

Social Media & eGovernment



Vandy BERTEN

Section Recherche



Table des matières

Introduction

Use case 1 : Recherche par mots-clés

Use case 2 : Page Facebook

Use case 3 : Network Analytics

Conclusions



Introduction



Table des matières

Introduction

Contexte

eGov belge

À l'étranger

Réseaux sociaux

Uses cases

Mots-clés

Page Facebook

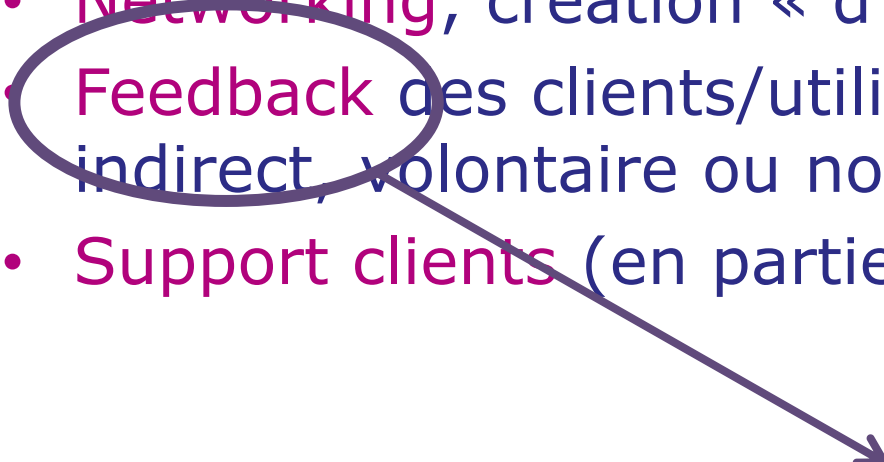
Network Analytics

Conclusions



Contexte

- Réseaux/médias sociaux : au centre de la **communication** de la grande majorité des « communicants » (marques, entreprises, personnalités, presse ...)
- **Diffusion** de publicités, campagnes...
- **Networking**, création « d'ambassadeurs »
- **Feedback** des clients/utilisateurs/fans... (direct ou indirect, volontaire ou non)
- **Support clients** (en partie par les clients)



Focus de l'exposé : médias sociaux en tant que source d'information



Informations ?

- Expression **d'avis**, de sentiment, de protestation, de mécontentement...
- **Amplification** de messages (presse, « amis », personnalité...)
- **Rumeurs**, hoax, dénigrement, désinformation...
Canal idéal, rapide et **(quasi) incontrôlable** !
- **Évènement** plus vite sur les réseaux sociaux que sur les « médias traditionnels »



« Dérapage » faciles, sans contrôle



Yves Leterme @YLeterme · 10 min.

Rôle emploi actu

US Airways @USAirways

Matignon @Matignon

In da face S&P! #bim RT @lemondefr : "Perte du AA+ : le prix Nobel d'économie Paul Krugman défend la France" lemonde.fr/tiny/3511260/

10/11/13 01:29

Perte du AA+ : le prix Nobel d'économie Paul Krugman défend la France

Le prix Nobel d'économie Paul Krugman prend la défense de la France face à la dégradation de sa note par l'agence de notation Standard and Poor's.



 Le Monde @lemondefr



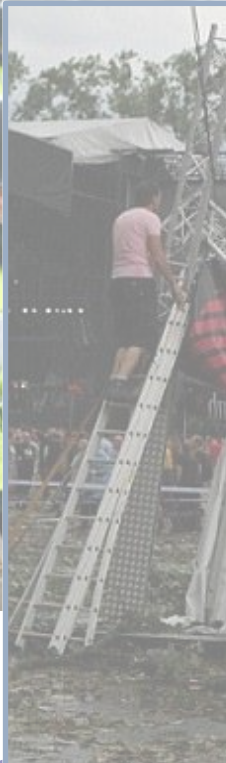
Médias sociaux et actualité



Émeute

Rôle majeur

de rumeur



© AFP/GETTY IMAGES



Pukkel Daesh (IS), 2014

Réseau : Outil principal de propagande

Message pour les proches, hébergement, ...



Situation actuelle eGov Belge

- Services actifs : Intégration sociale, ONP, ONSS (+student@work), Fedasil, AFSCA, Finances, armée...
- Services présents, mais inactifs : ONEM
- Présence partielle : CPAS Forest
- Police : grande « fragmentation »
 - @PolFed_presse, @police_temoin, @Polsupport_F, @HELI_FEDPOLFED, @polinfo_fr, @Jobpol_F (+NL)
 - @PolitieLeuven, @Politie_Halle, @PZSintNiklaas, @Politie5418, @ZPNamur, @polzonepuy, @PZWLW
- Qui fait réellement du monitoring/webcare ?
Difficile à savoir...



À l'étranger

- Dans presque tous les pays, des institutions sont présentes
- Quelques pays ont une vraie politique « médias sociaux »
- **Nouvelle-Zélande** (référence pour Gartner) : **lignes de conduites** (générales, en cas de problème, ...), **toolbox**...
- Canada : beaucoup d'organismes présents
- ...

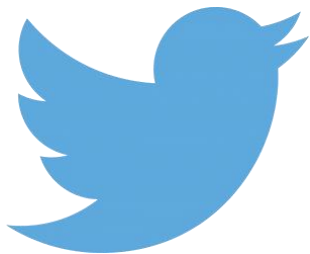


Cette présentation est/n'est pas ...

- Cette présentation **n'est pas** :
 - Une intro à « comment démarrer sur les médias sociaux »
 - → SPF « **Personnel & Organisation** »
 - Littérature abondante
 - Comment écrire un tweet percutant, réussir une campagne...
 - Un catalogue d'outils de monitoring
 - Un tutorial des API
- Focus :
 - Identification de **problèmes**, plus que mesure de performance d'une campagne/d'une marque
 - Présentation de **techniques**, plus que des outils



Réseaux sociaux



Réseaux sociaux



Facebook

- 1,3 milliards d'utilisateurs actifs
- 1 Belge sur 2 (5.6 millions) !
- Présence massive du politique, des entreprises, des people...
- Beaucoup d'expression d'opinion
- Une API permettant de récolter beaucoup de données, mais uniquement pour ce qui est « public » (pages + utilisateurs d'application)
- Énormément d'informations personnelles



Réseaux sociaux

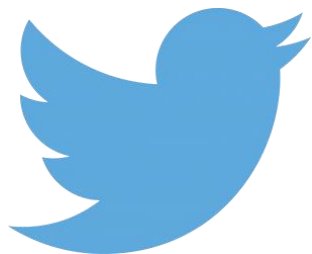


LinkedIn

- 300 millions d'utilisateurs actifs
- 2.2 millions de Belges
- Plus un espace de networking que d'expression
- API disponible



Réseaux sociaux



Twitter

- 200 millions d'utilisateurs
- 1 million en Belgique (?)
- Beaucoup d'opinion exprimée
- Relations plus basées sur l'intérêt que l'amitié
- Tout est public ! (sauf direct messages)
- API complète, mais limitations temporelles (requête/minutes et passé restreint)



Réseaux sociaux

WhatsApp

- Outil de communication, similaire au SMS
- Uniquement des messages privés, rien n'est public
- Pas d'API



Google+



Réseaux sociaux

Google+

- 1 millions de Belges, mais peu actifs
- Vient avec un compte Gmail
- API disponible
- Public très « branché » [opinion !]



Réseaux sociaux

Snapchat

- Similaire à WhatsApp
- Outil de communication
- Uniquement des messages privés, rien n'est public
- Tout s'efface en quelques secondes ... pas d'API par définition !



Réseaux sociaux

Netlog

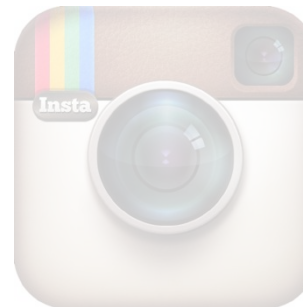
- Similaire à Facebook
- Principalement pour les adolescents
- Réseau belge, a été populaire en Belgique, surtout en Flandre,
- Est « en voie d'extinction » !
- API disponible, mais très mal documentée



Réseaux sociaux

Pinterest

- Plateforme de partage de photos
- Pas beaucoup d'opinion exprimée
- API disponible



Réseaux sociaux

Instagram

- Plateforme de partage de photos
- Idem que Pinterest, pas beaucoup d'opinion exprimée (en Belgique ?)
- API disponible



Uses cases

- **Use case 1 – Twitter** : recherche par **mots-clé** ; Comment identifier les « hot topics » ? Sur un sujet particulier ? Comment collecter l'information et la traiter ? Comment identifier des pics de discussion ? De quoi parle-t-on ?
- **Use case 2 – Page Facebook** : Quels sont les sujets de discussion ? Peut-on identifier des moments de tension ? Les citoyens sont-ils critiques ? Qui sont les citoyens qui s'expriment ?
- **Use case 3 – Network analytics** : À partir d'un ensemble d'informations publiques, que peut-on reconstruire ?



Table des matières

Introduction

Use case 1: Introduction
Mots-clé

Hot topics

Recherche ciblée

Détection d'évènement

Page Facebook

Network Analytics

Conclusions



Mots-clés

INTRODUCTION

Contexte

- Twitter : tous les tweets **publics** ! (FB : pages publiques)
- Limite des 140 caractères : on va à l'essentiel
- Très utilisé pour identifier les **sujets populaires**
- Contraintes API Twitter :
 - **Temporelle** : # requêtes/minutes limité
 - **Contenu** : accès (borné à 1%) au 7 derniers jours
- Certains outils ont un accès complet (Firehose)



Mots-clés

HOT TOPICS

« Hot topics »

- Identification de sujets « chauds » sans savoir ce qu'on cherche
- À faire en parallèle à des recherches plus ciblées, pour identifier les nouveaux sujets
- Se base sur les mots qui reviennent le plus souvent dans une zone géographique précise



Twitter.com

- Sur la page d'accueil



- <https://twitter.com/i/discover>







Mots-clés

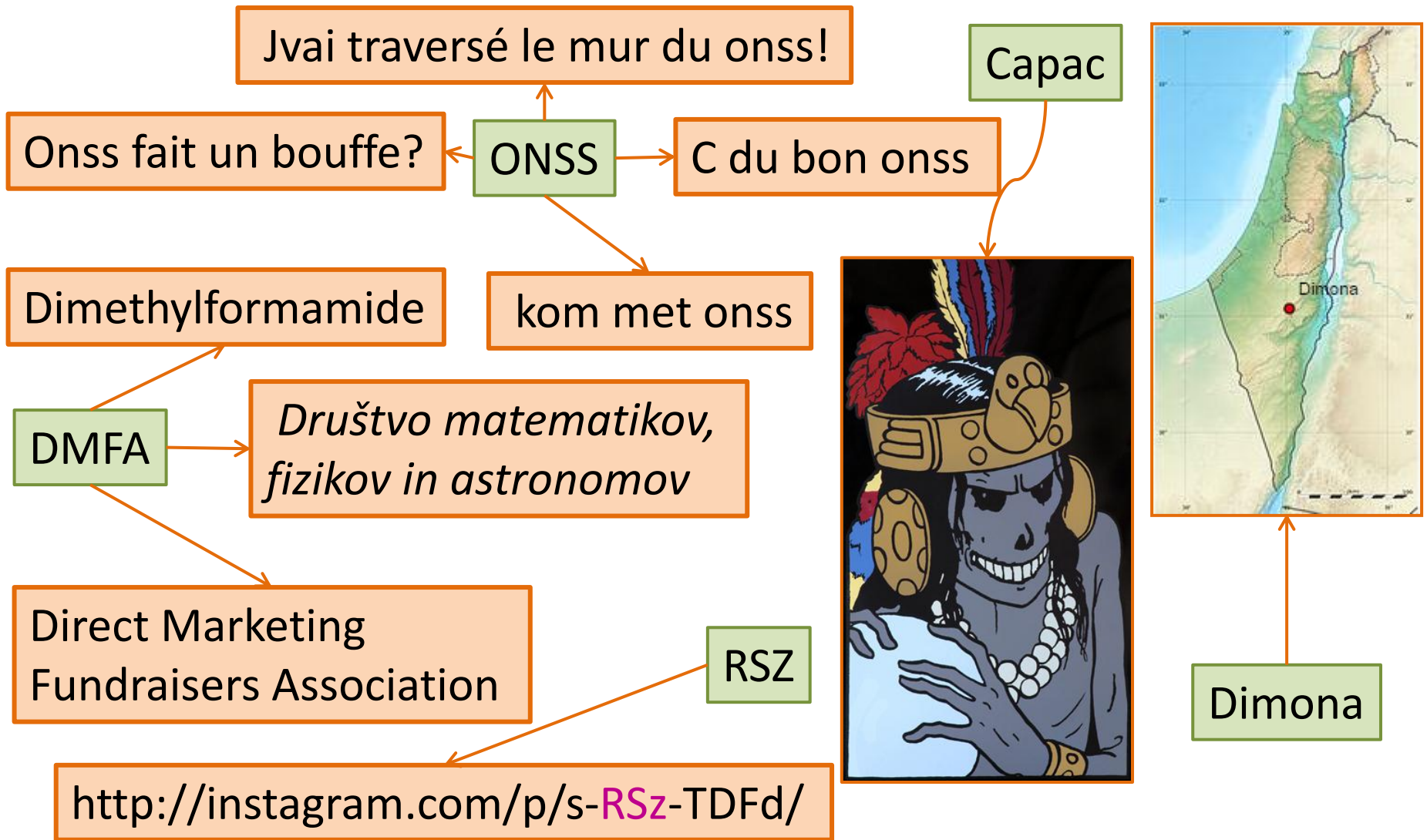
RECHERCHE CIBLÉE

Recherche ciblée

- Si on connaît les « mots-clés » à examiner, il existe de nombreux outils
- Certains se basent sur le passé, d'autres nécessitent la mise en place d'une « écoute »
- Un certain nombre de moteurs de recherche permettent juste de lister les tweets répondant à une requête → limite l'intérêt
- Pour la suite, on s'intéresse surtout à Twitter



Difficultés



Contourner les difficultés

- Critère linguistique : « dimona lang:fr »
 - langue du profil, pas du message
- Rajouter des mots-clés : « déclaration dimona »
 - limite Israël, mais trop restrictif (6 tweets !)
- Exclure des mots : « dimona -israel -israélien -palestinien -palestine -nucléaire -gaza -missile »
 - Pas encore suffisant !
- Critère géographique : « dimona near:brussels within:150km »
 - Doit être renseigné (profil ou tweet)



Méthodes d'extraction

Plusieurs méthodes d'extraction de tweets :

- Configurer un **outil** (Engagor, Mention.com...) avec une recherche programmée (passé souvent limité)
- Utiliser **l'API** :
 - « **search** » : tweets jusqu'à une semaine en arrière
 - « **stream** » : « réveillé » dès qu'un nouveau tweet arrive
- Web **crawler** : version « Web » de Twitter donne plus de résultats que l'API (plus anciens) → possibilité d'extraction, mais laborieux



« Tax-on-web » : Google trends

tax on web
Terme de recherche

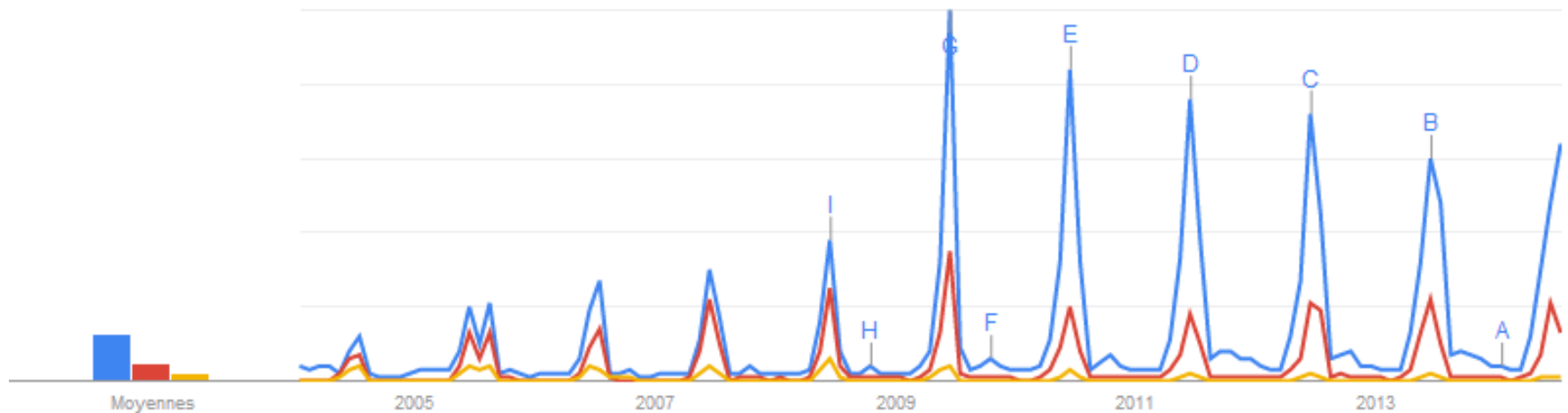
taxonweb
Terme de recherche

tax-on-web
Terme de recherche

+ Ajouter un terme

Évolution de l'intérêt pour cette recherche ?

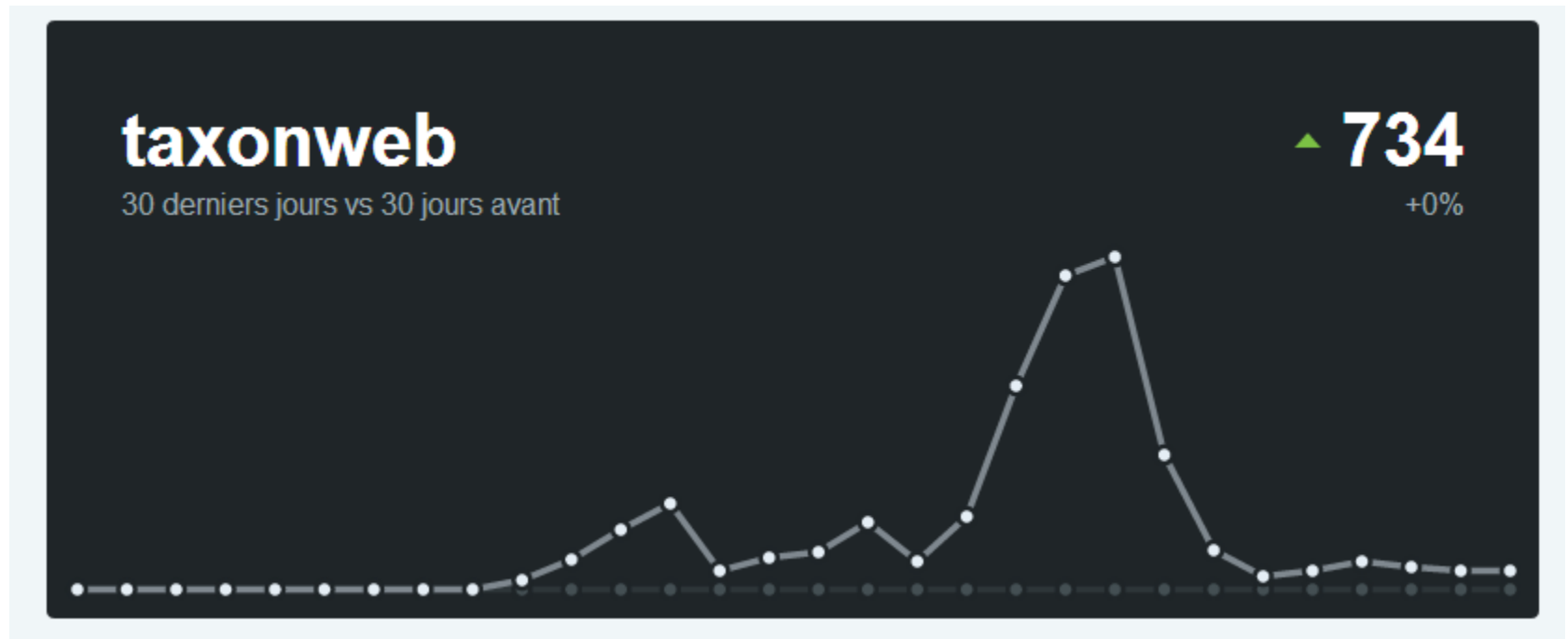
Titres des actualités Prévisions ?



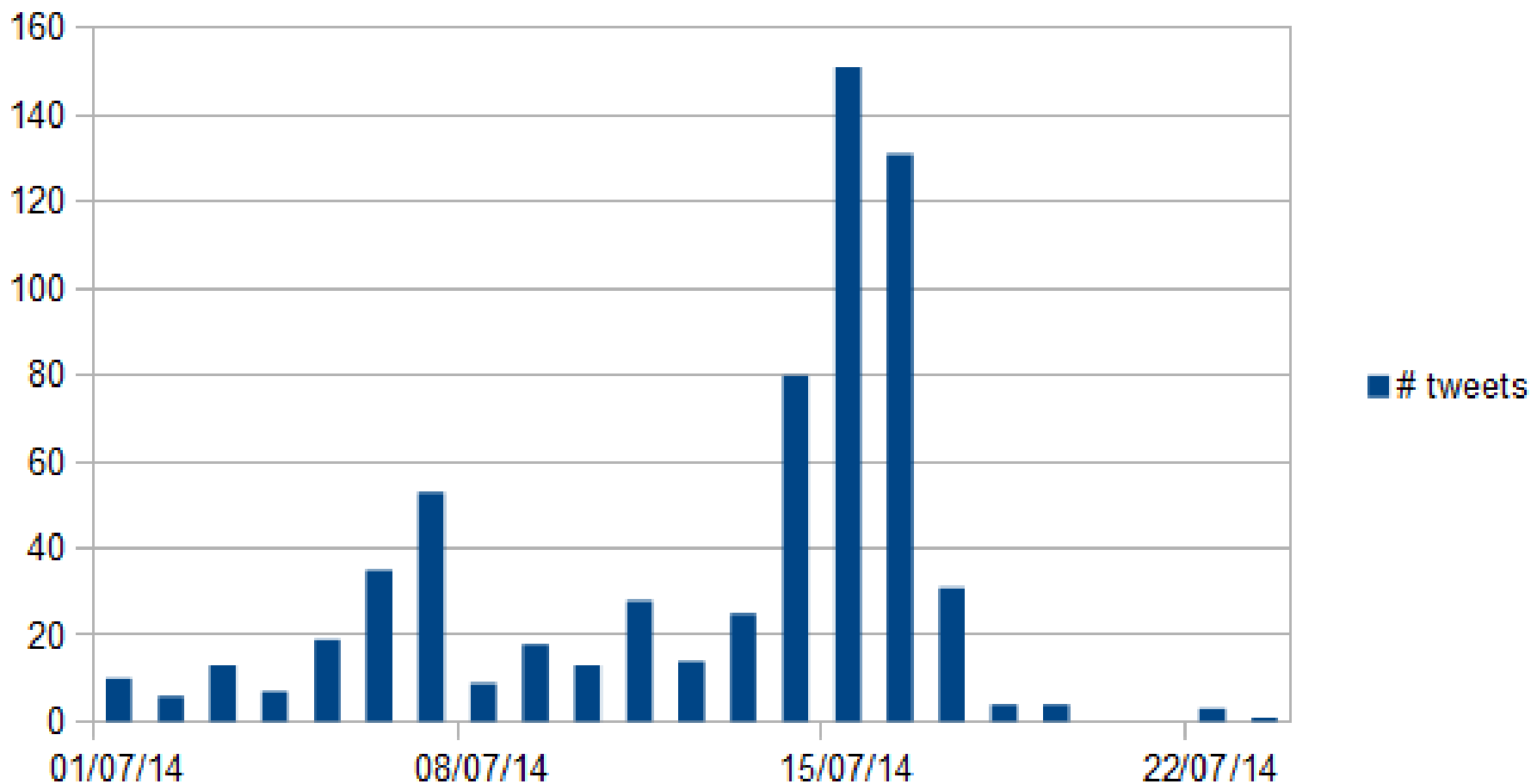
« Tax-on-web » : Topsy.com



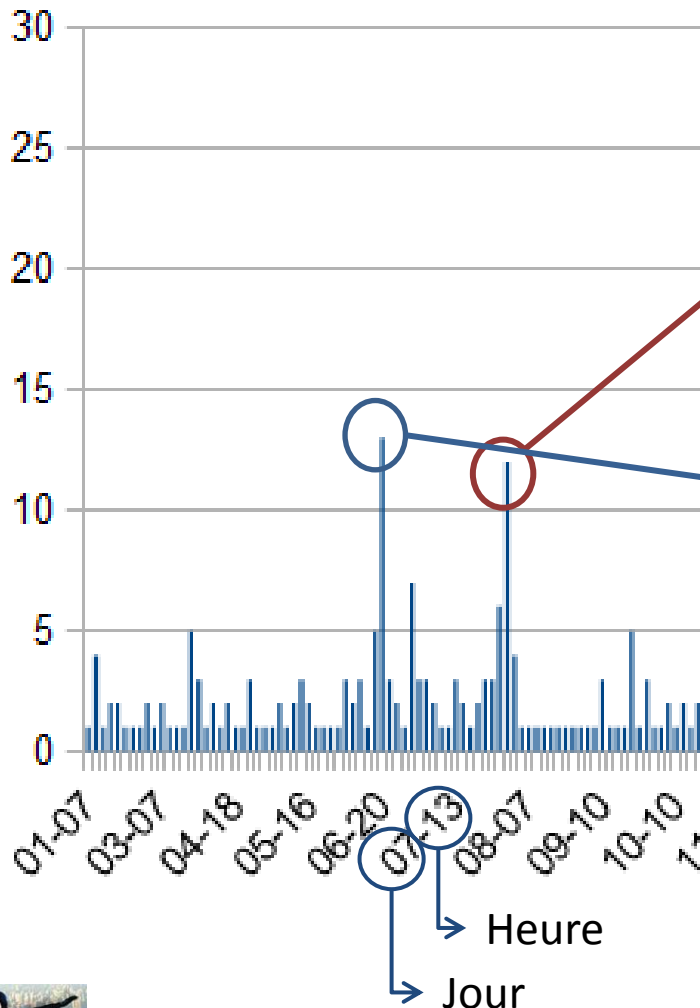
« Tax-on-web » : Mention.com



« Tax-on-web » : API + tableur (/jour)



« Tax-on-web » : API + tableur (/heure)



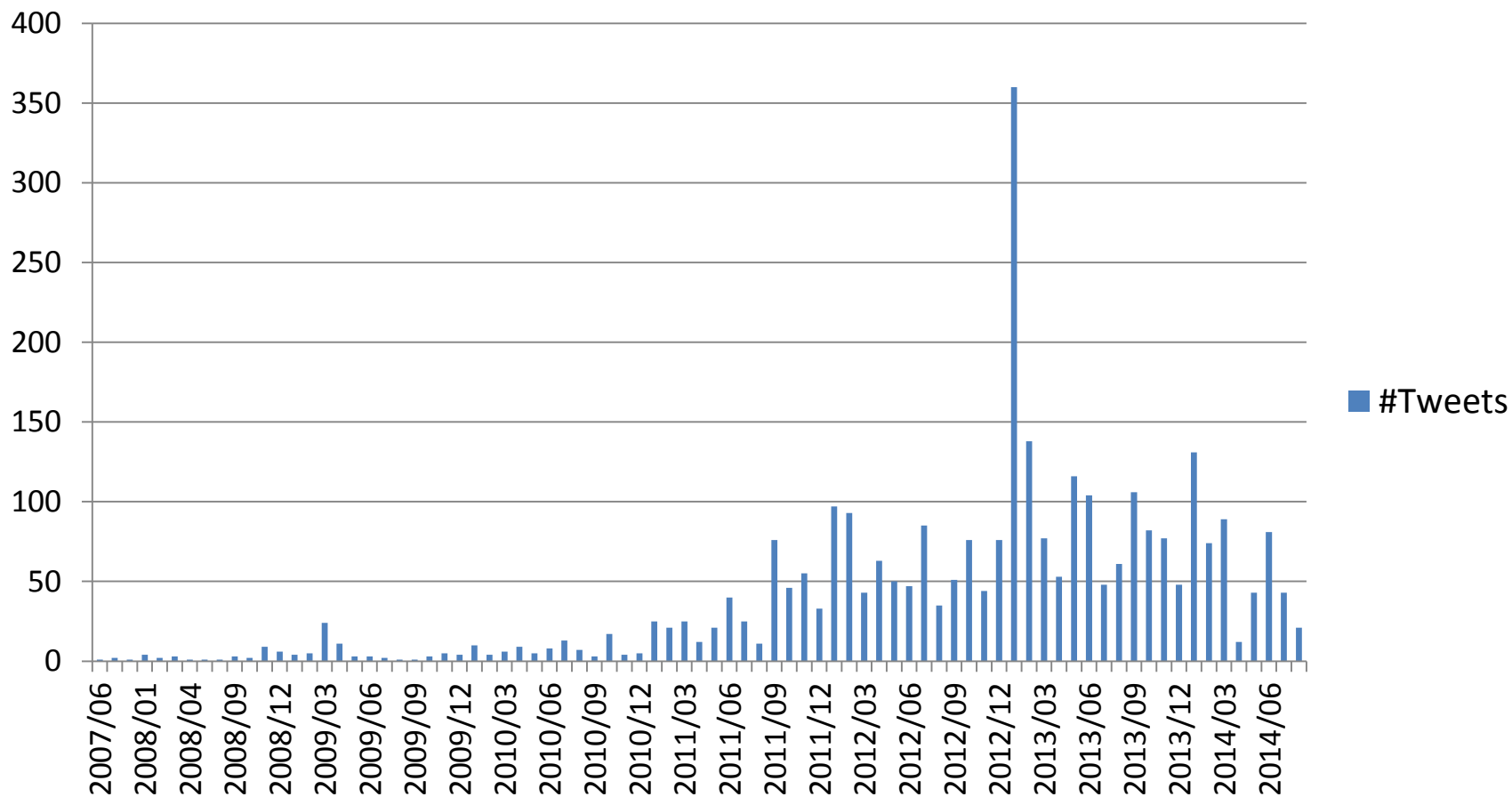
• @EOD Finance :

« Tax-on-web **down**... » ; « Lap: #taxonweb is **down**. » ; « Bijna zo **down** als de mensen die er gebruik van wilden maken. » ; « Typisch : [...] aan zijn #taxonweb dan ligt het systeem weer **plat**. #fb » ; « Heeft iedereen [...] **#crash #serverdown** » ; « Een "groene" dag op #taxonweb , maar helaas: site **niet beschikbaar** :-(. Elk jaar een probleem ... volgend jaar dan maar terug op papier? » ; « Lap #taxonweb **#down**. Even geduld dus » ; « pfff #taxonweb est **down**, juste quand j'allais m'y mettre ! »

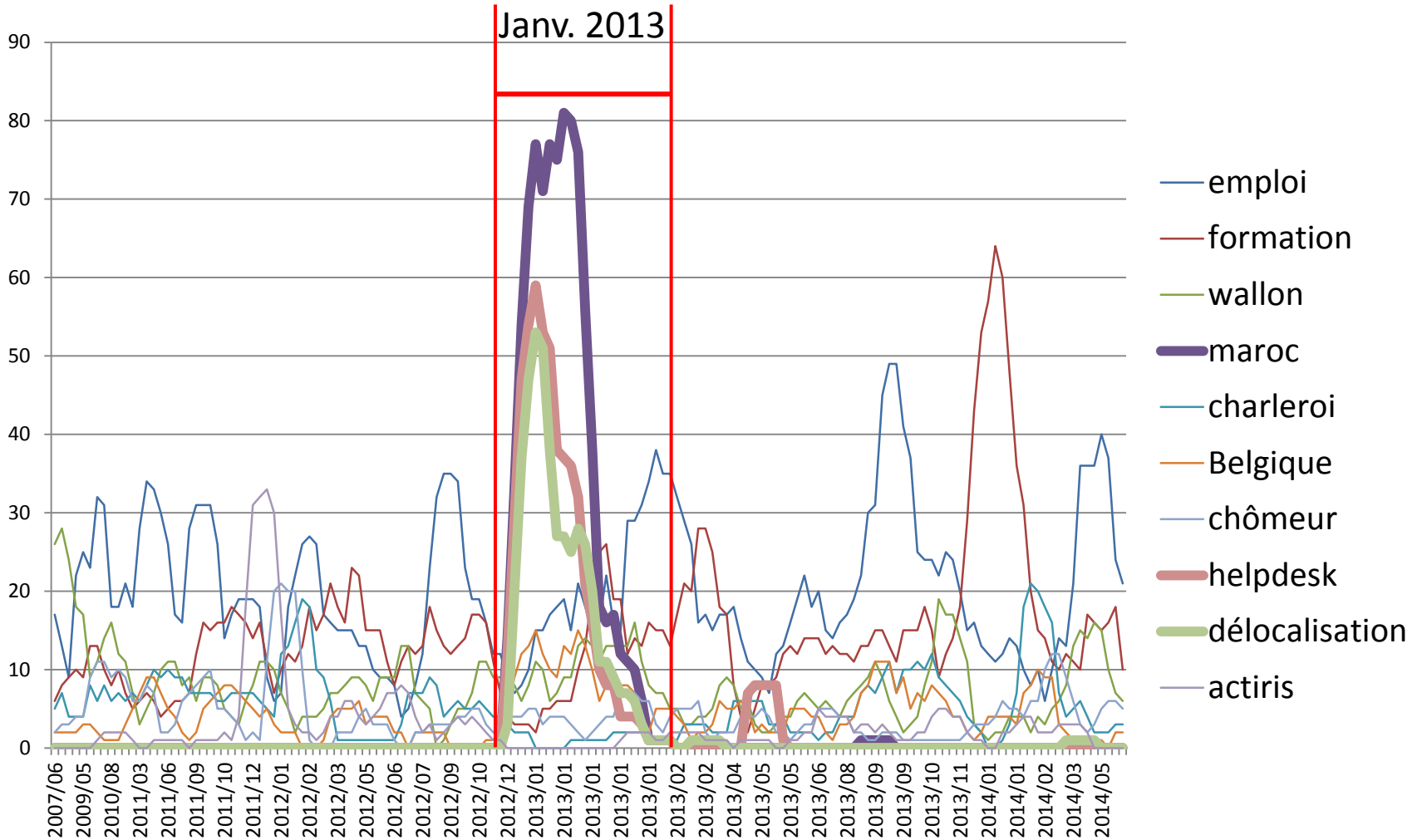


Forem (web crawling)

Forem (fr)



Forem (Twitter) : topic extraction



Mots-clés

DÉTECTION D'ÉVÈNEMENTS

Détection d'évènement

- De plus en plus, les **témoins d'un incident** le tweet directement
- Un évènement majeur (attentat, catastrophe, ...) génère un **grand nombre de tweets localisés**, avec des **mots en commun**
- Certains outils (**Dataminr**, **Twitcident**) s'en servent pour automatiquement détecter un évènement sur base de critères
- Pas encore d'outil gratuit ?



Mots-clés: l'essentiel

Basé sur des noms de projet, d'organisme, de personne

Difficile sur des acronymes courts, peu uniques

Y réfléchir quand on crée un projet ? Créer des comptes ou hashtags ?

Identification + compréhension des pics de tension

Support « non structuré » (≠ page FB)
bruit à filtrer



Mots-clés: les outils

Recherche ciblée (listing):

- Twitter.com
- Tweetdeck
- Topsy
- Socialmention

Recherche ciblée (listing +stats) :

- Engagor (15j+€)
- Hootsuite (subset+€)
- Simplymeasured (15j+€)
- Topsy (€)
- ...

Hot topics :

- Trendsmap (subset+€)
- Geofeedia (€)
- Trends topics (Twitter)

Pour les plus branchés :

- API search & stream
- Selenium (Web crawling)

- Détection évènement

- Dataminr (€)
- Twitcident (€)
- Google Trends





Smals
Community & Government

Create Page

- Recent
- 2014
- 2013
- 2012
- 2011
- 2010
- 2009
- Founded

Timeline About Photos Likes More ▾

PEOPLE >

★★★★★
498 likes
216 visits

and like this.

Invite your friends to like Smals

-
-
-

Invite

Invite

Post Photo / Video

Write something on this Page...

Smals shared a link.
2 hours ago

Al je wachtwoorden beheren via een online tool? Dat klinkt erg veilig, maar het blijft openen. Tania Martin van Smals Research beschrijft vijf OPM-tools en waarvan 5 met een veiligheidsprobleem.

Use case 2 : Page Facebook

Table des matières

Introduction

Mots-clé

Use case 2 :

Page Facebook

Introduction

Analyse de fréquence

Topic extraction

Sentiment Analysis

Network Analytics

Network Analytics

Conclusions



Page Facebook

INTRODUCTION

Contexte

- Page Facebook : **Vitrine** d'une société, d'un organisme, d'un projet (≠ profil)
- Tout est **public**, connexion pas nécessaire (→ indexé par Google)
- Contenu accessible via API
- Partie centrale : **Timeline** (Journal), composée de **posts** et de **commentaires**
- 3 niveaux d'interaction :
 - Posts par le gestionnaire, commentaires bloqués
 - Posts par le gestionnaire, commentaires acceptés
 - Posts par tout le monde, commentaires acceptés



Contexte

- **AFSCA** : Agence Fédérale pour la Sécurité de la Chaîne Alimentaire (= FAVV)
- Compétente pour **l'alimentation** (magasin, restaurants, usines, fêtes...) mais également pour les **animaux** (vaccins, commerce...) domestiques et d'élevage
- Page Facebook : annonce de contrôle, rappels de produits, annonces diverses
- Comme toute agence de contrôle : pas que des amis !
- Mais enjeux et responsabilités considérables !



Contexte : exemple d'annonce



 **Agence alimentaire (AFSCA)** shared a link.
September 10

Communiqué de Belki :
Rappel de certains lots de viande de poulet.
Problématique : Présence de salmonelle.

 **AFSCA - Communiqué de Belki**
www.favv-afscsa.fgov.be

Like · Comment · Share ↻ 57 Shares

 Fabrice Michel Lindley, Luc Leens, Zdenek Sotnyk and 6 others like this. Top Comments ▾

 **Philippe Bille** c bon , faites semblant , c vous qui autoriser ça , , y a la bouffe des démunis et la bouffe de ceux qui ont les moyens , alors on fait quoi là , pq vous laisser des élevages de merde avec des milliers de poulets ,qui n'auront meme jamais vu le ciel ... [See More](#)
👍 1 · September 11 at 4:54am

 **Yves** J'avais l'habitude d'acheter depuis 7 ans des yaourts et du lait dans des récipients en verre avec consigne chez mon fermier. Ce système de consigne, je trouvais que c'était une bonne alternative pour diminuer les déchets plastiques. De plus ce sont d... [See More](#)
September 12 at 4:17am

 **Agence alimentaire (AFSCA)** Nous demandons que le récipient en verre réutilisable soit suffisamment nettoyé (température suffisante pour une désinfection physique (> 70°C) ou produit désinfectant alimentaire autorisé avec un rinçage suffisant). Si le fermier ne peut pas le nettoyer correctement, il est préférable de ne pas se lancer dans un système de consigne.
September 12 at 5:50am

 [View 1 more comment](#)



Commentaires

A [redacted] e Madame, Monsieur les fonctionnaires de l'AFSCA.

Ce n'est pas une raison de tuer les chiots



L [redacted] a c est criminel de faire une chose comme ça pour c animaux X
l administration est grave il n ont aucune compation

See Translation



A [redacted] t <http://www.contrepoints.org/.../156797-belgique-quand...>
Qu'est-ce que c'est que cette histoire ??? Je m'oppose tout à fait à l'abattage de
ces pauvres chiens, c'est scandaleux et inhumain !

See Translation

Like · Reply ·  13 · February 16 at 9:45pm




arrivez à dormir la nuit????

See Translation



C [redacted] a Je suis choquée ! De quel c
laisser ces pauvres chiens dans une situation au
vousJe m'oppose à ce massacre !

See Translation

Like · Reply ·  3 · February 18 at 8:18pm · Edited

- 209 posts
- 590 commentaires
- 335 commentateurs
- 302 post likes
- 974 comment likes



Questions

- Peut-on « mesurer » cette situation ?
- À quel point est-elle problématique ?
- Est-ce un phénomène passager ?
- Les « contestataires » forment-ils une « communautés »
- Sont-ils représentatifs de la société en général ?

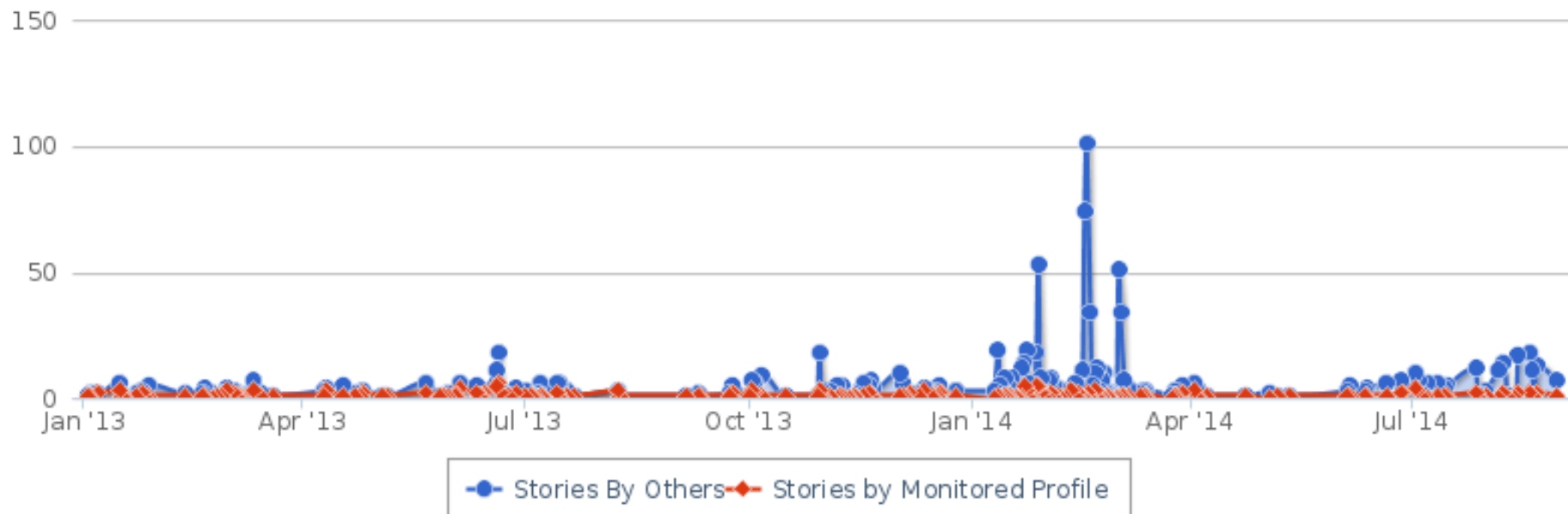


Outils disponibles

- Pour sa **propre page** Facebook : **nombreux outils** (gratuits ou non) disponibles
- Pour une **page quelconque** : **quelques outils...** pourtant le contenu d'une page est **public** !
 - Engagor
 - Quintly
 - SimplyMeasured
 - Aucun gratuit ?
- Une autorité peut avoir besoin de monitorer les pages de ses différentes entités...
- ... ou de suivre des pages liées à « l'opposition »



Engagor



Page AFSCA, posts (rouge) et commentaires (bleu) par jour



Engagor

Top Stories (by Interactions)

 **Agence Alimentaire (afsca)** on Agence Alimentaire January 29, 2014
(AFSCA)

Vous pouvez laisser des commentaires, vos questions, mais il y a quelques règles à respecter !
L'AFSCA reçoit les critiques dans le but de s'améliorer, mais...

 [More...](#)

AFSCA | Shares **16** | Comments **345** | Likes **37** |  **360** | 

 **Agence Alimentaire (afsca)** on Agence Alimentaire January 12, 2014
(AFSCA)

La rage est une maladie endémique en Espagne, plusieurs cas de rage y ont été détectés chez des animaux domestiques en 2013. La Belgique est indemne de rage et l'importation illégale d'animaux présente un risque élevé en terme de santé publique.

 [More...](#)

AFSCA | Shares **15** | Comments **124** | Likes **8** |  **111** | 



À quoi a-t-on affaire ?

- Une communauté **régulière** et active avec un **pic d'activité** ?
- 2-3 personnes qui se sont lancées dans un **débat sans fin** ?
- Des **visiteurs uniques** soudainement intéressés par un post ?
- Des gens qui se **connaissent mutuellement** ? Non liés entre eux ?
- Quel est le **sujet** du débat ?



Facebook API

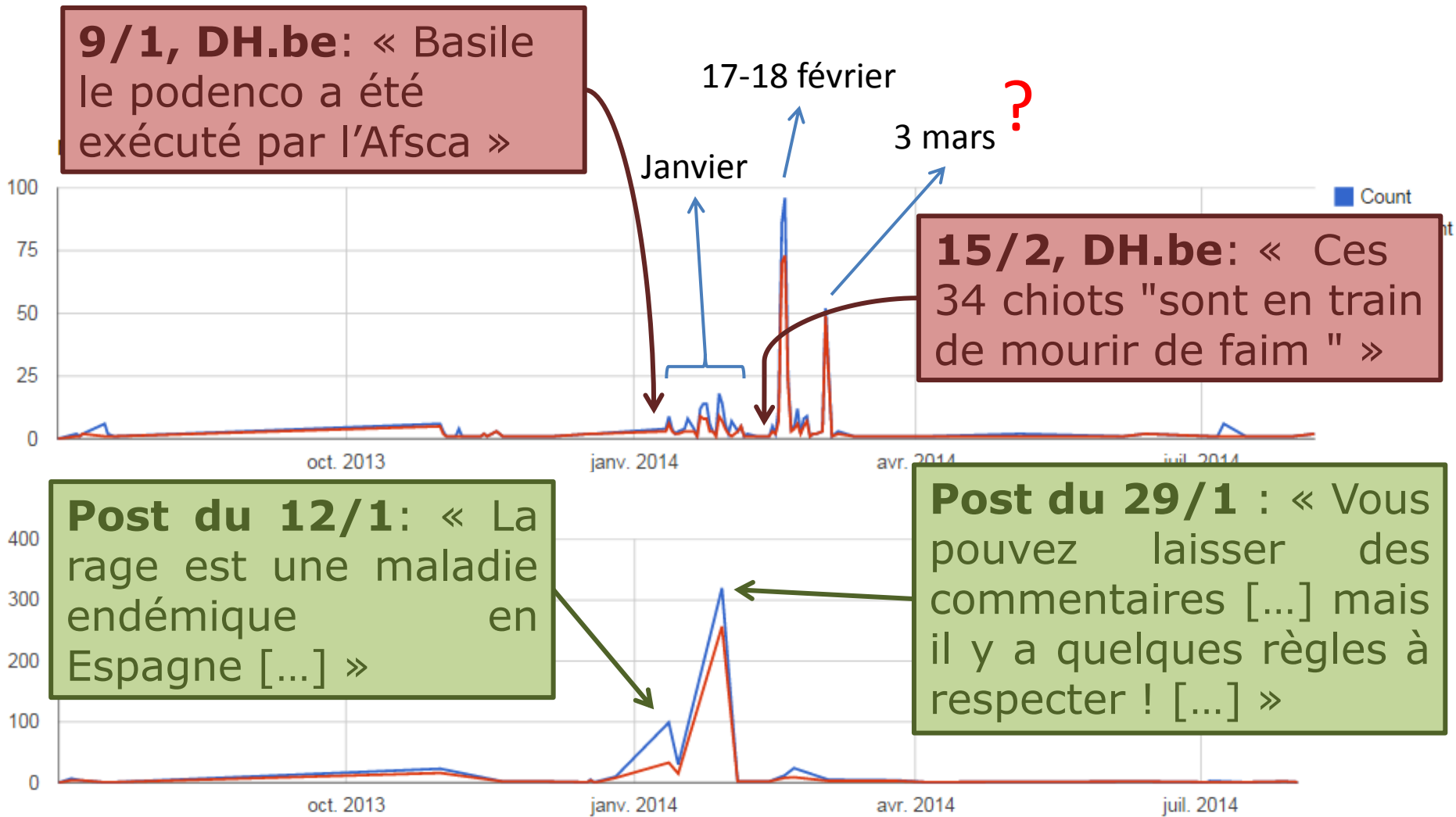
- Facebook permet d'extraire de l'information au travers d'**API** (Application Programming Interface) dans différents langages (principalement PHP et javascript)
- Authentification et délégation : oAuth
- Très **peu d'accès aux profils perso**, mais **accès complet aux pages**
- Certaines « application Facebook » (p.ex. Netvizz) permettent une extraction « toute faite »



Page Facebook

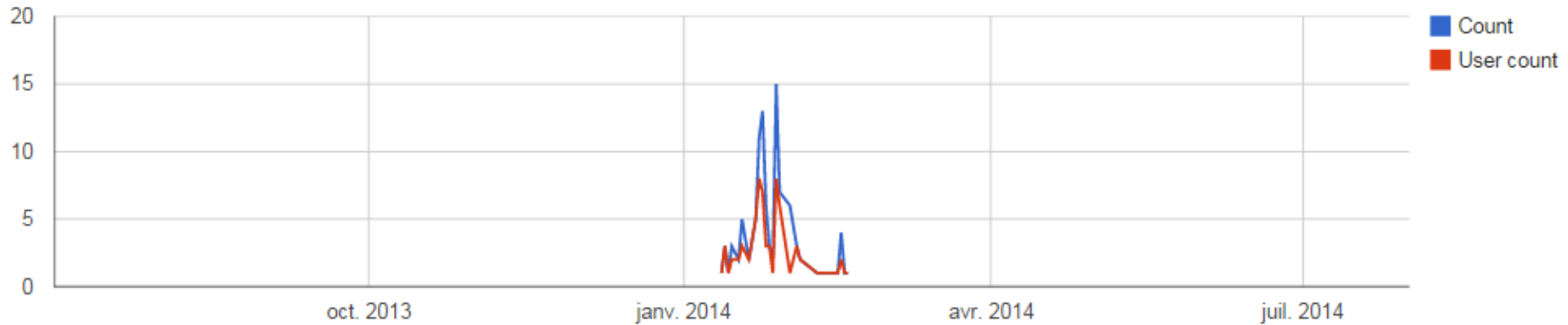
ANALYSE DES FRÉQUENCES

Facebook API + Google charts

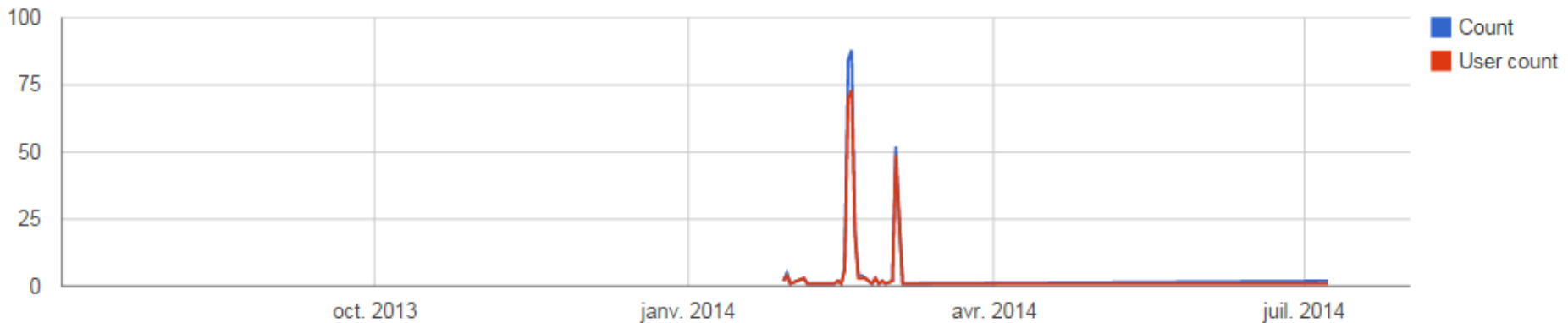


Facebook API + Google charts

Number of comments (post 12/1, date of the comment)



Number of comments (post 29/1, date of the comment)

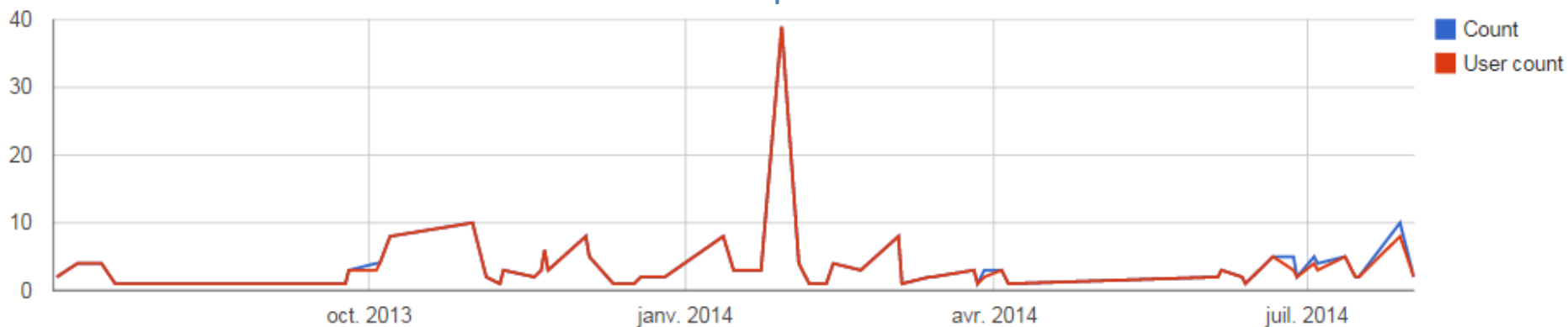


Facebook API + Google charts

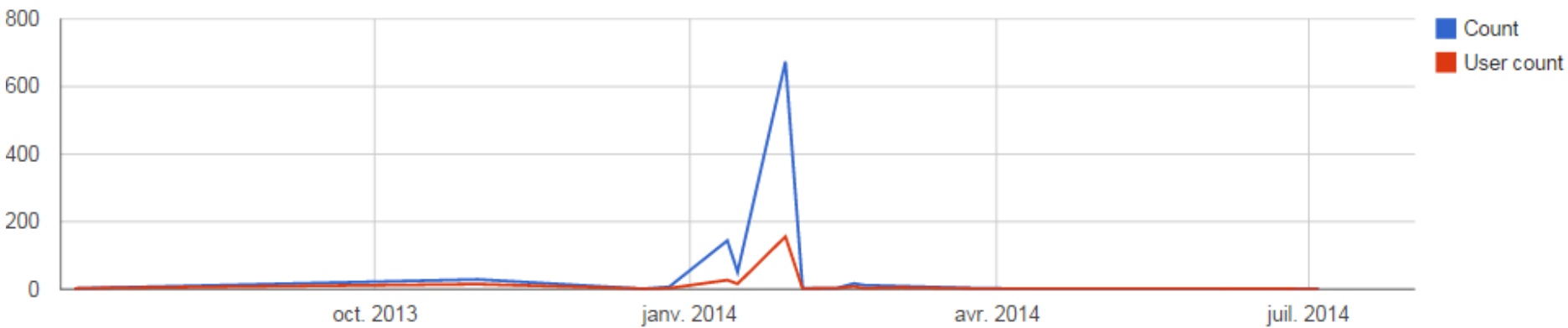
29 janvier



Number of post likes (date of the post)



Number of comment likes (date of the post)



Fréquence : lessons learned

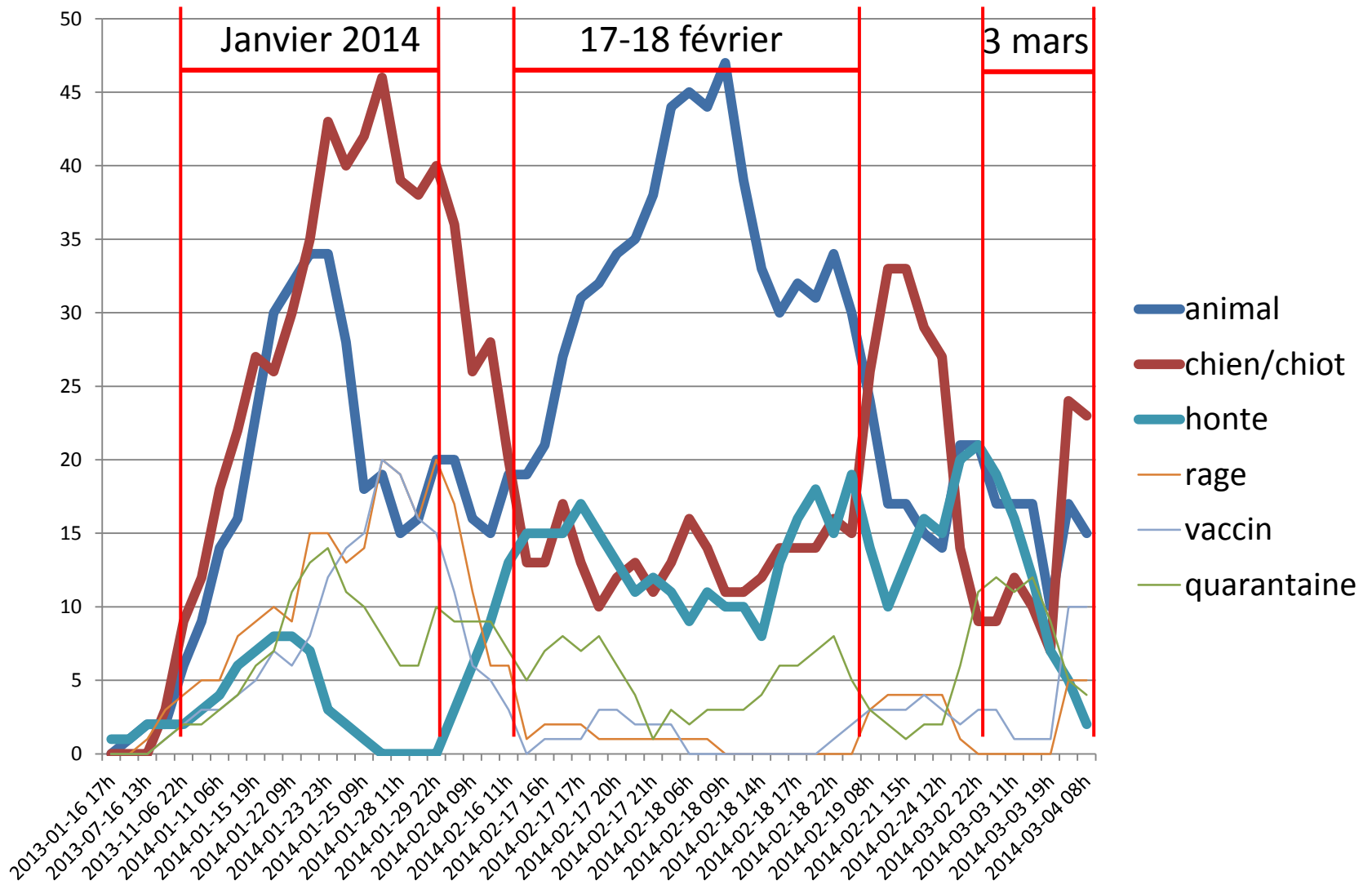
- Deux pics de commentaires, liés à **deux évènements identifiés** (Basile, quarantaine)
- Première phase : beaucoup de commentaires, peu de commentateurs
- Seconde phase : beaucoup de commentateurs
- Critiques **peu liées au post** commenté



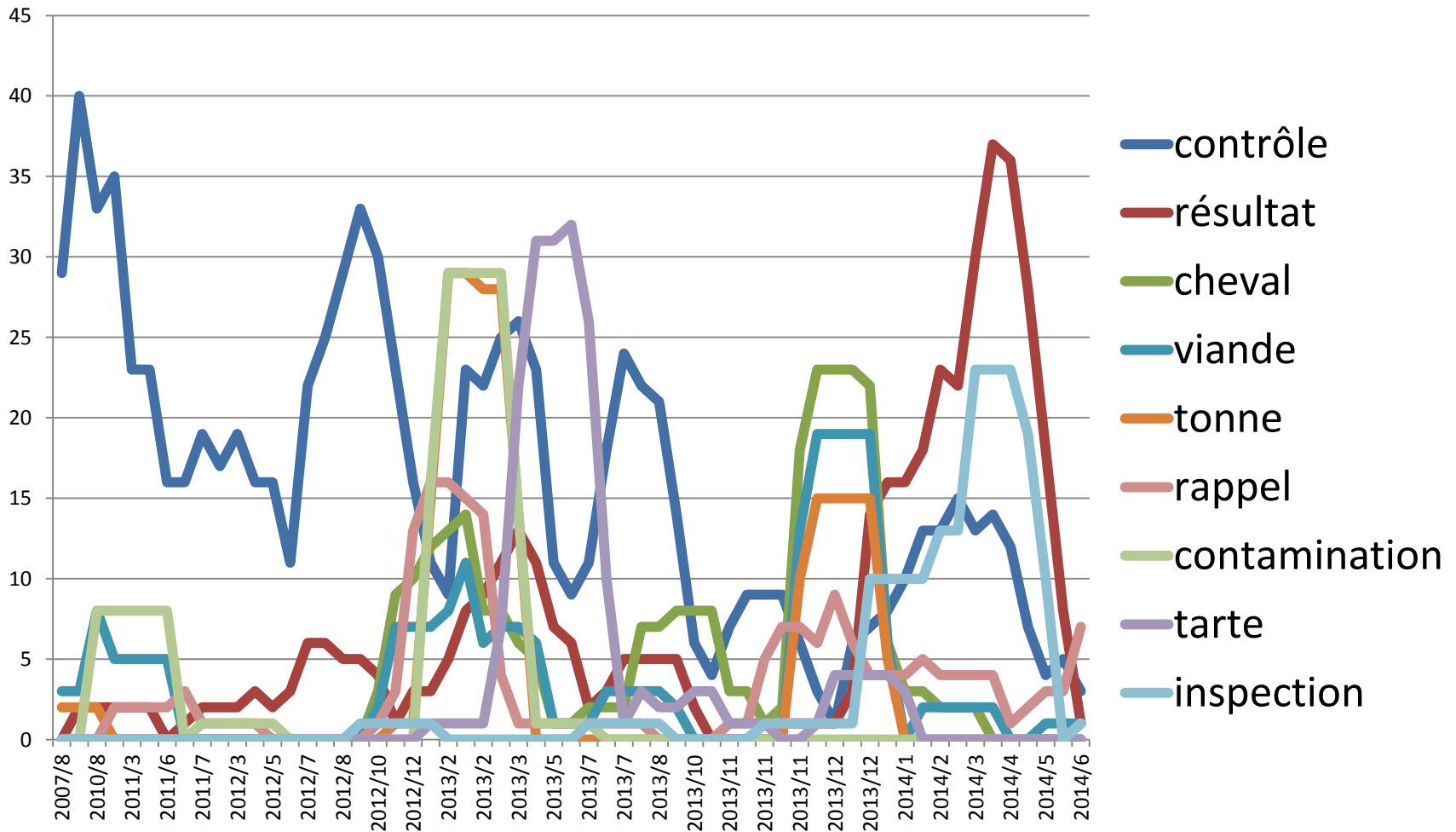
Page Facebook

TOPIC EXTRACTION

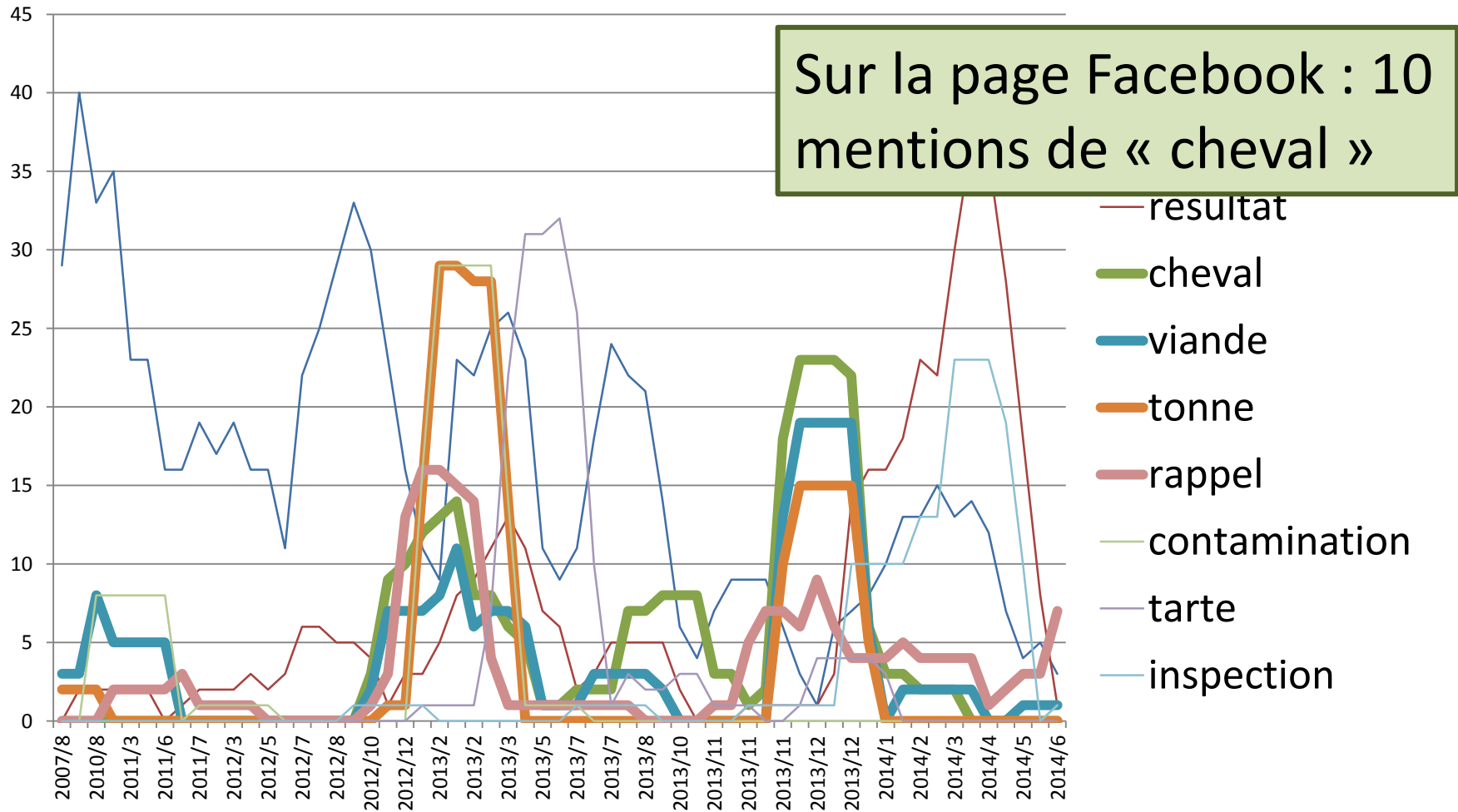
AFSCA (Facebook) : API + Excel



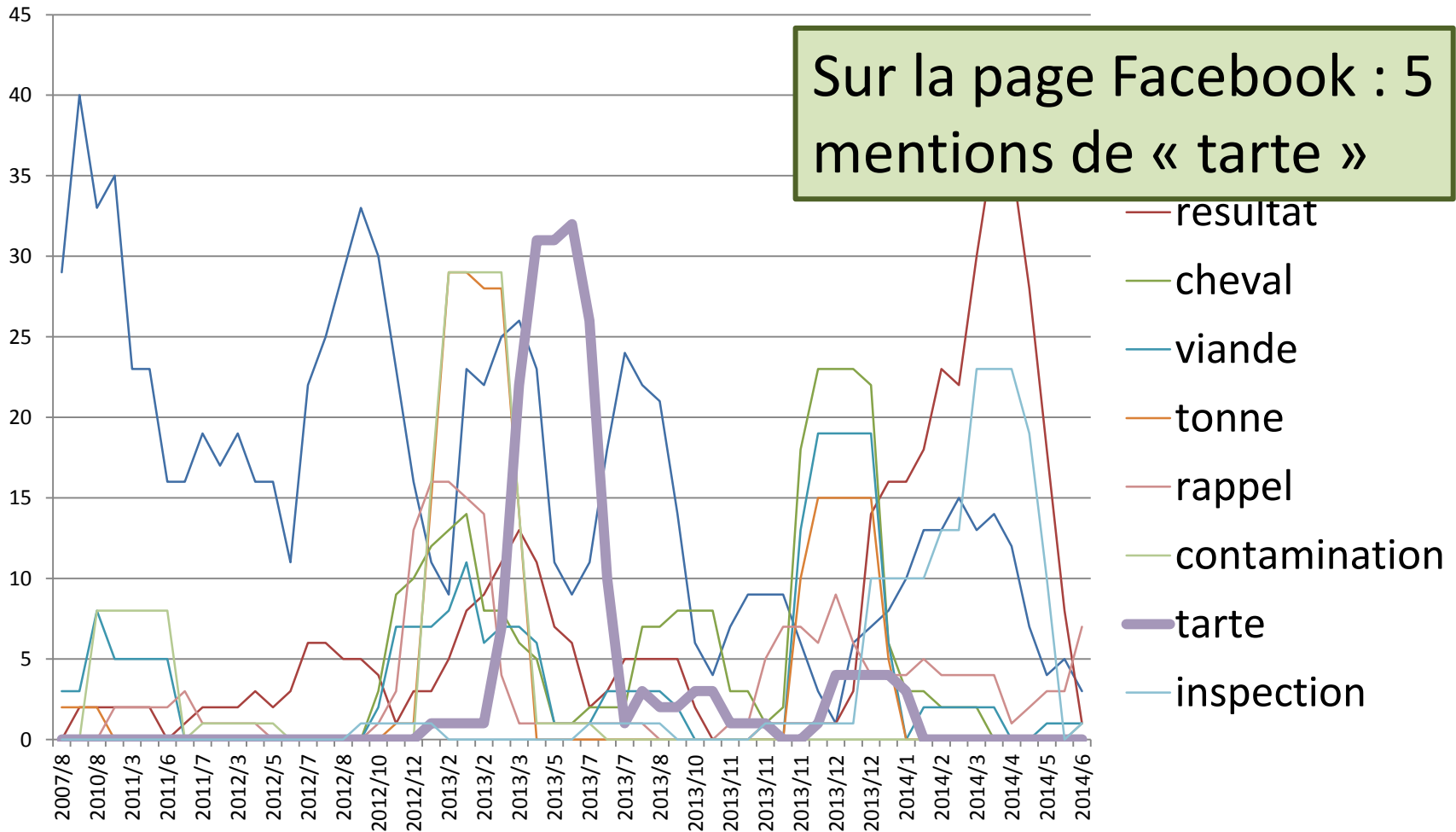
AFSCA (Twitter) : Crawler + Excel



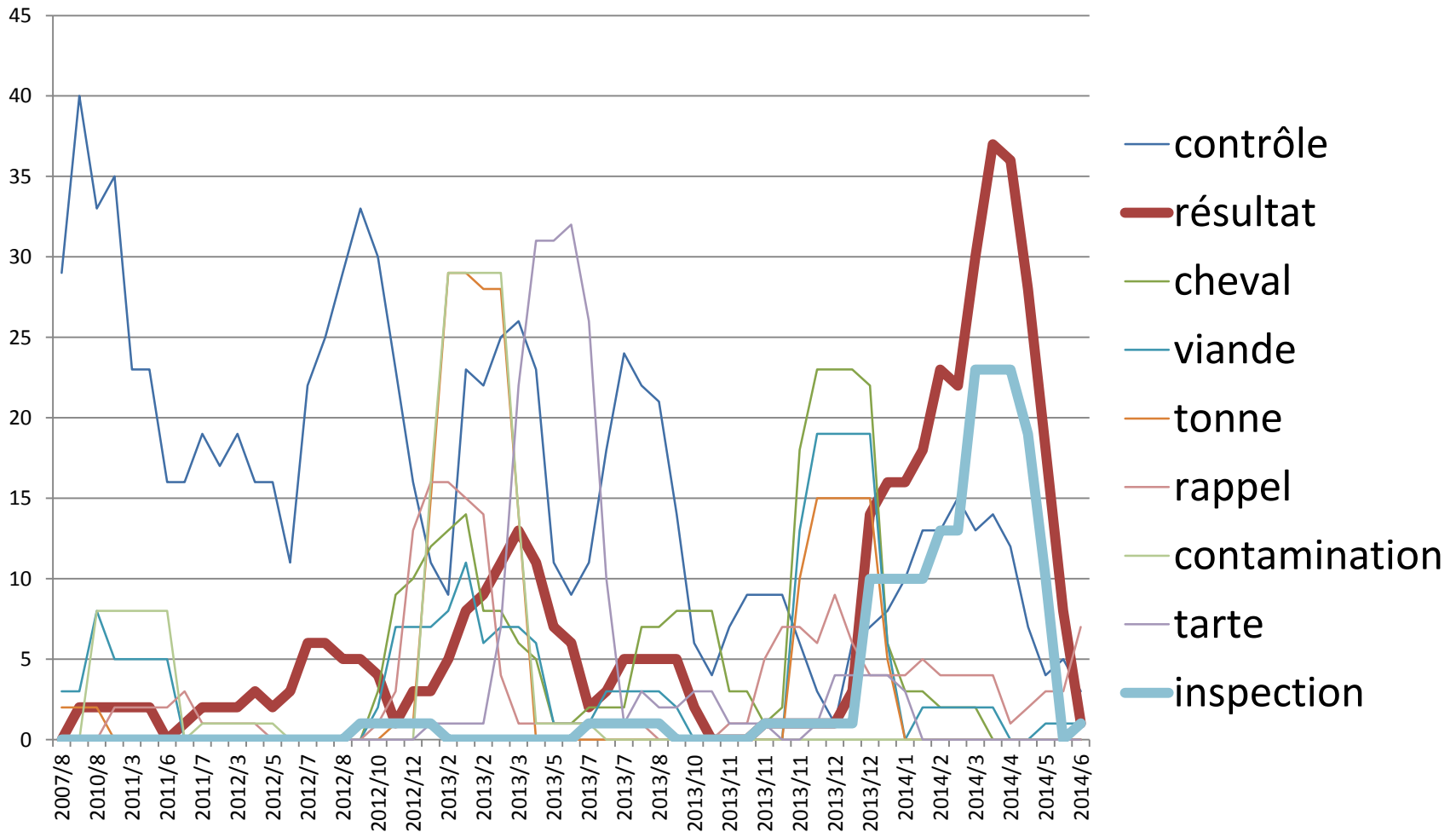
AFSCA (Twitter) : Crawler + Excel



AFSCA (Twitter) : Crawler + Excel



AFSCA (Twitter) : Crawler + Excel



Topic extraction: lessons learned

- **Facebook** : sujet = **animaux**
 - Première phase (Janvier) : rage (vaccin, quarantaine)
 - Seconde phase (Février) : animaux, honte
- **Twitter** : sujet = **alimentation**
 - Viande de cheval
 - Tarte
 - Publication des contrôles
- Permet **d'orienter** une campagne d'information vers le bon **public cible**



Page Facebook

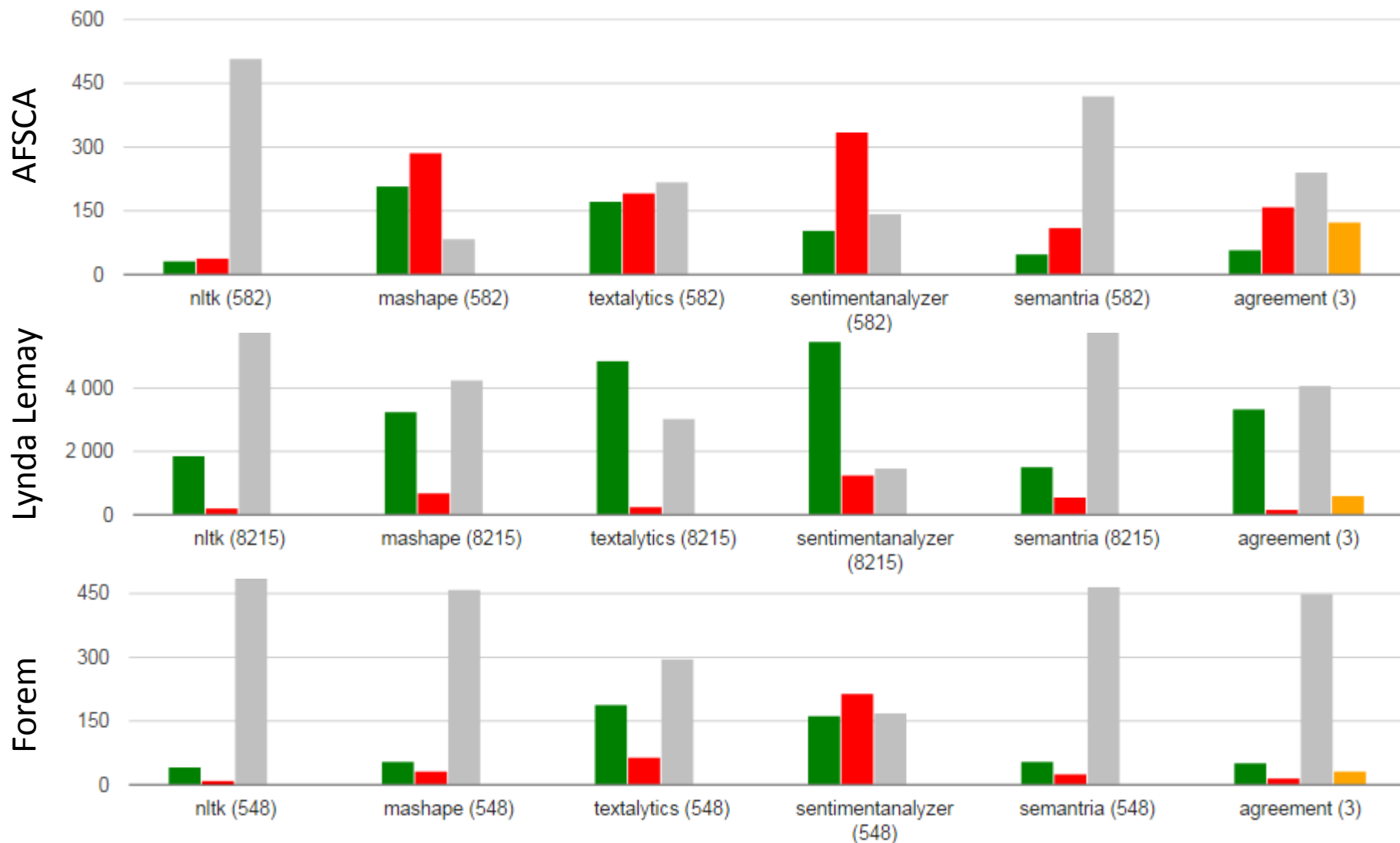
SENTIMENT ANALYSIS

Sentiment analysis

- Idée : déterminer si un texte est **positif**, **négatif** ou neutre
- Domaine de recherche issu de la *linguistique computationnelle*, surtout développé en **anglais**. Quelques outils en **français**, peu en **néerlandais**
- Gère mal les fautes d'orthographe, « langage sms », l'ironie
- Pas utile à l'échelle d'un message, mais pour des **tendances** plus larges, des comparaisons
- Remarque : on s'exprime plus pour râler que pour complimenter !



Sentiment analysis

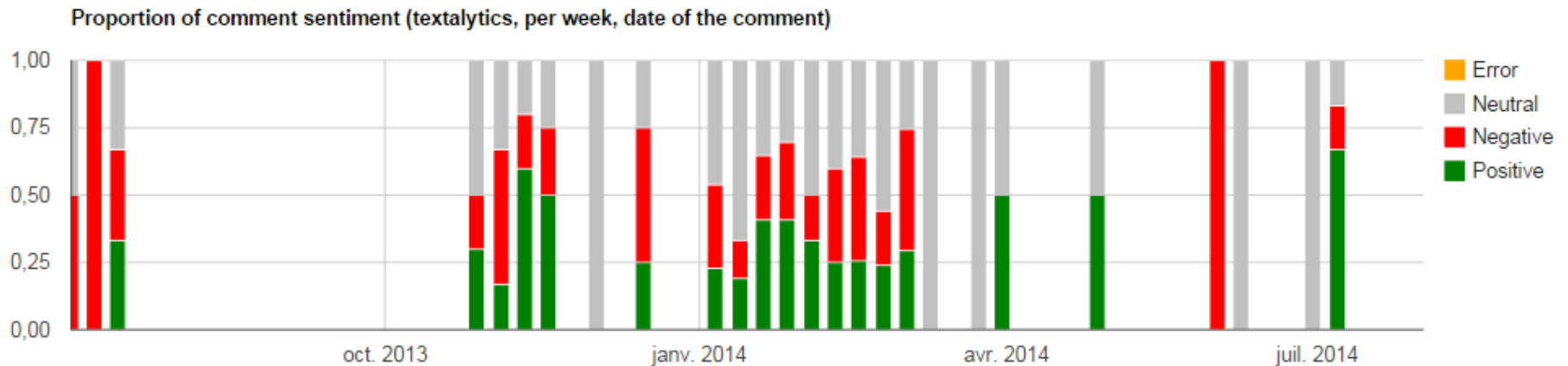
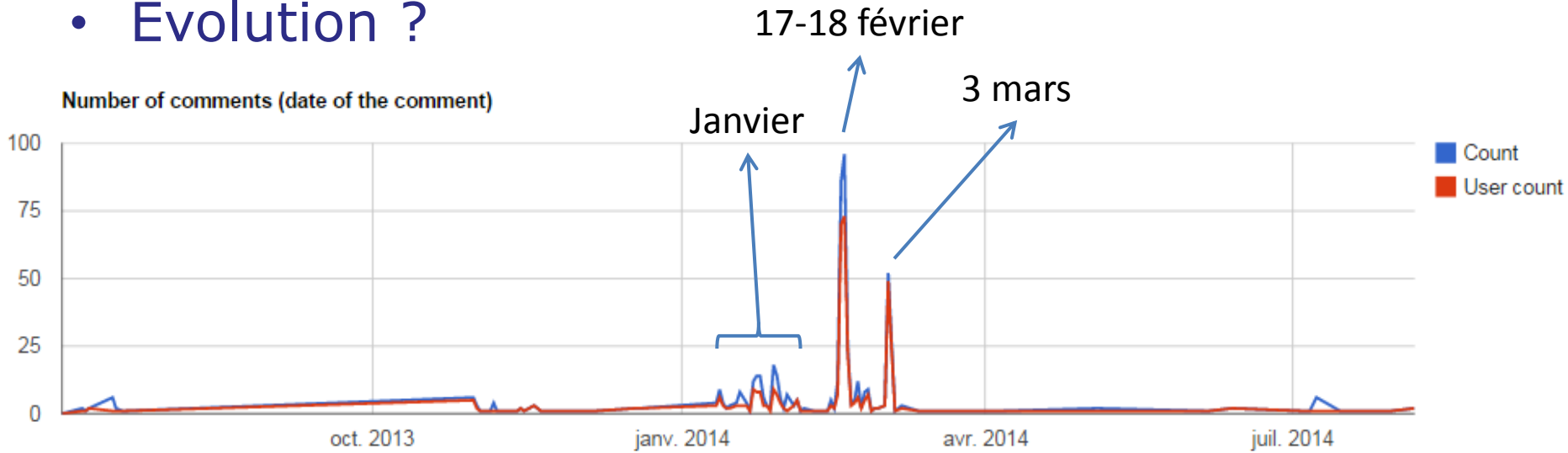


Positive Negative Neutral Error/Disagree



Sentiment analysis

- Évolution ?



Sentiment analysis: lessons learned

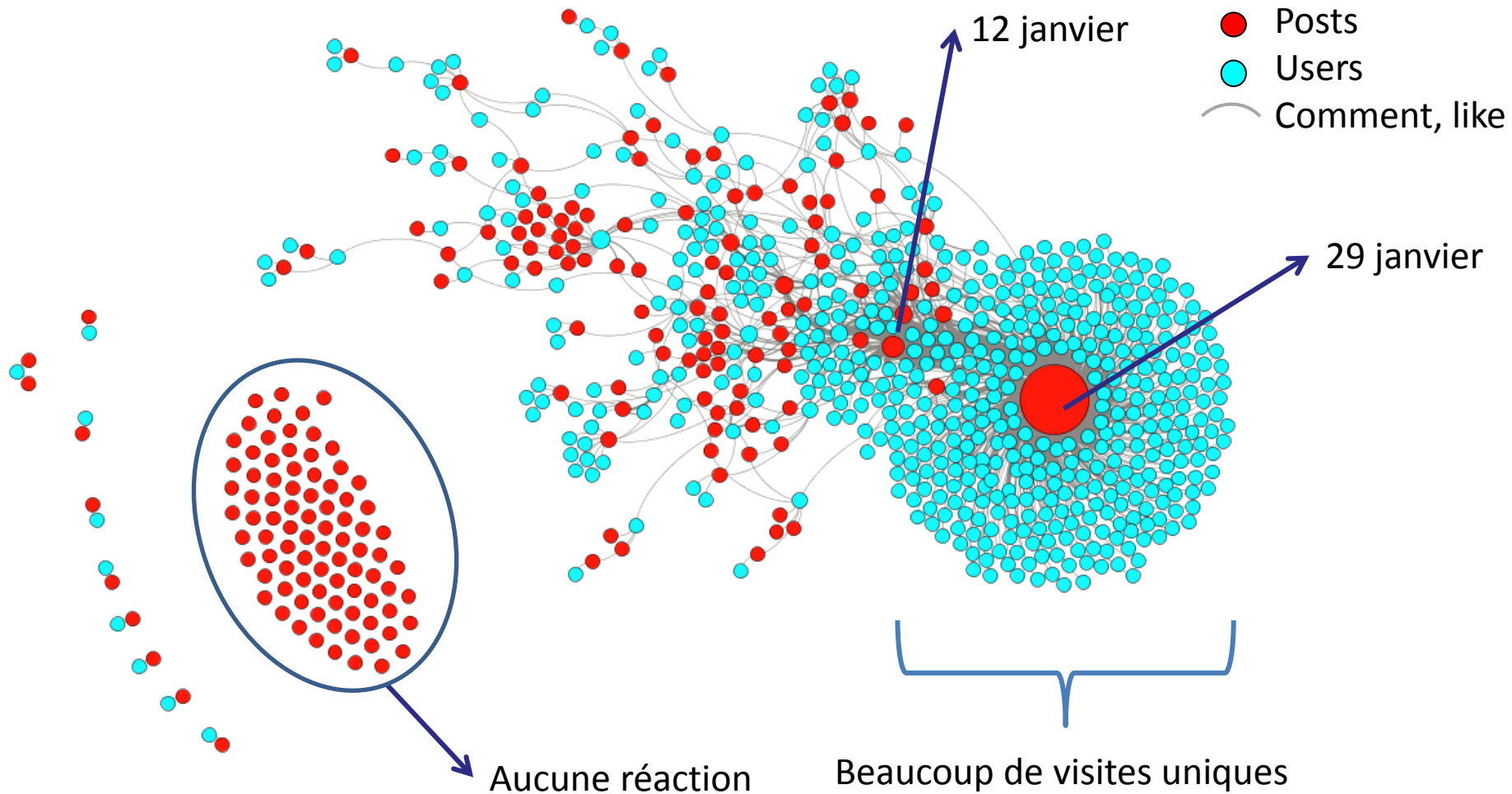
- Donne une information générale, mais technique pas encore très mature en français
- Pas utilisable plus « localement »



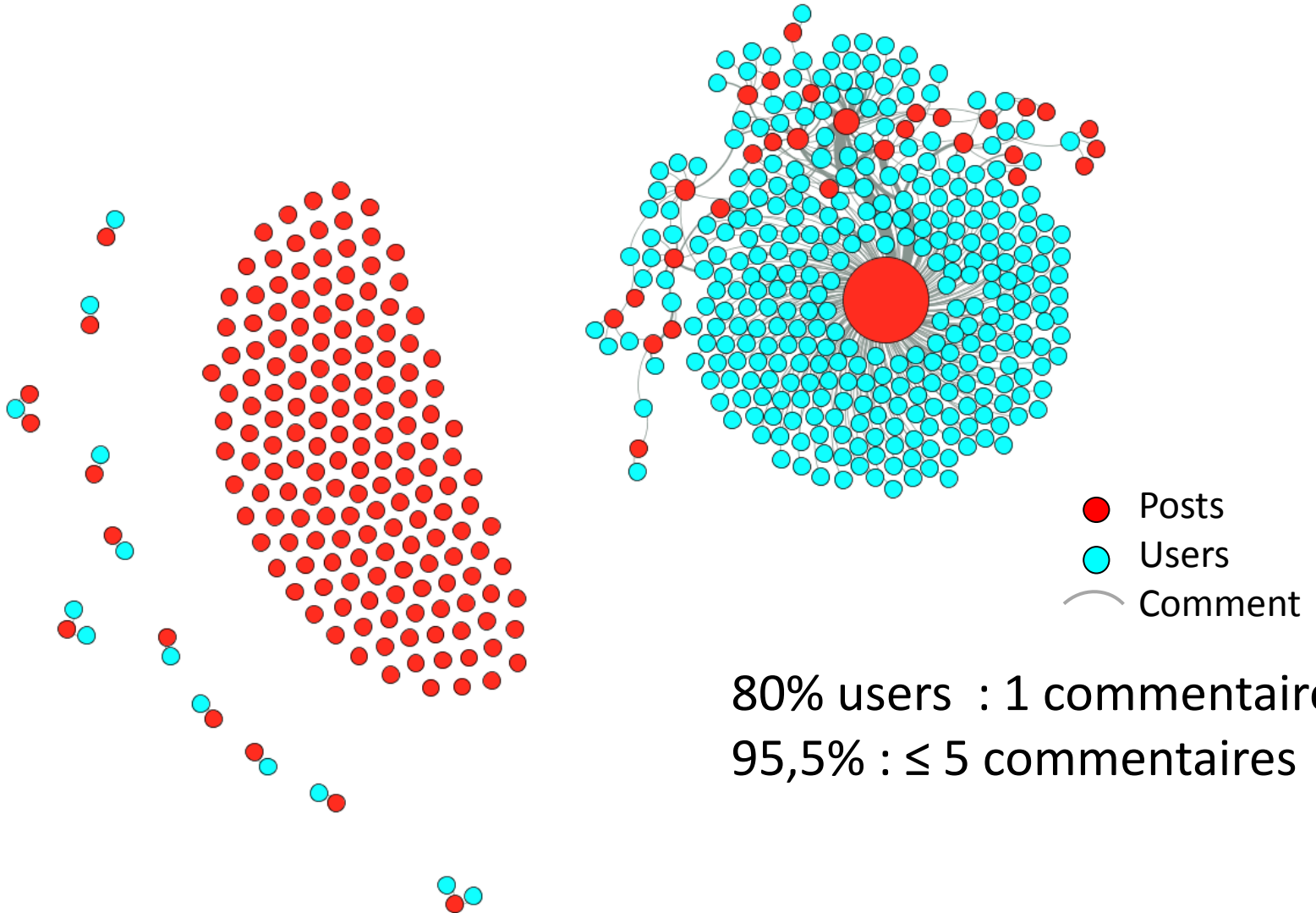
Page Facebook

NETWORK ANALYTICS

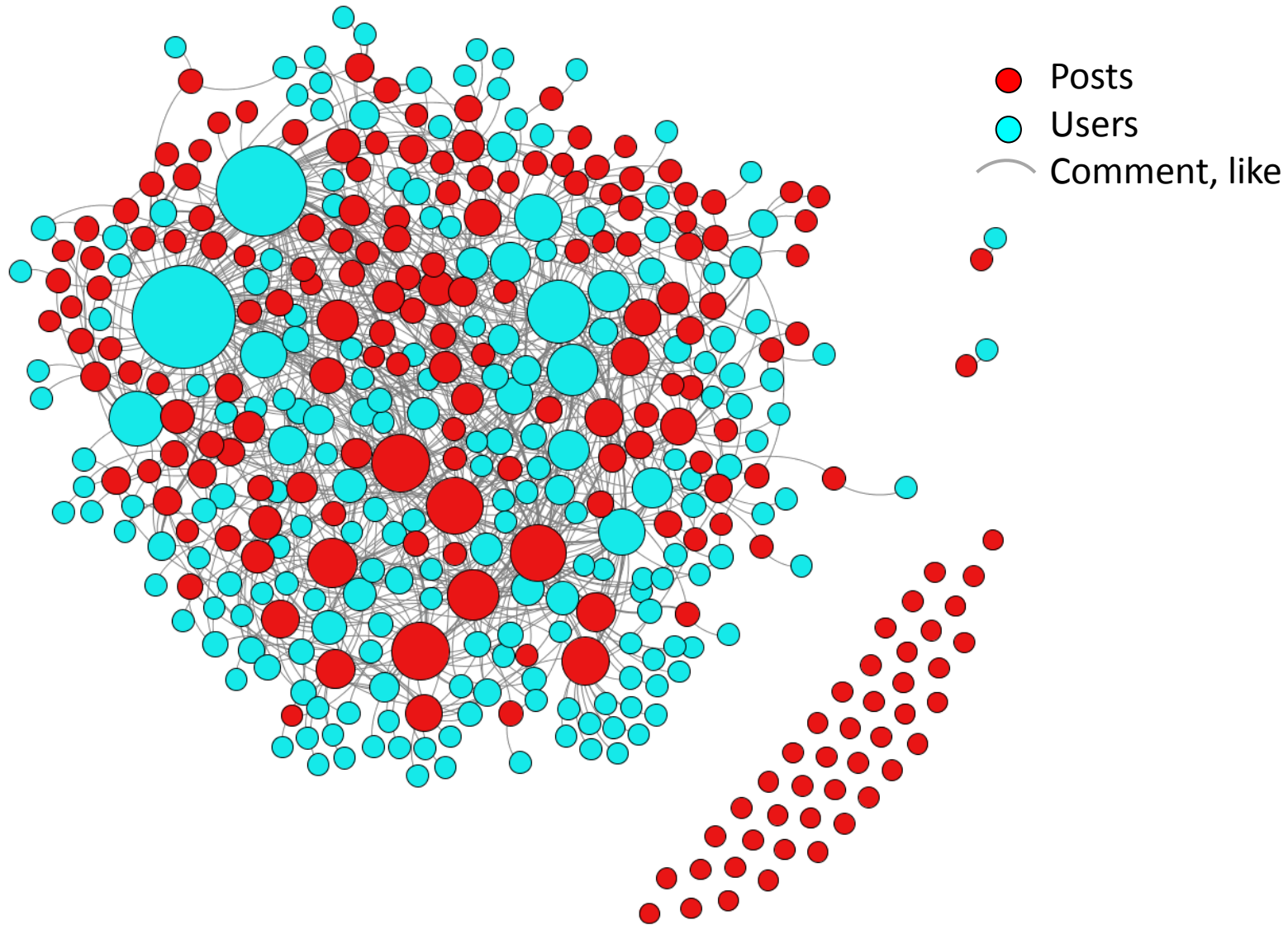
Network analytics : acteurs



Networks analytics : commentaires



Network analytics : Smals (acteurs)



Network analytics

- On pourrait aller plus loin, avec ce qui est accessible sur la **version Web** mais pas avec l'API
- On ne pourra trouver **qu'une partie de l'information**
- Exemple : **Relations d'amitié** entre les « acteurs » ou les « likers » de la page



Net

Approximation
(25% connex.
manquantes)

- Liste privée
- Liste publique
- Amitié



 Freelancer at Animal rights	 [Redacted Name]
 Works at Legum...	 CFPPA Nancy F...
 Animal Liberation Front	 Sy... T... Poitou-Charentes, France
 [Redacted Name]	 My... Works at Protection animale
 Les Dames de Marie	 Ensemble, battons nous ! Joy... Works at Animal rights
 Bourges	 V... Président Fondateur GIAPA
 [Redacted Name]	 [Redacted Name] Animal rights



Network analytics : liens d'amitié

Ap

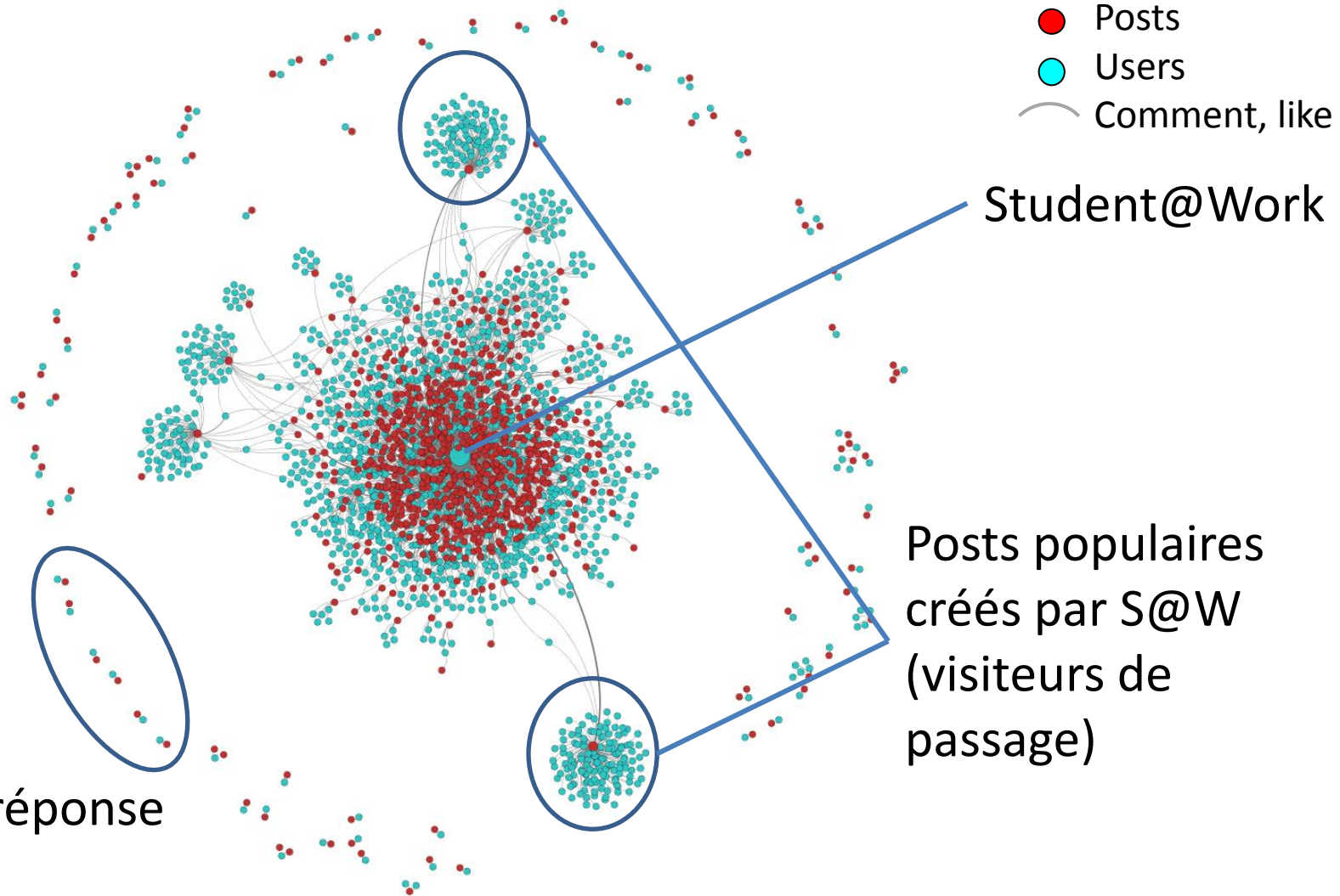


Autre exemple : Student@Work

- Student@Work : application de l'ONSS pour les étudiants jobistes
- Page Facebook sur laquelle tout le monde peut créer un post
- Sert à la fois d'outil de communication et de « service desk » pour les problèmes techniques

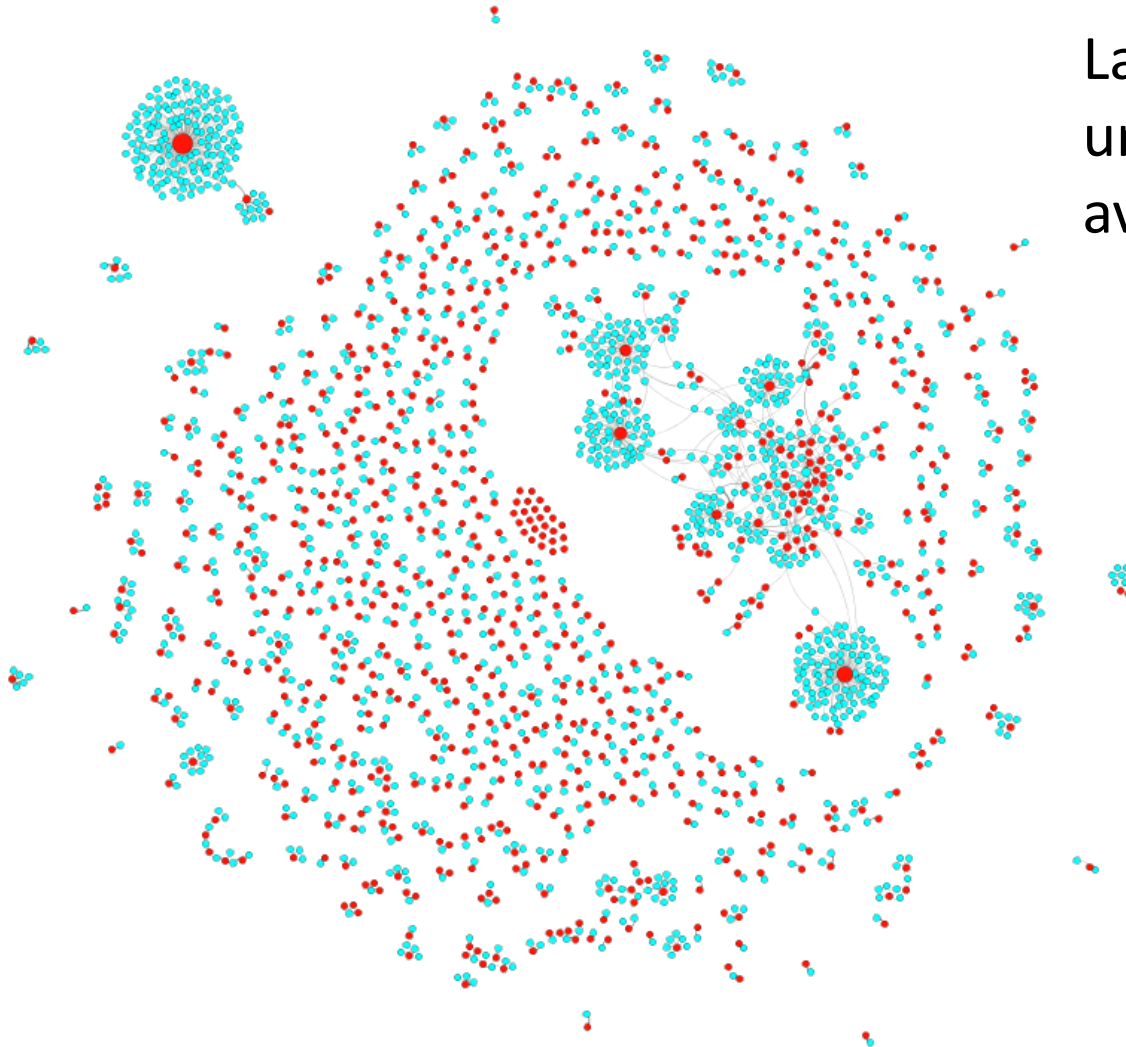


Student@Work : toutes les interactions



Student@Work : sans S@W

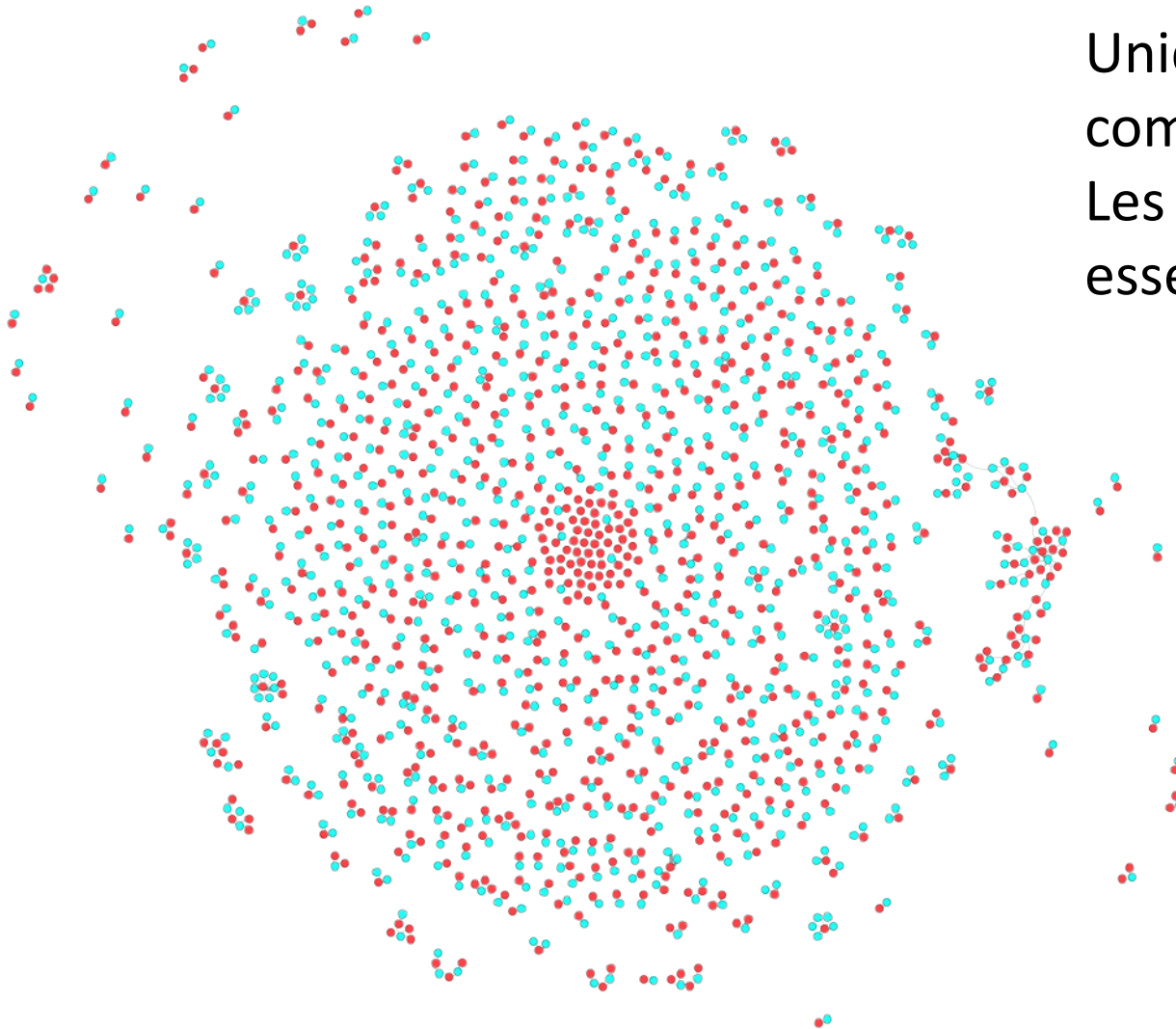
La plupart des interactions :
un user (qui ne revient pas)
avec un post



 Posts
 Utilisateurs



Student@Work : posts et commentaires



Uniquement création et commentaires (pas likes)
Les grappes : composées essentiellement de likes



Network analytics: lessons learned

- La grande majorité des commentaires viennent de gens de passage...
- ... et est condensée sur un seul post
- Partie non négligeable des utilisateurs très sensible à la cause animale → pas représentatif de la population en général



Page Facebook : l'essentiel

Analyse de fréquence :
identification des pics
de réaction et des
événements liés

Topic extraction :
De quoi parle-t-on
+ évolution

Sentiment analysis :
polarité du message.
OK pour analyse
macro uniquement

Network analytics :
qui réagit, à quoi



Page Facebook : les outils

Pour les gestionnaires de page :

- Facebook
- Hootsuite (subset+€)

Pour les plus branchés : API

Pour analyser d'autres pages :

- Engagor (15j+€)
- SimplyMeasured (15j+€)
- Quintly (15j+€)

Outils classiques :
orientés
« performances »,
plus qu'analyse de
problèmes



Questions ?



Pause !



A hand in a light blue shirt holding a white marker, pointing towards a network diagram. The diagram consists of several circular icons, each containing a white silhouette of a person, connected by thin blue lines. The background is a light blue gradient.

Use case 3 : Network Analytics

Table des matières

Introduction

Page Facebook

Mots-clé

Use case 3 :
Network
Analytics

Introduction

Structure

Centralité

Inférence

Reconstruction

Évolution

Conclusions



Network Analytics

INTRODUCTION

Contexte

- Les médias sociaux (Facebook en particulier) regorgent **d'informations** sur leurs utilisateurs :
 - Des informations qu'ils fournissent **eux-mêmes**
 - Des informations fournies par leurs « **amis** »
 - Des informations « **implicites** » (induites ou structurelles)
- On ne peut pas totalement **contrôler** cette information :
 - Parce qu'elle est **divulguée par d'autres**
 - Parce qu'elle est **implicite**, et qu'on peut la **reconstruire** à partir d'autres éléments
 - Parce qu'on en est en général **pas conscient**



Intérêt

- Conscientiser sur l'information personnelle disponible
- Informations très précieuses pour le hacking par « social engineering », l'usurpation d'identité
- Mieux connaître sa communauté (Facebook, Twitter...)
- Améliorer la lutte contre la fraude et la criminalité

Nous utiliserons des informations publiques, sans aucune technique de hacking ou phishing

Avec certaines techniques, il est facile de « dépasser les limites » (P/R vie privée ou EULA Facebook)... à utiliser avec précaution !



Médias sociaux et délits

- Les médias sociaux de plus en plus souvent utilisés dans des **enquêtes** (policière, fiscales...)
- Rarement un **preuve**, aide à **orienter une enquête**
- Techniques similaires basées sur les **réseaux** (liens entre entreprises, clients...) déjà utilisées (Fisc, ONSS)
- Exemples :
 - Vol de billet à la Banque de France (sept 2014) : publication de photo de vacances, nouvelle cuisine...
 - De nombreux délinquants publient eux-mêmes les photos de leur méfaits
- Facebook ne fournit pas toujours des données à la justice (Belgique : +/- 30% de refus) !

<https://govtrequests.facebook.com/>



Médias sociaux et fraude

- Hypothèse : si on identifie quelques fraudeurs/suspects dans un groupe « **fortement connecté** », s'intéresse à l'ensemble du groupe
- Parmi une population de fraudeurs, identifier les **plus influents**, les leaders, les « **connecteurs** »
- Recherche d'un « chemin » entre deux personnes (ami, ami d'ami...)
- Recherche des liens d'amitié entre un groupe de personnes
 - Identifiés hors Facebook
 - Membre d'un groupe, fan d'une page



Fil rouge


- Pour la suite, on se place dans la peau d'un « **attaquant** » visant la « **cible** » vandy.berthen :
 - N'a **jamais publié de photo** (hormis le profil), ne s'y est **jamais taggué** (mais l'a été par d'autres)
 - A « sécurisé » son compte de façon à **cacher ses photos, son mur, sa liste d'amis** aux « inconnus »
 - N'a rien mis de **visible sur son profil** (parcours scolaire...)
- On va reconstituer :
 - Ses groupes d'amis
 - Ses amis proches, sa famille
 - Son parcours scolaire, (une partie de) ses loisirs
- À partir d'un compte « **bidon** », sans aucun ami



Préambule : « Graph Search »

- « Graph Search » : (nouvel) outil de Facebook permettant de faire des **recherches** (version US)
- Par exemple :
 - « Photos of XXX »
 - « Photos commented on by XXX »
 - « XXX's friends who work at Smals »
 - « People tagged in XXX's photos »
 - « XXX's events », « Events that were attended by XXX »
 - « People who like Hitler Adolph and work at Smals »
- Trouve des **informations déjà accessibles**, mais noyées dans la masse jusqu'ici



 **Vandy Smals**
Berten Smals
[Edit Profile](#)

- Welcome
- News Feed
- Messages
- Events
- Find Friends

- DEVELOPER
- Manage Apps
 - Insights
 - BeTizenConnect
 - SmalsResearch

- APPS
- Games
 - Socialbakers.com
 - netvizz
 - Photos
 - Music

 Update Status  Add Photos/Video

What's on your mind?

Post



Agence alimentaire (AFSCA)

September 5 at 10:23am · 🌐

Communiqué de Carrefour Market Groupe Mestdagh :
Rappel de salade thon et de salade jambon.
Problématique : présence de Listeria monocytogenes.

[See Translation](#)



AFSCA - Communiqué de Carrefour
Market Groupe Mestdagh

AFSCA.BE

TRENDING

-  **Perth:** 1st pe
Apple Store
on sidewalk
 -  **Ron Washin**
manager rev
cheating on
 -  **Deadpool:** M
favorite hits 1
- [See More](#)

RECOMMENDED

-  **Netw**
73,35

-  **The E**
6,453

-  **Dinne**
14,37

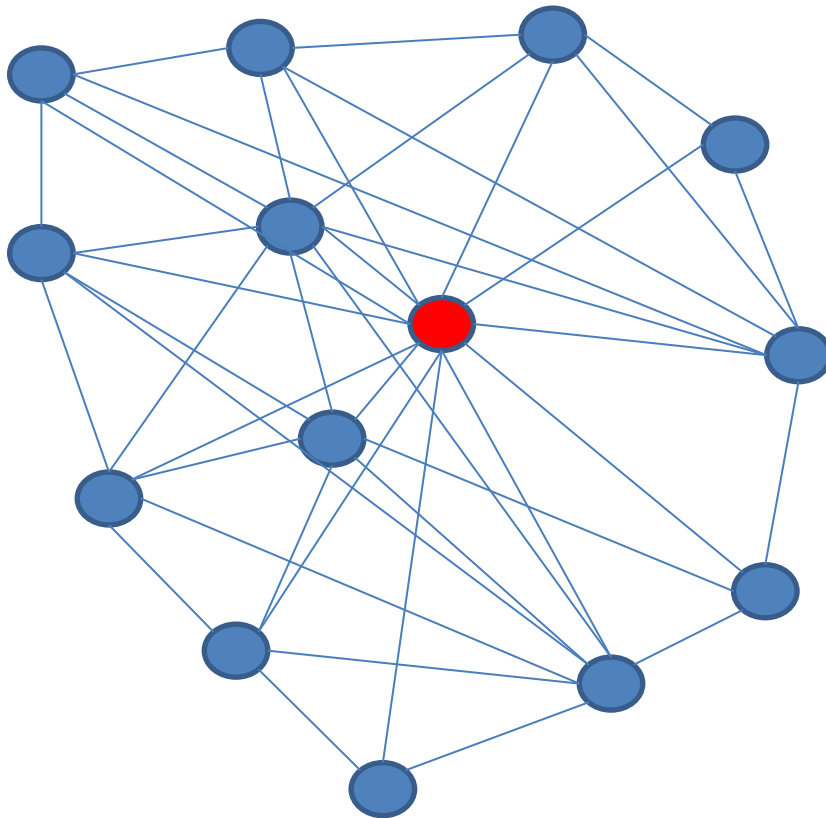

 Chat

Liste d'amis : très instructif

- Liste d'amis : ressource très précieuse
- Permet d'identifier les **groupes sociaux** : famille, travail, étude, loisir...
- Permet identifier les **liens « forts »**, ainsi que **l'influence/la popularité** (notion de centralité)
- Même masquée, la liste d'amis peut être **reconstruite** en grande partie
- On peut étudier les « structures » autour d'une **personne**, d'une **page** ou d'un **groupe**



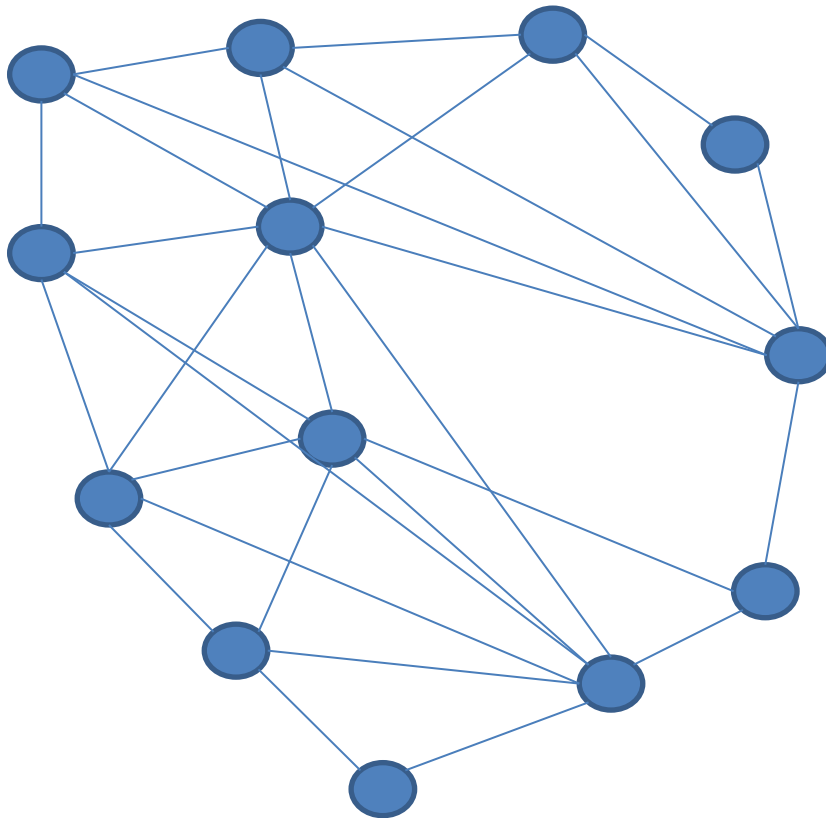
Réseau personnel



- Réseau de niveau 1.5 : la « cible », ses amis et les connexions entre eux
- Pas les amis de amis (non-mutuels)
- Facebook : relation symétrique (\neq Twitter)
- Via API : uniquement son propre réseau



Réseau personnel



Suite :

- Structure
- Centralité
- Inférence
- Reconstruction
- Évolution

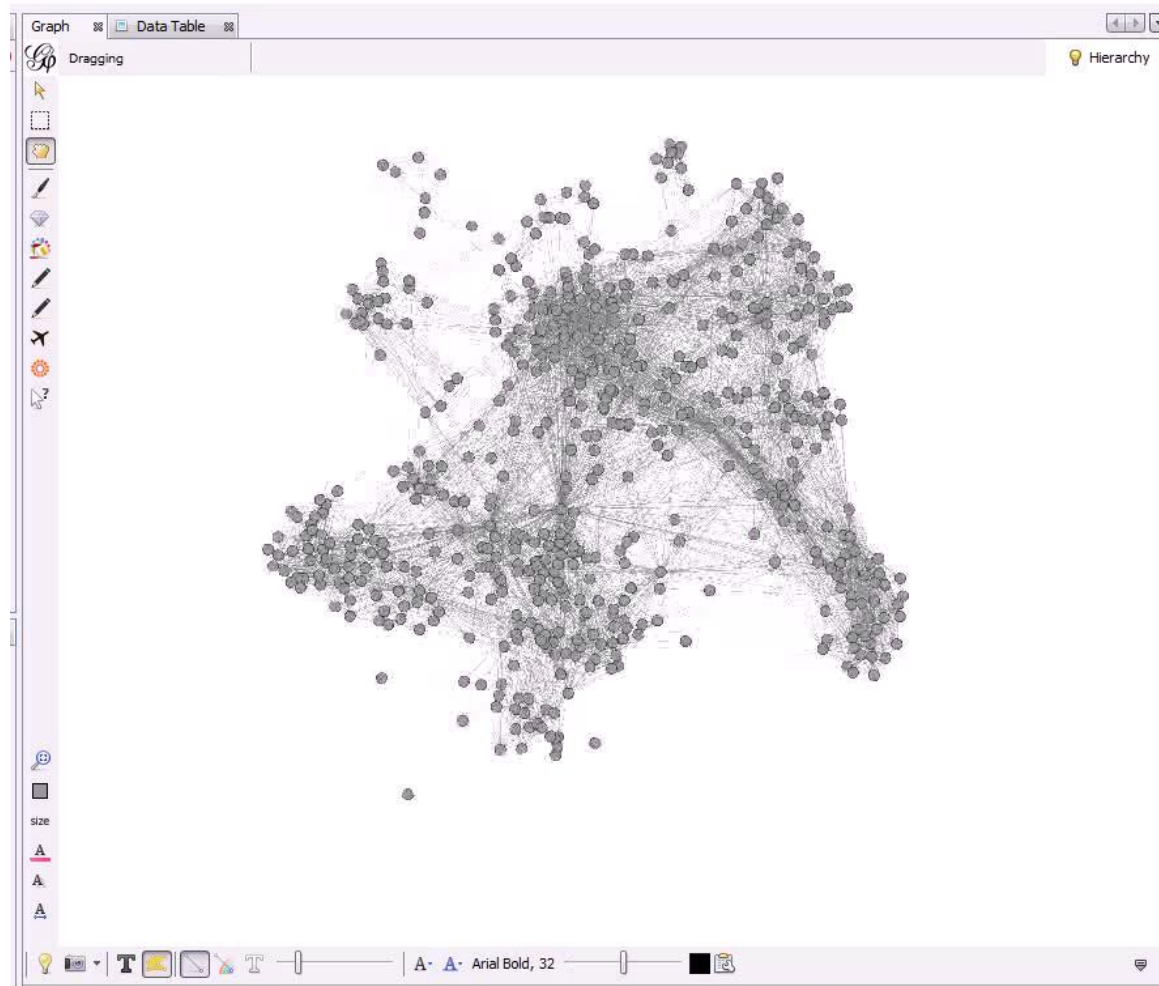


Outils

- On va se servir de l'outil « **Gephi** » de manipulation/visualisation de **graphes**
- 1^{er} étape : importation d'un compte personnel via l'API (ou une application comme « **Netvizz** »)
- Une partie de la suite se base sur un « crawling » de la version Web



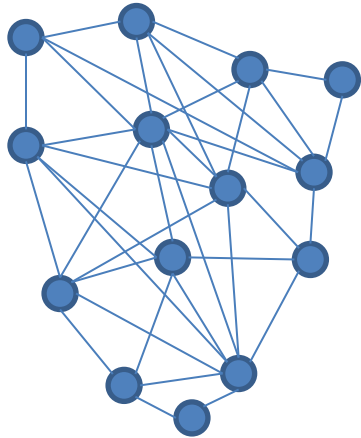
Gephi



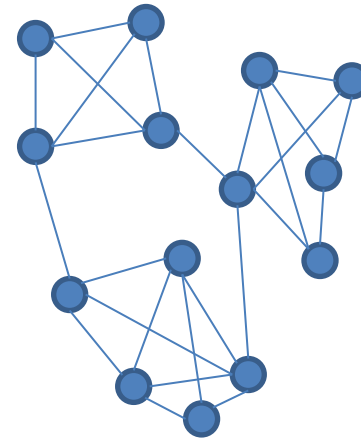
Network Analytics

STRUCTURE

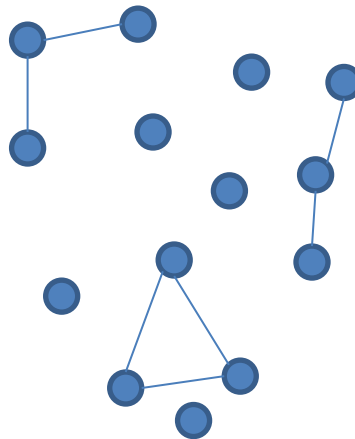
Structures de communauté



Communauté « connexe »



Agglomérat de petits groupes



Communauté peu « connexe »

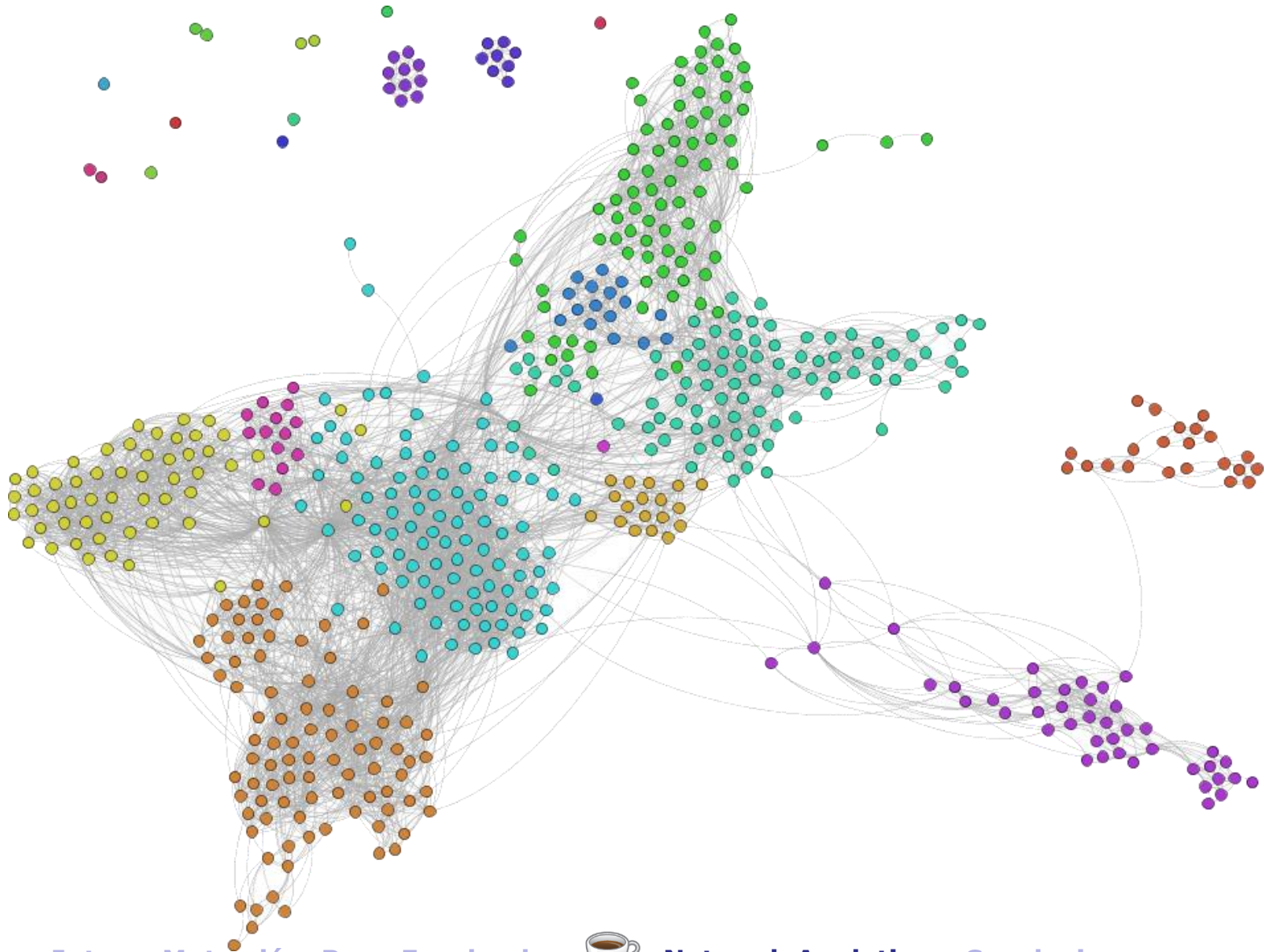


Structures de communauté

- À partir d'un graphe, on peut identifier des « **partitions** » (ou **communauté**, ou *cluster*) : sous-ensemble particulièrement connecté
- Correspond en général à des « **groupes sociaux** » dans la réalité : famille, amis des études, d'un loisir, ...
- Pour comprendre à quoi correspondent ces groupes, il faudra faire de **l'inférence**



Partitions



Network Analytics

CENTRALITÉ

Mesure de centralité

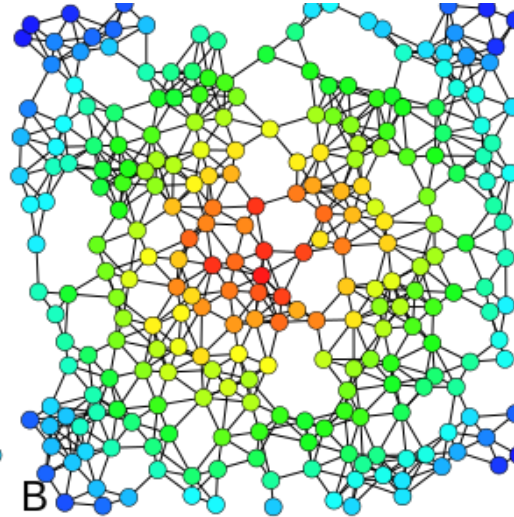
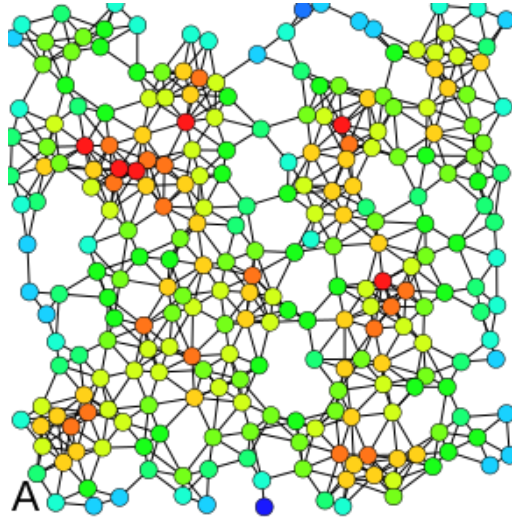
- Différentes façons de mesurer l'importance, la popularité, l'influence, la centralité... d'un acteur (personne, page, post, tweet...)
- Une personne peut être importante « dans le monde », mais pas dans un groupe social particulier... et vice-versa
- Permet aussi d'identifier les « connecteurs » entre plusieurs groupes sociaux



Mesure de centralité

Degree centrality :

#voisins, globalement (#likes, friends, followers...) ou localement (#mutual friends)

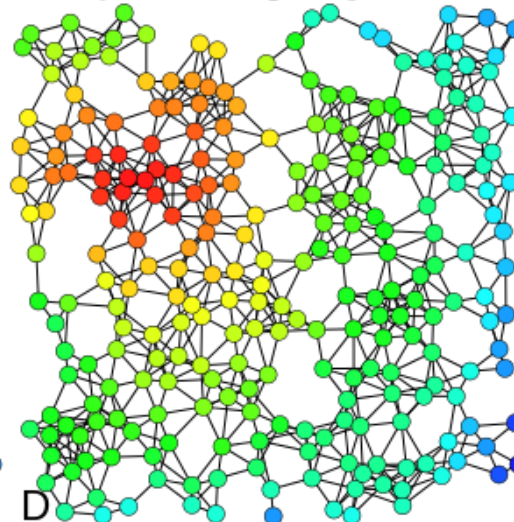
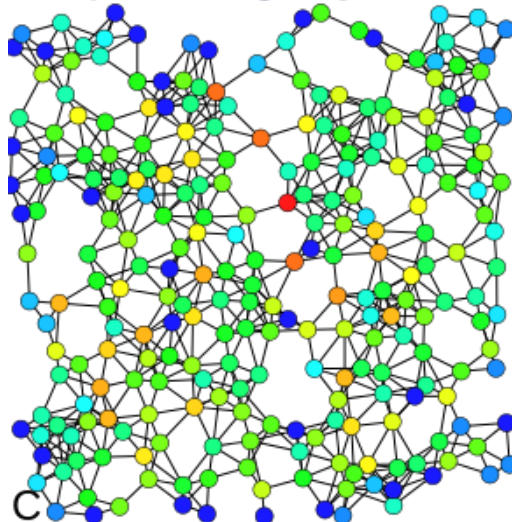


Closeness centrality :

Inverse de la distance moyenne à tous les autres nœuds

Betweenness centrality :

importance en tant qu'intermédiaire (# de shortest path passant par le nœud)

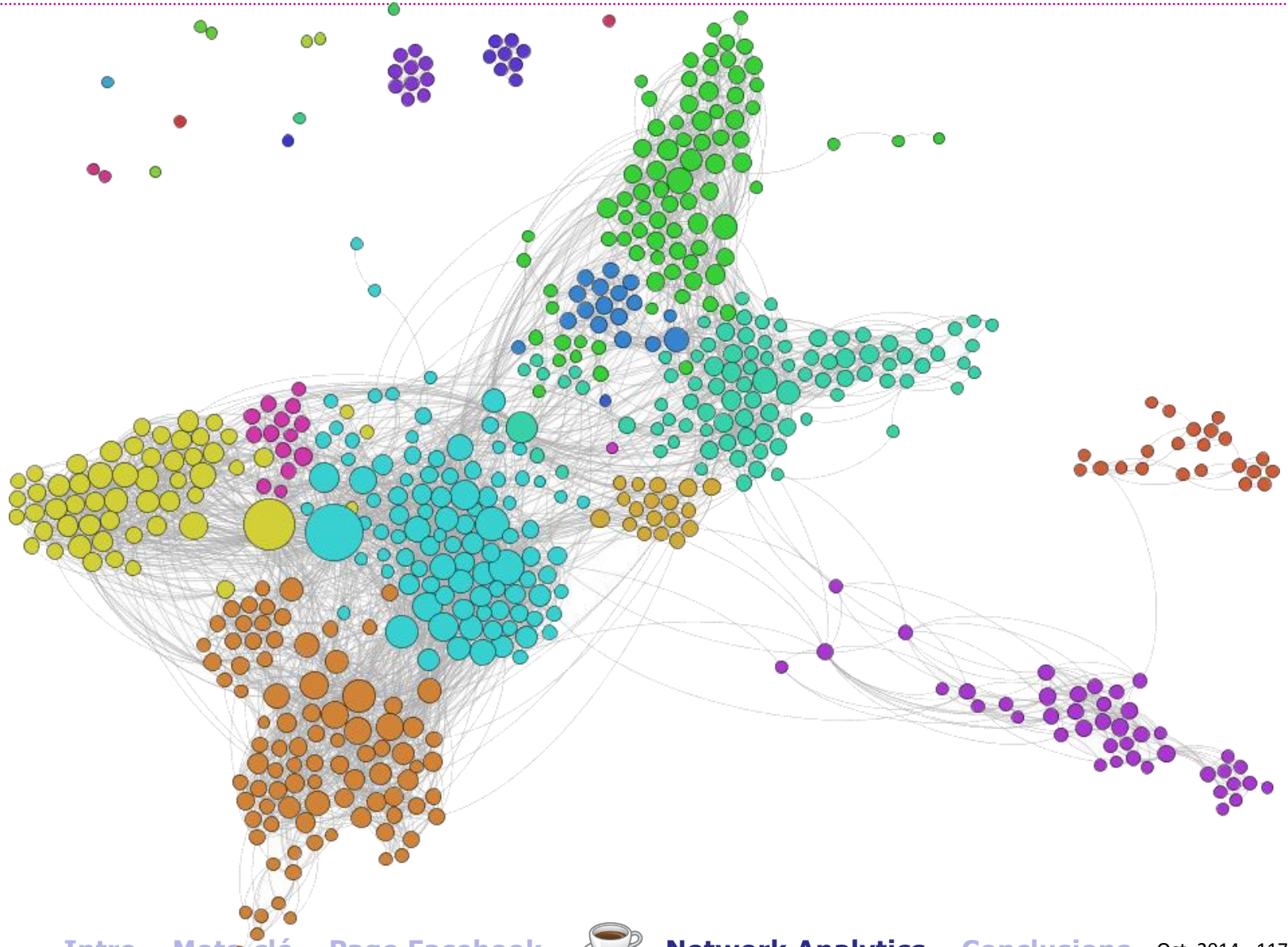


Eigenvector centrality :

PageRank de Google



Mesure de centralité



Network Analytics

INFÉRENCE

Inférence

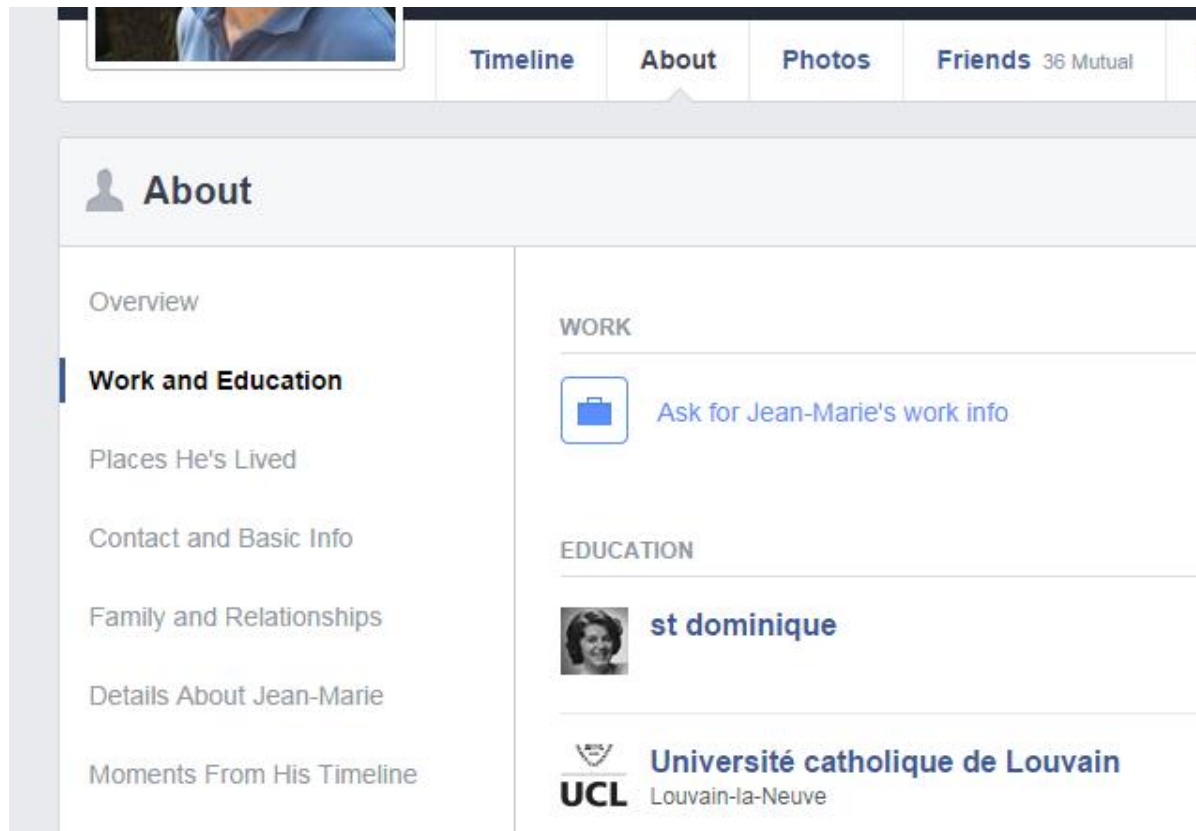
- Idée : repérer au sein d'une « partition » des informations (publiques) similaires pour **caractériser la partition**
- Puisque la « cible » **fait partie de toutes ses partitions**, elle possède (probablement) la **même caractéristique**
- Exemple :
 - 2% des « amis » travaillent (« publiquement ») chez « Smals » → pas significatif
 - Partition « P » : 40 % travaillant (publiquement cachée chez les autres), l'absence de travail est probable
 - Partition « P » : probablement pas de travail chez « Smals » ... comme la partition « P »

- Nécessite des techniques avancées, hors API, pas accessibles en quelques clics !
- On sort du cadre des EULA de Facebook !



Inférence (About / Work & Education)

- Section « About / Work & Education » : publique dans +/- 40%
- Pas accessible via l'API, mais sur la version web



Timeline About Photos Friends 36 Mutual

About

Overview

Work and Education

Places He's Lived

Contact and Basic Info

Family and Relationships

Details About Jean-Marie

Moments From His Timeline

WORK

Ask for Jean-Marie's work info







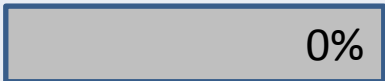



EDUCATION

st dominique

Université catholique de Louvain
UCL Louvain-la-Neuve



Inférence (About / Work & Education)

	Saint-Dominique	ULB	Taille partition	% publique
Total	 12%	 12%	562	38%
Partition 1	 4%	 3%	142	53%
Partition 2	 53%	 25%	83	46%
Partition 3	 0%	 57%	42	50%
Partition 4	 8%	 5%	95	37%



Inférence (Groupes)

- Idem pour les groupes : appartenance à un groupe (non secret) toujours publique
- Graph search : « Vandy Berten's groups »
- Exemple : présence de « Croix-Rouge », « CRB » ou « CR » dans les noms de groupe :
 - En général : 13%
 - 3 partitions avec 28%, 35% et 50%
 - Toutes les autres : 0-1%
- Alternative possible : se baser sur les « likes »



Inférence (Famille)

- En regardant les **noms de famille**, on peut parfois identifier le réseau « familial »
- En général, peu de personnes ont **le même nom** de famille dans un groupe social ... sauf s'il s'agit d'une famille !
- Si le top 3 des noms de famille d'un groupe comprend une proportion élevée, c'est sans doute le groupe familial



Network Analytics

RECONSTRUCTION

Reconstruction

- On pense parfois pouvoir **contrôler** ce qui se trouve à **son propos** sur Facebook
- Problème : pas de contrôle sur **les autres**
- Liste d'amis : deux méthodes pour reconstruire une « liste cachée »
 - Créer un profil « bidon », inviter la cible, puis regarder « People You May Know » (PYMK)
 - Grâce aux « mutuals friends »





Vandy Smals
Berten Smals
Edit Profile

Update Status

Add Photos/Video

What's on your mind?

Post

Welcome

News Feed

Messages

Events

Find Friends

DEVELOPER

More

Manage Apps

Insights

BeTizenConnect

SmalsResearch

APPS

More

Games

netvizz

Socialbakers.com

Photos



No posts to show

TRENDING

Foster F
poultry c
with goll

Zombiel
comedy

Cung Le
positive
receives

See More

RECOMMEND



Chat

Reconstruction

- Sur la version **Web**, on voit **plus de choses** qu'avec **l'API**
- Cas « simple » : liste d'amis visible
- Sinon : on apparait dans la liste d'amis de ses (certains de ses) amis !
- Dans quelques conditions, on peut obtenir les « **mutual friends** » de deux profils
- En partant de quelques amis connus, on peut **reconstituer une grande partie du réseau** d'un compte, en quelques minutes, sans en être ami



 **Vandy Smals**
Berten Smals
[Edit Profile](#)

- Welcome
- News Feed
- Messages
- Events
- Find Friends

DEVELOPER

- Manage Apps
- Insights
- BeTizenConnect
- SmalsResearch

APPS

- Games
- netvizz
- Socialbakers.com
- Photos

 Update Status  Add Photos/Video

What's on your mind?

Post

 **Agence alimentaire (AFSCA)**
16 hrs · 

Communiqué de Quality Nuts :
Rappel de "Céréale sel".
Problématique : Présence possible d'un allergène (soja).

[See Translation](#)



AFSCA - Communiqué de Quality Nuts

TRENDING

-  **Billy Cor**
final pho
Williams
-  **Play Sta**
video ga
release
-  **Bryan S**
latest in
franchis
- See More

RECOMMEND

-  C
2
-  T
L
1
-  F
2

 Chat

Reconstruction



- Liste d'amis de « T » privée
- Liste d'amis de « A » publique
- La liste d'amis de « T » est vide, mais sur la liste de « A », on voit « T »



 **VandySmals**
Berten Smals
[Edit Profile](#)

- Welcome
- News Feed
- Messages
- Events
- Find Friends

DEVELOPER

- Manage Apps
- Insights
- BeTizenConnect
- SmalsResearch

APPS

- Games
- Socialbakers.com
- netvizz
- Photos

 Update Status  Add Photos/Video

What's on your mind?

Post

 **Agence alimentaire (AFSCA)**
September 5 at 10:23am · 🌐

Communiqué de Carrefour Market Groupe Mestdagh :
Rappel de salade thon et de salade jambon.
Problématique : présence de Listeria monocytogenes.

[See Translation](#)



**AFSCA - Communiqué de Carrefour
Market Groupe Mestdagh**

TRENDING

-  Perth: 1
Apple st
sidewal
-  Ig Nobel
announ
and unu
-  JetBlue
slides a
emerge
- [See More](#)

RECOMMEND

-  Cooksongold
-  SEJ
- 

 Chat

Reconstruction



- On soupçonne **B** d'être proche de **T** (ami direct ou non)
- On demande à FB : « Mutual friends of **T** and **B** » ... il répond **A** → **A** et **T** sont donc amis
- On réessaye avec **A** : « Mutual friends of **T** and **A** »...

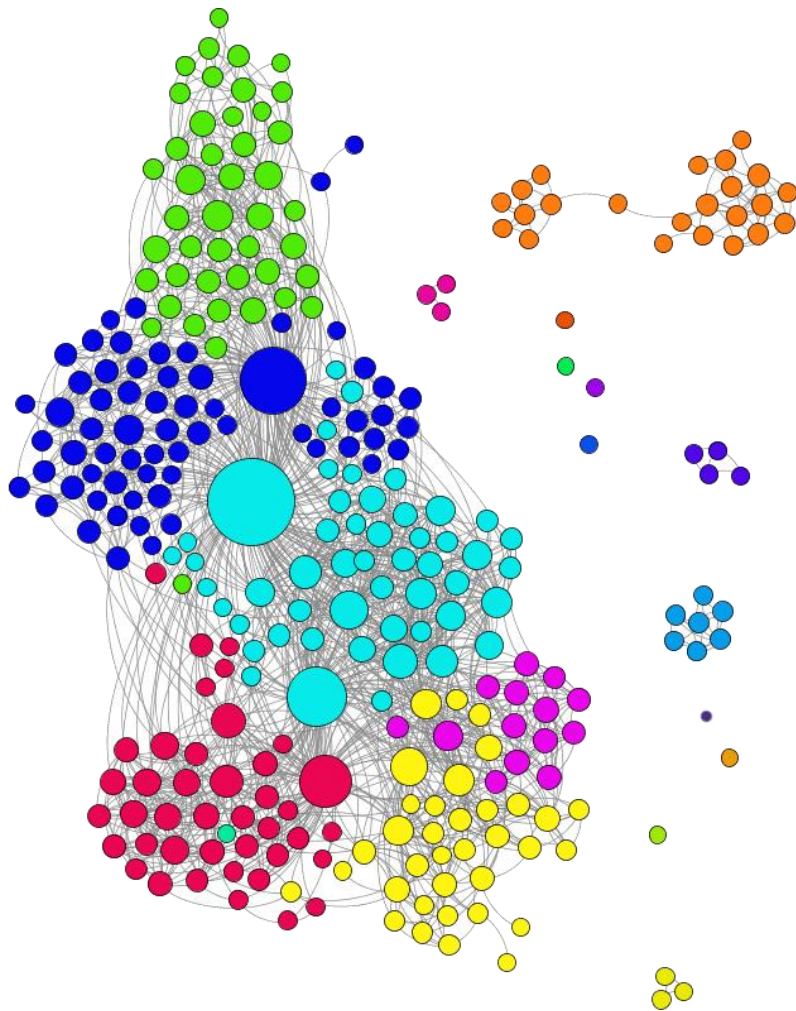


Reconstruction de vandy.berten

- Depuis ce compte (via API) :
561 nœuds (298 pub, 53%), 5059 connexions
- Si liste privée, « People tagged in... » + « PYMK » :
296 nœuds (52,9%, 99% de ce qui est public !)
1638 connexions (32%)
- Si liste publique, par « crawling » :
561 nœuds (100%), 4136 connexions (81%)

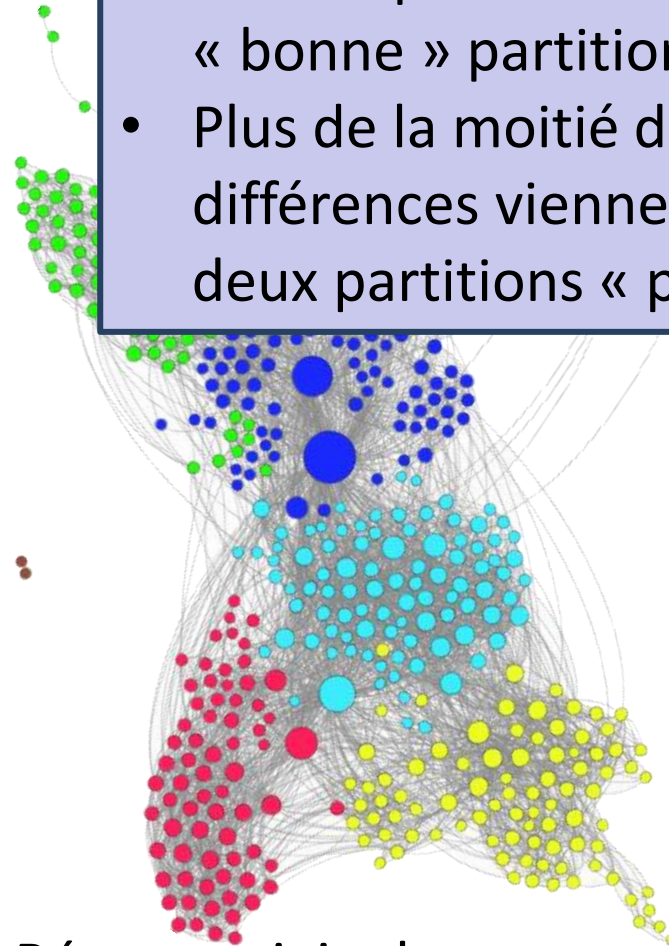


Reconstruction



Réseau reconstruit

- Entre 85 et 100% des nœuds ont été placés dans la « bonne » partition
- Plus de la moitié des différences viennent de deux partitions « proches »



Réseau original

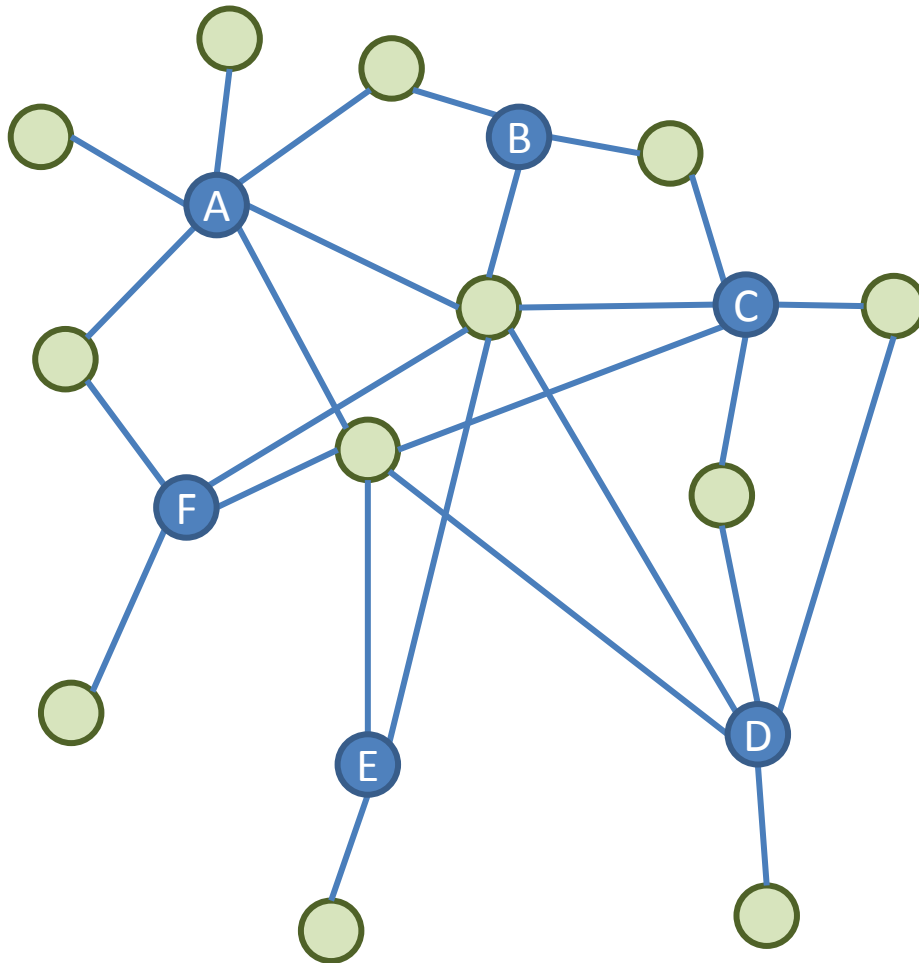




Reconstruction : deviner

- Peut-on « suspecter » d'autres personnes d'être amies avec la cible ?
- Hypothèse : deux personnes qui ont **beaucoup d'amis communs** ont des chances d'être également amis
- On recherche les amis d'amis fréquents
- Si beaucoup d'amis de T disent être amis avec A (même si A et T cachent leur liste), il y a des chances que A et T se connaissent



Reconstruction : deviner



-  Amis de la « cible »
-  Amis d'amis



Reconstruction : deviner

- Sur vandy.berten :

Seuil	Amis FB (cachés)	Connus (hors FB)	Inconnu/ vague	Total
15	61	8	3	72
10	120	32	21	173

- Impossible de le savoir en général !
- Autre possibilité : « Friendship page »
(<https://www.facebook.com/user1?and=user2>)

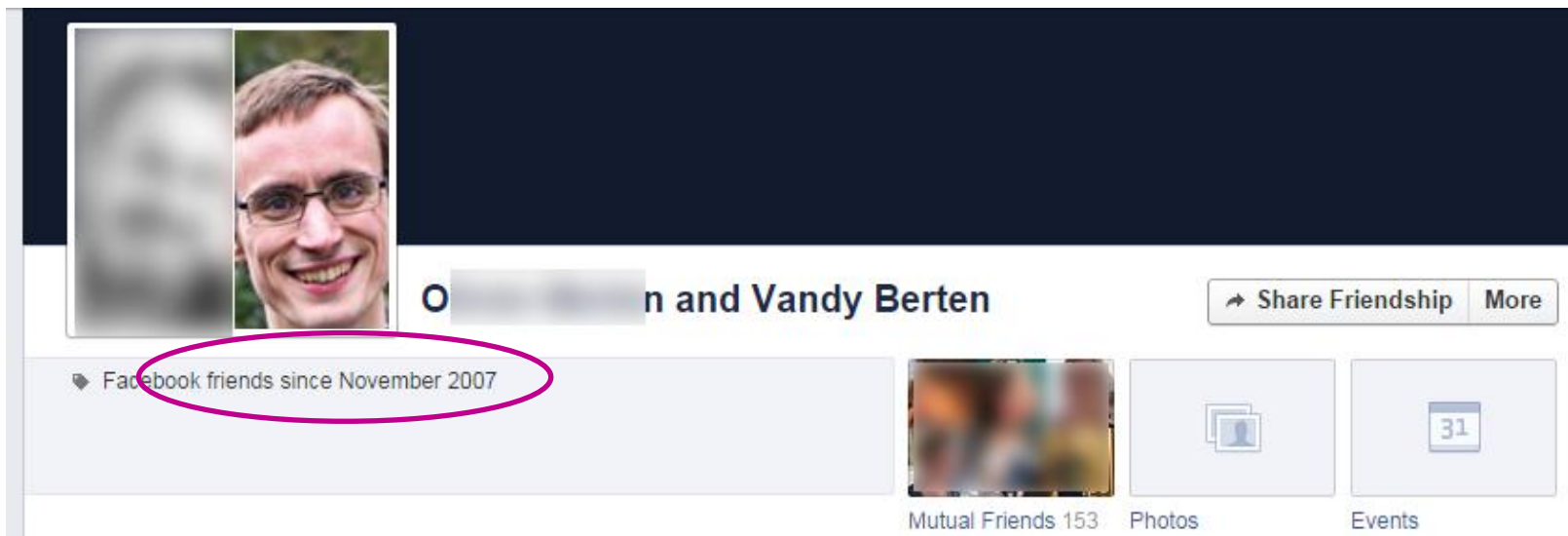


Network Analytics

ÉVOLUTION

Évolution

- On peut voir depuis quand deux personnes sont « Facebook friends »
- www.facebook.com/o*****n?and=vandy.berten :



- La dynamique des réseaux est très souvent instructive



Et alors ?

- On pourrait appliquer la même technique pour deviner la **religion**, les **préférences politiques** ou l'**orientation sexuelle** de quelqu'un
- Sans doute pas un problème chez nous ... pas anodin si on a des amis vivant dans des régimes plus « sévères »
- Même chez nous : un « cambrioleur » pourrait se servir des profils de vos amis pour savoir que vous n'êtes pas chez vous !
- Usurpation d'identité facilitée



Amélioration du phishing ?

Autre ami
ayant bcp
d'amis en
commun
avec YYY

Bonjour XXX
Ce week-end j'ai
discuté avec YYY et
ZZZ à la fête de AAA,
qui m'ont dit que tu
serais intéressé par
cette application/ce
site web ...

→ Votre nom

Un ami avec
une haute
centralité de
degré

Grâce à des photos ou
événements trouvés
par « Graph Search »



Network analytics: l'essentiel

Grande quantité
d'informations sur
des individus et sur
leur relations

Même avec une
« cible » méfiante

Petit exemple des
possibilités
montrées ici

Peut être long, mais
pas de matériel
complexe/puissant



Network analytics: les outils

Extraction compte :

- Netvizz
- NameGenWeb
- Twitter : NodeXL

Analyse de son compte :

- wolframalpha.com/facebook
- Touchgraph

Visualisation, partitions,
centralité... : Gephi

Pour les plus branchés :

- API Facebook
- Web crawling :
selenium (Python)





Conclusions

Conclusions

Les citoyens s'expriment sur les médias sociaux, il est nécessaire de les suivre

La quantité de données rend l'information difficile à comprendre... mais des outils sont disponibles

Prise de connaissance d'un problème/événement plus rapide qu'avec les médias « traditionnels »

Souvent, méta-information plus intéressante que l'info elle-même





Vandy Berten
02/787.57.32
vandy.berten@smals.be



More on Smals Research :
Website : www.smals.be
Blog : www.smalsresearch.be
Twitter : [@SmalsResearch](https://twitter.com/SmalsResearch)

Blog (www.smalsresearch.be/author/berten/)

- Contacter un citoyen sur Facebook ?
- La vie privée selon Facebook
- Ce qu'un réseau social peut nous apprendre
- Facebook : peut-on vraiment cacher sa liste d'amis ? [6/11]

