

# Search Neutrality debate: arguments and mathematical modeling

Patrick Maillé, Bruno Tuffin

Institut Mines-Telecom/Telecom Bretagne, Inria  
Rennes, France

Peyresq summer school - June 2013



# Search engines

- Major role of search engines in the Internet economy

# Search engines

- Major role of search engines in the Internet economy
  - ▶ most used way to reach webpages
  - ▶ about 20 billion requests treated per month from the US home and work desktops only

# Search engines

- Major role of search engines in the Internet economy
  - ▶ most used way to reach webpages
  - ▶ about 20 billion requests treated per month from the US home and work desktops only
- Search engines return a ranked list of links—the *organic results*—from a (set of) keyword(s)

# Search engines

- Major role of search engines in the Internet economy
  - ▶ most used way to reach webpages
  - ▶ about 20 billion requests treated per month from the US home and work desktops only
- Search engines return a ranked list of links—the *organic results*—from a (set of) keyword(s)

organic results are supposed to be based on **relevance**

# Search engines

- Major role of search engines in the Internet economy
  - ▶ most used way to reach webpages
  - ▶ about 20 billion requests treated per month from the US home and work desktops only
- Search engines return a ranked list of links—the *organic results*—from a (set of) keyword(s)

organic results are supposed to be based on **relevance**

Search engines are more and more suspected to tamper with the ranking

# Search engines

[Recherche avancée](#)Rechercher dans : ☒ Web ☐ Pages francophones ☐ Pages : FranceWeb [Afficher les options...](#)Résultats 1 à 10 sur un total d'environ **240 000 000** pour **used cars usa** (0,29 sec)

## Used Cars

Lien commercial

[MortizChevrolet.com/Used\\_Cars](http://MortizChevrolet.com/Used_Cars) Low Prices On **Used Cars** - Get A Free Internet Quote!

Conseil : [Recherchez des résultats uniquement en français](#). Vous pouvez indiquer votre langue de recherche sur la page [Préférences](#).

[New & Used Cars for Sale, Auto Dealers, Car Reviews and Car ...](#) - [ [Traduire cette page](#) ]

Search 2.6 million new & **used car** listings, price a new car, get a dealer quote, read expert reviews, or sell your car at thousands over trade-in.

[Advanced Search](#) - [Buy a Car](#) - [How Much Is Your Car Worth](#) - [SUVs](#)[www.cars.com/](http://www.cars.com/) - [États-Unis](#) - [En cache](#) - [Pages similaires](#)[Used Cars for Sale, New Cars, & Auto Buying Guide at AutoMallUsa.net](#) - [ [Traduire cette page](#) ]

View **used cars** for sale, read car reviews, or research new cars at AutoMallUsa.net, your auto buying guide online.

[Used Cars](#) - [Cars for Sale](#) - [Toyota](#) - [Nissan](#)[www.automallusa.net/](http://www.automallusa.net/) - [En cache](#) - [Pages similaires](#)[New Cars, Used Cars - Find Cars at AutoTrader.com](#) - [ [Traduire cette page](#) ]

Find **used cars** for sale at AutoTrader.com. With over 3.5 million cars, finding your ... I want to stay and look for cars on AutoTrader.com in the **U.S.A.** ...

[www.autotrader.com/](http://www.autotrader.com/) - [États-Unis](#) - [En cache](#) - [Pages similaires](#)[en stock aux Etats-Unis et au Canada - DENKER US CARS | Import New ...](#)

Sale New **US Cars**, **Used US-Cars**, Import Vehicles from **USA**, SUV, Offroad, 4x4, ... Collector **Car** - oldtimer. S'il faut remplir le ZIP Code, remplissez-le pour ...

[french.denker.cz/outre-mer/](http://french.denker.cz/outre-mer/) - [En cache](#) - [Pages similaires](#)

Liens commerciaux

## Used Car In Usa

100% Free : **Cars** Classified Ads!1500 American **Cars** For Sale[Voitures-Americales.com](http://Voitures-Americales.com)

## 4X4 & Cars for export

Tax free **cars** since 1973Japan, german **cars**, bus & trucks[www.transauto.be](http://www.transauto.be)

## Used Car In Usa

Trouvez **Used car** in **usa**Vous trouverez **Used car** in **usa** sur Ask ![FR.Ask.com](http://FR.Ask.com)

## Salvage Auto Auctions USA

Open 4 Public join to Bid and Buy immediately. Free Access!

[www.easyexport.us](http://www.easyexport.us)[Affichez votre annonce ici »](#)

# Search Neutrality



# Search Neutrality

- Term coined in 2009 by Adam Raff, co-founder of Foundem (a price-comparison company)
- Google admitted penalizing Foundem and other specialized search services “to protect users from spam”

# Search Neutrality

- Term coined in 2009 by Adam Raff, co-founder of Foundem (a price-comparison company)
- Google admitted penalizing Foundem and other specialized search services “to protect users from spam”
- But this can also be seen as hindering competition: Google offers similar services (Maps, Shopping, Images, videos...)

# Search Neutrality

- Term coined in 2009 by Adam Raff, co-founder of Foundem (a price-comparison company)
- Google admitted penalizing Foundem and other specialized search services “to protect users from spam”
- But this can also be seen as hindering competition: Google offers similar services (Maps, Shopping, Images, videos...)

Term close to *net neutrality*: limitations on users' access to all relevant services on the Internet

# Search Neutrality

- Term coined in 2009 by Adam Raff, co-founder of Foundem (a price-comparison company)
- Google admitted penalizing Foundem and other specialized search services “to protect users from spam”
- But this can also be seen as hindering competition: Google offers similar services (Maps, Shopping, Images, videos...)

Term close to *net neutrality*: limitations on users' access to all relevant services on the Internet

*Search neutrality* would impose that all contents have the same chances of being displayed

⇒ a ranking based on **relevance** (to be defined objectively)

# Search Neutrality: the debate

Arguments in favor:

Arguments against:

# Search Neutrality: the debate

Arguments in favor:

- Neutrality benefits to users, who get the most relevant results

Arguments against:

# Search Neutrality: the debate

Arguments in favor:

- Neutrality benefits to users, who get the most relevant results
- Neutrality elicits efforts from websites to improve their content quality

Arguments against:

# Search Neutrality: the debate

## Arguments in favor:

- Neutrality benefits to users, who get the most relevant results
- Neutrality elicits efforts from websites to improve their content quality
- Neutrality benefits to the global economy by facilitating the access to the best-performing actors/services
  - ⇒ new businesses can emerge more easily

## Arguments against:



# Search Neutrality: the debate

## Arguments in favor:

- Neutrality benefits to users, who get the most relevant results
- Neutrality elicits efforts from websites to improve their content quality
- Neutrality benefits to the global economy by facilitating the access to the best-performing actors/services
  - ⇒ new businesses can emerge more easily

## Arguments against:

- Users are interested in the differences among search engines ( "*Google users tend to prefer Google products*" )

# Search Neutrality: the debate

## Arguments in favor:

- Neutrality benefits to users, who get the most relevant results
- Neutrality elicits efforts from websites to improve their content quality
- Neutrality benefits to the global economy by facilitating the access to the best-performing actors/services  
⇒ new businesses can emerge more easily

## Arguments against:

- Users are interested in the differences among search engines ( "*Google users tend to prefer Google products*" )
- Search engines would not be able to manipulate rankings to prevent spam ⇒ worse results for users

# Search Neutrality: the debate

## Arguments in favor:

- Neutrality benefits to users, who get the most relevant results
- Neutrality elicits efforts from websites to improve their content quality
- Neutrality benefits to the global economy by facilitating the access to the best-performing actors/services  
⇒ new businesses can emerge more easily

## Arguments against:

- Users are interested in the differences among search engines ( "*Google users tend to prefer Google products*" )
- Search engines would not be able to manipulate rankings to prevent spam ⇒ worse results for users
- Imposing transparency of the ranking algorithms facilitates the job of spammers

# Search Neutrality: the debate

## Arguments in favor:

- Neutrality benefits to users, who get the most relevant results
- Neutrality elicits efforts from websites to improve their content quality
- Neutrality benefits to the global economy by facilitating the access to the best-performing actors/services
  - ⇒ new businesses can emerge more easily

## Arguments against:

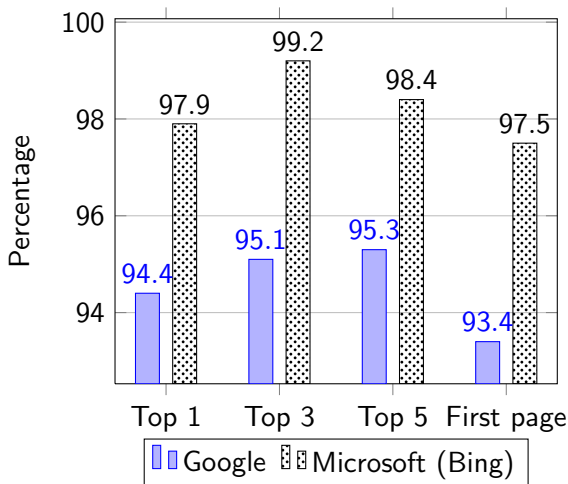
- Users are interested in the differences among search engines ( "*Google users tend to prefer Google products*" )
- Search engines would not be able to manipulate rankings to prevent spam ⇒ worse results for users
- Imposing transparency of the ranking algorithms facilitates the job of spammers
- Competition is just one click away...

# Do search engines return biased results?

# Do search engines return biased results?

**They do**

# Do search engines return biased results?



Percentage of Google or Bing search results with **own content not ranked similarly** by any rival search engine (Wright, 2012).

# Do search engines return biased results?

Comparison between Google, Bing, and Blekko (Wright, 2012):

- Microsoft content is 26 times more likely to be displayed on the first page of Bing than on any of the two other search engines
- Google content appears 17 times more often on the first page of a Google search than on the other search engines

Search engines do favor their own content



# Regulatory intervention

- The European Commission, is progressing toward an antitrust settlement deal with Google

*Google must be even-handed. It must hold all services, including its own, to exactly the same standards, using exactly the same crawling, indexing, ranking, display, and penalty algorithms.*

- The European Commission is running a market testing (started in April 2013) to estimate the extent to which the Google ranking algorithm respects these guidelines (Google may face a fine as large as \$5 billion)

# Building a mathematical model

For a given keyword, each webpage  $i$  is characterized by

# Building a mathematical model

For a given keyword, each webpage  $i$  is characterized by

- its relevance  $r_i$

# Building a mathematical model

For a given keyword, each webpage  $i$  is characterized by

- its relevance  $r_i$
- the gain  $g_i$  to the search engine if the link is clicked

# Building a mathematical model

For a given keyword, each webpage  $i$  is characterized by

- its relevance  $r_i$
- the gain  $g_i$  to the search engine if the link is clicked

Then users click on links depending on their *position* in the ranking  
clicking probabilities  $\theta_1 > \theta_2 > \dots$

# Building a mathematical model

For a given keyword, each webpage  $i$  is characterized by

- its relevance  $r_i$
- the gain  $g_i$  to the search engine if the link is clicked

Then users click on links depending on their *position* in the ranking  
clicking probabilities  $\theta_1 > \theta_2 > \dots$

Relevance of the ranking  $\pi$  ( $\pi_i = \text{position of webpage } i$ ):

$$r = \sum_i \theta_{\pi_i} r_i.$$

# Building a mathematical model

For a given keyword, each webpage  $i$  is characterized by

- its relevance  $r_i$
- the gain  $g_i$  to the search engine if the link is clicked

Then users click on links depending on their *position* in the ranking  
clicking probabilities  $\theta_1 > \theta_2 > \dots$

Relevance of the ranking  $\pi$  ( $\pi_i$  = position of webpage  $i$ ):

$$r = \sum_i \theta_{\pi_i} r_i.$$

Expected gain per search:

$$g = \underbrace{\beta}_{\text{from ads}} + \sum_i \theta_{\pi_i} g_i$$

# Trade-off relevance *versus* revenue



# Trade-off relevance *versus* revenue

- Favoring revenue-yielding webpages increases the revenue per search

# Trade-off relevance *versus* revenue

- Favoring revenue-yielding webpages increases the revenue per search
- But users are interested in relevance: they may switch to another search engine

# Trade-off relevance *versus* revenue

- Favoring revenue-yielding webpages increases the revenue per search
- But users are interested in relevance: they may switch to another search engine

Model: a number of searches per time unit  $\lambda(r)$ , increasing with the relevance  $r$

# Trade-off relevance *versus* revenue

- Favoring revenue-yielding webpages increases the revenue per search
- But users are interested in relevance: they may switch to another search engine

Model: a number of searches per time unit  $\lambda(r)$ , increasing with the relevance  $r$

$$\Rightarrow \text{expected revenue per time unit} = \lambda(r) \times g$$

# Neutral and non-neutral rankings

# Neutral and non-neutral rankings

- Neutral ranking: based on relevance  $(r_i)_i$

# Neutral and non-neutral rankings

- Neutral ranking: based on relevance  $(r_i)_i$   
⇒ maximizes the average relevance, and thus the number of requests

# Neutral and non-neutral rankings

- Neutral ranking: based on relevance  $(r_i)_i$   
⇒ maximizes the average relevance, and thus the number of requests
- Non-neutral ranking: aimed at maximizing the revenue  $\lambda(r) \times g$



# Neutral and non-neutral rankings

- Neutral ranking: based on relevance  $(r_i)_i$   
 $\Rightarrow$  maximizes the average relevance, and thus the number of requests
- Non-neutral ranking: aimed at maximizing the revenue  $\lambda(r) \times g$   
Maximization over the set of permutations

# Neutral and non-neutral rankings

- Neutral ranking: based on relevance  $(r_i)_i$   
 $\Rightarrow$  maximizes the average relevance, and thus the number of requests
- Non-neutral ranking: aimed at maximizing the revenue  $\lambda(r) \times g$

Maximization over the set of permutations

Say with  $m$  candidate pages and  $\lambda(r) = r$

$$\max_{\text{permutations } \pi} \left( \sum_{i=1}^m \theta_{\pi_i} r_i \right) \cdot \left( \beta + \sum_{i=1}^m \theta_{\pi_i} g_i \right)$$

# Neutral and non-neutral rankings

- Neutral ranking: based on relevance  $(r_i)_i$   
 $\Rightarrow$  maximizes the average relevance, and thus the number of requests
- Non-neutral ranking: aimed at maximizing the revenue  $\lambda(r) \times g$

Maximization over the set of permutations

Say with  $m$  candidate pages and  $\lambda(r) = r$

$$\max_{\text{permutations } \pi} \left( \sum_{i=1}^m \theta_{\pi_i} r_i \right) \cdot \left( \beta + \sum_{i=1}^m \theta_{\pi_i} g_i \right)$$

not an easy task...

## An example

One keyword, three pages,  
click probabilities  $\theta_i = \frac{1}{2^i}$   
 $\lambda(r) = r$

| $i$ | Relevance $r_i$ | Gain $g_i$ |
|-----|-----------------|------------|
| 1   | 3               | 0          |
| 2   | 2               | 0          |
| 3   | 1               | 2          |

| Ranking | Relevance<br>( $r$ ) | Engine revenue per<br>time unit |
|---------|----------------------|---------------------------------|
| 1; 2; 3 | 2.125                | $2.125 (\beta + \frac{1}{4})$   |
| 1; 3; 2 | 2                    | $2 (\beta + \frac{1}{2})$       |
| 3; 1; 2 | 1.5                  | $1.5 (\beta + 1)$               |

Depending on the revenues from ads (value of  $\beta$ ),  
each of these three can be the best one

## The case of a general set of keywords

Users do not change search engines for each keyword: they build a “reputation” of the search engines (average relevance of their results)

# The case of a general set of keywords

Users do not change search engines for each keyword: they build a “reputation” of the search engines (average relevance of their results)

Model:

- keywords treated as random: for each search we have
    - ▶ a (random) vector of relevances  $R$
    - ▶ a (random) vector of revenues  $G$
- $\Rightarrow$  one ranking  $\pi(R, G)$  to perform for each search

# The case of a general set of keywords

Users do not change search engines for each keyword: they build a “reputation” of the search engines (average relevance of their results)

Model:

- keywords treated as random: for each search we have
  - ▶ a (random) vector of relevances  $R$
  - ▶ a (random) vector of revenues  $G$

$\Rightarrow$  one ranking  $\pi(R, G)$  to perform for each search
- but the ranking chosen affects the reputation of the search engine

# The case of a general set of keywords

Users do not change search engines for each keyword: they build a “reputation” of the search engines (average relevance of their results)

Model:

- keywords treated as random: for each search we have
  - ▶ a (random) vector of relevances  $R$
  - ▶ a (random) vector of revenues  $G$

$\Rightarrow$  one ranking  $\pi(R, G)$  to perform for each search
- but the ranking chosen affects the reputation of the search engine

$$\text{Revenue} = \lambda \left( \mathbb{E} \left[ \sum_{i=1}^m \theta_{\pi_i} R_i \right] \right) \cdot \left( \beta + \mathbb{E} \left[ \sum_{i=1}^m \theta_{\pi_i} G_i \right] \right)$$

We have a few results regarding that problem



# What we want to do

# What we want to do

- Analyze the non-neutral ranking

# What we want to do

- Analyze the non-neutral ranking
- Compare the performance and neutral and non-neutral policies

# What we want to do

- Analyze the non-neutral ranking
- Compare the performance and neutral and non-neutral policies
  - ▶ Cost of non-neutrality: loss of relevance for users

# What we want to do

- Analyze the non-neutral ranking
- Compare the performance and neutral and non-neutral policies
  - ▶ Cost of non-neutrality: loss of relevance for users
  - ▶ Cost of neutrality (for search engines): loss of revenue for search engines

# What we want to do

- Analyze the non-neutral ranking
- Compare the performance and neutral and non-neutral policies
  - ▶ Cost of non-neutrality: loss of relevance for users
  - ▶ Cost of neutrality (for search engines): loss of revenue for search engines
- Discuss the need for regulation

This topic (and many others), in a book to appear:

