SciTo trends: visualising scientific topic trends

Serafeim Chatzopoulos[1,2], Panagiotis Deligiannis[1], Thanasis Vergoulis[2], Ilias Kanellos[2,3], Christos Tryfonopoulos[1], Theodore Dalamagas[2]

[1]Univ. of the Peloponnese, Dep. of Informatics & Tel/tions, Tripoli, Greece
[2]IMSI, "Athena" Research & Innovation Center, Athens, Greece
[3]NTUA, School of Electrical & Computer Engineering, Athens, Greece

ATHENA Research & Innovation Information Technologies

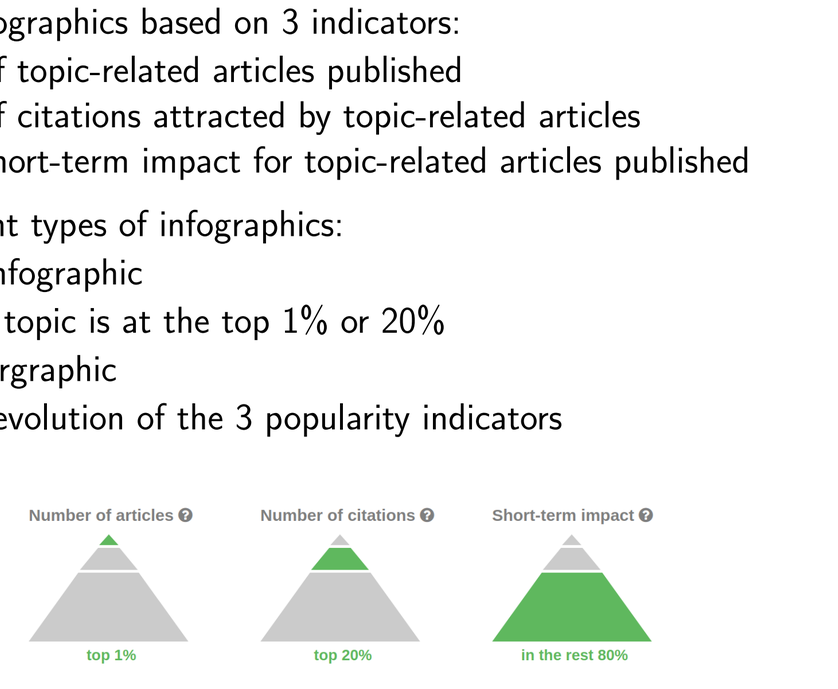
Data collection and processing

- **Article citation data**
  - OpenCitations COCI dataset[1]
  - $\sim$ **450***M* citations for $>$ **45***M* articles
- **Article impact scores**
  - Citation counts (overall impact)
  - RAM [1] scores (short-term impact)
  - Both impact scores are gathered by BIP! Finder [2]
- **Article abstracts**
  - $\sim$ **12***M* abstracts were collected from Open Academic Graph[2] [3, 4] and Crossref API[3]
  - An **LDA** [5] model was trained using the gensim[4] library

[1] https://opencitations.net/download
[2] https://www.openacademic.ai/oag/
[3] https://www.crossref.org/services/metadata-delivery/rest-api/
[4] https://radimrehurek.com/gensim/

Scito's search interface

- Keyword-based search interface based on Apache Solr[1] that facilitates scientific topics exploration



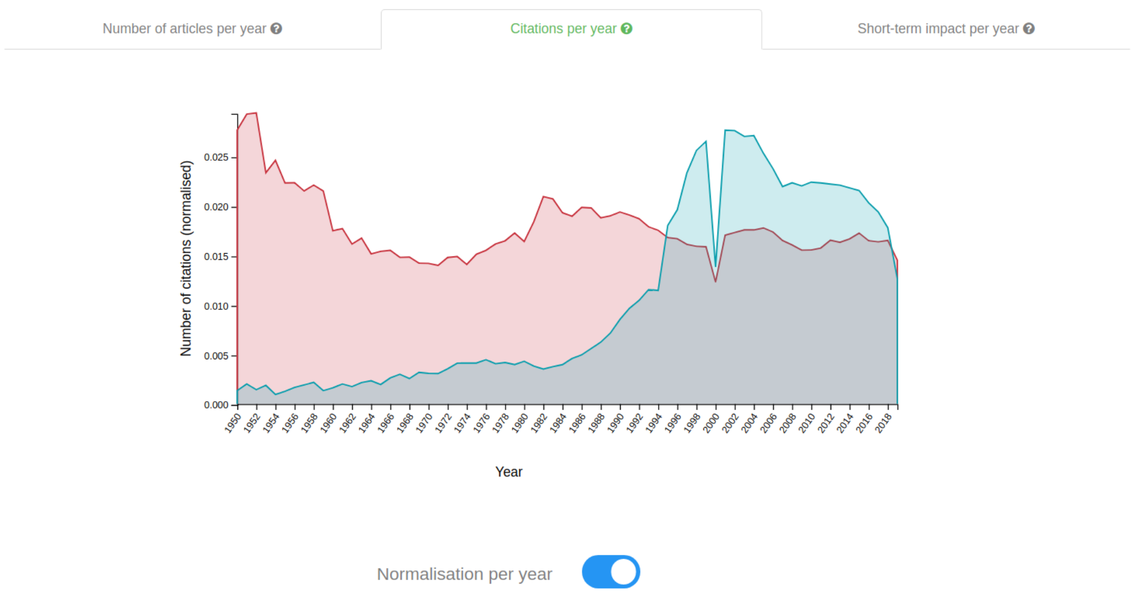
[1] https://lucene.apache.org/solr/

Visualisations for topic details

- Popularity trends for a specific topic
- Intuitive infographics based on 3 indicators:
  - Number of topic-related articles published
  - Number of citations attracted by topic-related articles
  - Average short-term impact for topic-related articles published
- Two different types of infographics:
  - Pyramid infographic
    - Shows if topic is at the top 1% or 20%
  - Trend inforgraphic
    - Dispays evolution of the 3 popularity indicators

Visualisations for topic comparison

- 2 or more topics can be selected for comparison
- The comparison view contains the Trend infographic for all topics under comparison
- Comparison scenario:
  - A researcher explores topics related to the keyword "gene". She finds a rather popular topic containing the terms "mrna", "rna" and "transcript". (blue time-series)
  - Then, she wants to compare this topic with the research field studying drug effects. She identifies a topic from life sciences containing the terms "drug", "treatment" and "effect" (red time-series).
  - In the comparison view, the Trend infographic reveals that although the "drug"-related topic was traditionally more popular, after **1995** the "gene"-related topic started to become equal or more popular (depending on the indicator used)



References

[1] Rumi Ghosh, Tsung-Ting Kuo, Chun-Nan Hsu, Shou-De Lin, and Kristina Lerman.
Time-aware Ranking in Dynamic Citation Networks.
In *IEEE ICDMW*, pages 373–380. IEEE, 2011.

[2] Thanasis Vergoulis, Serafeim Chatzopoulos, Ilias Kanellos, Panagiotis Deligiannis, Christos Tryfonopoulos, and Theodore Dalamagas.
Bip! Finder: Facilitating scientific literature search by exploiting impact-based ranking.
CIKM, 2019 (to appear).

[3] Arnab Sinha, Zhihong Shen, Yang Song, Hao Ma, Darrin Eide, Bo-june Paul Hsu, and Kuansan Wang.
An Overview of Microsoft Academic Service (MAS) and Applications.
In *WWW*, pages 243–246, 2015.

[4] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su.
ArnetMiner: Extraction and Mining of Academic Social Networks.
In *ACM SIGKDD*, pages 990–998, 2008.

[5] David M. Blei, Andrew Y. Ng, and Michael I. Jordan.
Latent Dirichlet Allocation.
JMLR, 2003.

Funding: *We acknowledge support of this work by the project "Moving from Big Data Management to Data Science" (MIS 5002437/3) which is implemented under the Action "Reinforcement of the Research and Innovation Infrastructure", funded by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).*

European Union, European Regional Development Fund
HELLENIC REPUBLIC, MINISTRY OF ECONOMY & DEVELOPMENT, SPECIAL SECRETARY FOR ERDF & CF
EPAnEK 2014-2020 OPERATIONAL PROGRAMME COMPETITIVENESS ENTREPRENEURSHIP INNOVATION
ΕΣΠΑ 2014-2020 Partnership Agreement 2014-2020

Co-financed by Greece and the European Union

diwis@imis.athena-innovation.gr