

UNIVERSITÀ DEGLI STUDI DI MILANO

Selezione pubblica per n.2 posti di Ricercatore a tempo determinato ai sensi dell'art.24, comma 3, lettera a) della Legge 240/2010 nell'ambito del Piano Nazionale di Ripresa e Resilienza (PNRR), per il settore concorsuale 01/B1 INFORMATICA, settore scientifico-disciplinare INF/01 INFORMATICA presso il Dipartimento di informatica Giovanni degli Antoni (avviso bando pubblicato sulla G.U. n. 81 del 11/10/2022) Codice concorso 5113.

Torino, 25-10-2022



# CV. - Anna Sapienza

## Center for Social Data Science (SODAS) University of Copenhagen (UCPH)

Letters in the present CV refer to the list of titles in Art. 12 "Valutazione dei titoli e del curriculum", Bando per Ricercatore Tipo A - PNRR - s.c. 01/B1, ssd INF/01 (Codice 5113)



### Email

ansa@sodas.ku.dk



### Website

annasapienza.it

## SOCIAL



### Twitter

@Anna\_Sapienza



### Linkedin

[anna-sapienza-299ab510a](#)

## GOOGLE SCHOLAR

Citations - 443

h-index - 12

## CURRENT POSITION

c-f)

### Qualification

Postdoctoral researcher

### Department and Institution

Center for Social Data Science,  
University of Copenhagen, Copenhagen

### Other collaborators

Dept. of Applied Mathematics and  
Computer Science, Technical University of Denmark

### Start-End date

1/4/2020-Today

## PROFILE

I am a Postdoctoral researcher at the Copenhagen Center for Social Data Science (SODAS) of the University of Copenhagen, working in collaboration with the department of Applied Mathematics and Computer Science at the Technical University of Denmark (DTU). My research interests stay at the intersection of computational social science and data science, with focus on studying human behaviors through digital traces. My current research aims at investigating the use of digital technologies and smartphones to understand their effects on human behaviors.

I previously worked as a Postdoctoral researcher at the USC Information Sciences Institute, where my research aimed at monitoring and modeling human behaviors in online environments, with specific focus on performance and engagement of users in online games.

I hold a PhD in Applied Mathematics from the Polytechnic University of Turin, where I worked in collaboration with the Data Science group at the ISI Foundation of Turin. Here, I focused on the development of tensor factorization techniques and their applications to high-dimensional data of human-proximity.

## PROFESSIONAL EXPERIENCE

c-f)

- Postdoc Dpt. of Applied Mathematics and Computer Science, Technical University of Denmark, Copenhagen, DK. 1/3/2019 - 29/2/2020
- Postdoc Machine Intelligence and Data Science (MINDS), University of Southern California - Information Sciences Institute, Marina del Rey, CA, USA. 1/4/2017 - 15/2/2019
- Research Assistant Machine Intelligence and Data Science, University of Southern California - Information Sciences Institute, Marina del Rey, CA, USA. 1/1/2017 - 31/3/2017
- Intern System Biology Group, IRCC Candiolo Cancer Institute, Candiolo (TO), ITA. 1/10/2012 - 31/3/2013

## ORGANIZATION OF SCIENTIFIC EVENTS

f)

- SODAS PHD READING CLUB, UNIVERSITY OF COPENHAGEN, SPRING 2022
- ERC DISTRIBUTE SEMINARS, UNIVERSITY OF COPENHAGEN, SPRING 2022
- SODAS POSTDOC SALON, UNIVERSITY OF COPENHAGEN, 19/8/2021
- SODAS PANEL DISCUSSION SERIES ON FAIR SCIENCE, UNIVERSITY OF COPENHAGEN, SPRING 2021
- COPENHAGEN DATA BEER, 2ND EDITION 23/9/2019
- COPENHAGEN DATA BEER, 1ST EDITION 23/5/2019
- SUMMER SCHOOL ON COMPUTATIONAL SOCIAL SCIENCE 30/7/2019 - 4/8/2019
- ICDMW ACUMEN: DATA SCIENCE FOR HUMAN PERFORMANCE IN SOCIAL NETWORKS. 18/11/2017

## EDUCATION

a)

- PhD, Applied Mathematics Polytechnic University of Turin, thesis and research funded by the ISI Foundation, Turin. 1/1/2014 - 30/3/2017
- M.Sc., Mathematical Modeling and engineering. Polytechnic University of Turin, thesis in collaboration with the Molecular Biotechnology Center of Turin. 1/9/2011 - 10/10/2013
- B.Sc., Mathematical Modeling and engineering. Polytechnic University of Turin, thesis in collaboration with the Mauriziano Umberto I Hospital of Turin. 1/10/2008 - 13/12/2011

## COLLABORATIONS

c-f)

- Visiting research stay at the Computer Science department, University of Turin, IT to study the mobile app ecosystem via network analysis. 10/10/2022-11/11/2022
- Visiting research stay at the ISI Foundation, Turin, IT to collaborate with the Data Science and Computational Social Science group. 1/7/2019-24/7/2019

## PROJECTS INVOLVEMENT

c-e-f)

- **ERC DISTRIBUTE** (PI Morten Axel Pedersen) The Political Economy of Distraction in Digitized Denmark. 2020-Today
- **DARPA YFA MITIGATE** (PI Emilio Ferrara) Modeling Individual Trajectories and Incentives in Gamified Techno-social Environments. 2017-2019
- **IARPA CAUSE EFFECT** (PI Kristina Lerman) Cyber-attack Automated Unconventional Sensor Environment. 2017-2019
- **DARPA SOCIALSIM COSINE** (PI Emilio Ferrara) Cognitive Online Simulation of Information Network Environments. 2018-2019

## TEACHER TRAINING

b)

### DTU - LearningLab

University Teacher Training Programme: approximately 250 hours of workload, including the redesign of a university course as a capstone project, 2021-2022.

**Final product:** Teaching portfolio and Capstone Project Poster.

### DTU - LearningLab

Workshop - Small group learning strategies, 12/5/2022 & 28/6/2022.

**Final product:** Group contract.

### DTU - LearningLab

Workshop - Strategies for teaching large classes, 30/3/2022.

**Final product:** Plan of a Teaching Session.

## TEACHING ACTIVITIES

b)

### DTU - Department of Applied Mathematics and Computer Science.

Teacher in the Individual course - Measuring behavioral change within mobile app usage: M.Sc level, 10 ECTS, 18/4/2022-30/6/2022.

### DTU - Department of Applied Mathematics and Computer Science.

Teacher in 02805 Social Graphs and Interactions: M.Sc level, 10 ECTS, 13 weeks, Fall 2021, 172 students .

### DTU - Department of Applied Mathematics and Computer Science.

Teacher in 02806 Social Data Analysis and Visualization: M.Sc level, 5 ECTS, 13 weeks, Spring 2022, 215 students.

## MENTORING ACTIVITIES b)

### DTU Compute

**Co-supervision of master students:** Clara Andersen and Long Qin (M.Sc. Thesis: Modeling Motifs in Mobile Phone Application Usage - 23/8/2021-27/03/2022), Josefine Dyhr Thomsen and Signe Nedergaard Mikkelsen (M.Sc. Thesis: Analyzing Global-scale Smartphone Data to Understand App-fatigue 24/1/2022-24/6/2022).

### USC Information Sciences Institute

**Co-supervision of summer interns:** Mert Ozer (2018, Ph.D student at Arizona State University), Zhao Wang (2018, Ph.D student at Illinois Institute of Technology), Hao Peng (2017, Ph.D student at Indiana University), Sindhu K. Ernala (2017, Ph.D student at Georgia Tech University), Susan Fennell (2017, Ph.D at University College Cork).

### USC Information Sciences Institute

**Co-supervision of Ph.D students:** Yiley Zeng, Akira Matsui, Di Huang (2017-2018).

### High school E. Majorana, Moncalieri, To, Italy

Intervention for educational guidance (2015)

## INVITED TALKS h)

### Understanding Large-scale Smartphone Use

University of Eastern Piedmont, DISIT 20/10/2022

### Understanding Human Behavior and Engagement in Smartphone Usage

Polytechnic of Milan, DEIB 18/10/2022

### Escape from Covid Island

HOPE Seminar, Aarhus University, 24/11/2021.

### Understanding Large-scale Smartphone Usage

CUDAN Open Lab seminars, 1/2/2021.

### Understanding Large-scale Smartphone Usage

DISTRACT Seminar, University of Copenhagen, 11/12/2020

### Escape from Covid Island

SODAS Data Discussion, University of Copenhagen, 19/6/2020

### How Millions of Individuals Use the Apps on their Smartphones

Machine Anthropology Workshop, University of Copenhagen, 27-28/1/2020.

### Understanding Human Behavior and Engagement in Mobile App Usage

IT University of Copenhagen, 28/11/2019.

### Mining Human Behaviors and Performance in Online Platforms

ISI Foundation of Turin, 20/2/2019.

## COMMUNITY SERVICE

### Member of the Program Committee for:

**WSDM** - ACM International WSDM Conference, 2022-2023 (Demo Track PC). (GGS rating: A+)

**ICWSM** - The International AAAI Conference on Web and Social Media, 2020, 2022, 2023. (GGS rating: A)

**ICWSM Workshop** - Data for the Wellbeing of Most Vulnerable 2022. (GGS rating: A)

**ASONAM** - The international conference series on Advances in Social Network Analysis and Mining, 2019-2022. (GGS rating: B)

**CoG** - IEEE Conference on Games, 2019-2022. (CORE rating: C)

**FDG** - International Conference on the Foundations of Digital Games, 2022. (CORE rating: C)

### Member of the Technical Committee :

**Elsevier** - Online Social Networks and Media - Scopus: 82th percentile "Computer Networks and Communications"

**Reviewer for:** Applied Network Science, PLOS ONE, EPJ Data Science, Big Data, International Journal of Human-Computer Interactions, ICWSM - The International AAAI Conference on Web and Social Media, 2021.

## RELEVANT TALKS & POSTERS h)

**DISCOVER : mining online chatter for emerging cyber threats** - TheWebConf18, 23-27/04/2018.

**Performance Dynamics and Success in Online Games**

IEEE ICDMW17, 18-21/11/2017

**Early Warnings of Cyber Threats in Online Discussions**

IEEE ICDMW17, 18-21/11/2017

**A tensor decomposition-based method to estimate the spreading process outcomes for temporal networks with incomplete information** - Conference on Complex Systems 17, 18-22/09/2017

**Human behavioural patterns in online games**

Conference on Complex Systems 17, 18-22/09/2017

**Generative model of temporal networks for studying dynamical processes**

Conference on Complex Systems 16, 20-22/09/2016

**Tensor decomposition techniques for temporal graph mining** - European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD) PKDD16, 19/09/2016

**Joint factorization for analysing multi-dimensional data** - Complex Networks, 11-13/07/2016

**Detecting anomalies in time-varying networks using tensor decomposition** - IEEE ICDMW15, 14-17/11/2015

**Anomaly detection in temporal graph data: an iterative tensor decomposition and masking approach** ECML PKDD15, 07-11/09/2015

## ARTICLES (SUBMITTED)

- Litla M., **Sapienza A.**, Lehmann S, and Alessandretti L. The Urban-Rural Divide in Smartphone Usage. Submitted to: Nature

## ARTICLES (PEER-REVIEWED)

- **Sapienza, A.**, & Lehmann, S. A view from data science. Big Data & Society, 8(2), 20539517211040198 (2021) - Scopus: 87th percentile "Computer Science Applications"
- Ozer, M., **Sapienza, A.**, Abeliuk, A., Muric, G., Ferrara, E. Discovering patterns of online popularity from time-series. Expert Systems with Applications, 113337 (2020) - Scopus: 95th percentile "Computer Science Applications"
- Matsui, **A.**, **Sapienza, A.**, & Ferrara, E. Does streaming esports affect players' behavior and performance? Games and Culture, 15(1), 9-31 (2020) - Scopus: 58th percentile "Computer Science Human-Computer Interaction"
- **Sapienza, A.**, Goyal, P., Ferrara, E. Deep neural networks for optimal team composition. Frontiers in Big Data, 2.14 (2019) - Scopus: 58th percentile "Computer Science"
- Zeng, Y., **Sapienza, A.**, Ferrara, E. The Influence of Social Ties on Performance in Team-based Online Games. IEEE Transactions on Games (2019) - Scopus: 68th percentile "Computer Science: Artificial Intelligence"
- Ozella, L., Gauvin, L., Carenzo, L., & ..., D., **Sapienza, A.**, Kalimeri K., Della Corte, F., Cattuto, C. Wearable proximity sensors for monitoring a mass casualty incident exercise: feasibility study. Journal of medical internet research 21.4: e12251 (2019) - Scopus: 88th percentile "Health Informatics"
- **Sapienza, A.**, Barrat, A., Cattuto, C. & Gauvin, L. Estimating the outcome of spreading processes on networks with incomplete information : a dimensionality reduction approach. Physical Review E 98.1 : 01237 (2018) - Scopus: 88th percentile "Statistics and Probability"
- Kobayashi, T., **Sapienza, A.** & Ferrara, E. Extracting the multi-timescale activity patterns of online financial markets. Scientific Reports (2018) - Scopus: 91th percentile "Multidisciplinary"
- **Sapienza, A.**, Zeng, Y., Bessi, A., Lerman, K. & Ferrara, E. Individual performance in team-based online games Royal Society Open Science, 5.6 : 180329. (2018) - Scopus: 89th percentile "Multidisciplinary"
- **Sapienza, A.**, Bessi, A. & Ferrara, E. Nonnegative tensor factorization for human behavioral pattern mining in online games. Information 9.3 : 66 (2018) - Scimago: 62th percentile "Computer Science: Information Systems"
- Campa, C. C., Germena, G., Ciraolo, E., Copperi, F., **Sapienza, A.**, Franco, I., ... & Perino, A. Rac signal adaptation controls neutrophil mobilization from the bone marrow. Sci. Signal., 9(459), ra124-ra124 (2016) - Scopus: 93th percentile "Biochemistry, Genetics and Molecular Biology"

## CONFERENCE PROCEEDINGS (PEER-REVIEWED)

- Wang, Z., **Sapienza, A.**, Culotta, A., Ferrara, E. Personality and behavior in role-based online games. Proceedings of the IEEE Conference on Games (2019) - CORE Rating: C
- Blythe, J., Bollenbacher, J., Huang, D., Hui, P. M., Krohn, R., Pacheco, D., Goran, M., **Sapienza, A.**, ... & Flammini, A. Massive Multi-agent Data-Driven Simulations of the GitHub Ecosystem. International Conference on Practical Applications of Agents and Multi-Agent Systems pp. 3-15, Springer, Cham (2019) - CORE rating: B
- Blythe, J., Ferrara, E., Huang, D., Lerman, K., Muric, G., **Sapienza, A.**, ... & Hui, P. M. The DARPA SocialSim Challenge: Massive Multi-Agent Simulations of the Github Ecosystem. Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (2019) - GGS Rating: A
- **Sapienza, A.**, Goyal, P. & Ferrara, E. Recommending teammates with deep neural networks. Proceedings of the 29th ACM conference on Hypertext and Social Media. pp. 57-61 (2018) - GGS rating: B
- **Sapienza, A.**, Ernala, S. K., Bessi, A., Lerman, K. & Ferrara, E. DISCOVER : mining online chatter for emerging cyber threats. Companion of The Web Conference 2018 pp. 983-990 (2018) - GGS rating: A++
- **Sapienza, A.**, Bessi, A. & Ferrara, E. Early warnings of cyber threats in online discussions. Proceedings of ICDMW: Workshop of IEEE International Conference on Data Mining. pp. 667-674 (2017) - GGS rating: A++
- **Sapienza, A.**, Peng, H. & Ferrara, E. Performance dynamics and success in online games. Proceedings of ICDMW: Workshop of IEEE International Conference on Data Mining. pp. 902-909 (2017) - GGS rating: A++
- **Sapienza, A.**, Panisson, A., Wu, J., Gauvin, L., & Cattuto, C. Detecting anomalies in time-varying networks using tensor decomposition. Proceedings of IEEE ICDMW: International workshop on data mining in networks (2015) - GGS rating: A++
- **Sapienza, A.**, Panisson, A., Wu, J. T. K., Gauvin, L., & Cattuto, C. Anomaly detection in temporal graph data: An iterative tensor decomposition and masking approach. In International Workshop on Advanced Analytics and Learning on Temporal Data, AALTD (2015).

## PRESS RELEASES

- The corona lockdown has changed our gaming patterns. KU Social Sciences (2020)
- New study finds taking breaks boosts team performance. USC Viterbi. (2018)
- Online chatter could give warning of incoming cyber attacks. NewScientist. (2018)

### European Research Council DISTRACT - PI: Morten Axel Pedersen

The project brings together social science and data science methods to explore the political economy of distraction in the post-digital age.

#### Efforts:

- **Data collection** of large-scale GitHub data and **data access** via **PySpark**.
- Design of **observational experiments** to study the **causal effects** of attention and the influence of different types of **network structures** on developers behaviors.
- Use of **network analysis**, including the extraction of the main structures and **components of the network**, to study the **GitHub dependency ecosystem** and its robustness.

### Understanding Large-Scale Smartphone Usage - Supervisor: Sune Lehmann

The project aims at understanding engagement in smartphone use and its effects on online and offline users' behavior.

#### Efforts:

- **Data processing** of high-dimensional longitudinal smartphone data, including mobile app information via **PySpark**.
- **Data analysis** via **statistical** and **deep learning** methods to model **multi-dimensional time-series** of mobile app use. The method includes the use of CNN and attention layers to predict smartphone use.
- Design of **matching experiments** to study the effect of smartphones on users' online and offline behaviors.

### DARPA (Defense Advanced Research Projects Agency) YFA MITIGATE - PI: Emilio Ferrara

The project aimed at modeling behavioral trajectories and incentives in gamified techno-social environments.

#### Efforts:

- **Data collection** of massive datasets of Multiplayer Online Battle Arena games using **APIs and crawlers**.
- Design of a **machine learning framework** (based on **tensor factorization**) to model user performance.
- Development of a **recommender system** of teammates in online games based on a **deep autoencoder** model to study the **co-play performance network** of players.

### IARPA (Intelligence Advanced Research Projects Agency) CAUSE EFFECT - PI: Kristina Lerman

The project aimed at developing and testing new automated methods to forecast **and detect cyber-attacks**.

#### Efforts:

- **Data collection** of cyber-security related Tweets and Blogs using **APIs and crawlers**.
- Data storage and database query through **ElasticSearch and Kibana engines**.
- **Time series analysis** of reported cyber-attacks to predict new threats.
- Development of an **algorithm that predicts new cyber threats** from social media and online discussions.

### DARPA (Defense Advanced Research Projects Agency) SOCIALSIM COSINE - PI: Emilio Ferrara

The project aimed at modeling and simulating the dynamics and evolution of social network platforms, such as GitHub and Twitter.

#### Efforts:

- **Data collection** of GitHub data, and data access through the big data processing engine **Apache Spark via PySpark**.
- Model GitHub system with AI and machine learning **embedding methods**, such as **Graph Factorization** and **Deep Neural Networks (e.g., LE, HOPE)**.
- Development of a **multi-dimensional time-series clustering** algorithm for content popularity in Twitter and GitHub.