

Komputacijska etnografija „deepfakea“

Vojnić Kortmiš, Kristijan

Master's thesis / Diplomski rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zadar / Sveučilište u Zadru**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:162:942130>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-01-31**



Sveučilište u Zadru
Universitas Studiorum
Jadertina | 1396 | 2002 |

Repository / Repozitorij:

[University of Zadar Institutional Repository](#)



Sveučilište u Zadru

Odjel za etnologiju i antropologiju

Diplomski sveučilišni studij etnologije i antropologije (dvopredmetni)



Kristijan Vojnić Kortmiš

Komputacijska etnografija „deepfakea“

Diplomski rad

Zadar, 2024.

Sveučilište u Zadru

Odjel za etnologiju i antropologiju

Diplomski sveučilišni studij etnologije i antropologije (dvopredmetni)

Diplomski rad

Student/ica:

Kristijan Vojnić Kortmiš

Mentor/ica:

izv. prof. dr. sc. Senka Božić Vrbancić

Zadar, 2024.



Izjava o akademskoj čestitosti

Ja, **Kristijan Vojnić Kortmiš**, ovime izjavljujem da je moj **diplomski** rad pod naslovom **Komputacijska etnografija „deepfakea“** rezultat mojega vlastitog rada, da se temelji na mojim istraživanjima te da se oslanja na izvore i radove navedene u bilješkama i popisu literature. Ni jedan dio mojega rada nije napisan na nedopušten način, odnosno nije prepisan iz necitiranih radova i ne krši bilo čija autorska prava.

Izjavljujem da ni jedan dio ovoga rada nije iskorišten u kojem drugom radu pri bilo kojoj drugoj visokoškolskoj, znanstvenoj, obrazovnoj ili inoj ustanovi.

Sadržaj mojega rada u potpunosti odgovara sadržaju obranjenoga i nakon obrane uređenoga rada.

Zadar, 2. srpanj 2024.

SAŽETAK

Komputacijska etnografija „deepfakea“

U ovom se radu fokusiram na mrežne reprezentacije pisanog teksta kao otvorene sustave značenja na digitalnoj društvenoj platformi *Twitter*. Mrežne reprezentacije teksta u metodološkom smislu pružaju alternativu linearnom iščitavanju teksta. Drugim riječima, društvenim mrežama pristupam kao specifičnu terenu koji proizvodi određena značenja i prakse inherentne digitalnom formatu. Za analizu sam isključivo koristio izjave s društvene mreže *Twitter*, a posebnu pozornost posvećujem *hashtagovima* i njihovoj supojavnosti. Cilj je ovoga rada istražiti na koji se način tema *deepfakea* predstavlja i percipira na društvenoj mreži *Twitter*. Istraživanjem diskursa društvenih mreža o pojmu *deepfake* tehnologije možemo bolje razumjeti šire implikacije digitalnih tehnologija na društvo i kulturu, otvarajući put za bolje razumijevanje isprepletenosti percipiranih prijetnji.

Ključne riječi: grafovi mreža, *hashtag*, društvene mreže, analiza sentimenta, deepfake, digitalna antropologija, digitalna etnografija, komputacijska etnografija, strojna obrada jezika, *Twitter*, vizualizacija

ABSTRACT

Computational ethnography of „deepfake“

Algorithms and new digital technologies on one hand actively shape human everyday experience, while simultaneously providing the foundation for the use and development of new methodologies in researching human everyday life. My primary interest in this work is to explore how discussions about deepfakes unfold on the social network Twitter. Inspired by the research opportunities emerging in the field of digital anthropology, as well as the possibilities of computational ethnography, discourse theories, and network graph theories, methodologically, I aim to develop an approach that utilizes various natural language processing techniques. Natural language processing is employed for both quantitative and qualitative research, and I am particularly interested in how it can contribute to qualitative research, especially discourse analysis.

Key words: network graphs, hashtag, social networks, sentiment analysis, deepfake, digital anthropology, digital ethnography, computational ethnography, natural language processing, Twitter, visualization

SADRŽAJ

1. UVOD.....	1-3
2. Digitalna antropologija, digitalna etnografija i teorija grafova.....	3-9
2.2. Mreža supojavnosti hashtagova	9-11
2.3. Digitalne platforme i algoritmi	11-13
3. <i>Deepfake</i>	13-18
3.1. Analiza sentimenta	19-20
3.2. Prikupljanje građe za analizu i strojna obrada jezika	20-24
4. Rezultati	24-43
5. Zaključak	43-45
6. Popis literature	46-53
7. Prilozi	52

1. UVOD

Od 2017. godine, od kada se prvi put pojavio, *deepfake* je obilježio internetsku kulturu. Termin *deepfake* (Doss i dr., 2023: 1) potječe od složenice *deep learning* i *fake* te se u užem smislu odnosi na sintetički medij u kojem korisnik poput lutkara upravlja videosadržajem druge osobe. Obradeni sadržaj ostavlja dojam da su izjave korisnika stvarne iako su odraz digitalne manipulacije. Ovakva tehnologija u širem smislu omogućuje generiranje nepostojećih krajobraza, kao i lica. Prema podacima tvrtke za kibernetičku sigurnost *Deeptrace*, 2019. godine je na internetu pronađeno 14 698 *deepfake* videa¹. U usporedbi s 2018., kada je pronađeno a 7 964 videa, uočljiv je rast od 84%². U kolikoj je mjeri uporaba *deepfakea* u porastu govori i podatak da je u 2023. detektirano 95 820 videa³. Ovaj rast naglašava sve veću sofisticiranost i dostupnost *deepfake* tehnologije, što s jedne strane stvara strah zbog mogućnosti krađe identiteta i širenja dezinformacija, a s druge strane *deepfake* tehnologija omogućava i kreativnu primjenu, pogotovo u području popularne kulture (film, reklame, industrija kompjuterskih igri itd.) i edukacije (virtualni predavači, povijesne rekonstrukcije,...). Negativne konzekvence uporabe *deepfake* tehnologija uglavnom se odnose na lažne vijesti i političke manipulacije, kompromitirajuće pornografske videozapise i identitetske prevare. U svakom slučaju, *deepfake* tehnologija na nov način otvara pitanja o tome što je to „stvarnost”, a što „laž”, i kao što to tvrde Paris i Donovan (Paris i Donovan 2017: 2) - kada govorimo o *deepfaku* kao lažnom, često zanemarujemo činjenicu da niti jedan audiovizualni sadržaj, kao niti bilo koja druga prezentacija, nikad i nije bio stabilan. U tom smislu ono što je ključno za razumijevanje *deepfaka* nije njegova „lažnost” ili „istinitost”, već kontekst u kojem nastaje, kao i značenja koja generira.

Cilj je ovoga rada istražiti na koji se način tema *deepfakea* predstavlja i percipira na društvenoj mreži *Twitter*. Društvene mreže, kao što je to *Twitter*, imaju veliku ulogu u širenju informacija, a *Twitter* kao jedna od najutjecajnijih platformi omogućava korisnicima diljem svijeta slanje i čitanje kratkih poruka koje su ograničene na samo 280 znakova. Upravo ta ograničenost potiče brzu razmjenu informacija, kao i sažeto izražavanje. Kao i na drugim društvenim mrežama, na *Twitteru*

¹ „The staggering rise of deepfake videos in 2019“, CNN Business, 7.10.2019., pristupljeno 1.7.2024. <https://edition.cnn.com/2019/10/07/tech/deepfake-videos-increase/index.html>.

² „The staggering rise of deepfake videos in 2019“, CNN Business, 7.10.2019., pristupljeno 1.7.2024. <https://edition.cnn.com/2019/10/07/tech/deepfake-videos-increase/index.html>.

³ „24 Deepfake Statistics - Current Trends, Growth, and Popularity“, prosinac 2023., Content Detector AI, pristupljeno 1.7.2024., <https://contentdetector.ai/articles/deepfake-statistics>.

korisnici mogu pratiti jedni druge, a *Twitter* koristi i oznake (*hashtag* - #) kako bi se poruke koje imaju sličnu temu grupirale, što olakšava praćenje specifičnih tema, a ujedno ukazuje i na određene trendove u komunikaciji, tj. u objavama (usp. Cristiano 2024: 2). Ono što je značajno za *hashtag* oznaku je njegova funkcija mrežnog uokviravanja (usp. Meraz 2017:143-145), stoga mu u radu pristupam kao spremniku korisničkih izjava kako bih analizirao formiranje diskursa u vezi ključne riječi *deepfake*. S obzirom na to da je u trenutku kada je započet proces prikupljanja podataka za ovaj rad (26. ožujka 2023.- 2. travnja 2023.) *Twitter* pružao besplatnu opciju prikupljanja podataka za studente i istraživače, upravo to je postalo glavnim razlogom da se *Twitter* iskoristiti kao izvor građe za analizu o tome kako se u diskursu društvenih mreža progovara o *deepfakeu*. Metodološki gledano, u radu kombiniram komputacijsku etnografiju i grafove mreža s kvalitativnom analizom sentimenta kako bih objasnio razne nijanse u izražavanju emocija kad se govori o *deepfakeu*, a primarno me zanima razumijevanje šireg konteksta na koji se analizirani *tweetovi* odnose.

Teorijski gledano, inspiriran sam kulturnim obratom u antropologiji u čijoj je podlozi poveznica između jezika i značenja te ću se u analizi fokusirati na pitanja kako jezik konstruira značenje, kakva je njegova povezanost s kulturom te kako jezik omogućava kolektivno razumijevanje odnosno konstruiranje kulture u kojoj svi članovi interpretiraju svijet na sličan način. *Deepfake* promatram kao tehnološku inovaciju koja nije odvojena od društveno-kulturnog konteksta unutar kojeg je nastala; upravo suprotno - ta je inovacija producirana društveno-kulturnim kontekstom, ali ga istovremeno i oblikuje. U kategorizaciju sadržaja na lažne ili stvarne upisano je nešto inherentno ljudski, a to je sintaksa jezika zasnovanog na binarnim opozicijama koja nužno vodi do hegemonijskih „igara” raznih diskursa. Pod diskursom ovdje mislim na foucaultovsku definiciju diskursa kao određene forme znanja koja omogućava način na koji razumijemo i govorimo o svijetu oko nas: „diskurs referira na grupu izjava koje strukturiraju način na koji je nešto mišljeno, kao i način na koji djelujemo na temelju određenog mišljenja ... to je znanje o svijetu koje oblikuje način na koji razumijemo svijet i ono što se u njemu događa” (Rose, 2001: 136). Prema Rose, jedna od glavnih tema diskurzivne analize je analiza organizacije samog diskursa, stoga me u ovom radu zanima kako je diskurs *deepfakea* strukturiran? Kako *deepfake* (re)producira određeno znanje? Kako je *deepfakeom* društveno-kulturna različitost konstruirana, tj. reprezentirana? Za Stuart Halla (1997:173-202) koncept reprezentacije odnosi se na korištenje jezika (i slika) radi kreiranja značenja. Foucaultovski gledano, reprezentacija je ujedno i koncept i praksa. Hall (ibid.) naglašava da reprezentacija nije isto što i prezentacija. Prezentacija kao koncept

uključuje vjerovanje u realizam, u istinitost i preslik onog što je predstavljeno. Za Halla (ibid.) realizam je profesionalna ideologija koja počiva na uvjerenju da slike ne lažu i da postoji jedan pravi istiniti događaj koji se onda predstavlja na istinit ili lažan, tj. manipulativan način. Kao i ostali konstruktivisti, Hall (ibid.) polazi od uvjerenja da stvari po sebi nemaju značenja; značenja konstruiraju korisnici služeći se reprezentacijskim sustavima. Značenje se ne izvodi iz gologa postojanja materijalnog svijeta, već iz simboličkih praksi (kao npr. jezika) koje pridružujemo svijetu kako bismo ga osmislili, (pre)uredili, odnosno – konstruirali. No pitanje koje se postavlja je kako pristupiti analizi suvremene svakodnevice koja je obilježena *deepfake* sadržajima? Kako analizirati značenja koja proizlaze iz *deepfakea*? U ovom je radu ishodište analize mrežna reprezentacija diskursa društvenih medija na temu *deepfakea*, no u tijeku analize mreža se obogaćuje poveznicama s drugim diskursima poput, na primjer, medijskog. Rose (2001:136-137) naglašava značaj intertekstualnosti na način da se pri analizi ne propituje samo značenje određenog teksta već i druga značenja drugih tekstova u vezi njega. U tom smislu intertekstualnost u mom pristupu odnosi se na automatski generirane mreže *supojavnosti* na društvenoj mreži *Twitter*, ali i poveznice stvorene iščitavanjem tekstova izvan okvira te društvene mreže.

Moj rad spada u područje digitalne antropologije. Za Daniel Millera (2018:2) digitalna antropologija podjednako uključuje proučavanje upotrebe novih tehnologija kao i učinke njihove upotrebe. Digitalna antropologija, navodi Miller (ibid.), može se odnositi na posljedice uspona digitalnih tehnologija za određena društva, korištenje ovih tehnologija u kontekstu antropološke metodologije ili proučavanje specifičnih digitalnih tehnologija. Međutim, ova tema svakako može postaviti i šira pitanja o mogućnostima suvremene antropologije kad su u pitanju digitalne tehnologije.

2. Digitalna antropologija, digitalna etnografija i teorija grafova

Za Millera i Horst (2012., prema ibid.) „digitalno“ označava svaki ishod binarnog kodiranja iako postoje i alternative, poput „fokusa na uspon kibernetičkih sustava ili virtualnih svjetova“ (Boellstorff i dr., 2012., prema ibid.). Miller (ibid.) se opredjeljuje za binarnu definiciju digitalnog jer u njoj prepoznaje bitne implikacije tehnologije, poput mogućnosti identične reprodukcije i dijeljenja. Upravo ove dvije stavke omogućuju rapidno širenje novih digitalnih tehnologija, što uzrokuje promjenu fokusa javnosti i akademske zajednice na teme poput interneta, virtualnog, društvenih medija, umjetne inteligencije, *big data* i drugo, te se taj fokus iz godine u godinu

učestalo mijenja. Uzimajući u obzir da se ove tehnologije javljaju u kontekstu upotrebe i učinaka upotrebe za određenu populaciju, one podliježu procesima kulturne diferencijacije pa se tako kineske društvene mreže razlikuju od, primjerice, brazilskih ili američkih. Prema Milleru (ibid.) dok u jednoj regiji pronalazimo internet u učestaloj debati koja propitkuje koje su digitalne forme kompatibilne s islamom, u drugoj pak regiji možemo promotriti kako je internet omogućio uspon feminističkih agenda (npr. Pokret #metoo⁴). Pojam *digitalno* u kontekstu antropologije otvara nove svjetove i populacije. Tako je, na primjer, Gabriela Coleman (2012.,2014., prema ibid.:4) istraživala zajednice hakera i grupe poput hakerske grupe *Anonymus*, a Jenna Burell (2012., prema ibid.) istraživala je zapadnoafričke internetske prevarante, čime je nastojala pružiti dublji uvid u prekarnu poziciju samih prevaranata. Digitalno se u kontekstu antropologije također može tumačiti kroz prizmu neoliberalizma, pa su tako Nardi i Kow (2010. prema ibid.) istraživali praksu *gold farming*, gdje su kineski radnici intenzivno igrali kompjutorske igre kako bi stekli blaga koja su zatim prodavali igračima iz drugih dijelova svijeta.

Istraživanja digitalnog imaju i performativnu dimenziju, a tu Miller navodi primjer Ilane Gershon (2017., prema ibid.) koja je istraživala mrežu *LinkedIn* kao digitalni prostor u kojemu su radnici primorani izvoditi različite prakse, isključivo zbog reprezentacije, kako bi dobili posao. Fokusiranje na pojedinačne i specifične platforme predstavlja još jedan od načina kojim su antropolozi pristupili ovim novim svjetovima. Jaime Coates je tako pružio perspektivu fenomena slika koje postaju viralne, „kao što je slučaj s rapidnim širenjem memeova“ (2017., prema ibid.:5). Zajedničko svim ovim navedenim primjerima jest značaj etnografije kao antropološke metode. Miller (ibid.) ističe u kontekstu digitalne antropologije značaj etnografskog pristupa u razumijevanju kako ljudi percipiraju svoju povezanost s digitalnom tehnologijom u svakodnevnom životu, koji nije nužno digitalan. Miller smatra da etnografija pruža intimniji uvid u ono što ljudi stvarno vjeruju jer ostale discipline u istraživanju digitalnih medija ovise o javno dostupnim informacijama poput *Twittera*, što može rezultirati političkim *echo chamberima* „gdje čujemo samo mišljenja slična našima“ (ibid.). Etnografija, s druge strane, u kontekstu digitalnog može imati pristup privatnim medijima poput *WhatsAppa*. Raniji radovi Wilsona i Petersona (2002. prema ibid.:9) su se u kontekstu etnografskog pristupa digitalnom fokusirali na istraživanje *online*

⁴ Me too (ili „meetoo“) je izraz koji se koristi kao hashtag (#MeToo) na društvenim medijima kako bi se podigla svijest o seksualnom nasilju i uznemiravanju, potičući žrtve da podijele svoja iskustva i podrže jedni druge. ChatGPT, „Definicija #meetoo,“ generirano s OpenAI, 2024.

zajednica, dok je kasnije, s Colemanom (2010. prema *ibid.*), fokus istraživanja prešao na etnografski pristup digitalnim medijima. Horst i Miller „istražili su raznovrsna područja studija, od lokacije do politike i svakodnevnog života, kao i implikacije za teoriju i antropologiju općenito. (2012., prema *ibid.*)”

Pomoću novih mehanizama bilježenja i analize informacija digitalnu je etnografiju prema Milleru potrebno promatrati istovjetno uz razvoj etnografije digitalnog. Miller kao jedan od ključnih autoriteta u polju digitalne etnografije navodi antropologinju Sarah Pink. Pink etnografiju ne smatra kao svrhu samoj sebi, već „ona postaje korisna isključivo kada se primjenjuje u okvirima određene paradigme disciplinarnog ili interdisciplinarnog istraživanja i u relaciji naspram praksi i ideja u istraživačkom procesu“ (2016:2). Jednom kada etnografija postane digitalnom, ona je uvjetovana našim shvaćanjem načina na koji su digitalni mediji postali djelom svakodnevice (*usp. ibid.*). Drugim riječima, kao što Pink ističe pozivanjem na O' Reilleya, ona dovodi u pitanje definiciju „etnografije kao direktnog i održavanog kontakta s ljudskim dionicima u kontekstu njihove svakodnevice i kulture.“ (O' Reilly, 2005. prema *ibid.*:3) Većina je etnografskih aktivnosti primjenjiva u kontekstu digitalne etnografije, no etnografska se praksa mijenja. Etnografsko pisanje može biti zamijenjeno videosadržajima, fotografijom ili *blogiranjem*, stoga, kao što i autorica ističe - digitalna etnografija može uključivati različite poglede na svijet, ovisno o istraživanju. Pink (*ibid.*: 8-14) navodi da postoji pet ključnih principa digitalne etnografije:

1. Višestrukost

Istraživanje je jedinstveno na način da ovisi o istraživačkom pitanju. Potrebno je uzeti u obzir kako su digitalne tehnologije i mediji usko povezani s infrastrukturom svakodnevice i kako pristup određenim medijima nije univerzalan.

2. Digitalno nije centralno za digitalnu etnografiju

Fokus istraživanja zahtjeva određen odmak od samog medija kako bi otvorio prostor za tumačenja kojima on postaje nerazdvojan od drugih aktivnosti, tehnologija, materijalnosti i osjećaja. Pink sugerira kako je indirektan pristup često plodonosniji. Primjerice, u njezinu istraživanju o energetske zahtjevima za korištenje digitalnih medija, Pink je sa sudionicima istraživala njihove svakodnevne rutine koje su implicirale upotrebu energije i digitalnih medija.

3. **Otvorenost**

Pod otvorenošću Pink podrazumijeva koncept koji se često naziva digitalnom kulturom koja je zasnovana na ideji otvorenog koda i kolaboracije, što znači da istraživač stvara znanje i načine spoznavanja s drugima, umjesto u izolaciji.

4. **Refleksivnost**

Digitalna etnografija podrazumijeva refleksivnost koja velikim dijelom potječe iz kulturnog obrata koji su predvodili Marcus i Clifford 1986. godine. Refleksivnost podrazumijeva da u obzir uzimamo način na koji mi kao istraživači proizvodimo znanje susretima s drugim ljudima i stvarima. Refleksivnost također podrazumijeva etičke prakse priznavanja kolaboracije.

5. **Neortodoksnost**

Digitalno se priznaje kao dio svakodnevice ili samo predmet istraživanja. U kontekstu digitalne antropologije, digitalna je etnografija uobičajena praksa koja uključuje studije slučaja, intervju, ankete, promatranje sa sudjelovanjem i participaciju u određenom vremenskom razdoblju, bilo *online* ili u fizičkom svijetu.

Razvojem društvenih mreža i digitalno posredovane komunikacije istraživači su dobili pristup ogromnim korpusima tekstova, što je dalo zamah i istraživanjima utemeljenim na kvalitativnim metodama tekstualne analize, koje metodološki nužno ne uključuju promatranje sa sudjelovanjem o kojem govori Pink kada uvodi principe digitalne etnografije. Digitalni format teksta omogućuje izvršavanje programskih upita nad građom. Geiger⁵ (2017.) govori o komputacijskoj etnografiji kao novom području koje podrazumijeva širenje dijapazona alata etnografske kvalitativne metodologije na komputacijske alate. Tako, na primjer, pri istraživanju društvenih mreža vrijedi spomenuti Marwick i Boyda (2011:1) koji su istraživali komentare na *Twitter* platformi, one koji su bili vezani za popularnu kulturu, tj. poznate osobe (*celebrity*) Mariah Carey, Miley Cyrus i Paris Hilton, kako bi pokušali elaborirati aspekte odnosa moći inherentne interakciji između obožavatelja. Nadalje, Bruns i Burgess (2011:1) su upotrebom *inovativnih* metodologija (računalno generiranje grafova mreža) istraživali komentare na *Twitteru* fokusirajući se na upotrebu *hashtagova* u političkim debatama te načinima kako isti doprinose formiranju *ad hoc* javnosti u vezi specifičnih tema. Iako se u primjeru Brunsova i Burgessova (ibid.) istraživanja

⁵ Geiger, R.S., „Computational Ethnography and the Ethnography of Computation“, 14.9.2017., pristupljeno 13.12.2023. <https://newweb.stuartgeiger.com/talks/2017-09-14-computational-ethnography/>

korištenju grafova mreža pripisuje epitet „inovativnosti“, upotreba grafova mreža u antropologiji nije novost. Ono što je novo su alati koji omogućuju njihovo generiranje, kao i načini interpretacije.

Teorija grafova se u antropologiji koristila i 60-ih godina prošlog stoljeća kada je Per Hage u suradnji s matematičarom Frankom Hararyjem napisao niz članaka u kojima su osmisli jedanaest teorema i šest algoritama koji su im pomagali pri analizi društva. U članku „Graph Theory as a Structural Model in Cultural Anthropology“ Hage i Harary (1979: 130) tvrde da mrežni grafovi predstavljaju strukture unutar antropoloških interesa koji se sastoje od odnosa koji su nebinarni, probabilistički ili višestruki. Jenkins (2008: 2-7) navodi kako su Hage i Harary upotrebom teorije grafova istraživali su mnoštvo tema, poput Puluwateske navigacijske tehnike i mnemonike, besciljnih traganja Polinezije, Orkavske razmjene darova, Majanske ceremonijalne arhitekture, melanazijske društvene strukture, srodstva, saveza i statusnih struktura različitih mjesta Nove Gvineje, Tikopije, Tonge, Truka i Solomonskog otočja te zapadne Karoline, Indije i Meksika. Hage i Harary (1984:9) navode četiri prednosti korištenja teorije grafova:

1. Grafovi mreža su ikonički, oni izgledaju kao ono što i prezentiraju, pa su stoga i „povezani sa stvarnošću“ (Berge, 1962 prema ibid.), za razliku od algebarskih modela.
2. Teorija grafova nudi precizne i bogate definicije koncepata poput povezanosti, transformacije, dualnosti i centralnosti, a sama se analiza može proširiti pojašnjenjem česte antropološke metafore.
3. Teorija grafova također sadrži i tehnike kvantitativnih aspekata kulture, poput dostupnosti u komunikacijskoj mreži, zatim udaljenosti u mrežnoj razmjeni te tranzivnosti u statusnim sistemima.
4. Također, grafovi sadrže i teoreme koji omogućuju izvlačenje zaključaka o svojstvima strukture iz znanja o njenim svojstvima.

Prema Hageu se stoga može reći kako upotreba teorije grafova i algoritama antropolozima omogućava otkrivanje uzoraka društvene interakcije mapiranjem odnosa između pojedinaca i grupa. Ono što je problematično u njihovu pristupu je univerzalizacija grafova. Inspirirani strukturalizmom koji kao teorijska škola teži analizi struktura koje djeluju nesvjesno (strukture jezika, psihe, društva) te na razne načine proizvode značenja, daju prednost strukturi, što onemogućuje analizu promjena strukture. Kao što postukturalisti naglašavaju, nemoguće je opisati potpun ili neproturječan označiteljski sustav budući da su sustavi uvijek promjenjivi.

Harcourt (2007:17) ističe kako poststrukturalističke teorije priznaju strukturalističke principe prema kojima su značenja izvedena iz relacija različitosti; da su ona velikim dijelom nesvjesna i da formiraju strukturu, ali se odbacuje potraga za univerzalijama kao ontološkim istinama o ljudima i društvu. Umjesto potrage za generaliziranim zakonitostima, naglasak se stavlja na razdore i nejasnoće u strukturama značenja.

Prema Harcourtu (ibid.), u poststrukturalističkom pristupu strukturalistička je paradigma okrenuta naglavačke. Cilj nije pronaći pravilnosti i uzorke, već propitati što to znači da nalazimo uzorke i zatvorene sisteme značenja. Djelomično inspiriran Hageovom i Hararayevom teorijom grafova, ali uzimajući u obzir i poststrukturalističku kritiku iste, u ovom se radu fokusiram na mrežne reprezentacije pisanog teksta kao otvorene sustave značenja na digitalnoj društvenoj platformi *Twitter*. Mrežne reprezentacije teksta u metodološkom smislu pružaju alternativu linearnom iščitavanju teksta. Drugim riječima, društvenim mrežama pristupam kao specifičnu terenu koji proizvodi određena značenja i prakse inherentne digitalnom formatu. Za analizu sam isključivo koristio izjave s društvene mreže *Twitter*, a posebnu pozornost posvećujem *hashtagovima* i njihovoj supojavnosti.

2.1. *Hashtag* konstruktivizam

Hashtag je riječ ili fraza kojoj prethodi znak # i koja se koristi na raznim internetskim platformama, uključujući *Twitter*, kako bi se olakšala pretraga i grupirale slične teme. Ichau, Firssen i Haensen u svom istraživanju pružaju sažet pregled pristupa *hashtagovima* kao mrežnom uokviravanju:

„Korisnici Twittera (Wang et al., 2011.) (...) označavaju objave ključnim riječima kako bi kategorizirali, kontekstualizirali ili istaknuli teme, koristeći znak '#' kao prefiks odabranoj ključnoj riječi ili frazi (npr. *#photooftheday*). One se organski pretvaraju u hiperlinkove koji preusmjeravaju na sav drugi sadržaj na platformi označen istim *hashtagom*. Budući da *hashtagovi* omogućuju korisnicima *samokuriranje* tematskog sadržaja (Meraz, 2017.), praksa *hashtagiranja* proizvodi dinamičnu, korisnički vođenu taksonomiju *online* sadržaja (Highfield i Leaver, 2015.). U ovom procesu ad hoc javnosti (Bruns i Burgess, 2015.) pozivaju se u postojanje kada se *hashtagovi* (npr. *#ows* [Occupy Wall Street], *#egypt*) koriste za artikuliranje individualnih i društvenih identiteta te sudjelovanje u društvenim pokretima (Meraz, 2017.; Meraz i Papacharissi, 2013.). (...)

Sve je više literature u kojoj se proučavaju narativna svojstva *hashtagova*. Meraz (2017.) te Meraz i Papacharissi (2013.) konceptualiziraju *hashtagove* kao okvire. Uokviravanje je, kako je definirao Entman (2007.), proces diskurzivne konstrukcije u kojem se odabiru određeni elementi problema i predstavljaju na način koji sugerira koherentnost i predlaže tumačenje. Mrežno uokviravanje događa se kada 'određene definicije problema, uzročne interpretacije, moralne procjene i/ili preporuke za postupanje steknu značajnost putem praksi *crowdsourcinga*' (Meraz i Papacharissi, 2013.: 159). To je organski i iterativni proces u kojem problemi dobivaju značaj i značenje kolektivnom upotrebom *hashtagova* (Meraz, 2017.; Meraz i Papacharissi, 2013.). Drugi autori proučavali su oznake kao sredstva za komuniciranje afekta (Wang et al., 2011.) i kao metakomunikativne uređaje. (...) Zappavigna (2015.) razlikuje tri glavne funkcije *hashtaga*: tumačenje iskustva (kao što je označavanje sadržaja), izvođenje odnosa (kao što je označavanje stajališta) i označavanje metadiskursa. (...) studije ukazuju da su *hashtagovi* potencijalno moćni agenti za kolektivnu proizvodnju značenja.“ (Ichau, Firssen i Haensen, 2019: 6-7)

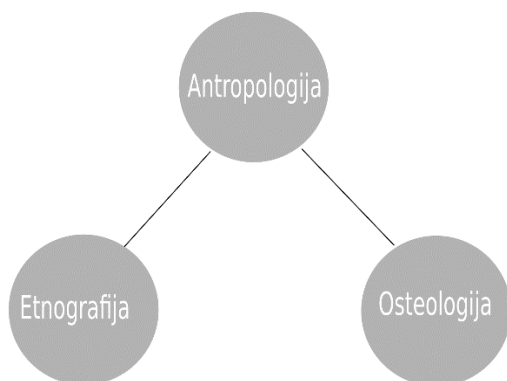
Ili, drugim riječima, „*hashtagovi* aktivno doprinose oblikovanju društvene konstrukcije koncepata na platformi.“ (Loukianov, Burningham i Jackson, 2022: 65).

2.2. Mreža supojavnosti *hashtagova*

Uzimajući u obzir gore spomenutu *hashtag* funkciju mrežnog uokviravanja, pojedinačnom *hashtagu* pristupam kao spremniku korisničkih izjava. Pritom je kolokacija ili supojavnost više *hashtagova* u pojedinačnoj izjavi korisnika veza između tih spremnika definirana od strane korisnika. Vizualizacija ovih veza pruža uvid u *intertekstualnost* i formiranje diskursa u vezi ključne riječi *deepfake* na društvenoj mreži *Twitter*. Nastojat ću ovo ilustrirati jednostavnim primjerom nevezanim za moju temu. U slučaju da imamo dvije izjave koje sadrže sljedeće *hashtagove*:

1. #antropologija #osteologija
2. #antropologija #etnografija

mreža supojavnosti *hashtagova* bila bi vizualizirana kao u primjeru donjeg prikaza (Prikaz 1.).



Prikaz 1. mreža sa tri čvora

Pritom je vidljivo kako je antropologija na temelju supojavnosti povezana s osteologijom i etnografijom, ali osteologija i etnografija nisu povezane. Antropologija se javlja dva puta, ali je predstavljena kroz jedan čvor jer svaki je čvor jedinstvena pojava najmanje jedinice građe, dok veze mogu biti višestruke. Ponavljanje čvora u građi može biti pripisano atributu čvora, a ponavljanje veza atributu veze. Atributu čvora također može biti pripisan i rezultat procjene sentimenta izjave. Ovo je korisno iz razloga koji će biti pojašnjeni u poglavlju koje se bavi metodologijom. Kada se radi sa većim korpusima teksta, ovakve mreže mogu rasti u kompleksnosti; pri njihovu iščitavanju i filtriranju korisne su metode grafova mreža. Čvorovi mogu imati stupanj centralnosti koji ovisi o broju veza koje posjeduje ili mjeru „spajanja“ pri čemu se lociraju čvorovi koji imaju minimalan broj veza, ali su povezani sa čvorovima koji imaju velik broj veza. Ovakvo bilježenje akumulira veze u pojedinačnim korisničkim komentarima te vizualizira cjelokupan korpus odjednom. Tako djeluje kao prikaz kolektivnog supstrata korisnički stvorenih veza. Pritom se u obzir uzimaju sve izjave neovisno o tome jesu li korisnici bili u interakciji. Značaj koji pridajem ovakvoj vizualizaciji zasigurno jednim dijelom proizlazi i iz nesvjesne internalizacije episteme digitalizacije društva. Ipak, najvećim dijelom moj naglasak na mrežnom iščitavanju teksta s društvenih medija proizlazi iz nastojanja da se pruži otpor praksama algoritmične regulacije društva. Specifična vidljivost mrežnog iščitavanja teksta barem dijelom izbjegava nevidljivosti linearnog iščitavanja teksta koje najčešće leži u rukama vlasnika platforme čije je kuriranje sadržaja obilježeno željom za ostvarivanjem profita reklamiranjem. To ne znači da sadržaj korpusa ne sadrži izjave s finansijskim interesima, ali takav utjecaj operira na mikro, a ne na makro razini. U nastavku ću se pozvati na istraživanja koja se bave pitanjem kako algoritmi

društvenih platformi uvjetuju pristup sadržaju, kako bih istaknuo čemu pružam otpor svojim pristupom.

2.3. Digitalne platforme i algoritmi

Digitalne društvene platforme poput *Facebooka* ili *Twittera* oslanjaju se na algoritme koji određene sadržaje čine vidljivima, tj. vidljivijima, ili nevidljivima, tj. manje vidljivima, i tako oblikuju „načine na koje doživljavamo svijet, kako ga činimo smislenim te kako na njega djelujemo“ (Deuze, 2012: 5). Prema Mackenzie algoritmi, imaju selektivnu funkciju „poticanja određenog poretka nauštrb drugog“ (2006: 44). Bucher (2018: 69), inspirirana foucaultovskim konceptom mikrofizike moći, propituje kako i kada moć djeluje kroz algoritme. Pozivajući se na Foucaultovu arhitekturu dijagrama moći iz knjige „Nadzor i kazna“, Bucher uspostavlja analitički okvir kroz koji se obavijesti na *Facebooku* mogu shvatiti kao primjer društva programiranog putem algoritmičnog medija. Bucherin (Lash, 2007; Thrift, 2005 prema ibid.:72) interes za pitanje moći proizlazi iz načina na koji se moć materijalizira kroz softver i kod, postajući dijelom nesvjesnog tehnološkog aspekta svakodnevnog života. Za Bucher (ibid.:73) algoritmi imaju produktivne kapacitete stvaranja razlika kako su društvene formacije oblikovane i informirane. U poglavlju posvećenom algoritmu *Facebook* obavijesti, Bucher (ibid.:77) se poziva na objašnjenje Arija Steinberga iz 2010. godine⁶. Naime, prema Steinbergu, *Facebook* algoritam traži veze u grafovima kako bi rangirao obavijesti. Svaka stavka na *Facebooku*: poruka, status, slika ili video, u samom se grafu koncipira kao čvor. Svaka interakcija s čvorom, primjerice kroz *like* ili komentar, veza je između dvaju čvorova. Veza u ovom slučaju može imati tri atributa: vrijeme kada se veza dogodila, tip veze (komentar ili *like*) te korisnika koji je kreirao vezu. Dio formule koji računa težinu veze pripisuje težinu svakom tipu veze. Primjerice, komentar može biti teži od *likea* itd. Kasnijim razvojem ovakva jednostavna distinkcija na tri tipa veza nije bila moguća. Prosječno, svaki korisnik ima oko 1500 *storyja* (priča) koje može vidjeti, a sam algoritam je taj koji odlučuje koje se priče (*stroyji*) najrelevantnije, uzimajući pritom u obzir odnose prijateljstva, frekvenciju interakcije, broja *likeova* i dijeljenja objava. „*Facebook* je sposoban pripisati svakoj *Facebook* objavi tzv. ocjenu značajnosti, koja je specifična za određenog korisnika“ (Oremus⁷, 2016, prema

⁶ „F8 Techniques,“ Livestream, 2010. pristupljeno 2.7.2024, http://www.livestream.com/f8techniques/video?clipId=pla_5219ce25-53c6-402d-8eff-f3f8f7a5b510.

⁷ Oremus, W. „Who controls your Facebook feed.“ Slate, pristupljeno 2.7.2024. http://www.slate.com/articles/technology/cover_story/2016/01/how_facebook_s_news_feed_algorithm_works.html

ibid.:78). Pomoću obrnutog inženjeringa vlastitog *news feeda* (vijesti u vremenskoj crti), Bucher (ibid.:89) je otkrila da algoritmi *Facebooka* ne distribuiraju vidljivosti jednako svim korisnicima, već su radije ažurirali *feed* sadržajem koji bi potencijalni angažman i interakciju korisnika učinio vidljivijim. Nadovezujući se na Foucaultov koncept panoptizma i preokrećući ga, Bucher (ibid.: 89-91) tvrdi da je tzv. *news feed* utjelovio *prijetnju nevidljivosti* u discipliniranju korisnika kako bi kao reakciju izazvao *participativnu subjektivnost*. Discipliniranje korisnika *Facebook* radi s pomoću triju mehanizama: kažnjava neparticipaciju nevidljivošću, normalizira participaciju te učestalo mjeri vidljivost popularnošću korisnika (usp. ibid). Pomoću tih praksi vidljivo je kako politika i moć mogu operirati u domeni tehnološke infrastrukture *Facebooka* i drugih sličnih platformi poput *Twittera* (usp ibid.). Nadovezujući se na Bucher i njezin uvid u kažnjavanje neparticipacije nepopularnošću, novi se pojam ispoljava. Prema Thrift i French (2002: 312) u situaciji gdje je algoritam skrojen tako da navodi korisnika da aktivno sudjeluje i doprinosi sadržaju na platformi, javlja se pojam *tehnološke nesvjesnosti* te on održava prisustvo kojem se ne može pristupiti; međutim, ono ima vrlo očit učinak kao tehnički supstrat nesvjesnog značenja i aktivnosti.

Prema Ziewitz (2015: 5) *nedokučivost* je jedna od ključnih karakteristika algoritama. Komputacijski procesi strojnog učenja, posebice u slučaju učenja stroja da prepozna ili sintetizira novo lice, nadilaze kapacitete ljudskih kognitivnih sposobnosti. Za Seavera, problem znanja o *tajnim algoritmima*, poput *Facebookova* ili *Twitterova*, leži u činjenici da pokušavamo doći do znanja „koje inženjeri već kontroliraju i koje odlikuje iznimna dinamičnost, konstantne izmjene i multiplikacija“ (2019: 417). Algoritmi su u kulturi suvremenog kapitalizma dosegli status mita koji „nosi određene sličnosti s dugostojećim mitologijama poput nevidljive ruke Adama Smitha ili 'prirodne selekcije' Charlesa Darwina, oni 'se preklapaju s ovom linijom zavodljivih priča o porijeklu reda'“ (Kennedy 2009. prema Ziewitz, 2015:5). No, Seaver (2019.:413) u svojem radu predlaže da algoritme promislmo kao društvene konstrukte kojima doprinosimo kao *outsajderi* i kritičari. Ovakvim pristupom algoritmima Seaver (ibid.:418) u suvremenom mitu nastoji otvoriti prodore koji omogućuju rekonstruiranje njihova prividnog determinizma. Svijest o mehanizmima koji na društvenim mrežama reguliraju korisnički pristup sadržajima na temelju automatskog profiliranja, sve više prodiru javni diskurs. Sve su češći zahtjevi za transparentnosti, no kao i drugi konstruktivisti, Seaver (ibid:417) kritizira ideju da bi transparentnost sama po sebi pridonijela većoj demokratičnosti novih medija. Seaver (ibid.:415) naglašava procese kroz koje se znanje

dostiže u produkciji kontigentnih, parcijalnih i temporalnih istina, umjesto skidanja zastora kako bi se otkrilo stvarno stanje stvari. Prema Seaver (2017.:7) znanje koje bi se trebalo otkriti o algoritmima često ne postoji ni unutar same korporacije, stoga ovaj rad ne govori o algoritmima samo u tehničkom smislu, već o algoritmičnim sustavima čijih je kod dio. Takve sustave održavaju različiti ljudi, dok se osobe koje rade na određenom kodu mijenjaju, a usložnjavanjem takvih izmjena niti sami inženjeri često nisu u stanju uperiti prst na određen komad kôda koji je generirao pogrešku ili predrasudu.

Algoritmi i nove digitalne tehnologije s jedne strane aktivno oblikuju ljudsko iskustvo svakodnevice, dok istovremeno pružaju osnovu za upotrebu i razvoj novih metodologija u istraživanju ljudske svakodnevice. Moj primarni interes u ovom radu je istražiti način na koji se govori o *deepfakeu* na društvenoj mreži *Twitter*. Inspiriran mogućnostima istraživanja koja se otvaraju u domeni digitalne antropologije, kao i mogućnostima komputacijske etnografije, teorijama diskursa i teorijom grafova mreža, u radu, metodološki gledano, pokušavam razviti pristup koji koristi različite tehnike strojne obrade jezika. Strojna obrada jezika koristi se i za kvantitativna kao i kvalitativna istraživanja, a mene zanima na koji način ona može doprinijeti kvalitativnim istraživanjima, posebice analizi diskursa. Prije nego što detaljno objasnim kako koristim strojnu obradu jezika, važno je dati kratak pregled o samoj pojavi *deepfake* tehnologije i načinima njezina korištenja.

3. *Deepfake*

Kao što sam već naveo u uvodu, *deepfake* je nastao kao tehnologija umjetne inteligencije koja koristi napredne algoritme kako bi generirala falsificirane sadržaje. Iako falsificiranje sadržaja nije novost, pri generiranju *deepfakeova* koriste se moćne tehnike strojnog učenja s velikim potencijalom za prevaru percepcije čovjeka. Najčešće korišteni modeli strojnog učenja odnose se na generativne neuralne mreže poput autoenkodera ili generativne suparničke mreže. Generativne suparničke mreže možemo zamisliti kao opoziciju falsifikatora i policajca te ih usporediti s ulogom policajca u igri „mačke i miša”, gdje falsifikator uči proizvesti vjerodostojne falsifikate, a policajac ih uči prepoznati. Oba su procesa pritom dinamična. Policajac uči od falsifikatora, a falsifikator uči od policajca kako proizvesti vjerodostojniji falsifikat. Algoritam u inicijalnoj fazi prima niz fotografija ekstrahiranih iz videosadržaja ciljane osobe, a zatim nastoji generirati nove vizualne sadržaje ciljane osobe, pri čemu ih klasificira kao stvarne ili lažne. Mnoštvom iteracija, učeći

samog sebe, klasifikator doseže točku gdje ne može razaznati original od generiranog falsifikata – i u tom je trenutku *deepfake* uspješno generiran. Ovaj jednostavan, ali moćan algoritam (GAN⁸) objavljen je 2014. kao *opensource code*⁹. Korištenje je takvog koda u početku bilo svedeno na generiranje pornografskih uradaka sa slavnim ličnostima u svrhu njihove distribucije i monetizacije, no ovakva tehnologija „ima potencijal za reklamiranje, umjetnost i ucjenu, pa čak i potencijalno uzrokovanje rata“¹⁰.

Prema Stephanie Beatz, *deepfakeovi* kao učinak imaju zbunjujuć odnos slike i stvarnosti. Kada je videouratkom moguće producirati svaki fiktivan svijet, čin takve produkcije u društvenom kontekstu postaje političkim¹¹. No, kao što Vaccari i Chadwick naglašavaju: „političko krivotvorenje ne dovodi nužno u zabludu pojedince, već može stvoriti nesigurnosti koje zauzvrat mogu smanjiti povjerenje u vijesti na društvenim mrežama“ (Vaccari i Chadwick 2020: 9).

Ovaj rad svakako naginje u smjeru Vaccarijeve i Chadwickove tvrdnje, no ovdje se nameće pitanje koja sve značenja korisnici društvenih mreža pridaju *deepfakeovima* i ako povjerenje u vijesti već jest narušeno, je li ono ponovno uspostavljeno nekim novim kanalima na društvenim mrežama?

Sami mediji o fenomenu *deepfakea* najčešće govore kao o potencijalnoj opasnosti političkih falsifikata, čime se opravdava uvođenja restriktivnih politika. U listopadu 2020. *Microsoft* je izdao alat nazvan *Video Autentifikator* koji bi trebao ograničenom korisničkom tijelu omogućiti uvid u bodovni prag vjerodostojnosti ili postotak moguće manipulacije medijskog sadržaja posredstvom umjetne inteligencije. Alat je iz sigurnosnih razloga trenutno dostupan kao nekomercijalna usluga isključivo za medijske kuće te u svrhu procjena vjerodostojnosti političkih kampanja. Na mrežnoj stranici *Video Autentifikatora* posjetitelj će se susresti s ambicioznim opisom projekta: „Branitelj stvarnosti 2020. (sila protiv *deepfakea*). Nepristran i nekomercijalan napor za pomoć reporterima

⁸ GAN algoritam u kontekstu *deepfake-a* je tehnika koja koristi generativni model za stvaranje uvjerljivih lažnih medijskih sadržaja poput slika ili videozapisa. ChatGPT, „Definicija *GAN-a*,“ generirano s OpenAI, 2024.

⁹ Open source je model razvoja softvera koji omogućuje besplatan pristup izvornom kodu, doprinos zajednice i slobodno redistribuiranje softvera. ChatGPT, „Definicija *open source*,“ generirano s OpenAI, 2024.

¹⁰ Dr. Beatz, „Episode 43: Deepfakes,“ podcast audio, 5.8.2019. , pristupljeno 2.7.2024., <https://thefamiliarstrange.com/2019/08/05/ep-43-deepfakes/>.

¹¹ Dr. Beatz, „Episode 43: Deepfakes,“ podcast audio, 5.8.2019. , pristupljeno 2.7.2024., <https://thefamiliarstrange.com/2019/08/05/ep-43-deepfakes/>.

i kampanjama u očuvanju istine i etičkih standarda.”¹² Prema logici diskursa, prije nego što uopće možemo govoriti o sigurnosti i cenzuri, *deepfake* mora biti predstavljen kao problem. Nesigurnost i neizvjesnost u binarnim okvirima jezika impliciraju potrebu za sigurnošću. Stoga, kako bi se sigurnost omogućila, potrebna je artikulacija novih orijentacija u znanosti; potrebno je prikupljati specifično znanje koje će konstrukcijom novog autoritativnog slučaja polučiti učinke moći. Tome svjedočimo, gornjim primjerom „Branitelja stvarnosti”, ali i naglim razvojem novih grana znanosti, poput digitalne forenzike i multimilijunskih projekata prepoznavanja i analize *deepfakeova* kakve sprovode DARPA¹³ i europski ured protiv prevare¹⁴ te automatizirani algoritmi društvenih mreža. U društvu čija se komunikacija velikim dijelom odvija u digitalnoj sferi, ovakvi su mehanizmi razumna reakcija na prijetnje krađe identiteta, financijskih prevara, osvetničke pornografije itd.

No društvena je stvarnost *deepfakea* raznolika i slojevita. Tehnologiju *deepfakea* koriste mnogi dionici u raznim sektorima, a njezini učinci i etičke implikacije variraju ovisno o kontekstu uporabe. U širem akademskom diskursu o *deepfaku* se u glavnom govori u kontekstu političkih i pornografskih *deepfakeova*. Politički *deepfakeovi* najčešće se vežu za potencijalne prijetnje. U svom istraživanju nisam uspio locirati niti jedan politički *deepfake* koji prethodno nije bio označen kao takav. S druge strane, pornografski *deepfakeovi* čine najveći udio ovakvog sadržaja; često nisu označeni, a oko sprječavanja njihove cirkulacije, prema mojim saznanjima, nisu artikulirane nikakve restriktivne politike.

Cirkulacija *deepfakea* ipak otvara šira pitanja od političkih i pornografskih. Primjer jednog takvog *deepfakea* korišten je u audiokampanji za podizanje svijesti o problemu malarije u zemljama u razvoju. U glasovnoj kampanji „Malarija mora umrijeti“¹⁵, iza lica Davida Beckhama progovara šest osoba različitog spola i etniciteta. Svaka od osoba na vlastitu jeziku izgovara segment cjelovite poruke u vezi problema malarije u zemljama u razvoju. Ovaj asamblaž sadržava

¹² Reality Defender, „Reality Defender — Enterprise-Grade Deepfake Detection,” pristupljeno 01.09.2020., <https://www.realitydefender.com/>.

¹³ DARPA, „DARPA News & Events,” 14.03.2024., pristupljeno 2.7.2024., <https://www.darpa.mil/news-events/2024-03-14>.

¹⁴ European Anti-Fraud Office (OLAF), „Digital Forensics,” OLAF, pristupljeno 2.7.2024., https://anti-fraud.ec.europa.eu/investigations/digital-forensics_en.

¹⁵ Campaign Live, „Deepfake Voice Tech Used for Good in David Beckham Malaria Campaign,” pristupljeno 2.7.2024., <https://www.campaignlive.co.uk/article/deepfake-voice-tech-used-good-david-beckham-malariacampaign/1581378>.

pripadnike neeuropskih zemalja koji zadržavaju svoj glas i jezik, ali gube svoju boju kože noseći masku Davida Beckhama koji je plavooki, plavokosi bijelac. Tvrtka *Synthesia*, koja je producirala kampanju koristeći *deepfake*, multinacionalnim tvrtkama nudi mogućnost generiranja „digitalnih osoba“. Koristeći *deepfake*, generirane su nepostojeće digitalne osobe različitih etničkih karakteristika. Korisnik upisuje tekst u sučelje, a digitalna osoba odabranog spola i etniciteta iščitava tekst. Na ovaj način *Synthesia* nastoji omogućiti korporacijama produkciju jeftinih videoporuka koje simuliraju etnicitet zaposlenika¹⁶.

Mogućnost generiranja realističnih falsifikata osoba ne staje isključivo na simuliranju etniciteta. Izraelska tvrtka D-ID¹⁷ nudi svojim korisnicima mogućnost deidentifikacije osoba. Snimke s kamera za video nadzor zbog zaštite podataka ne mogu biti pohranjene i monetarizirane u procesima rudarenja podataka. Upotrebom *deepfake* algoritma tvrtka D-ID generira falsifikate snimljenih osoba. Falsifikat zadržava ključne karakteristike lica osobe, istovremeno brišući osobne karakteristike koje omogućuju strojnu identifikaciju snimljene osobe. Falsifikat ljudskom oku izgleda gotovo identično originalu, međutim, stroj ne prepoznaje falsifikat. Na ovaj način korporacije mogu legalno monetarizirati podatke ekstrahirane iz lica osoba, poput: starosti, spola, orijentacije pogleda, etniciteta. *Deepfakeovi* se također koriste i u popularnim aplikacijama poput kineskog „Zao“¹⁸, aplikacije koja omogućuje umetanje lica korisnika u popularne filmske scene. Nejasno je što tvrtka radi sa pohranjenim podacima korisnika, međutim, kao što je časopis *Forbes* svojedobno pisao (2019), privatnost milijuna korisnika potencijalno je ugrožena korištenjem aplikacije.

Osim prethodno navedenih primjera, *deepfakeovi* su također bili i korišteni u svrhu zaštite progonjenih skupina, što kao primjer pokazuje dokumentarni uradak „Welcome to Chechnya“ reportera David Franca¹⁹. Dokumentarac prikazuje kako čečenski politički vođa Ramzan Kadyrov u jednom od intervjua opisuje homoseksualne osobe kao *subhuman* (podljudi). Pripadnici LGBTQ zajednica mučeni su i ubijani u koncentracijskim kampovima, što je prikazano u dokumentarcu.

¹⁶ Synthesia, „Ethical Deepfake Maker | Use AI Deepfake for Good | Synthesia,“ pristupljeno 2.7.2024., <https://www.synthesia.io/tools/deepfake-video-maker>.

¹⁷D-ID, „D-ID Creative Reality™,“ pristupljeno 2.7.2024., <https://www.d-id.com>.

¹⁸Zak Doffman, „Chinese 'Best Ever' Deepfake App Zao Sparks Huge FaceApp-Like Privacy Storm,“ *Forbes*, 2.9. 2019., pristupljeno 2.7.2024., <https://www.forbes.com/sites/zakdoffman/2019/09/02/chinese-best-ever-deepfake-app-zao-sparks-huge-faceapp-like-privacy-storm>.

¹⁹ IMDb, „Welcome to Chechnya (2020),“ pristupljeno 2.7.2024., <https://www.imdb.com/title/tt11394200/>.

Uz pomoć *deepfake* tehnologije, identitet žrtava koje u dokumentarcu pružaju ispovijesti o političkom progono, zaštićen je generiranim hiperrealističnim maskama nepostojećih osoba.

Deepfake također uključuje i različite forme političkog aktivizma gdje autori amateri produciraju *deepfakeove* slavni i utjecajni osobe poput Donalda Trampa ili Marka Zuckerberga, koji zatim u svom *deepfake* izdanju progovaraju o prikivenim istinama koje nikada ne bi izrekli. Tako *deepfake*se također koristi i na ironičan način kako bi razotkrio društvenu „istinu”. Donald Trump²⁰ u *deepfake* formatu kritizira belgijske ekološke politike, dok Mark Zuckerberg²¹ korisnicima *Facebook* mreže poručuje kako se on zapravo bogati na ukradenim osobnim podacima korisnika te popularne društvene mreže.

U ovom radu *deepfake* odjeljujem od lažnih vijesti. Glavna razlika između *deepfakeova* i lažnih vijesti nije u njihovu potencijalu za obmanu, jer mnogi autori tvrde da su klasične medijske taktike dezinformacija još uvijek učinkovitije u tome, niti u samoj tehnologiji koja se koristi za generiranje raznih prikaza nepovezanih s ljudskim tijelom. Glavni fokus njihove produkcije je određeno područje ljudskog tijela – ljudsko lice. Deleuze i Guattari tvrde da je ljudsko lice „nehumano u ljudskim bićima“ (1987: 171). „(...) određeni sklopovi moći zahtijevaju proizvodnju lica, dok drugi ne“ (ibid.: 175). „(...) razlog je jednostavan (...) lice nije univerzalno (...) nije čak ni lice bijelog čovjeka (...) lice je tipični Europljanin (ibid.: 176). „(...) lice je politika (ibid.: 181)”. U mojoj je analizi lice diskurzivno. Ono je kulturno specifičan artefakt koji označava i identificira ljudsko tijelo, ali i oblik koji mutira unutar specifičnog kulturno-povijesnog konteksta. Kao što Jenny Edkins sumnja: „politika koja stvara lice je politika koja proizvodi osobu kao objekt (...) Osoba se poziva da bude (...) kategorizirana, poznata.“ (2015: 7)

Davne 1996. Thomas Macho dijagnosticirao je kako živimo u facijalnom društvu koje neprekidno proizvodi lica (usp. Macho, 1996: 26). Stoga je *deepfake* moguće povezati s izmijenjenom vizualnom kulturom koja je prema Beltingu (2017: 214-303), u okvirima tehnoloških promjena i poststrukturalističke krize identiteta, poljuljala obećanje autentičnosti i

²⁰ Politico, „Belgian Socialist Party Circulates 'Deep Fake' Donald Trump Video,“ pristupljeno 2.7.2024., <https://www.politico.eu/article/spa-donald-trump-belgium-paris-climate-agreement-belgian-socialist-party-circulates-deep-fake-trump-video/>.

²¹ Artnet News, „Artists Create a Sinister 'Deepfake' of Mark Zuckerberg to Teach Facebook (and the Rest of Us) a Lesson About Digital Propaganda,“ 12.6.2019., pristupljeno 2.7.2024, <https://news.artnet.com/art-world/mark-zuckerberg-deepfake-artist-1571788>.

individualnosti ljudskog lica; masovni mediji doprinijeli su vizualnom prezasićenju preslikanim licima čija je singularnost i moć djelovanja obezvrijeđena programima za obradu fotografija poput *photoshopa* ili *morfinga*²². Danas živimo u društvu za koje je karakterističan koncept koji Agamben naziva „identitet bez osobe“ (2011:46); ovaj novi identitet, lišen osobe, nije određen prepoznavanjem zajednice, već državnih institucija:

U trenutku kada su pojedinci prikovani za čisto biološki i asocijalni identitet, obećava im se i mogućnost da preuzmu sve maske i sve druge i treće živote moguće na Internetu, nijedan od kojih im zaista ne može pripadati. Što su više građani metropole izgubili intimu jedni s drugima, što su više postali nesposobni gledati se međusobno u oči, to je utješnija virtualna intima s aparatom (...) Ovdje sam ako me Stroj prepoznaje ili barem vidi; živ sam ako Stroj, koji ne poznaje ni san ni budnost, već je na vječnom oprezu, jamči da sam živ. Nisam zaboravljen ako Veliko Pamćenje zabilježi moje numeričke ili digitalne podatke (ibid.:53).

Ova afirmacija života nipošto nije vitalistička, već nihilistička.

Ovim kratkim pregledom nastojao sam *deepfake* artikulirati u širim okvirima postrukturalističke krize identiteta. Iako sam svjestan činjenice da se radi o izuzetno kompleksnom fenomenu koji varira ovisno o kontekstu, pa samim tim ima i razna značenja, zanima me u kojoj mjeri analiza strojne obrade jezika može doprinijeti razumijevanju načina na koji korisnici društvenih mreža doživljavaju *deepfake*, koja mu značenja pridaju, u kojim kontekstima se pojavljuje u diskusijama te na koje kontekste se odnosi. Posebnu pozornost posvećujem analizi sentimenta i, iako se taj pristup često koristi u kvantitativnim istraživanjima, u ovom ga radu kombiniram s komputacijskom etnografijom, teorijom grafova i analizom mreže supojavnosti *hashtagova* kako bih ukazao na razne diskurse.

3.1 Analiza sentimenta

Bing Liu (2015:1-10) definira analizu sentimenta ili *opinion mining* kao područje istraživanja mišljenja, emocija i stavova ljudi izraženih u tekstu u vezi različitih entiteta. Ovo polje

²² Scott London, „The Face of Tomorrow: Reflections on Diversity in America,“ pristupljeno 2.7.2024., <https://scott.london/articles/newface.html>.

prate brojni povezani zadatci, a počecima analize sentimenta se smatraju radovi Nasukawa i Yi (2003. prema *ibid.*) te Davea i drugih (2003. prema *ibid.*) Ranija istraživanja od Wiebe (2000. prema *ibid.*) i Das and Chen (2001. prema *ibid.*) postavila su temelje ove metode iako im prethode istraživanja koja su doprinijela razvoju analize sentimenta, poput interpretacije metafora, ekstrakcije sentimenta pridjeva, afektivne komputacije, analize subjektivnosti, gledišta i afekta (Wiebe, 1990, 1994; Hearst, 1992; Hatzivassiloglou and McKeown, 1997; Picard, 1997; Wiebe et al., 1999, prema *ibid.*). Prema Liu (*ibid.*) postoje tri distinktivne razine analize sentimenta: analiza na razini dokumenta, pri čemu se procjenjuje negativni ili pozitivni emotivni naboj čitavog dokumenta; analiza na razini rečenice, pri čemu se procjenjuje pozitivan, negativan ili neutralan emotivni naboj rečenice te analiza na razini aspekta. Analiza na razini aspekta u obzir uzima da rečenica može sadržavati kontradiktorne emotivne naboje u vezi različitih entiteta. Osim različitih razina analize, važno je istaknuti i različite pristupe. Analiza sentimenta u tehničkom smislu moguće je sprovesti leksičkom analizom ili analizom baziranoj na modelima strojnog učenja. Leksička analiza bazirana je na predefiniranim rječnicima u kojima su riječima pripisani bodovi emotivnog naboja, najčešće u rasponu od -1 do 1, pri čemu su rezultati bliži -1 procijenjeni kao negativni, bliži 0 kao neutralni, a bliži 1 kao pozitivni. Ovisno o razini analize, kalkulira se prosječna vrijednost bodovanih riječi. Pristup baziran na modelima strojnog učenja, s druge strane, znatno je kompleksniji, ali u bodovanje uključuje i iščitavanje konteksta, pa je stoga i kompleksniji. U međuvremenu je lingvistika kao znanstvena disciplina pružila znatni doprinos razradi leksikona za analizu sentimenta. Nissim (2017:33) navodi primjere koji uključuju rješenja kojima se u obzir uzima višeznačnost, primjerice da epitet „lud” može u različitim kontekstima biti shvaćen kao pozitivan ili negativan, višeznačnosti na razini riječi, primjerice da u engleskom jeziku „novel” može označavati pridjev novosti ili imenicu roman, proširenja kategorija pozitivnog i negativnog na različite emotivne kategorije.

U ovom radu korišten je *VADER*²³ leksikon (*Valence Aware Dictionary and Sentiment Reasoner*), prilagođen za analizu sentimenta na društvenim mrežama. *Vader* se kao alat često koristi u programskom jeziku *Python* te pomaže pri određivanju sentimenta teksta, a popularan je i u analizi osjećaja na društvenim mrežama.

²³ Cjhutto, „vaderSentiment,” GitHub, pristupljeno 2.7.2024., <https://github.com/cjhutto/vaderSentiment>.

3.2 Prikupljanje građe za analizu i strojna obrada jezika

U svrhu prikupljanja građe za ovaj radi korištena je verzija 2 *Twitterova restfull API-a*²⁴. *API* ili *Application Program Interface* programerska je apstrakcija koja označava most između dviju aplikacija. Jedna aplikacija preko *API-a* može komunicirati s drugom, bez potrebe integracije sirovog koda. Epitet *REST* označava protokol komunikacije između korisnika, preglednika i poslužitelja na kojem su pohranjeni podatci kojima korisnik želi pristupiti. *Restfull API* su stoga oni koji korisnicima omogućuju komunikaciju s različitim internetskim servisima. Svaka internetska stranica nalazi se na određenoj lokaciji koju nazivamo *URI (unique resource identifier)*. Upisom adrese */URL* u pretražnik korisnik izvršava zahtjev poslužitelju, a taj zahtjev vraća podatke koje pretražnik interpretira u formatu internetske stranice gdje korisnik kroz vizualne elemente navigira sadržajem stranice. *Restfull API* ima predodređenu strukturu i *URI-ove* koji se nazivaju *endpointovima*. Kada na određeni *endpoint* izvršimo zahtjev, poslužitelj vraća podatke u predefiniranoj strukturi koja se koristi za daljnju obradu u aplikaciji ili kao građa za analizu. S obzirom na činjenicu da je u ovom radu korištena besplatna verziju *API-a*, ona je imala određena ograničenja te se dodatno u *Python codu* moralo definirati petlje koje će ciklički paliti i gasiti postavljene zahtjeve. Dobivena građa nalazi se u formatu *JSON*²⁵, koji je ugniježđen. *JSON* struktura sastoji se od parova ključeva i vrijednosti ključeva. Primjerice, možemo imati jedan ključ pod nazivom tekst te će svaki ključ imati drugu vrijednost teksta određenog *posta*. Ugniježđena struktura označava da pod jednim elementom možemo imati niz drugih elemenata, poput poglavlja i potpoglavlja od potpoglavlja. Takav je format bilo potrebno pohraniti u tablični format – zbog toga je kreirana *XSLT shema*²⁶ s pomoću koje su izvučeni svi elementi podrazina na istu razinu. *JSON* je prvo pretvoren u *XML* te je zatim na *XML* primijenjena *XSLT shema* kako bi se normaliziran *XML*, uz pomoć *Pandas* knjižnice²⁷, učitala u tablični format *Pandas* podatkovnog okvira. *Pandas* knjižnica jedan je od paketa u programskom jeziku *Python*. *Pandas* omogućuje

²⁴ „API reference index,“ Twitter Developer, pristupljeno 2.3.2023., <https://developer.twitter.com/en/docs/api-reference-index>.

²⁵ *JSON* (JavaScript Object Notation) je lagan format za razmjenu podataka koji se koristi za strukturiranje i pohranu informacija u čitljivom obliku. ChatGPT, „Definicija *JSON-a*,“ generirano s OpenAI, 2024.

²⁶ *XSLT* (Extensible Stylesheet Language Transformations) je jezik za transformaciju *XML* dokumenata u druge formate poput *HTML-a* ili drugih *XML* struktura. *XML* (Extensible Markup Language) je označni jezik koji se koristi za označavanje struktura podataka radi lakše razmjene i obrade informacija između različitih računalnih sustava. ChatGPT, „Definicija *XSLT-a*,“ generirano s OpenAI, 2024.

²⁷ *Pandas* je popularna *Python* knjižnica za analizu i manipulaciju podataka, koja pruža moćne alate i strukture podataka za rad s tabličnim podacima. ChatGPT, „Definicija *Pandasa*,“ generirano s OpenAI, 2024.

pohranjivanje i obradu podataka u tabličnom formatu. Kôd koji je korišten, kao i samo objašnjenje kôda, dostupno je na *Github*²⁸ repozitoriju kreiranom za ovaj rad. Osim *API-a*, postoji i druga mogućnost prikupljanja podataka, a ona se naziva *scraping*. Prilikom *scrapinga*²⁹ programski se simulira internetski pretražnik i upiti korisnika. Ova je metoda, iako popularna, etički upitna, i stoga nije iskorištena u svrhu ovog istraživanja. Odabrana je verzija 2 *Twitterova API-a* jer struktura povratnih informacija sadrži korisne metapodatke, poput anotacije *threada* u kojem se javlja pojedinačni post korisnika. Drugi korisni metapodatci uključuju kontekstualne anotacije, vrijeme objave, lokacije objave, jezik objave, identifikacijski broj objave, identifikacijski broj korisnika, broj *retweetova*, broj *likeova*, citirane *postove*, identifikacijski broj referiranih *postova* te poveznice na vizualnu građu uključenu u poruku korisnika. Svaki red tablice čini jednu izjavu korisnika u koloni pod nazivom tekst; preostale kolone sadrže metapodatke o tom tekstu.

Nakon što sam prikupio podatke svih korisničkih unosa unazad pet dana od postavljanja upita (26. ožujka 2023.- 2. travnja 2023.), napisao sam funkcije u *Pythonu* koje iz kolone „tekst“ ekstrahiraju u novu kolonu pod nazivom „hashtags“ *hashtag* vrijednosti iz teksta. Nakon toga sam filtrirao tablicu da se zadrže samo oni unosi koji sadrže više od jednog *hashtaga*. Napisao sam novu funkciju koja bilježi sve moguće kombinacije *hashtagova* iz pojedinačnog reda u parove te broj ponavljanja parova. Dobivene sam parove uz pomoć *Pythona* ispisao u tablicu pod nazivom „veze“. U ovoj tablici u kolonu „source“ upisana je prva vrijednost *hashtag* para, a u kolonu „target“ druga vrijednost pojedinačnog *hashtaga* para; u kolonu „type“ je za svaki red tablice upisana vrijednost „undirected“ s obzirom na to da smjer veze nije određen; u kolonu „weight“ upisana je brojčana vrijednost frekvencije ponavljanja parova. U drugu tablicu pod nazivom „čvorovi“ uz pomoć *Pythona* u kolonu „id“ upisao sam jedinstvenu pojavu *hashtaga*; u kolonu „label“ su kopirani unosi iz kolone „id“; u koloni „weight“ upisana je brojčana vrijednost frekvencije ponavljanja jedinstvene pojave *hashtaga* u ukupnom korpusu; u kolonu „color“ upisana je heksidecimalna šifra za boju pripisana kalkulaciji vrijednosti sentimenta pojedinačnog *hashtaga* (zeleno za pozitivne vrijednosti, crveno za negativne i ljubičasto za neutralne).

²⁸ GitHub je platforma za smještaj i upravljanje razvojem softvera pomoću sustava za verzioniranje koda Git, omogućujući suradnju, praćenje promjena i upravljanje projektima. ChatGPT, „Definicija *GitHub-a*,“ generirano s OpenAI, 2024.

²⁹ Scraping je proces automatiziranog prikupljanja podataka s web stranica, često pomoću programskih skripti ili alata, radi analize ili spremanja informacija u strukturirani oblik. ChatGPT, „Definicija *scrapinga*,“ generirano s OpenAI, 2024.

Vrijednosti sentimenta iskalkulirane su upotrebom *VADER* knjižnice. S obzirom na to da su se kalkulacije sentimenta odnosile na rečenice koje sadrže više od jednog *hashtaga*, napisao sam dodatnu funkciju u *Pythonu* koja kalkuliira medijan vrijednosti pojedinačnog *hashtaga* s obzirom na distribuciju vrijednosti u cijelom korpusu. Vrijednosti sentimenta prevedene su u pripadajuću heksidecimalnu šifru za boju s obzirom na predodređene pragove vrijednosti za pozitivno, negativno i neutralno konotirani sadržaj. Pozitivnom vrijednosti smatra se sve iznad 0.05, negativnom sve manje od -0.05, a neutralnom sve između tih dviju vrijednosti. Kreirao sam tablice „čvorovi“ i „veze“ kako bih građu mogao učitati u *open-source* program *Gephi*³⁰ radi mrežne vizualizacije građe. Zatim sam napisao funkciju koja iz primarnog podatkovnog okvira koji sadrži izvornu građu svaku vrijednost iz reda kolone „tekst“ ispisuje u zasebni *.txt* dokument nazvan po indeksu pripadajućeg reda. Ovaj niz tekstualnih datoteka učitao sam u *opensource* program *AntConc*³¹ koji je alat korišten za korpusnu lingvistiku. Ovo je učinjeno kako bih pri iščitavanju čvorova iz mreže generirane u *Gephi-u*, u *AntConc* tražilicu mogao upisati *hashtag* koji me zanima i dobiti uvid u sve izjave koje su označene tim *hashtagom*. S obzirom na to da je svaki dokument u *AntConc* korpusu nazvan po indeksu korespondirajuće vrijednosti originalnog podatkovnog okvira, moguće je u iščitavanju provjeriti metapodatke iščitane izjave, poput broja *retweetova*. Bitno je napomenuti kako sam vrijednosti *hashtaga* normalizirao pisanjem funkcije koja slične izraze s razlikama u sufiksu ili razlike velikih i malih slova svodi na istu vrijednost, kako bi se izbjeglo ponavljanje čvorova.

Kada je prikupljena građa vizualizirana kao graf mreže supojavnosti *hashtagova* sadržanih u *tweetovima* u kojima se spominje termin *deepfake*, u *Gephiu* sam podesio parametre za vizualizaciju. Veličina čvora određena je vrijednošću kolone „*weight*“ iz tablice „čvorovi“. Korištenjem *ForceAtlas* algoritma čvorovi su vizualno razmješteni ovisno o vrijednosti kolone „*weight*“ iz tablice „veze“. Češće ponavljanje veze znači veću težinu i veću silu privlačenja u generiranoj vizualizaciji. U procesu navigiranja građom *Gephi* sučelje omogućuje vizualnu preglednost podmreža. Prelaskom sa strjelicom računalnog miša preko odabranog čvora, vizualno se ističu samo direktno povezani čvorovi. Tako je moguće „šetati“ grafom u druge nepovezane

³⁰ „The Open Graph Viz Platform,“ Gephi, pristupljeno 2.3.2023., <https://gephi.org/>.

³¹ Anthony, Laurence. „AntConc,“ Laurence Anthony's Website, pristupljeno 2.3.2023., <https://www.laurenceanthony.net/software/antconc/>.

podmreže ili u povezane podmreže podmreža. S obzirom na to da su svi čvorovi mreže neminovno povezani s terminom *deepfake*, njegov je čvor uklonjen. Kada bih uočio čvorove čija me poveznica zanimala u kontekstu drugih čvorova, odabrani *hashtag* bih upisao u tražilicu generiranog *AntConc* korpusa. Kombinirano korištenje vizualizacije grafa mreže uz kvalitativno iščitavanje komentara u *AntConc* korpusu omogućilo mi je nelinearno iščitavanje građe.

Osim toga, kvalitativnim iščitavanjem komentara provjeravao bih je li rezultat kalkulacije vrijednosti sentimenta donekle točan. „Donekle točan“, jer je generiran rezultat sentimenta, vrijednost medijana uz težine pripisane pojedinom *hashtagu* s obzirom na učestalost javljanja u korpusu. Korekcije građe manualno su zabilježene u zasebnu kolonu originalnog podatkovnog okvira jer mogu biti korišteni za usavršavanje procjene vrijednosti sentimenta korištenjem modela mašinskog učenja. Dok je *Gephi* korišten za vizualno iščitavanje, a *AntConc* za tekstualno, bilješke su pohranjene u programu *Obsidian*³² (*LogSeq*³³ je *open-source* alternativa). Ovaj program omogućuje upotrebom jednostavnih prefiksa i sufiksa markirati tekstualne elemente bilješki kao veze ili oznake te ih kasnije vizualizirati kao graf mreže. Svaka je bilješka nazvana po imenu *hashtaga* koji se interpretira. Ovako je u bilješkama moguće markirati poveznice s drugim diskursima koji se uočavaju u procesu iščitavanja *Twitter* mreže te bilješke vizualizirati kao novu mrežu ili je spojiti s postojećom mrežom. Mrežu je supojavnosti *hashtagova* također moguće obogatiti kroz automatiziranu ekstrakciju organizacija na koje se korisnici referiraju u izjavama označenim određenim *hashtagom*. Pritom je organizacija ubačena u mrežu kao novi čvor u relaciji s *hashtagovima* na koje se referira. Na ovaj način kombiniranjem automatizirane obrade teksta i manualne anotacije mogu se generirati mreže koje na smislen i predefiniiran način različite razine i tipove veza izvlače na istu razinu; takve reprezentacije su pljosnate, postoji samo površina bez dubine. Iako je generirana mreža pokušaj prikaza ukupnog korpusa, njezino je iščitavanje selektivno. U nastavku ću interpretirati primjere proizišle iz „šetnje“ mrežom.

³² „Obsidian - Sharpen your thinking,“ Obsidian, pristupljeno 2.3.2023., <https://obsidian.md/>.

³³ „A privacy-first, open-source knowledge base,“ Logseq, pristupljeno 2.7.2023., <https://logseq.com/>.

4. Rezultati



Prikaz 2. vizualizacija mreže supojavnosti *hashtagova*, čvor: „ugccommunity“

U odabranoj mreži *UGC community* čvor (Prikaz 2.) vizualno se ističe u odnosu na ostale zbog svoje veličine. Njegova isključiva povezanost s manjim čvorom *faceswap*³⁴ druga je distinktivna karakteristika u odnosu na konstalaciju ukupne mreže. Atribut crvene boje sugerira negativnu percepciju u kontrastu s pozitivnom vrijednošću (zeleni atribut) *faceswap* čvora. Iščitavanjem sadržaja komentara u vezi korisnički generiranog sadržaja razvidno je isticanje njegova potencijalnog izumiranja uslijed napretka u *deepfake* tehnologiji.

„Jednostavan korisnički generirani sadržaj kakav poznajemo nestati će zbog deepfakea.“ (4544)

Naime, u društvenim medijima korisnički generirani sadržaj kombinira tekstualne, auditivne i/ili vizualne elemente. Tekst služi kao primarna poruka čiji sadržaj može biti generiran upotrebom modela poput *GPT*³⁵ (relevantan čvor prisutan u mreži). Generirani sadržaj često se nadograđuje u auditivnu i/ili vizualnu formu. *Deepfake* tehnologija ovo dodatno komplicira, simulirajući glasove i lica. Drugim riječima, kombinacija vizualnih i *audiodeepfakeova* s nedavnim naprecima generativnih modela umjetne inteligencije otvara mogućnost dodatne ekspanzije onoga što Yancey

³⁴ Zamjena lica je tehnologija koja digitalno zamjenjuje lice jedne osobe s licem druge osobe na slikama ili u videozapisima. ChatGPT, „Definicija *faceswap*,“ generirano s OpenAI, 2024.

³⁵ GPT, ili Generativni Pre-trenirani Transformer, je vrsta naprednog jezičnog modela razvijenog od strane OpenAI koji koristi duboko učenje za generiranje teksta sličnog ljudskom na temelju primljenog unosa. ChatGPT, „Definicija *GPT-a*,“ generirano s OpenAI, 2024.

Strickler (2024.) naziva *mračnom šumom*. Teorija mračne šume opisuje *online* prostore preplavljene *botovima*³⁶, *trollovima*³⁷, generičkim korisničkim sadržajem i algoritmičnom manipulacijom lišenom „prave“ ljudske interakcije. Kako bi ovo izbjegli, ljudi se povlače u zatvorene platforme (poput privatnih -*WhatsApp* i *Discord* grupa) da prikriju svoju prisutnost u javnim prostorima i da održe „autentičnu“ komunikaciju. U tom smislu, većina korisničkog sadržaja na digitalnim društvenim platforma pripada površinskom sloju Interneta čija primarna svrha i izvor financijske opstojnosti čini reklamiranje i prodaja proizvoda i usluga. U tom smislu epitet „jednostavnog“ korisnički generiranog sadržaja implicira nostalgični sentiment za korisnički generiranim sadržajem prije napretka u generativnoj tehnologiji umjetne inteligencije i *deepfaka*. Ipak, *deepfake* je u ovakvom kontekstu segment šire problematike tekstualnosti u interakcijama korisnika digitalnih platformi. Tekst, bilo u pisanom ili audiovizualnom formatu, jednom kada je objavljen, može lako biti prerađen, modificiran, recikliran i redistribuiran dvojakim utjecajem odnosa publike i kreatora sadržaja. *Deepfake* i umjetna inteligencija mogu amplificirati navedene aspekte tekstualnosti digitalnih sadržaja, no problem autorstva u generativnom korisničkom sadržaju proizlazi iz šireg konteksta koji je prethodio pojavi *deepfake* tehnologije. Godine 1969. Foucault je predvidio ovakvu situaciju i primijetio kako:

„Možemo lako zamisliti kulturu u kojoj bi se diskursi širili bez potrebe za autorom. Diskursi, bez obzira na njihov status, oblik ili vrijednost, i bez obzira na način na koji ih obrađujemo, razvijali bi se u općoj anonimnosti. Više nema zamornih ponavljanja: 'Tko je pravi autor?' 'Imamo li dokaze o njegovoj autentičnosti i originalnosti?' (..) 'Nova pitanja će glasiti: Koji su načini postojanja ovog diskursa?' 'Odakle dolazi; kako se širi; tko ga kontrolira?' 'Koji su položaji određeni za moguće subjekte?' 'Tko može ispuniti ove različite funkcije subjekta?'“ (Foucault, 1977: 138)

U slučaju mojih primjera, izjava: „bez potrebe za autorom“ implicira da ne postoji niti jedan subjekt koji nije okaljen svojom okolinom i ona njime. Ovakva dinamika korespondira s konceptom koji Deleuze (1995: 141) naziva *haeccity*. Autor postoji isključivo kao asamblaz drugih autora, algoritmičnih/programskih i materijalnih arhitektura koje postavljaju okvire društvene

³⁶ Botovi su automatizirani programi dizajnirani za obavljanje specifičnih zadataka ili simuliranje ljudske aktivnosti na internetu. ChatGPT, „Definicija *botova*,“ generirano s OpenAI, 2024.

³⁷ Trolovi na internetu su pojedinci koji namjerno provociraju i uznemiruju druge objavljivanjem provokativnih, irelevantnih ili uvredljivih komentara. ChatGPT, „Definicija *trollova*,“ generirano s OpenAI, 2024.

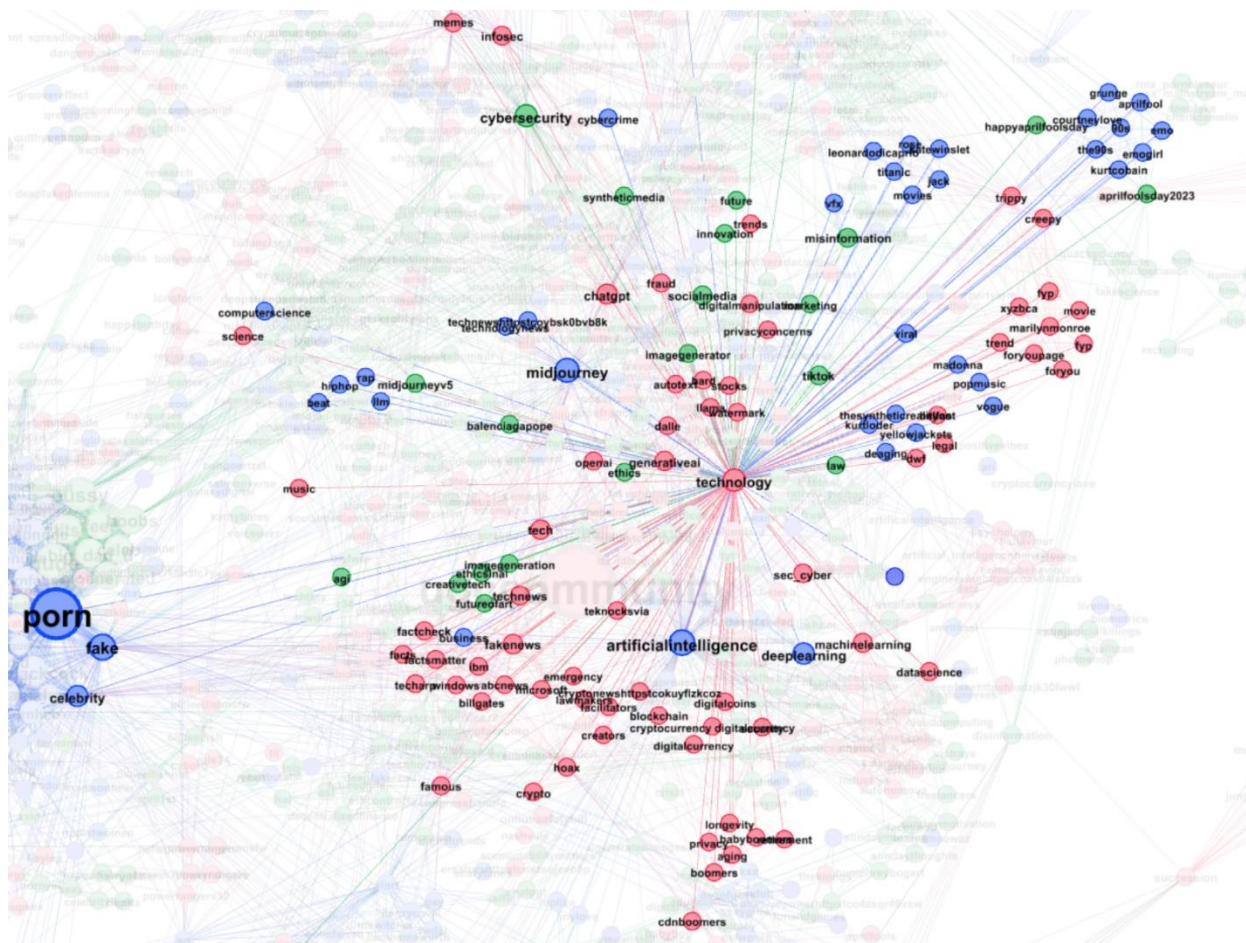
igre. U tom smislu, distinkcija između korisničkog sadržaja generiranog umjetnom inteligencijom i tradicionalnog autorskog sadržaja otkriva da obje forme sadržaja izazivaju održivost ideje jedinstvenog, koherentnog autora. Umjetna inteligencija kao nehumani agent omogućuje produkciju sadržaja pri čemu se fokus pomiče s identiteta kreatora na sâm sadržaj i nove modalitete njegova postojanja i cirkulacije. Ovakvi sustavi generiraju sadržaj koji odražava unesene podatke i treniranje isprepletено s kreativnosti pojedinca koje leži u novom području inženjerstva upita³⁸ (*prompt engineering*). Ovo proširuje problematiku tekstualnosti digitalnih interakcija na neljudske dionike, ali u smislu šire problematike autorstva postojalo je i prije pojave tehnologija umjetne inteligencije.

S druge strane, u korelaciji s negativno konotiranim korisnički generiranim sadržajem nailazimo na pozitivne komentare u vezi *faceswap* tehnologije:

„*Da, to je moje lice na licu Jennifer Lopez.*”(27678)

S obzirom na sadržaj odabranog komentara, poveznica pozitivnog i negativnog može proizlaziti iz percepcije metode zamjene lica kao inovativnog alata za obogaćivanje korisničkog sadržaja. U ovom kontekstu, pristup identitetu u odnosu na *deepfake* tehnologiju ukazuje na to da, iako zamjena lica omogućuje uvjerljive, ali neistinite prikaze, one istovremeno nose potencijal za kreativnost i izražavanje. Ovaj pristup kompromitira tradicionalno shvaćanje autentičnosti, naglašavajući da se ono ne nalazi isključivo u esencijalističkoj interpretaciji.

³⁸ Inženjering upita je proces dizajniranja i usavršavanja ulaznih upita kako bi se učinkovito usmjerili i optimizirali odgovori AI modela. ChatGPT, „Definicija *prompt engineeringa* (inženjerstvo upita),“ generirano s OpenAI, 2024.



Prikaz 3. vizualizacija mreže supojavnosti *hashtagova* čvor: „technology“

Sljedeći uočljiv primjer u mreži je čvor s oznakom „tehnologija“ (Prikaz 3.). Uz „pornografiju“, ovo je jedan od najpovezanijih čvorova u mreži iako znatno manji u odnosu na „korisnički generirani sadržaj“. S obzirom da se radi o vrlo općenitoj oznaci, supojavnosti uključuju širok raspon tematskih poveznica u raspravama na *Twitteru*. Prema procjenama sentimenta, to su pozitivno konotirane (zelene) oznake poput inovacije, budućnosti, marketinga, etike, zakona, budućnosti umjetnosti ili pak šaljivih konotacija poput prvog aprila. No, kalkulaciju sentimenta potrebno je uzeti s oprezom jer, kao što ću kasnije u radu pokazati, neke od procjena vrijednosti sentimenta nisu bile ispravne, primjerice za čvor budućnosti i zakona.

Neutralno kalkulirani čvorovi (plavi) su tematski konzistentni i u najvećoj mjeri sastoje se od sadržaja koji su vezani za rasprave o mogućnostima *deepfakea* da lica nepoznatih osoba prenesu na tijela slavnih glumaca u upečatljivim filmskim scenama ili pak na tijela slavnih pjevača i

pjevačica, pri čemu se posebno ističe nostalgija za devedesetima i žanrovima poput *grungea*. Kasnije u radu ću se detaljnije osvrnuti na ovakve primjere.

Iako su neutralno konotirani primjeri tematski konzistentni, uočljivo je kako je u raspravama znatno zastupljeniji (vizualno veći) čvor pornografije i slavni ličnosti. On u tehnološkom smislu dovodi u vezu, s jedne strane, benignu upotrebu poput umetanja nepoznatih osoba u poznate scene, a s druge strane, nekonzenzualnu *deepfake* pornografiju, na što ću se detaljnije osvrnuti kasnije u radu.

Negativno procijenjeni čvorovi u raspravama na društvenoj mreži u vezi tehnologije *deepfakea* u najvećoj mjeri su vezani za velike kompanije poput *IBM-a* i *Microsofta*, bogate pojedince poput Bill Gatesa, različite metode brze zarade putem interneta kao što su *kriptovalute*, ali i znanstvene discipline vezane za *deepfake* poput znanosti o podacima. Isčitavanjem komentara vidljivo je kako su negativno konotirani čvorovi u najvećoj mjeri vezani za incident u kojem je za *ABC* medijsku kuću Sarah Ferguson intervjuirala Bill Gatesa o kontroverznim pitanjima poput COVID-19, Elon Muskove kupnje *Twittera* i uspona umjetne inteligencije.

Određeni *TikTok* korisnik koji, prema izvještajima medija, promiče desničarske politike je ovu snimku preradio kao *deepfake* koji, prema navodima medija, nije bio satira, već pokušaj promicanja desno orijentiranih vrijednosti³⁹. Negativna konotacija tehnologije sa sobom nosi prevladavajući skepticizam javnog diskursa društvenih mreža u vezi pojma *deepfake*. Analiza odabranih komentara ukazuje na intersekciju tehnoloških napredaka, javnog mijenja i utjecaja *deepfake* sadržaja. To je vidljivo i iz sljedećeg komentara vezanog za čvor *factchecker* iz odabrane mreže.

„Fact checker⁴⁰ nije podjelio sa svijetom da je koristio Deepware tehnologiju osmišljenu isključivo da razotkrije deepfake VIDEO kao AUDIO. Bilo tko sa imalo tehnološkog znanja zna da će rezultati biti KRIVI. To je poput korištenja testa za HIV za testiranje COVID-19“ (151)

³⁹ "Was Bill Gates Destroyed In ABC News Interview?! - Tech ARP," Tech ARP, pristupljeno 2.7.2024, <https://www.techarp.com/facts/bill-gates-abc-interview-destroyed/>.

⁴⁰ Foundation for Investigative Journalism, „FIJ NG,“ pristupljeno 3.7.2024, <https://fij.ng/>.

Komentar kako agencije koje se bave provjerom činjenica nisu objavile korištenje programa *deepware* specijaliziranog za razotkrivanje zvučnih *deepfakeova* u svrhu razotkrivanja vizualnih, uvodi kritičku perspektivu ograničenja i potencijalnih zastranjivanja u trenutnoj infrastrukturi provjere činjenica. Analogija kojom se ovakav pristup veže uz provjeru HIV-a COVID testom, ističe neusklađenost između korištenih alata i izazova postavljenih produkcijom *deepfakea*. Nameće se pitanje o točnosti i pouzdanosti procesa utvrđivanja činjenica. S obzirom na to da komentar uvodi kritičku perspektivu, u svojoj analizi kontrastiram ga sa širom akademskom raspravom u vezi narušavanja institucionalnog autoriteta⁴¹ (Ajder, 2019, prema Cover 2023:80) i posljedične „infokalipse“ (Lenters, 2021. prema *ibid.*). „Infokalipsa“, kako ju tumači Lenters, korištena je da označi *deepfake* kao nadnesenu krizu ili puknuće koje bi moglo prouzročiti nadzemaljsku destrukciju civilizacije i kulturnih praksi kakve poznajemo. Drugi autori u tom smislu *deepfake* označuju kao nadolazeću distopijsku komunikacijsku ekologiju (Yadlin-Segal i Oppenheim, 2020. , prema *ibid.*) ili kao dio nepoželjne aplikacije odlučivanja o svakodnevnom životu preko umjetne inteligencije (Pavis, 2012., prema *ibid.*). No, kao što Cover (*ibid.*) ističe, implikacija „puknuća“ ili „novosti“ negira da je *deepfake* produkcija nastavak različitih društvenih praksi prisutnih prije njegova nastanka. Diskrepancija vjerodostojnosti praksi agencijskog ili institucionalnog provjeravanja činjenica i katastrofičnih vizija javne erozije povjerenja u te dionike ukazuje na evoluirajuće izazove i kompleksnost problematike prepoznavanja *deepfake* sadržaja. Pritom komentar naglašava hitnu potrebu za efikasnije pristupe.

„Mislim kako umjetna inteligencija treba imati žig koji će ljudima govoriti da je slika nastala umjetnom inteligencijom. Inače će ljudi koji koriste umjetnu inteligenciju prevariti ljude da misle kako su lažne slike stvarne. Ova deepfake stvar je uznemirujuća“
(5139)

„Ono što očajnički trebamo je da zakonodavci i tehnološke kompanije svoje kreatore i facilitatore drže odgovornima. Mnogo života će biti uništeno prije nego političari sjednu i ovo ozbiljno promisle“ (4393)

Sugestija kako je potrebno napraviti digitalne žigove na vizualnim sadržajima generiranim umjetnom inteligencijom, odražava proaktivan pristup suočavanja s potencijalno zavaravajućom

⁴¹ Ajder, H., Patrini, G., Cavalli, F. i Cullen, L., „The State of Deepfakes: Landscape, Threats, and Impact.“, 2019. pristupljeno 2.7.2024. https://regmedia.co.uk/2019/10/08/deepfake_report.pdf

prirodom takve produkcije. Ovakva perspektiva prepoznaje potrebu za transparentnom cirkulacijom sadržaja u skladu s vrijednostima iskrenosti i kolokvijalnog shvaćanja autentičnosti. Žig za umjetnu inteligenciju stoga postaje kulturološki marker, mehanizam koji informira i osnažuje pojedince u interakciji s digitalnim medijima. Na sličnom tragu, u drugim pod mrežama, uočavamo poveznicu ovakvih nastojanja s tehnologijom *Blockchain*, no u ovom radu neću ulaziti u dublju analizu takvih prijedloga jer je *Blockchain* prekompleksna tehnologija za pojašnjenje u ovom radu, a procijenio sam kako analiza ove pod mreže ne pruža značajan interpretativni doprinos ukupnom radu.

Sljedeći primjer proaktivnog pristupa u okvirima odabrane pod mreže je komentar:

„Zato pretvorimo duboku državu u memeove i produkcije.“ (12843)

Izjava kako je u cirkulaciju javnog diskursa potrebno uvesti produkciju *deepfake memeova*⁴² na temu *duboke države*⁴³ otvara sloj u diskursu koji u insurekcionističkoj maniri isprepliće *deepfake* s političkim narativima. Takva koncepcija podudara se sa postojećim praksama upotrebe humora za angažman u kompleksnim sociopolitičkim pitanjima. Utilizacija *deepfakea* za produkciju *memeova* označava specifičnu formu kulturne ekspresije s potencijalom subverzije struktura moći, pružajući jedinstvenu leću kroz koju pojedinci navigiraju političkom stvarnosti. Ovakvo tumačenje potrebno je uzeti s oprezom jer je pitanje *duboke države* kontroverzno. S obzirom na emancipirajući ton izjave, upitno je zašto je *Vader* knjižnica generirala negativni sentiment. Stoga je korekcija manualno unesena u podatkovni okvir građe kao pozitivna, shodno mojoj subjektivnoj procjeni. Ovakva žudnja za subverzijom kroz *deepfake*, između ostalog, odražava dva važna kulturološka trenutka: 1. obrat postmoderne tekstualnosti koja uključuje: „*bricolage*, pastiš, hiperrealnost i polisemiju“ (Fiske, 1989., prema *ibid.*:86) i „žudnje za angažmanom sukreacije u kontekstu digitalne interaktivnosti“ (Cover, 2006. prema *ibid.*). Žižek (2002. prema *ibid.*) ističe kako gledateljstvo više ne očekuje autentičnost ili realnost; više ne vidimo realno kao nešto nadređeno digitalnom. U ovom primjeru ostaje otvoreno pitanje pristupa narativu *duboke države* koji ću elaborirati u nastavku.

⁴² Memeovi su humoristični ili informativni sadržaji, često u obliku slika ili kratkih videozapisa, koji se brzo šire putem interneta. ChatGPT, „Definicija *memeova*,“ generirano s OpenAI, 2024.

⁴³ Duboka država je termin koji se koristi za opisivanje navodne skrivene i moćne strukture ili organizacije unutar vlade ili druge institucije koja navodno manipulira političkim procesima ili donosi odluke izvan javnog nadzora ili demokratske kontrole. ChatGPT, „Definicija *deep statea (duboka država)*,“ generirano s OpenAI, 2024.

„Šećući“ po povezanim čvorovima odabrane pod mreže, nailazimo i na negativno konotirani čvor „news“:

„Ovo je demografska skupina 'Real Raw News', već utjecajni dio republikanske baze. Oni gutaju deepfake vijesti...“ (13704)

Gornji primjer navodi „Real Raw News“ kao utjecajan medij republikanske baze na području SAD-a. Komentator iznosi stav kako ovoj demografskoj skupini nedostaje kritička sposobnost razlučivanja lažnog sadržaja te oni stoga „jedu deepfakeove“. Ovakav stav ukazuje na ulogu deepfake sadržaja u oblikovanju i poticanju političkih narativa u specifičnim demografskim grupama. Utjecaj ovakvog sadržaja na politički aktivnu bazu naglašava potencijal da se poljulja stavove glasača unutar specifičnih ideoloških zajednica. Ova opservacija preklapa se s Chakhoyanovom⁴⁴ (2018., prema ibid.:82) perspektivom. On predlaže da deepfake sadržaji nose potencijal intenzifikacije postojeće političke polarizacije kakvu vidimo u suvremenim demokracijama, primjerice tijekom izbora u SAD-u 2016. Simbiotska relacija deepfakea i političke kulture, kakva je prikazana u ovom primjeru, podudara se i sa širom javnom debatom koju je opisao Cover i drugi (2022., ibid.:80). Diskusija je centrirana oko implikacija tehnologije na reprezentaciju istine, autentičnosti i potencijalnog doprinosa kulturi dezinformacije, posebice u procesima glasanja.

U ovoj analizi kontrastirat ću pozitivno i negativno konotirani sadržaj posljednjih dvaju komentara u kontekstu političke polarizacije u SAD-u (uzevši u obzir da se nalaze u istoj pod mreži međusobno povezanih pojmova). Sadržaj prvog primjera predlaže moguću subverziju „duboke države“ kroz produkciju deepfakeova. Drugi primjer sugerira nedostatak kritičke prosudbe unutar republikanske desno orijentirane baze. Prihvatanje kontroverznog urotničkog narativa „duboke države“ može se stoga pripisati široko rasprostranjenom dojmu u medijskim reprezentacijama. Taj dojam inherentno povezuje takve narative s ekstremnom desničarskom frakcijom republikanske baze. Urotnički pogled na svijet često zanemaruje suptilno razumijevanje struktura moći. On previđa čimbenike poput kontingencije i metafizike moći, kako ih razumije Foucault. Zanemarivanje ovih čimbenika otežava dublje razumijevanje igre moći. Dok drugi primjer kritizira

⁴⁴ Chakhoyan, D., „Deep fakes could threaten democracy. What are they and what can be done?“ World Economic Forum. 2018., pristupljeno 2.7.2024., www.weforum.org/agenda/2018/11/deep-fakes-maydestroy-democracy-can-they-be-stopped/

republikansku bazu kao naivne potrošače *deepfakeova*, važno je prepoznati da kritizirana pozicija, unatoč mogućem propustu u mikrofizici moći, izražava odgovore na opipljive učinke dinamike moći unutar kapitalizma. Narativima povezanima s kontroverznom sintagmom *duboka država* pristupam kao mitopoetičkim trenutcima. Trenucima koji izražavaju stvarnosti ukorijenjene u životnim iskustvima boli, nejednakosti i nepravdi karakterističnih za kapitalističke sustave. Poput mitova prošlosti iz kojih dešifriramo društvenu dinamiku, urotnički se narativi mogu tumačiti kao suvremeni mitovi čije dešifriranje pruža uvide u društvene dinamike i odnose moći. Pristupanje urotničkim narativima kao mitopoetičnim trenutcima u kulturno antropološkoj analizi nosi potencijal dekonstrukcije okidača političke polarizacije. Ovaj pristup, izbjegavajući simplifikaciju, stigmatiziranim grupama pristupa kao kompleksnim pojedincima, otkrivajući zajednička ljudska iskustva. Analizirajući urotničke narative kao izraze stvarnih iskustava, umjesto jednostavnog označavanja kao laži, potiče razumijevanje i nadilazi antagonizme u suvremenom društvu.

Ovakva perspektiva podudara se s primjerom alternativnih kulturoloških pristupa teorijama zavjere kakvu zastupaju Boullier, Kontras i Siles (2022). Urotničke narative moguće je artikulirati kao „situirane ekspresije kulturnih bojazni (Dean, 1998; Goldberg, 2001; Knight, 2000.) ili, kako bi Aupers (2012) rekao: 'radikalne i generalizirane manifestacije nepovjerenja usidrena u kulturološkoj logici modernosti (ibid: 5).“ Jedan je primjer ovakve analize Harambama (2020) dvogodišnja etnografska studija u Nizozemskoj. Za njega su urotnički narativi kultura u smislu da reprezentiraju „širi društveni konflikt oko znanja i istine u suvremenim društvima, namećući propitivanje onoga što smatramo legitimnim znanjem i zašto“ (Harambam, 2020., prema ibid.). Drugim riječima, ovakvi narativi povezani su sa „širim nesuglasticama o autoritetu i vjerodostojnosti institucija u produkciji znanja i istine (Harambam i Aupers, 2015., prema ibid.)“.

Boullier, Kontras i Siles (ibid.:11-13) navode tri ključne strategije u ovakvoj analizi. Kao prvo, odbacivanje implicitne normativnosti predmeta istraživanja. Oni tvrde kako su izrazi „teorije zavjere“ i “urotništvo” generičke oznake koje služe diskreditiranju suprotstavljenih stajališta i diskvalificiranju dionika doživljenih kao devijantnih u odnosu na dominantnu normu. Govoreći o urotničkim narativima kao „neizvjesnom znanju“ autor nastoji izbjeći *apriori* diskvalifikaciju izjava koje se istražuju.

Druga strategija predlaže da se diskurse označene kao urotničke ne promatra kao homogenu kategoriju, već kao set pregovaranih epistemologija situiranih u kontinumu koji je kritičan prema institucijama odgovornim za definiranje istine. Pritom se nadovezuje na izjavu Steve Shapina (2019, prema *ibid.*:11) koji tvrdi kako naša društva ne proživljavaju u tolikoj mjeri „krizu istine” koliko krizu povjerenja u institucije, posebice znanstvene. Naglašava kako ove kritike ne zauzimaju ne znanstveni stav; dapače, Harambam i Aupers (2015, *ibid.*:12) skeptičnim su pristupom uspješno ukazali kako je specifično uvježbavanje znanstvenika, da svijet vide na određeni način, dovelo do toga da se određene interpretacije *apriori* odbacuju. U vezi s odbačenim interpretacijama, Shapin (2019., prema *ibid.*) zaključuje kako nepovjerenje u znanstveni autoritet mora biti shvaćeno u relaciji s „povećanim ispreplitanjem između znanosti i dvije velike sile: 'masovnog poslovanja' i 'države'” (Boltanski i Thévenot, 1991. prema *ibid.*).

U konačnici, treća strategija odnosi se na otvaranje prostora u analizi za promišljanje sociotehnoloških uvjeta proboja i cirkulacije „urota”. Pritom se slaže sa tvrdnjom France (2019., prema *ibid.*) koji smatra da je poveznicu učinka interneta na urotničke narative potrebno pažljivo proučiti u svrhu izbjegavanja simplističkih i ahistorijskih interpretacija.

U vezi s posljednjom strategijom i mojim prikupljanjem građe, važno je istaknuti *Twitter Files* skandal⁴⁵. *Twitter* datoteke (*Twitter Files*), kolekcija internih dokumenata i diskusija između zaposlenika bivšeg vodstva firme *Twitter*, objavljeni su u devetnaest dijelova, pružajući uvide u odluke tvrtke vezane za prakse moderiranja sadržaja korisnika platforme. Naime, nakon što je Elon Musk preuzeo *Twitter*, u javnosti obećanom nastojanju da razotkrije ideološki motivirano cenzuriranje sadržaja, omogućio je pristup internim datotekama tvrtke trojici novinara: Mattu Taibbiju, Bariju Weissu i Michaelu Shellenbergeru. Navedeni novinari analizom datoteka otkrivaju interne debate na određene teme, poput cenzure Trumpovih komentara 6.01.2021., alata korištenih za upravljanje prekršajima politike firme, odluke u vezi supresije priča o skandalu oko Hunter Bidenova laptopa te interferencije FBI-ja u interne odluke tvrtke. Bari Weiss ističe tajne akcije vodstva firme u ograničavanju vidljivosti određenih korisnika, često vezane za konzervativna stajališta republikanaca. Matt Taibbi pak naglašava izvanredne mjere supresije priče

⁴⁵ „The Defaming of Brandon Straka and #Walkaway,” *Twitter Files*, Substack, pristupljeno 2.7.2024, <https://twitterfiles.substack.com/>.

o Hunter Bidenovu laptopu (predstavnik demokrata), kojim su se blokirale poveznice i direktne poruke u vezi priče o Hunter Bidenovu laptopu (mjere rezervirane za krajnje slučajeve poput dječje pornografije). Šesti i sedmi dio baca svjetlo na FBI-jevo kontinuirano i opsežno uplitanje u odluke tvrtke od 2018. *Twitter* je predstavljen kao doušnik FBI-ja, pri čemu se ističe suradnja u svrhu navodnih dezinformacija. FBI-jeva široka definicija dezinformacija potkrijepljena je primjerima gdje su humoristični komentari s niskim dosegom publike cenzurirani. U javnim saslušanjima suđenja u vezi skandala često je korišten argument kako je FBI intervenirao zbog navodnih pokušaja ruskih *botova* da utječu na američke izbore. Bitno je istaknuti, kao što Schenberger tvrdi, da je FBI isplaćivao milijune dolara tvrtki za obavljanje posla. Iako je Musk ovaj skandal artikulirao kao prekršaj slobode govora, kritičari se pozivaju na činjenicu da se sloboda govora ne odnosi na usluge koje je pružila privatna tvrtka, a FBI tvrdi kako nisu postavljali ultimatum (usp. Wikipedia, 2023.). *Twitter* datoteke stoga ukazuju na delikatni balans između slobode govora i državne intervencije, otvarajući bitna pitanja o učestalosti i prirodi državne inteferencije na platformama poput *Twittera* (je li poveznica masovnog poslovanja i države teorija urote ili društvena stvarnost?).

Skandal s *Twitter* datotekama naglašava ključne odlike određenih narativa, poput pejorativne oznake *duboke države* kao zavjere, istovremeno otkrivajući poveznice s dinamikama političke stvarnosti u SAD-u. Dodatno, nastojanje koje uključuje politički pristranu intervenciju vlade u način poslovanja privatne društvene platforme moguće je protumačiti kao manifestaciju psiholoških operacija (*PSYOPS*⁴⁶). Psihološke operacije su u medijskim reprezentacijama u području SAD-a najčešće povezane s republikanskom bazom biračkog tijela SAD-a. U analitičkom okviru poseban je naglasak potrebno staviti na detaljno proučavanje korisničkih komentara relevantnih za tu temu, radi sveobuhvatnijeg razumijevanja. U nastavku ću se osvrnuti na ovakve primjere.

„Trebali biste više čitati povijest, provjeriti oči zbog komunizma. Bolujete od psihološke operacije, ništa od ovoga nije stvarno, čak je i Biden AI deepfake.“ (9479)

U odabranom primjeru komentator predlaže povijesnu neinformiranost primatelja poruke te nagovještaj komunističke prijete povezan s potencijalno preuveličanom tvrdnjom prezasićenosti

⁴⁶ Psihološke operacije (psyops) uključuju uporabu psiholoških tehnika za utjecanje na stavove, uvjerenja i ponašanje, često primijenjene u vojnim ili vladinim kontekstima. ChatGPT, „Definicija *psyopsa* (psihološke operacije),“ generirano s OpenAI, 2024.

javnosti sadržajima produciranim u kontekstu psiholoških operacija. Stav je korisnika kako ništa nije stvarno. Negativno konotirana sugestija kako je čak i Biden „*deepfake*” ne samo da propituje autentičnost političkih figura već implicira i namjernu manipulaciju javnog mijenja. Poveznica psiholoških operacija s komunizmom podcrtava ideološku dimenziju i povijesni kontekst u kojem su političke ideologije pretvorene u oružje za psihološki utjecaj nad određenom populacijom.

„*Pripremite se za porast antisocijalizma.*” (14985)

S druge strane, nailazimo na komentare koji sugeriraju suprotan učinak, pri čemu će *deepfake* u kontekstu psiholoških operacija biti korišten za propagandu protiv socijalnih mjera države. I u ovom primjeru moguće je uočiti odraz šireg ideološkog konteksta karakterističnog za SAD. Frakcije jedne strane (homogenizirane medijske reprezentacije), socijalne mjere poput javnog zdravlja doživljavaju kao štetnu interferenciju države i naganjanje pema prijetnji komunizma. Druga strana (homogenizirana medijska reprezentacija) zastupa stajalište kako su socijalne politike nužnost. Suprotstavljene se pozicije upisuju u *deepfake* kao svojevrsan amplifikator već postojeće polarizacije SAD-a. Pritom je bitno istaknuti njegov potencijal da označava kontradiktorne narative istovremeno. U radu ne namjeravam dublje ulaziti u dekonstrukciju ovakvog prijevora, već ga istaknuti kao uzorak koji se ponavlja u građi. U građi su prisutna i skeptična stajališta poput donjeg primjera:

„*Svi su razotkriveni. Nijedna AI deepfake maska nije dovoljno velika da ih sakrije.*

PSIHOLOŠKA OPERACIJA OSUĐENA NA PROPAST.” (20924)

Korisnik sugerira propast psiholoških operacija, čime implicitno prihvaća ovakvu praksu kao normalizirani aspekt stvarnosti. Svi su odabrani komentari u vezi psiholoških operacija negativno konotirani.

U vezi mog iščitavanja rezultata sentimenta i mreže supojavnosti *hashtagova*, bitno je istaknuti kako Raymond Williamsov (1977.) koncept *strukture osjećaja* pruža vrijedan okvir za razumijevanje rezultata sentimenta upisanih u fenomen *deepfakea*. *Struktura osjećaja* prati emocionalne aspekte stavova i raspoloženja koji određuju specifičan povijesni trenutak. Odnosno to je način da se shvati osjećaj određenog povijesnog trenutka, prije nego što se oblikuje u definiranije ideološke ili institucionalne oblike. Ovakav pristup odražava kolektivnu svijest i relacije unutar određene kulture. Kao što sam dosad u radu pokazao, a što ću u nastavku produbiti,

deepfake doprinosi promjenjivim, često ambivalentnim emocionalnim krajobrazima izazivajući uspostavljene norme istinitosti i autentičnosti. Emocionalni odgovori na *deepfake*, koji variraju od zabave do anksioznosti, odražavaju razvijajući sentiment šireg sociokulturnog miljea.

Uzimajući u obzir fenomen psiholoških operacija kojim se manipulira psihološkim faktorima radi vršenja utjecaja na pojedince ili grupe, *struktura osjećaja* preuzima stratešku dimenziju. Osjećaji su pretvoreni u oružje koje oblikuje percepcije i ponašanja, odražavajući se na kolektivno raspoloženje unutar zajednice. Intersekcija *deepfake* tehnologije, *Twitter* datoteka i psiholoških operacija čine kompleksan emocionalni teren, gdje se osjećaji povjerenja i skepticizma isprepliću. Razumijevanje *strukture osjećaja* u eri *deepfake* i informacijskog ratovanja⁴⁷ uključuje dekodiranje emocionalnog tkanja koji predstoji javnim reakcijama i institucionalnim mjerama. Kombinirana metoda analize sentimenta i mreža supojavnosti *hashtagova* uz kvalitativno iščitavanje i manualne korekcije rezultata bitan je element u ovakvu dešifriranju. Pritom je analizi prijepora, kakvi se javljaju u ovakvoj građi, korisno pristupiti u okvirima strategija *neizvjesnog znanja* (jer trenutni pristupi ignoriraju povezanost tržišta i vlasti, odnosno moć lišavaju odgovornosti).

Iako se generirana mreža pretežno sastoji od neutralnih i negativno konotiranih čvorova, u pozitivnom spektru mreže nalazimo čvorove poput: *future, innovation, cybersecurity, balanciagapope, agi, etichsinai, futureofart, law, Tiktok, socialmedia, aprilfoolsday*. Ovi primjeri proizvode narative koji priznaju potencijalne pozitivne aspekte korištenja *deepfake* tehnologije, naglašavajući dvojaku percepciju unutar diskursa.

„*Deepfake ili photoshop, ali možda i pogled u budućnost.*“ (9954)

„*Suočavanje s prijjetnjom deepfakea ključno je za očuvanje integriteta naše političke rasprave i izgradnju sigurnije budućnosti za sve građane, kako bi se osjećali sigurno ne samo u svakodnevnom životu nego i u kibernetičkom prostoru.*“ (21813)

Primjer koji označava *deepfake* kao pogled u budućnost podudara se s uvidima analize sentimenta prve podmreže koji ovu tehnologiju predstavlja u okvirima transformativne sile, s kapacitetom da preoblikuje različite aspekte kreiranja sadržaja digitalnih medija. Istovremeno,

⁴⁷ Informacijsko ratovanje je strategija koja koristi informacije i komunikacijske tehnologije kako bi se postigao vojni ili politički cilj, često uključujući širenje dezinformacija, manipulaciju javnim mišljenjem ili napade na informacijske sustave suparnika. ChatGPT, „Definicija *informacijskog ratovanja*,“ generirano s OpenAI, 2024.

poziv da se suočimo s prijetnjom *deepfakea* zauzima proaktivan stav, naglašavajući nužnost očuvanja političkog diskursa i sigurne budućnosti kako u fizičkom tako i u digitalnom svijetu. Ovakav se stav, naglašavajući sjecište tehnološkog napretka, sigurnosnih pitanja i društvenih vrijednosti, podudara s Taylorovom (2021.) analizom. Taylor (2021., prema Cover, 2023:81) utvrđuje eskalaciju javnog i akademskog diskursa koji promovira sekuritizaciju *deepfake* tehnologije kroz državnu regulaciju. No, što uopće mislimo pod „sigurnošću“ (uzevši u obzir gornje uvide o intereskciji *deepfakea*, *Twitter fileova* i fenomena psiholoških operacija)? Koji su mogući učinci promišljanja „sigurnosti“?

„Ne smijemo dopustiti da ovaj zakon prođe, on uništava naše korištenje interneta, vrlo moćnog alata. Imajte na umu da još nije donesen savezni zakon za uklanjanje deepfake stranica streamera/slavnih. Umjesto da se fokusiraju na to, oni se fokusiraju na zabranu svega što smatraju 'nesigurnim'..“ (15056)

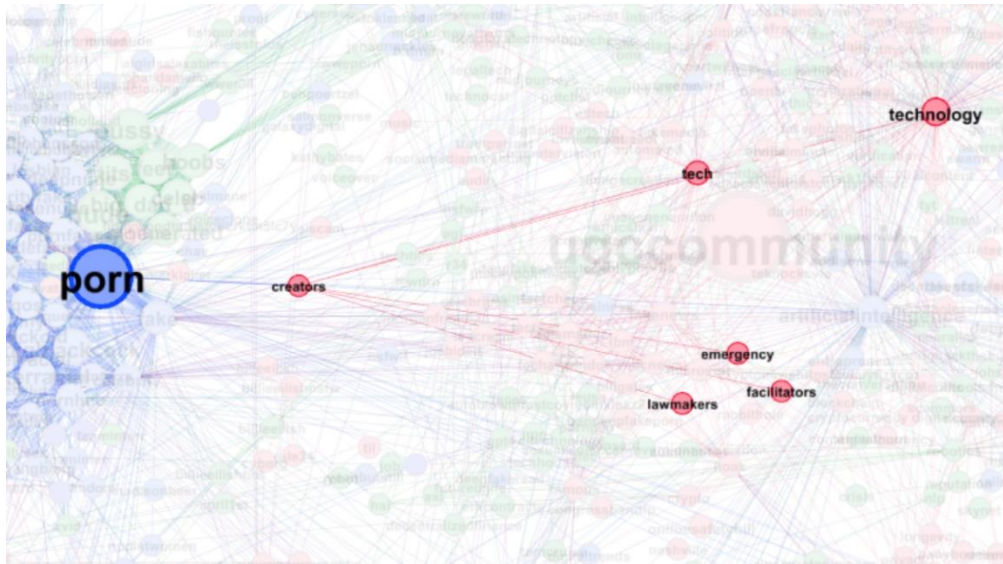
Gornji primjer uvodi kritičku perspektivu u javni diskurs o sigurnosti i regulaciji *deepfakea* koji se podudara s alternativnom perspektivom u akademskom diskursu kakvu zastupa Flew. Prema Flew (2021, prema *ibid.*) umjesto da ovakvu tehnologiju percipiramo kao vanjski entitet koji iziskuje regulativne mjere poput cenzure i uklanjanja sadržaja, što bi je potencijalno moglo transformirati iz nečeg štetnog u ilegalno, predloženo je *deepfakeu* pristupiti kao inherentnoj komponenti kulture. Cover (*ibid.*) stoga zaključuje, kako je *deepfake* karakteriziran kao „kulturni artefakt“ i „kreativna praksa“ u organskom nastajanju unutar kulturnog miljea. Drugim riječima, pristup regulaciji *deepfakea* kao novoj/stranoj tehnologiji koju treba ukinuti (što je tehnološki gotovo nemoguće) ne doprinosi artikulaciji regulacija koje će spriječiti upotrebe koje su društveno nepoželjne. Ovo se podudara sa stajalištima iz prethodnih primjera koji ističu višeznačan utjecaj *deepfakea*, ne samo na političke narative već i na ključne aspekte društvene reprezentacije. Ovakva je tehnologija nastavak različitih društvenih praksi koje su postojale i prije njezine pojave.

Posljednji primjer uvodi čvor slavnih ličnosti u korelaciji s čvorom budućnosti, koji pripada istoj pod mreži.

„ još nije donesen savezni zakon za uklanjanje deepfake stranica streamera/slavnih “

Čvor slavnih ličnosti je pozicioniran iznimno blizu čvora pornografije, dok je znatno udaljeniji od čvorova koji su vezani za zamjenu lica nepoznatih osoba sa slavnim ličnostima. Bliskost u

ovom slučaju sugerira da se u raspravama na *Twitteru* puno češće ponavlja poveznica između slavnih ličnosti i pornografije, što je vidljivo u donjem Prikazu 4.. Na Prikazu 4., pomaknuo sam se s oznake slavnih ličnosti na povezani čvor kreatora. U supojavnostima čvorova vezanih za kreatora, vidljivo je kako su u *Twitter* raspravama *deepfake* kreatori često povezani s dominantnom temom pornografije. Blizina čvoru pornografije sugerira često ponavljanje ove veze. Poveznica sa tehnologijom, zakonodavstvom, hitnoćom i facilitatorima te negativna konotacija sugeriraju kako je ovo u raspravama percipirano kao značajan problem.



Prikaz 4. vizualizacija mreže supojavnosti *hashtagova* čvor: „women“

To potvrđuje i sljedeći komentar isčitani iz odabranih čvorova:

Deepfake AI pornografska industrija operira nam pred očima. (7774)

Pritom se korisnik referira na internetski članak o mogućnostima jeftine kupovine *deepfake* pornografije.

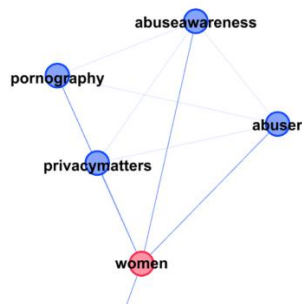
„Digitalno uređeni pornografski videozapisi s licima stotina žena koje nisu dale pristanak privlače desetke milijuna posjetitelja na web stranicama, od kojih se jedna nalazi na vrhu rezultata pretraživanja na *Googleu*. Osobe koje stvaraju ove videozapise naplaćuju svega 5 dolara za preuzimanje tisuća isječaka s licima

slavnih osoba, pri čemu prihvaćaju plaćanje putem *Visa* kartica, *Mastercarda* i kriptovaluta.“⁴⁸

U okvirima generiranog grafa razvidno je kako su slavne ličnosti u najvećoj mjeri povezane s pornografijom i sa širokim dijapazonom pornografskih fetiša koji dominiraju graf. Prevlast pornografskog sadržaja podudara se s nalazima Wangine⁴⁹ (2019. prema *ibid.*:89) studije, koja je utvrdila da je više od 90% internetskih *deepfakeova* pornografskog sadržaja. Čvor je pornografije u kontekstu rezultata analize sentimenta iznimno heterogen; podjednako su zastupljeni negativno pozitivno i neutralno konotirani sadržaji. Kvalitativnim iščitavanjem komentara potvrdio sam si da, iako etički upitan, mnogi korisnici ovakav pornografski sadržaj doživljavaju pozitivno. U stotinama komentara, u „moru“ pornografskih *deepfakeova* slavni ženskih glumica u heteroseksualnom snošaju, primjeri pornografskih uradaka sa slavnim muškim glumcima bili su rijetkost. Komentari koji su se ticali pornografskih *deepfakeova* s muškim slavim ličnostima odnosili su se isključivo na snimke homoseksualnog snošaja cis-rodni muškaraca (Colin Farrell, Tom Cruise, Liam Hemsworth, Henry Cavill, Marki Plier). U akademskom diskursu, u vezi s temom pornografskih *deepfakeova*, duže je vrijeme općeprihvaćena teza kako su ovakvom produkcijom u najvećoj mjeri pogođene žene (*ibid.*:82). Bilo u kontekstu pornografskih uradaka slavni ličnosti ili osvetničke pornografije u kojoj bivši partneri (uglavnom muškarci) umeću lica svojih partnerica u pornografske uratke s pornografskim glumicama, pa čak i pedofilskih uradaka prikupljanjem *Tik-Tok* videa maloljetnih individua ubačenih u porno uratke porno glumica *petite* kategorije (kategorija punoljetnih glumica koja fiziološki nosi sličnosti s maloljetnim), nema sumnje kako su najpogođenija skupina žene. U ovom dijelu rada, kada se referiram na muškarce ili žene, mislim na njihov spol, a ne nužno i rod.

⁴⁸ Brandy Zadrozny, „Found through Google, bought with Visa and Mastercard: Inside the deepfake porn economy,“ NBC News, 13.4.2023, pristupljeno 2.7.2024. <https://www.nbcnews.com/tech/internet/deepfake-porn-ai-mr-deep-fake-economy-google-visa-mastercard-download-rcna75071>.

⁴⁹ Wang, C., „Deepfakes, revenge porn and the impact on women.“ Forbes, 1.11.2019., pristupljeno 2.7.2024., www.forbes.com/sites/chenxiwang/2019/11/01/deepfakes-revenge-porn-and-the-impact-on-women/?sh=6092fe8a1f53.



Prikaz 5. vizualizacija mreže supojavnosti *hashtagova* čvor: „women“

Ovakvu predodžbu podupiru i brojni komentari u prikupljenoj građi koji ističu ranjivost žena u vezi s *deepfakeovima*. U Prikazu 5. vidljiva je manja mreža koja povezuje oznaku žene s zlostavljačima, pornografijom, ali i potrebom za povećanjem svijesti o problemu zlostavljanja te poštovanjem privatnosti žena u digitalnom dobu. Iako kalkulacije sentimenta na prvi pogled djeluju problematično, detaljnim pregledom građe postaje očito kako su neutralno konotirani čvorovi povezani s porukom o potrebi podizanja svijesti o zlostavljanju. Što je vidljivo u donjem primjeru.

„Još više razloga zašto ljudi trebaju razmišljati o svojoj privatnosti na društvenim medijima. #pornografija #žene #važnoprivatnost #zlostavljač #svijestoozlostavljanju.“ (4688)

Pritom je oznaka žena procijenjena kao negativna, što se podudara s građom jer se najčešće javlja u negativno konotiranim raspravama gdje korisnici mreže ističu zabrinutost za njihovu ranjivu poziciju. Ovaj primjer je vrijedno istaknuti, jer da ga nisam povezo s pregledom građe, moglo bi se naslutiti da je procjena sentimenta kriva ili da je u raspravama prisutan govor mržnje prema ženama.

„Kozmetičke“ mjere regulacije ovakvih sadržaja, primjerice cenzura *deepfake* pornografskih uradaka na *PornHubu*⁵⁰, mjerom kojom je *deepfake* zabranjeno upisati u tražilicu, ali dozvoljeno objaviti s drugačijim oznakama i izvedenicama, ukazuje na prevlast monetizacijskih interesa naspram etičkih. Ovakvom iščitavanju ide u prilog komentar povezanog čvora.

⁵⁰ „Pornhub promised to ban deepfakes videos and it failed miserably,“ *The Next Web*, pristupljeno 2.7.2024., <https://thenextweb.com/news/pornhub-promised-to-ban-deepfakes-videos-and-it-failed-miserably>

„Industrija pornografije - sviđalo se to vama ili ne - pionir je mnogih značajki koje imamo na internetu. Slike, videozapisi, e-trgovina i društvene mreže sve su izgrađene od strane njih. Nije iznenađujuće vidjeti ovo, što je zabrinjavajuće jer nisu pioniri etike.“ (19963)

Ovakva dinamika ukazuje na *deepfake* kao kulturološki amplifikator postojećih odnosa moći u kontekstu binarnih kategorija muškarac - žena te monetizacijskih aspekata ovakve dinamike. Simetriju između *deepfake* produkcije i postojećih odnosa moći u odnosu muškarac - žena nemoguće je poreći, no nameće se pitanje kako interpretirati marginalne primjere muških žrtava. Uzevši u obzir da su primjeri muških žrtava isključivo homoseksualnog karaktera podudara se s nalazima Ramsayine (2017:16) studije u kojoj je tematskom analizom polustrukturiranih intervjua razotkrila kako su ispitane heteroseksualne žene u SAD-u uglavnom uživale u muškoj homoseksualnoj pornografiji. S obzirom na to da *Twitter* API ne nudi mogućnost ekstrakcije informacija o rodu ili spolu korisnika izjava objavljenih na platformi, teško je bez sumnje konstatirati sudjeluju li u ovakvoj produkciji i konzumaciji samo muškarci (heteroseksualni i homoseksualni) ili i statistički podzastupljene heteroseksualne žene. U svakom slučaju, čak i u primjerima pornografskih *deepfakeova* gdje su meta muškarci, oni su postavljeni u kontekst muške homoseksualnosti koja je učestalo medijski reprezentirana kao društveno marginalizirana skupina.

Bitno je napomenuti kako u grafu čvor slavne ličnosti gravitira prema pornografskom roju, ali je, uvjetno rečeno, također povezan s nepornografskim sadržajima poput politike. Ovakvi primjeri uključuju satirične sadržaje poput falsificirane izjave Morgana Freemana kako je Joe Biden *budala* jer priča o sladoledu nakon incidenta pucnjave u Nashvillu⁵¹. S obzirom na sadržaj (isticanje apsurdnog ponašanja predsjednika SAD-a), nameće se pitanje zašto je kreator *deepfakea* posegnuo baš za licem Morgana Freemana? Mogući odgovor na ovo pitanje leži u pristupu koji lice promatra kao društvenu konstrukciju. Kao što bi Deleuze i Guattari rekli: „lice je politika“ (1987:181). Ako lice tumačimo kao društvenu konstrukciju, otvaramo mogućnost da mu pripišemo svojstva poput idiomatičnosti. Morgan Freeman jedan je od rijetkih glumaca u Hollywoodu koji je imao priliku reprezentirati kršćanskog Boga, primjerice u filmu „Bruce Almighty“ (2003.). U serijalu „The Story of God with Morgan Freeman“ (2016.) lik Morgana Freemana istražuje religije u kontekstu različitih kultura, pri čemu je kao narator preuzima ulogu etičkog autoriteta. Teško je empirijski

⁵¹ „Fact check: Biden, the Nashville shooting, and ice cream," DW, pristupljeno 2.7.2024., <https://www.dw.com/en/fact-check-bidens-gaffe-about-ice-cream-and-nashville/a-65159845>.

dokazati ovakvu tvrdnju, no moguće je kako je zbog naravi uloga koje Morgan Freeman preuzima njegovo lice postalo idiom za etičnost. Stoga odabir baš njegova lica u svrhu kritiziranja neprimjerene izjave američkog predsjednika može biti protumačeno kao logičan odabir za intertekstualni pastiš.

U istoj pod mreži nailazimo i na primjere koji su nepornografski i nepolitički:

„Jeste li znali da je Viggo Mortensen zapravo umro kada je udario u kacigu u Dvije kule i da su preostale scene Aragorna snimljene s Nicolasom Cageom i najmodernijom tehnologijom deepfakea?“ (22248)

Iako lice Nicolas Cagea nikada nije bilo iskorišteno u scenama *Kralja prstenova*, a glumac Viggo Mortensen još nije preminuo, ovaj fiktivan primjer progovara o mogućnostima kreativne upotrebe *deepfakea* u filmskoj industriji. Prema Alexanderu, „tipična kreativna upotreba za profesionalnu filmsku produkciju pokušaj je da se stvore scene koje djeluju kao da je glumac u određenom prostoru, u određenoj radnji koja se naizgled uklapa u snimljeni film, što uključuje i prikaze preminulih glumaca“ (2017., prema Cover, 2023:85). Na sličnom tragu Mihailova (2021, prema *ibid.*) ističe i primjere upotrebe tehnologije u muzejskog produkciji, gdje se publici predstavljaju naizgled autentični preminuli autori. Cover (*ibid.*:90) zaključuje kako ovakva upotreba tehnologije odstupa od dominantnog narativa o *deepfake* tehnologiji isključivo kao prevari. Za njega je ona odraz tehnologije koja ispunjava kreativne i kulturne potrebe. No čak i kada govorimo o nepornografskoj, nezavaravajućoj i nepolitičkoj upotrebi *deepfakea*, daljnji primjeri u pod mreži dodatno kompliciraju ovakav stav.

„Novi chatbot koji koristi "deepfake" tehnologiju za pretvaranje teksta u glasove poznatih osoba. Sa 25 dostupnih glasova poznatih ličnosti do sada i planovima za proširenje na preko 100, mogućnosti su beskrajne.“ (16232)

Upotreba *deepfake* tehnologije u javno dostupnim aplikacijama podudara se s uvidima kako je ovakva tehnologija producirana u kontekstu „ranije uspostavljenih potreba društva, uključujući one u filmskoj industriji u svrhu poboljšanja video efekta.“ (Naruniec i dr., 2020. prema *ibid.*:86) te šire društvene potrebe za intertekstualnom igrom, humorom, kreativnošću kao i nepoželjnim kreativnim praksama poput dezinformacije i prevare (*ibid.*:80). Ovakva je upotreba *deepfakea*, slično kao s ranije obrađenim primjerom prijedloga produkcije subverzivnih *deepfake memova*,

usklađena sa širim kulturološkim obratom u kojem žudnja za hiperrealnim prevladava. Na prisutnost ovakve potrebe ukazuju i daljnji primjeri u pod mreži koji uključuju čvorove poput *music*, *grunge*, *90s nostalgia*, *kurt cobain*, *madonna*, u kojem korisnici najčešće s pozitivnim sentimentom označavaju umetanje svog lica na snimke popularnih ličnosti. Ovakvi primjeri tvore zasebnu skupinu koju je moguće podijeliti na dvije vrste: aplikacije za produkciju *deepfakeova* gdje se u određene scene umeću lica korisnika ili, kao što je slučaj s gornjim primjerom, produkcija *deepfakeova* gdje se manipulira glas ili lice poznate osobe. Činjenica da se ovakve aplikacije monteraziraju otvara bitna pitanja o tome kako u eri generativne umjetne inteligencije pristupiti regulaciji i zaštiti lica. Pitanje lica kao osobnih podataka u kontekstu europskih legislativa, poput GDPR-a, ostaje nedefiniran problem.

S gornjim primjerom prekinuo sam iščitavanje mreže i njezinih pod mreža.

5. Zaključak

Prikupljenu građu ovog rada sačinjavaju 27 804 unosa objavljenih na društvenoj mreži *Twitter* u razdoblju od 26. ožujka 2023. do 2. travnja 2023. Polazna pretpostavka kako je uz pomoć strojne obrade jezika moguće provesti kvalitativnu analizu ukupne građe u okvirima pristupa digitalne antropologije i digitalne etnografije u najboljem je slučaju moguće okarakterizirati kao ambiciozan pokušaj, a u najgorem kao opasnu tendenciju ka kompjutorski potpomognutom generaliziranju kompleksne društvene pojave. Mrežna vizualizacija građe omogućila mi je uvid u guste relacije značenja koje su stvorili korisnici društvene platforme u vezi pojma *deepfake*, ali sam, unatoč bogatstvu tih relacija, morao prihvatiti vlastitu subjektivnost te odbaciti mnoge veze i odabrati i privilegirati određene. Nažalost, uslijed privatizacije *Twittera* izgubio sam mogućnost daljnjeg prikupljanja podataka. Generiranje odvojenih mreža u dužem vremenskom intervalu pružilo bi mi mogućnost uočavanja promjena koje ovise o širem društvenom kontekstu. Osim toga, leksička se analiza sentimenta pokazala iznimno problematičnom. Kalkulacije pojedinačnih izjava bile su nepouzdanе, što je utjecalo i na kalkulacije prosječnih vrijednosti sentimenta vezanih za pojedinačne *hashtagove*. Manualnim iščitavanjem uočio sam da su određeni rezultati bili donekle ispravni, ali kategorizacija sentimenta korisničkih izjava na pozitivne, negativne i neutralne, u mnogo slučajeva bila je nemoguća. Nailazio sam na primjere koji negativnim tonom progovaraju o pozitivnim težnjama ili pozitivnim tonom progovaraju o negativno konotiranim pojavama poput

nekoncezualne *deepfake* pornografije. Isprobavao sam i naprednije modele koji su bolje prepoznavali ironiju i sarkazam, ali rezultati nisu pružali značajan doprinos interpretaciji. Ipak, kontrastiranje subjektivne interpretacije i kalkulacija sentimenta pružilo mi je reflektivne uvide o problematičnosti mehanizama koji se koriste u društveno-tehnološkoj svakodnevnici za procjenu javnog mišljenja, zadovoljstva kupaca i emocionalnih reakcija na proizvode, usluge ili događaje. Osim toga, u usporedbi s drugim građama koje sam prikupljao, *hashtagovi* u vezi *deepfakea* nisu bili lingvistički raznoliki. U drugim građama bilo je moguće uočiti razlike u upotrebi složenica, pejorativa, oznaka za institucije, stranke, osobe, konkretne i apstraktne imenice. *Deepfake* je u tom smislu specifičan jer je njegovo postojanje reducirano na digitalno.

No, kao što Daniel Miller (2018:2) ističe, digitalna antropologija istovremeno proučava upotrebu tehnologije i njezine implikacije. Stoga način istraživanja u ovom radu, unatoč manjkavostima, može pomoći u razvijanju kompleksnije analize određenih pojava u suvremenom društvu. Metodološka poveznica strojne obrade jezika, komputacijskih alata, teorija diskursa i teorija grafova mreža, otvara nove mogućnosti za digitalnu antropologiju i etnografiju. Ovakav pristup omogućava analiziranje kako jezik konstruira značenje i društvenu stvarnost, proširujući reflektivnost na digitalne i tehnološke čimbenike. Kao što Sarah Pink (2016:2) tvrdi, digitalna etnografija postaje značajna tek kada je primijenjena unutar određenih disciplinarnih ili interdisciplinarnih paradigmi. Društvene mreže, kao dinamični tereni značenja i praksi, zahtijevaju metodologije koje nadilaze koncepcije linearnosti teksta. Platforme poput *Twittera*, normalizacijom participacije, mjerenjem popularnosti i algoritmičkim kuriranjem vidljivosti sadržaja, pokazuju kako politika i moć djeluju unutar tehnološke infrastrukture. Kontrastiranje vidljivosti kakvu nudi platforma s grafovima mreža može pružiti vrijedne uvide učinka društvenih platformi. Nažalost, privatizacija takvih medija nameće pitanje po koju cijenu kao društvo plaćamo *besplatne* usluge.

Deepfake tehnologija sa svojim implikacija u vezi autentičnosti i povjerenja također potvrđuje potrebu za ovakvim analizama. Kontroverze koje *deepfake* producira ukazuju na društveno pregovaranu koncepciju stvarnosti i reprezentacije u digitalnoj eri.

Moja analiza *deepfakea* otkriva određene ključne aspekte vezane uz ovu tehnologiju i njezine učinke u suvremenom društvu. Distinkcija između sadržaja generiranog umjetnom inteligencijom

i tradicionalnog autorskog sadržaja izaziva ideju jedinstvenog autora, pomičući fokus na sâm sadržaj i njegove nove modalitete postojanja i cirkulacije. Umjetna inteligencija kao nehumani agent omogućuje stvaranje sadržaja koji odražava unesene podatke i kreativnost pojedinca, proširujući problematiku tekstualnosti digitalnih interakcija na neljudske sudionike. Performativni pristup identitetu u kontekstu *deepfake* tehnologije ukazuje na potencijal za kreativnost i izražavanje, kompromitirajući tradicionalno shvaćanje autentičnosti. Unatoč novitetu *deepfake* tehnologije, ona se može promatrati kao nastavak postojećih društvenih praksi, ističući složenost prepoznavanja takvog sadržaja. Proizvodnja *deepfake memeova* u političkom kontekstu integrira humor u društveno-politički angažman. Žudnja za subverzijom posredstvom *deepfakea* odražava ključne kulturne trenutke poput prevlasti hiperrealnosti i želje za sukreacijom u digitalnom kontekstu. Označavanje *teorija zavjere* služi diskreditiranju suprotstavljenih stajališta i marginalizaciji devijantno percipiranih sudionika rasprava. *Deepfake* doprinosi političkoj polarizaciji u SAD-u kroz političku instrumentalizaciju. *Deepfake* tehnologija pridonosi promjenjivim emocionalnim krajobrazima, izazivajući reakcije od zabave do anksioznosti. Intersekcija *deepfake* tehnologije, *Twittera* i psiholoških operacija stvara složen emocionalni teren gdje se povjerenje i skepticizam isprepliću. Umjesto da se tehnologija *deepfake* vidi kao vanjski entitet koji zahtijeva cenzuru, može se percipirati kao inherentni dio kulture. Ova dinamika naglašava *deepfake* kao pojačivač postojećih odnosa moći, posebno u kontekstu rodni binarnosti i aspekata monetizacije, iako korištenje tehnologije odražava i kreativne i kulturne potrebe. *Deepfakom* su posebno pogođene žene, no po tom pitanju se ne poduzimaju ikakve preventivne mjere. Mjere za političke deepfakove su razrađene ali u praksi oni pretežno predstavljaju potencijalnu opasnost. Istraživanjem diskursa društvenih mreža o pojmu *deepfake* tehnologije možemo bolje razumjeti šire implikacije digitalnih tehnologija na društvo i kulturu, otvarajući put za bolje razumijevanje isprepletenosti percipiranih prijetnji.

6. Literatura

- Agamben, Giorgio. 2011. *Nudities*. Kalifornija: Stanford University Press.
- Belting, Hans. 2017. *Face and Mask: A Double History*. New Jersey: Princeton University Press.
- Boullier, H., Kotras, B. i Siles, I. 2021. *Uncertain Knowledge. Studying “Truth” and “Conspiracies” in the Digital Age*. RESET .
- Bruns, A., i Burgess, J. 2011. *The Use of Twitter Hashtags in the Formation of Ad Hoc Publics*. Australia: Queensland University of Technology Brisbane/ARC Centre of Excellence for Creative Industries and Innovation.
- Bucher, T. 2018. *If...Then: Algorithmic Power and Politics*. Oxford Studies in Digital Politics. London: Oxford University Press.
- Coates, J. 2017. „So 'Hot' Right Now: Reflections on Virality and Sociality from Transnational Digital China.“ *Digital Culture & Society. Mobile Digital Practices* 3 (2): 77–97.
- Cover, R. 2023. *Identity and digital communication: Concepts, theories, practices*. London: Routledge.
- Cristiano Garcia. 2024. *Temporal Analysis of Drifting Hashtags in Textual Data Streams: A Graph-Based Application*. arXiv.
- Deleuze, G. 1995. *Negotiations. 1972-1990*. New York: Columbia University Press.
- Deleuze, Gilles, i Félix Guattari. 1987. *A Thousand Plateaus: Capitalism and Schizophrenia*. Minnesota: University of Minnesota Press.
- Deuze, M. 2012. *Media life*. London: Cambridge.
- Doss, C., Mondschein, J., Shu. D., Wolfson T., Kopecky D., Fitton-Kane VA, Bush L., Tucker C. 2023. *Deepfakes and scientific knowledge dissemination*. New York: Sci Rep.
- Edkins, Jenny. 2015. *Face Politics*. London: Routledge.
- Entman, R. M. 2007. *Framing Bias: Media in the Distribution of Power*. *Journal of Communication*, 57(1), 163-173.
- Foucault, Michel. 1994. *Nadzor i kazna: Rađanje zatvora*. Zagreb: Informator.

- Foucault, Michel. 1969. *LANGUAGE, COUNTER-MEMORY, PRACTICE* Selected Essays and Interviews. New York: Cornell University Press.
- Hage, P. 1979. Graph Theory as a Structural Model in Cultural Anthropology. *Annual Review of Anthropology*, 8(1), 115–136.
- Hage, P., Harary F. 1984. *Structural Models in Anthropology*. Cambridge: Cambridge University Press.
- Hall, S. 1997. The Work of Representation. In *Representation: Cultural Representations and Signifying Practices*. Los Angeles: Sage Publications.
- Harcourt, Bernard E. 2007. *An Answer to the Question: „What Is Poststructuralism?“* Chicago: University of Chicago Public Law & Legal Theory.
- Hatzivassiloglou, Vasileios, and Kathleen R. McKeown. 1997. "Predicting the Semantic Orientation of Adjectives." In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics (ACL-1997)*, 174-181.
- Ichau, Elke. Frissen, Thomas i Haenens, Leen. 2019. From #Selfie to #Edgy. Hashtag Networks and Images Associated with the Hashtag #Jews on Instagram. California: Telematics and Informatics.
- Jenkins, D. 2008. *Anthropology, Mathematics, Kinship: A Tribute to the Anthropologist Per Hage and His Work with the Mathematician Frank Harary*. Los Angeles: Human Complex Systems.
- LIU BING. 2015. *Sentiment Analysis Mining Opinions, Sentiments, and Emotions*. Chicago: University of Illinois.
- Loukianov, A., Burningham, K. i Jackson, T. 2022. The patterning of the discursive space in search for the #goodlife: A network analysis of the co-occurrence of Instagram hashtags, *The Information Society*, 39(1), 62–78.
- Macho, Thomas. 1996. *Medium Gesicht. Die faciale Gesellschaft. Ästhetik & Kommunikation*. Berlin: Elefant-Press Berlin.
- Mackenzie, A. 2006. *Cutting code: Software and sociality*. New York, NY: Peter Lang.

- Meraz, S., & Papacharissi, Z. 2013. "Networked gatekeeping and networked framing on #Egypt." *The International Journal of Press/Politics* 18(2):138-166.
- Meraz, Sharon. 2017. "Hashtag Wars and Networked Framing." In *Between the Public and Private in Mobile Communication*, edited by Antonio Sánchez Tellería, 303-323. New York: Routledge.
- Mihailova, Mihaela. 2021. "To dally with Dalí: deepfake (inter)faces in the art museum." *Convergence: The International Journal of Research into New Media Technologies* 27(4):882-898.
- Marwick, Alice, and danah boyd. 2011. "To See and Be Seen: Celebrity Practice on Twitter." *Convergence: The International Journal of Research into New Media Technologies* 17(2):139-158.
- Miller, D. 2018. *Digital anthropology*. London: The Open Encyclopedia of Anthropology.
- Paris i Donovan. 2019. *Deepfakes and Cheep fakes*. Data and Society.
- Pink, S., Horst, H., Postill, J., Hjorth, L., Lewis, T., Tacchi, J. 2016. *Digital Ethnography: Principles and Practice*. Los Angeles: SAGE.
- Ramsay, Guy. 2017. "Straight Women Seeing Gay Porn: 'He's Too Good Looking!'" *Porn Studies* 4, no. 2: 157–175.
- Rose, G. 2001. *Visual Methodologies: An Introduction to the Interpretation of Visual Materials*. London: Sage Publications.
- Seaver, N. 2017. Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society*, 4(2)
- Seaver, N. 2019. "Knowing Algorithms." In *Digital STS: A Field Guide for Science & Technology Studies*, edited by J. Vertesi and D. Ribes, 412–422. Princeton: Princeton University Press.
- Thrift, N. i French, S. 2002. *The Automatic Production of Space*. The Royal Geographical Society.
- Vaccari, Cristian i Chadwick, Andrew. 2020. *Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News*. *Social Media + Society*.
- Williams, Raymond. *Marxism and Literature*. Oxford University Press, 1977.
- Yancey, Stricker. 2024. *The Dark Forest Anthology of the Internet*. Metalabel.

Ziewitz, Malte. 2016. "Special Issue Introduction: Governing Algorithms: Myth, Mess, and Methods." *Science, Technology, & Human Values* 41, no. 1: 3–16.

Internetski izvori:

"Twitter Files." Wikipedia, Wikimedia Foundation, pristupljeno 3.7.2024., https://en.wikipedia.org/wiki/Twitter_Files.

7. Prilozi

Prikaz 1. mreža sa tri čvora

Prikaz 2. vizualizacija mreže supojavnosti hashtagova, čvor: „ugccommunity“

Prikaz 3. vizualizacija mreže supojavnosti hashtagova čvor: „technology“

Prikaz 4. vizualizacija mreže supojavnosti hashtagova čvor: „women“

Prikaz 5. vizualizacija mreže supojavnosti hashtagova čvor: „women“