

# Curriculum Vitæ et Studiorum

Francesco Gullo

*Date and Place of Birth:* March 17, 1982 — Cosenza, Italy

*Citizenship:* Italian

*Address:* UniCredit, Via Molfetta, 101, 00171 Rome - Italy (*office*)

*E-mail:* [gullof@acm.org](mailto:gullof@acm.org)

*Web:* <http://uweb.deis.unical.it/gullo>

<https://www.linkedin.com/pub/francesco-gullo/7/7a1/271>

[https://www.researchgate.net/profile/Francesco\\_Gullo3](https://www.researchgate.net/profile/Francesco_Gullo3)

<http://scholar.google.it/citations?user=ueZsy3wAAAAJ&hl=en>

<https://www.scopus.com/authid/detail.uri?authorId=22034355400>

## SHORT BIO

---

Francesco Gullo is a senior researcher at UniCredit, R&D department. He received his Ph.D. from the University of Calabria (Italy) in January 2010. During his Ph.D., in 2009, he was an intern at the George Mason University, Fairfax VA, USA. From January 2010 to August 2011 he was a postdoctoral research fellow at the University of Calabria. Before joining UniCredit, he spent four years (from September 2011 to June 2015) in Barcelona, working at Yahoo Labs, first as a postdoctoral researcher, and, starting from September 2013, as a research scientist.

As a researcher, he has been used to working in both industrial and academic environments, and, as such, he likes to focus on both *applied research* and *basic research*.

His research interests fall into the broad area of *data science*, with special emphasis on *data mining*, *machine learning*, and *databases*. His work is centered on real-world problems that require *large-scale* data processing, and he usually tackles such problems from a *combinatorial-optimization* and *algorithmic* perspectives. Specifically, his current research is mainly devoted to *graph analytics*, *graph querying* and *graph mining*, as well as (*social*) *web mining* and *NLP*, while in the past he has also focused on data mining/machine learning for *high-dimensional/multi-faceted data*, *uncertain/probabilistic data*, *spatio-temporal data*, *biological data*, and *XML*. His research has been published in premier venues such as SIGMOD, VLDB, KDD, ICDM, CIKM, EDBT, WSDM, ECML-PKDD, SDM, TODS, TKDE, TKDD, Machine Learning, DAMI, JCSS, Pattern Recognition.

He has also been active in serving the data-mining/databases scientific community, by, e.g., acting as a Workshop Chair of ICDM'16, organizing workshops/symposia (MIDAS @ECML-PKDD['16,'17,'18], MultiClust 2014 mini-symposium @SDM'14, 4th MultiClust workshop @KDD'13, 3Clust workshop @PAKDD'12), or being part of the program committee of major conferences (KDD, WWW, ICDM, CIKM, WSDM, ECML-PKDD, ICWSM, SDM).

## WORK EXPERIENCE

---

[Jul'15—Present]

**Researcher**

UniCredit, Rome, Italy

R&D department

[Sep'13—Jun'15]

**Research Scientist**

Yahoo Labs, Barcelona, Spain

Web Mining group

- [Sep'11—Aug'13]     **Postdoctoral Researcher**  
 Yahoo Labs, Barcelona, Spain  
 Web Mining group
- [Jan'10—Aug'11]     **Postdoctoral Researcher**  
 University of Calabria, Italy  
 Department of Electronics, Computer, and Systems Science (DEIS, currently DIMES)
- [Apr'09—Sep'09]     **Short-term Visiting Scholar**  
 George Mason University, Fairfax, VA (USA)  
 Data Mining & Machine Learning group headed by Prof. Carlotta Domeniconi
- [Jan'06—Dec'09]     **Research & Teaching Assistant**  
 University of Calabria, Italy  
 Department of Electronics, Computer, and Systems Science (DEIS, currently DIMES)

## EDUCATION

---

- [Jan'10]             **Ph.D.** in Computer and Systems Engineering [*with highest honors*]  
 University of Calabria, Italy  
 Thesis: *Overcoming Uncertainty and the Curse of Dimensionality in Data Clustering*  
 Advisor: prof. Sergio Greco
- [Dec'05]             **M.Sc.** in Computer Engineering [*with highest honors*]  
 University of Calabria, Italy  
 Thesis: *Querying and Repairing Inconsistent XML Databases*  
 Advisor: prof. Sergio Greco
- [Oct'03]             **B.Sc.** in Computer Engineering [*with highest honors*]  
 University of Calabria, Italy  
 Thesis: *Semistructured Data and XML*  
 Advisor: prof. Sergio Greco

## RESEARCH ACTIVITY

---

His research interests fall into the broad area of *data science*, with special emphasis on *data mining*, *machine learning*, and *databases*. His work is centered on real-world problems that requires *large-scale data* processing, and he tackles such problem mainly from a *combinatorial-optimization* and *algorithmic* perspectives. Specifically, his research activity is mainly focused on the following topics:

- Graph analytics, graph querying, graph mining [1, 6, 8, 18, 20, 21, 22, 25, 26, 27, 30, 31, 32, 33, 35, 36, 52]
  - graph clustering; dense subgraph discovery; managing/mining/querying uncertain graphs; reachability/distance queries on graphs; querying graph databases; graph pattern mining
- Managing and mining data on the (social) Web [2, 4, 7, 19, 24, 28, 34, 36, 56, 63]
  - (social) web mining; (social) recommender systems; information propagation in complex networks; community search/detection; personalization of online services; ranking and centrality
- Natural Language Processing (NLP) [23, 51, 54]
  - entity recognition and disambiguation; distributional semantic embeddings

- Clustering high-dimensional and multifaceted data [5, 10, 12, 37, 38, 40, 41, 42, 47, 55, 62]
  - projective/subspace clustering; clustering ensembles; projective clustering ensembles; document clustering leveraging multiple views/classifications
- Uncertainty in data mining & machine learning [3, 11, 13, 29, 39, 43, 53, 59, 66, 67]
- Spatio-temporal (time series) data management [16, 57, 68]
- Bioinformatics [14, 17, 58, 60, 61]
  - querying and mining biological networks; modeling, mining, and analyzing proteomic data; modeling, mining, and analyzing gene expression data
- Querying and repairing inconsistent XML databases
- XML document clustering [15, 50, 64, 65]

## INDUSTRIAL EXPERIENCE

---

### PATENTS

X. Bai, B. B. Cambazoglu, F. Gullo, A. Mantrach, F. Silvestri  
*Using Exogenous Sources for Personalization of Website Services*  
 Patent Granted [US9792372B2, 2017-10-10]

H. Vahabi, F. Gullo  
*Method and/or System for Recommender System*  
 Published Patent Application [US20170061310A1, 2017-03-02]

F. Bonchi, A. Gionis, F. Gullo, A. Ukkonen  
*Method and System for Computing Relationship-Constrained Shortest Paths in Large Networks*  
 Defensive Publication

F. Bonchi, A. Gionis, F. Gullo, C. Tsourakakis  
*Method and System for Extracting High-quality Subgraphs from Large Graphs*  
 Defensive Publication

### INDUSTRIAL PROJECTS (SELECTED)

*Network-based Invoice Factoring*  
 Product: UC Balance  
 Skills: graph analytics, network flow, circuit enumeration  
 Technologies: Scala

*Email Classification for Talent Acquisition*  
 Product: UniCredit Aladin  
 Skills: NLP, supervised learning  
 Technologies: Java, Weka, JavaMail, EWS Java API

*Identification of Crisis Events for Business Continuity*  
 Product: UniCredit Hermes Crisis  
 Skills: NLP, entity recognition, entity disambiguation, sentiment analysis, text categorization  
 Technologies: Scala, Spark, HBase

*Entity Recognition and Disambiguation for Financial News*  
 Product: UniCredit Hermes News  
 Skills: NLP, entity recognition, entity disambiguation, MinHash, locality-sensitive hashing (LSH)  
 Technologies: Scala, Spark, HBase

*Detecting Corporate Group Structure from Unstructured Data*

Product: UniCredit EGM (Enhanced Group Mapping)

Skills: web scraping, learning from positive and unlabeled examples, NLP, entity recognition, entity disambiguation, computer vision

Technologies: Scala, Spark, Weka, OpenCV

*News Categorization for Crisis Event Detection*

Product: UniCredit Hermes Crisis

Skills: NLP, entity recognition, entity disambiguation, text mining

Technologies: Scala, Spark

*Exploiting Search History of Users for News Personalization*

Product: Yahoo News

Skills: personalization of online services, query-log analysis, big data processing

Technologies: Hadoop

*Predicting the Next Application a User is Going to Use*

Product: Yahoo Aviate

Skills: machine-learning predictive analytics, recommender systems, mobile data management, feature engineering, big data processing

Technologies: Hadoop, Hive

*Churn Analysis on Flickr*

Product: Flickr

Skills: churn analysis, predictive analytics, big data processing

Technologies: Matlab, Hadoop

*Optimal User Engagement through Social-Diffusion-aware Recommender Systems*

Product: Tumblr

Skills: recommender systems, information diffusion in complex networks, big data processing

Technologies: Hadoop

*Ticker Similarity Search*

Product: Yahoo Finance

Skills: online similarity search, time series analysis, time series motif discovery, locally-sensitive hashing (LSH)

Technologies: Java

*Recommendation of Items and Streams in Heterogeneous Information Networks*

Product: Flickr

Skills: recommender systems, heterogeneous information networks

Technologies: Hadoop

*Identification of Historical Trends in Organic Search via Query-log Analysis*

Product: Yahoo Sales

Skills: query-log analysis, big data processing

Technologies: Hadoop

PROTOTYPES  
(CO-)DEVELOPED

Hermes – A distributed-messaging tool for NLP

<http://hermes.rnd.unicredit.it:9603>

Skills: NLP, graph analytics

Technologies: Spark, Kafka, Redis, ElasticSearch, Hbase

MSPtool – Mass Spectra Preprocessing tool

<http://polifemo.deis.unical.it/~gtradigo/jnlp/msptool/>

Skills: bioinformatics, mining of biological data, time series analysis

Technologies: JFreeChart, Java Web Start, J2EE

TSCtool – Time Series Clustering tool  
Skills: time series analysis, mining of spatio/temporal data  
Technologies: JFreeChart

JPCE – Java-based Projective Clustering Ensembles tool  
[http://uweb.deis.unical.it/gullo/?page\\_id=22](http://uweb.deis.unical.it/gullo/?page_id=22)  
Skills: clustering ensembles, subspace clustering, projective clustering ensembles  
Technologies: Java

## SCIENTIFIC COMMUNITY SERVICE

---

### ORGANIZATION

#### **Workshop co-Chair**

*16th IEEE International Conference on Data Mining (ICDM '16)*

#### **Program co-Chair**

*[1st, 2nd, 3rd] Workshop on Mining Data for financial applications (MIDAS ['16, '17, '18])*

*In conjunction with the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery (ECML-PKDD ['16, '17, '18])*

#### **Industrial Session co-Chair**

*[2nd, 3rd] International Conference on Machine Learning, Optimization and Big Data (MOD ['16, '17])*

#### **Program co-Chair**

*Mini-Symposium on Multiple Clusterings, Multi-view Data, and Multi-source Knowledge-driven Clustering (MultiClust '14)*

*In conjunction with the 2014 SIAM Int. Conf. on Data Mining (SDM '14)*

#### **Program co-Chair**

*4th MultiClust Workshop: Multiple Clusterings, Multi-view Data, and Multi-source Knowledge-driven Clustering*

*In conjunction with the 19th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD '13)*

#### **Program co-Chair**

*1st International Workshop on Multi-view data, High-dimensionality, External Knowledge: Striving for a Unified Approach to Clustering (3Clust '12)*

*In conjunction with the 16th Pacific-Asia Conf. on Knowledge Discovery and Data Mining (PAKDD '12)*

### PROGRAM COMMITTEES

*International World Wide Web Conference (WWW)*

*[2014, 2016–2019]*

*ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining (KDD)*

*[2014–2018]*

*ACM International Conference on Information and Knowledge Management (CIKM) (Research Track)*

*[2012–2018]*

*ACM International Conference on Information and Knowledge Management (CIKM) (Industry/Case Studies Track)*

*[2017–2018]*

*IEEE International Conference on Data Mining (ICDM)*

*[2014–2016, 2018]*

*SIAM International Conference on Data Mining (SDM)*  
[2012, 2014–2018]

*European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD) (Research Track)*  
[2016–2018]

*European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD) (Applied Data Science Track)*  
[2015–2018]

*International AAAI Conference on Weblogs and Social Media (ICWSM)*  
[2014, 2017–2018]

*IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining (ASONAM)*  
[2015–2018]

*International Conference on Knowledge Discovery and Information Retrieval (KDIR)*  
[2016–2018]

*International Conference on machine Learning, Optimization and big Data (LOD)*  
[2017–2018]

*International Conference on Mathematics and Statistics (ICoMS)*  
[2018]

*International Workshop on Mining Attributed Networks (MATNet)*, in conjunction with the *International World Wide Web Conference (WWW)*  
[2018]

*Workshop on Social Aspects in Personalization and Search (SoAPS)*, in conjunction with the *European Conference on Information Retrieval (ECIR)*  
[2018]

*IEEE International Conference on Big Knowledge (ICBK)*  
[2017]

*IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*  
[2016]

*ACM International Conference on Web Search and Data Mining (WSDM)*  
[2013–2014]

*3rd MultiClust Workshop: Discovering, Summarizing and Using Multiple Clusterings*, in conjunction with the *SIAM International Conf. on Data Mining (SDM '12)*

*Int. Conf. on Emerging Intelligent Data and Web Technologies (EIDWT '11)*

INVITED TALKS

*Mining Massive Complex Graphs: New Frontiers and Methods*  
ISI Foundation  
Turin, Italy, September 2015

*From Patterns in Data to Knowledge Discovery: what Data Mining can do*  
3rd International Conference Frontiers in Diagnostic Technologies (ICFDT '13)  
Frascati, Italy, November 25–27, 2013

*Projective Clustering Ensembles*  
Yahoo Labs  
Barcelona, Spain, September 2011

*Information-Theoretic Hierarchical Clustering of Uncertain Data*  
George Mason University  
Fairfax VA, USA, April 2009

EXTERNAL REFEREING	<p><b>Journals:</b> TKDE, TKDD, TPAMI, VLDBJ, Machine Learning, DAMI, TIST, Information Systems, TWEB, Pattern Recognition, Internet Mathematics, KAIS, Information Sciences, SAM, AI Communications, AI in Medicine, JOCS, Computational Intelligence, BDR, IJDSN, IJITDM, JINT, JIIS, EAAI, ADAC, Neural Networks, JDIQ, Information Fusion</p> <p><b>Conferences:</b> SIGMOD, KDD, WWW, ICDE, ICDM, WSDM, CIKM, EDBT, SDM, ECML PKDD, PAKDD, DaWaK, CIDM, EDB, IDEAS, SEBD</p>
CONFERENCE PARTICIPATION	<p>ECML-PKDD '17 [2, 44], WIMS '17 [51], IEEE BigData '16 [23], ECML-PKDD '16 [46], ECML-PKDD '14 [29], KDD '14 [30], EDBT '14 [32, 33], VLDB '12 [13], SEBD '11, SIGMOD '11 [38], ICDM '10 [40, 39], ICDM '09 [41], IDEAS '09 [57], SDM '09 [42], ICDM '08 [43], SUM '08 [59], CBMS '07 [61]</p>

## TEACHING ACTIVITY

---

BOARDS	<p><b>External Scientific Board Member</b>  <i>Information and Communication Technologies</i> PhD Program          University of Calabria (Italy), Department of Computer Science, Modeling, Electronics and Systems Engineering (DIMES)          Academic years: '18-'19 (XXXIV cycle), '17-'18 (XXXIII cycle)</p>
COURSES TAUGHT AS AN INSTRUCTOR	<p><b><i>Advanced Learning and Mining Problems in Big Graph Data</i></b> (6 hours/year)          University of Calabria (Italy)          Department of Computer Science, Modeling, Electronics and Systems Engineering (DIMES)          PhD Program in Information and Communication Technologies          Academic year: '17-'18</p> <p><b><i>Computer Science</i></b> (55–84 hours/year)          University of Calabria (Italy)          Faculties of Engineering, Pharmacy, and Political Sciences, B.Sc. 1st year (remedial class project)          Academic years: '09-'10, '08-'09</p> <p><b><i>Computer Science</i></b> (16–24 hours/year)          “Magna Græcia” University of Catanzaro, Crotono campus (Italy)          Faculty of Medicine and Surgery, B.Sc. in Nurse Sciences (3rd year)          Academic years: '10-'11, '09-'10, '08-'09, '07-'08</p> <p><b><i>Laboratory Activity in Computer Science</i></b> (16 hours/year)          “Magna Græcia” University of Catanzaro, Catanzaro campus (Italy)          Faculty of Medicine and Surgery, B.Sc. in Dental Health (3rd year)          Academic years: '09-'10</p>
COURSES TAUGHT AS A TEACHING ASSISTANT	<p><b><i>Foundations of Computer Science</i></b> (18 hours/year)          University of Calabria (Italy)          Faculty of Engineering, B.Sc. in Computer Engineering (1st year)          Academic years: '10-'11, '09-'10, '08-'09</p> <p><b><i>Computer Science</i></b> (20 hours/year)          University of Calabria (Italy)          Faculty of Political Sciences, M.Sc. in Political Sciences (1st year)          Academic years: '10-'11, '08-'09, '07-'08, '06-'07</p>

***Internet Algorithms and Cryptography*** (13 hours/year)

University of Calabria (Italy)

Faculty of Engineering, M.Sc. in Computer Engineering (2nd year)

Academic years: '10-'11, '09-'10

***Web-based Information Systems*** (13 hours/year)

University of Calabria (Italy)

Faculty of Engineering, M.Sc. in Computer Engineering (2nd year)

Academic years: '09-'08, '08-'07, '06-'07, '05-'06

***Data and Knowledge Bases*** (13 hours/year)

University of Calabria (Italy)

Faculty of Engineering, M.Sc. in Management Engineering (1st year)

Academic years: '07-'08, '05-'06

***Algorithms and Data Structures*** (13 hours/year)

University of Calabria (Italy)

Faculty of Engineering, B.Sc. in Computer Engineering (2nd year)

Academic years: '06-'07

***Foundations of Computer Science I*** (30 hours/year)

“Magna Græcia” University of Catanzaro (Italy)

B.Sc. interuniversity course in Computer and Biomedical Engineering (1st year)

Academic years: '10-'11, '09-'10, '08-'09, '07-'08

***Computer Science*** (8 hours/year)

“Magna Græcia” University of Catanzaro, Catanzaro campus (Italy)

Faculty of Medicine and Surgery, B.Sc. in Health Services (3rd year)

Academic years: '09-'10

THESIS  
(CO-)ADVISING

Giacomo Legnaro

Sapienza University of Rome, M.Sc. in Data Science (A.Y. '17-'18)

Thesis: *Efficient and Effective Streaming Algorithms for Network-based Invoice Factoring*

Giulia Gavazzi

Sapienza University of Rome, M.Sc. in Data Science (A.Y. '16-'17)

Thesis: *Mining Dense and Non-redundant Subgraphs, with an Application to Event-graph Exploration*

Fabrizio Granieri

University of Calabria, B.Sc. in Computer Engineering (A.Y. '09-'10)

Thesis: *Graph Partitioning for Clustering Ensembles*

Ronny Meringolo

University of Calabria, B.Sc. in Computer Engineering (A.Y. '09-'10)

Thesis: *Graph Partitioning for Clustering Ensembles*

Antonio Senno

University of Calabria, M.Sc. in Computer Engineering (A.Y. '08-'09)

Thesis: *Clustering Ensembles Methods*

Giuseppe Scrivano

University of Calabria, B.Sc. in Computer Engineering (A.Y. '06-'07)

Thesis: *Multidimensional Time Series: Similarity Detection and Clustering*

Emanuele Forlano

University of Calabria, B.Sc. + M.Sc. in Computer Engineering (A.Y. '05-'06)

Thesis: *Algorithms for Time Series Clustering*

RESEARCH INTERN  
(CO-)SUPERVISING

Giulia Gavazzi, Giacomo Legnaro, Hugo Maldini  
M.Sc. students at Sapienza University of Rome  
Research Interns at UniCredit, Winter 2016 (Data Science Industrial Liaison Program)  
Project: *News Collection for Emerging-Event Detection* [51]

Alessio Areni, Cristiano Di Crescenzo, Elena Troccoli  
M.Sc. students at Sapienza University of Rome  
Research Interns at UniCredit, Winter 2016 (Data Science Industrial Liaison Program)  
Project: *Event Detection in Temporal Graphs* [51]

Pranay Anchuri  
Ph.D. student at Rensselaer Polytechnic Institute  
Research Intern at Yahoo Labs Barcelona, Summer 2014  
Project: *Mining Interesting Patterns from Uncertain Graphs*

Daniele Ramazzotti  
Ph.D. student at University of Milano-Bicocca  
Research Intern at Yahoo Labs Barcelona, Summer 2014  
Project: *Social Influence Detection by Probabilistic Causation and Spatial Proximity*

Edoardo Galimberti  
M.Sc. student at Politecnico of Milan  
Research Intern at Yahoo Labs Barcelona, Summer 2014  
Project: *Efficient and Effective Community Search* [7]

Natali Ruchansky  
Ph.D. student at Boston University  
Research Intern at Yahoo Labs Barcelona, Summer 2014  
Project: *The Minimum Wiener Connector Problem* [27]

Davide Mottin  
Ph.D. student at University of Trento  
Research Intern at Yahoo Labs Barcelona, Summer 2013  
Project: *Query Reformulation in Graph Databases* [26]

Panos Parchas  
Ph.D. student at Hong Kong University of Science and Technology  
Research Intern at Yahoo Labs Barcelona, Summer 2013  
Project: *Extracting Representative Instances of Uncertain Graphs* [31]

Lucrezia Macchia  
Ph.D. student at University of Bari  
Research Intern at Yahoo Labs Barcelona, Winter 2013  
Project: *Mining Summaries of Propagations* [34]

Arijit Khan  
Ph.D. student at University of California at Santa Barbara  
Research Intern at Yahoo Labs Barcelona, Summer 2012  
Project: *Fast Reliability Search on Uncertain Graphs* [33]

Charalampos Tsourakakis  
Ph.D. student at Carnegie Mellon University  
Research Intern at Yahoo Labs Barcelona, Summer 2012  
Project: *Extracting optimal quasi-cliques with quality guarantees* [35]

## HONORS AND AWARDS

---

### **National Scientific Habilitation** (*Abilitazione Scientifica Nazionale (ASN)*)

Granted by: Italian Ministry of Education, Universities and Research (*MIUR*)

Role: Associate Professor (*II fascia*)

Sector 01/B1 - Computer science (*Informatica*), valid from Aug 7<sup>th</sup>, 2018 to Aug 7<sup>th</sup>, 2024

Sector 09/H1 - Information processing systems (*Sistemi di elaborazione delle informazioni*), valid from Jul 27<sup>th</sup>, 2018 to Jul 27<sup>th</sup>, 2024

### **Habilitation to practice as an Engineer**

Section A, Field of Information Technology

Qualifying exam passed on October 2006 (1st 2006 session), no expiration date

### **ABI IT Innovation Award** (“Premio Innovazione IT – Tecnologie Disruptive”)

Granted by “Associazione Bancaria Italiana” (ABI), the trade association of Italian banks

Awarded system: UC Balance (an innovative system for network-based invoice factoring)

Personal contribution: design, theoretical characterization, and implementation of the algorithmic core behind the system

### **WWW 2017 Outstanding Reviewer Award**

Granted to reviewers who went well beyond their call of duty, provided exceptional reviews, and contributed a lot to discussions

### **BDTA 2016 Best Paper Award**

Paper “*Handling Uncertainty in Clustering Art-exhibition Visiting Styles*” [53]

“**BC**” label (“*paper where the results are backed by code*”)

from the 2015 Arizona University Repeatability in Computer Science study

(<http://reproducibility.cs.arizona.edu/v2>)

Paper “*Uncertain Centroid based Partitional Clustering of Uncertain Data*” [13]

### **SIGMOD 2011 Repeatability and Workability Evaluation (RWE) label**

Paper “*Advancing Data Clustering via Projective Clustering Ensembles*” [38]

### **SDM 2009 Student Travel Award**

Paper “*Diversity-based Weighting Schemes for Clustering Ensembles*” [42]

### **Italian Innovation Award** (*Premio Nazionale per l’Innovazione*) - I edition (2009)

*Fraud Detection via Load Profiles* project within *EUREKA! An idea for energy* initiative

## PUBLICLY-FUNDED R&D PROJECTS

---

### **TYPES** (*Towards transparencY and Privacy in the onlinE advertising business*)

Funding entity: EU-H2020

Project number: H2020-DS-2014-1 – Project ID 653449

### **SUPER** (*Social sensors for secUrity Assessments and Proactive EmERgencies management*)

Funding entity: EU-FP7

Project number: FP7-Security-2013.6.1-1 (<http://super-fp7.eu/>)

### **Cenit Social Media**

Funding entity: Spanish Centre for the Development of Industrial Technology

CENIT program, project CEN-20101037 (<http://www.cenitsocialmedia.es/>)

**GeoPKDD** (*Geographic Privacy-aware Knowledge Discovery and Delivery*)

Funding entity: EU-FET-014915

Project number: IST-6FP-014915 (<http://www.geopkdd.eu>)

**LogNET** (*An Innovative Network to Improve Logistics in Gioia Tauro*)

Subproject LOGICA (*laboratory of LOGIstics in CALabria*)

Funding entity: POR Calabria 2000-2006

**EUREKA!** An idea for energy

Project: *Fraud Detection via Load Profiles*

Funding entity: Enel University, area “Distribution Nets of Electricity”

## SELECTED PUBLICATIONS

---

### JOURNALS

- [1] A. Khan, F. Bonchi, F. Gullo, and A. Nufer. Conditional Reliability in Uncertain Graphs. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, ONLINE, 2018 (DOI: <http://dx.doi.org/10.1109/TKDE.2018.2816653>).
- [2] A. Tagarelli, A. Amelio, and F. Gullo. Ensemble-based Community Detection in Multilayer Networks. *Data Mining and Knowledge Discovery (DAMI)*, 31(5):1506-1543, 2017 (DOI: <http://dx.doi.org/10.1007/s10618-017-0528-8>).
- [3] F. Gullo, G. Ponti, A. Tagarelli, S. Greco. An Information-Theoretic Approach to Hierarchical Clustering of Uncertain Data. *Information Sciences*, 402:199-215, 2017 (DOI: <http://dx.doi.org/10.1016/j.ins.2017.03.030>).
- [4] X. Bai, B. Barla Cambazoglu, F. Gullo, A. Mantrach, F. Silvestri. Exploiting Search History of Users for News Personalization. *Information Sciences*, 385-386:125-137, 2017 (DOI: <http://dx.doi.org/10.1016/j.ins.2016.12.038>).
- [5] I. Assent, C. Domeniconi, F. Gullo, A. Tagarelli, A. Zimek. MultiClust 2013: Multiple Clusterings, Multi-view Data, and Multi-source Knowledge-driven Clustering. *SIGKDD Explorations*, 18(1):35-38, 2016 (DOI: <http://dx.doi.org/10.1145/2980765.2980769>).
- [6] P. Parchas, F. Gullo, D. Papadias, F. Bonchi. Uncertain Graph Processing through Representative Instances. *ACM Transactions on Database Systems (TODS)*, 40(3):20:1-20:39, 2015 (DOI: <http://dx.doi.org/10.1145/2818182>).
- [7] N. Barbieri, F. Bonchi, E. Galimberti, and F. Gullo. Efficient and Effective Community Search. *Data Mining and Knowledge Discovery (DAMI)*, 29(5):1406-1433, 2015 (DOI: <http://dx.doi.org/10.1007/s10618-015-0422-1>).
- [8] F. Bonchi, A. Gionis, F. Gullo, C. Tsourakakis, and A. Ukkonen. Chromatic Correlation Clustering. *Transactions on Knowledge Discovery from Data (TKDD)*, 9(4):34:1-34:24, 2015 (DOI: <http://dx.doi.org/10.1145/2728170>).
- [9] F. Gullo. From Patterns in Data to Knowledge Discovery: What Data Mining Can Do. *Physics Procedia*, 62:18-22, 2015 (DOI: <http://dx.doi.org/10.1016/j.phpro.2015.02.005>).
- [10] F. Gullo, C. Domeniconi, and A. Tagarelli. Metacluster-based Projective Clustering Ensembles. *Machine Learning*, 98(1-2):181-216, 2015 (DOI: <http://dx.doi.org/10.1007/s10994-013-5395-y>).
- [11] F. Gullo, G. Ponti, and A. Tagarelli. Minimizing the Variance of Cluster Mixture Models for Clustering Uncertain Objects. *Statistical Analysis and Data Mining (SAM)*, 6(2):116-135, 2013 (DOI: <http://dx.doi.org/10.1002/sam.11170>).

CONFERENCES

- [12] F. Gullo, C. Domeniconi, and A. Tagarelli. Projective Clustering Ensembles. *Data Mining and Knowledge Discovery (DAMI)*, 26(3):452-511, 2013 (DOI: <http://dx.doi.org/10.1007/s10618-012-0266-x>).
- [13] F. Gullo, and A. Tagarelli. Uncertain Centroid based Partitional Clustering of Uncertain Data. *Proceedings of the VLDB Endowment (PVLDB)*, 5(7):610-621, 2012.
- [14] F. Gullo, G. Ponti, A. Tagarelli, G. Tradigo, and P. Veltri. A Time Series Approach for Clustering Mass Spectrometry Data. *Journal of Computational Science (JOCS)*, 3(5):344-355, 2012 (DOI: <http://dx.doi.org/10.1016/j.jocs.2011.06.008>).
- [15] S. Greco, F. Gullo, G. Ponti, and A. Tagarelli. Collaborative Clustering of XML Documents. *Journal of Computer and System Sciences (JCSS)*, 77(6):988-1008, 2011 (DOI: <http://dx.doi.org/10.1016/j.jcss.2011.02.005>).
- [16] F. Gullo, G. Ponti, A. Tagarelli, and S. Greco. A Time Series Representation Model for Accurate and Fast Similarity Detection. *Pattern Recognition*, 42(11):2998-3014, 2009 (DOI: <http://dx.doi.org/10.1016/j.patcog.2009.03.030>).
- [17] F. Gullo, G. Ponti, A. Tagarelli, G. Tradigo, and P. Veltri. MaSDA: A System for Analyzing Mass Spectrometry Data. *Computer Methods and Programs in Biomedicine (CMPB)*, 95(2 suppl.):S12-S21, 2009 (DOI: <http://dx.doi.org/10.1016/j.cmpb.2009.02.011>).
- [18] E. Galimberti, A. Barrat, F. Bonchi, C. Cattuto, F. Gullo. Mining (maximal) Spancores from Temporal Networks. In *Proceedings of the ACM International Conference on Knowledge and Information Management (CIKM '18)*. Turin, Italy, October 22-26, 2018 (TO APPEAR).
- [19] F. Bonchi, F. Gullo, B. Mishra, D. Ramazzotti. Probabilistic Causal Analysis of Social Influence. In *Proceedings of the ACM International Conference on Knowledge and Information Management (CIKM '18)*. Turin, Italy, October 22-26, 2018 (TO APPEAR).
- [20] I. Bordino, F. Gullo. Network-based Invoice Factoring. In *Proceedings of the ACM International Conference on Knowledge and Information Management (CIKM '18)*. Turin, Italy, October 22-26, 2018 (TO APPEAR).
- [21] E. Galimberti, F. Bonchi, F. Gullo. Core Decomposition and Densest Subgraph in Multilayer Networks. In *Proceedings of the ACM International Conference on Knowledge and Information Management (CIKM '17)*, pp. 1807-1816. Singapore, November 6-10, 2017.
- [22] N. Ruchansky, F. Bonchi, D. Garcia Soriano, F. Gullo, N. Kourtellis. To Be Connected, or Not to Be Connected: That is the Minimum Inefficiency Subgraph Problem. In *Proceedings of the ACM International Conference on Knowledge and Information Management (CIKM '17)*, pp. 879-888. Singapore, November 6-10, 2017.
- [23] I. Bordino, A. Ferretti, M. Firrincieli, F. Gullo, M. Paris, S. Pascolutti, G. Sabena. Advancing NLP via a distributed-messaging approach. In *Proceedings of the 2016 IEEE International Conference on Big Data (IEEE BigData '16)*, pp. 1561-1568. Washington D.C., USA, December 5-8, 2016.
- [24] H. Vahabi, I. Koutsopoulos, F. Gullo, M. Halkidi. DifRec: a Social-Diffusion-aware Recommender System. In *Proceedings of the ACM International Conference on Knowledge and Information Management (CIKM '15)*, pp. 1481-1490. Melbourne, Australia, October 19-23, 2015.

- [25] A. Khan, F. Gullo, T. Wohler, F. Bonchi. Top-k Reliable Edge Colors in Uncertain Graphs. In *Proceedings of the ACM International Conference on Knowledge and Information Management (CIKM '15)*, pp. 1851-1854. Melbourne, Australia, October 19-23, 2015.
- [26] D. Mottin, F. Bonchi, and F. Gullo. Graph Query Reformulation with Diversity. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*, pp. 825-834. Sydney, Australia, August 10-13, 2015.
- [27] N. Ruchansky, F. Bonchi, D. Garcia Soriano, F. Gullo, and N. Kourtellis. The Minimum Wiener Connector Problem. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '15)*, pp. 1587-1602. Melbourne, Victoria, Australia, May 31-June 14, 2015.
- [28] O. Balalau, F. Bonchi, T-H. Hubert Chan, F. Gullo, and M. Sozio. Finding Subgraphs with Maximum Total Density and Limited Overlap. In *Proceedings of the International Conference on Web Search and Data Mining (WSDM '15)*, pp. 379-388. Shanghai, China, February 2-6, 2015.
- [29] F. Gullo, G. Ponti, and A. Tagarelli. Be certain of how-to before mining uncertain data. In *Proceedings of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD '14)* (Nectar Track), pp. 489-493. Nancy, France, September 15-19, 2014.
- [30] F. Bonchi, F. Gullo, A. Kaltenbrunner, and Y. Volkovich. Core Decomposition of Uncertain Graphs. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '14)*, pp. 1316-1325. New York City, New York (USA), August 24-27, 2014.
- [31] P. Parchas, F. Gullo, D. Papadias, and F. Bonchi. The Pursuit of a Good Possible World: Extracting Representative Instances of Uncertain Graphs. In *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD '14)*, pp. 967-978. Snowbird, Utah (USA), June 22-27, 2014.
- [32] F. Bonchi, A. Gionis, F. Gullo, and A. Ukkonen. Distance oracles in edge-labeled graphs. In *Proceedings of the International Conference on Extending Database Technology (EDBT '14)*, pp. 547-558. Athens, Greece, March 24-28, 2014.
- [33] A. Khan, F. Bonchi, A. Gionis, and F. Gullo. Fast Reliability Search in Uncertain Graphs. In *Proceedings of the International Conference on Extending Database Technology (EDBT '14)*, pp. 535-546. Athens, Greece, March 24-28, 2014.
- [34] L. Macchia, F. Bonchi, F. Gullo, and L. Chiarandini. Mining Summaries of Propagations. In *Proceedings of the IEEE International Conference on Data Mining (ICDM '13)*, pp. 498-507. Dallas, Texas (USA), December 7-10, 2013.
- [35] C. E. Tsourakakis, F. Bonchi, A. Gionis, F. Gullo, and M. A. Tsarli. Denser than the densest subgraph: extracting optimal quasi-cliques with quality guarantees. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '13)*, pp. 104-112. Chicago, Illinois (USA), August 11-14, 2013.
- [36] F. Bonchi, A. Gionis, F. Gullo, and A. Ukkonen. Chromatic Correlation Clustering. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '12)*, pp. 1321-1329. Beijing, China, August 12-16, 2012.

- [37] F. Gullo, A. K. A. Talukder, S. Luke, C. Domeniconi, and A. Tagarelli. Multiobjective Optimization of Co-Clustering Ensembles. In *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '12)*, pp. 1495-1496. Philadelphia, PA (USA), July 7-11, 2012.
- [38] F. Gullo, C. Domeniconi, and A. Tagarelli. Advancing Data Clustering via Projective Clustering Ensembles. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data (SIGMOD '11)*, pp. 733-744. Athens, Greece, June 12-16, 2011.
- [39] F. Gullo, G. Ponti, and A. Tagarelli. Minimizing the Variance of Cluster Mixture Models for Clustering Uncertain Objects. In *Proceedings of the 10th IEEE International Conference on Data Mining (ICDM '10)*, pp. 839-844. Sydney, Australia, December 14-17, 2010.
- [40] F. Gullo, C. Domeniconi, and A. Tagarelli. Enhancing Single-Objective Projective Clustering Ensembles. In *Proceedings of the 10th IEEE International Conference on Data Mining (ICDM '10)*, pp. 833-838. Sydney, Australia, December 14-17, 2010.
- [41] F. Gullo, C. Domeniconi, and A. Tagarelli. Projective Clustering Ensembles. In *Proceedings of the 9th IEEE International Conference on Data Mining (ICDM '09)*, pp. 794-799. Miami, Florida (USA), December 6-9, 2009.
- [42] F. Gullo, A. Tagarelli, and S. Greco. Diversity-based Weighting Schemes for Clustering Ensembles. In *Proceedings of the 9th SIAM International Conference on Data Mining (SDM '09)*, pp. 437-448. Sparks, Nevada (USA), April 30-May 2, 2009.
- [43] F. Gullo, G. Ponti, A. Tagarelli, S. Greco. A Hierarchical Algorithm for Clustering Uncertain Data via an Information-Theoretic Approach. In *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM '08)*, pp. 821-826. Pisa, Italy, December 15-19, 2008.

## OTHER PUBLICATIONS

---

### EDITED VOLUMES

- [44] I. Bordino, G. Caldarelli, F. Fumarola, F. Gullo, T. Squartini. *Proceedings of the First Workshop on Mining Data for financial applicationS (MIDAS 2017)*, co-located with the 2017 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery (ECML-PKDD 2017), Skopje, Macedonia, September 18, 2017. CEUR Workshop Proceedings Vol-1941, CEUR-WS.org, 2017.
- [45] C. Domeniconi, F. Gullo, F. Bonchi, J. Domingo-Ferrer, R. Baeza-Yates, Z.-H. Zhou, X. Wu. *16th IEEE International Conference on Data Mining Workshops (ICDMW 2016)*, Barcelona, Spain, December 12, 2016. IEEE Computer Society, 2016, ISBN 978-1-5090-5472-5.
- [46] I. Bordino, G. Caldarelli, F. Fumarola, F. Gullo, T. Squartini. *Proceedings of the First Workshop on Mining Data for financial applicationS (MIDAS 2016)*, co-located with the 2016 European Conference on Machine Learning and Principles and Practice of Knowledge Discovery (ECML-PKDD 2016), Riva del Garda, Italy, September 19, 2016. CEUR Workshop Proceedings Vol-1774, CEUR-WS.org, 2016.
- [47] I. Assent, C. Domeniconi, F. Gullo, A. Tagarelli, and A. Zimek. *MultiClust '13: Proceedings of the 4th MultiClust Workshop on Multiple Clusterings, Multi-view Data, and Multi-source Knowledge-driven Clustering*, co-located with the KDD '13 conference, Chicago, Illinois (USA), August 11 - 14, 2013. ACM, 2013, ISBN 978-1-4503-2334-5.

- [48] T. Washio, J. Luo, P. Desikan, K.-W. Hsu, J. Srivastava, E.-P. Lim, M. Teisseire, M. Roche, C. Domeniconi, F. Gullo, A. Tagarelli, H. K. Tan, and W. C. Omn. *Emerging Trends in Knowledge Discovery and Data Mining - PAKDD 2012 International Workshops: DMHM, GeoDoc, 3Clust, and DSDM, Kuala Lumpur, Malaysia, May 29 - June 1, 2012, Revised Selected Papers*. Lecture Notes in Computer Science 7769, Springer, 2013, ISBN 978-3-642-36777-9.

BOOK CHAPTERS  
(REFEREED)

- [49] F. Bonchi, F. Gullo, A. Kaltenbrunner. Core Decomposition of Massive, Information-Rich Graphs. In *Encyclopedia of Social Network Analysis and Mining - 2 edition*, Springer, 2017 (DOI: [http://dx.doi.org/10.1007/978-1-4614-7163-9\\_110176-1](http://dx.doi.org/10.1007/978-1-4614-7163-9_110176-1)).
- [50] F. Gullo, G. Ponti, and S. Greco. Organizing XML Documents on a Peer-to-Peer Network by Collaborative Clustering. In *XML Data Mining: Models, Methods, and Applications*, IGI Global, 2012, pp. 449-466 (DOI: <http://dx.doi.org/10.4018/978-1-61350-356-0.ch018>).

OTHER  
INTERNATIONAL  
CONFERENCES

- [51] C. Di Crescenzo, G. Gavazzi, G. Legnaro, E. Troccoli, I. Bordino, and F. Gullo. HERMEVENT: A News Collection for Emerging-Event Detection. In *Proceedings of the Int. Conf. on Web Intelligence, Mining and Semantics (WIMS '17)*. Amantea, Italy, June 19-22, 2017.
- [52] F. Bonchi, I. Bordino, F. Gullo, and G. Stilo. Identifying Buzzing Stories via Anomalous Temporal Subgraph Discovery. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence (WI '16)*, pp. 161-168. Omaha, Nebraska, USA, October 13-16, 2016.
- [53] F. Gullo, G. Ponti, A. Tagarelli, S. Cuomo, P. De Michele, F. Piccialli, Handling Uncertainty in Clustering Art-exhibition Visiting Styles. In *Proceedings of the 7th EAI International Conference on Big Data Technologies and Applications (BDTA '16)*. Seoul, South Korea, November 17-18, 2016.
- [54] I. Bordino, A. Ferretti, M. Firrincieli, F. Gullo, M. Paris, G. Sabena. Hermes: A Distributed-Messaging Tool for NLP. In *Proceedings of the 2nd International Conference on Machine Learning, Optimization, and Big Data (MOD '16)*, pp. 402-406. Volterra, Italy, August 26-29, 2016.
- [55] S. Romeo, A. Tagarelli, F. Gullo, and S. Greco. A Tensor-based Clustering Approach for Multiple Document Classifications. In *Proceedings of the International Conference on Pattern Recognition Applications and Methods (ICPRAM '13)*, pp. 200-205. Barcelona, Spain, February 15-18, 2013.
- [56] A. Tagarelli, and F. Gullo. Evaluating PageRank Methods for Structural Sense Ranking in Labeled Tree Data. In *Proceedings of the Int. Conf. on Web Intelligence, Mining and Semantics (WIMS '12)*. Craiova, Romania, June 13-15, 2012.
- [57] F. Gullo, G. Ponti, A. Tagarelli, S. Iiritano, M. Ruffolo, and D. Labate. Low-voltage Electricity Customer Profiling based on Load Data Clustering. In *Proceedings of the 13th International Database Engineering & Applications Symposium (IDEAS '09)*, pp. 330-333. Cetraro, Italy, September 16-18, 2009.
- [58] F. Gullo, G. Ponti, A. Tagarelli, G. Tradigo, P. Veltri. Hierarchical Clustering of Microarray Data with Probe-level Uncertainty. In *Proceedings of the 22th IEEE International Symposium on Computer-Based Medical Systems (CBMS '09)*. Albuquerque, New Mexico (USA), August 3-4, 2009.
- [59] F. Gullo, G. Ponti, and A. Tagarelli. Clustering Uncertain Data via K-medoids. In *Proceedings of the 2nd International Conference on Scalable Uncertainty Management (SUM '08)*, pp. 229-242. Napoli, Italy, October 1-3, 2008.

- [60] F. Gullo, G. Ponti, A. Tagarelli, G. Tradigo, and P. Veltri. MSPtool: A Versatile Tool for Mass Spectrometry Data Preprocessing. In *Proceedings of the 21th IEEE International Symposium on Computer-Based Medical Systems (CBMS '08)*, pp. 209-214. Jyväskylä, Finland, June 17-19, 2008.
- [61] F. Gullo, G. Ponti, A. Tagarelli, G. Tradigo, and P. Veltri. A Time Series Based Approach for Classifying Mass Spectrometry Data. In *Proceedings of the 20th IEEE International Symposium on Computer-Based Medical Systems (CBMS '07)*, pp. 412-417. Maribor, Slovenia, June 20-23, 2007.
- [62] C. Domeniconi, F. Gullo, and A. Tagarelli. The Projective Clustering Ensemble Problem for Advanced Data Clustering. In *Proceedings of the 25th Italian Symposium on Advanced Database Systems (SEBD '17)*. Squillace Lido (Catanzaro), Italy, June 25-29, 2017.
- [63] A. Freire, M. Manca, D. Saez-Trumper, D. Laniado, I. Bordino, F. Gullo, and A. Kaltenbrunner. Graph-Based Breaking News Detection on Wikipedia. In *Proceedings of the 2016 International Wiki Workshop*, in conjunction with the *10th International AAAI Conference on Web and Social Media (Wiki@ICWSM '16)*. Cologne, Germany, May 17, 2016.
- [64] S. Greco, F. Gullo, G. Ponti, A. Tagarelli, and G. Agapito. Clustering XML Documents: a Distributed Collaborative Approach. In *Proceedings of the 18th Italian Symposium on Advanced Database Systems (SEBD '10)*, pp. 406-413. Rimini, Italy, June 20-23, 2010.
- [65] S. Greco, F. Gullo, G. Ponti, and A. Tagarelli. Collaborative Clustering of XML Documents. In *Proceedings of the 1st International Workshop on Distributed XML Processing: Theory and Practice (DXP '09)*, in conjunction with the *38th International Conference on Parallel Processing (ICPP '09)*. Vienna, Austria, September 22-25, 2009.
- [66] F. Gullo, G. Ponti, A. Tagarelli, and S. Greco. Information-Theoretic Hierarchical Clustering of Uncertain Data. In *Proc. of the 17th Italian Symposium on Advanced Database Systems (SEBD '09)*, pp. 273-280. Geneva, Italy, June 21-24, 2009.
- [67] F. Gullo, and G. Ponti. Hierarchical Clustering of Uncertain Data. In *Doctoral Symposium* in conjunction with the *14th GII Doctoral School on Advances in Databases*, Cetraro, Italy, September 19, 2009.
- [68] F. Gullo, G. Ponti, A. Tagarelli, and S. Greco. Accurate and Fast Similarity Detection in Time Series. In *Proceedings of the 15th Italian Symposium on Advanced Database Systems (SEBD '07)*, pp. 172-183. Bari, Italy, June 17-20, 2007.