

TextGraphs-11

**Graph-Based Methods for
Natural Language Processing**

Proceedings of the Workshop

August 3, 2017
Vancouver, Canada

©2017 The Association for Computational Linguistics

Order copies of this and other ACL proceedings from:

Association for Computational Linguistics (ACL)
209 N. Eighth Street
Stroudsburg, PA 18360
USA
Tel: +1-570-476-8006
Fax: +1-570-476-0860
acl@aclweb.org

ISBN 978-1-945626-60-9

Introduction to TextGraphs-11

Welcome to TextGraphs, the workshop on Graph-based Methods for Natural Language Processing. The eleventh edition of the workshop is being organized on August 3, 2017, in conjunction with the 55th Annual Meeting of the Association for Computational Linguistics (ACL-2017), being held in Vancouver in Canada.

For the past eleven years, the workshops in the TextGraphs series have published and promoted the synergy between the field of Graph Theory (GT) and Natural Language Processing (NLP). The target audience of our workshop has comprised of researchers working on problems related to either Graph Theory or graph-based algorithms applied to Natural Language Processing, social media, and the Semantic Web.

The TextGraphs workshop series addresses a broad spectrum of research areas within NLP. This is because, besides traditional NLP applications like parsing, word sense disambiguation, semantic role labeling, and information extraction, graph-based solutions also target web-scale applications like information propagation in social networks, rumor proliferation, e-reputation, language dynamics learning, and future events prediction, to name a few. Following this tradition, this year's TextGraphs also presents research from diverse topics such as semantics, word embeddings, text coherence, multi-lingual applications and summarization.

Previous editions of TextGraphs have featured special themes, such as "Cognitive and Social Dynamics of Languages in the framework of Complex Networks" and "Large Scale Lexical Acquisition and Representation". For TextGraphs 2017, we set a special focus on the usage of graph-based methods to interpret deep learning models for NLP tasks. Though deep learning models have displayed state-of-the-art performance on many NLP tasks, they are often criticized for not being interpretable (due to their various layers and large number of parameters). Through our theme, we hoped to spur a discussion on the development of methods for reasoning and interpretation of the layers used in deep learning models, given that a neural network is, from one point of view, nothing but a graph.

We are pleased to have two excellent invited speakers for this year's event. We thank Apoorv Agarwal and Michael Strube for their enthusiastic acceptance of our invitation. We also thank Verisk for sponsoring an invited speaker and the best paper award. Finally, we are thankful to the members of the program committee for their valuable and high quality reviews. All submissions have benefited from their expert feedback. Their timely contribution was the basis for accepting an excellent list of papers and making this edition of TextGraphs a success.

Martin Riedl, Swapna Somasundaran, Goran Glavaš and Ed Hovy
TextGraphs-11 Organizers
July 2017

Organizers:

Martin Riedl, Universität Hamburg, Germany
Swapna Somasundaran, Educational Testing Service, Princeton, USA
Goran Glavaš, University of Mannheim, Germany
Eduard Hovy, Carnegie Mellon University, USA

Program Committee:

Alan Akbik, Zalando, Germany
Sivaji Bandyopadhyay, Jadavpur University, Kolkata, India
Chris Biemann, Universität Hamburg, Germany
Pushpak Bhattacharyya, IIT Bombay, India
Tomáš Brychcín, University of West Bohemia, Czech Republic
Tanmoy Chakraborty, University of Maryland, USA
Monojit Choudhury, Microsoft Research, India
Asif Ekbar, Indian Institute of Technology, Patna, India
Stefano Faralli, University of Mannheim, Germany
Michael Flor, Educational Testing Services, USA
Marc Franco Salvador, University of Valencia, Spain
Tomáš Hercig, University of West Bohemia, Czech Republic
Ioana Hulpus, University of Mannheim, Germany
Roman Klinger, University of Stuttgart, Germany
Nikola Ljubešić, University of Zagreb, Croatia
Héctor Martínez Alonso, Inria & University Paris Diderot, France
Gabor Melli, VigLink, USA
Rada Mihalcea, University of Michigan, USA
Alessandro Moschitti, University of Trento, Italy
Animesh Mukherjee, IIT Kharagpur, India
Vivi Nastase, Fondazione Bruno Kessler, Italy
Roberto Navigli, “La Sapienza” University of Rome, Italy
Alexander Panchenko, Universität Hamburg, Germany
Simone Paolo Ponzetto, University of Mannheim, Germany
Steffen Remus, Universität Hamburg, Germany
Stephan Roller, UT Austin, USA
Shourya Roy, Xerox Research, India
Josef Steinberger, University of West Bohemia, Czech Republic
Anders Søgaard, University of Copenhagen, Denmark
Jan Šnajder, University of Zagreb, Croatia
Kateryna Tymosenko, University of Trento, Italy
Aline Villavicencio, F. University of Rio Grande do Sul, Brazil
Ivan Vulić, University of Cambridge, United Kingdom
Fabio Massimo Zanzotto, “Tor Vergata” University of Rome, Italy

Invited Speakers:

Apoorv Agarwal, Columbia University, USA
Michael Strube, HITS gGmbH, Heidelberg, Germany

Sponsor:

Verisk Analytics

Table of Contents

On the "Calligraphy" of Books

Vanessa Queiroz Marinho, Henrique Ferraz de Arruda, Thales Sinelli, Luciano da Fontoura Costa and Diego Raphael Amancio 1

Adapting predominant and novel sense discovery algorithms for identifying corpus-specific sense differences

Binny Mathew, Suman Kalyan Maity, Pratip Sarkar, Animesh Mukherjee and Pawan Goyal 11

Merging knowledge bases in different languages

Jerónimo Hernández-González, Estevam R. Hruschka Jr. and Tom M. Mitchell 21

Parameter Free Hierarchical Graph-Based Clustering for Analyzing Continuous Word Embeddings

Thomas Alexander Trost and Dietrich Klakow 30

Spectral Graph-Based Method of Multimodal Word Embedding

Kazuki Fukui, Takamasa Oshikiri and Hidetoshi Shimodaira 39

Graph Methods for Multilingual FrameNets

Collin Baker and Michael Ellsworth 45

Extract with Order for Coherent Multi-Document Summarization

Mir Tafseer Nayeem and Yllias Chali 51

Work Hard, Play Hard: Email Classification on the Avocado and Enron Corpora

Sakhar Alkhereyf and Owen Rambow 57

A Graph Based Semi-Supervised Approach for Analysis of Derivational Nouns in Sanskrit

Amrith Krishna, Pavankumar Satuluri, Harshavardhan Ponnada, Muneeb Ahmed, Gulab Arora, Kaustubh Hiware and Pawan Goyal 66

Evaluating text coherence based on semantic similarity graph

Jan Wira Gotama Putra and Takenobu Tokunaga 76

Conference Program

Thursday, August 3, 2017

- 9:00–9:10 *Opening remarks*
Swapna Somasundaran and Goran Glavaš
- 9:10–10:10 *Invited talk: How communication networks inform interpretation of language*
Apoorv Agarwal
- 10:10–10:30 *On the "Calligraphy" of Books*
Vanessa Queiroz Marinho, Henrique Ferraz de Arruda, Thales Sinelli, Luciano da Fontoura Costa and Diego Raphael Amancio
- 10:30–11:00 *Coffee break***
- 11:00–11:20 *Adapting predominant and novel sense discovery algorithms for identifying corpus-specific sense differences*
Binny Mathew, Suman Kalyan Maity, Pratip Sarkar, Animesh Mukherjee and Pawan Goyal
- 11:20–11:40 *Merging knowledge bases in different languages*
Jerónimo Hernández-González, Estevam R. Hruschka Jr. and Tom M. Mitchell
- 11:40–12:00 *Parameter Free Hierarchical Graph-Based Clustering for Analyzing Continuous Word Embeddings*
Thomas Alexander Trost and Dietrich Klakow
- 12:00–12:15 *Spectral Graph-Based Method of Multimodal Word Embedding*
Kazuki Fukui, Takamasa Oshikiri and Hidetoshi Shimodaira
- 12:15–14:00 *Lunch***
- 14:00–15:00 *Invited talk*
Michael Strube
- 15:00–15:15 *Graph Methods for Multilingual FrameNets*
Collin Baker and Michael Ellsworth
- 15:15–15:30 *Extract with Order for Coherent Multi-Document Summarization*
Mir Tafseer Nayeem and Yllias Chali

Thursday, August 3, 2017 (continued)

15:30–16:00 *Coffee break*

16:00–16:20 *Work Hard, Play Hard: Email Classification on the Avocado and Enron Corpora*
Sakhar Alkhereyf and Owen Rambow

16:20–16:40 *A Graph Based Semi-Supervised Approach for Analysis of Derivational Nouns in Sanskrit*
Amrith Krishna, Pavankumar Satuluri, Harshavardhan Ponnada, Muneeb Ahmed,
Gulab Arora, Kaustubh Hiware and Pawan Goyal

16:40–17:00 *Evaluating text coherence based on semantic similarity graph*
Jan Wira Gotama Putra and Takenobu Tokunaga

17:00–17:10 *Best paper award and closing remarks*
Swapna Somasundaran and Goran Glavaš