Hochschule Mainz: University of Applied Science
FB 07 Digitale Methodik in den Geistes-und Kulturwissenschaften
Seminar: Network Analysis in the Digital Humanities
Dozenten: Aline Deicke and Dr. Demival Vasques Filho
Sommersemester 2020

#BlackLivesMatter: Its Linguistic Construction through Twitter Activism

by Magdalena Stumm, Yohan Park and Susanne Schank

Keywords: Network Analysis, Linguistic Research, Twitter, Bigrams, Black Lives Matter, George Floyd, Police

1. Introduction

Police brutality, especially against Black people, has been a problem and topic of discussion in the United States for years (Chaney 481). The most recent case („Breonna Taylor: Timeline of black deaths caused by police“), at the time of conceptualizing the research for this paper, was the killing of George Floyd on May 25, 2020 by a policeman („George Floyd: What happened in the final moments of his life“). This tragic event brought the questions surrounding systematic racism and violence back into the public focus, not only in the United States but in many Western societies („George Floyd: What happened in the final moments of his life“).

Within a very short time, the hashtag *#BlackLivesMatter* was used on the platform Twitter where people tweeted about the death of George Floyd as well as police brutality (Anderson). According to Pew Research, the hashtag was used nearly 48 million times between May 26 and July 7 2020.

The hashtag *#BlackLivesMatter* goes back to the organization of the same name, *Black Lives Matter*, which was founded in 2013 after the African-American teenager Trayvon Martin was shot dead by a neighborhood watch volunteer (Simon) and the offender was acquitted. The main goals of the organization are to fight against white supremacy and institutionalized racism against Black people („About “).

On social media platforms, hashtags are a good way of assigning your posts to specific topics („Hashtag“). Therefore, it is reasonable to use a hashtag as a basis for data material for a linguistic analysis.

Due to the relevance, we chose Twitter as our data source because it is a popular social media platform with 328 million users worldwide (Clement, 2019) and an even larger global reach due to its content coverage on traditional media. and also provides comprehensiveness data and allows for accurate data retrieval. Posts on Twitter are public and do not require private access (e.g. as do Facebook groups or subreddits), which allows researchers to obtain complete data. To date, user-generated Twitter data have been considered as real-time “social barometer” of public opinion. The data offer rich details on contextual factors in social research, including textual information (e.g., the text message on a given matter), temporal information (e.g., the time at which the message is posted), and communication information (e.g., who talks with whom about the matter)

In the period we selected, from May 25 to May 30 2020, we received a tweet amount of 4466 tweets, which served as a basis for a linguistic network analysis.

Since a complete evaluation of all tweets would have been too extensive, it was decided to focus on and evaluate words that occur together within the tweets, so-called bigrams, which now serve as the smallest unit to be examined.

A linguistic evaluation of the tweets is relevant, since the topic has gained worldwide attention and the tweets provide a small insight into how the topic is perceived and spread, since many people have participated in the tweets. Hence the bigrams could serve as an atmospheric picture for the people who used the hashtag *#BlackLivesMatter* on Twitter.

By examining the bigrams, we will investigate whether certain word combinations appear particularly often within the tweets and whether recurring patterns can be found that relate to the topics of racism, police brutality or Black Lives Matter.

The concrete research questions are:

RQ1: How is activism linguistically defined within *#BlackLivesMatter* on Twitter?

RQ2: How is Blackness constructed within *#BlackLivesMatter* on Twitter?

2. From a Hashtag to a Social Movement

A social movement can fit into the mold of different social and cultural categories such as interest groups or protest events. In an attempt to study and understand what a social movement is and what it goals can be, different definitions can be of use. Diani (1), for example, defined a social movement as "consisting in networks of informal interaction between a plurality of individuals, groups, and/or organizations, engaged in a political and/or cultural conflict, on the basis of a shared collective identity", which highlights the organized but informal nature of social movements in relation to a collective nature of social movements. Horn (1) on the other hand defined social movements as "forms of collective action that emerge in response to situations of inequality, oppression and/or unmet social, political, economic or cultural demands" which puts the focus on power relations between socio-economic groups. Considering both these definitions, it can be outlined that, at the very least, social movements entail organization and collective voices to gain a specific outcome. In addition to this, the role of technology in the 21st century, especially social media, can be considered a vital

influence on social movements in order for the collective voices to be gathered, united and amplified. Rheingold stated that the Internet is more democratic than print media and enables anyone to speak to a mass audiences ("Electronic Democracy Toolkit"). While traditional movements relied more on well-known leaders and experts, new media technologies allow users to become not merely receivers of messages but also to be the creators and distributors of messages. This network of messages enables people to share grievances and express their opinions through connectivity (Carty).

An apt example for this process is the hashtag *#BlackLivesMatter* which was created between three friends on Facebook and was later on propelled into a social movement on a global scale as described in the introduction above.

3. Methods

In this section the generation process and description of the tweet-dataset are outlined and the approach for collecting and pre-processing the user-generated dataset to create the final dataset are shown.

3.1 Data Collection and Preprocessing

The Twitter dataset included 4466 English-language tweets which were collected between May 25 and 30. This time period begins with the death of George Floyd, an unarmed black man who died after a white police officer knelt on his neck for almost nine minutes in the US city of Minneapolis. Public reactions to the death of Floyd and police custody, relating to structural racism, emerged quickly on Twitter since then. In order to collect tweets that are associated with Black Lives Matter, the API-based data crawling tool Twint[1] was employed. To extract Black Lives Matter movement-related tweets, we build a set of keywords related to the usage of hashtags in the semantic sentence (e.g., #BLM, #BlackLives-Matter) in both lowercase and uppercase. The collected tweets were restricted to users originating from United States based locations.

Using Python, the data was saved as either CSV (Comma-separated values) or JSON (JavaScript Object Notation) format, then imported as DataFrame - a 2-dimensional

[1] https://github.com/twintproject/twint

labeled data structure with columns of potentially different types using the Pandas[2] library. The next steps consisted of cleaning and preprocessing the text data using the regular expression (Regex) - a special sequence of characters that define a search pattern and help match or find other strings or sets of strings. This process removed URLs, converted words to lowercase, expanded contractions, removed punctuation, and stripped whitespace. Tweets were also tokenized such that contracted forms were split up into their constituents. Finally, stop words were dropped via a list of syntactically functional words (articles, conjunctions, prepositions), to improve precision of information retrieval in our network analysis.

3.2 N-Grams as a Representation of Linguistic Relation

N-grams are contiguous strings of words, that appear in a stretch of discourse and are specified by the number of words of the string in question. N-grams of two words are called bigrams, while N-grams of three words are called trigrams, and N-grams of five words are called fivegrams and so on. N-grams allow the addressing of linguistic relations of co-occurrence among words (i.e., words that tend to appear frequently in the corpus shared thematic meaning). This network study is based on bigrams, which identify types of word co-occurrences and quantify the numbers of tokens of each co-occurrence type. The basic network data set is an n $\times$ n matrix S, where n equals the number of nodes (words) in the analysis and S_{ij} is the measured relationship between nodes i and j with the node serving as the unit of analysis. Here, the nodes are identified based on the weighted frequencies of the words. The measurement of word co-occurrence is the standard for creating links between words in a semantic network.

The network in our study is created using the python-package, Networkx[3].The top 700 bigrams by frequency were included in the network visualization,[4] which is created using the JavaScript library, D3.js[5].

---

[2] https://pandas.pydata.org/

[3] https://networkx.github.io/documentation/stable/index.html

[4] https://monsieur-park.github.io/BLM-Network/

[5] https://d3js.org/

3.3 Representing Bigrams in Network

Figure 1: Visualization of the semantic network from Twitter

In total, 514 nodes, 673 edges were created from bigrams. The average degree of the network was 2.6187 with a network density of 0.0051. The size of nodes indicated graphically how frequently the word occurred. The thickness of each link represented the weight or number of co-occurrence between two words. The more closely related the words were, the shorter the link distance.

The color of nodes reveals the different clusters within a network. The Louvain algorithm[6] segments in our network is used to distinct between different communities (e.g., if there are four different colors in the network visualization, it signifies the presence of four different communities). Network measures, such as degree (the number of links connecting each word), betweenness, closeness and eigenvector centrality were also calculated, while these centralities provide a depiction of how central and how connected words are within the network. A community in a network can also be understood as a "subnetwork" which is structured according to its included nodes.

---

[6] https://python-louvain.readthedocs.io/en/latest/

These are often connected by a number of edges (i.e. generally a high edge density can be found inside a community, while the edge density is low between communities).

4. Results

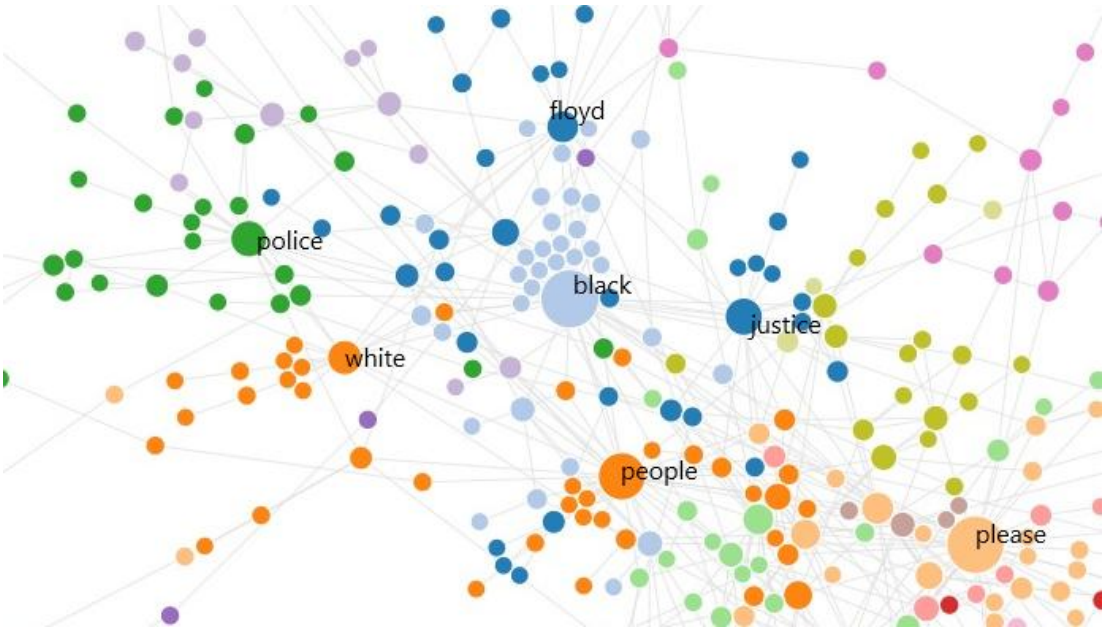

Figure 2: The largest nodes in the *#BlackLivesMatter* Network from Twitter

Since, this paper is limited in its research capacity, the focal point will be put on the three largest nodes in the network ("black", "please", "people") in connection to three smaller nodes ("justice, "police", and "white") as marked in Figure 2 above. Furthermore, it should be noted, that the node "floyd" is also highlighted in Figure 2 but will not be included in this section. Its linguistic contribution and implication will however be interpreted in the upcoming discussion section, wherein Figure 2 will serve as a reference point.

The edges of the three larger and three minor nodes work as the bases for the result interpretation in order to answer the research questions (RQ1 and RQ2) posed in the introduction. Therefore, the composition of the bigrams will be analyzed and in a final step the findings will be put in juxtaposition with one another, so as to compare the linguistic structure in each of the six nodes. In order to interpret the nodes in a linguistically relevant manner, the *iWeb: The 14 Billion Word Web Corpus* from the *English Corpora* which uses data collected from nearly 95,000 websites was consulted.

This source was deemed appropriate because of its collection of web based word usages and should be noted here as the source for the following linguistic analysis.

The node “black” is linked to a mixture of nouns, verbs, adverbs, adjectives, names, prefixes like “people”, “woman”, ”women”, “man”, “men”, “brothers”, “sisters”, “community”, “communities”, “culture”, “ppl” (short for ‘people’), “person”, “life”, “lives”, “americans”, “america”, “murder”, “justice”, “voices”, “friends”, “protesters”, “violence”, “fight”, “support”, “brown” (as in Michael Brown, not the adjective brown); “hear”, “see”, “matters”, “killing”, “stand”, “murdered”, “taken”, “mourn”, “kill”, “owned”, “unarmed”, “innocent”, “towards”, “non”, ”anti”, “trans” as well as the phrase “im” (short for ‘I’m’). The linguistic construction of these bigrams suggests that the node “black” is mostly used as an adjective itself (e.g. in the bigram ‘black culture’) and therefore serves as identity marker which suits to answer RQ2 specifically. As the list above shows, most of the connected nodes have negative connotations (e.g. “murder”, “violence”, “fight”), than neutral connotations (e.g. “person”, “voices”) and hardly possess any positive connotations (e.g. “friends”, “support”), which is an indicator of how blackness is experienced, as well as who or what performs blackness in the United States. When taking into consideration that the hashtag *#BlackLivesMatter* is closely linked as a phrase in itself (“black”, “lives”, and “matter”), the close surrounding bigrams can be read as an indicator that the movement highlights the systematic violence and repeated traumatic injustice experienced by black people (e.g. following the nodes “black”, “man”, and “murdered”).

Notably, “black” is only directly connected to the minor node “justice” and the larger node “people”, and through several different nodes indirectly linked to “white” but there is no connection of the first or second degree to “police”. The connection to “justice” can be interpreted as a literal call for “justice” as its surrounding nodes include for example “demanding”, “demand”, “get” “deserves” and “668366” (the number to which U.S. citizens could text the word “justice” to sign a petition and support the prosecution of the officer who killed George Floyd). This is an example of how *#BlackLivesMatter* constructed its online activism on Twitter by linking demands to civil action. As established above, being black is linguistically linked to being the victim of systematic violence and this can be juxtaposed whit the white experience in this network. While the node “white” is connected through some neutral nouns (“ppl”, “people”, “person”,

“man”, “reporter”) and even positive ones (“friends”, “protesters”) with the node “black”, the nodes linked to being white also suggests the role of the perpetrator in the *#BlackLivesMatter* movement: “supremacy”, “supremacist”, “racism”, “privilege”, “murdered”, “armed”.

In this context it should be noted that there are two minor nodes that are in close proximity and connected with one another and therefore proof to be of interest: the nodes “police” and “white”. Since, there is no direct link from either of the two nodes to the node “black” but there is a shared bigram construction through the node “racism” it suggests a stark divide in how blackness is perceived in its own community versus the community shared by the police and white people. The exclusion of the black identity but the inclusion of racism can be read as black people having no place in the shared society of white people and (white) police, or as the black identity not being acknowledged and subsequently its struggles ignored.

The node “please” is linked to mostly verbs and only a few names, and nouns: “tweet”, “retweet”, “rt” (short for ‘retweet’) “join”, “spread”, “sign”, “help”, “share”, “right”, “remember”, “stop”, “let”, “unfollow”, “go”, “see”, “keep”, “protesting”, “speak”, “donate”, “wear”, “consider”, “watch”, “hashtag”, “day”, “minneapolis”, “minnesota” “everyone”, “attention”, and “petition”. This node, more than any other, serves to answer RQ1. The bigram construction of please plus verb results in different suggestions like calls for actions (e.g. “speak”), joining (online) activism (e.g. “tweet”), educating oneself or others (e.g. “consider”, “remember”) or a mixture of those in links to third nodes (e.g. “please” “keep” “fighting” or “please” “keep” “signing”). In this stage the *#BlackLivesMatter* movement distinguishes itself by raising awareness and calling for civil action by highlighting what an individual person could do to further the cause. Interestingly, nearly no destructive language was used (an exception in this context would be the verb ‘fighting’). Destructive language would be marked by anger indicators like ‘mad’, call for violence, or the use of, for example, swear words. However, the activism language of the *#BlackLivesMatter* movement is largely constructed by pleas which can rather be interpreted as polite, urgent, and/or desperate.

The node “people” should be mentioned, considering it is one of the largest nodes (and even the largest node when betwenness centrality is selected). However, its linguistic impact cannot be considered meaningful on its own, which can be observed

when viewing the nodes to which “people” is linked. It is a mixture of adjectives, verbs, nouns, names and adverbs which have negative, neutral and positive connotations alike: “white”, “murdered”, “killed”, “minneapolis”, “say” “support”, “fighting”, “need”, “protest”, “help”, “time”, “protesting”, “keep”, “many”, “see”, “around”, “like”, “still”, “unfollowed”, “even”, “start”, “educate”, “lot”, “dying”, “color” and “innocent”. This is due to the fact that the noun ‘people’ serves as a neutral and general description of a collective. Hence, a conclusive interpretation can only be achieved through added adjectives in a bigram (e.g. ‘black’ or ‘white’). At the same time, the betwenness centrality clearly shows that the node “people” is needed to describe experiences; it just does not define those experiences. The interpretation of the nodes above with the help of the node “people” is thus more productive in answering RQ1 and RQ2.

As shown in this section, the *#BlackLivesMatter* activism on Twitter is linguistically defined by its general use of constructive language to call for societal change in the shape of joining (online) activism, calling for civil action as well as raising awareness. This is due to the fact that the black identity, as it is linguistically constructed in the *#BlackLivesMatter* movement, is shown to be the victim of racial injustice through systematic oppression and systemic racism found in connection with police and white people as the perpetrators.

5. Discussion

Finally, the paper’s further research potential can be highlighted in regards to the node “floyd” as shown in Figure 2. As mentioned in the results section above, due to the paper’s limited capacity, the implications of the node “floyd” and more importantly, its links leading to other nodes that form a chain of “breonna”, “arbery”, “tamir”, “oscar”, “eric” etc., could not be analyzed at length. These nodes represent victims of police brutality: Breonna Taylor, Ahmaud Arbery, Tamir Rice, Oscar Grant, Eric Garner and many more (“The killing of Ahmaud Arbery indicts all of us”). While there is no inherent chronological order to these names, one could easily be established and therefore a look into the history of police brutality and the Black Lives Matter movement and its surrounded wording could be analyzed. Keeping in mind, that the linguistic online discourse on Twitter, in May 2020, was marked by pleas and polite word choices, researchers could further look into the history of escalating, deescalating or consistent

language in the Black Lives Matter online activism. They might even be able to establish certain linguistic markers surrounding the victims before 2013 (e.g. Oscar Grant, who was killed in 2009) that could lead to the inception of the hashtag *#BlackLivesMatter*. However, the research is not limited to historic nature but can be tackled from various angles, such as cultural, literary, political and even economical. Arguably, these sciences are intertwined with linguistic research, which is why this paper has discussed some of these aspects, already. A deeper look into smaller nodes and their implications from various point of views could further the broad discourse surrounding the Black Lives Matter movement in a helpful manner and is therefore encouraged.

5. References

- „About.“ Black Lives Matter, blacklivesmatter.com/about/. n.d. Web. 23 Sept. 2020.
- Anderson, Monica et. al. “#BlackLivesMatter surges on Twitter after George Floyd’s death.” www.pewresearch.org/fact-tank/2020/06/10/blacklivesmatter-surges-on-twitter-after-george-floyds-death/. Pew Research Center, 10 Jun. 2020. Web. 23 Sept. 2020.
- “Breonna Taylor: Timeline of black deaths caused by police.” www.bbc.com/news/world-us-canada-52905408. BBC, n.d. Web. 23 Sept. 2020.
- Buckley, Stephen “The killing of Ahmaud Arbery indicts all of us - conservatives and liberals alike - who willfully ignore the systemic questions.“ https://www.tampabay.com/opinion/2020/05/12/the-killing-of-ahmaud-arbery-indicts-all-of-usconservatives-and-liberals-alikewho-willfully-ignore-the-systemic/. Tempa Bay Times, 12 May 2020. Web. 02. Dez. 2020.
- “Bundesweit Großdemos gegen Rassismus.” www.tagesschau.de/inland/georgefloyd-protest-deutschland-101.html. Tagesschau, 6 Jun. 2020. Web. 23. Sept. 2020.
- Carty, V.. *Social Movements and New Technology*. Boulder CO: Westview Press, 2015. Print.
- Chaney, Cassandra, and Robertson, Ray V. “Racism and Police Brutality in America”. *Journal of African American Studies*, vol. 17, no. 4, 2013, pp. 480-505. Web.
- Diani, Mario. “The Concept of Social Movement.” *The Sociological Review* 40, no.1 (1992): 1–25. doi:10.1111/j.1467-954X.1992.tb02943.x
- “George Floyd: What happened in the final moments of his life.” www.bbc.com/news/world-us-canada-52861726. BBC, 16 Jul. 2020. Web. 23. Sept. 2020.
- “Hashtag.” onlinemarketing.de, https://onlinemarketing.de/lexikon/definition-hashtag. n.d. Web. 23 Sept. 2020.

- Horn, Jessica. “Gender and Social Movements Overview Report.” *BRIDGE*. BRIDGE, UK: Institute of Development Studies, 2013. 1-131. *OPENDOCS*. Web. 29. Mar. 2021.
- Rheingold, Howard. “Electronic Democracy Toolkit.” https://people.well.com/user/hlr/electrondemoc.html. 02 Feb. 1999. Web. 29. Mar. 2021.
- Simon, Darran. “Trayvon Martin's death sparked a movement that lives on five years later.” https://edition.cnn.com/2017/02/26/us/trayvon-martin-death-anniversary/index.html. CNN, 27 Feb. 2017. Web. 10 Jun. 2020.
- *iWeb: The 14 Billion Word Web Corpus.* www.english-corpora.org/iweb/. n.d. Web. 27 Sept. 2020.
- Wortham, Jenna. “Black tweets matter: How the tumultuous, hilarious, wide-ranging chat party on Twitter changed the face of activism in America.” http://www.smithsonianmag.com/arts-culture/black-tweets-matter-180960117/. Smithsonian Magazine, Sept. 2016. Web. 29. Mar. 2021.

6. Data Archive and Network

- https://github.com/Monsieur-Park/BLM-Network (Data Collection and Code)
- https://monsieur-park.github.io/BLM-Network/ (Visualized Network)