



DIGITAL PROPAGANDA OR 'NORMAL' POLITICAL POLARIZATION?

Case study of political debate
on Polish Twitter
(IX-X 2017)

Summary

Digital propaganda or ‘normal’ political polarization? Case study of political debate on Polish Twitter (IX-X 2017) Summary

Authors:

Katarzyna Szymielewicz, Karolina Iwańska

Contributors:

Weronika Adamska, Maria Wróblewska

Collaborating researchers:

Emad Khazraee, Paweł Popiel

Data visualization: Marcin Antas

Cover: Nadia Snopek

Typesetting: Marianna Wybieralska

Panoptykon Foundation | panoptykon.org

March 2018

CC BY-SA 4.0



This is the abbreviated version of the report produced in Polish to summarize findings of the “Polish Political Debate on Twitter. Case study” research supported by the Internet Policy Observatory at the Annenberg School of Communication, University of Pennsylvania.

Table of contents

INTRODUCTION 4

KEY FINDINGS 5

I. Who's talking to whom? Who is creating trends? 5

Topics and engagement 7

Counter-narratives 8

II. Was the political debate on Twitter subject to manipulation? 11

Quality, not quantity 14

Suspicious accounts 15

Accounts that disappeared 17

CONCLUSIONS 18

At first glance 19

Going deeper 19

INTRODUCTION

It goes without saying that the new technologies which we use on a daily basis have a big impact on our social and political lives. Some correlations are palpable for us. We can tell that the public debate in newspapers differs from the one on Twitter, ruled by short and fast statements. We also understand that different issues arise from the disparate roles of traditional editorial boards and social media outlets. The former are responsible for the content they publish, while the latter have no established editorial policy as they use algorithms to provide their users with the content those users want to see.

We are right to be worried about the polarization of public debate, the radicalization of opinions, populism and filter bubbles. All of these phenomena can be used for propagandist purposes. However, we have to be cautious not to confuse observations with explanations.



At Panoptykon Foundation we came forward with the hypothesis that perhaps it is not entirely the new technology that is to be blamed for some unexpected political changes, such as the rise of populism. Rather than this, our assumption is that the emergence of what Eli Pariser called 'filter bubbles' reflects social and political dynamics and is therefore dependant on people, not the technology itself.

We decided to examine this hypothesis in the Polish backyard in collaboration with experienced researchers from the University of Pennsylvania - Dr Emad Khazraee and Paul Popiel.

KEY FINDINGS

I. Who's talking to whom? Who is creating trends?

Based on the social network analysis, we detected two main groups of accounts taking part in the political debate on Twitter between the 23rd of September and the 22nd of October 2017: the first one consisting of 13k accounts, and the second consisting of 10k accounts. After the examination of tweets posted by the most influential users in both groups, we drew the conclusion that the first group clusters actors with conservative leanings, who often **support the ruling government** (on graphs marked with **green colour**), while the second consists of users with liberal and left-wing leanings, far more **critical of the government's policies** (on graphs marked with **violet colour**).

To make it simple, we will refer to the first group as the one with **conservative leanings** and to the second as the one with **liberal leanings**. At the same time we want to make it very clear that our data does not allow us to make strong judgements on opinions and sentiments of all actors that were clustered (as a result of social network analysis) in each group. All we can say is that they communicated more often with influencers showing certain political inclinations (liberal or conservative). At the same time we cannot exclude that there are actors who were clustered in the group with conservative or liberal leanings only because they frequently engaged in the criticism of the views shared by other members of this group.

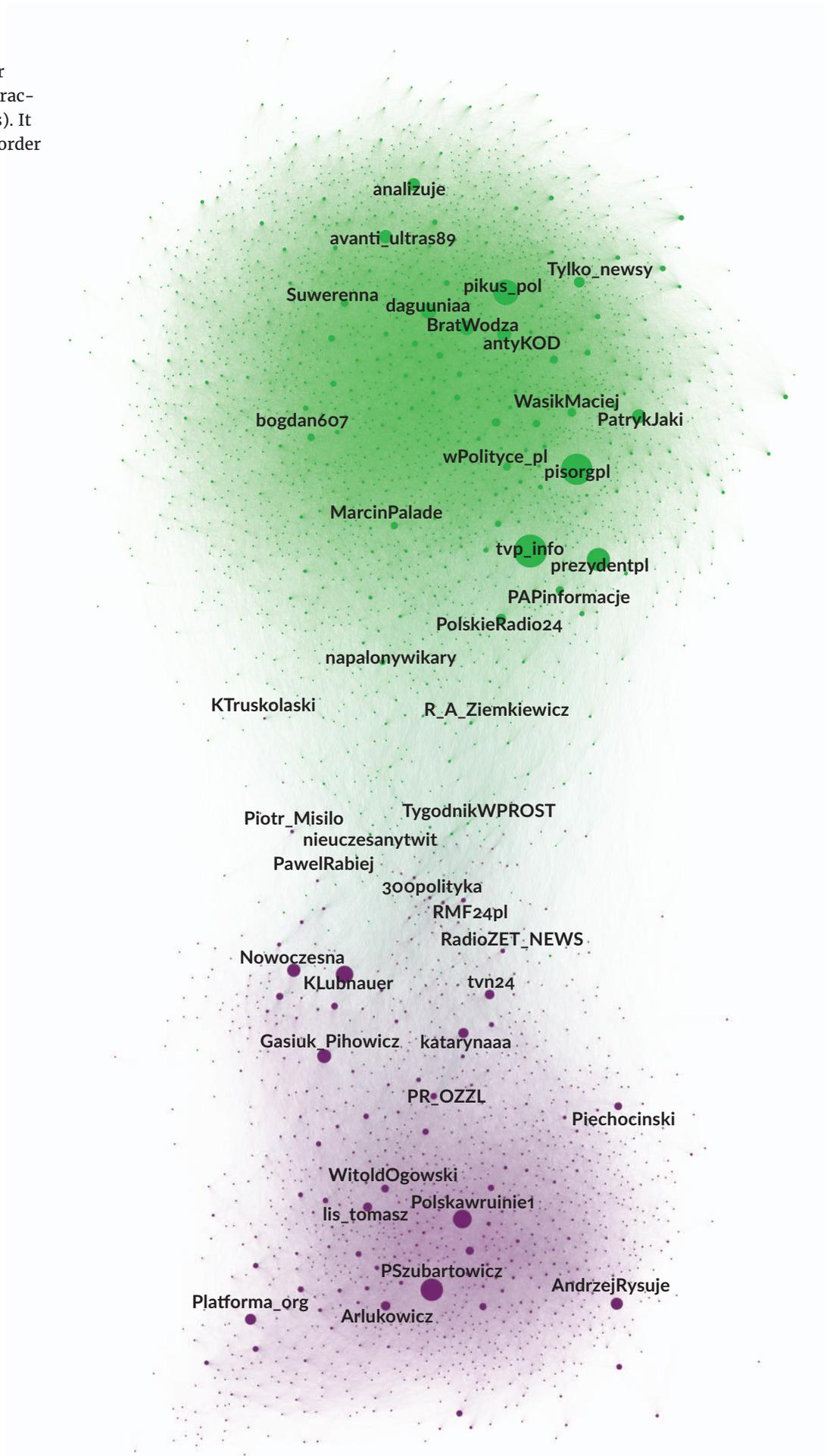
Defining terms – bots and false amplifiers

In a broad sense **a bot is** a computer program which interacts with other users on the basis of an algorithm. An automated account that mimics users on social media platforms. Bots can play a useful and supportive role (e.g. chatbots which give some simple advice or help installing an application). However, in the discussion on computational propaganda, bots are often described as so-called **false amplifiers** – programs which secretly control social media accounts and use them to disseminate certain content (e.g. propaganda or disinformation).

The term 'bot' is increasingly employed to describe semi-automated accounts, also known as **cyborgs**. Humans play an important role in their functioning: they create promoted content and define the activities of accounts controlled by them.

Having said that, what we really see in social networks are fake accounts, either hacked (as a result of identity theft) or artificially manufactured. Bots and cyborgs on the other hand are responsible for entire networks of connected fake accounts. This is why in our analysis we refer mostly to this kind of hacked or manufactured account and not bots/cyborgs which control them.

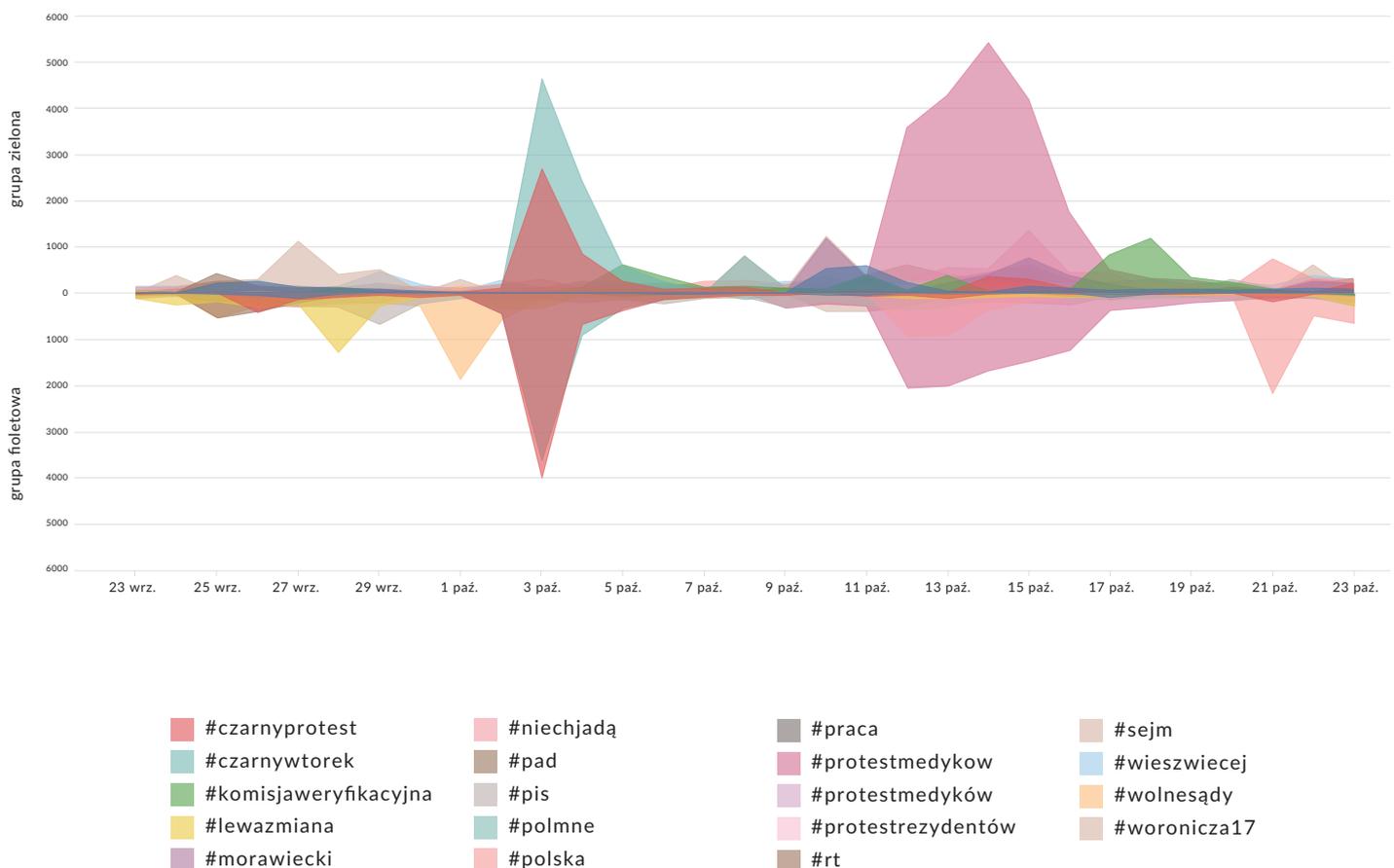
Graph 1. Network analysis for accounts with at least 50 interactions (posts, retweets, replies). It shows most influential and border accounts.



Topics and engagement

The most popular hashtags and keywords used during our period of interest reflected the most important political events in Poland of that time and corresponded to the hottest topics that were also receiving extensive coverage in traditional media, social media platforms (mostly Facebook), and in the streets. Examples include: women's protests against legal ban on abortion (#czarnyprotest), junior doctors' protest (#protestmedyków) and civic reaction to controversial reforms of the judiciary (#wolnesady).

Graph 2. Top 20 hashtags for both groups



Counter-narratives

Thanks to the topic modelling we were able to distinguish some counter-narratives within two key hashtags (#czarnyprotest, #protestmedykow). These counter-narratives constituted deliberate attempts at changing the course of discussions by shifting it from its original political demands to other topics. What enabled us to see such attempts was an emergence of keywords evidently detached from the mainstream discussion which – in the case of both hashtags – related to political demands: access to abortion and contraception in the case of the women’s protest (#czarnyprotest) and an increase in public spending on health services in the case of the junior doctors’ protest (#protestmedykow).

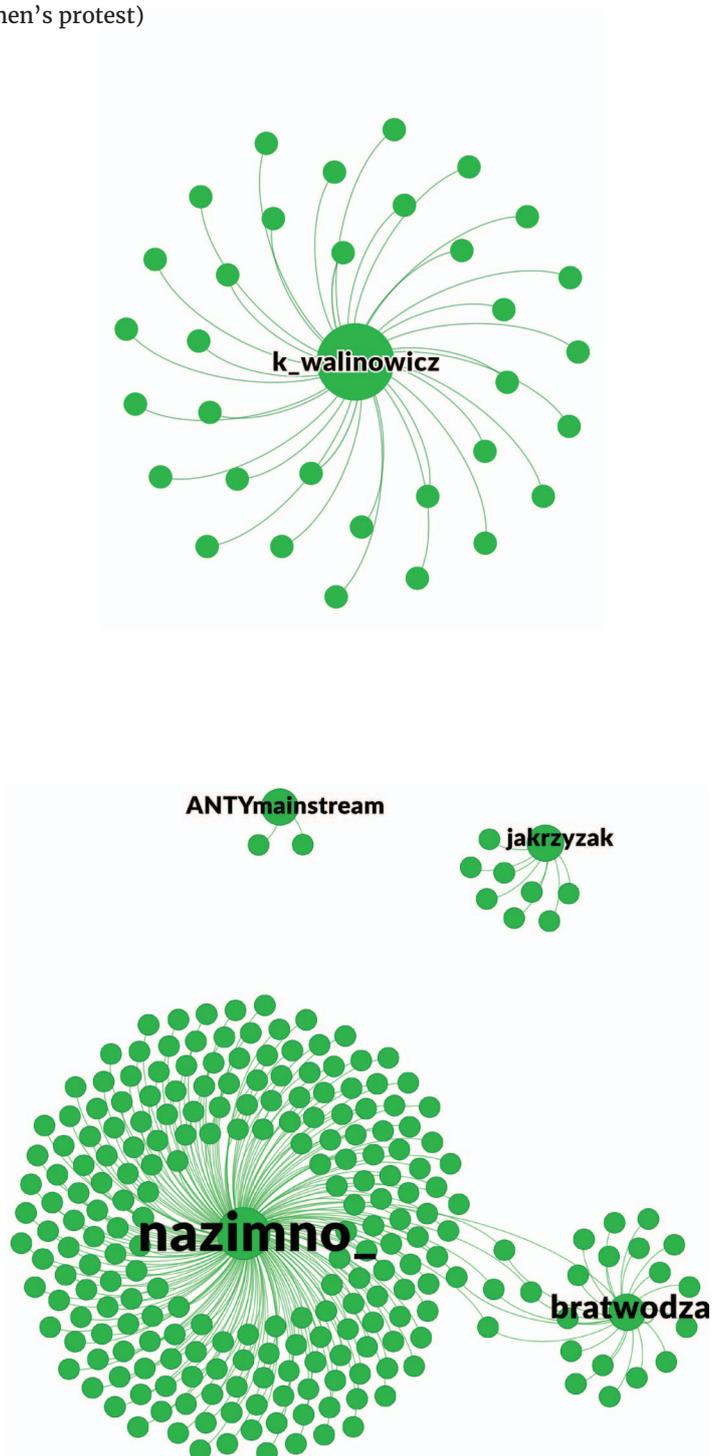
Depending on their political leanings, users either expressed their support for the protests and demands or were critical of them. The tweets we analyzed were often emotionally charged, and employed sarcasm, mockery, and even swear words. However, they still directly addressed the political demands.

Counter-narratives were promoted by members of the group with conservative leanings and were explicitly critical. However, that they did not respond to the protesters’ claims. The goal of the counter-narratives was to shift the discussion from the protesters’ claims to a different topic (eg. protesters themselves and their alleged motives). In some cases we noticed attempts to clutter hashtags by content unrelated to their intended use.

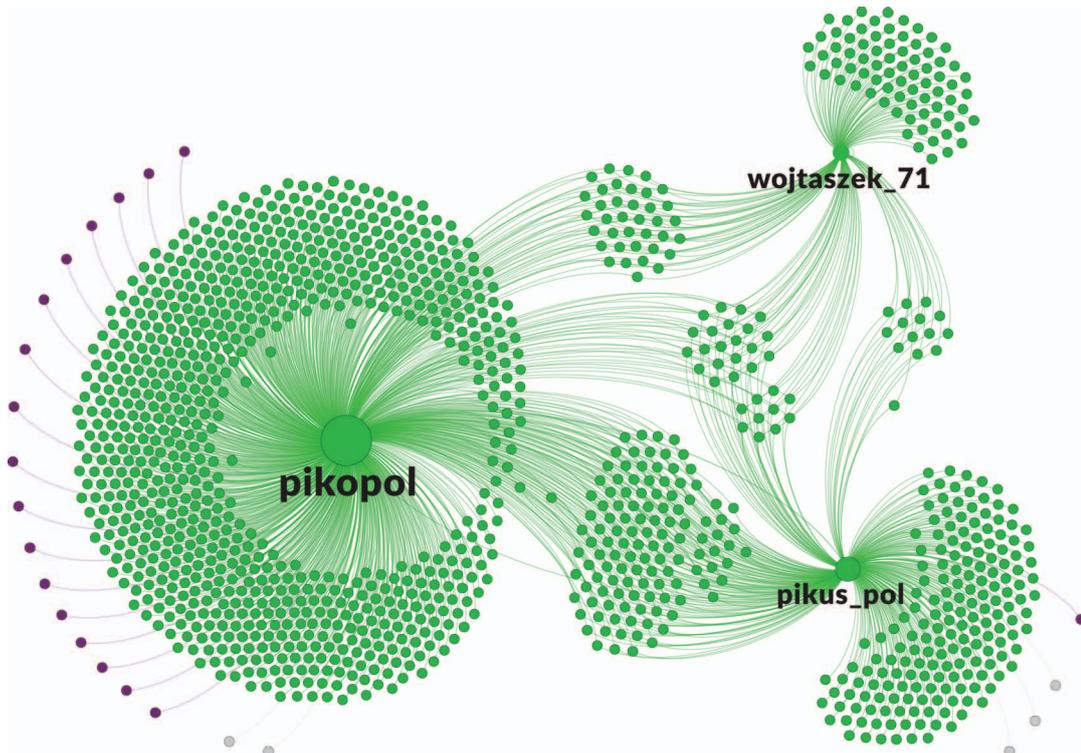
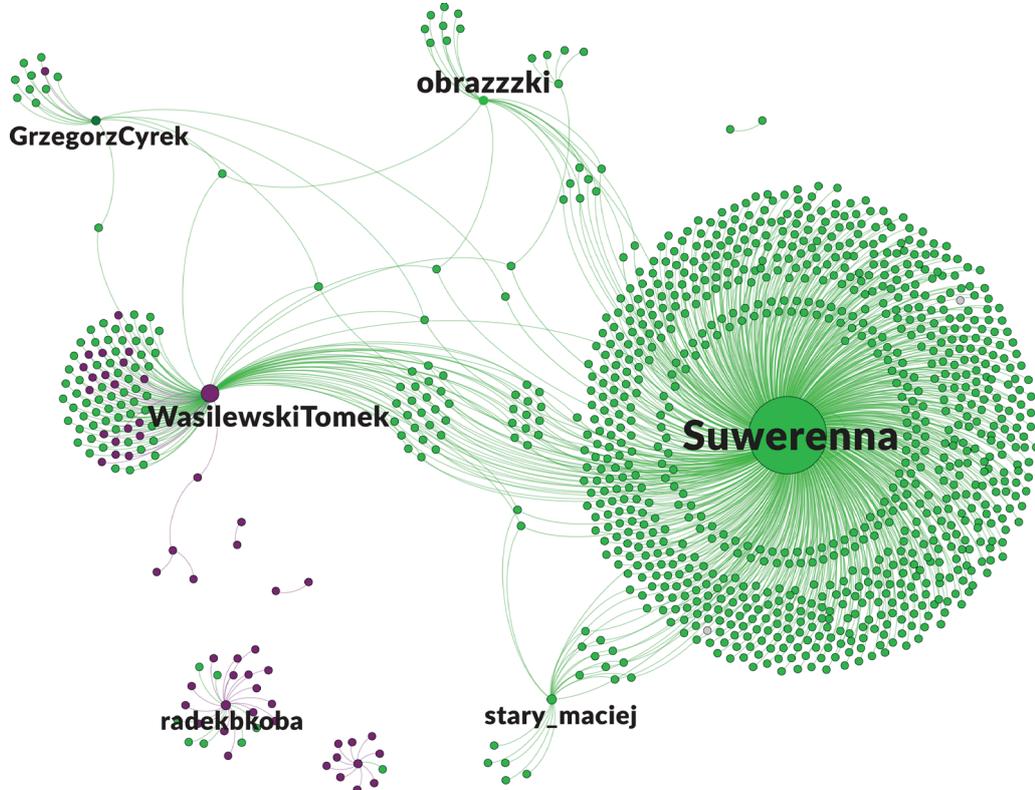
The counter-narratives that we examined had been initiated by influencers enjoying respect within groups of users with similar political views. They were further disseminated by authentic users, not fake accounts. The network analysis of accounts involved in sharing polemical tweets leads to the

conclusion that the counter-narratives initiated by users in the group with conservative leanings did not infiltrate the opposite group, which means the mainstream discourse did not change substantively.

Graph 3 and 4. Network analysis for counter-narratives within #czarnyprotest (women’s protest)



Graph 5 and 6. Network analysis for counter-narratives within #protestmedykow (junior doctors' protest)



The network analysis for key topics and counter-narratives within them shows that Twitter is not exactly a space for the confrontation of opinions between groups with diverging political views.



Despite a considerable number of retweets within the group with conservative leanings, none of the counter-narratives succeeded in the opposite group, with one exception - the counter-narratives which claimed that junior doctors, who went on strike in order to make the government increase spending on the national healthcare system and provide them with a pay rise, were in fact hypocritical and indulged themselves in luxurious holidays. What was supposed to represent exotic vacations turned out to be a humanitarian mission. This was the only case when the group with liberal leanings did engage in a discussion within counter-narratives, but only after a traditional, wide-reaching media outlet (TVP Info, a public news service) issued a feature supporting these allegations.

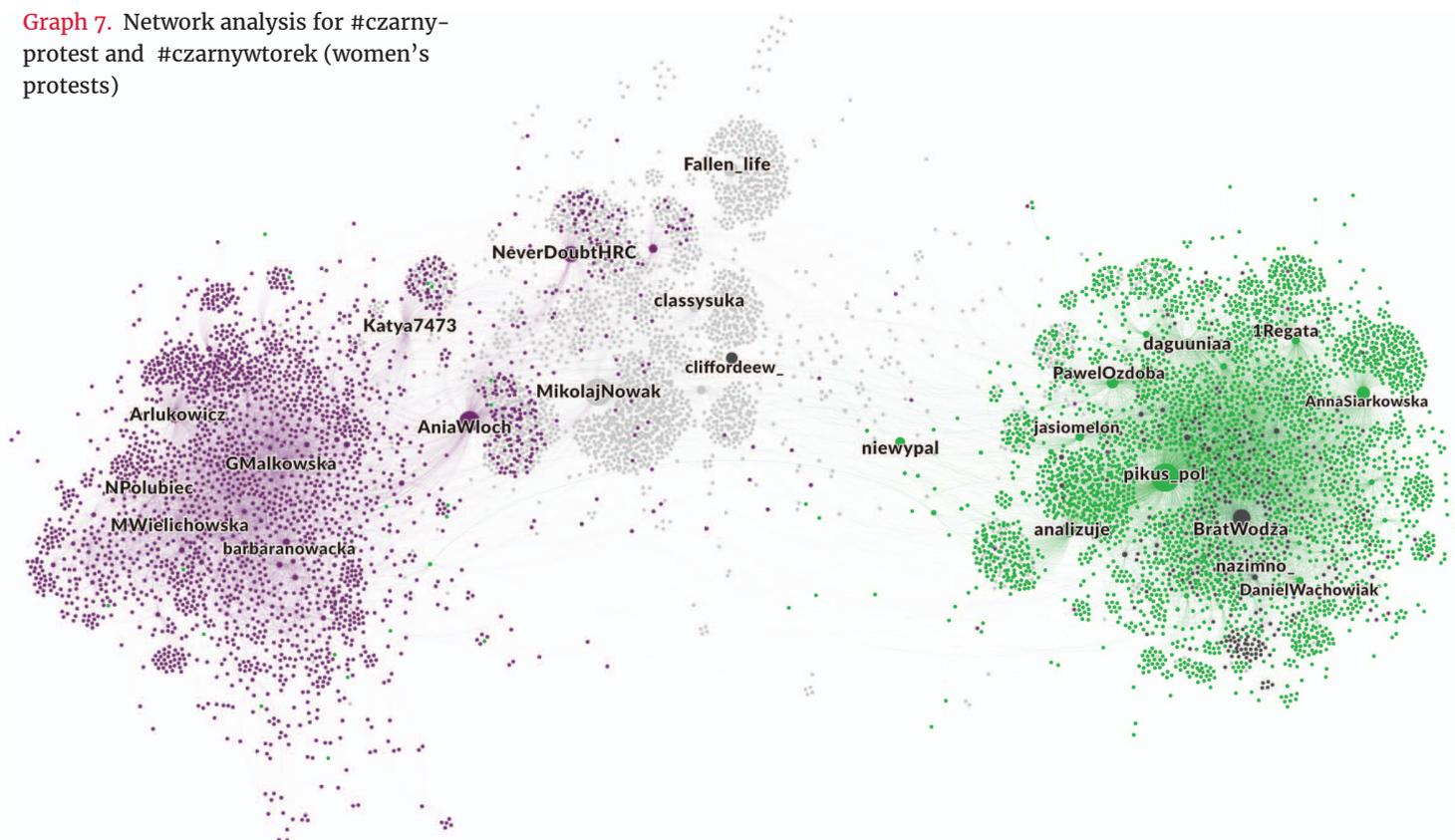
II. Was the political debate on Twitter subject to manipulation?

In our case study we did not observe any significant influence of fake accounts and false amplifiers on the trends in the debate or its strands. During the examined period key topics were introduced and promoted by users whose identity was not doubtful and whose prominence is well-established within their group (e.g. **politicians, journalists, main-stream media**).

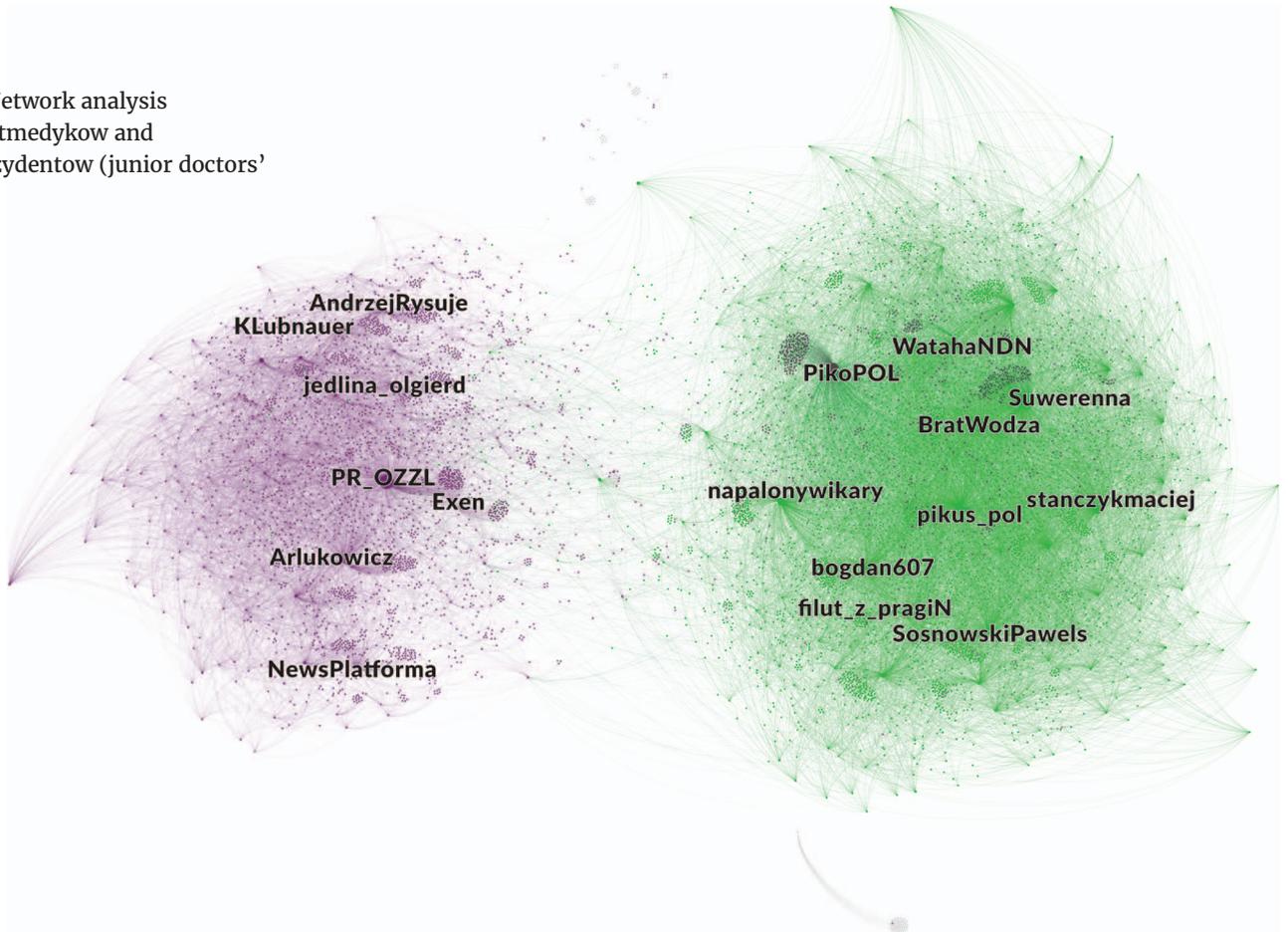
Among the most influential accounts responsible for shaping the political debate we also identified **influencers** with well-developed internet identities (e.g. PikuśPOL, daguunias, antyKOD, avanti_ultras89, Polskawruinie1). For this category of users it is typical to use

pseudonyms and present clear-cut political views. Rarely they reveal their professional or political affiliations. Therefore, it is significantly more difficult to verify their real-life identities. Without the historical analysis, consisting in the examination of the behaviour of these accounts from the moment they were created (in particular Twitter discussions they engaged in over time), **we are not able to determine whether or not there are other actors (with own agenda) behind those influencers**. We had no doubt, however, that these accounts are run by humans.

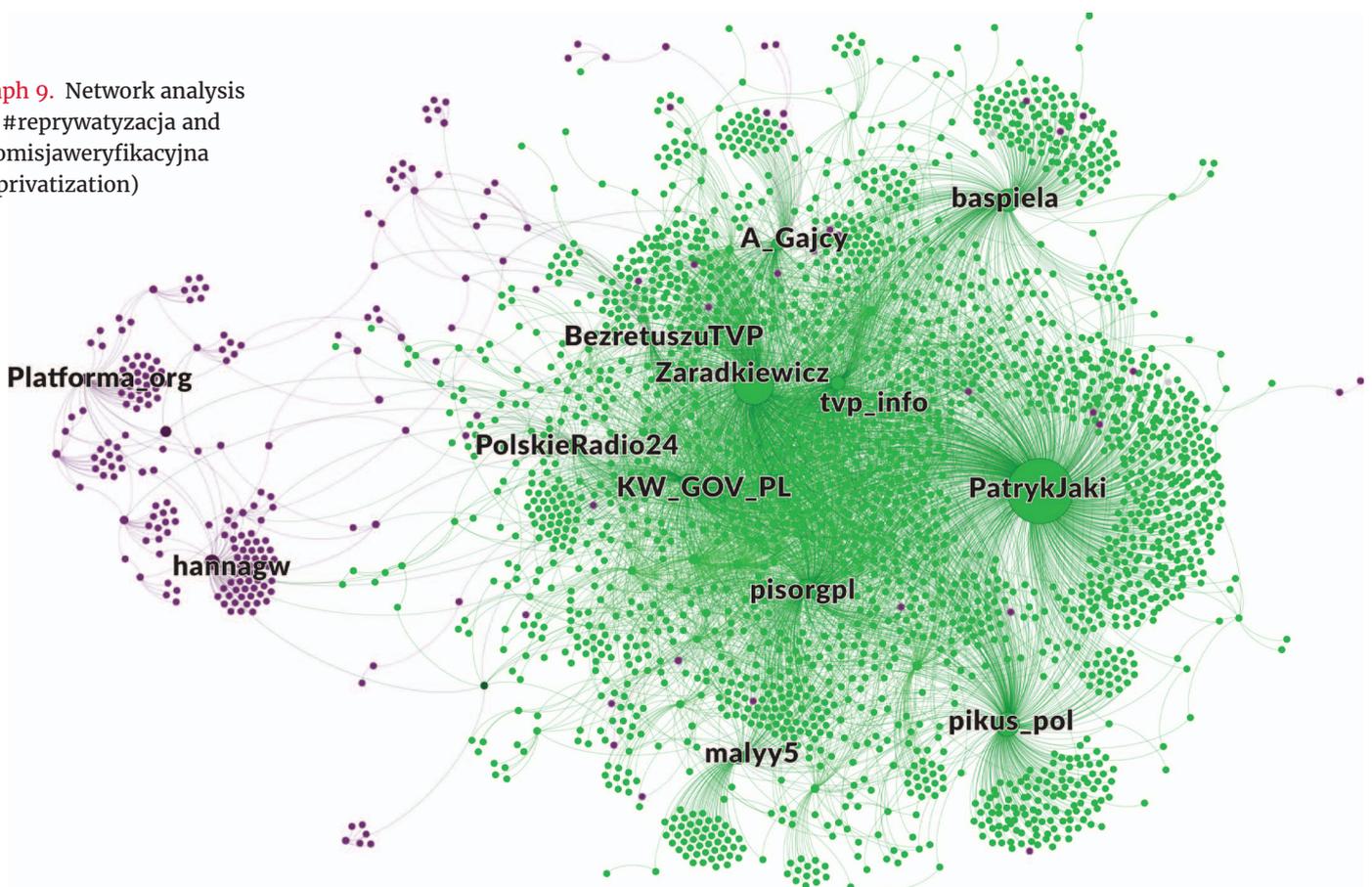
Graph 7. Network analysis for #czarny-protest and #czarnywtorek (women's protests)



Graph 8. Network analysis for #protestmedykow and #protestrezydentow (junior doctors' protest)

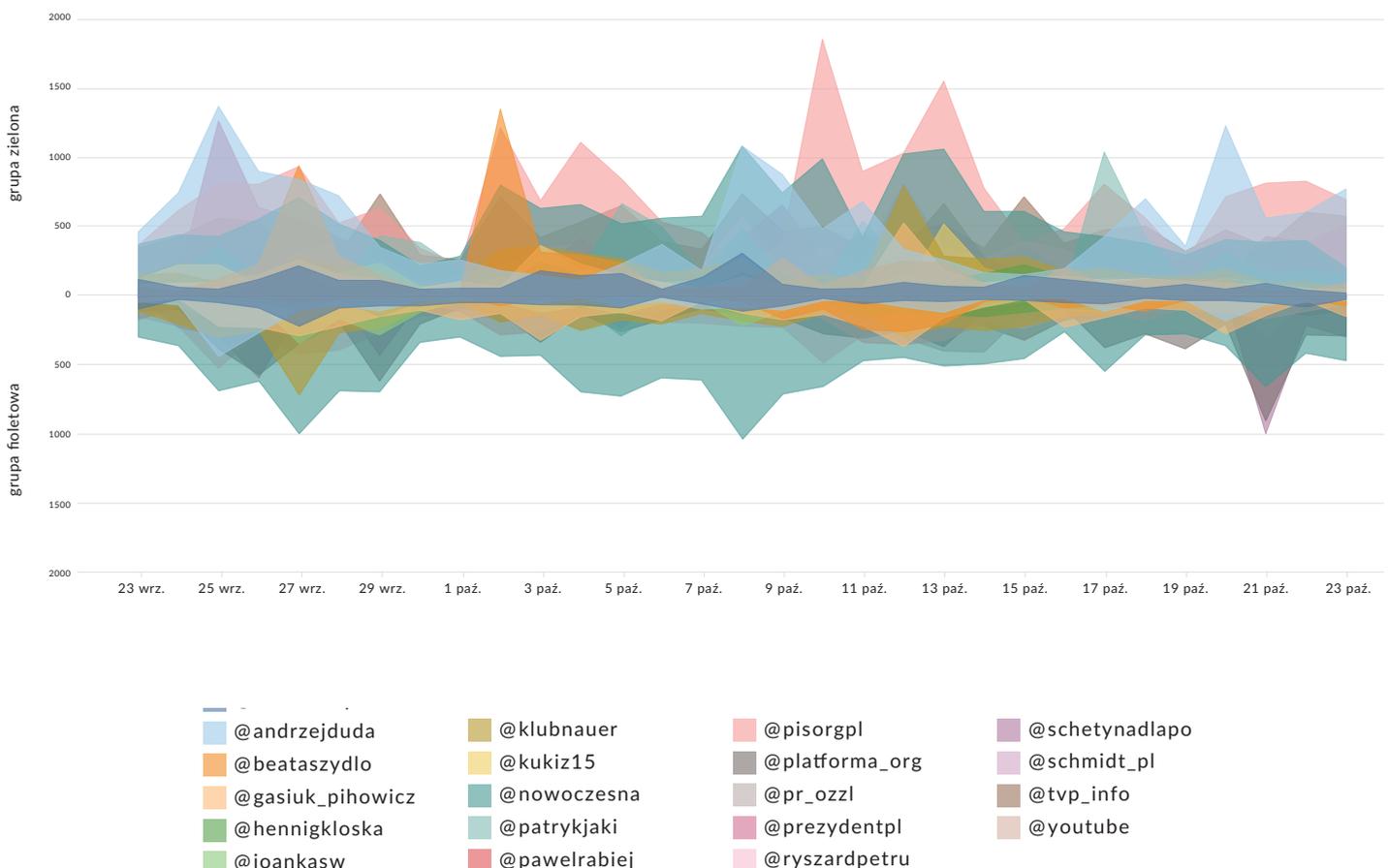


Graph 9. Network analysis for #reprivatyzacja and #komisjaweryfikacyjna (reprivatization)



The dynamics of counter-narratives was slightly different. They were rather created by Internet influencers familiar only to a limited group of users with similar views, and not by opinion-forming media or key politicians. This observation confirms our initial hypothesis that **it is well-established individuals and organisations, and not false amplifiers or fake accounts that have a significant influence** (namely being capable of reaching beyond groups of users with shared political views) on Twitter-based public debate. They are mainly politicians, journalists, media and political organisations - actors whose influence is not based only on their social media activity. The threads that they promote have the highest chance of spreading beyond a particular group.

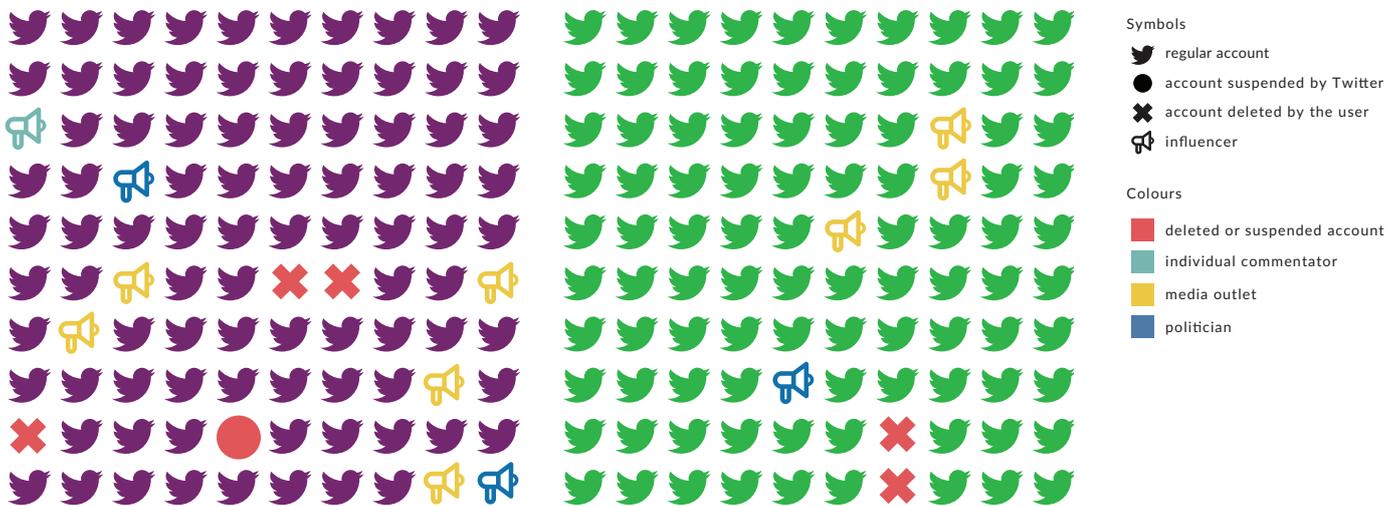
Graph 10. 20 most frequently mentioned accounts (for both groups) over time



Quality, not quantity

The comparison between the most active accounts and the accounts that are the most influential within the analyzed groups confirms that the frequency and number of posts that is easy to be artificially bumped up does not have an effect on the general impact, based on the number and the quality of interactions. In the case of both groups, **all accounts that were both the most active and the most influential belong to popular media outlets and politicians.** Their non-standard activities may be due to the fact that they are operated by multiple persons. It seems unquestionable, however, that their influence was defined by their social identity and the authority they are credited with. We did not identify a single account which - only by the fact of its increased activity (characteristic for bots) - would play a significant role in shaping the discussion on Twitter.

Graph 11. 100 most active users in both groups. The graph also shows additional characteristics of some accounts. The use of Twitter icon means that given account did not qualify for any additional characteristic.



Suspicious accounts

In our study we took a close look not only at the accounts defined as the most influential by the PageRank Centrality algorithm (i.e. the ones generating the highest number of quality interactions) as well as at those whose activity was increased (characteristic for bots), but also at the accounts with a behaviour pattern different from the one typical for real people. We were curious how many accounts with such characteristics were engaged in the public debate on Twitter and what political agenda they supported.

To identify suspicious accounts we used the bot detection API (Botometer) developed by the Observatory on Social Media Project at Indiana University. This algorithm has learned to identify false amplifiers by analyzing the behavioural pattern of accounts known to be hijacked by bots. In the process of machine learning researchers have used around a thousand different criteria to determine which ones best describe the behaviour of hijacked or controlled accounts. The algorithm relies on two types of criteria to predict the probability of an account being operated by a bot: language-dependent features (content and emotional charge) and language-independent features, which describe the position of an account within a social network (frequency of posts, number of interactions, type and number of followers, features of the account itself). The language-dependent features work only for English. Therefore, we were able to use only the language-independent criteria, which unavoidably lowers the efficiency of the algorithm.

Using this method, we identified over 500 accounts (out of approx. 50,000 engaged in the political debate during the analyzed period) which did not behave like regular Twitter users. The analysis of the content posted by these accounts showed us, however, that the results of the algorithm do not automatically lead to the conclusion that these accounts were hijacked by a botnet or were artificially manufactured. The analysis limited to the examination of how an account functions

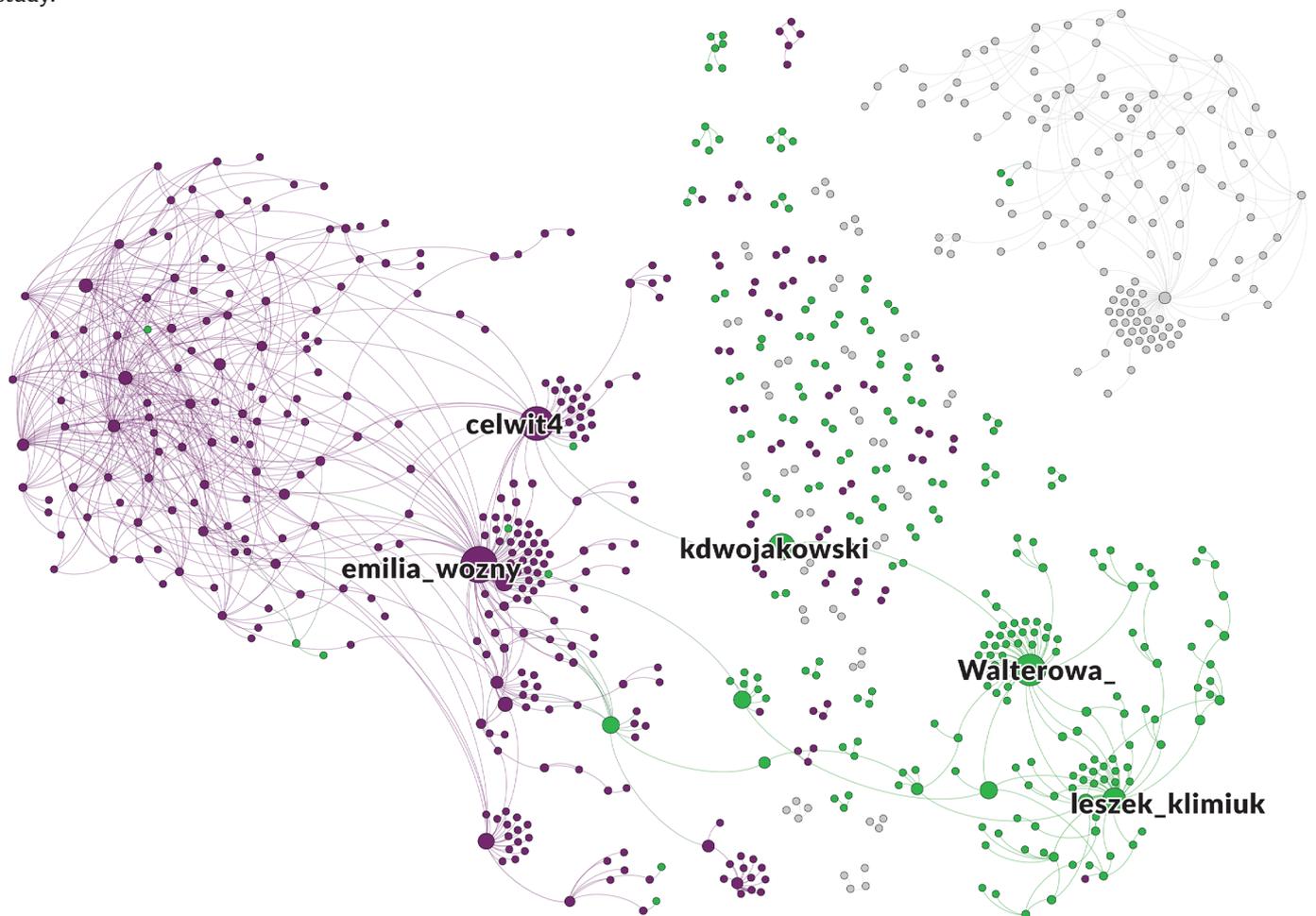
within a social network (e.g. how often it posts messages or interacts with others, who its followers are) does not enable us to make a distinction between accounts that have been hijacked or manufactured and accounts which are managed in a professional way.

Fairly often these accounts are operated by more than one person or they use extra software (e.g. Tweet Deck) that enables easier management of interactions. The accounts owned by news channels usually tweet frequently – every few minutes during the entire day. On the other hand, accounts of well-known politicians focus on responding quickly to mentions and comments. In the context of network analysis, both models of activity can easily be mistaken with the one of bots.

These observations do not lead to the conclusion that hijacked or manufactured accounts which distort the debate in the social media are impossible to be detected by algorithms. They do, however, caution against drawing some conclusions too quickly, especially if they are based solely on metadata and network analysis. Key factors that should be taken into account when identifying bots are: the content of communications, their context and emotional charge. Without going deeper we are unable to tell the difference between professionally managed accounts (or accounts with other atypical features) and false amplifiers.



Graph 12. Network analysis for suspected accounts. It shows their affiliation to one of the three groups of users identified in the study.



Accounts that disappeared

While analyzing atypical behaviour patterns in the public debate on Polish Twitter, we came across some accounts that were either influential or unusually active, and disappeared after a few months. This made it impossible for us to deeply examine their characteristics (e.g. applying the Botometer algorithm mentioned above). We established that out of nearly 50,000 accounts active in the debate as many as 4,056 disappeared by January 2018 (they were deleted by their owners or suspended by Twitter for violating the terms of use).

In order to see whether the activity of these accounts has particular features (e.g. it was shifted in time or promoted a different agenda) we compared their engagement and key words used within individual threads (hashtags) over the examined month with the activity of other accounts that participated in the political debate. This analysis did not bring any particular surprises. We discovered that, with rare exceptions, the behaviour pattern of deleted and suspended accounts within the examined period did not deviate from the general model.

What is interesting, the activity of deleted and suspended accounts is distributed evenly between the two politically polarized groups that we distinguished in the network analysis (2 out of 100 most active accounts in both groups and the ratio of retweets at 2.8%).

CONCLUSIONS

At first glance

We live in virtual bubbles: we prefer to talk to people who think the way we do, we look for confirmation of our own opinions, we share content we want to promote and not that which we question. This is quite obvious. That's why we were not surprised to discover two distinctive and polarized groups of users – one with conservative and another with liberal leanings. Users situated on the verge of two groups were mainly looking for confrontation with users from opposite groups and sharply disputed the narratives disseminated by them.

We were also not surprised that the central role in shaping public debate on Twitter is occupied by individuals and institutions that enjoy social or political respect (popular journalists, politicians, political parties, media outlets). New political or polemical threads that we called 'counter-narratives' are created mostly by online influencers who present more radical positions.

None of the influential roles in the public debate were played by hijacked or manufactured accounts. Such accounts do exist on Twitter and can be used for influence-boosting, but they serve as a claque rather than as key actors who would be in a position to change trends in public debate. At the same time our observations related to the influence of bots (false amplifiers) on the political debate do not allow us to reach general conclusions about the digital propaganda.



On the basis of the data we collected we can defend the assumption that fake or hijacked accounts did not have significant influence on the directions and dynamics of the debate happening on Polish Twitter. However, without a more comprehensive qualitative analysis and the examination of the history of the most influential accounts, we are not able to say whether or not the activity of key influencers (as revealed in the network analysis) bears features of organized propaganda.

Going deeper

One observation was perhaps not surprising but at least sobering: even though both groups discussed the same matters, dictated by current social and political events happening during the examined period, there was no place for polemics or confrontations of opinions between them. Counter-narratives - threads deviating from the main stream of discussion or aimed at reorienting the debate - did not manage to reach outside the group with conservative leanings which was responsible for generating them.

Not judging the main stream of the discussion or the counter-narratives, we do see it as a problem that the debate on social media does not lead to any confrontation of opinions between polarized groups. Such confrontation, if present, could build a political consensus or a sense of community.

■

Among the counter-narratives that we analysed, there was only one that managed to reach the opposite community. The reason for this was probably that the counter-narrative was taken up by a mainstream media outlet (TVP Info). Despite the fact that Polish mainstream media are, just like social media, subject to strong polarization and operate under the same mechanisms (i.e. they address the already convinced), their reach is wide enough for the audience of different political views to notice counter-narratives, thus provoking a discussion.

We still don't know enough

The conclusions which emerge from our case study find footing in the collected data but are not necessarily representative of the public debate in social media. That's why we present them not as the revealed truth, but as a trigger for an important discussion that we should be having in Poland on a deeper level and on the basis of data and research coming from various sources.

Social media are more and more influential in shaping our worldview and political opinions. The topics of their regulation and responsibilities are constantly present on the political agenda. We require more transparency from them when it comes to the techniques and criteria used for profiling content. At the same time, we do not pay enough attention to the social processes happening on social media, where it is the people who play the key role.

While looking for answers to questions which provoked our case study, **we noted a few areas which deserve further research:**

- The dynamics of interactions on the verge of the bubbles. What are the features of accounts on the border of polarised groups? What is their activity (do they reply or just retweet with a comment)? Are they trying to confront the other group with their own views or do they focus on building a strong position for themselves within their group (for example by making fun of the opponent)?
- If the conclusion that counter-narratives generally do not manage to reach beyond the bubble in which they were generated is justified, then what is the reason for this? Is it related to the way that information spreads in social media ('I do not see other opinions, therefore I don't react') or rather to decisions that users make themselves ('I see other opinions but I don't react because I don't agree with them')?
- Actions of bots. In Poland, is buying supporters a large-scale issue? If so, what is their influence on the discussion that is happening within real communities? Are they capable of shifting the stream of the debate even though they operate only occasionally and in a diffused way? How can we effectively identify these false amplifiers?
- Actions of real users using technological tools to distort the debate in the social media. How to explore this phenomenon, which seems to have a much more serious effect on the directions of the political debate than the activity of bots? How to reply to propaganda or disinformation distributed by real social media users, either hired by political groups or motivated by ideological reasons (or manipulated)?

About Panoptikon Foundation

Panoptikon Foundation is the only NGO in Poland which keeps an eye on those who collect and use personal data in order to influence people: public authorities, secret services, business corporations. Since 2009 we keep track of new legislation, we intervene to protect human rights, we explain how commercial and public surveillance tools work, we share our know-how. Personal data became a new currency and the most effective instrument of power. Pervasive surveillance feeds our fears and kills trust. Algorithm-based decisions reinforce stereotypes, leading to exclusion and discrimination. Therefore at Panoptikon we help people regain control over their data and build society that respects freedom. You can support us here: <https://en.panoptikon.org/support-us>

About Internet Policy Observatory

The Internet Policy Observatory (IPO) is a project at the Annenberg School for Communication at the University of Pennsylvania. The overarching goal of the program is to deepen the reservoir of researchers and advocates in regions where Internet freedom is threatened or curtailed and to support the production of innovative, high-quality, and impactful internet policy research. The IPO facilitates collaboration between research and advocacy communities, builds research mentorships between emerging and established scholars, and engages in trainings to build capacity for more impactful digital rights research and advocacy.



panoptykon.org

