

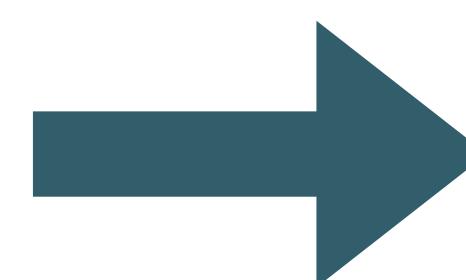
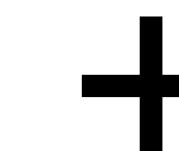
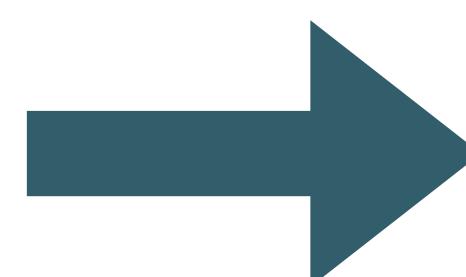
Examples of recommendation systems ?

What can be recommended?

How recommender systems arrived on the internet

From the street shop to the on-line shop
From the material content to the digital content

Storage, shipping, duplication, costs (shops and items) have evolved,
helping make large catalogues available



Properties of large catalogues

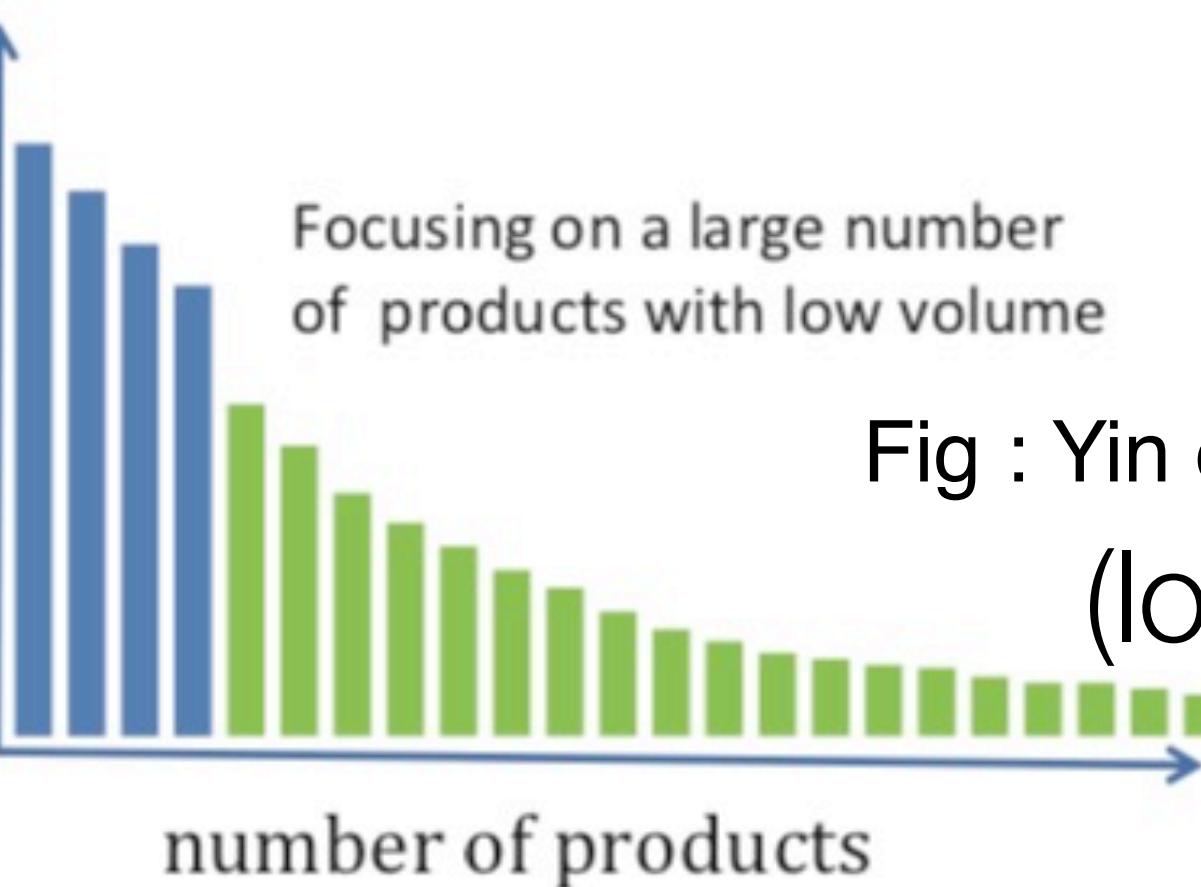
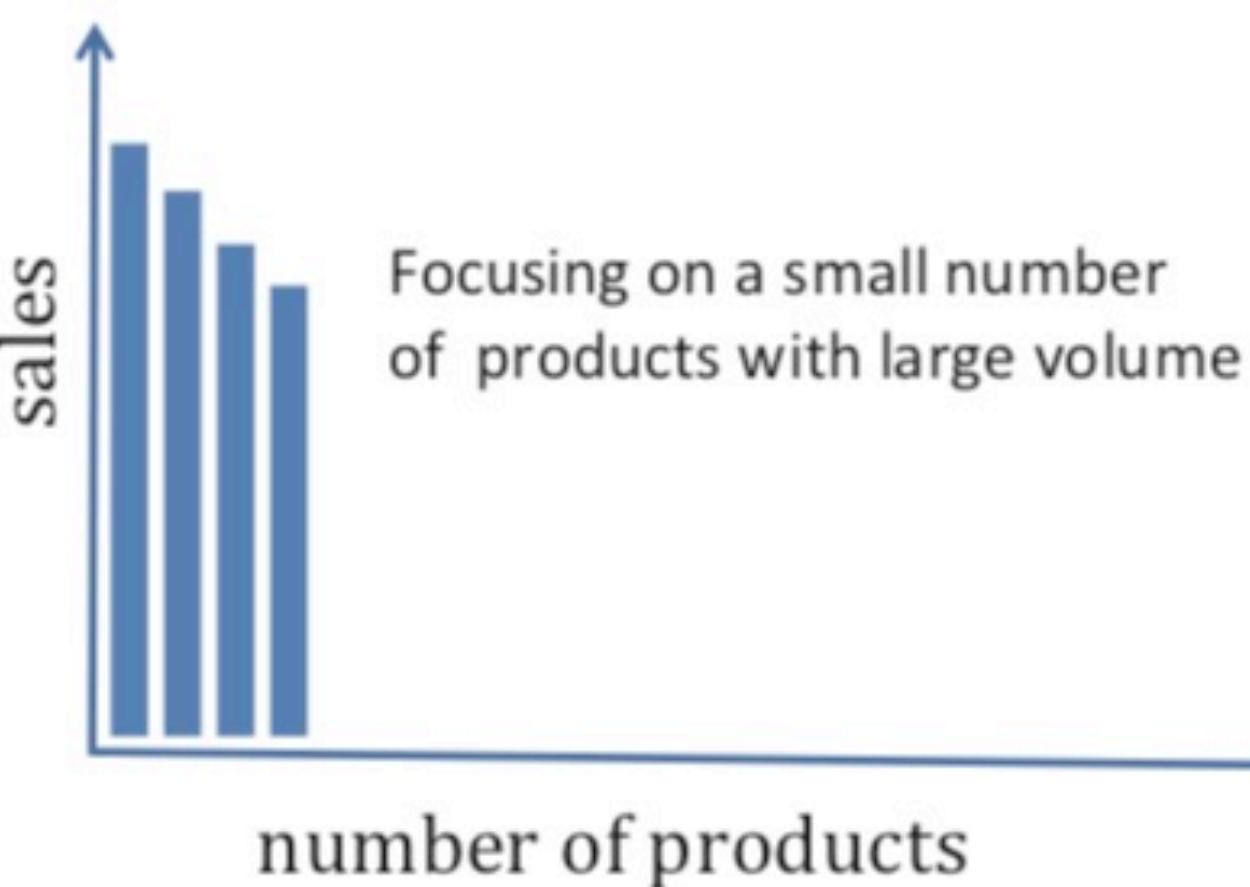


Fig : Yin et al. PVLDB 2012
(long tail)

Margin is higher on rare items (less competition)

Bonus of one-stop shopping

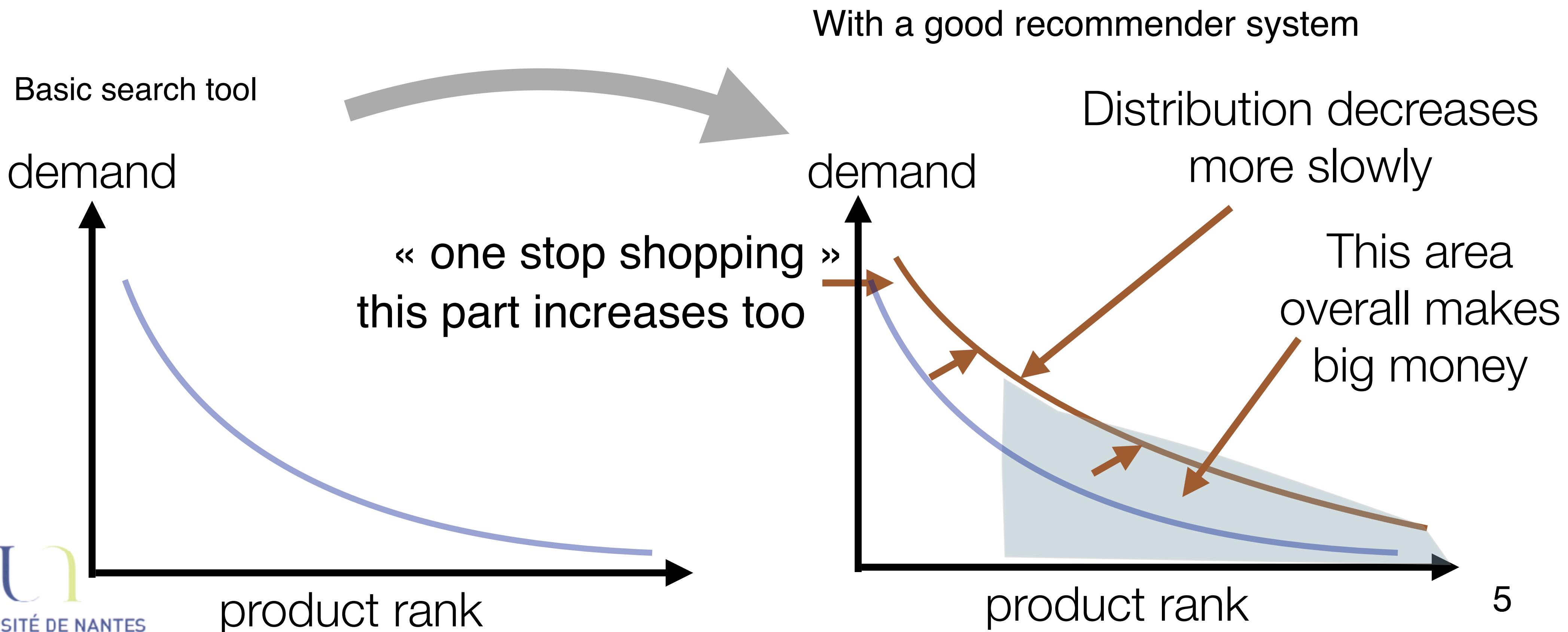
Nice to have a large catalogue (sub-contracted ?)

But need to be able to exploit it

(otherwise the long train might not be beneficial)

Large catalogue => hard to exploit to sales
because querying and browsing are not so easy

Experiments show you can make the demand much higher
by adding recommendation to browsing and querying



3 approaches to content selection

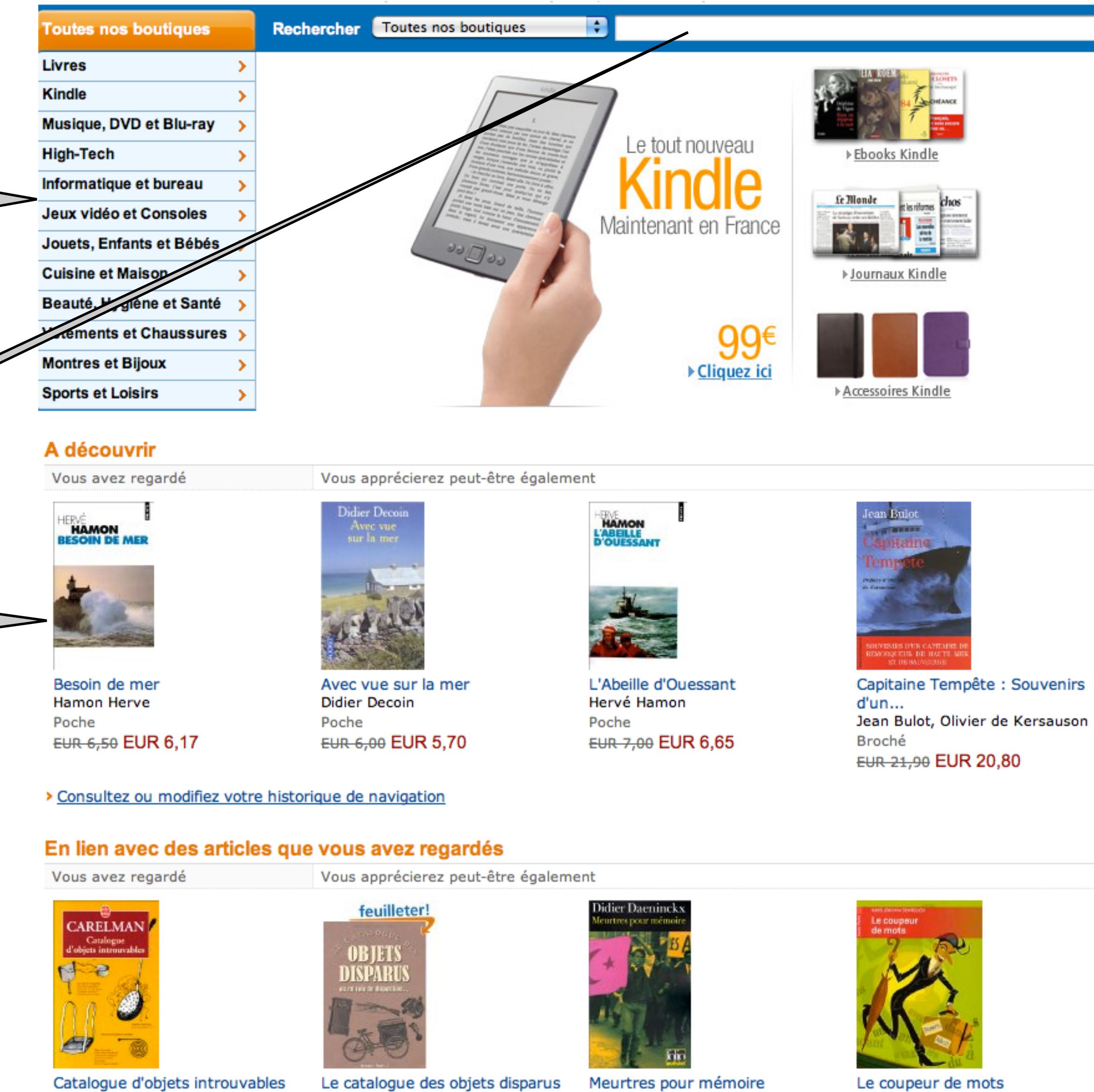
example : Amazon

Browsing

Querying

Recommendation

But wasn't
browsing/querying enough ?



The screenshot shows the Amazon France homepage with three callout boxes pointing to specific features:

- Browsing:** Points to the sidebar menu titled "Toutes nos boutiques" which lists categories like Livres, Kindle, Musique, DVD et Blu-ray, etc.
- Querying:** Points to the search bar at the top of the page.
- Recommendation:** Points to the "A découvrir" section, which displays recommended books based on previous viewing history.

The main content area features a large advertisement for the Kindle, a "You have viewed" section, and a "Books you may like" section.

Section	Content	Price
Books You've Viewed	Besoin de mer Hamon Hervé Poche EUR 6,50 EUR 6,17	
	Avec vue sur la mer Didier Decoin Poche EUR 6,00 EUR 5,70	
	L'Abeille d'Ouessant Hervé Hamon Poche EUR 7,00 EUR 6,65	
	Capitaine Tempête : Souvenirs d'un... Jean Bulot, Olivier de Kersauson Broché EUR 21,90 EUR 20,80	
Consultez ou modifiez votre historique de navigation		
En lien avec des articles que vous avez regardés		
Books Recommended	Catalogue d'objets introuvables CARELMAN	
	feuilleter! Le catalogue des objets disparus	
	Didier Daeninckx Meurtres pour mémoire	
	Le coupeur de mots	



Catalogue = a **search space**

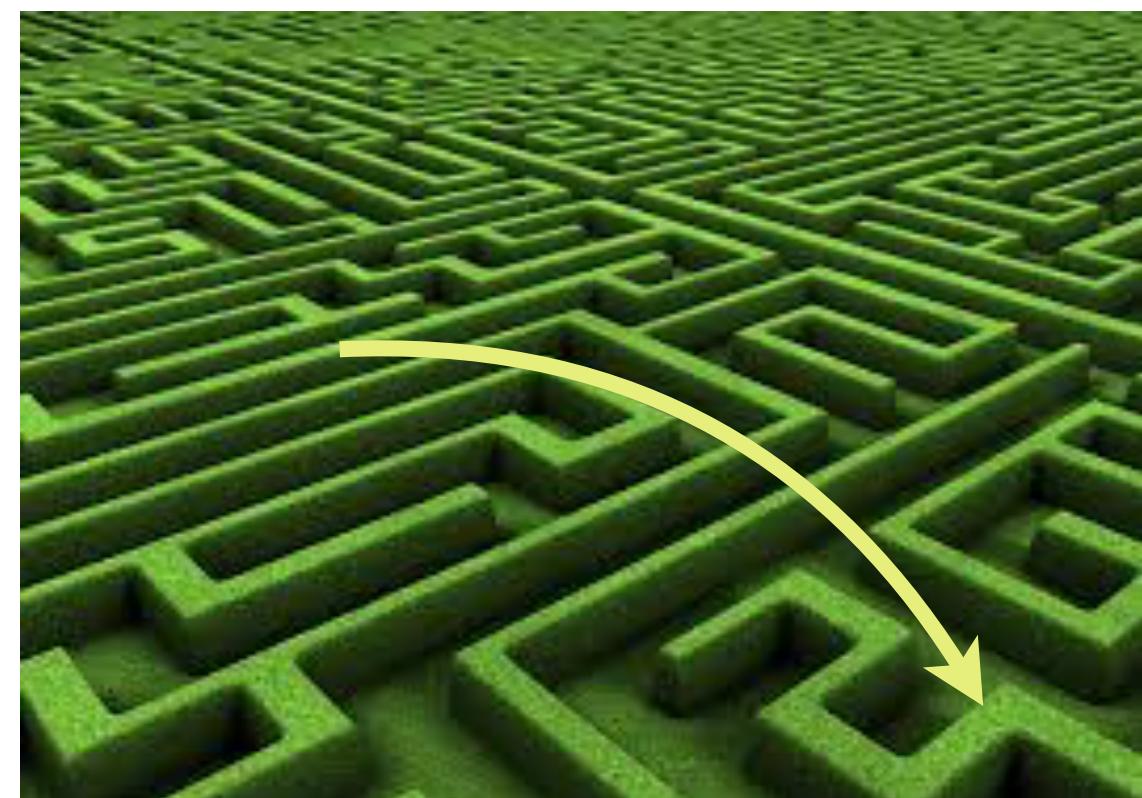
Querying, browsing (generalization) :
they enable exploring the space,
but most corners remain hard to reach
for most user knowledge

Serendipity :

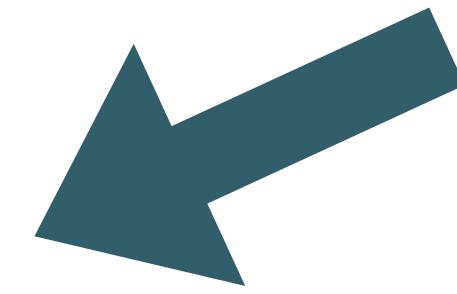
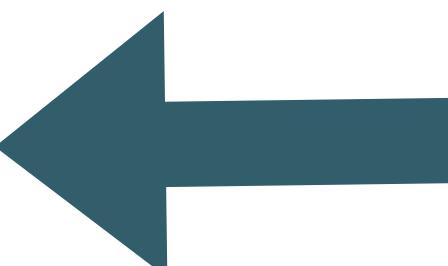
is a goal, a property (not a method) consisting in finding :

- interesting content
- that would not been found easily

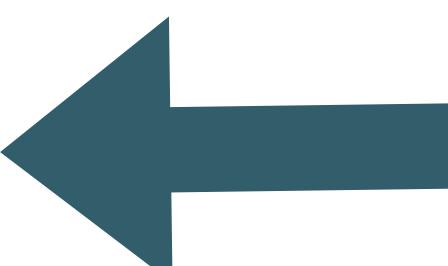
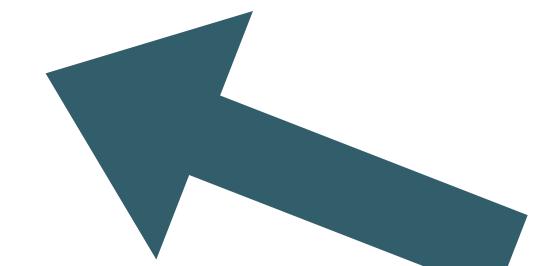
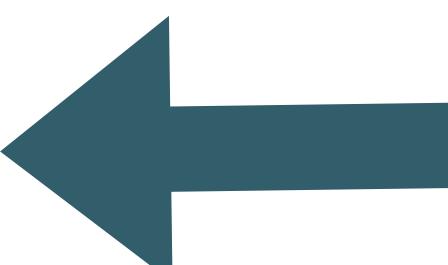
Ability to jump to areas hard to reach spontaneously
by querying/browsing : this is one of the goals
of recommendation



From the on-line shop to the intermediation platform



The catalogue becomes even larger
And platforms can share providers :



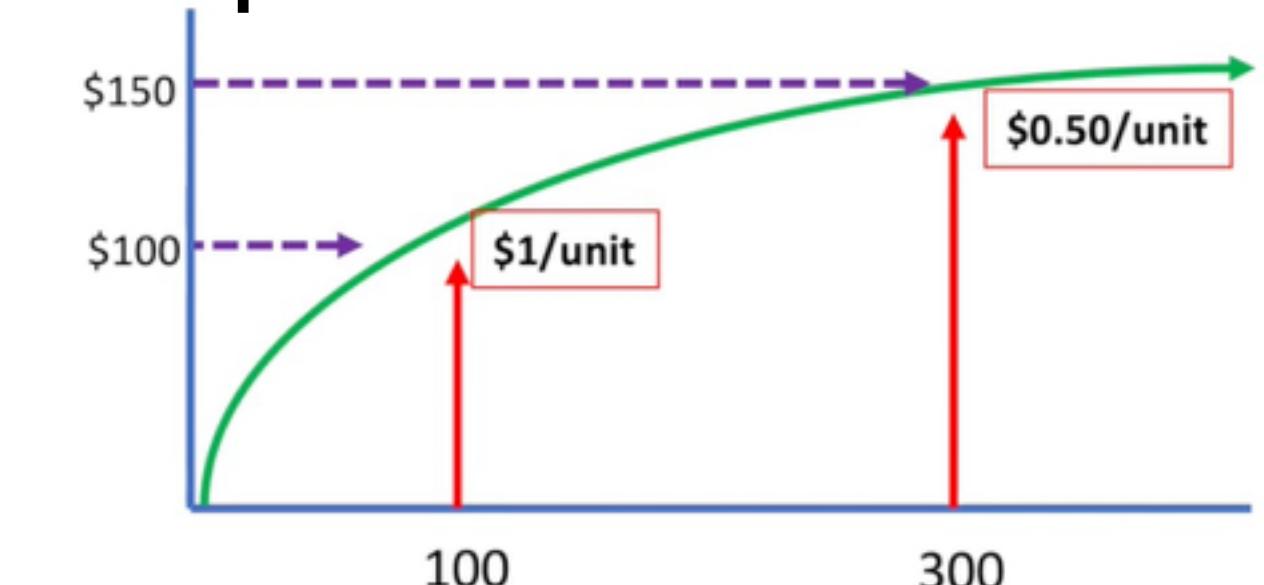
The value of the portal is the catalogue AND the power of the search tool (including recommendation)

Network effects on platforms

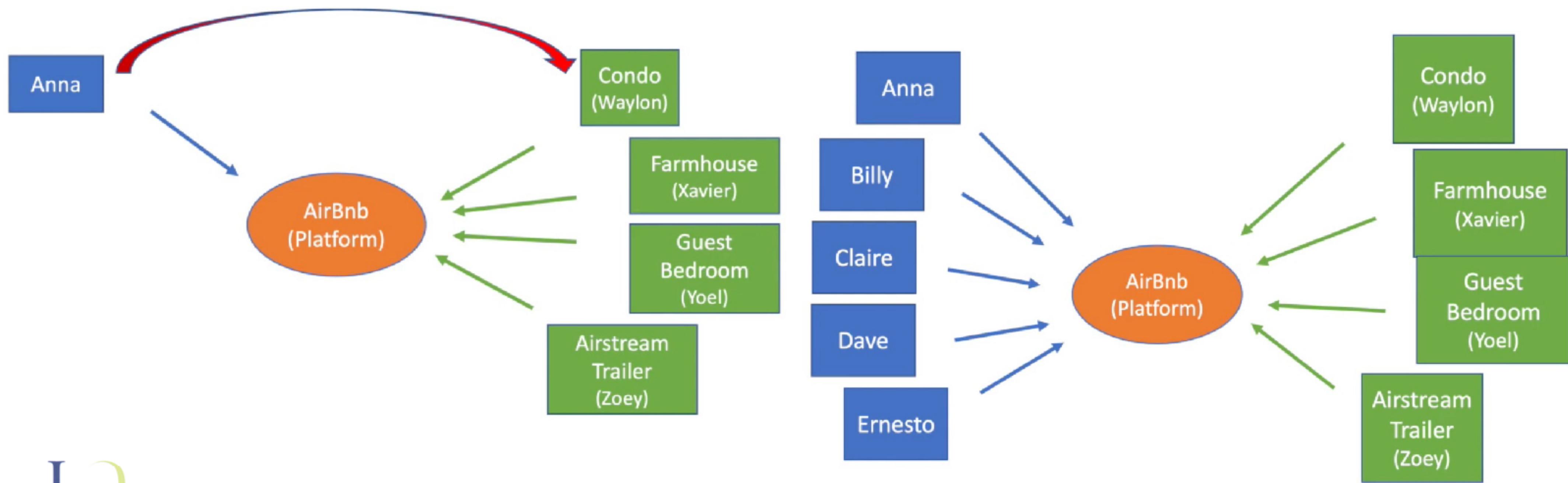


- Zoom
- MS Excel
- Instagram

cost for producer



Economy of scale



Offer, prices, competition

The platform is a two-sided market (an economics concept)
 (providers, consumers)
 with cross-side network effects

ex. Uber, Airbnb, Youtube :

- + more consumers => better for providers
- + more providers => better for consumers

making links = their business model !



Difficult to start a business (chicken and egg start), hence huge existing systems.

Free of charge for one side helps create dynamics on this side.
 Either one or the other side has to pay for the platform, sometimes both.

Recommender systems are often used for helping establish the link
 between customer and content from an external content provider,
 i.e. they operate on a platform.

The recommender system is a core tool for building the quality of a platform

Stakeholders

Who does Spotify serve ?

- End-users ?
- Singers ?
- Spotify ?

Optimizing recommendation is optimization a function involving all 3

Crowdsourcing, description, evaluation

A society of data crowdsourcing and « digital labor »

- recommender systems rely massively on content provided by the crowd of users, who « collaborate » to create a corpus that benefits to « all » fourniture de contenus à recommander
- Users provide:
 - content (stuff for others to read, view, buy,...)
 - meta-data over their content and other people's content (eg film topics)
 - Evaluations feed computation of reputation of everything on the platform
 - of content
 - of people (peers)

A society of data crowdsourcing and « digital labor »

- Platforms present themselves as a community of peers+content that cooperate, to hide a centralized+commercial control and goals and favour free work by users
- They provide incentives (non financial) :
 - offer social visibility (better than IRL) to people, contents, opinion
 - social recognition of expertise and taste
 - provide everyone with numerical indicators of social position
 - privileged access to rights, content, jobs,....
- Sometimes digital labor is a real job, poorly paid as micro-tasks controlled by automata.

Listen to :
(france culture)



Crowdsourcing

Is also a scientific topic about modelling, algorithms, maths, economics.

How to recruit and retain contributors ?

How to design and compute incentives ?

- evaluate quality of contributors and contributions
 - evaluation by peers, combined with
 - automated statistical analysis
 - how to combine contributions ?
 - need to aggregate evaluations and to take contributor reputation into account in this process
 - this is a joint problem that needs to be solved iteratively

Tagging is categorization
but categories may overlap (multiple tags)

Tagging

Simple Man-Machine Interaction :

- values but no attributes (even implicit)
- values from open vocabulary (often)
- requires little user competence
- puts little constraints on user

- > incentive for tagging
- > suitable for crowdsourced tagging

- Tagging your own items
- Tagging existing items

- > incentive = creators are motivated to tag their content well so that it gets retrieved by other people

Tags

Ecomarathon • Eurospeedway Lausitz •
Germany • Europe • 2010 • Polytech' Nantes •
Polyjoule • prototype



Collaborative tagging :

Tags

Tags given by user U1

Ecomarathon • Eurospeedway Lausitz •

Germany • Europe • 2010 • Polytech' Nantes •
PolyJoule • prototype

Tags given by user U2

No constraint vocabulary :
incremental collaborative construction :
not only are tags crowdsourced,
but also vocabulary construction and maintenance.



Camion mecede 809

Enjoliveur mercedez

Interieur cuir noir mescedes 190

MERCEDEE E 320 avangarde

Mercedesse



Such a vocabulary (:=folksonomy) can be unstable

- bad for matching
- but good for including new concepts with no central control
- overall, people tend to tag using popular tags (« rich get richer »)

Co-occurrence statistical analysis from tagged items

Goal : find pseudo-semantic relations between tags, by computing statistical co-occurrences.

Ex. : Polytech and Polyjoule frequently present together on Flickr

Ex. : When tag *scipy* on Stackoverflow, *python* also present 90% of time
but when *python* is present, *scipy* present only 12% of time
=> *scipy* is a subpart of *python* (*scipy* is a famous *python* package)

These statistics are more reliable
if there are few tags used many times rather than many tags used once

=> Use these findings for query expansion

further possibility : consider who tags in the analysis 18

Analyze the (temporal) dynamics of tags

Goal : identify short events
(vs. more stable entities)

ex. GreenCodingChallenge
in Nantes with
#GCLChallenge

highly active during 3 days

can also find relation
to (GPS-like) geolocation, if available.



IUTNantes-INFO @IUTNantes_INFO

#GCLChallenge @IUTNantes avec un peu de @CentraleNantes

25m



IUTNantes-INFO @IUTNantes_INFO

#GCLChallenge @IUTNantes

27m



Re-introducing some control over the vocabulary

To further stabilize the vocabulary

systems may suggest tags (ex. Amazon, Stackoverflow)

Mots-clés inspirés de produits similaires ([De quoi s'agit-il ?](#))
 Soyez le premier à ajouter un mot-clé pertinent (fortement associé à ce produit)

Cochez une des cases ou entrez vos propres mots-clés ci-dessous

<input type="checkbox"/> jacques attali (10)	<input type="checkbox"/> crise (6)	<input type="checkbox"/> nul (4)
<input type="checkbox"/> attali (9)	<input type="checkbox"/> escroquerie (6)	<input type="checkbox"/> fumisterie (3)
<input type="checkbox"/> dette publique (7)	<input type="checkbox"/> dette (4)	
<input type="checkbox"/> arnaque (6)	<input type="checkbox"/> littérature de chiotte (4)	

Vos mots-clés : [Ajouter](#)

(Appuyer deux fois sur la touche 'T' pour accéder rapidement à la fenêtre "Associer des mots-clés à ce produit".)

Attempts

I tried to achieve the following by adding the code:

```
mt_sum <- mtcars %>%
  group_by(am, gear) %>%
  summarise(n = n()) %>%
  spread(key = am, value = n) %>%
  mutate_each(funs(../rowSums(.)))
```

but it returns the following error:

Error: 'x' must be an array of at least two dimensions

Hence my question: **how can I add extra columns with row percentage values in dplyr ?**

Side points

- I would prefer blank values instead of NAs
- The table could be easily build with use of `CrossTable` in `gmodels` but I would like to stay in `dplyr` as I want to keep as many transformations as possible in one place

[r](#) [aggregate](#) [dplyr](#) [frequency](#) [crosstab](#)

[share](#) [improve this question](#)

asked 1 hour ago

 Konrad
1,953 ● 8 ● 22

Avoid divergence of vocabulary :

- You must use existing tag (stackoverflow when beginner)
- your question is more likely to be answered if you tag according to common practice

Describing objects :
automatically using external (web) resources
for processing annotations (correcting + extending). Ex. : wikipedia

Tag input



Google

fr.wikipedia.org chapelle sur edre

Environ 605 000 résultats (0,44 secondes)

▶ [La Chapelle-sur-Erdre - Wikipédia](#)  

La Chapelle-sur-Erdre est une commune française située dans le département de la Loire Atlantique et la région Pays de la Loire. ...

fr.wikipedia.org/wiki/La_Chapelle-sur-Erdre - En cache - Pages similaires

Situation

Tag corrected/
normalised

 [Vidéos](#)

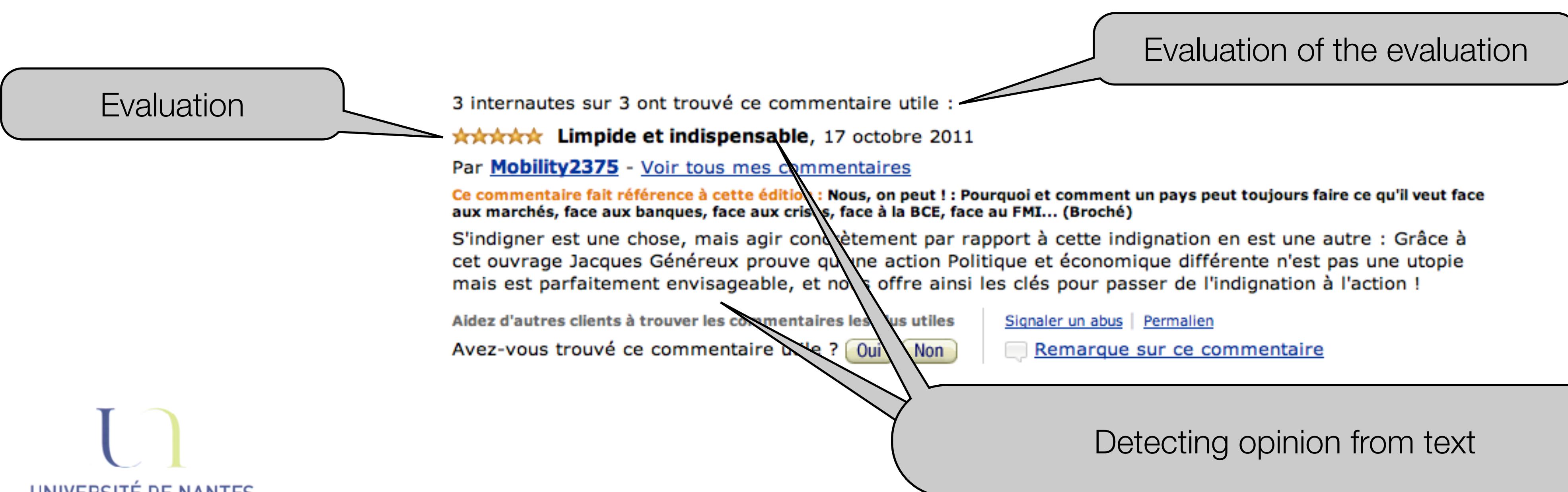
Additionnal tags
as an output

Cette ville est située à 13 km au nord de
Nantes, à la confluence de l'Erdre, du
Gesvres et de l'Hocmard.

Les communes limitrophes sont Nantes,
Carquefou, Sucé-sur-Erdre, Grandchamps-
des-Fontaines et Treillières.

Evaluating objects

- Ratings (1,2,3,4,5), good/not good ; ordering (« better than »)
- Text (algorithms for «sentiment analysis» on blogs/facebook posts/tweets/ amazon «book damaged vs. uninteresting»)
- Unary value (bought/read/like),
the alternative does not necessarily mean «bad» but maybe «no information»
(various possible interpretations : not seen, already bought elsewhere, ...)
- I agree, I don't agree (evaluating an evaluation)
- Aggregate multiple contributions (tags, evaluations) considering reputation



Evaluation

Evaluation of the evaluation

Detecting opinion from text

3 internautes sur 3 ont trouvé ce commentaire utile :
★★★★★ **Limpide et indispensable**, 17 octobre 2011
Par [Mobility2375](#) - [Voir tous mes commentaires](#)
Ce commentaire fait référence à cette édition : **Nous, on peut ! : Pourquoi et comment un pays peut toujours faire ce qu'il veut face aux marchés, face aux banques, face aux crises, face à la BCE, face au FMI... (Broché)**
S'indigner est une chose, mais agir concrètement par rapport à cette indignation en est une autre : Grâce à cet ouvrage Jacques Généreux prouve qu'une action Politique et économique différente n'est pas une utopie mais est parfaitement envisageable, et nous offre ainsi les clés pour passer de l'indignation à l'action !

Aidez d'autres clients à trouver les commentaires les plus utiles
Avez-vous trouvé ce commentaire utile ? [Signaler un abus](#) | [Permalink](#)

[Remarque sur ce commentaire](#)