SOCIAL MEDIA DATA MINING AND USER EMPOWERMENT

Ali Padyab, PhD

Information Systems

Department of Computer Science, Electrical and Space Engineering

LULEÅ UNIVERSITY OF TECHNOLOGY

AGENDA

- Introduction of myself
- Social media data mining
- Why do we need privacy? The role of awareness
- Introduction to USEMP project tools: DataBait (Collaboration between LTU and CEA)
- Demonstration of inferences in Facebook
- Implications for digitalization
- Q&A

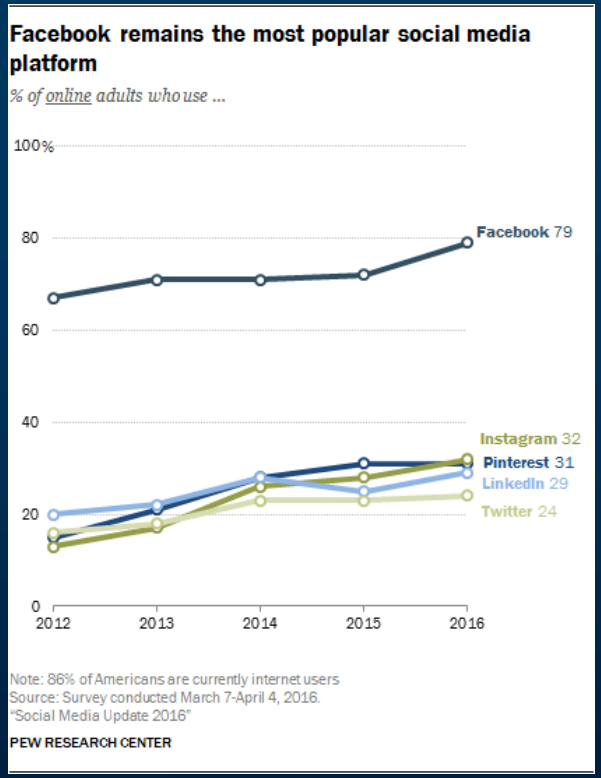
LULEÅ UNIVERSITY OF TECHNOLOGY

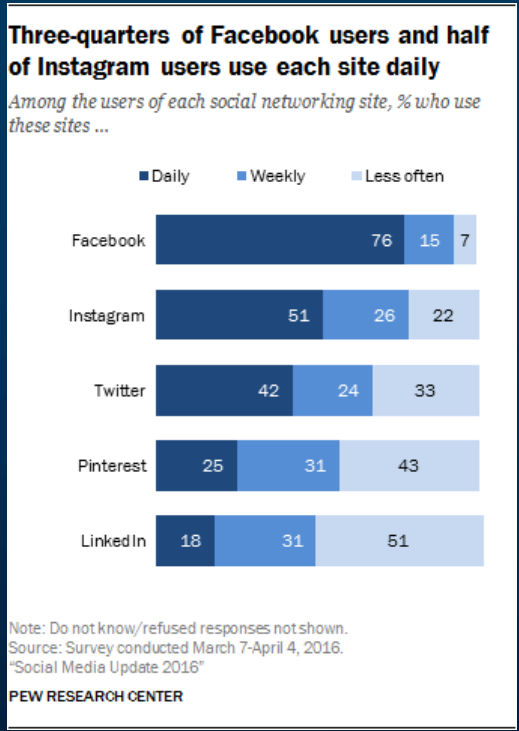
INTRODUCTION

- Ali Padyab, PhD
- Field of research: Information Privacy, end user attitude, Privacy Enhancing Tools, IoT, Smart city

LULEÅ UNIVERSITY OF TECHNOLOGY

SOCIAL MEDIA USERS

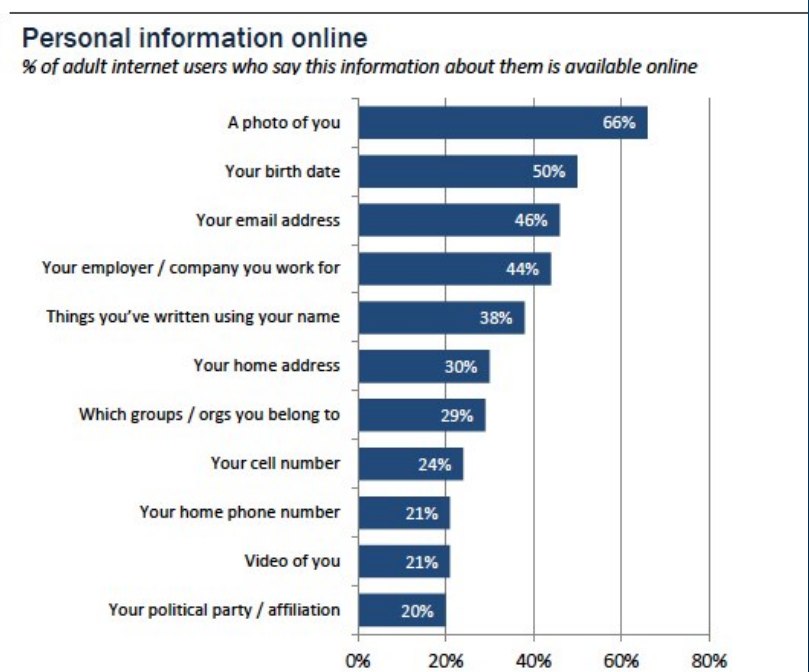




LULEÅ UNIVERSITY OF TECHNOLOGY

Source: http://pewrsr.ch/2fiOTBE

PERSONAL INFORMATION ONLINE

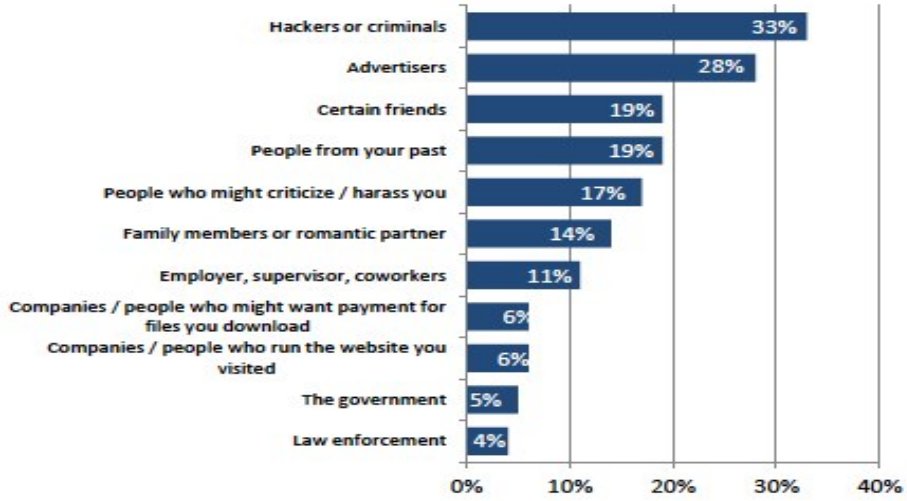


LULEÅ UNIVERSITY OF TECHNOLOGY

Source: http://pewrsr.ch/1iPhyaY

Who users try to avoid

% of adult internet users who say they have used the internet in ways to avoid being observed or seen by ...



Source: Pew Research Center's Internet & American Life Project Omnibus Survey, conducted July 11-14, 2013, on landline and cell phones. N=792 for internet users and smartphone owners. Interviews were conducted in English on landline and cell phones. The margin of error on the sample is +/- 3.8 percentage points.

LULEÅ UNIVERSITY OF TECHNOLOGY

Source: http://pewrsr.ch/1iPhyaY

ACCESS TO OSN DATA

Broad concern about government and third-party access to info on social networking sites

% Among adults ages 18 and older who use social networking sites

How concerned are you that some of the info you share on social networking sites might be accessed by ___ without your knowledge?

	The government	Third parties like advertisers or businesses
Very concerned	37	35
Somewhat concerned	34	45
Not too concerned	25	17
Not at all concerned	4	2

Source: Pew Research Privacy Panel Survey, January 2014. N=607 adults, ages 18 and older.

PEW RESEARCH CENTER

LULEÅ UNIVERSITY OF TECHNOLOGY

Source: http://pewrsr.ch/1EtEFQZ

WHAT INFORMATION COLLECTED

ACCORDING TO PRIVACY POLICIES

facebook

What kinds of information do we **collect**?

- Things you do and information you provide.
- Things others do and information they provide.
- Your networks and connections.
- Information about payments.
- Device information.
- Information from websites and apps that use our Services.
- Information from third-party partners.
- Facebook companies

Google

What kinds of information do we **collect**?

- Information you give us
- Information we get from your use of our services
  - Device information
  - Log information
  - Location information
  - Unique application numbers
  - Local storage
  - Cookies and similar technologies

LULEÅ UNIVERSITY OF TECHNOLOGY

WHAT ARE USES ACCORDING TO PRIVACY POLICIES?

facebook

How do we **use** this information?

- Provide, improve and develop Services.
- Communicate with you.
- Show and measure ads and services.
- Promote safety and security.

Google

How we **use** information we collect?

- …to provide, maintain, protect and improve them, to develop new ones, and to protect Google and our users. … to offer you tailored content – like giving you more relevant search results and ads.
- …to improve your user experience and the overall quality of our services.
- When showing you tailored ads, we will not associate an identifier from cookies or similar technologies with sensitive categories.

LULEÅ UNIVERSITY OF TECHNOLOGY

SOME EXAMPLES OF USE OUTSIDE OF DATA HOLDERS…

LULEÅ UNIVERSITY OF TECHNOLOGY

Source: https://tinyurl.com/3czsczw

SECTIONS SEARCH NEW YORK POST

METRO

Thieves steal $1M in jewels from Jerry Seinfeld's ex-girlfriend

By Larry Celona and Joe Tacopino

July 20, 2016 | 3:39am

Cops investigate a burglary at the home of Jerry Seinfeld's ex Shoshanna Lonstein Gruss.

LULEÅ UNIVERSITY OF TECHNOLOGY

Source: http://nyp.st/2at38iB

This is MONEY.co.uk
FINANCIAL WEBSITE OF THE YEAR

Going away? Don't tell your Facebook friends or risk having your insurance claims rejected

- **If you post photos of holidays you risk having insurance claims rejected if your home is burgled**
- **Allowing your Facebook page to reveal your location automatically can also be risky**

By RUTH LYTHE FOR MONEY MAIL
PUBLISHED: 00:10 BST, 22 April 2015 | **UPDATED:** 09:32 BST, 22 April 2015

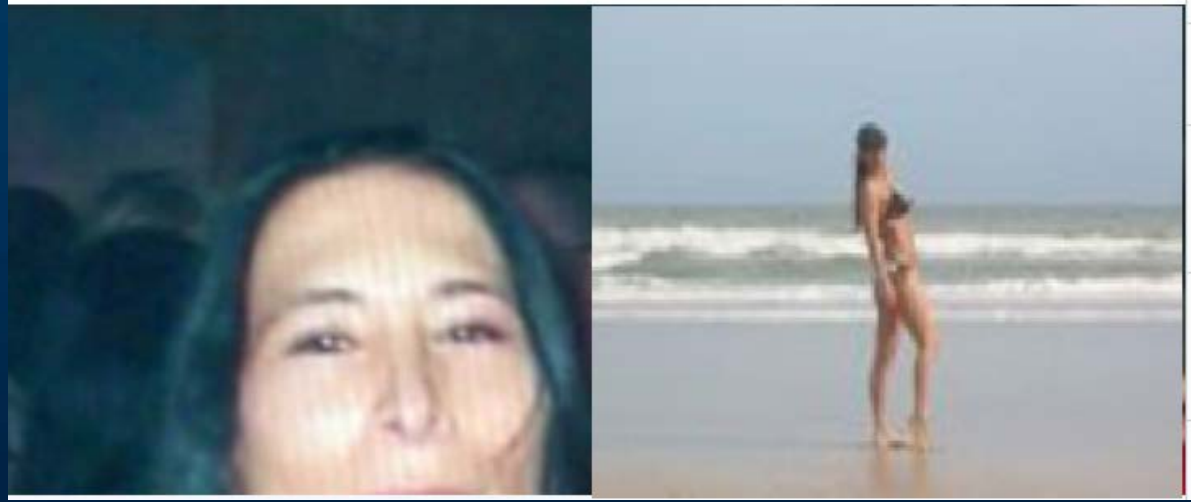
Source:
https://tinyurl.com/kb89zol

abc NEWS U.S. International Politics Lifestyle ...

Woman Loses Benefits After Posting Facebook Pics

By KI MAE HEUSSNER
Nov. 23, 2009

Share with Facebook Share with Twitter



LULEÅ UNIVERSITY OF TECHNOLOGY

Source: https://tinyurl.com/ybwo7b6

INDIRECT INFORMATION USE

Examples:

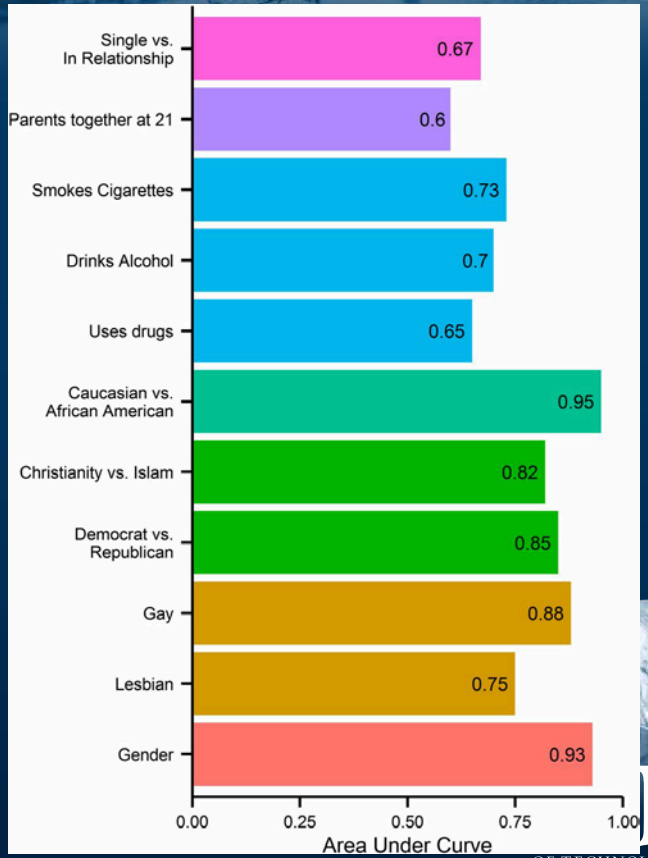
- a user who is interested in university/educational issues
  - → is very likely to be a young adult.
- in a dance club, people come together due to their common interest; in an office, people connect to each other because of similar professions
  - → to infer someone's attribute from the attributes of his/her friends.
- A martini or cigarette in your hand in 98% of your photos
  - → is very likely to get liver failure, lung cancer, and lowered life expectancy!

WHAT HAS RESEARCH SHOWN SO FAR?

Kosinski et al. (2013)

Analyzed 58,466 Facebook users :

- like history (170 likes/person on average)
- profile information
- the results of several psychometric tests



Source: Michal Kosinski et al. PNAS 2013;110:5802-5805

MORE EXAMPLES…

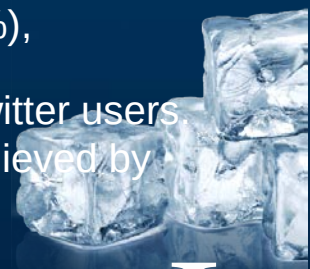
- Schwartz et al. (2013) analyzed text of 15.4 million status updates from a total of 74,941 Facebook users. Predicted gender with 92 % accuracy
- Backstrom and Kleinberg (2014) managed to predict whether a user is single or not with 68 % accuracy and whether he/she is single or married with 79 % accuracy.
- Jernigan et al. (2009) looked at sexual orientation and achieved an accuracy of 78 % by analyzing friendship associations.
- Zheleva and Getoor (2009), examined user attributes are the country, gender and political views.
- Rao et al. (2010) evaluated the accuracy of predicting gender (72 %), age (74 %), regional origin (77 %) and political affiliation (83 %) from Twitter messages.
- Conover et al. (2011) (95 % accuracy) on political views were obtained from Twitter users.
- Very good results on political views from Twitter (89 % accuracy) were also achieved by Penna et al. (2011)
- …

LULEÅ UNIVERSITY OF TECHNOLOGY

All references available in: http://link.springer.com/chapter/10.1007/978-3-319-45982-

MORE EXAMPLES…

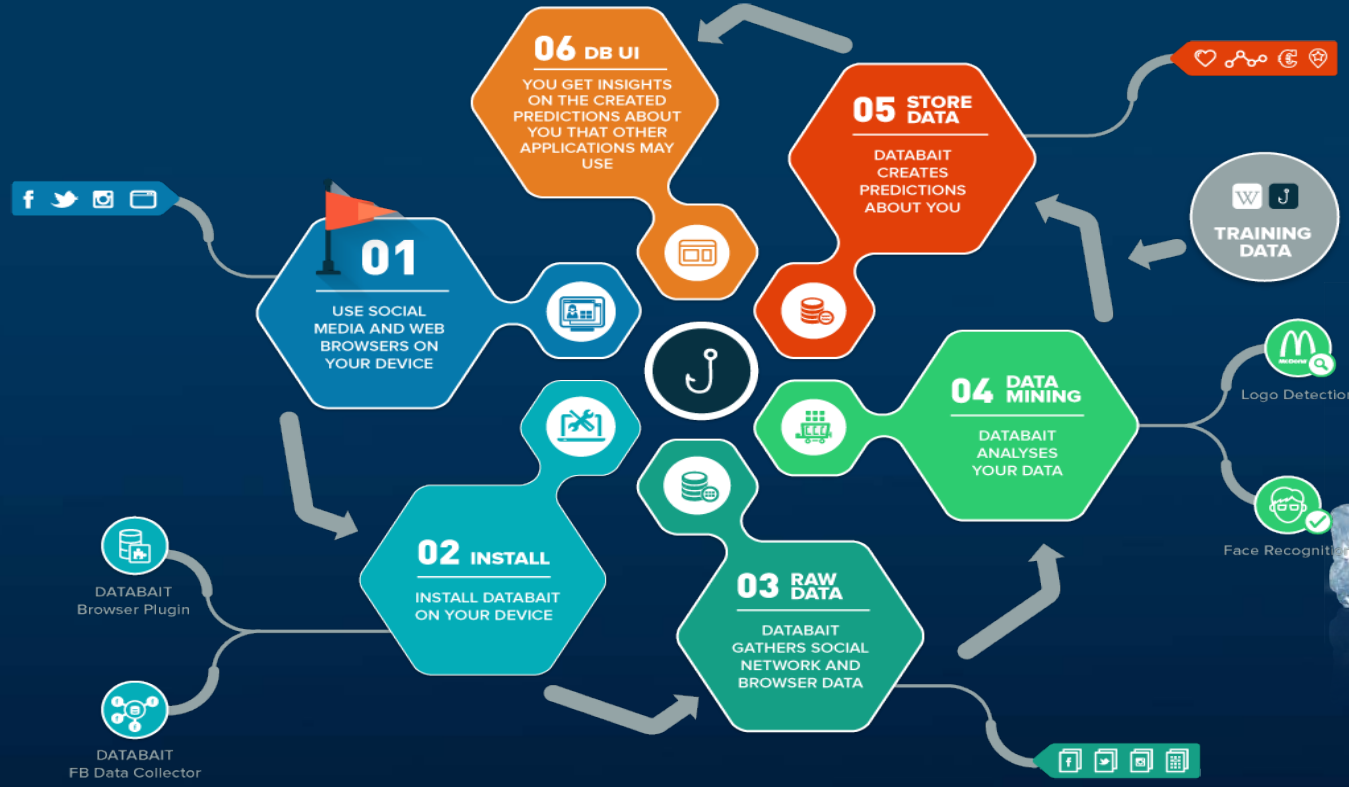
- Schwartz et al. (2013) analyzed text of 15.4 million status updates from a total of 74,941 Facebook users. Predicted **gender** with 92 % accuracy
- Backstrom and Kleinberg (2014) managed to predict whether a user is **single** or **not** with 68 % accuracy and whether he/she is **single** or **married** with 79 % accuracy.
- Jernigan et al. (2009) looked at **sexual orientation** and achieved an accuracy of 78 % by analyzing friendship associations.
- Zheleva and Getoor (2009), examined user attributes are the **country**, **gender** and **political views**.
- Rao et al. (2010) evaluated the accuracy of predicting **gender** (72 %), **age** (74 %), **regional origin** (77 %) and **political affiliation** (83 %) from Twitter messages.
- Conover et al. (2011) (95 % accuracy) on **political views** were obtained from Twitter users.
- Very good results on **political views** from Twitter (89 % accuracy) were also achieved by Penna et al. (2011)
- …



LULEÅ UNIVERSITY OF TECHNOLOGY

All references available in: http://link.springer.com/chapter/10.1007/978-3-319-45982-

THE DATABAIT TOOL



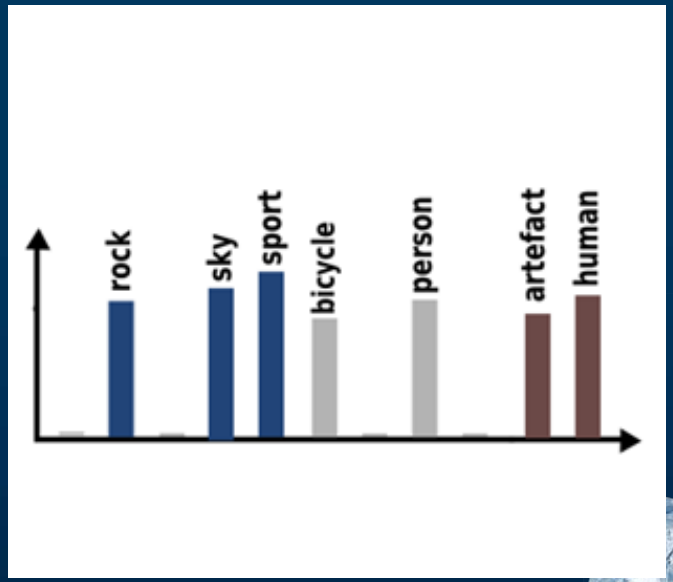
USEMP

DE LA RECHERCHE À L'INDUSTRIE
cea

LULEÅ UNIVERSITY OF TECHNOLOGY

IMAGE MINING TOOL

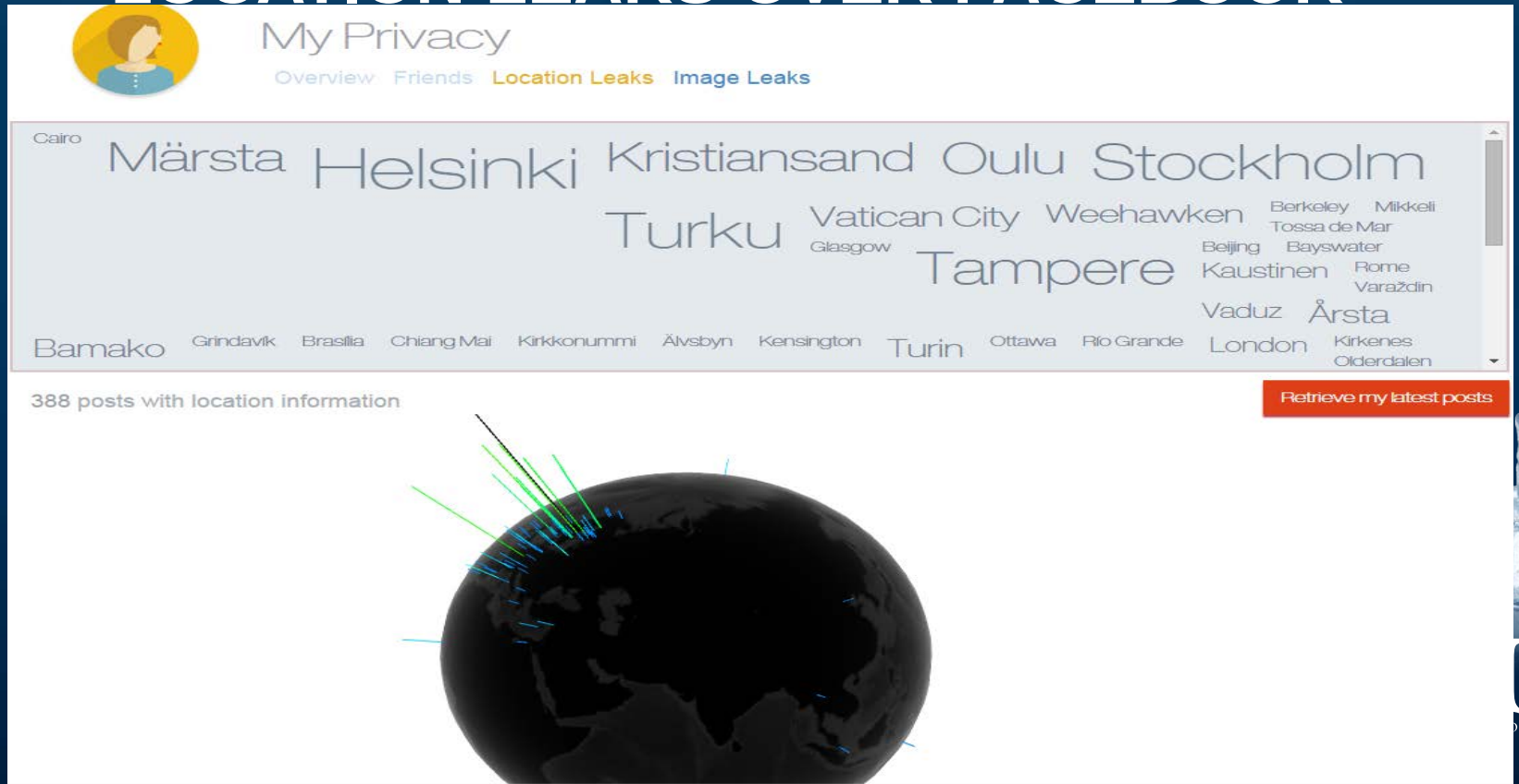
Input



Output

IMAGE LEAKS OVER FACEBOOK

LOCATION LEAKS OVER FACEBOOK

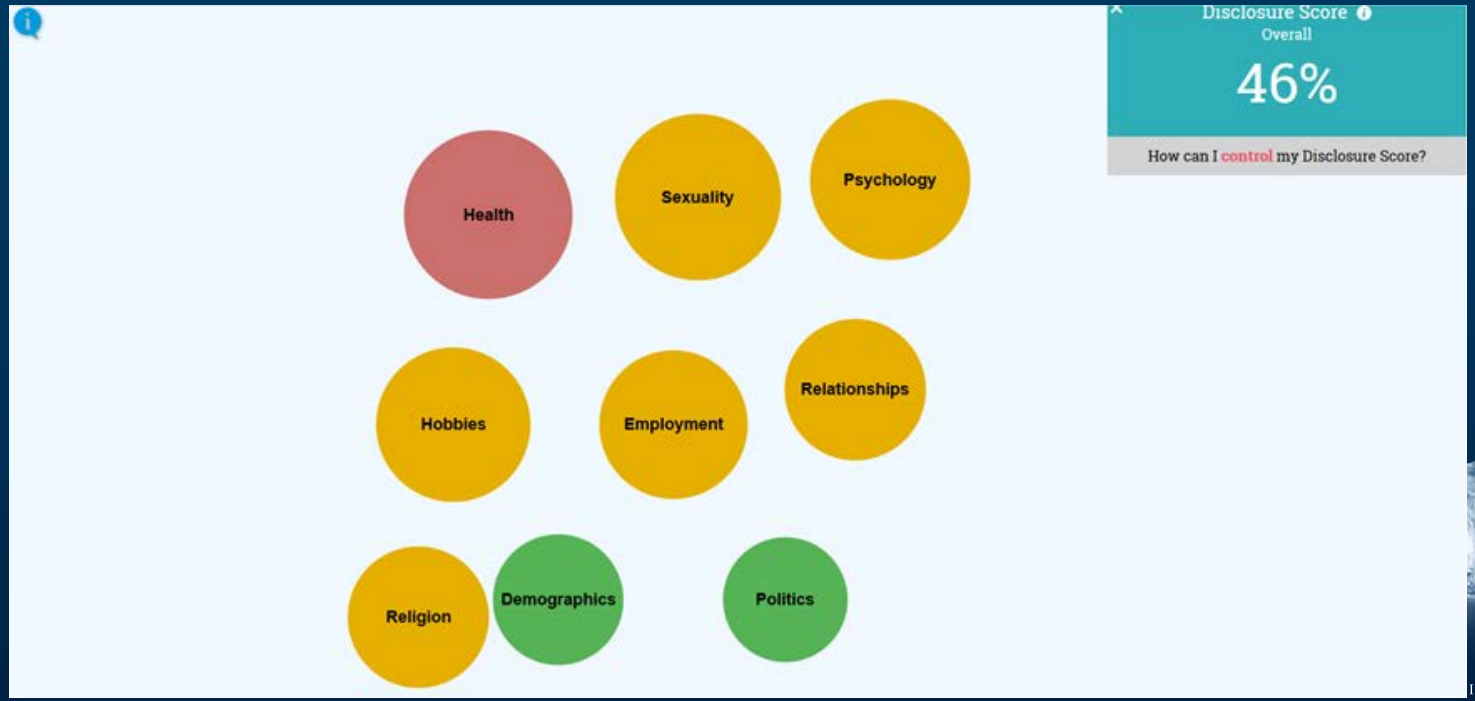


EXAMPLES OF 96 TOTAL CLASSIFIED ATTRIBUTES

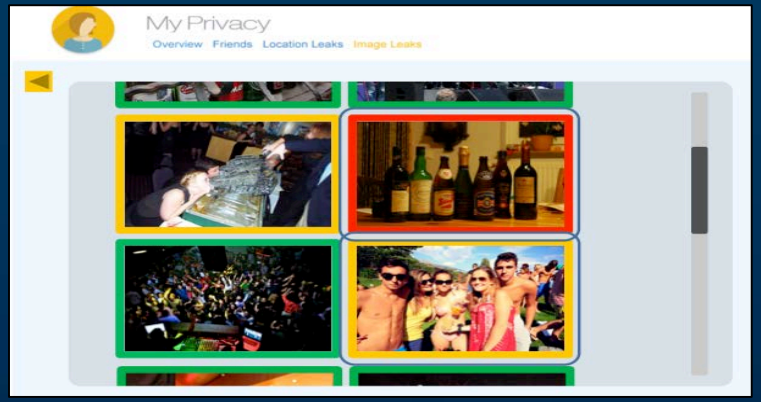
Health	Hobbies	Psychology
cannabis	Swimming	can-be-moody
energydrink	Bicycling	is-ingenious-a-deep-thinker
coffee	Listening-to-music	is-helpful-and-unselfish-with-
nosubstance	Reading	is-talkative
drinkingbehavior	Watching-series-or-movies-at-h	can-be-somewhat-careless
alcohol	Cooking	is-inventive
smokingbehavior	Going-to-the-movies	is-outgoing-sociable
cigarettes	Eating-out	is-curious-about-many-differen
healthstatus	Running	
bmiclass		

LULEÅ UNIVERSITY OF TECHNOLOGY

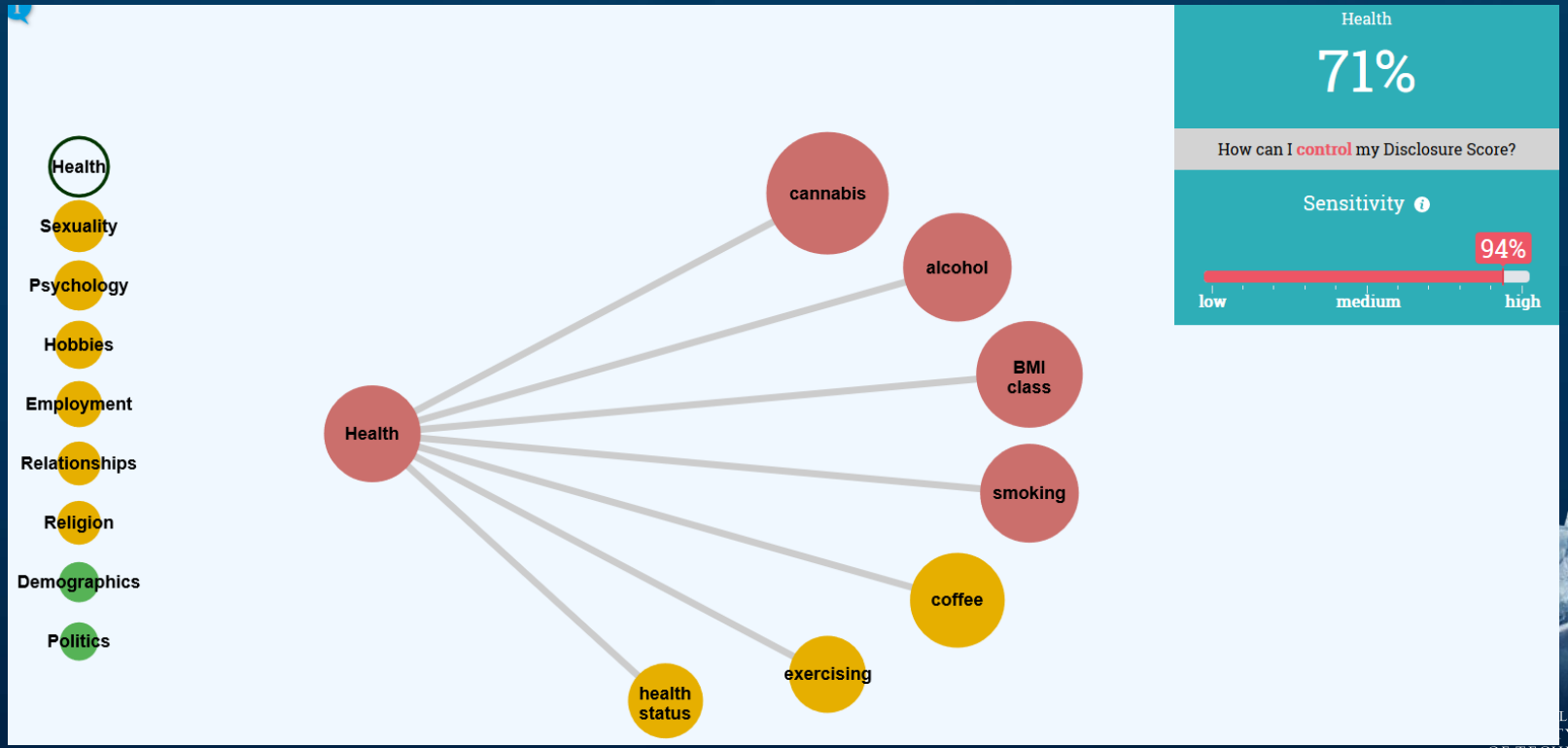
PERSONAL USER ATTRIBUTES WERE ORGANIZED INTO 9 CATEGORIES

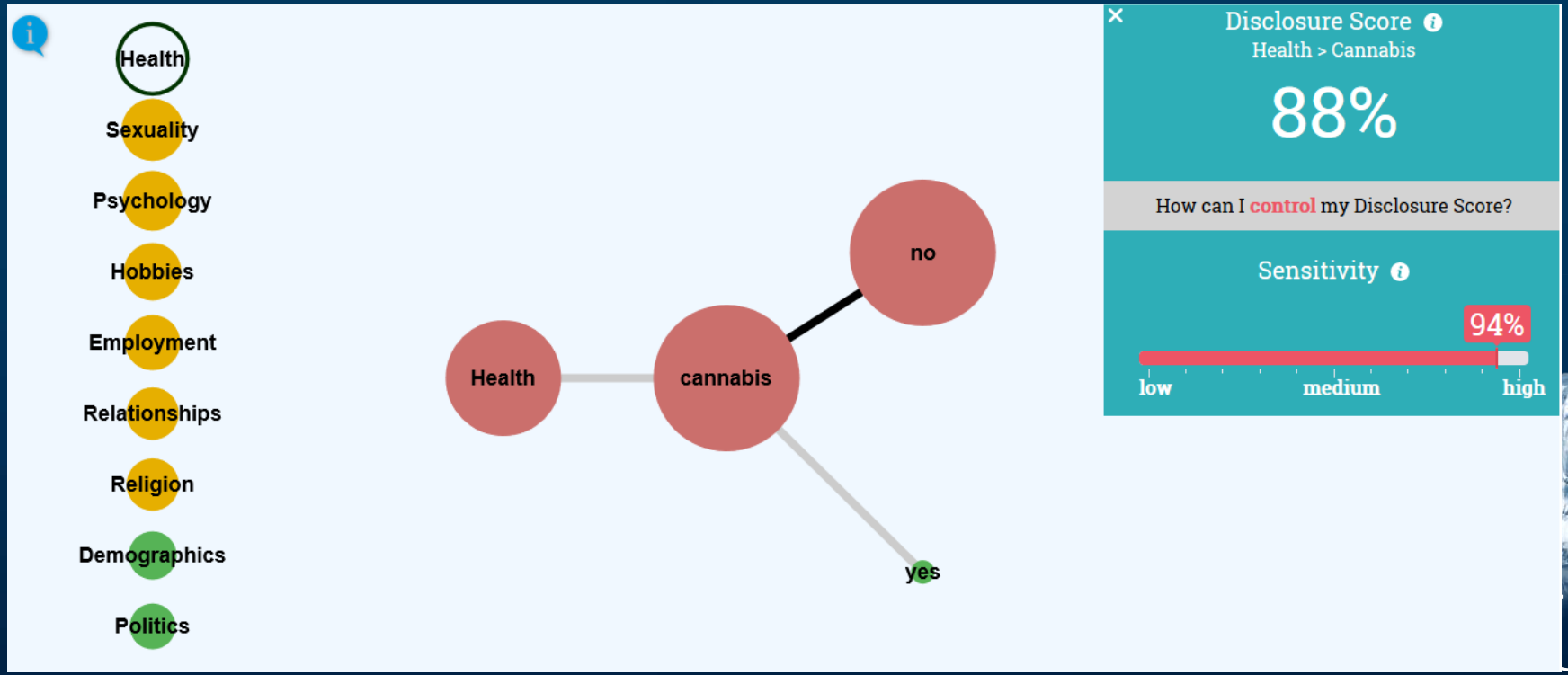


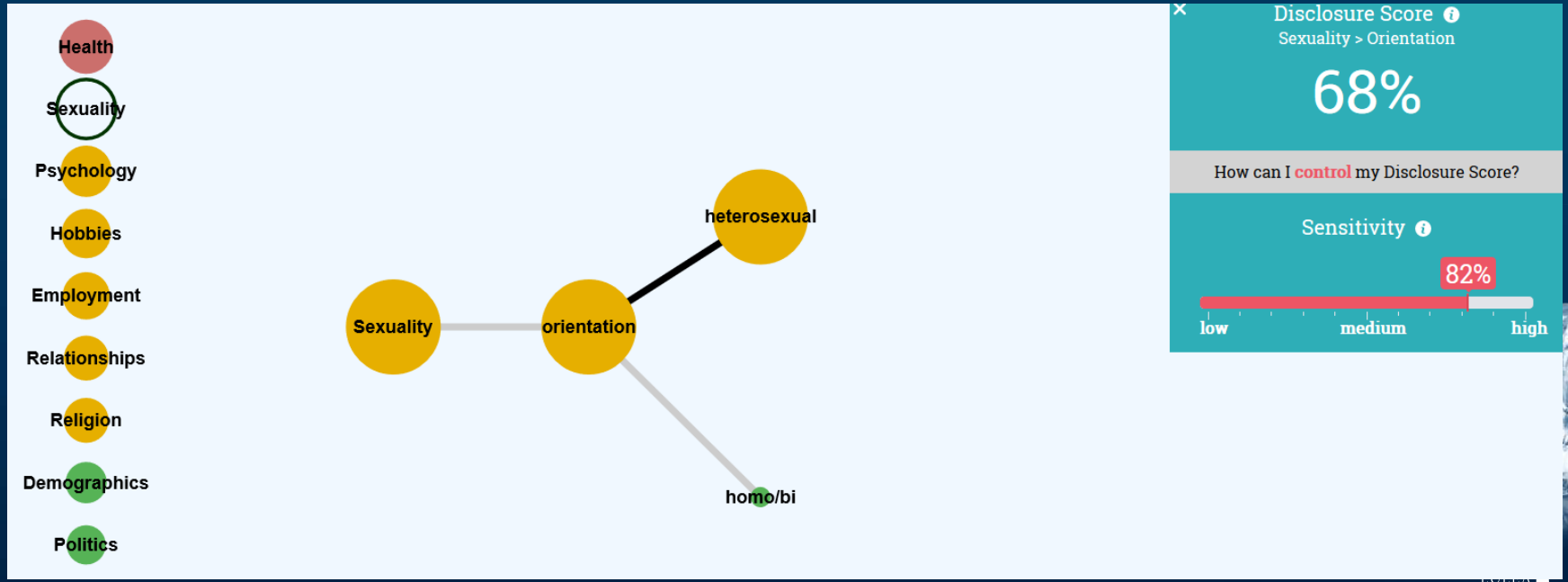
VISUALISATION PHOTO ALBUMS



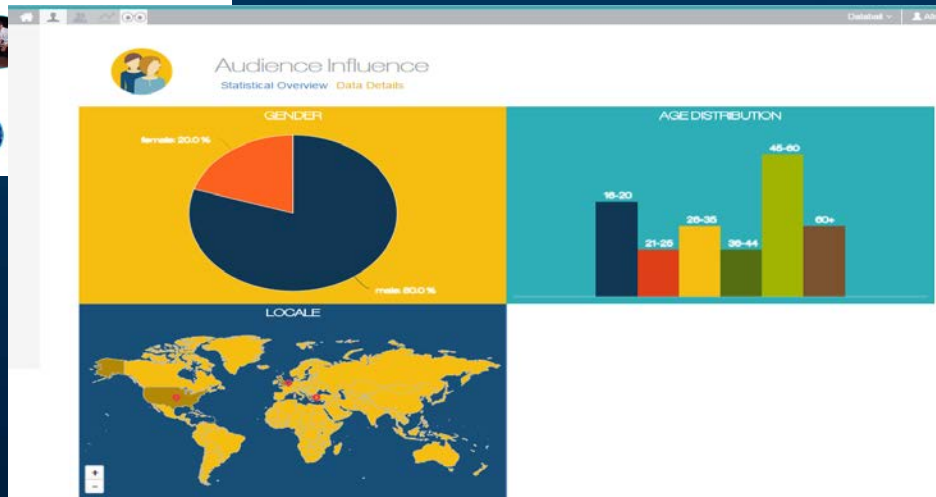
LULEÅ UNIVERSITY OF TECHNOLOGY


Health
Sexuality
Psychology
Hobbies
Employment
Relationships
Religion
Demographics
Politics
Health
cannabis
alcohol
BMI class
smoking
coffee
exercising
health status
Health
71%
How can I control my Disclosure Score?
Sensitivity
94%
low
medium
high


Health
Sexuality
Psychology
Hobbies
Employment
Relationships
Religion
Demographics
Politics
Health
cannabis
no
yes
Disclosure Score
Health > Cannabis
88%
How can I control my Disclosure Score?
Sensitivity
94%
low
medium
high


Health
Sexuality
Psychology
Hobbies
Employment
Relationships
Religion
Demographics
Politics
heterosexual
Sexuality
orientation
homo/bi
Disclosure Score
Sexuality > Orientation
68%
How can I control my Disclosure Score?
Sensitivity
82%
low
medium
high

AUDIENCE INFLUENCE



OUR STUDY IN LTU

- A mixture of participants:
  - occupation (13 students, 14 non-students)
  - educational background (5 high school, 11 BA and 11 MA level)
  - gender (15 male, 12 female)
  - cultural backgrounds (14 Swedish, 13 non-Swedish)
  - age (from 18 to 58 years old)

LULEÅ UNIVERSITY OF TECHNOLOGY

Source: Padyab et al. (2019), https://doi.org/10.1177%2F2056305118824199

SOME QUOTES FROM OUR PARTICIPANTS…

- *“I was recently looking for a job and I want to be very careful about what others tag me in and what I post about myself … if I knew that, then I would go and make changes so that I get things that only I think are ok to see.”*
- *“[Facebook] has all kinds of crazy programs and algorithms that analyze all your habits and likes. Facebook probably does worse than the developers of DataBait on a regular basis. Plus they are a massive multi-national company but DataBait is just a couple of developers in Europe.”*
- *“When you share your photos you share maybe, one, two, or three photos, and you kind of forget the ones you previously posted, and then when you see them all together, it gives you a kind of summary of the pictures that you are introducing of yourself, the profile that you are actually producing.”*

LULEÅ UNIVERSITY OF TECHNOLOGY

IMPLICATIONS FOR DIGITALIZATION

- If in this project we can infer, then others can do it too!
- How much are we aware of data gathered related to end users?
- How much personal information is susceptible to data mining?
- Indirect information disclosure
- User cognitive ability lags behind technological advances
- How can we empower users regarding data mining on the information they are sharing?

THANK YOU!

Ali.Padyab@ltu.se

linkedin.com/in/alipadyab

Acknowledgments

AFSR
Association franco-suédoise pour la Recherche
Svensk-franska forskningsföreningen

LULEÅ UNIVERSITY OF TECHNOLOGY

LULEÅ
UNIVERSITY
OF TECHNOLOGY