Twitter Publication Analysis of Higher Education Institutions at the top of the World Ranking

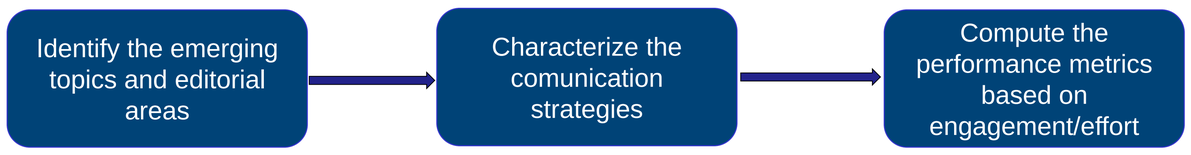
Tiago Coelho | Álvaro Figueira
Faculty of Sciences of the University of Porto

Introduction

- With the increase in competitiveness and the decrease in government funding, Higher Education Institutions (HEI) increasingly need to act in self-promotion and engagement with broader audiences.
- Social Networks like Twitter provide them the ability to increase brand recognition and improve public image [1].
- These institutions publish content according to a communication strategy to meet previously established business objectives when communicating directly with their audiences.
- It's important for them to find a good process that can evaluate their content strategy in absolute performance and relative performance terms, when in comparison to their competitors.

Objective

- The creation of an automatic system capable of discovering and classifying the communication strategy of a given HEI through the analysis of the communication patterns in a sample of 46429 Tweets gathered from different top-ranking institutions (together with UP) and the use of unsupervised machine learning techniques to perform topic detection and identify the editorial areas based on the publication's text content.



Conceptual Analysis Model

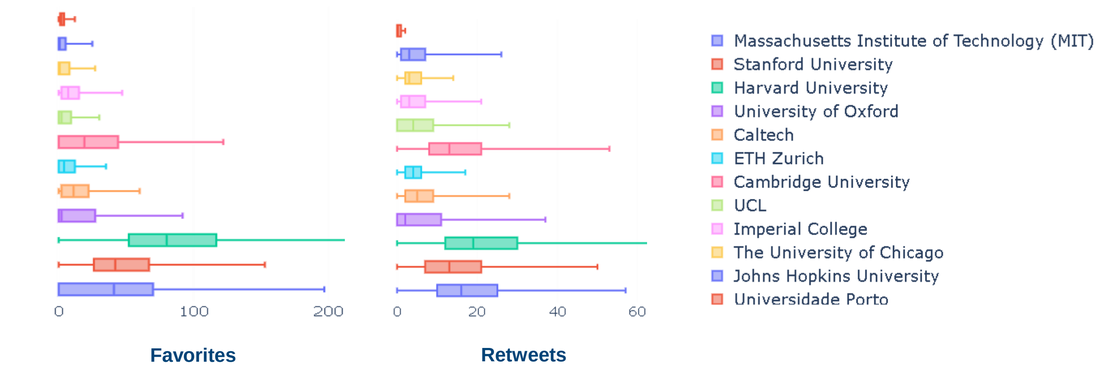
State-of-the-Art

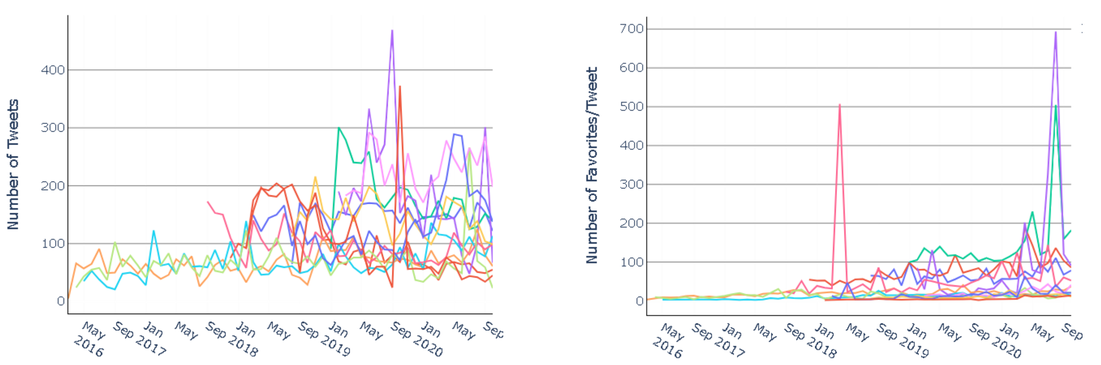
- The state-of-the-art revision methodology consisted in the study of articles indicated to me by my supervisor, completed by other papers discovered through proper research using the specific academic/scientific search engines.
- This revision resulted in the identification of best practices commonly used in this type of projects, the discovery of the theoretical and methodological approaches previously used to explore this topic and the selection of several sources to draw connections from and build upon.
- Based on the results of a comparative study between 16 different pre-processing techniques on Twitter data [2] we were able to determine the techniques to utilize when dealing with our data, namely lemmatization, URL and user mention replacement, number removal, among others .
- The experimental comparison between 4 popular Natural Language Processing libraries (NLTK, SyntaxNet, Stanford's Core NLP and spaCy) performed in [3] concluded that spaCy achieved the best performance in the tasks studied (tokenization and POS tagging). As such, we decided to choose this library to work with.
- Lastly, we adopted a performance evaluation approach derived from previous work [4] where performance is based on effort (number of posts) and response (a score based on a linear weighted function of favorites and retweets).

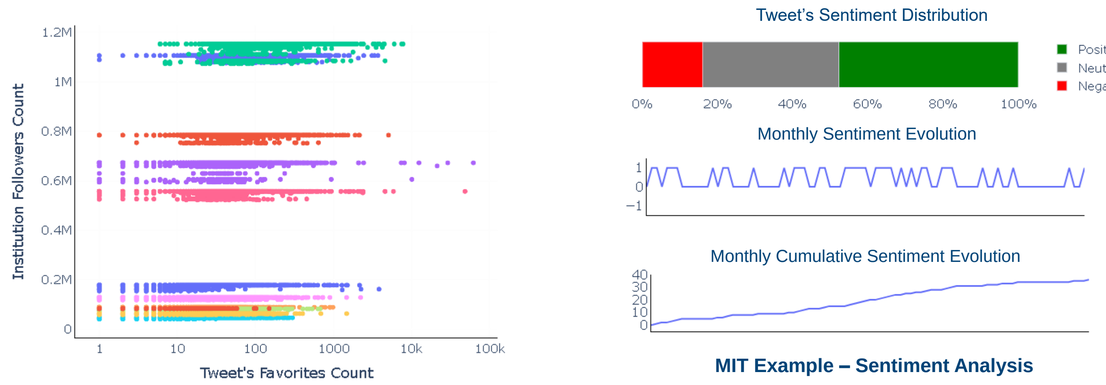
Workplan

- 2020/10/01 [3 weeks] : Perform a literature review.
- 2020/10/22 [1 week] : Acquire domain knowledge regarding the Twitter API.
- 2020/10/29 [6 weeks] : Conduct exploratory data analysis on the selected institutions and identify publication frequencies and patterns.
- 2020/12/10 [3 weeks] : Sentiment Analysis and Emotion Recognition
- 2021/12/31 [10 weeks] : Perform topic-model analysis using non-supervised algorithms.
- 2021/03/11 [4 weeks] : Correlate previous results with the publication strategy of each HEI.
- 2021/04/08 [10 weeks] : Write the Dissertation.
- 2021/05/17 [2 weeks] : Cushion Time

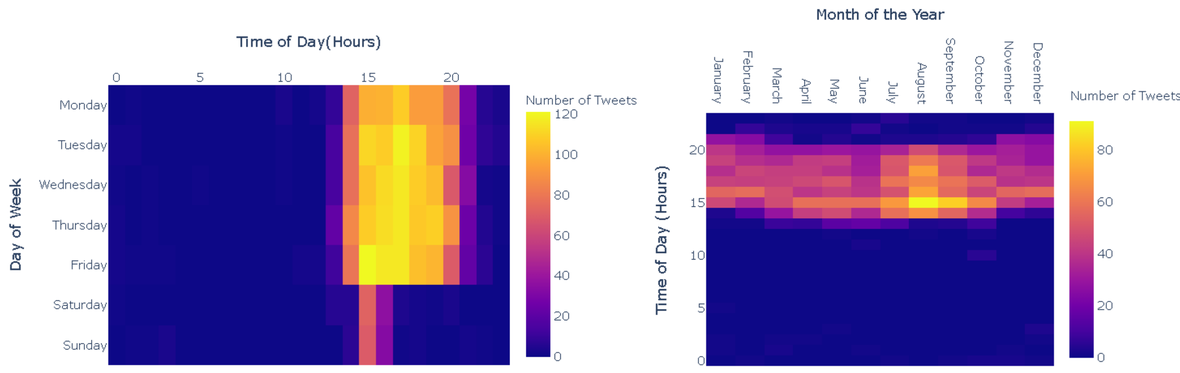
Preliminary Results







MIT Example – Sentiment Analysis



MIT Example – Publication Schedule Analysis

References

- [1] Erdoğmuş, E. and M. Çiçek.2012." The Impact of Social Media Marketing on Brand Loyalty." *Procedia - Social and Behavioral Sciences Volume 58.*
- [2] Symeonidis, S. D. Effrosynidis and A. Arampatzis.2018." A comparative evaluation of pre-processing techniques and their interactions for twitter sentiment analysis." *Expert Systems With Applications 110.*
- [3] Al Omran, F. N. A.., and C.Treude.2017."Choosing an NLP Library for Analyzing Software Documentation: A Systematic Literature Review and a Series of Experiments." *2017 IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)*
- [4] Oliveira, L. and A. Figueira.2018. "Measuring Performance and Efficiency on Social Media: A Longitudinal Study." ECSM 2018 5th European Conference on Social Media

U.PORTO
FACULDADE DE CIÊNCIAS UNIVERSIDADE DO PORTO
[dcc]
DEPARTAMENTO DE CIÊNCIA DE COMPUTADORES
FACULDADE DE CIÊNCIAS DA UNIVERSIDADE DO PORTO
MI:ERSI
Mestrado Integrado
Engenharia de Redes e Sistemas Informáticos