

# La communauté I.E. en France (2008)

Luc Quoniam<sup>\*,\*\*</sup>, Miguel Trigo<sup>\*\*</sup>, Arnaud Lucien<sup>\*</sup>

<sup>\*</sup>Université du Sud Toulon-Var (France)

<sup>\*\*</sup>Universidade Fernando Pessoa, Porto (Portugal)

L'analyse de la communauté "[Intelligence Economique](#)" (I.E.) en France présentée dans cet article a un but démonstratif du potentiel des traitements automatiques et cartographies interactives. Cette analyse est basée sur l'analyse de plusieurs communautés, celle décrite dans l'ouvrage "le top 100 de l'Intelligence Economique" éditée par Intelligence Online (2006), la blogosphère de l'I.E. française, les formations supérieures en I.E. en France

## A. Analyse exhaustive?

Cette analyse n'a pas la prétention d'être exhaustive sur l'I.E. en France. Pour approcher l'exhaustivité, il faudrait déjà au moins adjoindre une analyse de la production scientifique. Mais la collecte des données pour la production scientifique est plus délicate. En effet elle est souvent située dans le [web profond](#), invisible (Sherman & Price), ne s'affiche pas toujours I.E. (article d'application n'utilisant pas le mot I.E., publication en autre langue que le français où I.E. est beaucoup moins connu que le concept d'"intelligence compétitive", l'I.E. se montrant un concept franco-français. Pour le comprendre il suffit de s'amuser à faire une « *petite analyse* » de Wikipedia ([français](#), [anglais](#), [portugais](#)) avec l'excellent outil de création de cartes conceptuelles automatiques ([WikimindMap](#)). à partir de Wikipedia.

Mot clé appliqué sur		
<a href="#">Wikipedia Anglais</a>	<a href="#">Wikipedia français</a>	<a href="#">Wikipedia portugais</a>
Economic Intelligence (pas de page wikipedia anglaise)	<a href="#">Intelligence économique</a>	Inteligencia economica (pas de page wikipedia portugais)
<a href="#">Competitive intelligence</a>	Intelligence compétitive (pas de page en Wikipédia Français)	<a href="#">Inteligência competitiva</a>

## B. Communauté test

Par essence, l'Intelligence Economique est une activité en réseau et de réseau, donc étudier les réseaux de l'I.E. a du sens. Mais l'I.E. n'est pas le seul domaine qui se prête à ce genre d'analyse. En fait toute communauté (scientifique, financière, innovation, lobbying) se prête facilement à des analyses de réseaux humains, de travail, de compétences. L'analyse présentée ici a donc pour but de montrer le potentiel des méthodes d'analyse de réseaux, d'une façon générale tout autant que d'éclairer sur la communauté.

## C. Le point de départ

Une des phases délicates de ce genre d'analyse est la collecte des données, mais aussi le nettoyage et l'homogénéité des données pour obtenir des données analysables automatiquement. Cette analyse a été commencée avec la promotion 2006-2007 du [Master MIET de l'Université du Sud Toulon-Var](#). Profitant de la parution du livre "[le top 100 de l'Intelligence Economique](#)" éditée par Intelligence Online (2006) (Botbol), qui constituait une collecte consistante de données, un groupe d'étudiant a commencé à explorer les possibilités de logiciels d'analyse de graphe pour obtenir des visualisations permettant des interprétations. Cette étude a donné lieu à une publication scientifique (Quoniam, Pauluci, & Jorge, 2007). Au même moment [IE-Lobbying.info](#) (Le portail collaboratif de l'Intelligence économique et du Lobbying) réalisait [la première analyse de la Blogosphère de l'I.E. en France](#), offrant ainsi une deuxième collecte de données, ainsi qu'une représentation graphique des résultats. Mais ces travaux présentaient des imperfections notables:

- Collecte "manuelle des données", rendant le travail initial long et fastidieux.
- Structuration de graphes 2D plans non interactifs (bons pour le papier, mauvais pour le Web et l'interprétation des résultats)

Un travail d'amélioration et de recherche de solutions automatiques permet de présenter aujourd'hui quelques résultats intéressants.

## D. Représentation de petites communautés.

Dans le cas d'analyse de petites communautés, il est possible de recourir à des collectes manuelles et des représentations directes des données collectées. C'est le cas de la carte des [formations en IE en France](#) réalisée par [Christophe Jorge](#), disponible sur son site.

## E. La collecte automatisée

La collecte automatisée se justifie par l'économie de temps obtenue, indispensable surtout dans nos régions au coût de main d'œuvre élevé et à politique de réduction de coût, dès que le temps de collecte "manuel" dépasse quelques heures. Elle se justifie surtout car elle apporte à ces analyses un des critères indispensable à toute analyse "scientifique": la reproductibilité des résultats (Morse; Srinagesh). La collecte automatisée peut être obtenue avec des logiciels spécifiques qui vont collecter des informations publiées. Pour cette raison la communauté mise en évidence par l'ouvrage du top 100 se prête mal à une automatisation de la collecte sur le Web, très peu de personnes mentionnées dans le « top100 » ayant un site professionnel individuel. Pour des raisons évoquées plus haut le web de surface est privilégié, donc seule l'analyse sur la blogosphère a été "suivie". Tout acte de collecte produira des artéfacts. La collecte "manuelle" devra affronter le problème du risque de manque d'exhaustivité comme le risque de subjectivité du collecteur (Bond & Fox). La collecte automatisée produira une analyse de réseau qui est un mélange de l'analyse des liens réels, comme une analyse de stratégies (explicites ou implicites) de positionnement (par rapport aux moteurs de recherche ou par rapport à une communauté) (Smith & Chaffey). Il y a plusieurs façons d'envisager cette collecte automatisée.

# 1. Logiciels d'analyse de back links

Christophe Jorge, qui fit partie de "l'équipe initiale" a actualisée, en 2007, [l'analyse de la blogosphère IE en France](#), en utilisant l'analyse des back links (Clemm, Granville, & Stadler) et propose les résultats sur son site. L'inconvénient de ce mode de collecte est qu'il inclut, en plus de l'analyse de la production les auteurs, une analyse de leur réputation et de leur stratégie de référencement. Dans l'interprétation des graphes finaux, il devient difficile d'interpréter la part de chacun de ces facteurs.

## 2. L'analyse des liens directs

Ce type d'analyse permet de s'affranchir des inconvénients présentés plus haut en analysant uniquement les liens directs sans inclure l'artefact que constituent les "pages ranks". Pour l'actualisation 2008, c'est cette voie qui a été privilégiée. Nous avons utilisé l'excellent "*Link crawler for the social sciences*" que constitue [SocSciBot](#) du [Statistical Cybermetrics Research Group](#). Le point de départ de l'analyse est construit avec une liste de pages initiales *a priori*. Dans le cas de l'analyse présentée ici la liste de pages initiales a été constituée en sélectionnant les blogs qui se démarquaient dans les analyses antérieures et qui continuent leur activité de blog.

- <http://www.ouilsfroids.net/>
- <http://claro.motime.com/>
- <http://reputation.blogspirit.com/>
- <http://www.verbalkint.net/>
- <http://www.activeille.net/index.php>
- [http://affordance.typepad.com/mon\\_weblog/](http://affordance.typepad.com/mon_weblog/)
- <http://www.brainsfeed.com/>
- <http://www.vincentabry.com/>
- <http://www.actulligence.com/>
- <http://www.spyworld-actu.com/>
- <http://www.ie-lobbying.info/wiki/index.php/Accueil>

## F. La représentation des données

### 3. Du Positionnement des points du réseau

Il existe de très nombreuses équipes de recherche (au niveau mondial) qui ont travaillé et continuent de travailler sur le problème complexe de représentations de réseaux de grande taille ([théorie des graphes](#)). A cheval entre mathématiques appliquées, algorithmique, optimisation (du temps de calcul et de l'espace). Les solutions proposées cherchent toutes à identifier les particularités des réseaux analysés (centralité, densité, cliques fermées, point isolés, connections faibles, redondances). De la même façon il existe plusieurs solutions pour "réduire" ces réseaux, quand ils sont trop grands, trop complexes, et donc difficilement appréhendable dans leur totalité (enlever les liens trop "faibles" et/ou trop "forts", travailler clique par clique, enlever les isolés...) (Bornholdt & Schuster; Carrington, Scott, & Wasserman; Nooy, Mrvar, & Batagelj). Mais on comprend vite qu'une solution unique ne peut être que frustrante et partielle et que plusieurs "vues" risquent

d'être nécessaires. Pour en savoir plus il est possible de commencer par la page [Social network](#) de Wikipedia et de très bons logiciels "gratuits" sont disponibles qui font cela très bien ([netdraw](#), [pajek](#), [yED](#)).

## 4. De l'interactivité

La représentation des données de réseaux de liens entre site ou entre des personnes est frustrante quand elle n'est pas une interface de navigation pour/sur le web ou une base de données. De la même façon une interactivité pour déplacer les points ou changer la représentation est plus qu'utile. Ces deux formes d'interactions sont le moyen (partiel) de répondre à la nécessité de vues multiples pour interpréter les résultats. L'interprétation de ces réseaux n'est que partielle, incomplète et imparfaite sur des représentations sans ces deux type d'interaction, ce qui invalide l'utilisation de toute forme de représentation "papier" ou sous la forme de [fichiers "image"](#).

## 5. De la publication des résultats

Il vient du paragraphe précédent que seule une publication des résultats sur un site web va garantir l'interactivité sur les résultats au moins pour la forme d'interactivité que représente la navigation Web à partir du graphe. Pour garantir une bonne diffusion, il faut en plus garantir une mise à disposition dur Internet de façon non dépendante de la machine et du système d'exploitation de l'utilisateur final. Il faut donc avoir une publication des résultats sous forme d'applet Java ou de technologie similaire. Le ["toolkit" de visualisation Préfuse](#) permet de disposer d'un certain nombre d'algorithmes de positionnement permettant une publication interactive sur le Web. La encore, Christophe Jorge nous propose, dans son [actualisation 2007 de l'analyse de la Blogosphère IE](#) plusieurs représentations fort intéressantes en utilisant la bibliothèque Préfuse (graphe radial, graphe social, organigramme)

## 6. De l'augmentation du nombre de dimensions de la représentation.

La représentation 2D "plane", même avec des points bien positionnés montre rapidement ses limites. Il devient indispensable, rapidement d'augmenter le nombre de dimensions pour "aérer" la représentation pour comprendre les structures du réseau.

### a. Plus que la 2D: les hypergraphes

Les hypergraphes permettent d'atteindre ce but sans passer à la 3D. Mais les hypergraphes présentent plusieurs inconvénients. L'algorithme et sa transformation en logiciel sont brevetés, même si on trouve plusieurs routines "domaine public", donc pas si simple à utiliser d'un point de vue "légal". Les Hypergraphes sont en général faits pour des graphes dont tous les points sont interconnectés de façon hiérarchisée. Ils supportent mal les graphes composés de plusieurs cliques indépendantes. Une des façons de contourner cela est de créer des liens "artificiels invisibles" dans le seul but d'interconnecter les cliques. C'est le type de représentation et de subterfuge que nous avons utilisé dans la précédente version publiée des résultats sur le Top 100 de l'I.E. (avec [liens vers le Web](#) ou lien [vers une base de données](#)) et de l'analyse de la [blogosphère I.E.](#) (source [IE-Lobbying.info](#)).

## b. De la 3D au multidimensionnel

Pour aller plus loin, il est indispensable de recourir à la 3D. Mais elle n'est intéressante que si elle autorise aussi l'interactivité et positionnement mentionnés plus haut. Positionnement 3D optimisé, publiable sur le web sous forme de browser interactif et proposant une interaction avec la représentation (déplacement, couleur, manipulation de la forme). Une représentation "3D" qui correspondrait à tous ces critères, en fait, atteint rapidement la 4D (navigation web hypertextuelle) voir la 5D (modification interactive du positionnement de l'aspect...). Des logiciels ou solutions permettant une optimisation du réseau formé et de son positionnement en 3D deviennent beaucoup plus rare, il en est de même d'applet Java capable de publier sur le Web des graphes 3D en offrant Interaction avec la visualisation, navigation web à partir du graphe. Mais des solutions qui en plus acceptent de grandes masses de données collectées automatiquement, soient rapides au traitement et fluides à la représentation, cela devient vraiment très très rare. C'est ce qu'offrent le logiciel [Pajek](#) qui sait traiter de très grands réseaux, possède de nombreuses solutions de traitement et simplification de réseaux complexes et a implémenté l'algorithme de "Fruchterman Reingold" d'optimisation de position 3D. C'est ce que va offrir en complément la solution de visualisation 3D [Javaview](#) qui sait visualiser sur le web, avec plein d'options d'interactions, y compris la navigation Web hypertextuelle, des données "3D" de façon fluide et légère. La seule limite réside dans l'installation, par l'usager de l'application [Java](#). Par opposition nous pouvons montrer/signaler des [visualisation basées sur du format VRML](#), qui force à l'Installation de plug-ins, "viewers" spécifiques dont il n'existe pas forcément une version disponible pour tout type de machine et système d'exploitation et manquent de "fluidité" à la visualisation.

## G. Intégration ou Interopérabilité des solutions.

Obtenir une chaine de traitement qui fasse de façon performante la chaine entière des traitements, d'une collecte efficace et performante à la publication interactive sur le web en passant par d'excellentes optimisations des représentations passe par le développement d'un programme d'un niveau de complexité élevé. Réaliser un tel programme passe par la constitution et le maintien d'une équipe de développeur de pointe multi compétence qui rend ce genre de réalisation possible seulement si de gros investissements sont réalisés. Par contre, les retours sur investissement sont hasardeux dans la mesure où le nombre de "clients" pour ce genre d'application est relativement limité, rendant les prix de commercialisation rapidement prohibitifs. Au lieu de penser une intégration de toute la chaine dans un produit unique, il devient judicieux de penser à une mise en synergie de produits, chacun excellent dans sa spécificité. Pour réaliser cela, il est indispensable de rencontrer des produits "interopérables" que chaque entrée de logiciel sache utiliser la sortie du logiciel précédemment utilisé. Cet aspect est indispensable pour une optimisation des traitements et dispenser l'analyse de longues et fastidieuses phases de conversion des données à chaque passage entre un programme et un autre.

## H. Limitation de la publication scientifique.

Le monde académique et universitaire se meut en apportant foi à la valeur de la publication scientifique, le fameux "Publish or Perish". L'évaluation de nos systèmes de recherche, au niveau collectif et/ou individuel passe par l'évaluation des publications scientifiques selon plusieurs critères, chacun discutable, car dans ce domaine également il n'existe pas de "vérité absolue". Mais nous nous permettons d'en rajouter un: il n'existe pas, à notre connaissance de revue qui correspondent aux critères de "scienticité" communément admis (révision par les pairs: comité scientifique, prise en compte dans les catalogues de publications reconnus) mais qui en plus d'être des publications électroniques pour permettre l'interaction avec le Web, laissent publier en

incluant des applets java afin de garantir l'interaction du lecteur avec le jeu de données, permettant cette lecture des données multidimensionnelle que nous venons de décrire. Dit autrement on assiste, dans le domaine scientifique au même phénomène que l'on peut constater dans d'autres domaines: un décalage croissant entre les possibilités de la technique et la perception de ces possibilités par des communautés qui ne sont pas toujours au fait de la technique et de ce qu'elle peut apporter en termes de facilité pour l'interprétation de données sur un plan qui n'a rien à voir avec la technique. Il y a là un trou à analyser et essayer de combler. Dans l'attente la seule solution de publication est l'auto publication sur son propre site et la soumission à l'évaluation par la réaction, à la façon de faire des blogs, évaluation non reconnue dans les communautés scientifiques, ce qui correspond en fait à travailler « gratuitement » car de façon non « reconnue ».

## I. La blogosphère de l'I.E. 2008

L'[analyse de la Blogosphère de l'I.E.](#), mise à jour en 2008, qui sous-tend l'écriture de cet article, prend en considération toutes les remarques et recommandations faites au long de cet article, puisque les données présentées ont été collectées automatiquement par le logiciel SocSciBot, à partir de la liste initiale de blogs en I.E. sus présenté. La collecte (qui est la phase la plus longue) a généré 5924 pages interconnectées. Une réduction automatique du réseau à 860 sommets et un positionnement 3D ont été obtenus par le logiciel Pajek en quelques minutes, la publication est assurée par l'applet Javaview. Ces deux dernières phases étant très rapides elles auraient pu donner lieu à de nombreuses autres représentations possibles de la blogosphère. La visualisation dynamique des données et l'interprétation ne peut être que personnelle, dépend des algorithmes utilisés et du regard et des attentes qu'a l'utilisateur de ces représentations. Dans le graphe proposé, par exemple il apparaît clairement la redondance des sites cités par certains blogs (2 paires de blogs citent les mêmes pages). Cela pourrait être interprété, par exemple, soit comme une plus forte consistance des informations multi citée, soit un manque de spécificité de certains Blogs du domaine, mais de nombreuses autres lectures et interprétations sont possibles. Cela permet de montrer aussi que la multiplication des blogs sur un même sujet (sans recherche de spécificité et d'excellence) n'est pas forcément utile pour l'utilisateur, mais au contraire peut provoquer des pertes de temps inutiles.

## J. Développement ultérieurs

En utilisant une collecte différente il est possible d'obtenir rapidement d'autres représentations permettant d'autres analyses et interprétations aussi rapidement. Par exemple il est possible de réaliser des "SiteMaps" interactifs 3D permettant de donner une lecture d'un site basé sur le jeu de relations des pages du site. Par exemple il serait et sera possible de se livrer à une analyse de contenu mots clés en changeant juste la partie de collecte. Alors des graphes d'analyse de colloques, par exemple, seront possibles, comme celle que nous avons proposé des [mots clés](#) des communications du [3ème CONTECSI](#)

## K. Webographie/Bibliographie

Nous souhaitons présenter une webographie pour ce sujet (il s'y prête) collectée de façon automatique en utilisant le logiciel de collecte et présentation bibliographique [Zotero](#). Il nous a été impossible de le réaliser, la plupart des sites web n'ayant pas conscience de l'importance d'apparaître avec des formats standards permettant d'être « cité » en utilisant des logiciels de gestion de bibliographie automatisés, montrant encore le décalage entre ce que permet la technique et l'apprentissage. Le lecteur désireux de retrouver une

bibliographie "conventionnelle" pourra le faire très aisément en consultant les différents sites proposés dans cet article.

Bond, D. T. G., & Fox, C. M. *Applying the Rasch Model* (p. 340).

Bornholdt, D. S., & Schuster, H. G. *Handbook of Graphs and Networks* (p. 417).

Botbol, D. M. *France le top 100 de l'intelligence économique Intelligence online reports*.

Carrington, D. P. J., Scott, J., & Wasserman, S. *Models and Methods in Social Network Analysis* (p. 344).

Clemm, D. A., Granville, L. Z., & Stadler, R. *Managing Virtualization of Networks and Services* (p. 269).

Morse, D. J. M. *Critical Issues in Qualitative Research Methods* (p. 395).

Nooy, D. W. D., Mrvar, A., & Batagelj, V. *Exploratory Social Network Analysis with Pajek* (p. 334).

Quoniam, L., Pauluci, R., & Jorge, C. (2007). Networks of the economic intelligence in France : mapping and visualisation of knowledge. Dans (Vol. 11). Madrid (Espagne): Isabel Gomez. Retrouvé Juin 24, 2008, de <http://issi2007.cindoc.csic.es/>.

Sherman, D. C., & Price, G. *The Invisible Web* (p. 430).

Smith, D. P. R., & Chaffey, D. *Emarketing Excellence* (p. 360).

Srinagesh, D. K. *The Principles of Experimental Research* (p. 410).

## Remerciements

Nous tenons à remercier chaleureusement pour leur aide précieuse Michael Thelwall, responsable du laboratoire du [Statistical Cybermetrics Research Group](#) ainsi que Klaus Hildebrandt, équipe de développement de Javaview.