



**HAL**  
open science

## Méta-analyse ordinaire d'enquêtes d'opinion. Application aux usages de l'Internet des objets en entreprise

Rostand Affogbolo, Claire Gauzente, Alain Guénoche, Pascale Kuntz

### ► To cite this version:

Rostand Affogbolo, Claire Gauzente, Alain Guénoche, Pascale Kuntz. Méta-analyse ordinaire d'enquêtes d'opinion. Application aux usages de l'Internet des objets en entreprise. 18ème Conférence Internationale sur l'Extraction et la Gestion des Connaissances, Jan 2018, Paris, France. pp.11-22. hal-01997666

**HAL Id: hal-01997666**

**<https://hal.science/hal-01997666v1>**

Submitted on 29 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Méta-analyse ordinaire d'enquêtes d'opinion

## Application aux usages de l'*Internet des objets* en entreprise

Rostand Affoglobolo\*, Claire Gauzente\*, Alain Guénoche\*\*, Pascale Kuntz\*\*\*

\*Laboratoire d'Economie et de Management, Université de Nantes  
claire.gauzente@univ-nantes.fr  
rostand.affoglobolo@etu.univ-nantes.fr

\*\* Institut de Mathématiques de Marseille (AMU - CNRS)  
alain.guenoche@univ-amu.fr

\*\*\* Laboratoire des Sciences du Numérique, Université de Nantes  
pascale.kuntz@univ-nantes.fr

**Résumé.** La multiplicité des enquêtes d'opinion sur un même sujet nécessite la construction de synthèses qui agrègent les résultats obtenus dans des conditions indépendantes. Dans cet article, nous proposons une nouvelle approche ordinaire de méta-analyse qui consiste à rechercher un ordre consensus qui rend compte « au mieux » des ordres partiels entre les modalités issus des résultats des différentes enquêtes. Nous modélisons ce problème par une variante d'une recherche d'un ordre médian sur les sommets d'un graphe orienté pondéré et nous développons un algorithme de séparation-évaluation pour le résoudre. Notre approche est appliquée sur un ensemble d'enquêtes internationales portant sur les motivations et les freins à l'intégration de l'Internet des Objets dans les entreprises.

## 1 Introduction

La prolifération du nombre de publications scientifiques et d'enquêtes sur un sujet donné s'accompagne de l'essor croissant des méthodologies de synthèse des connaissances. Les méthodes agrégatives, souvent regroupées sous le terme de procédures « méta-analytiques », visent à combiner des résultats quantitatifs d'études indépendantes. Si l'historique de la méta-analyse remonte aux travaux de Pearson au début du XX<sup>ème</sup> siècle sur l'analyse de plusieurs études portant sur le vaccin contre la fièvre typhoïde, puis à ceux de Fisher et Cochran dans les années 30, les articles de Cohen (1962), Light et Smith (1971) et Glass (1976), qui a introduit le terme de méta-analyse dans le contexte statistique, ont contribué au développement du domaine. Aujourd'hui, la médecine reste un terrain d'application privilégié mais la méta-analyse connaît également un fort engouement en sciences sociales où elle complète et précise les synthèses issues de méthodes interprétatives (Laroche, 2015).

Dans son schéma classique issu des travaux historiques, la méta-analyse vise essentiellement à estimer le degré de relation entre des variables d'intérêt en tenant compte des variations observées entre différentes études. Cependant, ses définitions dans la littérature offrent souvent un cadre plus large qui met l'accent sur la combinaison de résultats quantitatifs de multiples

recherches pour produire une connaissance empirique sur un sujet donné (Littell et al., 2008). C'est dans ce cadre que se positionne cet article qui propose une nouvelle approche méta-analytique basée sur une analyse ordinaire pour contribuer à établir une synthèse de résultats d'enquêtes d'opinions. Plus précisément, on considère ici un ensemble  $E = \{e_1, e_2, \dots, e_m\}$  de  $m$  enquêtes portant sur une même thématique et  $X = \{x_1, x_2, \dots, x_n\}$ , l'ensemble de toutes les modalités considérées sur  $E$  qui correspondent à des points de vue ou opinions que les personnes interrogées peuvent approuver ou non. On dispose de la fréquence d'approbations  $P_A(x_i, e_k)$  de la modalité  $x_i$  dans l'enquête  $e_k$ . La distribution observée des  $P_A(x_i, e_k)$  permet de déduire un ordre sur les modalités pour chaque enquête :  $x_i$  est préférée à  $x_j$  dans l'enquête  $e_k$  si  $P_A(x_i, e_k) > P_A(x_j, e_k)$ . L'objectif est d'établir, à partir de ces fréquences, un ordre global sur l'ensemble  $X$  de toutes les modalités qui permettra d'obtenir une vue générale des opinions des plus aux moins approuvées. Notre démarche n'est pas seulement basée sur les ordres partiels induits sur chaque enquête, difficilement combinables, ni sur les sommes ou moyennes des fréquences d'approbation, très variables d'une enquête à l'autre, mais sur des comparaisons par paires des modalités comparables dans chaque enquête.

D'un point de vue opérationnel, deux questions délicates se posent. La première difficulté est méthodologique : les enquêtes ayant été menées indépendamment par des organismes différents, les modalités de  $X$  ne sont pas nécessairement présentes dans chaque enquête et de plus un codage préalable est nécessaire pour homogénéiser les réponses. La deuxième difficulté, plus importante, est algorithmique car la recherche d'ordres consensus se heurte à des problèmes de complexité (Barthelemy et Monjardet, 1981). S'appuyant sur des travaux antérieurs sur la recherche d'ordres médians dans un tournoi (Barthelemy et al., 1989; Charon et al., 1997), nous proposons dans cet article d'améliorer une méthode de Branch & Bound pour calculer efficacement un ordre total sur  $X$  qui soit le «plus compatible» avec les ordres sur les opinions exprimées pour chacune des enquêtes de  $E$ .

Nous appliquons notre approche sur un ensemble d'enquêtes portant sur les motivations et les freins à l'intégration de l'Internet des Objets (IdO) dans les entreprises. Si ce sujet est très présent dans l'actualité économique, car les entreprises cherchent à tirer partie de cette technologie transformatrice, il a été encore peu exploré dans les sciences sociales. Nous avons donc recueilli les résultats de huit enquêtes récentes réalisées par des organisations et entreprises de premier plan dans le conseil et la technologie (World Economic Forum, The Economist Intelligence Unit, etc) qui ont interrogé des interlocuteurs positionnés en haut de la chaîne hiérarchique (directeur de département, de business unit, etc) dans 6237 entreprises réparties dans le monde. Il s'agit à notre connaissance de la première méta-analyse sur ce sujet et les premiers résultats obtenus semblent questionner des opinions souvent diffusées dans les discours du moment ainsi que des hypothèses adoptées dans les « Business Models » récemment développés dans le contexte de l'industrie 4.0.

## 2 Modélisation du problème et état de l'art

Dans cette partie nous considérons que les modalités renvoyant à des concepts similaires mais rédigées avec des variations ont été recodées sous une même formulation et nous renvoyons au paragraphe 4 consacré à l'application pour les détails du codage. Néanmoins, les enquêtes ayant été menées indépendamment, toutes les modalités  $x_i$  n'apparaissent pas nécessairement dans toutes les enquêtes, et les ordres induits sur les modalités dans chaque enquête

$e_k$  ne sont donc pas directement comparables. Pour contourner cette difficulté intrinsèque à la nature des données, nous construisons préalablement un ordre partiel  $P$  sur  $X$  en comparant les modalités deux à deux ; ce qui nous permet de poser notre problème de méta-analyse comme un problème de recherche d'un ordre total «le plus compatible» avec l'ensemble ordonné  $(X, P)$ .

La recherche d'un ordre total compatible avec un ensemble ordonné a donné lieu à des travaux dans les domaines de la *Théorie du choix social* (vote) et de l'*Agrégation des préférences* (Monjardet, 1973; Barthelemy et Monjardet, 1981). Le problème classique se modélise à l'aide d'un *tournoi*  $T$ . Un tournoi est un graphe complet orienté, qui peut être pondéré ou non, dont les sommets sont les éléments de  $X$  (ici les modalités) et pour lequel il existe un arc entre deux sommets  $x_i$  et  $x_j$  si  $x_i$  est supérieur à  $x_j$  pour  $P$ .

La construction de  $P$  se déduit naturellement de la comparaison des fréquences d'approbation des modalités. Pour toute paire de modalités  $\{x_i, x_j\}$  de  $X$  on compare leurs différences de fréquence d'approbation sur les seules enquêtes dans lesquelles elles ont été proposées :

$$Dif(x_i, x_j) = \sum_{e_k | P_A(x_i, e_k) > 0 \wedge P_A(x_j, e_k) > 0} P_A(x_i, e_k) - P_A(x_j, e_k)$$

Ceci permet de munir l'ensemble des modalités  $X$  d'un ordre sur les paires. Ainsi,  $x_i$  est supérieur à  $x_j$  si  $Dif(x_i, x_j) > 0$  et on pose un arc dans  $T$  de  $x_i$  vers  $x_j$  de poids  $w(x_i, x_j) = Dif(x_i, x_j)$ . Si  $w(x_i, x_j) > 0$  on a  $w(x_j, x_i) = 0$  et réciproquement.

Si l'ensemble des relations entre paires n'induit pas de circuit, le tournoi est dit *transitif* et tout ordre qui respecte l'orientation des arcs de  $T$  est compatible avec l'ensemble ordonné  $(X, P)$ . En revanche, si il existe un circuit, par exemple  $x_i \rightarrow x_j \rightarrow x_k \rightarrow x_i$ , alors aucun ordre total n'est compatible, puisque l'un des arcs  $\{(x_i, x_j), (x_j, x_k), (x_k, x_i)\}$  est orienté en sens contraire de l'ordre. Ces arcs orientés à l'opposé d'un ordre total sont appelés *arcs-retour*.

## 2.1 Exemple

Considérons le cas de trois enquêtes qui portent sur cinq modalités dont la table des pourcentages d'approbation est donnée dans la Table 1. Les cases vides correspondent aux modalités non proposées dans les enquêtes.

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
$e_1$	40	50		20	
$e_2$		40	60		30
$e_3$		10		60	40

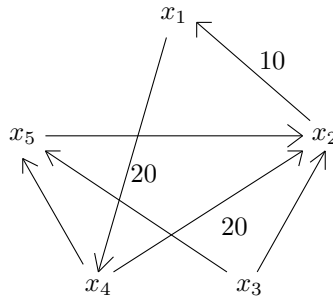
**Table 1** : Pourcentage d'approbation des 5 modalités dans les 3 enquêtes de l'exemple.

Les comparaisons par paires de ces modalités, sur les seules enquêtes où elle sont simultanément proposées donnent la Table 2 des poids des arcs du graphe orienté.

	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$
$x_1$	-	0	0	20	0
$x_2$	10	-	0	0	0
$x_3$	0	20	-	0	30
$x_4$	0	20	0	-	20
$x_5$	0	20	0	0	-

**Table 2** : Poids des arcs du graphe orienté issu des comparaisons par paires de modalités

Il lui correspond le graphe orienté de la Figure 1. Il ne contient qu'un seul circuit  $(x_1, x_4, x_2)$  dont les poids des arcs sont les seuls indiqués. L'arc de poids minimum est  $(x_2, x_1)$  de poids 10. C'est le seul arc retour de l'ordre optimal  $(x_3, x_1, x_4, x_5, x_2)$



**Figure 1** : Graphe orienté correspondant à la Table 2

## 2.2 Méthodologie

Dans le cas pondéré qui nous intéresse dans la suite, l'ordre le plus compatible avec l'ensemble ordonné  $(X, P)$  est celui pour lequel la somme des poids des arcs-retour est minimale. Cette somme est appelée *écart* à l'ensemble ordonné et, quand le tournoi modélise les comparaisons par paires d'éléments qui proviennent d'ordres totaux, un ordre associé à l'écart minimal est appelé *ordre médian* car il minimise la somme de la distance de Kendall aux ordres totaux. Ce n'est plus nécessairement vrai dans le cas d'ordres partiels que nous avons ici, mais nous conservons dans la suite le terme de médian pour désigner un ordre à écart minimum d'un ensemble ordonné quelconque.

La recherche d'un ordre médian, aussi connue sous le nom de problème de Kemeny (Kemeny, 1959), revient à rendre le tournoi transitif par retournement des arcs-retour dont la somme des poids est minimale. Ce problème a une longue histoire algorithmique, qui débute avec Slater (1961) pour les tournois non pondérés, puis s'est étendue aux tournois pondérés (Barthelemy et al., 1989; Charon et al., 1997). Le problème a été prouvé NP-difficile (Hudry, 1989).

En pratique, son calcul s'effectue par une méthode de Branch & Bound qui consiste à développer une arborescence dont les sommets sont les sections commençantes d'un ordre qui peut se prolonger en ordre médian (Guénoche, 1977). Cependant, lorsque le tournoi contient de nombreux circuits, le parcours de l'arborescence de recherche peut rapidement devenir inapplicable, même pour quelques dizaines de sommets (Barthelemy et al., 1989). En effet, la taille de l'arborescence peut dépasser le million de noeuds et le temps de recherche de la feuille à explorer pour prolonger la section commençante est pénalisant.

Dans notre contexte applicatif, l'ensemble ordonné  $(X, P)$  n'est pas un tournoi : environ 40% des paires de modalités ne sont pas ordonnées puisque toutes les modalités n'apparaissent pas dans toutes les enquêtes. Cependant, nous avons développé une méthode de Branch & Bound fortement inspirée de celle utilisée pour la recherche d'un ordre médian dans un tournoi pondéré et nous améliorons l'exploration de l'arborescence eu égard aux travaux antérieurs

en tenant compte du meilleur ordre construit sur chaque section commençante et d'un coût minimum de l'établissement d'un ordre sur les éléments non classés.

### 3 Algorithme pour un ordre médian

Dans la suite, nous considérons un graphe pondéré  $G$  associé à l'ensemble ordonné  $(X, P)$  construit selon le même principe qu'un tournoi où la pondération d'un arc entre deux modalités  $x_i$  et  $x_j$  est égale à  $w(x_i, x_j)$ . Nous cherchons un ordre total sur  $X$  dont la somme des poids des arcs-retour sur  $G$  soit minimale. Comme dans le cas de la construction d'un ordre médian sur un tournoi, nous développons une arborescence de recherche dont les sommets sont des *sections commençantes*, c'est à dire des débuts d'ordres totaux sur une partie de  $X$  que l'on prolonge aux étapes suivantes jusqu'à l'obtention d'un ordre total sur  $X$ . Dans la suite, on note  $O = (o_1, o_2, \dots, o_n)$ , avec  $o_i \in X$ , un ordre total sur  $X$ . Une feuille de l'arborescence courante est une section commençante  $(o_1, o_2, \dots, o_k)$  d'un ordre total sur une partie  $X_s \subset X$ . Le complément de  $X_s$  dans  $X$  est noté  $Y = X - X_s$ .

L'algorithme de B & B peut se décrire en trois étapes :

1. Etablir par une (ou plusieurs) heuristique(s) une borne supérieure, notée  $B_{sup}$  de l'écart d'un ordre total au tournoi ; elle doit être la plus faible possible ;
2. Initialiser les feuilles de l'arbre de recherche avec tous les sommets  $o_i$  tels que  $W^1(o_i) \leq B_{sup}$  où  $W^1(o_i)$  est le poids de  $o_i$  ;
3. Tant que la section commençante prolongée n'est pas un ordre total
  - Considérer une section commençante (une feuille) de poids minimum ;
  - Tester le coût de son extension avec chaque sommet non placé dans la section ;
  - Si le poids de la section étendue reste inférieur ou égale à  $B_{sup}$ , créer une nouvelle feuille dans l'arbre

#### 3.1 Evaluation des sommets de l'arborescence

L'efficacité de la recherche arborescente repose en grande partie sur le calcul d'une borne pour le poids  $W^k(o_1, o_2, \dots, o_k)$  d'une section commençante. Ce poids doit être inférieur ou égal au poids de tout ordre total ayant cette section commençante. Il est défini par la somme de trois termes :

- la somme des poids des arcs-retour du sous-graphe de  $G$  induit par les sommets de  $X_s$ . Cette somme est égale à  $\sum_{1 \leq i < j \leq k} w(o_j, o_i)$
- la somme des poids des arcs de  $G$  dont l'origine  $y$  est dans  $Y$  et l'extrémité dans  $X_s$  :  $\sum_{y \in Y} \sum_{j=1, \dots, k} w(y, o_j)$
- une borne inférieure  $B_{inf}(Y)$  du poids de l'ordre non encore déterminé sur la partie finissante  $Y$ .

Le calcul de la borne inférieure  $B_{inf}$  est une adaptation du calcul d'une borne inférieure pour le problème de Slater proposée dans (Charon et al., 1996). Rappelons que, dans le cas d'un tournoi qui est un graphe complet, pour supprimer tous les circuits et donc construire un ordre médian, il suffit d'éliminer les circuits de longueur 3 par retournement d'arcs. On peut alors évaluer une borne inférieure du coût de ces suppressions de la façon suivante. La suppression d'un 3-circuit coûte au moins le plus petit poids des trois arcs, noté  $w_{min}$ . Mais

si on retourne l'arc de poids  $w_{min}$ , on supprime également tous les 3-circuits qui contiennent cet arc. Il faudrait donc rechercher un ensemble d'arcs de poids minimum dans l'ensemble des 3-circuits arc-disjoints. A défaut d'optimum, ceci peut être approximé par un algorithme glouton. Il suffit de ranger tous les 3-circuits dans l'ordre des poids décroissants et de parcourir cette liste ; chaque fois que l'on retient un arc, on ignore les 3-circuits suivants qui contiennent cet arc. La somme des poids des arcs retenus est une borne inférieure de l'écart d'un ordre médian.

On remarque que cette procédure peut en fait s'appliquer à tout sous-graphe induit par un sous-ensemble de sommets du tournoi. On peut donc l'utiliser pour calculer une borne inférieure  $B_{inf}(Y)$  sur une partie finissante  $Y$ . Notons cependant que dans notre cas  $G$  n'étant pas un graphe complet, pour avoir la précision de la borne inférieure calculée sur un tournoi il faudrait considérer tous les circuits minimaux (sans corde) puisqu'il se peut qu'il n'y ait pas de 3-circuits. Ce calcul pouvant s'avérer très coûteux en temps, nous nous en tenons aux 3-circuits en acceptant une dégradation possible de la précision de la borne inférieure  $B_{inf}(Y)$ .

Au total, l'évaluation  $W^k(o_1, o_2, \dots, o_k)$  d'une section commençante  $(o_1, o_2, \dots, o_k)$  est donc définie par :

$$W^k(o_1, o_2, \dots, o_k) = \sum_{1 \leq i < j \leq k} w(o_j, o_i) + \sum_{y \in Y} \sum_{j=1, \dots, k} w(y, o_j) + B_{inf}(Y).$$

Puisqu'elle ne tient pas complètement compte de l'ordre non encore déterminé sur  $Y$ , elle est inférieure ou égale au poids de tout ordre total commençant par  $(o_1, o_2, \dots, o_k)$ .

Par rapport à l'algorithme détaillé dans (Guénoche, 2017), nous avons conservé la procédure d'optimisation locale (qui teste les transpositions des extrémités des arcs-retour) et la procédure de décomposition du tournoi. Elles permettent de calculer un ordre approché et donc une borne supérieure de l'écart au tournoi. Mais nous avons introduit deux améliorations qui permettent de limiter l'exploration de l'arborescence de recherche :

- la prise en compte de la borne inférieure du coût d'un ordre sur toute partie finissante, décrite ci-dessus ;
- la gestion du meilleur ordre sur une partie commençante.

Plus, précisément, lors du premier calcul d'une partie commençante on l'enregistre avec son évaluation et on prolonge l'arborescence. Si l'on retrouve cette partie, la décision dépend de la comparaison de son évaluation avec celle obtenue précédemment : si elle est plus élevée on rejette la partie et on ne prolonge pas l'arborescence ; si elle est plus faible on la conserve en mettant à jour l'évaluation et on prolonge l'arborescence, et en cas d'égalité si on ne cherche qu'un seul ordre médian on peut également la rejeter. Cette stratégie nécessite la construction d'une structure de donnée pour les fonctions caractéristiques des parties de  $X$  afin de mémoriser la valeur du meilleur ordre sur  $X_s$ .

Nous avons mesuré expérimentalement que ces procédures permettent un gain significatif en temps de calcul et en taille de l'arborescence. Cette efficacité se mesure par le rapport entre la taille de l'arbre sans les utiliser et la taille de l'arbre quand on les applique.

### 3.2 Efficacité des nouvelles procédures

Nous avons réalisé des simulations en tirant deux pourcentages pour chaque paire de modalités. La comparaison des valeurs permet de quantifier la préférence de l'une en faveur de

l'autre. On construit tout d'abord un tournoi en prenant comme poids de chaque arc la différence positive de ces pourcentages. Le tournoi généré est transitif car les arcs sont orientés dans le sens des indices croissant des modalités. Il est donc transitif et le seul ordre médian est l'ordre naturel.

Pour simuler le fait que toutes les modalités ne sont pas systématiquement évaluées, nous avons introduit un taux d'incomparabilité, noté  $Inc$ . Les  $Inc \times n(n-1)/2$  valeurs tirées au hasard sont mises à 0, et ainsi les deux modalités sont incomparables. Et pour graduer l'écart à la transitivité, nous bruitons le tournoi par des échanges aléatoires de valeurs de poids symétriques :  $w(x_j, x_i) \leftarrow w(x_i, x_j)$  et  $w(x_i, x_j) = 0$ . Plus ces échanges sont nombreux, plus on s'écarte de la transitivité et plus l'ordre médian est difficile à calculer et donc les arbres de recherche sont grands. Ce second paramètre est donc le taux d'échanges à partir du tournoi transitif noté  $Swap$  qui correspond au nombre de paires échangées, soit  $Swap \times n(n-1)/2$ .

Pour construire un ordre médian, on commence par calculer une borne supérieure, par application de l'heuristique d'optimisation locale, puis par décomposition en  $1 + \lfloor \frac{n}{10} \rfloor$  classes et enfin par une nouvelle optimisation locale, on calcule une borne supérieure de l'écart. C'est alors que démarre la construction de l'arborescence une fois avec les procédures à évaluer et l'autre fois sans ; cette dernière ne peut n'aboutir que si la place mémoire est suffisante. Notre programme limite la taille de l'arborescence à 2 000 000 de noeuds. La Table 3 indique, pour  $n = 20$ , le rapport moyen des tailles des arbres, sur 100 essais pour  $swap = 5$  ou 10, pour 50 essais pour  $swap = 15$  et pour 10 essais si  $swap = 20$ . On remarquera que nous n'avons pas pu établir de valeur moyenne de la taille de l'arbre sans appliquer les procédures pour  $swap = 20$  et  $Inc = 30$ , non plus que quand  $Inc = 40$  pour  $Swap = 15$  ou 20, parce que l'un des arbres, construit sans les procédures, dépassait les limites autorisées. Pourtant, la taille moyenne des arbres avec les procédures reste inférieure à 6000 noeuds !

$Inc Swap$	5	10	15	20
0	3,7	8,6	13,3	25,4
10	6,6	20,7	27,3	46,4
20	17,6	27,5	46,0	54,0
30	19,2	25,6	41,2	?
40	41,0	93,8	?	?

**Table 3 :** Valeurs moyennes des rapports entre les tailles des arbres de  $B\&B$  sans les procédures nouvelles et avec ces procédures.

Ces estimations, montrent que l'on gagne un facteur très important sur la taille de l'arborescence. Ce facteur va croissant quand l'écart à la transitivité augmente et aussi quand le taux d'incomparabilité devient plus important. Ceci est prévisible, du fait que nos graphes sont loin d'être complets et donc qu'il y a beaucoup d'ordres sur les sections commençantes qui sont ex-aequo. C'est en n'en conservant qu'un seul qu'on gagne en efficacité.

Pour vérifier que l'on peut calculer des ordres médians sur des graphes plus importants, nous avons testé des séries de 5 graphes aléatoires. Pour  $n = 30$ ,  $Swap = 10$ ,  $Inc = 30$ , un ordre médian est construit à l'aide d'un arbre de 65 000 noeuds en moyenne, le maximum observé étant 228 072. De même pour  $n = 40$ ,  $Swap = 5$ ,  $Inc = 40$ , la taille moyenne des arbres étant 181 493 et le maximum 361 458 noeuds.



## 4 Application

Notre cadre applicatif concerne le déploiement de l'*Internet des Objets* (IdO) dans les entreprises. Le terme Internet des Objets a émergé à l'Auto-ID Center du MIT à la fin des années 1990 pour désigner « une infrastructure intelligente mettant en lien des objets, de l'information et des humains à travers des réseaux d'ordinateurs, avec la RFID comme technologie de base pour sa réalisation » (Brock, 2001). Différentes définitions ont été proposées depuis, et l'essor de l'IdO démarre véritablement avec le rapport « Internet Reports – The Internet of Things » de l'ITU (International Telecommunication Union) en 2005 qui présente à la fois les technologies mobilisées par l'IdO et le potentiel du marché. Aujourd'hui, les capacités offertes par l'IdO dans le secteur industriel visent essentiellement quatre usages - le monitoring, le contrôle, l'optimisation et l'autonomie - et les exemples de déploiement se multiplient (Porter et Heppelmann, 2014). Cependant, selon une étude américaine récente du Boston Consulting Group (Rose et al., 2016), nombre d'industriels ne perçoivent pas encore l'étendue des opportunités offertes par l'industrie 4.0 ou n'en font pas un impératif même si ils en reconnaissent le potentiel. Les raisons profondes de ces différentes attitudes, qui oscillent entre l'adhésion enthousiaste et la réticence craintive, sont dues aux présentations plus orientées vers le « grand public » que vers le milieu industriel.

Pour éclaircir la question, avant de mener une étude de terrain plus approfondie, nous avons analysé les résultats proposés par des enquêtes internationales récentes et nous proposons ici une première méta-analyse. Ces enquêtes, menées sur un total de plus de 6 000 cadres de haut niveau, comportent deux familles de questions : celles concernant les facteurs incitant à l'adoption de l'IdO et celles concernant les barrières limitant leur adoption. Dans l'analyse menée dans cet article, nous traitons ces deux familles de façon indépendante et calculons donc un ordre consensus pour chaque ensemble de modalités associé à ces deux familles. La première (facteurs incitatifs) comporte au total 38 modalités et la seconde (facteurs de réticence) 45 modalités.

### 4.1 Codage préalable des modalités

D'un point de vue méthodologique, les enquêtes ayant été menées indépendamment, un codage des modalités de réponses a été nécessaire pour construire une base d'analyse homogène. Ce codage est basé ici sur les concepts déployés dans la construction des « Business Models » (BM) en sciences de gestion. De façon générale, un BM repose sur le triptyque création de valeur (l'offre), délivrance de valeur (le service aux clients) et capture de valeur (les revenus) (Teece, 2010). A partir de ces trois éléments de base, différentes architectures à la fois descriptives et explicatives ont été proposées et nous retenons ici celle proposée par Osterwalder et Pigneur (2010) qui est reconnue comme l'une des plus complètes. Elle est organisée autour de neuf blocs qui sont liés entre eux par des relations identifiées :

1. La proposition de valeur (offre de produits et services),
2. le segment de clientèle,
3. le réseau de distribution,
4. la relation client,
5. les activités clés ou configuration de valeur (arrangement des activités et ressources),

6. les ressources et compétences clés (nécessaires à la mise en œuvre du BM),
7. le réseau de partenaires ou réseau de valeur,
8. la structure de coûts, et
9. le modèle de revenus.

Ces blocs contiennent eux-mêmes des sous-blocs que nous ne détaillons pas ici. Nous avons utilisé cette structuration pour classer l'ensemble des modalités de réponses proposées dans les différents questionnaires. Puis, les modalités classées dans un même sous-bloc et jugées similaires ont été requalifiées en une seule modalité. Par exemple, les modalités « réaliser de la croissance sur des marchés connexes » et « adresser de nouveaux clients » ont été requalifiées en la modalité « pénétrer un nouveau marché » du bloc (2) « segment de clientèle ».

## 4.2 Résultats

A titre indicatif, la taille des arborescences de recherche est pour les modalités d'incitation de 13 742 sommets et pour les modalités de frein de 418 140 sommets.

Pour cette première analyse, nous avons retenu pour chaque ordre médian ses deux extrémités (modalités pas ou peu dominées et modalités très dominées). Cependant, en pratique, il était nécessaire de définir un indice qui puisse guider cette dichotomie. Nous n'avons pas trouvé de question équivalente dans la littérature et nous avons donc développé un nouvel indice de contribution des modalités à l'ordre qui se définit comme suit. Pour chaque modalité  $o_i$  de l'ordre médian  $O_M$ , sa contribution  $C(o_i)$  est égale à la somme des poids des arcs sortants de  $o_i$  ayant pour extrémités des sommets dominés (à droite de  $o_i$  dans  $O_M$ ) moins la somme des poids des arcs dominant  $o_i$ , ayant des sommets d'origine à gauche de  $o_i$  dans  $O_M$ .

$$C(o_i) = \sum_{j>i} w(o_i, o_j) - \sum_{j<i} w(o_j, o_i)$$

Plus il y a d'arcs sortant à droite plus la modalité est dominante et inversement plus il y a d'arcs venant de gauche plus la modalité est dominée. Les modalités qui sont associées à peu d'arcs dans le graphe  $G$  ou qui en ont autant venant de part et d'autre sont plus « neutres » eu égard à la force des opinions exprimées et sont donc situées au milieu de l'ordre.

Pour l'adoption de l'IdO, les modalités les plus dominantes concernent majoritairement deux blocs du BM : le modèle de revenus (9) avec les modalités « améliorer la rentabilité », « nouveaux revenus » et les ressources et compétences (6) avec les modalités « collecter de nouvelles données », « optimiser la productivité des actifs », « améliorer l'efficacité opérationnelle des ressources ». S'y ajoute le bloc structure de coût (8) avec la modalité « réduire les coûts récurrents ou opérationnels ». Ces résultats vont à rebours des discours souvent relevés dans la communication professionnelle ainsi que dans certaines conclusions de travaux théoriques récents sur les Business Models qui soutiennent que la proposition de valeur est le bloc le plus impacté par l'introduction de l'IdO au sein des activités de l'entreprise (Arnold et al., 2016; Dijkman et al., 2015). Cette priorité ne ressort pas de notre méta-analyse : les modalités portant sur la création de valeur (e.g. « créer une nouvelle proposition de valeur », « changer de proposition de valeur », etc) se retrouvent à la fois en fin de l'ordre médian associé aux facteurs incitatifs de l'IdO et en fin de celui associé aux difficultés d'adoption. En revanche, l'impact de l'IdO sur la transformation à travers la ré-ingénierie des processus business, également soulignée dans la littérature (Ferretti et Schiavone, 2016), est confirmé par l'analyse (e.g.

« améliorer l'efficacité opérationnelle des ressources », « optimiser la productivité des actifs »).

Pour les barrières rencontrées, le bloc (7) partenariats et environnement est le moins dominé avec les modalités « faiblesse de l'infrastructure publique », « absence de standard ouvert pour la technologie », « incertitude sur la stabilité des partenaires ». Ainsi, les entreprises conçoivent bien que l'IdO contribue à améliorer leurs revenus et leurs ressources et compétences, mais à condition que les interactions avec les parties prenantes externes soient plus efficaces.

## 5 Conclusions

Dans cet article nous avons proposé une nouvelle démarche de méta-analyse permettant d'extraire un ordre consensus sur les modalités évaluées dans un ensemble d'enquêtes d'opinion conduites indépendamment. Cet ordre permet de refléter les principales tendances exprimées et, en particulier, de distinguer les modalités dominantes dans l'ensemble des enquêtes de celles qui sont dominées, c'est-à-dire qui sont peu souvent préférées à d'autres. La recherche de l'ordre consensus a été posée ici comme une recherche d'un ordre médian des sommets dans un graphe orienté pondéré qui modélise les comparaisons par paire des modalités sur l'ensemble des enquêtes. Ce problème se ramenant à un problème NP-difficile nous avons adapté une approche de *Branch & Bound* pour le résoudre. Nous avons nettement amélioré un programme existant pour l'appliquer à un graphe orienté relativement peu dense, par rapport à un graphe orienté complet<sup>1</sup>. Il permet de construire un ordre optimal à l'aide d'arborescences de plusieurs centaines de milliers de sommets et s'avère bien adapté à la méta-analyse d'enquêtes où la taille des échantillons interrogés peut être grande mais où l'ensemble des modalités à ordonner ne dépasse pas quelques dizaines.

Avec peu de travaux préalables sur lesquels nous appuyer, pour commencer l'analyse, nous avons partitionné initialement l'ensemble des modalités de réponse en deux familles (facteurs incitatifs / facteurs de réticence) que nous avons traitées indépendamment. Cependant, les questions sur les « corrélations » entre ces différents facteurs se posent naturellement et nous prévoyons de les analyser prochainement. Elles soulèvent des problèmes statistiques non triviaux puisque nous ne disposons pas des réponses individuelles mais des réponses agrégées à l'échelle de chaque enquête.

Les premiers résultats obtenus concernant l'analyse des motivations et les freins à l'intégration de l'Internet des Objets dans les entreprises sont particulièrement intéressants car ils questionnent des hypothèses récentes de la littérature en sciences de gestion sur l'impact de l'IdO sur les « Business Models ». Par la mise en lumière de résultats contre-intuitifs et à rebours de la théorie, cette nouvelle approche de méta-analyse comporte des implications de deux ordres. Pour les chercheurs en sciences de gestion, elle invite à examiner de façon plus approfondie les modèles théoriques sur lesquels se fondent leurs observations et analyses des organisations. Pour les praticiens, elle suggère de revisiter les perceptions de l'IdO afin d'aller au-delà du discours général et de concentrer les efforts de transformation et d'accompagnement managérial sur les points les plus saillants. En sus de la méta-analyse, les questions posées devront être approfondies à travers une étude de terrain qui est en cours de planification.

---

1. Ce programme en C peut être obtenu par simple demande aux auteurs.

## Références

- Arnold, C., D. Kiel, et K. Voigt (2016). How the industrial internet of things changes business models in different manufacturing industries. *International Journal of Innovation Management* 20(08), 1640015.
- Barthelemy, J., A. Guenoche, et O. Hudry (1989). Median linear orders : Heuristics and a branch and bound algorithm. *European Journal of Operational Research* 42(3), 313 – 325.
- Barthelemy, J. et B. Monjardet (1981). The median procedure in cluster analysis and social choice theory. *Mathematical Social Sciences* 1(3), 235 – 267.
- Brock, D. (2001). The Electronic Product Code (EPC) - A Naming Scheme for Physical Objects. *MIT Auto-ID Center White Paper*.
- Charon, I., A. Guénoche, O. Hudry, et F. Woïrgard (1997). New results on the computation of median orders. *Discrete Mathematics* 165-166(Supplement C), 139 – 153. Graphs and Combinatorics.
- Charon, I., O. Hudry, et F. Woïrgard (1996). Ordres médians et ordres de Slater des tournois. *Mathématiques et Sciences humaines* 133, 23–56.
- Cohen, J. (1962). The statistical power of abnormal-social psychological research : A review. *The Journal of Abnormal and Social Psychology* 65(3), 145–153.
- Dijkman, R., B. Sprenkels, T. Peeters, et A. Janssen (2015). Business models for the internet of things. *International Journal of Information Management* 35(6), 672 – 678.
- Ferretti, M. et F. Schiavone (2016). Internet of things and business processes redesign in seaports : The case of Hamburg. *Business Process Management Journal* 22(2), 271–284.
- Glass, G. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher* 5(10), 3–8.
- Guénoche, A. (1977). Un algorithme pour pallier l'effet Condorcet. *RAIRO - Operations Research - Recherche Opérationnelle*, 11(1), 77–83.
- Guénoche, A. (2017). Analyse des Préférences et Tournois Pondérés. *Journal of Interdisciplinary Methodologies and Issues in Science Graphs and social systems*.
- Hudry, O. (1989). *Recherche d'ordres médians, complexité, algorithmique et problèmes combinatoires*. Ph. D. thesis, Ecole Nationale Supérieure des Télécommunications.
- Kemeny, J. (1959). Mathematics without numbers. *Daedalus* 88(4), 577–591.
- Laroche, P. (2015). *La méta-analyse – Méthodes et applications en sciences sociales*. Editions de Boeck.
- Light, R. et P. Smith (1971). Accumulating evidence : Procedures for resolving contradictions among different research studies. *Harvard Educational Review* 41(4), 429–471.
- Littell, J., J. Corcoran, et P. V. (2008). *Systematic reviews and meta-analysis, Pocket Guide to Social Work Research Methods*. Oxford University Press.
- Monjardet, B. (1973). Tournois et ordres médians pour une opinion. *Mathématiques et Sciences humaines* 43, 55–70.
- Osterwalder, A. et Y. Pigneur (2010). *Business Model Generation : A Handbook for Visionaries, Game Changers, and Challengers*. Hoboken : Wiley & Sons Inc.

## Méta-analyse ordinaire d'enquêtes d'opinion

- Porter, M. et J. Heppelmann (2014). How smart connected products are transforming competition. *Harvard Business Review* 92(11), 64–88.
- Rose, J., V. Lukic, T. Milon, , et A. Cappuzzo (2016). Sprinting to value in industry 4.0.
- Slater, P. (1961). Inconsistencies in a schedule of paired comparisons. *Biometrika* 48(3/4), 303–312.
- Teece, D. (2010). Business models, business strategy and innovation. *Long Range Planning* 43(2), 172 – 194. Business Models.

## Summary

The multiplicity of opinion surveys on a same topic requires the construction of summaries which aggregate results obtained in independent conditions. In this paper, we propose a new ordinal meta-analysis which consists in computing a consensus order reflecting the partial orders between modalities deduced from the results of the different surveys. We set this problem as a variant of the median linear order of vertices on a weighted digraph and we develop a branch and bound algorithm to solve it. Our approach is applied on a set of international surveys on drivers and concerns about the adoption of Internet of Things in companies.