VTT TECHNICAL RESEARCH CENTRE OF FINLAND LTD

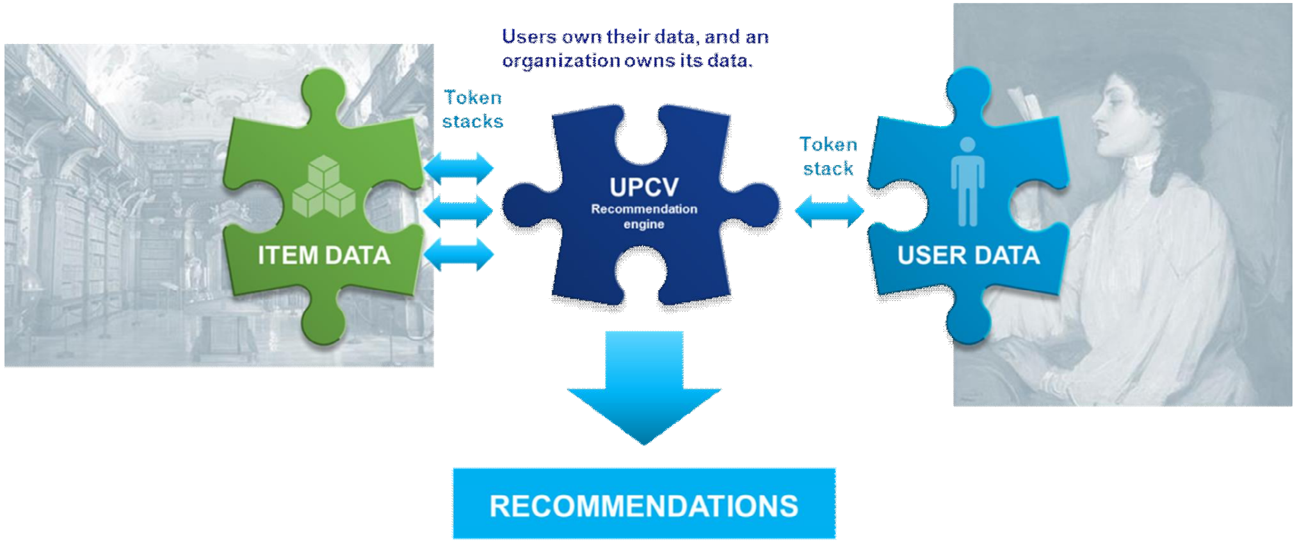
Evaluating the performance and privacy of a token-based collaborative recommender

INRA 2017, August 23rd, 2017, Leipzig

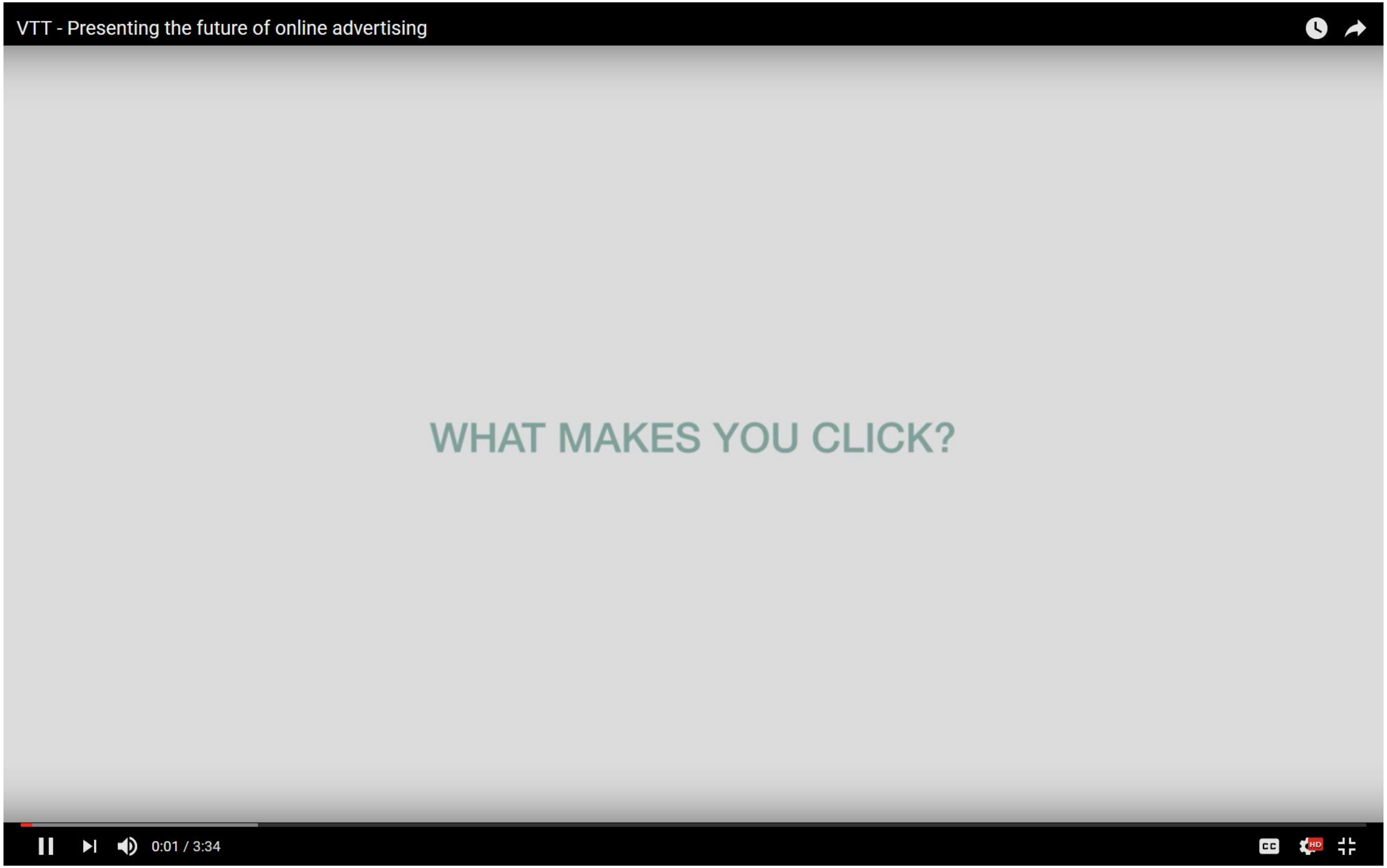
Ville Ollikainen & Valtteri Niemi

Background

- VTT has developed a **collaborative recommendation method** which is based on **exchanging random numbers, “tokens”**.
  - Each party can own their own data (tokens, that is) and control the use of it.
- This is in harmony with **General Data Protection Regulation** by EU
- Furthermore, these tokens do not carry any history; we claim that exchanging these tokens is safe from privacy point of view.



This is how it works… (a use case)



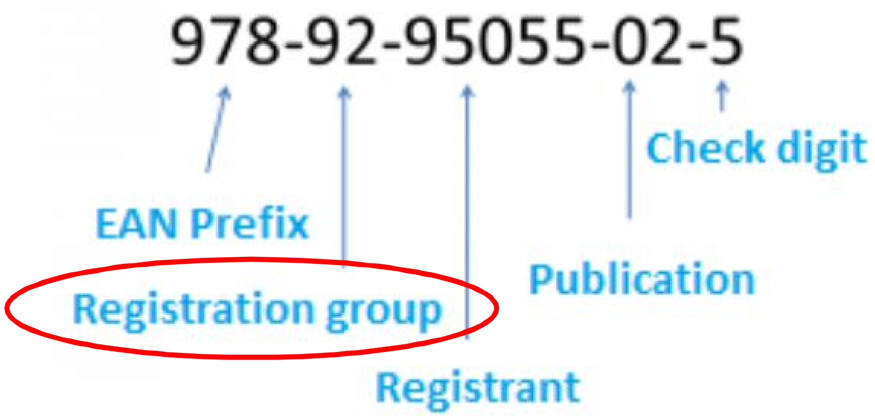
VTT - Presenting the future of online advertising

WHAT MAKES YOU CLICK?

Link to video: https://www.youtube.com/watch?v=jxCVR2CMKHA

That was about the recommendation method itself

ISBN (International Standard Book Number)



- First digit(s) after the prefix represent a registration group (”agency”):
  - Language code for English (0 and 1), French (2), German (3)
  - National agency otherwise (Finland, Norway, Netherlands, Seychelles…)
- Note: ISBN refers to manifestation of write art, not writer art itself
  - i.e. from ISBN you can NOT say who was the writer or single writer art
  - e.g. there are lots of ISBN’s for ”Adventures of Huckleberry Finn”, one for each publication - over decades.

Book-Crossing (”BX”) database; some issues

- Book-crossing is an **open community for exchanging second hand books**
- In general, books are left to a random location with an instruction sticker.
- Someone finds them and registers the action.

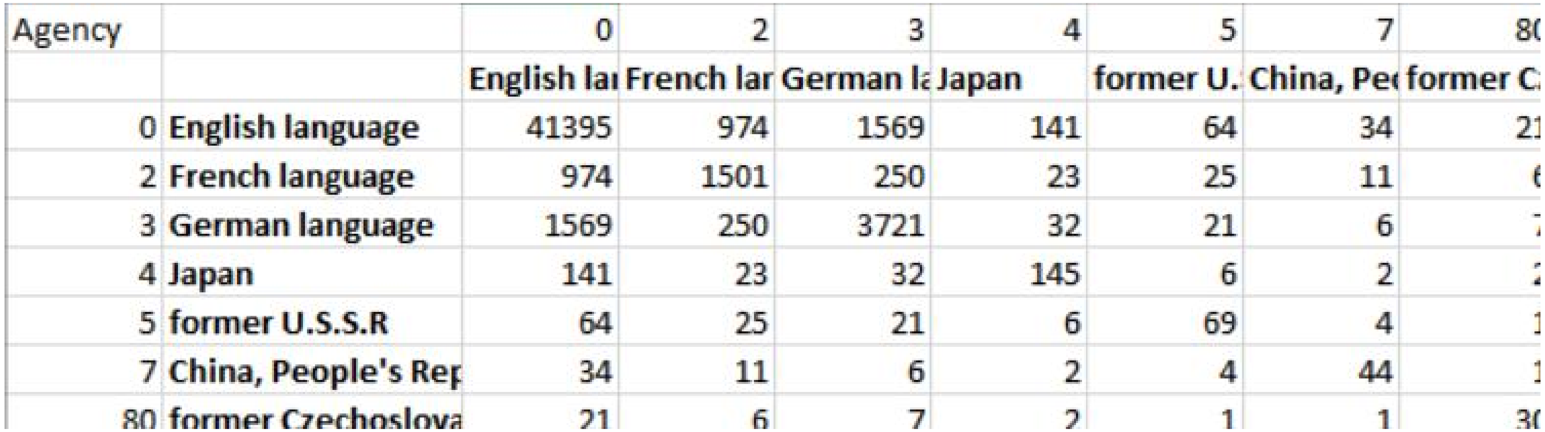
=> **Users pick the books by chance, not by selecting them.**

- When you have read the book you found, you MAY rate it 1..10
- If you don’t bother to rate it, the ”rating” is becomes ambiguous 0
  - A common practice is to **treat ’0’ ratings as missing data; that’s wrong**!
  - ’0’ ratings count for 62% of all ”ratings”
  - ’0’ does not tell, if you bothered to read it at all (truly ’0’), or just did not rate
  - The remaining ratings are highly biased with median value of 8

=> **Understand BX process before using BX dataset**

Step 1: What we found usable in BX - ISBN agencies

- Books are physical objects; **tendency to circulate within a region**.
- We created a permutation matrix containing agencies (no single-visit users):



Agency		0	2	3	4	5	7	80
		English la	French lar	German la	Japan	former U.	China, Pe	former C
0	English language	41395	974	1569	141	64	34	21
2	French language	974	1501	250	23	25	11	6
3	German language	1569	250	3721	32	21	6	7
4	Japan	141	23	32	145	6	2	2
5	former U.S.S.R	64	25	21	6	69	4	1
7	China, People's Rep	34	11	6	2	4	44	1
80	former Czechoslova	21	6	7	2	1	1	30

- Over all users: if a user had registered even a single e.g. English and even a single French book, the corresponding cell in the matrix was incremented.
- Each column was sorted in decrementing order

=> ground truth for any agency; other agencies in order of relevancy.

Step 2: Prepare transaction data sets A and B
Step 3: Create recommendations for A and B

- Create **transaction log: UserID – ItemID pairs**
  - ItemID was the agency
  - Shuffle it into random order
  - Divide into two halves: A and B
- First,
  - Train the recommender with transactions in A
  - Create Agency-Agency recommendations for each agency
  - Create Ground truth for B (previous slide)
  - Compare recommendations with the ground truth (how: next slide).
- Second,
  - Swap A and B and do the same

Step 4: Compare recommendations with ground truth

- Kendall Tau is a metric to compare the order of items in two lists
  - **Are each pair of two items in the same order in both lists?**
    - Does not matter how far they are, only their mutual order counts

$$\tau = \frac{(\text{number of concordant pairs}) - (\text{number of discordant pairs})}{n(n-1)/2}$$

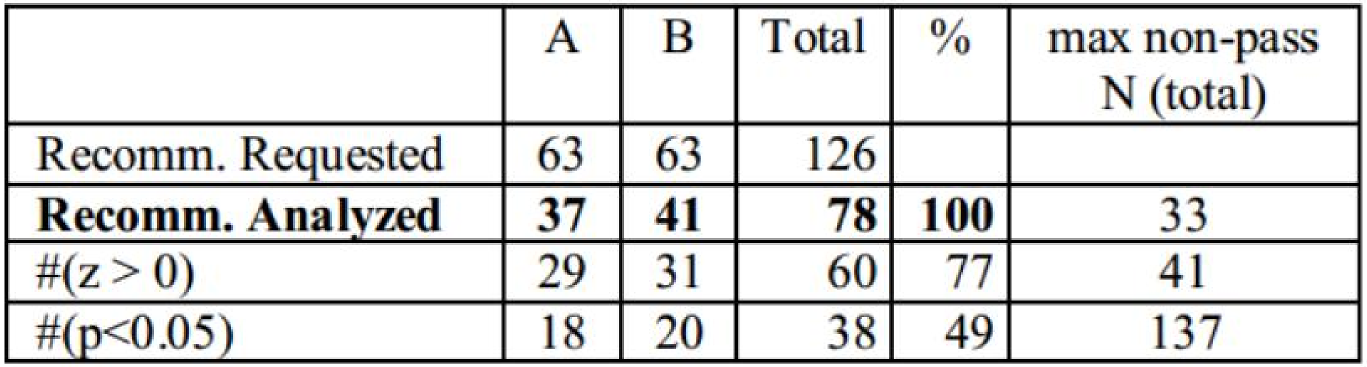
  - Tau-b has an adjustment for lists that contain ties (like ours)

  => calculate statistical significance

- One list is the **recommendation list** for an agency (set A or B)
- **Ground truth list** is based on the other set (set B or A, respectively)
- 63 agencies had users in both sets => 63 Tau-b's

Results 1/2

- 78 out of 126 recommendation requests were successful
  - Found similarities in token collections, if an agency had >33 users

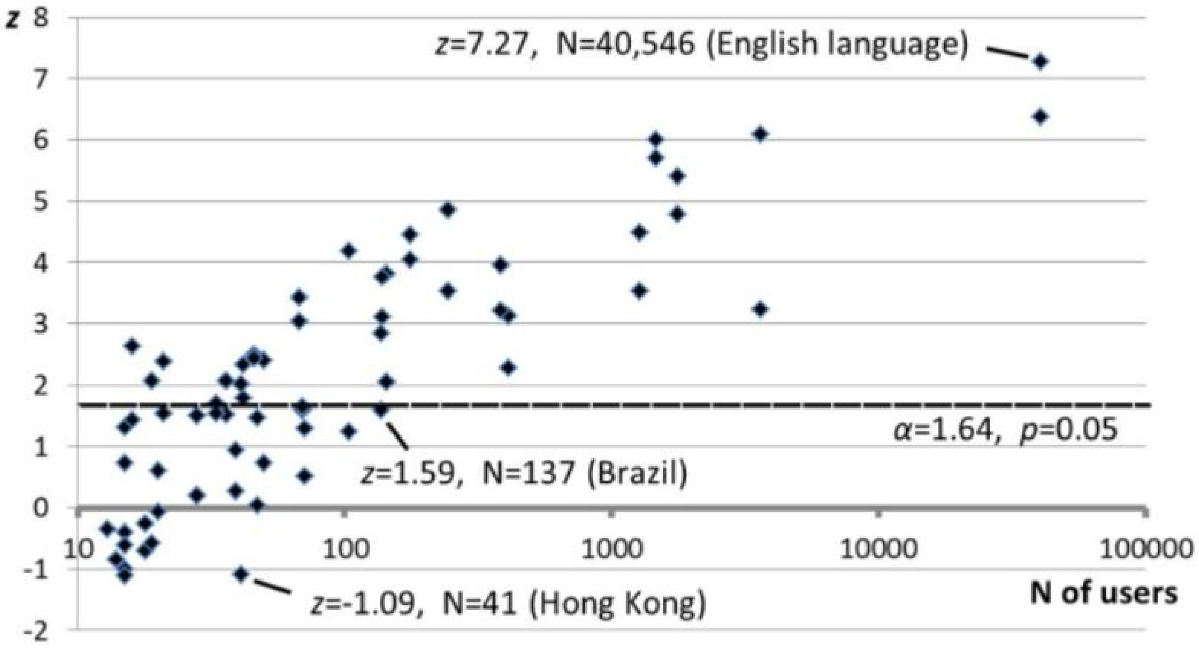


	A	B	Total	%	max non-pass N (total)
Recomm. Requested	63	63	126		
Recomm. Analyzed	**37**	**41**	**78**	**100**	33
#(z > 0)	29	31	60	77	41
#(p<0.05)	18	20	38	49	137

- All recommendations of agencies with > 41 users had positive correlation (Tau-b) with the ground truth

Results 2/2

- All recommendations of agencies with > 137 users passed p=0.05 significance test



Summary of privacy considerations

- Related to details presented in the paper…

- Recommendations are based on aggregating token collections.
- Tokens float around the system and are not associated with anything in the real world.
- Tokens are random values without any history data.
- If a token collection becomes disclosed to an adversary, the adversary is not able to deduce, where the token came from
  - **In the case of similarities with other users and items there is plausible deniability that the token has propagated from somewhere else.**

TECHNOLOGY FOR BUSINESS