

NAACL HLT 2018

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

Proceedings of the Twelfth Workshop

June 6, 2018
New Orleans, Louisiana

©2018 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 978-1-948087-25-4

Introduction

Welcome to TextGraphs, the Workshop on Graph-Based Methods for Natural Language Processing. The twelfth edition of our workshop is being organized on June 6, 2018, in conjunction with the 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, being held in New Orleans, Louisiana, USA.

The workshops in the TextGraphs series have published and promoted the synergy between the field of Graph Theory (GT) and Natural Language Processing (NLP) for over a decade. The target audience of our workshop comprises 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.

TextGraphs 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. Following this tradition, this year's TextGraphs also presents research from diverse topics such as lexical and computational semantics, text clustering and classification, and text compression and summarization, to name a few.

The selection process was competitive – we received 17 submissions (9 long and 8 short submissions) and accepted for presentation 8 of them (5 long and 3 short papers), resulting in the overall acceptance rate of 47%.

We are pleased to have two excellent invited speakers for this year's event. We thank Gabor Melli and Maximilian Nickel for their enthusiastic acceptance of our invitation. 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 the twelfth edition of TextGraphs a success.

Goran Glavaš, Swapna Somasundaran, Martin Riedl, and Ed Hovy
TextGraphs-12 Organizers
April 2018

Organizers:

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

Program Committee:

Željko Agić, IT University Copenhagen, Denmark
Chris Biemann, University of Hamburg, Germany
Tomáš Brychcín, University of West Bohemia, Czech Republic
Flavio Massimiliano Cecchini, University of Milano-Bicocca, Italy
Tanmoy Chakraborty, IIT Delhi, India
Monojit Choudhury, Microsoft Research Lab, India
Stefano Faralli, Università degli Studi di Roma Unitelma Sapienza, Italy
Michael Flor, Educational Testing Service, USA
Marc Franco Salvador, Symanto Group, Germany
Carlos Gomez-Rodriguez, University of A. Coruña, Spain
Binod Gyawali, Educational Testing Service, USA
Tomáš Hercig, University of West Bohemia, Czech Republic
Anne Lauscher, University of Mannheim, Germany
Nikola Ljubešić, University of Zagreb, Croatia
Héctor Martínez Alonso, Inria & University Paris Diderot, France
Gabor Melli, PlayStation Network, USA
Mohsen Mesgar, Heidelberg Institute for Theoretical Studies, Germany
Rada Mihalcea, University of Michigan, USA
Alessandro Moschitti, University of Trento, Italy
Animesh Mukherjee, IIT Kharagpur, India
Matthew Mulholland, Educational Testing Service, USA
Federico Nanni, University of Mannheim, Germany
Vivi Nastase, University of Heidelberg, Germany
Alexander Panchenko, University of Hamburg, Germany
Simone Paolo Ponzetto, University of Mannheim, Germany
Jan Wira Gotama Putra, Tokyo Institute of Technology, Japan
Steffen Remus, Universität Hamburg, Germany
Brian Riordan, Educational Testing Services, USA
Stephan Roller, UT Austin, USA
Natalie Schluter, IT University Copenhagen, Denmark
Khalil Sima'an, University of Amsterdam, Netherlands
Josef Steinberger, University of West Bohemia, Czech Republic
Kateryna Tymosenko, University of Trento, Italy
Dmitry Ustalov, University of Mannheim, Germany
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
Xiang Zhao, National University of Defense Technology, China

Invited Speakers:

Gabor Melli, Playstation Network, USA

Maximilian Nickel, Massachusetts Institute of Technology, USA

Table of Contents

<i>Scientific Discovery as Link Prediction in Influence and Citation Graphs</i> Fan Luo, Marco A. Valenzuela-Escárcega, Gus Hahn-Powell and Mihai Surdeanu	1
<i>Efficient Generation and Processing of Word Co-occurrence Networks Using corpus2graph</i> Zheng Zhang, Pierre Zweigenbaum and Ruiqing Yin	7
<i>Multi-hop Inference for Sentence-level TextGraphs: How Challenging is Meaningfully Combining Information for Science Question Answering?</i> Peter Jansen	12
<i>Multi-Sentence Compression with Word Vertex-Labeled Graphs and Integer Linear Programming</i> Elvys Linhares Pontes, Stéphane Huet, Thiago Gouveia da Silva, Andréa carneiro Linhares and Juan-Manuel Torres-Moreno	18
<i>Large-scale spectral clustering using diffusion coordinates on landmark-based bipartite graphs</i> Khiem Pham and Guangliang Chen	28
<i>Efficient Graph-based Word Sense Induction by Distributional Inclusion Vector Embeddings</i> Haw-Shiuan Chang, Amol Agrawal, Ananya Ganesh, Anirudha Desai, Vinayak Mathur, Alfred Hough and Andrew McCallum	38
<i>Fusing Document, Collection and Label Graph-based Representations with Word Embeddings for Text Classification</i> Konstantinos Skianis, Fragkiskos Malliaros and Michalis Vazirgiannis	49
<i>Embedding Text in Hyperbolic Spaces</i> Bhuwan Dhingra, Christopher Shallue, Mohammad Norouzi, Andrew Dai and George Dahl	59

Conference Program

Wednesday, June 6, 2018

9:00–9:10 *Opening Remarks*

Session 1

9:10–10:10 *Invited talk: Neural domain-specific wikification*
Gabor Melli

10:10–10:30 *Scientific Discovery as Link Prediction in Influence and Citation Graphs*
Fan Luo, Marco A. Valenzuela-Escárcega, Gus Hahn-Powell and Mihai Surdeanu

10:30–11:00 *Coffee break*

Session 2

11:00–11:20 *Efficient Generation and Processing of Word Co-occurrence Networks Using corpus2graph*
Zheng Zhang, Pierre Zweigenbaum and Ruiqing Yin

11:20–11:40 *Multi-hop Inference for Sentence-level TextGraphs: How Challenging is Meaningfully Combining Information for Science Question Answering?*
Peter Jansen

11:40–12:05 *Multi-Sentence Compression with Word Vertex-Labeled Graphs and Integer Linear Programming*
Elvys Linhares Pontes, Stéphane Huet, Thiago Gouveia da Silva, Andréa carneiro Linhares and Juan-Manuel Torres-Moreno

12:05–14:05 *Lunch break*

Wednesday, June 6, 2018 (continued)

Session 3

- 14:05–15:05 *Invited talk: Hierarchical Representation Learning on Graphs*
Maximilian Nickel
- 15:05–15:30 *Large-scale spectral clustering using diffusion coordinates on landmark-based bipartite graphs*
Khiem Pham and Guangliang Chen
- 15:30–16:00 *Coffee break*

Session 4

- 16:00–16:25 *Efficient Graph-based Word Sense Induction by Distributional Inclusion Vector Embeddings*
Haw-Shiuan Chang, Amol Agrawal, Ananya Ganesh, Anirudha Desai, Vinayak Mathur, Alfred Hough and Andrew McCallum
- 16:25–16:50 *Fusing Document, Collection and Label Graph-based Representations with Word Embeddings for Text Classification*
Konstantinos Skianis, Fragkiskos Malliaros and Michalis Vazirgiannis
- 16:50–17:15 *Embedding Text in Hyperbolic Spaces*
Bhuwan Dhingra, Christopher Shallue, Mohammad Norouzi, Andrew Dai and George Dahl
- 17:15–17.30 *Closing remarks*