



TOWARDS STATISTICAL FOUNDATIONS FOR DETECTING COORDINATED INAUTHENTIC BEHAVIOR ON FACEBOOK

Institute for Data,
Democracy & Politics

THE GEORGE WASHINGTON UNIVERSITY

David A. Broniatowski

TOWARDS STATISTICAL FOUNDATIONS FOR DETECTING COORDINATED INAUTHENTIC BEHAVIOR ON FACEBOOK

David A. Broniatowski

Department of Engineering Management and Systems Engineering
Institute for Data, Democracy, and Politics
The George Washington University
Washington, DC 20052
broniowski@gwu.edu

May 12, 2021

ABSTRACT

In this document we propose a novel method to detect coordinated inauthentic behavior in Facebook pages and groups. This method modifies a recently published technique [1], with the intention of establishing a firm statistical foundation for defining “simultaneous link sharing” behavior.

1 Introduction

Facebook’s community standards disallow “coordinated inauthentic behavior”, stating:

In line with our commitment to authenticity, we don’t allow people to misrepresent themselves on Facebook, use fake accounts, artificially boost the popularity of content, or engage in behaviors designed to enable other violations under our Community Standards. This policy is intended to protect the security of user accounts and our services, and create a space where people can trust the people and communities they interact with. (https://www.facebook.com/communitystandards/inauthentic_behavior)

Nevertheless, prior work suggests that this type of behavior is widespread on the platform. Building on this work [1], we seek to develop a technique that may be used to identify coordinated activity based on “near-simultaneous link sharing”. The intuition is that if two entities (Facebook Pages or Groups) routinely share the same URL at roughly the same time, these two entities are coordinated.

2 Approach

Our approach was to characterize the average time between successive shares of the same URL across multiple datasets, and therefore, to operationalize “near-simultaneous” link sharing in a manner that was robust to the specific query being used.

2.1 Datasets

We collected data from CrowdTangle,¹ for both Facebook Pages and Facebook Groups. We conducted multiple queries:

1. Posts containing the keyword “wellness” collected on March 31, 2021

¹CrowdTangle is a Facebook-owned tool that tracks interactions on public content from Facebook pages and groups, verified profiles, Instagram accounts, and subreddits. It does not include paid ads unless those ads began as organic, non-paid posts that were subsequently “boosted” using Facebook’s advertising tools. It also does not include activity on private accounts, or posts made visible only to specific groups of followers.

2. Posts containing vaccine-related keywords² but excluding keywords pertaining to pets and guns³ collected on March 31, 2021
3. Posts containing keywords pertaining to vaccine-preventable illnesses (the “vaccine stream” [2])⁴ collected on March 31, 2021
4. Posts pertaining to several different health conditions (the “health stream” [3])⁵ collected on March 31, 2021
5. Posts containing keywords pertaining to COVID-19 [4]⁶ collected on March 31, 2021
6. Posts collected from three separate blank searches collected on March 30, 2021 at 9:45pm EDT, 11:25pm EDT, and March 31, 2021 at 9:00am EDT. We checked for, and did not find, overlap between these three datasets.

CrowdTangle queries return a maximum of 300,000 posts per query. To ensure comparability between datasets for pages and groups for a given set of keywords, we removed all Page posts that dated earlier than the earliest group post and vice versa.

2.2 Model fitting

For each keyword query, we identified all posts containing the same URL, and calculated the interarrival time – i.e., the time difference – in seconds, between that URL’s successive shares. The distribution of interarrival times for each dataset is shown in Figure 1.

Treating each second as an opportunity for a given URL to be shared by an entity (group or page), we expect interarrival times to follow a Poisson distribution. We therefore used the REBMIX [5] package in R to fit a mixture of six Poisson distributions to each query. The number of components was selected by visual inspection of Figure 1.

Although there were some differences between the component values for blank searches and keyword searches, results show separate components with mean interarrival times of roughly 14 seconds and 40 seconds across all search queries (Table 1).

Averaging across queries, we constructed a mixture of two Poisson distributions with $\mu_1=14.15$, $\mu_2=39.94$, and respective weights of 59.25% and 40.75%. Per this mixture model, we define “near simultaneous” link sharing behavior as any pair of URLs that are shared within at most 25 seconds of one another. This number is comparable to the threshold defined by [1], but appears to be consistent across several search queries, including blank searches. More conservative estimates can also be generated, e.g., by focusing on the sub-second components found within all but the wellness and health queries.

²vaccine, vaxx, vaccines, vaccination, vaccinated, vax, vaxxed, vaccinations, jab

³gun, dog, cat, foster, adopt, shelter, vet, kennel, pet, chicken, livestock, kitten, puppy, paw, cow

⁴vaccine, vaccines, mmr, tdap, flushot, hpv, polio, rotavirus, chickenpox, smallpox, hepatitis, hepa, hepB, dtap, meningitis, shingles, vaccinate, vaccinated, vaccine, vaccines, vacine, vacines, tetanus, diphtheria, pertussis, whooping cough, dtp, dtwp, chicken pox, measles, mumps, rubella, varicella, diphtheria, haemophilus, papillomavirus, meningococcal, pneumococcal, rabies, tuberculosis, typhoid, yellow fever, immunizations, immunization, immunization, immune, imune, cholera, globulin, encephalitis, lyme, zika, zeka, zeeka, zica, zeca, zeeca, h1n1, cika, sika, cica, sica, h7n9, bird flu

⁵ill, sick, cold, body, pain, hurts, sore, nose, hospital, doctor, cancer, killing, stomach, headache, neck, ear, throat, chest, hurting, ouch, massage, burning, flu, exhausted, medicine, surgery, knee, cough, fever, doctors, insomnia, irritated, freezing, intense, emergency, dose, miserable, exercise, cure, eaten, dentist, vision, bedtime, physical, treatment, pills, coma, pounds, dealing, breathing, insurance, feelin, tooth, heal, appointment, ache, ankle, pill, numb, recovery, physically, wrist, depression, hungover, allergies, allergic, nurse, stroke, , cramps, woken, muscles, dizzy, clinic, pains, jaw, sneeze, lungs, swollen, puke, anxiety, appt, recover, severe, headaches, thirsty, vomit, tension, sneezing, caffeine, itchy, appetite, resting, coughing, infection, diabetes, migraine, sickness, uncomfortable, pounding, mild, aching, itching, hiccups, forehead, illness, recovering, hurtin, ribs, medication, aches, stuffy, advil, sneezed, symptoms, prescription, nyquil, drained, asthma, lung, anxious, itch, remedy, elbow, infected, sinus, kidney, allergy, torn, rash, chronic, tumor, poisoning, pimples, crutches, diagnosed, tylenol, nauseous, stiff, bladder, splitting, fatigue, lump, bruised, puking, germs, sunburn, relieve, runny, rehab, paracetamol, panadol, stomachache, watering, faint, toothache, icky, blisters, throbbing, veins, dehydrated, spine, heartburn, dental, nausea, needles, watery, puffy, yucky, surgeon, colds, antibiotics, vomiting, skull, shivering, acne, sniffles, healed, throats, painkillers, contagious, vitamins, stomach, strep, tiredness, benadryl, sinus, congestion, ibuprofen, withdrawal, arthritis, migraines, pneumonia, recovered, cured, cravings, tonsils, ulcer, remedies, limping, fluids, aspirin, battling, swelling, snot, unbearable, drowsy, acl, coughs, exhaustion, steroids, tissues, unwell, diarrhea, hives, hacking, congested, mri, chemo, achy, drugged, scratchy, dieting, sicker, sprain, h1n1, spasms, excedrin, infections, codeine, pollen, backache, nasal, bronchitis, seasonal, ick, viral, sunscreen, vaccine, gastric, coughed, vicodin, inhaler, excruciating, groggy, motrin, bypass, clogged, morphine, dizziness, acupuncture, dehydration, glands, earache, migrane, vertigo, ulcers, lasik, thyroid, feverish, achey, hayfever, respiratory, tonsillitis, sniffing, aleve, fevers, wheezing, ugh

⁶coronaoutbreak, coronavirus, covid19

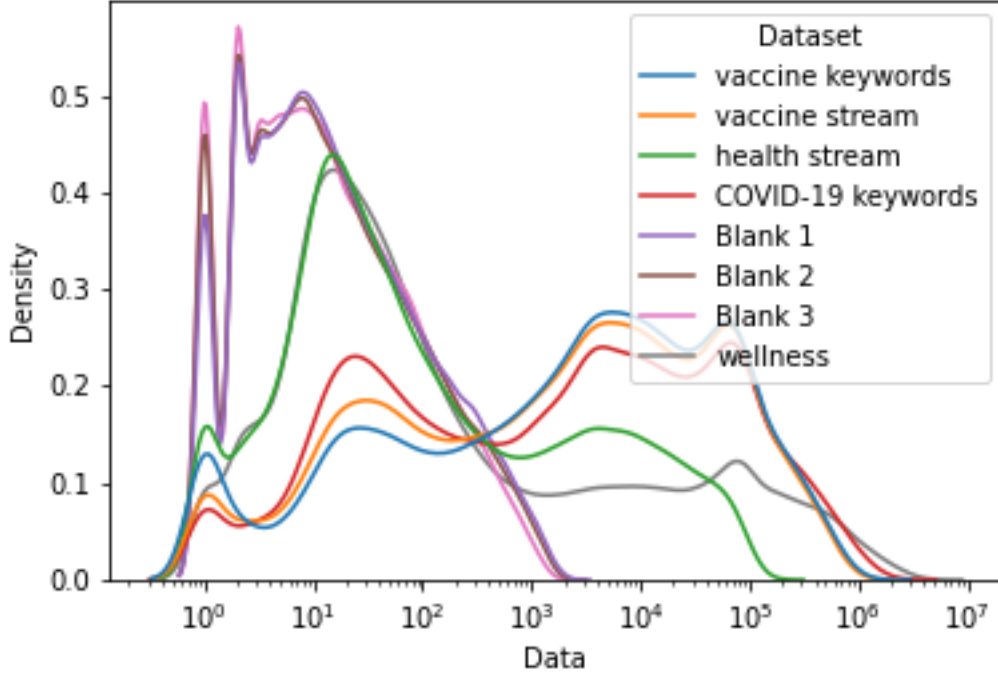


Figure 1: Density plots of interarrival times for each dataset. A one-second smoothing parameter has been added.

Table 1: Poisson Distribution Components

Wellness	Vacc. Keywords	Vacc. Stream	Health	COVID-19	Blank 1	Blank 2	Blank 3
	0.38 (22)	0.57 (15)		0.65 (12)	0.70 (19)	0.34 (10)	0.33 (10)
3.66 (29)			2.42 (26)		3.87 (31)	2.38 (35)	2.36 (35)
					13.92 (10)	7.00 (10)	7.08 (10)
14.92 (27)	13.31 (30)	12.93 (33)	12.72 (29)	13.21 (37)	15.72 (17)	15.32 (22)	15.06 (21)
38.45 (22)	43.65 (19)	42.44 (22)	35.24 (23)	40.77 (22)	41.85 (13)	39.17 (14)	37.96 (14)
80.40 (11)	97.67 (12)	92.15 (13)	78.78 (10)		91.20 (10)	82.23 (9)	79.01 (9)
142.00 (7)	178.52 (9)	166.33 (10)	140.12 (7)	152.99 (9)			
265.26 (4)	377.35 (8)	376.84 (7)	225.97 (5)	245.57 (7)			

2.3 Constructing coordinated networks

We consider a pair of entities coordinated if they are more likely to share links in a coordinated manner than in a non-coordinated manner. To make this determination, we first used the derived Poisson mixture model to calculate the log-odds ratio, $\log(P_1(x)) - \log(P_2(x))$, that any link that was shared at least twice was shared in a coordinated manner, where

$$P_n(x) = e^{-\mu_n} \frac{\mu_n^x}{x!}$$

and x is the amount of time between successive shares of the same URL. For each pair of entities in our dataset, we next created a list of interarrival times, and corresponding log-odds ratios, between successive URL shares. Finally, we used t-tests to test the hypothesis that the average log-odds ratio of coordinated link sharing was greater than zero. Edges between entities were retained if they were significant at the $p < 0.05$ level after controlling for multiple comparisons using the Holm-Bonferroni procedure.

2.4 Results

An example of a coordinated network for the keyword “wellness” is shown in Figure 2.

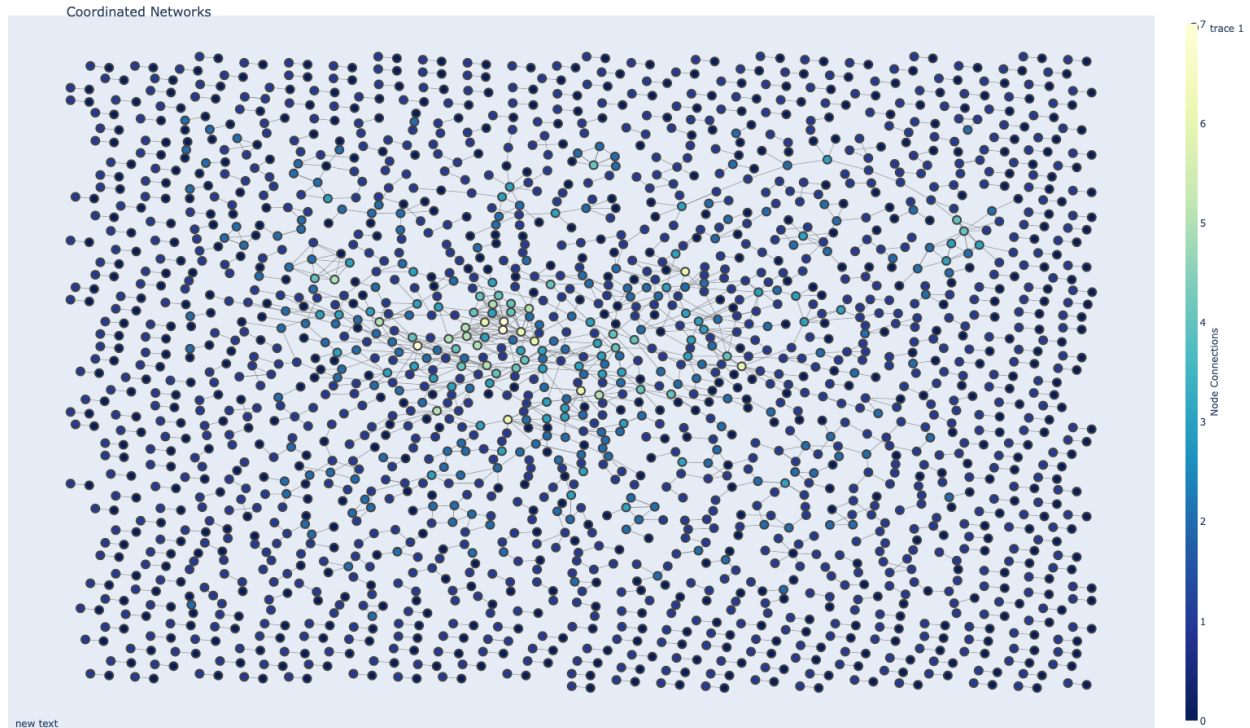


Figure 2: Coordinated network associated with “wellness” keywords. The largest cluster contains several Indonesian and Malaysian health and wellness groups. Other clusters include German health and beauty groups and several Facebook gamers groups.

Table 2 shows 5 nodes with the largest degrees – i.e., most coordinated connections – in each graph. In each case, we see high degrees of redundancy between entity names, or variations on a theme. For example, vaccine and COVID-19 networks seem to contain several seemingly local news organizations that are nevertheless posting URLs in a coordinated manner, suggesting top-down content dissemination.

References

- [1] Fabio Giglietto, Nicola Righetti, Luca Rossi, and Giada Marino. It takes a village to manipulate the media: coordinated link sharing behavior during 2018 and 2019 italian elections. *Information, Communication & Society*, 23(6):867–891, 2020.
- [2] Mark Dredze, David A Broniatowski, Michael Smith, and Karen M Hilyard. Understanding vaccine refusal: why we need social media now. *American journal of preventive medicine*, 50(4):550, 2016.
- [3] Michael J Paul and Mark Dredze. Discovering health topics in social media using topic models. *PLoS one*, 9(8):e103408, 2014.
- [4] David A. Broniatowski, Daniel Kerchner, Fouzia Farooq, Xiaolei Huang, Amelia M. Jamison, Mark Dredze, and Sandra Crouse Quinn. Debunking the misinfodemic: Coronavirus social media contains more, not less, credible content. *arXiv:2007.09682 [physics]*, Jan 2021. arXiv: 2007.09682.
- [5] Marko Nagode. Finite mixture modeling via rebmix. *Journal of Algorithms and Optimization*, 3(2):14–28, 2015.

Table 2: Selected Highly-Connected Entities in Each Coordinated Network

Search Terms	Selected Entitles
Vaccine Keywords	Micromedia Publications The Brick Times The Manchester Times The Jackson Times The Toms River Times
Vaccine Stream	Current in Zionsville Current in Geist Current in Westfield Current in Noblesville Current in Carmel
Health Stream	Pimple Popping Videos - The BEST Pimple Poppers Delight Pimple Popping Videos The Best Pimple Popping, Blackheads, Cyst Removal Pimple Popping Video and Acne Treatment
COVID-19 Keyword	Gi National News: Gi Dudley News Gi Leicester News Gi Oxford News Gi Aberdeen News
Blank	Hoi Tàì Xiu Game Bài (Ask Small Game Cards) Hoi Tai Xiu Game Bai Tp.Hcm Hoi Tai Xiu !! !!Game Bai Hoi Tàì Xiu [Telephone Emoji] Game Bài Hoi Tàì Xiu Game Bài [Grimacing Face Emoji] [Grimacing Face Emoji] [Grimacing Face Emoji]
Wellness	PENDUDUK SETIAWANGSA/AMPANG/KUALA LUMPUR (RESIDENTS OF SETIAWANGSA/AMPANG/KUALA LUMPUR) group iklan kecantikan & kesihatan (beauty & health ad group) warga Keramat, Setiawangsa (residents of Keramat, Setiawangsa) PENDUDUK AMPANG/KERAMAT/SETIAWANGSA/MELAWATI/WANGSA MAJU/SETAPAK/HULU KLANG (RESIDENTS OF AMPANG/KERAMAT/SETIAWANGSA/MELAWATI/WANGSA MAJU/SETAPAK/HULU KLANG) warga sentul <3 <3 (sentul citizens <3 <3)