

# Comment Devenir Cybercondriaque ?

Erick Stattner, Martine Collard

Laboratoire LAMIA  
Université des Antilles et de la Guyane, France  
{estattne, mcollard}@univ-ag.fr

**Résumé.** Nous avons tous déjà eu l'occasion d'effectuer des recherches d'ordre médical sur Internet. Si certains sites spécialisés se refusent à tout diagnostic en ligne, préférant le renvoi vers des professionnels de santé, d'autres en revanche conduisent souvent à des déclarations alarmistes faisant état de situations humaines difficiles. Dans ce travail, nous étudions l'ampleur de ce phénomène et montrons que quel que soit le syndrome recherché, les résultats obtenus conduisent toujours à l'énoncé des mots "cancer" ou "tumeur".

## 1 Introduction

Internet fournit une gigantesque base de connaissances qui permet aujourd'hui d'obtenir des informations sur tous les sujets du monde. C'est par exemple le cas des sites dédiés aux problèmes médicaux, pour lesquels les utilisateurs peuvent aisément soumettre des problèmes de santé et être au centre d'espaces de discussion. Si certains sites se refusent à tout diagnostic en ligne, d'autres en revanche font souvent étalage de situations individuelles difficiles qui conduisent bien souvent à l'énoncé de diagnostics par des non-professionnels, dont certains évoquant les mots "cancers" ou "tumeurs", ce qui peut avoir un impact psychologique fort sur les visiteurs à la recherche d'informations.

Dans ce travail, nous étudions ce phénomène et montrons que quel que soit le syndrome recherché, les résultats conduisent toujours à l'énoncé des mots "cancer" ou "tumeur". Nous générons pour cela des syndromes en associant des symptômes à différentes parties du corps et analysons les résultats du moteur de recherche et le réseau de pages web sous-jacent (Watts, 2004). Des travaux proches ont par exemple été menés par M. Godwin sur les forums de type Usenet (Godwin, 1994), pour montrer que plus une discussion s'étend dans le temps et plus la probabilité d'y trouver une comparaison impliquant des analogies extrêmes s'approche de un.

La Section 2 détaille notre objectif et la méthodologie suivie. La Section 3 présente les résultats expérimentaux. La Section 4 conclut l'article et présente nos travaux futurs.

## 2 Objectif et méthodologie

Notre objectif est d'étudier, pour une recherche donnée, la probabilité qu'apparaissent les mots "cancer" ou "tumeur" en analysant les résultats renvoyés par Google. Deux études sont ainsi menées : (i) La première vise à évaluer le nombre de cliques nécessaires à l'apparition

## Comment Devenir Cybercondriaque ?

de ces termes si les résultats sont consultés dans leur ordre d'apparition. (ii) La seconde, plus générale, vise à dépasser l'aspect temporel de la navigation, afin d'évaluer la probabilité que ces termes apparaissent quelle que soit la chronologie de consultation des résultats. Nous analysons pour cela l'ensemble du réseau de pages web sous-jacent et évaluons la probabilité de présence des termes sur deux niveaux.

La méthodologie que nous avons adoptée est la suivante. Nous générons un ensemble de syndromes en associant 5 symptômes à 10 parties du corps humain :

– **Symptômes** : Douleur, Irritation, Tache, Grosseur, Saignement

– **Parties du corps** : Tête, Yeux, Nez, Langue, Ventre, Bras, Main, Sexe, Jambe, Pied

50 requêtes sont ainsi obtenues par combinaisons de symptômes et de parties du corps. Nous effectuons ensuite une recherche automatique de ces syndromes à l'aide du moteur de recherche Google et analysons, selon les approches (i) et (ii), les sites renvoyés. Dans cette première approche du travail, nous nous intéressons uniquement à la présence des termes et ne prenons pas en compte le contexte dans lequel ils apparaissent. On peut en effet supposer que leur seule présence peut avoir un impact psychologique fort chez l'utilisateur qui se sait malade.

## 3 Résultats expérimentaux

### 3.1 Nombre de cliques

Dans l'hypothèse où l'utilisateur parcourt les résultats les uns après les autres, la question est de savoir combien de cliques seront nécessaires pour parvenir à une page affichant les mots "cancer" ou "tumeur". En analysant l'ensemble des résultats renvoyés par le moteur de recherche, nous avons identifié le nombre de cliques nécessaires pour atteindre ces termes. La Figure 1 montre la distribution du nombre de cliques obtenus avec les différentes requêtes.

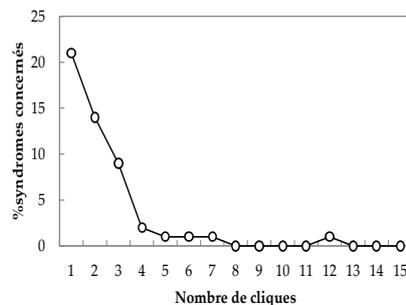


FIG. 1 – Distribution du nombre de cliques nécessaires pour les différentes requêtes

Nous pouvons observer que la majeure partie des requêtes effectuées conduit à des sites présentant les mots "cancer" ou "tumeur" en très peu de cliques. En moyenne, quel que soit le syndrome, 2.28 cliques sont nécessaires pour que l'utilisateur soit dirigé vers une page affichant ces termes. Si ce nombre paraît très faible, il s'explique par le fait que les sites placés en tête de liste sont souvent des sites à très grandes audiences, affichant énormément d'informations tels que des menus, des forums ou des sujets connexes, augmentant ainsi la probabilité que les mots recherchés y soient présents.

### 3.2 Vers une analyse réseau

Pour aller plus loin dans notre étude, nous nous sommes intéressés dans un second temps à la probabilité qu'apparaissent les mots "cancer" ou "tumeur" quel que soit l'ordre de consultation des résultats. Pour cela, nous collectons l'ensemble des résultats renvoyés et générons le réseau de pages web sous-jacent (Barabasi, 2002; Borner et al., 2007) sur deux niveaux (résultats directs et page directement liées). La figure 2 montre un exemple de ce réseau.

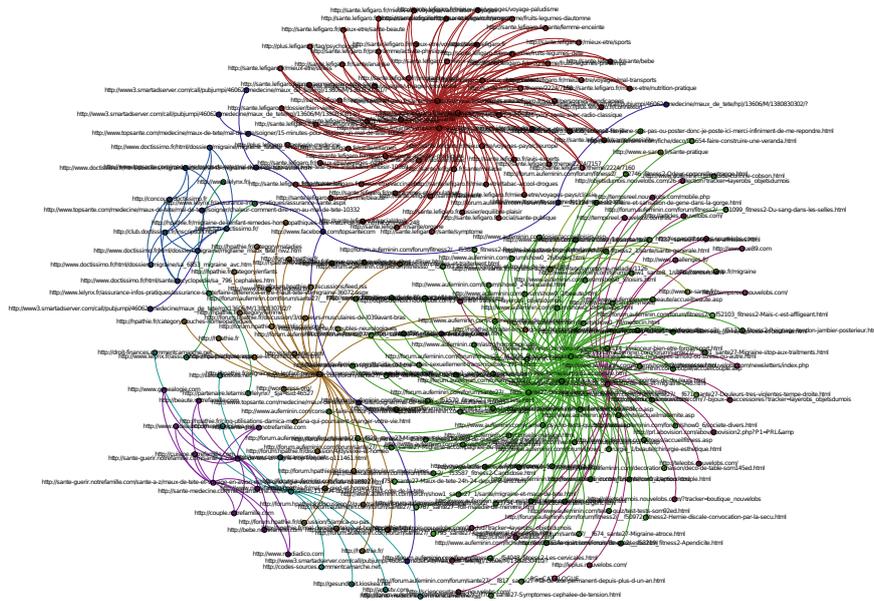


FIG. 2 – Exemple du réseau de pages web généré à partir de la recherche "douleur tête"

Ainsi, en générant et analysant ce réseau pour chacune des requêtes, nous avons calculé la probabilité de présence des mots "cancer" ou "tumeur" aux deux niveaux du réseau, c'est-à-dire le pourcentage de pages présentant au moins l'un des termes à chacun des niveaux. La Figure 3 montre la distribution de cette probabilité (a) au niveau 1 et (b) au niveau 2.

Comme attendu, on observe qu'un fort pourcentage de requêtes a donné lieu à des taux de probabilité de présence relativement élevés au premier niveau du réseau (cf Figure 3(a)). La moyenne des résultats obtenus pour ce premier niveau est de 0.47. Ainsi, quels que soient les syndromes recherchés, la probabilité qu'apparaissent les mots "cancer" ou "tumeur" dès les premiers résultats renvoyés par le moteur de recherche est de 0,47.

L'observation la plus intéressante concerne les résultats obtenus pour le niveau 2 du réseau (cf Figure 3(b)). Nous pouvons en effet observer que globalement, la probabilité que soient présents les mots "cancer" ou "tumeur" ne varie pas radicalement au second niveau. La moyenne des résultats obtenus pour ce second niveau est de 0.39. Quels que soient les syndromes recherchés, la probabilité que soient présents les mots "cancer" ou "tumeur" dans les pages citées par les résultats du moteur de recherche est de 0.39. Nous pouvons expliquer cette différence par le fait que les pages web font souvent référence à des sites qui ne sont pas nécessairement liés sémantiquement au sujet abordé. Les liens publicitaires, les liens vers les sponsors ou les

## Comment Devenir Cybercondriaque ?

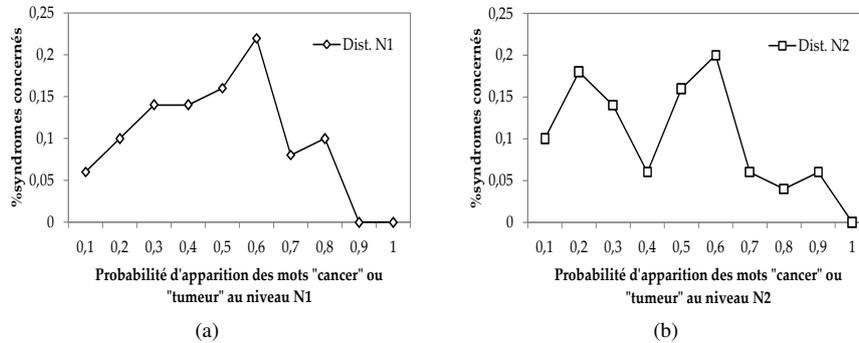


FIG. 3 – Distribution de la probabilité de présence des mots "cancer" ou "tumeur" lors des différentes requêtes (a) au niveau 1 et (b) au niveau 2 du réseau de pages web

partenaires sont par exemple des situations qui peuvent expliquer que le score soit inférieur au niveau 2.

## 4 Conclusion et travaux futurs

Nous avons montré que quel que soit le syndrome recherché sur Internet, il existe toujours dans le réseau de pages web sous-jacent des pages énonçant les mots "cancer" ou "tumeur". Dans ce travail préliminaire, nous avons considéré uniquement la présence du mot sans prendre en compte l'environnement dans lequel il intervient. L'étude pourra être complétée à la fois en tenant compte du contexte et en éliminant les cas de faux positifs qui peuvent être déclenchés par de nombreux éléments parasites sur les pages.

## Références

- Barabasi, A. L. (2002). *Linked : The New Science of Networks*. Perseus Books.
- Borner, K., S. Sanyal, et A. Vespignani (2007). Network science. *Blaise Cronin (Ed) Annual Review of Information Science and Technology 41*, 537–607.
- Godwin, M. (1994). Cyber rights : Defending free speech in the digital age. *The International Journal on Media Management 7*(3&4), 165–166.
- Watts, D. J. (2004). The 'new' science of networks. *Annual Review of Sociology 30*, 243–270.

## Summary

We've all had the opportunity to conduct medical research on Internet. While some specialized sites refuse any online diagnostics, preferring to return to health professionals, others by contrast often lead to alarmist statements alleging difficult human situations. In this work, we study the extent of this phenomenon and we show that whatever is the syndrome searched on the Internet, the results always lead to the occurrence of the words "cancer" or "tumor."