Social Mining & Big Data Analytics

SoBigData

RESEARCH INFRASTRUCTURE ++

Deliverable D8.3

Filling the gaps:

Emerging new analytical technologies 1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871042

DOCUMENT INFORMATION

PROJECT	
PROJECT ACRONYM	SoBigData-PlusPlus
PROJECT TITLE	SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics
STARTING DATE	01/01/2020 (60 months)
ENDING DATE	31/12/2024
PROJECT WEBSITE	http://www.sobigdata.eu
TOPIC	INFRAIA-01-2018-2019 Integrating Activities for Advanced Communities
GRANT AGREEMENT N.	871042

DELIVERABLE INFORMATION	
WORK PACKAGE	WP8 JRA1 - Social Mining and Big Data Resource Integration
WORK PACKAGE LEADER	LUH
WORK PACKAGE PARTICIPANTS	CNR, USFD, UNIPI, UT, IMT, LUH, SNS, AALTO, ETHZ, CNRS, CEU, URV, BSC, UPF, UvA
DELIVERABLE NUMBER	D8.3
DELIVERABLE TITLE	Filling the gaps: Emerging new analytical technologies 1
AUTHOR(S)	Giulio Rossetti (CNR), Francesca Pratesi (CNR), Michela Natilli (CNR)
CONTRIBUTOR(S)	
EDITOR(S)	Beatrice Rapisarda (CNR), Valerio Grossi (CNR)
REVIEWER(S)	Jesús A. Manjón Paniagua (URV), Alessio Rossi (UNIPI)
CONTRACTUAL DELIVERY DATE	30/06/2021
ACTUAL DELIVERY DATE	06/07/2021
VERSION	V1.1
TYPE	Report
DISSEMINATION LEVEL	Public
TOTAL N. PAGES	27
KEYWORDS	Machine learning, AI, complex networks, human mobility

EXECUTIVE SUMMARY

This deliverable gives a complete overview of all research activities related to WP8, in particular on the completed and ongoing methods, tools, and datasets integration activities at Month 18. It provides an up-to-date description of the algorithmic and data resources that are - or are planned to be - integrated within the SoBigData++ research infrastructure. The provided description has to be considered an incremental view of the resources available within the SoBigData++ research infrastructure that extends what is already reported in deliverables D8.1 and D8.2.

DISCLAIMER

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 871042.

SoBigData++ strives to deliver a distributed, Pan-European, multi-disciplinary research infrastructure for big social data analytics, coupled with the consolidation of a cross-disciplinary European research community, aimed at using social mining and big data to understand the complexity of our contemporary, globally-interconnected society. SoBigData++ is set to advance on such ambitious tasks thanks to SoBigData, the predecessor project that started this construction in 2015. Becoming an advanced community, SoBigData++ will strengthen its tools and services to empower researchers and innovators through a platform for the design and execution of large-scale social mining experiments.

This document contains information on SoBigData++ core activities, findings and outcomes and it may also contain contributions from distinguished experts who contribute as SoBigData++ Board members. Any reference to content in this document should clearly indicate the authors, source, organisation and publication date.

The content of this publication is the sole responsibility of the SoBigData++ Consortium and its experts, and it cannot be considered to reflect the views of the European Commission. The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

GLOSSARY

EU	European Union
EC	European Commission
H2020	Horizon 2020 EU Framework Programme for Research and Innovation
RI	Research Infrastructure

TABLE OF CONTENTS

1 Relevance to SoBigData++ 7
1.1 Purpose of this document 7
1.2 Relevance to project objectives 7
1.3 Relation to other work packages 7
1.4 Structure of the document 7
2 WP8 activities: organization and global statistics 9
2.1 Micro-projects 9
2.1.1 Activities summary 9
2.2 Integration modalities 10
3 Concluded integration activities 11
3.1 Task 8.3: Text and Social Media Mining services design and integration 11
3.2 Task 8.4: Complex Network Analysis Mining services design and integration 11
3.3 Task 8.5: Human Mobility Analytics services design and integration 12
3.4 Task 8.6: Web Analytics services design and integration 13
3.5 Task 8.11: Filling the gaps: emerging new analytical technologies 13
4 Ongoing/Planned integration activities 15
4.1 Task 8.1: Data Management and Integration of Social Data resources 15
4.2 Task 8.2: Social media observatory and crowd-sensing design and integration 15
4.3 Task 8.3: Text and Social Media Mining services design and integration 16
4.4 Task 8.4: Complex Network Analysis Mining services design and integration 16
4.5 Task 8.7: Visual Analytics services design and integration 17
4.6 Task 8.8: Privacy Enhancing Technology and Discrimination preventing services design and integration 18
4.7 Task 8.9: Explainable AI services design and integration 20
4.8 Task 8.10: Scalable machine learning services design and integration 21
4.9 Task 8.11: Filling the gaps: emerging new analytical technologies 21
5 Conclusions 22
Appendix A. References 23

1 Relevance to SoBigData++

1.1 Purpose of this document

The deliverable outlines a description of the newly integrated methods and datasets within the SoBigData++ research infrastructure and the planned/ongoing integration activities of M18. Provided descriptions include pointers to related WP3 and WP10 activities and a discussion on the specific level of integration for each of the reported resources.

1.2 Relevance to project objectives

This document provides an incremental view of WP8 integration activities and an outlook on the future resources that will be made available within the SoBigData++ research infrastructure.

1.3 Relation to other work packages

Work package 8 is part of “social mining research infrastructure building,” one of three axes the SoBigData++ work plan comprises. It is therefore strongly connected to the other work packages within the same axis, namely WP9 (“SoBigData e-Infrastructure and supercomputing network”) and WP10 (“Exploratories”). They are aimed at building the project core and infrastructure as well as advancing research in social mining.

Additionally, WP8 is connected to work packages in the “community building” axis, such as WP2 (“Critical Data Literacy, Ethics, and Legal Framework”), WP3 (“Dissemination, Impact, and Sustainability”), and WP4 (“Training”), as they go hand in hand with the creation of the platform and infrastructure. Finally, WP8 maintains connections to the work packages in the “user accessibility” axis, WP6 (“Transnational Access”) and WP7 (“Virtual Access”), as those dealing with providing access to the integrated resources.

This deliverable is intended to report on the methods/datasets resources that: are made available for exploitation in the connected WPs, have been integrated as a result of related WPs research activities, are outcomes of scientific publications within the consortium.

1.4 Structure of the document

The document is organized into four main sections:

- **Section 1:** provides an introduction to the aim of the deliverable and its relation with the other work packages;
- **Section 2:** describes how research activities are organized within the work package and introduces the different level of methods/datasets integration within the RI;
- **Section 3:** reports on the methods and tools integrated so far and their relations with WP3 and WP10 activities.

- **Section 4:** describes ongoing and planned activities related to integrating analytical methods/datasets within the research infrastructure.

2 WP8 activities: organization and global statistics

WP8 focuses on the integration of algorithmic and data resources within the SoBigData++ research infrastructure. This section briefly describes the two main aspects that underlie the research activities carried out within this work package: micro projects (2.1) and resources integration modalities (2.2).

2.1 Micro-projects

To better organize the activities related to the integration of algorithmic resources and datasets within the research infrastructure, WP8 leverages the concept of micro-projects.

A micro project is a commitment from a partner or more partners of the consortium over a period - typically 1-6 months - to produce a tangible outcome (dataset, method) to be made available to the community through the SoBigData platform.

Micro projects allow (i) partners to plan and organize their efforts explicitly, and (ii) task/work package leaders to timely track the ongoing research activities.

WP8 micro-projects target the integration of novel algorithmic resources/datasets within the RI and upgrade existing resource functionalities.

2.1.1 Activities summary

Since their first introduction in January 2021, 29 micro-projects involving WP8 have been submitted, and 14 have been completed. Figure 1 offers a breakdown of the micro-projects per task.

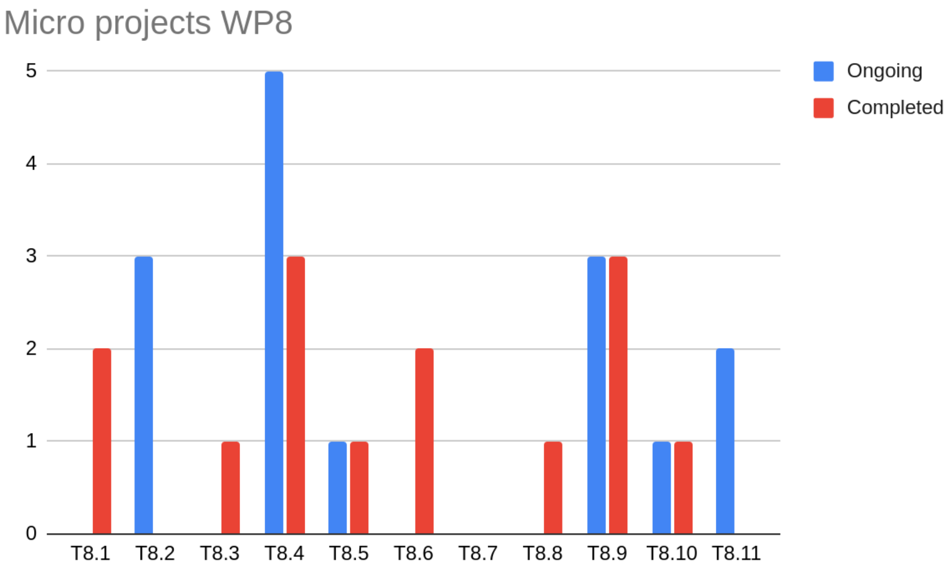

Figure 1. WP8 micro projects: task breakdown.

2.2 Integration modalities

As previously stated, WP8 micro projects' expected outcomes are methods/datasets to be integrated within the SoBigData++ RI. As underlined in D8.1, the integration of datasets will involve creating a dedicated entry in the Catalog.

However, the algorithmic resources can be integrated following different modalities within the RI. In particular, we devise three different integration levels:

A. Base integration: entry in the SoBigData++ Catalog;
B. Experiment prototyping: integration within the SoBigData++ Jupyter Hub
C. Engine integration: integration within the SoBigData++ Method Engine.

The "base integration" is the minimum requirement for a resource to be findable within the RI. It consists of a resource description through a fixed set of metadata and a link to its implementation and documentation.

The "experiment prototyping" integration level makes the algorithmic resource available for live experimental purposes within a SoBigData++ dedicated Jupyter Hub instance. This integration level allows RI users to prototype and execute their experiments using programming libraries developed within the consortium in a standard data science environment. Jupyter Hub integration is a novel feature for the SoBigData++ RI that has been introduced to ease the development of new social mining algorithms and methods. It provides access to a computing cluster that supports 80 concurrent users with 8 GB RAM per Jupyter notebook.

Finally, the "engine integration" level allows RI users to instantiate integrated methods using a visual interface - thus abstracting from code specificity - and run them on a dedicated experiment cluster.

In SoBigData++, we aim to integrate each resource t into at least two of the levels mentioned above (level A being mandatory).

3 Concluded integration activities

In this section, we report, for each task, the resources integrated within the RI as of M18. For each resource, it is specified its nature (either method or dataset), a brief description, the related WP10 exploratory (if any), the integration level, and the main/relevant references (as a link toward dissemination and impact, WP3).

3.1 Task 8.3: Text and Social Media Mining services design and integration

Method	Type	Description	WP10 Exploratory	Integration Level	Main Reference	Related Papers
DCI-based cross-language text classification	Methods	Cross-language text classification is a machine learning method that has been proposed to be published in the VRE.		A,B		

3.2 Task 8.4: Complex Network Analysis Mining services design and integration

Method	Type	Description	WP10 Exploratory	Integration Level	Main Reference	Related Papers
CDlib	Methods	Community Discovery library		A,B	[1]	[2-6]
NDlib	Methods	Network Diffusion Library		A,B	[56,57]	[58-61]
DyNetX	Methods	Dynamic Network analysis Library		A,B		[56,57]
Configuration Model	Methods	The code implements several variants of the entropy-based model known as Configuration Model: in particular, it allows for its implementation on monopartite (binary, weighted, undirected, directed) as well as on bipartite (binary, undirected) networks.	Economy & Finance 2.0	A,B	[7]	[8]

Bipartite Configuration Model and Validated Projection	Methods	The code implements the Bipartite Configuration Model and employs it to project any (binary, undirected) bipartite network over the layer of interest.	Economy & Finance 2.0	A,B	[9]	[10]
Generalized Network Dismantling	Methods	Implements network dismantling problem: dismantle into isolated subcomponents, thereby disrupting the malfunctioning of a system or containing the spread of misinformation or an epidemic		A	[11]	

3.3 Task 8.5: Human Mobility Analytics services design and integration

Method	Type	Description	WP10 Exploratory	Integration Level	Main Reference	Related Papers
STS-EPR	Methods	Implementation of a generative mobility model	Sustainable Cities for Citizens	A,B		
GeoSim	Methods	Implementation of the GeoSim generative mobility model	Sustainable Cities for Citizens	A,B	[12]	
Ground truth evaluation of home location detection algorithms	Methods	Evaluation of the accuracy of home detection algorithms and quantification of the amount of data	Economy & Finance 2.0	A		

		needed to carry out successful home detection for different mobile phone record streams.				
Mobility index for local quarantines in Chile	Dataset	Epidemiologically relevant metrics describing the mobility within and between comunas (sort of municipalities) in Chile.	Sustainable Cities for Citizens	A		
Mobility-emissions	Methods	Methods to estimate emissions starting from vehicles' GPS trajectories	Sustainable Cities for Citizens	A		

3.4 Task 8.6: Web Analytics services design and integration

Method	Type	Description	WP10 Exploratory	Integration Level	Main Reference	Related Papers
BoilerNet	Methods	Boilerplate removal (content extraction) from web pages		A, C		[13]

3.5 Task 8.11: Filling the gaps: emerging new analytical technologies

Method	Type	Description	WP10 Exploratory	Integration Level	Main Reference	Related Papers
SDNN24 from HR	Methods	Method to obtain SDNN24 from semi-continuous HR data obtaine by wrist-worn wearable	Sports Data Science	A	[14]	

Men vs women in soccer	Methods	Explain difference between men and women soccer players by a machine learning approach	Sports Data Science	A	[15]	
Injury forecaster	Methods	Framework of big data analytics to predict injuries in soccer from GPS data	Sports Data Science	A	[16]	
RPE predictor	Methods	Application of a Machine Learning model to predict the Rate of Perceived Exertion (RPE) from GPS data.	Sports Data Science	A	[17]	
Estimation of RHR	Methods	Method to estimate the resting heart rate (RHR) from wrist-worn devices.	Sports Data Science	A	[18]	
PassNet	Methods	Method for automatic event detection on soccer broadcasting video	Sports Data Science	A	[19]	

4 Ongoing/Planned integration activities

In this section, we report, for each task, the resources whose integration is ongoing/planned. For each resource, it is specified its nature (either method or dataset), a brief description, the related WP10 exploratory (if any), and the main/relevant references (as a link toward dissemination and impact, WP3).

4.1 Task 8.1: Data Management and Integration of Social Data resources

Method	Type	Description	WP10 Exploratory	Main Reference	Related Papers
Data Sprint	Methods	A technique or more precisely a format for organise workshops with and around data		[20]	

4.2 Task 8.2: Social media observatory and crowd-sensing design and integration

Method	Type	Description	WP10 Exploratory	Main Reference	Related Papers
Controversy Mapping	Methods	A series of techniques to explore and visualize sociotechnical debates	Societal Debates and Misinformation Analysis	[21]	[22]
RetweetCascadeGraph	Methods	Estimation of the retweet cascade graph	Societal Debates and Misinformation Analysis	[23]	

4.3 Task 8.3: Text and Social Media Mining services design and integration

Method	Type	Description	WP10 Exploratory	Main Reference	Related Papers
Classification of Wikipedia articles	Methods	Identify the human-labeled high-quality articles, e.g., "featured" ones, and differentiate them from the popular and controversial articles.	Societal Debates and Misinformation Analysis	[24]	

4.4 Task 8.4: Complex Network Analysis Mining services design and integration

Method	Type	Description	WP10 Exploratory	Main Reference	Related Papers
Human-Bot hybrid game	Methods	Human-Bot coordination game played on a virtual network		[25]	[26]
Conformity	Methods	Multi scale homophilic measure for attributed graphs		[27]	
CDlib	Methods	Community Discovery library update that includes bug fixes and more than 20 novel algorithms		[1]	[2-6]
Factor analysis methods for daily rhythms	Methods	Factoring and classification of chronotype using geolocated communication data	Migration Studies	[28]	
Tie-stability prediction using clustering	Methods	Clustering methods to study the effect of long-distance residential move within the	Migration Studies	[29]	

		country on mobile phone communication			
Fractal-network generator	Methods	Generating adjacency for regular fractal-like networks		[30]	
Visual Network Analysis	Methods	A technique to analyse networks by spatializing them through a force-directed algorithm and reading the resulting layout		[31]	
Library on network centrality measures	Methods	Python library on centrality measures in static and temporal networks and multiplexes. The library also includes the case of non-instantaneous link travel time , such as in transportation networks and multiplexes.		[32-35]	[36]
Library for statistical models of temporal networks	Methods	Python library for the inference, simulation, and forecasting of several models of temporal networks		[37,38]	[39]

4.5 Task 8.7: Visual Analytics services design and integration

Method	Type	Description	WP10 Exploratory	Main Reference	Related Papers
Visual analytics for social media research (or visual media analytics for short)	Methods	Collection of tools and techniques for visual analytics for social media research including how-to guide	Societal Debates and Misinformation Analysis		

4.6 Task 8.8: Privacy Enhancing Technology and Discrimination preventing services design and integration

Method	Type	Description	WP10 Exploratory	Main Reference	Related Papers
Privacy risk assessment	Methods	Library for computing the risk of re-identification of users in a dataset of sequential data (e.g., GPS trajectories). It is based on a methodology for assessing both the empirical privacy risk associated with users represented in the data, and the data quality guaranteed only with users not at risk.		[41,42]	[43-45]
k-anonymity via microaggregation	Methods	k-anonymity is a method to protect the privacy of individuals in a dataset while preserving the utility of the anonymized data. Microaggregation is a natural approach to satisfy k-anonymity. Microaggregation consist of two steps: i) partition of data into clusters and ii) aggregation of the values of each cluster		[46]	[47]
t-closeness through microaggregation	Methods	t-closeness improves k-anonymity protecting the dataset against attribute disclosure (attribute disclosure occurs if the confidential attribute is too similar for all k individuals in a cluster). This method uses microaggregation to generate k-anonymous and t-close data sets.		[48]	

Differential privacy via individual ranking	Methods	Differential privacy offers more robust privacy guarantees than k-anonymity and its extensions, at the cost of the utility of the anonymized data. To preserve the utility of the protected data, this method builds on microaggregation applied to each individual attribute. In this way, it is reduced the amount of noise needed to satisfy differential privacy		[49]	
dd	Methods	A library for discrimination discovery and sanitization		[50]	
Discrimination prevention method	Methods	Discrimination consist of unfairy treating of people in basis of their belonging to a specific group. Discrimination can be direct or indirect if decisions are based, respectively, on sensitive attributes or non-sensitive attributes strongly correlated with biased sensitive attributes. Antidiscrimination techniques include discrimination discovery and prevention. This method evaluates and treats the data set removing direct and/or indirect discrimination biases preserving data quality.		[51]	

4.7 Task 8.9: Explainable AI services design and integration

Method	Type	Description	WP10 Exploratory	Main Reference	Related Papers
XAILib	Methods	Library of explanation of machine learning models	Social Impacts of AI and Explainable Machine Learning		
XAILib-LORE	Methods	Rule-based local model agnostic explanation method	Social Impacts of AI and Explainable Machine Learning	[52]	
XAILib-LIME	Methods	Feature-based local model agnostic explanation method	Social Impacts of AI and Explainable Machine Learning		
XAILib-SHAP	Methods	Feature-based local model agnostic explanation method	Social Impacts of AI and Explainable Machine Learning		
XAILib-INTGRAD	Methods	Saliency-map-based DNN explanation method	Social Impacts of AI and Explainable Machine Learning		

4.8 Task 8.10: Scalable machine learning services design and integration

Method	Type	Description	WP10 Exploratory	Main Reference	Related Papers
Full Network Embedding	Methods	Extract features from a pre-trained CNN for posterior use on other models. May be integrated with SVM for classification.		[53]	[54]
ALIR	Methods	Asynchronous Training of Word Embeddings for Large Text Corpora		[55]	

4.9 Task 8.11: Filling the gaps: emerging new analytical technologies

Method	Type	Description	WP10 Exploratory	Main Reference	Related Papers
Pysoccer	Methods	Python library to unify soccer data and algorithms	Sports Data Science		
SoccerLogger	Methods	Data collection tools for soccer video	Sports Data Science		
Estimating countries' peace index with GDELT	Methods	Method for estimating peacefulness through the Global Peace Index (GPI), through the information extracted from Global Data on Events, Location, and Tone (GDELT) digital news database.	Economy & Finance 2.0	[40]	

5 Conclusions

This deliverable reports on the concluded and ongoing methods and datasets integration activities involving the SoBigData++ RI. The list of available algorithmic resources and datasets will be continuously updated throughout the project's lifetime.

References

[1] Rossetti, G., Milli, L., Cazabet, R. “CDLIB: a python library to extract, compare and evaluate communities from complex networks." Applied Network Science 4.1 (2019): 1-26.

[2] Citraro, S., Rossetti, G. "Eva: Attribute-aware network segmentation." International Conference on Complex Networks and Their Applications. Springer, Cham, 2019.

[3] Rossetti, G. "ANGEL: efficient, and effective, node-centric community discovery in static and dynamic networks." Applied Network Science 5.1 (2020): 1-23.

[4] Cazabet, R, Boudebza, S., Rossetti, G. "Evaluating community detection algorithms for progressively evolving graphs." arXiv preprint arXiv:2007.08635 (2020).

[5] Rossetti, G. "Exorcising the Demon: Angel, Efficient Node-Centric Community Discovery." International Conference on Complex Networks and Their Applications. Springer, Cham, 2019.

[6] Citraro, S., Rossetti, G. "Identifying and exploiting homogeneous communities in labeled networks." Applied Network Science 5.1 (2020): 1-20.

[7] Cimini, G., Squartini, T., Saracco, F., Garlaschelli, D., Gabrielli, A., & Caldarelli, G. The statistical physics of real-world networks. Nature Reviews Physics (2019), 1(1), 58-71.

[8] Vallarano, N., Bruno, M., Marchese, E., Trapani, G., Saracco, F., Squartini, T., ... & Zanon, M. Fast and scalable likelihood maximization for Exponential Random Graph Models. arXiv preprint (2021) arXiv:2101.12625.

[9] Saracco, F., Di Clemente, R., Gabrielli, A., & Squartini, T. Randomizing bipartite networks: the case of the World Trade Web. Scientific reports (2015), 5(1), 1-18.

[10] Saracco, F., Straka, M. J., Di Clemente, R., Gabrielli, A., Caldarelli, G., & Squartini, T. Inferring monopartite projections of bipartite networks: an entropy-based approach. New Journal of Physics (2017), 19(5), 053022.

[11] Ren, X. L., Gleinig, N., Helbing, D., Antulov-Fantulin, N. Generalized network dismantling. Proceedings of the national academy of sciences (2019), 116(14), 6554-6559.

[12] Toole, J. L., Herrera-Yaqüe, C., Schneider, C. M., González, M. C. Coupling human mobility and social ties. Journal of The Royal Society Interface (2015), 12(105), 20141128.

[13] Leonhardt, J., Anand, A., Khosla, M. Boilerplate Removal using a Neural Sequence Labeling Model. In Companion Proceedings of the Web Conference 2020 (pp. 226-229).

[14] Morelli, D., Rossi, A., Bartoloni, L., Cairo, M., Clifton, D. SDNN24 Estimation from Semi-Continuous HR Measures. Sensors, 2021

[15] Pappalardo, L., Rossi, A., Pontilio, G, Natilli, M., Cintia, P. Explaining the difference between men's and women's football. arXive, 2020 https://arxiv.org/abs/2101.01662

[16] Rossi, A., Pappalardo, L., Cintia, P., Iaia, F., Fernandez, J., Medina, D. Effective injury forecasting in soccer with GPS training data and machine learning. PLOS ONE (2018), vol. 13, p. 1-15, ISSN: 1932-6203, doi: 10.1371/journal.pone.0201264

[17] Rossi, A., Perri, E., Pappalardo, L., Cintia, P., Iaia, F. Relationship between External and Internal Workloads in Elite Soccer Players: Comparison between Rate of Perceived Exertion and Training Load. APPLIED SCIENCES (2019), vol. 9, p.1-11, ISSN: 2076-3417, doi: 10.3390/app9235174

[18] Morelli, D., Bartoloni, L., Rossi, A., Clifton, D. A computationally efficient algorithm to obtain an accurate and interpretable model of the effect of circadian rhythm on resting heart rate. PHYSIOLOGICAL MEASUREMENT (2019), vol. 40, ISSN: 1361-6579, doi: 10.1088/1361-6579/ab3dea

[19] Sorano, D., Carrara, F., Cintia, P. Falchi, F., Pappalardo, L. Automatic Pass Annotation from Soccer Video Streams Based on Object Detection and LSTM, ECML-PKDD 2020

[20] Venturini, T., Munk, A.,Meunier, A. Data-Sprint: a Public Approach to Digital Research. In Interdisciplinary Research Methods. Abingdon: Routledge. (2018)

[21] Venturini, T. Diving in Magma: How to Explore Controversies with Actor-Network Theory.Public Understanding of Science 19(3): 258–73. (2010). http://pus.sagepub.com/content/19/3/258.short

[22] Venturini, T. Building on Faults: How to Represent Controversies with Digital Methods. Public Understanding of Science 21(7): 796–812. (2012)

[23] Zola, P., Cola, G., Mazza, M., Tesconi, M. Interaction Strength Analysis to Model Retweet Cascade Graphs. Appl. Sci. 2020, 10(23), 8394; https://doi.org/10.3390/app10238394

[24] Ogushi, F., Kertesz, J., Kaski, K., Shimada, T., Ecology in the digital world of Wikipedia. arXiv:2105.10333v1 [physics.soc-ph] 21 May 2021

[25] Bhattacharya, K., Takko, T.,Monsivais, D., Kaski, K. Group formation on a small-world: experiment and modelling. Journal of the Royal Society Interface 16.156 (2019): 20180814.

[26] Takko, T., Bhattacharya, K., Monsivais, D., & Kaski, K. (2021). Human-agent coordination in a group formation game. Scientific Reports, 11(1), 1-10.

[27] Rossetti, G., Citraro, S., Milli, L.ù. "Conformity: A path-aware homophily measure for node-attributed networks." IEEE Intelligent Systems 36.1 (2021): 25-34.

[28] Roy, C., Monsivais, D., Bhattacharya, K., Dunbar, R. I., Kaski, K. Morningness-eveningness assessment from mobile phone communication analysis. bioRxiv (2021). (Accepted in Scientific Reports)

[29] Fudolig, M. I. D., Monsivais, D., Bhattacharya, K., Jo, H. H., Kaski, K. Internal migration and mobile communication patterns among pairs with strong ties. EPJ Data Science, 10(1), 1-21 (2021).

[30] Monsivais-Velazquez, D., Bhattacharya, K., Barrio, R. A., Maini, P. K., & Kaski, K. K. Dynamics of hierarchical weighted networks of van der Pol oscillators. Chaos: An Interdisciplinary Journal of Nonlinear Science, 30(12), 123146 (2020).

[31] Venturini, T., Jacomy, M., Jensen, P. What Do We See When We Look at Networks an Introduction to Visual Network Analysis and Force-Directed Layouts. Social Science Research Network (2020)

[32] Benzi, M., Boito, P. Matrix functions in network analysis, GAMM-Mitteilungen, Volume 43, Issue 3, (2020)

[33] Benzi, M., Klymko, C Total communicability as a centrality measure. Journal of Complex Networks, Volume 1, Issue 2, Pages 124–149, (2013)

[34] Zaoli, S., Mazzarisi, P., Lillo, F. Trip Centrality: walking on a temporal multiplex with non-instantaneous link travel time. Scientific Reports, 9, Article number: 10570 (2019),

[35] Zaoli, S., Mazzarisi, P., Lillo, F. Betweenness centrality for temporal multiplexes, Scientific Reports volume 11, Article number: 4919, (2021)

[36] Benzi, M., Chen, I., Chang, H., Hertzberg, V., Dynamic communicability and epidemic spread: a case study on an empirical dynamic contact network, Journal of Complex Networks, Volume 5, Issue 2, (2017)

[37] Williams, O., Lillo, F., Latora, V. Effects of memory on spreading processes in non-Markovian temporal networks. New Journal of Physics, Volume 21,

[38] Williams, O., Lillo, F., Latora, V. How auto- and cross-correlations in link dynamics influence diffusion in non-Markovian temporal networks. https://arxiv.org/abs/1909.08134v1

[39] Mazzarisi, P., Barucca, P., Lillo, F., Tantari, D. A dynamic network model with persistent links and node-specific latent variables, with an application to the interbank market. European Journal of Operational Research,

[40] Voukelatou, V., Pappalardo, L., Miliou, I., Gabrielli, L., Giannotti, F. Estimating countries' peace index through the lens of the world news as monitored by GDELT. International Conference on Data Science and Advanced Analytics DSAA 2020

[41] Pratesi, F., Monreale, A., Trasarti, R., Giannotti, F., Pedreschi, D. Yanagihara, T. PRUDEnce: a system for assessing privacy risk vs utility in data sharing ecosystems, Transactions on Data Privacy 11 (2018) 139–167

[42] Domingo-Ferrer, J., Torra, V. Disclosure risk assessment in statistical microdata protection via advanced record linkage. Statistics and Computing 13, 343–354 (2003). https://doi.org/10.1023/A:1025666923033

[43] Pellungrini, R., Pratesi, F., Pappalardo, L., Assessing privacy risk in retail data. International Workshop on Personal Analytics and Privacy, 17-22;

[44] Pellungrini, R., Pappalardo, L., Pratesi, F., Monreale, A., Analyzing Privacy Risk in Human Mobility Data. Federation of International Conferences on Software Technologies: Applications and Foundations

[45] Martínez, S., Sánchez, D., Valls, A., Batet, M. Privacy protection of textual attributes through a semantic-based masking method. Information Fusion, Volume 13, Issue 4, (2012)

[46] Domingo-Ferrer, J., Torra, V. Ordinal, Continuous and Heterogeneous k-Anonymity Through Microaggregation. Data Min Knowl Disc 11, 195–212 (2005). https://doi.org/10.1007/s10618-005-0007-5

[47] Sánchez D, Martínez S, Domingo-Ferrer J, Soria-Comas J, Batet M. μ -ANT: semantic microaggregation-based anonymization tool. Bioinformatics. 2020 Mar 1;36(5):1652-1653. doi: 10.1093/bioinformatics/btz792. PMID: 31621826.

[48] Soria-Comas, J., Domingo-Ferrer, J., Sánchez, D., Martínez, S. t-Closeness through Microaggregation: Strict Privacy with Enhanced Utility Preservation. In IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 11, pp. 3098-3110, 1 Nov. 2015, doi: 10.1109/TKDE.2015.2435777

[49] Sánchez, D., Domingo-Ferrer, J., Martínez, S., Soria-Comas, J. Utility-Preserving Differentially Private Data Releases Via Individual Ranking Microaggregation. arXiv:1512.02897

[50] Ruggieri, S.. Using t-closeness anonymity to control for non-discrimination. Transactions on Data Privacy. Vol. 7, Issue 2, August 2014, 99-129.

[51] Hajian, S., Domingo-Ferrer, J. A Methodology for Direct and Indirect Discrimination Prevention in Data Mining. In IEEE Transactions on Knowledge and Data Engineering, vol. 25, no. 7, pp. 1445-1459, July 2013, doi: 10.1109/TKDE.2012.72.

[52] Guidotti, R., Monreale, A., Giannotti, F., Pedreschi, D., Ruggieri, S., Turini, F. Factual and counterfactual explanations for black box decision making. IEEE Intelligent Systems 34.6 (2019): 14-23.

[53] Garcia-Gasulla, D., Vilalta, A., Parés, F., Ayguadé, E., Labarta, J., Cortés, U. and Suzumura, T., An out-of-the-box full-network embedding for convolutional neural networks. In 2018 IEEE International Conference on Big Knowledge (ICBK) (pp. 168-175). IEEE.

[54] Garcia-Gasulla, D., Parés, F., Vilalta, A., Moreno, J., Ayguadé, E., Labarta, J., Cortés, U. and Suzumura, T., 2018. On the behavior of convolutional nets for feature extraction. Journal of Artificial Intelligence Research, 61, pp.563-592.

[55] Anand, A., Khosla, M., Singh, J., Zab, J., Zhang, Z. Asynchronous Training of Word Embeddings for Large Text Corpora. ACM International Conference on Web Search and Data Mining 2019 https://doi.org/10.1145/3289600.3291011

[56] Rossetti, Giulio, et al. "NDlib: a python library to model and analyze diffusion processes over complex networks." International Journal of Data Science and Analytics 5.1 (2018): 61-79.

[57] Rossetti, G., Milli, L., Rinzivillo, S., Sirbu, A., Pedreschi, D., & Giannotti, F. (2017, October). Ndlib: Studying network diffusion dynamics. In 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA) (pp. 155-164). IEEE.

[58] Sîrbu, A., Pedreschi, D., Giannotti, F., & Kertész, J. (2019). Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. PloS one, 14(3), e0213246.

[59] Milli, L., Rossetti, G., Pedreschi, D., & Giannotti, F. (2018). Active and passive diffusion processes in complex networks. Applied network science, 3(1), 1-15.

[60] Milli, L., & Rossetti, G. (2019, December). Community-Aware Content Diffusion: Embeddedness and Permeability. In International Conference on Complex Networks and Their Applications (pp. 362-371). Springer, Cham.

[61] Grossi, V., Giannotti, F., Pedreschi, D., Manghi, P., Pagano, P., & Assante, M. (2021). Data science: a game changer for science and innovation. International Journal of Data Science and Analytics, 11(4), 263-278.