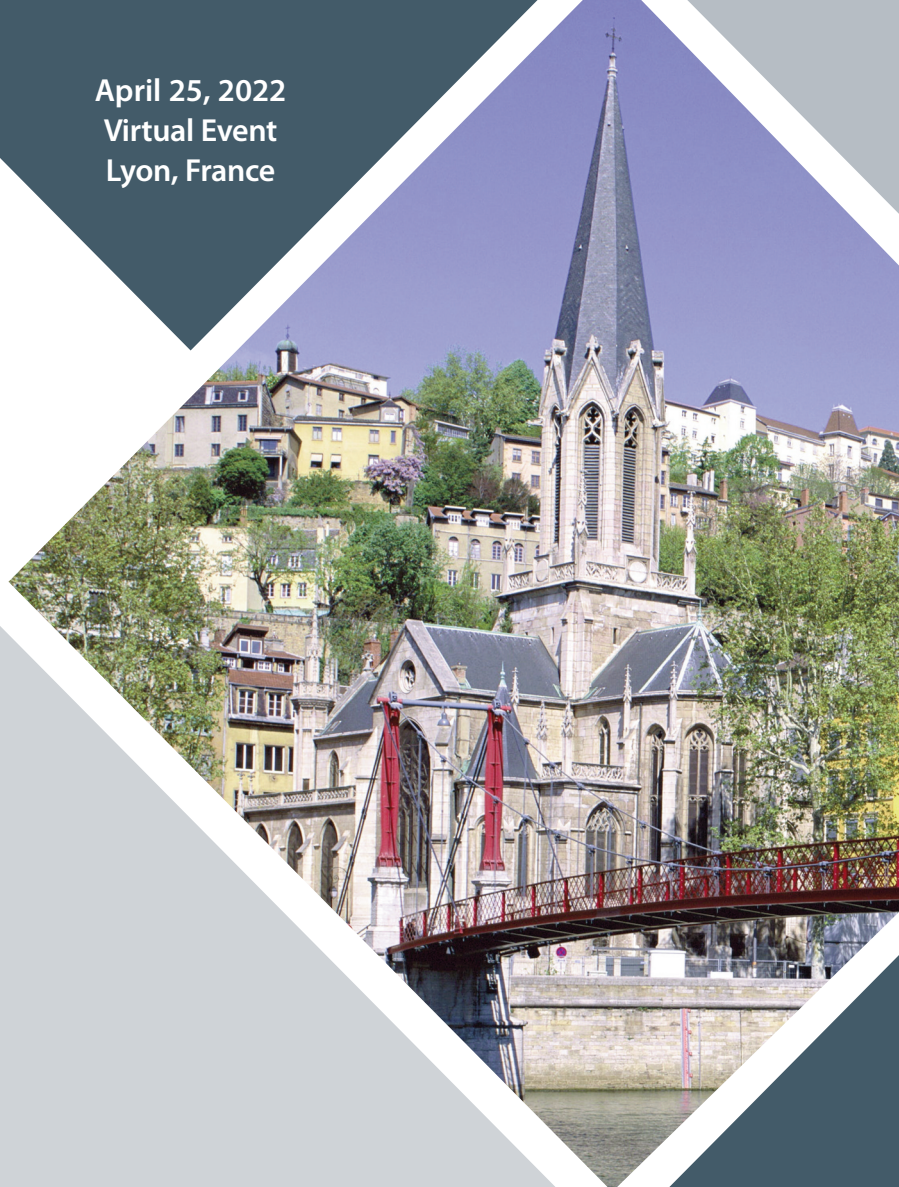




Association for  
Computing Machinery

April 25, 2022  
Virtual Event  
Lyon, France



# WWW '22 Companion

Companion Proceedings of the  
Web Conference 2022

Sponsor:

**ACM SIGWEB**

General Co-Chairs:

**Frédérique Laforest, INSA Lyon, France**

**Raphaël Troncy, EURECOM, France**

Program Co-Chairs:

**Elena Simperl, King's College London, UK**

**Deepak Agarwal, Pinterest, USA**

**Aristides Gionis, KTH Royal Institute of Technology, Sweden**

Proceedings Co-Chairs:

**Ivan Herman, W3C / retired**

**Lionel Médini, Université Lyon 1, France**



**The Association for Computing Machinery**  
1601 Broadway, 10<sup>th</sup> Floor  
New York, NY 10019-7434

Copyright © 2022 by the Association for Computing Machinery, Inc. (ACM). Permission to make digital or hard copies of portions of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyright for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permission to republish from: [permissions@acm.org](mailto:permissions@acm.org) or Fax +1 (212) 869-0481.

For other copying of articles that carry a code at the bottom of the first or last page, copying is permitted provided that the per-copy fee indicated in the code is paid through [www.copyright.com](http://www.copyright.com).

**ISBN:** 978-1-4503-9130-6

Additional copies may be ordered prepaid from:

**ACM Order Department**  
PO Box 30777  
New York, NY 10087-0777, USA

Phone: 1-800-342-6626 (USA and Canada)  
+1-212-626-0500 (Global)  
Fax: +1-212-944-1318  
E-mail: [acmhelp@acm.org](mailto:acmhelp@acm.org)  
Hours of Operation: 8:30 am – 4:30 pm ET

Printed in the USA

# Foreword from the General Chairs

It is our great pleasure to welcome you to *The ACM Web Conference 2022 held online on 25-29 April 2022 and hosted by Lyon, France*. It is the 31<sup>st</sup> edition of a series of yearly international conferences on the future directions of the Web.

This series began in 1994 at CERN, and it was formerly known as the International World Wide Web Conference, abbreviated as WWW. This conference has been the premier venue to present and discuss progress in research, development, standards, and applications of the topics related to the Web. The conference assembles scholars, researchers, policymakers, practitioners, and end-users with one unifying goal: to envision and create the future of the Web.

Over the past three decades, The Web Conference has been the forum where some of the most fundamental Web technologies have been introduced, such as the Anatomy of a Large Scale Web Search Engine in 1998 prefiguring Google, the EigenTrust algorithm in 2003 influencing reputation management distributed networks, and the YAGO knowledge base in 2007 anticipating the numerous developments over knowledge graphs.

The program of this year's edition includes three distinguished keynote talks by world-renowned experts: **Prabhakar Raghavan** of Google, **Virginia Dignum** of Umeå University and **Jaime Teevan** of Microsoft. The high-quality program is made of 323 selected papers spread in 11 research sessions, in addition to posters and demonstrations, a PhD symposium for the junior scholars, an Industry track, a Web developer and W3C track for the practitioners as well as 20 tutorials and 21 workshops reflecting the novel topics of interests for the Web. A journal track built in coordination with the TWEB journal also offers the possibility to authors of 3 journal papers to come and present their work to the participants. The scientific team has cooked some special tracks presenting new emerging topics such as eSports and online gaming, a dedicated track for the Web for good, inclusiveness and fairness, as well as a track discussing the History of the Web, its evolution and impact. Finally, following the tradition, the W4A (Web for All) conference is a co-located event highlighting the importance of researching methods and tools to build an accessible Web.

The various calls for papers have attracted a record number of submissions from all over the world, which have been thoroughly reviewed by a very large program committee. We are very grateful to **Deepak Agarwal** of Pinterest, **Aristides Gionis** of KTH Royal Institute of Technology and **Elena Simperl** of King's College London, our amazing Program Chairs, and the entire Organizing Committee, the 32 track chairs, the senior program committee members and all reviewers who are the ones who have worked very hard in selecting this high quality content:

<i>Track</i>	<i>Reviewed</i>	<i>Accepted</i>	
Research tracks	1822	323	17.7%
Special tracks	198	41	20.7%
Poster and demo	107	35	33%
PhD Symposium	22	10	45%
Web developer and W3C	5	4	80%
Industry track	80	25	31%

This year's motto is ***Let's make the Web Better, Faster, Stronger*** in order to emphasize the scientific enhancements the Web still requires as well as our collective obligation to pay attention to the effects of the Web on society, human beings and life in general.

This year is finally the moment of the transfer from the **International World Wide Web Conference Committee** (IW3C2) to the **ACM SIGWEB** community for the management of the Conference series, to ensure the future success of the conference.

We wish to thank IW3C2, ACM SIGWEB, all our supporters (Orange, TII, Pinterest, LinkedIn, Inria, Amazon, Baidu, Bloomberg, Google, Microsoft, eBay, Spotify, Megagon Labs, Booking.com, Meta, data.world) for their support, as well as the local organizing committee (Pierre-Antoine Champin, Yingtong Dou, Laurent Flory, Coralie Grégoire, Ivan Herman, Yoelle Maarek, Luc Mariaux, Lionel Médini), our institutional partners (INSA Lyon, EURECOM, LIRIS, Université de Lyon), and our PCO (INSAValor) for their hard work in making the 2022 edition an outstanding conference.

*The Web Conference 2022 General co-chairs,*

**Frédérique Laforest**

*INSA Lyon, France*

**Raphaël Troncy**

*EURECOM, France*

# The ACM Web Conference 2022 Organization Committee

**General Co-Chairs** Frédérique Laforest (*INSA Lyon, France*)  
Raphaël Troncy (*EURECOM, France*)

**Program Co-Chairs** Elena Simperl (*King's College London, UK*)  
Deepak Agarwal (*Pinterest, USA*)  
Aristides Gionis (*KTH Royal Institute of Technology, Sweden*)

## Research Tracks Chairs

**Economics, Monetization,  
and Online Markets** Vahab Mirrokni (*Google Research, USA*)  
Kamesh Munagala (*Duke University, USA*)

**Search** Ismail Sengor Altingovde (*Middle East Technical University, Turkey*)  
Claudio Lucchese (*Università Ca' Foscari Venezia, Italy*)

**Security, Privacy, and Trust** Enrico Gerding (*University of Southampton, UK*)  
Andrew Paverd (*Microsoft, UK*)

**Semantics and Knowledge** Axel Polleres (*Business University Vienna, Austria*)  
Valentina Presutti (*University of Bologna, ISTC CNR, Italy*)

**Social Network Analysis and Graph  
Algorithms** David Gleigh (*Purdue University, USA*)  
Tang Jie (*Tsinghua University, China*)  
Danai Koutra (*University of Michigan, USA*)

**Social Web** Lora Aroyo (*Google Research, USA*)  
Alessandro Bozzon (*University of Delft, The Netherlands*)  
Ioana Manolescu (*Inria and Institut Polytechnique de Paris, France*)

**Systems and Infrastructure** Ernesto Jimenez Ruiz (*City University of London, UK*)  
Fabrizio Silvestri (*Sapienza University of Rome, Italy*)

**User Modeling, Personalization and  
Accessibility** Nadia Fawaz (*Pinterest, USA*)  
Rui Li (*Pinterest, USA*)

**Web and Society** Pauline Leonard (*University of Southampton, UK*)  
Robert West (*EPFL, Switzerland*)

**Web Mining and Content Analysis** Brian D. Davison (*Lehigh University, USA*)  
Magdalini Eirinaki (*San Jose State University, USA*)  
Julian McAuley (*University of California San Diego, USA*)

**Web of Things, Ubiquitous and Mobile  
Computing** John Davies (*BT, UK*)  
Andreas Harth (*University Erlangen-Nuremberg, Germany*)

## Special Track Chairs

**Esports and Online Gaming** Casper Hartevelde (*Northeastern University, USA*)  
Magy Seif El-Nasr (*University of California at Santa Cruz, USA*)

**History of the Web** Dame Wendy Hall (*University of Southampton, UK*)  
Luc Mariaux (*École Centrale de Lyon, France*)

**Web for Good** Ricardo Baeza-Yates (*Northeastern University, USA*)  
Jeanna Matthews (*Clarkson University, USA*)  
Chiara Renso (*ISTI Institute of CNR, Italy*)

### Alternate Tracks Chairs

**Tutorials Co-Chairs** Senjuti Basu Roy (*New Jersey Institute of Technology, USA*)  
Riccardo Tommasini (*Politecnico di Milano, Italy*)

**Workshops Co-Chairs** Nathalie Hernandez (*University of Toulouse Jean Jaurès, France*)  
Preslav Nakov (*Qatar Computing Research Institute, Qatar*)

**PhD Symposium Co-Chairs** Elena Demidova (*University of Bonn, Germany*)  
Hala Skaf-Molli (*University of Nantes, France*)

**Posters and Demos Co-Chairs** Ana Lisa Gentile (*IBM Research Almaden, USA*)  
Pasquale Lisena (*EURECOM, France*)

**Industry Co-Chairs** Petra Selmer (*Neo4j, UK*)  
Joshua Shinavier (*LinkedIn, USA*)

**Web developer and W3C Co-Chairs** Dominique Hazael-Massieux (*W3C*)  
Thomas Steiner (*Google LLC, Germany*)

**Journal Chair** Ryen White (*Microsoft, USA*)

**Sponsorship Co-Chairs** Yi Chang (*Jilin University, China*)  
Liane Lewin-Eytan (*Amazon, Israel*)  
Juan Sequeda (*Data.world, USA*)

**Video Conference Chair** Yingtong Dou (*University of Illinois at Chicago, USA*)

**Volunteers Chair** Sébastien Forget (*TD Bank Group, Canada*)

### Local Organization Committee

**Local Organization Committee Co-Chairs** Laurent Flory (*Université Lyon 1, France*)  
Luc Mariaux (*École Centrale de Lyon, France*)

**Local Organization Committee Members** Pierre-Antoine Champin (*Inria / Université Lyon 1, France*)

#### W3C liaison

Coralie Grégoire (*INSAVALOR, France*) **Communication Chair**

Ivan Herman (*W3C / retired*) **Proceedings Co-Chair**

Yoelle Maarek (*Amazon, Israel*) **IW3C2 liaison**

Lionel Médini (*Université Lyon 1, France*) **Proceedings Co-Chair**



# Table of Contents

Foreword from the General Chairs .....	iii
The ACM Web Conference 2022 Organization Committee .....	v
WWW 2022 Companion Reviewers.....	xxiv
The ACM Web Conference 2022 Sponsors & Supporter .....	xxxv

## Alternate Track: Industry Track

• <b>WISE: Wavelet based Interpretable Stock Embedding for Risk-Averse Portfolio Management</b> .....	1
Mengying Zhu ( <i>Zhejiang University, China</i> ), Yan Wang ( <i>Macquarie University, Australia</i> ), Fei Wu, Mengyuan Yang, Cheng Chen, Qianqiao Liang, Xiaolin Zheng ( <i>Zhejiang University, China</i> )	
• <b>Cyclic Arbitrage in Decentralized Exchanges</b> .....	12
Ye Wang ( <i>ETH Zurich, Switzerland</i> ), Yan Chen ( <i>Zhejiang University, China</i> ), Haotian Wu ( <i>Xi'an Jiaotong University, China</i> ), Liyi Zhou ( <i>Imperial College London, UK</i> ), Shuiguang Deng ( <i>Zhejiang University, China</i> ), Roger Wattenhofer ( <i>ETH Zurich, Switzerland</i> )	
• <b>An Exploratory Study of Stock Price Movements from Earnings Calls</b> .....	20
Sourav Medya ( <i>Northwestern University, USA</i> ), Mohammad Rasoolinejad ( <i>Signature Consultants, USA</i> ), Yang Yang ( <i>Syracuse University, USA</i> ), Brian Uzzi ( <i>Northwestern University, USA</i> )	
• <b>DC-GNN: Decoupled Graph Neural Networks for Improving and Accelerating Large-Scale E-commerce Retrieval</b> .....	32
Chenchen Feng ( <i>Tsinghua University, China</i> ), Yu He, Shiyang Wen, Guojun Liu, Liang Wang, Jian Xu, Bo Zheng ( <i>Alibaba Group, China</i> )	
• <b>Multilingual Semantic Sourcing using Product Images for Cross-lingual Alignment</b> .....	41
Sourab Mangrulkar, Ankith M S, Vivek Sembium ( <i>Amazon, India</i> )	
• <b>Short, Colorful, and Irreverent! A Comparative Analysis of New Users on WallstreetBets During the Gamestop Short-squeeze</b> .....	52
Ehsan-Ul Haq, Tristan Braud ( <i>The Hong Kong University of Science and Technology, Hong Kong SAR</i> ), Lik-Hang Lee ( <i>Korea Advanced Institute of Science and Technology, South Korea</i> ), Anish K. Vallapuram, Yue Yu, Gareth Tyson, Pan Hui ( <i>The Hong Kong University of Science and Technology, Hong Kong SAR</i> )	
• <b>PEAR: Personalized Re-ranking with Contextualized Transformer for Recommendation</b> .....	62
Yi Li ( <i>Tsinghua University, China</i> ), Jieming Zhu, Weiwen Liu ( <i>Huawei Noah's Ark Lab, China</i> ), Liangcai Su ( <i>Tsinghua University, China</i> ), Guohao Cai, Qi Zhang, Ruiming Tang ( <i>Huawei Noah's Ark Lab, China</i> ), Xi Xiao ( <i>Tsinghua University, China</i> ), Xiuqiang He ( <i>Huawei Noah's Ark Lab, China</i> )	
• <b>FastClip: An Efficient Video Understanding System with Heterogeneous Computing and Coarse-to-fine Processing</b> .....	67
Liming Zhao, Siyang Sun, Yanhao Zhang, Yun Zheng, Pan Pan ( <i>Alibaba Group, China</i> )	
• <b>Modeling Position Bias Ranking for Streaming Media Services</b> .....	72
Matteo Ruffini, Vito Bellini, Alexander Buchholz, Giuseppe Di Benedetto, Yannik Stein ( <i>Amazon Music ML, Germany</i> )	
• <b>Fair Effect Attribution in Parallel Online Experiments</b> .....	77
Alexander Buchholz, Vito Bellini, Giuseppe Di Benedetto, Yannik Stein, Matteo Ruffini ( <i>Amazon Music ML, Germany</i> ), Fabian Moerchen ( <i>Amazon Music ML, USA</i> )	
• <b>On Reliability Scores for Knowledge Graphs</b> .....	84
Thomas Grubb ( <i>University of California, USA</i> ), Bill Andersen, Omar Alonso ( <i>Instacart, USA</i> )	

• <b>ROSE: Robust Caches for Amazon Product Search</b> .....	89
Chen Luo, Vihan Lakshman, Anshumali Shrivastava, Tianyu Cao, Sreyashi Nag, Rahul Goutam, Hanqing Lu, Yiwei Song, Bing Yin ( <i>Amazon Search, USA</i> )	
• <b>Semantic IR Fused Heterogeneous Graph Model in Tag-based Video Search</b> .....	94
Liming Gao ( <i>Sun Yat-sen University, China</i> ), Dongliang Liao, Gongfu Li ( <i>Tencent Inc., China</i> ), Jin Xu ( <i>South China University of Technology, China</i> ), Hankz Hankui Zhuo ( <i>Sun Yat-sen University, China</i> )	
• <b>Beyond NDCG: Behavioral Testing of Recommender Systems with RecList</b> .....	99
Patrick John Chia ( <i>Coveo, Canada</i> ), Jacopo Tagliabue ( <i>Coveo Labs, USA</i> ), Federico Bianchi ( <i>Bocconi University, Italy</i> ), Chloe He ( <i>Stanford University, USA</i> ), Brian Ko ( <i>KOSA AI, USA</i> )	
• <b>Privacy-Preserving Methods for Repeated Measures Designs</b> .....	105
Kevin Liou, Wenjing Zheng, Sathya Anand ( <i>Netflix, USA</i> )	
• <b>DCAF-BERT: A Distilled Cacheable Adaptable Factorized Model for Improved Ads CTR Prediction</b> .....	110
Aashiq Muhamed, Jaspreet Singh ( <i>Amazon, USA</i> ), Shuai Zheng ( <i>Amazon Web Services, USA</i> ), Iman Keivanloo, Sujan Perera, James Mracek, Yi Xu, Qingjun Cui, Santosh Rajagopalan, Belinda Zeng, Trishul Chilimbi ( <i>Amazon, USA</i> )	
• <b>A Multi-Task Learning Approach for Delayed Feedback Modeling</b> .....	116
Zhiqiang Huangfu, Gong-Duo Zhang, Zhengwei Wu, Qintong Wu, Zhiqiang Zhang, Lihong Gu, Jun Zhou, Jinjie Gu ( <i>Ant Group, China</i> )	
• <b>Search Filter Ranking with Language-Aware Label Embeddings</b> .....	121
Jacek Golebiowski ( <i>Amazon Research, Germany</i> ), Felice Antonio Merra ( <i>Politecnico di Bari, Italy</i> ), Ziawasch Abedjan ( <i>Leibniz University Hannover, Germany</i> ), Felix Biessmann ( <i>Berlin University of Applied Sciences, Germany</i> )	
• <b>Multi-task Ranking with User Behaviors for Text-video Search</b> .....	126
Peidong Liu ( <i>Tsinghua University, China and Tencent Inc.</i> ), Dongliang Liao ( <i>Tencent Inc.</i> ), Jinpeng Wang ( <i>Tsinghua University, China and Tencent Inc.</i> ), Yangxin Wu, Gongfu Li ( <i>Tencent Inc.</i> ), Shu-Tao Xia ( <i>Tsinghua University, China and Peng Cheng Laboratory</i> ), Jin Xu ( <i>South China University of Technology, China</i> )	
• <b>Deriving Customer Experience Implicitly from Social Media</b> .....	131
Aditya Kumar, Sneha Gupta, Ankit Sahu, Mayank Kant ( <i>Flipkart, India</i> )	
• <b>A Cluster-Based Nearest Neighbor Matching Algorithm for Enhanced A/A Validation in Online Experimentation</b> .....	136
Yan He, Lin Yu, Miao Chen, William Choi, Don Matheson ( <i>Yahoo Inc, USA</i> )	
• <b>Informative Integrity Frictions in Social Networks</b> .....	141
Lluís Garcia-Pueyo, Samantha Guthrie, Bernardo Santana Schwarz, Baoxuan Xu ( <i>Facebook, USA</i> )	
• <b>Personalized Complementary Product Recommendation</b> .....	146
An Yan ( <i>University of California San Diego, USA</i> ), Chaosheng Dong, Yan Gao, Jinmiao Fu, Tong Zhao, Yi Sun, Julian McAuley ( <i>Amazon, USA</i> )	
• <b>Spot Virtual Machine Eviction Prediction in Microsoft Cloud</b> .....	152
Fangkai Yang, Bowen Pang, Jue Zhang, Bo Qiao, Lu Wang ( <i>Microsoft Research</i> ), Camille Couturier ( <i>Microsoft 365</i> ), Chetan Bansal ( <i>Microsoft Research</i> ), Soumya Ram ( <i>Microsoft Azure</i> ), Si Qin ( <i>Microsoft Research</i> ), Zhen Ma ( <i>Microsoft 365</i> ), Íñigo Goiri ( <i>Microsoft Research</i> ), Eli Cortez ( <i>Microsoft Azure</i> ), Senthil Baladhandayutham, Victor Rühle, Saravan Rajmohan ( <i>Microsoft 365</i> ), Qingwei Lin, Dongmei Zhang ( <i>Microsoft Research</i> )	
• <b>Unsupervised Customer Segmentation with Knowledge Graph Embeddings</b> .....	157
Sumit Pai ( <i>Accenture Labs, Ireland</i> ), Fiona Brennan ( <i>Accenture, Ireland</i> ), Adrianna Janik ( <i>Accenture Labs, Ireland</i> ), Teutly Correia ( <i>Accenture, UK</i> ), Luca Costabello ( <i>Accenture Labs, Ireland</i> )	



## Track: Poster and Demo Track

- **From Discrimination to Generation: Knowledge Graph Completion with Generative Transformer**..... 162  
Xin Xie, Ningyu Zhang, Zhoubo Li, Shumin Deng (*Zhejiang University & AZFT Joint Lab for Knowledge Engine, China*), Hui Chen, Feiyu Xiong, Mosha Chen (*Alibaba Group, China*),  
HuaJun Chen (*Zhejiang University & AZFT Joint Lab for Knowledge Engine, China*)
- **Web Mining to Inform Locations of Charging Stations for Electric Vehicles**..... 166  
Philipp Hummler, Christof Naumzik (*ETH Zurich, Switzerland*), Stefan Feuerriegel (*LMU Munich, Germany*)
- **XAlign: Cross-lingual Fact-to-Text Alignment and Generation for Low-Resource Languages** ..... 171  
Tushar Abhishek, Shivprasad Sagare, Bhavyajeet Singh, Anubhav Sharma, Manish Gupta,  
Vasudeva Varma (*IIIT Hyderabad, India*)
- **Hypermedea: A Framework for Web (of Things) Agents** ..... 176  
Victor Charpenay, Antoine Zimmermann, Maxime Lefrançois, Olivier Boissier  
(*Univ Clermont Auvergne, France*)
- **GraphReform<sup>CD</sup>: Graph Reformulation for Effective Community Detection in Real-World Graphs**..... 180  
Jiwon Hong, Dong-hyuk Seo, Jeewon Ahn, Sang-Wook Kim (*Hanyang University, Republic of Korea*)
- **GraphZoo: A Development Toolkit for Graph Neural Networks with Hyperbolic Geometries**..... 184  
Anoushka Vyas, Nurendra Choudhary, Mehrdad Khatir, Chandan K. Reddy (*Virginia Tech, USA*)
- **QuatRE: Relation-Aware Quaternions for Knowledge Graph Embeddings** ..... 189  
Dai Quoc Nguyen (*Oracle Labs, Australia*), Thanh Vu (*AEHRC, CSIRO, Australia*), Tu Dinh Nguyen  
(*VinAI Research, Vietnam*), Dinh Phung (*Monash University, Australia*)
- **Universal Graph Transformer Self-Attention Networks**..... 193  
Dai Quoc Nguyen (*Oracle Labs, Australia*), Tu Dinh Nguyen (*VinAI Research, Vietnam*), Dinh Phung  
(*Monash University, Australia*)
- **A Two-stage User Intent Detection Model on Complicated Utterances with Multi-task Learning** ..... 197  
Shuangyong Song, Xiangyan Chen, Chao Wang, Xiaoguang Yu, Jia Wang, Xiaodong He  
(*JD AI Research, China*)
- **Know Your Victim: Tor Browser Setting Identification via Network Traffic Analysis**..... 201  
Chun-Ming Chang, Hsu-Chun Hsiao (*National Taiwan University, Taiwan*), Timothy Lynar (*University of New South Wales, Australia*), Tatsuya Mori (*Waseda University, Japan*)
- **HybEx: A Hybrid Tool for Template Extraction**..... 205  
Julián Alarte, Josep Silva (*Universitat Politècnica de València, Spain*)
- **Knowledge Distillation for Discourse Relation Analysis**..... 210  
Congcong Jiang, Tiejun Qian (*Wuhan University, China*), Bing Liu (*University of Illinois at Chicago, USA*)
- **User Donations in Online Social Game Streaming: The Case of Paid Subscription in Twitch.tv** ..... 215  
Jiajun Yu, Adele Lu Jia (*China Agricultural University, China*)
- **COCTEAU: An Empathy-Based Tool for Decision-Making**..... 219  
Andrea Mauri (*Delft University of Technology, Netherlands*), Andrea Tocchetti (*Politecnico di Milano, Italy*),  
Lorenzo Corti, Yen-Chia Hsu, Himanshu Verma (*Delft University of Technology, Netherlands*),  
Marco Brambilla (*Politecnico di Milano, Italy*)
- **Personal Attribute Prediction from Conversations** ..... 223  
Yinan Liu, Hu Chen, Wei Shen (*Nankai University, China*)

• <b>Multi-task GNN for Substitute Identification</b> .....	228
Tong Jian ( <i>Northeastern University, USA</i> ), Fan Yang, Zhen Zuo, Wenbo Wang, Michinari Momma, Tong Zhao, Chaosheng Dong, Yan Gao, Yi Sun ( <i>Amazon, USA</i> )	
• <b>Unsupervised Post-Time Fake Social Message Detection with Recommendation-aware Representation Learning</b> .....	232
Shao-Ping Hsiao, Yu-Che Tsai, Cheng-Te Li ( <i>National Cheng Kung University, Taiwan</i> )	
• <b>Cross-Language Learning for Product Matching</b> .....	236
Ralph Peeters, Christian Bizer ( <i>University of Mannheim, Germany</i> )	
• <b>A Graph Temporal Information Learning Framework for Popularity Prediction</b> .....	239
Caipiao Yang, Peng Bao, Rong Yan, Jianian Li ( <i>Beijing Jiaotong University, China</i> ), Xuanya Li ( <i>Baidu Inc., China</i> )	
• <b>PREP: Pre-training with Temporal Elapse Inference for Popularity Prediction</b> .....	243
Qi Cao ( <i>Chinese Academy of Sciences, China</i> ), Huawei Shen, Yuanhao Liu ( <i>Chinese Academy of Sciences, China</i> and <i>University of Chinese Academy of Sciences, China</i> ), Jinhua Gao ( <i>Chinese Academy of Sciences, China</i> ), Xueqi Cheng ( <i>Chinese Academy of Sciences, China</i> and <i>University of Chinese Academy of Sciences, China</i> )	
• <b>Supervised Contrastive Learning for Product Matching</b> .....	248
Ralph Peeters, Christian Bizer ( <i>University of Mannheim, Germany</i> )	
• <b>QAnswER: Towards Question Answering Search over Websites</b> .....	252
Kunpeng Guo ( <i>Laboratoire Hubert Curien, France</i> ), Clement Defretiere ( <i>Université Jean Monnet, France</i> ), Dennis Diefenbach ( <i>Laboratoire Hubert Curien, France</i> ), Christophe Gravier, Antoine Gourru ( <i>Laboratoire</i> <i>Hubert Curien, France</i> )	
• <b>Scriptoria: A Crowd-powered Music Transcription System</b> .....	256
Ioannis Petros Samiotis, Christoph Lofi, Shaad Alaka, Cynthia C.S. Liem, Alessandro Bozzon ( <i>Delft University of Technology, Netherlands</i> )	
• <b>SHACL and ShEx in the Wild: A Community Survey on Validating Shapes Generation and Adoption</b> .....	260
Kashif Rabbani, Matteo Lissandrini, Katja Hose ( <i>Aalborg University, Denmark</i> )	
• <b>Towards Knowledge-Driven Symptom Monitoring &amp; Trigger Detection of Primary Headache Disorders</b> .....	264
Mathias De Brouwer ( <i>Ghent University - imec, Belgium</i> ), Nicolas Vandebussche ( <i>Ghent University Hospital,</i> <i>Belgium</i> ), Bram Steenwinckel, Marija Stojchevska, Jonas Van Der Donckt, Vic Degraeve, Filip De Turck ( <i>Ghent University - imec, Belgium</i> ), Koen Paemeleire ( <i>Ghent University Hospital, Belgium</i> ), Sofie Van Hoecke, Femke Ongenae ( <i>Ghent University - imec, Belgium</i> )	
• <b>Using Schema.org and Solid for Linked Data-based Machine-to-Machine Sales Contract Conclusion</b> .....	269
Xinni Wang, Christoph H.-J. Braun ( <i>Karlsruhe Institute of Technology (KIT), Germany</i> ), Andreas Both ( <i>DATEV eG, Germany</i> ), Tobias Käfer ( <i>Karlsruhe Institute of Technology (KIT), Germany</i> )	
• <b>VisPaD: Visualization and Pattern Discovery for Fighting Human Trafficking</b> .....	273
Pratheeksha Nair, Yifei Li ( <i>McGill University &amp; Mila, Canada</i> ), Catalina Vajiac ( <i>Carnegie Mellon University,</i> <i>USA</i> ), Andreas Olligschlaeger ( <i>Marinus Analytics, USA</i> ), Meng-Chieh Lee, Namyong Park ( <i>Carnegie Mellon</i> <i>University, USA</i> ), Duen Horng Chau ( <i>Georgia Institute of Technology, USA</i> ), Christos Faloutsos ( <i>Carnegie</i> <i>Mellon University, USA</i> ), Reihaneh Rabbany ( <i>McGill University &amp; Mila, Canada</i> )	
• <b>ECCE: Entity-centric Corpus Exploration Using Contextual Implicit Networks</b> .....	278
Julian Schelb ( <i>University of Konstanz, Germany</i> ), Maud Ehrmann ( <i>École Polytechnique Fédérale de Lausanne,</i> <i>Switzerland</i> ), Matteo Romanello ( <i>University of Lausanne, Switzerland</i> ), Andreas Spitz ( <i>University of Konstanz,</i> <i>Germany</i> )	
• <b>Towards Preserving Server-Side Privacy of On-Device Models</b> .....	282
Akanksha Atrey ( <i>University of Massachusetts, USA</i> ), Ritwik Sinha, Somdeb Sarkhel, Saayan Mitra, David Arbour ( <i>Adobe Research, USA</i> ), Akash V. Maharaj ( <i>Adobe Inc., USA</i> ), Prashant Shenoy ( <i>University</i> <i>of Massachusetts, USA</i> )	

• <b>Demo: PhishChain: A Decentralized and Transparent System to Blacklist Phishing URLs .....</b>	286
Shehan Vidyakeerthi ( <i>University of Colombo School of Computing, Sri Lanka</i> ), Mohamed Nabeel ( <i>Qatar Computing Research Institute, Qatar</i> ), Charith Elvitigala, Chamath Keppitiyagama ( <i>University of Colombo School of Computing, Sri Lanka</i> )	
• <b>Technology Growth Ranking Using Temporal Graph Representation Learning.....</b>	290
Daniel Cummings, Ashrita Brahmaroutu, Marcel Nassar, Nesreen K. Ahmed ( <i>Intel Corporation, USA</i> )	
• <b>Linking Streets in OpenStreetMap to Persons in Wikidata .....</b>	294
Daria Gurtovoy ( <i>Universität Bonn, Germany</i> ), Simon Gottschalk ( <i>Leibniz Universität Hannover, Germany</i> )	
• <b>Effectiveness of Data Augmentation to Identify Relevant Reviews for Product Question Answering .....</b>	298
Kalyani Roy, Avani Goel, Pawan Goyal ( <i>Indian Institute of Technology Kharagpur, India</i> )	
• <b>Does Evidence from Peers Help Crowd Workers in Assessing Truthfulness?.....</b>	302
Jiechen Xu, Lei Han, Shaoyang Fan, Shazia Sadiq, Gianluca Demartini ( <i>The University of Queensland, Australia</i> )	
• <b>Graph-level Semantic Matching model for Knowledge base Aggregate Question Answering.....</b>	307
Ya Liu, Shaojuan Wu, Jiarui Zhang, Linyi Han, Xiaowang Zhang, Yongxin Yu, Zhiyong Feng ( <i>Tianjin University, China</i> )	

## Alternate Track: PhD Symposium

• <b>Query-Driven Graph Processing .....</b>	311
Angela Bonifati ( <i>Lyon 1 University, France</i> )	
• <b>Interactions in Information Spread .....</b>	313
Gaël Poux-Médard ( <i>Université de Lyon, France</i> )	
• <b>Canonicalisation of SPARQL 1.1 Queries .....</b>	318
Jaime Salas ( <i>Universidad de Chile, Chile</i> )	
• <b>Towards Automated Technologies in the Referencing Quality of Wikidata .....</b>	324
Seyed Amir Hosseini Beghaeiraveri ( <i>Heriot-Watt University, UK</i> )	
• <b>User Access Models to Event-Centric Information .....</b>	329
Sara Abdollahi ( <i>Leibniz Universität Hannover, Germany</i> )	
• <b>Geometric and Topological Inference for Deep Representations of Complex Networks .....</b>	334
Baihan Lin ( <i>Columbia University, USA</i> )	
• <b>Predicting SPARQL Query Dynamics.....</b>	339
Alberto Moya Loustaunau ( <i>University of Chile, Chile</i> )	
• <b>Personal Knowledge Graphs: Use Cases in e-learning Platforms .....</b>	344
Eleni Ilkou ( <i>Leibniz University Hannover, Germany</i> )	
• <b>Enhancing Multilingual Accessibility of Question Answering over Knowledge Graphs.....</b>	349
Aleksandr Perevalov ( <i>Anhalt University of Applied Sciences, Germany</i> )	
• <b>Enhancing Query Answer Completeness with Query Expansion based on Synonym Predicates.....</b>	354
Emetis Niazmand ( <i>Leibniz Universität Hannover, Germany</i> )	
• <b>Comprehensive Event Representations using Event Knowledge Graphs and Natural Language Processing .....</b>	359
Tin Kuculo ( <i>Leibniz Universität Hannover, Germany</i> )	

## Alternate Track: Web Developer and W3C

- **Web Audio Modules 2.0: An Open Web Audio Plugin Standard** ..... 364  
Michel Buffa (*University Côte d'Azur, France*), Shihong Ren (*Université de Saint-Etienne, France*),  
Owen Campbell (*USA*), Tom Burns (*Canada*), Steven Yi (*USA*), Jari Kleimola, Oliver Larkin (*Finland, UK*)
- **Walks in Cyberspace: Improving Web Browsing and Network Activity Analysis with 3D Live Graph Rendering** ..... 370  
Lionel Tailhardat (*Orange, France*), Raphaël Troncy (*EURECOM, France*), Yoan Chabot (*Orange, France*),
- **JSREHAB: Weaning Common Web Interface Components from JavaScript Addiction** ..... 376  
Romain Fouquet (*Inria / Univ. Lille, France*), Pierre Laperdrix (*CNRS / Univ. Lille / Inria, France*),  
Romain Rouvoy (*Univ. Lille / Inria, France*)
- **With One Voice: Composing a Travel Voice Assistant from Repurposed Models** ..... 383  
Shachaf Poran, Gil Amsalem, Amit Beka, Dmitri Goldenberg (*Booking.com, Israel*)

## Alternate Track: Journal Track

- **Exploiting Anomalous Structural Nodes in Dynamic Social Networks** ..... 388  
Huan Wang (*Huazhong Agricultural University*), Chunming Qiao (*University at Buffalo, The State University of New York*), Xuan Guo (*University of North Texas*), Ying Sha (*Huazhong Agricultural University*), Zhiguo Gong (*University of Macau*)
- **Cross-Site Prediction on Social Influence for Cold-Start Users in Online Social Networks** ..... 389  
Qingyuan Gong, Yang Chen, Xinlei He (*Fudan University, China*), Yu Xiao (*Aalto University, Finland*),  
Pan Hui (*University of Helsinki, Finland and Hong Kong University of Science and Technology, Hong Kong*),  
Xin Wang (*Fudan University, China*), Xiaoming Fu (*University of Göttingen, Germany*)
- **Who Has the Last Word? Understanding How to Sample Online Discussions** ..... 390  
Gioia Boschi, Anthony P. Young, Sagar Joglekar (*King's College London, UK*), Chiara Cammarota (*Sapienza Università di Roma, Italy and King's College London, UK*), Nishanth Sastry (*University of Surrey, UK and King's College London, UK*)

## Alternate Track: Tutorials

- **Accepted Tutorials at the Web Conference 2022** ..... 391  
Riccardo Tommasini (*INSA de Lyon - CRNS LIRIS, France*), Senjuti Basu Roy (*New Jersey Institute of Technology, USA*), Xuan Wang, Hongwei Wang, Heng Ji, Jiawei Han (*University of Illinois at Urbana-Champaign, USA*), Preslav Nakov (*HBKU, Qatar*), Giovanni Da San Martino (*University of Padova, Italy*), Firoj Alam (*Qatar Computing Research Institute, Qatar*), Markus Schedl (*Johannes Kepler University, Austria*), Elisabeth Lex (*Graz University of Technology, Austria*), Akash Bharadwaj, Graham Cormode (*Meta AI, USA*), Milan Dojchinovski (*DBpedia Association, Germany & Czech Technical University in Prague, Czech Republic*), Jan Forberg, Johannes Frey (*DBpedia Association, Germany*), Pieter Bonte (*Ghent University, Belgium*), Marco Balduini, Matteo Belcao (*Quantia Consulting, Italy*), Emanuele Della Valle (*Politecnico di Milano, Italy*), Junliang Yu, Hongzhi Yin, Tong Chen (*The University of Queensland, Australia*), Haochen Liu, Yiqi Wang (*Michigan State University, USA*), Wenqi Fan (*The Hong Kong Polytechnic University, Hong Kong*), Xiaorui Liu, Jamell Dacon (*Michigan State University, USA*), Lingjuan Lye (*Sony AI, Japan*), Jiliang Tang (*Michigan State University, USA*), Aristides Gionis, Stefan Neumann (*KTH Royal Institute of Technology, Sweden*), Bruno Ordozgoiti (*Queen Mary University of London, UK*), Simon Razniewski, Hiba Arnaout, Shrestha Ghosh (*Max Planck Institute of Informatics, Germany*), Fabian Suchanek (*Institut Polytechnique de Paris, France*), Lingfei Wu (*JD.COM Silicon Valley Research Center, USA*), Yu Chen (*Meta AI, USA*), Yunyao Li (*IBM Research AI, USA*), Bang Liu (*University of Montreal, Canada*), Filip Ilievski (*University of Southern California, USA*), Daniel Garijo (*Universidad Politécnica de Madrid, Spain*), Hans Chalupsky, Pedro Szekely (*University of Southern California, USA*), Ilias Kanellos (*Athena Research Center, Greece*), Dimitris Sacharidis (*Université Libre de Bruxelles, Belgium*), Thanasis Vergoulis (*Athena Research Center, Greece*), Nurendra Choudhary (*Virginia Tech, USA*), Nikhil Rao, Karthik Subbian, Srinivasan Sengamedu (*Amazon, USA*), Chandan K. Reddy (*Virginia Tech, USA*), Friedhelm Victor (*Technical University of Berlin, Germany*), Bernhard Haslhofer (*AIT - Austrian Institute of Technology, Austria*), George Katsogiannis-Meimarakis, Georgia Koutrika (*Athena Research Center,*

Greece), Shengmin Jin (*Syracuse University, USA*), Danai Koutra (*University of Michigan, USA*), Reza Zafarani (*Syracuse University, USA*), Yulia Tsvetkov (*University of Washington, USA*), Vidhisha Balachandran, Sachin Kumar (*Carnegie Mellon University, USA*), Xiangyu Zhao (*City University of Hong Kong, Hong Kong*), Bo Chen, Huifeng Guo (*Huawei Noah's Ark Lab, Hong Kong*), Yejing Wang (*University of Science and Technology of China, China*), Ruiming Tang (*Huawei Noah's Ark Lab, Hong Kong*), Yang Zhang (*University of Science and Technology of China, China*), Wenjie Wang (*National University of Singapore, Singapore*), Peng Wu (*Peking University, China*), Fuli Feng, Xiangnan He (*University of Science and Technology of China, China*)

## Workshop: DataLit – 3rd International Workshop on Data Literacy

- **3rd Data Literacy Workshop** ..... 400  
Manuel León-Urrutia (*University of Southampton, UK*), Davide Taibi (*National Research Council of Italy, Italy*)
- **Culturally Responsive Data Literacy: An Emerging and Important Construct**..... 402  
Ellen B. Mandinach (*WestEd, USA*)
- **One Year of DALIDA Data Literacy Workshops for Adults: A Report** ..... 403  
Christophe Debruyne (*University of Liege, Belgium*), Laura Grehan (*Dublin City University, Ireland*), Mairéad Hurley (*Trinity College Dublin, Ireland*), Anne Kearns (*Dublin City University, Ireland*), Ciaran O'Neill (*Trinity College Dublin, Ireland*)
- **Towards Benchmarking Data Literacy** ..... 408  
Calum Inverarity, David Tarrant, Emilie Forrest (*Open Data Institute, UK*), Phil Greenwood (*Glacis, UK*)
- **Towards Digital Economy through Data Literate Workforce**..... 417  
Maja Miloradov, Slavko Rakic, Ugljesa Marjanovic (*University of Novi Sad, Serbia*)

## Workshop: BeyondFacts – 2nd International Workshop on Knowledge Graphs for Online Discourse Analysis

- **BeyondFacts'22: 2nd International Workshop on Knowledge Graphs for Online Discourse Analysis** ..... 423  
Konstantin Todorov (*University of Montpellier, France*), Pavlos Fafalios (*Information Systems Laboratory, Greece*), Stefan Dietze (*GESIS & Heinrich-Heine-University Düsseldorf, Germany*)
- **Accurate and Explainable Misinformation Detection: Too Good to be True?**..... 426  
Jose Manuel Gomez-Perez (*Expert.ai, Spain*)
- **Have You been Misinformed?: Computational Tools and Analysis of Our Interactions with False and Corrective Information** ..... 428  
Harith Alani (*The Open University, UK*)
- **Incorporating External Knowledge for Evidence-based Fact Verification**..... 429  
Anab Maulana Barik, Wynne Hsu (*National University of Singapore, Singapore*), Mong Li Lee (*NUS Centre for Trusted Internet and Community, Singapore*)
- **Geotagging TweetsCOVID19: Enriching a COVID-19 Twitter Discourse Knowledge Base with Geographic Information** ..... 438  
Dimitar Dimitrov (*GESIS - Leibniz Institute for the Social Sciences, Germany*), Dennis Segeth (*Heinrich Heine University, Germany*), Stefan Dietze (*GESIS - Leibniz Institute for the Social Sciences, Germany and Heinrich Heine University, Germany*)
- **Towards Building Live Open Scientific Knowledge Graphs**..... 443  
Anh Le-Tuan (*Technical University Berlin, Germany*), Carlos Franzreb (*Weizenbaum Institute, Germany and Fraunhofer Institute for Open Communication Systems, Germany*), Danh Le Phuoc (*Technical University Berlin, Germany and Fraunhofer Institute for Open Communication Systems, Germany*), Sonja Schimmler (*Weizenbaum Institute, Germany and Fraunhofer Institute for Open Communication Systems, Germany*), Manfred Hauswirth (*Technical University Berlin, Germany and Weizenbaum Institute, Germany*)



- **Towards Analyzing the Bias of News Recommender Systems Using Sentiment and Stance Detection** ..... 448  
Mehwish Alam (*Karlsruhe Institute of Technology, Germany*), Andreea Iana (*University of Mannheim, Germany*), Alexander Grote (*Karlsruhe Institute of Technology, Germany*), Katharina Ludwig, Philipp Müller, Heiko Paulheim (*University of Mannheim, Germany*)
- **Methodology to Compare Twitter Reaction Trends between Disinformation Communities, to COVID related Campaign Events at Different Geospatial Granularities** ... 458  
Debraj De, Gautam Thakur, Drahomira Herrmannova, Carter Christopher (*Oak Ridge National Laboratory (ORNL), USA*)

## **Workshop: CAAW – International Workshop on Cryptoasset Analytics Workshop**

- **CAAW’22: 2022 International Workshop on Cryptoasset Analytics** ..... 464  
Friedhelm Victor (*Technische Universität Berlin, Germany*), Bernhard Haslhofer (*Complexity Science Hub Vienna, Austria*), Sourav Sen Gupta (*Nanyang Technological University, Singapore*), Malte Möser (*Chainalysis, USA*), Rainer Böhme (*University of Innsbruck, Austria*)
- **How Much is the Fork? Fast Probability and Profitability Calculation during Temporary Forks** ..... 467  
Aljosha Judmayer, Nicholas Stifter, Philipp Schindler (*SBA Research, Austria and University of Vienna, Austria*), Edgar Weippl (*University of Vienna, Austria*)
- **Analysis of Arbitrary Content on Blockchain-Based Systems using BigQuery** ..... 478  
Marcel Gregoriadis (*Humboldt University of Berlin, Germany*), Robert Muth (*Technical University of Berlin, Germany*), Martin Florian (*Weizenbaum Institute, Germany*)
- **Characterizing the OpenSea NFT Marketplace** ..... 488  
Bryan White, Aniket Mahanti (*University of Auckland, New Zealand*), Kalpdrum Passi (*Laurentian University, Canada*)

## **Workshop: CLEOPATRA – 3rd International Workshop on Cross-lingual Event-centric Open Analytics**

- **CLEOPATRA’22: 3rd International Workshop on Cross-lingual Event-centric Open Analytics** ..... 497  
Elena Demidova (*University of Bonn, Germany*), Sherzod Hakimov (*TIB - Leibniz Information Centre for Science and Technology, Germany*), Jane Winters (*University of London, UK*), Marko Tadić (*University of Zagreb, Croatia*)
- **Geospatial Interlinking with JedAI-spatial** ..... 499  
Manolis Koubarakis (*National and Kapodistrian University of Athens, Greece*)
- **Exploring Cross-Lingual Transfer to Counteract Data Scarcity for Causality Detection** ..... 501  
Sebastian Reimann (*Uppsala University, Sweden and Ruhr-Universität Bochum, Germany*), Sara Stymne (*Uppsala University, Sweden*)

## **Workshop: COnSeNT – 2nd International Workshop on Consent Management in Online Services, Networks and Things**

- **COnSeNT 2022: 2nd International Workshop on Consent Management in Online Services, Networks and Things** ..... 509  
Paulina J. Pesch (*Karlsruhe Institute of Technology, Germany*), Harshvardhan J. Pandit (*Trinity College Dublin, Ireland*), Vitor Jesus (*Aston University, UK*), Cristiana Santos (*Utrecht University, Netherlands*)
- **Consent of the Governed** ..... 514  
Robin Berjon (*The New York Times, USA*)
- **A Policy-Oriented Architecture for Enforcing Consent in Solid** ..... 516  
Laurens Debackere, Pieter Colpaert, Ruben Taelman, Ruben Verborgh (*Ghent University - imec, Belgium*)



- **Internalization of Privacy Externalities through Negotiation: Social Costs of Third-party Web-analytic Tools and The Limits of the Legal Data Protection Framework**..... 525  
Nils Wehkamp (*Digital Society Institute of European School of Management and Technology, Germany*)
- **Context, Prioritization, and Unexpectedness: Factors Influencing User Attitudes About Infographic and Comic Consent** ..... 534  
Xengie Doan (*Université du Luxembourg, Luxembourg*), Annika Selzer (*Fraunhofer Institute for Secure Information Technology, Germany*), Arianna Rossi, Wilhelmina Maria Botes, Gabriele Lenzini (*Université du Luxembourg, Luxembourg*)

## Workshop: EMDC – 2nd International Workshop on the Efficiency of Modern Datacenters

- **Second International Workshop on the Efficiency of Modern Data Centers (EMDC) Chairs’ Welcome**..... 546  
Eyhab Al-Masri (*University of Washington, USA*), Di Wang (*Microsoft, USA*)
- **Optimizing Data Layout for Training Deep Neural Networks** ..... 548  
Bingyao Li, Qi Xue (*University of Pittsburgh, USA*), Geng Yuan (*Northeastern University, USA*), Sheng Li (*University of Pittsburgh, USA*), Xiaolong Ma, Yanzhi Wang (*Northeastern University, USA*), Xulong Tang (*University of Pittsburgh, USA*)
- **Security Challenges for Modern Data Centers with IoT: A Preliminary Study**..... 555  
Zhen Zeng, Chun-Jen Chung, Liguang Xie (*Futurewei Technologies, USA*)
- **Streaming Analytics with Adaptive Near-data Processing**..... 563  
Atul Sandur (*University of Illinois at Urbana-Champaign, USA*), ChanHo Park (*UNIST, South Korea*), Stavros Volos (*Microsoft Research, UK*), Gul Agha (*University of Illinois at Urbana-Champaign, USA*), Myeongjae Jeon (*UNIST, South Korea*)
- **Powering Multi-Task Federated Learning with Competitive GPU Resource Sharing**..... 567  
Yongbo Yu, Fuxun Yu, Zirui Xu (*George Mason University, USA*), Di Wang, Minjia Zhang (*Microsoft, USA*), Ang Li (*Duke University, USA*), Shawn Bray, Chenchen Liu (*University of Maryland, USA*), Xiang Chen (*George Mason University, USA*)

## Workshop: FinWeb – 2nd International Workshop on Financial Technology on the Web

- **An Overview of Financial Technology Innovation**..... 572  
Chung-Chi Chen (*National Institute of Advanced Industrial Science and Technology, Japan*), Hen-Hsen Huang (*Academia Sinica, Taiwan*), Hiroya Takamura (*National Institute of Advanced Industrial Science and Technology, Japan*), Hsin-Hsi Chen (*National Taiwan University, Taiwan*)
- **A Generative Approach for Financial Causality Extraction**..... 576  
Tapas Nayak, Soumya Sharma, Yash Butala (*IIT Kharagpur, India*), Koustuv Dasgupta (*Goldman Sachs, India*), Pawan Goyal, Niloy Ganguly (*IIT Kharagpur, India*)
- **Rayleigh Portfolios and Penalised Matrix Decomposition** ..... 579  
Francois Buet-Golfouse (*University College London, UK*), Hans Roggeman, Islam Utyagulov
- **FiNCAT: Financial Numeral Claim Analysis Tool**..... 583  
Sohom Ghosh, Sudip Kumar Naskar (*Jadavpur University, India*)
- **Understanding Financial Information Seeking Behavior from User Interactions with Company Filings** ..... 586  
Ben Trovato, G.K.M. Tobin (*Institute for Clarity in Documentation Dublin, USA*), Lars Thørväld (*The Thørväld Group Hekla, Iceland*), Valerie Béranger (*Inria Paris-Rocquencourt*)
- **FinRED: A Dataset for Relation Extraction in Financial Domain** ..... 595  
Soumya Sharma, Tapas Nayak, Arusarka Bose, Ajay Kumar Meena (*IIT Kharagpur, India*), Koustuv Dasgupta (*Goldman Sachs, India*), Niloy Ganguly (*IIT Kharagpur, India and LU Hannover, Germany*), Pawan Goyal (*IIT Kharagpur, India*)

- **SEBI Regulation Biography** ..... 598  
Sathvik Sanjeev Buggana, Deepti Saravanan, Shravya Kanchi, Ujwal Narayan, Shivam Mangale,  
Lini T. Thomas, Kamalakar Karlapalem (*IIIT Hyderabad, India*), Natraj Raman (*J.P. Morgan AI Research, UK*)
- **Numeral Tense Detection in Chinese Financial News** ..... 604  
Yuxuan Liu, Maofu Liu, Mengjie Wu (*Wuhan University of Science and Technology, China*)
- **Detecting Regulation Violations for an Indian Regulatory Body through Multi Label Classification** ..... 610  
Ujwal Narayan, Pulkit Parikh, Kamalakar Karlapalem (*International Institute of Information Technology, Hyderabad, India*), Natraj Raman (*J.P. Morgan AI Research, UK*)
- **Improving Operation Efficiency through Predicting Credit Card Application Turnaround Time with Index-based Encoding** ..... 615  
Jing Xiang Toh, Kay Jan Wong, Samarth Agarwal, Xuejie Zhang, John Jianan Lu (*DBS Bank, Singapore*)
- **TweetBoost: Influence of Social Media on NFT Valuation** ..... 621  
Arnav Kapoor, Dipanwita Guhathakurta (*IIIT Hyderabad, India*), Mehul Mathur, Rupanshu Yadav (*IIIT Hyderabad and IIIT Delhi, India*), Manish Gupta, Ponnurangam Kumaraguru (*IIIT Hyderabad, India*)
- **Graph Representation Learning of Banking Transaction Network with Edge Weight-Enhanced Attention and Textual Information** ..... 630  
Naoto Minakawa, Kiyoshi Izumi, Hiroki Sakaji, Hitomi Sano (*The University of Tokyo, Japan*)

## Workshop: LocWeb – 12th International Workshop on Location and the Web

- **LocWeb2022: 12th International Workshop on Location and the Web** ..... 638  
Dirk Ahlers (*NTNU - Norwegian University of Science and Technology, Norway*), Erik Wilde (*Axway, Switzerland*)
- **Anonymous Hyperlocal Communities: What Do They Talk About?** ..... 639  
Jens Helge Reelfs, Oliver Hohlfeld (*Brandenburg University of Technology, Germany*), Niklas Henckell (*The Jodel Venture GmbH, Germany*)
- **Exploiting Geodata to Improve Image Recognition with Deep Learning** ..... 648  
Christian Arbing, Martin Bullin, Andreas Henrich (*University of Bamberg, Germany*)
- **Predicting Spatial Spread on Social Media** ..... 656  
Rimjhim, Sourav Dandapat (*Indian Institute of Technology Patna, India*)

## Workshop: MAISoN – 8th International Workshop on Mining Actionable Insights from Social Networks, Special Edition on Mental Health and Social Media

- **MAISoN'22: 8th International Workshop on Mining Actionable Insights from Social Networks Special Edition on Mental Health and Social Media** ..... 660  
Ebrahim Bagheri (*Ryerson University, Canada*), Diana Inkpen (*University of Ottawa, Canada*), Christopher C. Yang (*Drexel University, USA*), Fattane Zarrinkalam (*University of Guelph, Canada*)
- **Utilizing Pattern Mining and Classification Algorithms to Identify Risk for Anxiety and Depression in the LGBTQ+ Community During the COVID-19 Pandemic** ..... 663  
Josephine Bierbaum, Melissa Lynn, Louis Yu (*Gustavus Adolphus College, USA*)
- **Supporting People Receiving Substance Use Treatment during COVID-19 through a Professional-Moderated Online Peer Support Group** ..... 673  
Ou Stella Liang, Christopher C. Yang, Anne Glenney, Richard E. Pointer, Sharlene Irving, Barbara Schindler, David Bennett (*Drexel University, USA*)
- **A Large-scale Temporal Analysis of User Lifespan Durability on the Reddit Social Media Platform** ..... 677  
Amirhossein Nadiri (*Sharif University of Technology, Iran*), Frank W. Takes (*Leiden University, Netherlands*)

- **“I’m Always in so Much Pain and no one will Understand” – Detecting Patterns in Suicidal Ideation on Reddit** ..... 686  
Michal Monselise, Christopher C. Yang (*Drexel University, USA*)

## **Workshop: MUWS’2022 – 1st International Workshop on Multimodal Understanding for the Web and Social Media**

- **MUWS’22: 1st International Workshop on Multimodal Understanding for the Web and Social Media** ..... 692  
Sherzod Hakimov, Gullal Singh Cheema (*TIB - Leibniz Information Centre for Science and Technology, Germany*), Marc A. Kastner (*National Institute of Informatics, Japan*), Rajiv Ratn Shah (*Indraprastha Institute of Information Technology, Delhi, India*), Karan Sikka (*SRI International, USA*)
- **Visual Persuasion in COVID-19 Social Media Content: A Multi-Modal Characterization**..... 694  
Mesut Erhan Unal, Adriana Kovashka, Wen-Ting Chung, Yu-Ru Lin (*University of Pittsburgh, USA*)
- **Improving and Diagnosing Knowledge-Based Visual Question Answering via Entity Enhanced Knowledge Injection**..... 705  
Diego Garcia-Olano, Yasumasa Onoe, Joydeep Ghosh (*University of Texas at Austin, USA*)
- **ViNTER: Image Narrative Generation with Emotion-Arc-Aware Transformer** ..... 716  
Kohei Uehara, Yusuke Mori (*The University of Tokyo, Japan*), Yusuke Mukuta, Tatsuya Harada (*The University of Tokyo / RIKEN, Japan*)
- **Leveraging Intra and Inter Modality Relationship for Multimodal Fake News Detection** .... 726  
Shivangi Singhal, Tanisha Pandey, Saksham Mrig, Rajiv Ratn Shah (*IIIT Delhi, India*), Ponnurangam Kumaraguru (*IIIT Hyderabad, India*)

## **Workshop: Sci-K – 2nd International Workshop on Scientific Knowledge Representation, Discovery, and Assessment**

- **Sci-K 2022 – International Workshop on Scientific Knowledge: Representation, Discovery, and Assessment** ..... 735  
Paolo Manghi, Andrea Mannocci (*ISTI-CNR, Italy*), Francesco Osborne (*The Open University, UK*), Dimitris Sacharidis (*Université Libre de Bruxelles, Belgium*), Angelo Salatino (*The Open University, UK*), Thanasis Vergoulis (*“Athena” RC, Greece*)
- **The Semantic Scholar Academic Graph (S2AG)** ..... 739  
Alex D. Wade (*Allen Institute for Artificial Intelligence, USA*)
- **Data Models for Annotating Biomedical Scholarly Publications: The Case of COVID-19** ..... 740  
Houcemeddine Turki, Mohamed Ali Hadj Taieb (*University of Sfax, Tunisia*), Alejandro Piad-Morffis (*University of Havana, Cuba*), Mohamed Ben Aouicha (*University of Sfax, Tunisia*), René Fabrice Bile (*University of Maroua, Cameroon*)
- **Sequence-Based Extractive Summarisation for Scientific Articles** ..... 751  
Daniel Kershaw, Rob Koeling (*Elsevier Ltd, UK*)
- **Assessing Network Representations for Identifying Interdisciplinarity** ..... 758  
Eoghan Cunningham, Derek Greene (*University College Dublin, Ireland*)
- **Personal Research Knowledge Graphs** ..... 763  
Prantika Chakraborty (*Indian Association for the Cultivation of Science, India*), Sudakshina Dutta (*Indian Institute of Technology Goa, India*), Debarshi Kumar Sanyal (*Indian Association for the Cultivation of Science, India*)
- **Quantifying the Topic Disparity of Scientific Articles** ..... 769  
Munjung Kim, Jisung Yoon, Woo-Sung Jung (*Pohang University of Science and Technology, Republic of Korea*), Hyunuk Kim (*Boston University, USA*)
- **Beyond Reproduction, Experiments Want to be Understood** ..... 774  
Jérôme Euzénat (*Grenoble INP, LIG, France*)

- **GraphCite: Citation Intent Classification in Scientific Publications via Graph Embeddings**..... 779  
 Dan Berrebbi, Nicolas Huynh (*Ecole Polytechnique, Institut Polytechnique de Paris, France*),  
 Oana Balalau (*Inria, Institut Polytechnique de Paris, France*)
- **A Study of Computational Reproducibility using URLs Linking to Open Access Datasets and Software** ..... 784  
 Lamia Salsabil, Jian Wu, Muntabir Hasan Choudhury (*Old Dominion University, USA*), William A. Ingram,  
 Edward A. Fox (*Virginia Polytechnic Institute and State University, USA*), Sarah M. Rajtmajer,  
 C. Lee Giles (*Pennsylvania State University, USA*)
- **Semi-automated Literature Review for Scientific Assessment of Socioeconomic Climate Change Scenarios**..... 789  
 Vanessa Jine Schweizer (*University of Waterloo, Canada*), Jude Herijadi Kurniawan (*University of Waterloo, Canada and Institute for Advanced Sustainability Studies e.V., Germany*), Aidan Power  
 (*University of Waterloo, Canada*)
- **SciNoBo: A Hierarchical Multi-Label Classifier of Scientific Publications** ..... 800  
 Nikolaos Gialitsis, Sotiris Kotitsas, Haris Papageorgiou (*Athena Research and Innovation Center, Greece*)
- **Examining the ORKG towards Representation of Control Theoretic Knowledge – Preliminary Experiences and Conclusions** ..... 810  
 Carsten Knoll (*TU Dresden, Germany*)

## Workshop: SeBiLAn – International Workshop on Semantics-enabled Biomedical Literature Analytics

- **The International Workshop on Semantics-enabled Biomedical Literature Analytics (SeBiLAn)** ..... 818  
 Faezeh Ensan (*Ryerson University, Canada*), Halil Kilicoglu (*University of Illinois Urbana-Champaign, USA*),  
 Bridget McInnes (*Virginia Commonwealth University, USA*), Lucy Lu Wang (*Allen Institute for AI, USA*)
- **Powering Semantic Analysis with Bio-ontologies** ..... 821  
 Olivier Bodenreider (*National Institutes of Health, USA*)
- **Why Bother Enabling Biomedical Literature Analysis with Semantics?** ..... 822  
 Karin Verspoor (*RMIT University, Australia*)
- **Exploring Representations for Singular and Multi-Concept Relations for Biomedical Named Entity Normalization** ..... 823  
 Clint Cuffy, Evan French, Sophia Fehrmann, Bridget T. McInnes (*Virginia Commonwealth University, USA*)
- **Graph Convolutional Networks for Chemical Relation Extraction** ..... 833  
 Darshini Mahendran, Christina Tang, Bridget T. McInnes (*Virginia Commonwealth University, USA*)
- **Biomedical Word Sense Disambiguation with Contextualized Representation Learning** .... 843  
 Mozhgan Saeidi, Evangelos Milios, Norbert Zeh (*Dalhousie University, Canada*)

## Workshop: SocialNLP – 10th International Workshop on Natural Language Processing for Social Media

- **SocialNLP’22: 10th International Workshop on Natural Language Processing for Social Media** ..... 849  
 Cheng-Te Li (*National Cheng Kung University, Taiwan*), Lun-Wei Ku (*Academia Sinica, Taiwan*), Yu-Che Tsai  
 (*National Taiwan University, Taiwan*), Wei-Yao Wang (*National Yang Ming Chiao Tung University, Taiwan*)
- **Multi-Context Based Neural Approach for COVID-19 Fake-News Detection** ..... 852  
 Arkadipta De, Maunendra Sankar Desarkar (*Indian Institute of Technology Hyderabad, India*)
- **Measuring the Privacy Dimension of Free Content Websites through Automated Privacy Policy Analysis and Annotation** ..... 860  
 Abdulrahman Alabduljabbar, David Mohaisen (*University of Central Florida, USA*)

- **KAHAN: Knowledge-Aware Hierarchical Attention Network for Fake News Detection on Social Media**..... 868  
Yu-Wun Tseng, Hui-Kuo Yang, Wei-Yao Wang, Wen-Chih Peng (*National Yang Ming Chiao Tung University, Taiwan*)
- **Hoaxes and Hidden Agendas: A Twitter Conspiracy Theory Dataset: Data Paper** ..... 876  
Samantha C. Phillips, Lynnette Hui Xian Ng, Kathleen M. Carley (*Carnegie Mellon University, USA*)
- **Influence of Language Proficiency on the Readability of Review Text and Transformer-based Models for Determining Language Proficiency** ..... 881  
Salim Sazzed (*Old Dominion University, USA*)
- **Making Adversarially-Trained Language Models Forget with Model Retraining: A Case Study on Hate Speech Detection**..... 887  
Marwan Omar, David Mohaisen (*University of Central Florida, USA*)

## Workshop: TempWeb – 12th International Workshop on Temporal Web Analytics

- **12<sup>th</sup> Temporal Web Analytics Workshop (TempWeb) Overview** ..... 894  
Marc Spaniol (*Université de Caen Normandie, France*), Ricardo Baeza-Yates (*EAI, Northeastern University, USA*), Omar Alonso (*Northeastern University, USA*)
- **Temporal Question Answering in News Article Collections**..... 895  
Adam Jatowt (*University of Innsbruck, Austria*)
- **Semantic Modelling of Document Focus-Time for Temporal Information Retrieval** ..... 896  
Lirong Zhang, Hideo Joho, Hai-Tao Yu (*Univeristy of Tsukuba, Japan*)
- **Analytical Models for Motifs in Temporal Networks** ..... 903  
Alexandra Porter (*Stanford University, USA*), Baharan Mirzasoleiman (*University of California, USA*), Jure Leskovec (*Stanford University, USA*)
- **Multi-touch Attribution for Complex B2B Customer Journeys using Temporal Convolutional Networks**..... 910  
Aniket Agrawal (*Adobe Systems, India*), Nikhil Sheoran (*University of Illinois at Urbana-Champaign, USA*), Sourav Suman (*Adobe Systems, India*), Gaurav Sinha (*Adobe Research, India*)
- **Diachronic Analysis of Time References in News Articles**..... 918  
Adam Jatowt (*University of Innsbruck, Austria*), Antoine Doucet (*University of La Rochelle, France*), Ricardo Campos (*LIAAD-INESCTEC, Ci2 - Polytechnic Institute of Tomar, Portugal*)
- **Detection of Infectious Disease Outbreaks in Search Engine Time Series Using Non-Specific Syndromic Surveillance with Effect-Size Filtering**..... 924  
Oded Ovadia, Oren Elisha (*Microsoft Research and Development Israel, Israel*), Elad Yom-Tov (*Microsoft Research Israel, Israel*)
- **A Bi-level Assessment of Twitter Data for Election Prediction: Delhi Assembly Elections 2020**..... 930  
Maneet Singh, S.R.S. Iyengar (*Indian Institute of Technology Ropar, India*), Akrati Saxena (*Eindhoven University of Technology, Netherlands*), Rishemjit Kaur (*CSIR-Central Scientific Instruments Organisation, India and Academy of Scientific and Innovative Research, India*)

## Workshop: DLWoT – The 2nd International Workshop on Deep Learning for the Web of Things

- **DLWoT'22: 2nd International Workshop on Deep Learning for the Web of Things**..... 936  
Wenzhong Guo (*Fuzhou University, China*), Chin-Chen Chang (*Feng-Chia University, Taiwan*), Eyhab Al-Masri (*University of Washington Tacoma, USA*), Chi-Hua Chen (*Fuzhou University, China*), Haishuai Wang (*Fairfield University, USA*), Qichun Zhang (*University of Bradford, UK*), K. Shankar (*Federal University of Piauí, Brazil*)
- **Adaptively Offloading the Software for Mobile Edge Computing** ..... 940  
Xing Chen (*Fuzhou University, China*)



- **Word Embedding based Heterogeneous Entity Matching on Web of Things**..... 941  
Xingsi Xue (*Fujian University of Technology, China*), Jianhua Guo (*Taiyuan University of Technology, China*)
- **A Spatio-Temporal Data-Driven Automatic Control Method for Smart Home Services**..... 948  
Jinrong Chen, Zheyi Chen (*Fuzhou University, China and Fujian Provincial Key Laboratory of Network Computing and Intelligent Information Processing, China*), Longhai Zheng (*Fuzhou University, China*), Xing Chen (*Fuzhou University, China and Fujian Provincial Key Laboratory of Network Computing and Intelligent Information Processing, China*)
- **Discovering Top-k Profitable Patterns for Smart Manufacturing**..... 956  
Shicheng Wan, Jiahui Chen, Peifeng Zhang (*Guangdong University of Technology, China*), Wensheng Gan, Tianlong Gu (*Jinan University, China*)
- **Fast RFM Model for Customer Segmentation** ..... 965  
Shicheng Wan, Jiahui Chen, Zhenlian Qi (*Guangdong University of Technology, China*), Wensheng Gan (*Jinan University, China*), Linlin Tang (*Harbin Institute of Technology, China*)
- **Mining with Rarity for Web Intelligence**..... 973  
Yijie Gui, Wensheng Gan, Yao Chen, Yongdong Wu (*Jinan University, China*)
- **A Hand Over and Call Arrival Cellular Signals-based Traffic Density Estimation Method**..... 982  
Yizhuo Zhang, Hao Fang, Chi-Hua Chen (*Fuzhou University, China*)
- **Fraship: A Framework to Support End-User Personalization of Smart Home Services with Runtime Knowledge Graph** ..... 987  
Zhiming Huang, Jiawen Chen, Liwei Shen, Xing Chen (*Fuzhou University, China*)
- **MI-GCN: Node Mutual Information-based Graph Convolutional Network**..... 996  
Lei Tian, Huaming Wu (*Tianjin University, China*)

## Workshop: GraphLearning – The First International Workshop on Graph Learning

- **GraphLearning'22: 1st International Workshop on Graph Learning**.....1004  
Feng Xia (*Federation University Australia, Australia*), Renaud Lambiotte (*University of Oxford, UK*), Charu Aggarwal (*IBM T. J. Watson Research Center, USA*)
- **Structure-based Large-scale Dynamic Heterogeneous Graphs Processing: Applications, Challenges and Solutions**.....1006  
Wenjie Zhang (*University of New South Wales, Australia*)
- **Graphs in Computer Vision Then and Now: How Deep Learning has reinvigorated Structural Pattern Recognition**.....1007  
Donatello Conte (*Université de Tours, France*)
- **Revisiting Neighborhood-based Link Prediction for Collaborative Filtering** .....1009  
Hao-Ming Fu (*Carnegie Mellon University, Snap Inc., USA*), Patrick Poirson, Kwot Sin Lee, Chen Wang (*Snap Inc., USA*)
- **MarkovGNN: Graph Neural Networks on Markov Diffusion** .....1019  
Md. Khaledur Rahman, Abhigya Agrawal, Ariful Azad (*Indiana University Bloomington, USA*)
- **Multi-view Omics Translation with Multiplex Graph Neural Networks**.....1030  
Costa Georgantas, Jonas Richiardi (*Lausanne University Hospital and University of Lausanne, Switzerland*)
- **Improving Bundles Recommendation Coverage in Sparse Product Graphs**.....1037  
Saloni Agarwal (*The University of Texas at Dallas, USA*), Aparupa Das Gupta, Amit Pande (*Target Corporation, USA*)
- **Unsupervised Superpixel-Driven Parcel Segmentation of Remote Sensing Images Using Graph Convolutional Network**.....1046  
Fulin Huang (*University of Toronto, Canada*), Zhicheng Yang, Hang Zhou, Chen Du, Andy J.Y. Wong, Yuchuan Gou, Mei Han, Jui-Hsin Lai (*PAII Inc., USA*)



• <b>Deep Partial Multiplex Network Embedding .....</b>	1053
Qifan Wang ( <i>Facebook AI, USA</i> ), Yi Fang ( <i>Santa Clara University, USA</i> ), Anirudh Ravula, Ruining He, Bin Shen ( <i>Google Research, USA</i> ), Jingang Wang ( <i>Meituan NLP Center, China</i> ), Xiaojun Quan ( <i>Sun Yat-sen University, China</i> ), Dongfang Liu ( <i>Rochester Institute of Technology, USA</i> )	
• <b>Graph Augmentation Learning.....</b>	1063
Shuo Yu, Huafei Huang ( <i>Dalian University of Technology, China</i> ), Minh N. Dao, Feng Xia ( <i>Federation University Australia, Australia</i> )	
• <b>Scaling R-GCN Training with Graph Summarization.....</b>	1073
Alessandro Generale ( <i>Vrije Universiteit Amsterdam, Netherlands</i> ), Till Blume ( <i>Ernst &amp; Young GmbH WPG - R&amp;D, Germany</i> ), Michael Cochez ( <i>Vrije Universiteit Amsterdam, Netherlands and Elsevier Discovery Lab, Netherlands</i> )	
• <b>RePS: Relation, Position and Structure aware Entity Alignment.....</b>	1083
Anil Surisetty, Deepak Chaurasiya, Nitish Kumar, Alok Singh, Gaurav Dhama, Aakarsh Malhotra, Ankur Arora ( <i>Mastercard, India</i> ), Vikrant Dey ( <i>Indian Institute of Technology Roorkee, India</i> )	
• <b>CCGG: A Deep Autoregressive Model for Class-Conditional Graph Generation .....</b>	1092
Yassaman Ommi ( <i>Amirkabir University of Technology, Iran</i> ), Matin Yousefabad, Faezeh Faez, Amirmojtaba Sabour, Mahdiah Soleymani Baghshah, Hamid R. Rabiee ( <i>Sharif University of Technology, Iran</i> )	
• <b>JGCL: Joint Self-Supervised and Supervised Graph Contrastive Learning .....</b>	1099
Selahattin Akkas, Ariful Azad ( <i>Indiana University Bloomington, USA</i> )	
• <b>Surj: Ontological Learning for Fast, Accurate, and Robust Hierarchical Multi-label Classification.....</b>	1106
Sean T. Yang, Bill Howe ( <i>University of Washington, USA</i> )	
• <b>A Triangle Framework among Subgraph Isomorphism, Pharmacophore and Structure-function Relationship.....</b>	1115
Mengjiao Guo ( <i>Swinburne University of Technology, Australia</i> ), Hui Zheng ( <i>Nanjing University of Posts and Telecommunications, China</i> ), Tengfei Ji ( <i>Peking Union Medical College, China</i> ), Hu Fa, Jing He ( <i>Nanjing University of Posts and Telecommunications, China</i> )	
• <b>Understanding Dropout for Graph Neural Networks .....</b>	1128
Juan Shu, Bowei Xi ( <i>Purdue University, USA</i> ), Yu Li ( <i>Chinese University of Hong Kong, China</i> ), Fan Wu ( <i>Purdue University, USA</i> ), Charles Kamhoua ( <i>US Army Research Laboratory, USA</i> ), Jianzhu Ma ( <i>Peking University, China</i> )	
• <b>Mining Homophilic Groups of Users using Edge Attributed Node Embedding from Enterprise Social Networks .....</b>	1139
Priyanka Sinha ( <i>Tata Consultancy Services Limited, India and IIT Kharagpur, India</i> ), Ritu Patel, Pabitra Mitra ( <i>IIT Kharagpur, India</i> ), Dilys Thomas, Lipika Dey ( <i>Tata Consultancy Services Limited, India</i> )	
• <b>Mining Multivariate Implicit Relationships in Academic Networks.....</b>	1148
Bo Xu, Bowen Chen ( <i>Dalian University of Technology, China</i> ), Tianyu Zhang ( <i>65447, China</i> ), Jiaying Liu, Chunke Liao, Zhehuan Zhao ( <i>Dalian University of Technology, China</i> )	
• <b>SchemaWalk: Schema Aware Random Walks for Heterogeneous Graph Embedding .....</b>	1157
Ahmed E. Samy, Lodovico Giarretta, Zekarias T. Kefato, Šarūnas Girdzijauskas ( <i>KTH Royal Institute of Technology, Sweden</i> )	
• <b>Multi-Graph based Multi-Scenario Recommendation in Large-scale Online Video Services .....</b>	1167
Fan Zhang, Qiuying Peng ( <i>OPPO Research Institute, China</i> ), Yulin Wu ( <i>OPPO AI&amp;Data Engineering System, China</i> ), Zheng Pan ( <i>OPPO Research Institute, China</i> ), Rong Zeng, Da Lin ( <i>Bytedance Ltd., China</i> ), Yue Qi ( <i>OPPO Research Institute, China</i> )	

## Workshop: UserNLP – International Workshop on User-centered Natural Language Processing

- **UserNLP'22: 2022 International Workshop on User-centered Natural Language Processing** .....1176  
 Xiaolei Huang (*University of Memphis, USA*), Lucie Flek (*University of Marburg, Germany*),  
 Franck Dernoncourt (*Adobe Research, USA*), Charles Welch (*University of Marburg, Germany*), Silvio Amir  
 (*Northeastern University, USA*), Ramit Sawhney (*Tower Research Capital, India*), Diyi Yang (*Georgia Institute of  
 Technology, USA*)
- **Personalization and Relevance in NLG** .....1178  
 Shiran Dudy (*University of Colorado Boulder, USA*)
- **Stylistic Control for Neural Natural Language Generation** .....1179  
 Shereen Oraby (*Amazon Alexa AI, USA*)
- **Concept Annotation from Users Perspective: A New Challenge** .....1180  
 Souvika Sarkar, Shubhra Kanti (Santu) Karmaker (*Auburn University, USA*)
- **Detecting Addiction, Anxiety, and Depression by Users Psychometric Profiles** .....1189  
 Anna Monreale (*University of Pisa, Italy*), Benedetta Iavarone (*Scuola Normale Superiore, Italy*),  
 Elena Rossetto (*University of Pisa, Italy*), Andrea Beretta (*ISTI-CNR, Italy*)
- **Expressing Metaphorically, Writing Creatively: Metaphor Identification for  
 Creativity Assessment in Writing** .....1198  
 Dongyu Zhang, Minghao Zhang (*Dalian University of Technology, China*), Ciyuan Peng,  
 Feng Xia (*Federation University Australia, Australia*)
- **A Decision Model for Designing NLP Applications** .....1206  
 Eason Chen, Yuen-Hsien Tseng (*National Taiwan Normal University, Taiwan*)
- **Do Not Read the Same News! Enhancing Diversity and Personalization of  
 News Recommendation** .....1211  
 Seonghwan Choi (*Seoul National University, South Korea*), Hyeondey Kim, Manjun Gim (*Bigpearl Inc.,  
 South Korea*)

## Workshop: WebAndTheCity – 8th International Workshop on Web and Smart Cities

- **WebAndTheCity'22: 8th International Workshop on the Web and Smart Cities** .....1216  
 Leonidas Anthopoulos (*University of Thessaly, Greece*), Marijn Janssen (*Delft University of Technology,  
 Netherlands*), Vishanth Weerakkody (*University of Bradford, UK*)
- **Human Centric Design in Smartcity Technologies: Implications for the Governance,  
 Control and Performance Evaluation of Mobility Ecosystems** .....1218  
 Rob Christiaanse (*Technical University, Netherlands*)
- **Citizens as Developers and Consumers of Smart City Services: A Drone Tour  
 Guide Case** .....1228  
 Christian Muck, Alexander Voelz, Danial M. Amlashi, Dimitris Karagiannis (*University of Vienna, Austria*)
- **A Human-Centered Design Approach for the Development of a Digital Care Platform  
 in a Smart City Environment: Implications for Business Models** .....1237  
 Jelena Bleja, Sara Neumann, Tim Krüger, Uwe Grossmann (*University of Applied Sciences and Arts Dortmund,  
 Germany*)
- **Sensor Network Design for Uniquely Identifying Sources of Contamination in  
 Water Distribution Networks** .....1245  
 Kaustav Basu, Arunabha Sen (*School of Computing and Augmented Intelligence, USA*)

- **Enhancing Crowd Flow Prediction in Various Spatial and Temporal Granularities**.....1251  
Marco Cardia (*University of Pisa, Italy*), Massimiliano Luca (*Free University of Bolzano, Italy* and Bruno Kessler Foundation, Italy), Luca Pappalardo (*National Research Council (CNR), Italy*)
- **A Framework to Enhance Smart Citizen Science in Coastal Areas** .....1260  
Maria Papoutsoglou (*University of Cyprus, Cyprus*), Konstantinos Markakis (*Foundation for Research and Technology - Hellas (FORTH), Greece*), Loukas Chatzivasili, Georgia Kapitsaki (*University of Cyprus, Cyprus*), Kostas Magoutis (*Foundation for Research and Technology - Hellas (FORTH), Greece* and *University of Crete, Greece*), Leonidas Katelaris (*University of Cyprus, Cyprus*), Chryssoula Bekiari (*Foundation for Research and Technology - Hellas (FORTH), Greece*)
- **Multi-Tenancy in Smart City Platforms** .....1266  
Ioannis Nikolaou, Leonidas Anthopoulos (*University of Thessaly, Greece*)

## Workshop: Wiki – 9th International Wiki Workshop

- **WIKI’2022: 9<sup>th</sup> Annual Wiki Workshop** .....1271  
Leila Zia (*Wikimedia Foundation, USA*), Robert West (*EPFL, Switzerland*), Miriam Redi (*Wikimedia Foundation, USA*), Srijan Kumar (*Georgia Tech, USA*), Emily Lescak (*Wikimedia Foundation, USA*)
- **Rows from Many Sources: Enriching Row Completions from Wikidata with a Pre-trained Language Model**.....1272  
Carina Negreanu, Alperen Karaoglu, Jack Williams (*Microsoft Research, UK*), Shuang Chen (*Harbin Institute of Technology, China*), Daniel Fabian, Andrew Gordon (*Microsoft Research, UK*), Chin-Yew Lin (*Microsoft Research, China*)
- **Are Democratic User Groups More Inclusive?**.....1281  
Anass Sedrati (*INPT and Wikimedia Morocco, Morocco*), Reda Benkhadra (*Al Akhawayn University and Wikimedia Morocco, Morocco*)
- **A Map of Science in Wikipedia**.....1289  
Puyu Yang, Giovanni Colavizza (*University of Amsterdam, Netherlands*)
- **Improving Linguistic Bias Detection in Wikipedia using Cross-Domain Adaptive Pre-Training**.....1301  
Karthic Madanagopal, James Caverlee (*Texas A&M University, USA*)
- **Anchor Prediction: A Topic Modeling Approach** .....1310  
Jean Dupuy, Adrien Guille, Julien Jacques (*Université de Lyon, France*)
- **The Gender Perspective in Wikipedia: A Content and Participation Challenge**.....1319  
Núria Ferran-Ferrer (*Universitat de Barcelona, Spain*), Marc Miquel-Ribé (*Tecnocampus-Universitat Pompeu Fabra, Spain*), Julio Meneses, Julià Minguillón (*Universitat Oberta de Catalunya, Spain*)
- **Going Down the Rabbit Hole: Characterizing the Long Tail of Wikipedia Reading Sessions** .....1324  
Tiziano Piccardi (*EPFL, Switzerland*), Martin Gerlach (*Wikimedia Foundation, USA*), Robert West (*EPFL, Switzerland*)
- **Building a Public Domain Voice Database for Odia**.....1331  
Subhashish Panigrahi (*O Foundation, India*)

**Author Index**..... 1339

# The Gender Perspective in Wikipedia: A Content and Participation Challenge

Núria Ferran-Ferrer  
nferranf@ub.edu  
Universitat de Barcelona  
Barcelona, Catalonia

Julio Meneses  
Universitat Oberta de Catalunya  
Barcelona, Spain  
jmenesesn@uoc.edu

Marc Miquel-Ribé  
Tecnocampus-Universitat Pompeu Fabra  
Mataró, Catalonia  
mmiquelr@tecnocampus.cat

Julià Minguillón  
Universitat Oberta de Catalunya  
Barcelona, Spain  
jminguillona@uoc.edu

## ABSTRACT

Wikipedia is one of the most widely used information sources in the world. Although one of the guiding pillars of this digital platform is ensuring access to the diversity of human knowledge from a neutral point of view, there is a clear and persistent gender bias in terms of content about or contributed by women. The challenge is to include women as equal partners in the public sphere, in which Wikipedia is developing a central role as the most used educational resource among students, professionals, and many other profiles. In this paper, we introduce the gender perspective in the analysis of the gender gap in the content and participation of women in Wikipedia. While most studies focus on one of the two dimensions in which the gender gap has been observed, we review both approaches to provide an overview of the available evidence. Firstly we introduce how the gender gap is framed by the Wikimedia Movement strategy, then we evaluate the gender gap on content and participation, especially regarding editor practices. Finally, we provide some insights to broaden the discussion about the consequences of not addressing the gender gap in Wikipedia, and we provide some research topics that can support the generation of recommendations and guidelines for a community that needs both equity and diversity.

## CCS CONCEPTS

• **Human-centered computing** → **Collaborative content creation**; • **Information systems** → **Social networks**.

## KEYWORDS

Gender gap, Gender bias, Equality, Editors, Peer production, Participation, Retention, Digital inequality, Open content, Women biographies, Public sphere, Notability

## ACM Reference Format:

Núria Ferran-Ferrer, Marc Miquel-Ribé, Julio Meneses, and Julià Minguillón. 2022. The Gender Perspective in Wikipedia: A Content and Participation Challenge. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion)*, April 25–29, 2022, Virtual Event, Lyon, France. ACM, New York, NY, USA, ?? pages. <https://doi.org/10.1145/3487553.3524937>

## 1 INTRODUCTION

Information and communication technologies seem to have opened up a scenario in which pluralism and new forms of expression are more possible than ever before, giving rise to an interconnected public sphere [5, 10, 11, 48, 53]. Wikipedia is part of the current public sphere and has the unique potential to facilitate a more equitable production of knowledge [23] through commons-based peer production and the provision of virtual spaces for discussion. Wikipedia defends itself as a kind of intellectual democracy because it is ostensibly based on the principles of rational deliberation, consensus, and negotiation [8].

Wikipedia has transformed the way in which information is produced and distributed through open collaboration [51], but doubts have been raised about how decentralized, flexible, and open these new opportunities are, given that they are generated within a pre-existing economic, social, and political model [28]. Wikipedia has content policies such as neutral point of view, no original research, and verifiability which might pose unintended difficulties to several collectives in order to reach access to the public sphere and include their contribution to knowledge construction in the encyclopedia.

Through agreed-upon rules for debates, it is said that the community of Wikipedians can reach consensus, constituting an ideal speech community in Habermas's terms [21]. However, this notion of consensus on Wikipedia often results in exclusions, thereby failing to meet Wikipedia's ideal goals of being the most comprehensive encyclopedia and including a plurality of points of view [17].

Moreover, it is well known that Wikipedia suffers from a strong and persistent gender bias in different ways [27, 37, 54]. This issue has been raised both in the community involved in the editing process (most editors are men) and in its available content (biographies of men outnumber those of women and tend to be more extensive). In this short paper, we will introduce the gaps in terms of the gender perspective, in its contents and its participation, in order to ensure that Wikipedia contents reflect the true composition of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France.

© 2022 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-9130-6/22/04...\$15.00

<https://doi.org/10.1145/3487553.3524937>

society and eliminate stereotypes as recommended by UNESCO in its Gender-Sensitive Indications for Media [20]. In this regard, it must be noted that the only element that appears to be fairly gender equally distributed is the readership of Wikipedia [25].

The organization of this paper is as follows. In section 2, we will explain how the gender gap is framed by the Wikimedia Movement strategy conversations over the years. In section 3, we will deepen into what we call the Gender perspective, which is the gender gap both in terms of content and participation, paying special attention to editor editing practices and the first stages of the editor's life cycle measured by the metric retention. In section 4, we will conclude with some reflections on the negative consequences of such a gap and the urgency to address it. The aim of this paper is to introduce scholars to an unsolved long-term problem that affects the most consumed knowledge repository and encourage research on both its causes and the ways in which they could be addressed.

## 2 THE GENDER GAP IN THE WIKIMEDIA MOVEMENT AND THE STRATEGY CONVERSATIONS

Gender inequality, both in Wikipedia's communities and their available contents, has been recognized by the Wikimedia Foundation, which has been working to address it. In 2011, Sue Gardner, former executive director of the Wikimedia Foundation, suggested some reasons that may prevent women from editing Wikipedia, such as difficulties with the interface, lack of time, lack of self-confidence, encountering a misogynistic atmosphere, and the absence of a culture of social interaction [16].

Currently, the Wikimedia Movement, which includes the Wikimedia Foundation and all the language and territory-based affiliates, has agreed on a new strategic direction towards the 2030 horizon. This agreement sets the goal of delivering knowledge as a service ("become a platform that serves open knowledge to the world across interfaces and communities") and knowledge equity ("focus our efforts on the knowledge and communities that have been left out by structures of power and privilege [...]. We will break down the social, political, and technical barriers preventing people from accessing and contributing to free knowledge"). Accordingly, the strategic direction has collaboratively set the principles<sup>1</sup> of inclusivity and people-centeredness (i.e., addressing people's needs) as the core of its recommendations, and has established the goal to eliminate the gender gap and focus on the inclusion of underrepresented groups by 2030<sup>2</sup>.

## 3 THE GENDER PERSPECTIVE IN WIKIPEDIA

Several authors have analyzed the gender gap in Wikipedia, both in terms of content and participation. The gaps in both levels suggest that there is still a lack of gender perspective in Wikipedia. In the next two sections, we will review and describe its aspects.

### 3.1 The gender gap in content

Regarding the research conducted on the gender gap in content, less than 20% of biographies are from women in the Spanish Wikipedia [15]. One of the handicaps faced to develop new content in the Wikipedia is that an entry must be notable, that is to say, it must comply with the notability requirements: articles should be based on third party, non-affiliated sources, with some degree of editorial overview, as a guarantee of neutrality and quality. In this sense, Wikipedia's vision of knowledge conflates cultural significance with visibility in secondary media sources [17]. This is especially relevant to women, who do not easily attract mainstream media interest and therefore are less frequently covered. Only 24% of news sources are related to women; that is to say, most of the people seen, heard, or read about in the media are men. Additionally, the news topics in which women are the most visible gather the least coverage [36]. The case of Spain is even worse, only 7.3% of women are represented as main characters in media [38].

Men and women also focus on different content areas in Wikipedia, thus reinforcing content imbalance between 'male' and 'female' topics [32]. The Wikipedia Gender Inequality Index measures the gender gap in the content of Wikipedia, following the analysis in time and languages [31] of the available biographies in different languages, and provides a quantitative evaluation of the gap. Other initiatives such as the Wikiproject 'Countering Systematic Bias' [56] aim at broader goals, including gender bias. There are specific studies about the content gender gap, such as the analysis of the number of female porn actresses' biographies vs. the number of female poets' biographies [30], attributing the differences to editors' interests and showing how the lack of female perspective is present in the "sum of all knowledge", the Wikipedia's prime objective. Other analyses compare the born-digital encyclopedia with other traditional encyclopedias such as the Britannica [45], and previous studies aim at the Britannica bias as well [18].

Recently, a theoretical study, from the cyberfeminism approach, has analyzed the empirical studies on content gap in order to provide a framework for understanding this complex phenomenon which produces an asymmetric collective perception of knowledge. This research mentions at least ten types of biases in Wikipedia content: related to length of writing, centrality in the hyperlink network, lexicon used, number of sources, quantity and quality of images, multilingual notability, topics mentioned, classification of articles, among others [7].

### 3.2 The gender gap in participation

Regarding the gender gap in participation or community composition, Glott et al. [19] analyzed a survey to characterize Wikipedia's contributors and found that only 12.64% were women. Lam et al. [32] tried to determine the nature of this gender imbalance. Analyzing users' profiles, they found that female editors only comprised 16.1% of the profiles and only accounted for 9.0% of edits. In a previous study based on a massive content codification, we showed that female editors represent a small minority of the editors involved, accounting for just 11.3% of all the editors of the Spanish Wikipedia [39, 40]. This evidence is similar to that of the English Wikipedia. Although till now the data has been obtained by survey methods,

<sup>1</sup>[https://meta.wikimedia.org/wiki/Strategy/Wikimedia\\_movement/2018-20/Recommendations/Movement\\_Strategy\\_Principles](https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2018-20/Recommendations/Movement_Strategy_Principles)

<sup>2</sup>[https://meta.wikimedia.org/wiki/Strategy/Wikimedia\\_movement/2018-20/Recommendations/Introduction](https://meta.wikimedia.org/wiki/Strategy/Wikimedia_movement/2018-20/Recommendations/Introduction)

it has been reported that the percentage of female editors is between 10% and 15% or even lower in the English Wikipedia [1, 32]. Nevertheless, in our previous quantitative research, we found that for those editors with a large number of edits over time, such differences were not statistically significant, or even reversed, with females outperforming male editors [40]. According to this evidence, engaging, participating, and persisting in Wikipedia appears to be a much more complex process for most female editors than for men.

**3.2.1 The gender gap in editing practices.** When researching the gender gap in Wikipedia, different editing practices have been observed among men and women. Women are more exposed to conflicts such as reversion and blocking than men, pushing them towards dropping out. In 2017, Ford and Wajcman [14] described the gender gap as the result of a knowledge institution governed by power issues. Since power is male-dominated, those who do not fit in with what Wikipedia recognizes as knowledge are excluded.

Access to the public sphere is not equally available to everyone. According to Juliano [29], whoever is in the power position in hierarchical societies such as ours determines what constitutes legitimate discourse and what does not, as well as who has the right to express their point of view or must remain silent. Therefore, to understand the production of content, we need to better understand the culture behind it and, particularly, how exclusion practices are performed [14]. Juliano reminds us that women have represented the group that has been most systematically denied the right to talk in the public sphere throughout history. Missing individuals and collectives in the public sphere subtract plurality from the human condition. As Arendt [2] stated, if someone is prevented from accessing the public space, they are not only deprived of seeing and hearing others, but also of seeing and hearing themselves. For Arendt, the public sphere is the place where plurality concurs and has the double meaning of equality and distinction. The public arena is where individuals are differentiated by each other, and reality is produced from the intertwining of perspectives of all those who fit in the arena occupying different positions. Thus, a common is created where identities are distinguished and recognized.

Habermas defined the public sphere as the realm of our social life, in which public opinion can be formed. Therefore, the public sphere must permit the circulation of information to enable communicative interaction among individuals, to form a public consensus, and to facilitate decision-making. For Habermas, the success of the public sphere was founded on rational-critical discourse. In the public sphere, everyone is an equal participant, and the power of argument is the supreme communication skill [22]. Obviously, liberal values were placed upon patriarchal substrates, the subordination of women, who were kept in the private sphere to tend merely to domestic matters. Therefore, achieving the equality needed in the public sphere requires the epistemological and philosophical foundations of liberalism to be broken [43].

The Habermasian conception of the public sphere also understands deliberative processes as opportunities for consensus, although agonistic pluralism criticizes this idea, as consensus can only be built at the expense of the dissenters' voices. Mouffe [42] believes that society is irreducibly plural, therefore the idea that

all identities may deliberate based on shared communicative rationality is implausible. Also, the idea that an unadulterated and unbiased ideal speech situation should serve as a model is unrealistic. Mouffe's main critique is that Habermas understands a rational consensus specifically in liberal terms, and this excludes individuals and collective identities who do not fully identify with these values.

In Wikipedia, the power relations, its ideology, and its culture are produced by infrastructures [14]. The rationales of the origin of Wikipedia are found in the modern encyclopedic tradition and the free software movement. Wikipedia is built on an installed base [49] and represents knowledge and how this knowledge is produced [57]. The encyclopedic tradition produces a specific kind of knowledge and has its roots in colonialism according to predefined norms and logics for the scientific knowledge [55]. Contrary to an ideal design, this knowledge is not open to everybody.

Furthermore, low female participation rates in Science, Technology, Engineering, and Maths (STEM) are part of the academic tradition [9]. Some critics blame this problem on women's attitudes, rather than pointing to the lack of welcoming spaces and methods in STEM disciplines and institutions. Works questioning whether technical knowledge and expertise are masculine skills [3, 6, 35, 46, 55] reveal the exclusion of women in the academic and encyclopedic tradition. These issues in science and technology extend to computer science and the free software tradition in which Wikipedia is rooted.

On the one hand, according to Ford and Wajcman [14], Wikipedia is extending the epistemologies of previously male-dominated technoscientific projects such as D'Alembert and Diderot's encyclopedia, which contained articles and drawings explaining the world, culture, and society, among others, and were written by philosophers and academics. The gender gap is not just related to Wikipedia, there are also examples of how women's writings were invisible, hidden or underrated throughout history [47].

On the other hand, Wikipedia is based on hacker culture [34], follows the hacker ethics [26], and it is based on an open culture of collaboration. At the same time, it has been built within a culture of white men, online geeks and programmers [45] where the masculine identity has its roots in computational culture [52]. The research on this topic suggests that the lack of participation of women is due to their lack of skills, confidence, and fear of criticism and conflict [24], but Ford and Wajcman's point of view identified a culture that denies the space for women's participation. As Aaron Swartz, a well-known hacker, wrote: 'a website is not based only on the technology but on its community' [50].

Bryce Peake suggests that the change needed is not the incorporation of women, but "the creation of a space of multiple points of view" [44]. Achieving this change would first require a major cultural shift among Wikipedians. Although the ideal solution could be to expand the public sphere by including women, this strategy is not receiving support from feminist authors such as Landes and Young, who agree that the public sphere already has a gender bias. For them, the exclusion of women from the public sphere is no accident, they think that universality is homogeneous to appeal to impartiality and is therefore a way of silencing differences and plurality [33, 58]. In this sense, Benhabib [4] asserts that it is only possible to include women in the public sphere if they enter with full rights in the universal dialogue, and their specificity as women



is recognized. In contrast to Habermasian's universal communication community, Benhabib argues for the need for a community that shows solidarity, taking into consideration the needs for both equality and differentiation. This means that the public sphere needs to be feminized. However, the point of view of feminist criticism, women do not identify with the cultural patterns that respond to the male experience of the world, they do not feel comfortable with the requirements of public time and space, which demand exclusive dedication, meaning that others will have to take care of domestic demands [12].

**3.2.2 The gender gap in retention.** Several studies have shown that women seem to stop editing Wikipedia sooner than men [32, 40]. In the case of newcomers, the mechanisms used to attract and retain new editors are not enough to retain them, especially in the case of women. Some authors have pointed out that creating a safe and warm environment is a key factor to attract and retain new editors [13, 41]. Participation in WikiProjects<sup>3</sup> or edit-a-thons<sup>4</sup> are two examples of these spaces. Also being a Wikipedian in a smaller Wikipedia helps, as are said to provide a quiet and safer environment for participation [13]. Research in such spaces can be useful to better detect and understand possible signs of attrition among newcomers. This research could provide some basic recommendations and guidance to both editors and communities, fostering such spaces and smaller communities.

## 4 CONCLUSIONS

Adopting the gender perspective, both in content and participation, is crucial to ensure that Wikipedia reflects the true composition of society and eliminate stereotypes in one of the most used information/learning resources in the world.

In this sense, there is the need that women could effectively access, edit, and transform Wikipedia. To do so, on the one hand, we see the need for measuring their effective participation as well as to characterize their editing practices, observing the gender differences in content creation, modification, and participation in the discussion processes, among others. The assessment of editing practices can provide a set of different types of good and bad practices for the rest of the community that could contribute to the generation of a space of plurality and diversity once implemented.

Additionally, studying the grounds of the notability criteria, that serve as a basis to determine the value of articles through the analysis of their information sources, is critical for women's biographies. The absence of sources or citations in an article does not always indicate that a subject is not notable. This is the case for women's biographies, as they suffer an exclusion from mainstream media that makes it harder to provide the traditionally accepted sources. Alternative sources or new evidence for achieving the notability criteria should be suggested.

On the other hand, we foresee the need to deepen our knowledge on the retention of women editors, to be able to identify the critical points for dropping out and provide recommendations to foster their engagement. Using a gender-based perspective, this research could expand on previously examined concepts and theories, translating them into specific social actions. Collaboration between Wikipedian

groups and projects, who share aspirations for reducing gender bias, with other entities such as libraries, schools, and universities, could provide fruitful opportunities for positive change.

## ACKNOWLEDGMENTS

This study is funded by the Plan Nacional I+D+I of the Ministry of Science and Innovation of Spain (Ref. PID2020-116936RA-I00).

## REFERENCES

- [1] Judd Antin, Raymond Yee, Coye Cheshire, and Oded Nov. 2011. Gender differences in Wikipedia editing. In *Proceedings of the 7th international symposium on wikis and open collaboration (WikiSym '11)*. Association for Computing Machinery, New York, NY, USA, 11–14. <https://doi.org/10.1145/2038558.2038561>
- [2] Hanna Arendt. 1992. *The human condition*. University of Chicago Press, Chicago, IL.
- [3] Anne Marie Balsamo et al. 1996. *Technologies of the gendered body: Reading cyborg women*. Duke University Press.
- [4] Seyla Benhabib. 1992. *Situating the self: Gender, community, and postmodernism in contemporary ethics*. Routledge, New York.
- [5] Yochai Benkler. 2008. *The wealth of networks*. Yale university press, New Haven, CT. <https://doi.org/10.12987/9780300127232>
- [6] Anne-Jorunn Berg and Merete Lie. 1995. Feminism and constructivism: Do artifacts have gender? *Science, Technology, & Human Values* 20, 3 (1995), 332–351.
- [7] Pablo Beytia and Claudia Wagner. 2022. Visibility layers: A framework for facing the complexity of the gender gap in wikipedia content. <https://doi.org/10.31235/osf.io/5ndkm>.
- [8] Laura Black, Ted Welsler, Jocely DeGroot, and Daniel Cosley. 2008. "Wikipedia is not a democracy": Deliberation and policy-making in an online community. In *Annual meeting of the International Communication Association*. International Communication Association, Montreal, Quebec, Canada, 1–50. <http://www.cs.cornell.edu/~danco/research/papers/democracy-black-ica2008.pdf>
- [9] Peter Burke. 2012. *A Social History of Knowledge, volume II: from the Encyclopedia to Wikipedia*. Polity Press, Cambridge, UK.
- [10] Manuel Castells. 2008. The New Public Sphere: Global Civil Society, Communication Networks, and Global Governance. *The ANNALS of the American Academy of Political and Social Science* 616, 1 (2008), 78–93. <https://doi.org/10.1177/000216207311877>
- [11] Anna Clua, Nuria Ferran-Ferrer, and Ludovic Terren. 2018. Youth impact on the public sphere in Press and Twitter: The dissolution of the Spanish Youth Council. *Comunicar* 55, 1 (2018), 49–55. <https://doi.org/10.3916/C55-2018-05>
- [12] Jodi Dean. 2007. *Feminism, Communicative Capitalism, and the Inadequacies of Radical Democracy*. Palgrave Macmillan, London, UK, 226–245. [https://doi.org/10.1057/9780230592469\\_13](https://doi.org/10.1057/9780230592469_13)
- [13] Núria Ferran-Ferrer, Patricia Castellanos-Pineda, Julià Minguiñón, and Julio Meneses. 2021. The gender gap on the spanish wikipedia: Listening to the voices of women editors. *Profesional de la Información* 30, 5 (2021), e300516. <https://doi.org/10.3145/epi.2021.sep.16>
- [14] Heather Ford and Judy Wajcman. 2017. 'Anyone can edit', not everyone does: Wikipedia's infrastructure and the gender gap. *Social Studies of Science* 47, 4 (2017), 511–527. <https://doi.org/10.1177/0306312717692172>
- [15] Wikimedia France. 2020. Gender Gap in Spanish Wikipedia. <https://denelezh.wmcloud.org/gender-gap/?project=eswiki/>.
- [16] Sue Gardner. 2011. Nine reasons women don't edit Wikipedia (in their own words). <https://suegardner.org/2011/02/19/nine-reasons-why-women-dont-edit-wikipedia-in-their-own-words/>.
- [17] Maude Gauthier and Kim Sawchuk. 2017. Not notable enough: feminism and expertise in Wikipedia. *Communication and critical/cultural studies* 14, 4 (2017), 385–402. <https://doi.org/10.1080/14791420.2017.1386321>
- [18] Thomas Gillian. 1992. A Position to Command Respect: Women and the Eleventh Britannica. <https://archive.org/details/positiontocommand0000thom>.
- [19] Ruediger Glott, Philipp Schmidt, and Rishab Ghosh. 2010. Wikipedia survey—overview of results. , 1158–1178 pages.
- [20] Alton Grizzle. 2012. *Gender-sensitive indicators for media: Framework of indicators to gauge gender sensitivity in media operations and content*. United Nations Educational, Scientific and Cultural Organization, Paris, France. <https://unesdoc.unesco.org/ark:/48223/pf0000217831>
- [21] Jürgen Habermas. 1981. *Theory of communicative action*. Beacon Press, Boston, MA), Boston.
- [22] Jürgen Habermas, Sara Lennox, and Frank Lennox. 1974. The public sphere: An encyclopedia article. *New German Critique* 3 (1974), 49–55.
- [23] Eszter Hargittai and Aaron Shaw. 2015. Mind the skills gap: the role of Internet know-how and gender in differentiated contributions to Wikipedia. *Information, communication and society* 18, 4 (2015), 424–442. <https://doi.org/10.1080/>

<sup>3</sup><https://en.wikipedia.org/wiki/Wikipedia:WikiProject>

<sup>4</sup><https://en.wikipedia.org/wiki/Edit-a-thon>

- 1369118X.2014.957711
- [24] Susan C Herring, Joseph Reagle, Justine Cassell, Terri Oda, Anna North, Jessamyn West, and Marina Ranga. 2011. Where are the women in Wikipedia. *The New York Times (Feb 02)* (2011). <https://www.nytimes.com/roomfordebate/2011/02/02/where-are-the-women-in-wikipedia>
  - [25] Benjamin Mako Hill and Aaron Shaw. 2013. The Wikipedia gender gap revisited: Characterizing survey response bias with propensity score estimation. *PLoS ONE* 8, 6 (2013), e65782. <https://doi.org/10.1371/journal.pone.0065782>
  - [26] Pekka Himanen. 2002. *The hacker ethic and the spirit of the information age*. Random House Trade Paperbacks, New York, US.
  - [27] Marit Hinnoosar. 2019. Gender inequality in new media: Evidence from Wikipedia. *Journal of economic behavior & organization* 163 (2019), 262–276. <https://doi.org/10.1016/j.jebo.2019.04.020>
  - [28] Nina Hood and Allison Littlejohn. 2018. Hacking history: redressing gender inequities on Wikipedia through an editathon. *International Review of Research in Open and Distributed Learning* 19, 5 (2018).
  - [29] Dolores Juliano. 2017. *Tomar la Palabra. Mujeres, discursos y silencios*. Edicions Bellaterra, Bellaterra, Spain.
  - [30] Jenny Kleeman. 2015. The Wikipedia wars: does it matter if our biggest source of knowledge is written by men? <https://www.newstatesman.com/uncategorized/2015/05/wikipedia-has-colossal-problem-women-dont-edit-it>
  - [31] Piotr Konieczny. 2016. Teaching with Wikipedia in a 21st-century classroom: Perceptions of Wikipedia and its educational benefits. *Journal of the Association for Information Science and Technology* 67, 7 (2016), 1523–1534. <https://doi.org/10.1002/asi.23616>
  - [32] Shyong (Tony) K Lam, Anuradha Uduwage, Zhenhua Dong, Shilad Sen, David R Musicant, Loren Terveen, and John Riedl. 2011. WP: clubhouse? An exploration of Wikipedia's gender imbalance. In *Proceedings of the 7th international symposium on Wikis and open collaboration*. ACM Press, Mountain View, CA, US, 1–10. <https://doi.org/10.1145/2038558.2038560>
  - [33] Joan B Landes. 1992. Jürgen Habermas, the structural transformation of the public sphere: A feminist inquiry. *Praxis International* 12, 1 (1992), 106–127.
  - [34] Andrew Lih. 2009. *The Wikipedia revolution: How a bunch of nobodies created the world's greatest encyclopedia*. Hachette Books, Paris, France.
  - [35] J Scott Long and Mary Frank Fox. 1995. Scientific careers: Universalism and particularism. *Annual review of sociology* 21, 1 (1995), 45–71.
  - [36] Sarah Macharia. 2020. *Global Media Monitoring Project (GMMP)*. Wiley Online Library, 1–6. <https://doi.org/10.1002/9781119429128.iegmc074>
  - [37] Paolo Massa and Asta Zelenkauskaitė. 2014. *Global wikipedia: International and cross-cultural issues in online collaboration*. Vol. 85. Rowman & Littlefield, Maryland, US, Chapter Gender Gap in Wikipedia, 49–68.
  - [38] M Pilar Matud, Inmaculada Espinosa, and Carmen Rodríguez Wangüemert. 2021. Women and men portrayal on television news: a study of Spanish television newscast. *Feminist Media Studies* 21, 2 (2021), 298–314. <https://doi.org/10.1080/14680777.2019.1681489>
  - [39] Julià Minguiñón, Eduard Aibar, Maura Lerga, Josep Lladós, and Antoni Meseguer-Artola. 2018. Wikipedia in academia as a teaching tool: from aversive to proactive faculty profiles. <https://doi.org/10.48550/ARXIV.1801.07138>
  - [40] Julià Minguiñón, Julio Meneses, Eduard Aibar, Núria Ferran-Ferrer, and Sergi Fàbregues. 2021. Exploring the gender gap in the Spanish Wikipedia: Differences in engagement and editing practices. *PLoS ONE* 16, 2 (2021). <https://doi.org/10.1371/journal.pone.0246702>
  - [41] Jonathan T. Morgan, Siko Bouterse, Sara Stierch, and Heather Walls. 2013. Tea & sympathy: Crafting positive new user experiences on wikipedia. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. ACM, New York, US, 839–848. <https://doi.org/10.1145/2441776.2441871>
  - [42] Chantal Mouffe. 2000. *The democratic paradox*. Verso, London, UK.
  - [43] Carol Pateman. 1988. *The sexual contract*. Cambridge: Polity Press, Cambridge, UK.
  - [44] Bryce Peake. 2015. WP: THREATENING2MEN: Misogynist Infopolitics and the Hegemony of the Asshole Consensus on English Wikipedia. *Journal of gender, new media and technology* 7 (2015). <https://adanewmedia.org/2015/04/issue7-peake/>
  - [45] Joseph Reagle and Lauren Rhue. 2011. Gender bias in Wikipedia and Britannica. <https://ijoc.org/index.php/ijoc/article/view/777>. *International Journal of Communication* 5 (2011), 1138–1158.
  - [46] Hilary Rose. 1983. Hand, brain, and heart: A feminist epistemology for the natural sciences. *Signs: journal of Women in Culture and Society* 9, 1 (1983), 73–90.
  - [47] Joanna Russ. 2018. *How to suppress women's writing*. University of Texas Press, Texas, US.
  - [48] James Slevin. 2000. *Internet and society*. Blackwell Publishers, Inc., New Jersey, US.
  - [49] Susan Leigh Star. 1999. The ethnography of infrastructure. *American behavioral scientist* 43, 3 (1999), 377–391. <https://doi.org/10.1177/000276499921955326>
  - [50] Aaron Swartz. 2016. *The boy who could change the world: The writings of Aaron Swartz*. The New Press, New York, US.
  - [51] Nathaniel Tkacz. 2014. Wikipedia and the Politics of Openness.
  - [52] Sherry Turkle. 2005. *The second self: Computers and the human spirit*. MIT Press.
  - [53] Jan Van Dijk. 2012. *The network society*. Sage Publications Ltd., Newbury Park, US.
  - [54] Claudia Wagner, Eduardo Graells-Garrido, David Garcia, and Filippo Menczer. 2016. Women through the glass ceiling: gender asymmetries in Wikipedia. *EPJ Data Science* 5 (2016), 1–24. <https://doi.org/10.1140/epjds/s13688-016-0066-4>
  - [55] Judy Wajcman. 2004. *TechnoFeminism*. Polity Press, Cambridge, UK.
  - [56] Wikipedia. 2020. Countering Systematic Bias. [https://en.wikipedia.org/wiki/Wikipedia:Systemic\\_bias](https://en.wikipedia.org/wiki/Wikipedia:Systemic_bias)
  - [57] Sally Wyatt, Anna Harris, and Susan E Kelly. 2016. Controversy goes online: Schizophrenia genetics on Wikipedia. *Science & Technology Studies* 29, 1 (2016), 13–29. <https://doi.org/10.23987/sts.55407>
  - [58] Iris Marion Young. 1997. Feminism and the public sphere. *Constellations* 3, 3 (1997), 340–363. <https://doi.org/10.1111/j.1467-8675.1997.tb00064.x>