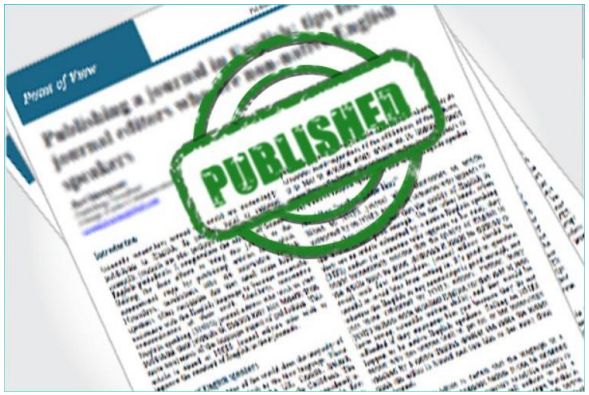
GROUP 32
Alban Bornet
Gizay Ceylan
Wei-Hsiang Lin
Lukas Vogelsang

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Network Tour of Data Science
EE-558

Term Project:
Twitter for publication success? A computational investigation

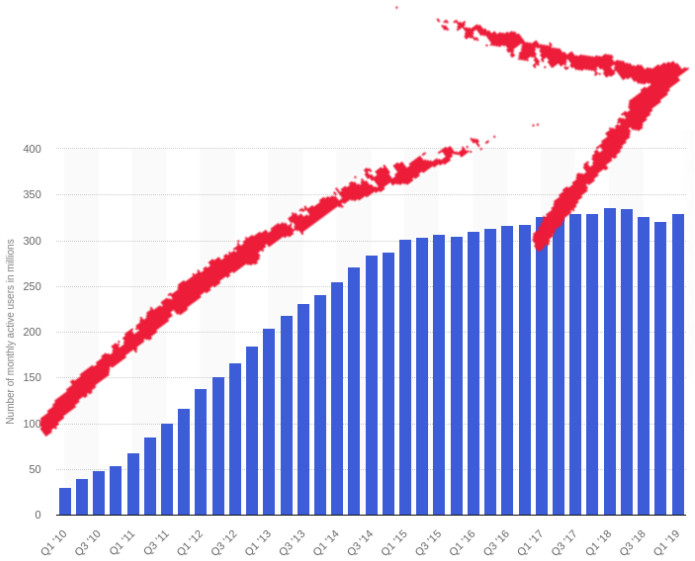
Introduction



Introduction

Introduction

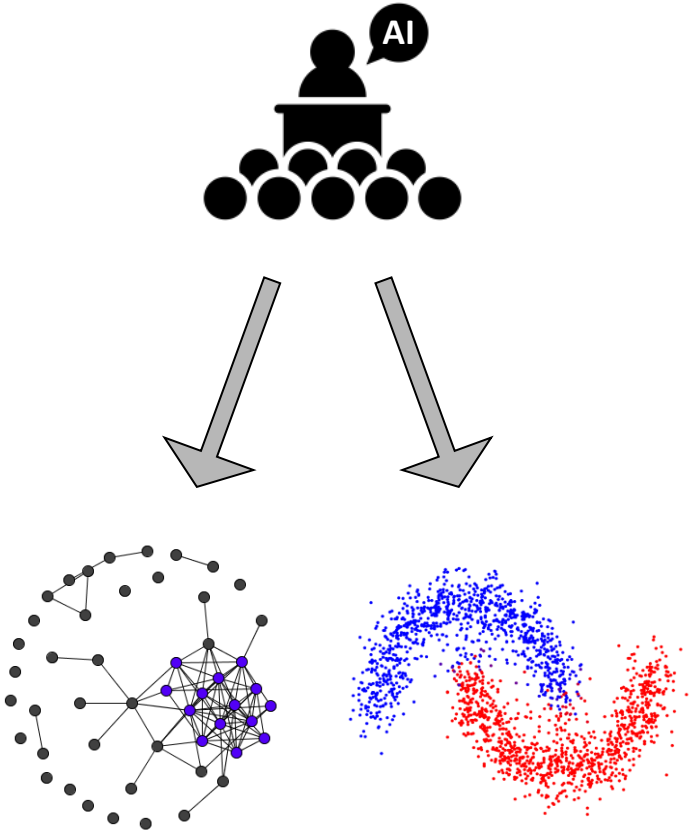
- Is it a general trend



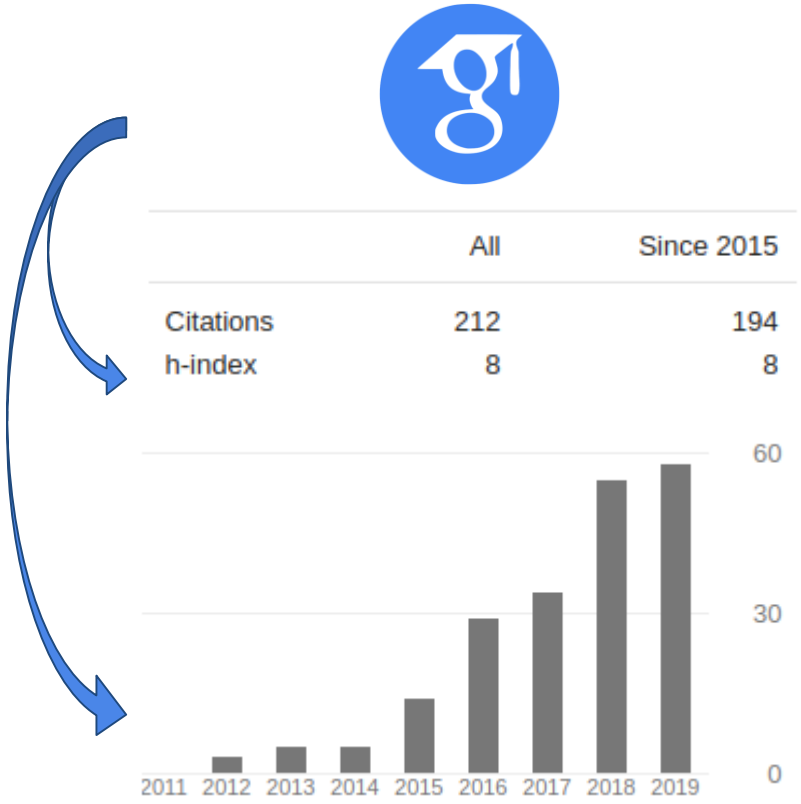
- Is it a useful strategy

Introduction

- PART I

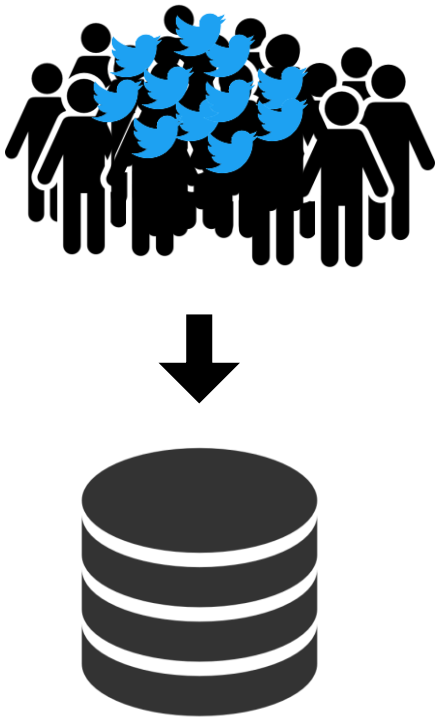


- PART II



I – Basic analysis of Twitter dataset

8605 researchers



17 Features:

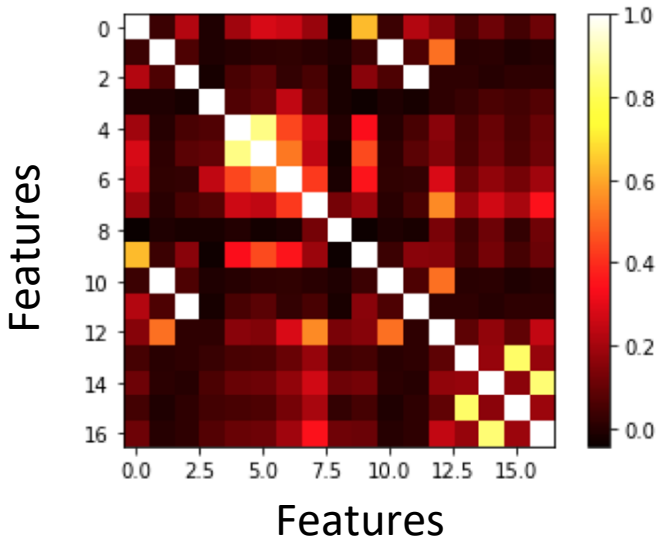
# of tweets
# of followers
# of friends
.
.
.
% of male-female researcher
% of phd-prof researcher
.
.
.

(Hadgu & Jäschke, 2014)

I – Basic analysis of Twitter dataset

Correlation of features

Correlation matrix



Strong correlation:
#hashtags and #conf_hashtags

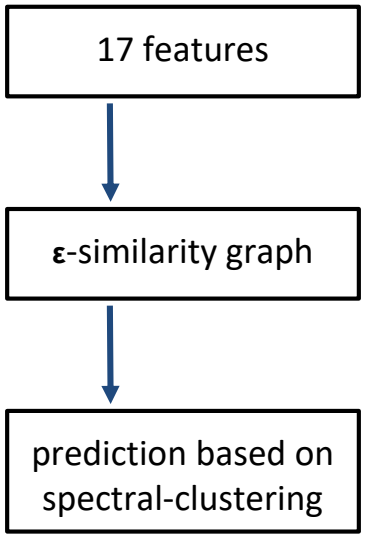
No correlation:
#publications and #followers
→ confounds: e.g., age?

Options:
a) different method?
b) different data?

I – Basic analysis of Twitter dataset

Spectral-clustering prediction

Approach:



Prediction results:

PhD vs. **Prof** ?

	precision	recall
PhD (N=911)	0.75	0.68
Prof (N=570)	0.56	0.64

✓

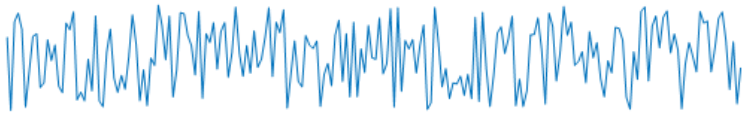
Male vs. **Female** ?

	precision	recall
Male (N=4147)	0.83	0.72
Female (N=953)	0.22	0.35

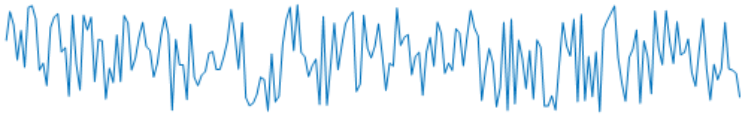
✗

→ Academic position can be predicted (above chance) from Twitter data

II – Twitter and Scholar time-series



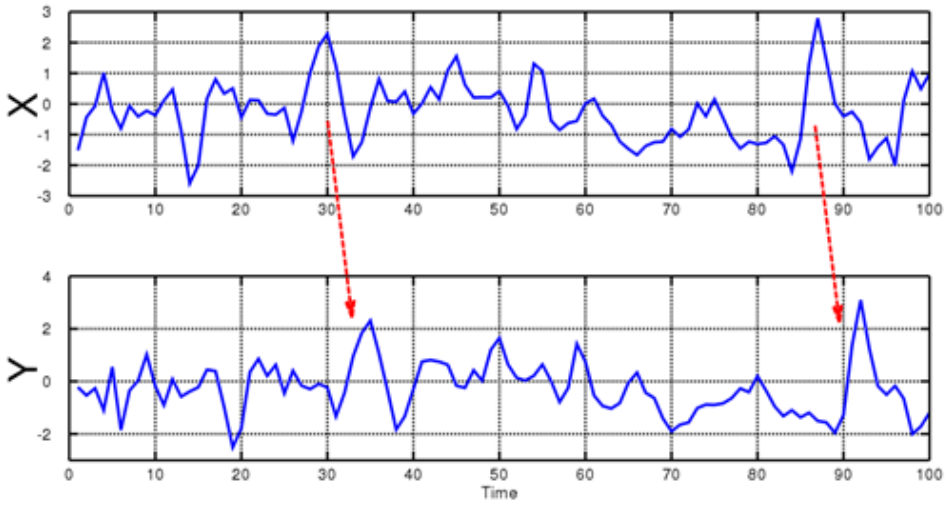
Twitter activities

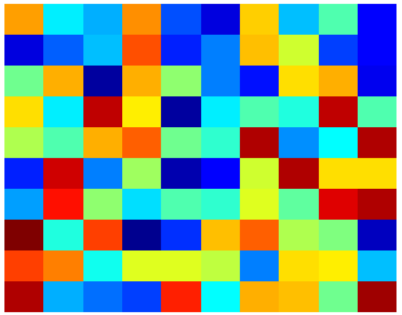


Citations numbers

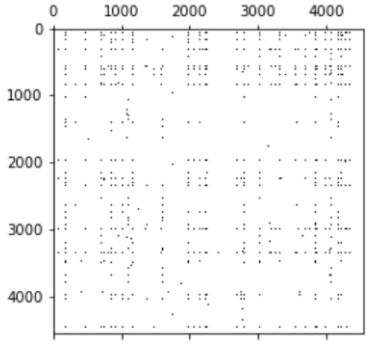
Granger causality

Granger causality





Graph properties



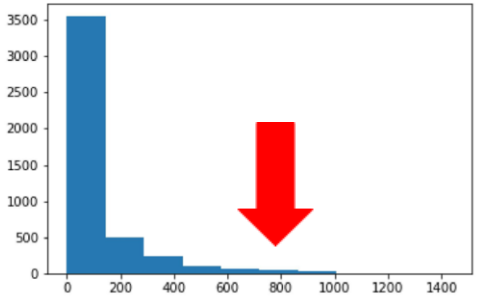
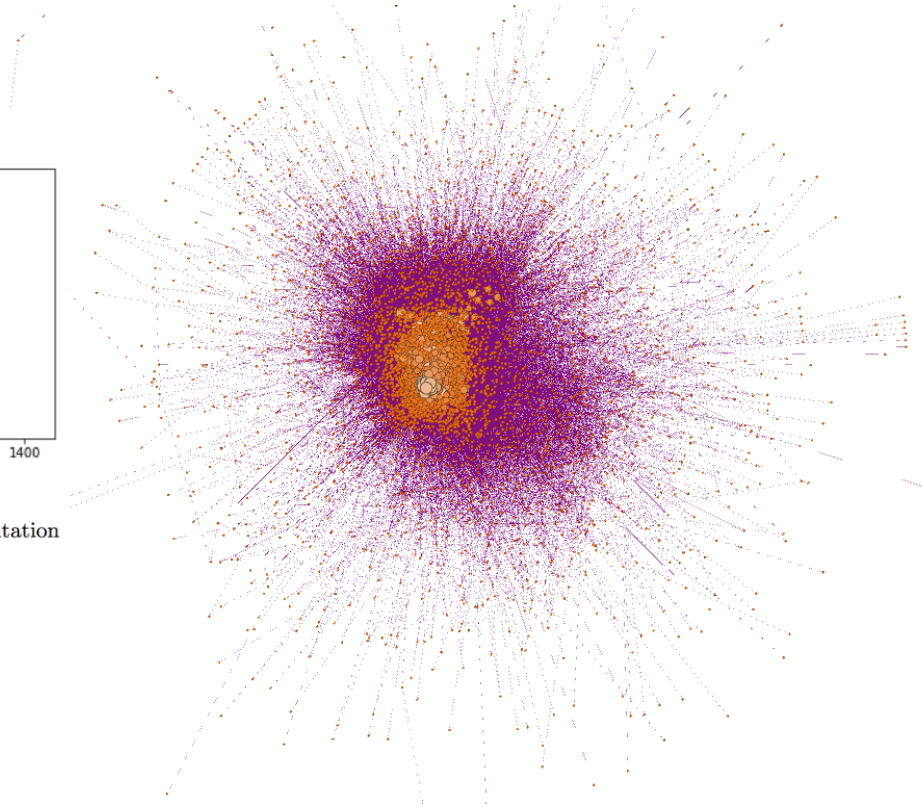


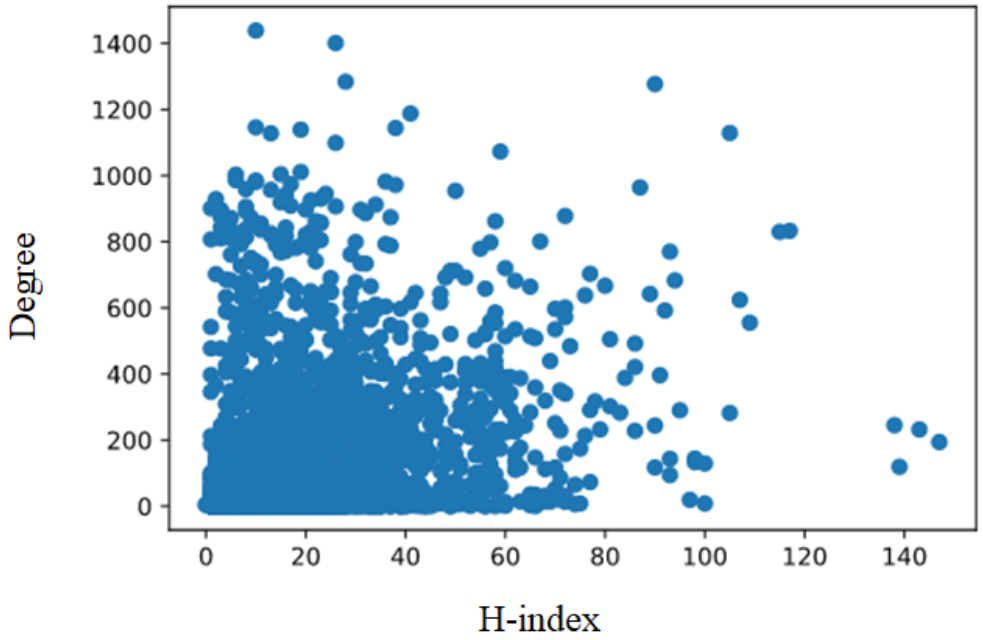
Figure 2: Left: Granger causality matrix from Twitter activity time-series to Google Scholar citation time-series Right: Degree distribution of the Twitter activity matrix



Number of connected components: 463
Average clustering coefficient: 0.155

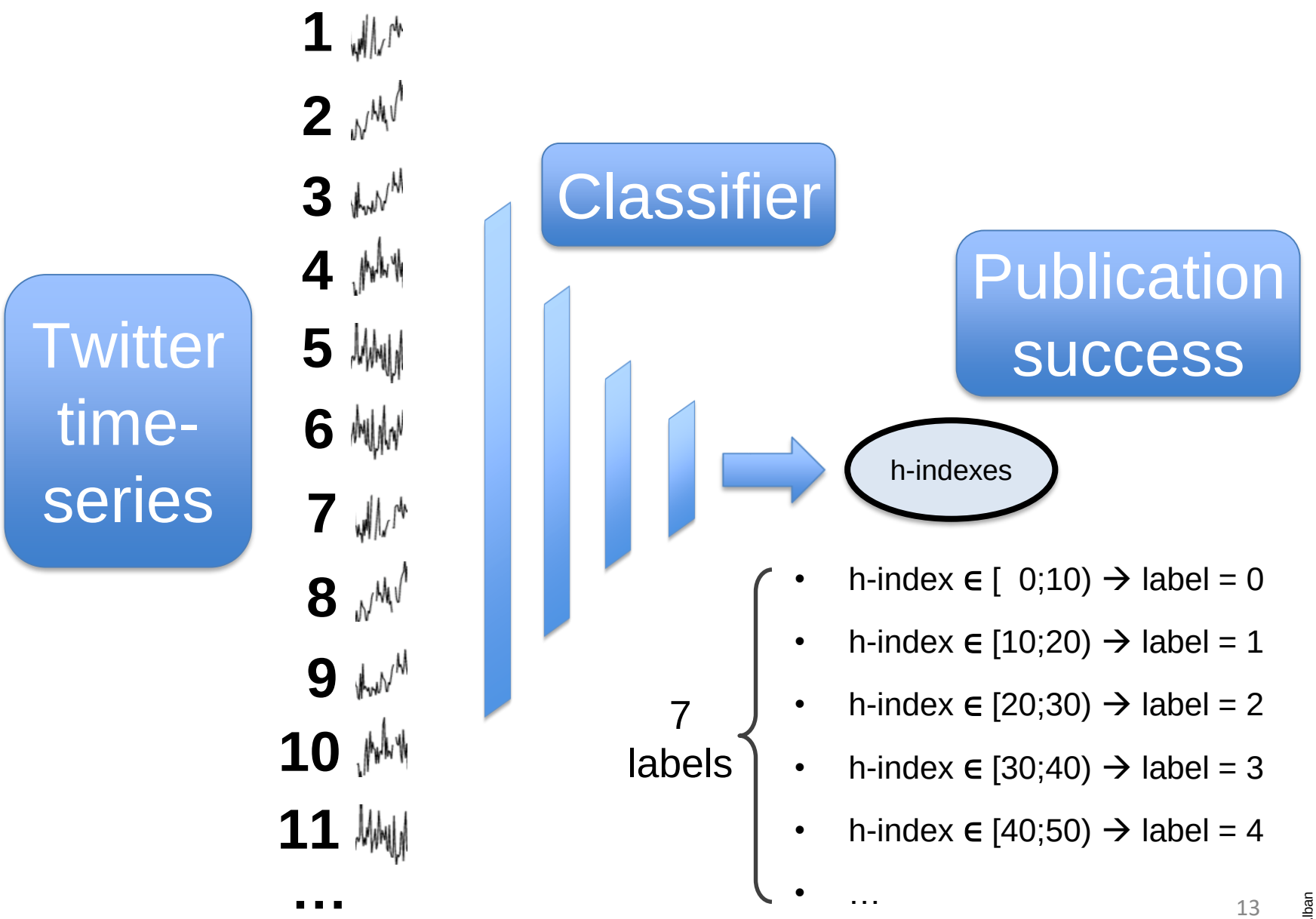
Wei

Correlation between H-index and degree of nodes



Correlation coefficient: 0.32
P-value: p<0.001

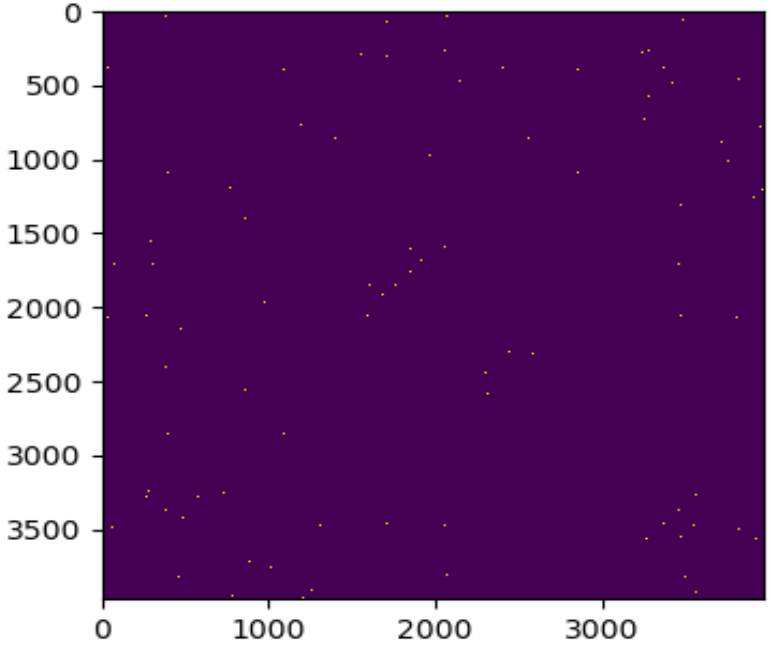
Probing Twitter time-series predictive power

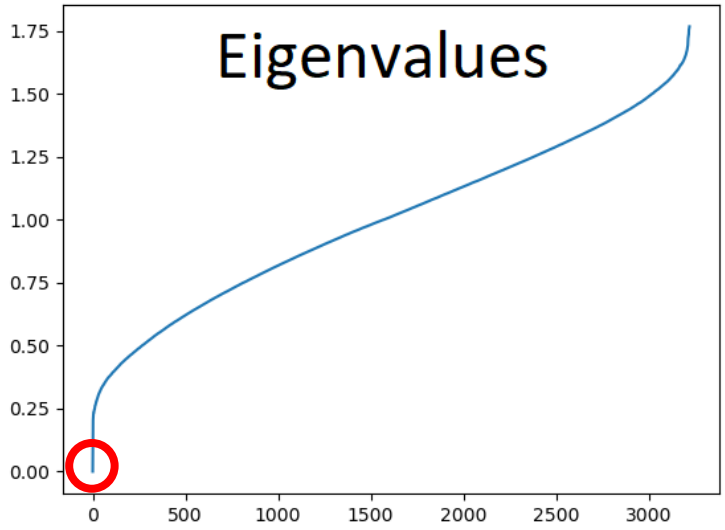


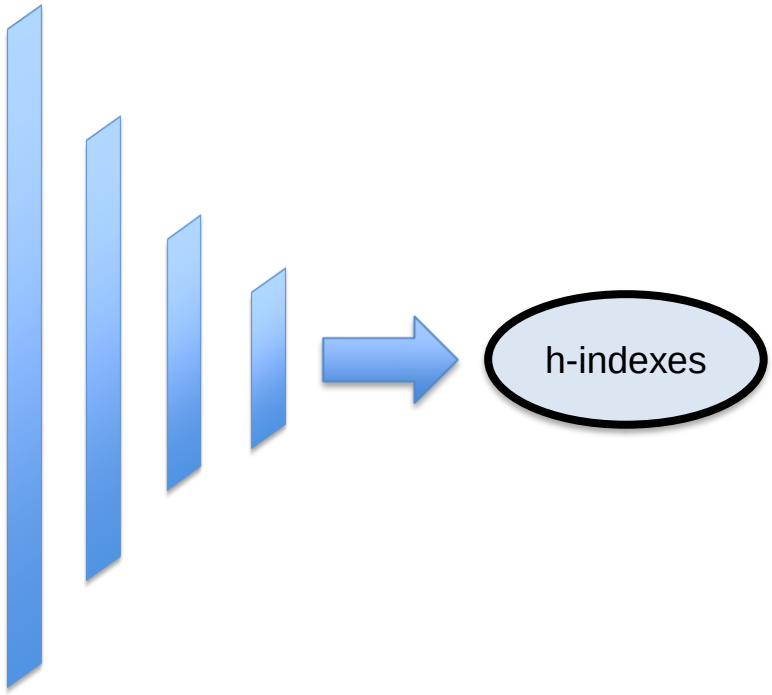
Alban

Feature enhancement with graph filtering

1
2
3
4
5
6
7
8
9
10
11
...



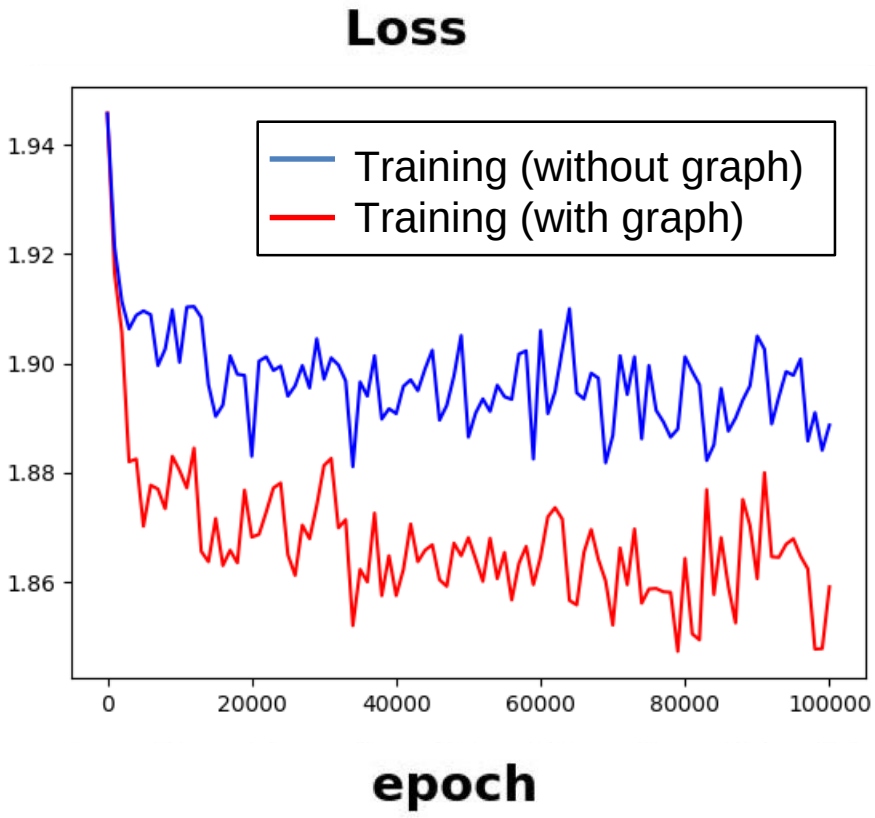


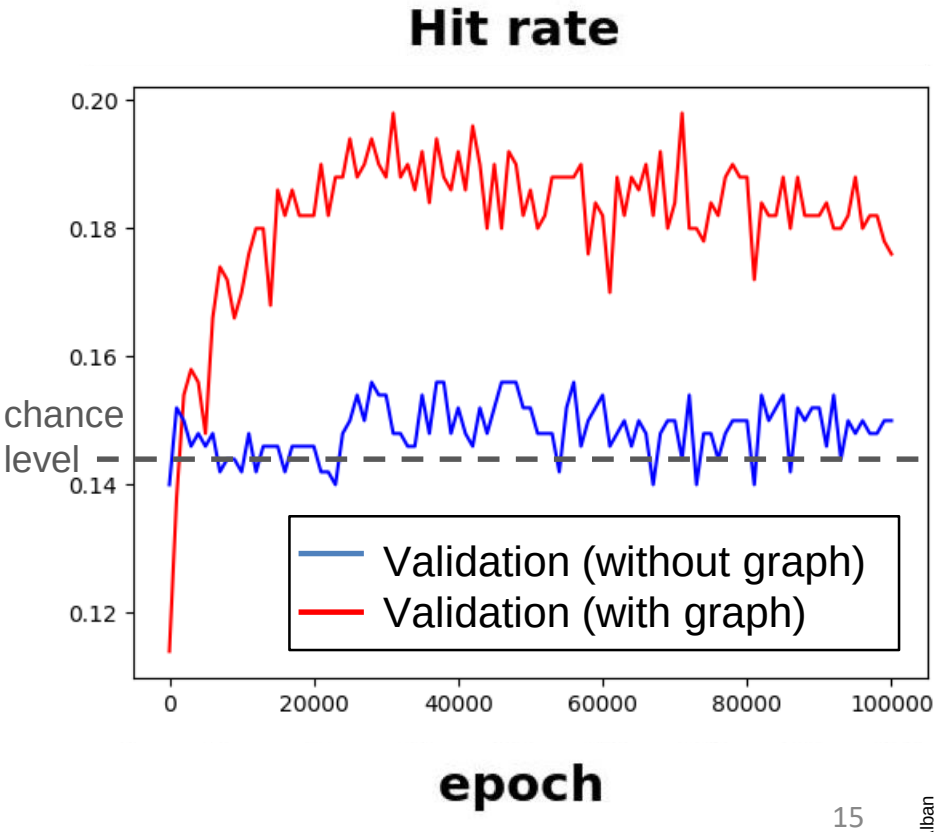


Alban

Results

- Twitter time-series are not able to predict publication success on their own, better than chance level.
- Graph filtering allows Twitter time-series to “unleash” their (small) predictive power.





Alban

Conclusion/Discussion

1. Twitter features have predictive power for identifying academic status.
2. Node-degrees in the Granger causality matrix from Twitter to publication time-series is correlated to publication success.
3. Pure Twitter time-series can predict publication success only if they are framed, using graph filtering, in the network they live in.


Thank you
for following!