Comparison of Search Engines Non-Neutral and Neutral Behaviors

P. Coucheney, P. Maillé, Bruno Tuffin

INRIA Rennes - Centre Bretagne Atlantique

W-PIN
London
June 2012
Outline

1. Introduction: why investigating search neutrality?

2. Model

3. Numerical comparison

4. Case of a general set of keywords
Outline

1. Introduction: why investigating search neutrality?

2. Model

3. Numerical comparison

4. Case of a general set of keywords
Network neutrality, a key political issue...

- A very hot topic in the past few years, at the same time from political, economic, daily-life points of view

- **Reason of the debate:**
  - Major CPs have their traffic flowing through their networks
  - but do not pay any fee for that
  - while CPs’ revenue represents an important (and increasing) part of the total network-related money.

- **Threat** to ask side payments to those CPs, to cut their access, or to downgrade their quality of service.

- This violates the **neutrality principle**, stating that all consumers are entitled to reach meaningful content, and that packets should not be differentiated.

- A lot of protests from CPs and users associations claiming, among other reasons, that it would be an impeachment of freedom of speech.

Network neutrality an issue... but not the only one...
What about search engines?

- Strangely forgotten in the debate
- **Key role in the Internet:**
  - Content often reached by dialing a keyword on a search engine
  - A biased search engine could somewhat “cut" CPs from end users if omitted or put low in the list of displayed links.
- **Search engine advertising** an increasing business, thanks to **clearly declared sponsored links**
  - Combined revenue of Yahoo! and Google: more than $11 billion in 2005
  - Business is expected to count for about 40% of total advertising revenue
- **Success** of search engines: **organic links**, i.e., non-commercial links displayed by search engines and ranked according to their appropriateness/relevance related to the keyword(s).
- A **neutral ranking**, based only on relevance is expected!
Non-neutrality and our goal

Possible **non-neutral** behaviors:

- push, in organic rankings, some paying content ahead of more relevant links
- favor some content that it owns

Recent reproach to Google, accused to favor Youtube content

Google’s CEO was forced to testify in front of the US senate, and is facing to be dismantled because of that.

Our **goals** here:

- model and understand a non-neutral behavior
- compare it with a neutral one.
Outline

1. Introduction: why investigating search neutrality?
2. Model
3. Numerical comparison
4. Case of a general set of keywords
Players

- **Users**
  - represented by a rate of requests $\lambda(r)$ per unit of time that depends on $r$, the (average) relevance of displayed links (to be detailed later).

- **Content providers (CPs)**
  - $m$ CPs for a typical search
  - CP $i$ relevance is $R_i$
  - CP $i$ average gain per click is $G_i$

- **Search engine (SE)**
  - displays all $m$ possible CPs (no limited amount of slots)
  - Impact of the ranking through the click-through-rate (CTR)
  - Let $\theta_j$ be the CTR of rank $j$, with $\theta_1 > \theta_2 > \cdots > \theta_m$.

- **Consider first a single keyword**
  - $\lambda(r)$ arrival rate of requests for that specific keyword
  - The relevance of the ranking can be defined as
    \[
    r = \sum_{i=1}^{m} \theta_{\pi(i)} R_i
    \]
    where $\pi(i)$ is the rank of CP $i$. 

P. Coucheney, P. Maillé, B. Tuffin (Inria)
Search Engine Neutrality
June 2012 8 / 18
Utility functions

- **Content providers**: avg revenue from clicks

  \[ U_i^{\text{CP}} = \lambda(r) \theta_{\pi(i)} G_i. \]

- **Search engine**
  - neutral revenue from sponsored links: \( h(\lambda(r)) \) (\( h \) assumed increasing and concave: the more visits, the more chances advertisement slots are clicked)
  - Two non-neutral options:
    1. \( m' < m \) links \( \{1, \ldots, m'\} \) owned by the SE itself. Additional revenue \( \sum_{i=1}^{m'} \lambda(r) \theta_{\pi(i)} G_i. \)
    2. The SE puts “sponsored” links in the organic ranking and gets a proportion \( \alpha \) of the CP revenue: \( \alpha \sum_i \lambda(r) \theta_{\pi(i)} G_i. \)
Total relevance UR for users (corresponding to a total user satisfaction)

- Rewrite $\lambda(r) = \Lambda \Phi(r)$ with
  - $\Lambda = \lambda(\infty)$ maximal arrival rate
  - $\Phi(r)$ probability that an arriving request accepts a relevance $r$, i.e., requires $r$ or less.
- Conversely, a rate $\lambda$, corresponds to requests asking for an average relevance level up to $\Phi^{-1}(\lambda/\Lambda)$.
- If $\phi$ is the density associated to distribution $\Phi$,
  $$UR = \int_0^r y\phi(y)dy.$$
Outline

1. Introduction: why investigating search neutrality?

2. Model

3. Numerical comparison

4. Case of a general set of keywords
Illustrative framework

- \( m = 3 \) relevant links, with \((G_1, R_1) = (3, 2); (G_2, R_2) = (2, 3) \) and \((G_3, R_3) = (1, 1)\)
- \( \theta_i = 1/2^i \), and \( \lambda(r) = \sqrt{\min(r, 4)} \).

This gives
- \( \phi(y) = 1/(4\sqrt{y})1_{\{y\in(0,4]\}} \) and \( UR = r^{3/2}/6 \)
- Let \( h(\lambda) = \beta\lambda \).

The neutral case:
- rank first CP 2, then CP 1 and CP 3
- Thus \( r = \sum_{i=1}^{3} \theta_{\pi(i)}R_i = 17/8 \), \( \lambda = \sqrt{17/8} \approx 1.4577 \), \( UR \approx 0.5163 \), \( U_{1\text{CP}} \approx 1.0933 \), \( U_{2\text{CP}} \approx 1.4577 \) and \( U_{3\text{CP}} \approx 0.1822 \).
Non-neutral: the engine does not own any content but gets a proportion $\alpha$ of the revenue of each CP as a tax

- Neutral situation: maximize revenue $h(\lambda(r))$, hence $r$
- Non-Neutral: max

$$U^{SE} = h(\lambda(r)) + \alpha \lambda(r) \sum_{i=1}^{3} \theta_{\pi(i)} G_i = \sqrt{\sum_{i=1}^{3} \theta_{\pi(i)} R_i \left( \beta + \alpha \sum_{i=1}^{3} \theta_{\pi(i)} G_i \right)}.$$ 

- Non-neutral preferred if $\beta < 1.996\alpha$. 

![Graph of Revenues and UR](https://via.placeholder.com/150)

P. Coucheney, P. Maillé, B. Tuffin (Inria)  
Search Engine Neutrality  
June 2012 13 / 18
CP 1 is owned by the engine

\[ U^{SE} = h(\lambda(r)) + \lambda(r)\theta_{\pi(1)} G_1 = (\beta + 3\theta_{\pi(1)}) \sqrt{\sum_{i=1}^{3} \theta_{\pi(i)} R_i}. \]

- Neutral case: \( U^{SE} = \sqrt{17/8}(\beta + 3/4). \)
- Non-neutral preferred if \( \beta < 3/2(\sqrt{15/17} – 1/2)/(1 – \sqrt{15/17}) \approx 10.8633. \)
Outline

1. Introduction: why investigating search neutrality?
2. Model
3. Numerical comparison
4. Case of a general set of keywords
\[ \lambda(r) \text{ the total rate of requests, not for a single keyword} \]

- Each time only \( m \) CPs are candidate to be displayed (can be generalized)
- The relevance of any seen CP comes from a distribution \( F \) (independent relevance between CPs)
- The gain of any seen CP comes from a distribution \( G \) (independent between CPs)

**New measures (neutral case):**

- Average relevance \( r \)
  \[
  r = \sum_{j=1}^{m} \theta_j \int y \frac{m!}{(m-j-1)!j!} F(y)^{m-j-1}(1 - F(y))^j f(y) dy.
  \]

- Average CP revenue: \( \mathbb{E}[U^{\text{CP}}] = \lambda(r) \mathbb{E}[\theta_J G] \), with \( J \) random rank of the CP (uniform on \( \{1, \ldots, m\} \), which depends only on the relevance).
Non-neutral situations

Ranking depends in general on both the relevance (through the rate $\lambda(r)$) and the gains.

Let $\pi$ optimal family of permutations, defined for each possible configuration $((R_1, G_1), \ldots, (R_m, G_m))$

- If the SE gets a proportion $\alpha$ of CPs’ revenues:

\[
\pi = \arg\max_{\text{permutations}} \delta \ h(\lambda(r)) + \alpha \lambda(r) \mathbb{E} \left[ \sum_{j=1}^{m} \theta_{\delta(j)} G_j \right]
\]

with $r$ the average over such optimally chosen permutations of relevance.

- If the SE owns a CP:

\[
\pi = \arg\max_{\text{permutations}} \delta \ h(\lambda(r)) + \lambda(r) \mathbb{E} \left[ \sum_{j=1}^{m'} \theta_{\delta(j)} G_j \right].
\]
Can be solved using dynamic programming

Difficulty: the “reward” associated to a transition depends on the optimal policy through $\lambda(r)$, a non-classical assumption.