Recherches Internet - UE

—

[redacted]

Perspective de l'UE sur la collecte d'informations sur Internet

OSint ou RoSO?

Quelques éléments

Pour savoir de quoi on parle....

- OSIF
- Pas uniquement internet (presse, édition...)
- Information librement accessible
- Pas de stratagème pour l'obtenir

Applications

En matière de terrorisme ou de crimes contre l'humanité...

- Le Mandat d'arrêt CPI Werfalli (Libye) - Bellingcat

- Exemple en juridiction - Dossier Sheddadi.

Difficultés

Il y en a...

- Sécurité (Virus, Manipulation...)
- Juridiques (qques jurisprudences...)
- Technique (ToS, volumétrie des data)
- Ethique (Fichiers, data de tiers...)

Etat des lieux dans l'UE

Régime Juridique

Trois grandes tendances qui cohabitent au sein de l'UE

- Régime totalement contrôlé (ex : Portugal)
- Régime différenciation ponctuel/habituel (ex Belgique)
- Régime "Open-Bar" (France, Finlande, etc...)

Action de l'UE

Rôle normatif et de protection

Deux exemples parmi d'autres....

- RGPD (GDPR)(1016/679)

  Conséquences sur l'OSint (databreach, whois....)

- E-evidence (*Cloud Act EU*)

  Conséquences notamment sur l'accès au contenu des réseaux sociaux...

Rôle Formation et coopération

Trois exemples...

- CEPOL (Formation)
- EUROPOL (Opérationnel)
- ENLETS (Réseau prospective, échanges et bonnes pratiques)

En conclusion...

Contexte important

Internet est notre assistant personnel...

- Accès à la preuve

  Conflits longue distance, temporalité

- Risque de disparition de la preuve

  Transfert de charge des Etats vers les plateformes pour la suppression du contenu

Investigations sur Twitter - Temps réel

—

[redacted]

LinkedIn is for the people you know.
Facebook is for the people you used to know.
Twitter is for people you want to know.

~ Inconnu

Twitter...

Twitter

Quelques chiffres et notions...

- 500 millions de tweets sont publiés chaque jour (5.787t/s)
- 326 millions de personnes utilisent Twitter chaque mois
- Réseau asymétrique
- Terrorisme et désinformation
- "Réseau de l'instant"

Twitter

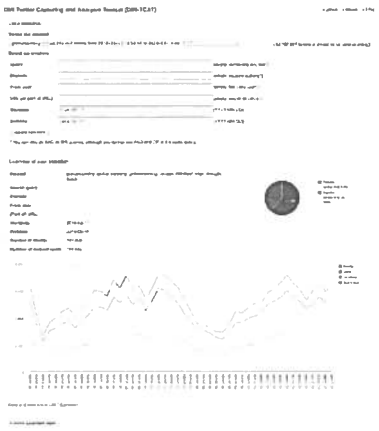
Aspects techniques

- Une API très documentée...
- ... mais avec ses limites...
- Beaucoup d'outils disponibles
- Beaucoup de bibliothèques
- Mon choix : deux outils libres, gratuits et très efficaces pour le monitoring temps réel : DMI-TCAT et Gephi

Monitoring temps réel

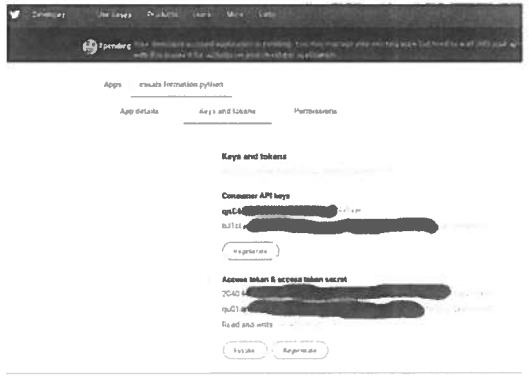
Monitoring en temps réel

1. Possible via l'API de Twitter
2. Pratique si on veut suivre un événement prévu
3. Suivi par hashtag, mot-clefs, utilisateurs
4. Rendu sous forme de base de données et de graphes
5. Attention : ne fait pas le passé!



Etape UN - clef d'API

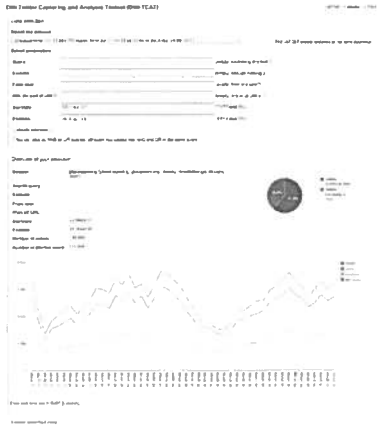
1. Créer un compte développeur Twitter
2. Prévoir un email et un numéro de gsm
3. Récupérer keys et token
4. Attention : Twitter est devenu pénible avec les comptes développeurs pour cause de bot russes...



Installation de DMI-TCAT...

DMI-TCAT

1. Installation en une ligne de commande sous linux
2. Architecture Client/Serveur : Un serveur de collecte, un serveur de requête
3. Suivi par hashtag, mot-clefs, ou utilisateurs
4. Conseils d'utilisation et utilisation



Installation de Twitter Streaming Importer pour Gephi...

Twitter Streaming Importer

1. Un plugin pour Gephi
2. Collecte et analyse en temps réel
3. Suivi par hashtag, mot-clefs, ou utilisateurs
4. Conseils d'utilisation et installation

Intérêt des outils de SNA

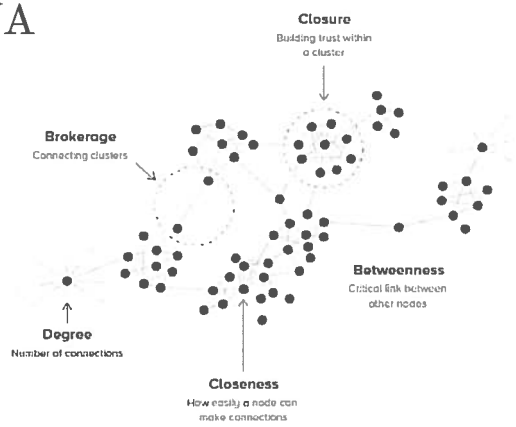
Les unités de mesures quantitatives ne sont pas forcément les plus pertinentes...

L'intérêt d'une analyse est d'identifier rapidement des acteurs-clefs : Auteur initial, influenceurs, etc....

Certains concepts de la SNA sont à connaître et vous seront utiles...

- La modularité
- La centralité intermédiaire

Elles permettent de dégager rapidement des tendances et des éléments marquants.



Conclusion

Le monitoring en temps réel efficace de Twitter est possible gratuitement à l'aide d'outils open-source.

L'Analyse des Réseaux Sociaux (SNA) permet quant à elle de s'affranchir des biais traditionnels de l'approche volumétrique de l'analyse en améliorant la précision des *metrics*.

Le plus difficile reste juste... à obtenir une clef d'API!

Investigations sur Twitter -
Retour vers le futur
—
[redacted]

Comment étudier
le passé sur
Twitter?

L'expérience

Question

Qui est le premier média à avoir donné le nom du terroriste de l'attentat de Nice? Quand et à quelle heure?

- Articles modifiés en temps réel
- Beaucoup de tweets (>17.000)
- Besoin d'archivage
- Besoin d'exhaustivité et de preuve
- Solution : Twint

Twint

Un scraper et une bibliothèque python qui se passe de l'API

- Facile à installer et très rapide
- Pas de limite de temps
- Fonctions de recherche avancées
- Extensible et intégrable
- Archivage

Installation et usage

Installation en une ligne de commande...

pip3 install --upgrade -e git+https://github.com/twintproject/twint.git@origin/master#egg=twint

Usage simple....

- Syntaxe claire
- Export dans divers formats (csv, json, sql...)
- Utilisable via TOR

Exemples d'utilisation

A vous de jouer...

Quelques exemples

Géolocalisés

Chercher des tweets sur l'EI, dans un rayon de 20km autour de Rabat.

Temporalisés

Complexifier la requête en limitant la période à septembre/décembre 2016

Exemple de Nice

Résoudre la question initiale sur Mohamed Lahouej Bouhlel.

Que faire des data?

Conclusion

Twint est un outil indispensable lors d'une enquête judiciaire.

Il permet de :

- S'affranchir des contraintes de temps et des limitations de l'API.
- Préserver et archiver la preuve
- Gagner du temps
- S'ouvrir des possibilités d'analyse statistiques complémentaires.

OSINT - Préparation du poste de travail

[redacted]

Objectif : travailler confortablement et efficacement...

Le navigateur

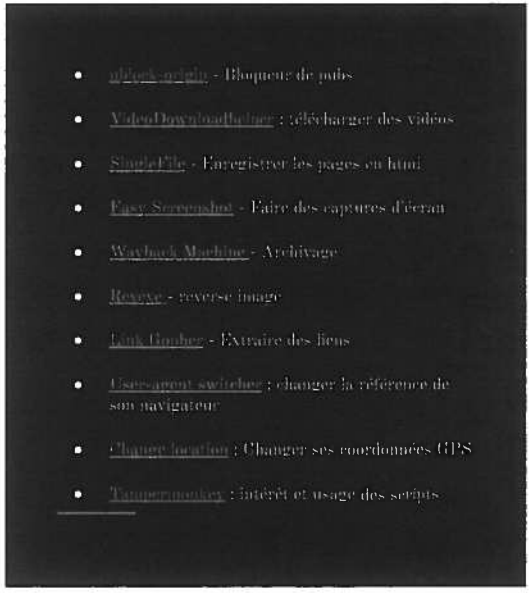
Firefox

Le bon choix de navigateur !

- Libre et gratuit
- Rapide
- Modulaire et extensible
- Mis à jour régulièrement
- Respect de la vie privée

Firefox

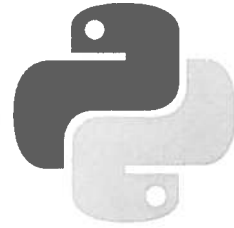
Des plugins très utiles...



Autres Outils....

1. GoogleEarth : géolocalisation
2. Hugin : création de panoramas
3. Python3 : langage de programmation
4. TOR Browser : un peu d'anonymat
5. Gephi : analyse de graphes

Cette liste n'est pas exhaustive!...



Les accessoires...
Indispensables!

Commencez une
double vie...

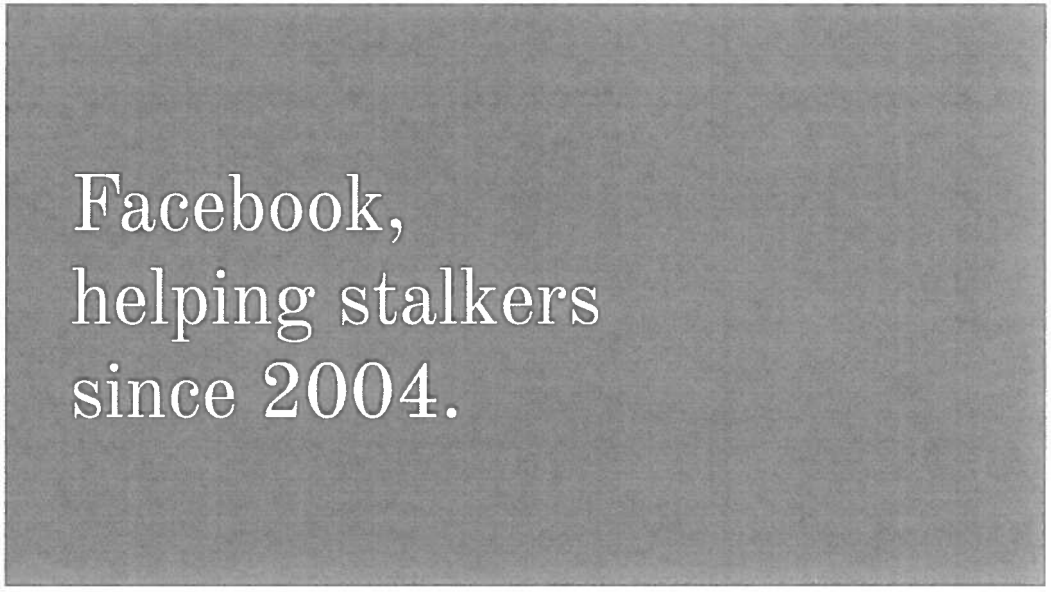
Autres Outils....

1. Un ou plusieurs burner-phones
2. Un VPN avec plusieurs points de sortie
3. Des comptes sur les réseaux sociaux
4. Des adresses mails
5. Un bloc-note et des crayons!
6. de la curiosité et de la patience!

Conclusion

- Pas besoin de beaucoup d'argent pour faire de l'OSINT!
- Bien préparer son poste de travail permet de gagner du temps!


Aller plus loin sur Facebook
—
[redacted]
CEPOL


Facebook,
helping stalkers
since 2004.

Facebook en chiffres

Facebook

Le premier réseau social.
(Des vieux...)

- 2,234 milliards d'utilisateurs actifs par mois
- 1,49 milliard d'utilisateurs actifs quotidiens
- Un moteur de recherches : le GRAPH
- Population qui vieillit
- Contrôle très fort des publications

Le Graph Facebook

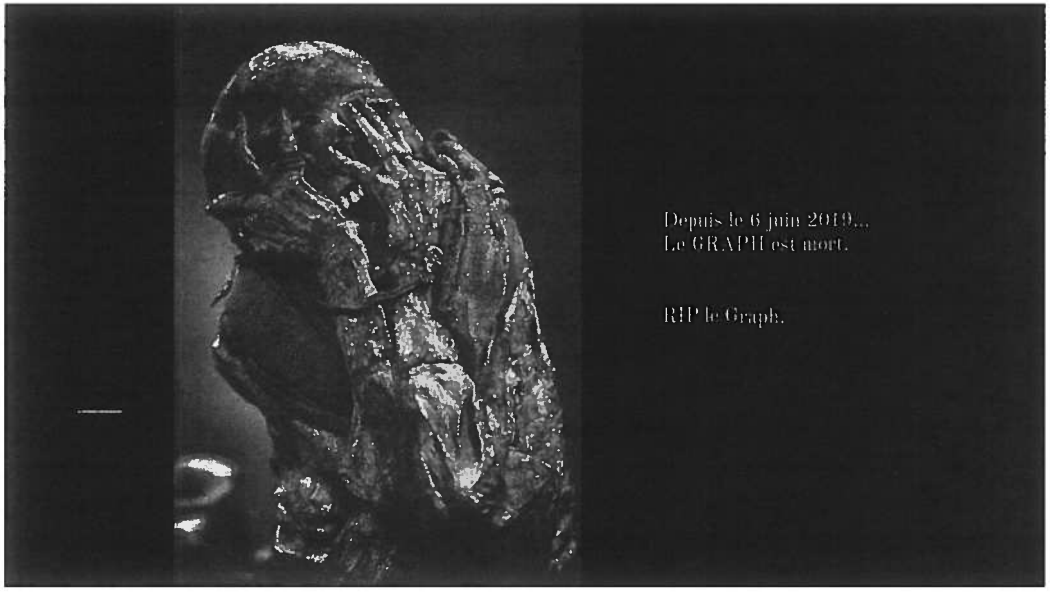
Quelques prérequis...

Un profil fake en anglais...

- Un profil fake est un bien précieux
- Il faut l'entretenir, comme une orchidée
- VPN oui mais tard. TOR? bof.
- Il faut le paramétrer en anglais
- Attention : on travaille en sources ouvertes, on ne devient pas amis avec sa cible.
- On ne travaille pas avec son compte personnel. Jamais.

Une syntaxe particulière...

- en Anglais simplifié
- via des opérateurs
- un guide important : Paul Myers
- Des applications pour vous aider :
  - Stalkscan
  - Whopostedwhat
  - PeopleFind Thor
  - Facebook Matrix



Quelques prérequis...

Un profil fake en anglais...

- Un profil fake est un bien précieux
- Il faut l'entretenir, comme une orchidée
- VPN oui mais tard. TOR? bof.
- Il faut le paramétrer en anglais
- Attention : on travaille en sources ouvertes, on ne devient pas amis avec sa cible.
- On ne travaille pas avec son compte personnel. Jamais.

Une syntaxe particulière depuis juin 2019

- en Anglais simplifié
- via des opérateurs
- https://whopostedwhat.com/
- https://sowdust.github.io/fb-search/
- https://mtg-bi.com/content/facebook-graph-search-workaround
- (Facebook Matrix)

Fake profile

Un profil fake en anglais...

- Un profil fake est un bien précieux
- Une adresse mail et un numéro de téléphone.
- Pas de “nouveau mobile”, préférer un smartphone déjà utilisé
- Pas de VPN les premiers temps
- Ajouter des amis de temps en temps
- Poster des liens (youtube)
- Adopter un comportement “normal”.

Quelques exemples...

Transposition à Instagram

Instagram a été racheté par Facebook.

La fusion des objets et entités est en cours.

Exemple sur les lieux :

Hotel Ibis Agdal 231797643558338 -> Sur Instagram

Espace Hassan à Rabat -> Sur Instagram

Conclusion

Le Graph est un moteur de recherches très puissant.

Sa syntaxe est complexe, il est régulièrement mis à jour dans le plus grand secret par Facebook.

Le suivi régulier de quelques spécialistes permet toutefois de se tenir à jour de ces évolutions.

Le Graph est une source primordiale d'information en ligne.

Facebook - Graphes relationnels

—

[redacted]

Introduction : Construire un graphe simple

—

Un graphe

Quelques notions simples sur les graphes

- Les noeuds et arêtes (nodes & edges)
- LibreOffice ou Excel
- Gephi

Un graphe dans Gephi

Comment symboliser une relation

- Relation entre A et B
- Notion de sens de la relation
- Comment le symboliser dans Excel
- Colonnes Source, Target, Type et commentaire

Hypothèse Facebook 1 : la liste d'amis est publique

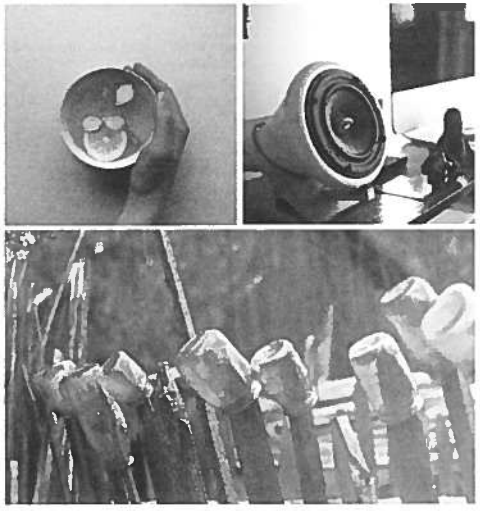
Matériel

Trouvé sur notre bureau !

- Un compte Facebook
- Linkgopher
- LibreOffice ou Excel
- Gephi

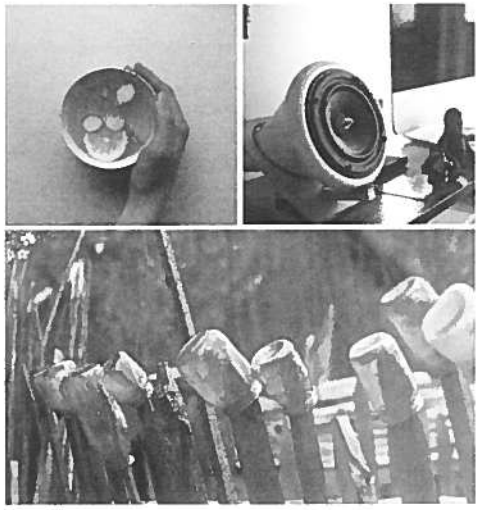
Procédure 1

1. Afficher la liste d'amis de la cible et scroller jusqu'au dernier
2. A l'aide de LinkGopher, récupérer les liens d'amitiés (pb&hc_location=friends_tab)
3. Copier/coller les liens dans une feuille de calcul, dans une colonne "Target"
4. Ajouter une colonne "Source" avec l'ID de votre cible.
5. Ajouter une colonne "Type" avec pour valeur "Undirected"

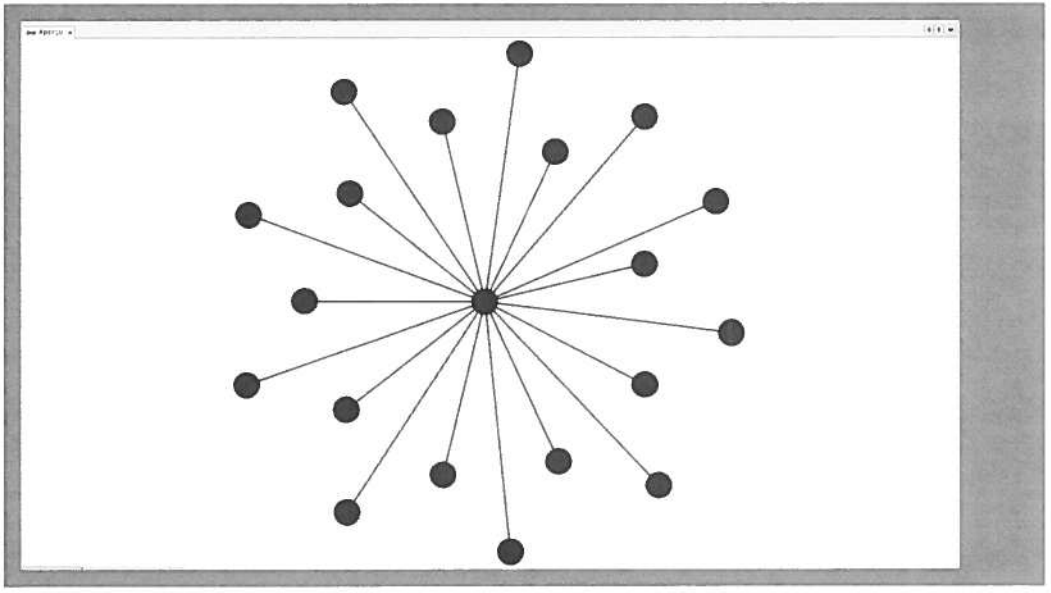


Procédure 2

1. Pour chaque amis, récupérer l'ID via lookup-id. compléter le tableau
2. La colonne Source doit contenir l'ID ou le pseudo de votre cible
3. La colonne Target doit contenir l'ID ou le pseudo de ses amis.
4. Enregistrer le fichier au format csv.
5. Importer ce fichier dans Gephi.



Un premier résultat



C'est un premier graphe, allons plus loin....

Une illustration des liens d'amitiés à faible valeur ajoutée...

Certes, notre cible est liée à ses amis...

https://www.facebook.com/profile.php?id=100009332073971

Mais il manque une information essentielle pour comprendre ce réseau, les liens entre les amis.

A est ami avec B et C. Est-ce que B et C sont amis?

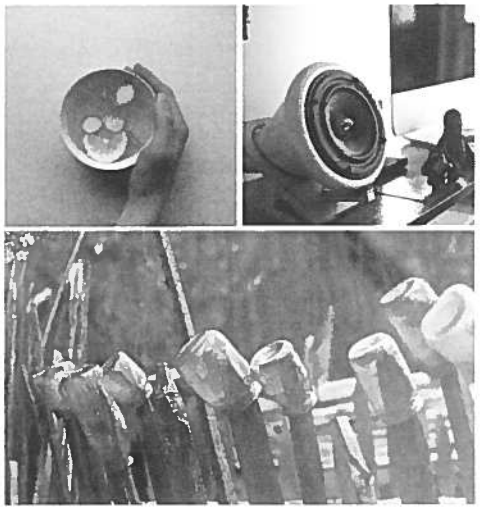
Sur Facebook, il est possible d'obtenir facilement cette information.

https://www.facebook.com/browse/mutual_friends/?uid=ID_Cible&node=ID_Target

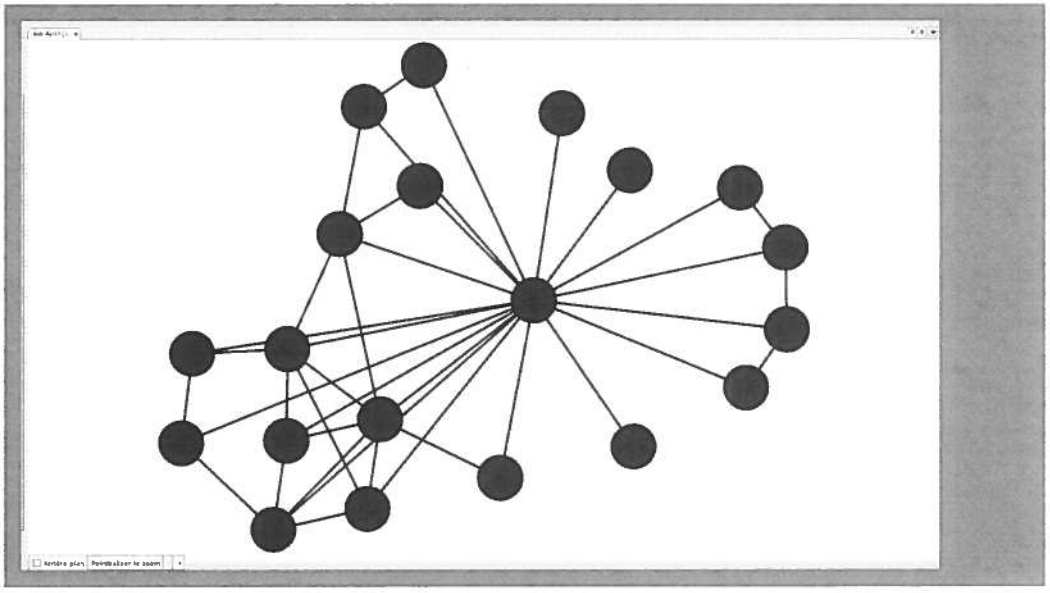
Limite : si les deux amis masquent leurs amis, il n'y a pas de résultats.

Procédure

1. Composer une URL avec l'id de votre cible et son premier ami
2. Les amis communs apparaissent alors à l'écran
3. Compléter le tableau avec en "Target", les profils obtenus, et en "Source" le nom du premier ami.
4. Faire de même pour les amis suivants



Un résultat plus précis

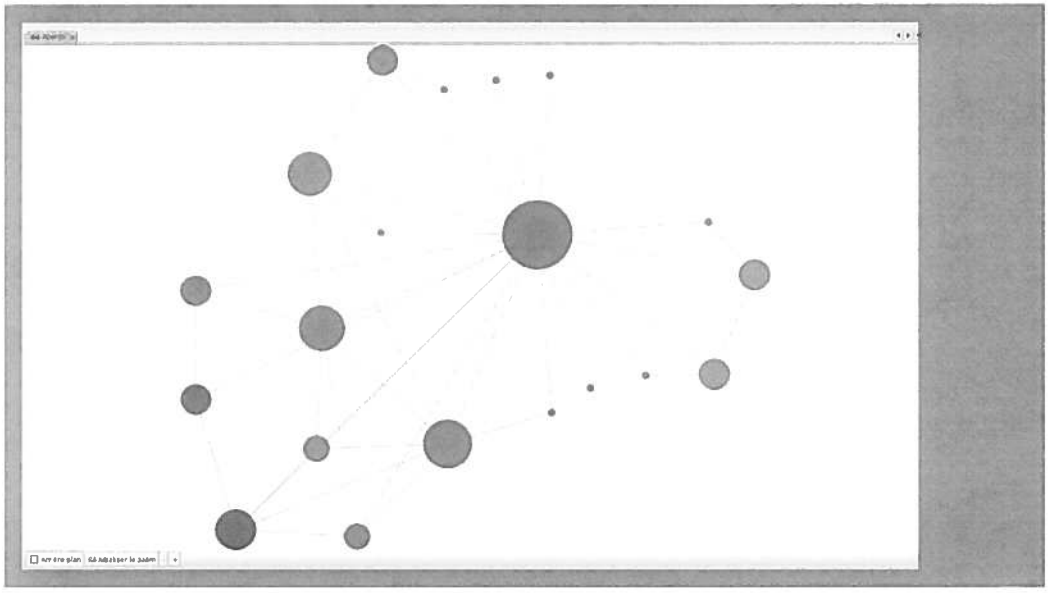


Affinons le résultat

Social Network Analysis

Calculons quelques valeurs !

- Les sous-communautés par Modularité.
- Les influenceurs par la centralité Betweenness
- Forme et taille des noeuds
- Spatialisation



Accélérer le processus

C'est un peu lent, en effet....

- Facebook n'aime pas l'automatisation
- Mais le processus peut être automatisé
- Seule technique à ce jour, le scraping.

Hypothèse Facebook 2 : la liste d'amis est privée

Idée : Si tu likes mes publications, tu es mon ami.

Vérifions cette idée...

- Ouvrir les publications publiques
- Afficher les likes et commentaires
- Récupérer leurs liens
- Construire une première liste d'amis potentiels
- Tester les amis communs
- Réitérer

Exemple : Farid

Transposition : Réseau Social VK et automate

Conclusion

Nous avons vu comment créer un graphe relationnel des amis d'une cible sur Facebook.

A l'aide de Gephi, il est possible de donner du sens à ce graphe et de dégager des pistes de travail, beaucoup plus facilement que par une simple lecture du profil.