

ISTITUTO GIUSEPPE TONIOLO  
DI STUDI SUPERIORI  
ENTE FONDATORE DELL'UNIVERSITÀ CATTOLICA DEL SACRO CUORE



fondazione  
cariplo



UNIVERSITÀ  
CATTOLICA  
del Sacro Cuore



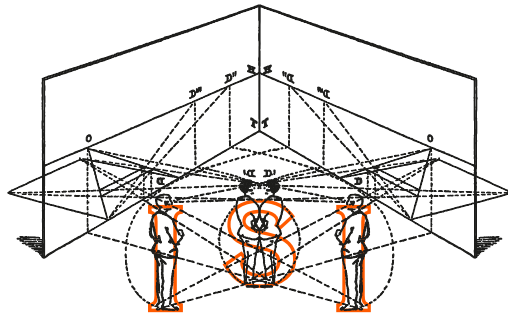
# KYRIAKI KALIMERI

researcher - Computational Social Science - @ISI Foundation

Milano, 1 Dicembre 2017

# Inferring Demographic and Psychological Attributes from Digital Data for Behavioural Nudging

---



ISI Foundation

**Kyriaki Kalimeri**

November 30, 2017

Data Science Laboratory @ ISI Foundation



# Overview

---

Impact of computer science on other disciplines gave rise to interdisciplinary fields like **computational social science** and **digital humanities**.

More complex research questions, more complex data, not unambiguous validation.

The portrait of who we are is unveiled through the **digital traces** of **our everyday activities** in ever growing detail.



ISI Foundation

# Overview

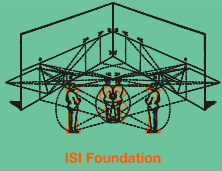
---

## Inferring basic demographics (Age & Gender):

- **web browsing** (Hu et al. [5], Weber et al. [9], Weber et al.[10])
- **smartphone data calls and apps** (Ying et al. [12], Dong et al. [2], Felbo et al. [3], Seneviratne et al. [8], Malmi et al. [7])

## Inferring Personality Traits (Big5):

- **smartphone data calls** (De Montjoye [1]) and **applications** (Xu et al. [11])
- **Facebook Likes** (Kosinski et al. [6], Youyou et al. [13])



# Youth Unemployment Understanding via Social Media

---



- Risk for social marginalisation
- limits income and skill development, and also their likelihood of later employability.

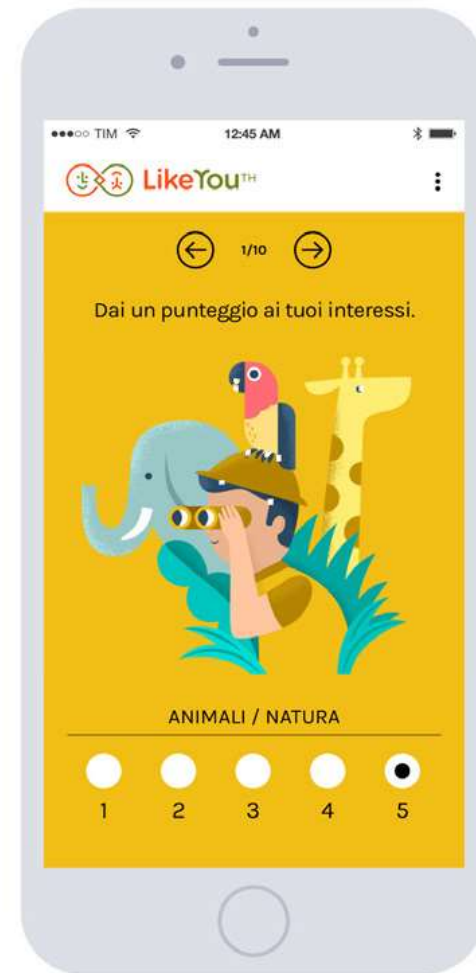


ISI Foundation

# Aims

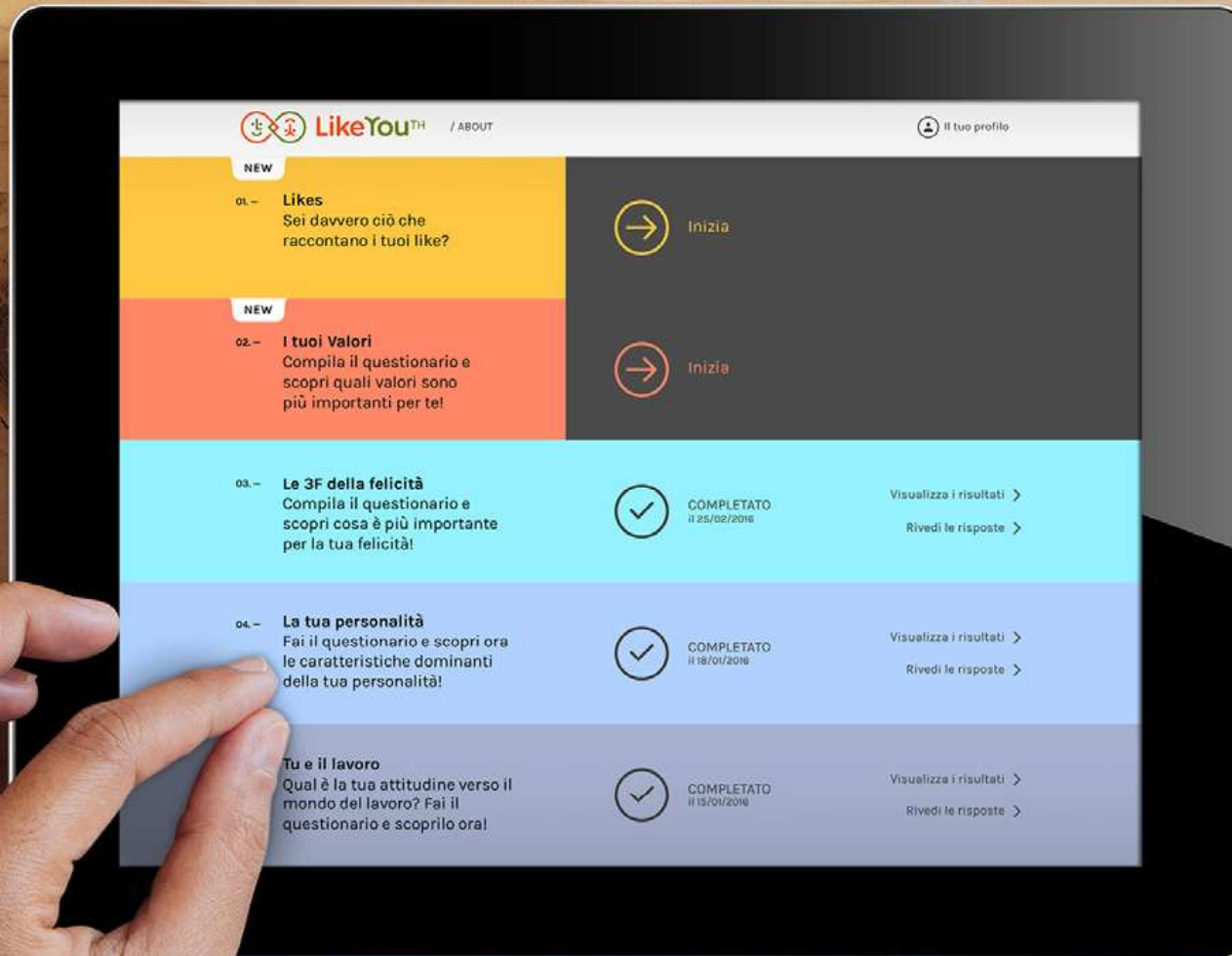
---

- to **automatically identify the NEET population** inferring from their **online digital traces**
  - to **uncover digital behaviours** of the NEET community easily accessible from online social platforms, which can then be used as indicators of the most **privileged communication channels** for unemployment or educational **advertising campaigns**.

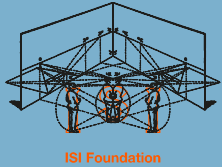


Try it now! [www.likeyouth.org](http://www.likeyouth.org)

Try it now! [www.likeyouth.org](http://www.likeyouth.org)







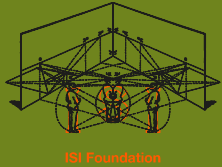
# LikeYouth

---

Advantages with respect to traditional surveys:

- Extendable
- Scalable
- Targeted/Customised population
- Relatively limited cost





# LikeYouth

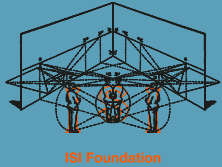
Public profile  
information

Digital information  
Facebook Pages' Likes

Psychometric  
questionnaires



Try it now!  
[www.likeyouth.org](http://www.likeyouth.org)

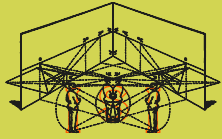


# Random Forest Classification

Automatic prediction model trained on participants' "Likes" on Facebook Pages and Categories

- Employment 0.61 (0.01)
- Neet 0.63 (0.03)
- Education 0.59 (0.03)
- Gender 0.82 (0.01)





ISI Foundation

# Where do we find the NEET?

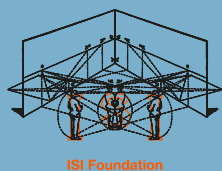
---

## Non-NEET

University  
Musician/Band  
Arts/Entertainment/Nightlife  
Concert Venue  
Concert Tour

## NEET

Food/Beverages  
DimmiCosaCerchi  
Smart Shoppers Italia  
Consulting/Business Services  
Retail and Consumer Merchandise



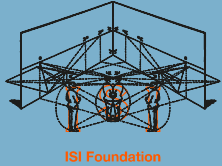
# Psychometric Attribute Prediction for Behavioural Nudging

---

Employ webpage history data and mobile application usage, fused with Census and Open Data (Foursquare, Google Places etc) to automatically infer advanced demographic attributes such political orientation, income and education level.

We also inferred the moral constructs of a person employing the Moral Foundation Theory [4].

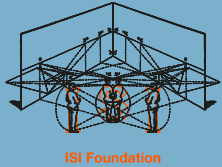
Insights on the population's behaviour and preferences.



# Key Elements

---

- Medium scale but demographically representative population
- Platform Independence
- Fusion of web-browsing and application data as well as open source data (US Census, Foursquare etc)
- Inference of advanced demographic attributes
- Inference of complex psychometric attributes
- Predictive models validated on questionnaires answered by the participants

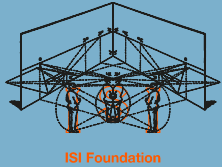


# Prediction of Moral Foundations

	Desktop Web	Mobile Web	Mobile Apps	Mobile Fusion
Authority	<b>0.66 (0.02)</b>	0.63 (0.03)	0.60 (0.02)	0.64 (0.05)
Care	<b>0.62 (0.02)</b>	0.55 (0.05)	0.57 (0.04)	0.55 (0.03)
Fairness	<b>0.58 (0.01)</b>	0.57 (0.04)	0.55 (0.01)	0.55 (0.02)
Loyalty	<b>0.63 (0.02)</b>	0.60 (0.03)	0.55 (0.04)	0.58 (0.05)
Purity	0.64 (0.01)	0.62 (0.02)	<b>0.67 (0.04)</b>	0.66 (0.04)
Ind/Bind	<b>0.66 (0.01)</b>	0.65 (0.02)	0.64 (0.02)	<b>0.66 (0.02)</b>

Prediction of moral foundation attributes from the extracted behavioural features of desktop web browsing, mobile browsing and mobile application usage.

Random Forest Classification, 5-fold cross validation

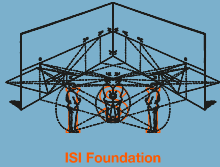


# Prediction of Demographics

	Desktop Web	Mobile Web	Mobile Apps	Mobile Fusion
Age	<b>0.71 (0.01)</b>	0.68 (0.03)	<b>0.71 (0.03)</b>	<b>0.71 (0.02)</b>
Education	<b>0.59 (0.01)</b>	0.57 (0.01)	<b>0.59 (0.02)</b>	<b>0.59 (0.01)</b>
Ethnicity	0.73 (0.02)	0.69 (0.02)	0.72 (0.05)	<b>0.74 (0.02)</b>
Exercise	0.61 (0.02)	0.59 (0.02)	<b>0.63 (0.04)</b>	0.60 (0.03)
Gender	0.86 (0.01)	0.88 (0.02)	<b>0.90 (0.02)</b>	0.89 (0.02)
Income	<b>0.60 (0.01)</b>	0.55 (0.02)	0.60 (0.02)	0.58 (0.01)
Marital Status	<b>0.67 (0.02)</b>	0.61 (0.02)	0.64 (0.01)	0.63 (0.03)
Parent	0.71 (0.01)	0.66 (0.02)	<b>0.72 (0.04)</b>	0.69 (0.04)
Politics	0.58 (0.01)	0.58 (0.02)	0.59 (0.03)	<b>0.60 (0.02)</b>
Smoker	0.63 (0.06)	0.59 (0.06)	0.62 (0.03)	<b>0.64 (0.04)</b>
Wealth	<b>0.66 (0.01)</b>	0.60 (0.02)	0.61 (0.02)	0.62 (0.02)
Weight	<b>0.62 (0.02)</b>	0.58 (0.02)	0.58 (0.03)	0.60 (0.02)

**Gender** and **Age** comparable to the state-of-the-art prediction scores.



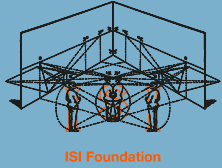


# Insights on Predictors

## Binders vs Individualists

	Desktop	Mobile browsing and APPS
1	google (I)	huffingtonpost (I)
2	foxnews (B)	FOX NEWS (B)
3	dailykos (I)	BIBLE (B)
4	yelp (I)	google (I)
5	imdb (I)	accuweather (B)
6	cnsnews (B)	HANGOUTS (I)
7	wikipedia.org (I)	EMERGENCY ALERTS (B)
8	mrctv.org (B)	facebook (I)
9	theblaze (B)	GOSPEL LIBRARY (B)
10	thepetitionsite (I)	wikipedia.org (I)

Top web sites of the 10 most predictive Desktop and the Mobile (browsing & apps).



# Conclusions

---

Automatic inference of moral views and values offers a great potential for personalising web services, advertising and fine-tuning communication strategies.

Weighted AUROC values are below standard values from other areas, however, the complexity of the learning task is high.

Insights on most efficient communication channels



ISI Foundation

# Thank you!

---

Contact Info:  
[kalimeri@ieee.org](mailto:kalimeri@ieee.org)

@KyriakiKalimeri





**Yves-Alexandre de Montjoye, Jordi Quoidbach, Florent Robic, and Alex Sandy Pentland.**

Predicting personality using novel mobile phone-based metrics.

*In International Conference on Social Computing, Behavioral-Cultural Modeling, and Prediction*, pages 48–55. Springer, 2013.



**Yuxiao Dong, Yang Yang, Jie Tang, Yang Yang, and Nitesh V. Chawla.**

Inferring user demographics and social strategies in mobile social networks.

*In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14*, pages 15–24, New York, NY, USA, 2014. ACM.



**Bjarke Felbo, Pål Roe Sundsoy, Alex Pentland, Sune Lehmann, and Yves Alexandre de Montjoye.**

Using deep learning to predict demographics from mobile phone metadata.

*CoRR*, abs/1511.06660, 2015.



**Jonathan Haidt and Craig Joseph.**

Intuitive ethics: How innately prepared intuitions generate culturally variable virtues.

*Daedalus*, 133(4):55–66, 2004.



**Jian Hu, Hua-Jun Zeng, Hua Li, Cheng Niu, and Zheng Chen.**

Demographic prediction based on user's browsing behavior.

*In Proceedings of the 16th International Conference on World Wide Web, WWW '07*, pages 151–160, New York, NY, USA, 2007. ACM.



**Michal Kosinski, David Stillwell, and Thore Graepel.**

Private traits and attributes are predictable from digital records of human behavior.

*Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.



**Eric Malmi and Ingmar Weber.**

You are what apps you use: Demographic prediction based on user's apps.

*CoRR*, abs/1603.00059, 2016.



**Suranga Seneviratne, Aruna Seneviratne, Prasant Mohapatra, and Anirban Mahanti.**

Your installed apps reveal your gender and more!

*ACM SIGMOBILE Mobile Computing and Communications Review*, 18(3):55–61, 2015.



**Ingmar Weber and Carlos Castillo.**

The demographics of web search.

In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '10*, pages 523–530, New York, NY, USA, 2010. ACM.



**Ingmar Weber and Alejandro Jaimes.**

Who uses web search for what: and how.

In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 15–24. ACM, 2011.



**Runhua Xu, Remo M. Frey, Elgar Fleisch, and Alexander Ilic.**

Understanding the Impact of Personality Traits on Mobile App Adoption – Insights from a Large-Scale Field Study".

*Computers in Human Behavior*, 62:244–256, sep 2016.



**Josh Jia-Ching Ying, Yao-Jen Chang, Chi-Min Huang, and Vincent S Tseng.**

Demographic prediction based on users mobile behaviors.

*Mobile Data Challenge*, 2012.



**Wu Youyou, Michal Kosinski, and David Stillwell.**

Computer-based personality judgments are more accurate than those made by humans.

*Proceedings of the National Academy of Sciences*, 112(4):1036–1040, 2015.