

CONCEPTION DE PROFILS VISUELS D'UTILISATEURS A PARTIR DE RESEAUX EGOCENTRIQUES (CAS DE FACEBOOK)

Dieudonné TCHUENTE (*), Nadine BAPTISTE-JESSEL (*), Marie-Françoise CANUT (*)
tchuente@irit.fr, baptiste@irit.fr, canut@iut-blagnac.fr

(*) [Institut de Recherche en Informatique de Toulouse](#),
Systèmes d'Information Généralisés/Documents et Données Semi Structurées
118, Route de Narbonne, 31062, Toulouse, France.

Mots clefs :

Réseaux sociaux, réseaux égocentriques, profil utilisateur, réseaux sémantiques, fouille de texte, Facebook.

Keywords:

Social networks, egocentric networks, user profiles, semantic networks, text-mining, Facebook.

Résumé

En 2010, plus de 72% des internautes dans le monde (77% des internautes en France) utilisent les réseaux sociaux numériques. Ces environnements sont devenus de véritables « systèmes d'exploitation » proposant une infinité d'interfaces applicatives aux utilisateurs et aux développeurs. Des masses de données sur la structure des réseaux et surtout sur les activités des utilisateurs dans ces environnements du Web 2.0 deviennent de plus en plus disponibles. Analyser des données mixtes (structures et activités) issues de ces environnements nécessite des outils proposant des fonctionnalités allant au-delà de simples analyses structurales des réseaux sociaux. Dans cet article, nous nous intéressons à deux principales problématiques : celle de l'accès à l'information dans les réseaux sociaux numériques et celle de l'enrichissement des analyses structurales des réseaux sociaux par l'analyse des données issues des interfaces d'usage. Un cas pratique est présenté pour la plate-forme Facebook avec des analyses statiques et évolutives de réseaux égocentriques (centrés sur l'utilisateur) basées sur la fouille de texte, à l'aide de la plate-forme de veille technologique Tétralogie. Les résultats observés démontrent le potentiel de la plate-forme Tétralogie pour l'analyse de l'influence du réseau social sur le profil d'un utilisateur ou pour la visualisation de l'évolution des centres d'intérêts de l'utilisateur dans le temps (profils à court-terme et profils à long-terme).

1 Introduction

L'analyse des réseaux sociaux est menée dans le domaine des sciences sociales depuis les années 1930 au moyen d'éléments de la théorie des graphes ou de la sociométrie [13]. Les données collectées permettaient surtout de construire la structure des réseaux dans des outils d'analyse dédiés (Ucinet, Vizter, Stocnet, Pajek etc.) [18, 19]. Les métriques extraites à partir de la structure des réseaux sociaux (degré, centralité, proximité, etc.) sont les principaux indicateurs permettant de caractériser un individu au sein du réseau (nœud central, nœud intermédiaire, etc.) [15].

Avec l'avènement du Web 2.0, les plates-formes de réseaux sociaux numériques (RSN) (Facebook, MySpace, Skyrock, etc.) fournissent en plus de la structure des réseaux, beaucoup d'informations propres aux individus et à leurs interactions au moyen d'applications multiples et variées (photos, vidéos, tags, blogs, murs, liens, etc.). Plusieurs techniques ont également été développées pour extraire des informations sur les usages des internautes dans ces environnements. Ainsi, dans les RSN, il est possible d'enrichir les graphes structuraux relatifs à l'analyse des réseaux sociaux avec les centres d'intérêts des utilisateurs. Rajouter les centres d'intérêts des utilisateurs aux analyses structurales existantes pourrait faciliter ou améliorer plusieurs tâches telles que la détection de communautés basée sur des centres d'intérêts, la détection de réseaux d'influences, le contrôle de l'e-réputation ou de l'identité numérique, la prise en compte du réseau social d'un utilisateur dans la conception de son profil, etc. Dans cet article nous nous intéressons (au moyen d'un cas pratique) à la visualisation du réseau social des utilisateurs (dans les RSN) en prenant en compte la construction de leurs centres d'intérêts par la fouille de texte. Deux principaux enjeux sont particulièrement abordés : la sécurité d'accès aux données dans les RSN et les dispositifs d'analyse permettant de concilier la conception de profils d'utilisateurs et la visualisation de réseaux sociaux.

La suite de cet article se présente comme suit. Dans la section 2, nous faisons une brève revue de littérature sur les méthodes d'accès aux données dans les RSN et sur l'analyse de réseaux sociaux. Dans la section 3 nous présentons notre méthodologie et quelques résultats d'une expérimentation menée conjointement sur Facebook et dans l'outil de veille technologique Tétralogie. La section 4 présente la conclusion et les perspectives de nos travaux.

2 Etat de l'art

Les nouvelles interfaces de communication qu'apportent les RSN pour créer des liens et les interactions sociales entre les individus poussent les sociologues à mettre en œuvre de nouveaux dispositifs propres à l'analyse des RSN [1, 2]. Du point de vue purement informatique, le tout premier challenge pour l'analyse de RSN consiste au recueil de l'information à analyser dans le respect de la vie privée des internautes. Certains auteurs utilisent les données représentées sous des vocabulaires RDF tels que FOAF ou SIOC par les plates-formes Web 2.0 [3, 4]. Cependant, ces vocabulaires sont implémentés majoritairement par des plates-formes de type blog (*LiveJournal* par exemple) et les informations qui y sont réellement représentées concernent surtout la structure (relations entre individus) du réseau d'utilisateurs. D'autres auteurs s'intéressent plutôt à l'extraction des relations entre utilisateurs contenues dans le code source HTML de leur profil public (profil accessible même aux internautes non inscrits sur le RSN) [5]. Si cette méthode d'extraction fonctionne sur certains sites de RSN dont les profils publics des utilisateurs contiennent assez d'information (MySpace par exemple), la plupart des sites de RSN donne l'accès à très peu d'information dans le profil public de leurs utilisateurs (Facebook, FriendSter, etc.). Une autre méthode d'accès aux données utilisateurs dans les RSN consiste à utiliser les API (Facebook API, OpenSocial de Google, etc.) fournies par les sites de RSN [6] aux parties tierces, transformant ainsi les RSN en véritables systèmes d'exploitation pouvant héberger tous types d'applications Web. Bien que ces API puissent poser plusieurs problèmes de sécurité [7, 8, 9] (exécution de scripts malicieux chez les clients, reconstruction des réseaux complets à partir des fragments de réseaux, collecte massive de données à des fins douteuses, etc.), elles sont aussi utilisées pour renforcer la sécurisation des données utilisateurs [10, 11]. Pour la collecte de données, les fonctionnalités de ces API offrent le moyen d'accéder au réseau égocentrique (centré sur l'utilisateur, par exemple l'utilisateur et ses amis) ainsi qu'à plusieurs données sur les activités dans ce réseau [12].

De manière générale, l'analyse des réseaux sociaux (une fois les données collectées) s'appuie sur des éléments de la sociométrie et de la théorie des graphes [13]. Les algorithmes de génération de graphes dans les réseaux sociaux s'intéressent à la caractérisation de la structure de ces réseaux, des rôles des individus dans ces réseaux (centralité, intermédiarité, etc.), des sous-réseaux (communautés extraites à partir des cliques par exemple), des activités dans le réseau (nombre de connexions, nombre de messages postés, etc.) [14, 15, 16]. Toutes ces caractéristiques de l'analyse des réseaux sociaux peuvent être conceptualisées au moyen d'ontologies [17] et sont intégrés dans la majorité des outils d'analyse [18, 19]. Les données des réseaux entiers n'étant pas très souvent accessibles, certains auteurs se sont plutôt penchés sur le calcul des mesures de centralités définies dans l'analyse structurale des réseaux sociaux (l'intermédiarité par exemple) dans le contexte particulier des réseaux égocentriques [27]. Les résultats expérimentaux de ces travaux montrent une forte corrélation entre les mesures obtenues pour un même utilisateur dans le réseau social entier d'une part, et dans un réseau égocentrique d'autre part. Toutefois aucune démonstration théorique ne confirme ce résultat pour le moment. Dans les RSN, des applications sont développées pour permettre aux internautes une meilleure vue sur leur réseau égocentrique (cas de l'application FriendWheel [26] de Facebook). Le graphe sur la figure 1 présente un exemple de réseau égocentrique (sociogramme) d'un profil Facebook dans lequel on peut clairement identifier plusieurs clusters (famille, profession, études, etc.) [6].

Dans cet article, nous envisageons d'enrichir sémantiquement ce type de sociogramme en y rajoutant les centres d'intérêts des utilisateurs. On obtient alors des profils d'utilisateurs sous forme de graphes sémantiques [20], mais avec l'avantage de pouvoir expliquer les centres d'intérêts de l'utilisateur à partir de son réseau social. Cette étude permettrait alors de pouvoir dériver des réseaux d'influence à partir des profils d'utilisateurs. De plus, nous représentons l'évolution de ces réseaux d'influence dans le temps (pour y détecter les profils à court-terme et les profils à long-terme). Notre méthodologie et les résultats de notre expérimentation sont présentés dans les sections qui suivent.

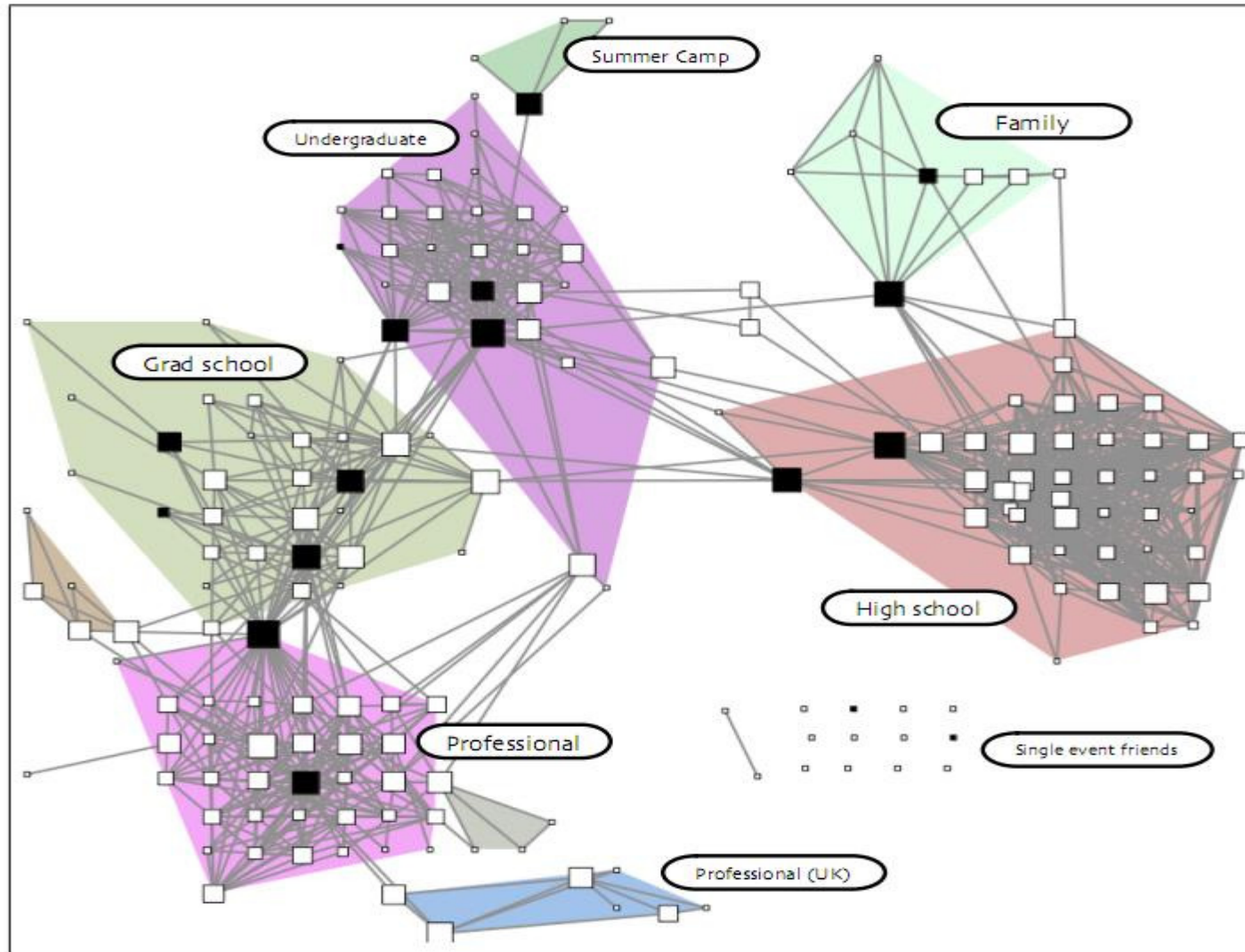


Figure 1 : Réseau égo-centrique (sociogramme d'un utilisateur Facebook [6]).

3 Méthodologie de construction des profils utilisateurs

Notre méthodologie se décompose en 3 grandes étapes : extraction de données (dans Facebook ici), prétraitement et traitement de données textuelles (pour construire les centres d'intérêts et la structure du réseau égocentrique de chaque utilisateur), visualisation des réseaux construits (graphes bipartis 2D ou 3D) (figure 2). Chacune de ces étapes est décrite dans les sections qui suivent.

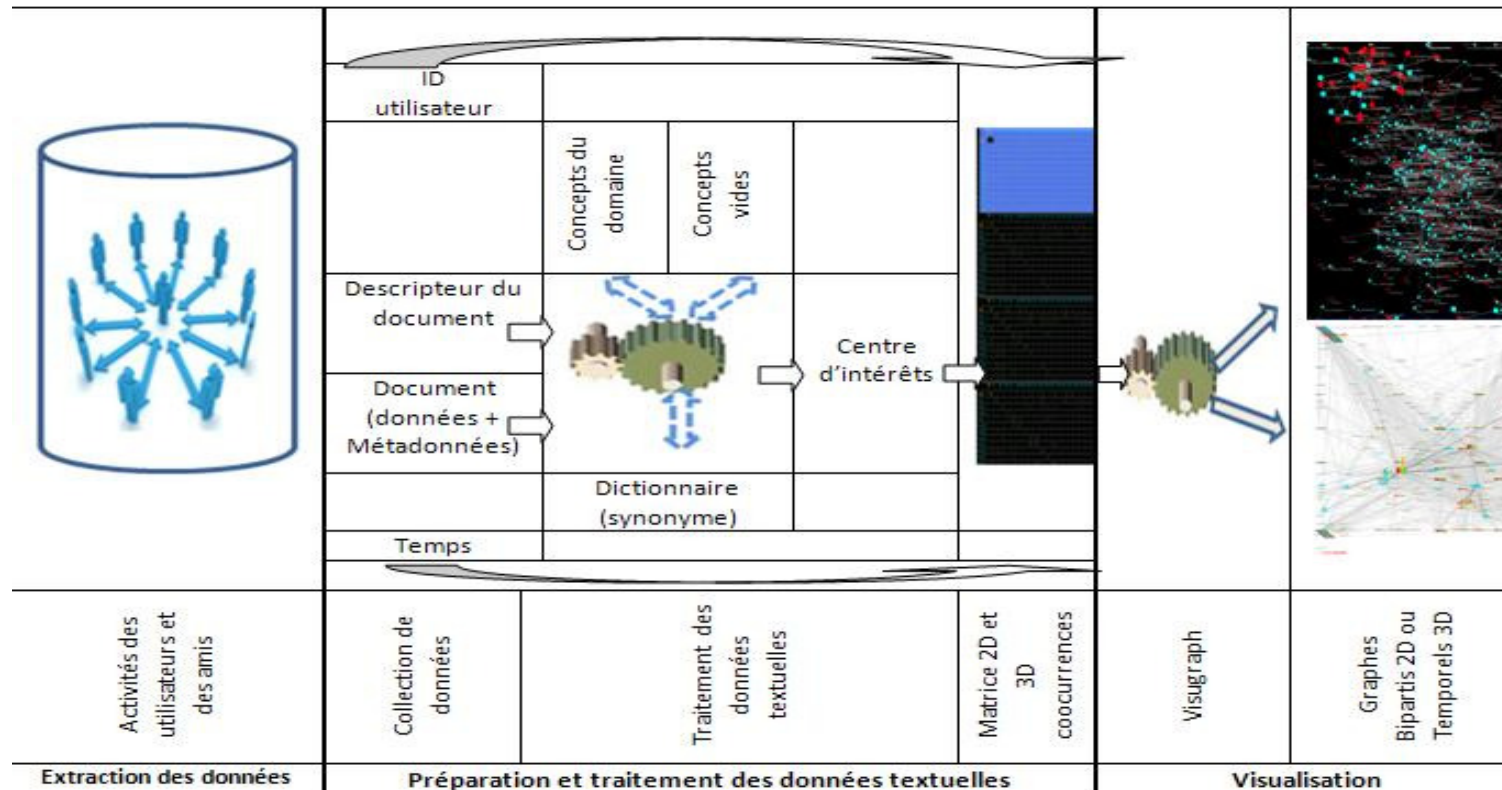


Figure 2 : Méthodologie d'extraction, d'analyse et de visualisation des centres d'intérêts et des réseaux égocentriques d'utilisateurs.

3.1 Extraction de données

Comme indiqué dans l'état de l'art, les API de RSN permettent d'accéder à des informations supplémentaires sur les profils et les activités des utilisateurs. Nous nous intéressons dans notre expérimentation au site de RSN Facebook qui est, d'une part, le plus utilisé dans le monde (et en France) et d'autre part, le site hébergeant le plus d'applications développées par des tiers au moyen de son API [22].

Nous avons développé une application visant uniquement à extraire les données sur les profils et les activités des utilisateurs (<http://apps.facebook.com/analyze-network>). Nous demandons explicitement aux utilisateurs de cette application de nous donner le droit de stocker et d'analyser les données disponibles sur leurs profils et leurs activités au moyen de l'API Facebook. Cette application étant uniquement destinée à des fins de recherche, elle n'a pas été disponible au grand public et seuls les utilisateurs qui se sont portés volontaires ont pu y accéder (85 utilisateurs de l'application développée, 7081 profils (utilisateurs et amis) accédés au total). Avec l'API Facebook, une application développée par un tiers peut accéder aux informations chez chacun de ses utilisateurs ainsi que chez tous leurs amis [12]. Chaque utilisateur peut :

- Limiter ses informations (profil et activités) accessibles par une application tierce. Cependant très peu d'utilisateurs restreignent l'accès à ces informations, et sont même pour la plupart inconscients des informations accessibles sur leur profils et leurs activités [9, 23].
- Donner des autorisations supplémentaires à une application tierce. Par exemple l'utilisateur peut permettre à l'application d'écrire des informations sur son mur, de mettre à jour ses statuts, de créer et envoyer des e-mails ou des sms au nom de l'utilisateur, d'accéder aux informations de l'utilisateur même lorsqu'il n'est pas connecté, de lire les flux d'activités disponibles sur le profil de l'utilisateur, etc.

Pour construire les centres d'intérêt des utilisateurs, nous nous intéressons à leurs activités (informations qu'ils publient et applications qu'ils utilisent dans le RSN). Ainsi, pour enrichir les analyses du réseau égocentrique d'un utilisateur, nous collectons les informations sur ses activités ainsi que celles de ses amis. Pour accéder aux informations sur les activités des amis de l'utilisateur, nous lui demandons une autorisation supplémentaire pour l'accès à ses flux d'activités (Figure 3). Les flux d'activités disponibles sur le profil (mur) d'un utilisateur rapportent un bon nombre d'activités de ses amis.

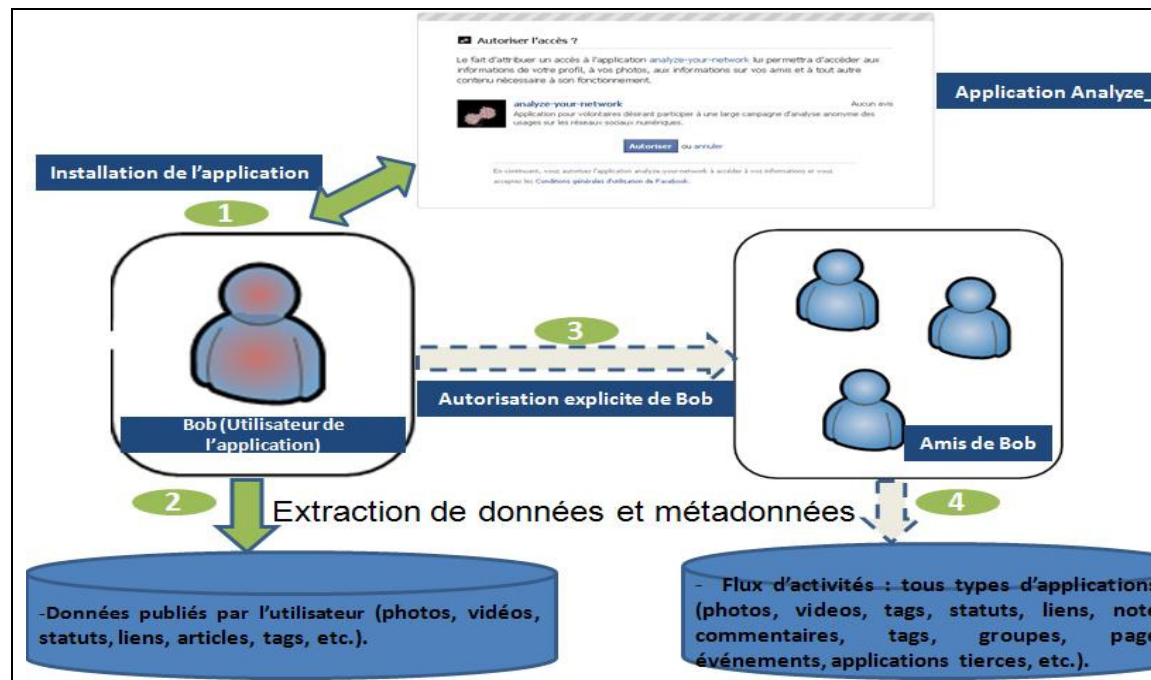


Figure 3 : schéma d'exécution de l'application d'extraction de données (utilisateurs et amis)

Pour chacune des activités de l'utilisateur, l'application extrait les données (contenu d'un article publié par exemple) et les métadonnées associées (titre ou date de publication de l'article exemple).

3.2 Prétraitement des données textuelles

Les données ou métadonnées extraites sur les activités des utilisateurs sont essentiellement textuelles. Elles sont prétraitées et analysées (fouille de textes) dans la plate-forme de veille technologique Tétralogie [21] qui offre également les possibilités de visualisation de graphes sémantiques. Les données extraites sont représentées sous forme d'un document textuel comprenant les informations sur chaque activité ou flux d'activités associés au profil de l'utilisateur (figure 4A). Les identifiants d'utilisateurs sont cryptés avec la méthode MD5. Les entités textuelles sont extraites à partir d'un descripteur de document indiquant les séparateurs de texte (blancs, signes de ponctuations, retour à la ligne, etc.).

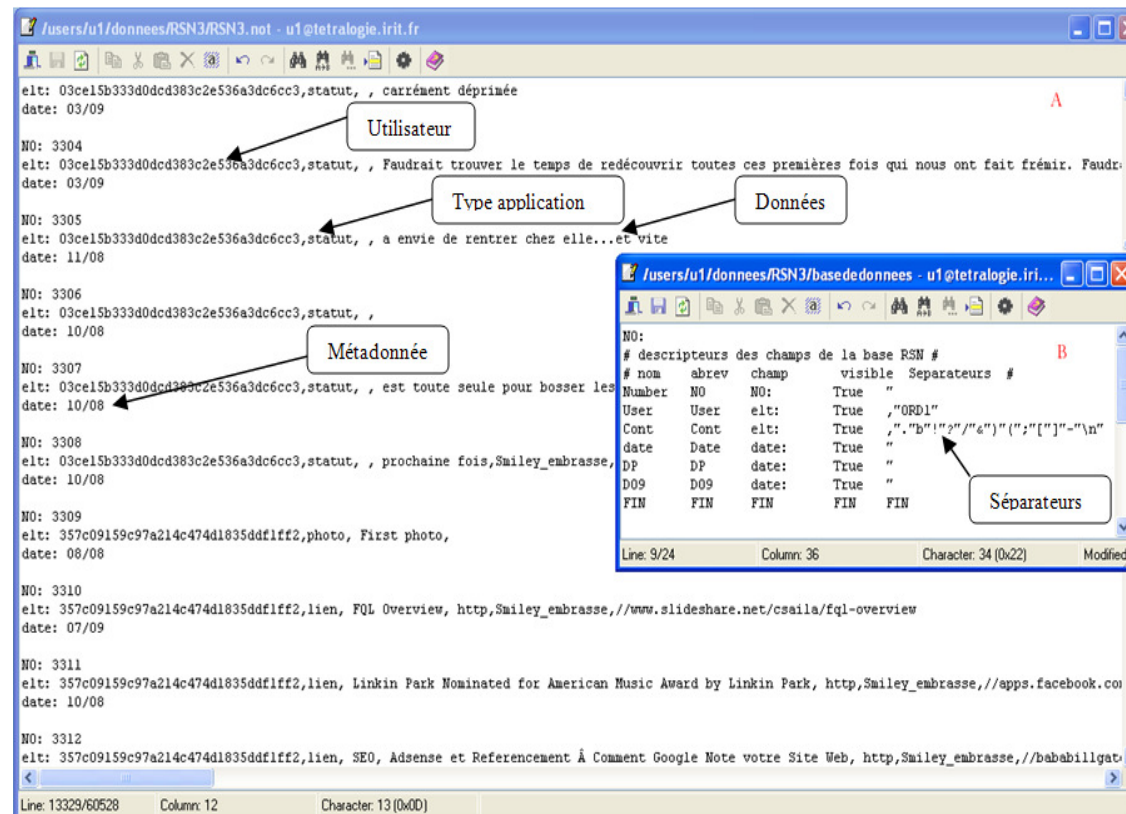


Figure 4. (A) Extrait du document contenant les données extraites. (B) Descripteur du document.

L'extraction des centres d'intérêts se fait ensuite par comptage de cooccurrences de termes dans le document. Pour extraire uniquement les centres d'intérêts pertinents, nous utilisons des filtres et des dictionnaires :

- Les filtres positifs (concepts du domaine) : il s'agit d'une liste de termes du domaine à extraire exclusivement dans le texte. Pour l'analyse d'un domaine particulier, les termes peuvent être fournis a priori par les analystes (experts) de ce domaine ou par une projection sur une ontologie du domaine étudié. Dans notre cas, nous souhaitons étudier le vocabulaire employé par les utilisateurs de Facebook dans sa généralité et non dans un domaine précis. Ainsi, nous avons construit un filtre non spécifique à un domaine précis, et contenant uniquement les termes ayant une fréquence supérieure à 1 dans l'ensemble du document (pour simplifier l'étude). La figure 5A montre un extrait de filtre positif contenant les termes retenus dans nos analyses.
- Les filtres négatifs (concepts vides) : il s'agit d'une liste de termes vides de sens dans le contexte de l'étude. Les articles du (des) langage(s) étudié(s) sont des cas simples de termes très souvent inclus dans les filtres négatifs. En fonction des langages, plusieurs types de filtres sont fournis par défaut et peuvent être réutilisés ou enrichis. La figure 5B présente un extrait de filtre négatif utilisé dans notre cas.
- Les dictionnaires de synonymes : il s'agit de dictionnaires permettant de mettre en relation les termes de même sens dans le contexte étudié. Si de tels dictionnaires peuvent facilement être constitués dans un domaine bien précis, la tâche s'avère plus compliquée dans les documents à contenu très varié comme dans les RSN. Nous avons constitué un fichier de synonymie prenant en compte les termes dans les différentes langues étudiées (français et anglais notamment) ainsi que les abréviations et les caractères spéciaux (les smileys par exemple) (figure 5C).

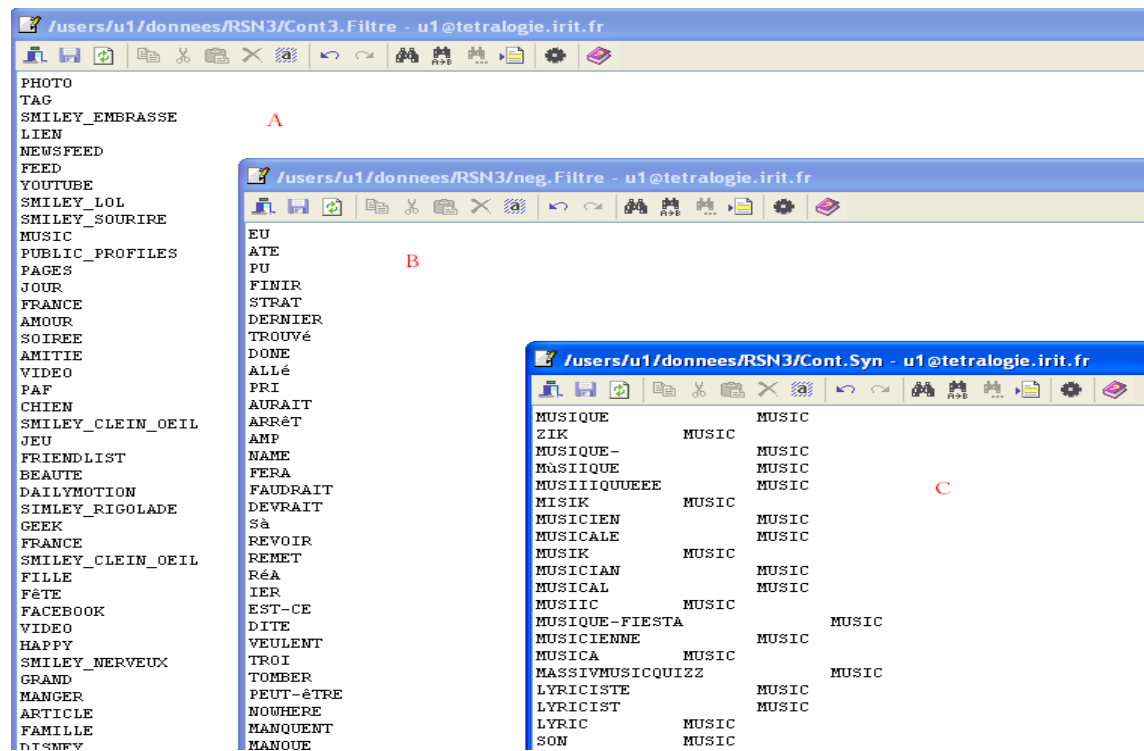


Figure 5. (A) extrait d'un filtre positif, (B) extrait d'un filtre négatif, (C) extrait d'un dictionnaire de synonymes.

Une fois les centres d'intérêts construits, ils sont mis en relation avec les utilisateurs par des calculs de cooccurrences afin de produire des matrices 2D ou 3D (lorsque le temps est également pris en compte) contenant les poids des liens entre utilisateurs et centres d'intérêt (voir section ci-dessous).

3.3 Conception des profils

Nous construisons les profils sémantiques d'utilisateurs au travers de matrices 2D et de matrices 3D afin de réaliser des visualisations statiques et des visualisations évolutives (dynamiques).

Les profils statiques sont construits sous forme de graphe biparti $G = (U, I, L)$ dans lequel :

- U représente l'ensemble des utilisateurs (un utilisateur et ses amis dans notre cas).
- I représente l'ensemble des centres d'intérêt des utilisateurs.
- L représente l'ensemble des arêtes (liens) entre un utilisateur (élément de U) et un centre d'intérêt (élément de I).

Il est également possible d'inclure dans L les liens entre centres d'intérêts (éléments de I) afin de construire des profils sémantiques d'utilisateurs [20]. Pour des raisons de clarté, nous ne considérons pas ces liens ici.

Les poids d'un centre d'intérêt (P_i) et le poids d'un lien entre un utilisateur et un centre d'intérêt ($P_{i,u}$) sont définis comme suit :

$$\left\{ \begin{array}{l} \mathbf{P}_i = \mathbf{Occ}(i, \mathcal{D}) \\ \text{Nombre d'occurrences (Occ) du centre d'intérêt dans le document } (\mathcal{D}) \text{ (figure 4A). Sur le graphe, plus le poids du centre d'intérêt est élevé, plus le nœud associé apparaît grand (figure 6B).} \\ \\ \mathbf{P}_{i,u} = \mathbf{Cooc}(i, u, \mathcal{D}) \\ \text{Nombre de cooccurrences (Cooc) entre } i \text{ et } u \text{ dans le document } (\mathcal{D}) \text{ (figure 4A). Sur le graphe, plus le poids du lien est élevé, plus le lien entre } i \text{ et } u \text{ apparaît fort (figure 6B).} \end{array} \right.$$

Les profils dynamiques sont également construits sous forme de graphe biparti et temporel $G = (U, I, L, T)$ dans lequel :

- T représente les différentes périodes de temps (t_1, t_2, \dots, t_n). Les périodes considérées dépendent du niveau de granularité souhaité dans les analyses (jour, semaine, mois, trimestre, etc.)
- U, I, et L sont tels que définis pour les graphes statiques plus haut.

Cependant, la représentation d'un utilisateur ou d'un centre d'intérêt fait ressortir son poids pour chacune des périodes de temps (figure 7). Globalement le poids d'un centre d'intérêt (P_i) et le lien entre un utilisateur et un centre d'intérêt ($P_{i,u}$) sont définis relativement à chaque période de temps, mais restent agrégés dans un même nœud (histogramme) :

$$\left\{ \begin{array}{l} \mathbf{P}_i = \sum_{i=1,n} [\mathbf{Occ}(I, \mathcal{D}, t_i)] \\ \text{Somme du nombre d'occurrences (Occ) du centre d'intérêt dans le document } (\mathcal{D}) \text{ (figure 4A) pour chaque période de temps. Sur le graphe, plus le poids du centre d'intérêt est élevé pour une période de temps, plus la barre de l'historgramme correspondant à cette période de temps est élevée, et le nœud se rapproche de cette période de temps (figure 7).} \\ \\ \mathbf{P}_{i,u} = \sum_{i=1,n} [\mathbf{Cooc}(U, I, \mathcal{D}, t_i)] \\ \text{Somme du nombre de cooccurrences (Cooc) entre } i \text{ et } u \text{ dans le document } (\mathcal{D}) \text{ (figure 4A) pour chaque période de temps. Sur le graphe, plus le poids du lien est élevé, plus le lien entre } i \text{ et } u \text{ apparaît fort (figure 7).} \end{array} \right.$$

4 Visualisation des profils

Une fois que les centres d'intérêts des utilisateurs sont extraits par fouille de textes (figure 2), les réseaux égocentriques des utilisateurs peuvent être visualisés à partir des matrices 2D (réseaux statiques) et 3D construites (réseaux dynamiques) via l'algorithme de génération de graphes *Visugraph* [24] dans Tétralogie.

Nous présentons sur la figure 6 (A) les différents clusters du réseau égocentrique d'un utilisateur. On peut ainsi, voir des *clusters* (segments d'amis) isolés (*clusters* 2, 4, 12, 15, 16) qui représentent des cercles fermés d'amis de cet utilisateur (amis rencontrés lors d'une conférence par exemple), ou des *clusters* centraux (*clusters* 1, 3, 9,10, 9) qui jouent le rôle de ponts entre plusieurs cercles d'amis de l'utilisateur.

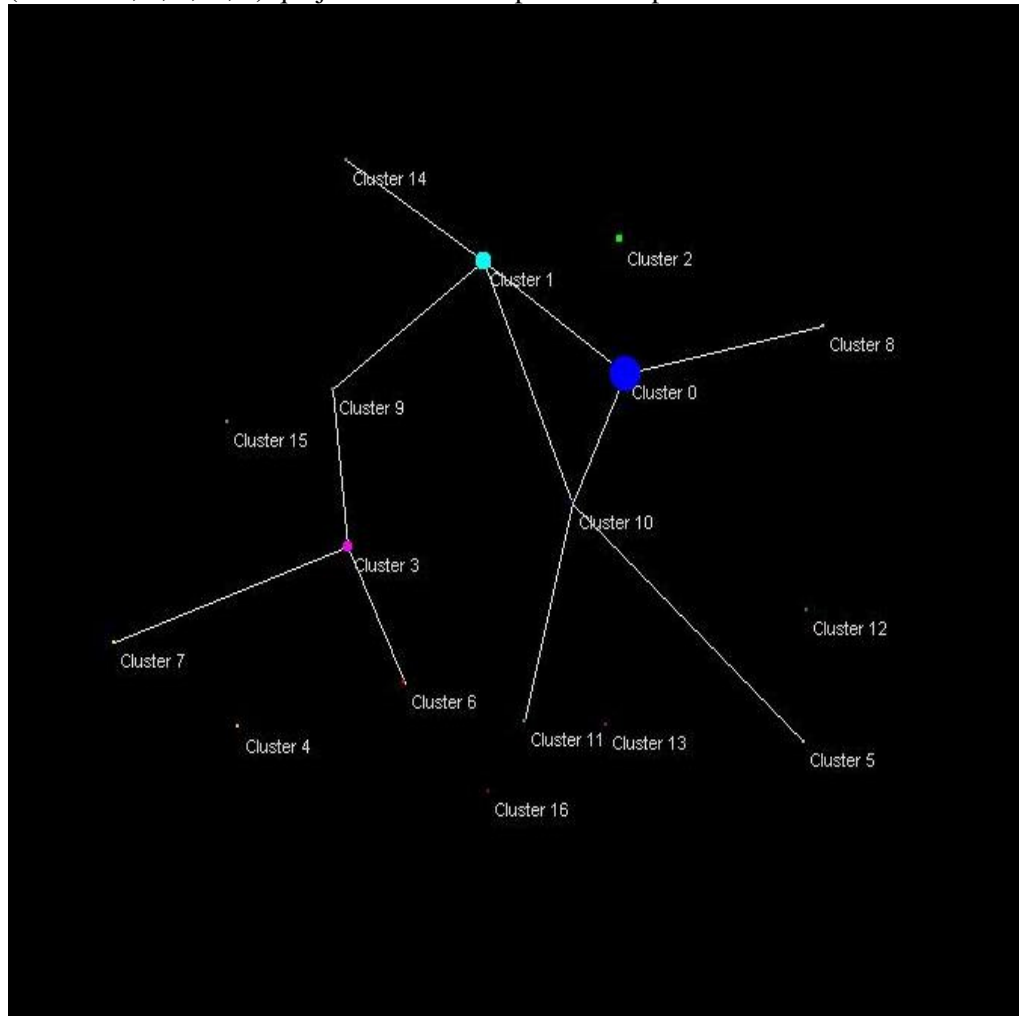


Figure 6 (A) : Clusters d'un réseau égocentrique

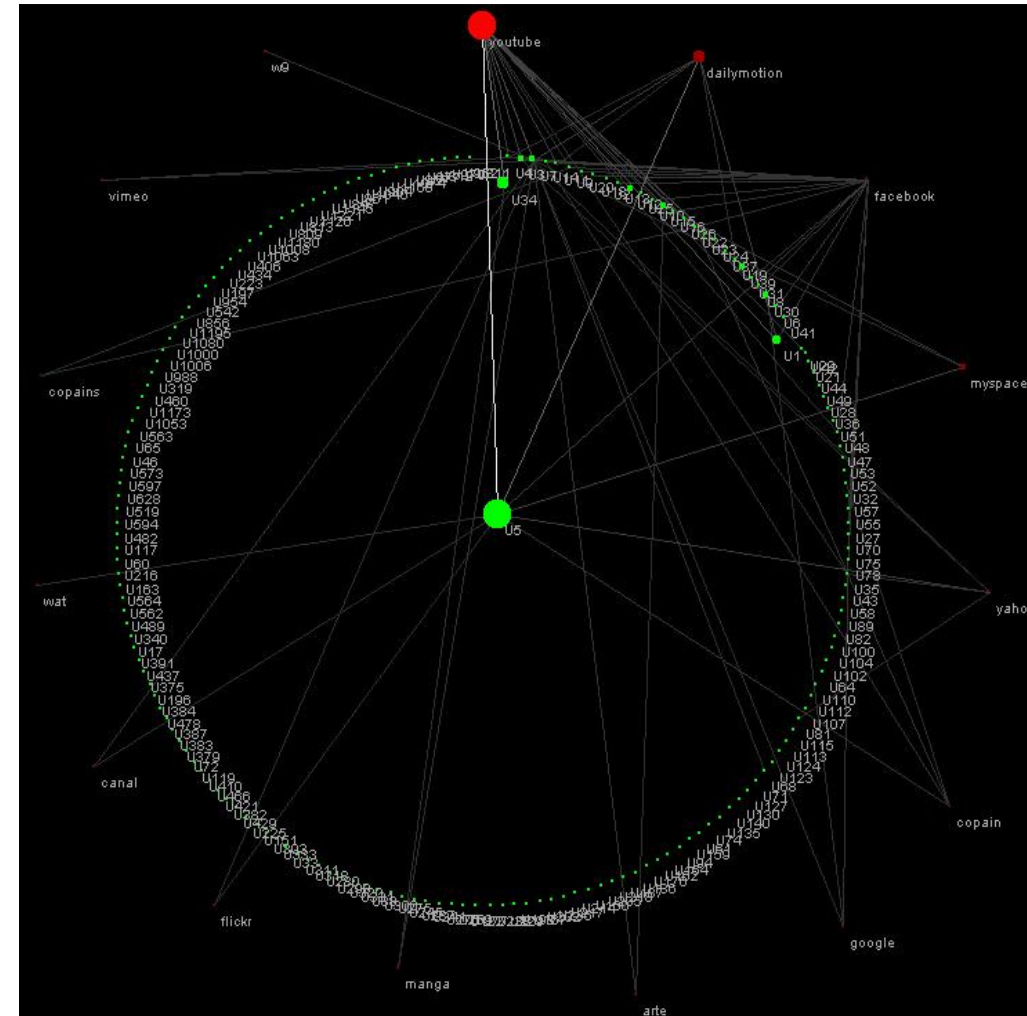


Figure 6 (B) Réseau égocentrique enrichi par les centres d'intérêts liés aux médias

Les analyses approfondies des attributs (ou des profils) des utilisateurs dans chaque cluster peuvent nous donner plus d'informations quant à la sémantique de chaque cluster par rapport à l'utilisateur (famille, profession, études, etc.) telle que celles étudiées dans [6]. Ici, nous nous intéressons plutôt à la visualisation du profil sémantique de l'utilisateur, dans lequel on peut identifier l'influence des utilisateurs de son réseau égocentrique. Ainsi la figure 6(B) présente un exemple de graphe biparti (utilisateurs - centres d'intérêt) correspondant à la méthodologie de conception définie dans la section 3. Les centres d'intérêt retenus ici ont été construits avec un filtre positif incluant les termes représentant des médias (chaînes de télévisions, médias sociaux, etc.).

Les nœuds en vert représentent les utilisateurs, l'utilisateur étudié ici est le nœud central U5 et les autres nœuds (en vert) représentent ses amis dans Facebook. Les nœuds en rouge représentent les centres d'intérêt de U5 et de ses amis construits par fouille de textes (cf section 3) à partir de leurs activités. On observe que U5 s'intéresse particulièrement aux sites de partage de vidéos (Youtube et Dailymotion notamment). Bien que l'utilisateur U5 ait plus de 1000 amis, seuls ses amis U1 et U34 s'intéressent aux mêmes médias sociaux que lui. Il pourrait ainsi exister une relation d'influence sur U5 (par rapport aux usages de médias sociaux) de la part des utilisateurs U1 et U34 ou inversement. Des sous-graphes restreints uniquement à des utilisateurs précis peuvent être extraits pour mieux analyser les centres d'intérêt communs entre plusieurs utilisateurs dans le graphe construit. Pour visualiser l'évolution des centres d'intérêts des utilisateurs dans le temps, il est possible de générer des graphes évolutifs.

Les réseaux évolutifs des utilisateurs sont construits selon la même approche que les réseaux statiques, à la différence que le poids des centres d'intérêts et des cooccurrences entre utilisateurs et centres d'intérêt est calculé pour plusieurs périodes de temps (jours, semaines, mois, trimestres, années, etc.) dans des matrices 3D. Chaque centre d'intérêt est maintenant représenté par un histogramme dont chaque barre représente le poids du centre d'intérêt sur une période de temps [24]. Chaque utilisateur est également représenté par un histogramme dont chaque barre représente la fréquence des activités de l'utilisateur pour une période de temps. Le réseau égocentrique et évolutif d'un utilisateur (Up) est présenté sur la figure 7.

En fonction de la fréquence des activités, nous avons regroupé le temps en quatre semestres, du semestre 2 de l'année 2008 jusqu'au semestre 1 de l'année 2010. La succession des barres pour chaque nœud se fait dans le sens des aiguilles d'une montre. Par exemple, pour l'utilisateur Up, la barre rouge représente sa fréquence d'activités au semestre 1 de l'année 2009, la barre orange représente sa fréquence d'activités au semestre 1 de l'année 2010, la barre jaune représente sa fréquence d'activités au semestre 2 de l'année 2008, et la barre verte représente sa fréquence d'activités au semestre 2 de l'année 2009.

Pour observer plus de centres d'intérêt dans ce graphe, aucun filtre positif (termes d'un domaine précis) n'a été utilisé. Tous les termes ayant des fréquences supérieures à 1 ont été considérés comme des centres d'intérêt pertinents. Les centres d'intérêts (respectivement les utilisateurs) dont les poids (respectivement les fréquences d'activités) sont quasi constants durant toutes périodes se rapprochent du centre du graphe et représentent les centres d'intérêts à long-terme de l'utilisateur (respectivement les amis de l'utilisateur qui sont actifs à long-terme). Sur le graphe (cadre en vert, figure 7), on peut ainsi identifier les centres d'intérêt à long-terme (photos, article, champagne, vidéos, musique, jeu, etc.) où seul l'utilisateur (U29) est constamment actif (voire influent) dans le réseau égocentrique de Up sur toutes les périodes de temps étudiées.

Les centres d'intérêts (respectivement les utilisateurs) dont le poids sur une seule période (respectivement la fréquence d'activités sur une seule période) est particulièrement très élevé relativement aux poids sur les autres périodes (respectivement aux fréquences d'activités sur les autres périodes), se rapprochent vers la période considérée et représentent les centres d'intérêt à court-terme de l'utilisateur (respectivement les amis de l'utilisateur qui sont actifs ou qui l'influencent uniquement sur la période considérée). Sur le graphe (cadres en rouge, figure 7), on peut ainsi identifier les centres d'intérêts (ou utilisateurs actifs) à court-terme du semestre 1 de 2009 (films, Renault, sociologie, U57), du semestre 1 de 2010 (U3, U14, U32, etc.), du semestre 2 de 2008 (Michael Vendetta, U37, etc.), du semestre 2 de 2009 (le site web *lemonde*).

A la différence des méthodes de calcul de valeurs pour les centres d'intérêt et les périodes associées aux profils à long-terme et aux profils à court-terme [25], les réseaux évolutifs construits dans la plateforme Tétralogie permettent ainsi de détecter graphiquement ces mêmes informations. De plus, les individus actifs ou influents dans le réseau égocentrique de l'utilisateur peuvent également être détectés pour chacune des périodes définies via l'extraction et l'analyse approfondie de sous-graphes limités à un ces utilisateurs.

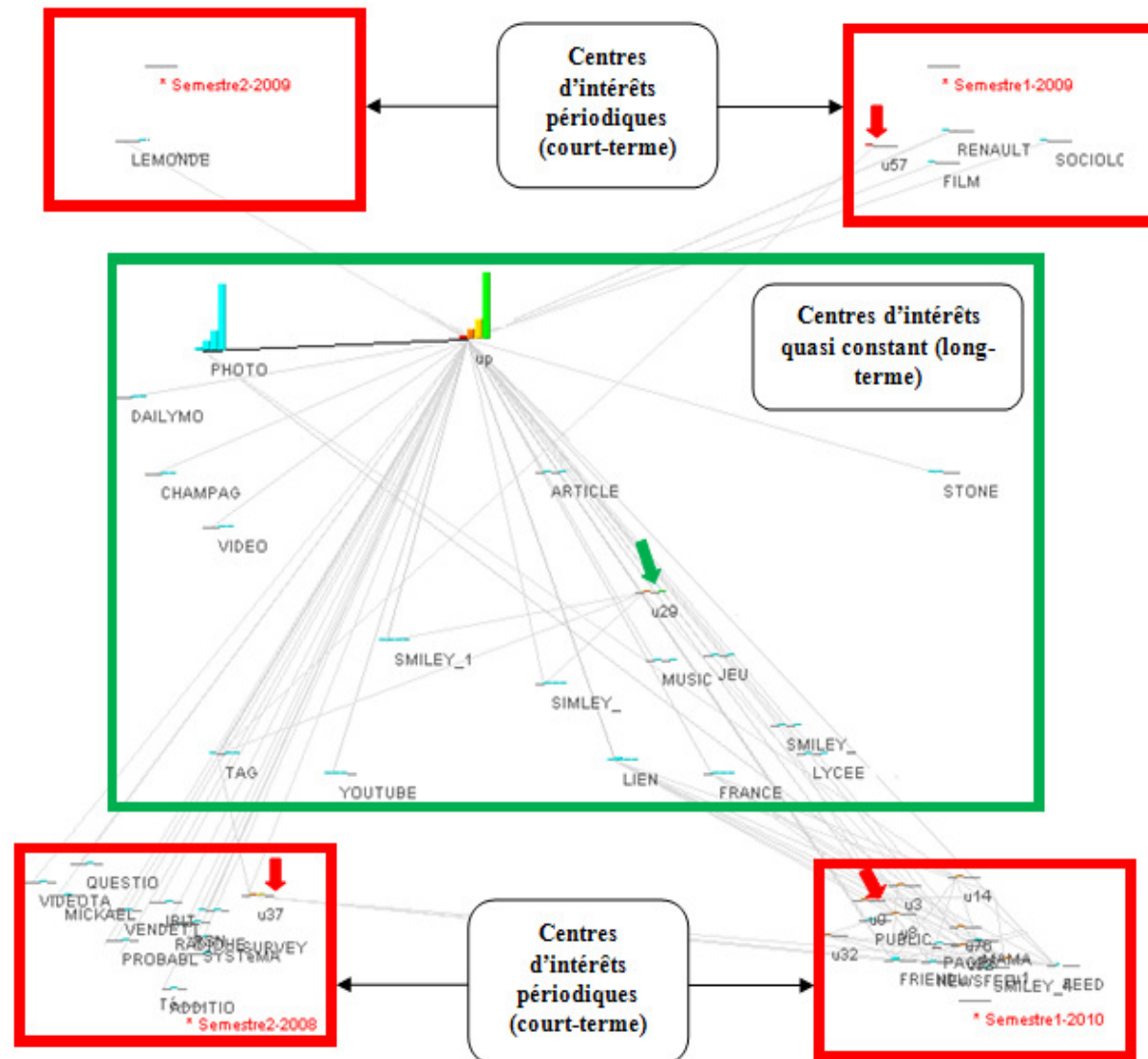


Figure 7. Profil évolutif d'un utilisateur (Up) sur quatre semestres.

5 Conclusion

Dans cet article, nous nous sommes intéressés à l'analyse des profils utilisateurs via l'enrichissement des réseaux égocentriques (avec les centres d'intérêts des utilisateurs) dans le contexte des RSN. Pour ce faire, nous nous sommes intéressés aux problématiques d'accès aux données utilisateurs dans les RSN ainsi qu'aux outils existants dédiés à l'analyse des réseaux sociaux en général. Nous avons mené une expérimentation sur la plate-forme de veille technologique Tétralogie qui a l'avantage d'offrir des outils permettant à la fois de construire les centres d'intérêt d'utilisateurs (par fouille de textes) et de visualiser les réseaux sociaux. L'analyse statique des profils construits permet de visualiser les centres d'intérêt d'un utilisateur, ainsi que des conjectures sur l'influence du réseau égocentrique de l'utilisateur sur chacun de ses centres d'intérêts. Toutefois, ces analyses ne démontrent qu'une piste de solution qui devra être enrichie (usage d'ontologies par exemple pour améliorer la qualité des centres d'intérêt par rapport à un domaine, intégration d'outil de génération de sous-graphes cohésifs, etc.) afin de mieux évaluer graphiquement l'influence du réseau égocentrique de l'utilisateur sur son profil. A notre connaissance, ce travail est le tout premier à s'intéresser à l'identification visuelle des profils à long-terme et à court-terme des utilisateurs. Ces profils sont réalisés par analyse temporelle des traces d'activités des utilisateurs et par la visualisation des graphes évolutifs à partir du prototype *Visugraph* [24] de la plateforme Tétralogie. Ainsi, cet outil est en mesure de simplifier énormément la tâche d'un analyste qui souhaiterait suivre l'évolution dans le temps, des comportements individuels d'une population (clientèle, visiteurs d'un site, employés, collaboration avec des partenaires, etc.) à un niveau de détail très élevé. Pour terminer, une perspective importante de ce travail est d'aller au-delà des réseaux égocentriques, et d'analyser les réseaux dans leur globalité pour la détection des segments d'utilisateurs basés sur des centres d'intérêt, ou sur le suivi de l'évolution globale d'un réseau dans le temps. Il serait également intéressant d'enrichir les algorithmes de la plate-forme Tétralogie afin de mieux interpréter les graphes construits via l'implémentation de métriques usuelles de l'analyse structurale des réseaux sociaux (mesures de centralité, de proximité, de groupes cohésifs, etc.).

6 Bibliographie

- [1] STENGER T., COUTANT A., "Social Network Sites vs Social Network Analysis: do they match? Definition and methodological issues", *Sunbelt XXIX, Annual Conference of the INSNA (International Network for Social Network Analysis)*, San Diego, California, USA, 10-15 March. 2009. BOURCIER-DESJARDINS R., MAYERE A., MUET F. et SALAÛN J.M., *Veille technologique : revue de la littérature et étude de terrain*, CERSI, 1990
- [2] Danah Boyd, Nicole B. Ellison, "Social Network Sites: Definition, History, and Scholarship", *Journal of Computer-Mediated Communication*, 13 (1), article 11, October 2007. <http://jcmc.indiana.edu/vol13/issue1/boyd.ellison.html>, Vu le 23/09/2008. DRUCKER P., *The new society of organizations*, Harvard Business Review, Sept-Oct 1992, p 95-104
- [3] Jennifer Golbeck, Matthew Rothstein, "Linking Social Networks on the Web with FOAF: A Semantic Web Case Study", Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence (2008). GRUNDSTEIN M. et ROSENTHAL-SABROUX C., *Système d'information et Capitalisation des connaissances : un enjeu nouveau pour la conception de système d'information*, Prédit, Octobre 1999
- [4] Uldis Bojars, Alexandre Passant, Richard Cyganiak, John Breslin, "Weaving SIOC into the Web of Linked Data", LDOW2008, April 22, 2008, Beijing, China. REDST suydtfshbdf.
- [5] Sophia Alim, Ruquya Abdul-Rahman, Daniel Neagu, Mick Ridley, "Data Retrieval from Online Social Network Profiles for Social Engineering Applications", *Internet Technology and Secured Transactions, ICITST 2009*, pp 1-5.
- [6] Bernie Hogan, "A comparison of on and offline networks through the Facebook API", Electronic paper available at <http://ssrn.com/abstract=1331029>.
- [7] Patsakis, C.; Asthenidis, A.; Chatzidimitriou, A.; « Social networks as an attack platform: Facebook case study", *Networks*, 2009. ICN '09. Eighth International Conference Page(s):245-247, 2009.
- [8] Tianjun Fu; Hsinchun Chen; "Analysis of cyberactivism: A case study of online free Tibet activities" *Intelligence and Security Informatics*, 2008. ISI 2008. IEEE International Conference Page(s):1 – 6, 2008.

- [9] Joseph Bonneau, Jonathan Anderson, George Danezis, “Prying data out of a Social Network” In 2009 Advances in Social Network Analysis and Mining, pp. 249- 254.
- [10] Baatarjav, E.-A.; Dantu, R.; Yan Tang; Cangussu, J.; “BBN-Based Privacy Management System for Facebook”, Intelligence and Security Informatics, 2009. ISI '09. IEEE International Conference Page(s):194 – 196, 2009.
- [11] Saikat Guha, Kevin Tang, and Paul Francis, “NOYB: Privacy in Online Social Networks”, WOSN'08, August 18, 2008, Seattle, Washington, USA.
- [12] Nadine Jessel, Dieudonné Tchuenta, Marie-Francoise Canut, André Peninou, “Quelle modélisation des profils utilisateurs des réseaux sociaux numériques?” In Colloque International EUTIC - Usages et Enjeux des TIC (EUTIC 2009), Bordeaux, 18/11/2009-20/11/2009.
- [13] John Breslin and Stefan Decker, “The Future of Social Networks on the Internet”, Published by the IEEE Computer Society, December 2007.
- [14] Scott, J.: Social Network Analysis, a handbook, 2nd edn. Sage, Thousand Oaks (2000).
- [15] Mohsen Jamali, Hassan Abolhassani, « Different Aspects of Social Network Analysis », Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI'06).
- [16] Angélique DIMITRACOPOULOU, Éric BRUILLARD, « Enrichir les interfaces de forums par la visualisation d'analyses automatiques des interactions et du contenu », Article de Recherche, Revue Sticef, Volume 13, 2006.
- [17] Guillaume Erétéo , Michel Buffa , Fabien Gandon , and Olivier Corby, “Analysis of a Real Online Social Network Using Semantic Web Frameworks”, ISWC 2009, LNCS 5823, pp. 180–195, 2009.
- [18] UCINET: Social Network Analysis Software. <http://analytictech.com/>
- [19] Jeffrey Heer, Danah Boyd, “Vizster: Visualizing Online Social Networks”, [Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on](#) 23-25 Oct. 2005 Minneapolis,MN, pp. 32 –39.
- [20] Susan Gauch, Mirco Speretta, Aravind Chandramouli, Alessandro Micarelli, “User profiles for Personalized Information Access” in The Adaptive Web , Vol. 4321 (2007), pp. 54-89.
- [21] Josiane Mothe, Claude Chrisment, Taoufiq Dkaki, Bernard Dousset, Said Karouach, “Combining mining and visualization tools to discover the geographic structure of a domain”, In Computers, Environment and Urban Systems, Elsevier, Numéro spécial Geographic Information Retrieval, Vol. Hors-série N. 4, p. 460-484, juillet 2006.
- [22] Facebook statistics 2009, <http://www.facebook.com/press/info.php?statistics>
- [23] Catherine Dwyer, Starr Roxanne Hiltz, and George Widmeyer Pace University, New Jersey Institute of Technology, “Understanding Development and Usage of Social Networking Sites: The Social Software Performance Model”, Proceedings of the 41st Hawaii International Conference on System Sciences – 2008.
- [24] Brigitte Gay, Eloïse Loubier, “Dynamics and Evolution Patterns of Business Networks”, In 2009 Advances in Social Network Analysis and Mining, pp. 290-295.
- [25] Lin Li, Zhenglu Yang, Botao Wang, Masaru Kitsuregawa, “Dynamic Adaptation Strategies for Long-Term and Short-Term User Profile to personalized search,” In APWeb/WAIM, Vol. 4505 (2007), pp. 228-240.
- [26] Friend Wheel application on Facebook, <http://apps.facebook.com/friendwheel/>
- [27] Martin Everett, Stephen P. Borgatti, “Ego network betweenness”, In Social Networks, Elsevier, Vol. 27, No. 1. (January 2005), pp. 31-38.