

13^{ème} journée Jveille, Genève

early
makers

em
lyon
business
school

Passer à l'échelle du réseau global de Twitter

Clément Levallois

La data science à em Lyon Business School

- Partenaire Smarter Business School avec IBM
 - Objectif: personalized learning
- Renforcement de l'équipe professorale
 - Jean Savinien, math and machine learning
 - Lynn Cherny, data scientist
- En construction: Master DATA en alternance
 - « Data Science and Advanced Technologies for Business Aministration »
 - blockchain, nouveaux business models, IoT, deep learning, etc.

Plan

- Twitter, quelle pertinence pour l'exploration de données ?
- Echantillons de travaux de recherche sur Twitter
- Les solutions commerciales
- Le projet en cours

Twitter, quelle pertinence?

- 117 millions d'utilisateurs actifs
 - 50,000 à 100,000 en Suisse?
 - ~ 2 millions en France
- en 2015, 500 millions de tweets par jour



Un tweet

Auteur du tweet

le tweet lui-même

Un aperçu du lien,
affiché
automatiquement

la date et l'heure

qui a répondu?
qui a retweeté?
qui a liké?
d'où le tweet a-t-il été posté?

une « mention »

des « hashtags »

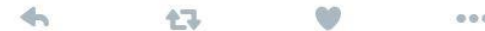


#Mozart et #Schubert en concert ce soir avec @Orchestresavoie et Nicolas Chalvin à La Rampe bit.ly/24v0jDj ← un lien

View translation

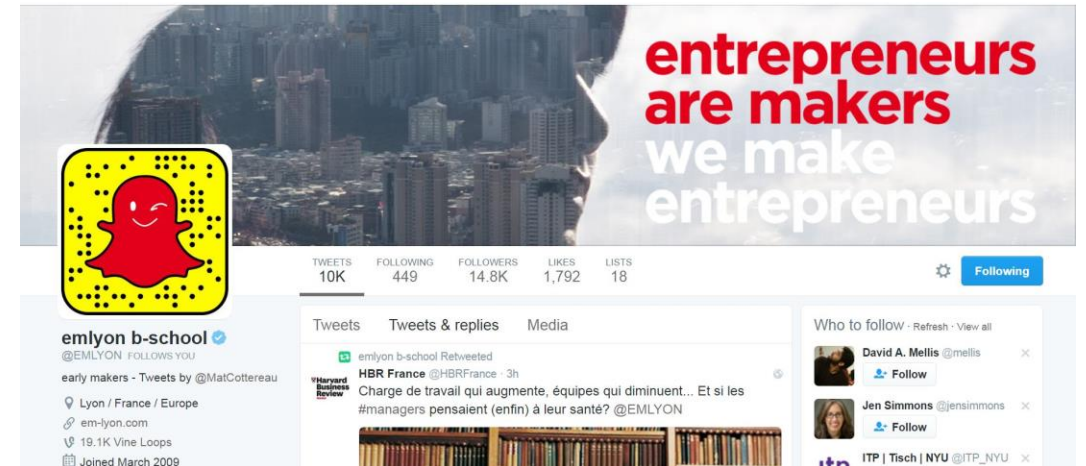


9:29 AM - 3 May 2016



Un utilisateur

- Nom
- Lieu
- Image de profil
- Date d'inscription
- Langue
- Nombre de tweets
- Followers
- Friends
- ...



Twitter: pertinent, potentiellement

Tweets: **unique** car ce sont des messages **publics**, en **temps réel**, sur tous les sujets, accessibles facilement et en **grande quantité** par voie logiciel.

... encore faut-il « faire parler » cette donnée.

Comment les chercheurs font-ils parler ces données?

- Mieux comprendre les villes

Marta Severo et al, 2016: Twitter data for urban policy making

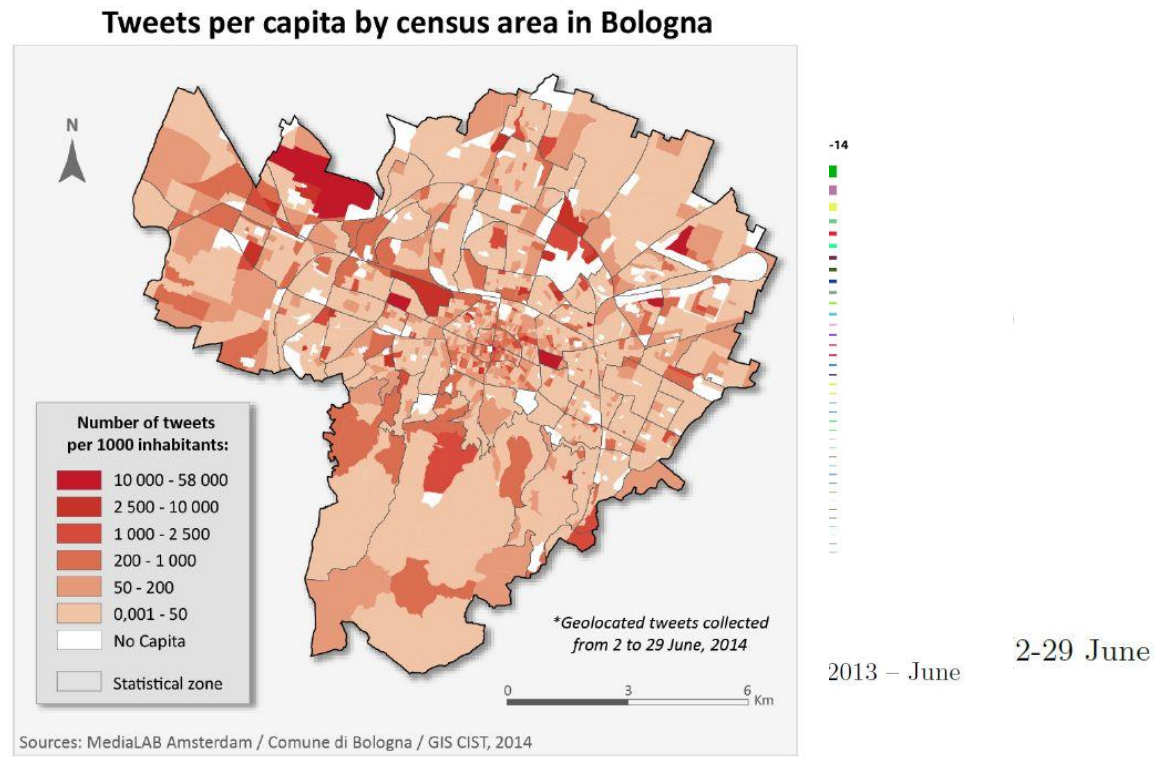


Fig. 9: Stre
2014). Spc



Mining Brand Perceptions from Twitter Social Networks

Aron Culotta

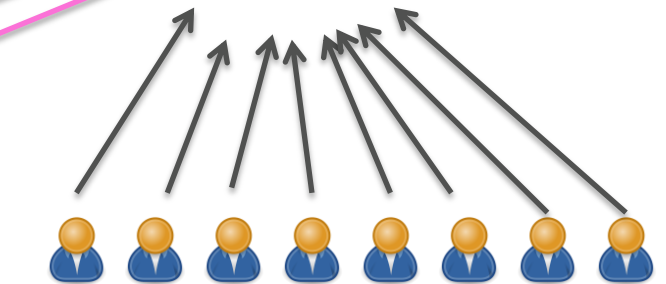
Department of Computer Science, Illinois Institute of Technology, Chicago, Illinois 60616

Jennifer Cutler

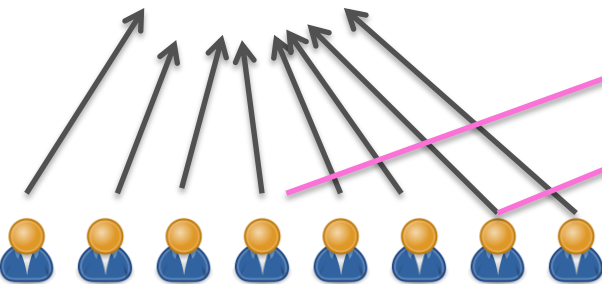
Kellogg School of Management, Northwestern University, Evanston, Illinois 60208

Ex: Is NISSAN perceived to be a green brand?

 @Greenpeace



 @Nissan



Higher
numbers
=
greener

Solutions commerciales (1)



TweetDeck

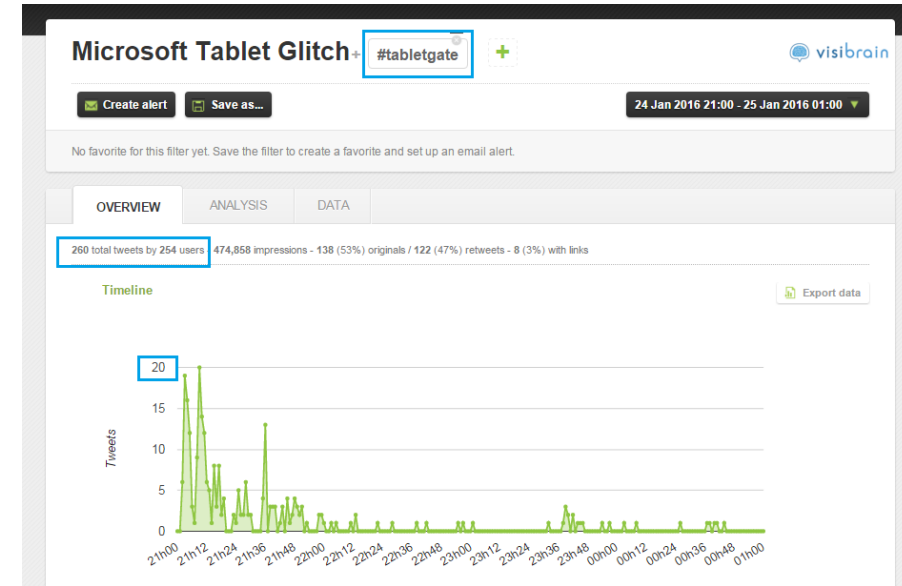
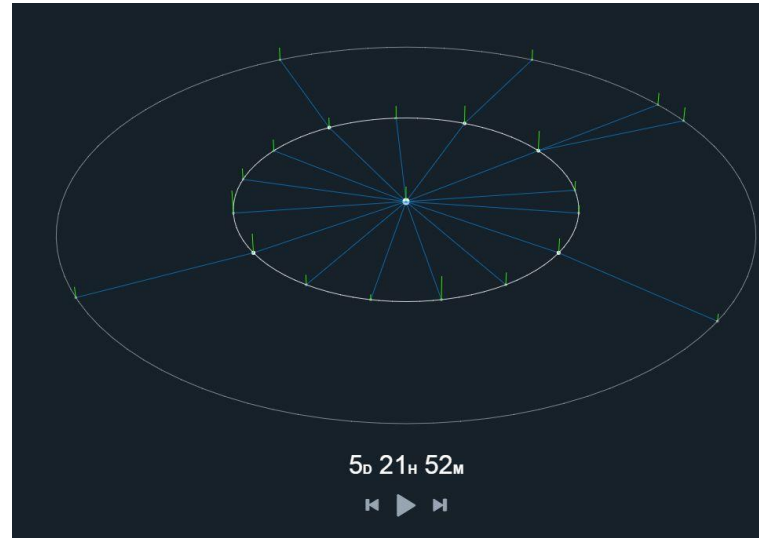
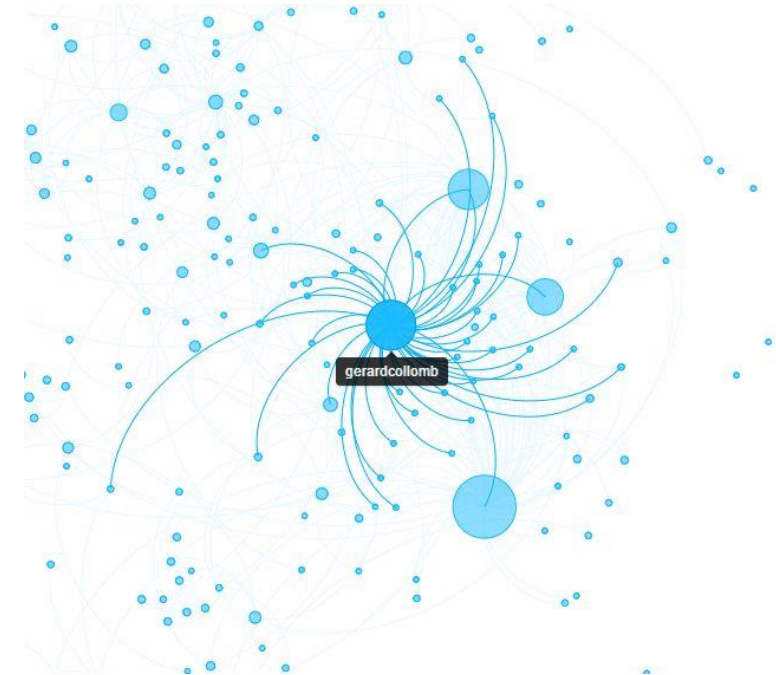


buffer

hootsuite™



Solutions commerciales (2)



bluenod

mfg labs

Visibrain

Un nouvel outil universitaire

OSOME Observatory on Social Media

70 milliards de tweets!



Le projet

Pour aller plus loin que la « lampe torche »

70 milliards de tweets

100+ millions d'utilisateurs

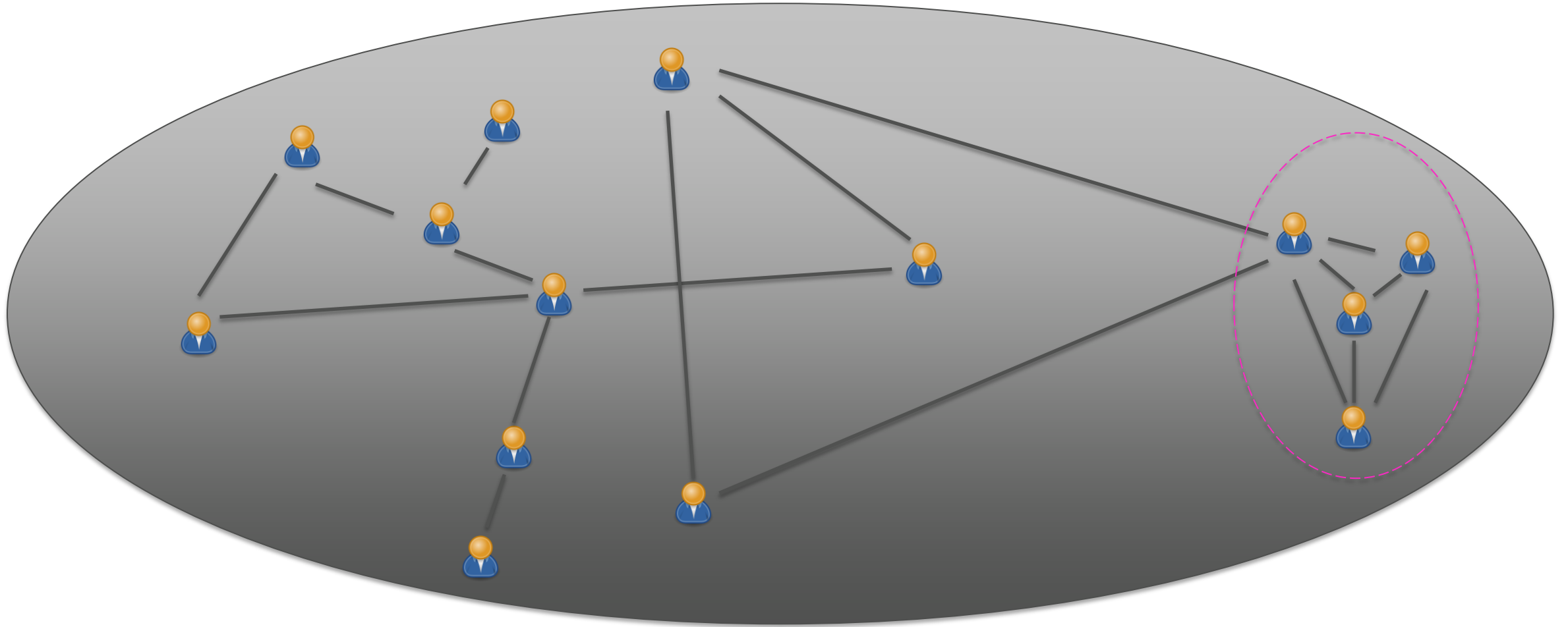
Les milliers de tweets
Les centaines
d'utilisateurs
de notre cas d'étude

(ou la goutte dans l'océan)

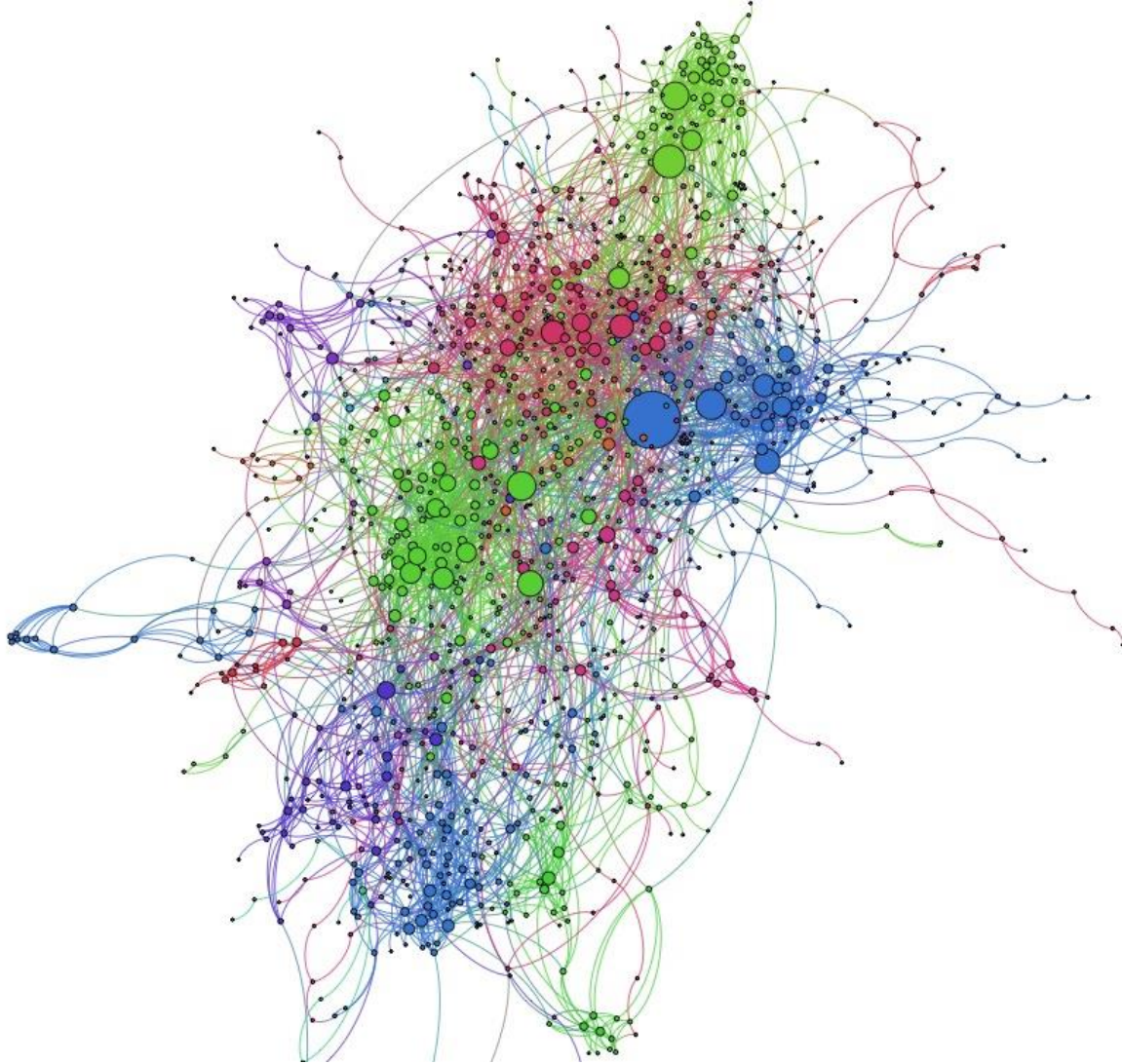


la communauté
d'utilisateurs Twitter que
nous avons identifiée

Objectif: Obtenir une carte complète – un atlas de Twitter



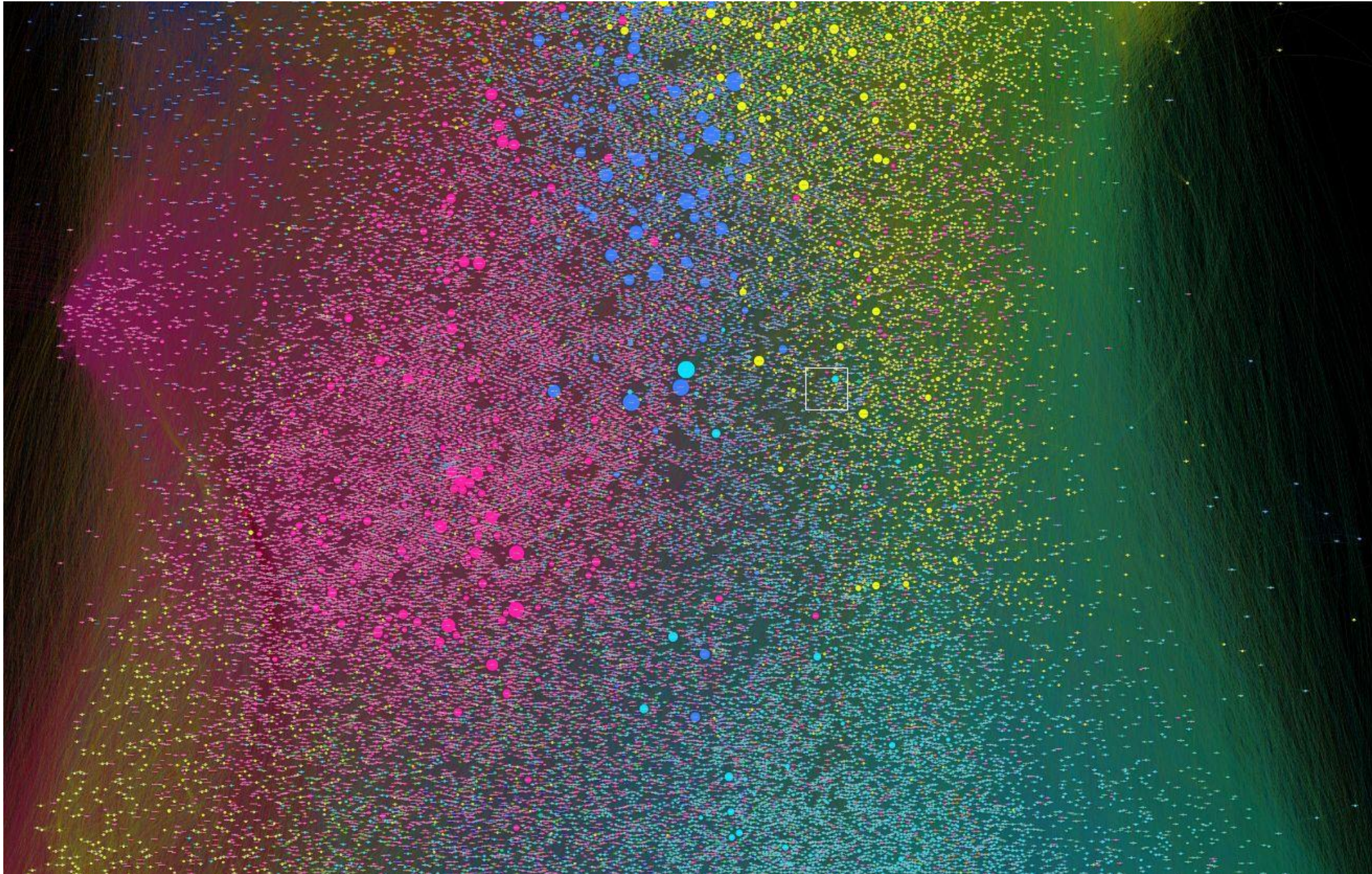
Puis détecter les communautés à cette échelle globale



NB: procédure classique, mais jamais réalisée à l'échelle du réseau entier de Twitter

Ces questions deviendront explorables:

- Comment une nouvelle / une pub / une rumeur / un bad buzz / une mode se propage-t-ils dans le monde?
- Une discussion est-elle locale ou globale, en termes de nombre et taille de communautés qu'elle affecte?
- Quelles communautés discutent de ma marque dans le monde, et ces communautés sont-elles éparses ou en relation entre elles?
- Quel est la longueur du chemin le plus court entre deux communautés? Quels individus, quelles communautés sont sur ce chemin le plus court?
- Dans quelle mesure les chercheurs de tel champ scientifique se connectent-ils à leur industrie locale? Est-ce différent de la façon dont les scientifiques d'autres pays sont en rapport avec leurs propres industries?
- Dans quelles communautés mon entreprise est-elle inscrite, et quelles sont les communautés voisines?
- Quels sont les 100 prospects les plus intéressants, dans une communauté donnée ou dans 100 communautés différentes?
- Quel est l'empreinte de ma marque / mon produit sur cette carte, comment cette empreinte se compare-t-elle à celle de mes concurrents?



<http://twittercensus.se/graph2014/>

90,000
Suède
image statique



50 million
Monde
dataset explorable

Difficultés et solution

- Les données relationnelles entre utilisateurs Twitter sont difficiles à acquérir
 - Il faudrait des années pour récolter toutes les relations « qui suit qui »
- Des substituts ont été imaginés, mais sont insatisfaisants
 - Les relations « qui mentionne qui », faciles à acquérir, sont peu fiables
- Une solution: considérer s'intéresser aux métadonnées.
 - Donnée rapide à acquérir, et très informationnelle

Prochaines étapes

- Finaliser le setup de la solution technique
- Monitorer la récolte des données
- Date prévue de fin de collecte: été 2016
- Mettre en place une procédure de mise à jour régulière et incrémentale de l'atlas

Deliverables

- La carte du réseau, accessible au public via site web ou API
 - En accord avec la politique de confidentialité de Twitter
- Exploration de partenariats avec entreprises pour développer cas d'études spécifique sur cette carte

Equipe

- Pierre-Alexandre Fonta, data scientist, HES-SO
- Thomas Gauthier, professeur, HES-SO

- Contact:

Clément Levallois / @seinecle / levallois@em-lyon.com

Merci!

Questions?

early
makers

em
lyon
business
school