Capstone Project Final Report

Analyzing the Social Dynamics of the 2020 New York and Milano Fashion Weeks

New York University
Di Xin (dx489)
John Hughes (jth414)
Khouloud El Alami (kea383)

Politecnico di Milano
Andrea Pronzati
Antonio Esposito
Francesco Alongi

Urban Intensive II
July 20th, 2020

Abstract

As a semi-annual event, Fashion Week succeeds in gathering the interest of a massive audience from all corners of the world. New York City and Milan have long been at the center of this event as well as the fashion industry as whole. This research seeks to determine how this event affects peoples' behavior in the online world by looking at the 2020 Fashion Week event in each city, focusing on what insights Fashion Brands can draw from this in order to better engage and understand their audience, and ultimately developing a comprehensive view of the events as a whole from the perspective of these brands, in the form of an interactive tool.

We will observe the various communities that arise out of the diversity of interests garnered from the unique ideas that exist in the fashion ecosystem. For our analysis of the different interactions and behaviors of these people and communities that come into play with regards to fashion, we will turn to social media data from Instagram and Twitter and employ several machine learning techniques such as clustering, topic modeling, and sentiment analysis.

Introduction

Through history, fashion has become an important reflection of social and cultural change as outlets for self-expression. It is this ability of fashion to convey human voice that has made Fashion Week (FW) one of the most celebrated events in the world. Through this event, brands have been able to reach the public through strong messages and images to convey different ideas and values to build awareness.

Fashion Week has transcended the urban world and become one of the most followed international events on social media. The data generated by different social media users is revealing about the impact city-scale events have on shaping the image of brands. Understanding how brands interact with the public through social media helps produce an image of the underlying impact fashion week has on shaping the relationship they nurture with their audience. Most importantly, it can reveal how New York and Milan distinctly behave and respond to fashion weeks. As we seek to better understand the relationship between brands and their online consumers, brands can use this information to gather a better understanding of their followers and reach, and learn about the characteristics of successful brands at Fashion Week. As a result, this research will analyze the impact of Fashion Week in terms of social response, and observing characteristics of popular urban events on social media platforms, based on the 2020 Milan and New York Fashion Weeks.

Literature Review

Regarding the use and analysis of Social Media by Fashion Brands and during Fashion Week, publications such as the one by two of our advisors at Politecnico, help us understand how Social Media use is spread over the geographic area of a city during a live event such as fashion week [1]. Certain publications also apply similar methods to those we have implemented. We can observe prior work that performs a linguistic and visual analysis on twitter and instagram posts in order to analyze the impact of social media on fashion brands [2]. Methods such as LDA are also used in Twitter analysis to see how the media operates as a news source as well as how we can measure people's reactions [3][4], and visualizing this related to relevance of topics is well described using the tool LDAvis [5]. For the topic modeling performed in our project, we used a paper that outlines the process and a visual tool for it in Termite [6]. For our language and location analysis, we found a useful work which performs a geo-spatial analysis of tweets to identify global migration patterns [7]. These will help us to gain insights on the actual events in terms of how people respond to them.

Problem Statement

Our goal is to understand the role of Fashion Brands, FW itself, and the attendees in order to provide fashion brands with insights that lets them better understand their online users by measuring the reaction and interaction on social media to their events during FW. Our analysis would help them better identify and target their audiences, but also understand the strategies that worked best during FW to improve them.

Comparing the social reactions to FW in both cities could provide us with unique insights on the behavior of each, i.e. how events are perceived by the public and whether there exists communities that share similar reactions. Throughout this research, we will demonstrate how FW influences the behavior of people online and how this information can benefit brands.

How can we measure the impact of fashion week on brands through online communities?

RESEARCH QUESTIONS	EXPECTED OUTCOMES
Which are the main entities present on social media and how do they operate?	Identify the most successful brands and users during FW, and the most discussed topics.
What makes a fashion show successful online?	Identify the trendiest brands online and the key elements contributing to their online success.
Which communities interact with brands?	Identify the different communities of users that interact with each brand as well as the differences and similarities across brands, in Milan and New York.
Does the event influence brands' popularity on social media?	Understand whether and how Fashion Week events affect brand popularity
What is being talked about during FW and what are the reactions?	Identify the key topics related to each brand during FW and identify reactions via sentiments.

Data

Data was sourced using Instagram and Twitter APIs as these best measure discussion, reaction, and networks for such events. The data generated from Twitter was collected using the open source tool DMI T-Cat [8], which uses Twitter's open APIs. The Twitter data was collected for all of February (NYFW: February 4-12 and MFW: February 18-24) on tweets that contained a number of key hashtags related to FW such as *#NYFW*. The keywords dataset was based on tweets containing the following: *Fashion Week, MFW,* or *NYFW*. These datasets contain: *username, timestamp, id of the tweet, content of the tweet, retweet count, favorite count, source* *of the tweet, location, link to user profile, followers count, language,* and other features. Being that the lat/long fields came back null, for the geo-location analysis, we had to geocode the tweets using the textual location, e.g. Madrid. Being a computationally expensive process, we geocoded a 10% random sample for each week.

Instagram data was collected using a custom-built script, utilizing Instagram's APIs. The data was collected during New York and Milan FW and includes information on the basic profiles from which we have decided to collect, such as *Follower Count, Following Count,* and *Biography*. It also lists all the posts for the entirety of FW, including the *user id* of everyone to like and comment, helping us identify communities. We later retrieved the *followers count, following count, media count* (number of published media by a given user from the first post to the last one), and *user biography* for a 20% sample of the reached users in order to perform the community detection. The list of accounts to collect was gathered from domain experts from Politecnico di Milano consulting with the team.

Methodology

Beginning with the Instagram analysis related to Brand influence, our EDA focused on the accounts reached with success[1], which can represent community engagement towards a specific brand, and raw barcharts reported which brands were able to reach more accounts. Community detection can further explain why some brands with less followers have been able to reach more accounts than brands with more followers. The community detection sought to analyze users on their intrinsic properties: nationality, gender, passions, etc, retrievable from the biography, more technical details are available in Appendix A.

---

[1] An account A is said to be reached with success by a brand B on day D if A liked at least one picture posted by B on day D.

For the Twitter analysis, topic modeling was performed using Latent Dirichlet Allocation (LDA). Data preprocessing included:

1. Cleaning of the tweets by removing stopwords, punctuation, emojis, # and @ symbols, and retweets;
2. Lemmatization based on tweet's language, as reported by Twitter API's, for the 10 most frequent languages, with tweets in less common languages assumed as English.
3. Creation of the vocabulary and corpus as input for the LDA algorithm. Optimal input achieved using Bag-of-Words.
4. Grid Search plus manual inspection of the optimal number of topics.

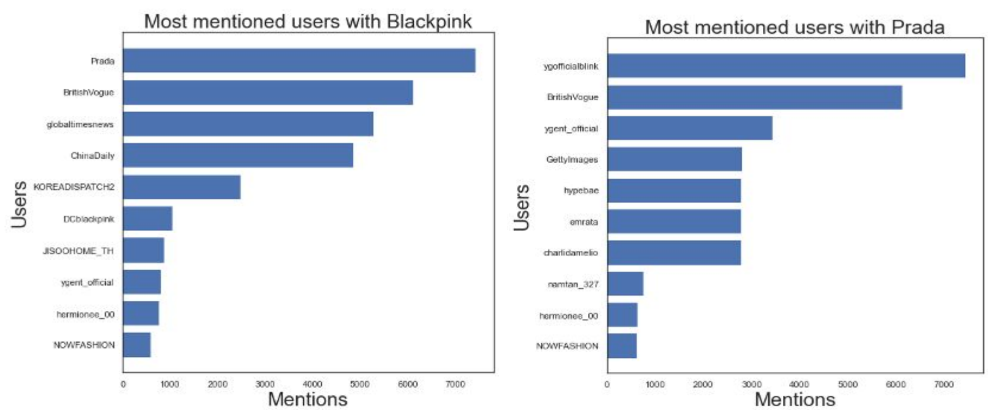
Additionally when getting the sentiment of the Tweets, we used Long Short Term Memory (LSTM) as it gave the best performance.

Results

Our first objective was to identify the most successful and influential brands by considering the number of mentions and hashtags on Twitter and number of accounts reached on Instagram. This contributed a better view over the event itself, and provided a starting point for further steps of our analysis.

On Instagram, we observed which brands were able to reach more accounts on average, throughout each FW [Fig. 1 & 2]. To explore brand influence, we noticed no correlation with the accounts reached and the number of followers per brand [Fig. 3 & 4]. On Twitter, we see a large variance in the mentions for each Brand [Fig. 5 & 6], leading us to identify the factors that have driven this engagement.

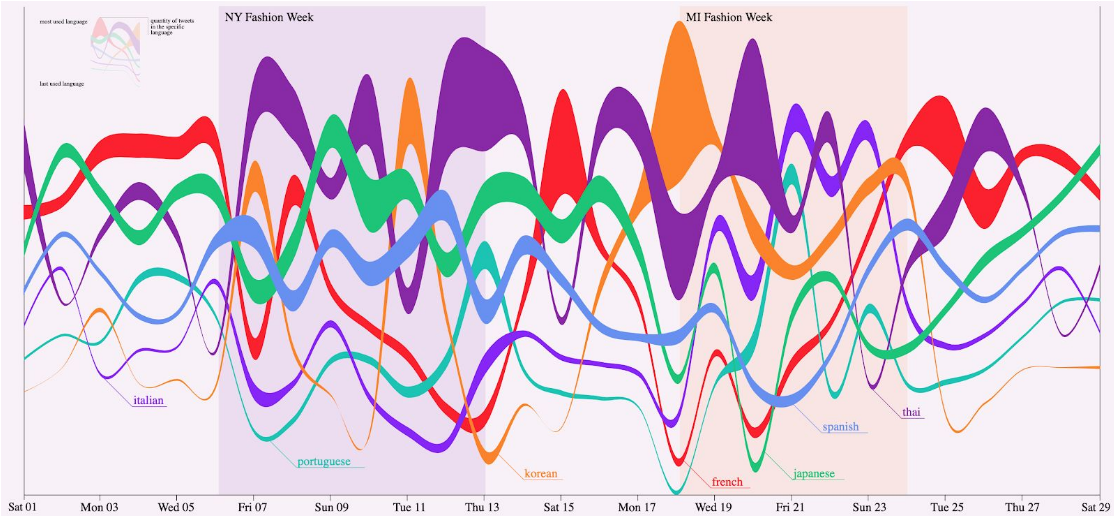
To better understand the characteristics of brands' online success on Twitter, we evaluated word trends associated with the most popular brands. Results showcased the participation of celebrities as a major contributor to online popularity [Fig. 7-9]. For instance, Prada (MFW) was mentioned 10 times more than any brand, which, through further investigation, this was attributed to the attendance of BlackPink K-Pop band members. By looking into the most frequent users mentioned together with Blackpink and Prada below, both of their official accounts were the most frequent co-mentioned accounts. Moreover, BritishVogue was the most common user between the two.



Twitter mentions for Prada (Fashion Brand) and Blackpink (K-Pop artist)

This was the initial demonstration of Asian users being a key audience, leading us to further analyze these global communities.

Compared to the entire population of Twitter [Fig.11], FW tweets show dramatic shifts, notably in Thailand and Korea [Fig. 11]. Initially attributed to the attendance of K-Pop and Thai celebrities, the timeline (below) for some of the most popular languages highlights engagement to be spread throughout the Fashion Week events for Thai users, demonstrating that this is a core audience of FW.

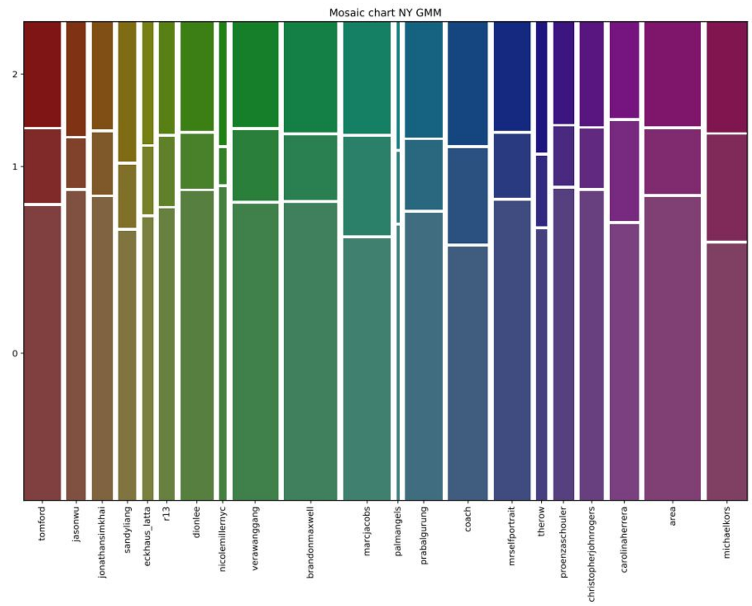


Time-Series Bump Chart for popular languages on Twitter

Next, we identify specific brands that have more engagement among individual countries. These results reveal what communities brands are connecting with and how certain strategies may strengthen this connection. We notice high engagement among the most common languages, especially among Thai users, to Prada and Coach, which were in the most popular brands for each of the common languages [Fig. 12].

Complimentary analysis of the geo locations of tweets allowed us to differentiate for countries that have the same languages, e.g. Portuguese was one of the most common languages of tweets, but this could be occurring in Portugal or Brazil, and it would be useful to form a distinction. Performing a visual inspection of the mapped tweets for each week confirmed previous findings with a large share of tweets occurring in Thailand for both NYFW and MFW in addition to more expected countries like the U.S. [Fig. 13].

To further understand this audience, we moved to community detection. Knowing the types of users surrounding an online entity could be used to mold online contents according to the audience who interact with them. To perform the community detection, we clustered on the biography of the users using a process outlined in Appendix A. These clusters are mapped to brands in a mosaic chart below.



Mosaic Chart of NY Brand communities

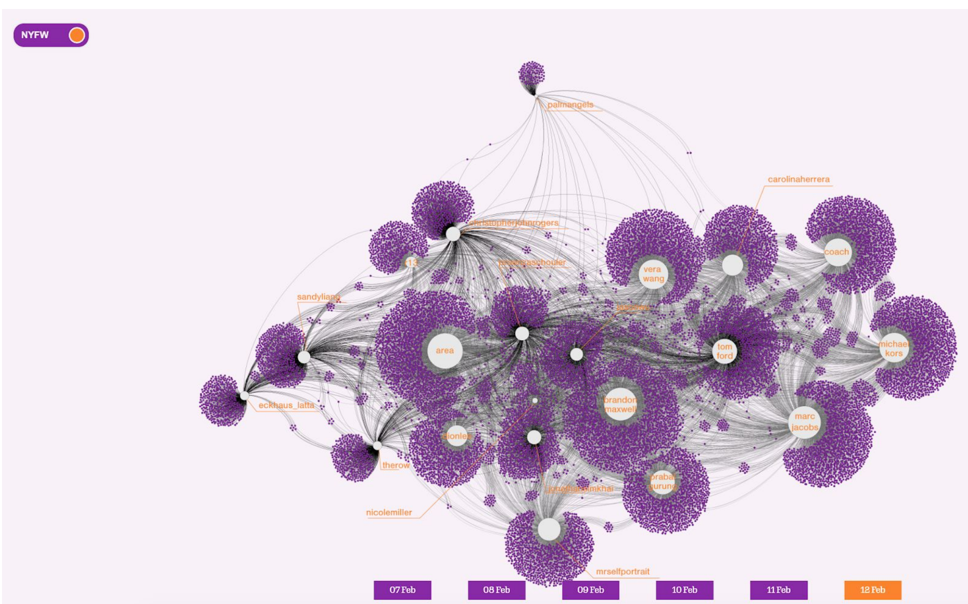
There are three main communities both for Milan and for New York, and both events present nearly the same type of users. For NYFW, the share of communities among brands is constant, while for MFW the share varies from brand to brand.

For MFW we notice that there are some brands, like marni and marcodevincenzo, for which the cluster of users related to fashion (cluster 1) is dominant, as opposed to brands that are more engaged by users belonging to cluster 0 (users which seem to be italian students) or cluster 2 (users which seem to be international).

Concerning NYFW, we notice that for every brand, the majority of users belong to cluster 0 (users related to the fashion world), compared to international users (cluster 2) which has almost the same size for each brand. Finally, there is the presence of users belonging to the business-world cluster, which is variable for each brand.

To further understand how these brands might share certain communities of users, we turn to a network analysis. Understanding the shared audience among brands is useful for strategic brand collaborations, which are very likely to happen among brands which share goals, business direction but also the target market.

We plotted the Instagram users in a network related to Brands for each week, and below we can see an example of how the NYFW community was represented on the final day of the event.



Network of Instagram Users during NYFW

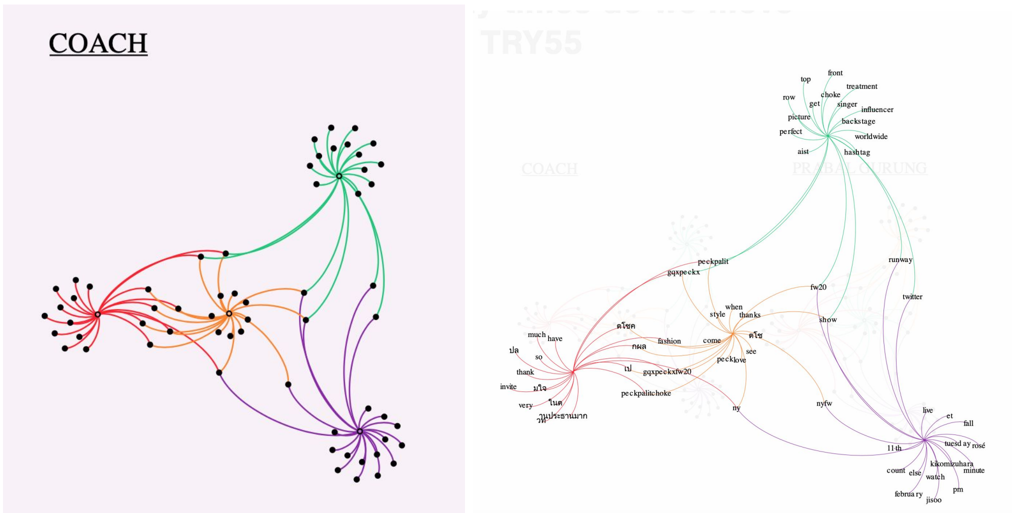
Brands in the middle of the network share the most users with other brands, while the ones which are on the edges share the less users with other brands.

Concerning the New York network, brands such as prabalgurung and mrselfportrait have their own dynamics and do not share much users with other brands, while other users instead share lots of users with others, and consequently they are placed in the middle of the network, such as proenzashouler.

The same behaviour can be seen in Milan network, in which brands such as filaeurope and philippplein are very far from the center of the network, where there are brands like maxmara and albertaferretti that share almost the same slice of users.

With a better understanding of who was engaging with FW and enhanced knowledge of Brands' audience, we also sought to gain a deeper knowledge of how these users reacted to the individual events of FW itself so that Brands could identify popular areas to target their approach. To do this, we identify the most relevant topics that are being discussed and which ones might have caused an increase in engagement for the different brands of the fashion week.

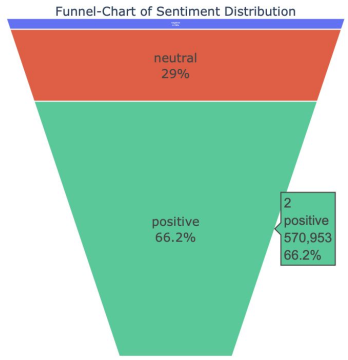
To better understand what was discussed about each brand, Topic Modeling was performed. This enabled us to identify not only the 'topics' discussed for each brand, but also similarities between topics. In the visualization below, results of the Topic Modeling are shown for the brand Coach as an example of a successful brand of NYFW. The different 'clusters' represent words that are similar and that often appear together in tweets.



Interactive tool to explore topics for most popular on Twitter

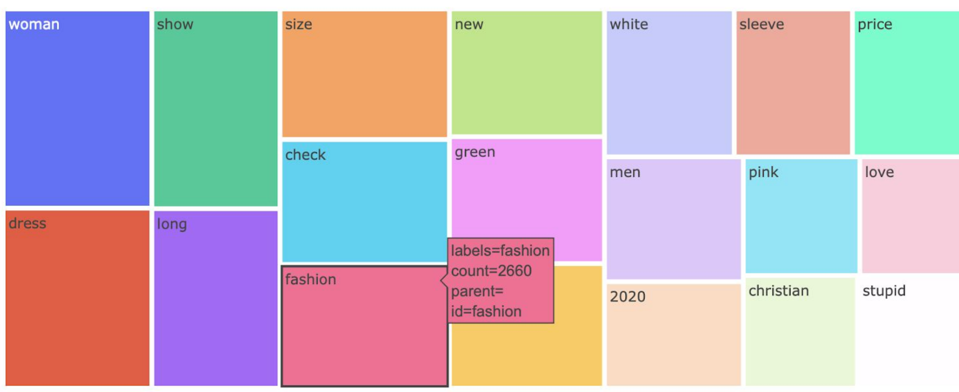
It is noteworthy how, again, a common factor among almost all the brands is the presence of topics related to the involvement of music artists (see Miley Cyrus and Nicki Minaj for Marc Jacobs) and more in particular of K-Pop artists (Blackpink for prada, Peck Palitchoke for Coach). Another important result is the appearance of a topic related to 'coronavirus' for Armani (who did its event closed-doors), and Prada, whose event was held on the same day as the beginning of the spread of COVID-19 in Italy.

To learn not only the topics, but how users reacted, we analyzed the FW reactions. For this, we can see in the figure below that the positive sentiment posts occupied the most at 66.2% of tweets, and negative sentiments posts are 4.78%.

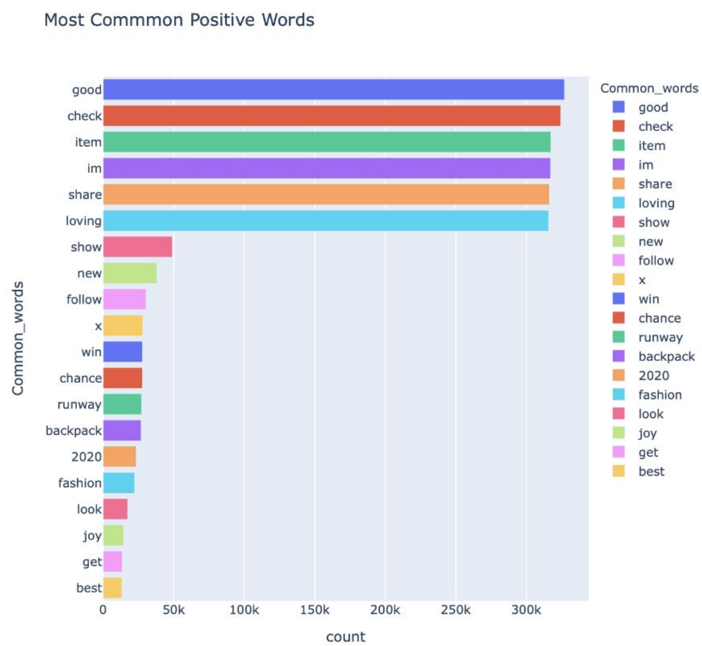


Sentiment of our collected Tweets

By taking a further look at which words appear the most in different sentiment posts, we observe the two diagrams below.



Most frequent words appearing in Negative Sentiment posts



Most frequent words appearing in Positive Sentiment posts

To let Brands explore and find insights relevant to them, we have hosted our results in the form of an interactive tool at https://datalife2020.github.io/index.html

Conclusion

Future work on this topic might seek to offer more predictive modeling for Brands to explore. For example it could build a tool by using the sentiment of reactions and FW success of events to predict how a Brand will perform in the following season.

Through our research we have gained key insights for Brands during Fashion Week. We have discovered online communities through analysis of Instagram bios and Twitter locales. This revealed the global reach of Fashion Week, finding communities clustered around international keywords and the particularly interesting community found in Asia and Thailand. We found the brands that resonate within these communities and what drove their success during FW, such as Prada and the influence of K-Pop singers at the event. Lastly, we looked at topics being discussed at these events, and how people reacted to events of FW. This further revealed that global celebrities drive much engagement during these two events. In addition, we built a website that includes interactive tools for Brands and users to explore many of these questions on their own, allowing them to find things which they find relevant to them. This research provides a comprehensive understanding of the two FW events that brands can utilize to better understand their online audience and what drives success at the event.

Team Roles

- Di Xin carried out the sentiment analysis to determine reactions to FW events.
- John Hughes analyzed Twitter communities in the form of nationalities and locations.
- Khouloud El Alami evaluated Twitter data to identify the top brands and the elements contributing to making a brand successful on social media.
- Andrea Pronzati assisted in all areas related to Data Visualization and built the project website.
- Antonio Esposito scraped data for Twitter, and analyzed the impact brands have on social media as well as the topics discussed.
- Francesco Alongi scraped data for Instagram and evaluated the influence of the trendiest online entities through networks and community detection.

References

[1] Brambilla, M., Ceri, S., Daniel, F., & Donetti, G. (2017). Temporal Analysis of Social Media Response to Live Events: The Milano Fashion Week. *Lecture Notes in Computer Science Web Engineering*, 134–150. doi: 10.1007/978-3-319-60131-1_8

[2] Manikonda, L., Venkatesan, R., Kambhampati, S., Li, B. (2015, December). Trending chic: Analyzing the influence of social media on fashion brands. Department of Computer Science, Arizona State University.

[3] Chrisler, J. C., Fung, K. T., Lopez, A. M., & Gorman, J. A. (2013). Suffering by comparison: Twitter users' reactions to the Victoria's Secret Fashion Show. *Body Image*, *10*(4), 648–652. doi: 10.1016/j.bodyim.2013.05.001

[4] Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E.-P., Yan, H., & Li, X. (2011). Comparing Twitter and Traditional Media Using Topic Models. *Lecture Notes in Computer Science Advances in Information Retrieval*, 338–349. doi: 10.1007/978-3-642-20161-5_34

[5] Sievert, C., Shirley, K. (2014). LDAvis: A Method for Visualizing and Interpreting Topics. *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, doi:10.3115/v1/w14-3110.

[6] Chuang, J., Manning, C., Heer, J. (2012). Termite: Visualization Techniques for Assessing Textual Topic Models. *Proceedings of the International Working Conference on Advanced Visual Interfaces - AVI '12*, doi:10.1145/2254556.2254572.

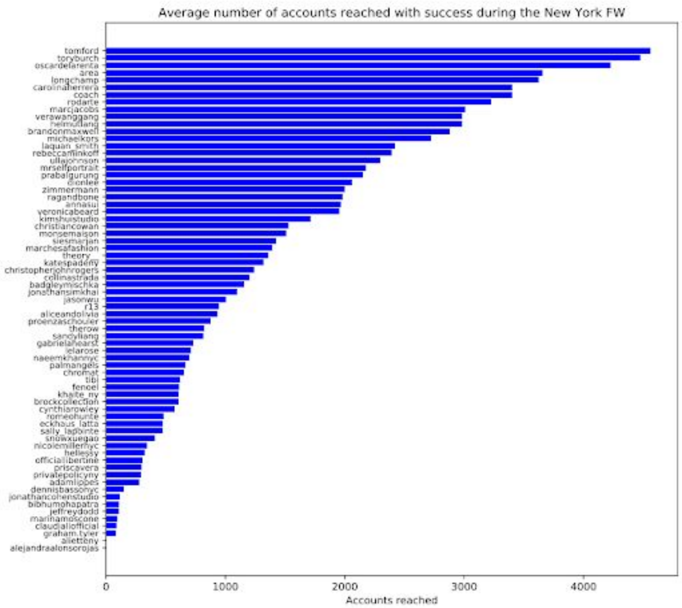
[7] Gaere, Edward. (2016). *Analysis of Language Mobility using Twitter Messages.* 10.13140/RG.2.1.2886.7448.

[8] Borra, E., Rieder, B. (2014) Programmed method: developing a toolset for capturing and analyzing tweets. *Aslib Journal of Information Management*, Vol. 66 Iss: 3, pp.262 - 278. http://dx.doi.org/10.1108/AJIM-09-2013-0094

[9] Clement, J. *Twitter: Most Users by Country. Statista*, 24 Apr. 2020, www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/.

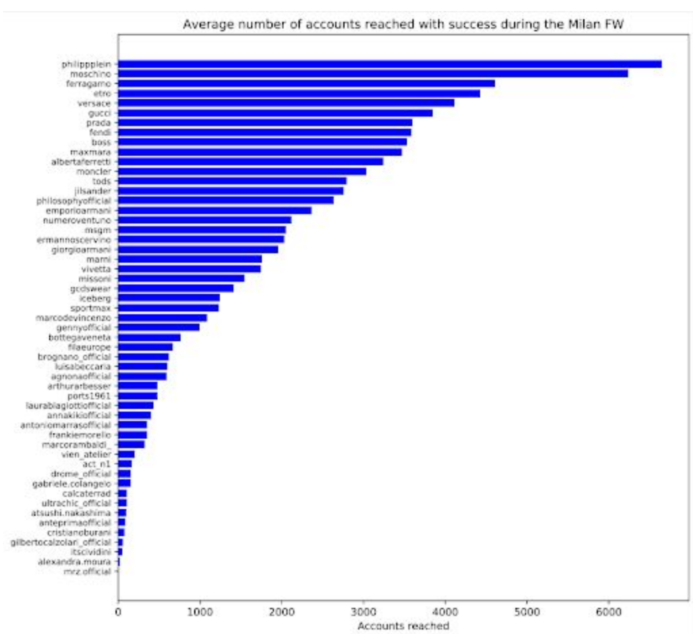
Figures

[Fig. 1]



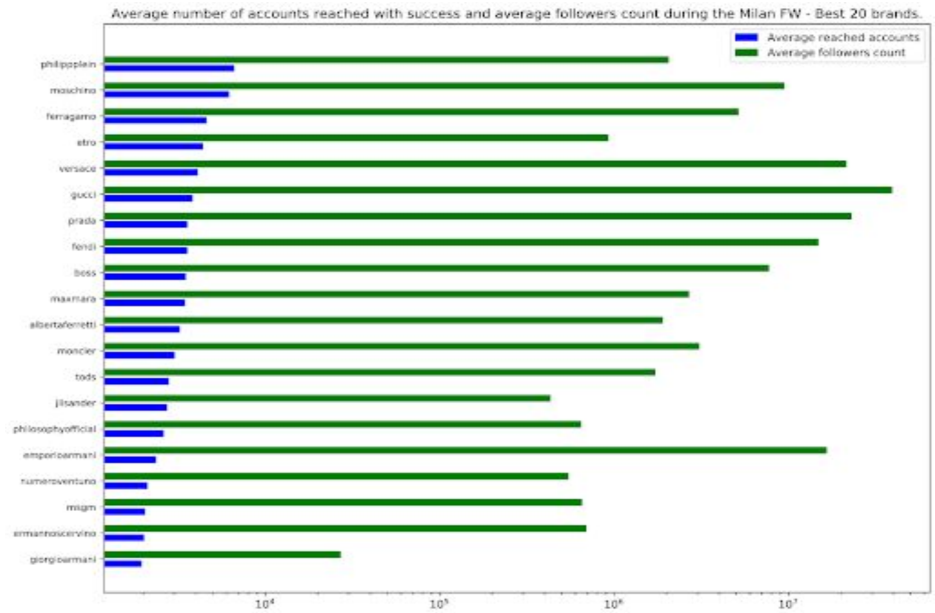
Accounts reached for NY Fashion Brands

[Fig. 2]



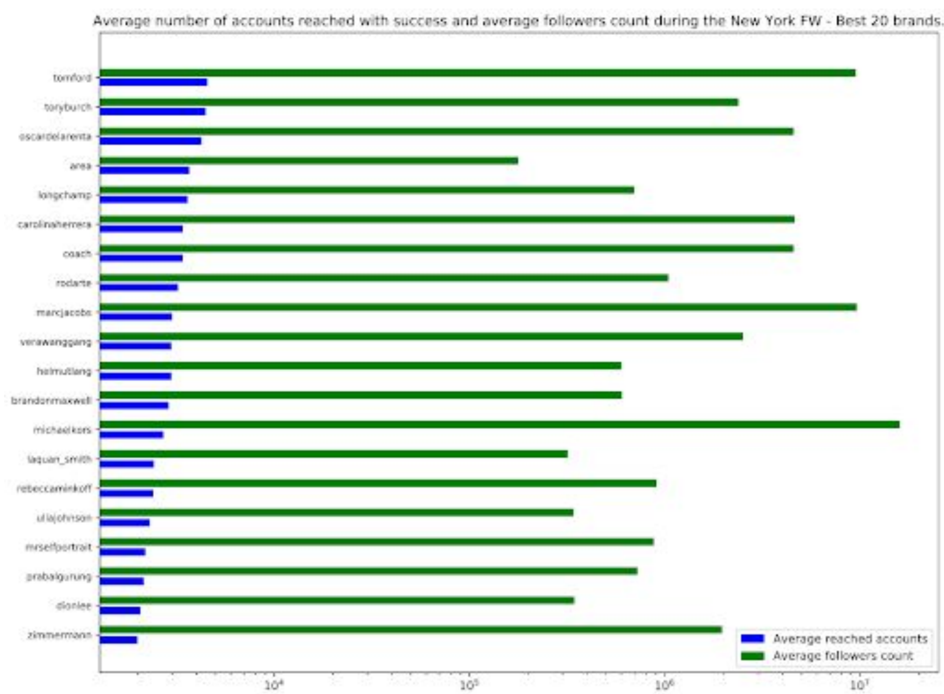
Accounts reached for Milan Fashion Brands

[Fig. 3]



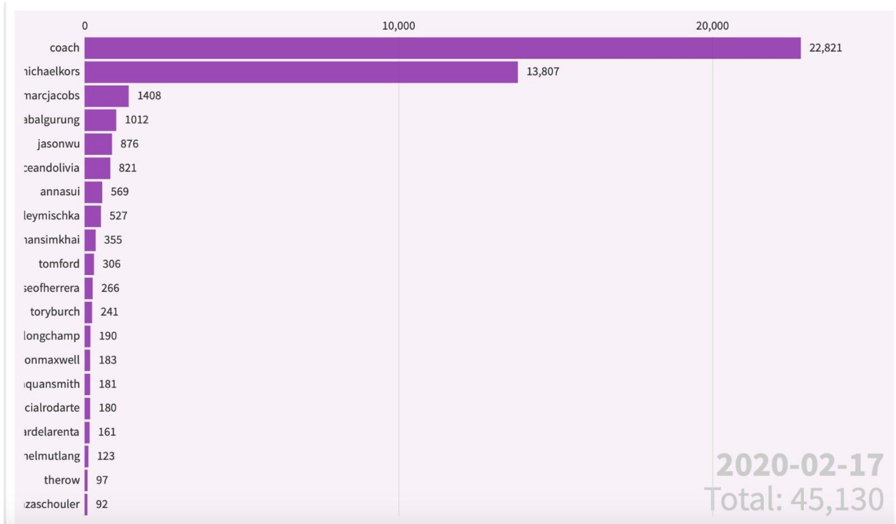
Milan Brands - Reached accounts during FW compared to total Followers

[Fig. 4]



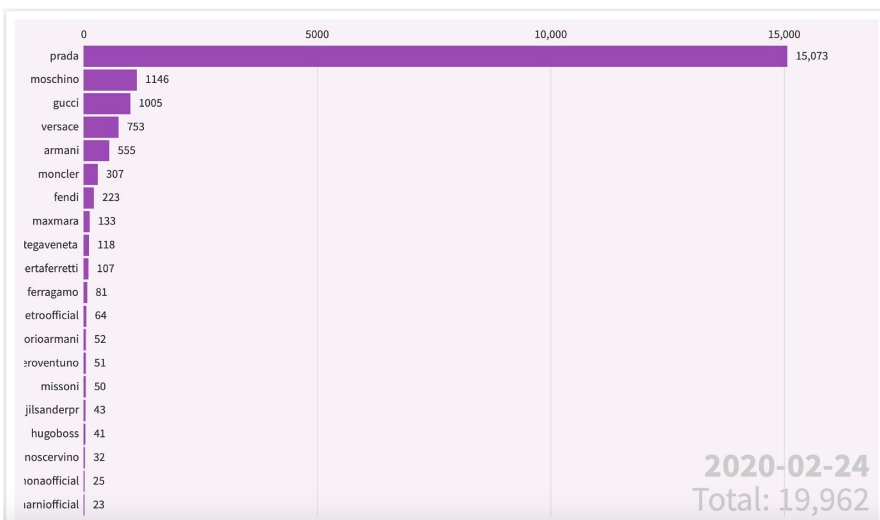
NY Brands - Reached accounts during FW compared to total Followers

[Fig. 5]



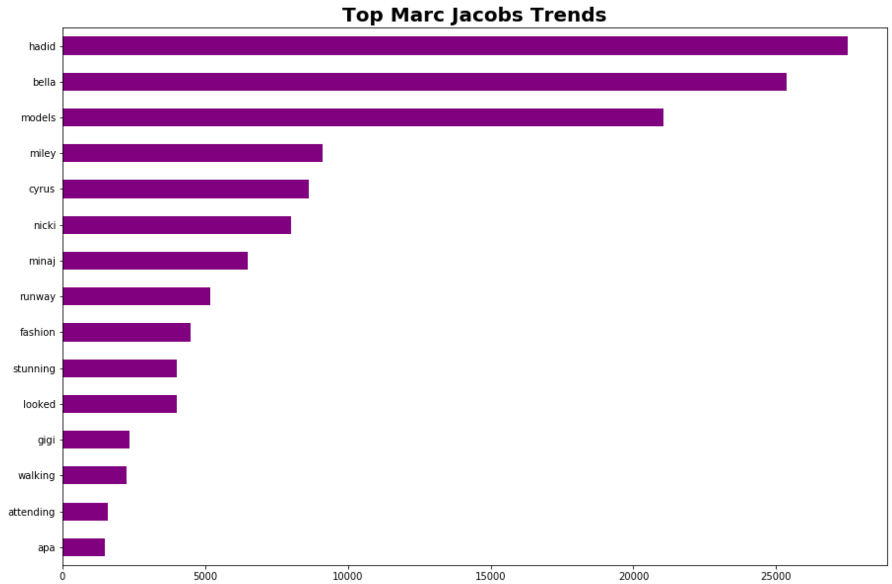
Most mentioned brands NYFW

[Fig. 6]



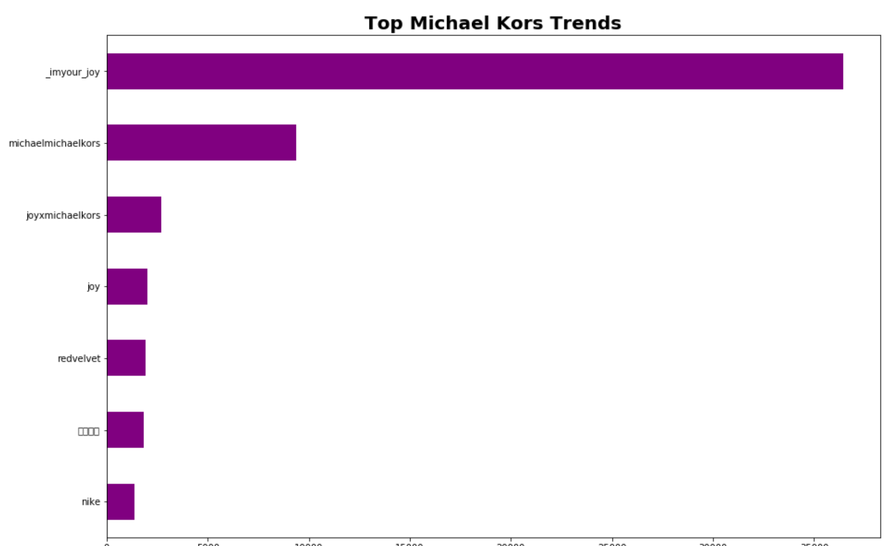
Most mentioned brands MFW

[Fig. 7]



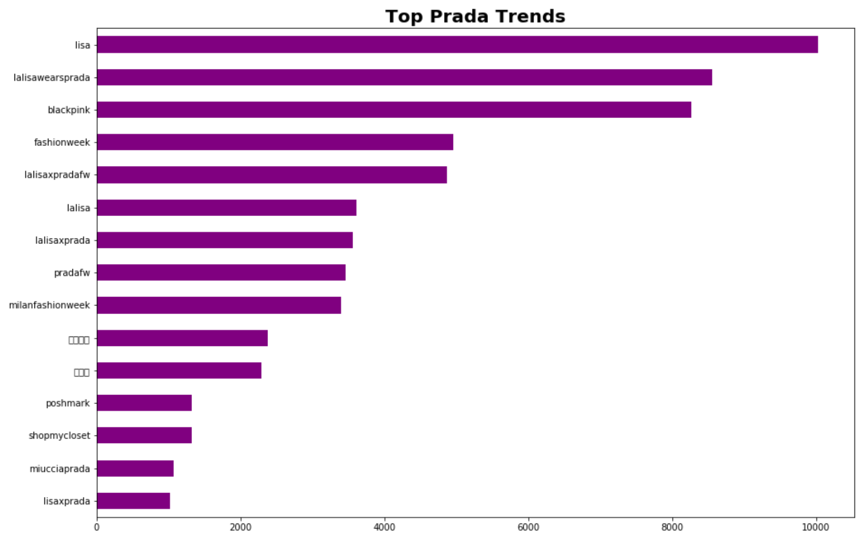
Top mentioned words for Marc Jacobs event

[Fig. 8]



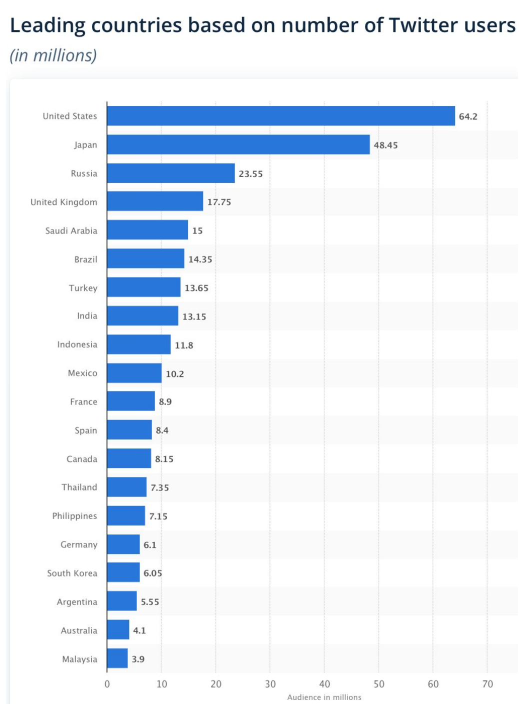
Top mentioned words for Michael Kors event

[Fig. 9]



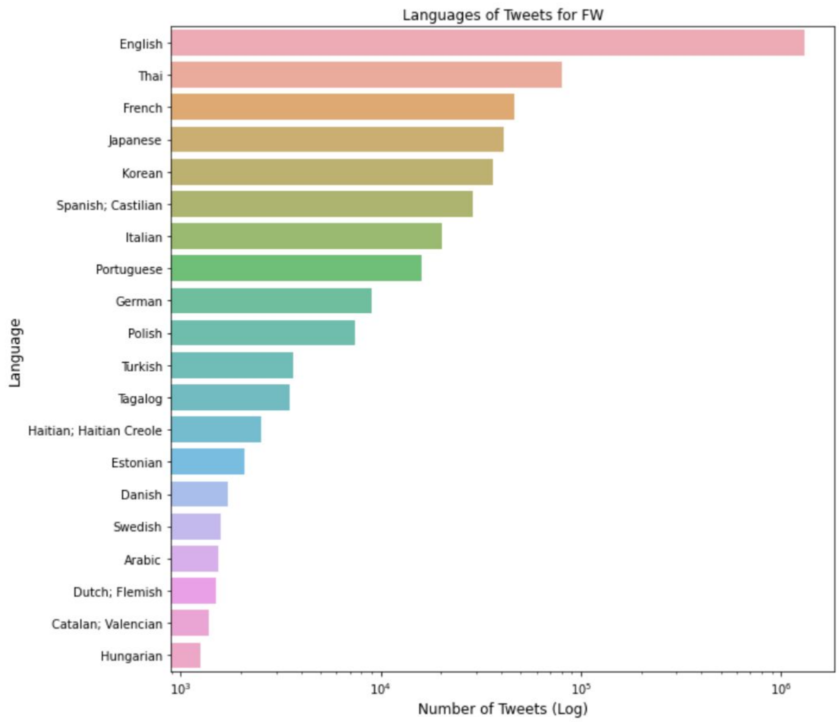
Top mentioned words for Prada event

[Fig. 10]



Twitter population, Source: Statista [9]

[Fig. 11]



Twitter demographics from our collected Tweets during FW

[Fig. 12]

Instance of the Treemap, filtered to MFW, Thai language users, this shows the popular Brands

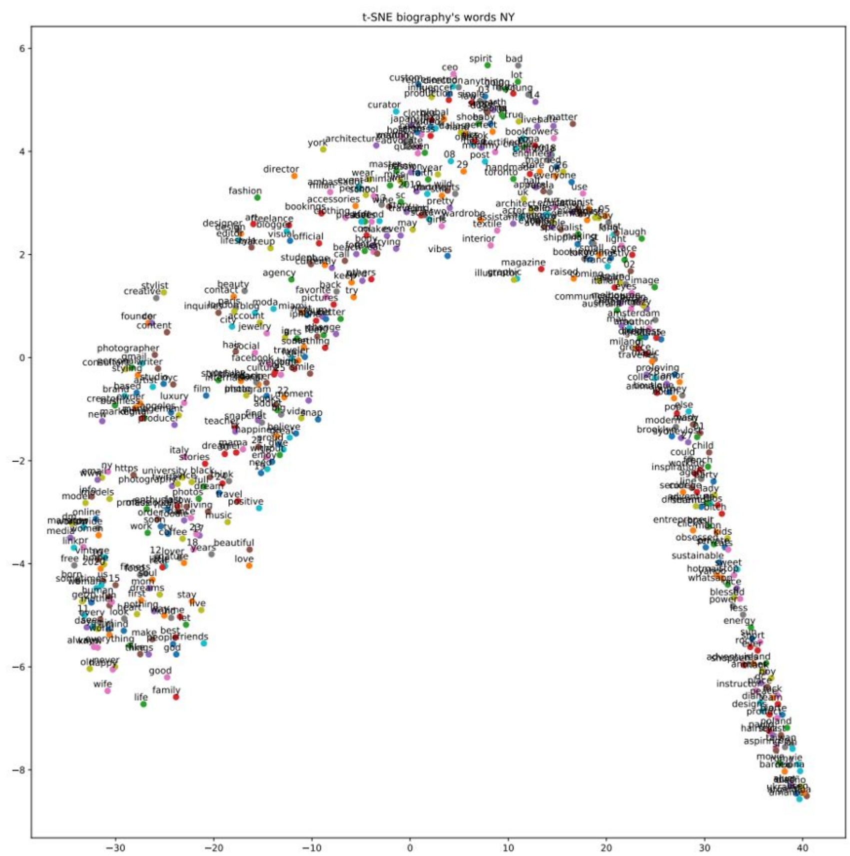
[Fig. 13]

Instance of WebMap showing locations of sample of Tweets from NYFW

Appendix A

In the attempt to find a better embedding of the biography, we performed Word2Vec, embedding all the words, removing the less common and the stop words, in a 100D destination space.

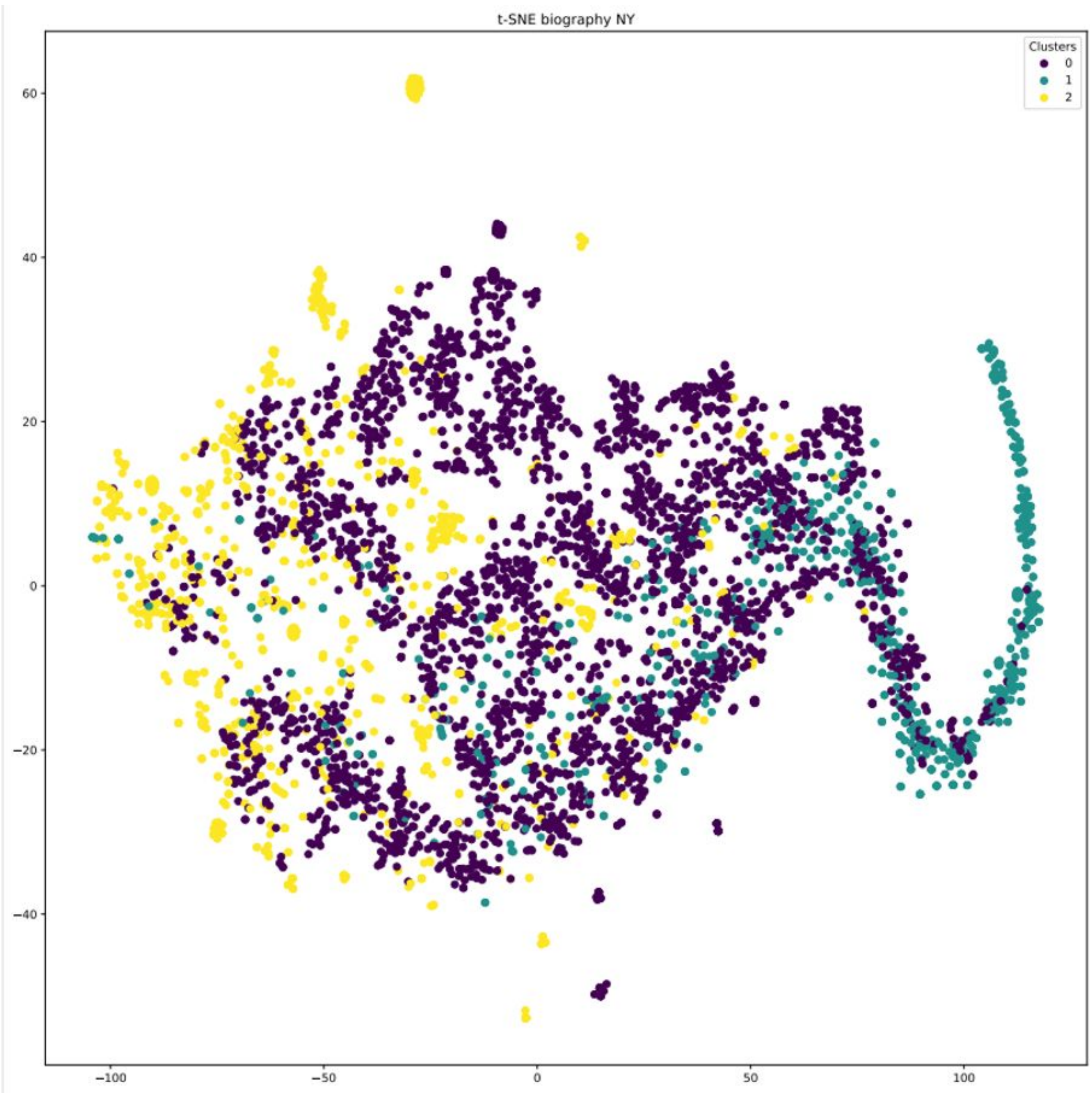
Using a t-distributed stochastic neighbor embedding, it is possible to visualize an approximation of the destination space in 2D:



Words2Vec Instagram biography embedding

The embeddings of the biographies have been found averaging the embedding vectors of each word contained in each biography.

It has been performed a clustering in the destination space, using a Gaussian Mixture algorithm, finding the following result:



GMM of Words2Vec embedding