



## Institute colloquium

# Word and graph embeddings for machine learning

**Steven Skiena**

Stony Brook University

Host: Herbert Edelsbrunner

Distributed word embeddings (e.g. word2vec) provide a powerful way to reduce large text corpora to concise features (vectors) readily applicable to a variety of problems in NLP and data science. I will introduce word embeddings, and apply them in variety of new and interesting directions, including:(1) Detecting Historical Shifts in Word Meaning -- Words like "gay" and "mouse" have substantially shifted their meanings over time in response to societal and technological changes. We use word embeddings trained over texts drawn from different time periods to detect changes in word meanings. This is part of our efforts in historical trends analysis.(2) Feature Extraction from Graphs -- We present DeepWalk, our approach for learning latent representations of vertices in a network, which has become extremely popular. DeepWalk uses local information on truncated random walks to learn embeddings, by treating walks as the equivalent of sentences in a language. It is suitable for a broad class of applications such as network classification and anomaly detection. We also introduce new graph embedding techniques based on random projections, which produce DeepWalk-quality embeddings thousands of times faster than previous algorithms.(3) Processes for Language and Knowledge Creation -- Can we uncover principles suggesting how vocabularies and other cultural concepts evolve, by studying the structure of their embedding spaces? We show that generative processes like preferential placement create point sets with properties suggestive of word embeddings.=====Biography: Steven Skiena is Distinguished Teaching Professor of Computer Science and Director of the Institute for AI-Driven Discovery and Innovation at Stony Brook University. His research interests include data science, bioinformatics, and algorithms. He is the author of six books, including "The Algorithm Design Manual", "The Data Science Design Manual", and "Who's Bigger: Where Historical Figures Really Rank", and over 150 technical papers. Skiena received his B.S. in Computer Science from the University of Virginia and his Ph.D. in Computer Science from the University of Illinois under Herbert Edelsbrunner in 1988. He is a Fellow of the American Association for the Advancement of Science (AAAS), a current and former Fulbright scholar, and recipient of the University of Virginia Engineering Distinguished Alumni Award (WahooWa!), the ONR Young Investigator Award and the IEEE Computer Science and Engineer Teaching Award. More info is available at <http://www.cs.stonybrook.edu/~skiena/>.

**Monday, June 19, 2023 11:30am - 12:30pm**

Raiffeisen Lecture Hall

---



This invitation is valid as a ticket for the ISTA Shuttle from and to Heiligenstadt Station.

Please find a schedule of the ISTA Shuttle on our webpage:

<https://ista.ac.at/en/campus/how-to-get-here/> The ISTA Shuttle bus is marked ISTA Shuttle (#142) and has the Institute Logo printed on the side.

[www.ista.ac.at](http://www.ista.ac.at) | Institute of Science and Technology Austria | Am Campus 1 | 3400 Klosterneuburg