



CYFROWA PROPAGANDA CZY „NORMALNA” POLITYCZNA POLARYZACJA?

Studium debaty politycznej
na polskim Twitterze
(wrzesień–październik 2017)

Cyfrowa propaganda czy „normalna” polityczna polaryzacja? Studium debaty politycznej na polskim Twitterze (wrzesień–październik 2017)

Opracowanie:

Katarzyna Szymielewicz, Karolina Iwańska

Współpraca: Weronika Adamska, Maria Wróblewska

Badacze współpracujący: Emad Khazraee,
Paul Popiel

Wizualizacja danych: Marcin Antas

Korekta: Urszula Dobrzańska

Ilustracja na okładce: Nadia Snopek

Projekt graficzny i skład: Marianna Wybieralska

Wydawca: Fundacja Panoptykon | panoptykon.org

Marzec 2018

Publikacja dostępna na licencji Creative Commons

Uznanie autorstwa – Na tych samych warunkach

4.0 Międzynarodowe



Zrzuty ekranu z portalu Twitter na s. 14, 16, 17, 18,
19, 29 wykorzystano na podstawie prawa cytatu

Raport stanowi podsumowanie projektu *Polska debata polityczna na Twitterze. Studium przypadku* i został sfinansowany ze środków Internet Policy Observatory Annenberg School for Communication działającej przy University of Pennsylvania

Spis treści

WPROWADZENIE 4

METODA BADANIA 6

Ustalenie pojęć 8

WNIOSKI 9

I. Kto z kim rozmawiał? Kto kreował trendy? 9

Tematy i zaangażowanie. O czym rozmawiano? 11

Temperatura dyskusji 12

Kontrnarracje: próby zmiany dyskursu 13

Co z tego wynika? 20

II. Czy debata polityczna podlegała manipulacji? 21

Nie boty, tylko ludzie 21

Nie ilość, ale jakość 24

Podejrzane konta 25

Konta, które zniknęły 27

Trzecia grupa kont 28

PODSUMOWANIE 30

WPROWADZENIE

W 2011 r. Eli Pariser spopularyzował pojęcie bańki informacyjnej (ang. *filter bubble*), które zmieniło sposób, w jaki myślimy o mediach elektronicznych. Okazało się, że nie czytamy już tej samej gazety, ale miliony spersonalizowanych gazet dynamicznie zmieniających się pod wpływem naszej aktywności. Nasz obraz rzeczywistości, kształtowany przez media społecznościowe, stał się relatywny i w dużej mierze uzależniony od światopoglądu naszych znajomych. Żyjąc w intymnej relacji z platformami internetowymi, oswajamy się z tym, że ceną za łatwy i tani dostęp do informacji jest zdanie się na ich „politykę redakcyjną”, wspieraną przez nieprzejrzyste i stale ewoluujące algorytmy.

Dla wielu osób, które zdążyły się poczuć całkiem komfortowo w świecie sprofilowanych mediów, dopiero polityczne zwroty przyniesione przez rok 2015 zadziałały jak sygnał alarmowy. Mało kto spodziewał się, że Donald Trump, uważany raczej za telewizyjnego ekscentryka niż za poważnego polityka, zostanie prezydentem USA. Kolejny wstrząs przyniosło referendum, w którym Brytyjczycy zdecydowali o opuszczeniu Unii Europejskiej, oraz fakt, że dopiero po podjęciu tej decyzji rozgorzała dyskusja o jej konsekwencjach. Te wydarzenia – uznane za syndrom wzbierania populistycznej fali w krajach o reputacji stabilnych demokracji – uruchomiły potrzebę znalezienia winnego.

W kontekście tego, jak szybko zmienia się rzeczywistość mediów internetowych, zrozumiała jest pokusa, by właśnie tam szukać wyjaśnienia nowych problemów społecznych i politycznych. W to zapotrzebowanie wpisuje się narracja o nowej internetowej propagandzie. Duże poruszenie w zachodnich mediach wywołały doniesienia o rosyjskiej fabryce

trolli, która posłużyła ingerencji w amerykańskie wybory, czy o sieci alt-rightowych mediów generujących nierzetelne treści i podbijających swoją widoczność dzięki botom. Również wybory we Francji w 2017 r. przebiegły w cieniu obaw przed internetowymi manipulacjami – na minuty przed ogłoszeniem ciszy wyborczej w pierwszej turze wyborów prezydenckich wyciekły tysiące maili działaczy partii Emmanuela Macrona, o co po raz kolejny oskarżono Moskwę.

Pojęcia takie jak postprawda (słowo roku 2016 wg redakcji *Oxford Dictionaries*) czy *fake news* przeszły ze słownika publicystyki do słownika nauki. Wpływ nowych technologii i mediów społecznościowych na kondycję demokracji i opinii publicznej stał się przedmiotem coraz odważniejszych badań. Wśród pionierów był Uniwersytet Oksfordzki, gdzie uruchomiono projekt Computational Propaganda Research Project (COMPROP). COMPROP analizuje zjawiska na styku działania algorytmów, automatycznej komunikacji i polityki. W ubiegłym roku Robert Gorwa, badacz z zespołu projektowego, przeanalizował polityczny ruch na polskim Twitterze pod kątem aktywności tzw. fałszywych wzmacniaczy (*false amplifiers*), w tym botów¹. Z jego raportu wynikało, że problem cyfrowej propagandy dotyczy także polskiego Internetu.

1 Robert Gorwa, *Computational Propaganda in Poland: False Amplifiers and the Digital Public Sphere*, <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/06/Comprop-Poland.pdf>.

Jak jest naprawdę? Czy skupiając się na fake newsach, botach i algorytmach filtrujących treści, nie tracimy z oczu rzeczywistych przyczyn kryzysu, za którym kryją się ludzie – wraz z ich dotąd lekceważonymi frustracjami i problemami?



Nie ulega wątpliwości, że narzędzia technologiczne, z których masowo korzystamy, mają wpływ na nasze życie społeczne i polityczne. Pewnych korelacji doświadczamy na własnej skórze. Czujemy, że dynamika debaty prowadzonej w tradycyjnej prasie jest zupełnie inna niż na Twitterze, gdzie rządzą krótkie i szybkie komunikaty. Rozumiemy, że inne problemy wiążą się z rolą tradycyjnych redakcji, ponoszących odpowiedzialność za publikowane treści, a inne – z działaniem platform internetowych, których „polityka redakcyjna” zmienia się z każdym dostosowaniem algorytmu do tego, co chce przeczytać użytkownik.

Słusznie nas niepokoi polaryzacja debaty publicznej, radykalizacja poglądów, populizm i efekt zamknięcia w bankach informacyjnych, który może być wykorzystany w celach propagandowych. Nie tracąc z oczu tych problemów, warto jednak zachować ostrożność w przechodzeniu od obserwacji („Coś ważnego się zmienia!”) do wyjaśnień („Dzieje się tak, ponieważ...”).

Jako Fundacja Panoptykon stawiamy hipotezę, że proporcje w wybuchowej mieszance, która doprowadziła w ciągu ostatnich dwóch lat do niespodziewanych skutków politycznych takich jak przejęcie władzy przez ugrupowania populistyczne, niekoniecznie są zdominowane przez komponent technologiczny.



Wychodzimy z założenia, że proces tworzenia się baniek informacyjnych i sposób rozprzestrzeniania się konkretnych poglądów w mediach społecznościowych odzwierciedla społeczną i polityczną dynamikę, a więc przede wszystkim zależy od ludzi, nie od technologii, z której korzystają. Tę hipotezę postanowiliśmy zbadać na własnym podwórku, we współpracy z doświadczonymi badaczami z Uniwersytetu Pensylwanii².

² W badaniu uczestniczyli – obok Katarzyny Szymielewicz i Karoliny Iwańskiej z Fundacji Panoptykon – dwaj badacze związani z Annenberg School of Communication przy Uniwersytecie Pensylwanii: dr Emad Khazraee oraz Paul Popiel. Ich rolą było zaprojektowanie metodologii, wybór narzędzi badawczych oraz pomoc przy analizie danych.

METODA BADANIA

Zdecydowaliśmy się na typ badania jakościowego, jakim jest studium przypadku (*case study*). Za jego przedmiot wybraliśmy debatę polityczną toczącą się po polsku na Twitterze w okresie od 23 września do 22 października 2017 r.

Wybór tego medium społecznościowego był podyktowany trzema czynnikami:

- dużą aktywnością polskich influencerów politycznych (polityków, działaczy, komentatorów) w tej przestrzeni;
- technicznymi ułatwieniami w dostępie do danych, które oferuje Twitter; oraz
- specyfiką organizacji treści na tym portalu, która ułatwia prowadzenie analizy sieciowej (np. możliwość łatwego identyfikowania kluczowych tematów przy pomocy hashtagów, a także śledzenia dynamiki toczącej się debaty treści dzięki funkcji retweet).

Wybrany zakres czasowy musiał być na tyle krótki, by umożliwić analizę o charakterze jakościowym (w wybranych wątkach). Jednocześnie udało się objąć zakresem badania wydarzenia polityczne o charakterze silnie polaryzującym: protesty kobiet przeciw zaostreniu prawa aborcyjnego, prace nad reformą sądownictwa, strajk lekarzy rezydentów i debatę na temat reprivatyzacji.

Projektując badanie, postawiliśmy sobie trzy cele:

- ustalenie, jacy aktorzy (osoby, organizacje, sieci społeczne) wywierają największy wpływ na debatę polityczną rozgrywającą się na polskim Twitterze;
- określenie podstawowych kierunków i opisanie dynamiki tej debaty – w szczególności interesowało nas, czy dochodzi do gwałtownych zmian w narracji (np.

dyskusja z afirmującej staje się krytyczna, pojawia się wyrazista kontrnarracja; dochodzi do zawłaszczenia hashtagu przez konkurencyjną społeczność albo zanieczyszczenia go przypadkowymi treściami);

- ustalenie, czy w debacie politycznej rzeczywiście mamy do czynienia z aktywnością i istotnym wpływem przejętych lub sztucznie wyhodowanych kont (botów i innych sztucznych wzmacniaczy), tak jak sugerował raport zespołu COMPROP z 2017 r.

Cele studium przypadku zdeterminowały wybór metod i narzędzi badawczych. Poniżej krótko opisujemy etapy badania i zastosowane metody.

Zebranie danych

Przy pomocy narzędzia Twitter Counter wyznaczyliśmy 100 polskich użytkowników Twittera o największej liczbie obserwujących. Spośród nich wybraliśmy konta, które zamieszczają treści związane z polityką. Na aktywność tej grupy nałożyliśmy filtr czasowy, wyodrębniając treści udostępnione w granicach czasowych badania. Zbiór wszystkich tweetów wygenerowanych przez tych użytkowników przeanalizowaliśmy pod kątem najczęściej pojawiających się hashtagów i słów kluczowych o charakterze politycznym. Taka metoda wyłonienia listy hashtagów i słów kluczowych do dalszej analizy pozwoliła nam zminimalizować ryzyko stronniczości.

W kolejnym kroku, przy pomocy narzędzia Streaming API, zaimportowaliśmy wszystkie tweety zawierające wybrane hashtagi i słowa kluczowe, które zostały zamieszczone na Twitterze między 23 września a 22 października 2017 r. Dodatkowo w okresie

pozyskiwania danych codziennie analizowaliśmy listę hashtagów i słów kluczowych, które w zebranych tego dnia tweetach pojawiły się przynajmniej 200 razy. Na tej podstawie sukcesywnie uzupełniliśmy listę wyszukiwanych terminów. Łącznie zebraliśmy nieco ponad 900 tys. tweetów do dalszej analizy.

Analiza sieciowa i analiza wpływu

Bazując na zebranych korpusie tweetów, prześledziliśmy relacje między użytkownikami na podstawie retweetów i bezpośrednich interakcji. Z pomocą algorytmu udało się wyodrębnić grupy, w ramach których użytkownicy najczęściej wchodziłi ze sobą w interakcje i które łączyły poruszane tematy. Do dalszej analizy wybraliśmy trzy największe grupy odpowiedzialne za niemal 90% wszystkich połączeń.

Do identyfikacji 100 najbardziej wpływowych użytkowników wykorzystaliśmy wskaźnik PageRank Centrality, który szereguje punkty węzłowe (centralnych użytkowników) w oparciu nie tylko o ilość, ale także jakość ich połączeń (interakcji). Algorytm za najbardziej wpływowe uznaje nie te konta, na które powoływała się największa liczba użytkowników, ale te, które były najczęściej cytowane (retweetowane) przez użytkowników cieszących się dużą popularnością i wpływem (mierzonym wg tego samego algorytmu).

Analiza dynamiki i tematów debaty

Dla całego korpusu tweetów stworzona została interaktywna aplikacja, która pokazywała częstotliwość pojawia się hashtagów i wzmianek na osi czasu. To pozwoliło nam zobaczyć, jakie tematy dominowały na polskim Twitterze w badanym okresie oraz jakie konta były najczęściej wspominane.

W kolejnym kroku dla całego zbioru przeprowadziliśmy modelowanie tematyczne, które miało pokazać, jakie dyskursy pojawiały się w treści tweetów, niezależnie od użytych hashtagów. Taka analiza (tj. rozkład hashtagów i wzmianek na osi czasu oraz modelowanie tematyczne) została powtórzona dla każdej z trzech wyodrębnionych grup w celu porównania dynamiki i tematów debaty.

Z całego zbioru hashtagów zdecydowaliśmy się wybrać do pogłębionej analizy trzy, które naszym zdaniem miały silny potencjał polaryzujący, a jednocześnie były wykorzystywane przez użytkowników o różnych poglądach i sympatiach politycznych: #czarnyprotest, #reprzywatyzacja, #protestmedyków. Dla każdego z tych hashtagów przeprowadziliśmy takie samo modelowanie tematyczne.

Próba zidentyfikowania botów

Następnie podjęliśmy próbę zidentyfikowania, które z kont występujących w badanym zbiorze wykazują cechy kwalifikujące je jako fałszywe wzmocniacze (np. boty). Wszystkie konta zostały przeanalizowane przy pomocy algorytmu Botometer, opracowanego przez badaczy z Observatory on Social Media, które działa przy Uniwersytecie Indiany. Poniżej bardziej szczegółowo opisujemy zasadę działania i ograniczenia tego algorytmu.

USTALENIE POJĘĆ

Żeby ułatwić dalszą lekturę, na wstępie wyjaśniamy nasze rozumienie kluczowych pojęć, do których będziemy się odwoływać.

Boty i fałszywe wzmacniacze

W najszerszym rozumieniu bot to rodzaj programu komputerowego, który wchodzi w interakcję z innymi użytkownikami sieci w oparciu o określony algorytm. Może pełnić pozytywną, wspierającą rolę (np. chatbot, który pomaga zainstalować aplikację albo udziela prostych porad). Jednak w dyskusji o cyfrowej propagandzie boty pojawiają się zwykle w negatywnej roli tzw. fałszywych wzmacniaczy, czyli programów, które w niej jawny sposób kontrolują konta w mediach społecznościowych i rozprzestrzeniają przy ich pomocy określone komunikaty (np. propagandę lub dezinformację).

Coraz częściej mianem botów określa się też konta półautomatyczne, nazywane inaczej cyborgami. Istotną rolę w ich działaniu pełni człowiek: tworzący propagowaną treść i determinujący rodzaj aktywności, którą podejmuje sterowane przez niego konta.

Przy czym to, co widać w sieci społecznościowej, to konta, które zostały przejęte (w wyniku kradzieży tożsamości) lub sztucznie wyhodowane, a nie sterujące nimi boty/cyborgi, które zwykle odpowiadają za całe sieci powiązanych ze sobą kont. To dlatego podczas analizy wyników badania częściej odnosimy się do przejętych lub sztucznie wyhodowanych kont niż do sterujących nimi botów/cyborgów.

Istotny wpływ na debatę

Przez istotny wpływ rozumiemy wygenerowanie nowych, nośnych wątków w dyskusji politycznej lub wyraźne oddziaływanie na kierunek już toczącej się dyskusji (np. na zmianę narracji, wypromowanie kontrnarracji, zniechęcanie zwolenników innej tezy do dalszego udziału w dyskusji). W tym sensie przejawem istotnego wpływu na debatę nie

jest większa widoczność danego konta w ramach sieci społecznościowej, jeśli wynika ona przede wszystkim z własnej, ponadstandardowej aktywności tego konta.

Kontrnarracje

Pisząc o kontrnarracjach, mamy na myśli próby zmiany dyskursu na dany temat poprzez wprowadzenie krytycznych lub polemicznych wątków w stosunku do głównego nurtu dyskusji. Tropem sygnalizującym pojawienie się kontrnarracji w danym temacie było wystąpienie słów kluczowych, które nie pasowały do głównego nurtu dyskusji. Przy czym badaliśmy tylko te kontrnarracje, które zostały podchwyczone przez innych użytkowników, generujących przynajmniej 20 retweetów lub nowych tweetów w tym samym wątku.

Zielona/fioletowa grupa

W raporcie opisujemy funkcjonowanie dwóch spolaryzowanych grup, które zdecydowaliśmy się oznaczyć neutralnymi kolorami: zielonym i fioletowym. Za tą decyzją stoi świadomość, że na podstawie samej analizy sieciowej (ilustrującej, w uproszczeniu, kto z kim rozmawiał i do kogo się odnosił) nie możemy wyciągać żadnych wniosków na temat poglądów tweetujących osób. Analiza jakościowa treści publikowanych przez najbardziej wpływowe konta w obu grupach pozwala jedynie stwierdzić, że w grupie zielonej przeważają sympatie konserwatywne i postawy prorządowe, a w grupie fioletowej – sympatie liberalne i postawy krytyczne wobec rządu. Jednocześnie nie możemy wykluczyć, że wiele kont znalazło się wewnątrz zielonej albo fioletowej grupy nie ze względu na spójność poglądów z tymi, które dominują w tej grupie, ale wyłącznie ze względu na częste interakcje (np. regularną krytykę podawanych dalej tweetów).

WNIOSKI

I. Kto z kim rozmawiał? Kto kreował trendy?

Na podstawie analizy sieciowej wyodrębniliśmy dwie główne grupy kont, które brały aktywny udział w debacie politycznej na Twitterze: pierwszą, liczącą ok. 13 tys. kont, i drugą, liczącą ok. 10 tys. kont. Analizując opisy i przykładowe tweety najbardziej wpływowych użytkowników w obu grupach³, ustaliliśmy, że pierwsza (oznaczona kolorem zielonym) skupia w dużej mierze osoby/podmioty o sympatiach konserwatywnych, często wyrażających poparcie dla polityki rządu. Z kolei w grupie drugiej (oznaczonej kolorem fioletowym) znalazło się wielu użytkowników o sympatiach liberalnych i lewicowych, zdecydowanie częściej wyrażających się krytycznie na temat polityki rządzącej partii. Ponieważ jednak analiza jakościowa została przeprowadzona wyłącznie dla 20 najbardziej wpływowych kont w każdej z grup, postanowiliśmy w dalszej części raportu posługiwać się neutralnymi określeniami **grupa zielona** i **grupa fioletowa**, co znajduje swoje odzwierciedlenie w kolorystyce wykresów. Przynależność poszczególnych kont do grupy zielonej lub fioletowej – określona za pomocą algorytmu na podstawie wszystkich zebranych tweetów – jest stała i nie zmienia się w zależności od tematu, o którym dyskutowano.

Wśród najbardziej wpływowych kont w grupie zielonej pojawiały się konta instytucji publicznych (Sejmu, prezydenta),

profile polityków związanych z PiS (Patryka Jakiego, Macieja Wąsika), prawników mediów i publicystów (TVP, Polskiego Radia, tygodnika Sieci, portalu wPolityce.pl, Marzeny Nykiel), a także konta internetowych influencerów wyrażających aprobatę dla polityki rządu (PikuśPOL, daguuniaa, antyKOD, avanti_ultras89, analizuje). Z kolei w grupie fioletowej najbardziej wpływowi byli publicyści i media krytycznie oceniające politykę partii rządzącej (TVN24, Onet, Radio ZET, RMF FM, portal 300Polityka), politycy związani z Platformą Obywatelską i Nowoczesną (Bartosz Arłukowicz, Kamila Gasiuk-Pihowicz) i krytyczni influencerzy (np. Polskawruinie1, katarynaaa, wsamraz).

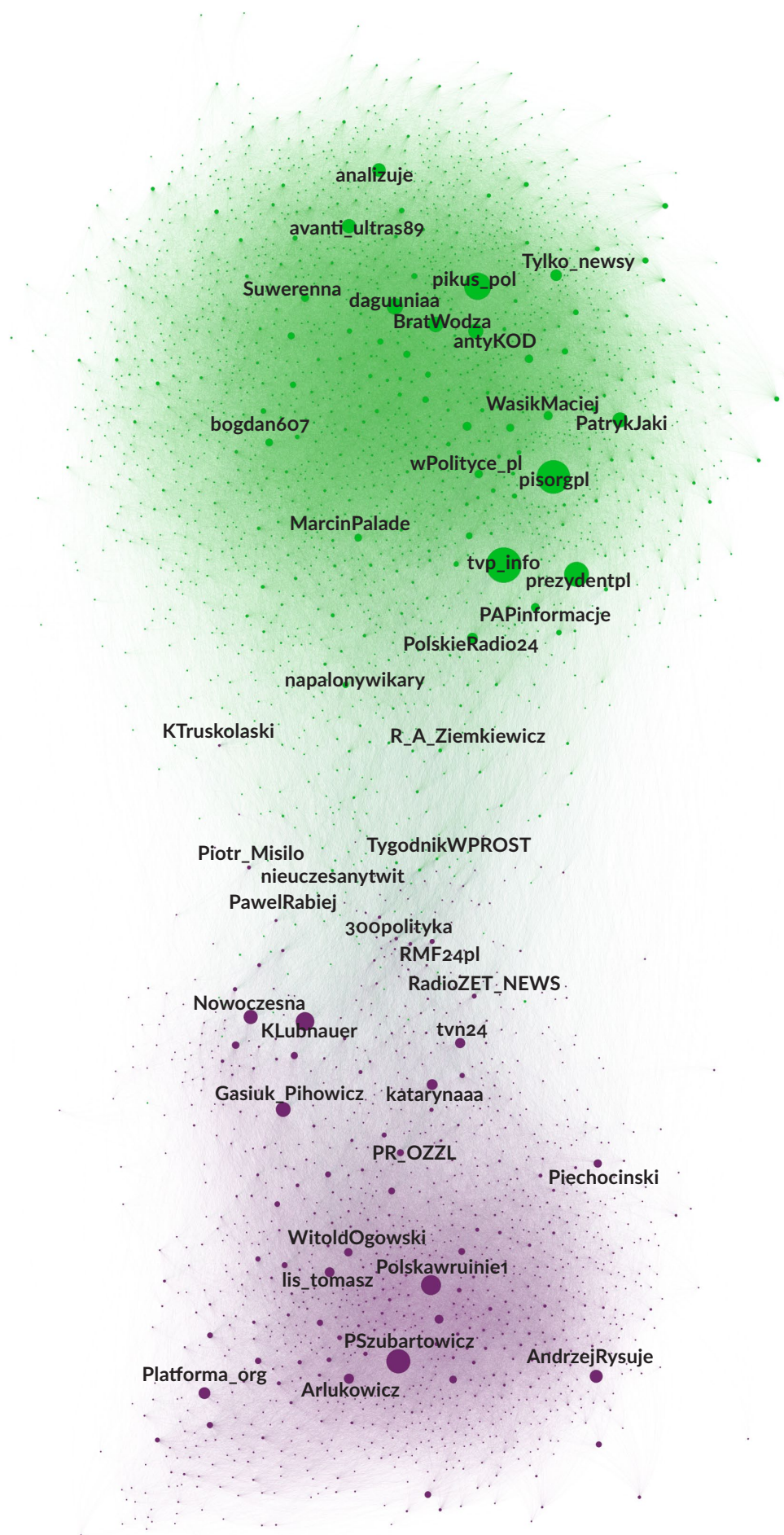
Analiza parametrów⁴ określających skalę wpływu badanych kont w ramach sieci społecznościowej wykazała, że konta aktywne w grupie zielonej mają większą widoczność w debacie politycznej i tworzą gęstszą (skuteczniejszą w oddziaływaniu) sieć powiązań niż najbardziej wpływowe konta w grupie przeciwnej.

Dynamika interakcji pomiędzy obiema grupami i analiza jakościowa reprezentatywnych tweetów wskazują na silną polityczną polaryzację. W obu badanych grupach konta zajmujące centralne miejsce charakteryzuje duża wyrazistość przekazu, który nie jest zniuansowany i nie odnosi się do argumentów

³ Najbardziej wpływowych użytkowników w obu grupach wyłoniliśmy w oparciu o listę 100 najbardziej wpływowych kont dla całego zbioru tweetów. Na tej liście znalazły się 33 konta należące do grupy zielonej i 21 kont należących do grupy przeciwnej.

⁴ W tej analizie zostały wykorzystane algorytmy PageRank i Eigenvector Centrality. Sposób działania tego pierwszego wyjaśniamy w części dotyczącej metody badania.

Wykres 1. Analiza sieciowa dla kont, które w badanym okresie weszły w przynajmniej 50 interakcji (tweety, retweety, odpowiedzi), z oznaczeniem kont najbardziej wpływowych i wybranych kont znajdujących się na styku grup



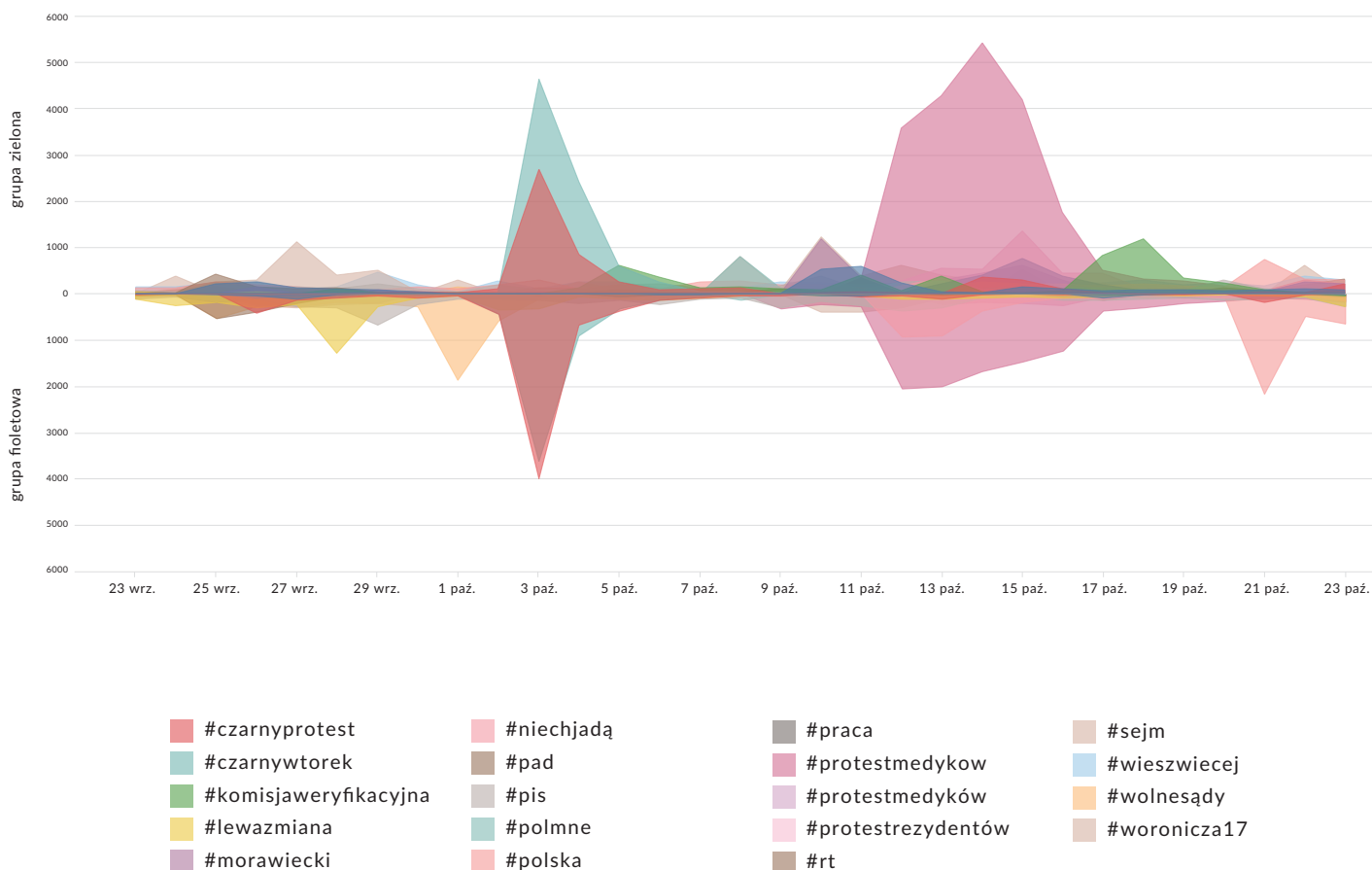
funkcjonujących w grupie przeciwnej. Konta znajdujące się (wg analizy sieciowej) na pograniczu nie są tak wpływowe jak konta umiejscowione głęboko wewnątrz grup. Nie byliśmy jednak w stanie ustalić, jaki typ kont dominował na styku grup – a znajdują się wśród nich zarówno media i politycy (głównie Nowoczesnej), jak i zwykli użytkownicy. Ci ostatni wchodzili w interakcje z przedstawicielami przeciwnej grupy, ostro polemizując z rozpowszechnianymi przez te osoby narracjami. Użytkownicy na styku grup nie szukają porozumienia i debaty, ale raczej konfrontacji.

Tematy i zaangażowanie. O czym rozmawiano?

Najpopularniejsze w badanym okresie hashtagi i słowa kluczowe odpowiadały najważniejszym wydarzeniom politycznym w Polsce i pokrywały się z najgorętszymi tematami w debacie, która toczyła się w tradycyjnych mediach, innych mediach społecznościowych (głównie na Facebooku) i na ulicy.

Dla obu grup odnotowaliśmy dwa wyraźne skoki aktywności: jeden w okolicach 3 października oraz drugi – pomiędzy 12 a 16 października. W pierwszym przypadku gwałtowny wzrost liczby zamieszczonych

Wykres 2. Aktywność obu grup w 20 najpopularniejszych hashtagach



tweetów wiązał się z odbywającymi się w całej Polsce protestami kobiet, co odzwierciedlały hashtagi #czarnywtorek i #czarny-protest. Z kolei na okres pomiędzy 12 a 16 października przypadł szczyt protestu głodowego lekarzy rezydentów, przejawiający się w popularności hashtagów #protestmedyków i #protestrezydentow. Poza tymi dwoma wydarzeniami rozkład hashtagów i związanych z nimi aktywności był dość równomierny.

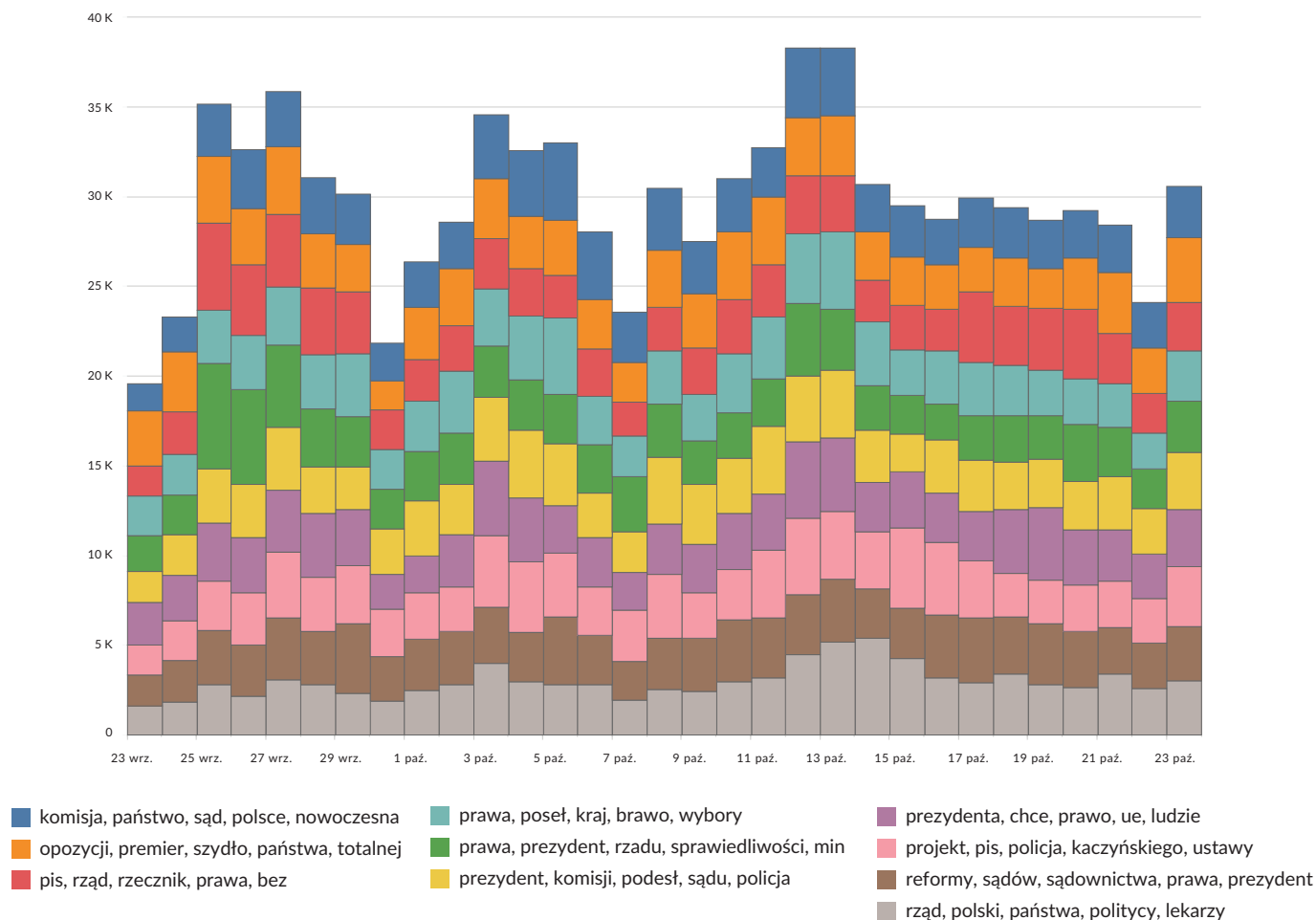
Zestawienie aktywności obu grup pokazuje, że grupa zielona wykazała dużo większą aktywność przy proteście rezydentów niż grupa przeciwna. W porównaniu do innych tematów w grupie fioletowej największe

zaangażowanie wywołała dyskusja o czarnym proteście, choć biorąc pod uwagę liczbę tweetów – ze względu na mniejszą liczbę kont – i tak nie przewyższała aktywności grupy zielonej.

Temperatura dyskusji

Projektując badanie, liczyliśmy na to, że automatyczna analiza danych pozwoli nam uchwycić także temperaturę dyskusji i główne tezy, jakie są lansowane przez konkurencyjne grupy. Temu miało służyć modelowanie tematyczne i analiza semantyczna, które zaowocowały wyodrębnieniem najbardziej reprezentatywnych tematów dla całego korpusu tweetów. Niestety, te działania nie

Wykres 3. Rozkład 10 najbardziej reprezentatywnych tematów wyodrębnionych w modelowaniu tematycznym dla wszystkich zebranych tweetów



przyniosły oczekiwanych rezultatów. Tematy wyodrębnione przez algorytm okazały się zbyt neutralne – tj. nie pokazywały zabarwienia emocjonalnego tweetów. W najczęściej pojawiających się słowach kluczowych dla poszczególnych grup zabrakło ocennych (krytycznych lub afirmujących) elementów. Z tego powodu na podstawie słów kluczowych nie byliśmy w stanie rzetelnie i stanowczo określić, które tematy były najsilniej polaryzujące i jakie emocje budziły po obu stronach twitterowej barykady. Wnioski dotyczące politycznej polaryzacji wyciągnęliśmy w oparciu o analizę jakościową tweetów generowanych przez najbardziej wpływowe konta w obu grupach.

Kontrnarracje: próby zmiany dyskursu

Modelowanie tematyczne przeprowadziliśmy także dla trzech tematów (wyodrębnionych na podstawie hashtagów w oparciu o kryteria opisane w części metodycznej): #czarnyprotest, #protestmedyków i #reprzywatyzacja. Podobnie jak w przypadku pełnego korpusu tweetów wyodrębnione przez algorytm wątki (słowa kluczowe) nie odzwierciedlały emocjonalnego zabarwienia komunikatów. Bez głębszej analizy jakościowej trudno byłoby ustalić stosunek autorów tweetów do danego tematu.

Natomiast dzięki modelowaniu tematycznemu w ramach hashtagów #czarnyprotest i #protestmedyków udało nam się dostrzec kontrnarracje, czyli świadome próby zmiany pola dyskusji poprzez przekierowanie jej z postulatów politycznych na inny temat. Tropem prowadzącym do kontrnarracji było pojawienie się kilku słów kluczowych wyraźnie oderwanych od głównego nurtu dyskusji, który w przypadku obu hashtagów dotyczył postulatów politycznych: dostępności aborcji i antykoncepcji w przypadku czarnego protestu i zwiększenia nakładów na służbę zdrowia w przypadku protestu rezydentów.

W zależności od ujawnianych sympatii politycznych użytkownicy albo wyrażali poparcie dla protestów i podnoszonych żądań, albo odnosili się do nich krytycznie. Analizowane

tweety niejednokrotnie charakteryzowały się silnym ładunkiem emocjonalnym, wykorzystywały sarkazm, kpinę, a nawet wulgaryzmy. Nadal jednak odnosiły się bezpośrednio do politycznych postulatów.

Kontrnarracje były propagowane przez użytkowników należących do grupy zielonej i miały wyraźnie krytyczny charakter. Natomiast nie były polemiczne, tj. nie odpowiadały na postulaty protestujących. W kilku przypadkach odnotowaliśmy próby zaśmiecenia hashtagu treściami niezwiązanymi z jego pierwotnym przeznaczeniem.

Zobacz kontrnarracje dla poszczególnych hashtagów. →

#czarnyprotest

W dyskusji związanej z protestami kobiet pojawienie się kontrnarracji sygnalizowały zwroty „zagraniczny kapitał”, „oszukiwały” oraz „dyktatury kobiet”. Dwa pierwsze zostały wprowadzone do debaty przez prawniczkę fundacji Ordo Iuris, działającej m.in. na rzecz zaostreżenia przepisów aborcyjnych, i dotyczyły podejrzanych – zdaniem autorki tej narracji – sposobów finansowania czarnego protestu z zagranicznych źródeł. Przedstawienie protestów jako walki dla idei było wg Ordo Iuris oszustwem, którego dopuściły się feministki wobec protestujących kobiet ①.

Druga kontrnarracja ② nawiązywała do przejęzyczenia posłanki Nowoczesnej, Joanny Scheuring-Wielgus, która podczas pierwszego czarnego protestu w październiku 2016 r. skandowała hasło „Dość dyktatury kobiet!”, mając na myśli zerwanie z dyktaturą mężczyzn. Potknięcie to posłużyło użytkownikom z grupy zielonej do promowania dyskursu ośmieszającego inicjatywę kobiet.

W porównaniu do ogólnej liczby tweetów

①

Karina Walinowicz @k_walinowicz · 3 Oct 2017
Milion zagranicznego kapitału na #czarnyprotest! Nie ma mowy o walce dla jakiegokolwiek idei, kobiety wręcz oszukano!

Translate from Polish



"Rzeczpospolita": milion na "czarny protest"
Organizacje z zagranicy chwają się, że ich pieniądze przyczyniły się do nacisków na polski rząd - informuje dzisiaj "Rzeczpospolita". Do polskich wiadomosci.onet.pl

6 25 51

Karina Walinowicz @k_walinowicz · 5 Oct 2017
Jutro na zaproszenie @EHoldynska podyskutujemy w @wPolscepl o tym, jak feministki oszukiwały Polki na #czarnyprotest! ordoiuris.pl/ochrona-zycia/...

Translate from Polish

6 14 40

②

Brat Wodza @BratWodza · 4 Oct 2017
#CzarnyWtorek #CzarnyProtest
"Dość dyktatury kobiet!"
youtu.be/yNupy_kHD9c przez @YouTube

Translate from Polish



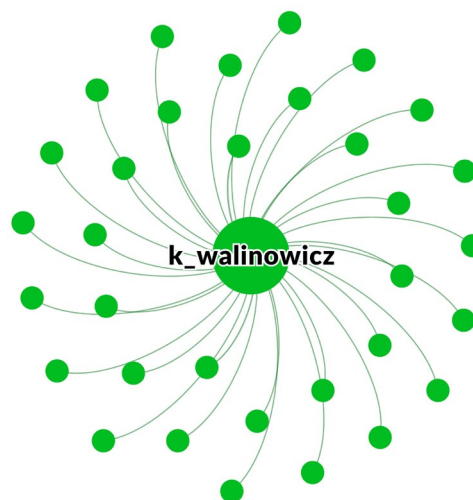
5 25 49

#czarnyprotest

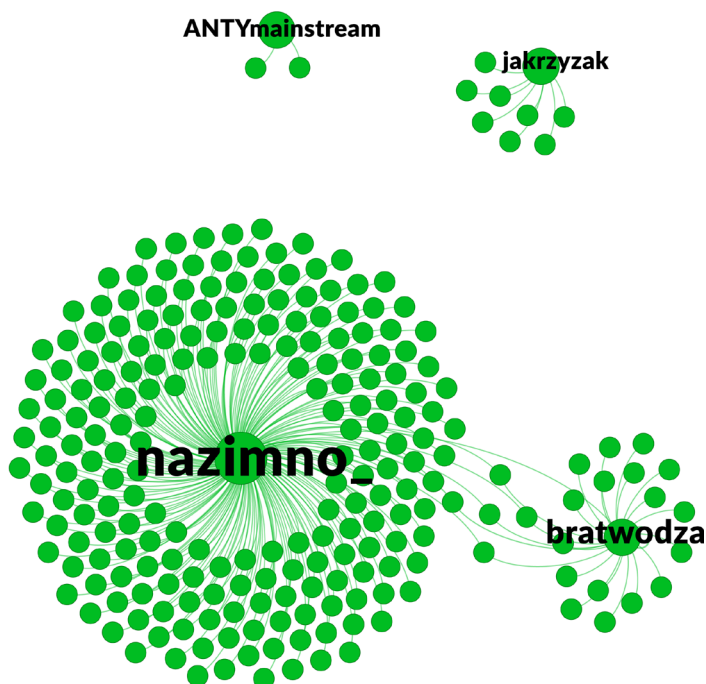
wygenerowanych na temat protestów kobiet kontrnarracje nie zyskiwały większej nośności. Oryginalny tweet o finansowaniu czarnego protestu przez zagraniczny kapitał został podany dalej zaledwie 25 razy. Nieco większą nośność zyskała ironiczna kontrnarracja wzywająca do zakończenia dyktatury kobiet – łącznie pojawiło się na ten temat 266 tweetów.

Analizowane przez nas kontrnarracje były zapoczątkowane przez internetowych influencerów, cieszących się wpływem w grupie użytkowników o zbliżonych poglądach. W ich dalszym rozpowszechnianiu brali udział autentyczni użytkownicy, a nie przejęte czy sztucznie wyhodowane konta. Analiza sieciowa przeprowadzona dla kont zaangażowanych w rozpowszechnianie polemicznych tweetów prowadzi dodatkowo do wniosku, że kontrnarracje nie wydostały się poza grupę zieloną, przez co nie wpłynęły istotnie na zmianę dyskursu na temat czarnego protestu.

Wykres 4. Analiza sieciowa dla kontrnarracji charakteryzującej się słowami: „zagraniczny kapitał” i „oszukały”



Wykres 5. Analiza sieciowa dla kontrnarracji charakteryzującej się słowami: „dyktatury kobiet”



#protestmedykow

W dyskusji prowadzonej w ramach hashtagu #protestmedykow zidentyfikowaliśmy trzy wyraźne kontrnarracje: komentującą rzekome luksusowe wakacje rezydentów w egzotycznych krajach, promującą iluzoryczność protestu głodowego oraz kwestionującą apolityczność całego protestu. Ich celem było przede wszystkim odwrócenie uwagi od postulatów rezydentów oraz pokazanie domniemanej hipokryzji protestujących lekarzy.

Największą nośność zyskała kontrnarracja dotycząca zagranicznych podróży rezydentów ①, które zostały przedstawione jako luksusowe wakacje, podczas gdy w rzeczywistości były związane z praktykami zawodowymi, wyjazdami na misje humanitarne lub – w nielicznych przypadkach – wyjazdami wakacyjnymi sprzed kilku lat. Najpopularniejsze tweety promujące ten dyskurs („Z życia Rezydenta: Meksyk, kurort narciarski w Dolomitach, 9 zł na godzinę” oraz „Z życia Rezydentki: Malta, Tanzania, Vegas, Rumunia, Kurdystan”) wykreował, należący do grupy zielonej, użytkownik o nicku Suwerenna, który został zawieszony w związku z naruszeniem regulaminu Twittera. W ramach tej kontrnarracji pojawiło się ponad 1000 podań dalej i tweetów w podobnym tonie.

Jako jedyna z analizowanych kontrnarracji właśnie ta wywołała również reakcje w grupie fioletowej. Było to najprawdopodobniej związane z podjęciem wątku z Twittera przez media głównego nurtu – 14 października TVP Info zamieściła na swoim portalu internetowym materiał *Kontrowersje wokół rezydentów. Narzekają na zarobki, a jedzą kanapki z kawiozem*. Ten artykuł Komisja Etyki TVP w listopadzie 2017 r. uznała za naruszający zasady etyki dziennikarskiej. W grupie fioletowej wszelkie retweety lub nawiązania do kontrnarracji polegały na konsekwentnym jej zbijaniu. Ten przykład pokazuje, że sam

①

Marcin Wyrwał @Wyrwal · 16 Oct 2017
 Dr Pikulska pracowała w Kurdystanie z @FundacjaPCPM. Byłem z nimi w Libanie, widziałem: to żadne wakacje, tylko ciężka orka. #ProtestMedyków



Translate from Polish

2 16 49

Tygodnik WPROST @TygodnikWPROST · 15 Oct 2017
 Fala hejtu na rezydentkę po publikacji TVP Info. Zarzucono jej... wyjazdy na misje medyczne #ProtestMedyków



Translate from Polish

Fala hejtu na rezydentkę po publikacji TVP Info. Zarzucono jej... wyja...
 Trwa protest lekarzy rezydentów. Po fiasku negocjacji z rządem, w internecie rozpoczęła się burza wokół tematu wakacyjnych wyjazdów nie...
 wprost.pl

14 8 19

Tomasz Wasilewski @WasilewskiTomek · 14 Oct 2017
 "Pokaż rezydencie z wakacji zdjęcie" to współczesna wersja "Pokaż lekarzu co masz w garażu" 🤔 #protestmedykow

Translate from Polish

28 127 385

MRCS @MRCS1977

Replying to @WasilewskiTomek

Jeśli wyjazd na misję medyczną do Kurdystanu to "wakacje", to pokazujemy takich zdjęć jak najwięcej! #ProtestMedykow

Translate from Polish

9:02 pm - 14 Oct 2017

#protestmedykow

fakt wydostania się kontrnarracji poza grupę, w ramach której została wypromowana, nie musi być jednoznaczny z jej rozpowszechnieniem ani tym bardziej uznaniem przez inne grupy. ②

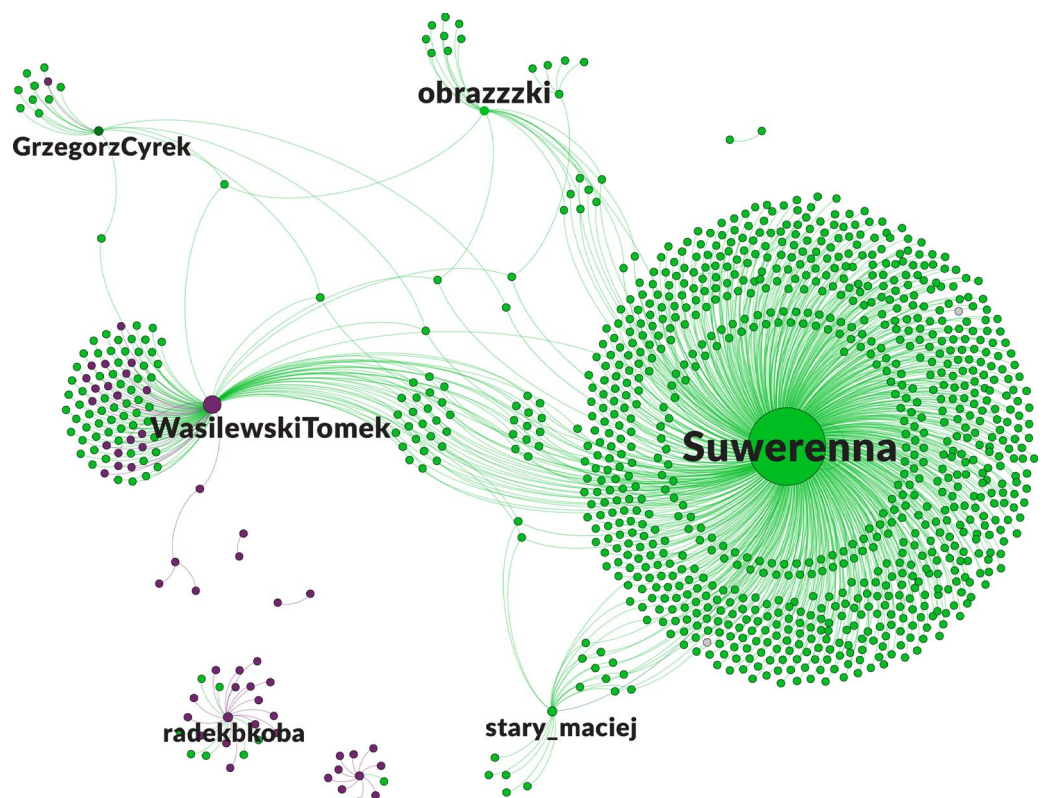
Druga kontrnarracja zmierzała do wypromowania przekonania, że prowadzony przez młodych lekarzy protest głodowy był iluzoryczny i został wykreowany na potrzeby medialne ②. W tym wątku pojawiła się m.in. insynuacja, że protestujący przerywali głodówkę, gdy tylko media głównego nurtu kończyły transmisję.

Jeden z najpopularniejszych tweetów pochodził od użytkownika o nicku PikoPOL, łudząco podobnym do jednego z kluczowych influencerów w grupie zielonej – Pikuś-POL. Nie jesteśmy w stanie zweryfikować charakteru konta PikoPOL (np. czy nie było



źródło: Twitter

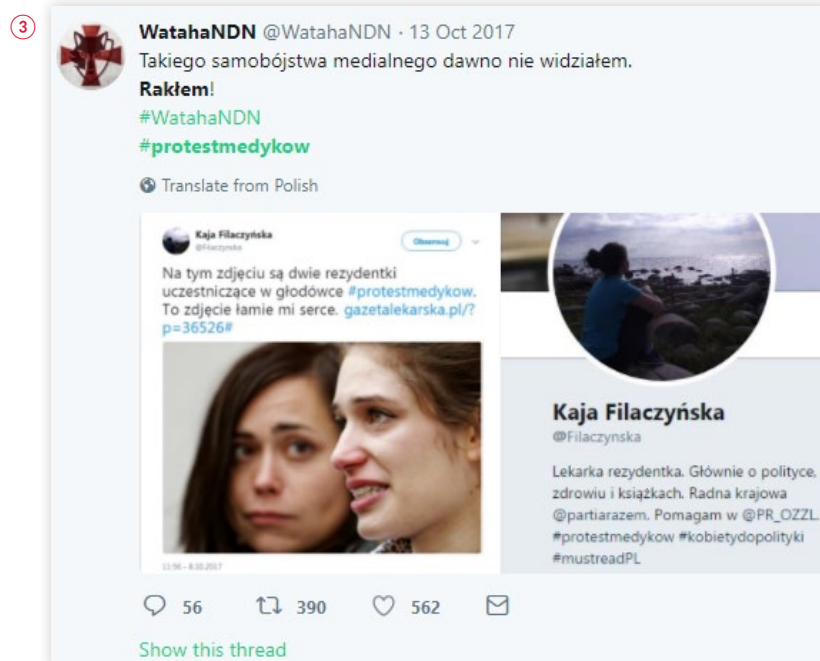
Wykres 6. Analiza sieciowa dla kontrnarracji charakteryzującej się słowami: „Dolomitach”, „Kurdystan”, „Rumunia”, „Vegas”, „Tanzania”, „kurort”, „Meksyk”, „wakacje”



#protestmedyków

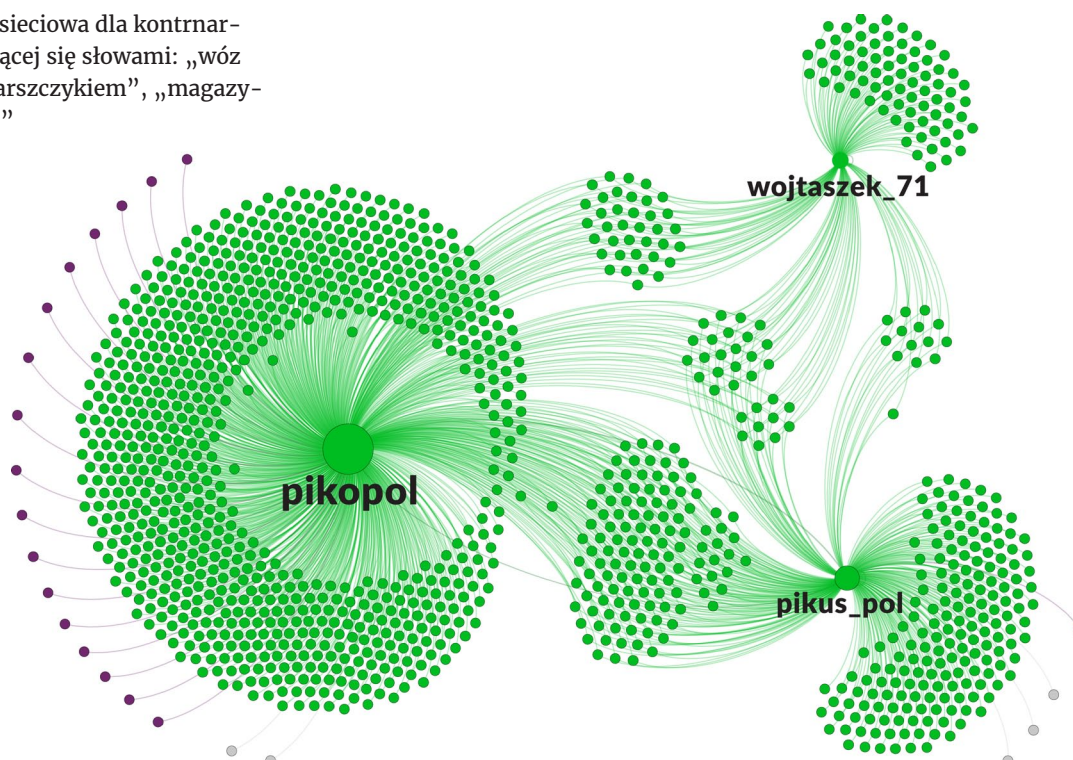
ono próbą podszycia się pod realnego użytkownika), ponieważ zostało usunięte przez właściciela. Treść zamieszczonego przez to konto tweeta brzmiała: „łapcie hit na dziś – mówią, że to protest głodowy, a tutaj za tymi panelami jest magazynek żywności”. Tweet został podany dalej ponad 1000 razy. Analiza sieciowa dla kont zaangażowanych w rozpowszechnianie tej kontrnarracji pokazuje jednak, że – z nielicznymi wyjątkami – nie przebiła się ona do grupy fioletowej.

Trzecia kontrnarracja miała na celu podważenie apolitycznego charakteru całego protestu poprzez powiązanie go z partią Razem ^③. Kluczowym argumentem dla promujących ją użytkowników było to, że jedna z protestujących rezydentek była członkinią partii



źródło: Twitter

Wykres 7. Analiza sieciowa dla kontrnarracji charakteryzującej się słowami: „wóz transmisyjny”, „barszczykiem”, „magazynek”, „spożywacie”



#protestmedykow

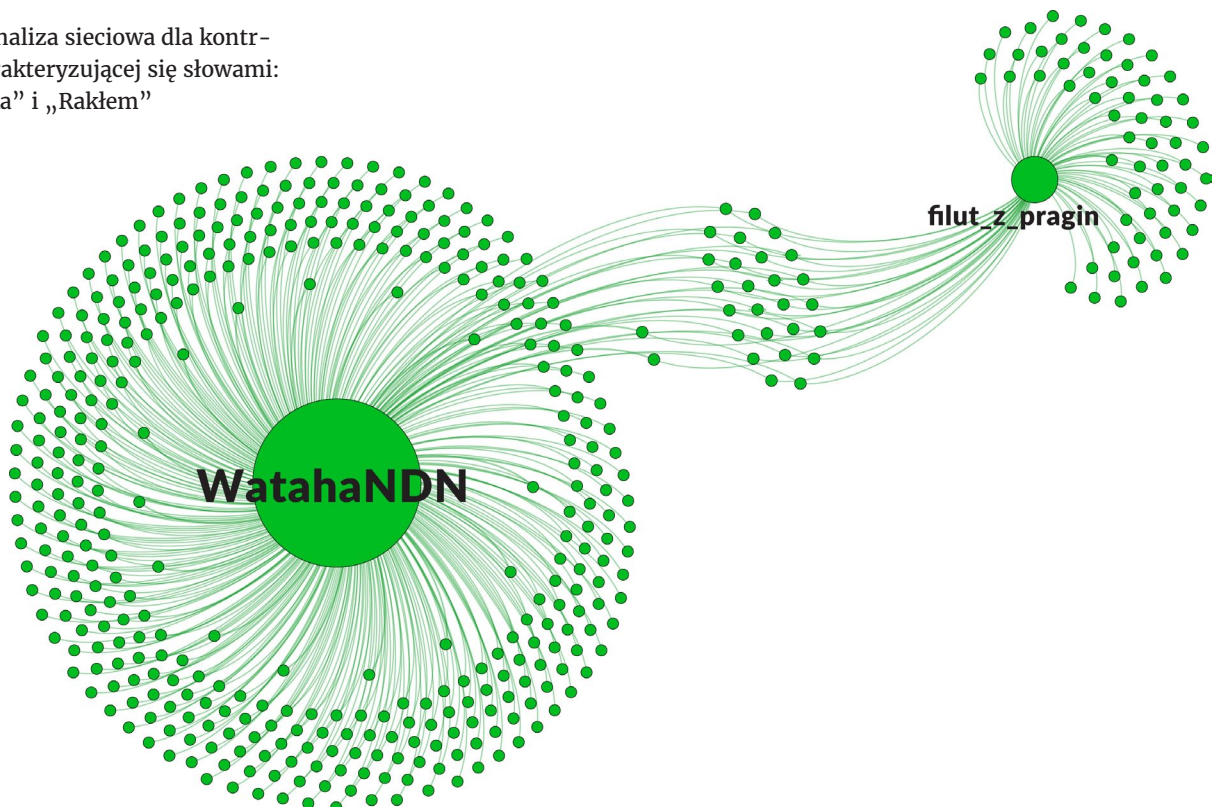
Razem, a sama partia publicznie wyraziła poparcie dla protestujących. Również ta kontrnarracja nie wywołała reakcji w grupie fioletowej.

Podsumowując: kontrnarracje związane z protestem medyków zyskały większy zasięg niż te wykreowane w ramach dyskusji o czarnym proteście. Jedna z nich, dotycząca rzekomych luksusowych wakacji rezydentów, okazała się wyjątkowo nośna, co znalazło wyraz w mediach głównego nurtu, a następnie zaangażowaniu użytkowników z grupy fioletowej. Jednak i w tym przypadku zaangażowanie poza grupą zieloną miało na celu wyłącznie zabicie polemicznych tez, a nie ich rozpowszechnianie.



źródło: Twitter

Wykres 8. Analiza sieciowa dla kontrnarracji charakteryzującej się słowami: „towarzyszka” i „Rakłem”



Co z tego wynika?

Analiza sieciowa przeprowadzona dla wiódących tematów oraz generowanych w ich ramach kontrnarracji pokazuje, że na Twitterze w zasadzie nie ma miejsca na ucieranie poglądów, szczególnie tych krytycznych, pomiędzy grupami o odmiennych sympatiach politycznych. Mimo znacznej liczby retweetów wewnątrz grupy zielonej żadna z krytycznych kontrnarracji nie zyskała nośności w grupie przeciwnej. Przykład tej jednej, która wywołała reakcje w grupie fioletowej, pokazuje, że do konfrontacji poglądów pomiędzy spolaryzowanymi grupami dochodzi dopiero w momencie zaangażowania tradycyjnych mediów, które mają szerszy zasięg.

Nasze obserwacje prowadzą do wniosku, że dyskusja w mediach społecznościowych toczy się przeważnie w tzw. bańkach, rozumianych jako odrębne i nierzadko spolaryzowane albo wzajemnie sobą niezainteresowane społeczności.

■ Nie jesteśmy w stanie tej tezy potwierdzić bez dodatkowych danych (odzwierciedlających wyświetlenia poszczególnych tweetów konkretnym użytkownikom, należącym do różnych społeczności), jednak wydaje się prawdopodobne, że osoby funkcjonujące w różnych bańkach nie są konfrontowane z poglądami, które zyskały nośność w innych społecznościach (w tym z krytycznymi kontrnarracjami), mimo że dyskutują na ten sam temat.

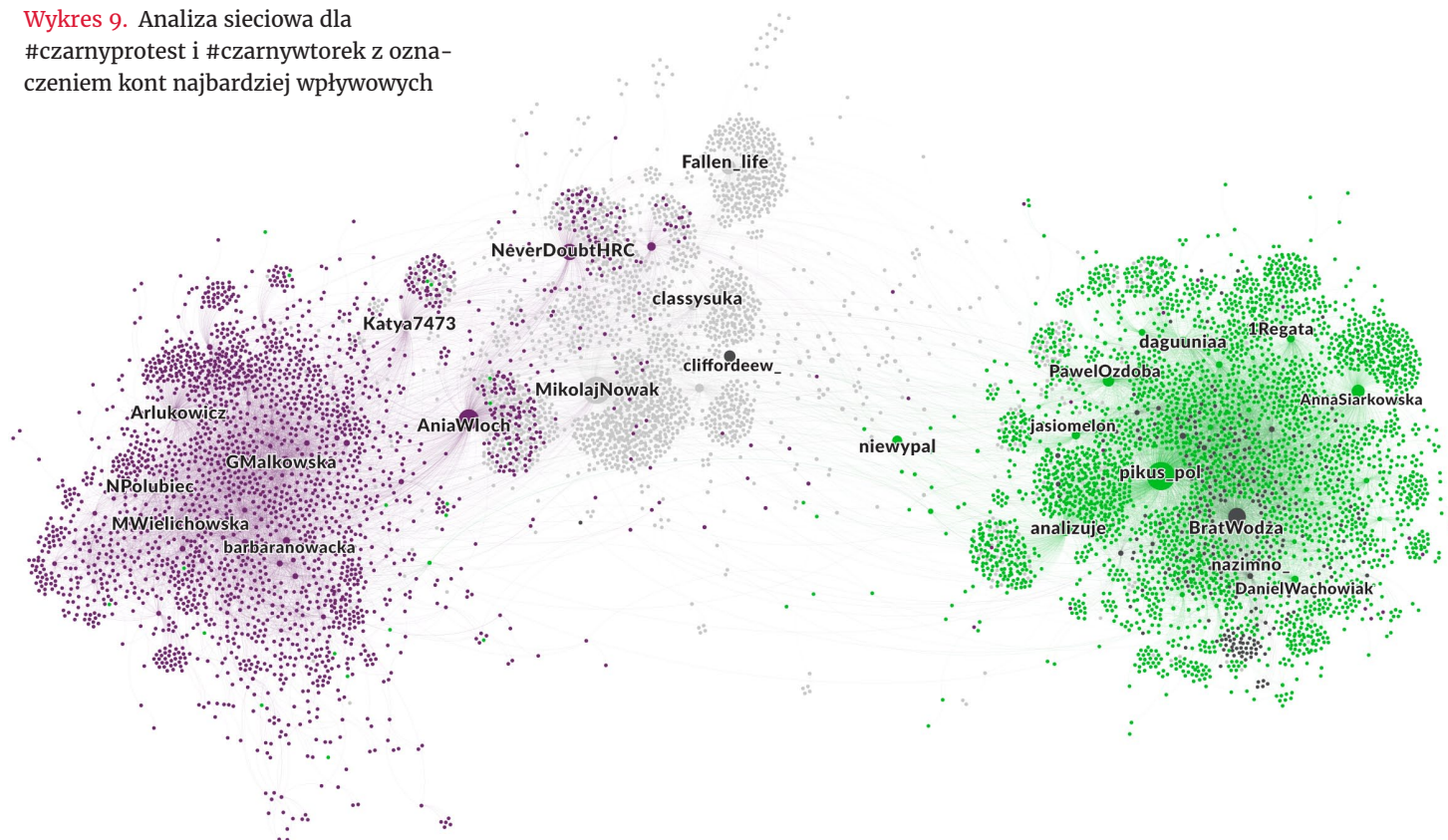
II. Czy debata polityczna podlegała manipulacji (np. z użyciem botów)?

Nie boty, tylko ludzie

Analizując debatę polityczną na polskim Twitterze, nie odnotowaliśmy istotnego wpływu przejętych lub sztucznie wyhodowanych kont na kierunki dyskusji i pojawiające się w niej wątki. W badanym okresie kluczowe wątki do debaty politycznej były wprowadzane i promowane przez użytkowników o ugruntowanym autorytecie w ramach danej grupy i możliwej do zweryfikowania tożsamości (np. **politycy, dziennikarze, media głównego nurtu**).

Nieco inna dynamika dotyczyła **kontrnarracji**, jakie pojawiały się w ramach wiodących hashtagów. Ich źródłem rzadziej były najbardziej opiniotwórcze media czy kluczowi politycy, a częściej **internetowi influencerzy** funkcjonujący niemal wyłącznie w ramach grupy o zbliżonych poglądach (np. Ordo Iuris, BratWodza, PikuśPOL, WatahaNDN). W efekcie również same kontrnarracje nie rozchodziły się poza obręb danej grupy, chyba że zostały podchwyczone przez media głównego nurtu i były przez konkurencyjną grupę konsekwentnie zbijane (por. omówiony w poprzednim punkcie przykład kontrnarracji, która zyskała największy zasięg w ramach dyskusji #protestmedyków).

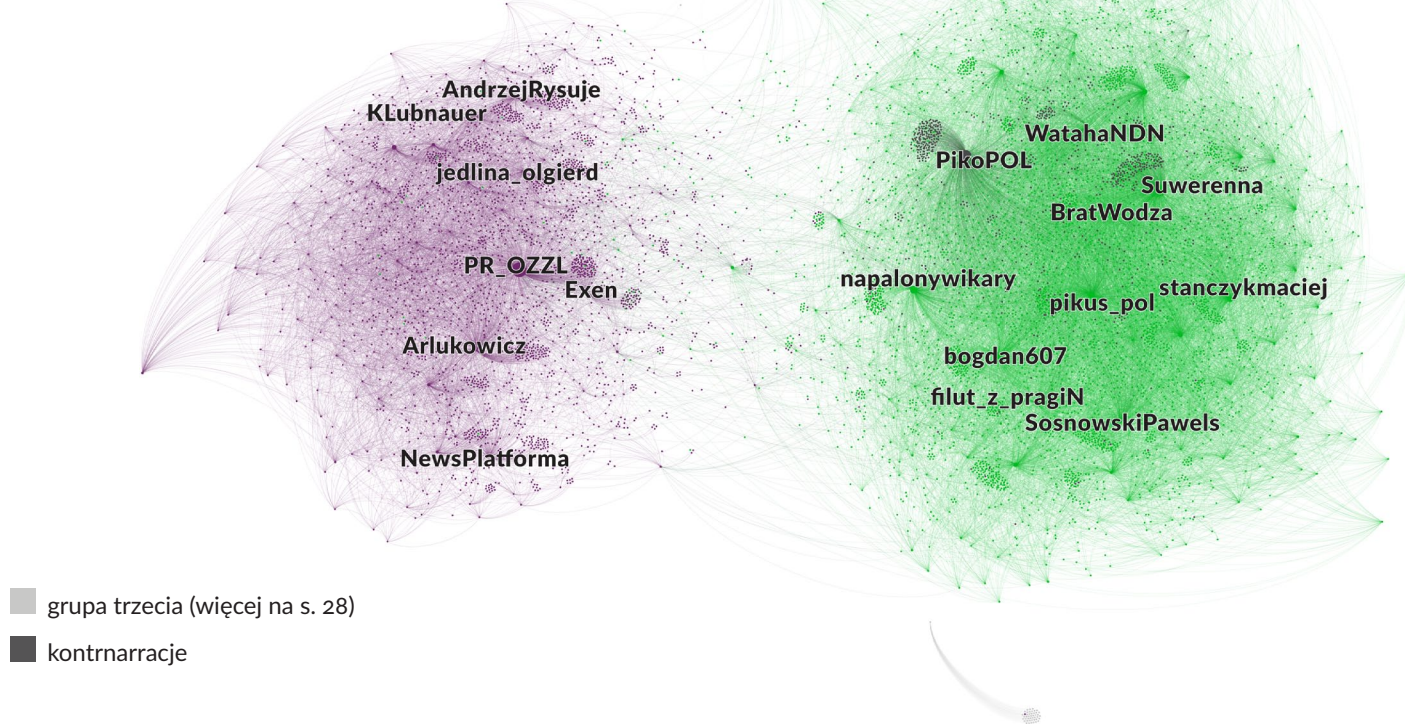
Wykres 9. Analiza sieciowa dla #czarnyprotest i #czarnywtorek z oznaczeniem kont najbardziej wpływowych



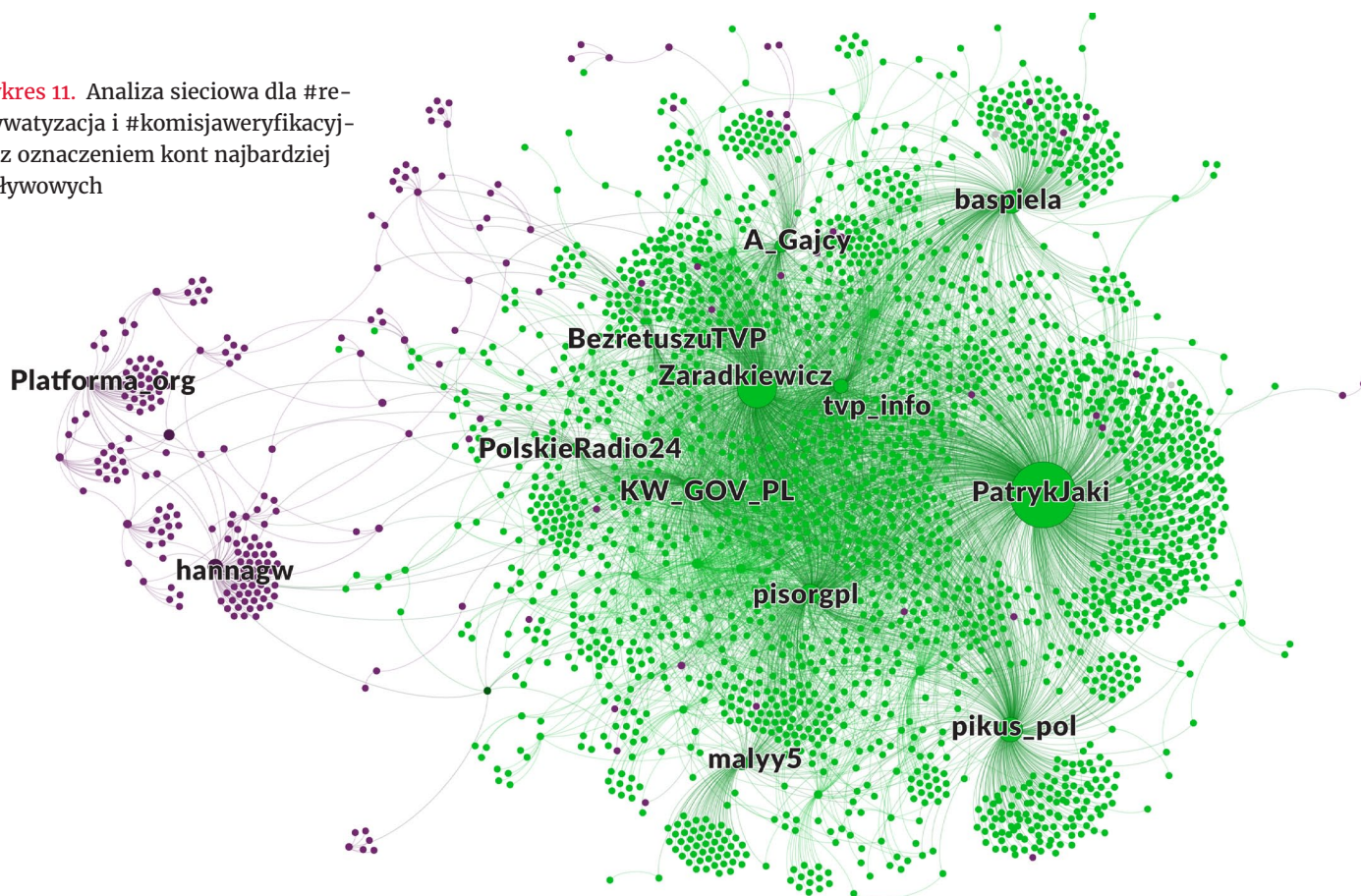
■ grupa trzecia (więcej na s.28)

■ kontrnarracje

Wykres 10. Analiza sieciowa dla #protestmedyków i #protestrezydentów z oznaczeniem kont najbardziej wpływowych



Wykres 11. Analiza sieciowa dla #re-privatyzacja i #komisjaweryfikacyjna z oznaczeniem kont najbardziej wpływowych



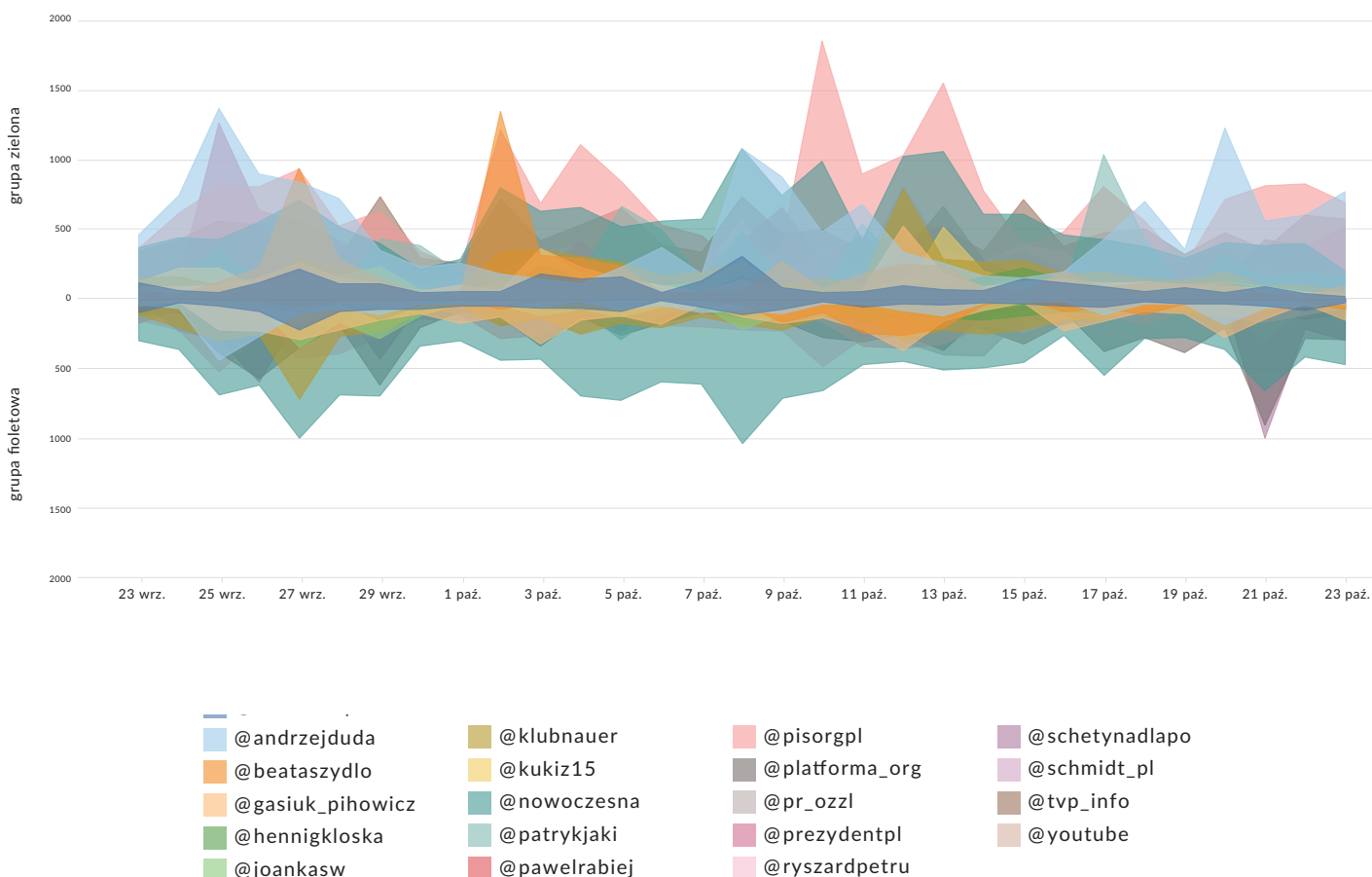
Ta obserwacja potwierdza naszą wyjściową tezę, że istotny wpływ na kierunek debaty politycznej na Twitterze (tj. ponad podziałem na grupy użytkowników o zbieżnych poglądach) mają przede wszystkim osoby i organizacje o ustalonym, społecznym autorytecie, a nie konta przejęte lub sztucznie wyhodowane.



Najczęściej są to politycy, dziennikarze oraz same media i organizacje polityczne, a więc aktorzy, których pozycja nie bazuje wyłącznie na aktywności w mediach społecznościowych. Wątki promowane przez tych aktorów mają największą szansę przyjąć się poza określoną grupą. Oczywiście, to nie oznacza braku polaryzacji w ramach danego tematu, ale jest wymiernym przejawem skali oddziaływania głównych aktorów.

Poniżej omawiamy sposób dojścia do tej konkluzji i najważniejsze czynniki, jakie wzięliśmy pod uwagę w ramach badania.

Wykres 12. 20 najczęściej wspomnianych kont dla obu grup – analiza częstotliwości wzmianek w badanym okresie



Nie ilość, ale jakość

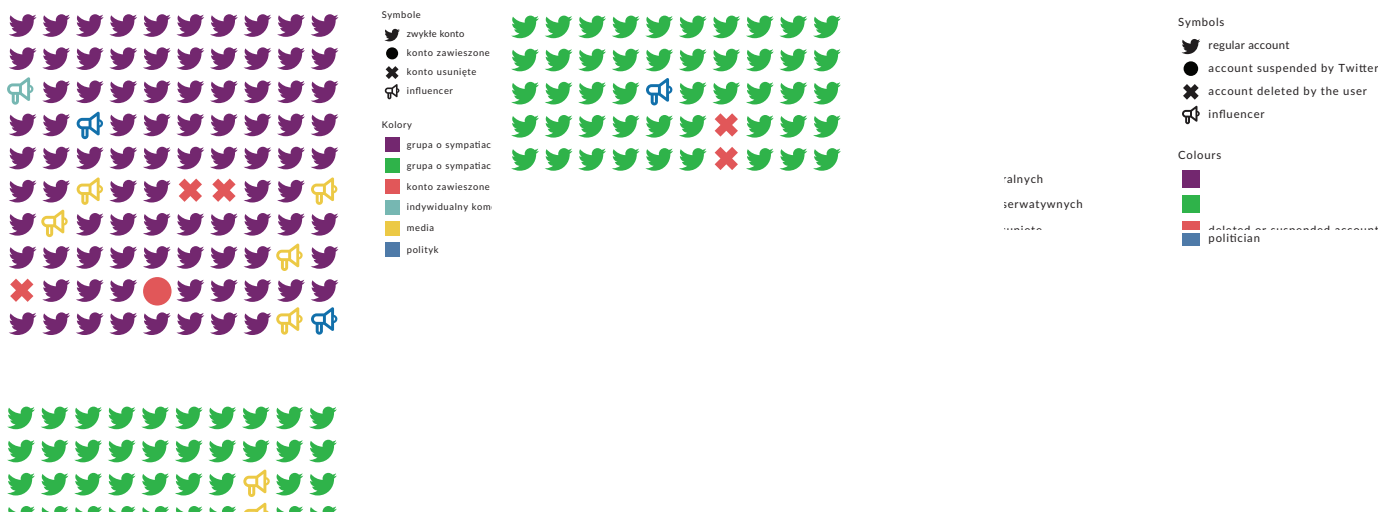
Porównanie kont najbardziej aktywnych i wpływowych dla obu analizowanych grup (graf poniżej, w zestawieniu z analizą sieciową pokazującą najbardziej wpływowych użytkowników por. Wykres nr 1) potwierdza, że częstotliwość i liczba wysyłanych komunikatów, którą łatwo podbić w sztuczny sposób, nie przekłada się na wpływ (określony na podstawie liczby i jakości interakcji).

W przypadku obu badanych grup wszystkie konta, które znalazły się zarazem w zbiorze najbardziej aktywnych i wpływowych, należą do rozpoznawalnych mediów i polityków.

Ich ponadstandardowa aktywność może wynikać z faktu, że są obsługiwane przez więcej niż jedną osobę. Wydaje się jednak niewątpliwe, że o ich wpływie w ramach danej grupy zdecydowała społeczna tożsamość i przypisywany im autorytet. Nie zidentyfikowaliśmy ani jednego przypadku konta, które przez sam fakt swojej wzmożonej aktywności (charakterystycznej dla botów) miało istotny wpływ na kierunek dyskusji toczącej się na Twitterze.

Wśród najbardziej wpływowych kont kształtujących debatę polityczną w ramach badanych przez nas grup pojawiają się też **influencerzy o wyrazistej internetowej tożsamości** (np. PikuśPOL, daguuniaa, antyKOD, avanti_ultras89, Polskawruinie1). W tej kategorii pojawiają się stałe pseudonimy i jednoznaczna reprezentacja poglądów politycznych, ale rzadziej zawodowa czy partyjna afiliacja. To sprawia, że zweryfikowanie prawdziwej tożsamości twitterowych influencerów może być nieco trudniejsze. Bez analizy historycznej, czyli przyjrzenia się zachowaniu tych kont od momentu założenia (w szczególności temu, w jakie dyskusje angażowały się na przestrzeni czasu), nie możemy również stwierdzić, czy za postaciami influencerów nie stoją inni mocodawcy o określonej agendzie. Nadal jednak nie ulega wątpliwości, że są to realne osoby, których wpływ na debatę ma stabilne oparcie w budowanych przez nie relacjach w ramach sieci społecznościowej. Ten wniosek potwierdza zarówno jakościowa analiza generowanych przez te konta interakcji (język i treść komunikatów, zawartość emocjonalna), jak i analiza sieciowa (to, z kim i w jaki sposób wchodzi w interakcje).

Wykres 13. 100 najbardziej aktywnych użytkowników w obu grupach (z oznaczeniem kont usuniętych lub zawieszonych oraz influencerów) wraz z ich przynależnością do określonej kategorii



Podejrzane konta

Wzmocniona aktywność pochodząca z podejrzanych kont na Twitterze jest stosunkowo łatwa do wykrycia i zneutralizowania. Z perspektywy innych użytkowników konto, za którym nie stoi żaden społeczny autorytet, a które nagle uaktywniło się w debacie politycznej, po prostu nie będzie wiarygodne i raczej nie zyska widoczności wykraczającej poza własny strumień komunikatów. W związku z tym skuteczniejszym – i trudniejszym do wykrycia – mechanizmem realizowania wpływu na Twitterze byłoby wykorzystanie przejętych lub sztucznie wyhodowanych kont do zwiększania widoczności i zasięgu prawdziwych kont, za którymi stoją popularni politycy lub media.

Wychodząc z tego założenia, w toku badania przyglądaliśmy się nie tylko tym kontom, które zostały uznane za najbardziej wpływowe wg wskaźnika PageRank Centrality (tj. generowały najwięcej jakościowych interakcji), oraz tym, które wykazywały wzmocnioną aktywność (charakterystyczną dla botów), ale wszystkim, których zachowanie odbiegało od schematu typowego dla realnych osób. Byliśmy ciekawi, ile takich kont było aktywnych w debacie politycznej na Twitterze i jaką agendę polityczną wspierały.

W poszukiwaniu podejrzanych kont skorzystaliśmy z metody opracowanej przez Obserwatorium on Social Media działające przy Uniwersytecie Indiany. Opracowany przez ten zespół algorytm nauczył się rozpoznawać fałszywe wzmacniacze, analizując zachowanie kont, co do których wiadomo, że zostały przejęte przez boty. W procesie uczenia maszynowego wykorzystano ok. tysiąca cech i kryteriów, aby ustalić, które z nich najlepiej opisują zachowanie przejętych lub w inny sposób sterowanych kont.

Wypracowana w ten sposób klasyfikacja opiera się na dwóch kategoriach czynników: takich, które wykorzystują analizę językową (treść, ładunek emocjonalny), i takich, które opisują sposób działania oraz pozycję danego konta w ramach sieci społecznościowej

(częstotliwość komunikatów, liczbę interakcji, typ i liczbę followersów, długość życia i charakterystykę samego konta). Czynniki uzależnione od analizy językowej można wykorzystywać tylko dla języka angielskiego, dlatego naszą próbę wykrycia podejrzanych kont mogliśmy oprzeć jedynie na czynnikach z drugiej kategorii. To ważne zastrzeżenie, ponieważ skuteczność działania algorytmu jest w związku z tym odpowiednio niższa.

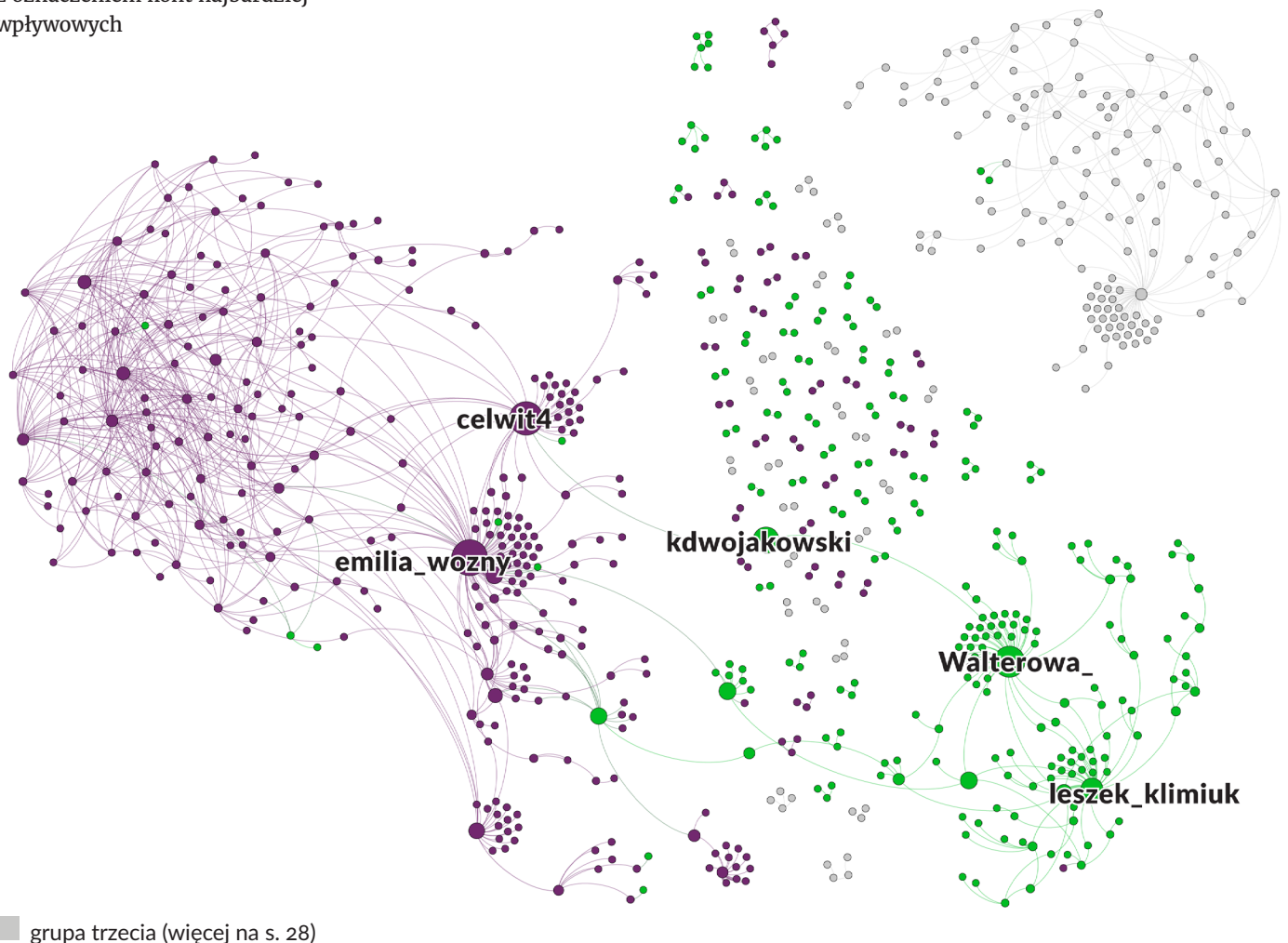
Posługując się tą metodą, zidentyfikowaliśmy ponad 500 kont (na ok. 50 tys. biorących udział w debacie politycznej w badanym okresie), które nie zachowywały się jak zwyczajni użytkownicy Twittera. Jak pokazała analiza jakościowa, od tej ogólnej konstatacji jest jeszcze długa droga do stwierdzenia, że dane konto zostało przejęte przez botnet albo zostało w sztuczny sposób wyhodowane. Analiza ograniczona do badania sposobu, w jaki dane konto funkcjonuje w ramach sieci społecznościowej (np. jak często wysyła komunikaty, czy wchodzi w interakcje z innymi, kto je śledzi), właściwie nie pozwala odróżnić kont przejętych lub wyhodowanych od takich, które są zarządzane w profesjonalny sposób.

Z perspektywy analizy sieciowej zachowanie kont należących do mediów, organizacji politycznych i polityków (zarządzanych w sposób profesjonalny) przypomina zachowanie botów.



Często są to konta obsługiwane przez więcej niż jedną osobę, wykorzystujące dodatkowe oprogramowanie (np. Tweet Deck), które pozwala sprawniej zarządzać interakcjami (np. bardzo szybko lub konsekwentnie reagować na zaczepki). W przypadku mediów informacyjnych standardem jest częste nadawanie (tweetowanie co kilka minut przez cały dzień). Z kolei konta znanych polityków są nastawione nie tylko na nadawanie, ale także szybkie wyłapywanie zaczepki i konsekwentne reagowanie. Z punktu widzenia analizy

Wykres 14. Analiza sieciowa przeprowadzona dla podejrzanych kont, pokazująca ich przynależność do trzech badanych grup na Twitterze, z oznaczeniem kont najbardziej wpływowych



sieciowej oba modele aktywności łatwo pomylić z działaniem odpowiednio zaprogramowanych botów.

Najprawdopodobniej właśnie z tych powodów na listę 504 podejrzanych kont zidentyfikowanych przez algorytm Uniwersytetu Indiany trafiły m.in. konta: J_Lewandowski, JanRokita, donaldtusk, KMAGpl, Zycie_Warszawy, EP_ThinkTank, PO_Suwalki, NCK_PL, PodcastTOKFM, TVP3Warszawa. Te przypadki dobitnie pokazują ograniczenia zastosowanej przez nas metody badania. Analiza sieciowa przeprowadzona dla podejrzanych kont pokazuje również, że ich rozłożenie pomiędzy analizowanymi grupami biorącymi aktywny

udział w debacie politycznej jest w miarę równomierne, z nieznaczną przewagą w grupie fioletowej.

Zachowania typowe dla botów, które mogą być zorientowane zarówno na ciągłe nadawanie, jak i na częste interakcje z innymi użytkownikami, na poziomie analizy sieciowej trudno odróżnić także od zachowania internetowych trolli. Tym bardziej że – jak wyjaśniamy w definicji pojęcia bota – w operacjach propagandowych wymagających większej finezji naturalnym wyborem będzie większe zaangażowanie czynnika ludzkiego w tworzenie treści lub wybór taktyki ich rozprzestrzeniania (tzw. cyborgi). Ten problem

został dostrzeżony przez komentatorów (m.in. przez analityka z serwisu Mashable⁵) po tym, jak Twitter opublikował 200 tys. tweetów, które wcześniej sklasyfikował jako rosyjską propagandę rozpowszechnianą przez boty w czasie amerykańskiej kampanii prezydenckiej w 2016 r.⁶ Analiza jakościowa potwierdziła, że wiele z zawieszonych przez Twittera kont było obsługiwanych przez ludzi, a nie przez zaprogramowane maszyny.

Te obserwacje nie prowadzą jeszcze do wniosku, że kont przejętych lub sztucznie wyhodowanych, które zniekształcają debatę w mediach społecznościowych, nie da się wykryć przy użyciu algorytmów. Ale na pewno przestrzegają przed szybkim wyciąganiem wniosków, szczególnie w oparciu o same metadane i analizę sieciową.

Kluczowymi czynnikami, które powinny być wzięte pod uwagę przy próbie identyfikacji botów, są: treść wysyłanych komunikatów, ich ładunek emocjonalny oraz kontekst. Bez zejścia na ten głębszy poziom nie jesteśmy w stanie odróżnić kont zarządzanych w sposób profesjonalny (lub z innego powodu wykazujących nietypowe cechy) od fałszywych wzmocniaczy.

W przeprowadzonym badaniu nie podjęliśmy próby przeprowadzenia takiej jakościowej analizy dla całego zbioru danych, ponieważ dostępne na niekomercyjnych zasadach algorytmy są trenowane tylko w języku angielskim. Jednak jak pokazują doświadczenia innych badaczy oraz krytyka, z jaką spotkał się sam Twitter – klasyfikując komunikaty

wysyłane przez realnych użytkowników jako działanie botów – analizowanie treści, ładunku emocjonalnego i kontekstu tweetów w sposób automatyczny jest bardzo trudne i obciążone (zbyt) dużym marginesem błędu.

Konta, które zniknęły

Analizując nietypowe schematy zachowania w debacie politycznej na polskim Twitterze, natrafiliśmy też na trop wpływowych lub niestandardowo aktywnych kont, które po kilku miesiącach zniknęły, uniemożliwiając nam głębsze zbadanie ich charakterystyki (przede wszystkim zastosowanie algorytmu do identyfikacji podejrzanych kont, opisanego w poprzednim punkcie). Ustaliliśmy, że na niemal 50 tys. kont, które brały udział w debacie politycznej w badanym okresie, do stycznia br. zniknęło aż 4056 (zostały usunięte przez właścicieli lub zawieszono w związku z naruszeniem regulaminu Twittera).

Aby sprawdzić, czy aktywność tych konkretnych kont wykazywała inną charakterystykę (np. była przesunięta w czasie albo promowała inną agendę), porównaliśmy ich zaangażowanie i wykorzystywane przez nie słowa kluczowe w ramach poszczególnych wątków (hashtagów) na przestrzeni czasu z aktywnością pozostałych kont, które brały udział w debacie politycznej. Ta analiza nie przyniosła większych zaskoczeń. Okazało się, że – z nielicznymi wyjątkami – schemat zachowania usuniętych i zawieszonych kont w badanym okresie nie odbiegał od ogólnego wzorca.

5 https://mashable.com/2018/02/14/examples-of-russian-twitter-bots/?utm_cid=hp-r-3#5DOWLwD6_PqZ.

6 https://www.nbcnews.com/tech/social-media/now-available-more-200-000-deleted-russian-troll-tweets-n844731?cid=sm_npd_nn_tw_ma.

Jednym z wyjątków, które odnotowaliśmy, była wzmożona aktywność tych kont w dwóch momentach: 27 września i 12 października 2017 r. Jednocześnie w generowanych przez te konta tweetach pojawiała się wzmianka konta *kosiniakkamysz*, zapewne w związku z dyskusją, jaką wywołało głosowanie nad wotum nieufności dla ministra rolnictwa. Ten wątek nie był równie widoczny w ogólnej debacie, jaka w tym czasie toczyła się na Twitterze. Drugą specyfiką związaną z usuniętymi i zawieszonymi kontami była nadreprezentacja w tym zbiorze słów kluczowych o charakterze religijnym i przekleństw.

Co ciekawe, aktywność usuniętych i zawieszonych kont w spolaryzowanych politycznie grupach, jakie wyodrębniliśmy w ramach analizy sieciowej, rozkłada się zaskakująco równomiernie (po 2 konta na 100 najbardziej aktywnych dla każdej grupy; liczba retweetów na poziomie 2,8%).

Te obserwacje nie dały nam podstaw do dalej idących wniosków. Nawet przy założeniu, że wśród usuniętych i zawieszonych kont znalazły się konta przejęte lub sztucznie wyhodowane, które były wykorzystywane w debacie politycznej na jesieni 2017 r., ich aktywność nie miała znaczącego wpływu na kierunki tej debaty, a ewentualne próby wygenerowania odrębnych narracji nie okazały się skuteczne (np. w wyniku ich działania na Twitterze nie pojawiły się nowe hashtagi).

Trzecia grupa kont

Ostatnim tropem, który badaliśmy w poszukiwaniu prób oddziaływania na debatę polityczną na Twitterze poprzez przejęte lub sztucznie wyhodowane konta, była punktowa aktywność trzeciej grupy, która nie pasowała do schematu politycznej polaryzacji i odróżniała się – pod względem zainteresowań – od dwóch wcześniej opisywanych grup. Na podstawie opisów 20 najbardziej wpływowych kont i analizy jakościowej ich tweetów ustaliliśmy, że tę grupę tworzą młode osoby zainteresowane rozrywką, popkulturą i własną popularnością.

W badanym okresie te konta nie były regularnie aktywne w debacie politycznej, natomiast uruchamiały się punktowo, w związku z konkretnymi wydarzeniami: między 3 a 5 października w temacie *#czarnyprotest*, a między 11 a 13 października w związku z zamieszczonym przez konto *republikaTV* tweetem krytycznie oceniającym możliwość adopcji dzieci przez pary homoseksualne. Najbardziej prawdopodobne wydawało się, że był to wynik spontanicznej politycznej aktywności osób należących do tej grupy, a nie efekt przejęcia lub sklonowania niektórych kont w celu wywarcia wpływu na debatę polityczną. Niemniej zdecydowaliśmy się sprawdzić obie hipotezy.

Jedną z inspiracji do tej analizy był materiał opublikowany w styczniu br. przez New York Times⁷, który opisuje rozpowszechnioną w Stanach Zjednoczonych praktykę klonowania kont społecznościowych należących przeważnie do nastolatków po to, by zamaskować działanie botnetu, wykorzystywanego w dowolnych celach marketingowych. Dodatkowy asumpt do podejrzeń, że w tej grupie mogły się pojawić przejęte lub sztucznie hodowane konta, dał wysoki stosunek retweetów do autorskich tweetów (na poziomie 85%) oraz większa aktywność kont usuniętych i zawieszonych (odpowiadały za 1 950 z 11 328 autorskich tweetów na tematy polityczne).

Analiza jakościowa tweetów na tematy polityczne, które zostały wygenerowane przez tę grupę, potwierdziła hipotezę o spon-tanicznym, punktowym zaangażowaniu konkretnych osób. Zdecydowana większość komunikatów oznaczonych hashtagiem #czarnyprotest lub #czarnywtorek wyrażała osobisty stosunek ich autorów do trwających protestów albo była próbą dowiedzenia się czegoś więcej ^①. Z kolei materiał opublikowany przez konto republikatv sprowokował młodych do krytycznych komentarzy wobec tego – ich zdaniem nietolerancyjnego – komunikatu ^②.

①



②



źródło: Twitter

7 New York Times, *The Follower Factory*, 27 stycznia 2018 r., <https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html>.

PODSUMOWANIE

Na pierwszy rzut oka

W sieci żyjemy w bańkach: najchętniej rozmawiamy z ludźmi, którzy myślą podobnie, szukamy potwierdzenia własnych poglądów, dzielimy się tymi informacjami, które chcemy promować, a nie tymi, które mamy potrzebę zbić. To truizm. Dlatego w wynikach przeprowadzonego badania nie zaskoczyło nas odkrycie dwóch wyraźnie wyodrębnionych i spolaryzowanych grup użytkowników – jednej o sympatiach konserwatywnych, drugiej – liberalnych. Na ich styku znaleźli się przede wszystkim użytkownicy, którzy szukali konfrontacji z przedstawicielami grupy przeciwnej, ostro polemizując z rozpowszechnianymi przez nich narracjami.

Nie zaskoczyło nas również to, że centralną rolę w kształtowaniu debaty politycznej na Twitterze odgrywają osoby i instytucje o wysokim społecznym lub politycznym autorytecie (znani dziennikarze, politycy, partie, redakcje informacyjne). Z kolei za generowanie nowych – krytycznych lub polemicznych – wątków, które nazwaliśmy kontrnarracjami, odpowiadają przede wszystkim internetowi influencerzy, prezentujący bardziej radykalne pozycje.

W żadnej z wpływowych ról nie występowały konta przejęte albo sztucznie wyhodowane (sterowane przez boty). Takie konta funkcjonują na Twitterze i mogą być wykorzystywane do budowania pozycji wpływu, ale raczej w roli klaki niż kluczowych aktorów, których działanie byłoby w stanie zmienić kierunek debaty politycznej.



Nasze ustalenia, dotyczące wpływu botów (fałszywych wzmacniaczy) na debatę polityczną, jaka toczy się na Twitterze, nie pozwalają na dalej idące wnioski dotyczące cyfrowej propagandy. Na podstawie zebranych danych możemy bronić tezy, że sztucznie wyhodowane lub przejęte konta w analizowanym okresie nie miały istotnego wpływu na kierunki i dynamikę debaty. Jednak – bez głębszej analizy jakościowej oraz przyjrzenia się historii najbardziej wpływowych kont – nie możemy stwierdzić, czy działalność internetowych influencerów, którzy wybijają się w analizie sieciowej, nie nosi znamion zorganizowanej propagandy.

Analiza sieciowa pozwala w stosunkowo łatwy sposób zidentyfikować społeczności i funkcjonujących w ich ramach liderów opinii dla danego tematu, który polaryzuje opinię publiczną.

Taką wiedzę media publiczne mogłyby z powodzeniem wykorzystać do realizowania swojej misji, odegrywając rolę swoistego pomostu między nierozmawiającymi ze sobą społecznościami.



Schodząc głębiej...

Jedna obserwacja okazała się dla nas jeśli nie zaskakująca, to przynajmniej trzeźwiąca: mimo tego, że obie grupy rozmawiały na te same tematy (dyktowane przez wydarzenia polityczne i społeczne, jakie miały miejsce w badanym okresie), nie dochodziło między nimi do polemik ani ucierania poglądów. Kontrnarracje, czyli wątki odbiegające od głównego nurtu dyskusji lub zorientowane na zmianę jej kierunku, w zasadzie nie przedostawały się poza grupę zieloną, która je generowała.

Nie oceniając ani głównego nurtu dyskusji, ani wątków polemicznych, dostrzegamy problem w tym, że debata w mediach społecznościowych nie prowadzi do ścierania się poglądów, co w efekcie mogłoby budować polityczny konsensus lub poczucie wspólnoty.

■
Wśród analizowanych przez nas kontrnarracji tylko jedna została dostrzeżona przez grupę przeciwną. Powodem jej przebiccia się poza bańkę najprawdopodobniej było to, że została przechwycona przez media głównego nurtu (TVP Info). Mimo tego, że media głównego nurtu w Polsce również uległy silnej polaryzacji i działają w oparciu o podobny mechanizm (tj. mówią do przekonanych), ich zasięg jest na tyle szeroki, że promowane przez nie kontrnarracje trafiają do świadomości odbiorców o innych poglądach, prowokując do polemiki.

Nadal wiemy za mało

Mamy świadomość, że opisując jeden moment i jeden wąski wycinek z medialnej rzeczywistości, zajrzeliśmy jedynie pod powierzchnię. Brakuje narzędzi do badania i opisywania świata nowych mediów, który zmienia się zbyt szybko i wytwarza zbyt dużo materiału dla klasycznych socjologicznych czy antropologicznych metod. Sięgając po automatyczne metody analizy danych (uczenie maszynowe i algorytmy), musieliśmy zmierzyć się z ryzykiem tych samych błędów i uproszczeń, które krytykujemy w publicznej debacie na temat wykorzystywania *big data* w modelowaniu ludzkich zachowań.

Wnioski, jakie wyłaniają się z naszego studium przypadku, mają oparcie w zebranych danych, ale nie muszą być reprezentatywne dla debaty publicznej w mediach społecznościowych. Dlatego prezentujemy je nie jako prawdę objawioną, ale jako przyczynek do ważnej dyskusji, którą powinniśmy w Polsce toczyć na głębszym poziomie, w oparciu o rozmaite badania i dane pochodzące z różnych źródeł.

■
Media społecznościowe w coraz większym stopniu kształtują nasz obraz świata i poglądy polityczne. Temat ich regulacji i odpowiedzialności nie schodzi z agendy politycznej. Domagamy się od tych platform większej przejrzystości pod względem stosowanych technik i kryteriów profilowania treści. A jednocześnie w niewielkim stopniu zajmujemy się badaniem zachodzących w nich procesów społecznych, których głównymi aktorami są ludzie.

Szukając odpowiedzi na pytania, które sprotokowały to studium przypadku, dostrzeżliśmy **kilka obszarów, które warto zbadać głębiej.**

- Dynamika interakcji na styku baniek. Jakie cechy wyróżniają osoby funkcjonujące na styku spolaryzowanych społeczności? Jaka aktywność podejmują (odpowiadają czy raczej podają dalej z własnym komentarzem)? Czy próbują konfrontować drugą stronę z własnymi poglądami, czy raczej skupiają się na budowaniu silnej pozycji w swojej grupie (np. wyśmiewając przeciwnika)?
- Jeśli uprawnione jest stwierdzenie, że krytyczne kontrnarracje co do zasady nie przedostają się poza bańkę, w której zostały wygenerowane, to dlaczego tak się dzieje? Czy jest to związane ze sposobem rozchodzenia się informacji w sieciach społecznościowych („Nie docierają do mnie inne poglądy, więc na nie nie reaguję”), czy raczej z decyzjami, które podejmują sami użytkownicy („Widzę, ale nie reaguję, bo się nie zgadzam”)?
- Działanie botów. Czy w Polsce mamy do czynienia z kupowaniem sympatyków na większą skalę, a jeśli tak, jaki jest ich wpływ na dyskusję, która się toczy w ramach prawdziwych społeczności? Czy są w stanie oddziaływać na jej kierunek, mimo że działają w punktowy i rozproszony sposób? Jak skutecznie można wykrywać ten typ fałszywych wzmocniaczy?
- Działanie realnych użytkowników wykorzystujących narzędzia technologiczne w celu zniekształcenia debaty w mediach społecznościowych. Jak badać to zjawisko, które wydaje się mieć o wiele poważniejszy wpływ na kierunki debaty politycznej niż sama aktywność botów? Jak sensownie odpowiadać na propagandę lub dezinformację, którą rozprzestrzeniają realni użytkownicy mediów społecznościowych, wynajęci w tym celu przez grupy polityczne albo działający z pobudek ideologicznych?
- W jaki sposób na polskim Twitterze są rozpowszechniane tzw. fake newsy? Czy ich widoczność i zasięg podbijają konta sterowane przez boty czy raczej konta realnych politycznych influencerów? A jeśli te drugie – czy robią to świadomie?



FUNDACJA PANOPTYKON



DOŁÓŻ SWÓJ 1% PODATKU
KRS:0000327613



**INTERNET
POLICY
OBSERVATORY**