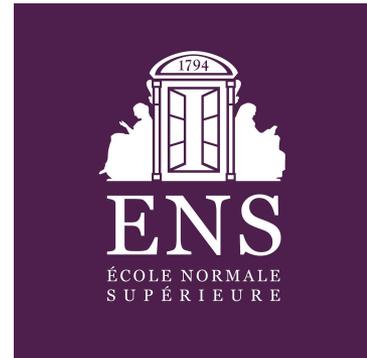


# Algorithmes et Concurrency

Serge Abiteboul



# Données numériques : grandes promesses, risques élevés

- Améliorer la vie des gens
- Accélérer les découvertes scientifiques
- Booster l'innovation
- Transformer l'industrie, le gouvernement, la société
- Optimiser le commerce
- Comportements criminels: sites racistes, terroristes, pédophiles...
- Marketing intrusif, personnalisation et prix cryptiques
- Surveillance abusive de la NSA et de ses homologues européens

## Et pour le commerce ?

- Commerce en ligne
- Choix et comparaisons de produits
- Recommandations d'autres clients
- Meilleur service (relation client)
- Marketing personnalisé/prédicatif
- Marchés du « gratuit »
- Publicités intrusives
- Techniques de *pricing* abusives
- Concentration
  - effet réseau – *winner takes all*
- Marchés « multiface »

# PARTIE A : QUATRE ASPECTS

1. Marchés multiface
2. Marchés du gratuit
3. La transformation du management
4. Un autre *business model* : le PIMS

# PARTIE B : LES BIAIS EN ANALYSE DE DONNÉES

# 1) Marchés « multiface »

Parfois

- Des services gratuits pour des clients en ligne
- Des services payants de publicité
- Des services propriétaires
  - Biais pour favoriser ses propres services, ses clients
- La monétisation des données
  - Tension avec la vie privée

## 2) Les marchés du « gratuit »

- Un moteur de recherche comme Google
  - Un réseau social comme Facebook
  - La musique avec Spotify ou Deezer...
  - Le logiciel *open source*
  
  - Business model : publicité principalement ou abonnement
  - Economie de l'attention
    - Rengaine : si c'est gratuit, c'est que nous sommes le produit
    - Oui mais
      - Ces services sont géniaux
      - Nous y avons accès quels que soient nos revenus
- Des limites : accès Internet + moyens cognitifs de s'en servir

# Comment ça marche ?

- Des services très peu coûteux (par client)
  - En 2013, WhatsApp = 200 millions d'utilisateurs et 50 employés !
  - Biens quasiment non rivaux
- Tension avec la propriété intellectuelle sur livres, morceaux de musique, films...
  - La production ne coûte quasi rien : comment refuser un livre qui ne coûte rien à quelqu'un qui ne peut pas le payer
  - Remise en question du modèle classique de la rétribution des auteurs mais surtout du modèle classique de l'édition

# Homework

- **Quelle différence y-a-t-il entre les fonctionnements de deux services géniaux que vous utilisez peut-être chaque jour, Google et Wikipédia ?**
- Le business model
  - La « pub » (Google) contre la « souscription » (Wikipédia)
- Les objectifs
  - La « vérité » (Wikipedia) contre le « profit commercial » (Google)
- Les biais
  - Non transparents pour Google
  - Les biais des éditeurs sur Wikipedia, mais corrigés par d'autres éditeurs
- Ajoutez la **neutralité** (voir plus loin) et nous pouvons voir cela comme **deux services « publics » avec des modèles de revenus différents**

# (3) La transformation du commerce

- Aussi bien pour le management de l'entreprise que pour ce qui est des échanges avec clients est basé sur de **l'échange d'information**
  - Exemple de Dell
- Temps pré-numérique
  - Organisations très hiérarchiques
  - Transactions très lourdes
- Au temps des algorithmes
  - Il est possible de **numériser** et
  - d'« **agiliser** »
    - Prises de décisions plus partagées
    - Initiatives individuelles d'employés plus autonomes
    - Un peu l'analogie du e-gouvernement participatif

# Décentralisation vs. centralisation, qui va l'emporter ?

- De plus en plus de décentralisation parce que c'est une forme plus efficace d'organisation
- De plus en plus de centralisation avec l'effet réseau

# 4) Un autre *business model* : le PIMS

Les problèmes de concurrence viennent des business models. On peut les changer

*A Personal Information Management System is a cloud system that manages all the information of a person*



## Vos services web, chacun

- Sur des machines inconnues
- Avec du logiciel inconnu
- Vos données

## Votre PIMS

- **Votre machine**
- **Des logiciels choisis par vous**
  - Ou des *wrappers* sur des services externes
- **Avec vos données**
  - Peut-être des copies de vos données sur des systèmes que vous aimez

# **PARTIE B : LES BIAIS EN ANALYSE DE DONNÉES**

# La responsabilité de algorithmes

- L'analyse de données massives permet d'apprendre des données
  - Corrélations statistiques, tendances...
  - Détection de valeurs aberrantes, tendances, des évolutions
  - Technologie : statistique, exploration de données, fouille de données, big data, apprentissage machine...
- Les algorithmes influencent nos décisions (e.g., recommandations) et écident pour nous (voitures, juges, APB...)
- On doit leur imposer de se comporter de façon « responsable » :

**équité, transparence, loyauté, diversité, neutralité...**

# Équité



- **Les humains sont injustes et manquent de cohérence**
  - Les juges : condamnations plus lourdes pour les noirs aux US & application des peines aux US
  - Les commerciaux aussi ?
- **Un algorithme est équitable et cohérent *s'il a été conçu ainsi***
  - Être insensible à la race ou au genre, et à l'heure de la journée
- **Un algorithme peut aussi être injuste**
  - Vente d'agrafeuses sur le web

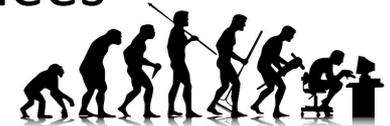


# Garantir la « responsabilité » dans l'analyse de données

Des logiciels pour **collecter et analyser** des données de manière responsable

– Responsabilité *by design*

– Notamment **tracer** la provenance des données



Des logiciels pour **vérifier** que certaines analyses ont été effectuées de manière responsable



# Différents aspects

## Tracer la provenance des données

- provenance et authentification
- Déjà, par ex, dans l'alimentation, la création vidéo...

## Vérification du code

- Audit
- Analyse automatique ( $\approx$  Preuve de théorèmes)

## Tester les effets

- Modèles mathématiques et analyses statistiques  
( $\approx$  Étude de phénomènes comme le climat)
- Ex : vérification de la transparence de Google Ads

# Google Ads Settings



## Control your Google ads

You can control the ads that are delivered to you based on your Google Account, across devices, by editing these settings. These ads are more likely to be useful and relevant to you.

### Your interests

- Action & Adventure Films
- Cooking & Recipes
- History
- Hygiene & Toiletries
- Mobile Phones
- Phone Service Providers
- Reggaeton
- Vehicle Brands

- Cats
- Fitness
- Hybrid & Alternative Vehicles
- Make-Up & Cosmetics
- Parenting
- Recording Industry
- Search Engine Optimization & Marketing

+ ADD NEW INTEREST

WHERE DID THESE COME FROM?

These interests are derived from your activity on Google sites, such as the videos you've watched on YouTube. This does not include Gmail interests, which are used only for ads within Gmail. [Learn more](#)

# Vérification de transparence

- Analyse d'AdFisher
- GAS fait bien ce qu'il dit
  - Si je précise que je ne veux plus recevoir de pubs sur les chats, je n'en reçois plus
- GAS ne se comporte pas comme il dit
  - Il base ses choix sur des infos qu'il ne dévoile pas
  - Par exemple, des attributs protégés comme le sexe
  - Les hommes reçoivent des pubs sur des jobs très bien payés plus que les femmes de manière significative

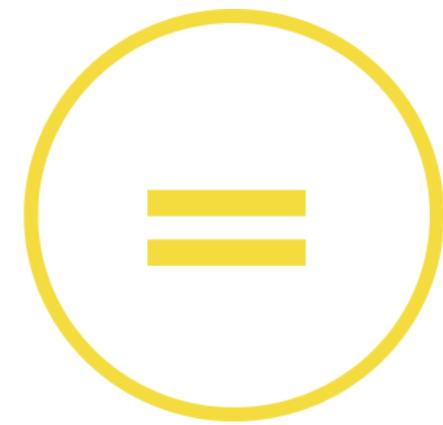
# Diversité



- La plupart des systèmes de recommandation sont basés sur la popularité
  - Ignore les informations moins communes (dans la *tail*) qui forment en réalité la majorité des données
  - Conduit à un manque de diversité, peut-être de la discrimination ou de l'exclusion
- Exemple
  - Une plateforme de rencontre comme Match.com
  - Une place de marché d'externalisation ouverte (aka *crowdsourcing*) comme Amazon Mechanical Turk
  - Une entreprise de financement participatif comme Kickstarter

*The rich gets richer, the poor poorer*

# Neutralité



## Rapport du CNNum

- **Neutralité des réseaux** : le réseau devrait transporter les données sans biais basé sur la source, la destination, et le contenu...
- **Neutralité des plateformes** : une grande plateforme de l'internet ne devrait pas discriminer en faveur de ses propres services

On ne demande pas cela au Monde ou à Carrefour, pourquoi ?

# Préserver la concurrence et la liberté de chacun

- Quelques compagnies sont en train de concentrer toutes les données du monde, et une énorme partie des puissances de calcul (oligopoles)
- **Menaces sur une vraie concurrence** entre les entreprises
  - Elles peuvent offrir de meilleurs services
  - Elles peuvent détruire leurs concurrents
  - Effet réseau
- **Menace sur nos libertés**
  - Elles guident nos décisions
  - Elles contrôlent les informations auxquelles nous sommes exposés

# Procès antitrust contre Google

**theguardian**

## European commission announces antitrust charges against Google

Inquiry will focus on accusations that internet search and tech multinational has unfairly used its products to oust competitors

Sam Thielman in New York

 @samthielman

Wednesday 15 April 2015 07.27 EDT



Le problème est aussi scientifique : détecter/vérifier la présence ou l'absence de biais

The [European Union](#) accused Google on Wednesday of cheating competitors by distorting Internet search results in favour of its Google Shopping service and also launched an antitrust probe into its Android mobile operating system.

# Conclusion (banale)

On peut utiliser les algorithmes pour rendre plus juste/transparente la concurrence... ou moins

Les outils dont nous disposons

- La **loi** et la **régulation**
- La **recherche et le développement** de meilleurs logiciels
- Les **associations d'utilisateurs**
- **L'éducation** des utilisateurs pour changer leurs comportements

le temps des  
ILS TRANSFORMENT LES SCIENCES, L' INDUSTRIE, LA SOCIÉTÉ... ILS BOULE-

algorithmes  
VERSENT LES NOTIONS DE TRAVAIL, DE PROPRIÉTÉ, DE GOUVERNEMENT,

serge abiteboul  
DE VIE PRIVÉE... ET D'HUMANITÉ. QUI N'A PAS ENTENDU PARLER DES ALGORITHMES ? ILS

et gilles dowek  
NOUS FACILITENT LA VIE, MAIS NOUS REDOUTONS ÉGALEMENT QU'ILS NOUS ASSERVISSENT...



Mars 17

SERGE ABITEBOUL  
VALÉRIE PEUGEOT

# Terra Data

Qu'allons-nous faire des données numériques ?

Le Pommier  
cité des sciences et de l'innovation