

 Social Mining & Big Data Ecosystem

SoBigData

RESEARCH INFRASTRUCTURE



Magazine

Editorial

Since its inception, SoBigData has emphasised the role of training, aiming to 'promote the education of the next generation of data science researchers'. In order to achieve this goal, training activities have been organised along two main threads: one devoted to different types of events aimed at engaging 'researchers in new research methodologies for social Big Data', and the other dedicated to establishing 'a joint training and education resource repository on social Big Data'.

As the SoBigData project reaches its final phase, it is possible to assess the objectives that were set in the Grant Agreement and offer an overview of all the activities that have taken place. Within Work Package 4, a number of tasks were set: Task 1 focused on the organisation of Summer Schools, which have involved all of the project's partners and aimed at 'introducing participants to techniques and methodologies for analyzing big data, in order to provide them with a solid background in the computational and mathematical theories behind algo

[Continues on pag. 3]

Inside this issue

- 03 EDITORIAL**
Mark Cotè and Marco Braghieri, King's College of London, UK
- 05 SOBIGDATA CONFERENCE ON AI @ EU PARLIAMENT**
Laura Pollacci, University of Pisa, Italy
- 10 NEXT SOBIGDATA EVENTS**
The editorial board
- 11 HIGHLIGHTS ON TRAINING EVENTS**
Multiple authors



SoBigData
e-Learning Area

Content

Editorial

.....3

News

Conference on Artificial intelligence:
the European approach for citizens' wellbe-
ing.....5
Next SoBigData events.....10

Highlights on Training events

GATE Summer School: open-source natural
language processing with an emphasis on
social media.....11
Summer School on Computational Misinfor-
mation Analysis12
Data Science Summer School..... 14
Complexity 72h: the 72 hours data science
marathon strikes back.....15
When soccer and science goes female: the
2nd Soccer Data Cup 16

ESME: the pro-res workshop about, ethics,
privacy and explainable ai.....17
The Data Science for Society (DS4S) work-
shop.....18
Computational Social Science Summer
School: 2nd edition of CSS summer school on
migration19

Editorial

SoBigData training for the new generation of data scientists

Mark Cotè, King's College London, UK | mark.cote@kcl.ac.uk

Marco Braghieri, King's College London, UK | k1346642@kcl.ac.uk

[continued]

rithmic tools for empowering their future research'. Task 2 focused on the creation of open source training modules, which have been integrated into the SoBigData Research Infrastructure, with a special focus on 'specific stakeholders that are currently not well-served by teaching materials on social Big Data, such as social scientists and humanities researchers'. Task 3

focused on events named 'datathons', a specific type of events, integrating an 'interactive classroom style with

According to the latest available data on training events, the SoBigData project has organised over 20 different training events, involving at least one thousand individuals. Such an intense activity has allowed the project to offer a wide variety of

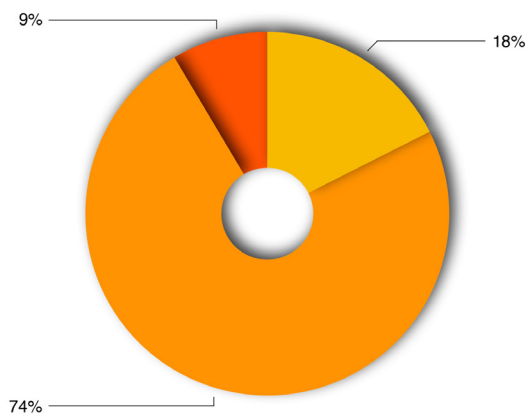
SoBigData explored and developed both conventional and unconventional training experiences for master students, PhD students and early career post-doctoral researchers as well as an academically interested general public.

training events, starting with Summer Schools (such as the recurring Lipari Summer School on Computational

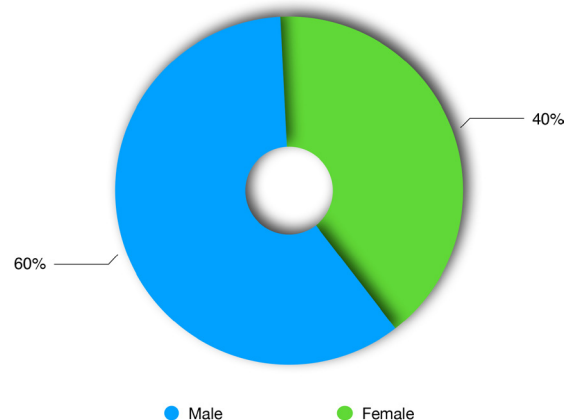
the project's 12 participants. Other type of events have been organised, such as datathons which provided a hands-on approach to participants, fostering real-time collaboration and production of tools and services. Moreover, other types of events such

as the PhD+ program in Italy, have been designed in order to foster innovation and entrepreneurial mind set in Master degree students, PhD students, PhDs and academics. Finally,

events have been especially tailored for high-school students, with themes ranging from data journalism cours-



● High School Students ● MA \ PhD Students \ Early Career Researchers ● Adults



● Male ● Female

datathons (= data + hackathons) thus inviting participants to 'create novel intelligent services on top of our infrastructure's big social data analytics methods'. The final task, Task 4, was devoted to addressing 'equality and diversity issues', especially concerning gender and younger generations, providing 'a series of training events tailored specifically for high-school students' and actions aimed at fostering participation from underrepresented groups.

Complex and Social Systems in Italy and the GATE Summer School in the UK and other summer schools such as the Web Science Summer School in Germany and the Digital Methods in Humanities and Social Sciences in Estonia). These summer schools have not only addressed a number of relevant themes for the SoBigData project but have also allowing the connection between research communities and strengthening bonds between

es to the Headstart cycle of events, which taught high-school students to use the GATE open-source platform and Twitter in order to move 'to drive Lego robots'. In order to foster participation from underrepresented groups, the SoBigData Project has also offered travel awards reserved to female applicants to attend events and conferences which regard the Big Data community. Up to September 2018, the SoBigData project has

founded 12 fellowships to attend two different events. These fellowships have been reserved to female early-career researchers.

Such a wide range of events has, over the years, attracted a composite pool of participants. More specifically, according to the latest available deliverable, there have been over one thousand individuals participating to these events. Master students, PhD Students, Early Career Researchers in the Young Adult category comprised 74% of participants, whereas high school students comprised 18% of participants and Adults comprised the remaining 9 percent. Regarding gender, 40% of all participants identified as female, whereas 60% identified as male. Thus, these figures demonstrate the high rate of success of SoBigData events within Master students, PhD Students, Early Career Researchers and a tendency towards balance between gender groups, although male participants represent the overall majority.

Aside from events, Training in the SoBigData project has also been focused on the construction of a dedicated area within the SoBigData

Research Infrastructure. The process of collection, harmonization and integration of dedicated training materials within the SoBigData RI included lecture slides, hands-on sessions, exercises, Jupyter Notebooks and interactive environments, which are now all available within the SoBigData e-Learning Area. Among these training materials, the most repre-

The overall impact of SoBigData training events, on the new generation of data scientists, involves more than 700 european students in various disciplines and age range. We also funded travel awards to 12 young female researchers in two of the summer school linked to SoBigData.

sented SoBigData Thematic Clusters were, in order, Text and Social Media Mining; Human Mobility Analytics; Social Network Analysis; Web Analytics; Social Data and Visual Analytics, whereas the most represented Exploratories indicated by partners in their metadata surveys were, in order, City of Citizens; Societal Debates; Sports and Data Science; Wellbeing & Economy and Migration Studies. The SoBigData project managed to integrate Training Materials directly into the SoBigData Catalogue, which hosts data and other resources of interest for the research community, specifically datasets and methods

which can the end user can explore and use in order to perform activities defined by the project Thematic Clusters. Moreover, the SoBigData Catalogue allows users to browse its content in different ways, such as a keyword search, a group search and a type search. A specific area within the SoBigData Infrastructure was created and named e-Learning Area, rendering retrieval and discovery of training materials easier and more immediate to the Research Infrastructure users. The e-Learning area now hosts 22 different training materials comprising more than 157 files. Moreover, this

area represents a collection that covers most topics of interest to the SoBigData community using a variety of tools, including hands-on and interactive training materials along with more traditional slides and courses.

Hence, it is possible to say that training within the SoBigData project has been developed both through live events and digital tools that have fostered collaboration and communication between different stakeholders.



Artificial intelligence: the European approach for citizens' wellbeing

An overview on the SoBigData final conference held at the European Parliament and Regione Toscana offices in Bruxelles, the 19th of November 2019

Laura Pollacci, University of Pisa, Italy | laura.pollacci@di.unipi.it

The paradigmatic shift imposed by the impact of Artificial Intelligence (AI) on humankind and decision-making has brought to light the need for profound reflection. If, on the one hand, Artificial Intelligence is already a reality; on the other hand, it becomes ever more urgent to address both regulatory and ethical issues. The challenge imposed by the growing ubiquity of AI involves several actors, who are today called to provide guidelines and address regulatory and ethical issues to assure a robust assurance process.

In this context and from the awareness of the epochal revolution that we are living, one wonders if a European approach is necessary to provide the respect of ethical values and to contribute to citizens' wellbeing.

To answer this question, the National Research Council of Italy, in collaboration with the European Parliament, has organized the "Conference on Artificial intelligence: the European approach for citizens' wellbeing."

The event, hosted in the Altiero Spinelli Building at European Parliament in Brussels, Belgium, involved EU Institutions, Research Institutions, and Industries. Influential representatives of these organizations discussed different visions and roles of Artificial Intelligence, as well as on its potential impact on the economy, social behavior, and political equilibrium.

The conference, chaired by Erika Widegren, Chief Executive of ReImagine Europa (RIE), opened with authoritative welcomes of Patrizia Toia,

Member of the European Parliament and Vice-Chair Committee on Industry, Research, and Energy; Daniela Rondinelli, Member of the European Parliament and Committee on Employment and Social Affairs; and Massimo Inguscio, president of CNR.

Later, some of the most authoritative voices in the field of AI-related research and industry contributed to the conference, such as Roberto Viola, EC Director-General DG-CONNECT, Marco Conti, Director of IIT-CNR, Jeroen van den Hoven, Professor at TU Delft EC European Group on Ethics EGE, Paul Lukowicz, German Research Center for Artificial Intelligence DFKI Humane-AI, Patrick Gatellier, R&D Manager at Thales AI4EU, Dino Pedreschi, Professor at University of Pisa, Kalina Bontcheva,



Professor at University of Sheffield, Aristides Gionis, Professor at Aalto University, Marlon Dumas, Professor at University of Tartu, and Fosca Giannotti, ISTI-CNR.

Finally, Peter Dröll, EC Directorate-General for Research and Innovation, and Paolo Borchia, Member of the European Parliament Committee on Industry, Research and Energy, completed the conference with their closing remarks.

The common denominator of these firsts speeches it was precisely the affirmative answer to the need for a shared European strategy. The European Union has already proved to be able to handle and coordinate ambitious projects thanks to the establishment of efficient governance founded on the collaboration between public authorities and scientific and industrial excellence. Just as no single industry should control technological development, no country might be able to manage this process. Based on this awareness, the European Union is an ideal frame to enclose an ecosystem capable of providing an effective critical mass and guaranteeing the transparency of the process respecting the citizens' ethical values.

The AI phenomenon still has obscure parts, as well as its future effects, which are not yet entirely clear. The revolution the AI is driving involves not only aspects linked to pragmatic functions but also the personal sphere, thus opening social relationships, and communication models to new scenarios. In this context, it becomes ever more urgent to ad-

dress the process following a shared line and interests. "Community" becomes a keyword in the discussion on the development of Artificial Intelligence and on the future of humanity. Governances must ignore national polycentrism and collaborate between themselves to provide guidelines at a community level that allows a generalized EU access. The European Parliament has the right and duty to trace these generalized guidelines for modeling an AI that benefits human beings, without retaliation against them.

The uncertainty about the real effects and consequences of AI cannot

key-concepts are shared professional retraining at the European level, the right to professional training throughout the career path, the establishment of a European minimum wage, and the right to disconnect, closely linked to the concept of ubiquitous work, which imposes substantial repercussions on social and family life. The president of the CNR, Massimo Inguscio, in this regard, cites Aristotle, "the ultimate goal of earthly life was the attainment of personal happiness and self-fulfillment." The European Union must be at the forefront, together with the industries, to favor the mixing of research fields for the benefit of the whole community.



Accessibility, multidisciplinary research, and knowledge sharing are among the main aims of the CNR, which is proposed today as an example not only in Europe but also worldwide. At the national level, CNR plays a crucial role in three initiatives: the Working Group on AI of the Italian Ministry for University and Research; the CNR Virtual Laboratory; the Artificial Intelligence and Intelligent Systems

and must not slow this process down. However, they must be taken into consideration and must be guided in a farsighted and conscious way. The technological innovation may lead to profound changes that start from the school system, up to involving training courses and the whole work-world. The school system must be remodeled, favoring accessibility at the European level, and adding new training courses able to provide increasingly deeper and shared knowledge. Together with school education, the industrial world must also be redesigned based on this technological process. In this context, the

(AIIS) laboratory of the Italian National Inter-university Consortium for Informatics (CINI). In particular, CNR is at the forefront in the recently established ministerial Working Group on AI, which aims to define long-term national strategies for integrating the Italian expertise in AI and developing a national program of Ph.D. courses. Regarding the sharing of knowledge, the CNR has established the national Virtual Laboratory to link researchers working on AI throughout its institutes and encouraging collaboration.

Roberto Viola, EC Director-General DG-CONNECT, also focuses on

research and technology during his keynote on the Commission's vision. The economic uncertainty of research and technological development due to non-EU policies represents a danger. In the current era of the global technological revolution, government policies, as well as European ones, must support and promote scientific research to lead society's growth and development. Artificial intelligence is a technology that must act as an amplifier of human abilities and can increase the current technologies already made available by scientific and industrial research. Knowledge, which is the basis of the current technological revolution, must be made available for the benefit of the entire society. In this new context, shared research platforms and infrastructures, such as Sobigdata, become crucial, since these favor the dissemination of knowledge, allowing even small industries to take advantage of otherwise inaccessible technology. The mission and, at the same time, the purpose of the AI is wellbeing as a general concept in an increasingly human-centric world.

To date, three main drivers can be identified behind the AI rise: stronger computational power, more sophisticated machine learning algorithms, and higher availability of vast amounts of exploitable data. These factors have already allowed essential benefits for individuals and society but could lead to mitigation or even the resolution of grand challenges, such as climate change or global health and wellbeing. Environment improvement, problem-solving, but also a better understanding of the phenomena and technologies that surround us, are some of the benefits made availa-

ble by this new process. For example, the use of drones in security systems has represented a real technological revolution in the areas of safety (flight safety, environmental safety, inspection of industrial production and energy plants), security (security of citizens for specialized companies and law enforcement or prevention), and privacy (impact on the processing and storage of data collected).

The positive impact of AI is partially clouded by the shadows cast by its potential threat, and current issues. As pointed out by Marco Conti, Director of IIT-CNR, the importance of the development of new artificial intelli-

other hand, it is necessary that the AI lead beneficial to humans and that adheres to European ethical values and social, cultural and political norms.

These are the topics discussed by Jeroen van den Hoven, Paul Lukowicz, Patrick Gatellier, and Dino Pedreschi.

Europe, today represents one of the three world technological powers, alongside America and China. However, the European ascent has not come to an end. It is, therefore, essential to increase efforts to establish an increasingly robust and reliable fair digital society. In this context, it be-

comes fundamental to rely on high-level expert groups working on the subject of AI trustability. In this sense, the European idea is based on two fundamental concepts, namely ethics by design, and responsive research. Ethics is an integral part of robust



and effective technology and must, therefore, represent a fundamental construct of artificial intelligence of the future, as pointed out by Jeroen van den Hoven.

Transparency of the World Wide Web, online equality, equal access, bias, security trust, distributed AI, and black boxes are some of the new challenges that artificial intelligence requires us to face. The challenge imposed by artificial intelligence is undoubtedly impressive from various points of view, but precisely for the same reason, it must be tackled effectively through a joint effort.

The discussion on the role of European governments is linked to that on the scientific aspects. If, on the one hand, it is crucial to achieving the technological breakthroughs needed to shape the AI revolution, on the

Regarding ethics, another major challenge is facing Europe: every combination of technology and ethics must be taken into account and must be effectively managed. The starting point seems to be the definition of the concepts that contribute to the construction of current and future artificial intelligence. To this end, a vademecum was created to shed light on conceptual aspects, aspects related to policies, and design.

The most authoritative voices currently agree on the central factors that must characterize artificial intelligence. Today's AI, but even more

important, tomorrow's AI must be the tool that allows people to carry out tasks and reach goals that were previously difficult or impossible to achieve. Artificial intelligence is everywhere and must have as its fulcrum the man himself following a Europeanist and human-centric vision. Research cannot "just understand and follow the learning and reasoning process of ai systems, but also seamlessly interact with it, guide it, and enrich it with uniquely human capabilities, knowledge about the world and the specific user's personal perspective," as highlighted by Paul Lukowicz. We need to work for a large-scale human AI, and this requires a multidisciplinary effort that guarantees robust scientific foundations and technologies.

The many challenges

cast various shadows on optimism about the work that research will have to do in the future, and a farsighted, responsible plight is more necessary than ever. The experts must be aware of the fact that not all aspects can be covered, but they must work in order to build a human-centric AI that is physical, explicable, verifiable, collaborative, and integrative.

Some of the excellences of scientific research have then discussed the open topics that contribute to the realization of the new vision of AI. Kalina Bontcheva, Professor at the University of Sheffield, discussed the importance of Europe's ecosystem for Social Mining and Big Data. The big data arising from the digital breadcrumbs of human activities represent today a powerful social microscope that must be harness for scientific advancement and the social good. The lack of a large-scale open ecosystem where big data and social mining re-

search can be carried out is today an obstacle to evolution. To this end, the SoBigData Research Infrastructure(RI) provides an integrated ecosystem for ethic-sensitive scientific discoveries. In particular, the infrastructure is a distributed data ecosystem and a distributed platform of interoperability methods, which exploits the techno-legal barriers to offer context-sensitive to ethics and, therefore, to the protection of personal information and respect for fundamental human rights.

Aristides Gionis, Professor at Aalto University, addressed the critical issue of bias and polarization reduction in online media.

and to mitigate its negative effects. The work to be done is still long, and there are many challenges still open, but, above all, today, the scientific community cannot avoid facing.

During his speech, Marlon Dumas, Professor at the University of Tartu, extended a topic already discussed during the conference, namely the importance of AI for Business Process Improvement. For contemporary organizations, the speed and efficiency of processes are vital. The main challenge in developing systems for business processes is a large amount of various parallel possibilities and solutions. The current digital revolution is already deeply rooted in business

processes, and AI is at the forefront of business technology. The benefits of new technologies towards companies are already numerous, such as those created by the automation of complex processes within dynamic technological environments. However, AI still offers nu-

merous application possibilities in the field of Business Process Improvement. Among these, such as the reduction of redundant activities, AI can also be used to learn, observe, and analyze data sets collected from various channels identifying patterns that the human eye cannot interpret.

Furthermore, the new systems can be used to improve customer experience and to help create transparent platforms of communication. Finally, the advancement in AI has introduced sophisticated algorithms capable of solving problems related to the description of properties of large data sets, but also of help in making decisions within a business process. AI-embedded Business Process Improvement solutions can help



While on the one hand, social media has led to the collapse of information barriers, favoring social connectivity and democratization, on the other hand, they involve numerous problems, including harassment, fake news, echo chambers, and polarization. The question arises as to whether it is possible to remove or at least reduce the bias and polarization, since how we handle disagreement is ever more essential to the democratic process. The phenomena of bias and polarization are closely connected and introduce significant interdisciplinary challenges. The scientific community has already made numerous efforts in order to gain an in-depth knowledge of the phenomenon that allows to identify the mechanisms behind it

companies overcome performance bottlenecks and therefore improve both work quality and efficiency. We need to look at artificial intelligence no longer as an intimidating technology, but as an essential tool for the improvement and accessibility of current systems.

Fosca Giannotti, ISTI-CNR, finally closed the series of interventions dedicated to scientific excellence and their contribution to realizing the AI vision, talking about Explainable AI. The need to explain comes from the principles on which the AI of the future must be founded, among others, acceptance and trust, and legality, in terms of compliance with ethical standards and fairness and right to be informed. Explainable AI “investigates methods to produce or complement AI models to make accessible and interpretable the internal logic and

the outcome of the algorithms, making such process understandable by humans,” as states Fosca Giannotti. It becomes essential to pursuing the idea of an explainable AI that empowers individuals against undesired effects of automated decision making, and that responds to the need to implement and spread the right to explanation. Explainable AI contributes significantly also in improving industrial standards that take into account the trust of both companies and consumers. Finally, the explanation helps people make better decisions and preserves human autonomy. Artificial Intelligence must, therefore, be (re) thought of as a cognitive exoskeleton that amplifies the abilities and abilities, also of man’s cognitive abilities, thanks to a close dialogue with it.

The event ended with the final greetings of Peter Dröll, EC Directo-

rate-General for Research and Innovation, and Paolo Borchia Member of the European Parliament in the Committee on Industry, Research and Energy, who had the daunting task of summarizing the vision of artificial intelligence long debated during the conference. The technological revolution has opened up considerable possibilities for the current improvement of the systems and, more generally, for the common wellbeing. Along with the significant benefits that AI can bring to society, some critical pitfalls cannot be underestimated. The road delineated by the experts is unanimous and conscious, both regarding dangers and extent of impacts. Work, dialogue, observation, and collaboration will be the keywords of this long but fundamental journey towards a technology centering on humans, their wellbeing, and their rights.



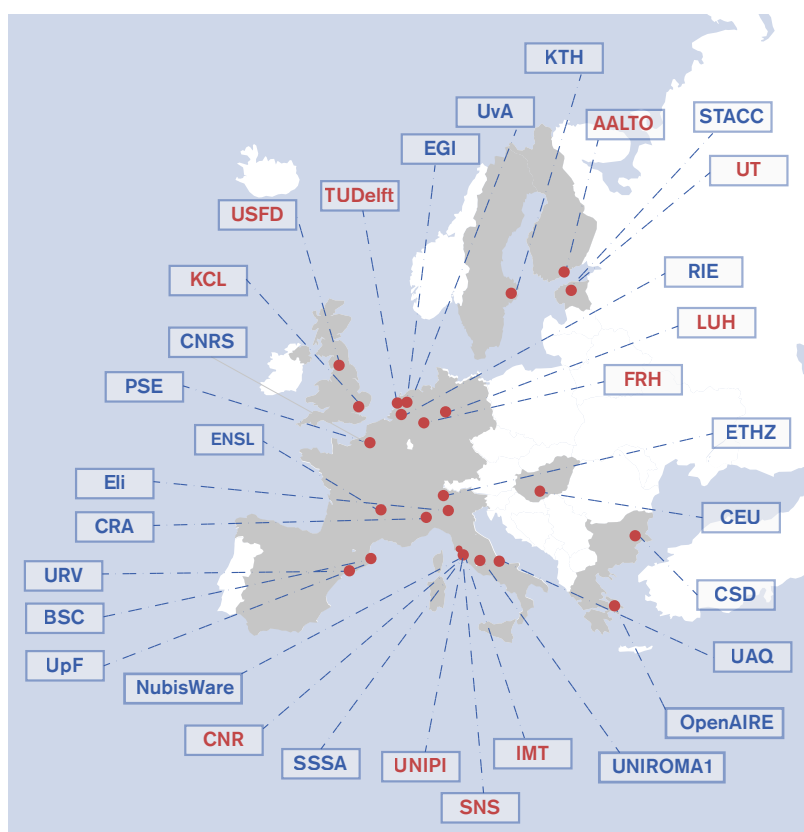
Next SoBigData events

SoBigData++

SoBigData++, a new H2020 INFRAIA project, is about to start!
New consortium, new ideas, renewed enthusiasm!

SoBigData++ consortium consists of 31 partners from 13 member countries of the European Union (Bulgaria, Belgium, France, Greece, Spain, Hungary, Italy, UK, Germany, Estonia, Finland, The Netherlands, Sweden) and Switzerland. Red ones are the old partners of SoBigData and blue ones are the new ones.

More info soon published on www.sobigdata.eu. Stay tuned!

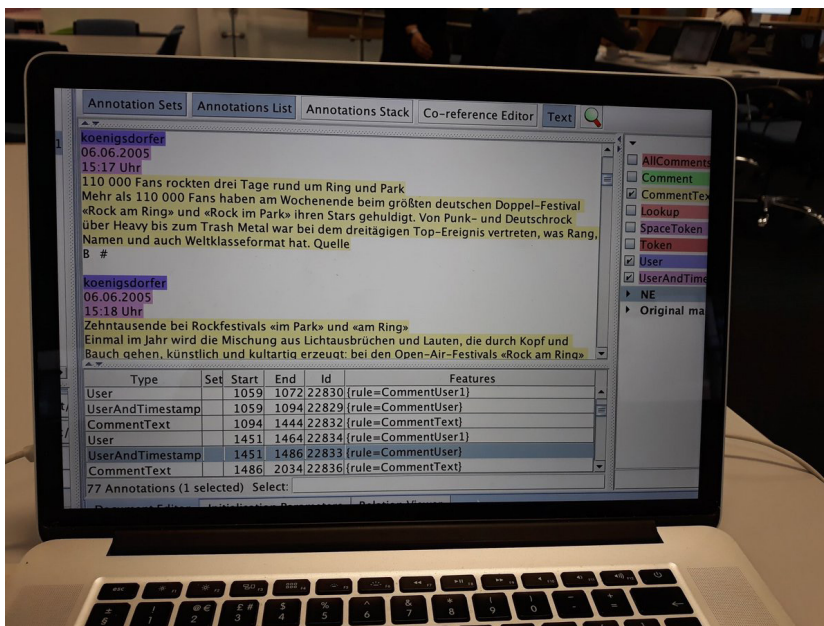


GATE Summer School: open-source natural language processing with an emphasis on social media

Adam Funk, University of Sheffield, UK | a.funk@sheffield.ac.uk

For over a decade, the GATE team has provided an annual course in using our technology. The course content and track options have changed a bit

We also disseminated work from several current research projects. From KNOWMAK (<https://www.knowmak.eu/>)



Processing German in GATE

over the years, but it always includes material to help novices get started with GATE as well as introductory and more advanced use of the JAPE language for matching patterns of document annotations.

The latest course also included machine learning, crowdsourcing, sentiment analysis, and an optional programming module (aimed mainly at Java programmers to help them embed GATE libraries, applications, and resources in web services and other “behind the scenes” processing). We have also added examples and new tools in GATE to cover the increasing demand for getting data out of and back into spreadsheets, and updated our work on social media analysis, another growing field.

and [knowmak.eu/](https://www.knowmak.eu/)) and RISIS (<https://www.risis2.eu/>), we presented our work on using semantic technologies in scientometrics, by applying NLP and ontologies to document categorization in order to contribute to a searchable knowledge base that allows users to find aggregate and specific data about scientific publications, patents, and research projects by geography, category, etc.

Much of our recent work on social media analysis, including opinion mining and abuse detection and measurement, has been done as

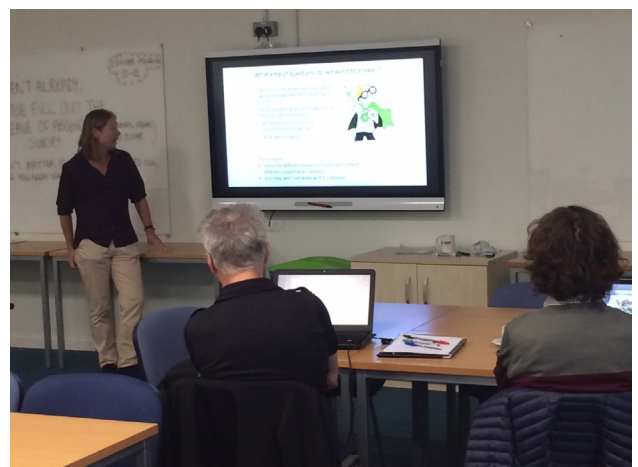
part of the SoBigData project.

The increasing range of tools for languages other than English links with our participation in the European Language Grid (<https://www.european-language-grid.eu/>), which is also supported further development of GATE Cloud, our platform for text analytics as a service.

The GATE software distributions, documentation, and training materials from our courses can all be downloaded from our website under open licences. Source code is also available from our github page.

Acknowledgements

The course included research funded by the European Union’s Horizon 2020 research and innovation programme under grant agreements No. 726992 (KNOWMAK), No. 654024 (SoBigData), No. 824091 (RISIS), and No. 825627 (European Language Grid); by the Free Press Unlimited pilot project “Developing a database for the improved collection and systematisation of information on incidents of violations against journalists”; by EPSRC grant EP/1004327/1; by the British Academy under the call “The Humanities and Social Sciences: Tackling the UK’s International Challenges”; and by Nesta.



Semantics in scientometrics

Summer School on Computational Misinformation Analysis

Beatrice Rapisarda, Wafi Group, IIT-CNR, Italy | beatrice.rapisarda@iit.cnr.it

Marco Braghieri, King's College London, UK | k1346642@kcl.ac.uk

The recent past has highlighted the influential role of social networks and other digital media in shaping public debate on current affairs and political issues.

Disinformation and the hyper-partisan media distort societal debates, increase polarisation, and threaten participatory democracy. For instance, the surprising success of Brexit and Trump's election has been, at least partially, attributed to the unprecedented weave of false information

that in both cases have polluted on-line debate before the vote.

Not only does misinformation get significant attention and shares, but also alternative narratives often try to gain credibility through reusing content from mainstream media, often framed so as to undermine reader confidence in the latter.

The aim of this summer school was, firstly, to set out the state-of-the-art and challenges in computational

misinformation analysis, followed by lectures and hands-on practical sessions on relevant methods, tools, and datasets.

9 male and 5 female attended, for a total of 15 people, mostly PhD students and early career researchers.

<https://gate-socmedia.group.shef.ac.uk/summer-school-comp-misinfo-analysis-2019/>

Interview with school participants

WHAT ARE THE MOST IMPORTANT RESULTS YOU FEEL HAVE BEEN OBTAINED BY THIS SUMMER SCHOOL?

"More knowledge and ideas for future testing and implementation"

"New information about research initiatives, about the most recent research in the field"

"The recent works in computational misinformation analysis"

"Network topology on information sharing, ML"

"Better learnt fake news, rumour evaluation and about the sobigdata project"

"Familiarised with Gephi and GATE, and to understand the research landscape for this area"

"Resources in presentations and networking"

"New ideas and a lot of good references to look more into – knowledge and explanation of work I did not know about before and which is relevant for my current studies"

"Practical hand on sessions for Gate and deep learning"

"Know more about about different sciences point of view in misinformation question. Have new ideas to my project"

"An overview of what is currently done in the field"

"Good understanding of the different efforts by different researchers, a view on a wide spectrum of techniques. Good networking"

WHICH SKILLS HAVE YOU DEVELOPED?

"Better understanding of certain issues like ML, social media analysis"

"How to use GATE system, how to detect controversy, bias, rumours, how to detect fake images in news"

"Gephi/GATE, the vocabulary for researching these topics"

"Better understand the ethics issues in computational social science"

"Gate ML"

"State of art researches on disinformation"

"Expanded knowledge in the filed that it is not my primary one"

"I think the course is to short to really develop skills. Also, GATE is not so relevant for my work, and it was difficult to follow the Gephi "tutorial" since we did not receive the data beforehand"

"Text Analysis (NLP)"

"Learn to use Gate, have new tools to help the development of my PHD project"

"Social network analysis, community detection, ego networks"

"Text processing with Gate"

WHAT IS THE MOST APRECIATED ASPECT OF THE SCHOOL?

"Variety of topics"

"Great quality of presentations, interesting discussions during them"

"Bias and diffusion"

"The variation and range of the talks, the variety of backgrounds of the attendants"

"Interdisciplinary"

"The subjects on rumour verification and social bot detection"

"Practical hand on sessions for Gate and deep learning"

"The opportunity to meet people around this subject"

"NLP, opinion diffusion models"

"Bot detection"

"I very much enjoyed the sessions, nothing in particular. A lot of new information since I come from communication and social science"

Data Science Summer School

Letizia Milli, University of Pisa, Italy | milli@di.unipi.it

Francesca Pratesi, ISTI-CNR, Italy | francesca.pratesi@isti.cnr.it

From 2 to 6 September the first edition of the Data Science Summer School took place in Pisa. Given the interdisciplinary nature of Data Science, this summer school offered lectures by high-level scholars from different domains, giving to the students the skills to exploit data and models for advancing knowledge in different disciplines, or across diverse disciplines (e.g. biology, economics, medicine, etc).

The main topics of the summer school are related to big data analytics, i.e., extraction of knowledge from big data, machine learning and artificial intelligence, i.e., providing an overview of the main techniques used to automatically learn and improve from experience, and complex systems, i.e., methods and technologies particularly related to network science. Moreover, lectures will highlight the ethical implications that data science could lead and the countermeasures that each data scientist can apply to perform analysis with respect to the individuals involved in the data.

The Data Science Summer School offered a broad multi-disciplinary perspective on the different pillars of data science featuring 10 half-day tutorial lectures held by top international scholars:

- János Kertész (CEU Budapest), Guido Caldarelli (IMT Lucca) lecturing on Network science
- Aristides Gionis (Aalto Univ. Helsinki) and Kalina Bontcheva (Univ. Sheffield)

lecturing on Social media analytics, information disorder and misinformation

- John Shawe-Taylor (UCL London) lecturing on Machine learning
- Jeroen van den Hoven (TU Delft) and Nello Cristianini (Univ. Bristol) lecturing on Ethical and social impacts of Data Science and Artificial

Intelligence

- Fosca Giannotti (CNR Pisa) and Dino Pedreschi (Univ. Pisa) lecturing on Explainable Artificial Intelligence

The summer school received a large number of applications; a total of 77 people participated in it, more than 25% were female, with a healthy diversity in the audience:

51 Ph.D. students, 13 postdocs/researchers, 5 MSc students, 8 from the industry. Their background is also diverse: 29 Data Science PhD students of heterogeneous provenance, 23 Computer Science & Engineering, 10 Economics and Management, 3 Statistics, 4 Physics & Math, 2 Biology, 4 Social Sciences.

The Data Science Summer School 2019 is jointly organized by the Data Science Ph.D. in Pisa and the Data Science Ph.D. in Rome provided by the following institutions: University of Pisa, Scuola Normale Superiore, Sant'Anna School of Advanced Studies of Pisa, IMT School of Advanced Studies of Lucca, Sapienza University, and National Research Council of Italy.

SPONSOR: SoBigData, EMbeDS - Economics and Management in the era of Data Science Department, L3S - Forschungszentrum Research Center - Leibniz Universität Hannover.

More information available at the webpage: <https://datasciencephd.eu/DSSS19/>



Complexity 72h: the 72 hours data science marathon strikes back

Angelo Facchini, IMT Lucca, Italy

Promote collaboration and build bridges between different disciplines, fostering teamwork and knowledge sharing.

This is the aim of Complexity 72h,



the second edition of the three days research marathon organized by IMT School for advanced Studies Lucca that took place in the School's Headquarters.

Following the example of "72 hours of Science" organised by the US Santa Fe Institute, Complexity72h opened on Monday June 17 in the San Francesco Complex, where over 40 researchers from European research institutions joined lectures, data tutorials and took on the challenge to pick up a problem, contribute to its solution, and publish a research paper within 72 hours.

The workshop has been introduced by the welcome speech of the Director, Professor Pietro Pietrini, who recalled how the IMT School has always favored a lively and harmonious dialogue between disciplines that are only apparently separate and distant. For over three intense days (72 hours, precisely), PhD students and young researchers worked together day and night on selected topics, under the guidance of teachers and vis-

iting professors of the IMT school. The presentation of the research projects has been introduced by a set of lectures and tutorials given by renowned scholars and researchers:

Marián Boguñá (University of Barcelona), Fabrizio Lillo (University of Bologna), Susanna Manrubia (National Biotechnology Center, Madrid), and Claudio Tessone (University of Zurich). The lectures covered cutting edge research topics ranging from cryptocurrency, evolutionary genomics, networks science.

Together with the lectures, a set of projects and case studies has been present-

ed by scholars and researchers working as tutors for the working groups of young scientists, that listened the talks given by Ennio Bilancini, professor of economics at IMT, Andrea Gabrielli, researcher at the CNR Institute of Complex Systems, Giovanni Petri and Michele Starnini, both researchers at the ISI Foundation in Turin. Case studies and projects covered a wide range of applications, ranging from social media analysis and misinformation, contagion dynamics, behavioural economics, interaction networks, and analysis of bio-physical data.

In the following three days the research groups focused on the project main research aspects, analysed data and under the tutorship of a senior researcher produced the results and drafted a

set of papers successfully uploaded to the ArXiv platform. The following papers has been submitted:

1. Billings et Al., Simplex2Vec embeddings for community detection in simplicial complexes
2. Balsamo et Al., Inside the Echo Chamber: Disentangling network dynamics from polarization
3. Bruno et Al., Community detection in the hyperbolic space
4. Bernini et Al., Evaluating the impact of PrEP on HIV and gonorrhea on a networked population of female sex workers
5. Aydin et Al., Shall we turn off the media? Global information can destroy local cooperation in the one-dimensional ring
6. Adam et Al., Maximum entropy approaches for the study of triadic motifs in the Mergers & Acquisitions network.

The Complexity72h Summer School has been organized by Rossana Mstrandrea and Tiziano Squartini, researchers at IMT, with Alberto Antonioni, researcher at the University of Madrid and Eugenio Valdano, researcher at the University of California, Los Angeles.



When soccer and science goes female: the 2nd Soccer Data Cup

Daniele Fadda, KDD Lab, CNR, Italy | fadda.daniele@gmail.com

Luca Pappalardo, KDD Lab, CNR, Italy | luca.pappalardo@isti.cnr.it

The Soccer Data Cup is an innovative initiative, unique of its kind, organized by the National Research Council of Italy (CNR) and University of Pisa, two partners of the SoBigData consortium, in collaboration with the Italian Ministry of Education (MIUR). The second edition was organized at L'Aquila (Italy) on 5-7 May 2019 and was **totally dedicated to women**.

In a first selection phase, six high schools were selected to represent six Italian regions: Abruzzo, Campania, Emilia-Romagna, Molise, Puglia, Umbria. Each high school then selected two teams: **a team of seven female soccer players and a team of two female "wannabe" data scientists**. Then, the competition was articulated in two parallel, but connected, competitions.

In the sports competition, the six teams of players faced each other in a futsal tournament: two Italian-style groups of three teams, semifinals and finals. During all matches, the players wore a device, produced by company Tracking4Fun, which monitored in great

detail the movements on the field. At the same time, matches were filmed so as to allow company Wyscout to detect, with the usage of its proprietary software, all the main spatio-temporal events that have occurred during the matches (passes, tackles, shots, fouls, etc.).

the data scientists coded in Python for three days, finally presenting their analyses in the last day to a committee of sports and science experts. The outcome of both the sports competition and the analytical one were used to decide the three finalists (Emilia-Romagna, Puglia and Umbria) that presented their analyses at the municipal theater.

The audience at the theater used an online app to vote for the best presentation, finally decreeing Puglia as the winner of the Soccer Data Cup.

The second edition of the Soccer Data Cup was a big success, suggesting a bright future for women both in soccer and the STEM disciplines. The young data scientists showed an enormous enthusiasm for data science and sports analytics as well as a great talent for science communication. This motivated SoBigData-

to organize a third edition where both men and women will attend in mixed teams, with the purpose of promoting the sound values of sports, STEM disciplines, and gender equality.



In parallel, the teams of data scientists analyzed the data produced during the matches, with the purpose of creating a critical analysis of the tournament. With the help of data scientists of the SoBigData infrastructure,



ESME: the pro-res workshop about, ethics, privacy and explainable ai

Francesca Pratesi, ISTI-CNR, Italy | francesca.pratesi@isti.cnr.it

Big data analytics and social mining

raise a number of ethical issues, especially as companies begin monetizing their data externally for purposes different from those for which the data was initially collected. The scale and ease with which analytics can be conducted today completely change the ethical framework. We can now do things that were impossible a few years ago, and existing ethical and legal frameworks cannot prescribe what we should do. Artificial Intelligence is becoming a disruptive technology, and resources for innovation are currently dominated by giant tech companies. To ensure European independence and leadership, we must invest wisely by bundling, connecting, and opening our AI resources having in mind ethical priorities such as transparency and fairness.



and they belong to 26 different institutions among 12 countries in total. The invited experts are from both academia and industry, and they have very different expertise and background (ranging from computer sci-

ence to law, from moral philosophy to sociology), in order to offer several different perspectives and to cover as many points of views as possible during the discussions. More detail about the invites and the event can be found at the website <https://kdd.isti.cnr.it/esme2019/>.

After that, participants openly discussed about ethical dilemmas, such as promoting ethics, reconciling ethical research and industry objectives, and analyzing the true applicability of the right to the explanation, starting from the not trivial problem to define what a good explanation is and how to measure it.

The collected material will be integrated into a supported guidance framework (<http://prores-project.eu/framework-portal/>), specifically built to provide a roadmap on ethics in research. In particular, the framework aims to offer user-friendly access to information, resources, existing guidelines and codes, case studies, and good practices.

Finally, the staff was composed of 14 researchers in total, composing the scientific committee, rapporteurs, and local committee. These researchers (7 male and 7 female) are members of ISTI - CNR, University of Pisa and Scuola Normale Superiore, and they are researchers of several career stages, from Ph.D. students to full professors.

The PRO-RES workshop about Ethics, Social Mining, and Explainable artificial intelligence (ESME) was held in Pisa, on July 8th and 9th, aiming to discuss the main open questions regarding privacy, explainability, and other ethical concerns. The ESME workshop was organized within the PRO-RES (PROmoting ethics and integrity in non-medical RESearch) EU project, with the support of SoBigData and AI4EU projects.

The event was scheduled in a day and a half, alternating presentations, parallel discussions, and high-level plenary discussions. 54 persons actively participated in the event; they are distinguished experts in the field,

The presentations (held in the first morning by 5 researchers, 2 female and 3 male) touched a variety of topics, such as trust in research, interdisciplinary differences as both resource and obstacle to the communication, positive-sum granting both ethics and utility, ethics-by-design, data life cycle, human-machine interaction, and moral machine experiments.



The Data Science for Society (DS4S) workshop

Laura Pollacci, University of Pisa, Italy | laura.pollacci@di.unipi.it

Within the 6th ACM Celebration of Women in Computing, womENCourage 2019, women engaged in STEAM areas have come together to enjoy a multidisciplinary program rich in both educational and networking activities. Thanks to this meeting opportunity, the Knowledge Discovery and Data Mining Laboratory (KDD Lab) organized the Data Science for Society (DS4S) workshop. The workshop allowed highlighting the need to fill the unequal distribution of men and women in STEAM fields. Works presented in the DS4S workshop are united by the combined use of Big Data with Social Data Mining techniques. Big data allows observing the ground truth of individual and collective behaviors. Then, social data mining provides the appropriate means to extract knowledge from big data offering a chance to understand the complexity of our contemporary, globally-interconnected society.

The workshop was allowed thanks to the effort of women involved in the European project SoBigData, a multidisciplinary European infrastructure for Big Data and Social Data Mining. SoBigData promotes gender equality fostering girl participation of women, e.g., providing scholarships in data-science summer schools. The DS4S workshop was attended by a heterogeneous audience for a total of about 80 officially enrolled. During the event, four different themes that aim to answer questions of increasing interest were explored: human migration, online debates, city of

citizens, and ethics.

Cristina Ioana Muntean presented a discussion about the perception of social phenomena through online social networks. With her work, she tried to answer to actual questions



about the evolution of the discussions about refugees' migration on Twitter, taking into account the sentiment of users across Europe concerning the refugee crisis.

Alina Sîrbu highlighted new possibilities introduced by big data to estimate migration stocks. She proposed a new index based on sentiment called Superdiversity that takes into

account geographical, language, and emotional dimensions. Alina Sîrbu also discussed the role of social networks and online media in shaping public debate and the problems related to algorithmic bias, i.e., fragmentation, polarization, and instability.

Diana Maynard described the language of political tweets by analysing social debates. The work she carried out shows that online abuse is increasing, and they might be particularly bad for particular communities, e.g., females, ethnic minorities, and LGBT.

Chiara Boldrini discussed possibilities posed by shared mobility. While car-sharing offers a wide range of benefits, its explosion caused some significant issues. An increasing of car-sharing depends on the possibility/ability to constantly study the phenomenon, e.g., through big data, as well as the ability to study effective solutions and apply them promptly.

Fosca Giannotti closed the workshop discussing the urgency of an open challenge on the Explicability, and the phenomena that it implies, that is both intelligibility (i.e., how does it work?) and accountability (i.e., who is responsible for?).

All presentations have been kindly provided by the speakers and can be found on the workshop site: Data Science for Society (DS4S).



Computational Social Science Summer School: 2nd edition of CSS summer school on migration

Jisu Kim, Scuola Normale Superiore, Italy | jisu.kim@sns.it

The Computational Social Science summer school was jointly organised and hosted by Jacobs University, Bremen International Graduate School of Social Sciences, University of Cagliari and Interuniversity Center for Social Science Theory and Methodology (ICS) and funded by Volkswagen Foundation. In addition, SoBigData generously offered two travel grants particularly aiming at supporting young female scholars.

The summer school was organised over three different but related themes over three years. Topic on migration was the second edition of the school for 2019. Previous year's edition focused on "Conflict" and the upcoming summer school, in 2020, is focused on "Social cohesion". It is a research incubator where it aims to promote collaborations of experts in computational social sciences to ad-

vance future research on three issues mentioned before using data-drive modeling.

During the summer school, participants were teamed up with experts/project leaders to work on a specific topic related to migration issues. They were expected to produce a draft manuscript at the end of school. Some of the projects proposed were; "Migration Data Innovation for Evidence-Informed Policymaking", "Digital and computational approaches to study migration" and "Analyzing the Social Integration of refugees through Text Analysis of Social media Data". The backgrounds of the participants varied across different subjects. It included economists, sociologists, computer scientists as well as engineers. The goal was to allow all participants interact across different disciplines. It allowed groups to exchange useful

feedbacks on projects from diverse perspectives.

Lectures and workshops were also organised by the summer school. The lectures were carried out by leading experts in the domain, giving participants to discover new and most recent works in the field. Workshops included tutorials on R and agent-based modelling programme, Netlogo.

The summer school provided participants very unique experiences and opportunities. It brought interesting ideas for research and is expected to be published soon.

BIGSSS Summer Schools in Computational Social Science

Research Incubators on Data-driven Modeling of Conflicts, Migration, and Social Cohesion





SoBigData Magazine is published under the
project N° 654024 | Programme: H2020 | INFRAIA

Duration: 1/09/2015 - 30/08/2019

Editorial Secretariat

info@sobigdata.eu

Editorial Board

Fosca Giannotti
Beatrice Rapisarda
Roberto Trasarti
Valerio Grossi

Layout and Design

Beatrice Rapisarda

Copyright notice

All authors, as identified in each article, retain copyright of their work.
The authors are responsible for the technical and scientific contents of their work.

Privacy statement

The personal data (names, email addresses...) and the other information entered in SoBigData Magazine will be treated according with the provision set out in Legislative Degree 196/2003 (known as Privacy Code) and subsequently integration and amendment.

Coordinator and Legal representative of the project: Fosca Giannotti | fosca.giannotti@isti.cnr.it

SOBIGDATA News is not for sale but is distributed for purposes of study and research and published online at
<http://www.sobigdata.eu/newsletter>

To subscribe/unsubscribe, please visit <http://www.sobigdata.eu/newsletter>