Uncovering correlated activity on French Twitter

Mike Bardet

Goals

- Use DeBot method to find correlated activity between users promoting content related to the Yellow vests movement in France
- Analyse the content promoted and the link between correlated activity and similarity in content shared
- Enhanced DeBot pipeline to handle smaller dataset

Workflow

1. From a dataset containing topic-related tweets, find suspicious users
   - Create time series for each user for each time window
   - Use DeBot hashing technique to find similar time series

Workflow

- From a second dataset containing the whole activity for around 20’000 users, find correlated users
  - Create time series for each user for each time window
  - Compute dynamic time warping distance and warping correlation between each user who have enough activity in the time window
  - Report all authors involved in a correlation above 0.995

Discussion

No correlated activity

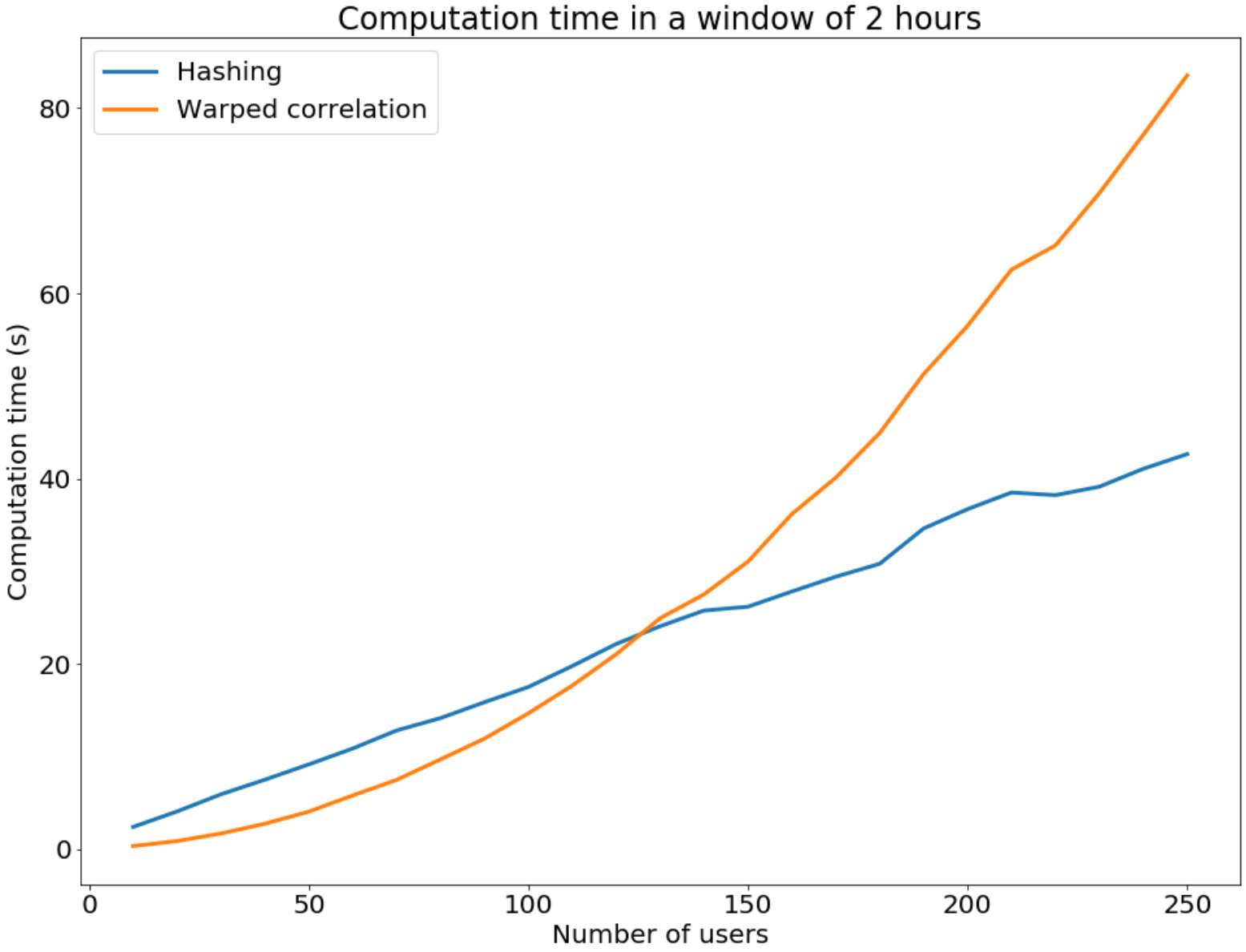
Lack of data : among all suspicious users, we have the whole timeline available for only around 3.5%

Old data : only the last 3’200 tweets of a given user are available using the Twitter API

DeBot limitations

Random projection

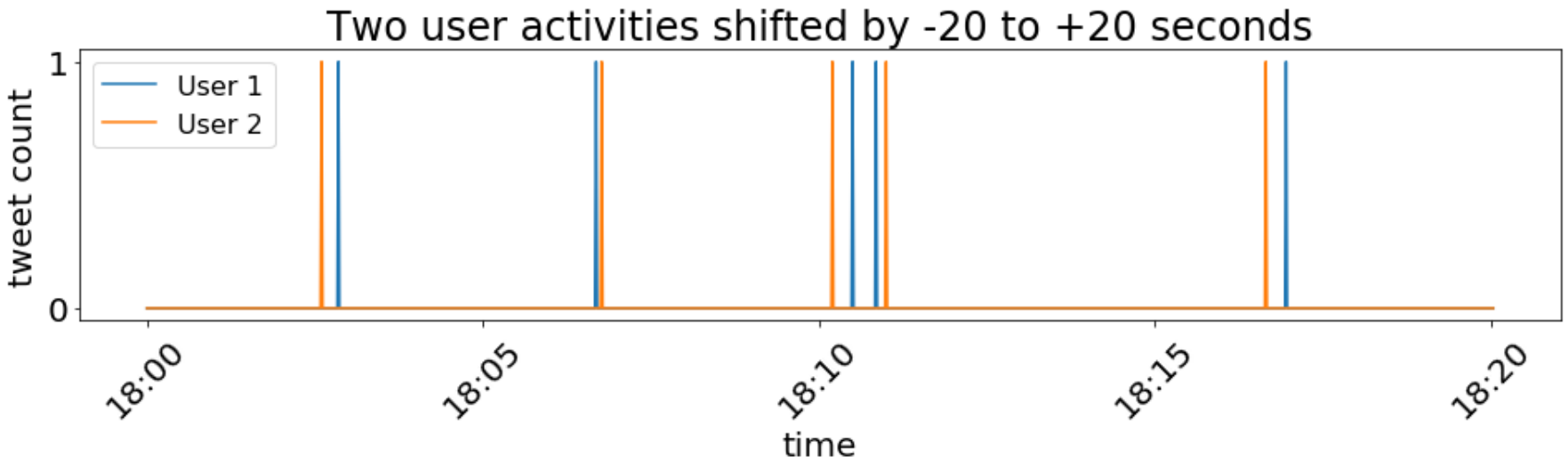
- The hashing part is O(n) but time consuming
- Random projection does not always work
  - Some similar time series may not be suspected
- Better to use DTW between all author without hashing in case the author count is low in a time window

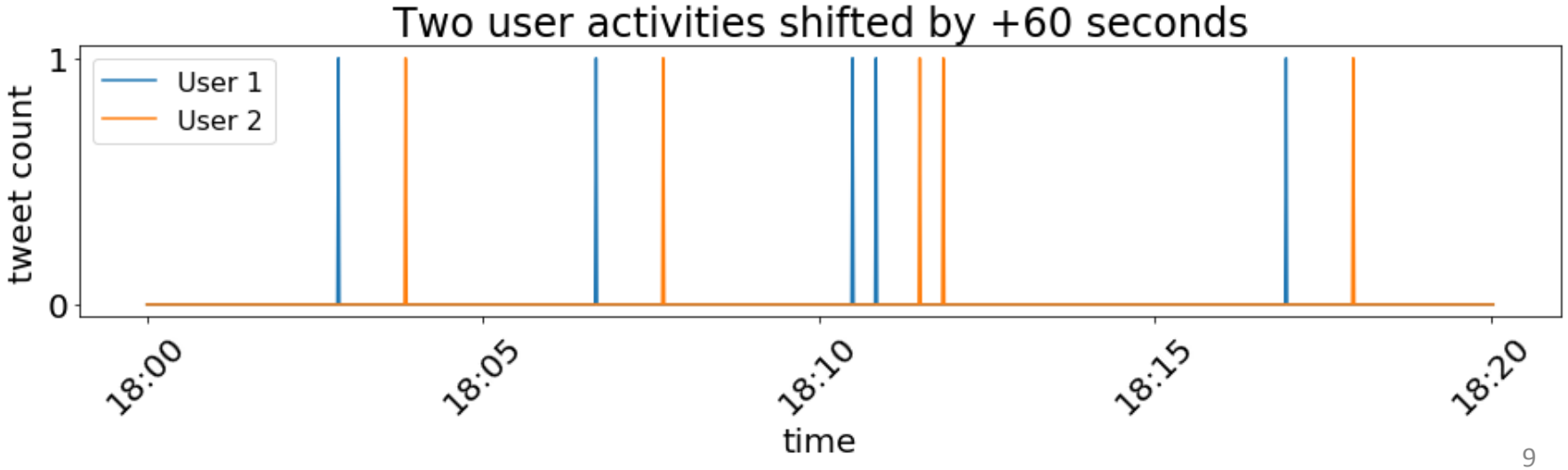

Computation time in a window of 2 hours
Hashing
Warped correlation
Computation time (s)
0
20
40
60
80
0
50
100
150
200
250
Number of users

Constrained dynamic time warping

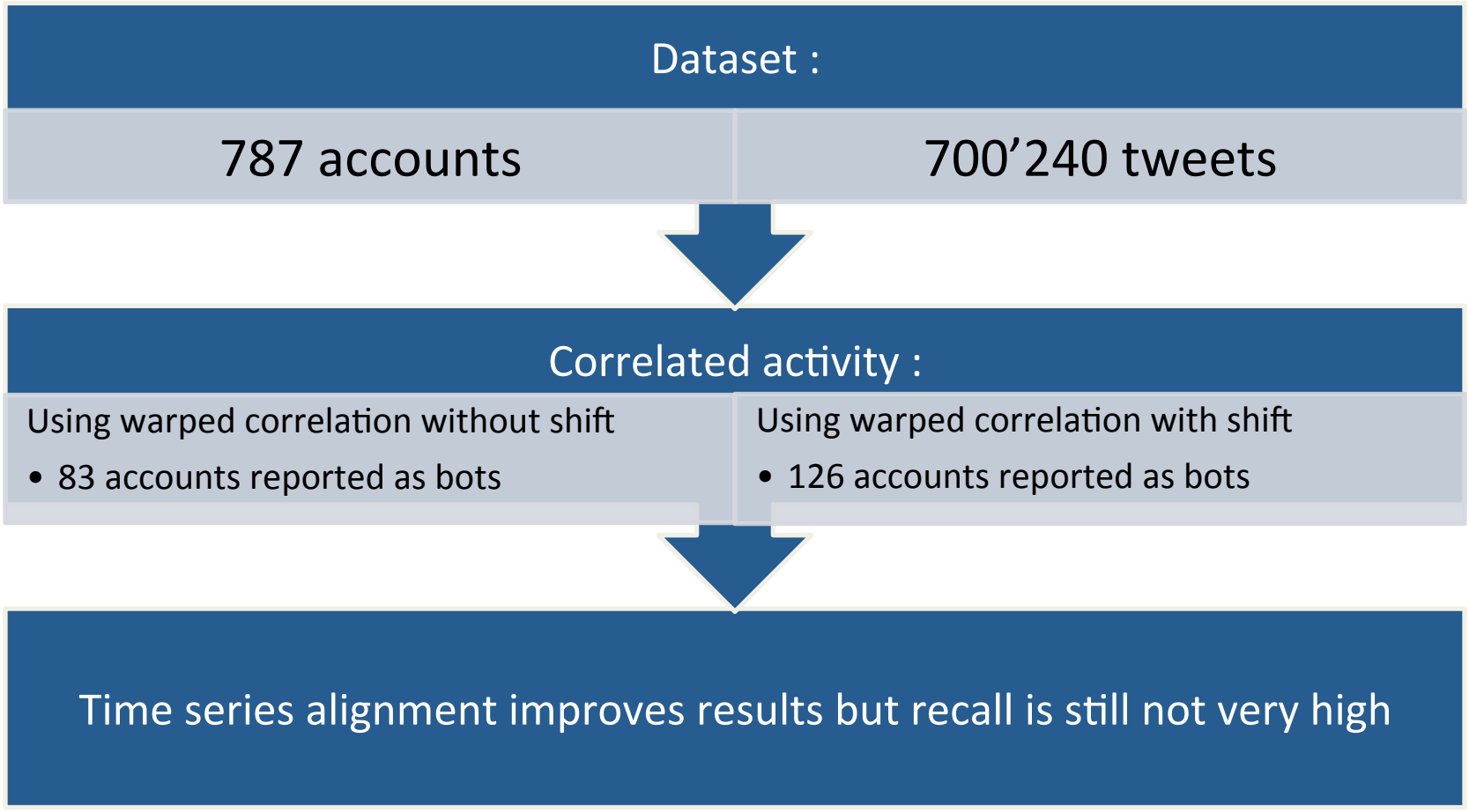
- Time series allow variations (shift and speed) but it is too permissive for our application (due to the sparsity of the time series)
- DTW is constrained using an allowable gap representing the maximum gap between two tweets from two different authors to match them

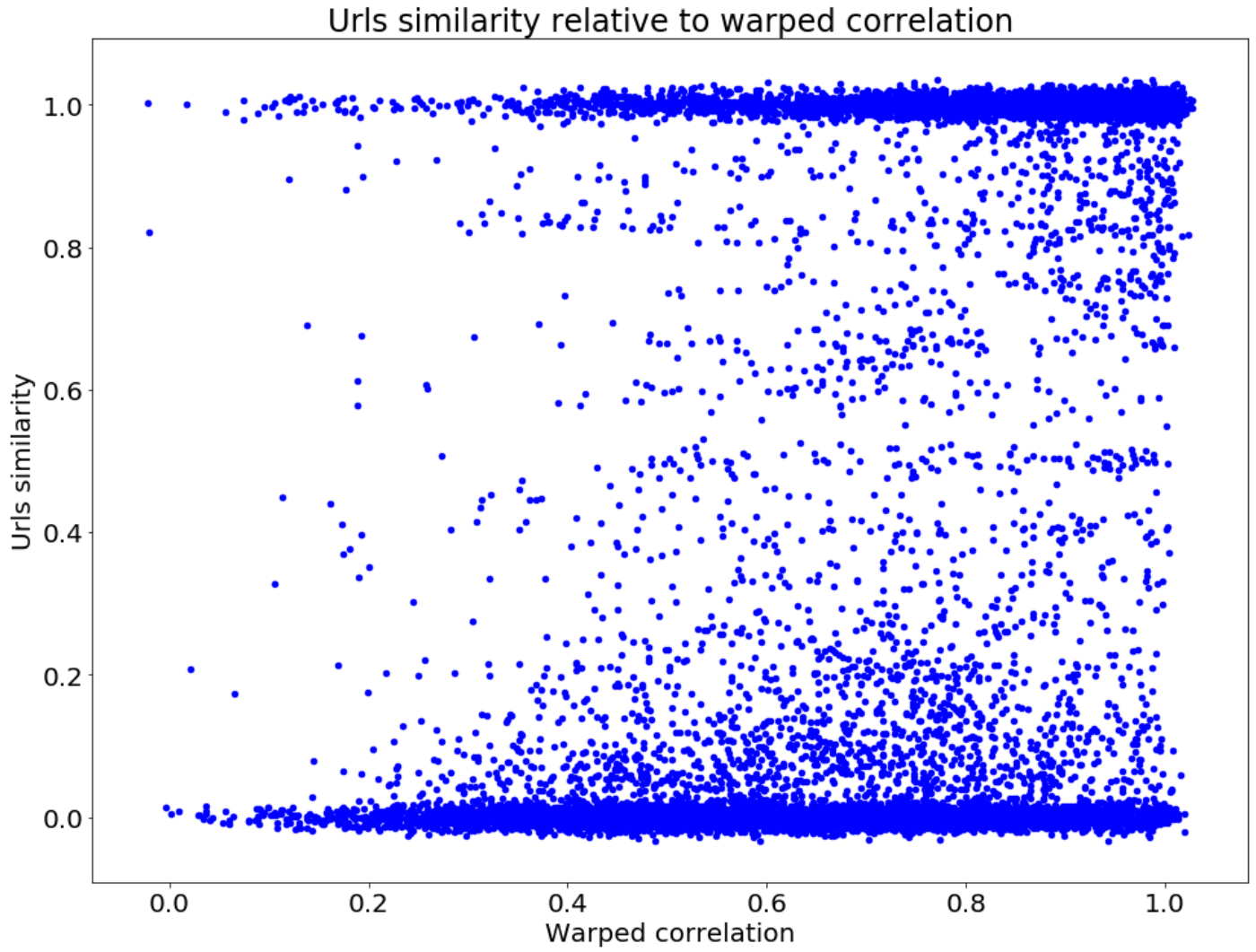
Caveats


Two user activities shifted by -20 to +20 seconds
User 1
User 2
tweet count
1
0
18:00
18:05
18:10
18:15
18:20
time

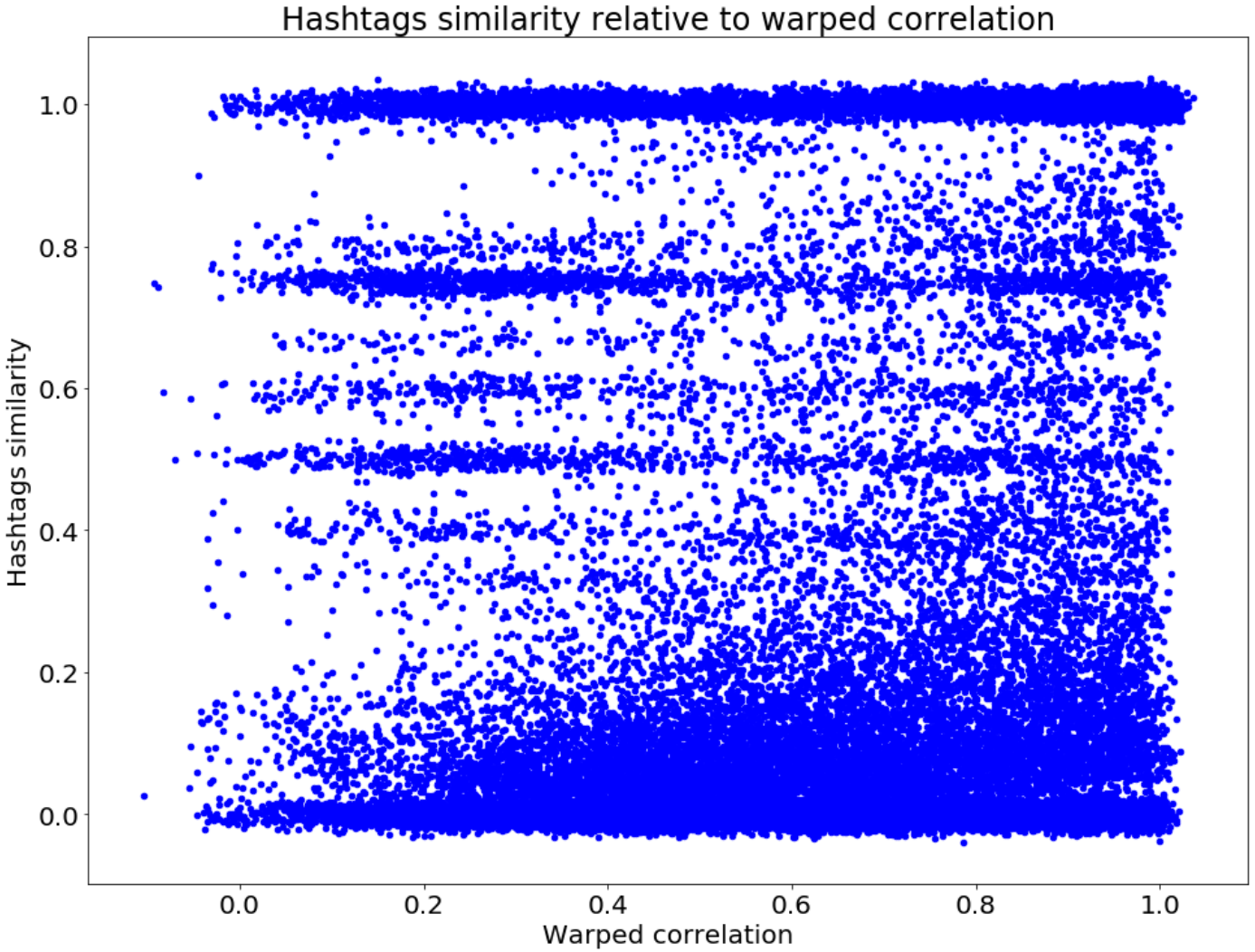

Two user activities shifted by +60 seconds
User 1
User 2
tweet count
1
0
18:00
18:05
18:10
18:15
18:20
time

Using a dataset of known bots





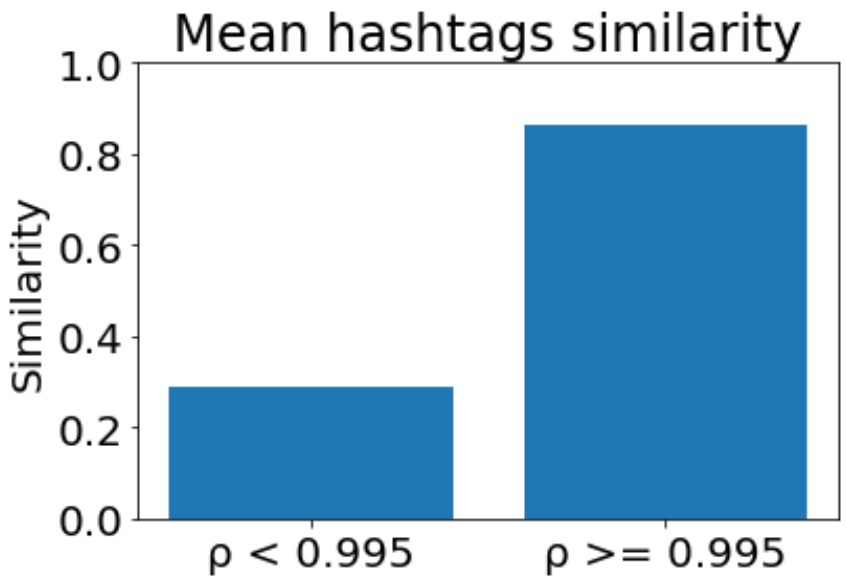
Pearson correlation = 0.49 p-value = 0.0

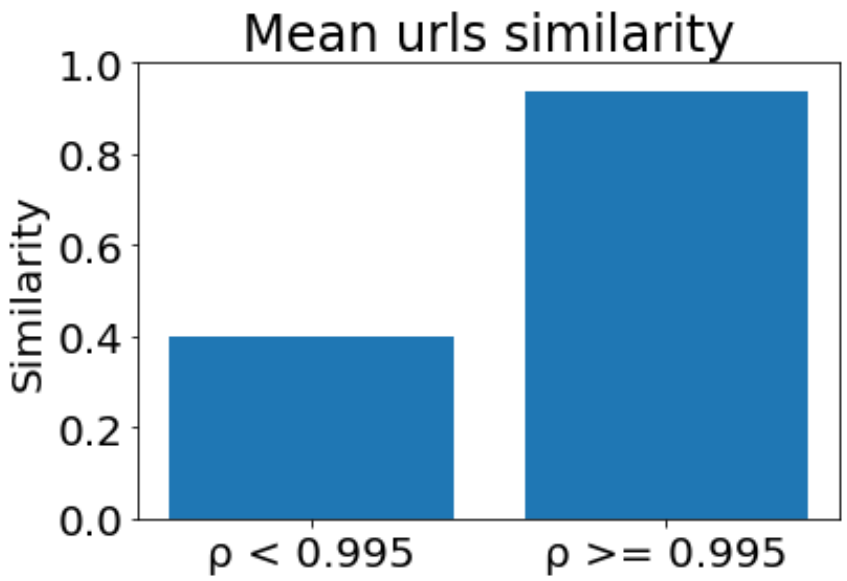


Pearson correlation = 0.16 p-value = 0.0

Content similarity

The content shared is very similar in case of correlated activity





Conclusion

- Creation of an enhanced version of DeBot
- Cope with specific topics and smaller datasets
- Choose whether to do random projection or not to speed up the process
- Fetch missing timelines if needed
- Add timeline alignment to catch more bots
- Analyse the content promoted

Thank you for your attention