



Geo-Awareness of Learnt Citations Prediction for Scientific Publications (Demo Paper)

*Ce Li Will Postler Ian Johnson Paul Brinkman
Evan Gossling Bailey Gorlewski Goce Trajcevski*

*Iowa State University,
Ames, IA, USA*

LocalRec (@ACM SIGSPATIAL) 2023

Hamburg, DE

LocalRec (@ACM SIGSPATIAL) 2023

Outline

- Background/Motivation
- System/Development
- Overview
- Conclusions

Background

(What are Information Cascades?)

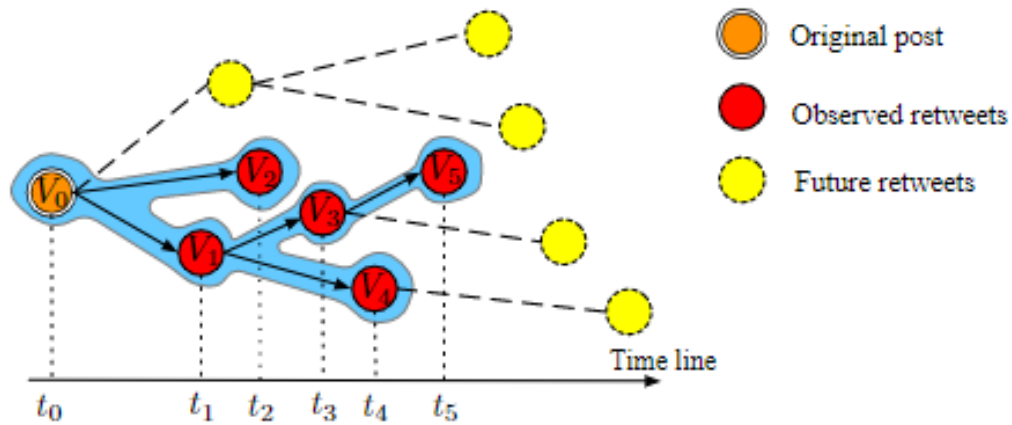


Fig. 1. The cascade graph of a post p_i . Node V_0 initiates the original message p_i .

Social Media Study



Market study experiment

Background

Domain:

- user-generated content (e.g., microblogs)
- online business (e.g., viral marketing and advertising)
- ...

Goal:

- Explore and exploit the “trajectories” and structures of the evolution of information cascades

In this work:

Domain: Scientific Literature

- ...

Goal: Quantifying and predicting the long-term impact of scientific papers (and individual authors)

- Important implications for many academic policy decisions (identifying emerging trends; assessing the merits of proposals for potential funding).

Motivation:

- **Develop a system** which, for a particular publication, would combine displaying:
 - Citation progress
 - Map visualization
 - Basic “demographics”.

- **Intended Users**
 - Individual Scientists
 - Funding Institutions
 - University Administrators



ML Foundations

- Jared David Tadeo Guerrero-Sosa, Víctor Hugo Menéndez-Domínguez, María-Enriqueta Castellanos-Bolaños, and Luis Fernando Curi Quintal. *Use of Graph Theory for the Representation of Scientific Collaboration*. In ICCCI, 2019.

- Song Jiang, Bernard Koch, and Yizhou Sun. 2021. *HINTS: Citation time series prediction for new publications via dynamic heterogeneous information network embedding*. In WWW, 2021.

- Xovee Du, Ting Zhong, Ce Li, Goce Trajcevski, Fan Zhou. *Heterogeneous Dynamical Network for Learning Scientific Impact Propagation*. Knowledge Based Systems, 238, 2022.

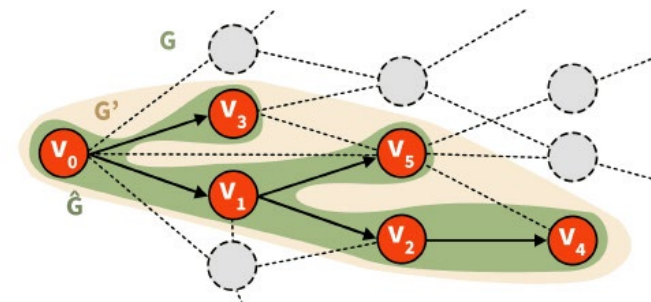
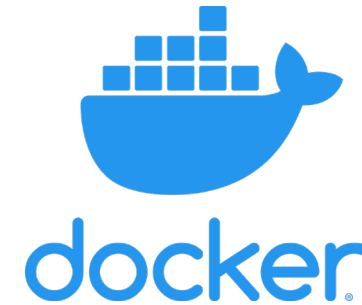
Dataset

APS (American Physical Society)

- 616K papers
- 17 journals
- 1893-2017

Broad set of tasks

- Design Infrastructure
- Initialize Database with dataset
- Implement UI Design
- Create Queries
- Implement backend ML logic
- Data Visualization
- CI/CD Pipeline

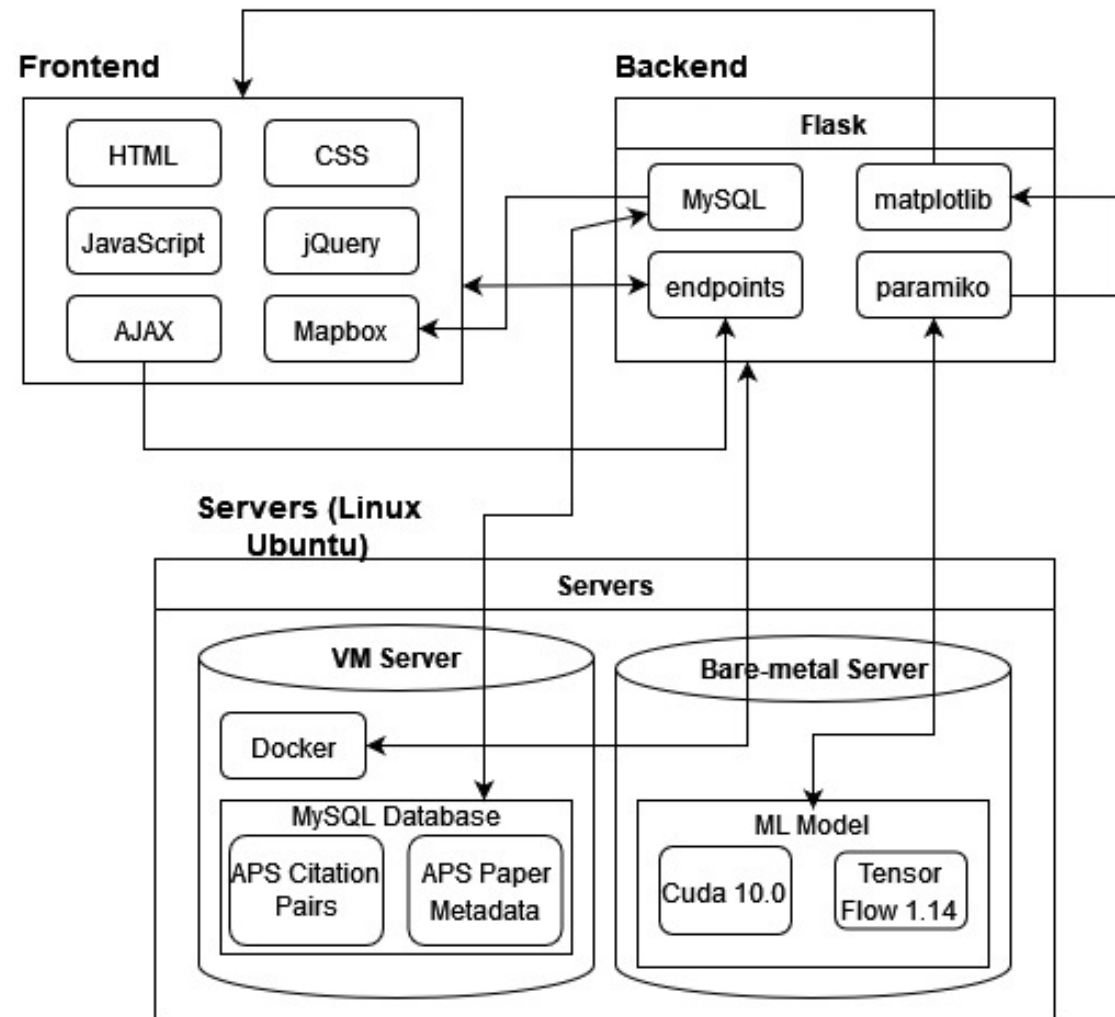


Software Architecture

- Main modules
 - Frontend
 - Backend
 - Servers

- Interfacing

- "Additional software"



Standards

- IEEE 830 - Software Requirements Specifications
- IEEE 1016 - Software Design Descriptions
- IEEE 12207 - Software Life Cycle Processes
- IEEE 1028 - Software Reviews and Audits



IEEE

Tools

- Python
- Flask
- HTML/Javascript/CSS
- JS Libraries (jQuery, Ajax)
- Mapbox
- MySQL
- Docker
- CI/CD
- Information Cascade Models



Flask



Testing

- Pytest unit testing
- Interface testing
- Integration testing
- Regression testing
- Acceptance testing
- System testing



Quick “Look-and-feel”

Future Work & Extensions

- User supplied datasets
 - Currently only APS
 - Training of models based on additional datasets
- New models/algorithms (VaCas, CasCN, etc.)
- More filtering options
 - Filtering based on region, general topics, etc.
- Additional queries
 - Query for n-th most prolific paper, 10 best papers (based on citations), etc.



Thank you!

Questions(?)