Facebook Algorithm Exposed

University of Amsterdam - Digital media initiative - Winter School 2019

Or, how to keep your bot alive

Team Members

Giovanni Rossetti, Bilel Benbouzid, Davide Beraldo, Giulia Corona, Leonardo Sanna, Iain Emsley, Fatma Yalgin, Hannah Vischer, Victor Pak, Mathilde Simon, Victor Bouwmeester, Yao Chen, Sophia Melanson, Hanna Jemmer, Patrick Kapsch, Claudio Agosti, Jeroen de Vos

Contents

1. Introduction

This Winter School project will kick off ALEX (Algorithms Exposed), an ERC-PoC funded research project, based at Media Studies (UvA), analyzing social media personalization algorithms through a collaborative data collection method. The test case is based on facebook.tracking.exposed, a browser extension that creates a copy of a Facebook user's timeline, with the aim to generate a critical mass of data for the analysis of News Feed algorithm behavior. The ultimate aim of the dataset being collected thus far is to analyze filter bubbles related to the upcoming European parliamentary elections. Embedded in the broader DATACTIVE project, we understand ALEX as 'data activism in practice', as it involves partners from the civil society and wants to devise a collaborative research method to tackle an issue of prominent societal relevance: democratizing the relation between users and the algorithms upon which their knowledge of the world critically depends.

This report is not mean for publication, it is an internal memo used by students to collect their findings.

2. Initial Data Sets

No existing data was recorded about user activities related to how their News Feeds are created by Facebook's algorithms. Therefore, new users (bots) needed to be created, and a browser extension was used to help collect and record what happens on a bot's News Feed. The data that was scraped contained a timeline ID, whether the impression came from followed pages, whether the posts were about Brexit (political content), Post ID, Link of the impression, Source (user name), description of the News Feed activity, Profile (to which bot the data belongs), and time of the impression on the News Feed.

We set up a clean page and relied on the data that the Facebook algorithm supplied which was captured using the fbTREX tool.

We will be working on an existing dataset of Facebook timeline posts (currently being collected). Each post contains the following metadata:
- post id;
- user id;
- text;
- post type;
- links;
- time stamps;
- n. Likes;
- n. Shares;
- n. Comments

The data are structured in JSON, with option to export in CSV; for more documentation, see the github page.
The dataset is being collected through the browser extension fbTREX (facebook.tracking.exposed), run by a number of volunteers in different countries. The tool acquires only public posts. We aim to adopt and refine a variety of data analysis methods, including descriptive statistics, inferential statistics, data visualization, network analysis. We also deem crucial to raise and tackle emerging ethical issue, devising an innovative data sharing protocol.

3. Research Questions

The research question for this project derives from one research question including two sub-questions leading for the operationalisation of the research:

1. How does Facebook's News Feed Algorithm filter out users experience in relation to (political) public content?
   a. Do 'emotional reactions' influence the ways in which the Facebook recommendation algorithm curates content that populates the News Feed?
   b. How does the Facebook algorithm react to emotion-reactions in friendship networks?
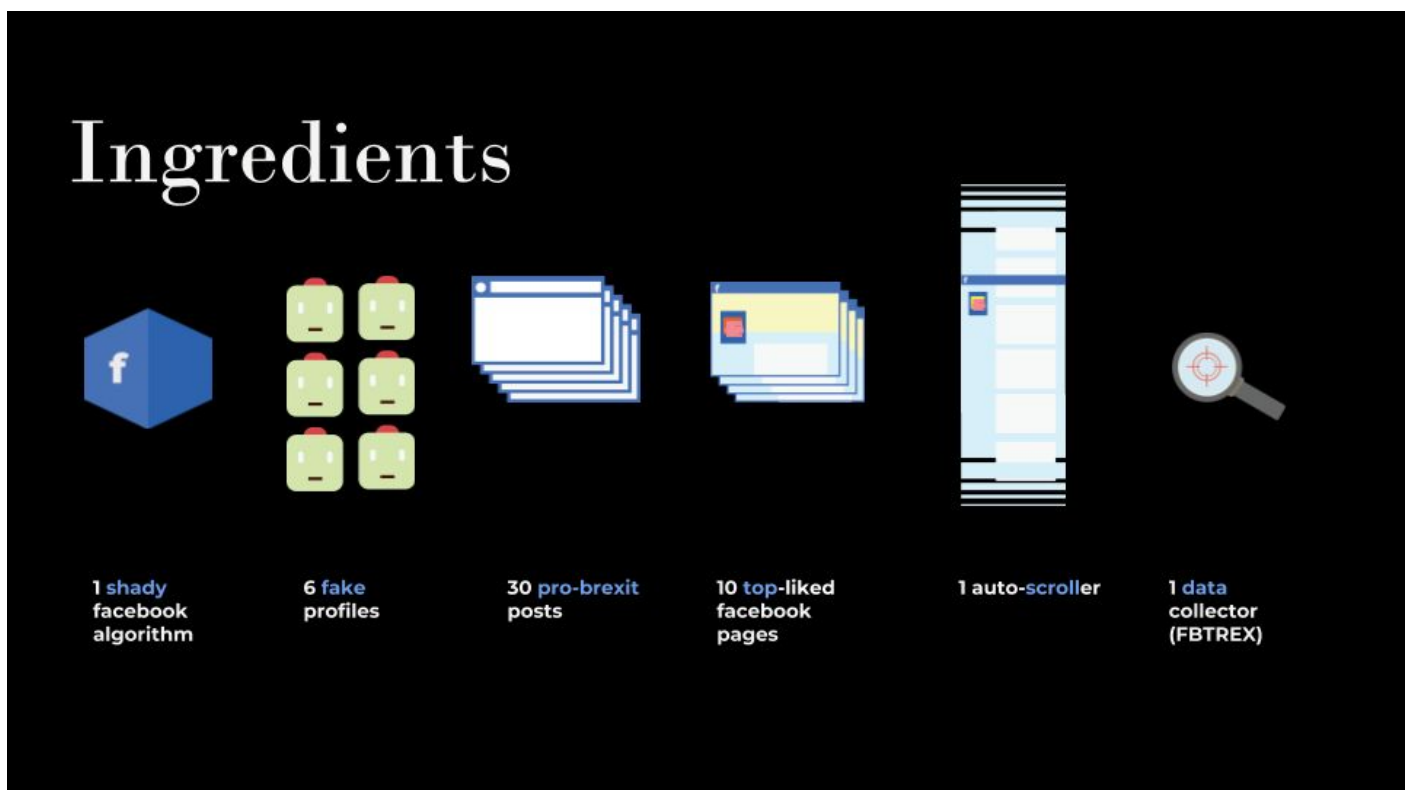
4. Methodology

Figure 1: The General ingredients for research

To inquire the role of emotional reactions in the curating process of a user's News Feed, it is necessary to be able to isolate as many variables as possible and, therefore, gain insights of how a specific setting function in the Facebook 'black box.' To do so, it was decided to employ an exploratory method - creating different bots that allowed us to research the change in their News Feed that followed the bots changes in their 'social behaviour'(reacting with emotions).

The image below depicts the main stages of the methodology. It starts with the creation of three bots, one a test bot, and the other two bots that react differently to political content (Brexit related posts). To analyse a news feed, a news

feed needs to be created, and that was done by the bots following 10 most popular Facebook pages. To explore whether different reaction produce a different news feed, 30 clearly Brexit related posts were chosen for the bots to react upon. An auto scroller was then used to gather data periodically. The installation of the fbTREX tool to the browser created a structured data set. The logic for all the steps is described in more detail below.
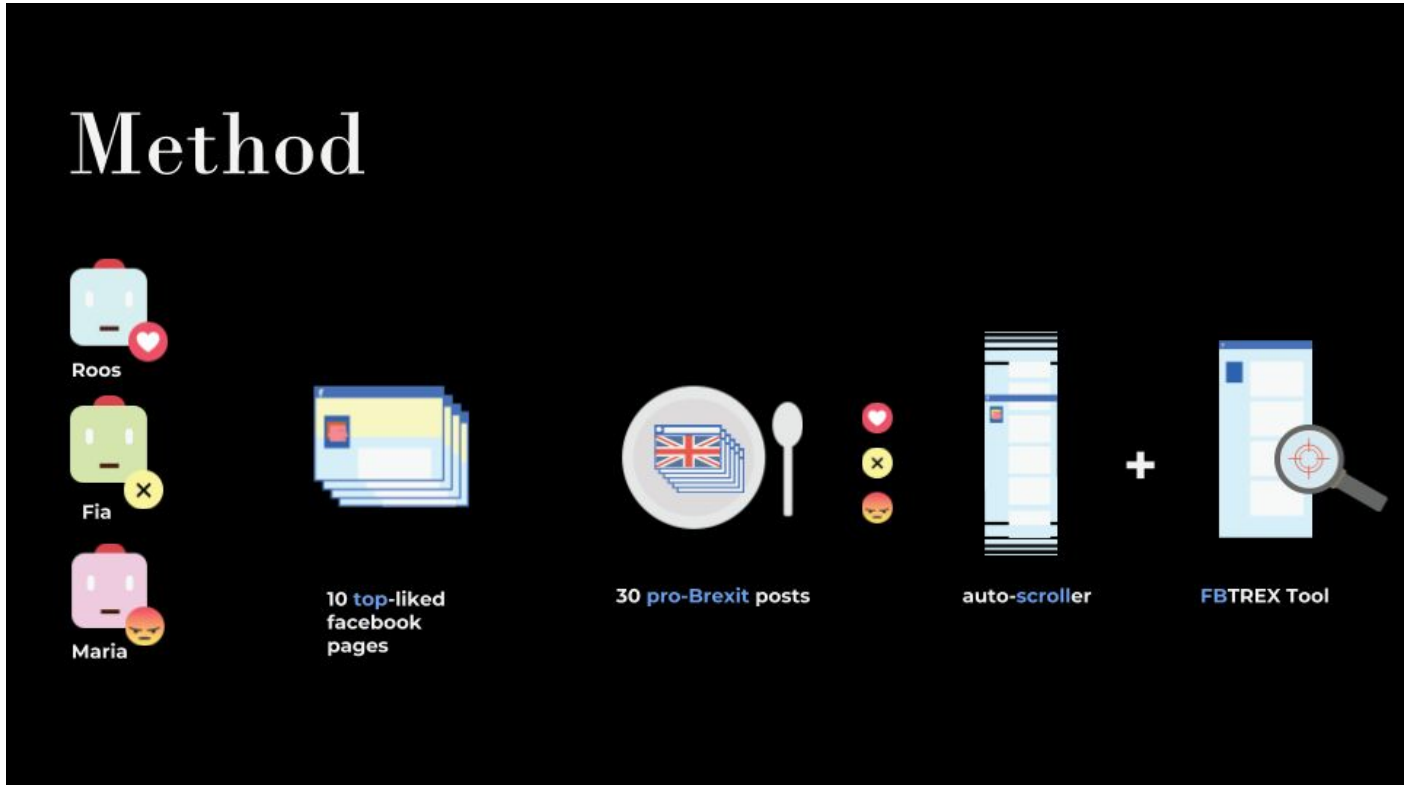

Figure 2: The procedure of research on emotions

1. New bots are born
To explore possible scenarios, we created a test bot before creating the two sample bots. The test bot went through three stages of tests to establish what actions lead to a News Feed (which is important to have when analysing one's curation later). The three stages were 1) reacting with emotions to posts 2) following bot friends 3) following neutral pages. The third step created a News Feed, and that insight was used to start the experiment with our two sample bots.

To avoid the interference from the records of the researchers' previous behaviours, we needed as clean and untainted bots as possible. To that end, we installed anonymous browsers, signed up for two new accounts (a pro-Brexit and an anti-Brexit bot) with new email addresses or phone numbers. Finally, to enable our research, two browser extensions were installed: the browser add-on 'Tampermonkey' was used to run a script that allowed us to automatically replicate the action of scrolling the News Feed; 'facebook.tracking.exposed' (fbTREX), a browser extension that provides all the public posts that appeared on the bots News Feeds, directly conveying all the data into a spreadsheet. In this moment, in order to generate a News Feed for each bot, both of them were made to like the top ten most liked pages of Facebook, that do not present any kind of political affiliation and were therefore deemed as 'neutral.'

Actual experiment
As for the actual test, we compared the differences in News Feeds of two bots before and after they consistently reacted to the same articles with the angry and love emoji. These actions can only indicate bots' political stances when these specific articles are clearly pro-Brexit. Therefore, two rounds of selections were conducted to guarantee the

effectiveness of this experiment. We first selected four pro-Brexit pages and then sampled 30 posts from these pages for bots to react to. Relevance and representativeness were considered in terms of page selection, which mean that these pages were the top results returned by the query "pro-Brexit" on Facebook and followed by more than 100,000 users. The posts were then collected chronologically within the pages and filtered by their relevance and by "pro-Brexit markers". The former referred to the titles of articles that should mention certain keywords (i.e. pro-Brexit, anti-EU, no deal, Brexiteer and leave the EU) and the latter denoted the clarity of their political stance by incorporating "pro-Brexit markers", such as call-to-action words and positive adjectives on Brexit (i.e. great, winning, successful etc.). Subsequently, two bots reacted differently (loving or hating reactions) to these 30 posts at the same time.
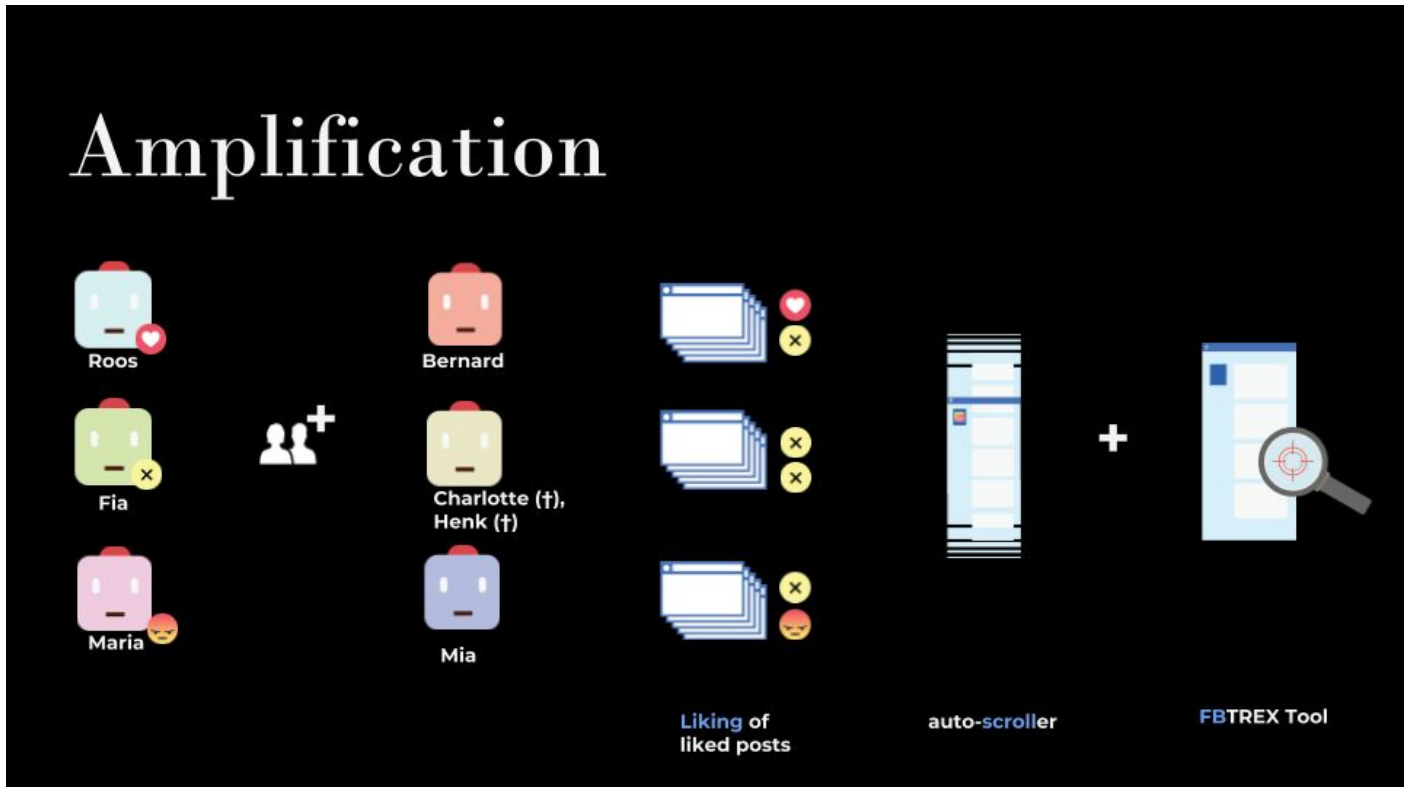

Figure 3: The friend network of bots

2. Amplification: bots Mia, Bernard, Charlotte and Henk

For an additional experiment focusing on amplification, three extra bots were created to simulate a friend network. These bots were named Mia, Bernard and Charlotte (followed by Henk). The goal of this experiment is to research to what extent Facebook is amplifying emotions, and to find if reacting angry to posts would result in a news feed that consists of relatively more posts that are 'angered' on with the like function. If so, in a further research we could investigate to what extent the Facebook algorithm has a radicalizing character.

The Facebook accounts for the bots were created using fake email addresses. All bots were created in an anonymised browser. However, Facebook detected irregularities and blocked Charlotte later that day. Another account (Henk) was created from the same browser that was cleaned and installed again, but was shut down by Facebook the next day, even though we used an unique telephone number to register it. The test bot was not operating and could not be used in the experiment.

The three additional bots friended Fia, Bernard, and Roos to create an initial friendship network, but did not befriend one another. In this way, they could not contaminate each other's news feed. The bots followed the top 10 Facebook pages to generate a neutral Facebook feed.

Figure 4: The neutral pages for bots to follow

Mia was selected to be the angry bot, Bernard the loving bot and Charlotte (later Henk) the neutral test bot. Charlotte, followed by Henk, was passive: the bot auto-scrolled but did not use the 'like' function at all.

Bernard reacted loving to posts that it's bot-friends loved, Mia reacted angrily to posts that it's bot friends hated, in a block of 10 at a time. The posts are here:

https://docs.google.com/spreadsheets/d/1UxYWsYMseQrTy97KmHnDNP8LgPxf0rKOep8rKpHDyp0/edit#gid=0

5. Findings

1. Emotions

Considering this research was conceived and executed within the timespan of five days, the findings are initial and justify further research on this or similar topics. Especially test environments that eliminate as many outside factors as possible should prove helpful in reaching a satisfactory answer to our research question. Separate from the central aim of this research, some interesting findings were that bot creation proved more difficult than initially thought, with Facebook showing quite apt in intervening in the creation process at different steps of the way. For example one of the first test bots was denied access for 24 hours, so caution is needed in replicating a similar bot experiment. Facebook's Help Desk also refers to the need for a user to provide ID in case there is doubt about the actual existence of a user (Facebook Help Center).

Another finding that stood out was that even though the bots are automated to scroll, a lot of the actions require human intervention in order to be used. Examples are the liking or following of pages, as well as reacting to posts with emotion markers. One could argue that the word 'bot' is only partially applicable, as it is much more a joint effort between a human and a bot. Although those findings are not directly related to answering our research question, they are important findings in understanding the inner workings of Facebook more generally. From that point onwards, we could start setting up and priming our sample bots to start collecting data by auto-scrolling the News Feed and extracting the posts using fbTREX.

Results of scrolling and extracting

In order for us to examine whether or not one bot showing a certain 'emotion' to posts would see different content than another bot showing an opposite 'emotion', we first had the bots collect 'neutral' data for a set amount of time, before having reacted to the selected pro-Brexit posts. We had previously established that the News Feed consists almost exclusively of content from pages that the user 'likes' or 'follows'. An interesting finding that we came across whilst setting up our actual research was that the friend bots added almost no additional content to our sample bots' News Feed. Only things such as a friend changing their profile picture showed up on our News Feed, whilst things such as them liking neutral pages, as well as them reacting to the posts of those pages were left out of our News Feed. Both bots seemed to be exposed to the same content, with roughly the same frequency. A necessary observation here is that some of the most popular pages post significantly less than others. For example, during our project the page 'Vin Diesel' posted no new content, whilst the page 'Tasty', which posts a variety of recipes, posts very frequently. No differences between the two bots and their News Feed could be observed at this point.

However, after the bots reacted to 30 pro-Brexit posts and 'followed' the four corresponding pages that had shared those 30 posts, some initial differences between the News Feeds could be observed. In line with our initial hypothesis, the bot that 'loved' pro-Brexit content started seeing posts from pro-Brexit pages more frequently. Consequently, the bot that was 'angry' at pro-Brexit content saw content - both overall pro-Brexit and unique pro-Brexit posts- from those pages less frequently. The pro-Brexit bot Roos saw 16% more pro-Brexit content in its News Feed taking account the overall number of pro-Brexit posts. Further, taking into account only the number of unique pro-Brexit posts, the pro-Brexit bot received 30% more political content in comparison to the anti-Brexit bot.
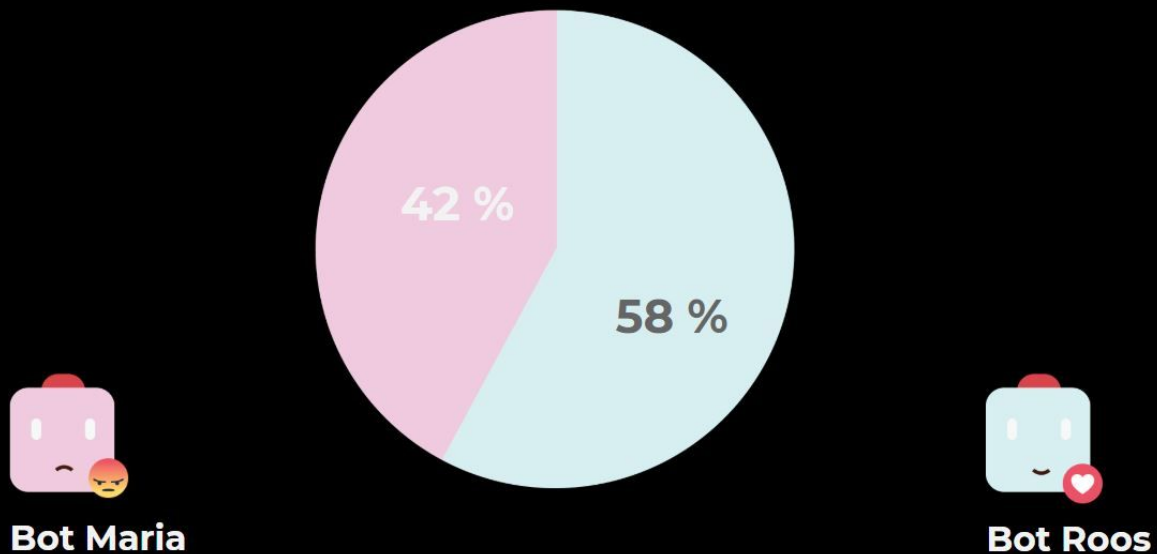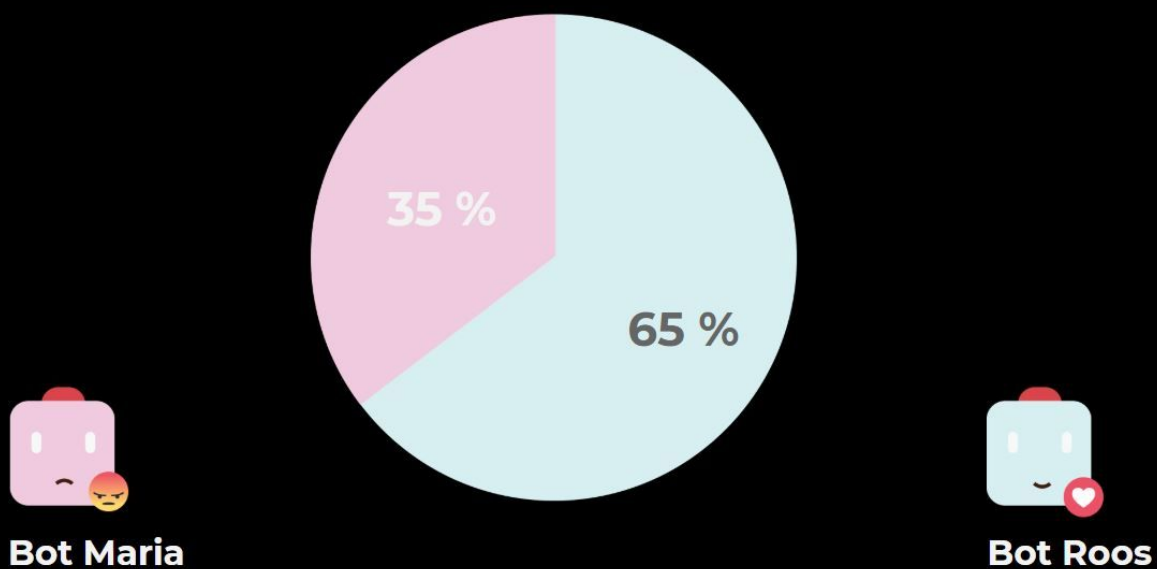
Figure 5: The frequency of pro-Brexit posts in the bots' News Feed (according to the total amount and unique posts)

2. Amplification

The initial finding in the angry bot Mia was that the Facebook algorithm provided more data from the pages most heavily angered, such as Tasty. This may be due to the posting times or the algorithm searching for older posts to keep the persona engaged.
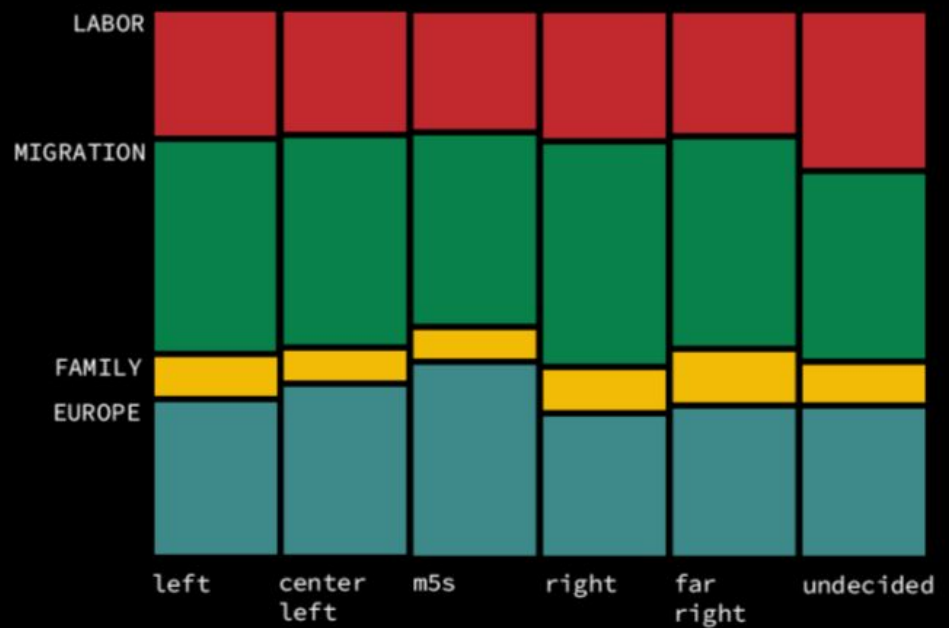
The Facebook algorithm was trying to find friends using recommendations from its social graph. Although no friends were chosen from this, it does show how the friendship algorithm works by looking for friends at one hop in the graph and using the number of links to that node, rather than any qualitative notion of friendship. The assumption is that algorithm is trying to create a wider social network from known links using a social graph. The quantitative aspect suggests that the algorithm could enhance relationships between more connected group members to share data and posts. Although this may be used to boost social engagement numbers, does it also point to a method of creating feedback loops in tight circles, such as radical politics?

The most surprising finding is the amount of friend requests gained. Mia got nearly 50 requests before being closed. Some are from pages that appear to be social engagement to boost friend counts. Some appear to be from accounts linked to pages that were liked, such as players or their family from the team pages liked.

On the initial page, the suggested pages appeared to be based on a geo-location based on IP address. These change slowly with the interactions and friendship links. A longer experiment to watch these changes over a period of time to see what it reflects and how Facebook is trying to guide the reader to pages that it thinks are linked by friendship and interactions.
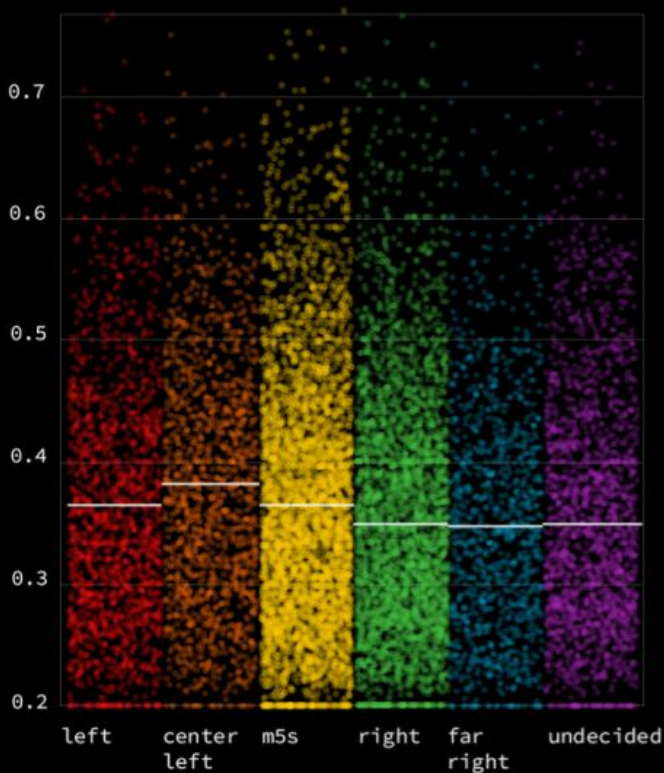
(These analysis have been left without any description)

Method *Semantic Analysis*

Inquiring **semantic similarity** with **doc2vec** (document embedding)

- 🟨 m5s
- 🟩 right
- 🟥 left
- 🟧 center-left
- ⬜ far right
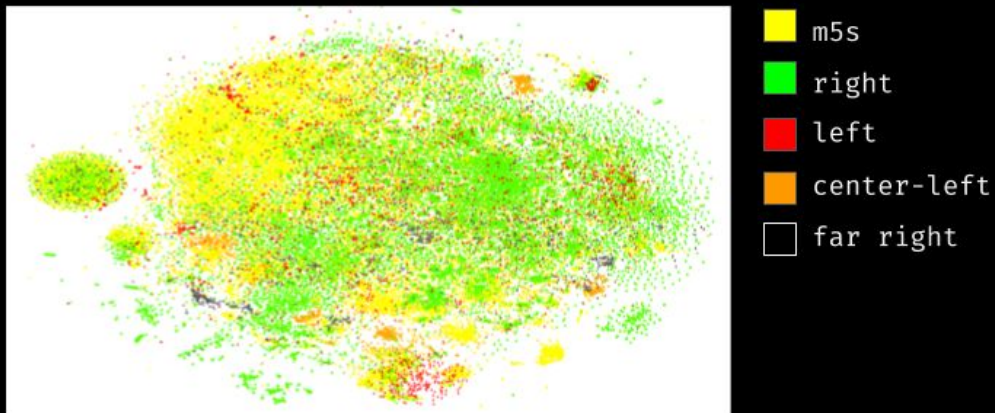
Results

- 🟨 M5s sources
- 🟦 M5s impressions (filter bubble)
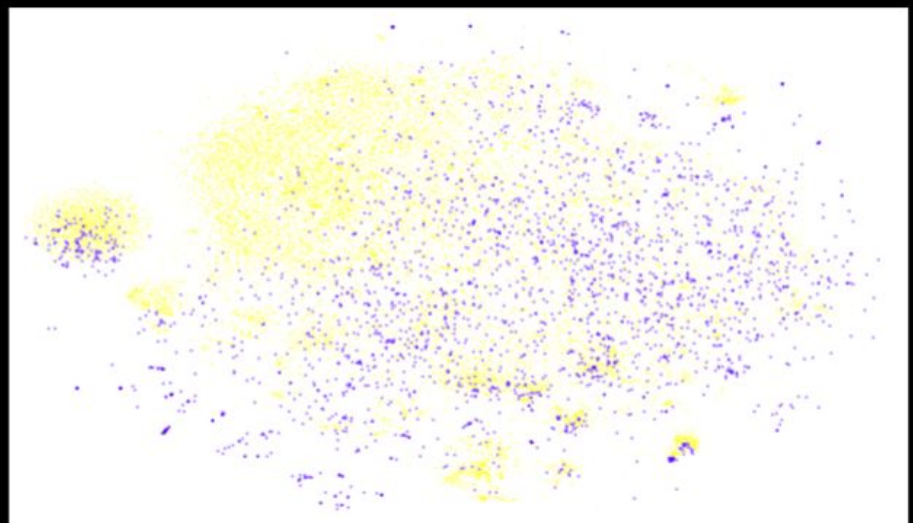
6. Discussion

Some reflections:
- Facebook's algorithm requires many variables, Which makes it hard to control your research setup
- There are many levels of obfuscation Facebook put in place, the only good research Facebook wants is the one which do not …
- It questions the role we can have as researchers questioning the politics of platforms with such public importances.
- Is it our role to break the ToS and prioritize research?
- Is it Facebook's responsibility to allow structural research to take place outside their own commercial realm?

Although the project is exploratory and the results need replicating with more time necessary for collection of data in the future, there are two initial findings that could represent a trend for what Facebook's algorithms in relation to emotional reactions stand for. The first observation is that the emotions markers became significant (in relation to News Feed filtering) only if associated with pages that are already followed/ liked. Borrowing from Bucher (2012), it can then be said that emotional reactions towards separate topical posts (pro-Brexit in this case) do not result in enough 'edge' for the algorithm to start offering more similar content or content from similar pages. News Feed functions only when a user actively 'follows' or 'likes' a certain content provider, thus 'edge' of emotional reactions is created together with a clearer link towards content.

The second observation is that when users (bots) already 'followed' pro-Brexit pages, the different reactions (love or hate in this case) did create a different News Feed meaning that our pro-Brexit bot started to receive slightly more pro-Brexit content into its News Feed than our anti-Brexit bot. This result can be seen to be in correlation with what Marwick says, namely, that platforms provide content that will keep users on the platform (503), so if one reacts to certain content positively, one is more likely to receive similar content in their News Feed later.

Further, the algorithm seems to semantically differentiate positive and negative reactions in providing content as our pro-Brexit bot did receive more Brexit related content. Our anti-Brexit bot had the same number of emotional reactions to the same content, and still received fewer Brexit related posts overall which suggests that reacting angrily can decrease content that has been angrily reacted upon. This is an interesting finding in relation to the notion of discourse creation by algorithms. These editorial decisions by an algorithm can directly influence possible future engagement with content meaning that if a user receives less content on a topic, their ability to interact with it is limited. It raises a question whether the angry reaction does really mean that the user wants less of similar content, they might not. This decision seems to be made for the user and signals the importance of algorithms in designing the way users interact with posts.

In addition, though this experiment was brief and exploratory, the initial findings on a contentious and polarising issue such as Brexit, also point towards Marres's (2018: 430) statement saying that platforms tend to be tilted towards exposing users to more sensationalist material. This, of course needs to be tested further, as the pro-Brexit pages that were in the end followed by our bots were not neutral pages and had a political sentiment. What can be repeated though, is that the black box of Facebook's News Feed algorithms treat users differently, and thus have the potential power to create filter bubbles. That said, a further exploration on what lies behind content curation is important in identifying how platforms shape broader discourses and how new power dynamics could be installed through the dominance of social media platforms.

Amplification
As we see in the results from angry bot Mia, the emotional response may have an effect on showing posts from friends and also sharing posts from friends. The new data arriving came from the increased interaction with news items. In part, this does show the algorithm working as suspected. It might be interesting, from a filter bubble perspective, to consider how the two forms of timeline - Most Recent and Top Stories - alter the data that is received. This could be

easily tested by using the auto scroller and fbTREX extension on the two URLs over a period of a week to see how that part of the algorithm changes data shown or creates a filter bubble.

The act of liking a page requires human intervention as might bots for other interactions. This raises questions of reproducibility and the sustainability. Once data is collected, then how can the research be repeated by others and can the method be reproduced or even sustained over a long term research project. The very act of interacting brings the human into the algorithmic loop, using them as a tool to act with the required variables. Arguably, this reveals Facebook's requirement on the human to derive its data, in particular the meaningful interactions.

The surprising aspect was the apparent use of the friends graph to create new links. This may be inline with a business model of sharing data but it appears to be based on a quantitative metric: the count of inward links. This suggests that there is a network of bots that either scrape pages to harvest the names of using an emotional response or using the API to identify accounts that like a page or content owned by one account and shared with others. It would be good experiment to accept the friend requests to see how they affect the graph and what else they generate. From this, it may be possible to begin identifying other bots, uses of the API, or whether troll farms are more common than thought. This may be a new study but it would also require humans work with machines for the parts that cannot be presently automated, as well as raising ethical issues.

As identified by the Spotify Teardown team, the ethical issues here need to be thought about. Clearly the terms and conditions are being broken but Facebook has blocked ways of trying to game their system. The emotion buttons will only exist when a human mouses over them and friend requests, in theory, have to be entered manually. There is also a checking process for accounts being created. However there does appear to be a bot community on the site and it is trying to use the emotional responses to game the system. I would suggest that this community is explored to see how it is using the algorithm and, potentially, the social graph to create a community. This may require new methods, depending on how far the experiment is to go, but more urgently it requires extensions to the ethical framework. This is needed not only to support researchers who have to benignly break the terms of service but also to consider the ethics of using bots to engage. One might hope that it would be with other bots, but potentially it may come across a person. Does this come under covert methods sections for ethics committees? Reflexively, does the benign researcher become a benign hacker if similar methods end up being employed? These ethical issues may only become more knotted as sites and methods more complex.

An epistemological question arises from such research. Given the secrecy surrounding the algorithm, how can we contextualise ourselves and the resulting research or ever know in what algorithmic period a piece of research applies to? How are these periods constructed: through major and minor changes? Does this not also apply to the tools used to study them? As such information is an industrial secret, then it can only be guessed at, leaving the historical periodisation of the algorithm open to change and re-interpretation.
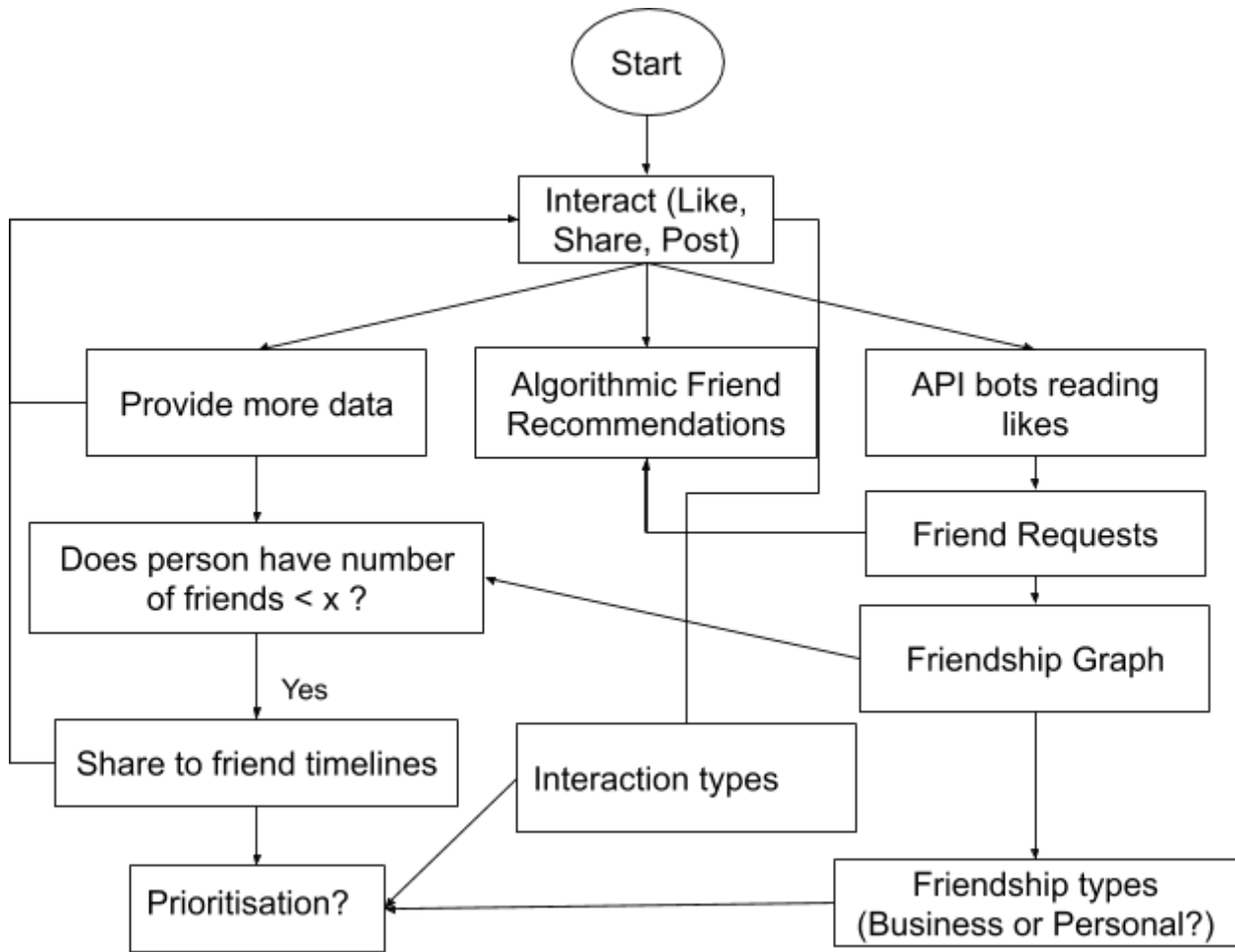
Figure 6: The flow that was observed

When a page is set up, the bot begins to interact with data either by liking pages or other interactions. While this is happening, Facebook is trying to find potential friends. When the page is new, one assumes that Facebook is reading the data and interactions to determine the potential friends. The Facebook algorithm prioritises comments over likes (and similar emotional interactions (Hootsuite 2018)) for posts) and shares. Likes do boost a post's reach but may not amplify as much as one hopes. As the social graph builds, this is prioritised over the data to determine graphs. Bots are also reading the API and using this to make friend requests based on the interactions. We show an initial flow in Figure 1, while also posing some questions that may need consideration, such as does Facebook infer types of relationship and does the size of the social graph affect the data being shared, or is this affected by the type of New Feed view?

Through both, the underlying social graph is increased but one might question at what cost? The bots are commercial and aimed at building their traffic so the interaction is perhaps less meaningful and the algorithmic recommendations will be guesses until a large enough graph can be built to recommend friends who exist at one hop difference.

By this, we mean the Facebook knows that Mia and Fia are friends and the Fia is friends with Roos, therefore Mia should be friends with Roos (Figure 2). This does not take into account the personal interactions by the original poster. In part, this might also be done by the size of the mutual links as well, suggesting a quantitative approach to friends rather than a nuanced reading of the graph.
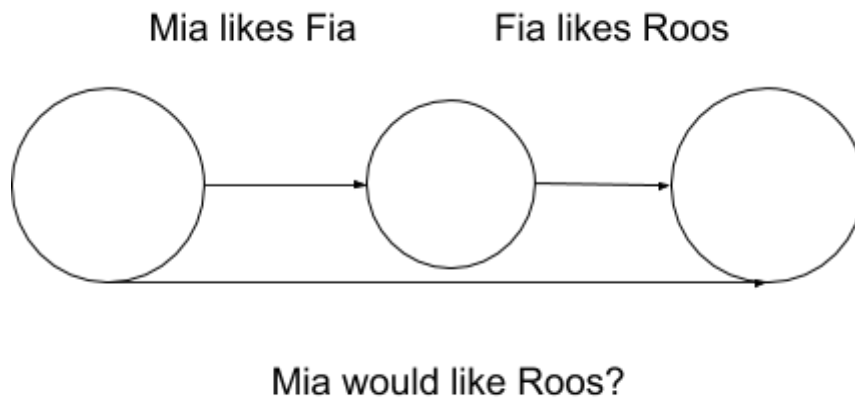
Figure 7: Recommended friends

As the interactions take place, more data is added to the timeline for interaction. We found that the data was added after interactions and a new question comes from this: does the amount of data echo the amount, pace, and type of interactions? A bot might be created to test this action and result.

It might thought that the meaningful applications move towards the building the wider social graph, so making Facebook's connections denser.

7. References

Bucher, T. (2012). Want to be on the top? Algorithmic power and the threat of invisibility on Facebook. *New Media & Society*, 14(7), pp. 1164-1180.

boyd D. and Crawford A. (2012). CRITICAL QUESTIONS FOR BIG DAT. Provocations for a cultural, technological, and scholarly phenomenon. Information, Communication & Society. 15(5).

Marres, N. (2018). Why We Can't Have Our Facts Back. *Engaging Science, Technology, and Society*, 4, pp. 423-443.

Marwick, A.E. (2018). Why Do People Share Fake News? A Sociotechnical Model of Media Effects. *Georgetown Law Technology Review*, 474. Retrieved Jan 14, 2019, from https://georgetownlawtechreview.org/why-do-people-share-fake-news-a-sociotechnical-model-of-media-effects/GLTR -07-2018/.

Pariser E. (2011). The Filter Bubble. What the Internet is Hiding From You. Viking - Penguin Books.

Pasquale F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard Univ. Press.

Sandvig C., Hamilton K., Karrie K., and Langbort C. (2014). Auditing Algorithms. Research Methods for Detecting Discrimination on Internet Platforms. Paper presented to "Data and Discrimination: Converting Critical Concerns into Productive Inquiry" a preconference at the 64th Annual Meeting of the International Communication Association. May 22, 2014; Seattle, WA, USA.

Hootsuite, 2018. https://blog.hootsuite.com/facebook-algorithm/, last accessed 11th January, 2019