Curriculum of scientific and research activity

Contents

1 Personal Data

Name and Surname:	Andrea Failla
Place of Birth:	Canicattì (AG), Italy
Date of Birth:	31/07/1998
Citizenship:	Italian
Email:	andrea.failla.ak@gmail.com
Website:	andreafailla.github.io
ORCID:	orcid.org/0009-0009-6162-0274

2 Positions Held

Current Position: PhD student in the National PhD Program in Artificial Intelligence "AI & Society", Department of Computer Science, University of Pisa, since 01/11/2022.

Research fellow at the Institute of Science and Information Technologies (ISTI), National Research Council (CNR), Italy under the project "Big Data Analytics and Artificial Intelligence 2.0 – BDAI - 2.0"

01/2023-10/2025: Associated as collaborator with the KDD Lab, ISTI-CNR, Pisa, within the research activity *Attributed Dynamic Hypergraph Modeling of Social Interactions* related to the SoBigData++ project[1].

09/2022-10/2022: Research Fellow at the Department of Computer Science, University of Pisa, within the project *Implementation of a web Genome browser for the visualization and analysis of 'omics' data for the study of Rett syndrome*. Prot. no. 1690/2022 of 31/08/2022. Terminated early due to incompatibility with PhD scholarship.

02/2022-04/2022: Intern at the Department of Computer Science, University of Pisa, within the project *Hypergraph Modeling of Social Interactions*. Tutor: Prof. Dino Pedreschi

3 Studies

3.1 Degrees

2020-2022: Master's Degree in Digital Humanities (LM-43).
Awarded by: University of Pisa, final grade 110/110 cum laude
Thesis title: "*Towards Attributed Stream-Hypernetwork Analysis: Structure, Features and Dynamics of Complex Social Systems*"

[1] http://www.sobigdata.eu

Date of defense: 29 September 2022
Advisors: Prof. Giulio Rossetti, Dr. Salvatore Citraro

2017-2020: Bachelor's Degree in Linguistic and Cultural Mediation (L-12)
Awarded by: University for Foreigners of Siena, final grade 110/110
Date of defense: 16/10/2020
Advisor: Prof. Liana Tronci

2011-2017: Scientific High School Diploma.
Awarded by the Scientific High School "A. Volta" in Canicattì (AG).

3.2 Training Courses and Studies

2022 – 2025: PhD courses offered by the National PhD Program in Artificial Intelligence, including: "Advanced Methods for Complex Systems" (20h), "Dynamics on Networks" (10h), "Philosophy of Science" (20h).

22-28/06/2025: Summer School "From Data to Social Innovation", SoBigData, Baratti, Italy

27-31/05/2024: Spring School "Complex Networks: Theory, Methods, and Applications", Lake Como School, Como, Italy;

5-9/06/2023: Summer School "AI & Society", University of Pisa, La Maddalena, Italy;

15-19/08/2022: Summer School "Introduction to Complex Systems", University of Utrecht, Utrecht, The Netherlands;

4 Research Activity

4.1 Topical classification of publications

- Analysis of complex networks and social mining
  - Complex Network Analysis [4, 6, 2, 8]
  - Dynamic Network Analysis [4, 3, 6, 5]
  - Higher-order Network Analysis [12, 11, 4, 10, **?**, 5]
  - Homophily and mixing [4, 19, 8, 10, 9, 7]
- Simulations of complex social processes:
  - Social simulations augmented with Large Language Models [19, 7, 18]
  - Social Digital Twins [20, 18]

4.2 Research interests and scientific activity

Andrea Failla's research activity began with his Master's Thesis, during which an initial contribution to complex network analysis was introduced. The work focused on the formalization and implementation of a model capable of capturing *high-order* (i.e., group) interactions, dynamic (i.e., changing over time) interactions, and attributed interactions (i.e., enriched with metadata) [17]. This framework, later developed and consolidated in the context of the PhD, has been applied to several **social mining** scenarios, with particular attention to the study of online and offline communities and to the description of their evolutionary processes [12, 11, 4, 3, 10].

During the PhD years, his activity has focused on defining analytical and computational methodologies aimed at describing, monitoring and predicting social phenomena observable on digital platforms, with an interdisciplinary approach that combines **network analytics**, **machine learning** and **natural language processing**. In particular, he has developed network analytics methodologies for the analysis of group dynamics, including an algorithm to estimate the level of segregation on social networks modeled as hypergraphs [10], and a framework for the characterization of the life cycle of social groups/communities [3].

A further line of research is devoted to the study of emerging and alternative *social media* platforms, analyzed both in functional terms and in terms of social dynamics. In particular, Andrea Failla has investigated architectural and functional differences with respect to major established social networks [2], with the aim of understanding how such features can affect interaction patterns and information diffusion. In parallel, he has examined the mechanisms of radicalization and polarization that can develop in these digital environments [**?**, 8]. These studies are part of the broader research stream aimed at understanding the impact of new social ecosystems on collective behavior and on the formation of online communities.

A final line of research concerns the integration of *Large Language Models* (LLMs) with network modeling methodologies, in order to simulate complex social processes. In this context, Andrea Failla has investigated opinion dynamics within synthetic societies augmented with LLMs, analyzing how different topological configurations can influence these dynamics [7]. Moreover, he has contributed to the development of a *digital twin* of a social media platform in which LLM agents simulate the behavior of real users in a controlled environment [20, 18].

4.3 Bibliometric Indicators

From 2023 to today, according to *Google Scholar*, Andrea Failla's articles have received a total of 85 citations, contributing to an *H-index* of 4 and an *i10-index* of 3.

4.4 Research stays abroad

03/2025 – 05/2025. Visit to the research group of Prof. Carlos Henrique Gomes Ferreira at the Department of Computer Science of the Federal University of Ouro Preto, Ouro Preto (MG), Brazil. The research activity focused on the collection and analysis of data from the decentralized social network *Bluesky*. In particular, the *starter packs* (i.e., lists of accounts compiled and shared by the users of the platform) were analyzed using network analysis and natural language processing methodologies;

11/2024 – 01/2025. Visit to the LIRIS (DM2L) research group on invitation by Prof. Remy Cazabet at CNRS/INSA Lyon/University of Lyon 1, Lyon, France. The research activity focused on the computational analysis of *alt-right* communities on the social media platforms *Reddit* and *Scored.co*. In particular, the temporal dynamics of their interactions and their evolutionary trajectories were analyzed.

05/2023. Visit to the LIRIS (DM2L) research group on invitation by Prof. Remy Cazabet at CNRS/INSA Lyon/University of Lyon 1, Lyon, France. The research activity focused on defining a methodology to track the evolution of communities/clusters.

4.5 Participation in research projects

SoBigData++: Social Mining and Big Data Ecosystem[2] (2022-2024) is a European project funded by the H2020 program ("INFRAIA-01-2018-2019 – Integrating Activities for Advanced Communities" Grant agreement #871042) in the framework of the construction of a Data Infrastructure aimed at supporting Big Data analysis and the exchange of scientific expertise.

The publications related to participation in this project are: [12, 11, 4, 9, 8, 6, 7]

SoBigData.it – Strengthening the Italian RI for Social Mining and Big Data Analytics (2023-2025) is an initiative funded by the European Union – NextGenerationEU, within the National Recovery and Resilience Plan (PNRR), prot. IR0000013, Call no.3264 of 28/12/2021. Purpose: to strengthen the Italian node of the SoBigData research infrastructure, enhancing interdisciplinary and innovative research on social complexity through combined data- and model-based approaches, and promoting responsible data science (ethics, FAIR/FACT).

The publications related to participation in this project are: [3, 9, 2, 8, 6, 7]

[2] http://www.sobigdata.eu

FAIR: Future AI Research[3] (2023-2025) is a European project funded by the H2020 program ("INFRAIA-1-2014-2015: Research Infrastructures" Grant agreement #654024) in the framework of the construction of a Data Infrastructure aimed at supporting Big Data analysis and the exchange of scientific expertise.

The publications related to participation in this project are: [3, 9, 10, 2, 8, 6, 7]

4.6 Conference Organization

Program Co-Chair of **SimSoc**[4]: *"Simulated Societies: From Agent-based Models to Digital Twins"*. Main conference: CCS 2025, Siena (Italy), 1-5 September 2025.

Program Co-Chair of **TENET (2nd ed.)**[5]: *"Satellite on Temporal Networks"*. Main conference: CCS 2025, Siena (Italy), 1-5 September 2025.

Program Co-Chair of **HyperSci (2nd ed.)**[6]: *"2nd International workshop on Theory and Applications of Hypernetwork Science"*. Main conference: ASONAM, Niagara Falls (Ontario, Canada), 25-29 August 2025.

Program Co-Chair of **SMS (2nd ed.)**[7]: *"The Social Media Sway Unraveling the Impact of Social Media on Human Behavior"*. Main conference: ASONAM, Niagara Falls (Ontario, Canada), 25-29 August 2025.

Session Chair of the track **"LLM in Network Analysis"**[8]. ASONAM, Niagara Falls (Ontario, Canada), 25-29 August 2025.

Chair of the tutorial **"Modeling and Mining High-order Interactions in Social Media Data"**. Main conference: ICWSM, Copenhagen (Denmark), 23-26 June 2025;

Chair of the tutorial **"LLM-powered Simulations of Social Media Environments"**. Main conference: HHAI, Pisa (Italy), 9-13 June 2025;

Program Co-Chair of **LSAI (2nd ed.)**[9]: *"International Workshop on Law, Society and Artificial Intelligence: Interdisciplinary perspectives on AI safety"*. Main conference: HHAI, Pisa (Italy), 9-13 June 2025.

[3] https://fondazione-fair.it/
[4] https://simulatedsocieties.github.io/
[5] https://sites.google.com/view/tenet-ccs25/home
[6] https://hypersci2025.github.io/
[7] https://sms-workshop.github.io/
[8] http://asonam.cpsc.ucalgary.ca/2025/
[9] lsai2025.github.io

Program Co-Chair of **TENET (1st ed.)**[10]: *"Satellite on Temporal Networks"*. Main conference: NetSci 2025, Maastricht (The Netherlands), 2-6 June 2025.

Program Co-Chair of **LSAI (1st ed.)**[11]: *"1st Doctoral Workshop on Law, Society and Artificial Intelligence"*. Independent event funded by the University of Pisa, Pisa (Italy), 2 December 2024.

Program Co-Chair of **HyperSci (1st ed.)**[12]: *"1st International Workshop on Theory and Applications of Hypernetwork Science"*. Main conference: ASONAM 2025, Rende (Italy), 2-5 September 2025.

4.7 Reviewing Activity

Andrea Failla has carried out reviewing activity for several national and international conferences and journals:

International Journals

TIST: ACM Transactions on Intelligent Systems and Technology

TWEB: ACM Transactions on the Web

BD&S: Big Data & Society

IJDSA: Interantional Journal of Data Science and Analytics

JCN: Journal of Complex Networks

JCSS: Journal of Computational Social Science

Network Science

PLoS ONE

Scientific Data

Scientific Reports

International Conferences

ASONAM: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining

HHAI: International Conference Series on Hybrid Human-Artificial Intelligence

[10] https://sites.google.com/view/tenet-netsci/home
[11] lsai-ws.github.io
[12] hypersci-workshop.github.io

NetSci: International School and Conference on Network Science

ACM SIGKDD Conference on Knowledge Discovery and Data Mining

ECML PKDD: European Conference on Principles and Practice of Knowledge Discovery in Databases

National Conferences

ITADATA: Italian Conference on Big Data and Data Science

FRCCS French Regional Conference on Complex Systems

5 Teaching

5.1 Teaching/Co-Teaching Activity

2024/25

International Master (equivalent to a Master's Degree) in Data Intelligence for Smart Systems (DISS) at Université Claude Bernard Lyon 1.
Course: module "Online Social Network Analysis and Mining" within the course "Machine Learning".
Total hours: 12.

5.2 Tutoring and Teaching Support Activity

2024-2025

Second-level Master "Big Data Analytics and Social Mining"[13] at the University of Pisa.
Course: "Social Network Analysis".
Course lecturers: Dr. Chiara Boldrini, Dr. Giulio Rossetti.
Total hours: 12.

2023-2024

Second-level Master "Big Data Analytics and Social Mining" at the University of Pisa.
Course: "Social Network Analysis".
Course lecturers: Dr. Chiara Boldrini, Dr. Giulio Rossetti.
Total hours: 10.

Tutor for "Incontra Informatica" at the University of Pisa.
Title: "Intro to Social Network Analysis"
Total hours: 2

[13] http://masterbigdata.it

5.3 Thesis Supervision and Co-Supervision

***ongoing*, Co-supervisor**. Cinzia Lestini, Thesis title: "Analysis of interactions on the social network Bluesky".
Master's Degree in Data Science and Business Informatics, University of Pisa.
Co-supervisor: Prof. Giulio Rossetti.

2025, Co-supervisor. Giuseppe Onesto, Thesis title: "Political classification of Reddit texts based on the multidimensional Political Compass model".
Master's Degree in Data Science and Business Informatics, University of Pisa.
Grade obtained by the candidate: 110/110.
Co-supervisor: Prof. Laura Pollacci.

2025, Co-supervisor. Biancamaria Bombino, Thesis title: "Politics and Mental Health: A Data-Driven Analysis of Toxicity and Moral Values in Public Discourse".
Master's Degree in Data Science and Business Informatics, University of Pisa. Grade obtained by the candidate: 110/110.
Co-supervisor: Prof. Giulio Rossetti.

2024, Co-supervisor. Veronica Mesina, Thesis title: "Analysis of Twitter data on public response to COVID-19 policies in Italy: which authorities and which influence?".
Master's Degree in Digital Humanities, University of Pisa.
Grade obtained by the candidate: 110L/110.
Co-supervisor: Prof. Giulio Rossetti.

2024, Co-supervisor. Davide Perra, Thesis title: "Tendency toward the extreme: Applicative case studies on Reddit".
Master's Degree in Data Science and Business Informatics, University of Pisa.
Co-supervisor: Prof. Giulio Rossetti.

2023, Co-supervisor. Francesco Di Cursi, Thesis title: "Peering into YouTube's Abyss: Analyzing NSFL Content and Empowering Users with Collaborative Risk Annotations".
Master's Degree in Digital Humanities, University of Pisa.
Grade obtained by the candidate: 110L/110.
Co-supervisor: Prof. Giulio Rossetti.

5.4 Invited Talks

2024, University of Pisa. Title: *"Group Dynamics in Online Social Networks: Communities and High-order Interactions in Time"*, 10/07/2024.

Within the seminar cycle "Colorful Seminars Series" organized by SoBigData.

2023, LIRIS/CNRS/Université Claude Bernard Lyon 1. Title: *"Temporal Modeling of Node-attributed High-order Interactions"*, 26/05/2023. Within "Seminaires LIRIS (DM2L)"

6 Scientific Awards

Andrea Failla's research activities have received support through scholarships and funding for international mobility, and have been recognized with scientific awards:

Young Researcher Award "Matteo Dellepiane" conferred on 18/11/2025 by ISTI-CNR for the scientific production related to the year 2024.

Best Poster Award conferred on 18/11/2025 by ISTI-CNR for the poster [13].

Best Project Award conferred on 28/06/2025 by Sobigdata Summer School for the project "A preparedness framework for predicting language needs of refugees arriving in Europe"

Best Paper Award conferred on 05/09/2024 by the Program Committee of ASONAM 2025 for the publication [10]

Best Oral Presentation Award conferred on 11/12/2024 by the Program Committee of Complex Networks and Their Applications 2025 for the presentation [20]

Best Poster Award conferred on 11/12/2024 by the Program Committee of Complex Networks and Their Applications 2025 for the poster [14]

GYM: Grants for Young Mobility funded by ISTI-CNR for the period 01/12/2024 – 31/01/2025.

Mobility Grant "Aide à la mobilité doctorale" funded by Université Claude Bernard Lyon 1 for the period 01/11/2024 – 01/12/2024.

Best Project Award conferred on 09/06/2023 by AI & Society Summer School for the project "IUS: Intelligent University System"

PhD Scholarship funded by the University of Pisa for the period 01/11/2022 – 31/10/2025.

Research Grant funded by the University of Pisa for the period 01/09/2022 – 28/02/2023.

7 Foreign Languages

Italian: Native speaker.

English: Excellent written and spoken proficiency (Self-certified CEFR level C2).

Portuguese: Basic written and spoken proficiency (Self-certified CEFR level A2).

Spanish: Basic written and spoken proficiency (Self-certified CEFR level A2).

8 Other Skills and Personal Competences

Programming languages: Python, Java, R

Databases: SQL, PySpark

AI/LLM Frameworks and Libraries: Ollama, Autogen, Transformers, LangChain

Web: HTML, Javascript, CSS

Development environments and Editors: VSCode, PyCharm, Eclipse, RStudio

Data Analysis Environments: Cytoscape, Gephi, KNIME

Operating systems: Mac OS X, GNU/Linux, Windows

9 Other Work Experience

2021 - 10/2022: Freelance Data Analyst, self-employed. Provided data mining solutions for clients on textual, tabular and network data.

05/2020 - 10/2022: Activist and collaborator of *Cademia Siciliana*[14]. Proofreading, translations, analysis.

09/2019 - 02/2020: Internship: Linguistic and Cultural Mediator at *ISS Sarrocchi*, Siena. Italian as a second language support, social integration.

[14] https://cademiasiciliana.org/

10 Software and Algorithms

The code developed by Andrea Failla is released as Free Software and can be found through the catalog of the SoBigData Research Infrastructure[15] and on GitHub[16]. The main software packages and algorithms resulting from Andrea Failla's research activity currently available online are:

ASH[17]: Library for the construction and analysis of temporal hypergraphs enriched with attributes. Introduced in [4] and used in [12, 11, 10, 5].

LifeCycles[18]: Library for the longitudinal analysis of clusters/communities, introduced in [3].

RWHS[19]: Algorithm to measure the level of local segregation on hypergraphs enriched with attributes, introduced in [10].

FairNet[20]: Algorithm to reduce the level of local and global marginalization in social networks, introduced in [9].

CDlib[21]: Library collecting community discovery algorithms, offering support for the evaluation/comparison of results. Used in [3].

11 Datasets

The data collected by Andrea Failla are released as Open Data according to the FAIR principles,[22] and can be found through the catalog of the SoBigData Research Infrastructure[23] and on Zenodo[24]. The main datasets collected in support of Andrea Failla's research activity currently available online are:

Bluesky Social Dataset[25]: collection of social interactions and user-generated content on Bluesky (over 4M accounts, 235M posts, and social ties). Introduced in [2]. 13K+ downloads.

Reddit Climate Change Debate Dataset[26]: Interactions in Reddit discussions on climate change, with users enriched with information on their position in the debate, estimated via neural model. Introduced in [8]. 100+ downloads.

[15] https://sobigdata.d4science.org/
[16] https://github.com/andreafailla
[17] https://github.com/GiulioRossetti/ASH
[18] https://github.com/andreafailla/LifeCycles
[19] https://github.com/GiulioRossetti/ASH
[20] https://github.com/andreafailla/fairnet
[21] https://github.com/GiulioRossetti/cdlib
[22] https://www.go-fair.org/fair-principles/
[23] https://sobigdata.d4science.org/
[24] https://zenodo.org/
[25] hhttps://zenodo.org/records/14669616
[26] https://zenodo.org/records/13603528

Scored.co Hypernetwork Dataset[27]: Higher-order interactions emerging from discussions on the American far-right social network Scored.co. Introduced in [**?**]. 100+ downloads.

Italian Covid-19 Retweet Network[28]: Italian retweet network on Covid-19 for the analysis of online debates. Introduced in [6]. 50+ downloads.

12 Publications

12.1 International Journals

[1] **A. Failla**, S. Citraro, G. Rossetti and F. Cauteruccio. Characterizing User Archetypes and Discussions on Social Hypernetworks. *Big Data and Cognitive Computing*, 9:236, 2025. MDPI, Basel, Switzerland. https://doi.org/10.3390/bdcc9090236

[2] **A. Failla** and G. Rossetti. "I'm in the Bluesky Tonight": Insights from a year worth of social data. *PloS ONE*, 19(11):e0310330, 2024. Public Library of Science, San Francisco, CA, USA.

[3] **A. Failla**, R. Cazabet, G. Rossetti, and S. Citraro. Describing group evolution in temporal data using multi-faceted events. *Machine Learning*, 113(10):7591–7615, 2024. Springer US, New York.

[4] **A. Failla**, S. Citraro, and G. Rossetti. Attributed Stream Hypergraphs: temporal modeling of node-attributed high-order interactions. *Applied Network Science*, 8(1):31, 2023. Springer International Publishing, Cham.

12.2 International Conferences

[5] F. Cauteruccio, S. Citraro, **A. Failla**, and G. Rossetti. Generalizing Hypergraph Ego Networks and Their Temporal Stability In *International Conference on Advances in Social Networks Analysis and Mining*, 2025 (in press).

[6] V. Mesina, **A. Failla**, V. Morini, and G. Rossetti. Whose Voice Matters? Authority and Influence in the Italian Twitter Debates on Covid-19. In *International Conference on Complex Networks and Their Applications*, pp. 352–363. Springer Nature Switzerland, Cham, 2024.

[7] E. Cau, **A. Failla**, and G. Rossetti. Bots of a Feather: Mixing Biases in LLMs' Opinion Dynamics. In *International Conference on Complex Networks and Their Applications*, pp. 166–176. Springer Nature Switzerland, Cham, 2024.

[27] https://zenodo.org/records/13142208
[28] https://zenodo.org/records/13909011

[8] D. Perra, **A. Failla**, and G. Rossetti. Quantifying Attraction to Extreme Opinions in Online Debates. In *International Conference on Discovery Science*, pp. 411–424. Springer, 2024.

[9] F. Mazzoni, **A. Failla**, and G. Rossetti. FairNet: A Genetic Framework to Reduce Marginalization in Social Networks. In *International Conference on Advances in Social Networks Analysis and Mining*, pp. 139–154. Springer, 2024.

[10] **A. Failla**, G. Rossetti, and F. Cauteruccio. Beyond Boundaries: Capturing Social Segregation on Hypernetworks. In *International Conference on Advances in Social Networks Analysis and Mining*, pp. 40–55. Springer Nature Switzerland, Cham, 2024.

[11] **A. Failla**, S. Citraro, and G. Rossetti. Attributed stream-hypernetwork analysis: homophilic behaviors in pairwise and group political discussions on reddit. In *International Conference on Complex Networks and Their Applications*, pp. 150–161. Springer International Publishing, Cham, 2022.

12.3 National Conferences

[12] **A. Failla**, S. Citraro, and G. Rossetti. Attributed Stream-Hypernetwork Analysis: a SocioPatterns Case Study. In *SEBD*, pp. 383–391, 2022.

12.4 Posters

[13] G. Rossetti, M. Stella, R. Cazabet, K. Abramski, E. Cau, S. Citraro, **A. Failla**, V. Mesina, V. Morini, and V. Pansanella. YSocial: an AI-powered Social Media Virtual Twin ISTI Day 2025, ISTI-CNR, Pisa, Italy.

[14] E. Cau, **A. Failla**, and G. Rossetti. Bots of a Feather: Mixing Biases in LLMs' Opinion Dynamics Complex Networks and Their Applications 2024, Istanbul, Turkey.

[15] **A. Failla**, F. Mazzoni, and S. Citraro. Attribute-aware Community Events in Feature-rich Dynamic Networks. Complex Networks and Their Applications 2022, Palermo, Italy.

[16] F. Mazzoni, **A. Failla**, G. Rossetti, *Quantifying and Reducing Marginalization in Networks: a Genetic Approach*. IC2S2 2023, Copenhagen, Denmark.

12.5 Theses

[17] Andrea Failla. *Towards Attributed Stream-Hypernetwork Analysis: Structure, Features and Dynamics of Complex Social Systems*. Master's Thesis in Digital Humanities, University of Pisa, Pisa, 2022.

12.6 Book Chapters (under review)

[18] E. Cau, **A. Failla**, V. Pansanella, and G. Rossetti. Social Simulations: From Agent-Based Modeling to Digital Twins In *Encyclopedia of Social Networks Analysis and Mining (3rd ed.)*

12.7 Preprints/Technical Reports

[19] V. Morini, V. Pansanella, K. Abramski, E. Cau, **A. Failla**, S. Citraro, and G. Rossetti. From perils to possibilities: understanding how human (and AI) biases affect online Fora. *arXiv preprint arXiv:2403.14298*, 2024.

[20] G. Rossetti, M. Stella, R. Cazabet, K. Abramski, E. Cau, S. Citraro, **A. Failla**, R. Improta, V. Morini, and V. Pansanella. Y social: an LLM-powered social media digital twin. *arXiv preprint arXiv:2408.00818*, 2024.