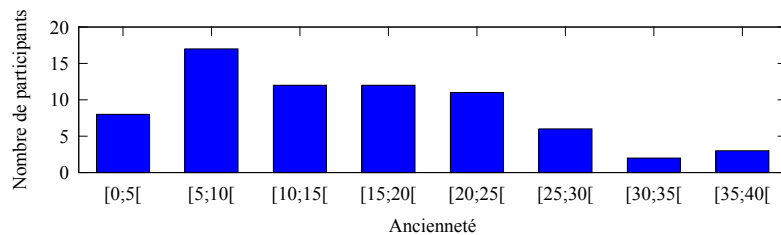
---

3. <http://www.linkedin.com>

4. cf. une démonstration sur <http://www.irit.fr/~Guillaume.Cabanac/expeSimT>



Au préalable à l'évaluation, les participants fournissaient leur âge et leur ancienneté (nombre d'années en recherche) afin de nous permettre de réaliser des analyses par catégories démographiques. Les 71 participants ont entre 24 ans et 61 ans, avec en moyenne 39 ans (écart-type 10 ans). Concernant leur ancienneté, illustrée dans la figure 2, les participants ont entre 1 et 39 années d'expérience en recherche, avec en moyenne 14 années (écart-type 9 années). L'analyse de leur production scientifique montre qu'ils ont en moyenne 21 publications répertoriées dans DBLP (figure 3). L'écart-type de 21 publications montre la variabilité de la production des participants. C'est une caractéristique intéressante de l'échantillon, car cela nous permettra de réaliser des analyses en fonction de leur nombre de publications.

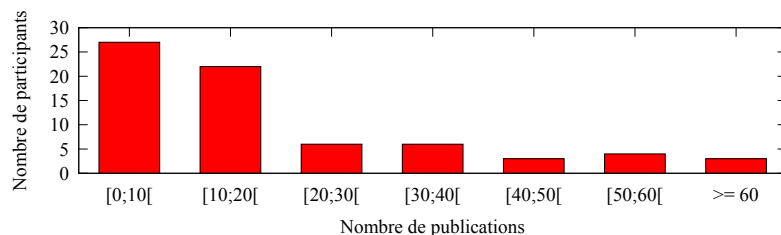


**Figure 2.** Répartition des 71 participants selon leur ancienneté.

#### 4.3.2. Qualité des mesures de similarité et apport de l'aspect social

Dans cette section, nous appelons « baseline » le système de référence produisant la liste T. Notre proposition correspond au système produisant la liste TS. Les résultats de notre analyse sont synthétisés dans le tableau 1. Un astérisque dans la colonne Gain indique une amélioration statistiquement significative ( $t$ -test avec  $p < 0,05$ ).

L'analyse des données au niveau global ①, sur les 71 participations, montre une performance de baseline élevée (0,7665). Notre proposition améliore cependant cette baseline de 8,49%. Cette amélioration est de surcroît statistiquement significative ( $p < 0,05$ ). Ces observations valident notre hypothèse : intégrer la dimension sociale en complément de la thématique dans les mesures de similarité améliore leur



**Figure 3.** Répartition des 71 participants selon de leur nombre de publications.

pertinence. En d'autres termes, la perception humaine des similarités entre chercheurs est mieux simulée.

Indicateurs	Cas	Critères	$ndcg(T, c, PH)$	$ndcg(TS, c, PH)$	Gain (%)
	①	global	0,7665	0,8316	8,49*
Nombre de publications	②	< 15 publications	0,6960	0,7683	10,39*
	③	≥ 15 publications	0,8313	0,8897	7,03*
Ancienneté en recherche	④	< 13 ans	0,7456	0,7941	6,50*
	⑤	≥ 13 ans	0,7857	0,8660	10,22*

**Tableau 1.** Différences de performances entre T et TS pour les 71 participants.

Afin d'approfondir ces analyses, nous avons constitué des catégories selon le nombre de publications (② et ③) et l'ancienneté en recherche (④ à ⑤) des participants. Chaque catégorie est divisée en deux groupes de même taille en fonction de la médiane (15 pour le nombre de publications et 13 pour l'ancienneté). Nous commentons les résultats obtenus en deux temps. Premièrement, l'amélioration est plus importante pour les chercheurs ayant peu publié ② que pour les chercheurs ayant beaucoup publié ③ (10,39 % vs 7,03 %). Toutefois, l'amélioration est robuste dans le deuxième cas uniquement (cf. *t*-test). Deuxièmement, l'amélioration est plus importante pour les chercheurs expérimentés ⑤ que pour les débutants ④ (10,22 % vs 6,50 %). Globalement, ces résultats s'expliqueraient par le fait que les moins expérimentés ou les moins publiants ont moins eu l'occasion de collaborer avec des personnes différentes, d'où le faible effet de l'intégration de l'aspect social. Par contre, les chercheurs qui sont plus expérimentés ou publient davantage (≥ 15 publications) ont forcément eu plus d'interactions sociales (co-conférenciers, etc.), d'où l'amélioration constatée.

Cette expérimentation a montré que l'apport de la dimension sociale est avéré pour toutes les catégories distinguées. Il est d'autant plus robuste que les chercheurs sont expérimentés et ont beaucoup publié.

## 5. Conclusion et perspectives

Dans cet article, nous avons considéré les mesures de similarité entre chercheurs d'un réseau social académique. La similarité la plus répandue dans la littérature, qualifiée de thématique, est basée sur l'analyse des termes issus des publications. Nous avons émis des critiques à son égard et proposé d'enrichir cette approche thématique en intégrant des aspects sociaux, tels que le degré de séparation, la force de connectivité et la notion de co-conférenciers. Afin de valider cette proposition, nous avons défini un protocole d'expérimentation et l'avons mis en œuvre avec la participation de 71 chercheurs bénévoles. L'analyse des données recueillies montre que notre approche améliore significativement l'approche thématique (+8,49 %). En d'autres termes, elle simule mieux la perception humaine de la proximité entre chercheurs. Par conséquent, sa prise en compte dans les systèmes de recommandation d'articles de recherche pourrait améliorer leur qualité.

De nombreuses perspectives sont envisagées à ces travaux. À court terme, il s'agit de confirmer nos résultats avec davantage de participants, notamment des chercheurs expérimentés qui sont les moins représentés dans notre étude. Le test d'autres combineurs et d'approches d'apprentissage (*machine learning*) est également à réaliser. À moyen terme, pour la dimension thématique, il faudrait indexer des concepts plutôt que des termes distincts (par exemple, l'expression « recherche d'information » plutôt que les mots qui la composent) et regrouper des expressions possédant la même sémantique (par exemple, « RI » avec « recherche d'information »). La considération de l'innovation au travers de l'identification des chercheurs précurseurs dans leur domaine (c'est-à-dire qui ont introduit des termes massivement repris par la suite, comme « ontologie ») est une autre piste. Pour le social, il faudrait identifier la sociabilité des chercheurs en matière de publication : publient-ils de façon solitaire, tribale (toujours avec le même groupe) ou éparse (avec de nombreux coauteurs distincts) ? Enfin, il faudrait prendre en compte la dimension temporelle : 1) pour l'aspect thématique car les chercheurs peuvent changer de domaine et se désintéresser de leurs thématiques passées ; 2) pour l'aspect social car la recommandation de chercheurs en cessation d'activité est discutable. De plus, les relations sociales les plus récentes sont éventuellement à privilégier. À long terme, il s'agit d'établir une typologie des besoins liés à la veille bibliographique. En effet, le processus de recommandation est à adapter, que l'on désire être notifié des événements dans son environnement scientifique proche (tâche d'*awareness*) ou, à l'opposé, que l'on souhaite découvrir des auteurs et des thématiques innovantes (tâche de prospective).

## Remerciements

Nous remercions vivement les 71 chercheurs bénévoles qui ont participé à l'expérimentation rapportée dans cet article. Leurs commentaires et suggestions pertinents ont été une source de réflexion stimulante.

## 6. Bibliographie

- Alonso O., Rose D. E., Stewart B., "Crowdsourcing for relevance evaluation", *SIGIR Forum*, vol. 42, n° 2, p. 9–15, 2008.
- Ben Jabeur L., Tamine L., Boughanem M., "A social model for Literature Access: Towards a weighted social network of authors", *RIAO'10: Proceedings of the 9th international conference on Information Retrieval and its Applications*, April, 2010.
- Biryukov M., "Co-author Network Analysis in DBLP: Classifying Personal Names", *MCO'08: Proceedings of the 2nd international conference on Modelling, Computation and Optimization in Information Systems and Management Sciences*, vol. 14 of *Communications in Computer and Information Science*, Springer, p. 399-408, 2008.
- Buckley C., Voorhees E. M., "Evaluating Evaluation Measure Stability", *SIGIR'00: Proceedings of the 23rd international ACM SIGIR conference*, ACM, New York, NY, USA, p. 33–40, 2000.
- Buckley C., Voorhees E. M., "Retrieval System Evaluation", in E. M. Voorhees, D. K. Harman (eds), *TREC: Experiment and Evaluation in Information Retrieval*, MIT Press, Cambridge, MA, USA, chapter 3, p. 53–75, 2005.

- Cazella S. C., Campos Alvares L. O., "Modeling User's Opinion Relevance to Recommending Research Papers", *UM'05: Proceedings of the 10th International Conference on User Modeling*, vol. 3538 of *LNCS*, Springer, p. 327–331, 2005.
- Elmacioglu E., Lee D., "On six degrees of separation in DBLP-DB and more", *SIGMOD Record*, vol. 34, n° 2, p. 33–40, 2005.
- Fox E. A., Shaw J. A., "Combination of Multiple Searches", in D. K. Harman (ed.), *TREC-1: Proceedings of the First Text REtrieval Conference*, NIST, Gaithersburg, MD, USA, p. 243–252, February, 1993.
- Hull D., "Using Statistical Testing in the Evaluation of Retrieval Experiments", *SIGIR'93: Proceedings of the 16th annual international ACM SIGIR conference*, ACM Press, New York, NY, USA, p. 329–338, 1993.
- Järvelin K., Kekäläinen J., "Cumulated gain-based evaluation of IR techniques", *ACM Trans. Inf. Syst.*, vol. 20, n° 4, p. 422–446, 2002.
- Karoui H., Kanawati R., Petrucci L., "COBRAS: Cooperative CBR System for Bibliographical Reference Recommendation", *ECCBR'06: Proceedings of the 8th European Conference on Advances in Case-Based Reasoning*, vol. 4106 of *LNCS*, Springer, p. 76–90, 2006.
- Klas C.-P., Fuhr N., "A new Effective Approach for Categorizing Web Documents", *Proceedings of the 22th BCS-IRSG Colloquium on IR Research*, April, 2000.
- Lee J. H., "Analyses of Multiple Evidence Combination", *SIGIR'97: Proceedings of the 20th annual international ACM SIGIR conference*, ACM Press, New York, NY, USA, p. 267–276, 1997.
- Ley M., "The DBLP Computer Science Bibliography: Evolution, Research Issues, Perspectives", *SPIRE'02 : Proceedings of the 9th international conference on String Processing and Information Retrieval*, vol. 2476 of *LNCS*, Springer, p. 1–10, 2002.
- Manning C. D., Raghavan P., Schütze H., *Introduction to Information Retrieval*, Cambridge University Press, 2008.
- McNee S. M., Kapoor N., Konstan J. A., "Don't look stupid: avoiding pitfalls when recommending research papers", *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, ACM, New York, NY, USA, p. 171–180, 2006.
- Milgram S., "The Small-World Problem", *Psychology Today*, vol. 1, n° 1, p. 61–67, 1967.
- Montaner M., López B., de la Rosa J. L., "A Taxonomy of Recommender Agents on the Internet", *Artif. Intell. Rev.*, vol. 19, n° 4, p. 285–330, 2003.
- Reips U.-D., "The methodology of Internet-based experiments", in A. N. Joinson, K. Y. A. McKenna, T. Postmes, U.-D. Reips (eds), *The Oxford Handbook of Internet Psychology*, Oxford University Press, New York, NY, USA, chapter 24, p. 373–390, 2007.
- Salton G., Wong A., Yang C. S., "A Vector Space Model for Automatic Indexing", *Commun. ACM*, vol. 18, n° 11, p. 613–620, November, 1975.
- Sanderson M., Zobel J., "Information Retrieval System Evaluation: Effort, Sensitivity, and Reliability", *SIGIR'05: Proceedings of the 28th annual international ACM SIGIR conference*, ACM, New York, NY, USA, p. 162–169, 2005.
- Yang Z., Hong L., Davison B. D., "Topic-driven Multi-type Citation Network Analysis", *RIAO'10: Proceedings of the 9th international conference on Information Retrieval and its Applications*, April, 2010.