

Czy ptaki są prawdziwe? – raport analizy sieci społecznościowej użytkowników Twittera udzielających się pod tagiem #BirdsArentReal

Wprowadzenie

Ruch "Birds Aren't Real" to satyryczna kampania, która wyśmiewa rozprzestrzenianie się teorii spiskowych, szczególnie wśród pokolenia Z. Głosi on, trochę na przekór, że ptaki to tak naprawdę drony wysłane przez rząd do inwigilacji ludzi. Przez swój absurdalny charakter, ten pomysł ma zwracać uwagę na to, jak łatwo ludzie przyjmują i rozpowszechniają niezweryfikowane informacje oraz na to, jak bardzo w dzisiejszych czasach można manipulować prawdą.

Mimo że wszystko to zaczęło się jako zabawa, wywołało niemałe zamieszanie. Niektórzy ludzie wzięli teorię ruchu dosłownie, co pokazuje, jak trudno w dzisiejszym świecie cyfrowym odróżnić fakt od fikcji. Inicjatywa ta ma na celu zachęcenie do refleksji nad tym, jak przyjmujemy i rozpowszechniamy informacje, a także jak technologie i media wpływają na nasze życie i postrzeganie rzeczywistości. W ten sposób "Birds Aren't Real" łączy humor, społeczne komentarze i pewnego rodzaju aktywizm, odzwierciedlając sposób, w jaki młode pokolenia wykorzystują możliwości internetu do dyskusji i zwiększania świadomości na temat ważnych kwestii współczesnego świata, takich jak prawda i dezinformacja.¹



Rysunek 1 Furgonetka głosząca hasło Birds aren't real wraz z grupą zainteresowanych

¹ www.nytimes.com/2021/12/09/technology/birds-arent-real-gen-z-misinformation.html (stan na 30.05.2024)

² https://www.reddit.com/r/vsauce/comments/1axfi9n/birds_are_real_or_are_they/ (stan na 30.05.2024)

Niniejszy raport przedstawia analizę sieci społecznościowej użytkowników platformy społecznościowej X/Twitter, którzy udzielali się w tematach związanych z ruchem "Birds Aren't Real", czyli wpisy z konta @birdsarenotreal na Twitterze i użytkowników wykorzystujących w swoich wrzutkach hashtagu #birdsarenotreal.

Dane niefiltrowane, pierwotnie pobrane przez użytkownika Gabriel Preda z wykorzystaniem biblioteki Tweepy, pozyskaliśmy ze zbioru udostępnionego przez niego na platformie kaggle.com pod tym linkiem:

<https://www.kaggle.com/datasets/gpreda/birds-arent-real> (stan na 30.05.2024)

Raport

Wykorzystany przez nas do analizy zbiór zawiera w sobie 1903 obserwacje 16 zmiennych. Każda obserwacja jest pojedynczym wpisem na tablicy twittera albo użytkownika @birdsarenotreal albo pod tagiem #birdsarenotreal i zawiera w sumie wpisy 1326 różnych użytkowników.

Wykorzystane do analizy biblioteki:

```
import pandas as pd
import numpy as np
import re
from itertools import combinations
import networkx as nx
import matplotlib.pyplot as plt
from wordcloud import WordCloud
from bertopic import BERTopic
import nltk
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from textblob import TextBlob
```

Zmienne w zbiorze:

```
print(twitter.columns)
```

- id – id każdego tweeta.
- user_name - Nazwa użytkownika, który opublikował tweet.
- user_location - Lokalizacja użytkownika.
- user_description – Opis na profilu użytkownika.
- user_created - Data utworzenia konta na platformie.
- user_followers - Liczba obserwujących użytkownika.
- user_friends - Liczba obserwowanych.
- user_favourites - Liczba tweetów, które użytkownik dodał do ulubionych.
- user_verified – Informacja czy konto użytkownika jest zweryfikowane przez Twittera
- text - Treść tweeta.
- hashtags – Hashtagi użyte w tweecie

- source - Aplikacja lub urządzenie, z którego został wysłany tweet.
- retweets - Liczba retweetów, które otrzymał tweet.
- favorites - Liczba polubień, które otrzymał tweet.
- is_retweet – Informacja czy tweet jest retweetem innego tweeta.

Wstępna deklaracja zmiennych wykorzystanych w analizie:

```
#Deklaracja zmiennych

#Ekstrakcja wzmianek użytkowników z treści tweetów i zapisanie w nowej kolumnie 'mentions'
twitter['mentions'] = twitter['text'].apply(lambda x: re.findall("@([a-zA-Z0-9_]{1,50})", str(x)))

#Inicjalizacja skierowanego grafu do modelowania relacji między użytkownikami
G = nx.DiGraph()

#Filtrowanie użytkowników zweryfikowanych
verified_users = twitter[twitter['user_verified'] == True]

#Filtrowanie użytkowników niezwyfikowanych (potencjalne boty)
bots = twitter[twitter['user_verified'] == False]

#Łączenie wszystkich hashtagów w jeden ciąg tekstowy, pomijając brakujące dane
hashtag_data = twitter['hashtags'].dropna().str.cat(sep=' ')

#Wyodrębnienie unikalnych użytkowników wraz z ich opisami i usunięcie duplikatów
unique_users = twitter[['user_name', 'user_description']].drop_duplicates()
unique_users.columns = ['name', 'description']

#Łączenie wszystkich opisów użytkowników w jeden ciąg tekstowy, pomijając brakujące dane
desc_data = unique_users['description'].dropna().str.cat(sep=' ')

#Tworzenie par wzmianek dla zweryfikowanych użytkowników
mention_pairs = verified_users['mentions'].apply(lambda mentions:
list(combinations(mentions, 2)))
```

Poprzez przeszukanie treści tweetów, identyfikowane są wszystkie wystąpienia wzmianek o innych użytkownikach Twittera (oznaczone znakiem '@'). Wyniki są zapisywane w nowej kolumnie mentions. To pozwala na analizę, jak użytkownicy są ze sobą powiązani poprzez bezpośrednie odniesienia w swoich tweetach.

Utworzenie skierowanego grafu (DiGraph) za pomocą biblioteki NetworkX umożliwia modelowanie i analizę sieci społecznych. W kontekście Twittera, węzły mogą reprezentować użytkowników, a krawędzie - interakcje między nimi, takie jak wzmianki czy odpowiedzi.

Podział użytkowników na zweryfikowanych i niezwyfikowanych pozwala na analizę różnic w zachowaniu tych dwóch grup. Użytkownicy zweryfikowani często są bardziej wiarygodni lub mają większe wpływy, podczas gdy konta niezwyfikowane mogą być bardziej podejrzane lub nawet botami.

Łączenie hashtagów z różnych tweetów w jedną dużą ciągłą tekstową formę ułatwia analizę najpopularniejszych tematów lub trendów na Twitterze.

Pozyskanie listy unikalnych użytkowników wraz z ich opisami pozwala na dalszą analizę charakterystyk demograficznych lub interesów użytkowników.

Analiza par wzmianek wśród zweryfikowanych użytkowników może pomóc zidentyfikować wzorce współpracy lub wspólnoty w sieci, pokazując, jak użytkownicy są ze sobą połączeni przez wspólne wzmianki.

Cały kod dołączony do raportu w osobnym pliku *ptaszki.py*

Analiza sieci społecznościowej

Gęstość, przechodniość, ilość pętli, średni współczynnik skupienia i korelacja stopni wraz z wizualizacją zweryfikowanych użytkowników:

Gęstość sieci (zweryfikowani): 0.23976608187134502

Przechodniość (zweryfikowani): 0.42745098039215684

Liczba pętli (zweryfikowani): 0

Średni współczynnik klastra (zweryfikowani): 0.4819045161924419

Współczynnik korelacji stopnia (zweryfikowani): -0.18443370141767376

Gęstość sieci (0.24): Gęstość sieci wynosi około 0.24, co oznacza, że około 24% możliwych połączeń między węzłami (użytkownikami) w sieci jest rzeczywiście obecnych. Jest to stosunkowo wysoka gęstość dla sieci społecznościowej, sugerująca, że sieć jest dość gęsto połączona i użytkownicy mają tendencję do interakcji z wieloma innymi użytkownikami w sieci.

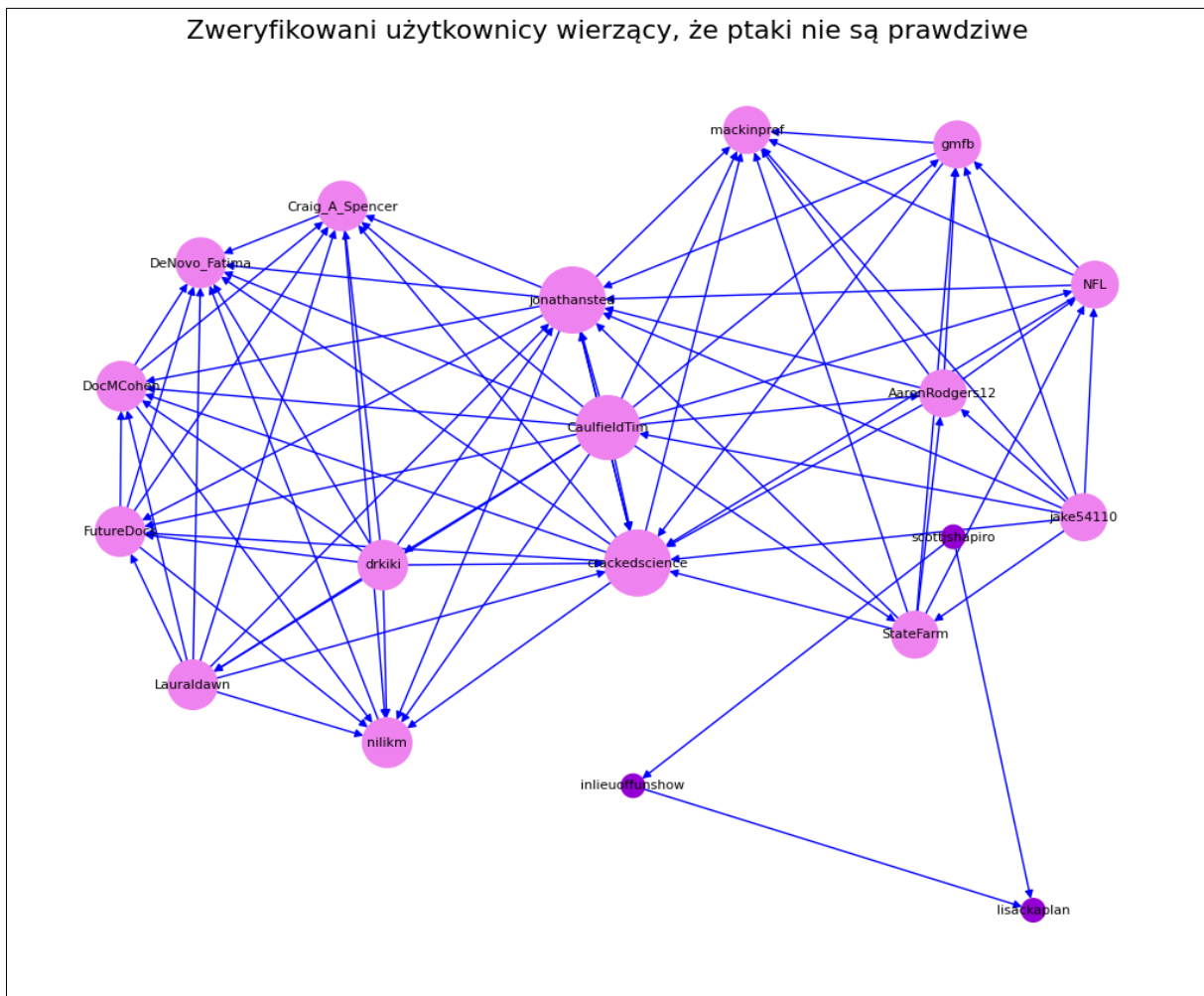
Przechodniość (0.43): Wartość przechodniości wynosi około 0.43. Oznacza to, że w około 43% przypadków, gdy użytkownik A wspomina użytkownika B, a użytkownik B wspomina użytkownika C, użytkownik A także wspomina użytkownika C. Jest to miara lokalnej spójności grup w sieci i wskazuje na istnienie grup użytkowników, którzy często wspominają się nawzajem.

Liczba pętli (0): W sieci nie ma żadnych pętli, co oznacza, że żaden użytkownik nie wspomina samego siebie w swoich tweetach. Jest to typowe dla większości sieci społecznościowych, gdzie interakcje są zazwyczaj skierowane do innych użytkowników.

Średni współczynnik klastra (0.48): Wartość średniego współczynnika klastra wynosi około 0.48. Oznacza to, że istnieje znaczna tendencja do tworzenia się zgrupowań w sieci, gdzie

użytkownicy, którzy są połączeni z tym samym użytkownikiem, mają również tendencję do wzajemnego połączenia. Wynik ten wspiera obserwacje z przechodniości i wskazuje na silne lokalne zgrupowania w sieci.

Współczynnik korelacji stopnia (-0.18): Ujemny współczynnik korelacji stopnia wskazuje, że w sieci występuje dysasortatywność, czyli użytkownicy o wysokim stopniu (duża liczba połączeń) mają tendencję do łączenia się z użytkownikami o niskim stopniu. To może sugerować, że bardziej wpływowi użytkownicy (z większą liczbą obserwujących) często wspominają mniej znane konta.



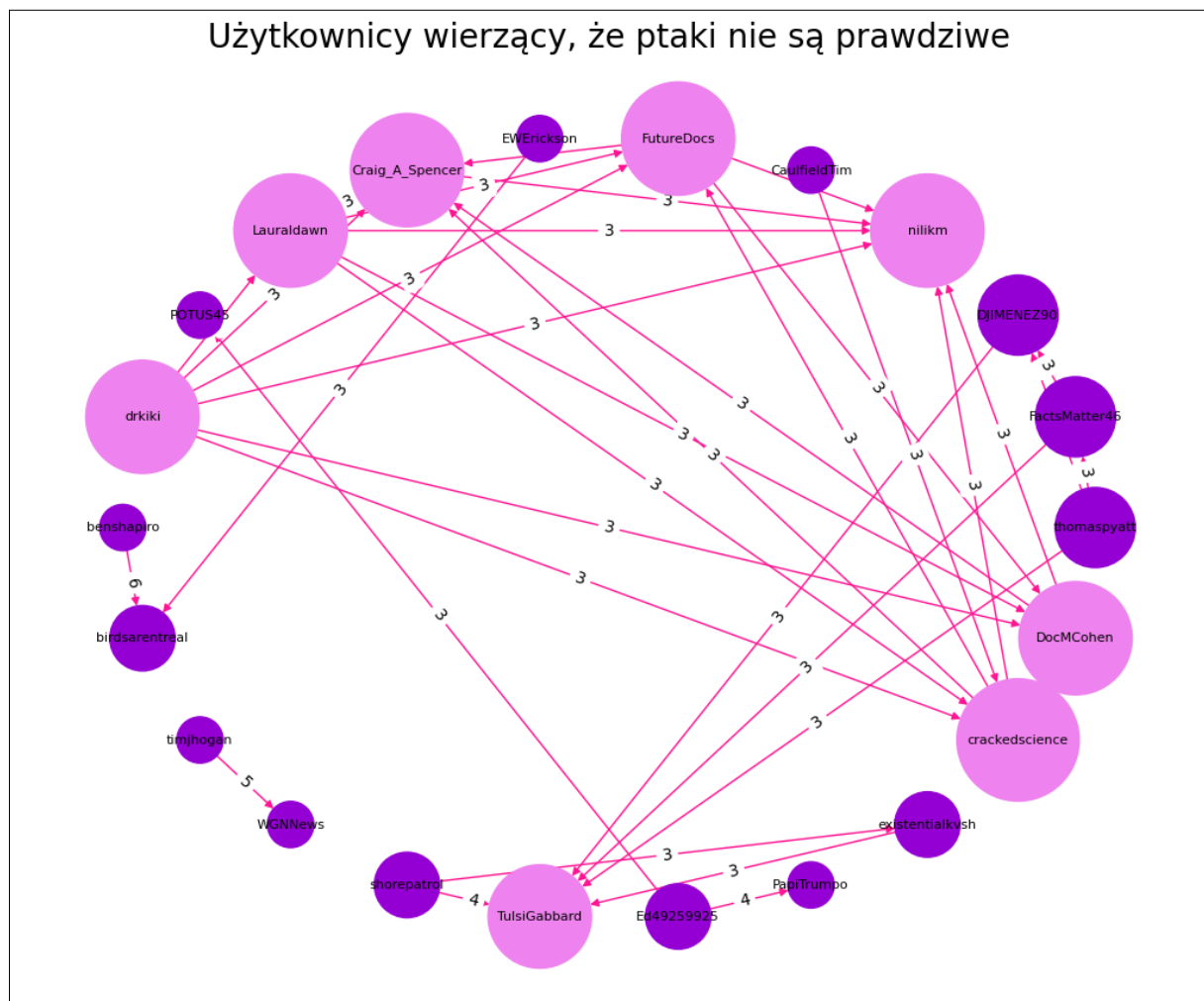
Widać, że niektóre węzły (użytkownicy) pełnią centralną rolę w sieci, co wskazuje na ich wysoki poziom aktywności lub wpływu. Przykładowo użytkownik o nicku *mackinprof* wydaje się być głównym węzłem, z którego wychodzi wiele krawędzi do innych użytkowników. Może to sugerować, że użytkownik ten jest liderem jakiejś opinii dotyczącej ruchu ptakosceptyków lub głównym propagatorem tematu w tej grupie. Można zauważyć, że niektórzy użytkownicy są silnie powiązani ze sobą, tworząc gęste subgrupy w sieci. Na przykład grupa skupiona wokół użytkowników takich jak *drkiki*, *Lauraldawn*, i *DeNovo_Fatima* pokazuje, że istnieje wysoki stopień współpracy lub wzajemnych wzmianek między nimi. Niektóre węzły, takie jak węzeł widoczny po lewej stronie łączący userów: *scottjshapiro*, *lisackaplan* i *inlieuoffunshow*, mają tylko jedno połączenie. Może to wskazywać na mniejszy stopień integracji z głównym

strumieniem dyskusji w sieci. Są to prawdopodobnie użytkownicy, którzy wspomnieli o temacie w mniej znaczący sposób lub tylko raz. Zgodnie również z poprzednią analizą, graf nie pokazuje żadnych pętli, co oznacza brak auto-wzmianek. Ogólnie rzecz ujmując, widoczna struktura może sugerować, że dyskusja lub wspomnienia są prowadzone głównie przez kilka kluczowych postaci, z dalszym rozprzestrzenianiem się tematu przez szerszą sieć mniejszych węzłów.

Wizualizacja wzmianek wszystkich użytkowników (również botów):

Po zgromadzeniu par wzmianek, przefiltrowaniu tylko tych, które wspomniały o sobie przynajmniej trzy razy, udało się utworzyć graf interakcji:

Tak samo jak w przypadku poprzedniego grafu, wielkość węzłów reprezentuje stopień

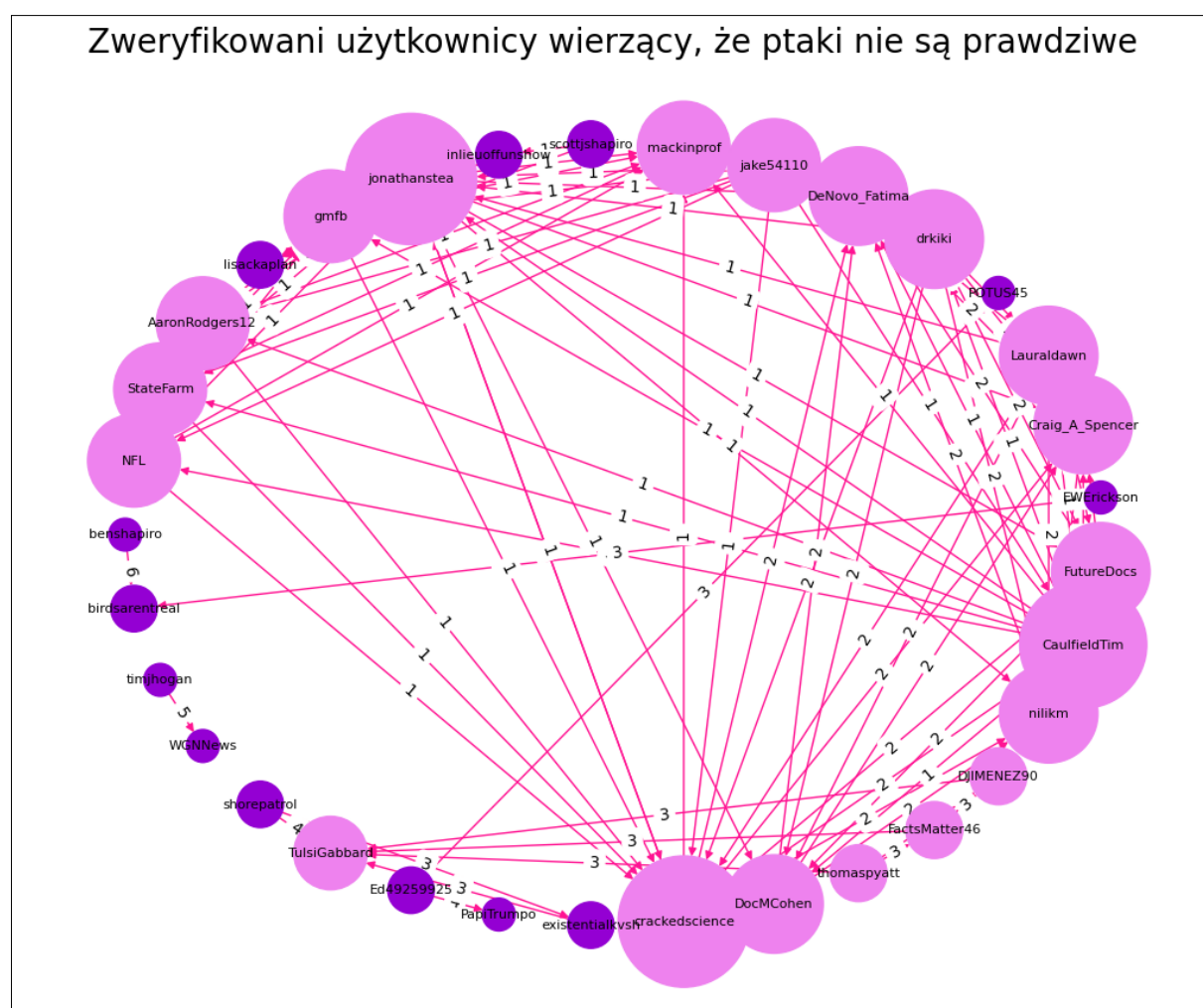


centralności każdego użytkownika, co w tym przypadku oznacza ilości bezpośrednich wzmianek danego użytkownika w tweecie innego. Większe węzły, jak np. *nilikm* czy *Lauraldawn*, wskazują userów, którzy są częściej wspominani w kontekście ptasiej teorii. Ponieważ w tej analizie zostali wybrani użytkownicy, którzy byli wspominani więcej niż 3 razy,

jest to ilustracja przedstawiająca twitterowiczów najbardziej aktywnych w tej kwestii. Być może któryś z nich to jeden z widocznych na fotografii zamieszczonej na pierwszej stronie! Z wykresu widzimy że najczęściej cytowani użytkownicy to: *drkiki*, *Lauraldawn*, *Craig_A_Spencer*, *FutureDocs*, *nilikm*, *DocMCohen*, *crackedscience* oraz *TulsiGabbard*. Wszyscy oni wchodzili w interakcje przynajmniej z 5 innymi użytkownikami przynajmniej 3 razy, co, patrząc przez pryzmat tego, że wszystkich badanych użytkowników było 1326 i znając realia funkcjonowania twittera jest naprawdę imponującym wynikiem.

Ta sama wizualizacja, lecz z uwzględnieniem wyłącznie zweryfikowanych użytkowników, tym razem bez filtrowania minimalnie trzech wspomnień:

Przez brak uwzględnienia przynajmniej trzykrotnej interakcji pomiędzy użytkownikami, na grafie widzimy wyłącznie zweryfikowanych użytkowników mających więcej niż trzy



połączenia ogólnie (wliczając różnych twitterowiczów). Z grafu możemy odczytać, że takich użytkowników jest całkiem dużo, o czym świadczy liczba węzłów i ich powiązań, co potwierdza wysoką gęstość (0,24) oraz przechodniość (0,43) wykazaną na początku analizy. Możemy też zauważyć nazwy użytkowników, które pojawiły się już na wcześniejszym grafie wskazującym wszystkich użytkowników, *drkiki*, *Lauraldawn*, *Craig_A_Spencer*, *FutureDocs*,

nilikm, *DocMCohen*, *crackedscience*, co wskazuje na to że najbardziej „rozchwytywani” są użytkownicy zweryfikowani. Użytkownicy o dużym węźle, ale generalnie niskiej liczbie wielokrotnych wzmianek jak np. *jonathanstea* zwracają uwagę. Może to świadczyć o np. jednej chwytliwej aktywności, pojedynczym wpisie, na który zareagowało wielu ptakosceptyków.

Wizualizacja społeczności #birdsarenoreal

Nazwa użytkownika	Centralność Stopnia	Betweenness Centrality	Centralność Bliskości	Centralność Wektorów Własnych
benshapiro	0.03125	0.0	0.0	5.39e-15
birdsarenoreal	0.0625	0.0	0.0625	3.93e-13
timjhogan	0.03125	0.0	0.0	5.39e-15
WGNNews	0.03125	0.0	0.03125	1.99e-13
shorepatrol	0.0625	0.0	0.0	5.39e-15
TulsiGabbard	0.15625	0.0	0.15625	5.30e-11
Ed49259925	0.0625	0.0	0.0	5.39e-15
PapiTrumpo	0.03125	0.0	0.03125	1.99e-13
existentialkvsh	0.0625	0.0	0.03125	1.99e-13
crackedscience	0.5	0.01277	0.28125	0.02702
DocMCohen	0.28125	0.0	0.23633	0.10807
thomaspyatt	0.09375	0.0	0.0	5.39e-15
FactsMatter46	0.09375	0.0	0.03125	1.99e-13
DJIMENEZ90	0.09375	0.0	0.0625	3.79e-12
nilikm	0.28125	0.0	0.29340	0.43229
CaulfieldTim	0.46875	0.00370	0.03125	1.99e-13
FutureDocs	0.28125	0.0	0.20833	0.05404
EW Erickson	0.03125	0.0	0.0	5.39e-15
Craig_A_Spencer	0.28125	0.0	0.26471	0.21615
Lauraldawn	0.28125	0.0	0.07031	4.56e-11
POTUS45	0.03125	0.0	0.03125	1.99e-13
drkiki	0.28125	0.0	0.04167	3.59e-12
DeNovo_Fatima	0.28125	0.0	0.32237	0.86454
jake54110	0.25	0.0	0.0	5.39e-15
mackinprof	0.25	0.0	0.26042	0.05404
scottjshapiro	0.0625	0.0	0.0	5.39e-15
inlieuoffunshow	0.0625	0.0	0.03125	1.99e-13
jonathanstea	0.5	0.01277	0.28125	0.02702
gmfb	0.25	0.0	0.15625	4.04e-09
lisackaplan	0.0625	0.0	0.0625	3.79e-12
AaronRodgers12	0.25	0.0	0.09375	4.92e-11
StateFarm	0.25	0.0	0.0625	3.79e-12
NFL	0.25	0.0	0.125	4.92e-10

Centralność Stopnia - Mówi o tym, ilu sąsiadów (bezpośrednich połączeń) ma dany węzeł. Na przykład, *crackedscience* ma najwyższą wartość (0.5), co oznacza, że jest bezpośrednio

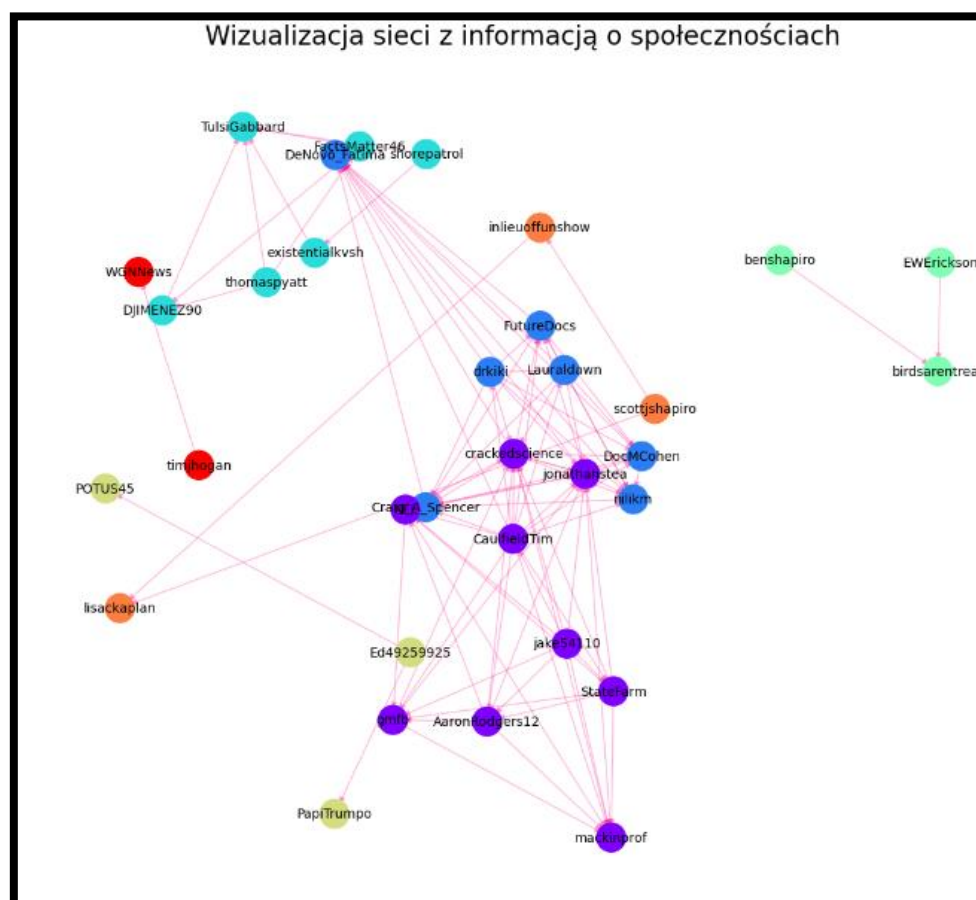
połączony z połową wszystkich innych węzłów w sieci. To może sugerować, że *crackedscience* jest bardzo aktywny lub wpływowy w tej sieci.

Betweenness Centrality - Pokazuje, jak często węzeł występuje na najkrótszych ścieżkach między parami węzłów. Wartości dla większości węzłów wynoszą 0, co oznacza, że te węzły nie są kluczowe w pośredniczeniu komunikacji między innymi węzłami. Wyróżniają się *crackedscience* i *jonathanstea*, które mają niezerowe wartości, sugerując ich kluczową rolę

Centralność Bliskości - Wskaźnik ten mówi o tym, jak blisko średnio znajduje się węzeł do wszystkich innych węzłów w sieci. Na przykład, *DeNovo_Fatima* ma najwyższą wartość (0.3223684210526316), co wskazuje na to, że średnio ma on najkrótsze ścieżki do innych węzłów, co może wskazywać na szybką zdolność rozprzestrzeniania informacji.

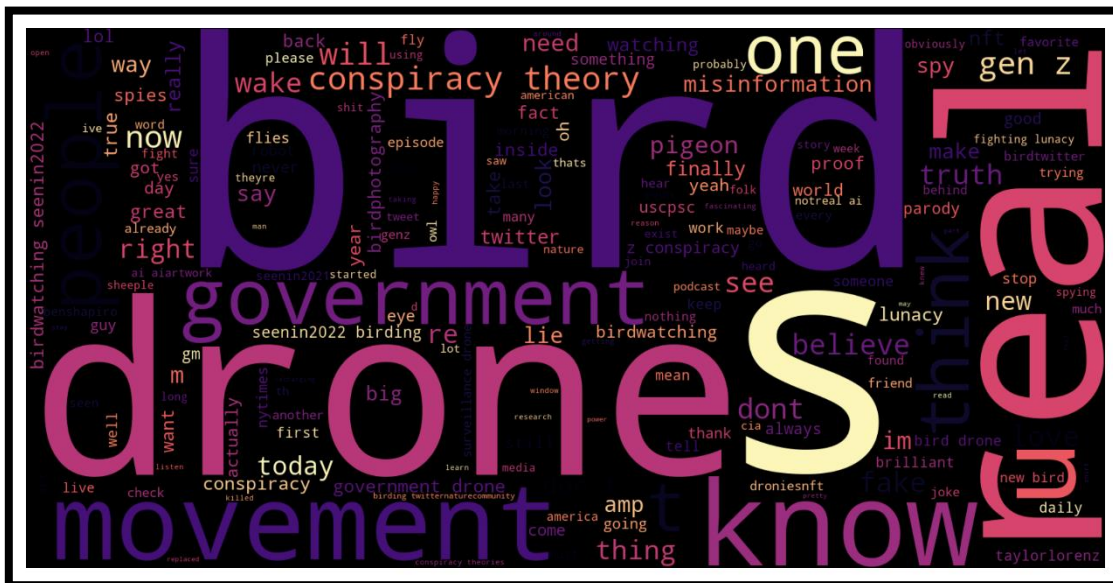
Centralność Wektorów Własnych - Ten wskaźnik określa wpływ węzła, biorąc pod uwagę wpływ jego sąsiadów. *DeNovo_Fatima* ma najwyższą wartość (0.8645391280180094), co sugeruje, że jest ona połączona z innymi wysoko wpływowymi węzłami, co jeszcze bardziej wzmacnia jej pozycję w sieci.

Przy wykorzystaniu biblioteki NetworkX zidentyfikowano 7 różnych społeczności w tej sieci, co wskazuje na istnienie różnorodnych grup użytkowników, które mogą się skupiać wokół wspólnych tematów lub zainteresowań. Każda społeczność może reprezentować grupę użytkowników z silniejszymi wzajemnymi interakcjami niż z użytkownikami poza grupą.

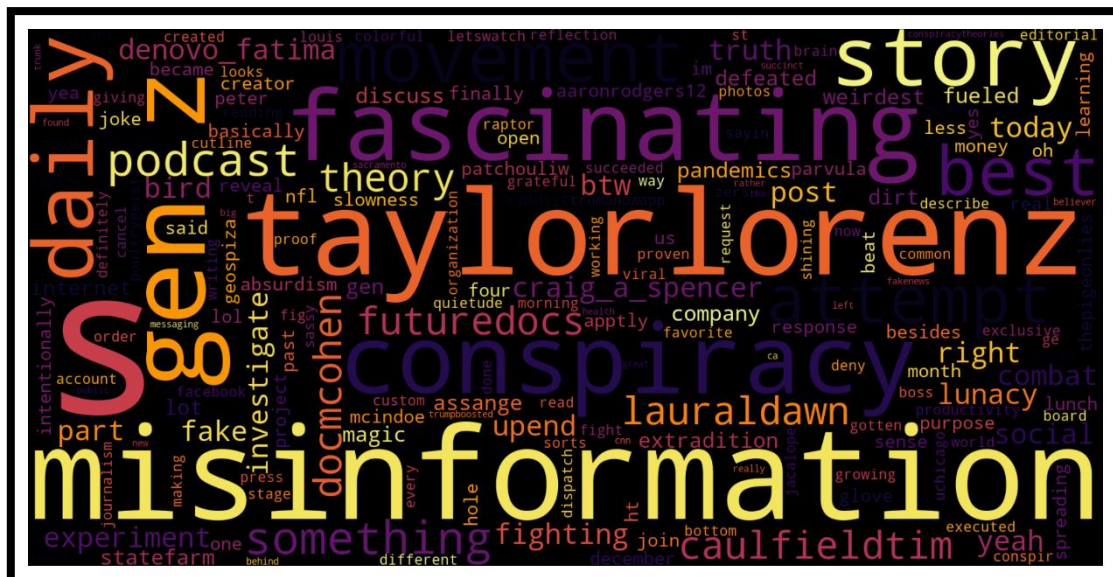


Analiza tekstów w badanej sieci społecznościowej

- treściach tweetów wszystkich użytkowników:



- treściach tweetów zweryfikowanych użytkowników:



- [illegible]

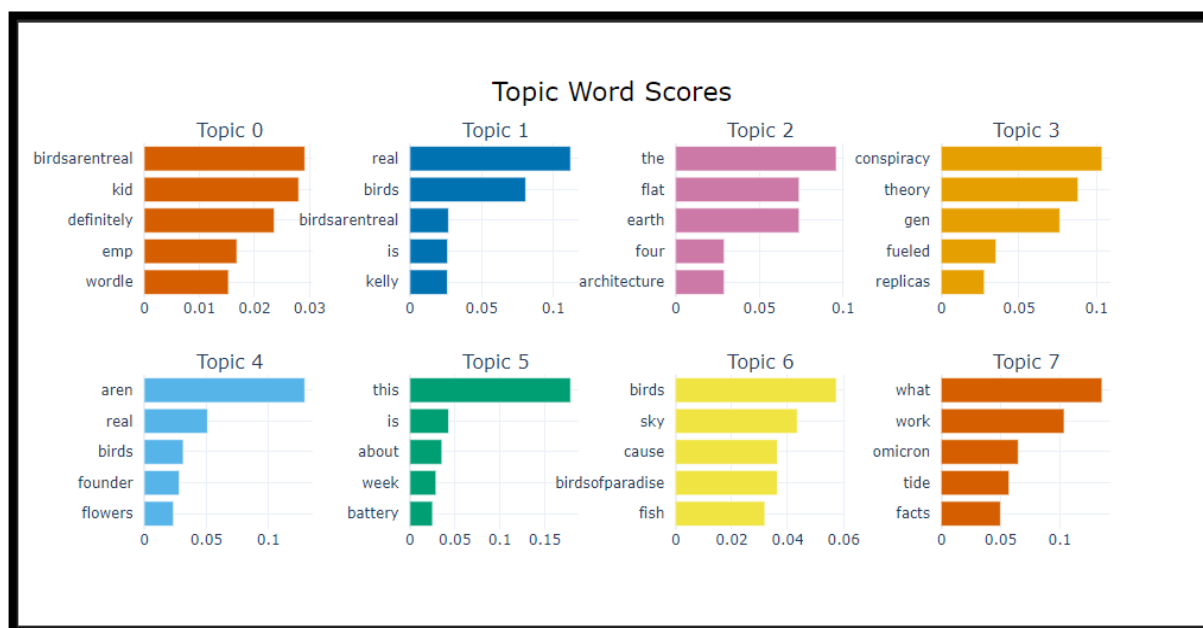
- [illegible]

-

Z powyższych chmur możemy wyciągnąć następujące wnioski:

- 12

Analiza tematyczna przy wykorzystaniu biblioteki BERTopic:



Model BERTopic wskazał na 8 tematów przewodnich, każdy z nich zawiera słowa kluczowe, które charakteryzują dany temat wraz z ich wynikami, pokazującymi, jak silnie dane słowo jest powiązane z tematem.

1. Temat 0 ("birdsarentreal") - Dominujące słowa: "birdsarentreal", "kid", "definitely", "emp", "wordle". Tema zdaje się koncentrować wokół przewodniej teorii spiskowej "birds aren't real" z niewielkim nawiązaniem do gry słownej (Wordle) i być może EMP (impuls elektromagnetyczny), co może wskazywać na zainteresowanie tematem od strony technicznej tego jak działają szpiegujące ptaki-drony.
2. Temat 1 ("real birds") - Dominujące słowa: "real", "birds", "birdsarentreal", "is", "kelly". Słowa te mogą odnosić się do dyskusji o rzeczywistych ptakach w kontraście do ptasiej teorii, z możliwym odniesieniem do konkretnej osoby lub kontekstu (Kelly).
3. Temat 2 ("theoretical concepts") - Dominujące słowa: "the", "flat", "earth", "four", "architecture". Temat ten prawdopodobnie łączy w sobie teorię ptako-dronów oraz znanej i lubianej teorii płaskiej Ziemi z ?elementami architektury?.
4. Temat 3 ("conspiracy theory") - Dominujące słowa: "conspiracy", "theory", "gen", "fueled". Bezpośrednio odnosi się do teorii spiskowych, być może tych zasilanych przez młodsze pokolenia (gen [Z?]).
5. Temat 4 ("birds in nature") - Dominujące słowa: "aren", "real", "birds", "founder", "flowers". Temat może dotyczyć natury i ptaków, ich realnego środowiska, a także kwiatów, co sugeruje być może tematy nawiązujące do naturalnego charakteru ptaków i ich bliskiego związku z naturą.

6. Temat 5 ("technical and temporal") - Dominujące słowa: "this", "is", "about", "week", "battery". Może odnosić się do technologii tego jakie baterie są wykorzystywane do zasilania ptako-dronów.
7. Temat 6 ("birds and environment") - Dominujące słowa: "birds", "sky", "cause", "birds of paradise", "fish". Skupia się na ptakach i ich środowisku naturalnym, z nawiązaniami do innych elementów przyrody, takich jak ryby i ptaki.
8. Temat 7 ("current affairs") - Dominujące słowa: "what", "work", "omicron", "tide", "facts". Może zawierać dyskusje na temat wydarzeń głośnych w sieci, takich jak działanie wariant wirusa SARS-CoV-2 Omicron, co było też dobrą pożywką dla fanów teorii spiskowych i szerszych wpływów tych wydarzeń na społeczeństwo ("tide", "facts").

Analiza sentymentu przy wykorzystaniu biblioteki TextBlob:

- Neutralne: 1000 tweetów
- Pozytywne: 664 tweety
- Negatywne: 237 tweetów

Zdecydowana większość tweetów (1000 z 1901 które zidentyfikował model, co stanowi około 53%) została zaklasyfikowana jako neutralna. To sugeruje, że większość treści nie niesie ze sobą wyraźnych emocji lub jest wystarczająco zrównowazona, by nie przesądzać o jednoznacznie pozytywnym czy negatywnym wydźwięku. Kolejna duża część tweetów (664, około 35%) ma pozytywny sentyment. To wskazuje na ogólnie pozytywną reakcję użytkowników Twittera na tematy związane z ptakosceptycyzmem. Relatywnie niewielka liczba tweetów (237, około 12%) została zaklasyfikowana jako negatywne. To wskazuje na to, że mniej treści wywołuje negatywne emocje takie jak złość, smutek, frustracja, czy krytyka. Mniejsza liczba negatywnych sentymentów w stosunku do pozytywnych może świadczyć o ogólnie pozytywnym lub przynajmniej neutralnym nastawieniu wśród użytkowników w kontekście ruchu #birdsarenoreal.