Accepted Tutorials at The Web Conference 2022

Riccardo Tommasini*
INSA Lyon - CRNS LIRIS
France

Senjuti Basu Roy*
New Jersey Institute of Technology
USA

Xuan Wang
University of Illinois at Urbana-Champaign
USA

Hongwei Wang
University of Illinois at Urbana-Champaign
USA

Heng Ji
University of Illinois at Urbana-Champaign
USA

Jiawei Han
University of Illinois at Urbana-Champaign
USA

Preslav Nakov
Qatar Computing Research Institute, HBKU, Doha
Qatar

Giovanni Da San Martino
University of Padova
Italy

Firoj Alam
Qatar Computing Research Institute
Qatar

Markus Schedl
Johannes Kepler University Linz
Austria

Elisabeth Lex
Graz University of Technology
Austria

Akash Bharadwaj
Meta AI
USA

Graham Cormode
Meta AI
USA

Milan Dojchinovski
DBpedia Association
Germany
FIT, Czech Technical University in Prague
Czech Republic

Jan Forberg
DBpedia Association
Germany

Johannes Frey
DBpedia Association
Germany

Pieter Bonte
Ghent University, IMEC, IDLab
Belgium

Marco Balduini
Quantia Consulting
Italy

Matteo Belcao
Quantia Consulting
Italy

Emanuele Della Valle
Politecnico di Milano, DEIB
Italy

Junliang Yu
The University of Queensland
Australia

Hongzhi Yin
The University of Queensland
Australia

Tong Chen
The University of Queensland
Australia

Haochen Liu
Michigan State University
USA

Yiqi Wang
Michigan State University
USA

Wenqi Fan
The Hong Kong Polytechnic University
Hong Kong

Xiaorui Liu
Michigan State University
USA

Jamell Dacon
Michigan State University
USA

Lingjuan Lyu
Sony AI
Japan

Jiliang Tang
Michigan State University
USA

Aristides Gionis
KTH Royal Institute of Technology
Sweden

Stefan Neumann
KTH Royal Institute of Technology
Sweden

Bruno Ordozgoiti
Queen Mary University of London
UK

Simon Razniewski
Max Planck Institute for Informatics
Germany

Hiba Arnaout
Max Planck Institute for Informatics
Germany

Shrestha Ghosh
Max Planck Institute for 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
Information Sciences Institute, University of Southern California
USA

Daniel Garijo
Universidad Politécnica de Madrid
Spain

Hans Chalupsky
Information Sciences Institute, University of Southern California
USA

Pedro Szekely
Information Sciences Institute, 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, VA
USA

Nikhil Rao
Amazon, CA, USA
USA

Karthik Subbian
Amazon, CA, USA
USA

Srinivasan Sengamedu
Amazon, CA, USA
USA

Chandan K. Reddy
Virginia Tech, VA
USA

Friedhelm Victor
Technical University of Berlin
Germany

Bernhard Haslhofer
AIT - Austrian Institute of Technology
Austria

George Katsogiannis-Meimarakis
Athena Research Center
Greece

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
Carnegie Mellon University
USA

Sachin Kumar
Carnegie Mellon University
USA

Xiangyu Zhao
City University of Hong Kong
Hong Kong

Bo Chen
Huawei Noah's Ark Lab
Hong Kong

Huifeng Guo
Huawei Noah's Ark Lab
Hong Kong

Yejing Wang
University of Science and Technology of China[19]
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
University of Science and Technology of China
China

Xiangnan He
University of Science and Technology of China
China

ABSTRACT

This paper summarizes the content of the 20 tutorials that have been given at The Web Conference 2022: 85% of these tutorials are lecture style, and 15% of these are hands on.

CCS CONCEPTS

• **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

KEYWORDS

tutorials, the web conference

ACM Reference Format:

Riccardo Tommasini, Senjuti Basu Roy, Xuan Wang, Hongwei Wang, Heng Ji, Jiawei Han, Preslav Nakov, Giovanni Da San Martino, Firoj Alam, Markus Schedl, Elisabeth Lex, Akash Bharadwaj, Graham Cormode, Milan Dojchinovski, Jan Forberg, Johannes Frey, Pieter Bonte, Marco Balduini, Matteo Belcao, Emanuele Della Valle, Junliang Yu, Hongzhi Yin, Tong Chen, Haochen Liu, Yiqi Wang, Wenqi Fan, Xiaorui Liu, Jamell Dacon, Lingjuan Lyu, Jiliang Tang, Aristides Gionis, Stefan Neumann, Bruno Ordozgoiti, Simon Razniewski, Hiba Arnaout, Shrestha Ghosh, Fabian Suchanek, Lingfei Wu, Yu Chen, Yunyao Li, Bang Liu, Filip Ilievski, Daniel Garijo, Hans Chalupsky, Pedro Szekely, Ilias Kanellos, Dimitris Sacharidis, Thanasis Vergoulis, Nurendra Choudhary, Nikhil Rao, Karthik Subbian, Srinivasan Sengamedu, Chandan K. Reddy, Friedhelm Victor, Bernhard Haslhofer, George Katsogiannis-Meimarakis, Georgia Koutrika, Shengmin Jin, Danai Koutra, Reza Zafarani, Yulia Tsvetkov, Vidhisha Balachandran, Sachin Kumar, Xiangyu Zhao, Bo Chen, Huifeng Guo, Yejing Wang, Ruiming Tang, Yang Zhang, Wenjie Wang, Peng Wu, Fuli Feng, and Xiangnan He. 2022. Accepted Tutorials at The Web Conference 2022. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3487553.3547182

1 INTRODUCTION

The Web Conference is pleased to host 20 tutorials for the 2022 edition chaired in Lyon, France.

In the attempt to foster interesting discussions as well as supporting the dissemination of prominent research areas, the tutorial chairs have selected high-quality contributions that cover a variety of topics. These range from fact-checking to natural language processing, deep learning, knowledge graphs, as well as the analysis of crypto assets. In particular, following the successful approach of previous years, this edition will host two tutorial formats:

- *lecture-style* (85%) tutorials will cover the state-of-the-art research, development, and applications in a specific web computing and related area, and stimulate and facilitate future work.
- *hands-on* (15%) tutorials will feature in-depth practical training on cutting-edge systems and tools of relevance to the web conference community. These sessions are targeted at novice as well as moderately skilled users

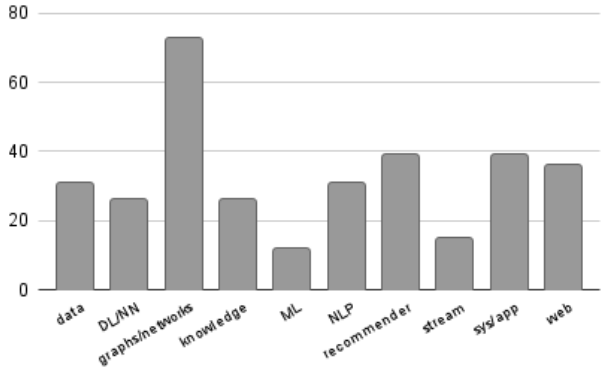

Figure 1: Top-10 Keywords in Tutorial Abstracts (clustered)

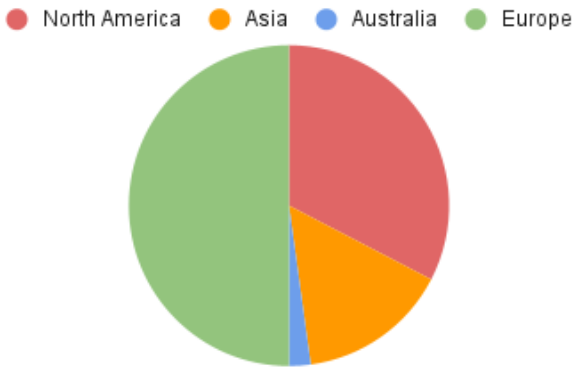

Figure 2: Geo-distribution of Authors' Affiliations

Figure 1 shows a qualitative analysis of the abstract of top-10 keywords. The general trend of TheWebConf 2022 tutorial goes around (knowledge) graphs and AI. Recommender systems and Natural Language Processing (NLP) remain to be important areas as well for the community.

Finally, the tutorial track of 2022 have received submissions from 39 unique institutions, 27 of which are universities. The majority of the submissions have come from Europe and North America, followed by Asia, and Australia. Notably, geographic distribution depicted in Figure 2 is based on the authors' affiliations at the time of submission. The tutorial track this year has received 28 submissions out of which 20 are accepted for presentation.

*Tutorial co-chair

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 ACM 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.

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

© 2022 Association for Computing Machinery.

ACM ISBN 978-1-4503-9130-6/22/04…$15.00
https://doi.org/10.1145/3487553.3547182

2 TUTORIAL 1: MODERN NATURAL LANGUAGE PROCESSING TECHNIQUES FOR SCIENTIFIC WEB MINING: TASKS, DATA, AND TOOLS

URL of the tutorial: https://xuanwang91.github.io/2021-12-19-www22-tutorial/
Type and Duration: lecture-style, 1.5 hours
Presenters name(s): Xuan Wang, Hongwei Wang, Heng Ji, and Jiawei Han
Abstract: This tutorial targets researchers and practitioners who are interested in natural language processing (NLP) technologies for scientific web mining. Exploring the vast amount of rapidly growing scientific literature available on the web is highly beneficial for scientific discovery. However, scientific web mining is particularly challenging due to the lack of specialized domain knowledge in natural language context, complex sentence structures in scientific writing, and multi-modal representations of scientific knowledge.

This tutorial presents a comprehensive overview of recent research and development on using NLP techniques for scientific web mining, focusing on the biomedical and chemistry domains. First, we introduce the motivation and unique challenges of web mining in the scientific domains. Then we discuss a set of methods that perform effective information extraction (named entity recognition, relation extraction, and event extraction), information retrieval (textual evidence retrieval, cross-modal molecule retrieval, and chemical reaction tracking) from scientific literature, and their applications on reaction prediction. Finally, we conclude our tutorial by demonstrating, on real-world datasets (COVID-19 and organic chemistry literature), how the information can be extracted and retrieved, and how they can assist further exploratory analysis. We also discuss the emerging research problems and future directions of using NLP techniques for scientific web mining.

3 TUTORIAL 2: FACT-CHECKING, FAKE NEWS, PROPAGANDA, AND MEDIA BIAS

URL of the tutorial: https://propaganda.math.unipd.it/www22-tutorial/
Type and Duration: lecture-style, 3 hours
Presenters name(s): Preslav Nakov, Giovanni Da San Martino and Firoj Alam
Abstract: Social media have democratized content creation and have made it easy for anybody to spread information online. However, stripping traditional media from their gate-keeping role has left the public unprotected against biased, deceptive and disinformative content, which could now travel online at breaking-news speed and influence major public events. The tutorial offers an overview of the emerging and interconnected research areas of fact-checking, disinformation, "fake news", propaganda, and media bias detection. We explore the general fact-checking pipeline and important elements thereof such as check-worthiness estimation, spotting previously fact-checked claims, stance detection, source reliability estimation, detection of persuasion techniques, and detecting malicious users in social media.

4 TUTORIAL 3: PSYCHOLOGY-INFORMED RECOMMENDER SYSTEMS: A HUMAN-CENTRIC PERSPECTIVE ON RECOMMENDER SYSTEMS

URL of the tutorial: https://socialcomplab.github.io/pirs-psychology-informed-recsys
Type and Duration: lecture-style, 3 hours
Presenters name(s): Markus Schedl and Elisabeth Lex
Abstract: Personalized recommender systems are essential tools to facilitate human decision making. Most state-of-the-art recommender systems use advanced machine learning techniques to model and predict user preferences from behavioral data. While such systems can provide useful recommendations, their algorithmic design does not incorporate the underlying psychological mechanisms that shape user preferences and behavior. In this interdisciplinary tutorial, we guide the attendees through the state-of-the-art in *psychology-informed recommender systems*, i.e., recommender systems that consider extrinsic and intrinsic human factors. We cover cognition-inspired, personality-aware, and affect-aware recommendation approaches; and we show how such systems can improve the recommendation process in a highly human-centric manner.

5 TUTORIAL 4: AN INTRODUCTION TO FEDERATED COMPUTATION

URL of the tutorial: https://akashb94.wixsite.com/website
Type and Duration: lecture-style, 1.5 hours
Presenters name(s): Akash Bharadwaj, Graham Cormode
Abstract: *Federated Computation* is an emerging area that seeks to provide stronger privacy for user data, by performing large scale, distributed computations where the data remains in the hands of users. Only the necessary summary information is shared, and additional security and privacy tools can be employed to provide strong guarantees of secrecy. The most prominent application of federated computation is in training machine learning models (federated learning), but many additional applications are emerging. This tutorial gives an overview of federated computation models and algorithms. It includes an introduction to security and privacy techniques and guarantees, and shows how they can be applied to solve a variety of distributed computations providing statistics and insights to distributed data. It also discusses the issues that arise when implementing systems to support federated computation, and open problems for future research.

6 TUTORIAL 5: DBPEDIA KNOWLEDGE GRAPH TUTORIAL

URL of the tutorial: https://www.dbpedia.org/events/tut-at-the-web-conf/
Type and Duration: lecture-style, 1.5 hours
Presenters name(s): Milan Dojchinovski, Jan Forberg, Johannes Frey
Abstract: DBpedia (https://www.dbpedia.org) is a crowd-sourced community effort to extract structured content from the information created in various Wikimedia projects. This structured information resembles an open knowledge graph, the DBpedia Knowledge

Graph, which is publicly available for everyone on the Web. The DBpedia Knowledge Graph has been under development for many years and is being improved to this day. In this tutorial, participants will gain basic information on the DBpedia Knowledge Graph and the DBpedia community. The tutorial will provide information on the complete DBpedia Knowledge Graph lifecycle, i.e. from extraction and modelling to publishing and maintenance of the DBpedia KG.

A particular focus will be put on the DBpedia Infrastructure, i.e. the DBpedia's Databus publishing platform and the associated DBpedia services (DBpedia Spotlight, DBpedia Lookup and the DBpedia service endpoints). The tutorial will also contain a session dedicated on the DBpedia's current motto, i.e. Global and Unified Access to Linked Data. In practical examples we will illustrate the potential and the benefit of using DBpedia in the context of the Web of Data.

7 TUTORIAL 6: WEB STREAM PROCESSING WITH RSP4J AND ONTOPSTREAM

URL of the tutorial: http://streamreasoning.org/events/web-stream-processing-tutorial-thewebconf
Type and Duration: hands-on, 3 hrs.
Presenters name(s): Pieter Bonte, Marco Balduini, Matteo Belcao and Emanuele Della Valle
Abstract: This tutorial provides a comprehensive introduction to Web stream processing, including some fundamental aspects of stream reasoning. Stream reasoning combines the ideas from stream processing with semantic web technologies, in order to process heterogeneous data streams as presented on the web. In particular, RDF Stream Processing (RSP) is a subfield of stream reasoning that focuses on the timely processing of streaming knowledge graphs in RDF format. Central to the tutorial is the recently published RSP4J library, which uniforms the interaction with existing RSP engines. Furthermore, the tutorial will detail OntopStream, a recent tool for Big Data processing of virtual streaming knowledge graphs, part of the Chimera suite library.

In practice, the tutorial will include (i) a survey on existing research outcomes from Stream Reasoning / RDF Stream Processing area, i.e., continuous querying, reactive reasoning over highly dynamic graph data; (ii) the introduction of the Stream Processing paradigm, for the processing streaming data, (iii) the introduction of Semantic Web technologies, for the integration of web data, (iv) an overview of how Stream Reasoning and RSP solve the challenges of web stream processing and processing of streaming knowledge graphs, (v) an introduction on how to include data that are not represented as knowledge graphs in web stream processing using virtual knowledge graphs, and (vi) the positioning of existing Web stream processing, e.g. RSP4J and OntopStream, to build and maintain Web Stream Processing applications. The tutorial will include several examples and exercises build around a relevant use case.

8 TUTORIAL 7: SELF-SUPERVISED LEARNING IN RECOMMENDATION: FUNDAMENTALS AND ADVANCES

URL of the tutorial: https://ssl-recsys.github.io/
Type and Duration: lecture-style, 1.5 hours
Presenters name(s): Junliang Yu, Hongzhi Yin, Tong Chen
Abstract: The neural architecture-based recommenders have demonstrated overwhelming advantages over their traditional counterparts. However, the highly sparse user behavioral data often bottlenecks deep neural recommendation models to take full advantage of their capacity for better performance. Recently, self-supervised learning (SSL), which can enable training on massive unlabeled data with automatic data annotation, has received tremendous attention across multiple fields including recommender systems. Self-supervised recommendation now becomes the latest trend, and it turns out that SSL can significantly improve the recommendation quality by designing pretext tasks to discover supervisory signals from the raw data, serving as a natural antidote to the data sparsity issue. In this tutorial, we will systematically introduce the methodologies of applying SSL to recommendation. The topics to be covered include: (1) foundation and overview of self-supervised recommendation; (2) a comprehensive taxonomy of existing SSL-driven recommendation methods which is constructed based on the characteristics of pretext tasks; (3) how to apply SSL to various recommendation scenarios where different types of data and multiple optimization objectives are involved; (4) limitations in current research and future research directions; (5) an opensource toolkit to facilitate empirical comparisons and methodological development of self-supervised recommendation methods.

9 TUTORIAL 8: TRUSTWORTHY AI: A COMPUTATIONAL PERSPECTIVE

URL of the tutorial: https://sites.google.com/msu.edu/trustworthy-ai/
Type and Duration: lecture-style, 3 hours
Presenters name(s): Haochen Liu, Yiqi Wang, Wenqi Fan, Xiaorui Liu, Jamell Dacon, Lingjuan Lyu and Jiliang Tang
Abstract: The past few decades have witnessed the rise of artificial intelligence (AI) technology. However, recent studies show evidence that AI algorithms may not be trustworthy. For example, they could be vulnerable to slight perturbations of input data; they could undermine fairness by showing bias and stereotypes towards certain groups of people; and their decisions could be hard to explain due to their opaque model architectures. With the widespread use of AI applications in our daily life, whether an AI algorithm is trustworthy or not has become a problem of great concern to researchers, developers and users. Recently, a great amount of research on trustworthy AI has emerged. In this tutorial, we aim to provide a comprehensive overview of the cutting-edge research progress on trustworthy AI from a computational perspective. Specifically, we focus on the six most important dimensions in realizing trustworthy AI: (i) Privacy, (ii) Safety & Robustness, (iii) Explainability (iv) Non-discrimination & Fairness, (v) Environmental Well-Being and (vi) Accountability & Auditability. We will introduce the latest technologies and real-world applications in each dimension according to a taxonomy, and discuss the accordant and conflicting interactions among various dimensions. Besides, we will discuss potential future research directions in this field. We expect that researchers and practitioners can gain a broad overview and a deep insight of trustworthy AI from this tutorial, so as to advance the progress of this field.

10 TUTORIAL 9: OPINION FORMATION IN SOCIAL NETWORKS: MODELS AND COMPUTATIONAL PROBLEMS

URL of the tutorial: https://sites.google.com/view/tutorial-opinion-formation/home
Type and Duration: lecture-style, 1.5 hours
Presenters names: Aristides Gionis, Stefan Neumann, Bruno Ordozgoiti
Abstract: Social networks are widely used nowadays to engage in conversations about a variety of topics. Over time, these discussions can have a significant impact on people's opinions. Works in sociology and other areas have provided mathematical models to simulate such opinion formation processes, and in the past decade it has become popular to consider the computational aspects of these models, motivated by the widespread use of online social networks. The goal is to obtain a better understanding of real-world phenomena, such as increasing polarization and filter bubbles. In this tutorial we aim to provide an overview of the opinion formation literature. We will present the most important opinion models that are studied in this context, and will discuss some of the computational challenges that have arisen recently. We will also reflect on emerging applications and directions for future work.

11 TUTORIAL 10: COMPLETENESS, RECALL, AND NEGATION IN OPEN-WORLD KNOWLEDGE BASES

URL of the tutorial: https://www.mpi-inf.mpg.de/www-2022-tutorial
Type and Duration: lecture-style, 1.5 hours
Presenters name(s): Simon Razniewski, Hiba Arnaout, Shrestha Ghosh, Fabian M. Suchanek
Abstract: General-purpose knowledge bases (KBs) are a cornerstone of the Semantic Web. Pragmatically constructed from available web sources, these KBs are far from complete, which poses a set of challenges in curation as well as consumption. In this tutorial, we discuss how completeness, recall and negation in KBs can be represented, extracted, and inferred. We proceed in 5 parts: (i) We introduce the logical foundations of knowledge representation and querying under partial closed-world semantics. (ii) We show how information about recall can be estimated via statistical patterns, and (iii) how it can be identified in KBs and in text. (iv) We show how salient negative knowledge can be identified, and (v) how recall can be targeted in a comparative notion.

12 TUTORIAL 11: DEEP LEARNING ON GRAPHS FOR NATURAL LANGUAGE PROCESSING

URL of the tutorial: https://tinyurl.com/dlg4nlp
Type and Duration: lecture-style, 3 hours
Presenters name(s): Lingfei Wu, Yu Chen, Heng Ji, Yunyao Li, Bang Liu
Abstract: There are a rich variety of NLP problems that can be best expressed with graph structures. Due to the great power in modeling non-Euclidean data like graphs, deep learning on graphs techniques (i.e., Graph Neural Networks (GNNs)) have opened a new door to solving challenging graph-related NLP problems, and have already achieved great success. Despite the success, deep learning on graphs for NLP (DLG4NLP) still faces many challenges (e.g., automatic graph construction, graph representation learning for complex graphs, learning mapping between complex data structures).

This tutorial will cover relevant and interesting topics on applying deep learning on graph techniques to NLP, including automatic graph construction for NLP, graph representation learning for NLP, advanced GNN based models (e.g., graph2seq and graph2tree) for NLP, and the applications of GNNs in various NLP tasks (e.g., machine translation, natural language generation, information extraction and semantic parsing). In addition, hands-on demonstration sessions will be included to help the audience gain practical experience on applying GNNs to solve challenging NLP problems using our recently developed open source library – Graph4NLP, the first library for researchers and practitioners for easy use of GNNs for various NLP tasks.

13 TUTORIAL 12: KGTK: TOOLS FOR CREATING AND EXPLOITING LARGE KNOWLEDGE GRAPHS

URL of the tutorial: https://usc-isi-i2.github.io/kgtk-tutorial-www-2022/
Type and Duration: hands-on, 3 hours
Presenters name(s): Filip Ilievski, Daniel Garijo, Hans Chalupsky, Pedro Szekely
Abstract: The Knowledge Graph Toolkit (KGTK) is a comprehensive framework for the creation and exploitation of large KGs, designed for simplicity, scalability, and interoperability. KGTK represents KGs in tab-separated (TSV) files with four columns: edge-identifier, head, edge-label, and tail. All KGTK commands consume and produce KGs represented in this simple format, so they can be composed into pipelines to perform complex transformations on KGs. The simplicity of its data model also allows KGTK operations to be easily integrated with existing tools, like Pandas or graph-tool. KGTK provides a suite of commands to import Wikidata, RDF (e.g., DBpedia), and popular graph representations into the KGTK format. A rich collection of transformation commands make it easy to clean, union, filter, and sort KGs, while the KGTK graph combination commands support efficient intersection, subtraction, and joining of large KGs. Its advanced functionality includes a query language variant of Cypher (called "Kypher"), which has been optimized for querying KGs stored on disk with minimal indexing overhead; graph analytics commands support scalable computation of centrality metrics such as PageRank, degrees, connected components, and shortest paths; lexicalization of graph nodes, and computation of multiple variants of text and graph embeddings over the whole graph. In addition, a suite of export commands supports the transformation of KGTK KGs into commonly used formats, including the Wikidata JSON format, RDF triples, JSON documents for ElasticSearch indexing, and graph-tool. Finally, KGTK allows browsing locally stored KGs using a variant of SQID; and includes a development environment using Jupyter notebooks that provides seamless integration with Pandas. KGTK can process Wikidata-sized KGs, with billions of edges, on a laptop computer. We have

used KGTK in multiple settings, focusing primarily on the construction of subgraphs of Wikidata, analysis of over 300 Wikidata dumps since the inception of the Wikidata project, linking tables to Wikidata, construction of a consolidated commonsense KG combining multiple existing sources, creation of an extension of Wikidata for food security, and creation of an extension of Wikidata for the pharmaceutical industry.

14 TUTORIAL 13: ASSESSING RESEARCH IMPACT BY LEVERAGING OPEN SCHOLARLY KNOWLEDGE GRAPHS

URL of the tutorial: https://athenarc.github.io/impact-assessment-tutorial/
Type and Duration: lecture-style, 1.5 hours
Presenters name(s): Ilias Kanellos, Dimitris Sacharidis, Thanasis Vergoulis
Abstract: Nowadays, the vast amount of existing published research works creates major obstacles regarding the traditional knowledge discovery required by common research processes and other relevant tasks. However, at the same time, the increased popularity of the Open Science movement, makes large amounts of scholarly metadata available through open Scholarly Knowledge Graphs (SKGs), paving the way for reliable research impact assessment processes, that can alleviate the aforementioned issue. The main objective of the tutorial is to inform and educate the audience about the opportunities and challenges that open SKGs create in the field of research impact assessment, presenting the respective state-of-the-art and highlighting common pitfalls.

15 TUTORIAL 14: HYPERBOLIC NEURAL NETWORKS: THEORY, ARCHITECTURES AND APPLICATIONS

URL of the tutorial: https://akirato.github.io/hyperbolic-networks-tutorial/
Type and Duration: lecture-style, 3 hours
Presenters name(s): Nurendra Choudhary, Nikhil Rao, Karthik Subbian, Srinivasan Sengamedu, Chandan K. Reddy
Abstract: Graphs are ubiquitous data-structures that are widely-used in a number of data storage scenarios, including social networks, recommender systems, knowledge graphs and e-commerce. This has led to a rise of GNN architectures to analyze and encode information from the graphs for better performance in downstream tasks. While preliminary research in the domain of graph analysis was driven by neural architectures, recent studies has revealed important properties unique to graph datasets such as hierarchies and global structures. This has driven research into hyperbolic space due to their ability to effectively encode the inherent hierarchy present in graph datasets. The research has also been subsequently applied to other domains such as NLP and computer vision to get formidable results. However, the major challenge to further growth is the obscurity of hyperbolic networks and a better comprehension of the necessary algebraic operations needed to broaden the application to different neural network architectures. In this tutorial, we aim to introduce researchers and practitioners in the web domain to the hyperbolic equivariants of the Euclidean operations that are necessary to tackle their application to neural network architectures. Additionally, we describe the popular hyperbolic variants of GNN architectures such as recurrent networks, convolution networks and attention networks and explain their implementation, in contrast to the Euclidean counterparts. Furthermore, we also motivate our tutorial through existing applications in the areas of graph analysis, knowledge graph reasoning, product search, NLP, and computer vision and compare the performance gains to the Euclidean counterparts.

16 TUTORIAL 15: CRYPTOASSET ANALYTICS: DATA, FUNDAMENTAL CONCEPTS, AND OPEN SOURCE TOOLS

URL of the tutorial: https://cryptoassetanalytics.net/tutorial/
Type and Duration: hands-on, 3 hours
Presenters name(s): Friedhelm Victor and Bernhard Haslhofer
Abstract: Cryptoassets are exchangeable virtual assets that use cryptography, represent an economic resource or value to someone, and are generally based on blockchain technology. From a socio-economic perspective, cryptoassets form a virtual ecosystem and represent a community of actors linked together through cryptoasset transfers. The goal of cryptoasset analytics is to develop and apply quantitative methods to understand the technical and socio-economic aspects of cryptoasset ecosystems. The tutorial provides hands-on experience on cutting-edge cryptoasset analytics systems and tools for analyzing cryptoasset ecosystems. It targets novice and moderately skilled users with a basic understanding of cryptocurrencies like Bitcoin and Ethereum, and assumes basic coding skills in Python. The tutorial will gently introduce the fundamental concepts behind cryptoassets, then focus on analyzing UTXO-model ledgers like Bitcoin and account-model ledgers like Ethereum, before providing an outlook on emerging trends like Decentralized Finance. Throughout the tutorial, attendees will work with open-source tools like GraphSense, ethereum-etl and Jupyter notebooks and work on data associated with real use cases.

17 TUTORIAL 16: DATA DEMOCRATISATION WITH DEEP LEARNING: AN ANALYSIS OF TEXT-TO-SQL SYSTEMS

URL of the tutorial: https://darelab.imsi.athenarc.gr/tutorials/text2sql_twc22/
Type and Duration: lecture-style, 1.5 hours
Presenters name(s): George Katsogiannis-Meimarakis, Georgia Koutrika
Abstract: In the age of the Digital Revolution, almost all human activities, from industrial and business operations to medical and academic research, are reliant on the constant integration and utilisation of ever-increasing volumes of data. However, the explosive volume and increasing complexity of data makes data querying all the more challenging even for experts. For this reason, numerous text-to-SQL systems have been developed that enable querying relational databases using natural language. The recent advances on deep neural networks along with the creation of two large datasets specifically made for training text-to-SQL systems, have paved the path for a novel and very promising research area. The purpose of

this tutorial is a deep dive into this area, covering state-of-the-art techniques for natural language representation in neural networks, benchmarks that sparked research and competition, recent text-to-SQL systems using deep learning techniques, as well as open problems and research opportunities.

18 TUTORIAL 17: INTERPRETABLE NETWORK REPRESENTATIONS

URL of the tutorial: https://shengminjin.github.io/tutorials/www2022
Type and Duration: lecture-style, 1.5 hours.
Presenters name(s): Shengmin Jin, Danai Koutra, Reza Zafarani.
Abstract: Networks (or interchangeably graphs) have been ubiquitous across the globe and within science and engineering: social networks, collaboration networks, protein-protein interaction networks, infrastructure networks, among many others. Machine learning on graphs, especially network representation learning, has shown remarkable performance in tasks related to graphs, such as node/graph classification, graph clustering, and link prediction. These tasks are closely related to the Web applications, especially social network analysis and recommendation systems. For example, node classification and graph clustering are widely used for studies on community detection, and link prediction plays a vital role in friend or item recommendation. Like performance, it is equally crucial for individuals to understand the behavior of machine learning models and be able to explain how these models arrive at a certain decision. Such needs have motivated many studies on interpretability in machine learning. Specifically, for social network analysis, we may need to know the reasons why certain users (or groups) are classified or clustered together by the machine learning models, or why a friend recommendation system considers some users similar so that they are recommended to connect with each other. Under such circumstances, an interpretable network representation is necessary and it should carry the graph information to a level understandable by humans.

In this tutorial, we will (1) define interpretability and go over its definitions within different contexts in studies of networks; (2) review and summarize various interpretable network representations; (3) discuss connections to network embedding, graph summarization, and network visualization methods; (4) discuss explainability in Graph Neural Networks, as such techniques are often perceived to have limited interpretability; and (5) highlight the open research problems and future research directions. The tutorial is designed for researchers, graduate students, and practitioners in areas such as graph mining, machine learning on graphs, and machine learning interpretability. Few prerequisites are required for "The Web Conference" participants to attend.

19 TUTORIAL 18: MITIGATING SOCIETAL HARMS OF LARGE LANGUAGE MODELS

URL of the tutorial: https://sites.google.com/view/societal-impacts-lm-tutorial
Type and Duration: lecture-style, 1.5 hours
Presenters name(s): Yulia Tsvetkov, Vidhisha Balachandran, Sachin Kumar
Abstract: Recent advancement in large language models have led to remarkable improvements in the capabilities of Natural Language Processing (NLP) models and resulted in an increasing adoption of language technologies in user-facing products and critical applications. However, several recent studies have highlighted their potential harms on people and society and proposed different solutions to mitigate them. With increasing deployment of model driven NLP tools, there is an urgent need to equip researchers and practitioners on knowledge about such social harms and methods and techniques to control models from adversely impacting people. However, this body of work is so far lacking a common framework and methodology. This tutorial aims to fill this gap.

This tutorial will provide an overview of potential social issues related to language technologies like bias, fairness and misinformation and methods to identify and mitigate them. While recent workshops and tutorials focus on highlighting societal harms of NLP tools and technologies, work on how to systematically identify them and research on techniques to eliminate them are not highlighted enough. This tutorial will briefly cover the potential harms of such technologies and focus more of techniques for evaluating and mitigating them.

20 TUTORIAL 19: AUTOMATED MACHINE LEARNING FOR RECOMMENDATIONS: FUNDAMENTALS AND ADVANCES

URL of the tutorial: https://advanced-recommender-systems.github.io/AutoML-Recommendations/
Type and Duration: lecture-style, 1.5 hours
Presenters name(s): Xiangyu Zhao, Wenqi Fan, Bo Chen, Ruiming Tang
Abstract: Recommender systems have become increasingly important in our daily lives since they play an important role in mitigating the information overload problem, especially in many user-oriented online services. Recommender systems aim to identify a set of objects (i.e., items) that best match users' explicit or implicit preferences, by utilizing the user and item interactions to improve the matching accuracy. With the fast advancement of deep neural networks (DNNs) in the past few decades, recommendation techniques have achieved promising performance. However, we still meet three inherent challenges to design deep recommender systems (DRS): 1) the majority of existing DRS are developed based on hand-crafted components, which requires ample expert knowledge of machine learning and recommender systems; 2) human error and bias can lead to suboptimal components, which reduces the recommendation effectiveness; 3) non-trivial time and engineering efforts are usually required to design the task-specific components in different recommendation scenarios.

In this tutorial, we aim to give a comprehensive survey on the recent progress of advanced Automated Machine Learning (AutoML) techniques for solving the above problems in deep recommender systems. More specifically, we will present feature embedding search, feature interaction search, and whole DRS pipeline search for the design of deep recommender systems. In this way, we expect academic researchers and industrial practitioners in related fields can get deep understanding and accurate insight into the spaces, stimulate more ideas and discussions, and promote developments of technologies in recommendations. A video teaser is made for the proposed tutorial. It can be found in https://bit.ly/3kxng8D.

21 TUTORIAL 20: CAUSAL RECOMMENDATION: PROGRESSES AND FUTURE DIRECTIONS

URL of the tutorial: https://causalrec.github.io/
Type and Duration: lecture-style, 3 hours.
Presenters name(s): Yang Zhang, Wenjie Wang, Peng Wu, Fuli Feng, Xiangnan He.
Abstract: Data-driven recommender systems have demonstrated great successes in various Web applications owing to the extraordinary ability of machine learning models to recognize patterns (*i.e.,* correlation) from the massive historical user behaviors. However, these models still suffer from several issues such as biases and unfairness due to spurious correlations. Considering the causal mechanism behind data can avoid the influences of spurious correlations brought by non-causal relations. In this light, embracing causal recommendation modeling is an exciting and promising direction. Therefore, causal recommendation is increasingly drawing attention in our recommendation community. Nevertheless, there lacks a systemic overview of this topic, leading to difficulties for researchers and practitioners to understand and keep up with this direction.

In this tutorial, we aim to introduce the key concepts in causality and provide a systemic review of existing work on causal recommendation. We will introduce existing methods from two different causal frameworks — the potential outcome framework and the structural causal model. We will give examples and discussions regarding how to utilize different causal tools under these two frameworks to model and solve problems in recommendation. A comparison between the two lines of work will be provided to facilitate understanding the differences and connections between them. Besides, we identify some open challenges and potential future directions for this area. We hope this tutorial could stimulate more ideas on this topic and facilitate the development of causality-aware recommender systems.