AITeGraD 2020-21

Advanced AI for Text and Graphs

Prof. M. Vazirgiannis

Data Science and Mining group (DaSciM)@LIX,

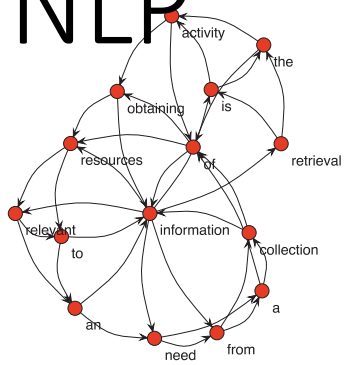
Ecole Polytechnique

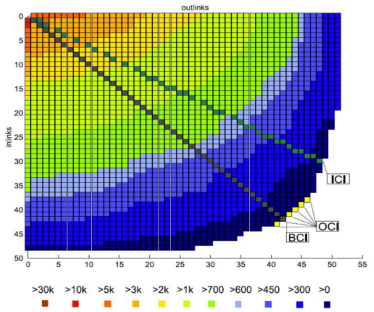
mvazirg@lix.polytechnique.fr

Scholar: https://tinyurl.com/yash58y6

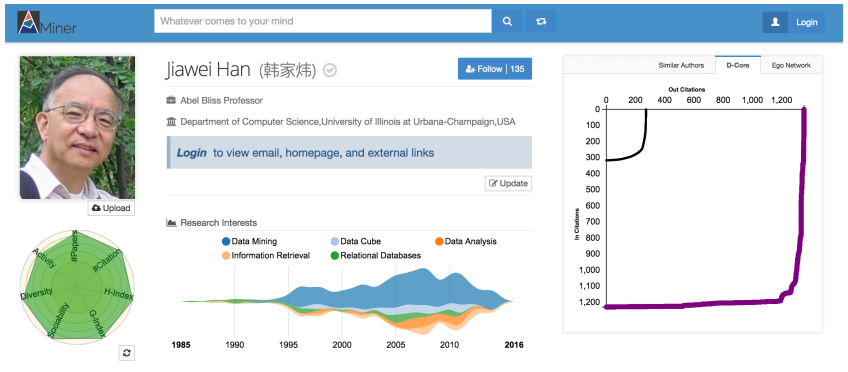
Tweeter: @mvazirg

DaSciM@X - ML for Graphs and NLP





- Graph Mining
  - Graph Degeneracy
  - Graph of Words – best paper ACM CIKM 2013
  - D-core metric for academic impact – adopted by Aminer.org
- Graph similarity
  - Kernels and DL – distinguished paper award IJCAI 2018
  - Influence Maximization – Nature/Scientific reports 2016
- Deep Learning for Graphs and Text data
  - Node embeddings – Social Nets
  - Graph classification (GNNs)
  - Unsupervised learning for graphs (Graph autoencoders)
  - Network architecture search - interpretability



- Strong relationships to
  - Industry (Google, Tencent, BNP, Airbus, AXA, MAIF, Tradelab, Deezer, Ericsson…)
  - Academia (Tsinghua, Jiaotong, KTH, Columbia, NTUA Athens)

ALTEGRAD (since 2014)

- Objectives
  - Provide state of the art research results and hands on experience
    - Text Mining NLP including DL methods
      - Categorization, opinion mining
      - Event detection in twitter
      - Keyword extraction, automated summarization, recommendations
    - Machine/Deep learning for graphs including
      - Community detection algorithms
      - Graph degeneracy for community detection
      - Deep Learning for node/edge/(sub)graph classification (GNNs)
      - Applications for social networks, biology, chemo-informatics, time series/finance

ALTEGRAD Syllabus 2020

TEXT/NLP – Graph based Text Mining
- Graph of Words - GoWvis
- Keyword extraction (TFIDF, TextRank, ECIR'15, EMNLP'16)
- extractive summarization (EMNLP'17)
- Sub-event detection in twitter streams (ICWSM'17)

- graph based document classification: TW-IDF (ASONAM'15), TW-ICW, subgraphs (ACL'15)
- abstractive summarization - ACL 2018 summarization

TEXT – NLP - Word & doc embeddings (P)
- Word embeddings: word2vec-glove models, doc2vec, subword, Latent Semantic Indexing, context based embeddings
- doc similarity metrics: Word Mover's distance, shortest path kernels (EMNLP16)

Deep learning for NLP
- CNNs, RNNs LSTMs for NLP, text classification

Meta-architectures
- Sequence to Sequence: Attention (HAN),
  Domains: summarization.
  Translation, image captioning
- Unsupervised word sense detection/disambiguation

ALTEGRAD Syllabus 2020

Graph kernels, community detection
Grakel python library - https://github.com/ysig/GraKeL/tree/develop.

Deep Learning for Graphs – node classification
- node embeddings (deepwalk & node2vec) for node classification and link prediction
- Supervised node embeddings (GCNN, …)

Deep Learning for Graphs – Graph classification
- graph CNNs
- message passing
- **Graph - Auto-encoders**

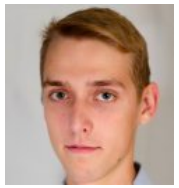
Sets embeddings – point clouds

Network Architecture Search – interpretability.

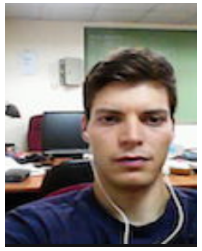
ALTEGRAD Team

- M. Vazirgiannis – Prof, DaSciM leader, LIX@X

- M. Kamal Eddine – PhD student in LIX – DL for NLP
- H. Abdine – PhD student in LIX – DL for NLP



- Dr. J. Lutzeyer – DL for Graphs
- Dr. G. Nikolentzos - DL for Graphs



best paper award in IJCAI 2018

ALTEGRAD Course format and logistics

- **7 sessions x 4 hours**
- **2h Lecture + 2h Lab**
- Data challenge (1 month …)

Evaluation
20% lab assignments
80% data challenge performance
(report/creativity/leaderboard score/)

Moodle: http://moodle.lix.polytechnique.fr/moodle/
Guest access: ALTEGRAD2020

VERY IMPORTANT: Register/enroll at:
https://tinyurl.com/ycsp2wcs
- get access to the teaching / lab material
- Receive our announcements

Schedule

17, 24 Nov,
1,8, 15 Dec,
12, 19 Jan 2021.
always **14:00 - 18:00**

Synchronous video classes
ZOOM Link: Topic: ALTEGRAD 2020 - 21
Join Zoom Meeting
https://ecolepolytechnique.zoom.us/j/89007143854?pwd=ZjAzcnFpMU9naTJaQjdGSDlqOG1vUT09
Meeting ID: 890 0714 3854
Passcode: 969359

SLACK channel for collaboration/messages:
https://join.slack.com/t/lixecolepolytechnique/shared_invite/zt-jczsdegh-bydxKK_x8q2dFFPep3MmJw

ALTEGRAD - Why choose this course

- State of the art AI ML/DL methods and software for the dominant data formats: Graphs, Text/NLP
- Acquire practical experience with large scale relevant problems
- Awesome applications: NLP, fraud detection, social media, Web, timeseries/financial
- Research Internship and/or PhD with DaSciM

- *Register/enroll at*: **https://tinyurl.com/ycsp2wcs**

THANK YOU!!

Michalis Vazirgiannis, mvazirg@lix.polytechnique.fr
Scholar: **https://tinyurl.com/yash58y6**
Tweeter: @mvazirg
Data Science and Mining Group@LIX. X
http://www.lix.polytechnique.fr/dascim/