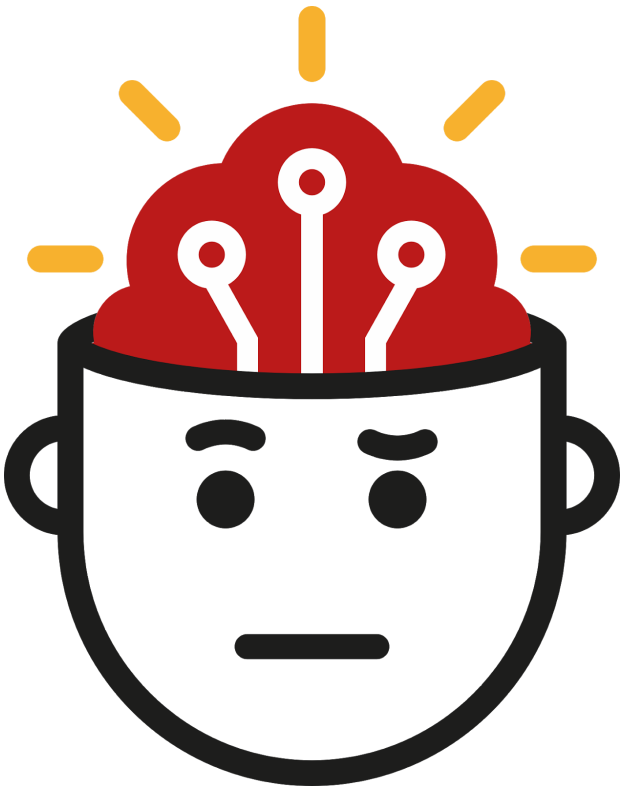
Verification Handbook

Disinformazione e manipolazione dei media

La guida definitiva per indagare sulle piattaforme e sugli account online e rivelare attività non autentiche e contenuti manipolati.



A cura di Craig Silverman

DataJournalism.com

European Journalism Centre

Craig Newmark Philanthropies

La traduzione italiana del Verification Handbook è stata realizzata da

Slow News.

slow-news.com

con il sostegno di

pagella politica

pagellapolitica.it

FACT.A

facta.news

Indice

A. Indagare sulla disinformazione e sulla manipolazione dei media

Scritto da Craig Silverman

***Craig Silverman** è il media editor di BuzzFeed News, per cui è responsabile della copertura a livello globale di temi riguardanti piattaforme, disinformazione online e manipolazione dei media. Ha curato il "Verification Handbook" e il "Verification Handbook for Investigative Reporting," ed è l'autore di "Lies, Damn Lies, and Viral Content: How News Websites Spread (and Debunk) Online Rumors, Unverified Claims and Misinformation."*

Nel dicembre del 2019, l'utente di Twitter @NickCiarelli condivise un video che, stando a quanto affermava, mostrava un balletto in cui i sostenitori della campagna presidenziale di Michael Bloomberg erano soliti cimentarsi. Lo scarso entusiasmo e la coreografia poco brillante del video fecero subito guadagnare al tweet una valanga di like e retweet, la maggior parte dei quali da parte di persone che si divertivano a riderci sopra. Alla fine il video fece oltre 5 milioni di visualizzazioni su Twitter.

Secondo la sua biografia su Twitter, Ciarelli lavorava come stagista per la campagna di Bloomberg. I suoi tweet successivi contenevano elementi che lo provavano, ad esempio lo screenshot di una email ricevuta da un presunto membro dello staff della campagna di Bloomberg che approvava il budget per il video.

Tuttavia, bastava cercare rapidamente il nome di Ciarelli su Google per scoprire che la persona in questione è un comico che in passato aveva già realizzato video parodistici. E quella email da parte di un membro dello staff di Bloomberg? Era stata inviata da Brad Evans, spesso spalla comica di Ciarelli. Anche questa informazione era alla portata di una semplice ricerca su Google.

Nonostante ciò, nei primi minuti e nelle prime ore dalla pubblicazione del tweet, qualcuno aveva creduto che quel video imbarazzante fosse un contenuto ufficiale prodotto dallo staff della campagna di Bloomberg.

Maggie Haberman, importante giornalista politica del New York Times, scrisse in un tweet che i giornalisti che si erano occupati in precedenza delle campagne di Bloomberg come candidato sindaco avevano le loro ragioni per non liquidare subito il video come un falso:

La conoscenza può assumere molte forme, e nel nuovo ambiente digitale i giornalisti devono essere cauti prima di fidarsi troppo di una qualsiasi fonte di informazioni, anche se questa dovesse essere la loro stessa e diretta esperienza.

A quanto pare, alcuni giornalisti che già conoscevano Bloomberg e il suo stile di far campagna elettorale credevano che quel video potesse essere vero. Allo stesso tempo, giornalisti che non sapevano nulla di Bloomberg e che avevano deciso di giudicare il video in base alla sua fonte avrebbero potuto trovare immediatamente risposta ai loro dubbi, semplicemente (in questo caso) cercando su Google il nome della persona che lo aveva condiviso.

La conclusione da tirare non è che essersi già occupati di Bloomberg rappresenti uno svantaggio, bensì che in qualsiasi momento corriamo il rischio di farci sviare da ciò che pensiamo di sapere. E che, in alcuni casi, il nostro bagaglio di conoscenze ed esperienze può persino giocare a nostro sfavore. Inoltre, possiamo rimanere fregati da segnali digitali come retweet, visualizzazioni e tentativi di manipolarli.

Come ha dimostrato il video di Bloomberg, ci vuole poco per creare segnali fuorvianti: ad esempio una biografia su Twitter o lo screenshot di un'email che sembra confermare un contenuto o un'affermazione. Questi elementi, a loro volta, aiutano i contenuti a diventare virali. E più questi elementi accumulano retweet e like, più convinceranno altre persone che il video a cui si riferiscono potrebbe essere vero.

Certo, ci sono esempi molto più subdoli. A differenza di Ciarelli, le persone che stanno dietro operazione informativa e campagne di disinformazione raramente svelano l'inganno. Ma questo caso di studio mostra quanto sia disorientante e frustrante per tutti, giornalisti compresi, navigare in un ambiente informativo pieno di indici di qualità e affidabilità facilmente manipolabili.

La fiducia è la base della società. Orienta e facilita gli scambi economici ed è la chiave dei contatti e delle relazioni tra esseri umani. Ma nell’ambiente digitale è pericoloso farsi guidare da un principio di fiducia a priori. Se dai per scontato che gli account Twitter che retwittano un video lo stiano facendo in maniera spontanea, verrai ingannato. Se pensi che le recensioni di un prodotto siano tutte scritte da veri clienti, sprecherai i tuoi soldi. Se pensi che il tuo newsfeed ti mostri soltanto una selezione imparziale di ciò che più ti serve sapere, finirai per essere disinformato.

Riconoscere questa realtà è importante per tutti, ma per i giornalisti è addirittura essenziale. Siamo presi di mira da campagne coordinate e ben finanziate che mirano a catturare la nostra attenzione, farci amplificare certi messaggi e piegarci alla volontà degli Stati e di altri poteri molto forti.

La buona notizia è che tutto crea opportunità per condurre indagini, nonché il dovere di farlo.

Questo manuale raccoglie le conoscenze e l'esperienza dei migliori giornalisti e ricercatori in circolazione, al fine di fornire indicazioni per condurre indagini sulla manipolazione dei media digitali, sulla disinformazione e sulle operazioni informative.

Lavoriamo in un ecosistema dell'informazione complesso e in rapida evoluzione. Ciò richiede un approccio a sua volta capace di evolvere, fondato sul mettere in dubbio le proprie supposizioni, sulla capacità di individuare e anticipare gli avversari e sull'uso delle migliori tecniche esistenti di investigazione open source e delle più efficaci tecniche di inchiesta tradizionale. Gli elementi di debolezza del nostro mondo digitale basato sui dati richiedono a noi giornalisti di saper mettere in discussione e analizzare ogni aspetto di questo mondo, e di sfruttare le nostre competenze per aiutare il pubblico a orientarsi verso informazioni accurate e affidabili. In questo contesto a noi giornalisti viene anche richiesto di riflettere su come possiamo finire per alimentare inconsapevolmente soggetti mossi da cattive intenzioni e campagne progettate per manipolarci, spingendoci a puntare il dito contro attori statali anche quando non ci sono prove per farlo.

L'obiettivo di questo manuale è fornire ai giornalisti le competenze e le tecniche necessarie per svolgere questo lavoro in modo efficace e responsabile. Al contempo illustra anche elementi basilari sulla teoria, sul contesto operativo e sulla forma mentis che permettono ai giornalisti di produrre un lavoro di alta qualità che informi il pubblico, faccia emergere gli attori mossi da cattive intenzioni e contribuisca al miglioramento del nostro ecosistema informativo. Tuttavia, la prima cosa da capire è che senza il giusto approccio a questo lavoro, strumenti e conoscenze concrete sono inutili.

"Giusto approccio" significa comprendere che nel mondo digitale ogni cosa può essere mistificata e manipolata, e saper riconoscere la grande varietà di persone e realtà che hanno interesse a farlo. Il bello di questo universo è che esiste spesso, sebbene non sempre, una scia di dati, interazioni, connessioni e altri indizi digitali da seguire. E la maggior parte di questi è disponibile in forma pubblica, se si sa dove andare a cercare.

Indagare nel mondo digitale significa non prendere mai nulla per buono. Significa capire che gli elementi che appaiono quantificabili e basati sui dati — like, condivisioni, retweet, interazioni, recensioni di prodotti e click sulla pubblicità – sono spesso, e facilmente, manipolati. Significa riconoscere che i giornalisti sono un elemento chiave nella manipolazione mediatica e nelle operazioni informative, sia quando sono considerati obiettivi da attaccare, sia quando sono il principale canale per diffondere disinformazione e informazioni inaccurate. E significa, per te e per i tuoi colleghi, armarsi della giusta mentalità, delle tecniche e degli strumenti necessari a garantire informazioni accurate e affidabili, evitando di fare da megafono a menzogne, contenuti manipolati e campagne di trolling.

Al centro di questa mentalità c'è il paradosso dell'indagine digitale: non fidandoci di nulla a prima vista, possiamo intraprendere un lavoro che ci porterà a scoprire a cosa non dobbiamo credere e quando invece possiamo farlo. Così produrremo un lavoro di cui le comunità di cui siamo al servizio potranno e vorranno fidarsi.

In aggiunta a questo, ci sono altri principi fondamentali che nel corso dei capitoli successivi e nei vari casi di studio verranno messi ripetutamente in risalto.

- **Pensa come il tuo avversario.** Ogni nuova funzionalità di una piattaforma o di un servizio digitale può essere sfruttata dal tuo avversario. Sapersi mettere nei panni di chi cerca di manipolare l'ambiente informativo per ragioni ideologiche, politiche, finanziarie o di qualsiasi altro tipo è cruciale. Quando hai a che fare con contenuti e messaggi digitali, chiediti sempre per quali ragioni sono stati creati e diffusi. Inoltre, è essenziale rimanere aggiornati sulle ultime tecniche utilizzate da chi manipola e diffonde disinformazione, da chi si occupa di marketing digitale e da altri attori il cui modello di business si basa sulla ricerca di nuovi modi per attirare l'attenzione e guadagnare nel mondo digitale.

- **Concentrati su attori, contenuti, comportamenti e network.** L'obiettivo è analizzare gli attori in campo e i loro contenuti, comportamento e caratteristiche per capire se e come si stiano muovendo in maniera coordinata, come un network. Mettendo insieme e confrontando questi quattro elementi inizierai a capire con che cosa hai a che fare. Come vedrai in più di un capitolo e in qualche caso di studio, iniziare da un singolo contenuto o da una singola realtà, ad esempio un sito web, ed esplorare il contesto in cui si inserisce per risalire dai suoi comportamenti e dalle sue connessioni a una rete più ampia è un approccio decisivo. Metterlo in pratica può voler dire esaminare il flusso dei contenuti e di chi li pubblica su più piattaforme, talvolta in lingue diverse.

- **Monitora e archivia.** Il modo migliore per riconoscere casi di manipolazione dei media e disinformazione è non smettere mai di cercarli. Monitorare e tenere costantemente traccia di argomenti oggetto di dibattito, delle attività di attori conosciuti e di comunità di interesse è essenziale. Conserva e organizza ciò che trovi in fogli di lavoro, raccolte di screenshot o anche utilizzando strumenti a pagamento, come Hunchly.

- **Attento con l'attribuzione.** A volte è impossibile dire esattamente chi ci sia dietro un determinato account, contenuto o a un'ampia operazione di intelligence. Una delle ragioni è che attori con moventi diversi possono essere portati a comportarsi in modo simile e produrre o amplificare gli stessi tipi di contenuto. Persino le piattaforme – che hanno accesso a molti più dati e che dispongono di molte più risorse – commettono errori di attribuzione. La dimostrazione più efficace e convincente di solito è quella che combina prove digitali e informazioni provenienti da fonti interne, un

> mix perfetto di lavoro investigativo online e tradizionale. Questo lavoro sta diventando sempre più difficile, in quanto sia le realtà istituzionali che le altre si evolvono e trovano sempre nuovi modi per nascondere le loro tracce. L'attribuzione è un'operazione complicata: farla in maniera errata minerà tutto il lavoro che ti ha portato fino a essa.

Per concludere, una nota sui due manuali che hanno preceduto questa edizione. Questo lavoro si fonda sulle basi della prima edizione del Verification Handbook e del Verification Handbook for Investigative Reporting. Entrambi forniscono competenze fondamentali per monitorare i social media, verificare immagini, video e account e identificare persone, aziende e altre realtà usando i motori di ricerca.

Molti dei capitoli e dei casi di studio di questo manuale sono stati scritti dando per assodato che chi legge possegga le conoscenze di base illustrate nei due manuali precedenti, in particolare nel primo Verification Handbook. Se fai fatica a seguire, ti consiglio di iniziare dalla prima edizione.

E ora, mettiamoci al lavoro.

B. L’epoca dell’information disorder

Scritto da: Claire Wardle

***Claire Wardle** è a capo della direzione strategia e della ricerca di First Draft, un'organizzazione internazionale non profit che supporta giornalisti, accademici ed esperti di tecnologia nell’affrontare le sfide legate alla fiducia e alla verità nell'epoca digitale. Ha fatto parte dello Shorenstein Center for Media, Politics and Public Policy della Harvard's Kennedy School ed è stata direttrice della ricerca presso il Tow Center for Digital Journalism della Graduate School of Journalism della Columbia University e direttrice dei social media per l'UNHR (l'agenzia delle Nazioni Unite per i rifugiati).*

Bugie, dicerie e propaganda non sono concetti nuovi, come tutti sappiamo. Gli esseri umani hanno sempre avuto la capacità di ingannare, e la storia offre esempi grandiosi di contenuti fabbricati ad arte per trarre in inganno il pubblico, destabilizzare i governi o far volare i mercati azionari. Quel che c'è di nuovo, oggi, è la facilità con cui ognuno di noi può creare contenuti falsi e ingannevoli del tutto credibili e la velocità con la quale questi contenuti riescono a diffondersi per il mondo.

Abbiamo sempre saputo quante complessità ponga il tema dell'inganno. Non esiste un criterio di valutazione standard. Per esempio, una bugia bianca detta per mantenere la pace durante una discussione in famiglia non è paragonabile alla dichiarazione ingannevole di un politico che cerca di ottenere più voti. Una campagna di propaganda sponsorizzata dallo Stato e una teoria cospiratoria riguardante l’atterraggio sulla Luna non sono la stessa cosa.

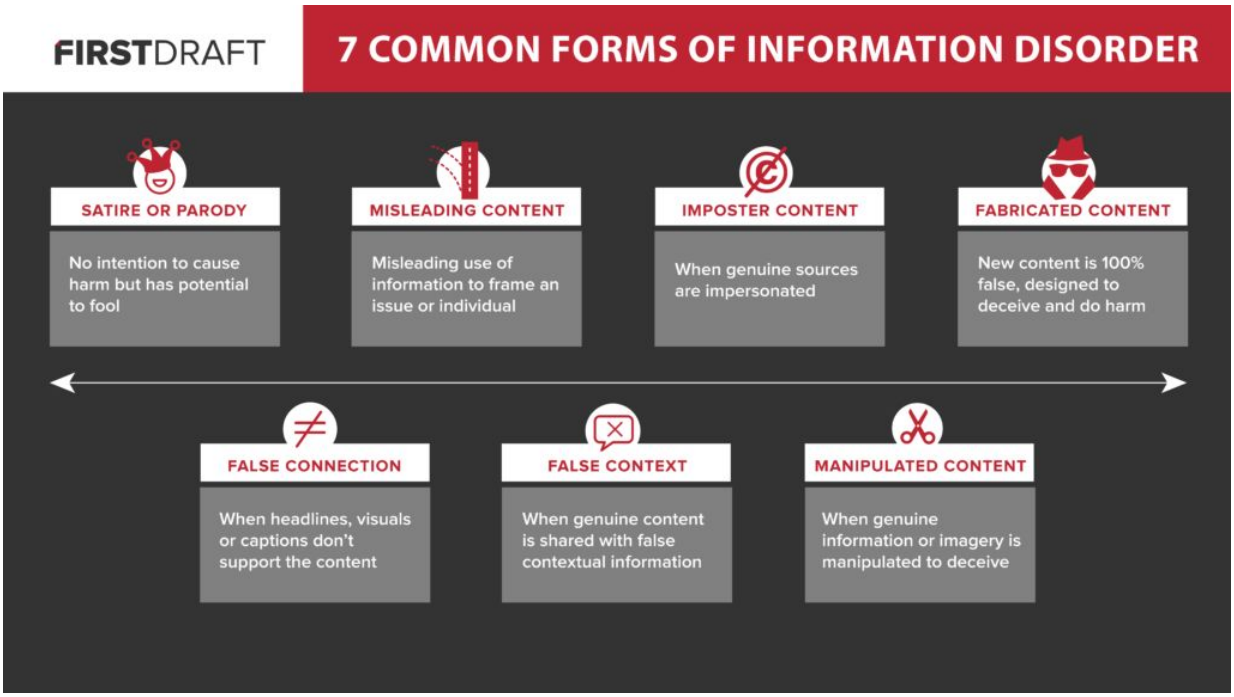
Sfortunatamente, negli ultimi anni tutto ciò che avrebbe potuto essere ricondotto a una delle categorie qui descritte è stato genericamente etichettato come “fake news”, termine semplice che si è diffuso in tutto il mondo, spesso senza bisogno di traduzione. E dico "sfortunatamente" perché si tratta di una formula miseramente inadeguata a descrivere la complessità che ci troviamo ad affrontare. La maggior parte dei contenuti ingannevoli in circolazione non finge nemmeno di essere una vera notizia. Si tratta di meme, video, immagini o attività coordinate su Twitter, YouTube, Facebook o Instagram, la maggior parte dei quali non è nemmeno falsa: sono fuorvianti o, ancor più spesso, autentici, ma usati fuori dal loro contesto.

La disinformazione più potente è quella che contiene un fondo di verità: è quel che accade ad esempio quando si prende qualcosa di vero e gli si dà un'etichetta fuorviante, o quando si diffonde un contenuto di tre anni prima spacciandolo per nuovo. L’aspetto forse più problematico è che il termine “fake news” è diventato un'arma, usata soprattutto dai politici e loro sostenitori per attaccare i mezzi di informazione professionali di tutto il mondo.

La mia frustrazione verso questa espressione mi ha portata, insieme al mio collega e co-autore Hossein Derakhshan, a coniare il termine information disorder. Nel 2017 abbiamo scritto insieme “Information Disorder”, un report in cui ci siamo occupati delle sfide poste dalla terminologia relativa a questo argomento. In questo capitolo illustrerò alcune definizioni chiave necessarie a comprenderlo e ad affrontarlo in maniera critica.

Le 7 categorie dell'information disorder

Nel 2017 ho creato le seguenti categorie per distinguere i diversi tipi di information disorder esistenti.



Satira/Parodia
Molte persone si sono dette comprensibilmente contrarie alla mia scelta di includere la satira in questa categoria, e io stessa ero combattuta all'idea di farlo. Sfortunatamente, però, chi fa disinformazione fa passare di proposito i suoi contenuti come “satira” per assicurarsi che non vengano sottoposti a verifica e per prendere le distanze da qualsiasi possibile danno da essi provocato. In un ecosistema informativo in cui le informazioni di contesto e i segnali, o scorciatoie mentali (euristiche), sono stati rimossi, aumentano le probabilità che i lettori si lascino confondere da un contenuto satirico. Un americano potrebbe sapere che The Onion è un sito satirico, ma sai quanti siti di satira ci sono al mondo? Wikipedia ne elenca 57. Se scorrendo velocemente il feed di Facebook ti passa davanti un sito di satira, è facile cascarci.

Ultimamente, Facebook ha preso la decisione di non verificare la satira, ma chi lavora in questo settore sa bene che l'etichetta "satira" viene usata con una precisa strategia. Ad esempio, nell'agosto del 2019, l'organizzazione americana di debunking Snopes ha scritto un articolo sui motivi per cui sottopone la satira a fact-checking. Un contenuto che si definisce satirico sfugge a chi si occupa di fact-checking e spesso, col tempo, se ne perde il contesto originale: la gente lo condivide e ricondivide pensando che sia vero, senza capire che si tratta di satira.

Collegamento ingannevole
Stiamo parlando di clickbait alla vecchia maniera, ovvero la tecnica di scrivere titoli che contengono affermazioni sensazionali e che, come si scopre, non hanno niente a che vedere con l'articolo o con il contenuto vero e proprio. È facile per i media pensare che il problema della disinformazione sia causato da chi la diffonde intenzionalmente, ma io sostengo che sia importante riconoscere che alle sfide dell'information disorder si aggiungono anche le cattive abitudini interne al giornalismo.

Contenuto fuorviante
Questo aspetto è sempre stato un problema sia nel giornalismo, sia nella politica. Che si tratti di citazioni riportate parzialmente, di statistiche citate per sostenere un'affermazione senza considerare come sono stati raccolti i dati per elaborarle o di foto tagliate per raccontare un evento in un certo modo, pratiche fuorvianti di questo tipo non sono certo una novità.

Contesto falso
Questa è la categoria a cui appartiene la maggior parte dei contenuti, la categoria chiamata in causa praticamente ogni volta che un'immagine autentica e pre-esistente viene ricondivisa come se fosse nuova. Accade spesso durante la copertura di un evento straordinario, quando si ricondivide qualche vecchia immagine, ma succede anche quando vengono rimessi in circolazione articoli vecchi come se fossero attuali, qualora il titolo si adatti ancora agli eventi contemporanei.

Contenuto impostore
Si creano quando il logo di un brand o un nome molto conosciuto vengono abbinati a un contenuto falso. Questa tattica è strategica perché gioca sull'importanza delle intuizioni euristiche. Uno dei metodi a cui più spesso ricorriamo per valutare un contenuto è vedere se è stato creato da un'organizzazione o da una persona di cui già ci fidiamo. Pertanto, prendendo il logo di una fonte di informazione affidabile e applicandolo a una foto o a un video, si fanno automaticamente aumentare le possibilità che il pubblico si fidi di quel contenuto senza controllare.

Contenuto manipolato
Sono contenuti manomessi o falsificati in qualche modo. Ne è un esempio un video di Nancy Pelosi del maggio 2019. La Presidentessa della Camera dei rappresentanti degli Stati Uniti venne filmata mentre teneva un discorso. Poche ore dopo, uscì un video di quel discorso in cui sembrava che fosse ubriaca. Il video era stato rallentato,

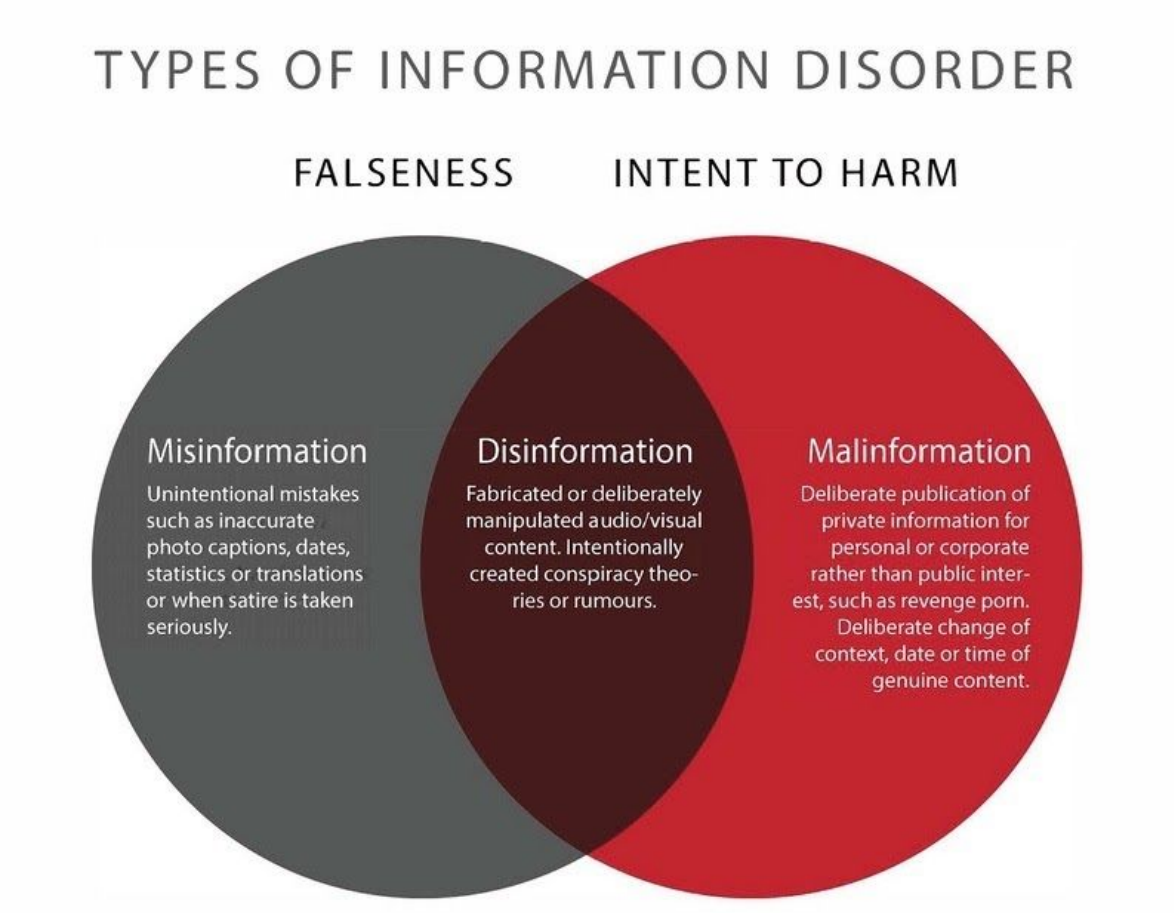
e per questo sembrava che l'oratrice biascicasse le parole. Questa tattica è molto potente, perché sfrutta video autentici: il pubblico, che sa che Nancy Pelosi ha tenuto un discorso in quell'occasione, è più propenso a credere al video manipolato.

Contenuto inventato
In questa categoria rientrano i contenuti inventati al 100%, ad esempio gli account falsi sui social media da cui si diffondono nuovi contenuti. Questa categoria include i cosiddetti deepfake, contenuti in cui, sfruttando l'intelligenza artificiale, si manipola un video o un audio per far sembrare che qualcuno dica o faccia cose che non ha mai detto o fatto.

Capire le intenzioni e gli scopi
Queste categorie sono utili per spiegare la complessità dell'ambiente informativo inquinato in cui ci muoviamo, ma non affrontano la questione delle intenzioni, che è un punto cruciale per comprendere il fenomeno.

Per occuparci di questo aspetto, Derakhshan e io abbiamo creato un diagramma di Venn per spiegare la differenza tra misinformation, disinformation e un terzo termine da noi creato, malinformation. Sia nel caso della misinformation che in quello della disinformation si hanno contenuti falsi. Ma nel caso della disinformation le informazioni false sono create e condivise da persone che intendono provocare dei danni, sia finanziari che di reputazione, politici o fisici. Anche nel caso della misinformation le informazioni sono false, ma le persone che le condividono non ne sono consapevoli. Questo accade spesso durante eventi da breaking news, quando le persone condividono voci non confermate o vecchie foto senza accorgersi che non si riferiscono agli eventi attuali.

Si parla di malinformation, invece, quando informazioni autentiche vengono condivise con l'intento di provocare dei danni. Un esempio di malinformation è la pubblicazione delle email di Hillary Clinton durante la campagna presidenziale statunitense del 2016. Anche il revenge porn rientra in questa categoria.

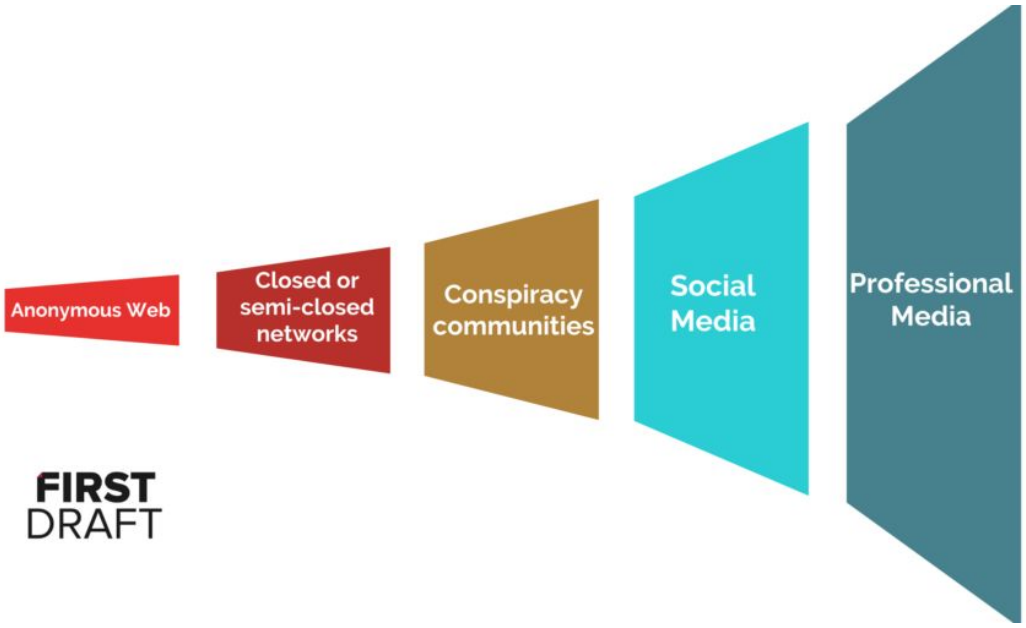


Queste categorie sono importanti, poiché l'intenzione con cui un'informazione viene condivisa fa parte degli elementi da considerare per capire come inquadrare quell'informazione. Ci sono tre ragioni principali che spingono a creare contenuti falsi e ingannevoli. La prima è di natura politica, sia interna che estera. È la ragione che spinge, ad esempio, un governo straniero a cercare di interferire con le elezioni di un altro paese, oppure, nel caso della politica interna, il motivo per cui in campagna elettorale si ricorre a tattiche "sporche" per infangare gli avversari.

La seconda ragione è finanziaria. Chi ha un sito può guadagnare soldi tramite la pubblicità, pertanto se pubblichi un articolo falso e sensazionale con un titolo iperbolico puoi guadagnare soldi finché riesci, tramite quel titolo, a portare click sulla tua URL. Persone di schieramenti politici opposti hanno raccontato di come abbiano creato siti di "notizie" inventate per portare click, e quindi entrate economiche. Infine, ci sono i fattori sociali e psicologici. Alcune persone sono spinte dal semplice desiderio di creare un polverone e vedere cosa ne tirano fuori; altre vogliono vedere se riescono a fregare i giornalisti, o creare un evento Facebook che porti la gente in strada a protestare o a prendersela con le donne e molestarle. Altre ancora si ritrovano a diffondere misinformation per il solo desiderio di apparire in un certo modo, come chi scrive "Non mi interessa se è vero o no, voglio solo sottolineare ai miei amici su Facebook quanto odio [nome del candidato politico]".

Il megafono dell'amplificazione
Per arrivare a comprendere fino in fondo questo vasto ecosistema, dobbiamo renderci conto di quanto profondamente sia aggrovigliato. Accade troppo spesso che chi nota un contenuto ingannevole o falso da qualche parte creda che quello sia il luogo dove è stato originato. Per nostra sfortuna, però, i più capaci nel produrre disinformazione sanno trarre vantaggio proprio dalla sua natura frammentaria.

Ricorda, inoltre, che voci non confermate, contenuti complottisti o informazioni false non potrebbero fare alcun male se non venissero condivisi. È proprio perché vengono condivisi che provocano così tanti danni. Per questo ho creato questa immagine, che io chiamo “il megafono dell’amplificazione", al fine di descrivere come chi fa disinformazione si coordina per far muovere le informazioni all'interno dell’ecosistema.



Molto spesso i contenuti che generano disinformazione vengono pubblicati in spazi come 4Chan o Discord (una app usata dai videogiocatori per comunicare), vale a dire spazi anonimi dove le persone possono pubblicare senza dover rispondere di quello che fanno. Spesso questi spazi vengono usati per comunicare dettagli su come coordinarsi nel diffondere i contenuti, ad esempio “cercheremo di far diventare di tendenza questo dato hashtag”, oppure “usate questo meme per reagire agli eventi di oggi su Facebook”.

Spesso il coordinamento si sposta poi da questi siti a grandi gruppi privati su Twitter o su WhatsApp, usati da coloro che fungono da nodi della rete per portare il contenuto a gruppi più grandi di persone. I contenuti così diffusi potrebbero quindi arrivare alle community di Gab, Reddit, YouTube e siti simili e da qui essere condivisi su siti più mainstream, come Facebook, Instagram o Twitter. Da tali

piattaforme vengono poi spesso ripresi da media professionisti; o perché questi, non riconoscendone la natura, li usano nei loro pezzi senza adeguati controlli, o perché decidono di sottoporli a verifica. In entrambi i casi, coloro che generano disinformazione considerano questo risultato come un successo. Cattivi titoli che riportano una voce non verificata o un'affermazione ingannevole, o articoli di verifica al cui interno è inserito un contenuto falso, giocano entrambi a favore di chi fa disinformazione, perché non fanno che amplificare il contenuto e alimentare le voci che gli nascono attorno.

In First Draft parliamo del concetto di punto di non ritorno. Per un giornalista occuparsi delle menzogne troppo presto significa dare ulteriore ossigeno a voci non verificate e potenzialmente dannose. Occuparsene troppo tardi, invece, significa che queste hanno avuto tempo di attecchire e che ormai c'è poco che si possa fare. La sfida è riconoscere quando arriva il punto di non ritorno e coglierlo. Il momento giusto dipende dal luogo, dall'argomento e dalla piattaforma.

Conclusioni

Le parole sono importanti. Il fenomeno di cui ci occupiamo è complesso e le parole che usiamo fanno la differenza. Sono già disponibili ricerche accademiche che mostrano che per un pubblico sempre più vasto le "fake news" coincidono con cattive abitudini giornalistiche messe in atto dai media professionisti.

Evitare di descrivere tutto come disinformazione, anche quando il contenuto potrebbe non essere realmente falso o quando è condiviso inconsapevolmente da utenti che non pensano sia falso, è un altro principio cruciale per comprendere cosa sta accadendo.

Viviamo in un'epoca di information disorder. E ciò sta creando nuove sfide per i giornalisti, per i ricercatori e per i professionisti dell'informazione. Scrivere un articolo o non scriverlo? Come formulare un titolo? Come smascherare efficacemente video e immagini false? Come sapere quando farlo? Come capire qual è il momento giusto? Queste sono le sfide che esistono oggi per coloro che lavorano nel mondo dell'informazione. Ed è complicato.

C. Il ciclo di vita della manipolazione dei media

Scritto da: Joan Donovan

*La **dott.ssa Joan Donovan** è direttrice della ricerca sui temi dei media, della politica e delle politiche pubbliche all'Harvard Kennedy's Shorenstein Center.*

In un'epoca in cui i mezzi tradizionali attraverso i quali la società viene informata sono stati scompaginati da un pugno di potenti piattaforme tecnologiche globali, le istituzioni sociali e politiche sono messe alla prova dalla manipolazione dei media e dalle campagne di disinformazione. Bufale e storie inventate vengono diffuse da un variegato insieme di attori politici, brand, movimenti sociali e "troll" senza affiliazione che hanno sviluppato e affinato nuove tecniche per influenzare la conversazione pubblica, generando il caos su scala locale, nazionale e globale.

C'è ampio consenso sul fatto che la manipolazione dei media e la disinformazione siano problemi importanti della società. Eppure, definire, riconoscere, documentare e disinnescare la manipolazione dei media e la disinformazione è ancora complicato, soprattutto perché gli attacchi sono trasversali a più settori professionali, tra i quali vi sono il giornalismo, il diritto e la tecnologia. Considerare la manipolazione dei media un'attività che risponde a un preciso schema è il primo, essenziale passo per studiarla, descriverla e combatterla.

Definire la manipolazione dei media e la disinformazione

Per definire la manipolazione dei media, prima di tutto dividiamo il termine in due parti. Nel loro significato più generale, i media sono artefatti della comunicazione. Ne sono un esempio i testi, le immagini, i materiali audio e video e i media digitali. Quando si studiano i media, ogni traccia da loro lasciata può essere usata come prova. La caratteristica fondamentale dei media è che vengono creati dalle persone allo scopo di comunicare. Sono, pertanto, portatori di un certo significato tra gli individui, ma l'interpretazione di quel significato è sempre relativa e dipendente dal contesto in cui il messaggio viene diffuso.

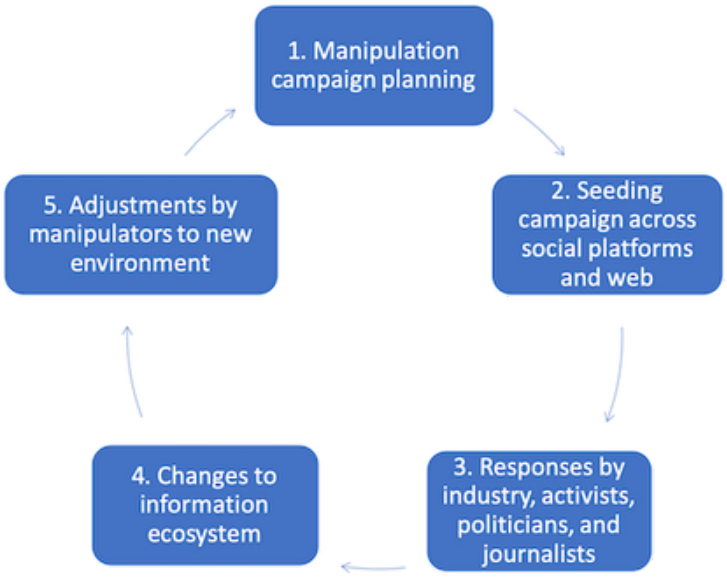
Affermare che i media sono manipolati significa spingersi oltre il semplice dire che sono fabbricati da qualcuno per trasmettere il messaggio che questo qualcuno intende esprimere. Il dizionario Merriam-Webster definisce il termine manipulation come "(l'atto di) modificare qualcosa con astuzia o con metodi scorretti per conseguire il proprio fine" (nota del traduttore: il vocabolario Treccani dà una definizione simile di "manipolare": "Adattare, volgere in senso favorevole a sé stessi, mediante imbrogli e intrighi, allo scopo di ottenere vantaggi personali").

Anche se può essere difficile conoscere il fine esatto per cui un dato artefatto è stato prodotto, chi indaga può individuare cosa, dove, come e da chi è stato creato: elementi che contribuiscono a capire se nel processo di diffusione del messaggio siano state adottate tattiche manipolatorie. Tra le tattiche manipolatorie rientrano ad esempio il fatto di nascondere la propria identità o la fonte dell'artefatto creato, modifiche volte a nascondere o cambiare il significato o il contesto di un artefatto e l'espediente di ingannare gli algoritmi tramite un'azione coordinata che sfrutti mezzi artificiali come bot o strumenti per spammare.

In questo contesto, la disinformazione è una sottocategoria della manipolazione dei media che indica la creazione e la diffusione intenzionali di informazioni false a fini politici. Professionisti della tecnologia, esperti, studiosi, giornalisti e politici devono concordare su questa distinta categorizzazione della disinformazione, perché gli sforzi per combatterla richiedono la cooperazione di tutti loro.

Da parte nostra, nel gruppo di ricerca del Technology and Social Change (TaSC) dell'Harvard Kennedy School's Shorenstein Center stiamo mappando il ciclo di vita della manipolazione dei media tramite un approccio basato sui casi di studio. Questo approccio metodologico si propone di analizzare l'organizzazione, le dimensioni e i fini di una campagna di manipolazione seguendo il percorso degli artefatti dei media nel tempo e nello spazio, ricomponendone le multiple relazioni per venire a capo di quella che si presenta come una matassa aggrovigliata. All'interno di questo lavoro abbiamo sviluppato una panoramica del ciclo di vita delle campagne di manipolazione, utile ai giornalisti che cerchino di riconoscere, ricostruire e portare alla luce la manipolazione dei media e la disinformazione.

Il ciclo di vita di una campagna di manipolazione dei media



1. Pianificazione della campagna di manipolazione
2. Disseminazione della campagna attraverso le piattaforme social e il web
3. Reazioni da parte del mondo industriale, degli attivisti, dei politici e dei giornalisti
4. Modifiche all'ecosistema dell'informazione
5. Adattamenti dei manipolatori al nuovo ambiente

Il ciclo di vita di una campagna di manipolazione si articola in cinque punti di azione; in ciascun punto le tattiche dei manipolatori possono essere documentate attraverso metodi qualitativi e quantitativi. Si noti che la maggior parte delle campagne di manipolazione non viene ricostruita secondo questo ordine. Quando si indaga riguardo una campagna di manipolazione si va alla ricerca di uno qualsiasi di questi punti, partendo da lì per poi ricostruire, in avanti o a ritroso, l'intero ciclo di vita della campagna.

Un caso studio: *'Blow the Whistle'*

Per vedere come si svolge una campagna di manipolazione dei media e come le scelte etiche di giornalisti e piattaforme in una delle prime fasi del suo ciclo di vita possano contribuire a vanificarne gli sforzi, esaminiamo l'attività dei social media sviluppatasi attorno alla denuncia di un informatore sulle attività collegate all'Ucraina del presidente Donald Trump.

Pianificazione e Disseminazione (Fasi 1 e 2) - Nella galassia dei media cospirazionisti, l'identità dell'informatore è già conosciuta e il suo nome circola su blog e su Twitter, Facebook, video di YouTube e forum di discussione. Una cosa importante: nomi specifici possono essere sostituiti con parole chiave e hashtag, che possono essere sfruttati come dati rintracciabili. C'è un'azione concertata per diffondere il presunto

nome e la foto della persona. Tuttavia, la circolazione del nome sembra ancora limitata entro i confini di un'eco-chamber di media online formata da soggetti e account cospirazionisti e di destra. Nonostante gli influencer cospirazionisti coordinino i loro sforzi per portare il presunto nome dell'informatore al grande pubblico, non riescono a farlo uscire dalle loro stesse bolle di filtraggio. Perché?

Reazione di giornalisti, attivisti, etc. (Fase 3) – Di contro, i media di sinistra e di centro non hanno reso pubblico il nome del presunto informatore, né amplificato dichiarazioni sul fatto che sia stato rivelato. I media mainstream si astengono dall'attirare l'attenzione sulla circolazione del nome negli ambienti dei social media, anche se si tratta di una storia che fa notizia per tutti quei giornalisti che si occupano di tecnologia e politica. Chi se ne occupa, spesso enfatizza il fatto che far circolare il nome sia un tentativo di manipolare la discussione sul contenuto della sua denuncia, ed evitano di diffondere ulteriormente il nome. Per questo occorre ringraziare in larga parte l'etica del giornalismo, che prevede il dovere da parte dei giornalisti di proteggere l'anonimato delle proprie fonti, anche degli informatori.

Modifiche all'ecosistema dell'informazione (Fase 4) – Anche se i giornalisti dei media mainstream omettono il presunto nome dell'informatore, "Eric Ciaramella" diventa una parole chiave. Cercandolo si trova un'ampia varietà di contenuti prodotti da un punto di vista influenzato dalla teoria cospirazionista. Ai giornalisti, che per etica non danno risalto a una storia che produrrebbe un traffico considerevole, si aggiungono le aziende gestrici di piattaforme, che iniziano a moderare attivamente i contenuti contenenti il nome come parola chiave. YouTube e Facebook rimuovono i contenuti che utilizzano il nome, mentre Twitter impedisce che diventi trending topic. Google permette ancora di fare una ricerca con il nome, restituendo come risultato migliaia di link a blog cospirazionisti.

Adattamenti dei manipolatori al nuovo ambiente (Fase 5) – Infastiditi da questi tentativi di impedire la diffusione della disinformazione, i manipolatori cambiano tattica. Invece di spingere su contenuti con il nome del presunto informatore, cominciano a far circolare immagini di un altro uomo bianco (con occhiali e barba) somigliante a quello che circolava in precedenza assieme al nome dell'informatore. Queste nuove immagini sono accompagnate da una narrativa cospirazionista che coinvolge le più alte sfere dello Stato e secondo la quale l'informatore sarebbe un amico dei democratici dell'establishment e, per questo, sarebbe motivato da ragioni di parte. L'immagine fatta circolare è un'immagine di Alexander Soros, il figlio del miliardario, affarista e filantropo George Soros, frequente bersaglio di teorie della cospirazione.

Quando anche questa tattica non riesce ad attirare l'attenzione dei media, l'account Twitter del Presidente Trump, @RealDonaldTrump, retwitta un articolo che rivela il presunto nome dell'informatore, sottolineando ai suoi 68 milioni di follower che "l'informatore della CIA non è un vero informatore!". Il tweet originale era stato pubblicato da @TrumpWarRoom, l'account ufficiale e verificato della campagna del presidente. Ne segue una copertura mediatica a cascata, in cui si inseriscono anche molti dei più importanti organi di stampa mainstream, i quali si premurano di togliere o coprire il nome del presunto informatore. Molte persone sui social media chiedono che l'informatore testimoni in Senato alle udienze per l'impeachment, dove il suo nome viene fatto insieme a quello di altri importanti potenziali testimoni, aumentando così la possibilità che altri si imbattano in lui mentre cercano gli altri. E così comincia un nuovo ciclo di manipolazione dei media.

Le ricerche contenenti il nome dell'informatore aumentano, mentre i blog abbondano di teorie del complotto circa le motivazioni personali e professionali che lo avrebbero spinto a fornire informazioni sulle attività di Trump. I giornalisti che coprono la vicenda di questi tweet oscillano tra due tendenze: alimentare il dibattito sul tema dell'intimidazione del testimone, sottolineando che un atto come questo possa dissuadere eventuali futuri informatori dal farsi avanti, e indulgere a una morbosa curiosità nel riportare i gossip sui motivi per cui Trump abbia fatto il nome del presunto informatore. Di per sé, è lodevole che alcuni media cerchino di richiamare le élite alle proprie responsabilità, ma la missione è impossibile se le aziende che gestiscono le piattaforme non affrontano la questione di come i loro prodotti siano diventati strumenti politici utili a manipolare i media e diffondere disinformazione.

Documentare il ciclo di vita di una campagna

Disseminando il nome e le foto del presunto informatore sui social media, dove le piattaforme potevano farli diventare di tendenza e facilmente trovabili affinché media più grandi e legittimati le amplificassero, i manipolatori dei media hanno cercato di far fare il salto di qualità alla notizia che volevano diffondere. Ma le decisioni e le azioni delle piattaforme e dei giornalisti hanno fatto fallire in gran parte il tentativo di affermare la presunta identità dell'informatore nell'immaginario

collettivo, almeno finché non è arrivata una spinta sull'argomento da parte di un personaggio che fa notizia. Mentre molti mezzi di informazione si impegnano ad attenersi a codici etici, i social media offrono a chi già è potente un'arma per imporre l'agenda mediatica e portare avanti complotti pericolosi.

Su un piano generale, tuttavia, questo esempio rappresenta un significativo miglioramento degli sforzi per fermare la diffusione della disinformazione rispetto a quanto non si facesse in passato, quando i giornalisti amplificavano le campagne di disinformazione nel tentativo di portarle allo scoperto, e le piattaforme non si sentivano in dovere di fornire al pubblico informazioni accurate. Questo cambiamento di tendenza è promettente, ma la responsabilizzazione delle élite è ancora debole. Quando ci si dedica a riconoscere, documentare e smontare le campagne di manipolazione dei media, sia come giornalisti che come ricercatori, la posta in gioco è alta. In questo periodo iperpolarizzato, ogni tentativo di denunciare una campagna di disinformazione può scatenare orde di troll e attenzioni non richieste. Occuparsi dei contenuti e del contesto della disinformazione significa, per tutte le figure coinvolte, documentare con rigore scientifico come una campagna comincia, evolve e finisce. E riconoscere che la presunta fine di una campagna può benissimo trasformarsi nell'inizio di una nuova.

1. Indagare sugli account nei social media

Scritto da: Brandy Zadrozny

***Brandy Zadrozny** è una giornalista investigativa per NBC News, dove si occupa prevalentemente di misinformation, disinformation e di estremismi su Internet.*

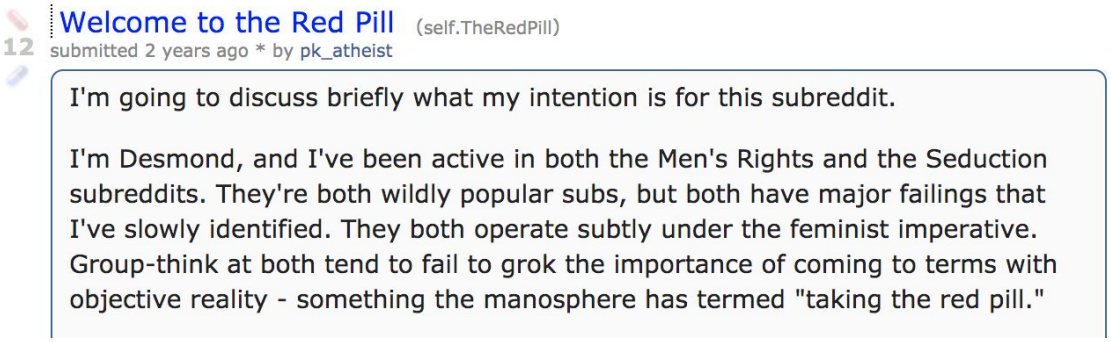
Praticamente tutte le storie di cui mi occupo comportano indagini sui social media. Che si tratti di ricostruire il background di un profilo, di breaking news o di inchieste più lunghe, le piattaforme di social media rappresentano uno dei modi migliori per scoprire la vera vita di una persona — la sua famiglia, gli amici, i lavori, le opinioni politiche, le associazioni – e anche una finestra sui pensieri segreti e le identità nascoste online.

Questo è un momento incredibile per fare i giornalisti: le persone vivono la propria vita sempre più online, e gli strumenti per cercare e indagare su un profilo social sono ovunque. Allo stesso tempo, sia le persone normali sia quelle intenzionate ad arrecare danni stanno diventando sempre più scaltre nel nascondere le proprie tracce. Parallelamente, tuttavia, piattaforme di social media come Facebook hanno reagito alla cattiva pubblicità - dovuta ad articoli riguardanti le loro falle nella privacy e la diffusione di ideologie pericolose tramite essi - chiudendo l'accesso agli strumenti da cui dipendevano giornalisti e ricercatori per scoprire storie e identificare persone.

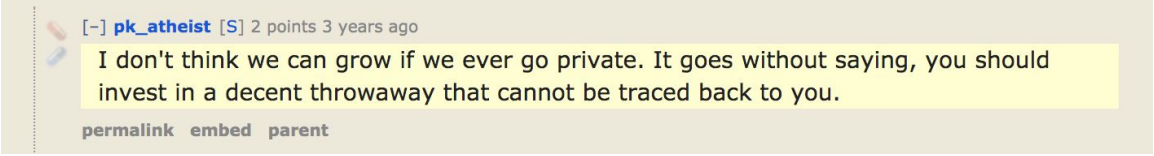
In questo capitolo mostrerò alcuni approcci fondamentali per fare indagini sugli account social. Gli strumenti sono quelli che uso in questo momento, ma presto verranno eliminati da Facebook o rimpiazzati da qualcosa di meglio. I reporter migliori in questo campo hanno ciascuno le proprie modalità e i propri strumenti, ma, come in ogni tipo di giornalismo, i risultati migliori sono quelli portati dall'ossessione e dai metodi old school. Preparati a leggere migliaia di tweet, a cliccare fino all'ultimo dei risultati di Google e a infilarti nel pozzo senza fondo dei social media se vuoi raccogliere i sottili indizi biografici che ti aiuteranno a rispondere alla domanda: "Chi è questa persona?"

I nomi utente

Talvolta un nome utente è l'unica cosa che abbiamo in mano, e va bene così, perché nella maggior parte dei casi è da lì che si inizia. Fu così, ad esempio, con il caso del rappresentante repubblicano del New Hampshire che mise in piedi una delle più popolari e odiose comunità maschili su Reddit. L'indagine che ha portato a smascherare l'architetto del subreddit The Red Pill, attualmente in quarantena, iniziò dal nome utente "pk_atheist."



Alcune persone mantengono lo stesso nome utente su varie piattaforme e provider di posta elettronica, apportando solo minime variazioni. Altre, più attente alla sicurezza, come appunto il rappresentante del New Hampshire, creano e poi abbandonano nomi utente a ogni nuova impresa.



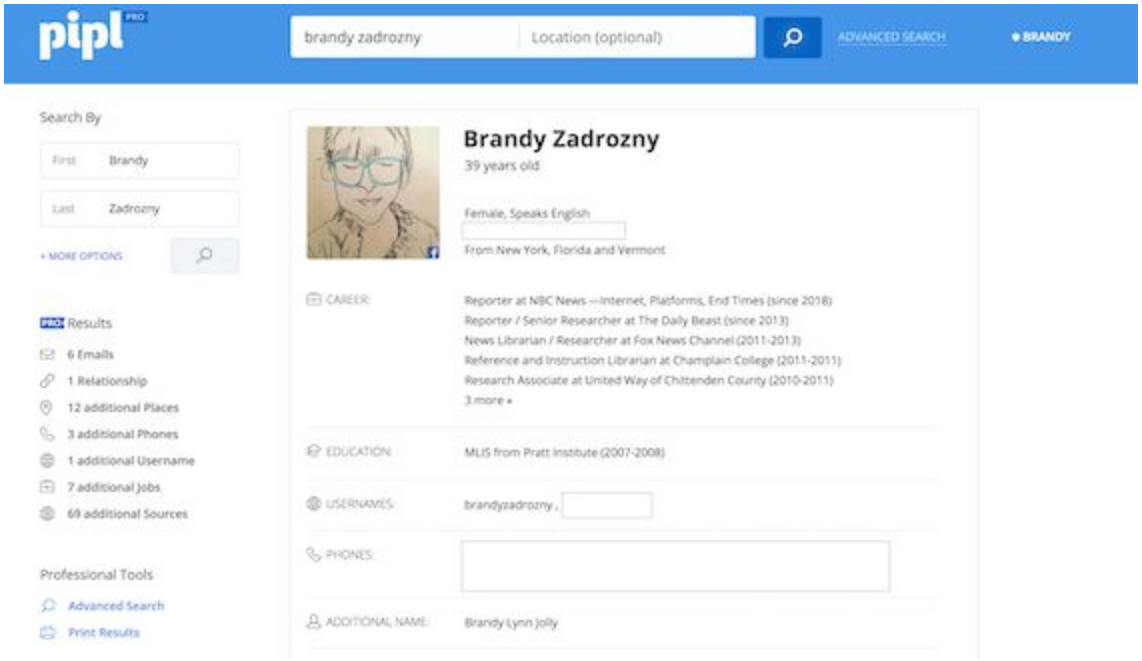
In ogni caso, esistono alcuni siti utili a cui dare in pasto il nome utente che stai cercando.

Ciò che faccio io per prima cosa è scrivere il nome utente su Google. Le persone, specialmente le più giovani, che rifuggono le grandi piattaforme social, tendono a lasciare tracce nei posti più inaspettati (come nelle sezioni dei commenti, nelle recensioni e nei forum), che possono condurti ad altre informazioni e altri account.

Oltre alla ricerca di Google, utilizza dei servizi privati. Hanno un costo, e la possibilità di accedervi dipende dal budget della tua redazione. La maggior parte delle redazioni ha Nexis, che è ottimo per i documenti pubblici e legali, ma purtroppo carente per quanto riguarda e-mail e username; in più, permette di fare ricerche solo su persone negli Stati Uniti. Pipl e Skopenow sono tra i migliori strumenti che ho trovato per fare confronti incrociati tra informazioni prese "dal mondo reale", come numeri di telefono e registri di proprietà, e dati online, come e-mail e nomi utente, ed entrambi funzionano in tutto il mondo.

Questi motori di ricerca a pagamento forniscono tabulati telefonici e registri di proprietà, ma possono anche rintracciare profili Facebook e LinkedIn, che rimangono anche dopo che l'account viene chiuso. Servono anche a ricollegarsi ad account di cui le persone si sono totalmente dimenticate, come vecchi blog o liste di desideri su Amazon, una miniera d'oro per scoprire cosa legge, compra e desidera una persona. Queste fonti restituiscono anche molti falsi positivi, per questo tendo a

iniziare le mie indagini con i loro risultati per poi continuare con altri metodi di verifica.



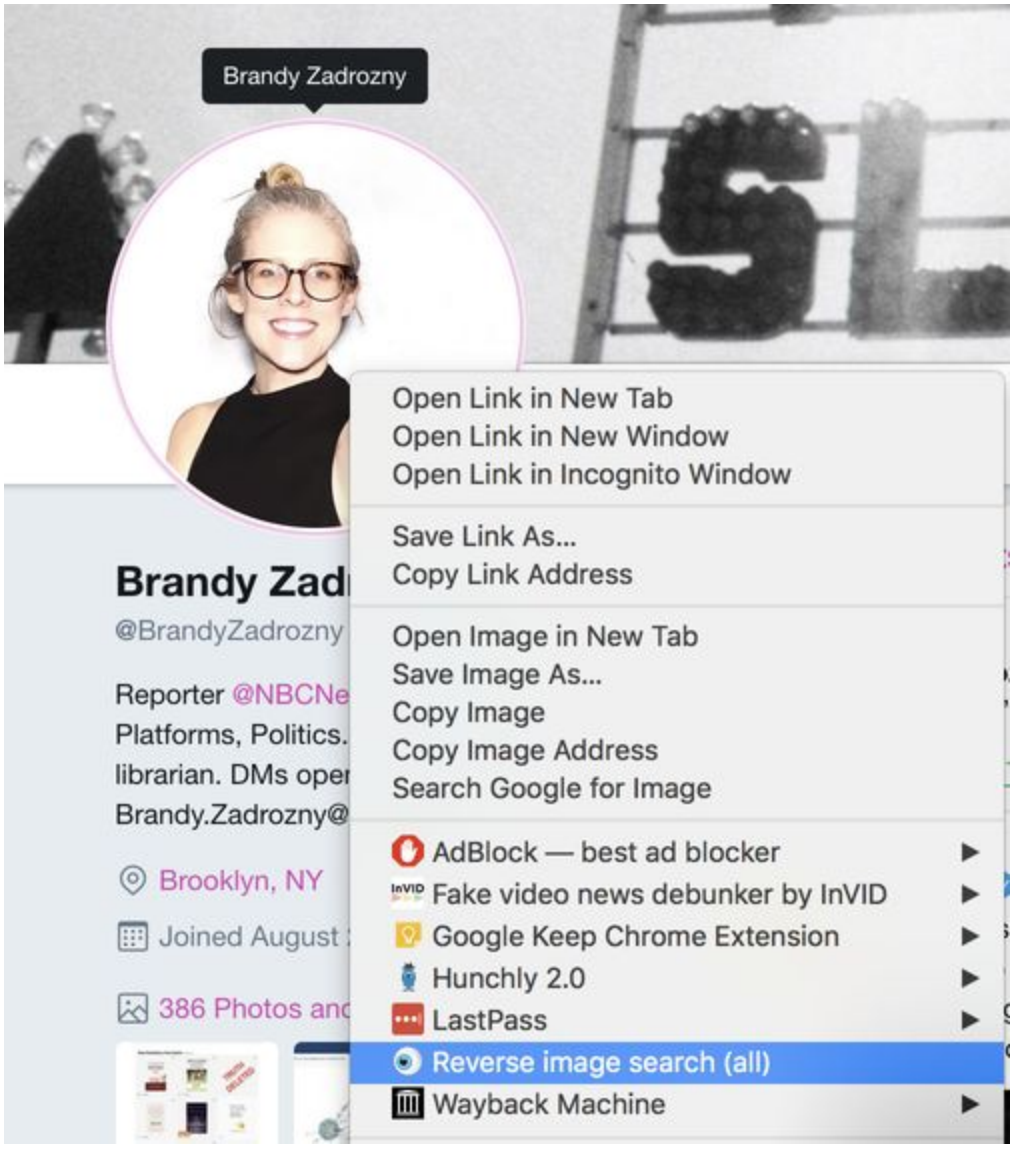
Quando scopro un nome utente o un indirizzo e-mail che penso possa appartenere al soggetto di cui mi sto occupando, lo inserisco in uno strumento online come namechk o namecheckr, che verifica la disponibilità dei nomi utente su diverse piattaforme. Questi strumenti sono progettati per aiutare chi si occupa di marketing a capire se un dato username che desideravano registrare sia ancora disponibile sulle piattaforme, ma sono molto utili anche per verificare se un nome utente su cui stai indagando esista anche da altre parti. Ovviamente, il fatto che uno stesso nome utente sia registrato su più piattaforme non significa che tutti gli account appartengano alla stessa persona. Tuttavia, è un ottimo punto di partenza per cercare in maniera trasversale su più piattaforme.

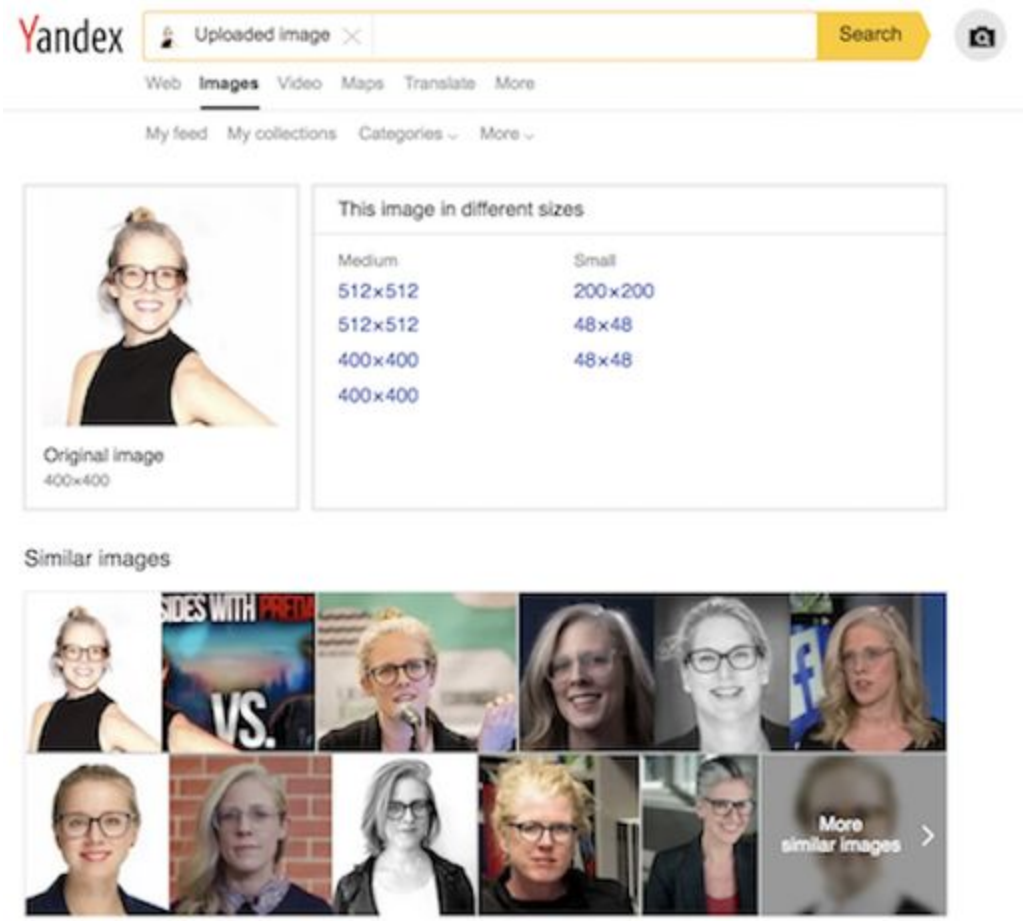
Per ulteriori controlli sui nomi utente sono utili anche haveibeenpwned.com e Dehashed.com, che permettono di verificare se ci sono state violazioni dei dati nelle informazioni degli utenti e possono rappresentare una via veloce per convalidare un indirizzo e-mail e arrivare a nuovi contatti.

Foto

Non sempre un nome utente è sufficiente per andare avanti nelle ricerche. E poi, nulla convince più di un'immagine. Le foto profilo sono un altro modo per verificare l'identità di una persona attraverso diversi account.

La ricerca inversa delle immagini di Google è buona, ma spesso altri motori di ricerca - specialmente il russo Yandex - riescono a dare risultati migliori. Io uso l'estensione di Chrome Reveye, che mi permette di cliccare con il tasto destro su un'immagine e cercarla su molteplici piattaforme, inclusi Google, Bing, Yandex e Tineye. L'estensione Cerca per immagine ha anche un'accurata funzione di cattura, la quale permette di avviare una ricerca a partire da un'immagine all'interno di un'altra immagine.

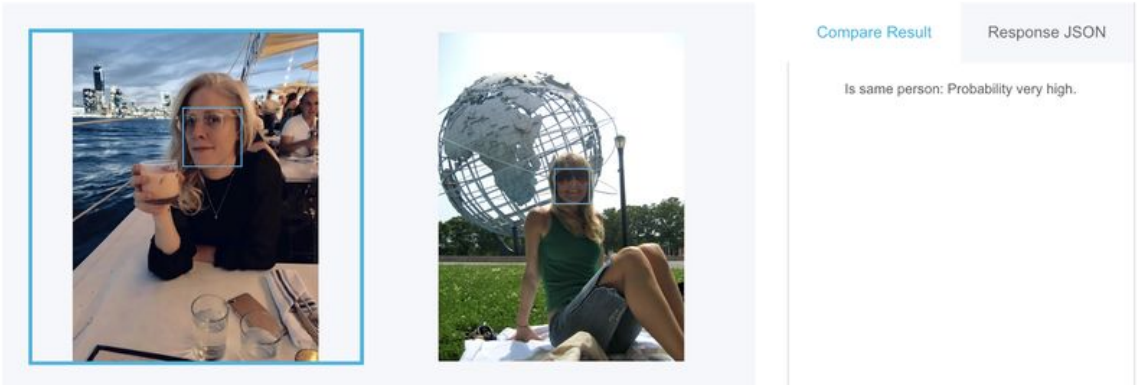

Brandy Zadrozny
Brandy Zad
@BrandyZadrozny
Reporter @NBCNe
Platforms, Politics.
librarian. DMs oper
Brandy.Zadrozny@
Brooklyn, NY
Joined August
386 Photos and
Open Link in New Tab
Open Link in New Window
Open Link in Incognito Window
Save Link As...
Copy Link Address
Open Image in New Tab
Save Image As...
Copy Image
Copy Image Address
Search Google for Image
AdBlock — best ad blocker
Fake video news debunker by InVID
Google Keep Chrome Extension
Hunchly 2.0
LastPass
Reverse image search (all)
Wayback Machine



Ovviamente, la ricerca inversa delle immagini pone dei problemi. I motori di ricerca citati sopra sono piuttosto scarsi nel cercare immagini su Twitter, e si rivelano praticamente inutili per ottenere risultati da siti come Instagram e Facebook.

La maggior parte delle volte mi ritrovo a guardare immagini di vari individui, e ho perso il conto delle volte in cui ho strizzato gli occhi davanti al monitor chiedendo ai miei colleghi: “Ma è la stessa persona?”.

Non mi fido dei miei occhi. Tuttavia, riuscire a riconoscere alcune caratteristiche da una foto all’altra, come i nei o la forma dei baffi o della barba, è di grande aiuto. Ultimamente trovo utile fare un controllo con strumenti di riconoscimento facciale come Face++, su cui si possono caricare due foto per ottenere un'indicazione di quanto sia probabile appartengano alla stessa persona. In questi esempi lo strumento è stato in grado di riconoscermi in foto scattate a distanza di dieci anni. Ha anche riconosciuto il mio collega Ben sulle foto dei profili social di Twitter e Facebook, rilevando, correttamente, che in effetti non si trattava di Ben Stiller.

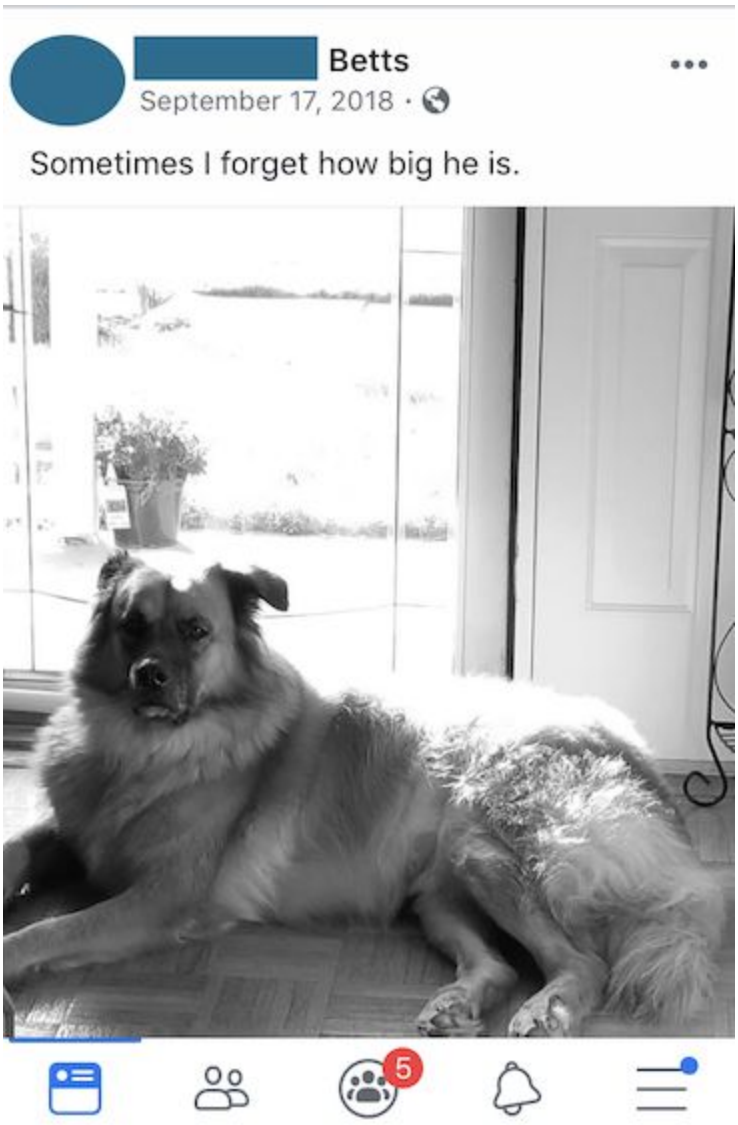


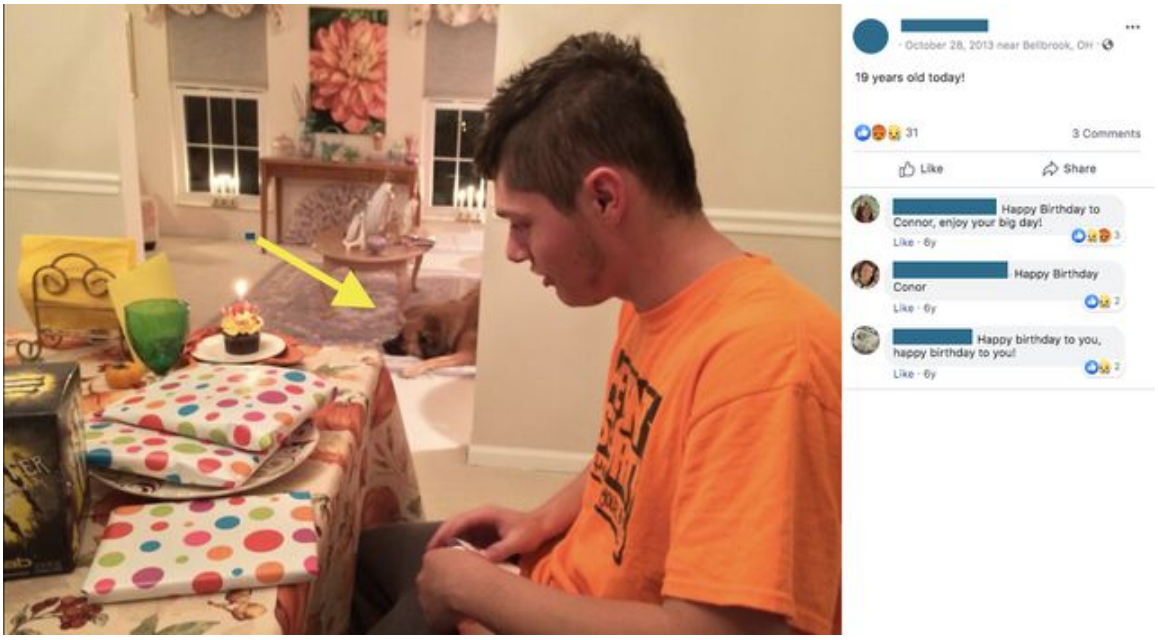
Se state dando la caccia a troll o truffatori, potreste scoprire che si sono impegnati a oscurare le proprio foto profilo o che usano delle foto false. In questi casi modificare la foto e capovolgerla può essere utile per ricostruire il loro processo all'inverso.

Ad ogni modo, le foto profilo non sono gli unici indizi a nostra disposizione. Anche quando diventano più consapevoli e attente alla propria privacy e a quella della propria famiglia, le persone continuano a condividere foto delle cose che le rendono fiere. Sono riuscita a identificare persone collegando tra loro foto di automobili, case, animali domestici e cose del genere. Le foto diventano quindi un mezzo per mettere in relazione alcuni account e le persone che li gestiscono con altri account,

rendendo possibile far emergere la rete attorno al tuo obiettivo. Si tratta di una pratica cruciale quando si indaga sugli account dei social media.

Un esempio: stavamo cercando di confermare l'appartenenza di alcuni account social a un uomo che aveva sparato e ucciso nove persone fuori da un bar a Dayton, in Ohio. Il suo account Twitter offriva degli indizi sulla sua ideologia politica, ma il suo pseudonimo, @iamthespookster, era unico e non assomigliava al suo vero nome, reso pubblico dalle autorità. Il fatto che una delle vittime fosse suo fratello, un uomo transessuale il cui nome non era riportato nei registri pubblici e non era ancora stato reso noto, rendeva ancor più complicato identificare le figure chiave. Ma sia nel profilo dell'omicida che in quelli della sua famiglia apparivano delle foto di un cane. L'animale era anche l'immagine del profilo del fratello transgender.





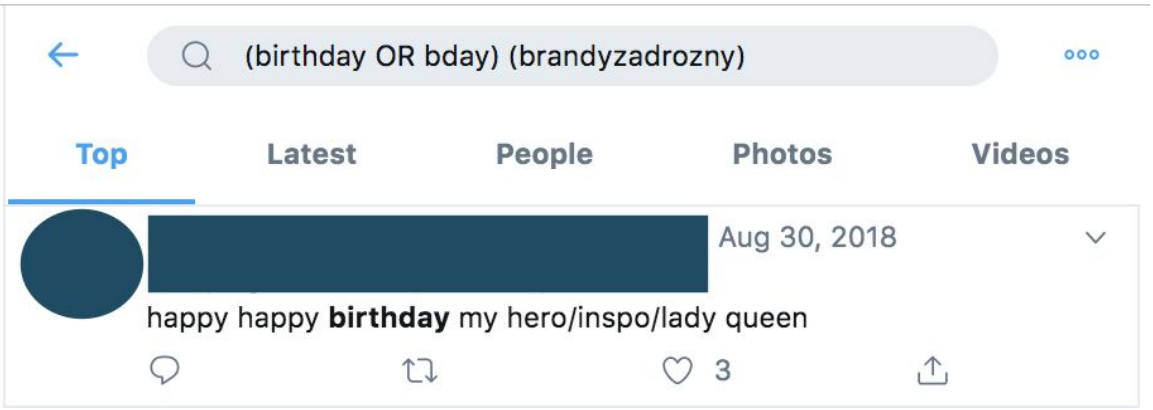
Nell'immagine qui sopra, il cane non era l'unico dettaglio utile. La foto proveniva dal padre del killer dell'Ohio, e ci aiutò a verificare i suoi account personali e quelli della sua famiglia.

Se hai un account su Facebook o Twitter probabilmente sono in grado di dirti in che giorno sei nato, anche se non hai condiviso l'informazione sul tuo profilo o non lo

hai detto tu stesso in un post. Spesso, quando arriva una breaking news, la data di nascita è uno dei primi dati identificativi di una persona rilasciati dalla polizia. Per questo motivo, un metodo affidabile per verificare un account social consiste nello scrollare la pagina del profilo fino al mese e al giorno in questione e controllare se ci sono degli auguri di compleanno. I genitori, come nel caso di Connor Betts qui sopra, scrivono spesso un post sul compleanno dei propri figli, anche se per il resto le loro pagine risultano vuote.

Lo stesso vale su Twitter. In fin dei conti, chi non ama i compleanni?

Trovare un post identificativo su Twitter è ancora più facile, perché il suo strumento di ricerca avanzata è uno dei migliori tra quelli offerti dalle piattaforme social. Io stessa, nonostante dica molto raramente, se non mai, quando è il mio compleanno, ho ritrovato il tweet di auguri di un gentile collega che mi smascherava.



I compleanni sono solo un esempio. Matrimoni, funerali, vacanze, anniversari, lauree: quasi ogni momento importante della vita viene celebrato sui social. E ciò apre delle strade per fare ricerche e indagini sugli account.

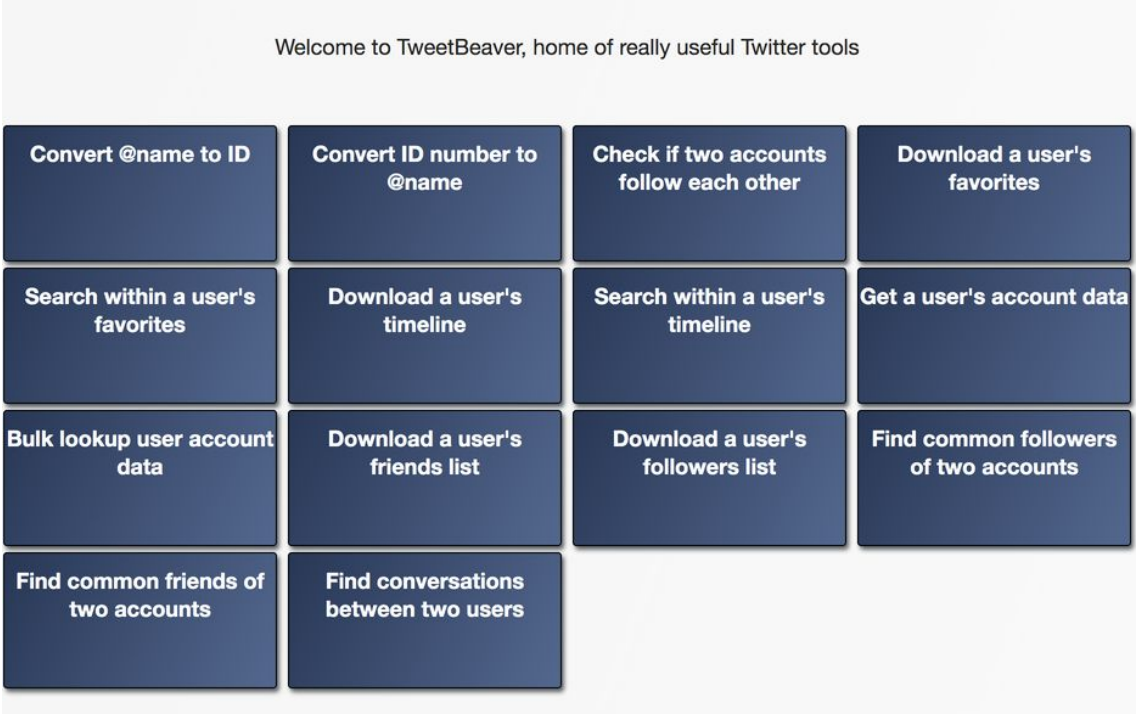
Con gli strumenti di ricerca per Facebook puoi fare una ricerca con queste parole chiave e applicare ulteriori filtri. Questi strumenti non riescono a spingersi così oltre come prima che Facebook annunciasse la sua svolta a favore della privacy, ma ci sono ancora. Uno dei miei preferiti è whopostedwhat.com.

Parentele

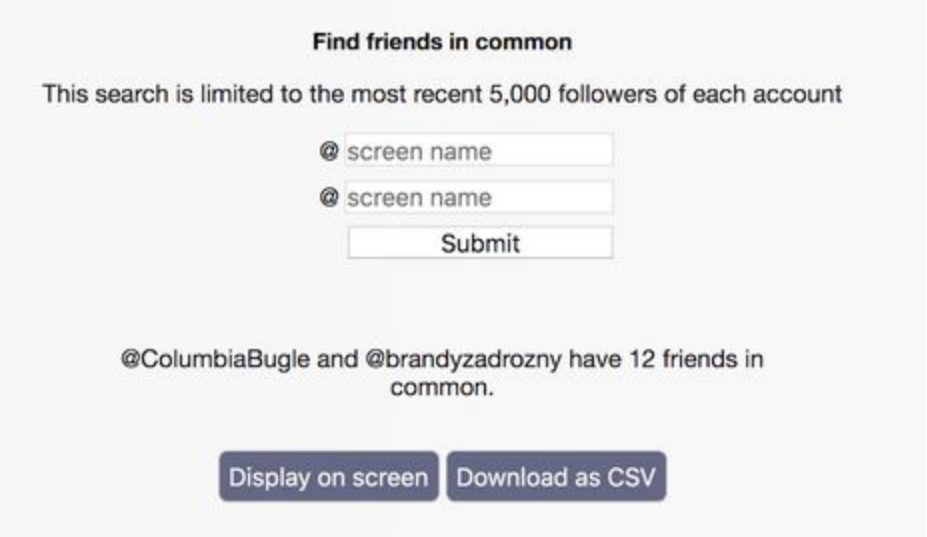
Puoi analizzare una persona a partire dalle frequentazioni che ha sui social media. Esaminando gli account con cui una persona interagisce online possiamo imparare moltissimo circa la sua vita e le sue inclinazioni.

Quando ho aperto il mio account su Twitter, ho fatto iscrivere anche mio marito e il mio migliore amico, così che potessero seguirmi. Ci penso ogni volta che indago su qualche account per lavoro. Nemmeno le piattaforme vogliono che tu rimanga solo, per questo quando apri un account si mette in moto un algoritmo che porta la piattaforma a suggerirti altri account da seguire. Per farlo, la piattaforma si orienta in base ai contatti che hai nella rubrica del telefono, alla tua presenza nelle liste di contatti di account esistenti, a dove ti trovi e ad altri fattori.

È quindi sempre molto interessante andare a guardare chi sono stati i primi follower e amici di un account. TweetBeaver è uno strumento utile per indagare le connessioni tra grandi account e per scaricare timeline, liste dei preferiti e altri dati simili riguardanti account più piccoli. Per banche dati più grandi, mi affido a sviluppatori con accesso API.



Prendiamo The Columbia Bugle, un account Twitter di estrema destra molto popolare che si vanta di essere stato retwittato due volte dall'account di Donald Trump.



I primi follower di Max Delarge, un account che dichiara di essere l'editor di The Columbia Bugle, sono fonti di notizie specifiche su San Diego e account sportivi della stessa città. Visto che molti tweet del Columbia Bugle comprendono video dei comizi di Trump a San Diego ed eventi dell'Università della California, che ha sede a San Diego, possiamo essere abbastanza sicuri del fatto che la persona che gestisce l'account viva vicino a San Diego.

Follow

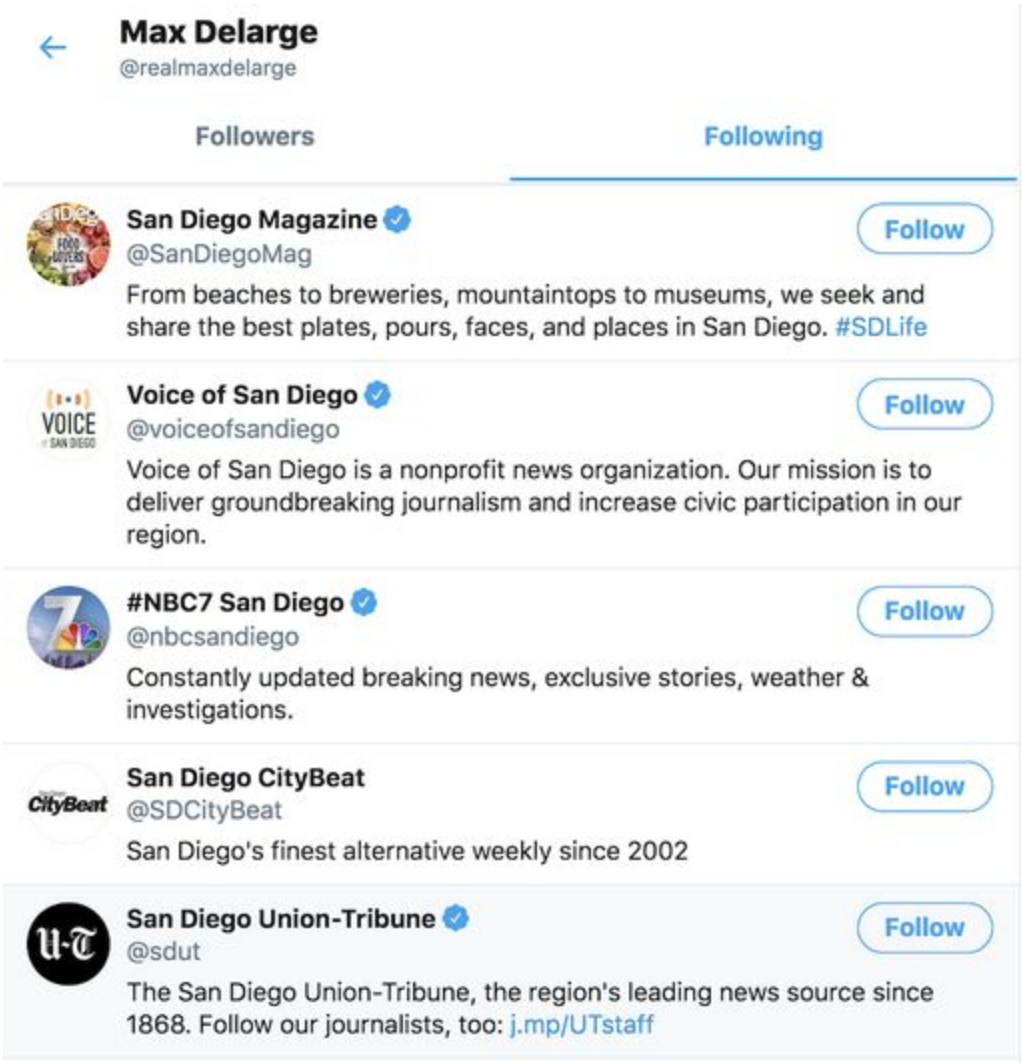
Max Delarge

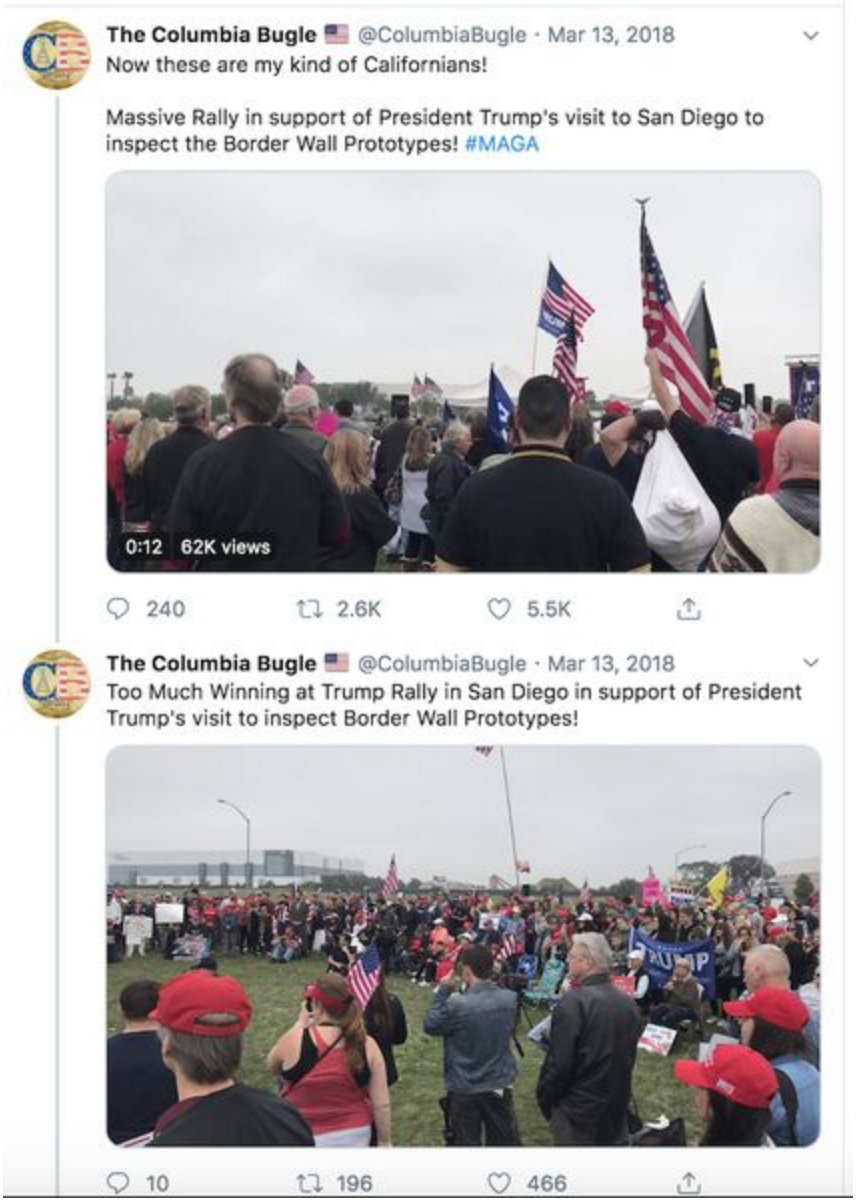
@realMaxDelarge

Co-Editor of The Columbia Bugle. Still got the scars of #NeverTrump, but im on the #TrumpTrain for good, unless he lights the train on fire

United States Joined July 2016

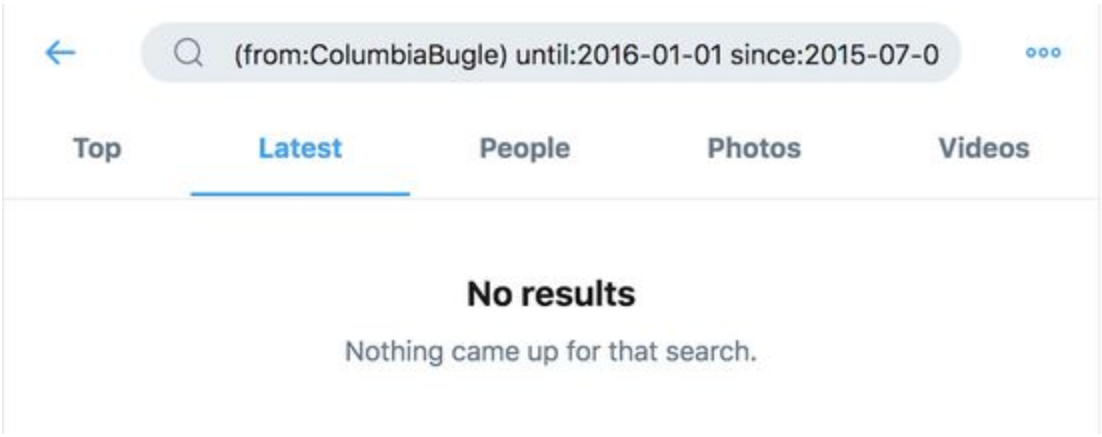
22 Following **0** Followers


Max Delarge
@realmaxdelarge
Followers
Following
San Diego Magazine
@SanDiegoMag
Follow
From beaches to breweries, mountaintops to museums, we seek and share the best plates, pours, faces, and places in San Diego. #SDLife
Voice of San Diego
@voiceofsandiego
Follow
Voice of San Diego is a nonprofit news organization. Our mission is to deliver groundbreaking journalism and increase civic participation in our region.
#NBC7 San Diego
@nbcsandiego
Follow
Constantly updated breaking news, exclusive stories, weather & investigations.
San Diego CityBeat
@SDCityBeat
Follow
San Diego's finest alternative weekly since 2002
San Diego Union-Tribune
@sdut
Follow
The San Diego Union-Tribune, the region's leading news source since 1868. Follow our journalists, too: j.mp/UTstaff



Quando comincio una nuova indagine, mi piace partire dalla creazione di un account su Twitter e indagare da quel punto in poi. Si può risalire a questo momento in maniera manuale, oppure tramite un'estensione di Chrome che scrolla automaticamente, o, ancora, usando la ricerca avanzata di Twitter per limitare la finestra temporale della ricerca ai primi mesi di vita dell’account.

Curiosamente, questo account non mostra nessun tweet nei primi sei mesi dalla sua creazione.

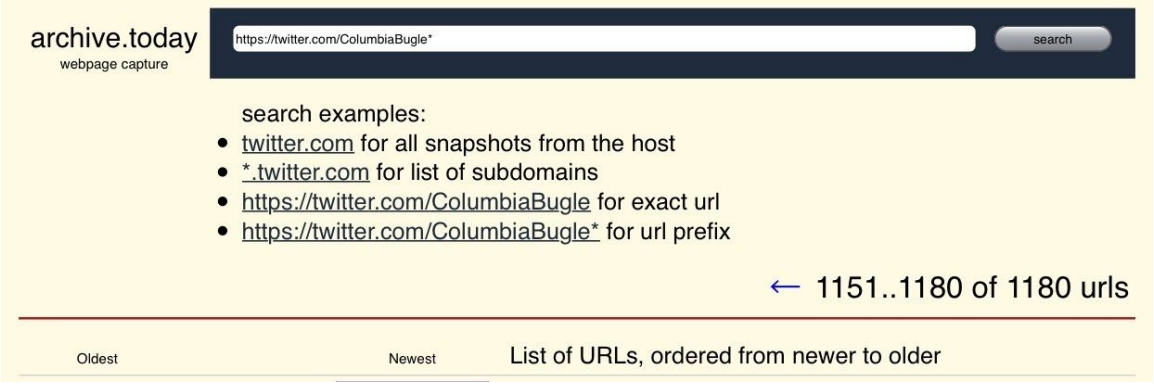


Ciò suggerisce che la persona dietro a The Columbia Bugle potrebbe aver cancellato i suoi tweet più vecchi. Per capire perché ciò possa essere accaduto, posso modificare la mia ricerca. Invece che cercare i tweet scritti da The Columbia Bugle, provo a cercare qualsiasi tweet che *lo menzioni*.

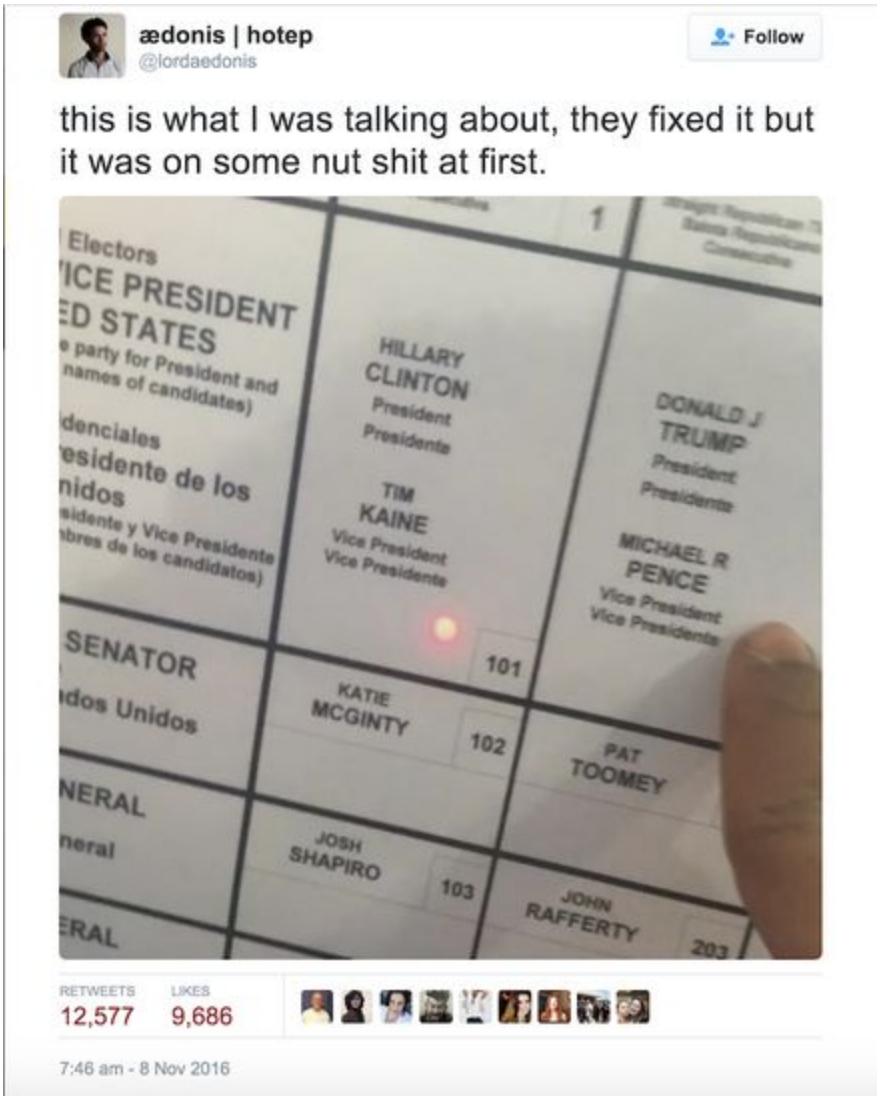
Queste conversazioni confermano che The Columbia Bugle ha cancellato i tweet del suo primo anno di vita, ma non ci dicono nulla sul perché. Inoltre, i primi account con i quali The Columbia Bugle aveva interagito non ci offrono particolari indizi.

Per trovare dei tweet cancellati di recente, puoi cercare nella cache di Google. Si può accedere a vecchi tweet cancellati anche usando la Wayback Machine dell'Internet Archive o un altro archivio. Il sito dell'archivio manuale archive.is restituisce molti tweet cancellati dai quali si evince che Columbia Bugle aveva partecipato a un evento durante il quale alcuni studenti del college avevano scritto messaggi pro Trump nei propri campus. Per consultare tutti i tweet di quell'account che possono

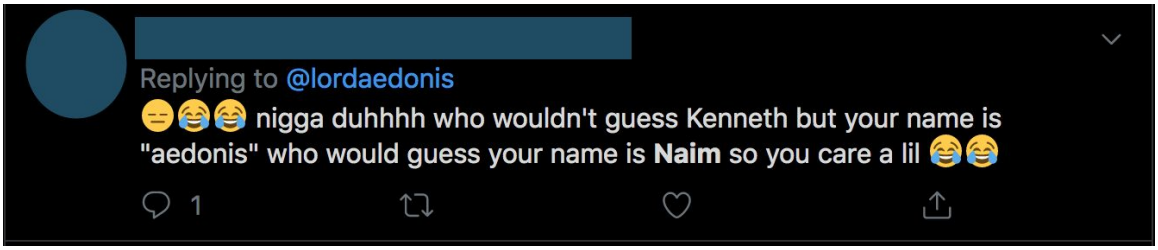
essere stati archiviati da qualcuno, puoi cercare inserendo il prefisso dell'URL e usando un asterisco dopo il nome dell'account, come ho fatto io per scovare questo tweet:



È raro che qualcuno riesca con successo a tenere separata la vita reale dalle attività online. Ad esempio, io e il mio collega della NBC News abbiamo raccontato la storia della più virale e fuorviante dichiarazione di frode elettorale dell'Election Day del 2016 grazie all'aiuto di un vicino del troll di estrema destra che l'aveva twittata.



Il tweet era stato pubblicato da un uomo conosciuto dai suoi follower come @lordaedonis. Tuttavia, i vicini del suo quartiere avevano risposto a vecchi tweet postati con il suo vero nome. Noi avevamo ricollegato questo nome al profilo di un imprenditore affamato di attenzioni il cui tweet era stato diffuso da un account Twitter sostenuto dal Cremlino, e che alla fine era stato visto da milioni di persone e promosso da quello che sarebbe diventato il presidente.

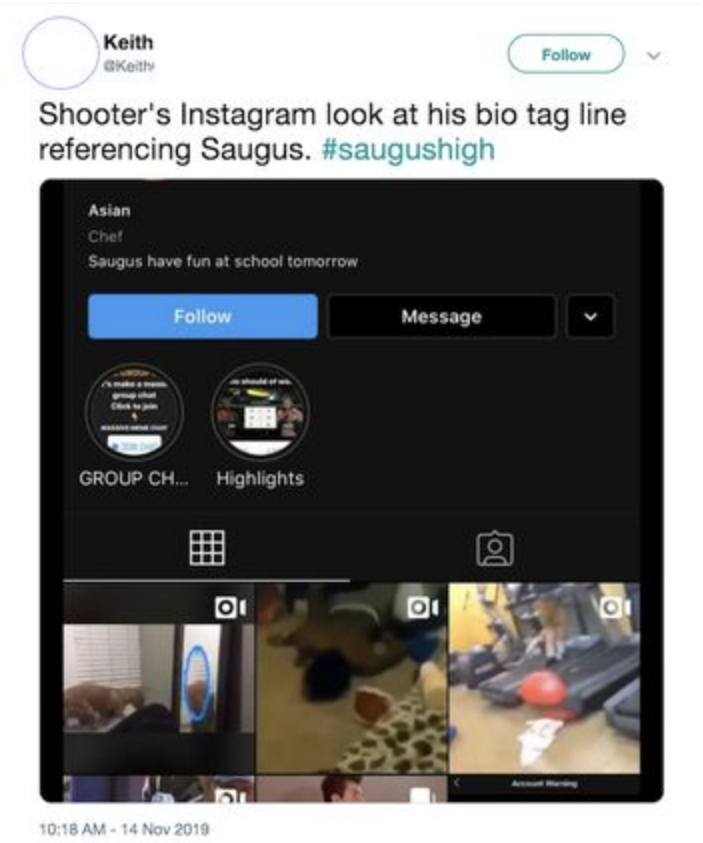


Le mie storie preferite sono quelle che riescono a rivelare chi sono le persone reali che si nascondono dietro influenti e anonimi account social. Gli account segreti di queste persone fanno poco affidamento sugli algoritmi, e sono gestiti con molta più attenzione al fine di sfuggire alla vita pubblica. Consentono a chi li manovra di tenere sotto controllo la situazione, comunicare con famiglia e amici senza ricorrere ai propri account pubblici ed esprimere idee e opinioni che, per ragioni personali o politiche, non osano dire ad alta voce.

La giornalista Ashley Feinberg è una maga in questo tipo di storie succose, quelle che smascherano gli account alternativi di personaggi di spicco come James Comey o Mitt Romney. Il suo segreto consisteva semplicemente nello scovare account minori di parenti che Comey o Romney avrebbero certamente voluto seguire e poi esaminarli tutti fino a trovare quello che poteva sembrare falso, ma il cui contenuto e i cui amici/follower corrispondevano a quelli delle persone reali.

Essere cauti con gli account fake

Ogni piattaforma ha la sua personalità, le sue capacità di ricerca e la sua utilità in differenti situazioni che riguardano l'informazione. Ma occorre essere cauti: vale sempre la regola "fidati, ma verifica". Ci sono persone elettrizzate dall'idea di ingannare i giornalisti. Soprattutto in situazioni da breaking news nascono continuamente account falsi, molti dei quali con messaggi minacciosi e terribili pensati solo per attirare i giornalisti. Questo account fake su Instagram ha usato il nome dell'autore di una strage ed è stato creato dopo la sparatoria alla Saugus High School in California. Ha ottenuto attenzione grazie a degli screenshot su Twitter, ma BuzzFeed News ha poi rivelato che non apparteneva all'omicida.



Per evitare di essere ingannati bisogna sempre verificare un account social indagando su chi lo gestisce, sulla sua famiglia e sugli amici e avvalendosi delle forze dell'ordine e/o dei Social Media PR. Un'ultima nota, forse la più importante di tutte: non esiste un giusto ordine per eseguire i passi che abbiamo descritto. Spesso vengo trascinata nella buca del coniglio e mi ritrovo con più schede aperte sul monitor di quante sarei disposta ad ammettere. Creare una procedura replicabile, ad esempio tenendo traccia di ogni tua azione su un Google Doc o pagando uno strumento come Hunchly per monitorare le tue ricerche, è la chiave per chiarire le connessioni tra le persone e le loro vite online, e trasformare le tue scoperte in storie.

1a. Caso di studio: Scoprire una rete coordinata di diffusione della propaganda nelle Filippine indagando su una serie di account Facebook

Scritto da Vernise Tantuco

e Gemma Bagayaua-Mendoza

Giornalista professionista da circa vent'anni, ***Gemma Bagayaua-Mendoza*** *è a capo della ricerca e della strategia di Rappler. Dirige l'unità di fact-checking e la ricerca di Rappler su disinformation e misinformation online.*

Vernise Tantuco *fa parte del gruppo di ricerca di Rappler, per cui si occupa di verifica dei fatti e di studiare i network di disinformazione nelle Filippine.*

Nell'autunno del 2016 John Victorino, analista degli investimenti, inviò a Rappler una lista di quelli che sosteneva essere 26 account Facebook sospetti provenienti dalle Filippine. Cominciammo a fare indagini e a monitorare gli account, e scoprimmo presto che le informazioni elencate nei profili erano false. Nel corso di settimane di indagine, questi 26 account ci portarono a scoprire una rete molto più vasta formata da pagine, gruppi e altri account.

Gli account in questione furono in seguito rimossi da Facebook, così come una serie di altre pagine e gruppi ai quali erano connessi. La vicenda inoltre spinse Rappler a creare Sharktank, uno strumento per monitorare come circolano le informazioni su Facebook. Questo lavoro gettò le basi per una serie di inchieste su come la propaganda e le operazioni informative su Facebook avessero influenzato la democrazia nelle Filippine. Tra le indagini della serie, una riguardava le attività dei 26 account falsi. Quest'ultima rappresentò l'inizio del nostro lavoro di copertura continua su come, nelle Filippine, Facebook sia stato trasformato in un'arma per diffondere disinformazione politica, molestare le persone e minare la democrazia nel Paese.

Questo caso di studio esamina come abbiamo indagato sui 26 account e come li abbiamo usati per scoprire una rete più grande.

Verificare le identità, smascherare gli account-fantoccio

Il primo passo nelle nostre indagini sul gruppo di account fu cercare di verificare se fossero collegati a persone vere. Per questa parte dovemmo ricorrere a del sano fact-checking vecchia maniera. Iniziammo creando delle tabelle per tenere traccia di

dettagli legati agli account, come ad esempio le informazioni personali che presentavano, le pagine a cui avevano messo like e altre informazioni.

Per esempio, l'utente di Facebook Mutya Bautista si descriveva come "analista programmatore" alla ABS-CBN, il più grande network televisivo delle Filippine. Rappler verificò con ABS-CBN, che confermò che la persona in questione non lavorava per loro.

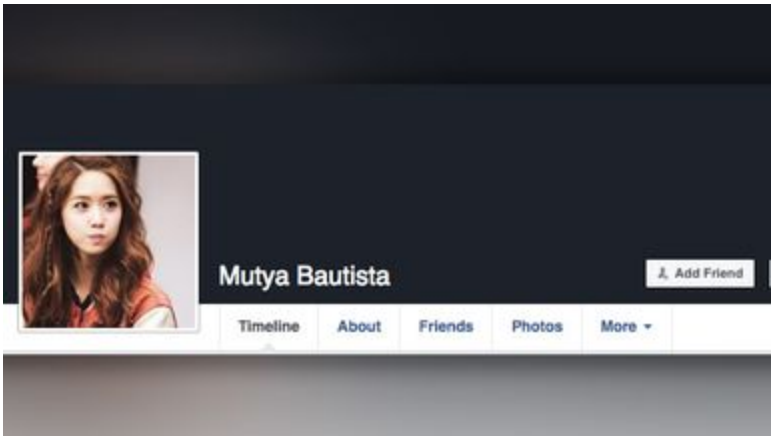
Personal Information	
Facebook ID	https://www.facebook.com/profile.php?id=10
Profile Name	Mutya Bautista
Occupation	Software Analyst
Current Company	ABS-CBN Corporation
Former Occupation 1	
Former Occupation 2	
Former Occupation 3	
Former Occupation 4	
Former Occupation 5	
Studied	Computer Engineering
Studied at	University of the Philippines
Went to	
Lives in	
Married to	
From	
Account Set-up Date	October 19, 2015

Liked Pages	Liked Pages Facebook ID
Okay Dito	https://www.facebook.com/vidtimestories/
The Philippine Pride	https://www.facebook.com/sirangplaka2/

Photos	Source of Photo
Profile Photo	Numerous sources. Im Yoona of SNSD
Cover Photo	

Usando gli strumenti di ricerca inversa delle immagini scoprimmo che molti dei 26 account usavano come immagine del profilo foto di celebrità o di personaggi noti.

Bautista, per esempio, usava un'immagine di Im Yoona, del gruppo pop coreano Girl's Generation. L'account di Lily Lopez (qui sotto) usava l'immagine dell'attrice coreana Kim Sa-rang.



Un altro account, Luvimin Cancio, usava come foto profilo un'immagine presa da softcorecams.com, un sito porno. Abbiamo scoperto che la foto veniva da lì usando lo strumento di ricerca inversa delle immagini di TinEye.

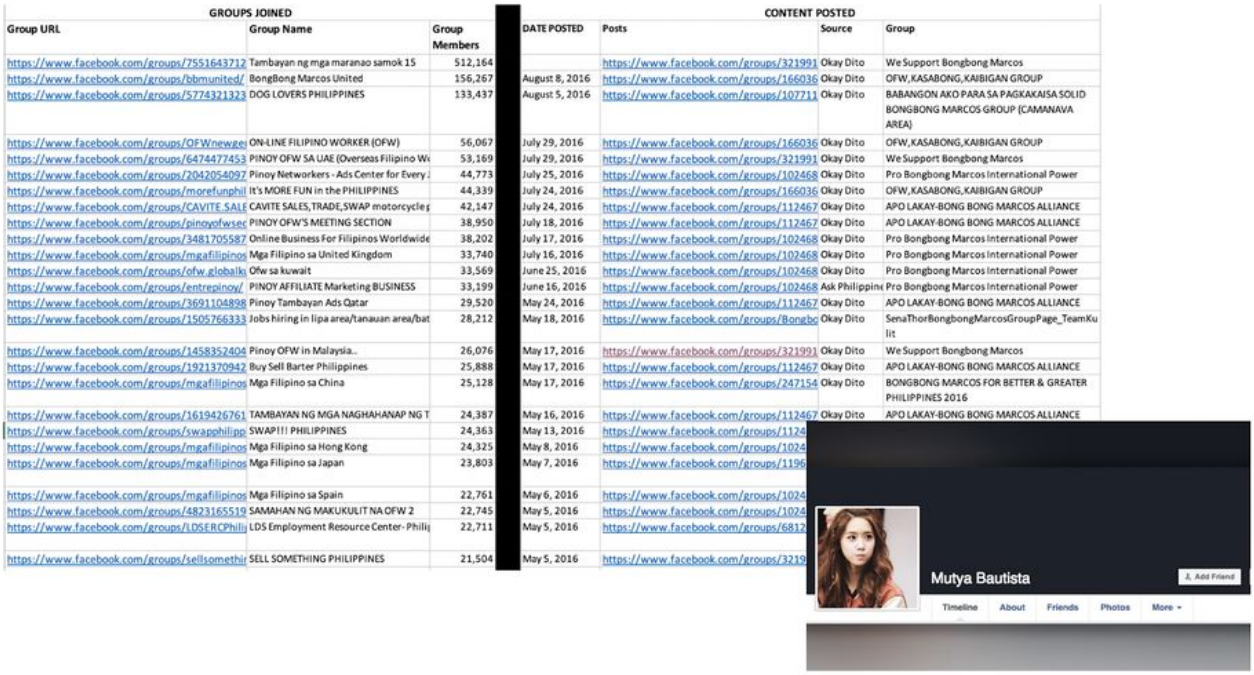
Inoltre, gli account utilizzavano sui loro profili immagini di copertina simili tra loro. Come si può vedere qui sotto, la foto di copertina dell'account di Jasmin De La Torre è la stessa di quella di Lily Lopez.

Notammo anche un'altra cosa curiosa: gli utenti dei 26 account avevano più gruppi che amici.

La cosa era insolita, perché nelle Filippine la maggior parte delle persone ha amici e familiari all'estero. Facebook serve fondamentalmente come canale di comunicazione per tenersi in contatto con amici e parenti. Per questo le persone tendono ad avere molti amici piuttosto che far parte di un gran numero di gruppi.

La lista degli amici di Bautista, che all'epoca era pubblica, contava solo 17 amici. Di fatto nel 2016, quando scoprimmo i 26 account, ciascuno di essi aveva meno di 50 amici.

Bautista, in ogni caso, faceva parte di oltre un centinaio di gruppi, inclusi gruppi che sostenevano l'allora vice candidato alla presidenza Ferdinand Marcos Jr., un certo numero di gruppi di comunità di filippini oltreoceano e anche gruppi di compravendita, ciascuno dei quali con un numero di membri che andava dalle decine alle centinaia di migliaia. Presi tutti insieme, questi gruppi contavano oltre 2,3 milioni di membri su Facebook. Qui sotto è riportata una lista di alcuni dei gruppi più grandi, con relativo numero di follower, e una lista dei post che Bautista vi aveva pubblicato.



GROUPS JOINED			CONTENT POSTED			
Group URL	Group Name	Group Members	DATE POSTED	Posts	Source	Group
https://www.facebook.com/groups/7551643712	Tambayan ng mga maranao samok 15	512,164		https://www.facebook.com/groups/321991	Okay Dito	We Support Bongbong Marcos
https://www.facebook.com/groups/bbmunited/	BongBong Marcos United	156,267	August 8, 2016	https://www.facebook.com/groups/166036	Okay Dito	OFW,KASABONG,KAIBIGAN GROUP
https://www.facebook.com/groups/5774321323	DOG LOVERS PHILIPPINES	133,437	August 5, 2016	https://www.facebook.com/groups/107711	Okay Dito	BABANGON AKO PARA SA PAGKAKAISA SOLID BONGBONG MARCOS GROUP (CAMANAVA AREA)
https://www.facebook.com/groups/OFWnewgec	ON-LINE FILIPINO WORKER (OFW)	56,067	July 29, 2016	https://www.facebook.com/groups/166036	Okay Dito	OFW,KASABONG,KAIBIGAN GROUP
https://www.facebook.com/groups/6474477453	PINOY OFW SA UAE (Overseas Filipino W	53,169	July 29, 2016	https://www.facebook.com/groups/321991	Okay Dito	We Support Bongbong Marcos
https://www.facebook.com/groups/2042054097	Pinoy Networkers - Ads Center for Every	44,773	July 25, 2016	https://www.facebook.com/groups/102468	Okay Dito	Pro Bongbong Marcos International Power
https://www.facebook.com/groups/morefunphil	It's MORE FUN in the PHILIPPINES	44,339	July 24, 2016	https://www.facebook.com/groups/166036	Okay Dito	OFW,KASABONG,KAIBIGAN GROUP
https://www.facebook.com/groups/CAVITE.SALE	CAVITE SALES, TRADE, SWAP motorcycle g	42,147	July 24, 2016	https://www.facebook.com/groups/112467	Okay Dito	APO LAKAY-BONG BONG MARCOS ALLIANCE
https://www.facebook.com/groups/pinoyofwsec	PINOY OFW'S MEETING SECTION	38,950	July 18, 2016	https://www.facebook.com/groups/112467	Okay Dito	APO LAKAY-BONG BONG MARCOS ALLIANCE
https://www.facebook.com/groups/3481705587	Online Business For Filipinos Worldwide	38,202	July 17, 2016	https://www.facebook.com/groups/102468	Okay Dito	Pro Bongbong Marcos International Power
https://www.facebook.com/groups/mgafilipinos	Mga Filipino sa United Kingdom	33,740	July 16, 2016	https://www.facebook.com/groups/102468	Okay Dito	Pro Bongbong Marcos International Power
https://www.facebook.com/groups/ofw.globalkı	Ofw sa kuwait	33,569	June 25, 2016	https://www.facebook.com/groups/102468	Okay Dito	Pro Bongbong Marcos International Power
https://www.facebook.com/groups/entrepinoy/	PINOY AFFILIATE Marketing BUSINESS	33,199	June 16, 2016	https://www.facebook.com/groups/102468	Ask Philippine	Pro Bongbong Marcos International Power
https://www.facebook.com/groups/3691104898	Pinoy Tambayan Ads Qatar	29,520	May 24, 2016	https://www.facebook.com/groups/112467	Okay Dito	APO LAKAY-BONG BONG MARCOS ALLIANCE
https://www.facebook.com/groups/1505766333	Jobs hiring in lipa area/tanauan area/bat	28,212	May 18, 2016	https://www.facebook.com/groups/Bongbo	Okay Dito	SenaThorBongbongMarcosGroupPage_TeamKulit
https://www.facebook.com/groups/1458352404	Pinoy OFW in Malaysia..	26,076	May 17, 2016	https://www.facebook.com/groups/321991	Okay Dito	We Support Bongbong Marcos
https://www.facebook.com/groups/1921370942	Buy Sell Barter Philippines	25,888	May 17, 2016	https://www.facebook.com/groups/112467	Okay Dito	APO LAKAY-BONG BONG MARCOS ALLIANCE
https://www.facebook.com/groups/mgafilipinos	Mga Filipino sa China	25,128	May 17, 2016	https://www.facebook.com/groups/247154	Okay Dito	BONGBONG MARCOS FOR BETTER & GREATER PHILIPPINES 2016
https://www.facebook.com/groups/1619426761	TAMBAYAN NG MGA NAGHAHANAP NG T	24,387	May 16, 2016	https://www.facebook.com/groups/112467	Okay Dito	APO LAKAY-BONG BONG MARCOS ALLIANCE
https://www.facebook.com/groups/swappilipp	SWAP!!! PHILIPPINES	24,363	May 13, 2016	https://www.facebook.com/groups/1124		
https://www.facebook.com/groups/mgafilipinos	Mga Filipino sa Hong Kong	24,325	May 8, 2016	https://www.facebook.com/groups/1024		
https://www.facebook.com/groups/mgafilipinos	Mga Filipino sa Japan	23,803	May 7, 2016	https://www.facebook.com/groups/1196		
https://www.facebook.com/groups/mgafilipinos	Mga Filipino sa Spain	22,761	May 6, 2016	https://www.facebook.com/groups/1024		
https://www.facebook.com/groups/4823165519	SAMAHAN NG MAKUKULIT NA OFW 2	22,745	May 5, 2016	https://www.facebook.com/groups/1024		
https://www.facebook.com/groups/LDSERCPhili	LDS Employment Resource Center- Phili	22,711	May 5, 2016	https://www.facebook.com/groups/6812		
https://www.facebook.com/groups/sellsomethir	SELL SOMETHING PHILIPPINES	21,504	May 5, 2016	https://www.facebook.com/groups/3219		

Mettendo insieme queste osservazioni e i dati a esse associati, concludemmo che gli account erano account-fantoccio: identità fittizie create per supportare un particolare punto di vista.

Network pro Marcos

Le date associate alle prime foto profilo e ai primi post ci portarono a dedurre che gli account erano stati creati nell'ultimo trimestre del 2015 per arrivare alle elezioni del maggio 2016. Scoprimmo anche che gli account promuovevano ripetutamente contenuti che negavano gli abusi, ampiamente documentati, compiuti negli anni Settanta sotto la legge marziale durante il regime di Marcos. Inoltre gli account attaccavano i rivali del figlio del vecchio dittatore, il candidato alla vice presidenza Ferdinand "Bongbong" Marcos Jr.

Nell'esempio riportato di seguito, l'utente Mutya Bautista dichiarava che la rivale di Bongbong, l'allora neo proclamata vice presidente Leni Robredo, fosse stata sposata con un attivista prima di unirsi al suo secondo marito, l'ex ministro dell'interno e segretario del governo locale Jesse Robredo (in seguito fu dimostrato che la cosa era falsa). Bautista pubblicò una storia intitolata "Leni Robredo è stata sposata con un ragazzo anti Marcos prima di incontrare Jesse?" nel gruppo "Pro Bongbong Marcos International Power" ("A favore del potere internazionale di Bongbong Marcos"), con il commento: "Kaya ganun na lamang ang pamemersonal kay [Bongbong Marcos], may root cause pala." ("Ecco il perché della sua avversione personale [a Bongbong Marcos], c'è una ragione sotto").

Lo stesso giorno, un altro account sospetto con il nome di Raden Alfaro Payas condivise esattamente lo stesso articolo nel gruppo “Bongbong Marcos loyalist Facebook warriors” ("Guerrieri di Facebook fedeli a Bongbong Marcos"), aggiungendo la stessa identica frase, parola per parola, fino all’ultimo segno di punteggiatura.

Gli account falsi vengono spesso usati per spammare link nei gruppi. Ogni tanto li si può beccare a usare lo stesso testo in più gruppi. All’epoca era ancora possibile usare il motore di ricerca di Facebook Graph per cercare tra i post pubblici degli utenti nei gruppi. Tuttavia nel 2019 Facebook ha chiuso molte delle funzionalità di ricerca di Graph, inclusa questa. Di conseguenza, ora per vedere cosa ha condiviso un certo utente in un gruppo bisogna cercare direttamente nel gruppo stesso.

Siti correlati

Analizzando i contenuti condivisi dagli utenti, ci accorgemmo che i 26 account fantoccio promuovevano gli stessi siti: Okay Dito, Ask Philippines e why0why.com, tra gli altri.

OKD2.com ha pubblicato un gran numero di bufale e altro materiale di propaganda a favore della famiglia di Marcos e del presidente Rodrigo Duterte. Ora si presenta come un sito di annunci pubblicitari, ma nel settembre del 2016 scoprimmo che i contenuti provenienti dal sito erano stati condivisi 11.900 volte su Facebook, in parte grazie agli account fantoccio. [Nota del traduttore: al momento della traduzione di questo manuale, andando su OKD2.com compare un messaggio che annuncia la cessazione dell'operatività del sito e la messa in vendita del dominio]

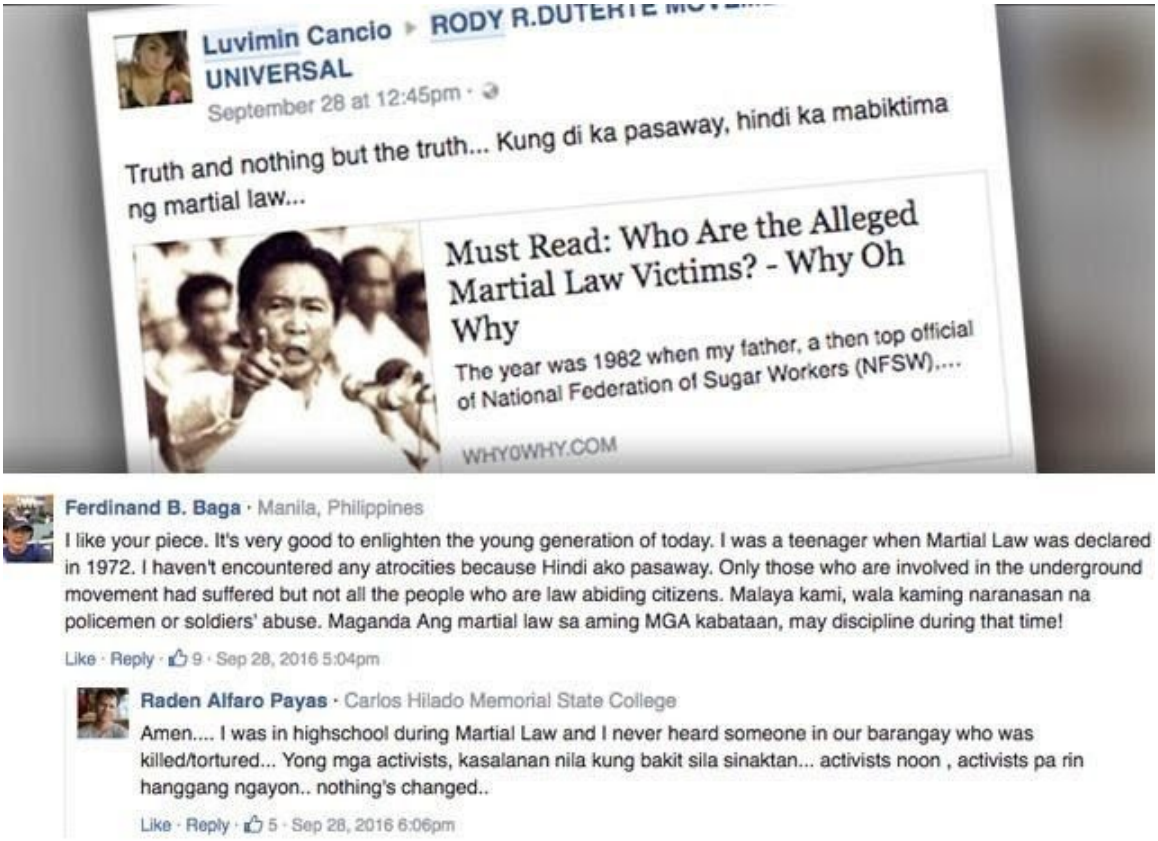
Attraverso questi siti Rappler rintracciò il potenziale burattinaio che muoveva i fili dei 26 account fantoccio: una persona di nome Raden Alfaro Payas.

Rintracciare i burattinai

Come accade con molti dei siti che Rappler monitora, i dati di registrazione del dominio di OKD2.com sono privati. Inoltre il sito non rivela i propri autori o proprietari, e non fornisce nessun contatto all'infuori di un modulo web.
Dallo storico delle registrazioni del dominio riuscimmo fortunatamente a identificare una persona associata al sito. Usando domaintools.com risalimmo al fatto che nel luglio del 2015 OKD2.com era registrato a nome di un tale Raden Payas, residente a Tanauan City, Batangas. Scoprimmo anche che l'ID di Google AdSense di OKD2.com era lo stesso di altri siti da cui i 26 account pubblicavano contenuti, ad esempio askphilippines.com e why0why.com. Individuammo l'ID di AdSense di questi siti visualizzando i codici sorgente delle loro pagine e cercando al loro interno serie numeriche che cominciavano con le lettere “ca-pub-”. A ogni account di Google AdSense è infatti assegnato un codice ID unico che comincia con “ca-pub-” e ogni pagina di un sito legato a quell’account ha al proprio interno quel codice.

Oltre ai dati di registrazione del dominio notammo anche che uno dei 26 account portava il nome di Raden Alfaro Payas (Unofficial). Scoprimmo inoltre che questo nome compariva anche nello username di un altro account, “realradenpayas”, il quale aveva interagito con alcuni degli account fantoccio.

L'account aveva ad esempio commentato un post di Luvimin Cancio in cui era linkata una storia che negava le atrocità avvenute sotto la legge marziale di Marcos. Il “vero” account di Payas scrisse che durante gli anni della legge marziale lui era al liceo e non aveva “mai sentito” di nessuno che fosse stato ucciso o torturato.



Alimentare lo Sharktank

La vicenda dei 26 account falsi e la loro portata d'azione spinsero Rappler a creare un database e una raccolta di dati automatica da gruppi pubblici e pagine pubbliche di Facebook. Ad agosto 2019 Rappler aveva tracciato circa 40.000 pagine con milioni di follower.

Da quella che era iniziata come un'indagine su un gruppo di account sospetti è nato uno studio continuativo su una rete di migliaia di account, gruppi e pagine, veri e finti, che diffondono disinformazione e propaganda, distorcendo lo scenario politico e indebolendo la democrazia di una nazione.

1b. Caso di studio: Come abbiamo dimostrato che la più grande pagina Facebook del movimento Black Lives Matter era un fake

Scritto da Donie O'Sullivan

Donie O'Sullivan *è un giornalista della CNN che si occupa delle aree in cui tecnologia e politica si incontrano. Fa parte del team Business della CNN e lavora a stretto contatto con l'unità investigativa dell'emittente per scoprire e ricostruire le campagne di disinformazione online dirette all'elettorato americano.*

Nell'estate e nell'autunno del 2017, mentre il mondo iniziava a scoprire i dettagli del vasto sforzo da parte della Russia di influenzare gli elettori americani tramite i social media, divenne chiaro che gli afroamericani e il movimento Black Lives Matter erano tra i principali bersagli della campagna del Cremlino per seminare divisione.

Io e i miei colleghi alla CNN abbiamo passato mesi a indagare e scrivere su come la Russia si fosse infiltrata in alcuni dei più grandi account social del movimento Black Lives Matter (BLM). Qualche volta, parlando con gli attivisti di BLM, mi sono sentito chiedere: "Sapete chi gestisce la più grande pagina Facebook di Black Lives Matter?"

Incredibilmente nessuno sapeva rispondere, nemmeno i più importanti attivisti di BLM del Paese e gli organizzatori sul territorio. Qualcuno aveva comprensibilmente sospettato che la pagina fosse gestita dalla Russia. Ma la nostra indagine ha dimostrato che dietro alla pagina non c'erano né i russi né gli americani, ma che era gestita da un uomo bianco che si trovava in Australia.

La pagina, intitolata semplicemente "Black Lives Matter", sembrava legittima. Nell'aprile 2018 contava quasi 700mila follower. Condivideva regolarmente link a storie sulle ingiustizie e sulle brutalità della polizia, lanciava raccolte fondi online e aveva anche uno store online che vendeva merchandise targato BLM.

Non è insolito che una pagina di queste dimensioni sia gestita in forma anonima. Alcuni attivisti preferiscono non mettere il proprio nome sulla pagina per non rischiare di attirare le attenzioni dei troll o i controlli delle forze dell'ordine che vogliono spegnere le proteste. Per gli attivisti fuori dagli Stati Uniti la possibilità di gestire pagine Facebook in forma anonima è stata essenziale per fare attivismo digitale, e ha avuto un ruolo chiave in alcuni movimenti (ed è esattamente ciò che la Russia ha sfruttato, facendo aumentare i sospetti che la pagina di BLM fosse a lei riconducibile).

Più o meno mentre iniziavo a prestare attenzione a questa pagina misteriosa, fui contattato da Jeremy Massler, investigatore freelance e incredibile segugio online, che mi diede una dritta. Massler aveva studiato i dati di registrazione dei domini dei siti a cui l'immensa pagina Facebook di BLM attingeva i contenuti che linkava regolarmente. Sebbene quasi tutti i domini fossero registrati in forma privata, scoprì che per un periodo del 2016 uno di loro era appartenuto a un certo Ian MacKay di Perth, in Australia. Un bianco.

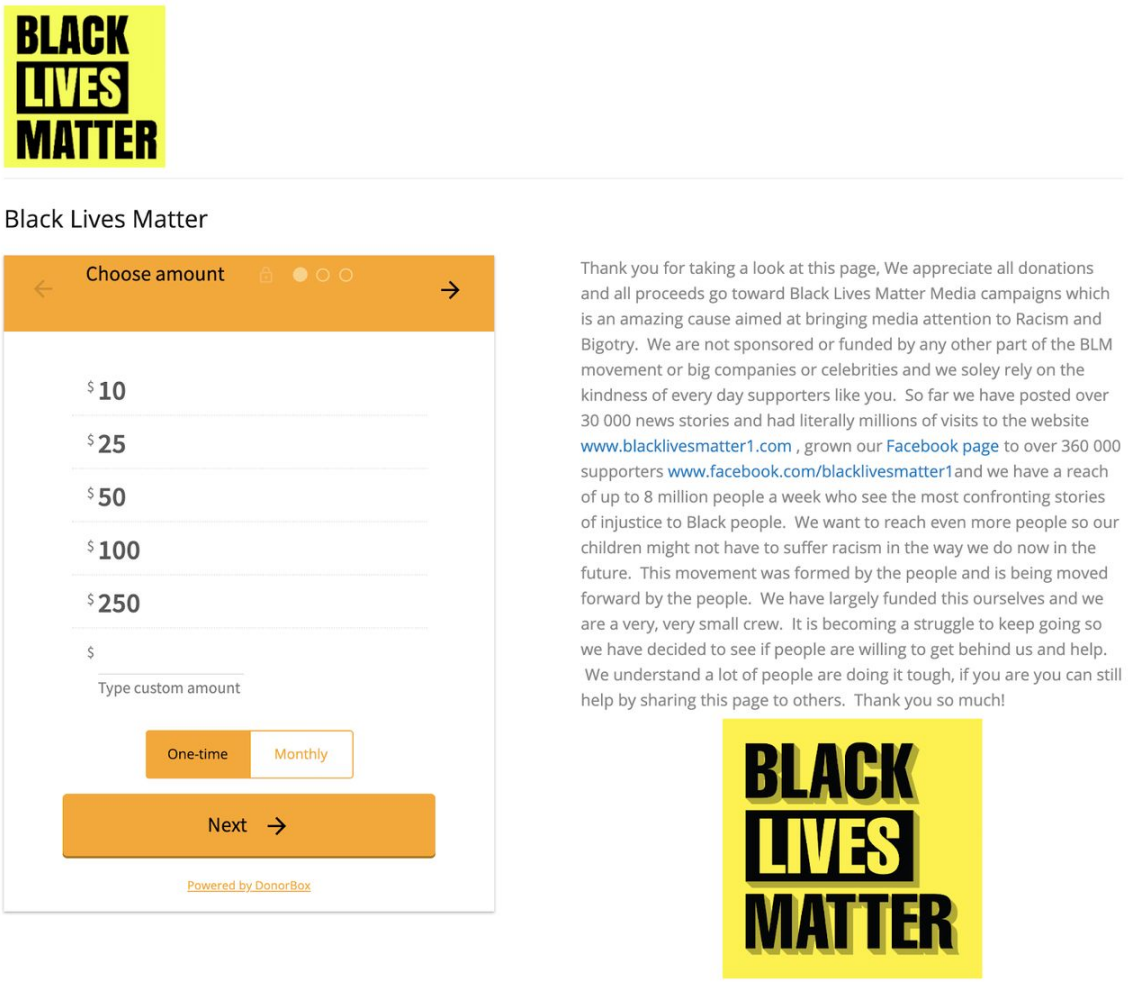
```
Domain Name: BLACKLIVESMATTERWEBSITE.COM
Registry Domain ID: 2065833077_DOMAIN_COM-VRSN
Registrar WHOIS Server: whois.launchpad.com
Registrar URL: LaunchPad.com
Updated Date: 2018-10-13T08:00:42Z
Creation Date: 2016-10-13T07:10:33Z
Registrar Registration Expiration Date: 2018-10-13T07:10:33Z
Registrar: Launchpad, Inc. (HostGator)
Registrar IANA ID: 955
Domain Status: clientTransferProhibited https://icann.org/epp#clientTransferProhibited
Registry Registrant ID: Not Available From Registry
Registrant Name: ian mackay
Registrant Organization: Website
Registrant Street: [redacted]
Registrant City: brisbane
Registrant State/Province: Queensland
Registrant [redacted]
Registrant
Registrant
Registrant
Registrant
Registrant Fax Ext:
Registrant Email: blacklivesmatter1@hotmail.com
```

Massler contattò MacKay, che gli disse che comprava e vendeva domini per hobby e che non aveva niente a che fare con quella pagina Facebook. Questa fu la stessa giustificazione che MacKay, funzionario sindacale di mezza età, diede a me quando lo raggiunsi al telefono qualche mese dopo. Ma nel frattempo avevamo scoperto che MacKay aveva registrato dozzine di domini web, molti dei quali correlati all'attivismo afroamericano.

Nonostante i miei dubbi su quella pagina e i sospetti che molti attivisti nutrivano su di essa, la spiegazione di MacKay non mi sembro assurda di per sé. I nomi dei domini possono avere un certo valore economico e sono oggetto di una costante compravendita. Il fatto che MacKay avesse registrato e venduto anche domini non correlati all'attivismo nero rendeva il suo caso ancor più credibile. Ma poi successe qualcosa di strano. Pochi minuti dopo la mia conversazione con MacKay, la pagina Facebook venne chiusa. Non era stata rimossa da Facebook, ma da chi la gestiva; e non era nemmeno stata del tutto eliminata, soltanto rimossa temporaneamente. Sembrava ci fosse qualcosa di sospetto, così Massler ed io cominciammo a scavare più a fondo.

Durante il suo periodo di attività la pagina Facebook (che tornò online nelle settimane successive alla mia chiamata con MacKay) aveva promosso campagne di raccolta fondi dichiaratamente a favore delle cause difese da BLM.

In un'occasione la pagina aveva dichiarato di essere impegnata a raccogliere fondi per gli attivisti di Memphis, in Tennessee. Ma quando parlai con gli attivisti di quella città, nessuno sapeva nulla della raccolta fondi, né su dove fosse finito il denaro raccolto. Altri attivisti ci raccontarono persino che avevano segnalato la pagina a Facebook, sospettando si trattasse di una truffa. Ma l'azienda non aveva preso nessuna iniziativa.



Quando iniziai a contattare le varie piattaforme online di pagamento e di raccolta fondi su cui la pagina si era appoggiata, le aziende in questione iniziarono a rimuovere le campagne fondi dicendo che avevano infranto le loro regole. Nessuna piattaforma, appellandosi alla privacy degli utenti, mi fornì informazioni sulla destinazione a cui era indirizzato il denaro. Si tratta di una difficoltà comune: chiamando in causa le loro regole sulla privacy, le piattaforme e i servizi digitali rivelano raramente alla stampa i nomi o i recapiti dei possessori dei loro account.

Successivamente seppi da una fonte che aveva dimestichezza con alcuni dei pagamenti processati che almeno uno degli account era legato al conto di una banca australiana e a un indirizzo IP anch'esso australiano. Un’altra fonte mi disse che erano stati raccolti circa 100.000 dollari. Dal momento che le sole informazioni open source non bastano a ricostruire molte storie, e che truffatori e malintenzionati stanno diventando sempre più abili, è sempre più importante dotarsi di fonti interne alle aziende tecnologiche disposte a darvi più informazioni di quanto le aziende stesse non siano pronte a fare.

Comunicai queste informazioni a Facebook per avere un commento sulla storia, dicendo loro che avevo prove che la pagina fosse collegata all'Australia, che le piattaforme di pagamento, dopo aver indagato, avevano rimosso le campagne di raccolta fondi e che sapevamo che parte del denaro era finita in Australia. Un portavoce di Facebook mi rispose che le loro indagini interne "non avevano trovato nulla che contravvenisse gli standard della nostra community".

Fu solo poco prima che la nostra storia venisse pubblicata, e solo dopo aver espresso a una figura di maggior livello nell'azienda le mie perplessità sulle indagini svolte da Facebook e sulla risposta ricevuta dal suo portavoce, che Facebook decise di agire e rimosse quella pagina.

Dopo l'uscita dell'articolo della CNN, il sindacato australiano per cui lavorava MacKay avviò un'indagine interna. Entro la fine della settimana aveva licenziato MacKay e un secondo funzionario che sosteneva essere a sua volta coinvolto nella truffa.

Un aspetto particolarmente rilevante in questa storia è la varietà di tecniche che Massler e io abbiamo usato per venirne a capo. Abbiamo fatto largamente ricorso ad archivi di siti online, come Wayback Machine, che ci hanno permesso di vedere come apparivano i siti a cui la pagina linkava e anche come si presentava la pagina stessa prima di finire nei nostri radar. Ciò si rivelò particolarmente utile, dal momento che da dopo il primo contatto di Massler con MacKay, chi gestiva la pagina iniziò a coprire le proprie tracce.

Per indagare sui siti che MacKay aveva registrato e trovare un contatto diretto con lui abbiamo sfruttato servizi che tengono traccia delle registrazioni dei domini, come DomainTools.com. Inoltre, per individuare i profili fake di Facebook creati per promuovere la pagina nei gruppi, Massler fece ampio uso dello strumento di Facebook Graph Search (non più disponibile). Le informazioni open source e gli strumenti di ricerca online, come quelli usati per accedere ai dati dei domini, sono strumenti cruciali, ma non sono gli unici.

Per smascherare questa truffa sono state fondamentali anche le tecniche del giornalismo tradizionale, come il semplice gesto di alzare il telefono per parlare con MacKay o coltivare fonti per ottenere informazioni che altrimenti non sarebbero state rese pubbliche.

2. Trovare il paziente zero

Scritto da Henk van Ess

***Henk van Ess** è consulente del Poynter's International Fact-Checking Network. Tirare fuori storie dai dati è la sua ossessione. Forma professionisti dei media di tutto il mondo sui temi delle ricerche su internet, dei social media e della multimedialità. Tra i suoi clienti ci sono NBC News, BuzzFeed News, ITV, Global Witness, SRF, Axel Springer, SRF e numerose ONG e università. I suoi siti, whopostedwhat.com e graph.tips, sono ampiamente usati per scandagliare i social media. Su Twitter è @henkvaness .*

Per decenni l'assistente di volo di Air Canada Gaëtan Dugas è stato ricordato come il "paziente zero", l'uomo che per primo portò l'AIDS negli Stati Uniti. Questa definizione, sostenuta da libri, film e un numero imprecisabile di articoli lo fece diventare il "supercattivo di un'epidemia che avrebbe ucciso più di 700.000 persone in Nord America".

Ma le cose non stavano così. Bill Darrow, un ricercatore impegnato con i Centers for Disease Control and Prevention, intervistò Dugas e lo catalogò come "Paziente O", dove la O stava per per "Out-of-California". Ma la O fu ben presto confusa con il numero 0, innescando una reazione a catena di cattiva informazione durata fino a tempi recenti.

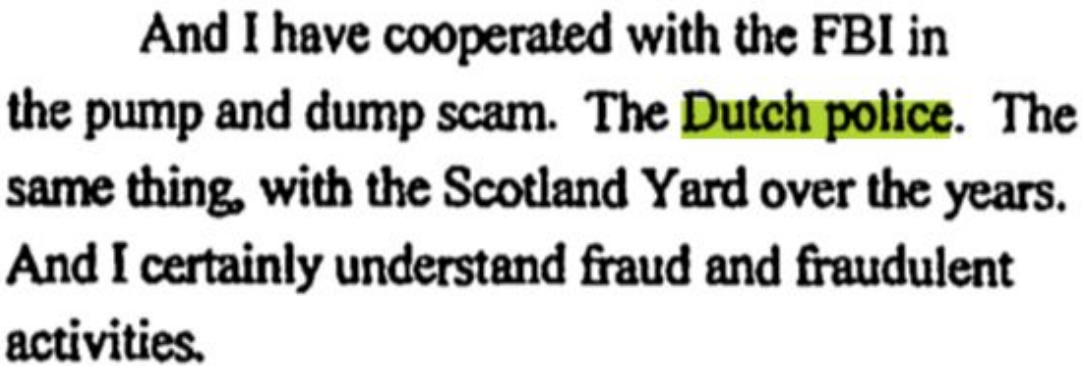
Anche a un giornalista può capitare di sbagliarsi nell'identificare il suo paziente 0, se non sa qual è la maniera corretta di cercarlo. Questo capitolo ti aiuterà a imparare a trovare online le fonti primarie di informazione, ignorando i risultati superficiali e scavando più a fondo nel web.

1. I rischi della consultazione di fonti primarie e come scongiurarli

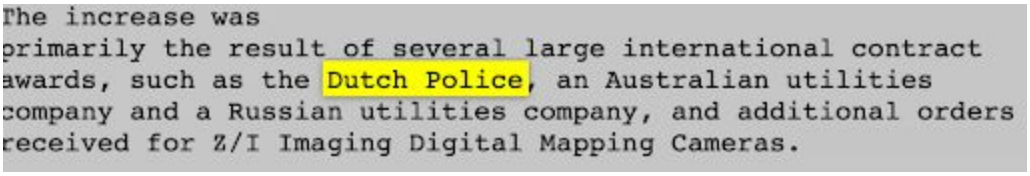
I giornalisti adorano le fonti primarie online. Le prove di prima mano che cercano possono trovarsi in un articolo di giornale, in uno studio scientifico, in un comunicato stampa, sui social o in qualsiasi altro possibile "paziente zero".

Facendo una ricerca base per parole chiave dentro un sito governativo ufficiale potresti essere portato a pensare che "quel che vedi è tutto quel che c'è". Ma spesso non è così, ed eccone un esempio. Andiamo sul sito della U.S. Securities and Exchange Commission, una fonte usata per cercare informazioni fiscali su cittadini statunitensi, ma anche su uomini d'affari di tutto il mondo. Poniamo di voler cercare la prima occorrenza dell'espressione "Dutch police" (polizia olandese) su sec.gov. Il motore di ricerca interno del SEC può esserci d'aiuto:

Otteniamo un solo risultato, un documento del 2016. Quindi la polizia olandese è citata nel SEC una sola volta, nel 2016, giusto?



Sbagliato. La prima menzione "Dutch police" su sec.gov risale al 2004, 12 anni prima, ed è contenuta in una mail desecretata e criptata:



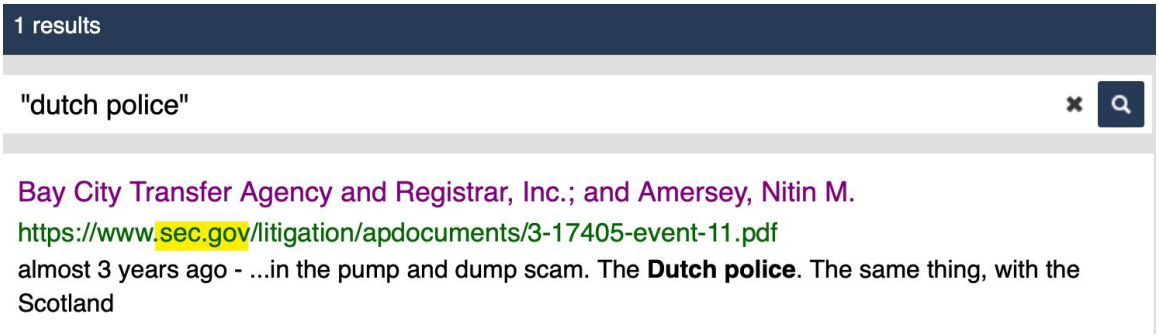
Non vedrai questo risultato tra quelli proposti dalla ricerca interna di sec.gov, nonostante l'informazione provenga proprio dallo stesso sito. Perché questa discrepanza?

Meglio non confidare mai nei motori di ricerca interni di siti che costituiscono fonti primarie. Possono dare una falsa impressione del vero contenuto di un sito e dei database associati. Il modo corretto per cercare è eseguire un primary source check, un controllo della fonte primaria.

Controllo della fonte primaria

Passo 1: Esamina il link che non funziona

La ricerca sul SEC ci ha restituito un solo risultato:



Proviamo a cavarne qualcosa di buono. Per prima cosa, leviamo di torno "https://www.", la prima parte del link. Concentriamoci sulla prima slash dopo quella parte, in questo caso prima della parola "litigation/".

Questa è la parte che ci serve: sec.gov.

Passo 2: Usa l'operatore di ricerca "site:"

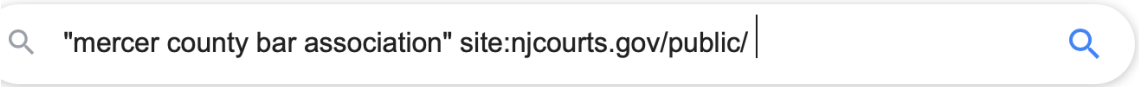
Apri un generico motore di ricerca. Digita per prima cosa la tua query ("Dutch police") e per ultima la parola "site:" seguita direttamente dall'URL (senza spazi). Questa formula serve a scoprire se la ricerca sul sito della fonte primaria ha mostrato tutto ciò che c'è nel sito:

Includere cartelle specifiche

A questo punto puoi adattare la "formula delle fonti primarie" alle tue esigenze. Andiamo nella sezione dei comunicati stampa del sito della New Jersey Courts. Immaginiamo che tu voglia scoprire quando la Mercer County Bar Association abbia sponsorizzato un programma per il Law Day (La Giornata del Diritto, celebrata il 1° maggio), ma non riesci a rintracciare l'informazione di prima mano in nessuno dei titoli dei comunicati stampa. La "Mercer County Bar Association" non compare in nessun titolo.

Ora guarda la URL di quella pagina piena di comunicati stampa mal indicizzati:

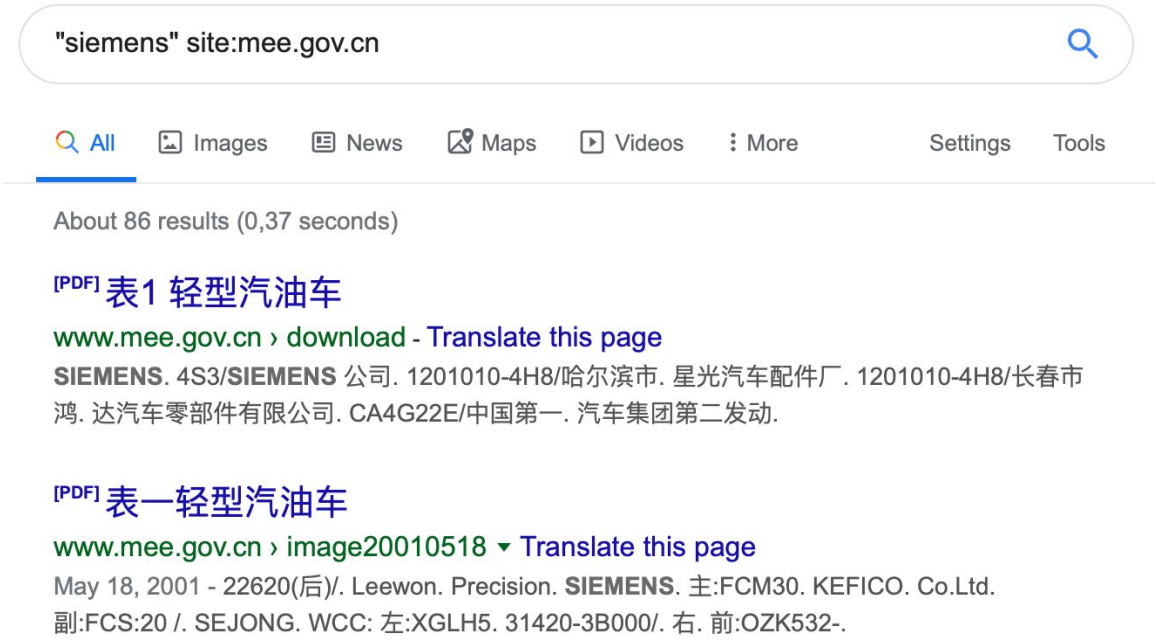
Il materiale delle pubbliche relazioni è archiviato nella cartella /public. Questa indicazione va inclusa nella tua ricerca su Google:



Ed ecco qua:

Sapere in quale cartella cercare

La Cina ha un Ministero dell'Ecologia e dell'Ambiente. Avranno documenti in inglese sull'azienda tedesca Siemens? Con la seguente formula la ricerca ti presenterà risultati sia in cinese che in inglese:



A questo punto occorre filtrare i risultati per vedere solo quelli in inglese. Non è che al Ministero hanno usato la parola English nel link? Fai una prova. Funziona:

2. Seguire le tracce dei documenti

A volte le informazioni che ci servono non sono contenute in una pagina web, ma in un documento ospitato in un sito. Ecco come seguire la traccia di un documento usando le formule di Google.

Ross McKitrick è un professore associato del dipartimento di Economia della University of Guelph, in Ontario. Nel 2014 tenne una presentazione per un gruppo di persone scettiche sui cambiamenti climatici. Proviamo a cercare l'invito a quell'incontro. Sappiamo che si è tenuto il 13 maggio del 2014, in occasione dell'11esimo pranzo annuale organizzato dal gruppo "Friends of Science (FOS)." Se cerchiamo su Google questi termini non troviamo nulla.

> No results found for **"Friends of Science 11th Annual Luncheon 2014" "invitation"**.

Perché? Perché la parola "invitation", o invito, non compare scritta su molti inviti. Lo stesso vale per la parola *interview:* molte interviste non contengono la parola *interview* o *intervista*. Anche su molte mappe non è esplicitamente scritta la parola *map*, o *mappa*. Il mio consiglio? Smetti di tirare a indovinare e segui questi passi.

Passo 1: Stabilisci il tipo di documento

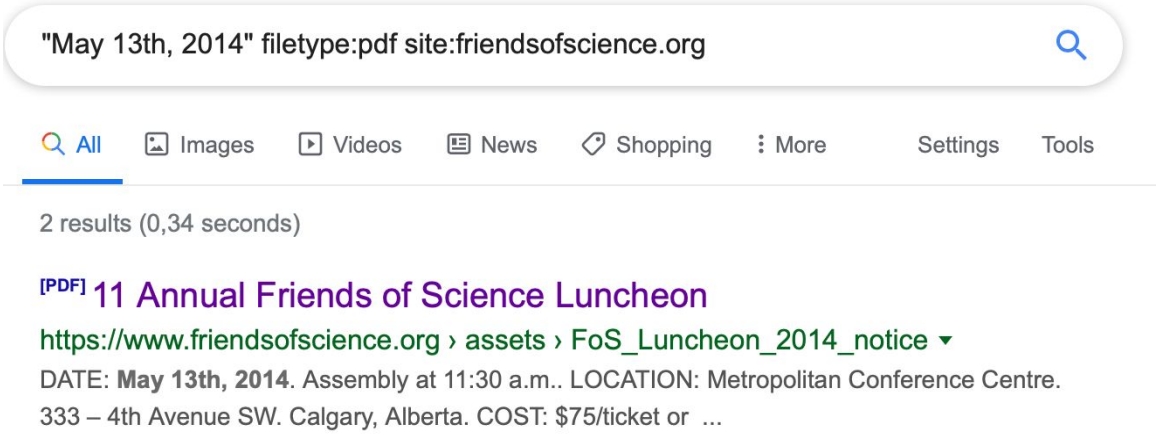
Cerca di stabilire qual è il denominatore comune di tutti gli inviti online: ovvero, che spesso si tratta di documenti PDF. Cerca proprio questo tipo di file usando l'operatore "filetype:pdf" e potresti avvicinarti a ciò che cerchi.

Passo 2: Sii (climaticamente) neutro

Anche se non conosci le esatte parole usate nell'invito, sai per certo che il video su YouTube mostrava un evento del 13 maggio 2014. Probabile che l'invito contenga questa data (se cerchi in inglese fa attenzione a cercare la data sia nella forma ordinale che in quella cardinale, ovvero May 13 e May 13th).

Passo 3: Chi è coinvolto?
Sappiamo che l'organizzatore dell'evento è "Friends of Science", e che il suo sito è friendsofscience.org.

Ecco come appare la ricerca su Google mettendo insieme questi tre passi:



Ed ecco che troviamo l'invito al primo colpo.

Il FOS, con sede a Calgary, è spesso considerato un gruppo negazionista del cambiamento climatico ed è in parte sovvenzionato dal settore dei combustibili fossili. Come dovremmo dunque costruire la nostra query per trovare più informazioni su di loro e sulla loro rete di sostenitori e finanziatori?

Passo 1: Restringi il campo al tuo target

“Friends of Science” ci porta troppi risultati, quindi includiamo anche “Calgary”.

Passo 2: Includi “filetype”

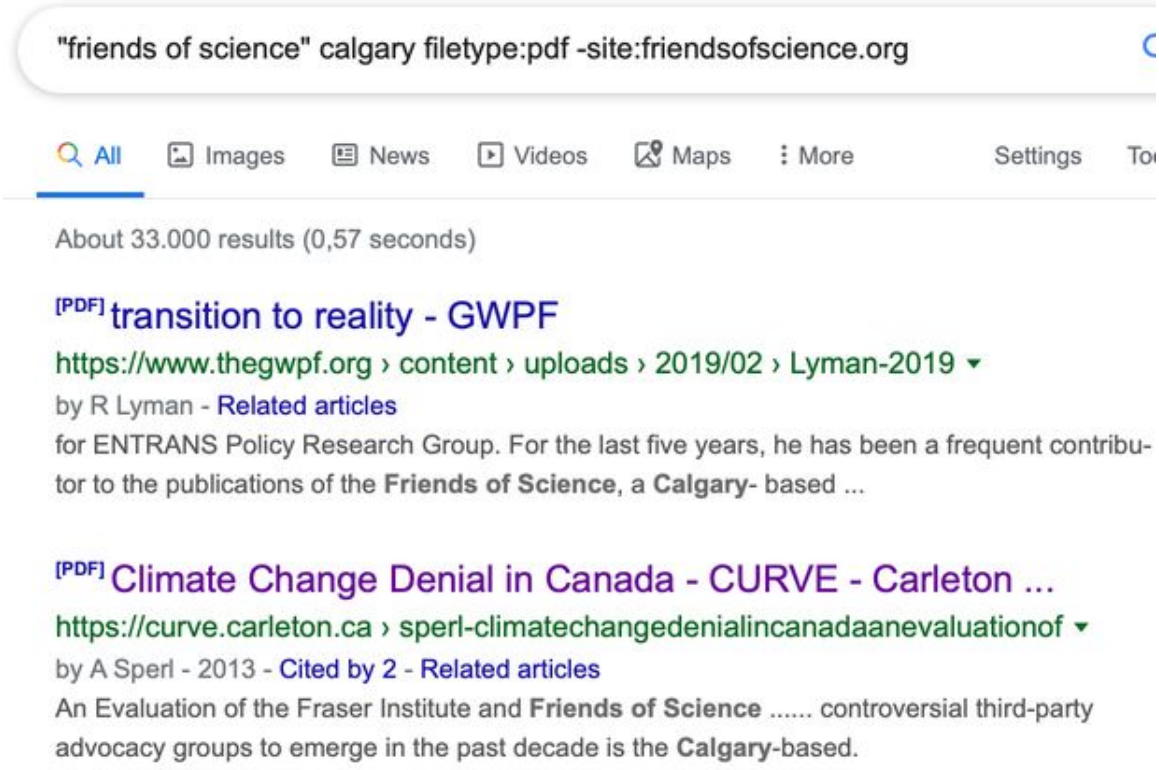
Usa l'operatore migliore per trovare i documenti ufficiali, “filetype:pdf”.

Passo 3: Escludi il sito del tuo target

Escludi dalla ricerca il sito del tuo target, in questo caso Friendsofscience.org, aggiungendo “-site:friendsofscience.org”. Così facendo troverai informazioni da altre fonti.

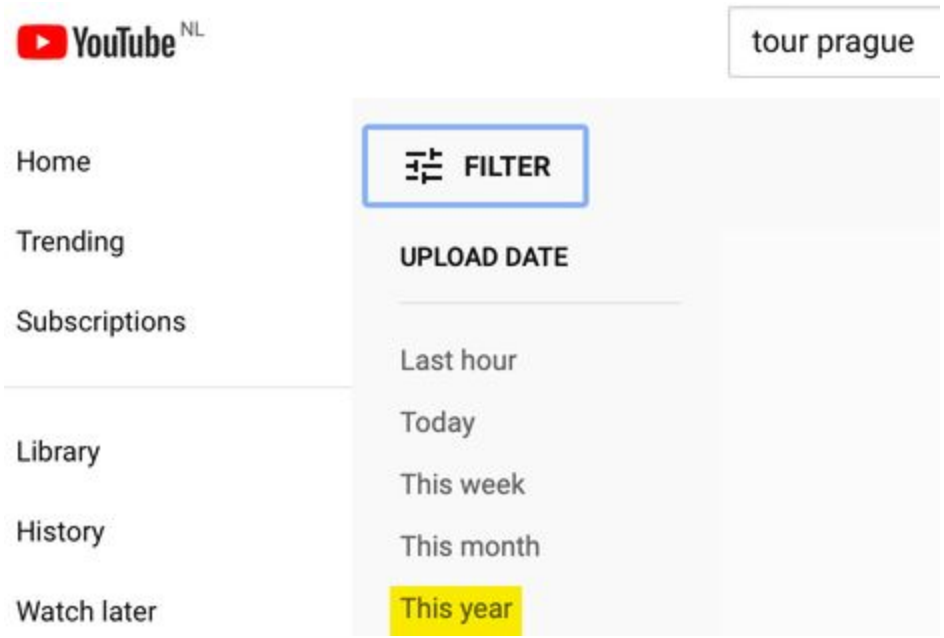
La query completa è questa:

Visto che hai impostato la ricerca per trovare il tuo target in documenti ufficiali, ma non in quelli del suo stesso sito, troverai risultati provenienti sia dai sostenitori di Friends of Science, sia dai critici verso l’organizzazione.

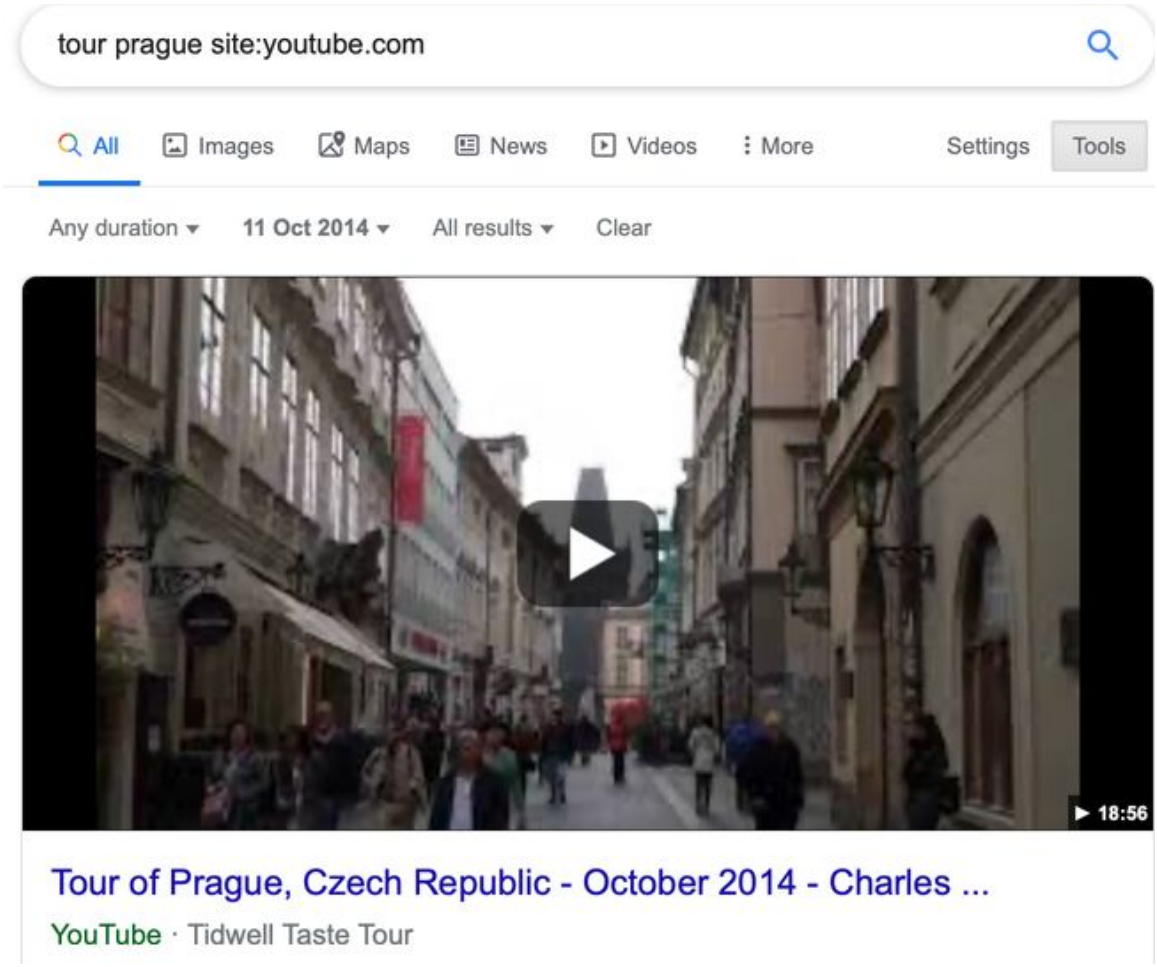


3. Scandagliare i social media alla ricerca di fonti primarie

YouTube
Lo strumento di ricerca interna di YouTube ha un problema: non permette di filtrare la ricerca per video più vecchi di un anno. Se vuoi trovare un video di un tour di Praga dell'11 ottobre 2014 ti troverai di fronte questo sbarramento:



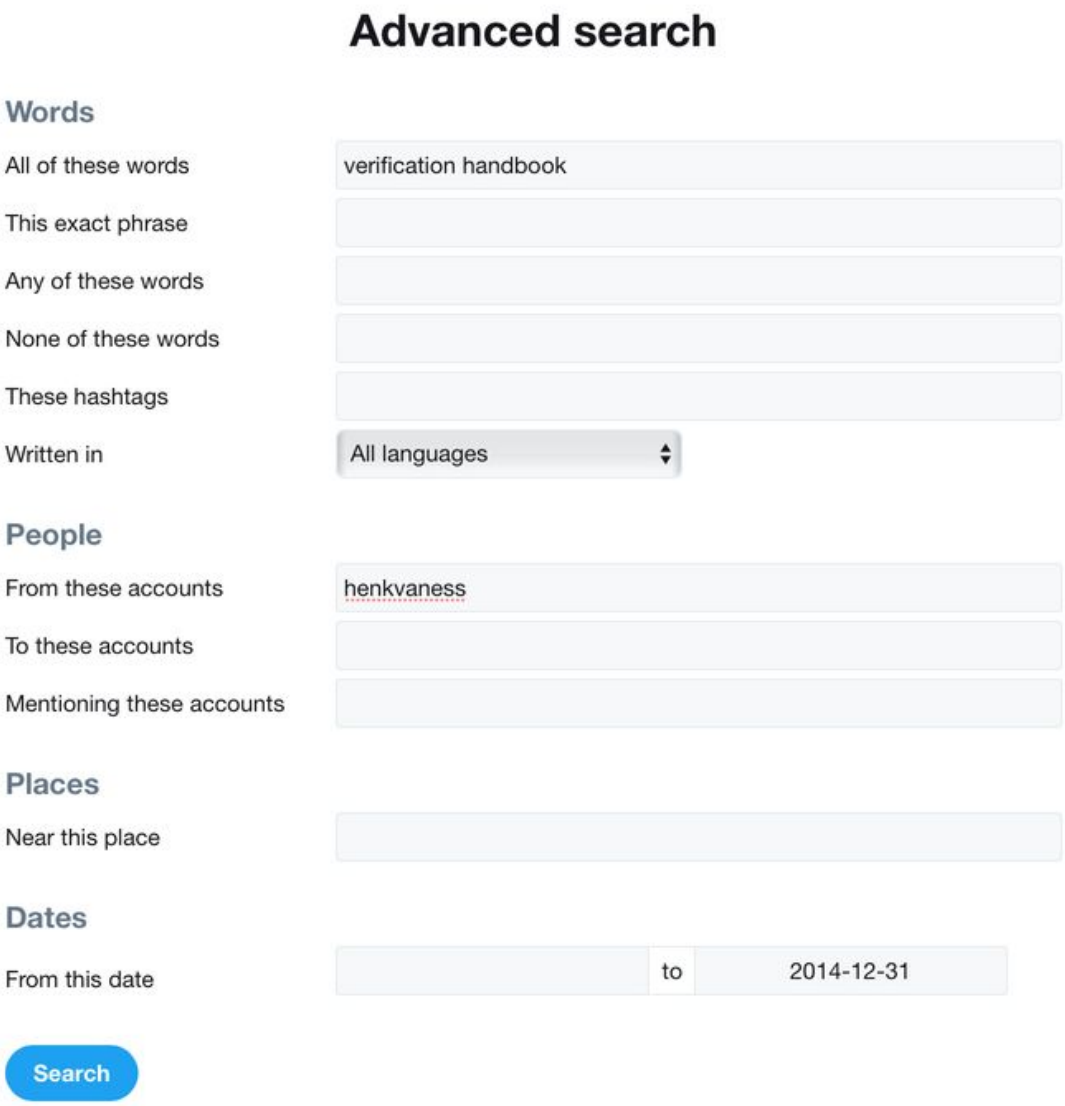
Per risolvere il problema, inserisci manualmente la data che stai cercando in una ricerca su Google.com. Vai su Google.com e clicca su "Strumenti", che trovi sotto la barra di ricerca sull'estrema destra. Dal menù che si apre seleziona “Qualsiasi data” e poi “Intervallo di date”. Inserisci l'intervallo che ti interessa. Ora abbiamo i risultati che ci servono.



Twitter
Nonostante le potenzialità dell'operatore di ricerca “site:”, resterai deluso usandolo su Google per cercare risultati dentro Twitter. Proviamo ad esempio a usare la seguente query per scoprire a quando risale il mio primo tweet sul Verification Handbook:

"verification handbook" site:twitter.com/henkvaness

Nel momento in cui scrivo, questa ricerca restituisce un solo risultato. Motori di ricerca generici come Google spesso fanno fatica a reperire risultati di qualità dalle migliaia di miliardi di post su Twitter o su grandi piattaforme come Facebook o Instagram. La soluzione per Twitter è usare la sua funzione di ricerca avanzata aggiungendo parole chiave, nomi utente e un periodo di tempo, come mostrato qua:



Non dimenticare di cliccare su “Recenti” sul menù in alto nella pagina dei risultati della ricerca, così potrai vedere i risultati in ordine cronologico, dal più recente al più vecchio. Per impostazione predefinita, infatti, Twitter ordina i risultati a partire da quelli che considera i più popolari.

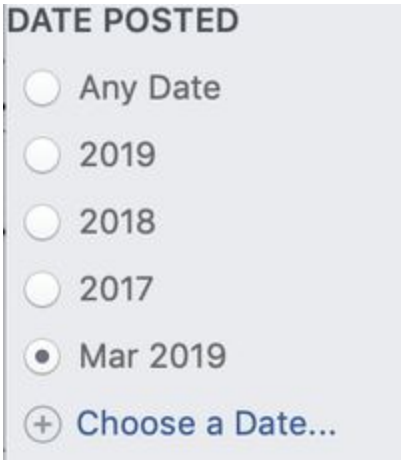
Facebook
Usare “site:” per scandagliare Facebook non è il massimo, ma possiamo fare in modo di adattare lo strumento di ricerca interno del social alle nostre esigenze. Immaginiamo ad esempio di voler trovare post sulle torte alla fragola scritti da persone che abitano a Brooklyn a marzo 2019. Segui questi passi:

Passo 1: Digita una query

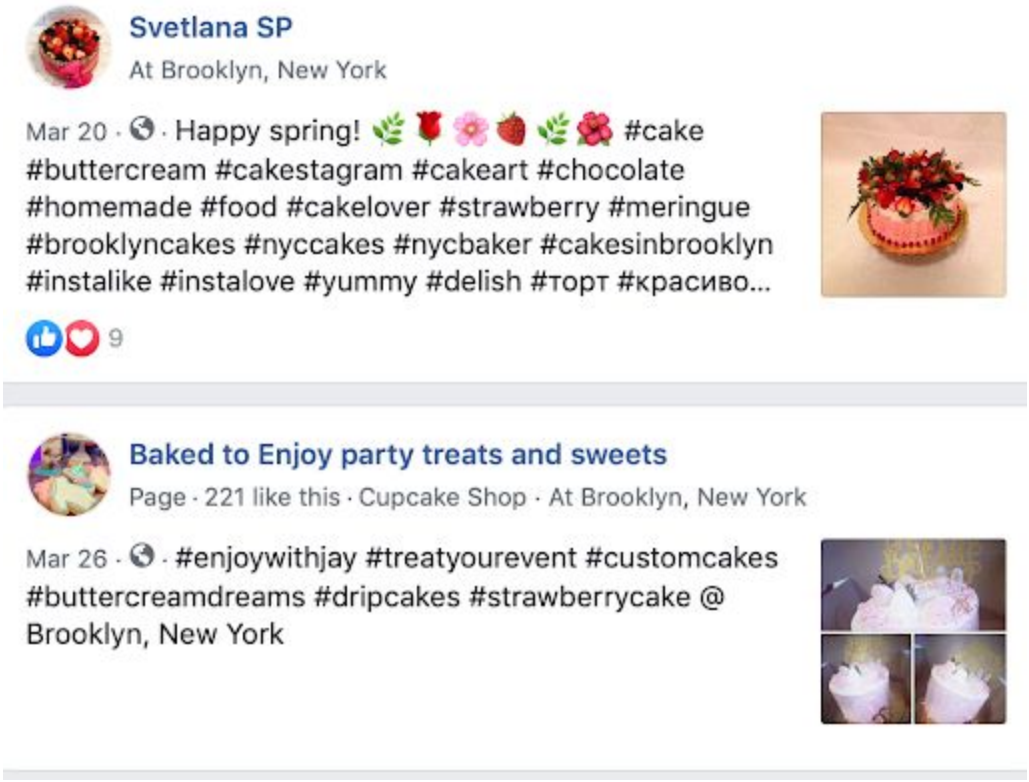
Passo 2: Clicca su "posts"

Passo 3: Definisci dove vuoi cercare

Passo 4: Scegli una data

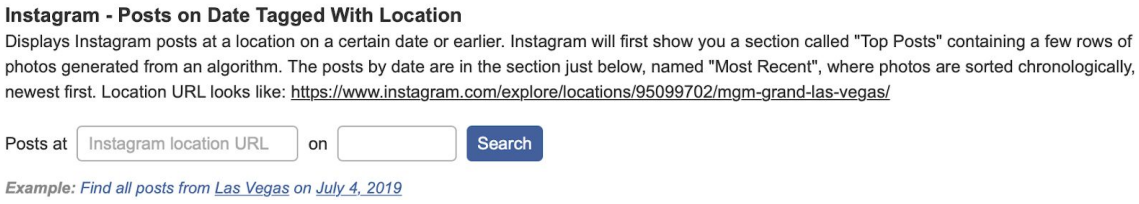


Ed ecco qua i tuoi post:



Instagram

Per cercare su Instagram post pubblicati in una data specifica in una precisa località puoi andare sul mio sito, whopostedwhat.com , e riempire i campi richiesti per costruire la tua ricerca:



3. Riconoscere bot, cyborg e attività non autentiche

Scritto da Johanna Wild e Charlotte Godart

***Charlotte Godart** è investigatrice e formatrice a Bellingcat. Prima di Bellingcat era impegnata presso lo Human Rights Center dell'Università di Berkeley, per cui lavorava all'interno dell'Investigations Lab e insegnava agli studenti a condurre ricerche open source sui conflitti globali per organizzazioni umanitarie internazionali.*

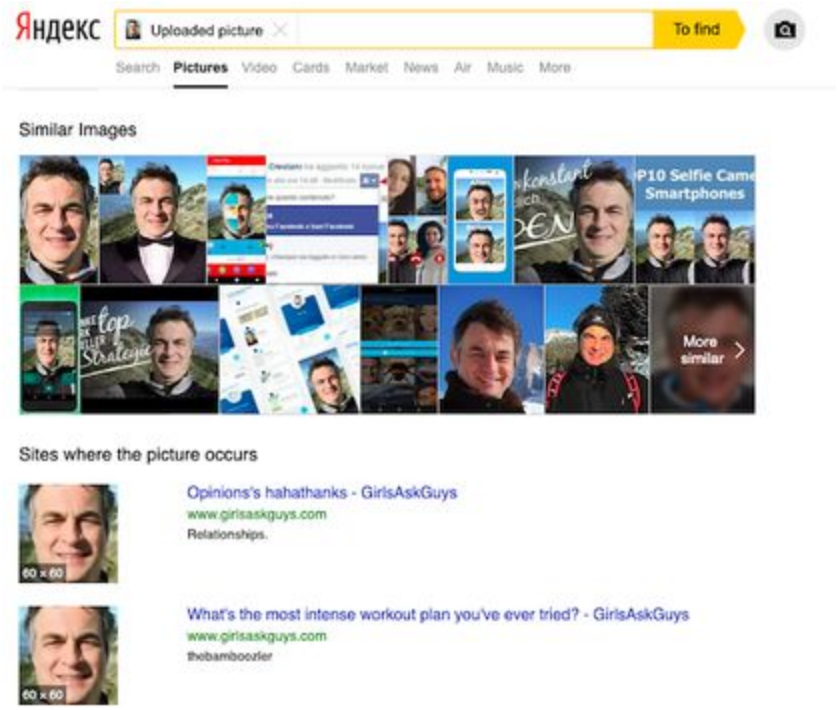
***Johanna Wild** è investigatrice open source a Bellingcat, dove si occupa anche di sviluppo di tecnologie e strumenti per indagini digitali. Proviene dal mondo del giornalismo online e precedentemente ha lavorato con altri giornalisti in zone di conflitto e post-conflitto. Tra i ruoli che ha ricoperto c'è stato quello di supporto a giornalisti impegnati in Africa Orientale a produrre trasmissioni e contenuti per The Voice of America.*

Alla fine dell'agosto 2019 Benjamin Strick, collaboratore di Bellingcat e di BBC Africa EYE, stava analizzando i tweet che diffondevano gli hashtag #WestPapua e #FreeWestPapua quando si rese conto che alcuni account avevano un comportamento anomalo. Questi account stavano tutti diffondendo messaggi a favore del governo indonesiano in un momento in cui il conflitto nella provincia della Papua Occidentale stava guadagnando visibilità internazionale: un movimento di indipendenza locale era sceso in strada per rivendicare la libertà dal controllo indonesiano, e tra polizia indonesiana e manifestanti stava scoppiando la violenza.

Gli account osservati da Strick avevano in comune molte stranezze. Il giornalista avrebbe ben presto capito che quelle somiglianze erano i primi segnali di un comportamento non autentico coordinato. Inizialmente, tuttavia, si limitò a notare alcune piccole cose.

Per prima cosa, molti degli account avevano foto profilo rubate. Guarda ad esempio questo account, che dichiara di appartenere a un certo Marco:

Usando lo strumento di ricerca inversa delle immagini di Yandex, Strick scoprì che la foto profilo dell'account era stata usata in precedenza in altri siti e sotto altri nomi. Nessuno degli account che usavano quella foto apparteneva ad una persona reale di nome Marco. Ciò provava che ciascuno di quegli account stava quantomeno fingendo una falsa identità.



Oltre alle false identità, Strick scoprì anche che gli account pubblicavano contenuti simili o addirittura identici, e che spesso si retwittavano a vicenda. Cosa ancor più sorprendente, alcuni degli account erano perfettamente sincronizzati nelle

tempistiche con cui pubblicavano i loro tweet. Per esempio, @bellanow1 e @kevinma40204275 pubblicavano i loro tweet al minuto 7 e al minuto 32 di ogni ora.

26/8/19	17:07:37	bellanow1	26/8/19	23:07:20	kevinma40204275
26/8/19	5:27:06	bellanow1	26/8/19	21:32:52	kevinma40204275
26/8/19	5:27:06	bellanow1	26/8/19	20:32:52	kevinma40204275
26/8/19	5:27:05	bellanow1	26/8/19	18:32:51	kevinma40204275
26/8/19	5:27:04	bellanow1	26/8/19	15:07:22	kevinma40204275
26/8/19	5:27:04	bellanow1	26/8/19	12:32:54	kevinma40204275
26/8/19	3:32:55	bellanow1	26/8/19	9:32:54	kevinma40204275
26/8/19	0:32:56	bellanow1	26/8/19	5:32:54	kevinma40204275
26/8/19	0:07:33	bellanow1	26/8/19	5:07:36	kevinma40204275
25/8/19	23:32:54	bellanow1	26/8/19	3:32:54	kevinma40204275
25/8/19	22:32:53	bellanow1	26/8/19	0:32:54	kevinma40204275
25/8/19	22:07:06	bellanow1	25/8/19	23:32:52	kevinma40204275
25/8/19	20:32:53	bellanow1	25/8/19	23:07:16	kevinma40204275
25/8/19	10:07:19	bellanow1	25/8/19	19:32:53	kevinma40204275
25/8/19	9:32:56	bellanow1	25/8/19	15:07:24	kevinma40204275
25/8/19	9:07:27	bellanow1	25/8/19	10:32:55	kevinma40204275
25/8/19	8:32:56	bellanow1	25/8/19	7:32:55	kevinma40204275
25/8/19	7:07:23	bellanow1	25/8/19	6:32:54	kevinma40204275
25/8/19	6:32:56	bellanow1	25/8/19	6:08:01	kevinma40204275
24/8/19	13:07:57	bellanow1	25/8/19	3:07:21	kevinma40204275
24/8/19	10:07:19	bellanow1	25/8/19	0:07:26	kevinma40204275
24/8/19	7:32:56	bellanow1	24/8/19	20:32:51	kevinma40204275
24/8/19	7:07:20	bellanow1	24/8/19	20:07:08	kevinma40204275
24/8/19	5:32:56	bellanow1	24/8/19	19:32:51	kevinma40204275
24/8/19	4:32:56	bellanow1	24/8/19	15:07:24	kevinma40204275
24/8/19	0:07:31	bellanow1	24/8/19	13:32:55	kevinma40204275
			24/8/19	10:07:17	kevinma40204275
			24/8/19	7:32:54	kevinma40204275
			24/8/19	7:07:18	kevinma40204275
			24/8/19	5:32:54	kevinma40204275
			24/8/19	1:32:54	kevinma40204275

È poco probabile che un essere umano segua un ritmo di pubblicazione del genere. La sincronizzazione tra molteplici account, combinata all'uso di foto ingannevoli, spingeva a pensare che gli account non fossero collegati a persone reali e che agissero in maniera automatizzata. Studiando il comportamento di account sospetti come questi, Strick scoprì infine che gli account in questione facevano parte di una rete di bot pro Indonesia su Twitter, che stavano diffondendo informazioni faziose e ingannevoli sul conflitto in Papua Occidentale (potrai leggere di più circa il più esteso network di cui facevano parte questi account nel capitolo 11b, dedicato al caso di studio "Indagare su una operazione informativa in Papua Occidentale").

Che cos'è un bot? La risposta è più complicata di quanto tu creda

Il caso della provincia della Papua Occidentale è lungi dall'essere l'unica operazione informativa a usare i bot sui social. Altre operazioni si sono diffuse e sono state

analizzate ben più ampiamente. Ad ogni modo, tra queste e il caso della Papua Occidentale ci sono comunque delle somiglianze operative.

Un bot è un'applicazione software che può automaticamente portare a termine dei compiti assegnati da esseri umani. Se un bot agisce bene o male dipende interamente dalle intenzioni del suo "proprietario".

I bot ai quali si fa riferimento nei dibattiti pubblici sono bot social, attivi su social network come Facebook, Twitter e LinkedIn. Su queste piattaforme possono essere usati per diffondere messaggi ideologici, spesso con l'obiettivo di far sembrare che ci sia un coro di voci a sostegno di un dato argomento, individuo, contenuto o hashtag.

I bot dei social media sono tendenzialmente riconducibili a tre categorie principali: i bot programmati, i bot di guardia e i bot amplificatori. È importante sapere qual è il tipo di bot che ti interessa, perché ciascuno ha scopi diversi. Ogni scopo determina il linguaggio e il modello comunicativo del bot. Nel contesto della disinformazione, i bot che ci interessano di più sono i bot amplificatori.

I bot amplificatori servono a fare esattamente ciò che suggerisce il loro nome: amplificare e diffondere contenuti con l'obiettivo di influenzare l'opinione pubblica online. Possono anche essere usati per dare l'impressione che un'azienda o una persona abbiano più follower di quanti ne posseggano in realtà. Il potere di questi bot sta nel loro numero. Un network di bot amplificatori può tentare di influenzare l'andamento di un hashtag, diffondere link o contenuti visivi, scatenare spam di massa o prendere di mira una singola persona online per screditarla o farla passare come figura controversa o sotto attacco.

Il fatto che i bot amplificatori concorrano in gran numero a uno stesso obiettivo fa sembrare più legittimo ciò che fanno, motivo per cui riescono effettivamente a influenzare il panorama dell'opinione pubblica online. I bot amplificatori che diffondono disinformazione operano soprattutto attraverso campagne di hashtag , oppure condividendo notizie sotto forma di link, video, meme, foto o contenuti di altro tipo.

Le campagne di hashtag coinvolgono bot che twittano costantemente lo stesso hashtag , o una serie di hashtag, in maniera coordinata. Spesso il loro obiettivo è ingannare l'algoritmo che regola i trend di Twitter per riuscire a far aggiungere un hashtag alla lista dei trending topic. Un esempio è l'hashtag #Hillarysick, ampiamente diffuso dai bot dopo l'episodio del settembre del 2016 (poco prima delle elezioni presidenziali) in cui Hillary Clinton era inciampata. Ad ogni modo, è importante notare che le campagne di hashtag non hanno necessariamente bisogno dei bot e che possono essere più efficaci senza di essi. Leggi questa inchiesta sulle "fabbriche umane" di hashtag in Pakistan pubblicata su Dawn.

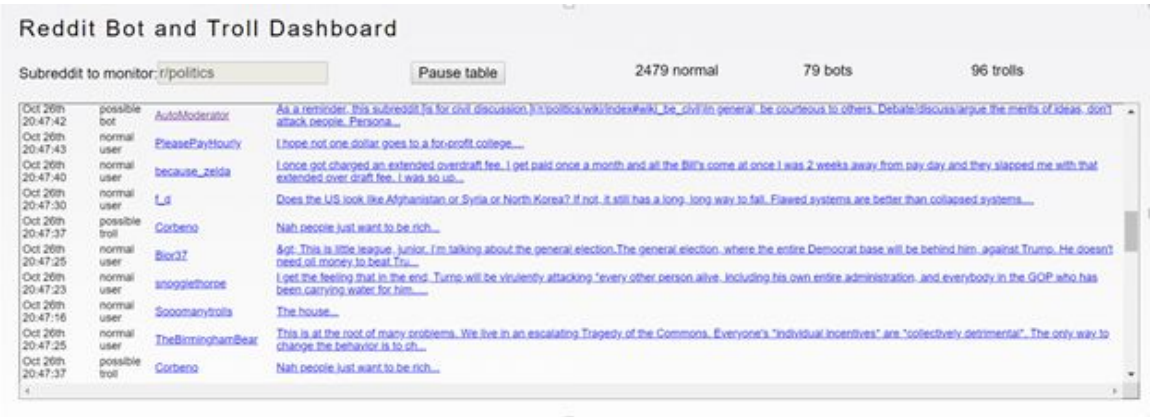
Comprare e creare bot è relativamente semplice. Ci sono innumerevoli siti che possono venderti il tuo esercito personale di bot per poche centinaia di dollari, o anche meno. È invece molto più difficile creare e mantenere una rete di bot sofisticata, che sembri umana.

Come riconoscere i bot

Sviluppatori e ricercatori hanno creato molti strumenti che aiutano a valutare se un account agisca in maniera automatizzata o meno. Questi strumenti possono essere molto utili per raccogliere informazioni, ma la valutazione fornita da un singolo strumento non può in nessun caso essere considerata definitiva, e non dovrebbe mai costituire l'unica base da cui trarre conclusioni o su cui scrivere un servizio giornalistico.

Uno degli strumenti più conosciuti è Botometer, creato dai ricercatori dell'Indiana University. Basandosi su vari criteri, lo strumento calcola un punteggio che indica la probabilità che un account Twitter e i suoi follower siano dei bot.

Jason Skowronski ha creato per Reddit una bacheca che si aggiorna in tempo reale la quale, dopo aver impostato il canale subreddit prescelto, stima se i commenti che vengono fatti sono opera di bot, troll o esseri umani.



Pur essendoci delle eccezioni, la maggior parte degli strumenti pubblicamente disponibili per riconoscere i bot è stata creata per Twitter. La ragione è che molti social network, incluso Facebook, limitano le funzionalità delle loro API (application programming interface) in maniera tale da rendere impossibile analizzare e usare i loro dati per creare strumenti del genere.

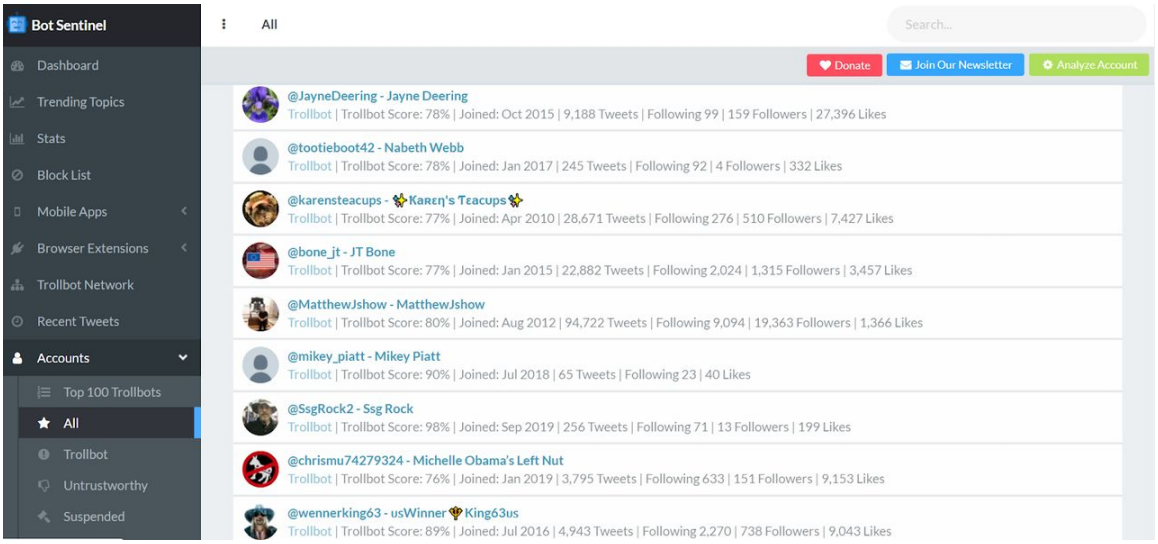
Come sottolineato in precedenza, gli strumenti per identificare i bot sono un ottimo punto di partenza, ma non dovrebbero essere l'unica prova su cui basarsi. Una delle ragioni per cui il loro livello di affidabilità è variabile è che, semplicemente, non esistono criteri universalmente validi che garantiscano di riconoscere un bot con una certezza del 100%. Inoltre, c'è scarso accordo su come classificare qualcosa come bot o meno. I ricercatori del Computational Propaganda Project dell'Oxford Internet Institute classificano come "altamente automatizzati" account che postano più di 50 volte al giorno. Il Digital Forensics Research Lab dell'Atlantic Council considera invece "72 tweet al giorno (uno ogni dieci minuti per dodici ore di seguito) come sospetti, e oltre 144 tweet al giorno come altamente sospetti".

Spesso può essere complicato stabilire se una campagna di disinformazione sia condotta da bot social o da essere umani che, per proprie ragioni o perché pagati, pubblicano un gran numero di contenuti su un determinato argomento. La BBC, per esempio, ha scoperto che gli account che postavano messaggi simili su Facebook per dare risalto a contenuti a favore di Boris Johnson nel novembre del 2019 appartenevano a persone vere che si facevano passare per bot.

Può anche capitare di avere a che fare con un cyborg, ovvero un account social in parte automatizzato e in parte gestito da esseri umani, il cui comportamento combina attività automatizzate e attività autentiche. I giornalisti devono evitare di etichettare erroneamente account sospetti come bot se non hanno prove e analisi adeguate per farlo, perché un'accusa sbagliata può minare la credibilità del giornalista.

Di fronte a questi diversi tipi di account - bot, cyborg e esseri umani iperattivi -, una strategia efficace per condurre la propria indagine è monitorare tutti i comportamenti non autentici o simili a quelli di un bot, evitando di fissarsi sul riconoscimento di uno solo di essi.

Ad esempio, Bot Sentinel fornisce un database, accessibile a tutti, di account Twitter (americani) dai comportamenti sospetti. I suoi creatori hanno deciso di raccogliere "account che hanno violato ripetutamente le regole di Twitter" invece che limitarsi a cercare specificatamente i social bot.



Come indagare comportamenti non autentici

Per identificare comportamenti non autentici e potenzialmente automatici sui social network, consigliamo in generale il seguente approccio:

1. Esaminare direttamente e manualmente gli account alla ricerca di comportamenti sospetti
2. Integrare il controllo manuale con strumenti o analisi più tecniche dei network.
3. Indagare l'attività e i contenuti degli account e il network degli altri account con cui interagiscono. Integrare quest'indagine con le tradizionali tecniche di investigazione, cercando ad esempio di mettersi in contatto con le persone mostrate negli account o con persone che affermano di conoscerle.
4. Consultare esperti esterni specializzati in bot e in attività non autentiche.

Per valutare gli account sospetti con un controllo diretto e manuale, è importante capire quali sono i segnali tipici che contraddistinguono il comportamento di un account automatico su Twitter o su altri social network.

Ogni bot sui social media ha bisogno di un'identità. I creatori di bot vogliono rendere i loro account più convincenti possibile, ma per costruire e mantenere dei profili credibili serve molto tempo, soprattutto se l'obiettivo è gestire un network esteso di bot. Più account si devono gestire, più tempo serve per crearli e gestirli in maniera tale che sembrino autentici. Ed è qui che si commettono gli errori: in molti casi, i creatori di bot fanno il minimo indispensabile per mettere in piedi il profilo dell'account, e un bravo investigatore sa riconoscere quando ciò avviene.

Ecco qualche elemento da cercare:

False immagini di profilo

Un'immagine di profilo rubata (come abbiamo visto nell'indagine di Benjamin Strick sulla Papua Occidentale) o l'assenza di un'immagine di profilo possono essere indicatori di non autenticità. Infatti, i creatori di bot vogliono creare molti account in una volta sola e per farlo hanno bisogno di una raccolta di foto da usare, foto che spesso copiano da altri siti. Tuttavia, ciò dà origine a delle incongruenze: un account con la foto di profilo di un uomo e un nome utente femminile potrebbe essere il segnale che c'è qualcosa che non va. Per ovviare a questo problema, molti creatori di bot scelgono come immagini di profilo cartoni animati o animali, ma questa tattica diventa a sua volta un elemento da tenere in considerazione per riconoscere gli account non autentici o i bot.

Nomi utente creati automaticamente

Passo successivo: fai attenzione ai nomi e ai nomi utente. Ogni pseudonimo su Twitter è unico, pertanto spesso il nome utente che vorresti è già preso. Questo per una persona normale è un inconveniente, ma diventa una vera e propria sfida se devi creare 50, 500 o 5000 account in poco tempo.

I creatori di bot spesso ricorrono a una strategia che li aiuta a trovare facilmente nomi utente disponibili e usano script (sequenze di istruzioni) come quelle riportate qui sotto:

Username followed by a 4 digit number	12 random characters in length which can consist of (a-zA-Z and 0-9)	Any first name followed by a random eight-digit number, indicating that the default username generated by Twitter has been used.
superman_1230 superman_2313 superman_9832 superman_3934 superman_4920	vP1tfI1ZoPG1 dNi29j2utANQ YQBrodhbPC84 TUq3R6GBWYyA XI87NreGshx8	Neil03121977 Sarah92839820 Claire02938593 John09340293 Stephen83749284

Se ti accorgi che molti account Twitter hanno pseudonimi composti dallo stesso numero di caratteri e cifre, puoi cercare manualmente nella lista dei follower di ogni account altri account il cui nome utente replichi quello stesso schema, e identificare così un potenziale network.

Nell'esempio qui sotto, gli account hanno in comune anche qualcos'altro: sono stati tutti creati nel settembre 2019. Questo dettaglio, quando si presenta in concomitanza ad altri, potrebbe essere un segnale che gli account sono stati creati nello stesso momento dalla stessa persona.

L'attività dell'account non corrisponde alla sua "età"

I sospetti aumentano se un account nuovo possiede un numero relativamente alto di follower o se ha già pubblicato un elevato numero di tweet in un breve periodo di tempo. La stessa cosa vale se un vecchio account ha molti pochi follower nonostante sia stato molto attivo.

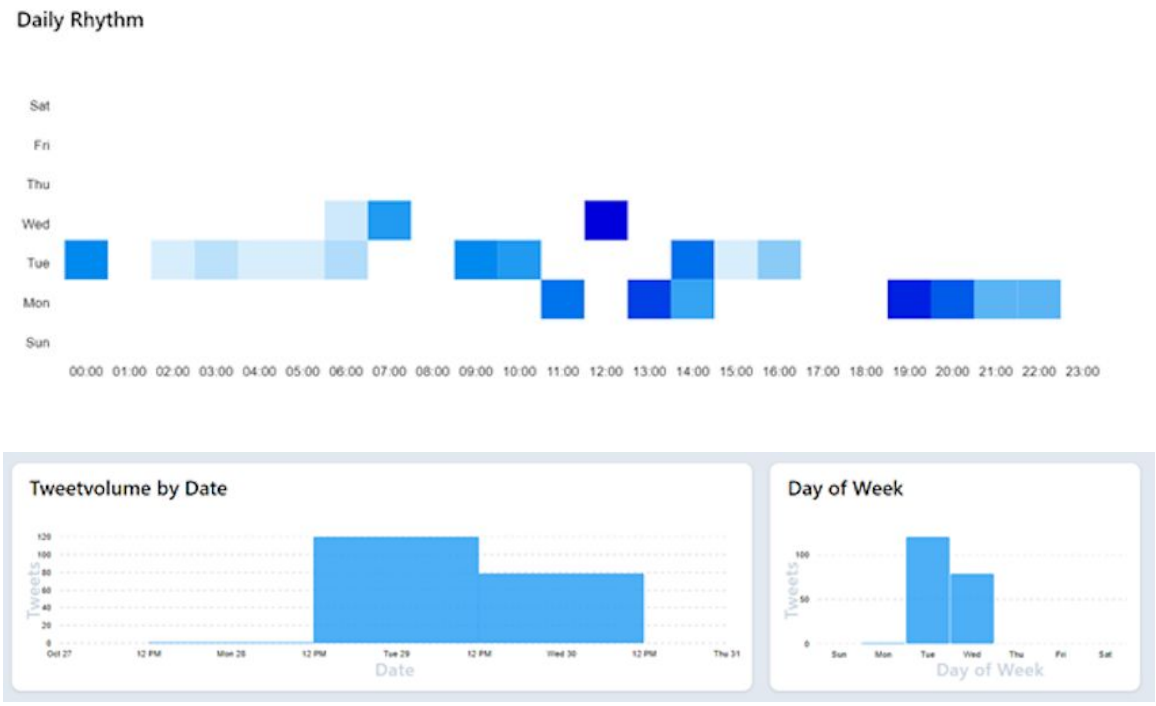
Se ti imbatti in un account del genere, analizza la sua attività in modo più approfondito. Prendi il numero di tweet che trovi in cima alla pagina e dividilo per il numero di giorni da cui è attivo l'account. Facciamo un esempio con un account creato il 15 agosto del 2019 che l'11 novembre dello stesso anno aveva 3489 tweet. Dividi 3489 per 89 (l'età dell'account) e ottieni 39,2 tweet al giorno.

Osservando i tweet pubblicati nel periodo di vita dell'account, il numero ottenuto ti sembra troppo alto, irrealistico o ingestibile?

Comportamenti sospetti dei tweet

Un altro elemento da esaminare è il ritmo con il quale un account pubblica dei tweet. Gli esseri umani possono mostrare delle leggere preferenze riguardo i giorni o le ore in cui sono soliti twittare, ma è difficile che una persona twitti regolarmente per un lungo periodo di tempo tutti i lunedì, martedì e mercoledì, sparendo del tutto negli altri giorni della settimana.

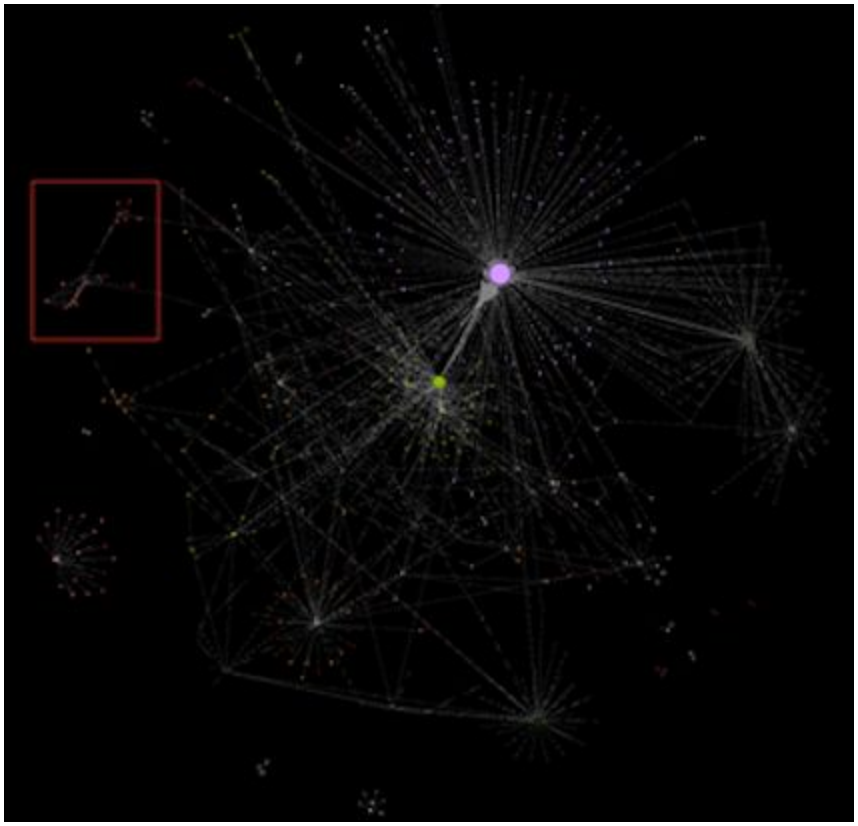
Se vuoi visualizzare graficamente i comportamenti relativi a uno specifico account, prova lo strumento di analisi degli account sviluppato da Luca Hammer:



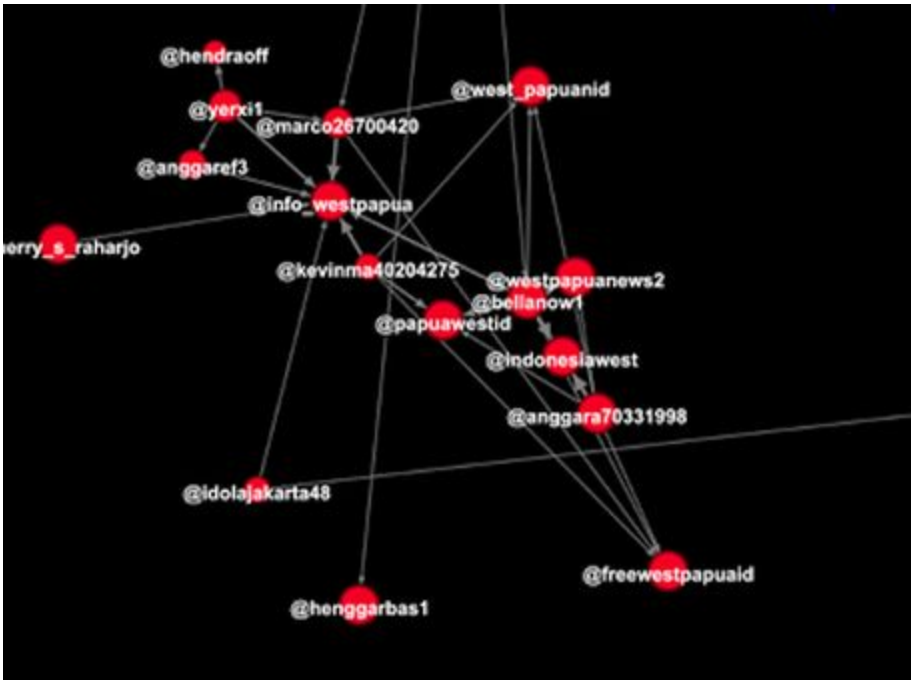
La visualizzazione come parte dell'indagine

Per capire meglio l'attività di un'intera rete di bot, puoi usare una piattaforma di visualizzazione come Gephi. Il collaboratore di Bellingcat Benjamin Strick usò proprio questo strumento per analizzare le connessioni tra gli account Twitter che appartenevano a una rete di bot a favore dell'Indonesia.

Osservando la rappresentazione grafica delle connessioni tra un gran numero di account Twitter, Strick si accorse che nella parte sinistra dell'immagine c'era una struttura che si distingueva dal resto (in rosso).



Ingrandendo l’area, riuscì a vedere quali erano gli account Twitter che componevano quella specifica struttura.



Ogni cerchio rosso rappresenta un account Twitter, e le linee rappresentano le relazioni tra gli account. Solitamente gli account più piccoli si dispongono intorno a un cerchio più grande che sta nel mezzo, il che significa che ciascuno degli account più piccoli interagisce con un account più influente. Gli account della struttura riportata qui sopra, tuttavia, interagivano tra loro in maniera diversa. Ciò spinse Strick ad analizzare questo comportamento anomalo.

Il futuro dei social bot: possiamo fregarli?

Negli ultimi anni la tecnologia che sta dietro i social bot si è fatta sempre più sofisticata, permettendo a queste piccole applicazioni software di diventare sempre più abili nel simulare il comportamento umano. Stiamo arrivando al punto di prevedere che utenti artificiali saranno in grado di gestire raffinate comunicazioni online senza che il loro interlocutore umano si accorga di intrattenere una lunga conversazione con un bot.

Tuttavia, per ora non ci sono prove che attestano l'esistenza e l'effettivo uso di social bot di così alto livello e capaci di apprendimento automatico. Attualmente, sembra che molte campagne di disinformazione siano ancora supportate da bot amplificatori molto meno complessi.

“Non penso che in circolazione ci siano molti social bot sofisticati al punto da essere in grado di avere una vera e propria conversazione con gli utenti e convincerli a sostenere determinate posizioni politiche”, ha affermato il dottor Ole Pütz, ricercatore del progetto “Unbiased Bots that Build Bridges” presso la Bielefeld University (Germania).

Secondo il ricercatore, il miglior modo per aiutare le persone a riconoscere un comportamento non autentico sui social network è usare un metodo di identificazione che cataloghi e soppesi tutti i fattori che rendono un account sospetto. Per fare un esempio, Pütz dice: “Questo account usa uno script per retwittare le notizie, segue automaticamente altri account e non usa mai strutture del discorso che gli esseri umani userebbero normalmente”.

Ad oggi, analizzare sistematicamente il comportamento, i contenuti, le interazioni e le abitudini degli account rimane il miglior approccio per individuare comportamenti non autentici.

Nel nostro capitolo dedicato al caso di studio spieghiamo in maniera più approfondita e tecnica come abbiamo analizzato i diversi fattori che ci hanno permesso di riconoscere, su Twitter, una rete sospetta di account collegata alle proteste di Hong Kong.

3a. Caso di studio: Trovare prove di attività automatizzata su Twitter durante le proteste di Hong Kong

Scritto da: Charlotte Godart, Johanna Wild

***Charlotte Godart** è un'investigatrice e formatrice di Bellingcat. Prima di Bellingcat era impegnata presso lo Human Rights Center dell'Università di Berkeley, per cui lavorava all'interno dell'Investigations Lab e insegnava agli studenti a condurre ricerche open source sui conflitti globali per organizzazioni umanitarie internazionali.*

***Johanna Wild** è un'investigatrice open source a Bellingcat, dove si occupa anche di sviluppo di tecnologie e strumenti per indagini digitali. Viene dal mondo del giornalismo online e precedentemente ha lavorato con altri giornalisti in zone di conflitto e post-conflitto. Tra i ruoli che ha ricoperto c'è stato quello di supporto ai giornalisti impegnati in Africa Orientale a produrre trasmissioni e contenuti per The Voice of America.*

Nell'agosto del 2019, Twitter annunciò la rimozione di migliaia di account che riteneva avessero contribuito a diffondere disinformazione sulle proteste di Hong Kong e fossero parte di una "operazione coordinata appoggiata dallo Stato". Ben presto, Facebook e YouTube rilasciarono delle dichiarazioni in cui comunicavano che anch'essi avevano rimosso account che mostravano un comportamento coordinato e non autentico in relazione alle proteste.

Diversamente da Facebook e da YouTube, Twitter pubblicò una lista degli account che aveva rimosso, offrendo l'opportunità di svolgere ulteriori indagini sulle loro attività. Insieme a una persona che aveva partecipato a un workshop di Bellingcat, il nostro team decise di indagare il materiale che rimaneva su Twitter relativo alle proteste di Hong Kong, per cercare di identificare segnali di comportamenti non autentici e coordinati.

Individuare attività sospette

Iniziammo cercando gli hashtag più rilevanti che riguardavano le proteste. Una semplice ricerca con la parola chiave "Hong Kong riots" fece emergere numerosi tweet, alcuni con molteplici hashtag.

Volevamo concentrarci su account e contenuti pro Cina, dal momento che Twitter aveva già riscontrato che erano questi a essere coinvolti in attività non autentiche. Provammo con formule di parole chiave come questa:

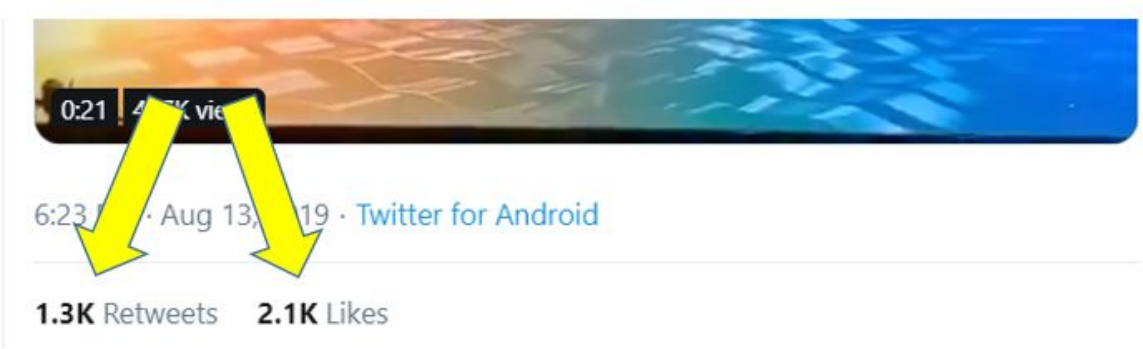
“Shame on Hong Kong” -police -government

Questa ricerca restituisce risultati che contengono la frase “Shame on Hong Kong” (Vergogna Hong Kong) ma non le parole "polizia" o "governo". L’obiettivo era escludere dai risultati tweet come “shame on Hong Kong police” (vergogna alla polizia di Hong Kong) e visualizzare invece tweet come “shame on hong kong protesters” (vergogna ai manifestanti di Hong Kong). Altri termini di ricerca erano “Hong Kong roaches” ("scarafaggi di Hong Kong") e “Hong Kong mobs" ("Gang di Hong Kong"), espressioni comunemente usate dagli account Twitter pro Cina per descrivere i manifestanti.

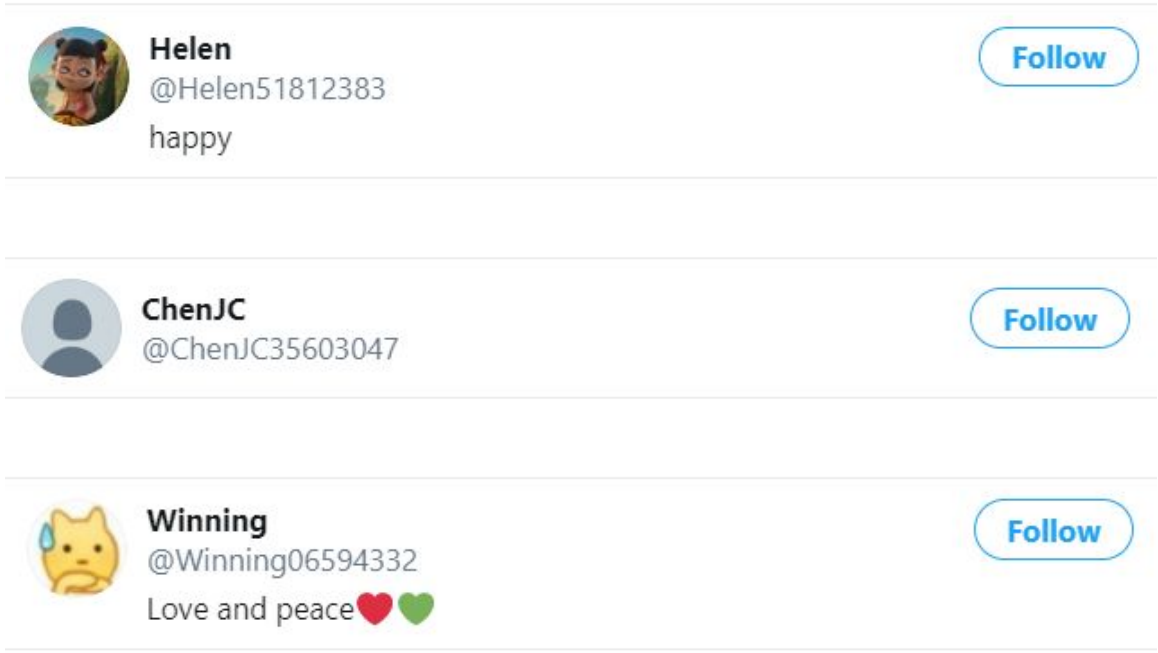
Dopo aver fatto una ricerca con questi e altri termini, esaminammo tweet recenti su Hong Kong che avevano ricevuto molti retweet e like. Puoi filtrare i risultati in base alle interazioni semplicemente aggiungendo “min_retweets:500” ("minimo 500 retweet) o “min_faves:500” (minimo 500 like) alla tua query. In questo modo otterrai soltanto tweet con almeno 500 retweet o 500 like.

Quindi ci concentrammo sugli account Twitter che avevano interagito con questi tweet. Per esempio, questo tweet dall’utente verificato Hu Xijin, direttore dell’edizione cinese e di quella inglese del Global Times, organo di stampa cinese controllato dallo Stato:

Cliccammo sugli iperlink “Retweets” e “Likes” che si trovano accanto al numero delle interazioni per visualizzare la lista degli account che avevano eseguito le azioni corrispondenti.

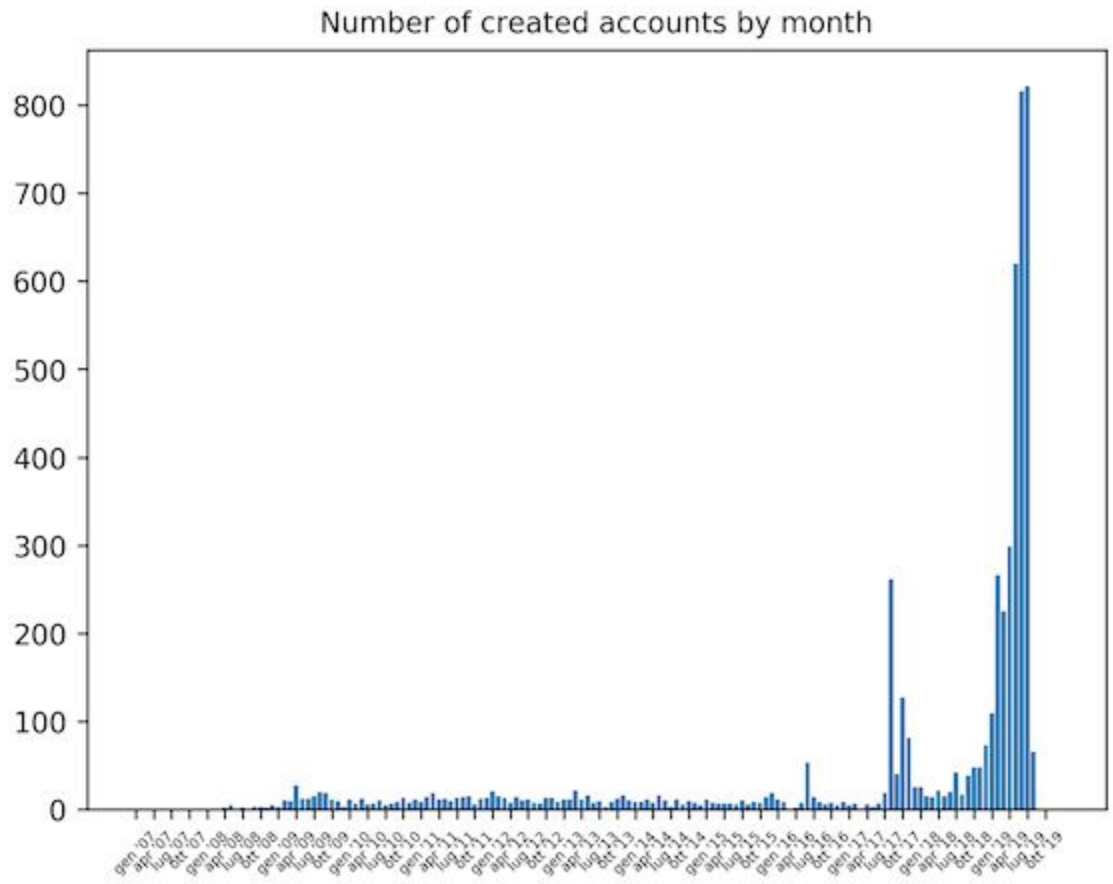


La nostra ipotesi era che account non autentici pro Cina amplificassero i tweet pubblicati dai collaboratori dell’importante organo di stampa cinese. Osservando le liste, ci accorgemmo che molti nomi utente si distinguevano dagli altri perché avevano un numero di otto cifre dopo il nome, il che indicava che l’utente aveva accettato il nome di default generato da Twitter al momento dell’iscrizione. Ciò ci motivò a svolgere ulteriori indagini sul comportamento e sulle caratteristiche di quegli account.



Esaminandoli, notammo che avevano solo pochi follower, che seguivano pochissimi account, che non avevano bio, che retwittavano tweet di altri senza praticamente pubblicare nulla di proprio e che promuovevano quasi esclusivamente contenuti in opposizione alle proteste.
Ci accorgemmo inoltre che questi account erano stati creati in date molto recenti, intorno all'agosto del 2019. Dato che Twitter aveva reso pubblica la lista di account a favore della Cina da lui rimossi, potemmo controllare le date di creazione di quegli account e appurare se anche per loro si verificava questa circostanza.

Con l'aiuto di Luigi Gubello, un programmatore attivo nelle community online di open source, usammo un semplice script Python (puoi trovare il codice sul suo profilo GitHub e più informazioni su di lui qui) per identificare possibili pattern nei dati sugli account. I grafici qui sotto mostrano che gli account rimossi erano stati tutti creati negli ultimi mesi, caratteristica che li accomunava al gruppo di account attivi su cui stavamo investigando.



Automatizzare il processo

Una volta individuato un campione di tweet che mostrava comportamenti e caratteristiche sospette, decidemmo di ampliare di molto la nostra indagine. Per farlo occorreva ricorrere a delle automazioni. Uno dei partecipanti a un workshop di Bellingcat aveva un background da sviluppatore di software, così scrisse un piccolo pezzo di codice JavaScript — l'espressione regolare (\w+\d{8}) — per svolgere due funzioni: estrarre i nomi utente degli account che avevano retwittato o messo un like a un dato tweet e poi filtrare velocemente la lista dei nomi utente per filtrare solo gli username che replicavano un dato pattern. Lo schema in base al quale volevamo filtrare gli username era nome+sequenza di otto cifre.

Caricando questo script sulla console degli strumenti per sviluppatori di Chrome, che mette a disposizione degli sviluppatori web strumenti direttamente nel browser, lo script avrebbe agito in background ogni volta che si cliccava sui link "Retweets" e "Likes" di un tweet specifico, mettendo in evidenza i nomi utente che corrispondevano al pattern che cercavamo. Clicca qui per vedere come si configura questa operazione.

A quel punto, potevamo usare lo script sviluppato dal nostro collaboratore per esaminare gli account che interagivano con altri importanti tweet pro Cina. Nel bel mezzo delle proteste di Hong Kong, l'attrice sino-americana Liu Yifei condivise su Weibo un post in sostegno alla polizia che indusse alcuni utenti dei social network a invocare un boicottaggio del suo nuovo film, "Mulan". Tuttavia, notammo anche che molti account Twitter supportavano l'attrice e il suo film usando l'hashtag #SupportMulan (anche la CNN ne parlò). Decidemmo di usare lo script per esaminare gli utenti che avevano retwittato o messo like ai tweet pro Mulan.

Raccogliemmo i nomi degli account che combaciavano con il nostro schema, e successivamente ne individuammo la data di creazione. Questa operazione rivelò che la maggior parte degli account era stata creata il 16 agosto.

https://twitter.com/monicaG62882882	created: 16 August, 20.07h
https://twitter.com/Min85741833	created: 16 August, 05.29h
https://twitter.com/cherry71737735	created: 16 August, 19.22h
https://twitter.com/Catheri57246362	created: 16 August, 06.13h
https://twitter.com/crystal09837022	created: 16 August, 04.16h
https://twitter.com/Suqing26464572	created: 16 August, 06.30h
https://twitter.com/Yates52905656	created: 16 August, 22.16h
https://twitter.com/hu02261927/	created: 16 August, 04.53h
https://twitter.com/xinjin66947005	created: 16 August, 19.18h
https://twitter.com/Ta99869608	created, 16 August, 21.15h

Ottenemmo la data e l'ora esatta di creazione degli account semplicemente posizionano il cursore sopra il link "Joined" (nella versione italiana di Twitter: "Iscrizione a") sul profilo, come mostrato qua sotto:

Con il gruppo di account davanti a noi, cominciammo ad analizzare manualmente i contenuti che avevano condiviso. Ben presto fu chiaro che gli account nella nostra lista avevano tutti twittato in favore di Yifei e contro i manifestanti di Hong Kong.

Molti degli account nella nostra lista divennero inattivi dopo il 17 o il 18 agosto, il che rappresentava un ulteriore elemento di coordinamento. Non sappiamo esattamente perché smisero di essere attivi, ma è possibile che Twitter abbia richiesto ai loro creatori di completare ulteriori passaggi di verifica a cui essi non potevano adempiere. Un'altra possibilità è che gli account smisero di twittare perché chi li aveva creati non voleva alimentare ulteriori sospetti dopo che Twitter aveva cominciato a sospendere gli account a favore della Cina.

In ogni caso, qualche mese dopo notammo che molti di quegli account erano di nuovo attivi. Questa volta diffondevano messaggi positivi su Yifei e sul suo film "Mulan".

Trovammo anche altri account a favore di “Mulan”, accomunati da nomi utente formulati secondo lo stesso schema e da stesse date di creazione che diffondevano continuamente messaggi a sostegno di Yifei. Lo scoprimmo cercando tweet che includevano hashtag come #SupportMulan o #liuyifei.

crystal_28cc @28ccCrystal · Dec 5

cool！ #disneyliveaction #SupportMulan
#Liuyifei #CrystalLiu
#mulan #liuyifei #yifei_cc #crystalliu #刘亦菲 #花木兰 #花木蘭 @yifei_cc

> **Disney** @Disney · Dec 5
>
> Loyal. Brave. True. I will bring honor to us all. Watch the brand new trailer for #Mulan. See it in theaters March 27, 2020.

3

crystal_28cc @28ccCrystal · Aug 16

The real thugs are the demonstrators, not the police.
We support the leading artist of mulan and the Hong Kong police.#Mulan
#LiuYifei #supportmulan

4 18 168

MULAN

Follow

Mulan Our pride. ❤️

@kongyuting1

Liu Yifei is a good girl. ❤️He has Mulan's qualities of justice and courage and patriotism.❤️ He is our pride. ❤️Be happy. 刘包子。 ❤️

Joined August 2019

53 Following **65** Followers

Follow

Cinderlance·icc

@cinderlance

cuz u sucked some

Joined December 2017

48 Following **64** Followers

Mulan Our pride. ❤ Retweeted

Choco @Choco_Xu · Aug 17

#SupportMulan #Mulan Democracy is not manifested by violence.Why can't people see the truth, she just stands on the side of justice?

18 31 114

Cinderlance·icc Retweeted

Choco @Choco_Xu · Aug 17

#SupportMulan #Mulan Democracy is not manifested by violence.Why can't people see the truth, she just stands on the side of justice?

18 31 114

Mulan Our pride. ❤️ @kongyuting1 · Sep 25

#mulan #liuyifei #supportmulan #LiuYiFei #花木蘭 ❤️

十五小甜心 @SNH48_15 · Sep 22

#liuyifei #Mulan Take you to know a real Liu Yifei (Mulan's actor). She always believes in one sentence, the harder she works, luckier she will be. I think one day, people will see the beauty of her bloom. twitlonger.com/show/n_1sr10p5

3

Sembrava che gli account avessero cambiato strategia, passando dal criticare i manifestanti di Hong Kong al promuovere l'attrice e il suo film, forse per evitare di essere bloccati da Twitter.

Questo caso di studio mostra come sia possibile combinare tecniche manuali e automatizzate per scoprire velocemente una rete di account Twitter sospetti. Contemporaneamente dimostra che è utile andare a cercare ulteriori account e attività sospette anche dopo che una piattaforma ha annunciato la sospensione di alcuni account.

In questo caso specifico, abbiamo usato delle semplici tecniche di ricerca e le informazioni degli account per identificare un più ampio gruppo di account che esibivano forti segnali di appartenenza a un'azione non autentica coordinata.

4. Monitorare bufale e *operazioni di informazione* durante le breaking news

Scritto da: Jane Lytvynenko

***Jane Lytvynenko** è senior reporter a BuzzFeed News, dove si occupa di disinformazione, cyber sicurezza e indagini online. Ha scoperto campagne di manipolazione sui social media, truffatori di criptovalute e soggetti che diffondevano disinformazione per interessi economici. Il suo lavoro è anche volto a rendere accessibile il fact-checking al grande pubblico durante periodi di crisi. Jane è di Kiev, in Ucraina, e attualmente risiede a Toronto, in Canada.*

Quando arriva una notizia, possono passare ore o persino giorni prima che i giornalisti e le autorità siano effettivamente in grado di dare senso a ciò che è accaduto. Mentre le prime prove e i primi dettagli iniziano a circolare sui social network e su altre piattaforme online, possono emergere soggetti intenzionati a seminare discordia o sospetto, o a far soldi veloci sfruttando l'attenzione di un pubblico di preoccupati consumatori di news. Gli stessi consumatori in buona fede che, assieme ad altre fonti, possono inconsapevolmente rendersi vettori di informazioni false o fuorvianti. Di fronte alla miscela di emozioni esagerate e informazioni che arrivano con il contagocce, come tipicamente avviene nei primi minuti e nelle prime ore dopo un evento, i giornalisti devono essere effettivamente in grado di monitorare, verificare e — quando necessario — smontare le breaking news. Per creare un tweet, un'immagine, un account social o un articolo falsi ci vogliono solo pochi minuti, e le informazioni vere fanno fatica a tenere il passo.

La chiave per riuscire a monitorare le informazioni e fare debunking durante le breaking news è costruire solide fondamenta prima che le breaking news arrivino. Ciò significa avere delle robuste basi nelle tecniche di verifica, come quelle illustrate nel primo Verification Handbook (), sapere come si monitorano i social network e le piattaforme, e come reagire nel caso in cui tu o i tuoi colleghi veniate presi di mira da attori che agiscono in malafede. I giornalisti non dovrebbero mai mettere la sicurezza online in secondo piano.

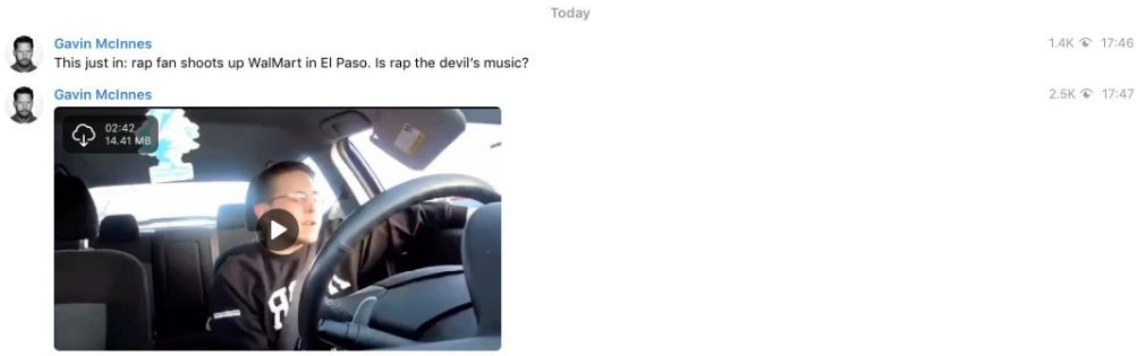
Quando arriva una breaking news, la prima cosa da fare è individuare le principali comunità colpite da quanto avvenuto. Durante la sparatoria del 2018 nel liceo di Parkland, in Florida, per trovare video di ciò che stava accadendo agli studenti intrappolati nelle aule i giornalisti setacciarono la mappa di Snapchat. Durante l'uragano Irma nel 2017 fu invece fondamentale concentrarsi su Facebook, il social su cui le persone colpite cercavano di trovare informazioni. Comprendere il funzionamento di ciascun social network e come esso venga usato durante un certo evento è essenziale.

Questo capitolo si concentra sugli strumenti che un giornalista può utilizzare per monitorare le breaking news e fare debunking quando si verificano. Gli strumenti descritti non vanno tutti bene per ciascuna delle situazioni che possono presentarsi: capire caso per caso chi sono le persone colpite dall'evento ti aiuterà a stabilire dove concentrare maggiormente le tue ricerche.

Tre cose da cercare

Mentre piattaforme e giornalisti lavorano duramente per combattere la disinformazione, i suoi artefici hanno migliorato le proprie tattiche per evitare di essere scoperti. Ciononostante, nei loro contenuti e comportamenti continuano ad emergere pattern ricorrenti.

1. Immagini manipolate o fuori contesto. La celebre immagine di uno squalo che nuota in un'autostrada allagata ha continuato per anni a fare il giro del web e a ingannare le persone (è stata anche oggetto di un caso di studio nel primo Verification Handbook). Chi si occupa di fact-checking e debunking chiama "bufale zombi" foto e video già smascherati ma ancora in circolazione. Tenere d'occhio le bufale zombi è importante. Sulle piattaforme digitali le immagini si diffondono molto più velocemente dei testi, quindi concentrarsi su di loro si rivela spesso proficuo.



Durante la sparatoria in un Walmart di El Paso nel 2019, esponenti di estrema destra cercarono di far passare per vero un vecchio video di YouTube che non aveva alcuna connessione con il sospettato.

2. False vittime o responsabili. Durante la sparatoria presso la sede di YouTube i social network erano disseminati di false dichiarazioni sui sospetti colpevoli, mentre nel corso delle elezioni statunitensi di metà mandato del 2018 il Presidente degli Stati Uniti diffuse false voci su schede elettorali depositate da immigrati irregolari: durante le più grandi breaking news appaiono sempre dei falsi responsabili.

Durante la sparatoria a Parkland, un account fake di Bill O'Reilly provò a diffondere un falso nome del sospettato.

3. Molestie e "brigading". Pur non trattandosi di disinformazione nel senso stretto del termine, accade comunemente che soggetti in malafede cerchino di molestare persone coinvolte in un evento che fa notizia per metterle a tacere. Quando si verifica una situazione del genere significa che un gruppo di persone sta prestando attenzione a un evento, e che potrebbe mettere in pratica diverse tattiche per ottenere ciò che vuole. Con "brigading" si indica l'attività di un gruppo di persone che collaborano tra loro per dare l'impressione che intorno a un certo fatto si siano scatenati un coro di reazioni e un'ondata di coinvolgimento, ad esempio promuovendo o bocciando dei contenuti o inondando un utente di commenti.

A seguito di un dibattito tra i leader dei Democratici nel 2019, più account anonimi diffusero lo stesso messaggio sull'etnia di Kamala Harris.

Buone pratiche per archiviare e pubblicare

Prima di metterti a cercare bufale, crea una cartella per i tuoi documenti e un foglio di lavoro per quello che trovi. Fai subito uno screenshot di ogni bufala e di ogni contenuto rilevante che trovi e poi archivia la pagina (l'estensione per browser di Archive.org (http://archive.org/) è uno strumento gratis, veloce e funzionale per archiviare contenuti). Abbi cura di archiviare nel tuo foglio di lavoro sia il link originale, sia la URL del contenuto archiviato. Ciò ti permetterà di tornare a lavorarci sopra non appena la bufera sarà passata, ed esaminare ciò che hai trovato alla ricerca di pattern ricorrenti.

Per evitare di favorire la diffusione di pagine associate a disinformation e misinformation, assicurati di linkare in ogni articolo o post sui social la URL del contenuto archiviato e non quella del contenuto originale. È buona abitudine anche apporre sulle tue immagini un watermark inequivocabile, come "Falso" o "Ingannevole", per assicurarti che vengano diffuse e indicizzate con l'appropriata

contestualizzazione. Se scrivi un articolo, enfatizza nel titolo e nel testo ciò che è vero, anziché dire in prima istanza ciò che è falso. Studi hanno dimostrato che ripetere le falsità può portare i lettori a ricordare le informazioni sbagliate.

Il tuo ruolo è quello di ripetere il meno possibile le cose false e guidare le persone verso informazioni accurate.

Trovare parole chiave e posizioni

Man mano che si sviluppano gli eventi, fai un elenco di luoghi e parole chiave rilevanti.
Per quanto riguarda il luogo, tieni in considerazione il Paese, la città e la regione, provincia o Stato in cui questa si trova, nonché tutti i nomi locali dei tuoi luoghi di riferimento, ad esempio il soprannome di una città o di un quartiere. Durante le elezioni usa anche il nome della circoscrizione elettorale. Queste informazioni servono a monitorare i post con geotag e a cercare le menzioni dei luoghi. Assicurati anche di individuare e monitorare gli account social di tutte le autorità locali competenti, come polizia e vigili del fuoco, e anche politici e testate locali.

Successivamente, individua le parole chiave. Queste possono includere parole come vittima, sospetto, tiratore, sparatoria, inondazione, incendio, i nomi confermati di chiunque sia coinvolto, ed espressioni più generiche come "sto cercando" o cose simili. Oltre alle parole chiave, pensa a come si esprimerebbero le persone in quella situazione. Se trovi account affidabili che affermano di essere nel bel mezzo dell'evento che stai seguendo, prendi nota dei loro nomi utente e scorri tutto il loro feed. Controllare tra i loro amici o follower può servire a trovare altre persone in zona coinvolte nella vicenda.

Ricordati che in situazioni stressanti può capitare di scrivere male i nomi di luoghi e persone. Per esempio, durante l’incendio di Kincade, in California, nel 2019, per colpa della correzione automatica qualcuno twittò #kinkaidfire. Includi nelle tue ricerche comuni errori di battitura e prova a immaginare possibili errori dovuti alla correzione automatica, scrivendo le parole chiave sui tuoi device e guardando cosa ti viene suggerito.

Inoltre, è anche il momento giusto per contattare tutte le fonti che conosci nel luogo interessato dalla notizia, oppure quelle appartenenti a comunità potenzialmente oggetto di molestie o disinformazione, e chiedere loro cosa hanno visto online. Puoi anche informare il tuo pubblico che stai cercando contenuti che disinformano, o comunque problematici, su ciò che sta accadendo. Coordinati con il team che si occupa dei social media nella tua redazione per spargere la voce sul lavoro che stai facendo e per chiedere loro se hanno notato qualcosa di importante.

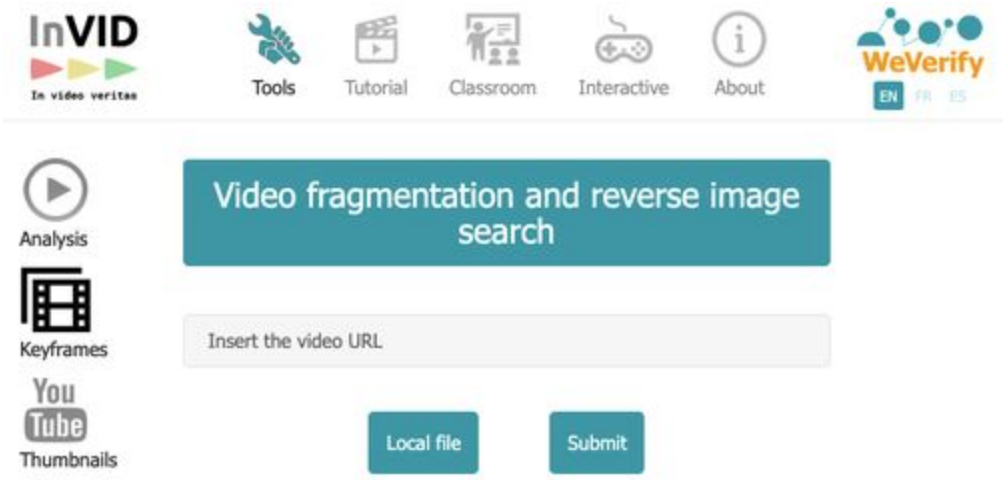
Principali strumenti per le immagini

1. Ricerca per immagini

La ricerca inversa delle immagini è uno strumento indispensabile. Cercare un'immagine su Google cliccando con il tasto destro del mouse su una fotografia e selezionando "Cerca l'immagine su Google" nel browser Chrome è un'operazione semplice, ma è sempre buona norma cercare un'immagine usando strumenti diversi. Se installi sul tuo browser l'estensione InVID, potrai cliccare sopra un'immagine con il tasto destro e cercarla tramite diversi strumenti. Questa tabella di comparazione tra motori di ricerca inversa delle immagini creata da Domain tools mostra i punti di forza e di debolezza dei vari strumenti:

	Elements Identified	Faces	Structures	Places	Digital/ Logos	Alternate Sizes	Flipped or Altered
Google	1	Neutral	Great	Great	Great	Good	Neutral
Yandex	2+	Great	Great	Great	Good	Good	Good
Bing	3+	Good	Good	Good	Good	Neutral	Great
TinEye	1	Neutral	Neutral	Neutral	Great	Great	Good

InVID
InVID è un'estensione per browser gratuita e rappresenta la migliore piattaforma per analizzare e verificare l'autenticità dei video. L'estensione permette agli utenti di incollare una URL nel suo motore di ricerca, che estrae le anteprime del video. Potrai poi eseguire delle ricerche inverse per immagini usando queste anteprime per vedere dove è apparso quel video in rete.



2. La ricerca su TweetDeck

Il modo migliore per fare ricerche su Twitter è usare TweetDeck, che permette di creare colonne dedicate a singole ricerche e liste.

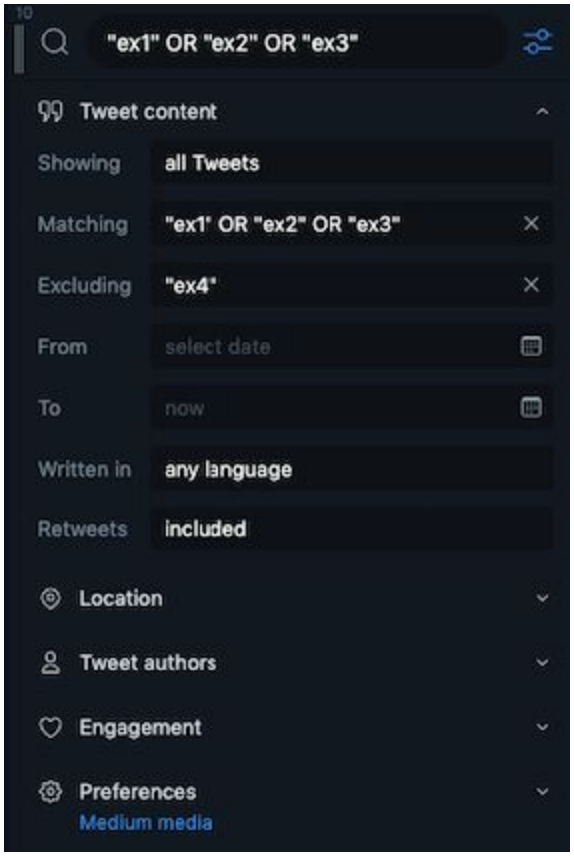
Trovare e duplicare liste importanti è fondamentale per rimanere aggiornati sulla situazione. Puoi usare Google per cercare liste di Twitter usando una semplice formula: scrivi site:twitter.com/*/lists nel motore di ricerca e poi aggiungi una parola chiave tra virgolette, ad esempio “Alabama reporters”. La stringa di ricerca finale si presenterà così:

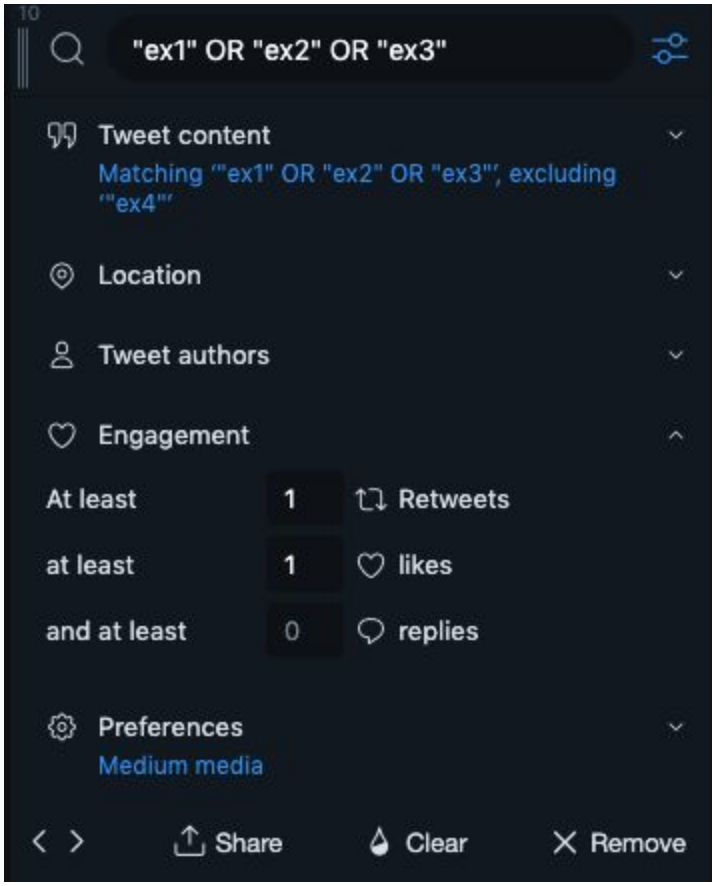
site:twitter.com/*/lists “Alabama reporters”

Così facendo troverai tutte le liste create da qualsiasi altro utente di Twitter nel cui titolo sia contenuta l'espressione “Alabama reporters”.

Una volta che hai trovato una lista utile ai tuoi scopi, occorre duplicarla per poterla aggiungere a TweetDeck. Usa questa app http://projects.noahliebman.net/listcopy/connect.php per duplicare tutte le liste che vuoi. È meglio duplicare una lista piuttosto che seguirla, perché in questo modo potrai aggiungere o rimuovere utenti a tuo piacimento.

Oltre a trovare e aggiungere elenchi alle colonne di TweetDeck, puoi creare colonne con filtri di ricerca specifici che ti consentano di monitorare rapidamente parole chiave, immagini e video. Per cercare più parole chiave, racchiudile tra virgolette e inserisci tra di loro "OR", ad esempio "Kincade" OR "Kinkade". Puoi anche escludere le parole che producono risultati irrilevanti. La maggior parte delle persone non tagga più i tweet in base alla posizione, quindi puoi lasciare il campo vuoto per ampliare le maglie della tua rete.


"ex1" OR "ex2" OR "ex3"
Tweet content
Showing
all Tweets
Matching
"ex1" OR "ex2" OR "ex3"
Excluding
"ex4"
From
select date
To
now
Written in
any language
Retweets
included
Location
Tweet authors
Engagement
Preferences
Medium media



Se vuoi restringere i risultati, imposta il campo "From" a uno o due giorni prima dello svolgimento dell'evento, in modo da non perdere tweet a causa di possibili differenze di fuso orario. Se i risultati che ottieni sono ancora troppi, prova a filtrarli in base alle interazioni per visualizzare solo i post che sono stati retwittati o che hanno ricevuto like. Puoi anche provare a suddividere le parole chiave in colonne separate. Ad esempio, inserisci le località in una colonna e altre parole chiave in un'altra. Io di solito creo una terza colonna per i nomi dei sospettati o delle vittime, comprese le loro versioni scritte con errori di battitura o di ortografia.

Infine, se noti una quantità molto elevata di tweet, può essere d'aiuto creare una nuova colonna con le parole chiave migliori e, dalla voce "Showing" del filtro "Tweet content", selezionare solo foto e video. Otterrai così un feed utile a individuare immagini virali o in crescita.

3. CrowdTangle

CrowdTangle è un'applicazione web messa gratuitamente a disposizione delle redazioni (contatta i produttori se la tua redazione non ha la possibilità di accedere).

È uno strumento potente che ti permette di configurare delle dashboard per monitorare contenuti su Facebook, Instagram e Reddit. Puoi anche fare ricerche usando parole chiave e impostare molti filtri, inclusa l'ora di pubblicazione, la lingua

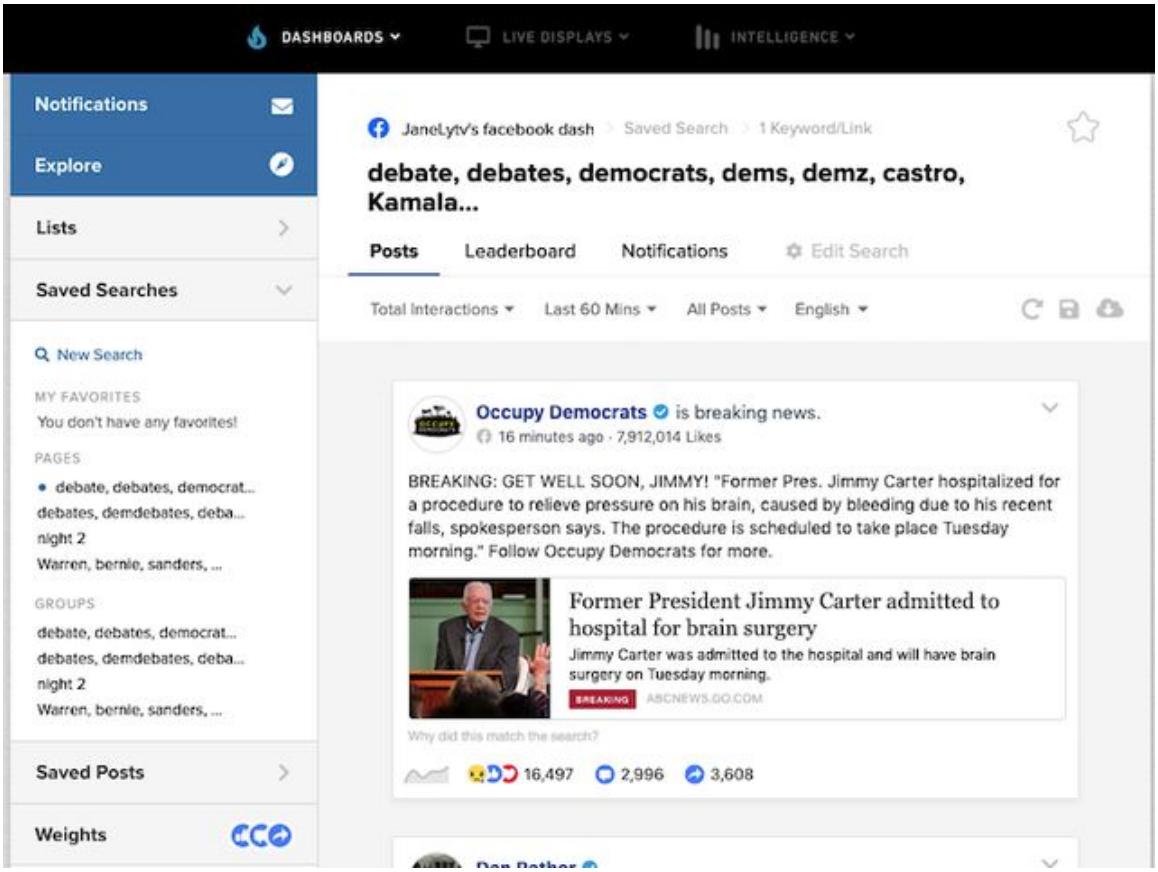
e le interazioni generate. CrowdTangle è utile in particolar modo per monitorare Facebook e controllare se una URL è stata pubblicata sui social. Una volta ottenuto l'accesso, per iniziare vai su app.crowdtangle.com e clicca su "Create New Dashboard". L'estensione per browser è gratis per tutti, anche per chi non ha l'accesso.

CrowdTangle: cercare post su Facebook

Clicca su "Saved Searches" sulla barra a sinistra e quindi su "New Search". Con Facebook hai due opzioni: cercare tra le pagine o cercare tra i gruppi. Io consiglio di fare entrambe le cose. Inserisci tutte le parole chiave che vuoi separandole con una virgola. A questo punto puoi scegliere come visualizzare i risultati, ad esempio partendo dal post più recente, da quello più popolare o da quello più "overperforming", un'etichetta usata per indicare i post che generano molte più interazioni della media di una pagina. Io scelgo tra le tre modalità a seconda della situazione, per essere sicuro di vedere i contenuti virali e quelli nuovi.

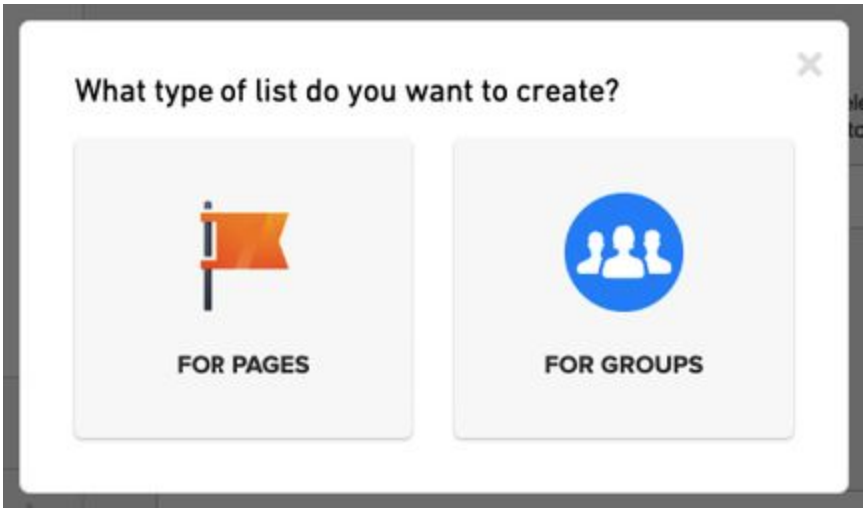
Puoi anche ordinare i post in base alla finestra temporale di pubblicazione e alla tipologia. Recentemente, CrowdTangle ha aggiunto la possibilità di cercare post in base alla località della pagina che li ha pubblicati. Cliccando su "English" e poi scegliendo "Country", puoi scegliere ad esempio di vedere solamente i post che arrivano da pagine che hanno dichiarato come propria posizione gli Stati Uniti. Puoi anche fare la ricerca opposta e cercare post pubblicati, ad esempio, da pagine basate in Iran, Russia, Arabia Saudita, Filippine o India. Presta particolare attenzione ai post con video o immagini, che tendono a diffondersi di più e a generare più interazioni.

Una volta configurata una ricerca che ti dà risultati rilevanti, ricordati di salvarla così da poterla recuperare.



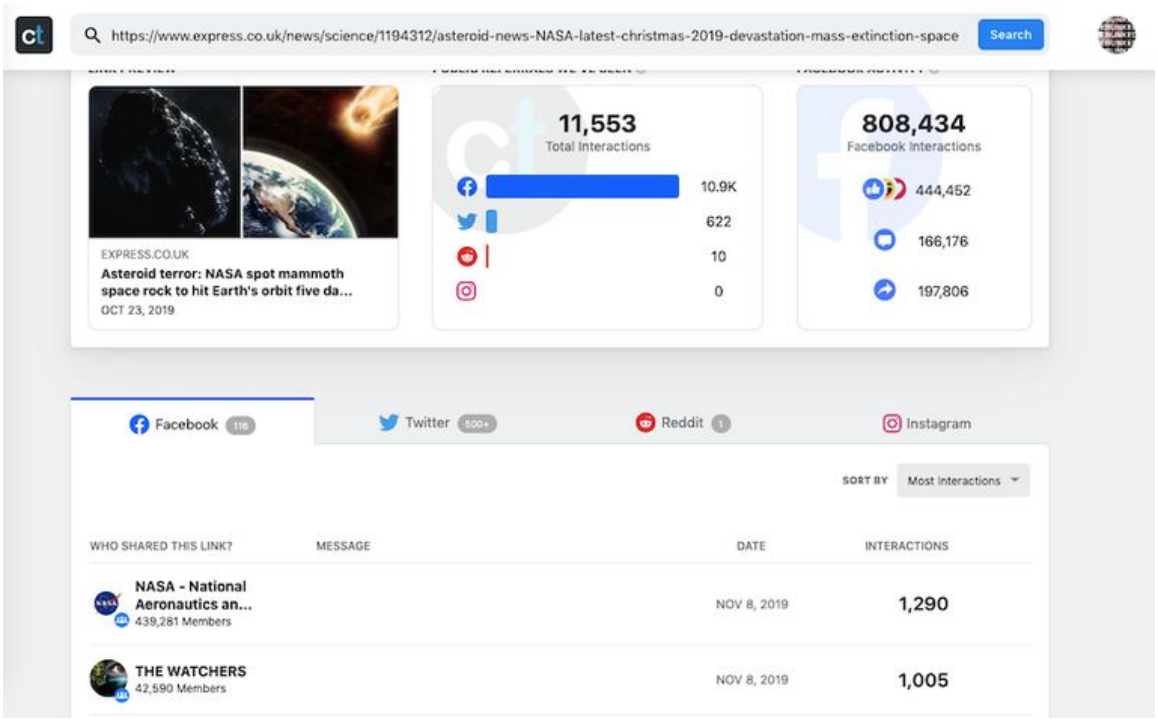
CrowdTangle: liste

Come TweetDeck, CrowdTangle ti permette di costruire liste di pagine e di gruppi pubblici di interesse. Cliccando su “Lists” sulla barra laterale a sinistra e poi su “Create list”, puoi monitorare pagine o gruppi che rispondono alle parole chiave che hai scelto o alle pagine delle quali possiedi le URL. CrowdTangle ha anche un numero di liste precompilate che puoi consultare cliccando sul tasto “Explore”. Come con Twitter, mettere insieme liste di pagine e di gruppi che parlano dell’evento che stai coprendo è un buon metodo per monitorare l’ecosistema informativo.



CrowdTangle: ricerca per link

Un'altra importante funzionalità di CrowdTangle è la ricerca per link. Vai su https://apps.crowdtangle.com/search/ e incolla la URL o le parole chiave del contenuto che ti interessa. CrowdTangle ti mostrerà le principali condivisioni pubbliche del link su Facebook, Instagram, Reddit e Twitter (nota che i risultati di Twitter sono limitati agli ultimi sette giorni). Questa ricerca ti aiuterà a capire come il contenuto si sta diffondendo, se ci sono gruppi o individui su cui dovresti indagare ulteriormente e se il contenuto è stato condiviso abbastanza da giustificare un debunk. Non esistono regole per stabilire quando fare debunking. Ci si può però porre alcune domande utili: la notizia si è diffusa al di fuori della rete iniziale di account che l'ha condivisa? È stata condivisa da figure autorevoli? Ha generato interazioni significative? Una nota su CrowdTangle: l'estensione gratuita per browser restituisce gli stessi risultati dello strumento di ricerca dei link, ed entrambi possono essere usati liberamente da chiunque senza un account CrowdTangle.

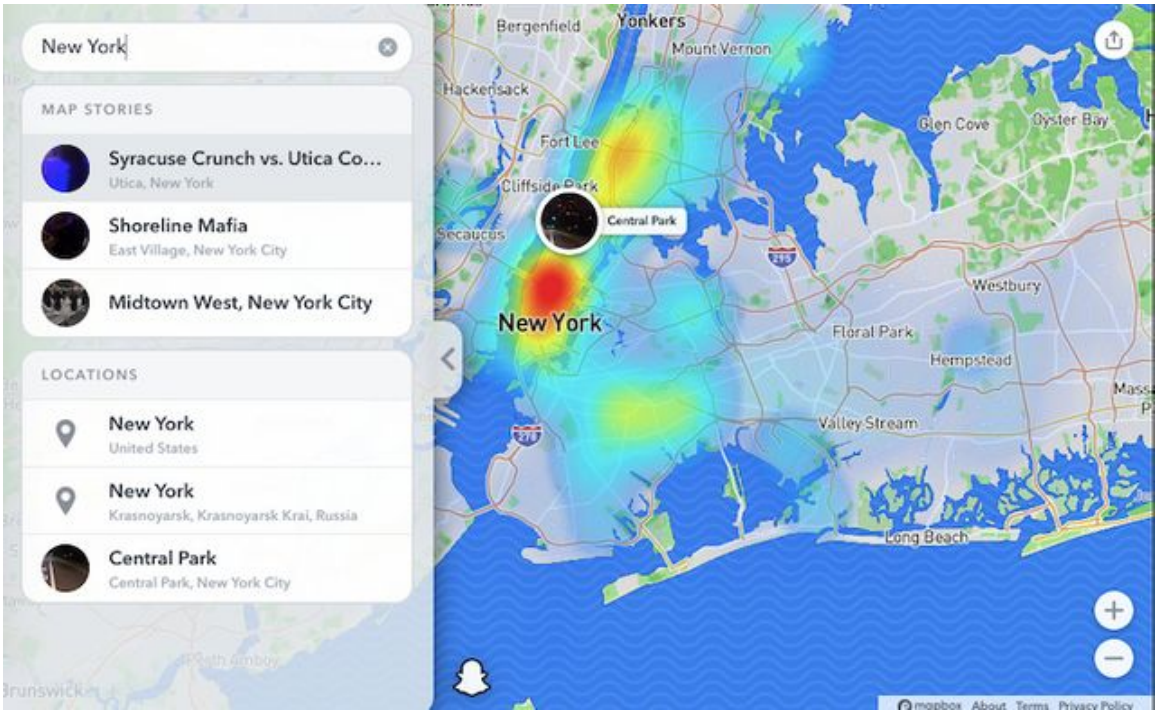


4. Instagram.com

Instagram è un luogo utile per monitorare hashtag e post geolocalizzati. Cerca luoghi chiave che gli utenti potrebbero aver taggato nelle foto, e ricorda che i tag di posizione possono includere anche quartieri e punti di riferimento sul territorio. Una volta trovato qualcuno (presumibilmente) coinvolto negli eventi, spostati sul suo account e vai a guardare le sue stories, che su Instagram sono molto più popolari dei normali post. Cerca tra i commenti altri potenziali testimoni e annotati ogni nuovo hashtag usato nei loro post. Se vuoi archiviare le storie di Instagram, puoi usare siti come storysaver.net per scaricarle.

5. SnapMap

Se da una parte è raro che Snapchat venga usato per disinformare, dall'altra è vero che la sua mappa pubblica è utile per verificare notizie o smascherare bufale. Per cominciare, vai su map.snapchat.com e scrivi una località che ti interessa. Ti verrà mostrata una mappa di calore che indica dove i contenuti sono stati pubblicati: più il colore della località è intenso, più snap provengono da quel punto. Per salvare snap utili, clicca sui tre puntini in alto a destra e seleziona “Share”. Potrai copiare la URL dello snap per ritrovarlo più tardi (assicurati anche di fare uno screenshot).



Mettere tutto insieme

Per evitare di fare confusione quando si presenta una breaking news, è essenziale fare un po' di pratica con ogni strumento prima che questa arrivi. Mentre cerchi in rete, ricordati che la disinformazione gioca sulle emozioni e trae vantaggio dalle lacune nella copertura delle notizie. Inoltre, ti imbatterai spesso in informazioni accurate che potrebbero servire ai tuoi colleghi. Per essere più veloce nel riconoscere le cose false, prendi nota di tutto ciò che sai con certezza essere vero, e non aver timore di chiedere aiuto ai reporter che la tua testata ha inviato sul campo.

È sempre utile ritornare sulle immagini e sui post che hai salvato dopo che la polvere si è depositata. Se durante la breaking news ti sei concentrato sul fare giornalismo di pubblica utilità segnalando le singole menzogne, successivamente è bene tornare sul materiale che hai raccolto per cercare pattern ricorrenti o tematiche rilevanti: ci sono state persone attaccate per la loro etnia o per il loro genere? Bufale nate da account piccoli e anonimi sono diventate mainstream? Ci sono stati social network che si sono comportati particolarmente bene o particolarmente male? Un articolo di riepilogo può aiutare i tuoi lettori a comprendere a fondo gli obiettivi e i metodi della diffusione di disinformazione, e servirà anche a te e alla tua redazione come strumento di ricerca, ricordandoti su cosa sarà utile concentrarsi quando arriverà la prossima breaking news.

5. Verificare e analizzare le immagini

Scritto da: Hannah Guy, Farida Vis, Simon Faulkner

***Farida Vis** è direttrice del Visual Social Media Lab e docente di Digital Media alla Manchester Metropolitan University. Il suo lavoro accademico e nell'ambito del data journalism si concentra sulla diffusione della disinformazione online. Ha lavorato per il Global Agenda Council on Social Media (2013-2016) e al Global Future Council for Information and Entertainment (2016-2019) del World Economic Forum ed è direttrice di Open Data Manchester.*

***Simon Faulkner** è professore di Storia dell'Arte e Cultura Visiva alla Manchester Metropolitan University. Le sue ricerche riguardano gli usi politici e i significati delle immagini, con un focus particolare sui movimenti di protesta e di attivismo. È anche co-direttore del Visual Social Media Lab, e nutre un forte interesse verso lo sviluppo di metodi efficaci per analizzare le immagini che circolano sui social media.*

***Hannah Guy** è dottoranda alla Manchester Metropolitan University e studia il ruolo delle immagini nella diffusione della disinformazione sui social media. È membro del Visual Social Media Lab, per cui attualmente porta avanti progetti che esaminano le immagini condivise su Twitter durante la comparsa del movimento Black Lives Matter, e della Visual Media Literacy per combattere la disinformazione nel contesto delle scuole canadesi.*

La comunicazione sui social media è ormai prevalentemente visiva. Foto e video sono persuasivi e avvincenti, riescono a innescare reazioni emotive potenti e oggi crearli è più facile che mai. Per questo, sono diventati potenti veicoli di misinformation e disinformation.

Ad oggi, la discussione sulle immagini nel contesto della misinformation e disinformation si è concentrata sulle tecniche di verifica o, più recentemente e in modo sproporzionato, sui cosiddetti video deepfake. Prima di considerare i deepfake, come faremo nel prossimo capitolo, è essenziale comprendere l'uso comune e a bassa tecnologia di foto e video fuorvianti, specialmente quelli mostrati fuori dal loro contesto.

Di fronte all'ampio utilizzo di materiale visivo allo scopo di influenzare e manipolare il dibattito pubblico, i giornalisti devono disporre sia delle conoscenze fondamentali di verifica delle immagini, sia della capacità di esaminarle e valutarle per capire come e perché vengono utilizzate. Questo capitolo si concentra sullo sviluppo di questa seconda categoria di capacità, basandosi su una risorsa che abbiamo sviluppato all'interno del Visual Social Media Lab.

Potenziare la verifica

Il nostro lavoro al Visual Social Media Lab si concentra sul comprendere il ruolo che le immagini ricoprono nella società. Ci occupiamo in larga misura di immagini fisse, categoria che comprende comunque una gamma di diversi tipi di immagine: foto, immagini composite, meme, grafiche e screenshot, per nominarne solo alcuni. Combattere la misinformation e la disinformation diffuse attraverso le immagini richiede l'attuazione di una serie di strategie specifiche. Fino a oggi, verificare un'immagine ha significato per i giornalisti soprattutto stabilire se quell'immagine sia realmente ciò che essi pensano che sia.

Nel primo Verification Handbook, Trushar Barot ha delineato quattro principi di base fondamentali per la verifica delle immagini, che rimangono di inestimabile valore. La First Draft Visual Verification Guide è un'altra risorsa utile che utilizza questi principi, focalizzandosi su cinque domande da porsi in merito alle foto e ai video:

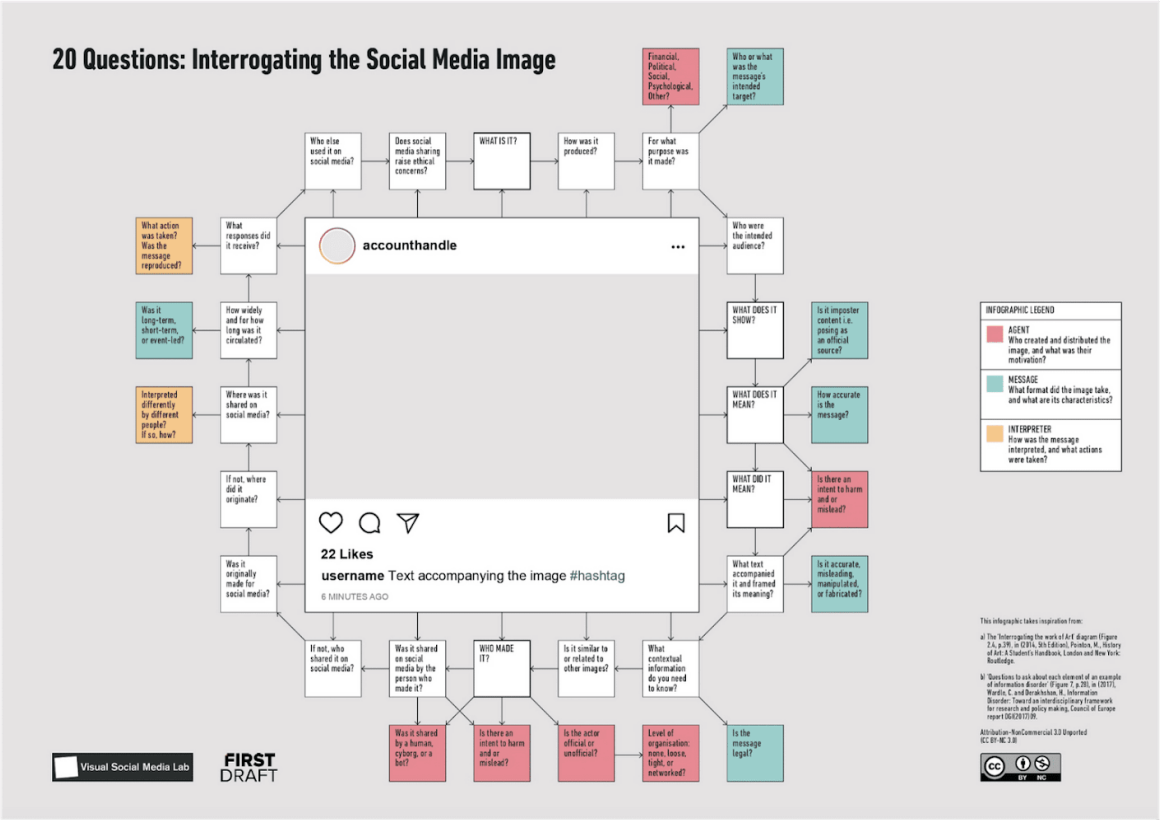
1. Quello che stai guardando è l'originale?
2. Sai chi ha fatto quella foto?
3. Sai dove è stata fatta quella foto?
4. Sai quando è stata fatta quella foto?
5. Sai perché è stata fatta quella foto?

Tra gli strumenti standard che possono aiutarti a svolgere indagini su foto e video ci sono InVID, Yandex Image Search, TinEye, Google Image Search e Forensically. Questi strumenti di verifica si concentrano sull'origine dell'immagine.

Fermo restando che questo approccio rimane decisivo, le strategie e tecniche frequentemente usate nell'ambito di misinformation e disinformation e quelle impiegate nelle varie forme di manipolazione dei media rendono importante considerare come e da chi le immagini sono usate e condivise, e anche quale ruolo possono avere i giornalisti nella potenziale ulteriore diffusione di immagini dannose.

Per andare oltre le forme standard di verifica delle immagini, abbiamo integrato metodi propri della storia dell'arte e domande appositamente ideate per analizzare i contenuti che diffondono disinformation e misinformation. La nostra procedura, "20 Questions for Interrogating Social Media Images" (20 domande per analizzare le immagini dei social media) co-ideata da First Draft e giornalisti, è un ulteriore strumento a disposizione dei giornalisti che hanno necessità di fare indagini su delle immagini.

Analizzare le immagini sui social media



Come suggerisce il titolo, la procedura è costituita da 20 domande che possono essere poste a proposito di qualsiasi immagine pubblicata sui social media (immagini statiche, video, gif, ecc.), e da altre 14 domande che mirano a scavare più a fondo nei diversi aspetti della misinformation e della disinformation. Le domande non hanno un ordine prestabilito, tuttavia ce ne sono cinque che è utile affrontare per prime:

1. Che cos'è?
2. Che cosa mostra?
3. Chi l'ha fatta?
4. Che cosa significava?
5. Che cosa significa?

Le domande dalla 1 alla 3 si avvicinano ai consueti approcci di verifica e sono volte a stabilire con che tipo di immagine si ha a che fare (una fotografia, un video, ecc.), cosa rappresenta e chi l'ha realizzata. Le domande 4 e 5, invece, ci portano da un'altra parte. Esse introducono considerazioni di significato che comprendono non solo ciò che l'immagine mostra, ma anche tutti i significati che si originano dall'uso dell'immagine, anche quelli derivati dalla sua errata identificazione. Considerate

insieme, le domande 4 e 5 ci portano a concentrarci sulla natura mutevole del significato delle immagini e su come il significato che assumono attraverso il loro riutilizzo diventi rilevante di per sé. Ciò non significa soltanto riflettere sul significato che nasce dall'usare le immagini in un contesto nuovo rispetto a quello originario e su come ciò porti a interpretarle erroneamente, ma anche considerare quali sono gli effetti di queste errate interpretazioni. Questo approccio esce dall'ambito della semplice verifica, avvicinandosi all'analisi dei significati delle immagini in discipline come la storia dell'arte e la teoria della fotografia.

Nella fase di sviluppo e di prima applicazione di questa procedura insieme ai giornalisti, abbiamo spesso sentito dire da parte loro che non avevano mai riflettuto sulle immagini in modo così dettagliato. Molti hanno affermato che la procedura li ha aiutati a riconoscere che le immagini sono forme complesse di comunicazione e che è necessario un metodo chiaro per indagare su di loro e sul loro significato.

La maggior parte delle volte non è necessario rispondere a tutte le 20 domande della procedura per comprendere appieno cosa sta succedendo a un'immagine. Le domande vanno considerate uno strumento da tirare fuori all'occorrenza. Nel nostro lavoro le abbiamo trovate particolarmente utili per confrontarci con immagini e video complessi e di alto livello divenuti oggetto di grande attenzione e analisi da parte dei media. Per mostrare tutto ciò nella pratica, ecco tre casi di studio con esempi a livello alto dal Regno Unito e dagli Stati Uniti.

Caso di studio 1: Breaking Point, giugno 2016

Che cos'è?

L'immagine che chiamiamo "Breaking Point" era un poster usato dallo UK Independence Party (UKIP) come parte della sua campagna durante il referendum europeo del 2016. Utilizzava una fotografia scattata dal fotoreporter Jeff Mitchell nell'ottobre 2015 e che parlava della crisi dei rifugiati.

Cosa mostra?

Una grande fila di rifugiati siriani e afghani scortati dalla polizia slovena dal confine tra Croazia e Slovenia al campo profughi di Brezice. Per il poster è stata utilizzata una versione ritagliata della foto, a cui è stato aggiunto il testo “BREAKING POINT: The EU has failed us all” ("BREAKING POINT: L'Ue ci ha delusi tutti") e “We must break free of the EU and take back control of our borders” ("Dobbiamo liberarci dell'Ue e riprendere il controllo delle nostre frontiere"). Visto che i rifugiati sembrano muoversi in massa verso lo spettatore, la foto ha un forte impatto visivo.

Chi l'ha fatta?

La società pubblicitaria di Edimburgo Family Advertising Ltd., ingaggiata dallo UKIP per la campagna Brexit.

Che cosa significava?

Lo UKIP non ha tentato di distorcere il contenuto dell'immagine, ma le ha conferito altri significati attraverso l'aggiunta degli slogan. Sfruttando il preesistente sentimento anti-immigrazione e razzista, la manipolazione punta a generare ulteriore paura dell'immigrazione e dei rifugiati sulla base di affermazioni e insinuazioni non comprovate riguardo la politica delle frontiere dell'Unione Europea.

Che cosa significa?

Nel novembre 2019, in vista delle elezioni generali nel Regno Unito, anche l'organizzazione Leave.EU della campagna elettorale ha utilizzato una versione ritagliata di questa stessa fotografia in un'immagine anti-immigrazione caricata su Twitter, facendo un chiaro riferimento al poster dello UKIP del 2016.

Quali altre domande è utile porsi?

Il responsabile è ufficiale o non ufficiale? L'attore chiave nella creazione e distribuzione dell'immagine, lo UKIP, è un partito politico ufficiale e non il genere di soggetto tipicamente associato alla disinformazione.

L'immagine è simile o collegata ad altre immagini? Alcuni hanno paragonato il manifesto alla propaganda nazista; esso rimanda sia a precedenti immagini anti-migranti, sia a una più lunga storia di manifesti politici britannici che mostrano

code, tra cui quello utilizzato dallo UKIP nel maggio 2016 per parlare di immigrazione dall'UE.

Tre punti chiave da ricordare:

- I politici e i partiti politici ufficiali possono essere responsabili della diffusione della disinformazione.
- Disinformare non significa necessariamente usare immagini false o identificare erroneamente ciò che mostrano. A volte le immagini possono essere usate a sostegno di un messaggio che rappresenta in modo errato una situazione più ampia.
- In certi casi bisogna spingersi oltre la verifica: occorre analizzare in modo critico come immagini reali vengano usate per manipolare, quali effetti hanno e cosa significano.

Esempi di copertura mediatica di questo caso:

Nigel Farage's anti-migrant poster reported to police (Il manifesto di Nigel Farage contro i migranti segnalato alla polizia) - The Guardian

Brexit: UKIP's 'unethical' anti-immigration poster - Al-Jazeera

Nigel Farage accused of deploying Nazi-style propaganda as Remain crash poster unveiling with rival vans (Nigel Farage accusato di ricorrere a una propaganda in stile nazista e i furgoni del Remain lo contrastano con un manifesto contrario) - The Independent

Caso di studio 2: La fotografia del ponte di Westminster, marzo 2017

Che cos'è?

Si tratta di un tweet pubblicato da un account Twitter, apparentemente gestito da un uomo bianco texano, che ha ricevuto una notevole attenzione mediatica. Successivamente si è scoperto che l'account era gestito dalla russa Internet Research Agency e usato per diffondere informazioni fuorvianti e false. Nel tweet si condivideva una fotografia scattata nei momenti successivi all'attacco terroristico al Westminster Bridge di Londra (22 marzo 2017).

Cosa mostra?

Una donna musulmana che passa davanti a un gruppo di persone e una persona a terra, ferita nell'attacco terroristico. Il testo ha connotazioni islamofobiche, sostiene infatti che la donna stia deliberatamente ignorando la persona ferita, ed è corredato da un hashtag apertamente anti-islamico.

Chi l'ha fatta?

L'addetto dell'Internet Research Agency che gestiva l'account Twitter di @SouthLoneStar, anche se al momento del tweet non era noto che l'account fosse gestito dall'Internet Research Agency. La foto è stata scattata dal fotoreporter Jamie Lorriman.

Cosa significava?

Nel marzo 2017 il tweet si presentava come il tweet di un utente texano di destra che interpretava il soggetto dello scatto come una donna musulmana che non si preoccupava della persona ferita. Il tweet suggeriva che la foto fosse un esempio e una dimostrazione di una verità più ampia sui musulmani.

Che cosa significa?

Ad oggi, il tweet è la prova che l'Internet Research Agency ha deliberatamente diffuso disinformazione islamofoba all'indomani di un attacco terroristico.

Quali altre domande è utili porsi?

Quali risposte ha ricevuto? Questo tweet ha ricevuto una risposta significativa da parte dei media mainstream. Ne hanno parlato decine di giornali britannici, in alcuni casi più di una volta. Anche se la maggior parte degli articoli usciti condannava @SouthLoneStar, la copertura ricevuta ha portato il tweet oltre i confini dei social media e lo ha posto all'attenzione di un pubblico mainstream. Dopo la diffusione dell'immagine la donna nella foto ha rilasciato una dichiarazione per dire che all'epoca era sconvolta per gli attacchi, e che "non solo sono stata devastata dal fatto di essere stata testimone dei momenti successivi a un attacco terroristico scioccante e paralizzante, ma ho anche dovuto affrontare lo shock di vedere la mia foto tappezzare tutti i social media, messa da chi non riusciva a guardare oltre il mio abbigliamento e ha tratto conclusioni basate sull'odio e la xenofobia".

L'immagine è simile o collegata ad altre immagini? L'immagine circolata per la maggior parte del tempo era una di sette fotografie scattate alla donna. Altri scatti mostravano chiaramente che era sconvolta, cosa che poche pubblicazioni hanno evidenziato.

Quanto e per quanto tempo è circolata l'immagine? Il fatto che l'immagine abbia raggiunto l'attenzione dei media mainstream significa che il tweet ha avuto larga diffusione. Tuttavia, nel giro di pochi giorni la circolazione del contenuto è notevolmente rallentata. Il contenuto è tornato in circolazione nel novembre 2017, quando si è scoperto che @SouthLoneStar era un account gestito dalla Internet Research Agency. Nei media mainstream questa seconda diffusione di novembre è stata notevolmente inferiore rispetto a quella di marzo.

Tre punti chiave da ricordare:

- La disinformazione visiva non ricorre sempre a immagini totalmente false e può integrare elementi tratti dalla realtà. La fotografia è vera, ma il suo contesto è stato manipolato e falsificato, e si fa affidamento sul fatto che il lettore/spettatore non sappia cosa stesse effettivamente pensando la donna in quel momento.
- I giornalisti dovrebbero riflettere attentamente sul fatto di attirare ulteriore attenzione su contenuti disinformativi che si appellano così tanto all'emotività, così controversi e potenzialmente dannosi, scrivendone nei loro articoli, anche se lo fanno con le migliori intenzioni.
- Si potrebbe dedicare maggiore attenzione a correggere notizie basate sulla disinformazione e a garantire che la vera cornice degli eventi emerga con la massima chiarezza. Il fatto che a novembre la copertura sia stata limitata significa che alcuni lettori potrebbero non aver mai scoperto che il tweet era un'operazione di disinformazione russa.

Esempi di copertura mediatica di questo caso:

'People are making alarming assumptions about this photo of 'woman in headscarf walking by dying man' (Dichiarazioni allarmanti da parte delle persone riguardo la foto della "donna con il velo che passa accanto a un uomo morente") - Mirror

'Who is the real monster?' Internet turns on trolls who criticised 'indifferent' Muslim woman seen walking through terror attack (Chi è il vero mostro? La rete attacca i troll che avevano criticato la donna musulmana "indifferente" vista passeggiare durante l'attacco terroristico) - Daily Mail

British MP calls on Twitter to release Russian 'troll factory' tweets (Parlamentare britannico chiede a Twitter di rilasciare i tweet dell'industria dei troll russa) - The Guardian

Caso di studio 3: il faccia a faccia al Lincoln Memorial, gennaio 2019

Che cos'è?

Un video di un gruppo di studenti della Covington Catholic High School alla manifestazione pro-vita March for Life e di un nativo americano, Nathan Phillips, che accompagnava altri nativi americani nella Indigenous People March, la Marcia dei Popoli Indigeni.

Cosa mostra?

Uno degli studenti della Covington Catholic High School faccia a faccia con Phillips. Le manifestazioni a cui partecipavano i due erano confluite nella stessa piazza, dove un grande gruppo di studenti della Covington con indosso dei cappelli MAGA (Make America Great Again) sembra affrontare Phillips. La foto restituisce l'immagine di un nativo americano che affronta da solo una massa di giovani bulli di estrema destra.

Chi l'ha fatta?

Il video fu caricato per la prima volta su Instagram da un partecipante della Marcia dei Popoli Indigeni. Venne visualizzato quasi 200.000 volte. Qualche ora dopo, il video fu caricato su Twitter, accumulando 2,5 milioni di visualizzazioni prima che l'account originale lo cancellasse. Il video fu poi ripubblicato su diversi social media, guadagnandosi così l'attenzione dei media mainstream. Nel giro di 24 ore vennero pubblicati molti articoli riguardanti il video.

Che cosa significava*?*
Inizialmente si diffuse online la narrativa secondo cui nel video Phillips e gli studenti si stessero fronteggiando e secondo cui gli studenti stessero intenzionalmente schernendo Phillips, dandosi man forte nel farlo.

Cosa significa*?*
Un video più lungo dell'incontro, emerso molti giorni dopo il primo, ricostruiva un quadro molto più complesso. Al memoriale si trovava anche un gruppo di Black Hebrew Israelites, gli ebrei neri israeliti, che schernivano i passanti, compresi gli studenti della Covington e i partecipanti alla Marcia dei Popoli Indigeni. Questa situazione sfociò in un acceso confronto tra i tre gruppi, con Phillips che, secondo quanto riportato, cercava di mettere pace. Il primo video inizia da questo momento della vicenda.

Che altre domande è utile farsi?

Di quali altre informazione di contesto hai bisogno?

Senza il video più lungo e senza sapere che nella piazza erano presenti anche i Black Hebrew Israelites e che stavano alimentando il conflitto, si perde l'intero contesto della vicenda. Il video riprende gli studenti mentre dicono cose razziste, ma la situazione era più complicata e non si risolveva semplicemente in un gruppo di giovani di estrema destra che se la prendeva con un anziano nativo americano.

Dove è stato condiviso sui social media?

Inizialmente il video fu condiviso su Instagram da un partecipante alla Marcia dei Popoli Indigeni, e in questa prima occasione ricevette un'attenzione limitata. Successivamente fu ricondiviso su Twitter e YouTube da altri utenti, il che amplificò grandemente la sua diffusione e assicurò l'attenzione dei media mainstream. L'attenzione si originò dunque dalle ricondivisioni, e non dal primo video condiviso su Instagram.

Tre punti chiave da ricordare:

- Quando contenuti visivi di una tale carica emotiva si diffondono online, è facile perdere di vista il contesto e lasciare che narrazioni superficiali e reazionarie abbiano la meglio.
- A posteriori, alcuni giornalisti sostennero che gli articoli iniziali erano serviti ad alimentare la controversia e dare ulteriore spinta alla narrazione sbagliata. Ciò suggerisce che, senza un'appropriata indagine, i media mainstream possono continuare a diffondere cattiva informazione pur senza volerlo.
- La velocità con la quale il video si diffuse online fece sì che molte testate mainstream abboccassero alla narrativa pompata sui social e non indagassero oltre. Quando la verità dei fatti venne a galla, molti siti di news

furono costretti a ritrattare o a correggere i loro articoli, e qualcuno fu anche perseguito.

Esempi di copertura mediatica di questo caso:

Native American Vietnam Vet Mocked And Surrounded By MAGA Hat-Wearing Teens (Veterano del Vietnam nativo americano deriso e circondato da adolescenti che indossano cappelli MAGA) - UNILAD

Outcry after Kentucky students in Maga hats mock Native American veteran (Proteste dopo che alcuni studenti del Kentucky con cappelli Maga hanno deriso un veterano nativo americano) - The Guardian

Fuller video casts new light on Covington Catholic students' encounter with Native American elder (Video completo getta nuova luce sullo scontro tra gli studenti cattolici di Covington e l'anziano nativo americano) — USA Today

Conclusione

Moltissimi dei contenuti condivisi sui social media sono contenuti visivi. I giornalisti devono essere dotati dell'abilità di analizzare e valutare in maniera critica le immagini per portare alla luce contenuti e intenzioni importanti. La velocità con la quale si diffonde la cattiva informazione visiva pone ulteriormente l'accento sulla necessità da parte dei giornalisti di procedere con cautela e indagare scrupolosamente le storie legate alle immagini prima di pubblicare. Le 20 domande per analizzare le immagini sui social media sono uno strumento supplementare che i giornalisti possono sfruttare quando fanno ricerche sulle immagini, specialmente quando la storia è fortemente incentrata su contenuti visivi. Le domande della procedura non sono tutte rilevanti per ogni immagine, ma le cinque domande principali rappresentano un solido punto di partenza, e si basano su tecniche di verifica di base con lo scopo di rendere possibile una copertura più accurata e approfondita.

APPENDICE

Qui sotto c'è la lista completa dei quesiti che compongono la procedura delle 20 domande, incluse 14 domande rapide pensate specificatamente per affrontare disinformation e misinformation. Le cinque domande in grassetto sono quelle che si prestano a essere considerate per prime, come abbiamo scritto sopra. Le domande rapide riguardano l'artefice o l'interprete del contenuto e il messaggio:

- Artefice (A) - Chi ha creato e condiviso l'immagine e qual era il suo scopo?
- Messaggio (M) - Che formato aveva l'immagine e quali erano le sue caratteristiche?
- Interprete (I) - Come è stato interpretato il messaggio e che azioni sono state intraprese?

1. Che cos'è?
2. Com'è stato prodotto?
3. Per quale scopo è stato creato?
    a. A - Economico, politico, sociale, psicologico o altro?
    b. M - Da chi o da cosa era composto il target a cui era indirizzato il messaggio?
4. Da chi era composto il pubblico indirizzato?
5. **Che cosa mostra?**
6. **Che cosa significa?**
    a. M - Si tratta di un contenuto impostore, ad esempio che finge di provenire da una fonte ufficiale?
    b. M - Quanto è accurato il messaggio?
7. **Che cosa significava?**
    a. A - Vi era l'intenzione di arrecare danni o fuorviare?
8. Qual era il testo che accompagnava l'immagine e ne inquadrava il contenuto?
    a. M - Il testo è accurato, fuorviante, manipolato o inventato?
9. Di quali informazioni di contesto hai bisogno?
    a. M - Il messaggio è legale?
10. L'immagine è simile o collegata ad altre immagini?
11. **Chi l'ha fatta?**
    a. A - L'artefice è ufficiale o non ufficiale?
    b. A - Grado di organizzazione: assente, scarsa, strutturata, in network
12. L'immagine è stata condivisa sui social da chi l'ha creata?
    a. A - È stata condivisa da un essere umano, da un cyborg o da un bot?
    b. A - C'è l'intenzione di arrecare danni o fuorviare?
13. Se no, chi ha condiviso l'immagine sui social?
14. L'immagine è stata originariamente creata per i social media?
15. Se no, dove è stata creata?
16. Dove è stata condivisa sui social media?
    a. I - È stata interpretata in maniera diversa da persone diverse? Se sì, come?
17. Quanto e per quanto è circolata?
    a. M - Era un contenuto a lungo termine, a breve termine o legate a un evento?
18. Quali risposte ha ricevuto?
    a. I - Quali azioni sono state intraprese? Il messaggio è stato riprodotto?
19. Chi altri ha usato il contenuto sui social media?
20. La condivisione sui social media genera problemi di natura etica?

La procedura è stata ispirata da:

1. The "Interrogating the work of Art" diagram (figura 2.4, p.39), in (2014, 5° edizione), Pointon, M. History of Art: A Student's Handbook, London and New York: Routledge.

2. "Questions to ask about each element of an example of information disorder" (figura 7, p. 28), in (2017), Wardle, C. and Derakshan, H., Information Disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe report DGI(2017)09.

6. I deepfake e le nuove tecnologie di manipolazione

Scritto da Sam Gregory

***Sam Gregory** è direttore della programmazione di WITNESS, una realtà che aiuta le persone a usare i video e la tecnologia per combattere a favore dei diritti umani. Premiato esperto di tecnologia e attivista, è un profondo conoscitore di nuove forme di misinformation e disinformation generate da intelligenze artificiali, e conduce un lavoro sulle opportunità e le minacce emergenti per gli attivisti e i giornalisti. È anche co-presidente di un gruppo di esperti della Partnership on AI che si occupa specificatamente di IA e media.*

Nell'estate del 2018 il professor Siwei Lyi, importante ricercatore specializzato in deepfake dell'Università di Albany, pubblicò un articolo che dimostrava come i soggetti dei video deepfake non battessero le palpebre con la stessa frequenza delle persone vere. Questo fatto fu subito riportato da Fast Company, New Scientist, Gizmodo, CBS News e da altre testate, portando molti a credere di aver trovato un modo efficace per riconoscere i deepfake.

Eppure, a poche settimane dalla pubblicazione dell'articolo, il ricercatore ricevette dei video in cui personaggi deepfake battevano le palpebre come veri esseri umani. Ad oggi, l'indicazione sul battito di ciglia non è più utile per scoprire i deepfake. Si è trattato di una momentanea debolezza degli algoritmi, causata dai dati con cui questi erano stati allenati. Ma nel giro di pochi mesi il metodo non era più valido.

Ciò dimostra un verità chiave sul riconoscimento e sulla verifica dei deepfake: ogni approccio tecnico a questa attività rimane valido fino a che le tecniche per produrre media artificiali (syntethic media, i media creati in maniera automatica da algoritmi o intelligenze artificiali) vi si adattano. Non esisterà mai un metodo perfetto per riconoscere i deepfake.

Quindi, come possono fare i giornalisti per verificare i deepfake e le altre forme di media artificiali?

Il primo passo è capire che la natura di questo lavoro è simile a una caccia del gatto al topo, ed essere attenti a come si evolve la tecnologia. In secondo luogo, per indagare se un contenuto è stato manipolato o generato artificialmente, i giornalisti devono apprendere e mettere in pratica l'uso di strumenti e tecniche fondamentali di verifica. Gli approcci alla verifica di immagini e video illustrati nel primo Verification Handbook, così come nelle risorse curate da First Draft per la verifica delle immagini, sono tutti validi. Infine, i giornalisti devono capire che ci muoviamo in un ambiente in cui accade sempre più spesso di sbagliarsi nell'etichettare un

contenuto come deepfake. Pertanto, saper dimostrare che una foto o un video sono stati manipolati è importante tanto quanto saperne verificare l'autenticità.

Questo capitolo approfondisce alcuni approcci fondamentali per verificare un deepfake. Ma prima di tutto è importante comprendere le basi del fenomeno dei deepfake e dei media artificiali.

Cosa sono i deepfake e i media artificiali?

I deepfake sono nuove forme di manipolazioni audiovisive che permettono di creare simulazioni realistiche del viso, della voce e delle azioni di una persona. Fanno sembrare che qualcuno abbia detto o fatto qualcosa che non ha mai né detto né fatto. Stanno diventando sempre più facili da realizzare, dal momento che servono sempre meno immagini di partenza per costruirli, e vengono commercializzati sempre più. Attualmente i deepfake hanno enormi conseguenze soprattutto sulle donne, perché vengono usati per mettere il volto di una persona in immagini e video a contenuto sessuale senza il suo consenso. Si teme tuttavia che i deepfake avranno un impatto ancora più grande sulla società, sulla raccolta delle notizie e sui processi di verifica.

I deepfake sono solo uno degli sviluppi possibili di una gamma di tecniche di creazione di media artificiali rese possibili dall'intelligenza artificiale. Questo arsenale di strumenti e tecniche permette di creare rappresentazioni realistiche di persone che dicono o fanno cose mai dette o mai fatte, persone o oggetti che non sono mai esistiti o eventi che non sono mai accaduti.

Attualmente, le forme di manipolazione consentite dalla tecnologia dei media artificiali sono le seguenti:

- Aggiungere o rimuovere oggetti in un video.
- Alterare le caratteristiche dello sfondo di un video. Per esempio, cambiare le condizioni meteorologiche per far sembrare che un video girato in estate sia stato girato in inverno.
- Simulare e controllare una rappresentazione video realistica delle labbra, delle espressioni facciali o dei movimenti del corpo di uno specifico individuo. Nonostante generalmente la discussione sui deepfake si concentri sui volti, tecniche simili sono state applicate ai movimenti di tutto il corpo o a specifiche parti del viso.
- Generare una simulazione realistica della voce di una specifica persona.
- Modificare una voce esistente con un filtro vocale di un genere diverso o di una specifica persona
- Creare una foto realistica, ma completamente falsa, di una persona che non esiste. La stessa tecnica si può applicare in maniera meno problematica per creare falsi hamburger, falsi gatti ecc.
- Trasferire realisticamente una faccia da una persona a un'altra, ovvero il cosiddetto deepfake.

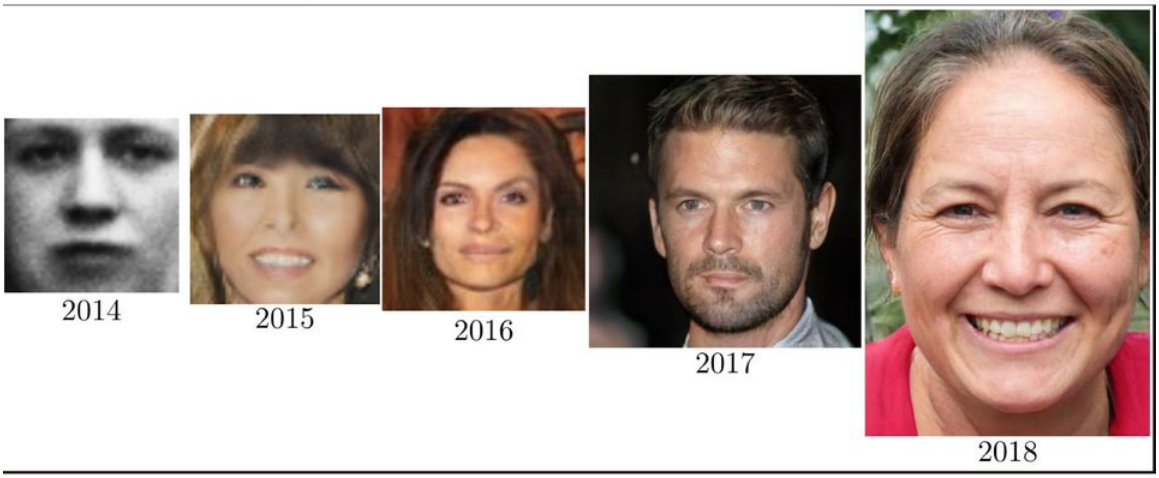
Queste tecniche si basano soprattutto, ma non esclusivamente, su una forma di intelligenza artificiale conosciuta come “deep learning”, e da quelle che vengono chiamate Generative Adversarial Networks (Reti Generative Avversarie), o GAN. Per generare un'unità di contenuto mediatico artificiale si inizia raccogliendo le fonti, immagini o video, della persone o dell’oggetto che si vuole falsificare. Un GAN sviluppa il contenuto artificiale — che sia una simulazione video di una persona reale o uno scambio di volto — usando due network.

Un network genera una ricostruzione plausibile dell’immagine di partenza, mentre un secondo network lavora per riconoscere le manipolazioni del primo. I dati di questo rilevamento vengono rimandati al network che crea le falsificazioni, aiutandolo a migliorare il proprio lavoro.

Ancora nel 2019 molte di queste tecniche — in particolare la creazione di deepfake — continuano a richiedere una potenza computazionale molto elevata, le conoscenze necessarie a mettere a punto il modello creato e, spesso, una significativa post-produzione nell'ambito della CGI per migliorare il risultato finale. Ad ogni modo, nonostante i limiti attuali, i media artificiali riescono già a ingannare gli esseri umani. Per esempio, una ricerca condotta dal progetto FaceForensics++ ha evidenziato che le persone non riescono a riconoscere con affidabilità i metodi attualmente in uso per modificare il movimento delle labbra allo scopo di far combaciare la bocca di qualcuno con una nuova traccia audio. Ciò significa che gli esseri umani non sono intrinsecamente capaci di accorgersi quando un media è stato manipolato artificialmente.

Bisogna anche aggiungere che la sintesi audio sta avanzando più velocemente del previsto, e sta arrivando sul mercato. Per esempio, le API di Google Cloud Text-to-Speech consentono di convertire una porzione di testo in un audio pronunciato con una voce umana realistica. Ricerche recenti si sono inoltre concentrate sulla possibilità di editare una video intervista partendo dal testo o da combinazioni testo/video.

A ciò occorre aggiungere che tutte le tendenze tecnologiche e commerciali indicano che produrre media artificiali diventerà sempre più facile e meno costoso. L’immagine qui sotto mostra, ad esempio, quanto velocemente sia progredita la tecnologia di generazione dei volti.



Credit: EFF

Data la natura del loro funzionamento, simile a un caccia tra gatto e topo, questi network migliorano nel tempo mano a mano che vengono nutriti di dati sulle manipolazioni riuscite e sui loro riconoscimenti. Per questo bisogna usare grande cautela quando si parla di "efficacia" dei metodi di riconoscimento.

Il panorama attuale dei deepfake e dei media artificiali

Deepfake e media artificiali non si sono ancora diffusi al di là della produzione di immagini sessuali non consensuali. Il report del DeepTrace Lab sulla loro diffusione indica che nel settembre 2019 più del 95 per cento dei deepfake apparteneva a questa categoria e ritraeva celebrità, attrici porno e persone comuni. A questo si aggiunge il fatto che le persone hanno iniziato a mettere in discussione i contenuti reali, liquidandoli come deepfake.

Nel corso di alcuni workshop organizzati da WITNESS abbiamo esaminato potenziali vettori di pericolo e lo abbiamo fatto insieme a una serie di esponenti della società civile, inclusi media partecipativi, giornalisti professionisti e fact-checker, ricercatori specializzati sul tema della disinformazione e specialisti OSINT (Open Source INTelligence).

Queste figure si sono concentrate sulle aree in cui le nuove forme di manipolazione potrebbero aggravare, modificare o rafforzare pericoli già esistenti, oppure introdurne di nuovi. Hanno identificato quali pericoli sussistono per giornalisti, fact-checker, investigatori open source e potenziali attacchi ai loro processi di lavoro. Hanno anche sottolineato quali sfide nascono dal fatto di usare la frase "È un deepfake" come frase fatta che fa il paio con "È una fake news".

In ogni contesto i partecipanti ai workshop hanno preso atto dell'importanza di adottare con i deepfake gli stessi approcci che già si adottano con la verifica delle fonti e il fact-checking. I deepfake e i media artificiali, hanno concluso, verranno integrati in campagne complottistiche e di disinformazione già esistenti, e per farlo si attingerà a tattiche (e risposte) in evoluzione in quell'ambito.

Ecco alcuni dei pericoli evidenziati:

- **La credibilità e la reputazione di giornalisti e attivisti verranno attaccate.** Ciò accrescerà forme già in atto di molestie online e violenza, rivolte soprattutto alle donne e alle minoranze. Alcune giornaliste sono già state prese di mira tramite video modificati, come nel caso della riconosciuta giornalista indiana Rana Ayyub.

- **I personaggi pubblici dovranno affrontare la diffusione di loro immagini non autorizzate a sfondo sessuale e contenenti violenza di genere, nonché altri impieghi di quelli che potremmo definire "alter ego credibili".** I politici locali potrebbero risultare particolarmente vulnerabili a tutto questo, data l'abbondanza di loro immagini in circolazione e la minore possibilità che hanno, rispetto ai politici di livello nazionale, di difendersi dagli attacchi appoggiandosi a strutture istituzionali. Inoltre, i politici locali sono spesso fonti chiave di storie che nascono come storie locali e in seguito si gonfiano fino a diventare nazionali.

- **Appropriazione di brand celebri** attraverso operazioni di video editing o altri metodi che rendono possibile applicare un brand giornalistico, governativo, aziendale o di una ONG a un determinato contenuto.

- **Tentativi di disseminare user generated content (contenuti generati dagli utenti, UGC) manipolati nel ciclo delle news**, in combinazione con altre tecniche come il source-hacking o la condivisione di contenuti manipolati con i giornalisti in momenti chiave. Generalmente, l'obiettivo è coinvolgere i giornalisti nella propagazione di quei contenuti.

- **Sfruttamento dei punti deboli dei processi di raccolta e copertura delle notizie**, come ad esempio quando si trasmette da remoto con telecamera singola (come ha riscontrato il team di Reuters che si occupa di UGC), oppure quando si raccoglie materiale in contesti di difficile verifica, come zone di guerra.

- Diventando sempre più comuni e facili da produrre in quantità, i deepfake **contribuiranno al fiume di falsità** che inonda le agenzie di verifica dei media e di fact-checkng con contenuti da verificare o smascherare, il che potrebbe sovraccaricare e sviare chi lavora in queste agenzie.

- **Le organizzazioni di raccolta e di verifica delle notizie subiranno pressioni per dimostrare che un contenuto è vero o che non è stato falsificato.** Figure di potere potranno dichiarare che un certo contenuto è un deepfake sfruttando il principio di plausibile negabilità

Un punto di partenza per verificare i deepfake.

Data la natura delle indagini sui media e delle emergenti tecnologie dei deepfake, bisogna accettare che l'assenza di prove di un'avvenuta manipolazione non sarà sufficiente a dimostrare che il contenuto in questione non sia stato effettivamente alterato.

Giornalisti e investigatori devono sviluppare un approccio di misurato scetticismo nei confronti di foto, video e audio. Devono essere consapevoli che con l'aumentare delle conoscenze e dei timori sui deepfake, questi media saranno messi in dubbio sempre più spesso. È inoltre essenziale sviluppare grande familiarità con gli strumenti per condurre indagini sui media.

Tenendo conto di tutto ciò, ecco le operazioni essenziali che non possono mancare quando si analizzano e verificano i deepfake e casi di manipolazione artificiali dei media:

1. Esaminare il contenuto alla ricerca di distorsioni o sfocature tipiche dei media artificiali.
2. Ricorrere a metodi esistenti di indagine e di verifica dei video.
3. Utilizzare, se disponibili, nuovi metodi emergenti basati sull'intelligenza artificiale e nuovi approcci di indagine.

Esaminare il contenuto per trovare di distorsioni o sfocature tipiche dei media artificiali.

Questo approccio è il meno solido per identificare deepfake e altre manipolazioni artificiali, in particolare se si considera la natura in divenire della tecnologia. Detto ciò, deepfake e contenuti artificiali fatti male possono mostrare prove visibili di errori. Tra gli elementi da cercare in un deepfake ci sono:

- Distorsioni sulla fronte o all'attaccatura dei capelli, o riscontrabili quando i movimenti del volto superano una determinata area di movimento.
- Mancanza di dettagli della dentatura.
- Pelle eccessivamente liscia.
- Assenza del battito delle palpebre.
- Oratore fermo, che parla senza alcun movimento reale della testa o senza cambiare espressione.

- Sfocature o glitch visibili quando una persona si volta dalla posizione di fronte a quella di lato.

Attualmente, è più probabile riuscire a vedere alcune di queste alterazioni analizzando il video fotogramma per fotogramma, motivo per cui può essere utile estrarre i vari fotogrammi per studiarli singolarmente. Questo suggerimento non è però valido per le sfocature dovute a cambi di posizione da frontale a laterale, che si colgono meglio se osservati in sequenza. Per questo è consigliabile adottare entrambi gli approcci.

Ricorrere a metodi esistenti di indagine e di verifica dei video.

Anche con i deepfake, così come quando si affrontano altre forme di manipolazione dei media e shallowfake (come video editati o estrapolati dal loro contesto), occorre ricorrere a ben consolidate pratiche di verifica. Le pratiche OSINT di verifica già esistenti non hanno perso la loro importanza e i capitoli e i casi studio del primo Verification Handbook dedicati alla verifica delle immagini e dei video sono un buon punto di partenza. Dal momento che attualmente la maggior parte dei deepfake e delle manipolazioni non è del tutto artificiale, bensì è il risultato di cambiamenti apportati a un video-fonte originale, la ricerca inversa delle immagini può rivelarsi utile per cercare altre versioni dei fotogrammi del video. Puoi anche verificare se il paesaggio o i punti di riferimento del video corrispondono alle immagini su Google Street View.

Inoltre, cercare di capire come il contenuto si è diffuso e per opera di chi può portare a scoprire informazioni utili per stabilire se sia ragionevole fidarsi o meno di quell'immagine o di quel video. Le conoscenze fondamentali per determinare fonte, data, orario e ragioni della creazioni di un contenuto sono imprescindibili per stabilire se le persone o gli eventi in esso rappresentati siano veri o meno (per una preparazione di base in merito a questo, guarda la guida di First Draft). Altra cosa che rimane fondamentale è contattare la persona o le persone presenti nel video per un commento, e per vedere se possono fornire informazioni concrete utili a confermare o negare l'autenticità del contenuto.

Governo, università, piattaforme e laboratori di innovazione nel campo del giornalismo stanno sviluppando nuovi strumenti per individuare i media artificiali e aumentare le risorse a disposizione di chi indaga. Nella maggior parte dei casi, questi strumenti vanno considerati integrativi rispetto a un metodo di verifica basato sulle buone pratiche.

Strumenti come InVID e Forensically sono di supporto sia nella verifica delle immagini a partire dalla loro provenienza, sia in limitate analisi forensi.

Tra gli strumenti gratuiti in questo campo di lavoro ci sono:

- FotoForensics: uno strumento di analisi forense delle immagini che include l'Error Level Analysis, l'analisi del livello di errore, per individuare i punti di un'immagine in cui potrebbero essere stati aggiunti degli elementi.

- Forensically: una serie di strumenti per la rilevazione della clonazione, l'analisi del livello di errore, i metadati delle immagini e diverse altre funzioni.

- InVID: un'estensione per browser web che consente di dividere i video in fotogrammi, eseguire la ricerca inversa delle immagini su più motori di ricerca, ingrandire ed esaminare fotogrammi e immagini attraverso una lente di ingrandimento e applicare filtri a scopo di indagine su immagini fisse.

- Reveal Image Verification Assistant: uno strumento che offre una gamma di algoritmi per rilevare la manipolazione delle immagini, analizzare i metadati, eseguire la geolocalizzazione tramite GPS, estrarre miniature EXIF ed eseguire la ricerca inversa delle immagini tramite Google.

- Ghiro: uno strumento digitale di analisi forense open source.

Occorre notare che quasi tutti questi strumenti sono stati progettati per la verifica delle immagini, non dei video. Nel mondo delle indagini questo è un punto debole, che comporta il fatto che per analizzare i video sia ancora necessario estrarre singole immagini, operazione che InVID può facilitare. Gli strumenti descritti si rivelano più efficaci con video ad alta risoluzione e non compressi in cui sono stati, ad esempio, rimossi o aggiunti oggetti video. La loro utilità diminuisce quanto più il filmato è stato compresso e quante più volte è stato ri-salvato o condiviso attraverso diversi social media o piattaforme di condivisione di video.

Se sei alla ricerca di nuovi strumenti di indagine per risolvere problemi già esistenti o esaminare deepfake, puoi andare a vedere quali sono gli strumenti condivisi dagli accademici. Uno dei più importanti centri di ricerca in questo campo, che si trova all'Università di Napoli, fornisce accesso online ai propri codici per eseguire una serie di operazioni, come scoprire le "impronte digitali" della fotocamera (Noiseprint), rilevare punti di giunzione nelle immagini (Splicebuster) e scoprire parte copiate e spostate o rimosse all'interno di un video.

Mano a mano che i media artificiali avanzano, nuove forme di indagine, sia manuali che automatiche, verranno affinate e integrate a strumenti di verifica già esistenti e usati da giornalisti e investigatori, e, potenzialmente, anche a metodi di verifica basati sulle piattaforme. È importante che i giornalisti si tengano aggiornati sugli strumenti disponibili, senza tuttavia diventarne troppo dipendenti.

Approcci emergenti basati sull'intelligenza artificiale e sull'indagine forense

All’inizio del 2020 non sono ancora in commercio strumenti testati di rilevamento basati sui GAN. Si può tuttavia prevedere che entro la fine dell'anno alcuni saranno disponibili ai giornalisti sul mercato, sia in forma di plugin che come strumenti integrati su piattaforme. Per una panoramica aggiornata sullo stato dell’arte del mondo delle indagini sui media, inclusi questi strumenti, si può leggere il contributo di Luisa Verdoliva "Media Forensics and Deepfakes: An overview".

Questi strumenti faranno generalmente affidamento sulla disponibilità di dati di apprendimento provenienti da media artificiali prodotti con i GAN, e saranno quindi in grado di usare questi dati per riconoscere altri contenuti prodotti usando la stessa tecnica o tecniche simili. Ad esempio, programmi di indagine come FaceForensic+++ generano dei fake utilizzando strumenti di deepfake già esistenti disponibili ai consumatori, per poi sfruttare questa grande massa di immagini false per addestrare gli algoritmi a identificare i fake. Ciò significa che potrebbero non essere efficaci con contenuti prodotti con i metodi e le tecniche di falsificazione più recenti.

Questi strumenti si presteranno molto meglio a rilevare media generati da GAN, rispetto alle attuali tecniche di indagine. Integreranno anche nuove forme di strumenti di analisi capaci di far fronte in maniera migliore ai progressi nella produzione di immagini artificiali. Ciononostante, dato il meccanismo di antagonismo che permette l'evoluzione dei media artificiali, non saranno infallibili. Un concetto chiave di cui far tesoro è che ogni segnale di produzione artificiale deve essere ricontrollato e corroborato da altri metodi di verifica.

I deepfake e i media artificiali si stanno evolvendo velocemente, e le tecnologie per crearli sono sempre più diffuse, commercializzate e facili da usare. Per produrre contenuti artificiali hanno bisogno di molti meno contenuti di partenza di quanto si possa immaginare. Mentre nuove tecnologie di identificazione emergono e vengono integrate nelle piattaforme e negli strumenti giornalistici e OSINT, il modo migliore per verificare video e immagini resta quello di ricorrere a metodi di analisi già esistenti e combinarli a strumenti di indagine in grado di riconoscere la manipolazione delle immagini. Fidarsi dell’occhio umano non è una strategia affidabile!

7. Monitorare e raccontare storie dai gruppi chiusi e nelle applicazioni di messaggistica

Scritto da: Claire Wardle

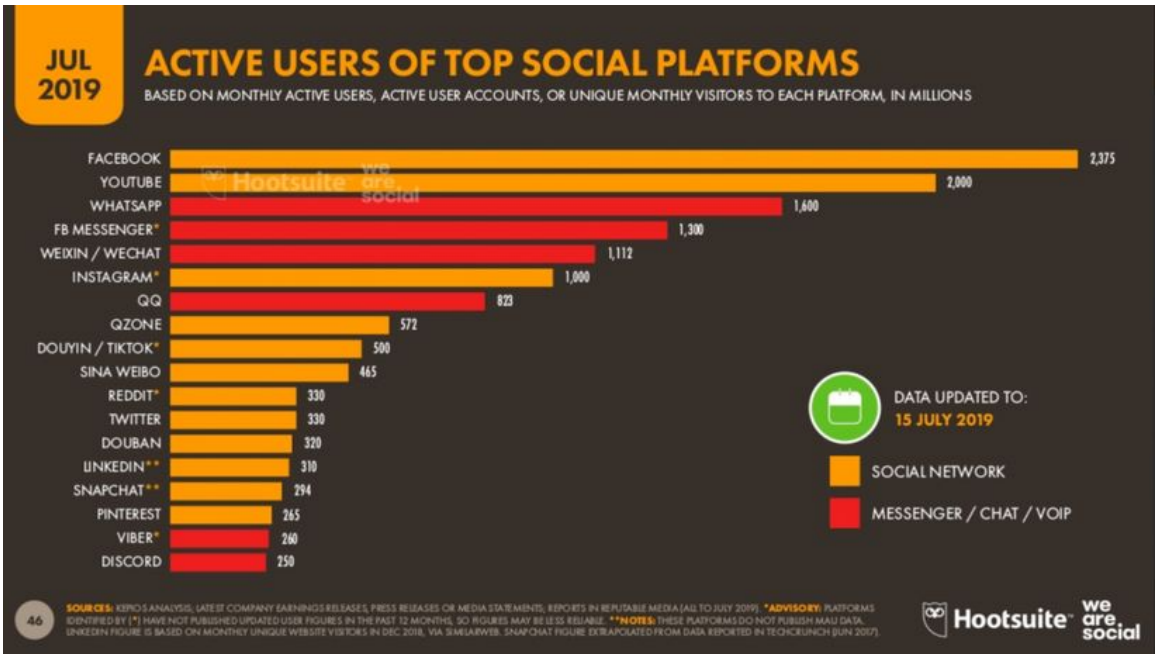
***Claire Wardle** è a capo della direzione strategica e della ricerca di First Draft, un'organizzazione internazionale non profit che supporta giornalisti, accademici ed esperti di tecnologia nell'affrontare le sfide legate alla fiducia e alla verità nell'epoca digitale. Ha fatto parte dello Shorenstein Center for Media, Politics and Public Policy della Harvard's Kennedy School ed è stata direttrice della ricerca presso il Tow Center for Digital Journalism della Graduate School of Journalism della Columbia University e direttrice dei social media per l'UNHR (l'agenzia delle Nazioni Unite per i rifugiati).*

Nel marzo del 2019 Mark Zuckerberg annunciò che Facebook avrebbe messo al centro la privacy; il che significava che l'azienda si avviava a dare più importanza ai gruppi Facebook, riconoscendo così la crescente tendenza tra le persone a comunicare con un numero ristretto di interlocutori e in spazi privati. Negli ultimi anni l'importanza dei gruppi ristretti nella comunicazione sui social si è resa evidente a chi di noi lavora in questi spazi.

In questo capitolo illustrerò le diverse piattaforme e applicazioni esistenti, parlerò della sfida di monitorare questi spazi e concluderò con osservazioni di carattere etico su ciò che questo lavoro comporta.

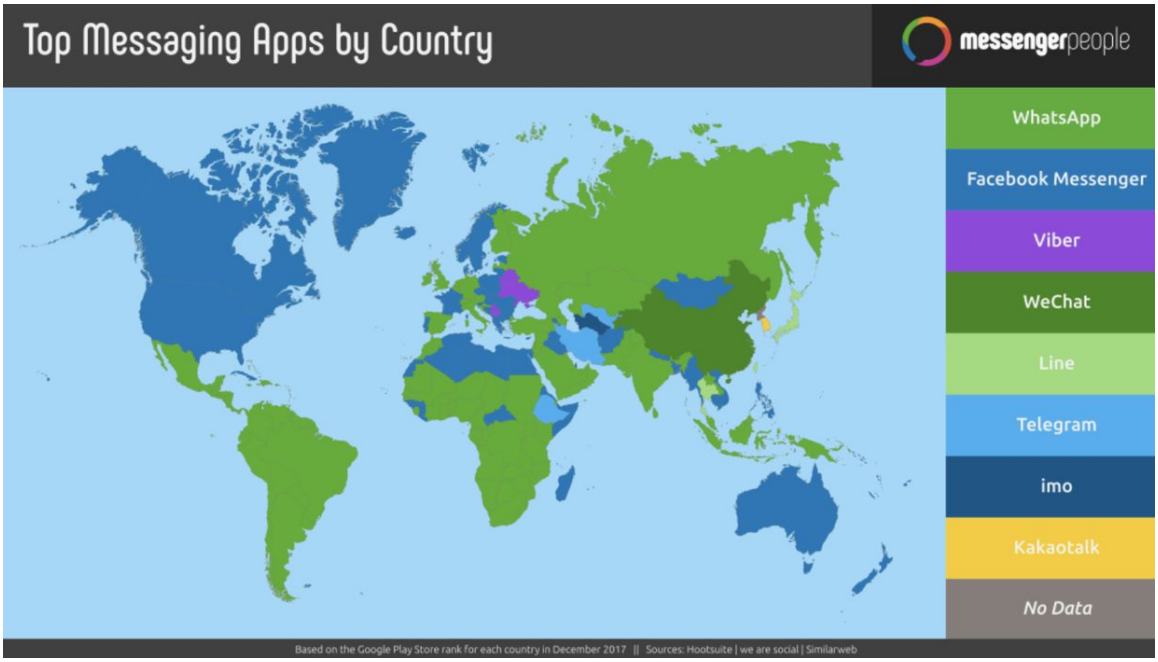
Piattaforme e applicazioni diverse

Recenti ricerche condotte da We Are Social mostrano che Facebook e YouTube continuano a predominare, seguite dalle tre altre piattaforme più popolari, ovvero WhatsApp, FB Messenger e WeChat.



Oggi in molte regioni nel mondo la principale fonte di notizie per la maggior parte dei consumatori sono le app di messaggistica, in particolare Whatsapp, che lo è, ad esempio, in Brasile, India e Spagna.

Whatsapp e FB Messenger sono certamente popolari in tutto il mondo, ma in alcuni paesi sono altre applicazioni a essere predominanti. Ad esempio, l'alternativa in Iran è Telegram, in Giappone è Line, in Corea del Sud KakaoTalk e in Cina WeChat.



Questi siti presentano tutti leggere differenze in termini di crittografia, gruppi, funzionalità broadcast e opzioni aggiuntive come, per esempio, e-commerce all'interno dell'app.

Gruppi chiusi di Facebook

Esistono tre tipologie di gruppi Facebook: aperti, chiusi e nascosti.

- I gruppi aperti possono essere trovati attraverso la ricerca di Facebook e chiunque può unirsi a essi.
- I gruppi chiusi possono essere trovati, ma per unirsi bisogna fare richiesta.
- I gruppi nascosti non possono essere trovati attraverso la ricerca di Facebook e vi si può entrare solo su invito.

Su Facebook i gruppi stanno riscuotendo sempre più successo, in parte perché è l'algoritmo di Facebook che spinge gli utenti a ritrovarsi nei gruppi, in parte perché al giorno d'oggi le persone preferiscono passare il proprio tempo con chi già conoscono o con chi condivide i loro punti di vista o interessi.

Discord

Secondo Statista, nel luglio 2019 Discord ha registrato 250 milioni di utenti attivi (per fare un paragone, Snap ne fece 294 milioni, Viber 260 e Telegram 200). Discord è popolare all'interno della community dei videogiocatori, ma negli ultimi anni si è fatto conoscere anche come sito dove gli utenti possono riunirsi in "server" (che su Discord corrispondono a delle specie di gruppi) per coordinare campagne di disinformazione.

Una delle caratteristiche di Discord e di qualche gruppo chiuso di Facebook è che prima di essere accettati nel gruppo occorre rispondere a delle domande. Queste domande possono riguardare la professione, la religione, le convinzioni politiche o le proprie posizioni nei confronti di alcune questioni sociali.

Crittografia, gruppi e canali

Una delle ragioni per cui queste piattaforme e queste applicazioni sono diventate così popolari è che offrono la possibilità della cifratura, a livelli diversi. Whatsapp e Viber offrono la crittografia "end to end" e sono attualmente le più sicure. Altre, come Telegram, FB Messenger e Line, eseguono la cifratura solo se questa funzione viene attivata.

Alcune app hanno gruppi o canali in cui le informazioni vengono condivise con un gran numero di persone. Su WhatsApp il gruppo più grande può avere al massimo 256 persone. I gruppi di FB Messenger arrivano a 250. Su Telegram, un gruppo può essere privato oppure rintracciabile attraverso lo strumento di ricerca e può avere fino a 200 persone. Una volta raggiunto quel numero può diventare un supergruppo,

al quale possono aggiungersi fino a 75mila persone. Telegram ha anche i canali, la soluzione di broadcast all'interno dell'app: ciò significa che l'utente può unirsi al canale e vedere quello che viene pubblicato al suo interno, ma non può pubblicare nessun contenuto in risposta.

Monitoraggio continuo

Non c'è dubbio sul fatto che le applicazioni di messaggistica privata siano canali di diffusione di disinformazione. È difficile stabilire se ci sia più disinformazione su queste piattaforme o sui siti di social media, perché non c'è modo di vedere quello che viene condiviso tramite le prime. Ma sappiamo che questo fatto rappresenta un problema, come hanno dimostrato casi eclatanti in India, Francia, Indonesia e negli Stati Uniti, dove durante le sparatorie di El Paso e di Dayton nell'agosto del 2019, ci furono casi di dicerie e menzogne circolanti su Telegram e su FB Messenger.

Bisogna chiedersi se giornalisti, ricercatori, fact-checker, operatori sanitari e umanitari debbano stare all'interno dei gruppi chiusi per monitorare così la circolazione della disinformazione. E, se sì, come dovrebbero portare avanti il loro lavoro per non venire meno all'etica e preservare la propria sicurezza?

Sebbene questo lavoro ponga grandi sfide, farlo è possibile. In ogni caso, ricorda che molti utenti scelgono queste app proprio per non essere intercettati. Le usano proprio perché sono criptate, e quindi si aspettano un certo livello di privacy. Questo aspetto deve rimanere centrale per chiunque lavori in questi ambienti. Nonostante si possa entrare in questi ambienti e monitorarli, è di primaria importanza essere consapevoli della responsabilità che si ha verso i membri dei gruppi, che spesso non capiscono cosa potrebbe accadere.

Tecniche di ricerca

Trovare questi gruppi può essere difficile dato che il procedimento cambia per ogni applicazione o piattaforma. Per trovare gruppi su Facebook puoi cercare l'argomento di interesse sul motore di ricerca del social e poi applicare il filtro per i gruppi. Se vuoi utilizzare operatori booleani più avanzati, cerca su Google usando le tue parole chiave e aggiungendo site:facebook.com/groups.

Con Telegram, se hai un telefono Android puoi cercare da dentro l'app, ma non lo puoi fare se hai un iPhone. Ci sono applicazioni per desktop come https://www.telegram-group.com/. Per Discord ci sono siti analoghi, come https://disboard.org/search.

Le decisioni da prendere quando si entra in gruppo

Come già accennato, l'ingresso in alcuni di questi gruppi è subordinato a una serie di domande. Prima di provare a entrare, confrontati con il tuo caporedattore o il tuo responsabile sulle risposte da dare. Dirai la verità su chi sei e sul perché ti stai

unendo al gruppo? C'è modo di rispondere ed entrare restando deliberatamente vaghi? Altrimenti, come puoi giustificare la decisione di nascondere la tua identità (questo potrebbe essere necessario se ti stai unendo a un gruppo che potrebbe mettere in pericolo la tua sicurezza se ti identifichi come giornalista)? Se ottieni l'accesso, darai un qualche contributo al gruppo o ti limiterai a "spiare" per trovare informazioni da verificare altrove?

Raccogliere contenuti dai gruppi in maniera automatica

È possibile risalire a gruppi "aperti" cercando link pubblicati su altri siti: link che poi appaiono sui motori di ricerca. È quindi possibile usare metodi automatici per raccogliere contenuti da questi gruppi. Questo è quanto hanno fatto ad esempio i ricercatori che hanno seguito le elezioni in Brasile e in India, e conosco storie di altre organizzazioni che hanno fatto un lavoro simile.

Questa tecnica permette di tenere sotto controllo più gruppi simultaneamente, cosa altrimenti spesso impossibile. Tuttavia, solo una piccola percentuale dei gruppi può essere trovata in questo modo. Solitamente si tratta dei gruppi che cercano di avere un pubblico il più ampio possibile, e che pertanto sono poco rappresentativi dei gruppi in generale. Oltre a ciò, questo procedimento pone dei problemi di natura etica. Tuttavia, si possono applicare delle restrizioni per tenere al sicuro i dati, non condividerli con altri e rendere i messaggi anonimi. Per fare questo tipo di lavoro servono procedure trasversali a più settori.

Tipline

Un'altra tecnica consiste nel mettere a punto una tipline, ovvero una linea di contatto con il pubblico per incentivare le persone a inviare contenuti. Per creare una tipline è fondamentale elaborare una call to action semplice e chiara e spiegare come si intende usare i contenuti che arriveranno. Serviranno solo a monitorare delle tendenze o farai anche debunking una volta che avrai indagato su ciò che ti è stato inviato?

Tornando alle questioni etiche, che hanno una rilevanza significativa quando ci si occupa di app di messaggistica private, è importante non limitarsi soltanto a "prendere" contenuti, ovvero a fare un lavoro estrattivo. Inoltre, mettendo per un attimo da parte l'etica, tutte le ricerche dimostrano che quando le persone non sanno come verranno usati i suggerimenti che viene chiesto loro di dare sono molto meno inclini a mandarne. Le persone aiutano più volentieri se si sentono trattate come collaboratori.

Un altro aspetto legato alle tipline, ad ogni modo, riguarda la facilità con cui ci si può prendere gioco di loro, mandando contenuti bufala o inviando lo stesso contenuto molte volte, da parte di un singolo individuo o di un piccolo gruppo, per far sì che sembri rappresentare un problema più grande di quanto non sia in realtà.

Scrivere storie con le informazioni ottenute da gruppi chiusi di messaggistica: questioni etiche

Trovati i contenuti di cui parlare, la domanda che si pone è come farlo. Bisogna essere trasparenti su come si è arrivati alle informazioni? Nelle indicazioni per la propria community, molti gruppi chiedono che ciò di cui si discute al loro interno non venga condiviso al di fuori. Se il gruppo è pieno di contenuti di disinformazione, quale impatto avrà la tua storia? Puoi confermare ciò che hai trovato in altri gruppi o in altri ambienti online? Parlandone metti a rischio la tua sicurezza, quella dei tuoi colleghi o della tua famiglia?

Ricordati che rivelare informazioni sensibili (o anche peggio) su giornalisti e ricercatori fa parte delle strategie di alcuni dei gruppi online più pericolosi.

Conclusioni

Fare giornalismo con informazioni da e su gruppi chiusi e app private di messaggistica è un'attività piena di sfide. Ciò non cambia il fatto che queste fonti diventeranno spazi sempre più importanti dove si condividono informazioni. Per prima cosa rifletti sulle questioni sottolineate in questo capitolo, parla con i tuoi colleghi e i tuoi responsabili e, se non avete linee guida all'interno della redazione per lavorare su questo tipo di contenuti, iniziate a elaborarle. Non ci sono regole standard su come agire. Dipende dalla storia, dalla piattaforma, dal giornalista e dalle linee editoriali della redazione. Ma è importante che vengano presi in considerazione tutti i dettagli prima di iniziare questo tipo di indagini.

7a. Caso di studio: Bolsonaro in ospedale

Scritto da: Sérgio Lüdtke

***Sérgio Lüdtke** è giornalista e redattore di Projeto Comprova, un gruppo di 24 testate che indaga in forma collaborativa sulle notizie non confermate che circolano in Brasile riguardo politica e ordine pubblico. Nel 2018 Comprova ha esaminato contenuti sospetti sulle elezioni presidenziali in Brasile, condivisi sui social media e sulle app di messaggistica.*

Il 6 settembre del 2018, un mese prima delle elezioni presidenziali in Brasile, il candidato di estrema destra Jair Bolsonaro tenne un evento della sua campagna presidenziale nel centro di Juiz de Fora, una città di 560.000 abitanti a 200 chilometri da Rio de Janeiro.

Era passata una settimana da quando Bolsonaro si era posizionato in testa ai sondaggi sul primo turno delle presidenziali in Brasile. Era diventato il favorito dopo che la candidatura dell'ex presidente Luiz Inácio Lula da Silva, precedentemente solo in testa ai sondaggi, era stata bocciata dalla Corte Elettorale Superiore.

Tuttavia, nelle simulazioni di ballottaggio Bolsonaro risultava perdere contro tre dei quattro candidati avversari a lui più vicini nei sondaggi.

La situazione di Bolsonaro era preoccupante, dal momento che il candidato aveva a disposizione soltanto due blocchi giornalieri di 9 secondi ciascuno per farsi pubblicità nelle trasmissioni elettorali televisive aperte a tutti. Il regolamento elettorale brasiliano prevede infatti che radio e televisione mettano gratuitamente del tempo a disposizione dei partiti politici per pubblicizzare i loro programmi.

Questo tempo viene distribuito a seconda di quanti seggi ciascun partito si è aggiudicato nelle precedenti elezioni alla Camera dei Deputati. La mancanza di seggi di Bolsonaro significava molto poco tempo a disposizione in onda gratuitamente. Per questo il candidato dovette puntare sui suoi sostenitori sui social network, e cercare il contatto diretto in strada con gli elettori.

A Juiz de Fora, come in altre città che aveva visitato in precedenza, Bolsonaro partecipò a una marcia in cui venne portato sulle spalle dai suoi sostenitori. Era seguito da una folla di ammiratori, quando la marcia fu improvvisamente interrotta. Nel mezzo della folla un uomo si era avvicinato al candidato e lo aveva accoltellato. Il coltello procurò a Bolsonaro una profonda ferita all'addome e scoperchiò il vaso di Pandora sui social network.

Si diffusero voci e teorie complottiste. C'era chi accusava Adélio Bispo de Oliveira, l'uomo che aveva accoltellato Bolsonaro, di essere legato al partito dell'ex presidente Dilma Roussef, rimossa dal suo incarico nel 2016. Foto false mostravano l'attentatore in piedi di fianco a Lula. Quel Bispo era stato iscritto nel partito di sinistra Partido Socialismo e Liberdade (PSOL), e il rifiuto da parte dei suoi avvocati di rivelare chi stesse pagando le loro parcelle non fece altro che alimentare le teorie del complotto.

Al contempo, sulle piattaforme social prendevano piede messaggi e video che cercavano di danneggiare Bolsonaro. Alcuni di questi contenuti sostenevano che l'accoltellamento fosse stato inscenato, che Bolsonaro in realtà era stato in ospedale per curare un tumore e che le foto che mostravano l'operazione chirurgica erano false.

L'accoltellamento diede a Bolsonaro una ragione per ritirarsi dalle le attività della campagna elettorale, ma contemporaneamente gli fecero guadagnare una migliore posizione nei sondaggi (alla fine, come sappiamo, Bolsonaro vinse le elezioni).

Il 19 settembre, quasi due settimane dopo l'aggressione, Eleições sem Fake, un programma di monitoraggio dei gruppi WhatsApp creato dall'Università di Minas Gerais, individuò una registrazione audio che stava venendo diffusa in giro. L'audio era stato condiviso da 16 dei circa 300 gruppi monitorati dal programma, alcuni dei quali sostenitori di Bolsonaro.

Quello stesso giorno Comprova, la nostra organizzazione, iniziò a ricevere, sempre tramite WhatsApp, richieste da parte dei lettori di verificare l'autenticità della registrazione.

Nell'audio, che durava poco più di un minuto, un uomo arrabbiato con una voce somigliante a quella di Bolsonaro discuteva animatamente con qualcuno che sembrava essere suo figlio, Eduardo, e si lamentava del fatto di essere trattenuto in ospedale. Nella registrazione, l'uomo diceva di non poter più sopportare "questo teatrino", suggerendo che l'intera faccenda fosse una montatura.

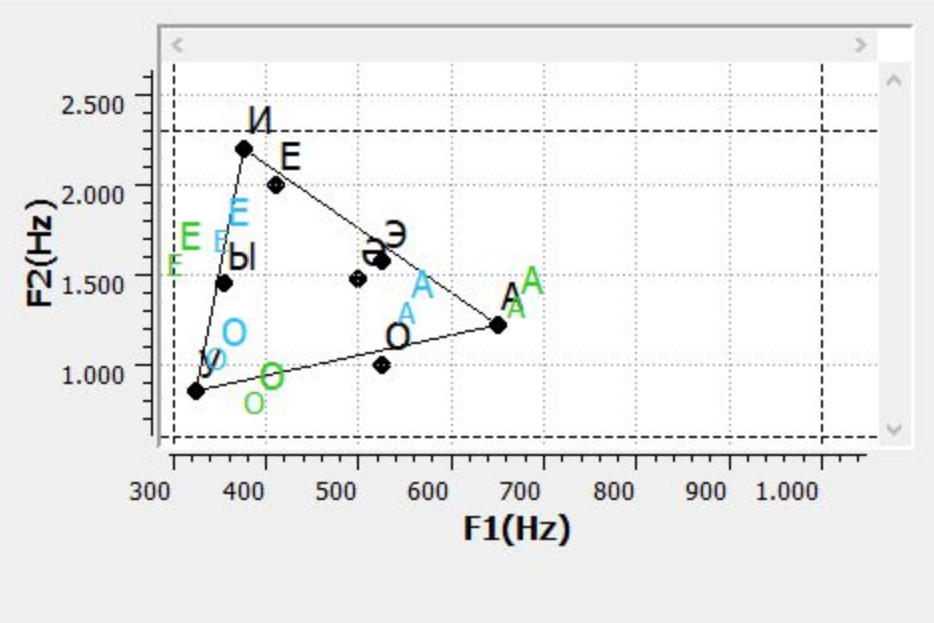
Quel giorno Bolsonaro era ancora ricoverato presso l'unità semi-intensiva dell'Albert Einstein Hospital di San Paolo. Il rapporto dei medici diceva che non aveva febbre, che veniva alimentato per via endovenosa e che aveva recuperando le funzioni intestinali.

Comprova non riuscì a trovare la fonte originaria della registrazione. L'audio si era diffuso principalmente su WhatsApp, dove all'epoca era ancora possibile condividere file in più di 20 conversazioni. Ciò fece sì che l'audio si diffondesse rapidamente e che si facesse ben presto strada sugli altri social network. Risalire alla fonte originaria divenne impossibile (da allora WhatsApp ha limitato il numero di gruppi ai quali si possono inoltrare messaggi).

Incapace di identificare l'autore (o gli autori) della registrazione, Comprova si orientò verso un'indagine più convenzionale e richiese una relazione da parte di un esperto dell'Instituto Brasileiro de Perícia (Istituto Forense Brasiliano). Gli esperti paragonarono la registrazione virale con la voce di Bolsonaro presa da un'intervista dell'aprile 2018, e conclusero che la voce che si sentiva nella registrazione condivisa sui social network non era la voce del candidato.

Gli esperti condussero un'analisi qualitativa dei segnali discorsivi e delle caratteristiche della voce e del parlato dell'uomo della registrazione, poi confrontarono questi parametri in ogni campione di voce e di parlato. In questa analisi, esaminarono i pattern di vocali e consonanti, il ritmo e la velocità d'eloquio, l'intonazione, la qualità della voce e le abitudini mostrate dall'oratore, così come l'uso di specifiche parole e regole grammaticali.

L'immagine qui sotto mostra ad esempio l'analisi della frequenza delle "formanti", termine con cui si indicano le frequenze di risonanza prodotte dalle vibrazioni del tratto vocale, la cavità in cui viene filtrato il suono prodotto a livello della laringe. L'aria dentro il tratto vocale vibra con frequenze diverse, a seconda della forma e del grado di apertura. L'immagine mostra un'analisi delle frequenze delle formanti usando le vocali "a", "e" e "o". Le vocali verdi corrispondono a estratti audio che abbiamo ottenuto da WhatsApp, e quelle blu a estratti presi dall'intervista rilasciata da Bolsonaro qualche giorno prima dell'aggressione a suo danno.



Ulteriori analisi evidenziarono il fatto che la persona che parlava nell'audio di WhatsApp aveva l'accento tipico di chi abita nella campagna dello stato di São Paulo. Ma ciò non veniva riscontrato nelle caratteristiche della parlata di Bolsonaro. Negli

estratti messi a confronti si individuarono differenze di risonanza, articolazione, velocità di eloquio e deviazioni fonetiche.

Comprova consultò un secondo esperto. Anche questo professionista concluse che la voce della registrazione differiva da quella di Bolsonaro per molte ragioni. Disse che il tono della voce appariva leggermente più acuto di quello di Bolsonaro. Notò inoltre che la velocità di pronuncia era maggiore di quella riscontrabile in un altro video del candidato dall'ospedale.

Un elemento che sosteneva ulteriormente la tesi della falsità dell'audio era la scarsa qualità della registrazione. Secondo specialisti con esperienza si tratta di un trucco tipico: abbassare la risoluzione di audio, video e foto li rende più difficili da analizzare.

Venendo alla reazione di Bolsonaro, i suoi figli, Flavio e Carlos, dichiararono sui social che l'audio era una "fake news".

Se l'audio fosse diventato virale oggi, probabilmente sarebbe stato più difficile credere che la voce apparteneva a Bolsonaro. Dati i soli 18 secondi al giorni concessi in televisione e l'assenza del candidato dai dibattiti elettorali per via del ricovero e delle cure, prima delle elezioni la voce dell'attuale presidente non era molto conosciuta. Ciò fece sì che la finta registrazione audio potesse ingannare molte persone.

Ad ogni modo, a distanza di più di un anno, risulta ancora difficile capire perché gruppi che appoggiavano Bolsonaro o che facevano campagna per la sua candidatura abbiano condiviso un audio che, se si fosse dimostrato autentico, avrebbe potuto distruggere la sua candidatura. Non conosceremo mai il vero motivo per cui quei gruppi condivisero con tanta foga quel contenuto. Ciononostante, questa vicenda rimane una forte prova del fatto che se c'è un contenuto esplosivo, questo si diffonderà rapidamente sui social media.

8. Indagare sui siti internet

Scritto da: Craig Silverman

***Craig Silverman** è il media editor di BuzzFeed News, per cui è responsabile della copertura a livello globale di temi riguardanti piattaforme, disinformazione online e manipolazione dei media. Ha curato il "Verification Handbook" e il "Verification Handbook for Investigative Reporting," ed è l'autore di "Lies, Damn Lies, and Viral Content: How News Websites Spread (and Debunk) Online Rumors, Unverified Claims and Misinformation."*

Chi si occupa di manipolazione dei media sfrutta i siti internet per generare profitti, raccogliere indirizzi email e altre informazioni personali, oppure per creare una testa di ponte online per le proprie operazioni. I giornalisti devono sapere come si indaga intorno a una presenza sul web e come la si ricollega, quando possibile, a operazioni più ampie che coinvolgano account social, applicazioni, aziende o altre realtà.

Ricorda che testi, immagini e persino interi siti possono sparire nel tempo, soprattutto dopo che hai iniziato a contattare persone e fare domande. Una buona abitudine, da considerare parte integrante del tuo lavoro, è usare la Wayback Machine per salvare pagine importanti del sito su cui stai indagando. Se non riesci a salvare correttamente la pagina su quello strumento, ricorri ad altri come archive.today. Quest'ultimo ti permette di generare link alle pagine archiviate, così potrai fornire prove di ciò che hai scoperto senza dover linkare direttamente il sito che diffonde disinformation e misinformation (Hunchly è un ottimo strumento a pagamento per creare automaticamente archivi personali delle pagine web mentre lavori). Questi strumenti di archiviazione sono fondamentali anche per osservare i cambiamenti nell'aspetto di un sito internet nel corso del tempo. Raccomando anche di installare l'estensione per browser della Wayback Machine, così da rendere molto semplice archiviare le pagine e guardarne versioni precedenti.

Un'altra utile estensione per browser è Ghostery, che segnala i tracker presenti su una pagina. Ciò ti aiuta a capire velocemente se un sito utilizza Google Analytics e/o ID di Google AdSense, il che ti aiuterà nell'uso di una delle tecniche illustrate a seguire.

Questo capitolo esaminerà quattro categorie da analizzare quando si fanno indagini su un sito internet: il contenuto, il codice, gli analytics, la registrazione e gli elementi a essa collegati.

Il contenuto

Nella maggior parte dei siti si trovano informazioni, anche minime, su ciò di cui si occupa quel sito. La pagina “Informazioni”, oppure le descrizioni presenti nel footer o altrove nel sito, sono buoni punti di partenza. Di contro, la mancanza di informazioni chiare potrebbe essere un indizio che il sito sia stato creato in fretta, o che si stia cercando di nascondere informazioni sul suo proprietario e sui suoi obiettivi.

Oltre a leggere qualsiasi cosa nel sito che stia sotto "Informazioni", "Su di noi" o titoli simili, passa in rassegna l'intero contenuto, cercando di capire chi lo gestisce, qual è il suo obiettivo e se fa parte di un più ampio network o iniziativa. Ecco alcune delle cose su cui concentrarsi:

- Nella pagina delle informazioni è riportata l'identità del proprietario o di una persona giuridica? Anche il fatto che il sito non abbia una pagina "Informazioni" è indicativo.

- C'è una nota di copyright a tutela di un'azienda o di una persona in fondo alla homepage o in qualsiasi altra pagina?

- Nella sezione relativa alla privacy, o ai termini e alle condizioni del servizio, sono elencati i nomi e i recapiti di qualcuno o i dati di una persona giuridica? Questi nomi sono diversi da quelli che hai trovato nel footer, nella pagina "Informazioni" o in altri parti del sito?

- Se il sito pubblica articoli, vai a guardare da chi sono stati scritti e se si può cliccare sul nome dell'autore. Se sì, controlla se il link porta a una pagina con più informazioni sull'autore, ad esempio una biografia o i link ai suoi account social.

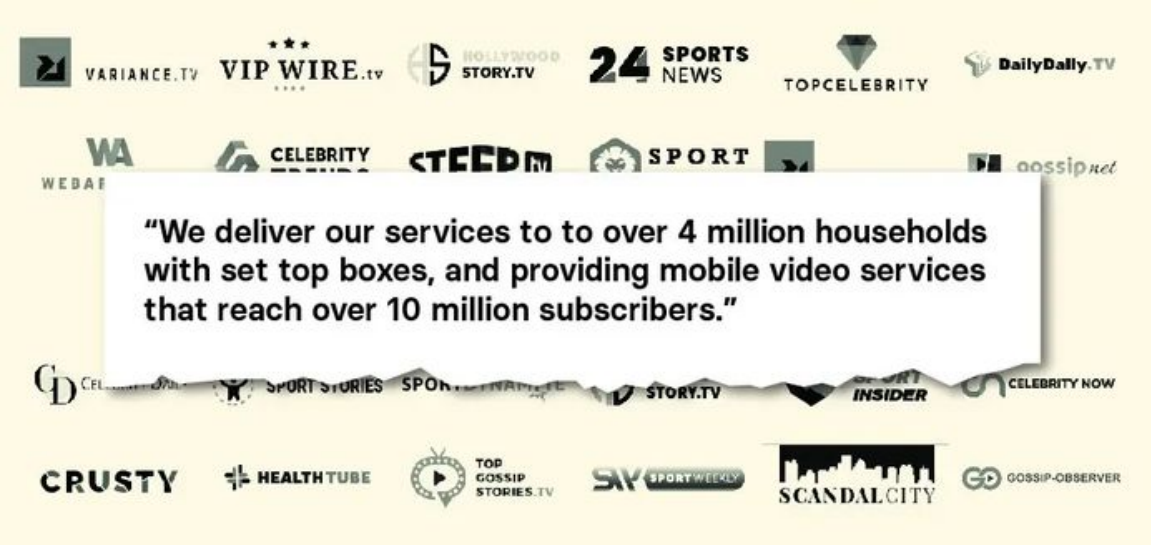
- Ci sono rimandi agli account social collegati al sito? Possono essere ad esempio piccole icone collocate in alto, in basso o a lato della homepage, oppure form incorporati per mettere like alla pagina Facebook. Se ci sono icone di piattaforme come Facebook o Twitter, posizionati con il mouse sopra di esse e guarda nell’angolo in basso a sinistra del browser per vedere a che URL rimandano: spesso quando si crea un sito in fretta non ci si cura di inserire nel template il link allo specifico profilo social, ma ci si limita a un link generico, ad esempio facebook.com/, senza l’indicazione dello username.

- Nel sito vengono mostrati prodotti, clienti, testimonial o altre persone o aziende che potrebbero essere collegati ad esso e su potrebbe valere la pena approfondire?

- Accertati di andare oltre la homepage. Clicca su tutti i menu principali e scrolla fino al footer per cercare altre pagine che valga la pena visitare.

Una parte importante dell'analisi dei contenuti è verificare se sono originali. I testi della pagina "Informazioni" o altri testi generici sono stati copiati da qualche altra parte? Il sito diffonde informazioni false o ingannevoli o che contribuiscono a promuovere una specifica agenda mediatica?

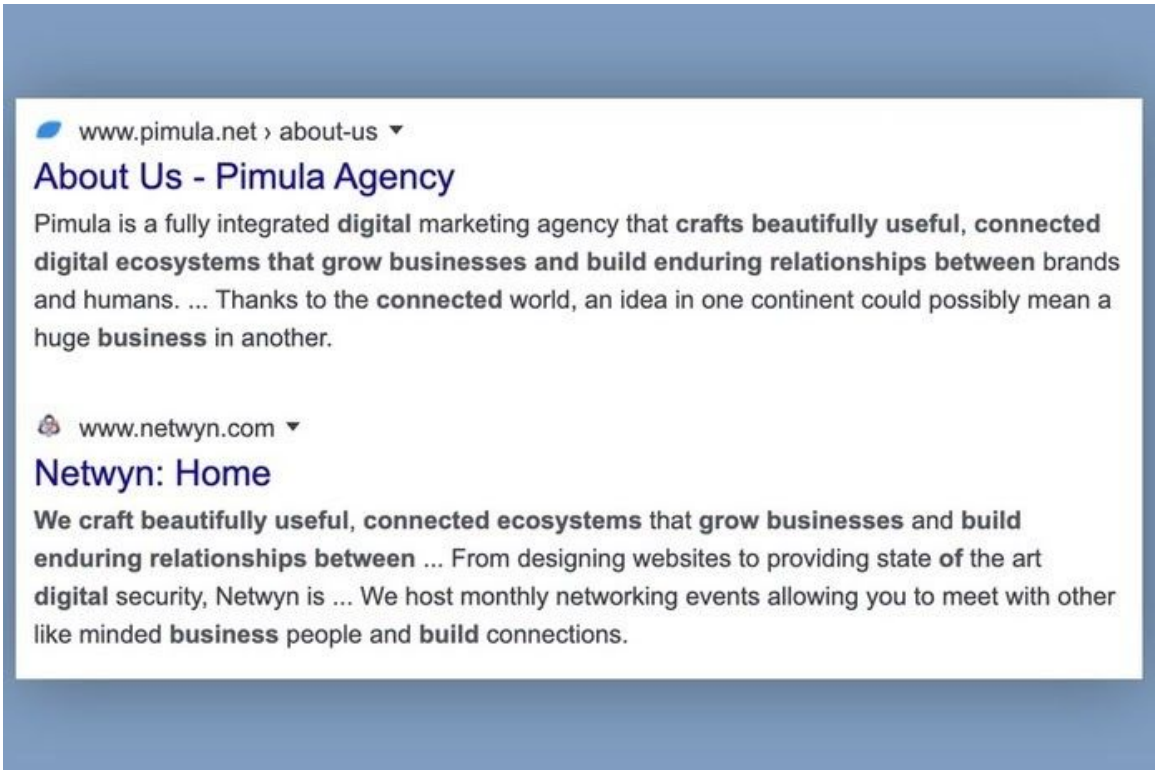
Nel 2018 indagai su un grande piano fraudolento di pubblicità digitale in cui erano coinvolti siti di contenuti e applicazioni mobile, nonché società di comodo, impiegati fantasma e aziende inesistenti. Alla fine trovai più di 35 siti collegati al piano. Molti di questi siti li trovai copiando e incollando il testo della pagina "Informazioni" di uno di essi nel campo di ricerca di Google. Trovai all'istante circa 20 siti che avevano lo stesso identico testo:



I truffatori dietro questa frode avevano anche creato siti delle loro società di copertura, che le facevano sembrare legittime agli occhi di potenziali partner di ad network che le visitavano per dovuta diligenza. Ad esempio, nella homepage della società chiamata Atoses comparivano le foto di numerosi dipendenti. La ricerca inversa delle immagini di Yandex (il miglior motore di ricerca per i volti) permise di scoprire velocemente che molte di queste foto erano immagini d'archivio.

Nel footer del proprio sito, Atoses riportava inoltre questo testo: “We craft beautifully useful, connected ecosystems that grow businesses and build enduring relationships between online media and users” (“Creiamo ecosistemi incredibilmente utili e connessi che fanno crescere le aziende e costruiscono relazioni durature tra i media online e gli utenti”).

Lo stesso testo appariva nei siti di almeno altre due agenzie di marketing:



Se un'azienda usa immagini d'archivio per mostrare i propri dipendenti e pubblica sul suo sito testi copiati da altri siti, non è ciò che dichiara di essere.

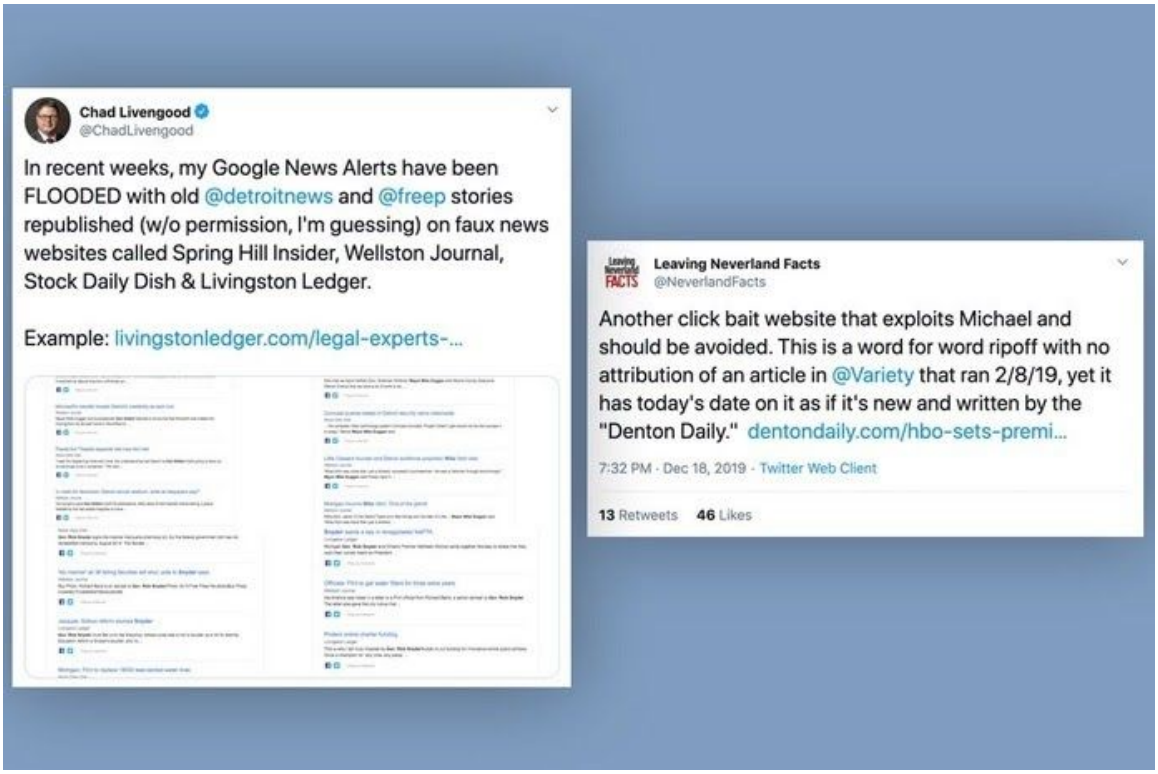
Altra cosa utile è prendere porzioni di testo da alcuni articoli su un sito e fare copia e incolla per cercarli su Google o tramite un altro motore di ricerca. Ci sono infatti siti che si presentano come fonti di informazione, ma che in realtà si limitano a copiare informazioni pubblicate da fonti vere.

Nel 2019 finii su un sito chiamato forbesbusinessinsider.com, che sembrava un sito di news dedicato all'industria tecnologica. In realtà, ciò che faceva era copiare massicciamente articoli da un'ampia gamma di altri siti. Tra questi, ironicamente, anche un mio articolo a proposito dei siti locali di fake news.

Altro passaggio base è prendere la URL di un sito e cercarla su Google. Per esempio “forbesbusinessinsider.com”. Questa operazione ti darà la misura di quante pagine del sito sono state indicizzate, e potrebbe portarti a scoprire altri casi di persone che si sono occupate del sito o che ne hanno parlato. Puoi anche controllare se il sito è presente nelle liste di Google News: carica la pagina principale di Google News e digita “forbesbusinessinsider.com” nel campo di ricerca.

Un altro suggerimento è prendere la URL del sito e copiarla nella barra di ricerca di Twitter.com o di Facebook.com. Scoprirai così se qualcuno pubblica dei link a quel sito. Durante un'indagine mi capitò di finire su un sito chiamato dentondaily.com. La sua homepage mostrava solo pochi articoli risalenti all’inizio del 2020, ma quando

cercai il suo dominio su Twitter scoprii che prima di quella data aveva diffuso contenuti plagiati, cosa che gli utenti avevano scoperto e di cui si erano lamentati. Queste vecchie storie erano state cancellate dal sito, ma i tweet provavano ciò che era stato fatto in passato.

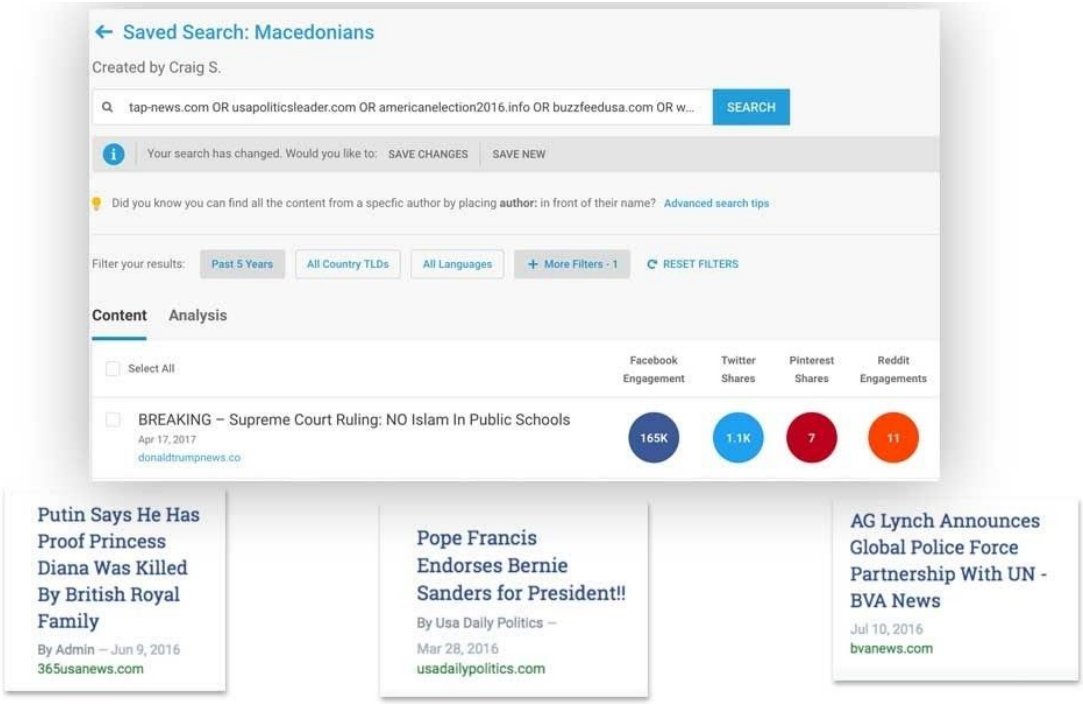


Una volta svolte indagini sui contenuti del sito, è tempo di capire come questi vengono disseminati. Esamineremo due strumenti per farlo: BuzzSumo e CrowdTangle.

Nel 2016 assieme a Lawrence Alexander, ricercatore, mi occupai di siti di informazione politica americana gestiti da oltreoceano. Presto le nostre indagini si focalizzarono su alcuni siti gestiti da Veles, una cittadina nel nord della Macedonia. Grazie ai dati di registrazione del dominio (tema che verrà approfondito a seguire) individuammo più di 100 siti di politica americana gestiti da quella località. Volevo avere la misura di quanto i contenuti di questi siti fossero popolari e che tipo di storie venissero pubblicate, allora presi la URL di molti dei siti in apparenza più attivi e impostai una ricerca su di loro su BuzzSumo, uno strumento che elabora liste dei contenuti dei siti ordinati secondo il grado di interazione generato su Facebook, Twitter, Pinterest e Reddit (c'è una versione gratuita, anche se quella a pagamento offre molti più risultati).

Mi accorsi immediatamente che gli articoli che avevano generato più interazioni su Facebook erano articoli completamente falsi. Ciò ci fornì un'informazione chiave e una prospettiva diversa da quella delle storie precedenti. L'immagine sottostante

mostra la schermata della ricerca base di BuzzSumo, dove sono elencati i livelli di interazione su Facebook, Twitter, Pinterest e Reddit riguardanti un sito specifico, e qualche esempio di storie false del 2016:

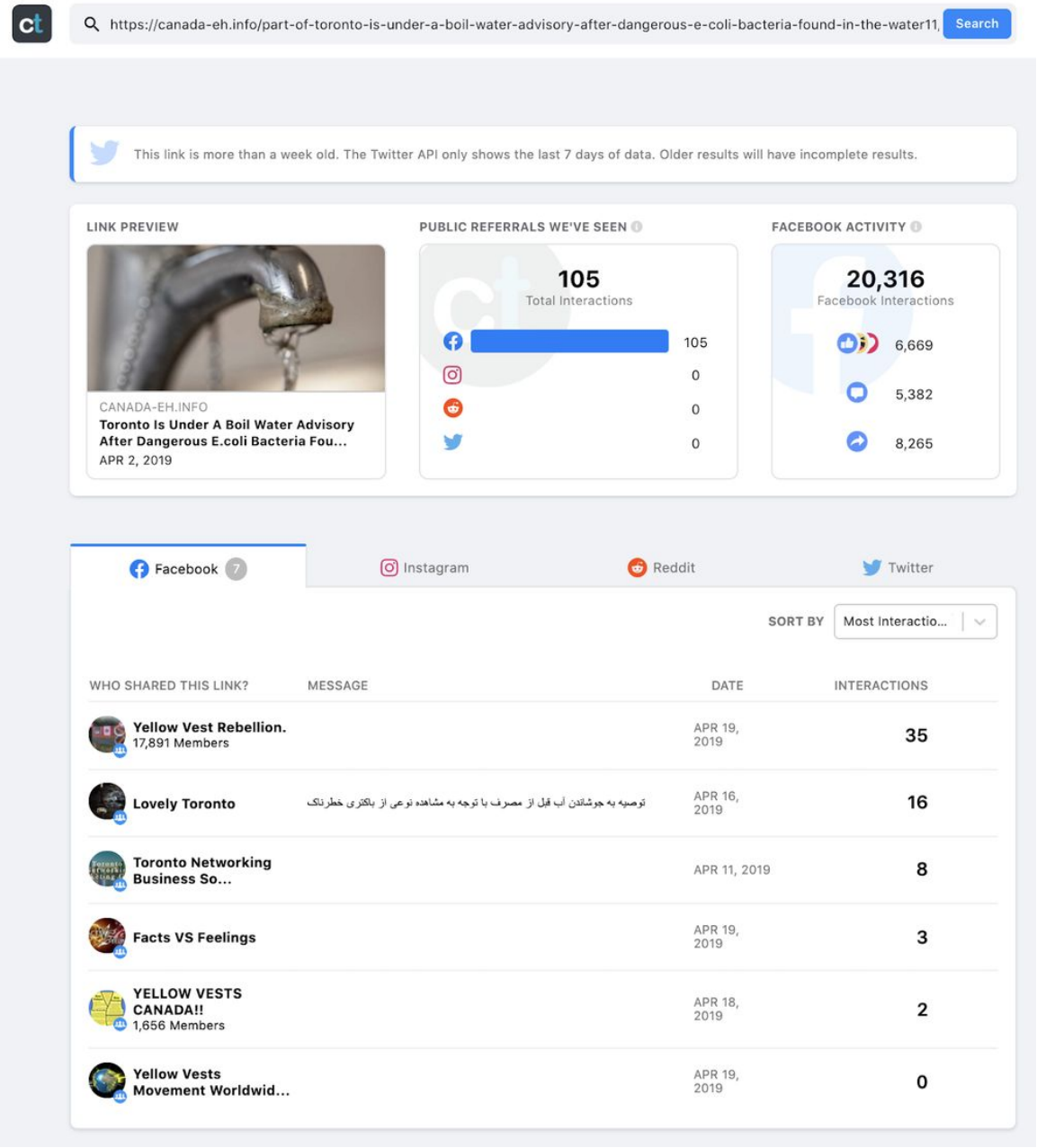


Un altro modo per capire come i contenuti di un sito si diffondono su Facebook, Twitter, Instagram e Reddit è usare l'estensione per browser gratuita di CrowdTangle o il suo strumento di ricerca dei link basato sul web. Entrambe le opzioni offrono le stesse funzionalità, ma concentriamoci sulla versione web (questi strumenti sono gratuiti, ma serve un account Facebook per accedervi).

La differenza chiave tra BuzzSumo e CrowdTangle è che in BuzzSumo puoi inserire la URL di un sito e ottenere automaticamente la lista dei contenuti di quel sito che hanno generato maggiore interazione. CrowdTangle serve invece a controllare una URL specifica all'interno di un sito. Quindi se inserisci buzzfeednews.com su CrowdTangle, ti verranno mostrati i livelli di interazione relativi soltanto alla homepage, mentre se inserisci la stessa URL su BuzzSumo lo strumento analizzerà l'intero dominio a partire dai contenuti più importanti. Un'altra differenza è che lo strumento di ricerca dei link di CrowdTangle e la sua estensione mostrano le interazioni su Twitter soltanto per i precedenti sette giorni. BuzzSumo invece conteggia tutte le condivisioni su Twitter degli articoli del sito.

Una volta, ad esempio, inserii nella barra di ricerca tramite link di CrowdTangle la URL di una vecchia storia falsa riguardante un avviso pubblico di contaminazione delle acque a Toronto (il sito più tardi cancellò la storia, ma nel momento in cui

scrivo la URL è ancora attiva). CrowdTangle ci dice che dal momento della sua pubblicazione la URL ha generato più di 20.000 reazioni, commenti e condivisioni su Facebook. Ci mostra inoltre alcune delle pagine e dei gruppi pubblici che hanno condiviso quel link e ci dà la possibilità di visualizzare dati simili per Instagram, Reddit e Twitter. Ricorda: la scheda di Twitter mostra soltanto i tweet degli ultimi sette giorni.



Come puoi notare, il breve elenco delle pagine e dei gruppi che vediamo non riflette realmente l'alto numero di interazioni su Facebook. Questo è in parte dovuto al fatto che Facebook rimosse alcune delle pagine chiave che avevano diffuso il link quando

questo fu pubblicato la prima volta. Ciò è utile a ricordarci che CrowdTangle mostra solo i dati degli account attivi e non mostra *tutti* gli account pubblici che hanno condiviso una determinata URL. Quella che ci offre è una selezione, che rimane comunque incredibilmente utile perché spesso rivela chiare connessioni tra un sito e determinati account social. Se una pagina Facebook condivide in maniera costante o esclusiva contenuti da uno stesso sito, ciò può voler dire che sito e pagina sono gestiti dalle stesse persone. Allora puoi esaminare più a fondo la pagina per confrontare le informazioni che vi trovi con quelle del sito, operazione che può potenzialmente portare a identificare le persone coinvolte e i loro moventi. Tra i risultati sulle condivisioni dei link su Facebook elencati da CrowdTangle, alcuni possono provenire da utenti che hanno condiviso l'articolo su un gruppo Facebook. Prendi nota degli account che hanno condiviso il link, e guarda se hanno condiviso altri contenuti dal sito. Anche qui potrebbe esserci un collegamento.

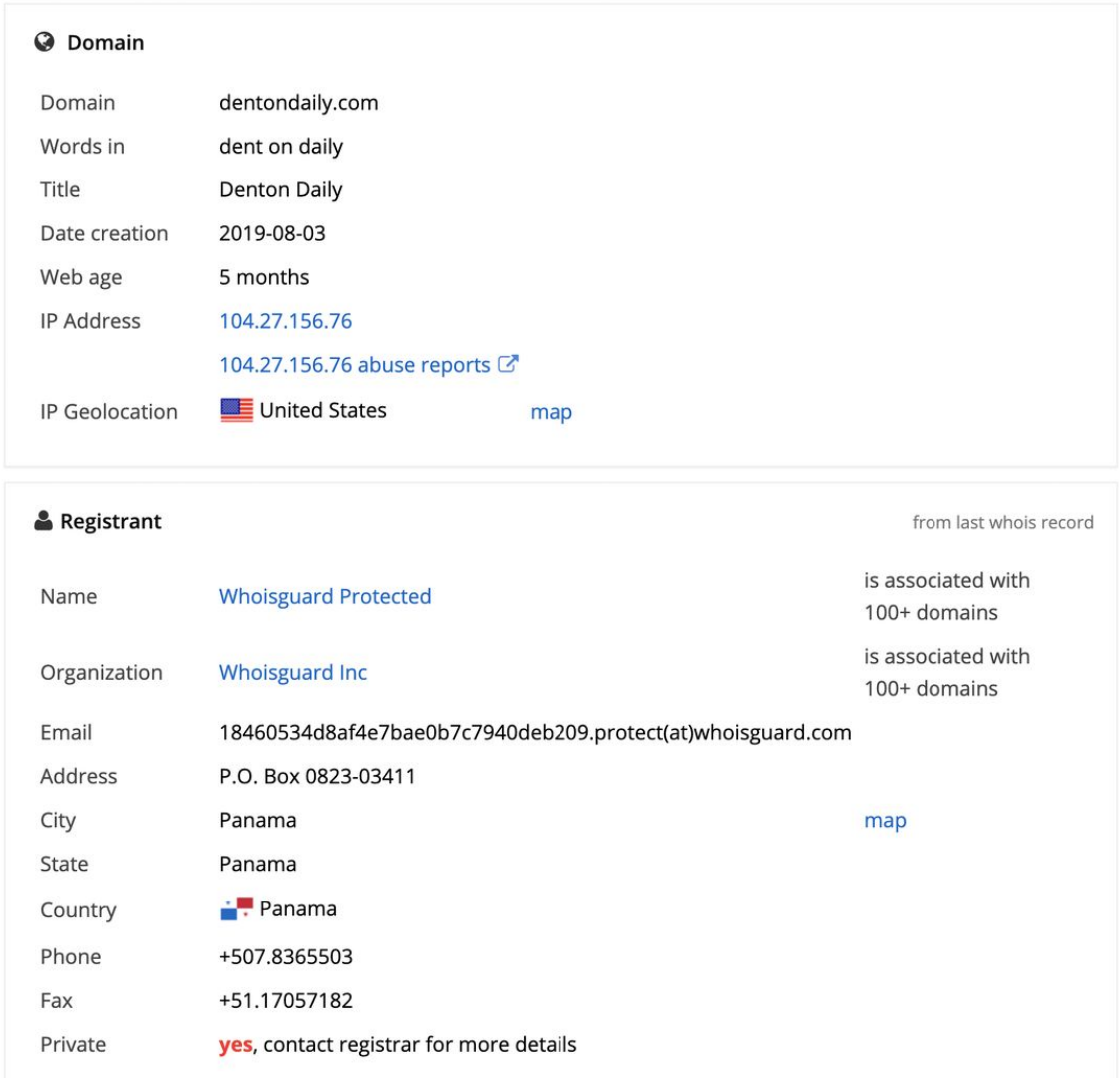
Registrazione

Le informazioni di base sulla creazione e sulla storia di ogni nome di dominio sul web sono memorizzate all'interno di un database centrale. In casi fortunati, in questi database troviamo anche informazioni su chi ha pagato per registrare il dominio. Queste informazioni possono essere recuperate attraverso una ricerca Whois, offerta da molti strumenti gratuiti. Tramite altri, ottimi strumenti gratuiti o a basso prezzo si possono ricavare anche altre informazioni: ad esempio chi sono stati i proprietari del dominio nel corso del tempo, su che server è stato ospitato e altri dettagli utili.

Va precisato che i costi da sostenere per proteggere la privacy dei dati personali quando si registra un dominio sono relativamente ridotti. Se eseguendo una ricerca Whois su un dominio vedi comparire tra i risultati una serie di diciture come "Registration Private," "WhoisGuard Protected," o "Perfect Privacy LLC", significa che la privacy è protetta. Potrai comunque vedere la data di registrazione del dominio più recente, la data di scadenza del dominio e l'indirizzo IP del server web su cui il sito è ospitato.

DomainBigData è uno dei migliori strumenti gratuiti per fare indagini su un nome di dominio e sulla sua storia. Su questo strumento puoi cercare anche inserendo, al posto di una URL, un indirizzo email, il nome di una persona o quello di un'azienda. Altri servizi economici di cui prendere nota sono DNSlytics, Security Trails e Whoisology. Alternativa ottima, ma più costosa, è la piattaforma per indagini Iris sviluppata da DomainTools.

Facciamo un esempio: se cerchiamo dentondaily.com su DomainBigData, vediamo che la privacy è stata protetta. Al posto del nome di chi ha effettuato la registrazione compare la dicitura "Whoisguard Protected." Per fortuna, possiamo comunque vedere che l'ultima registrazione risale all'agosto 2019.



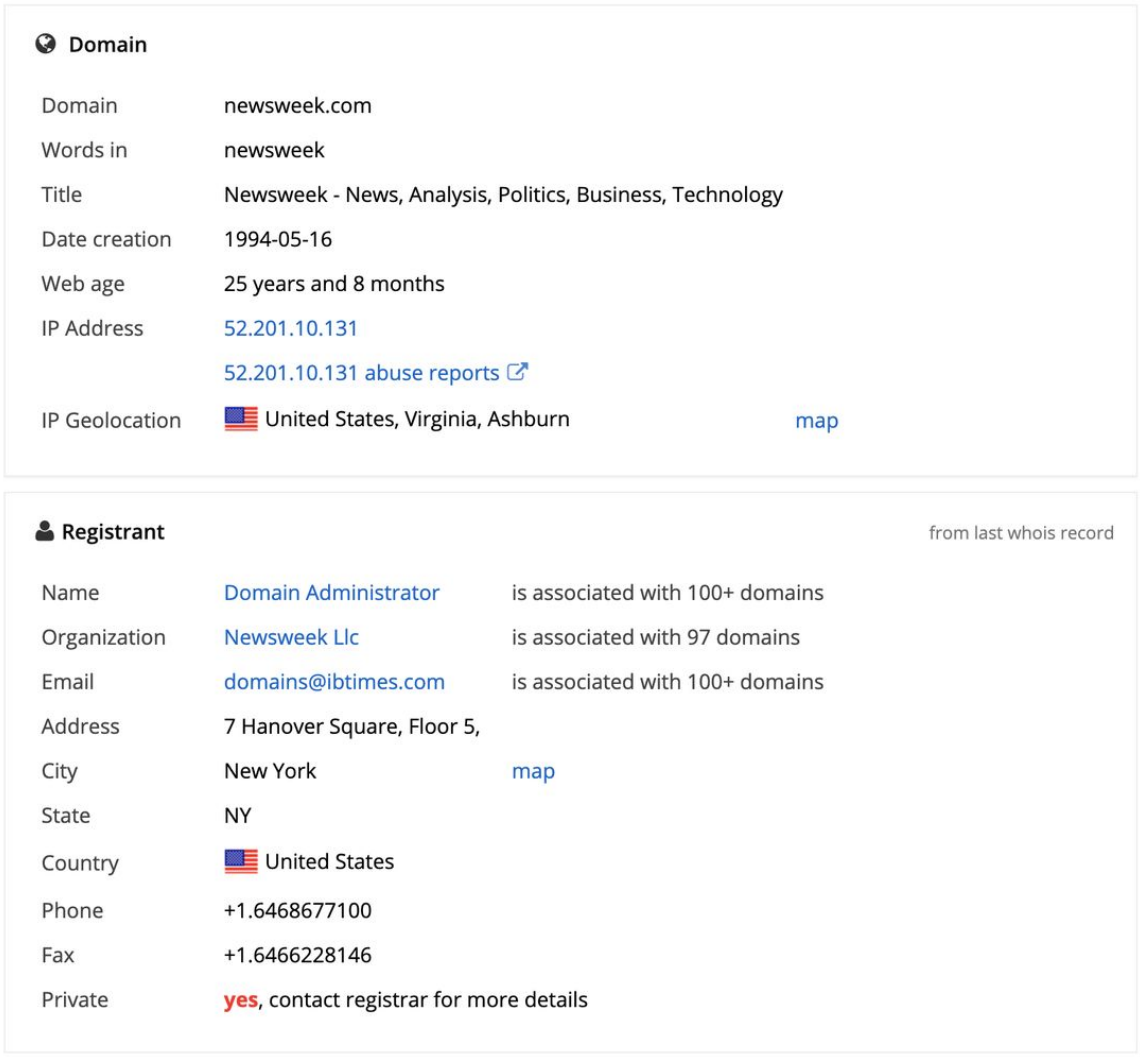
Domain

Domain	dentondaily.com
Words in	dent on daily
Title	Denton Daily
Date creation	2019-08-03
Web age	5 months
IP Address	104.27.156.76
	104.27.156.76 abuse reports
IP Geolocation	United States map

Registrant from last whois record

Name	Whoisguard Protected	is associated with 100+ domains
Organization	Whoisguard Inc	is associated with 100+ domains
Email	18460534d8af4e7bae0b7c7940deb209.protect(at)whoisguard.com	
Address	P.O. Box 0823-03411	
City	Panama	map
State	Panama	
Country	Panama	
Phone	+507.8365503	
Fax	+51.17057182	
Private	yes, contact registrar for more details	

Facciamo un altro esempio, cercando newsweek.com su DomainBigData. Notiamo subito che il proprietario non ha pagato per proteggere la sua privacy: c'è il nome della società, un indirizzo email e un numero di telefono e di fax.



Domain

Domain	newsweek.com
Words in	newsweek
Title	Newsweek - News, Analysis, Politics, Business, Technology
Date creation	1994-05-16
Web age	25 years and 8 months
IP Address	52.201.10.131
	52.201.10.131 abuse reports
IP Geolocation	United States, Virginia, Ashburn map

Registrant from last whois record

Name	Domain Administrator	is associated with 100+ domains
Organization	Newsweek Llc	is associated with 97 domains
Email	domains@ibtimes.com	is associated with 100+ domains
Address	7 Hanover Square, Floor 5,	
City	New York	map
State	NY	
Country	United States	
Phone	+1.6468677100	
Fax	+1.6466228146	
Private	yes, contact registrar for more details	

Vediamo anche che il soggetto in questione è proprietario del dominio dal maggio del 1994, e che il sito in questo momento è ospitato all'indirizzo IP 52.201.10.13. Notiamo anche che il nome della società, l'email e l'indirizzo IP sono evidenziati come link. Ciò significa che possono portarci ad altri domini appartenenti a Newsweek LLC, domains@ibtimes.com e altri siti ospitati allo stesso indirizzo IP. Queste connessioni sono incredibilmente importanti in un'indagine, per questo bisogna sempre andare a vedere di quali altri domini è proprietaria la persona o il soggetto a cui siamo risaliti.

Come per gli indirizzi IP, tieni presente che sullo stesso server possono essere ospitati siti completamente scollegati tra loro. Di solito ciò accade perché le persone ricorrono alla stessa società di hosting per i loro siti. Una regola generale è che minore è il numero di siti web ospitati sullo stesso server, maggiore è la probabilità che siano collegati. Ma non è certo.
Se vedi che un server ospita centinaia di siti, è possibile che non ci siano collegamenti tra questi a livello di proprietà del dominio. Ma se vedi che ce ne sono,

ad esempio, solo nove, e se le informazioni di registrazione di quello che ti interessa sono private, vale la pena eseguire una ricerca Whois sugli altri otto domini per verificare se questi hanno un proprietario comune e se questo è lo stesso del sito da cui sei partito. A volte, infatti, i proprietari pagano per proteggere la privacy del dominio di un sito e non per quella di altri.

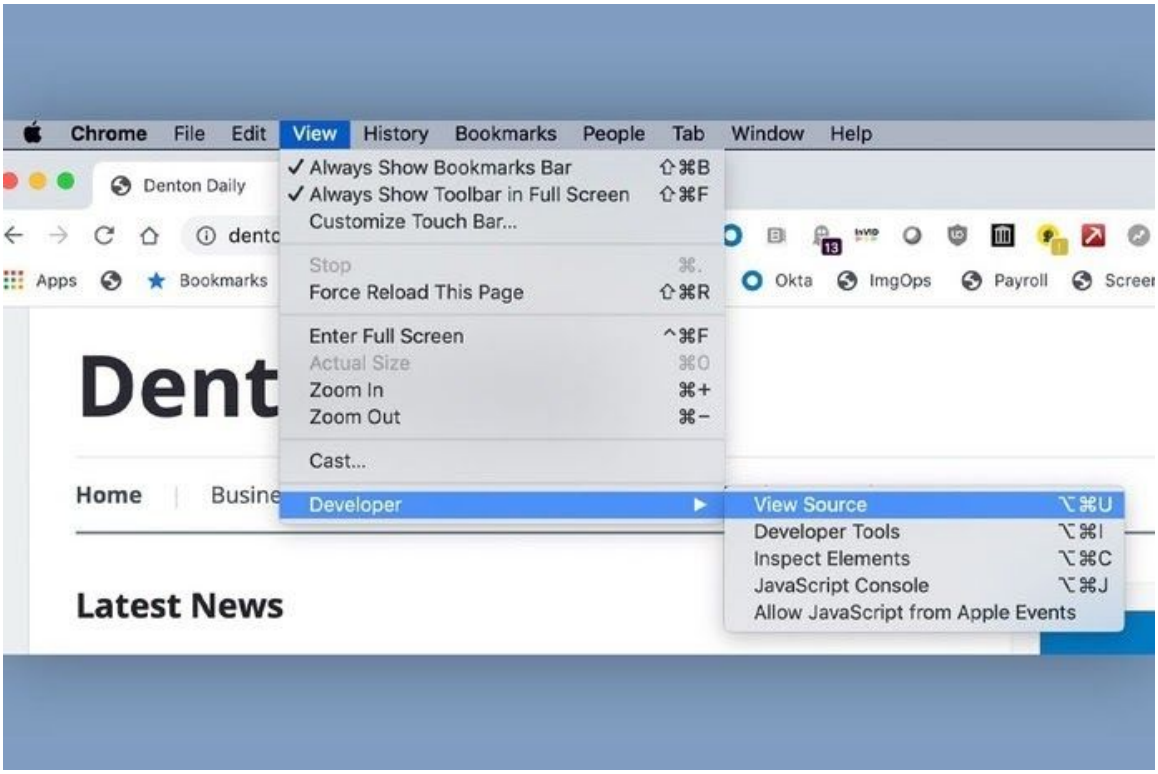
Ricostruire i collegamenti tra i siti sfruttando indirizzi IP e/o le informazioni di registrazione dei domini è fondamentale per identificare reti operative e chi ci sta dietro.

Esaminiamo ora un altro modo per risalire ai collegamenti tra i siti, ovvero andando ad analizzare il codice di una pagina web.

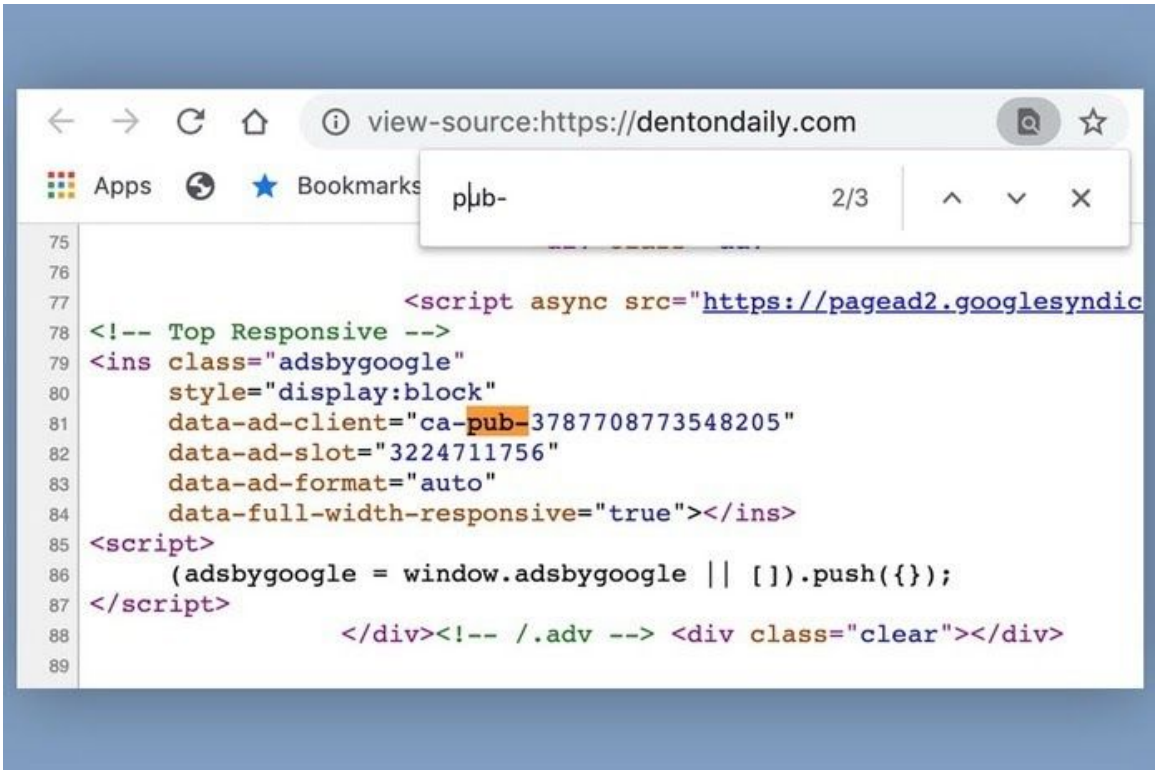
Codice e analytics

Questo metodo, definito per la prima volta da Lawrence Alexander, inizia dall'analisi del codice sorgente di una pagina web e da una ricerca al suo interno per trovarvi il codice Google Analytics e/o Google AdSense. Questi ultimi sono prodotti di Google estremamente popolari che permettono al proprietario di un sito, rispettivamente, di tenere traccia delle statistiche relative al sito e di guadagnare soldi con la pubblicità. Una volta integrati questi prodotti nel sito, a ogni pagina web viene assegnato un ID unico legato al proprietario dell'account di Analytics o AdSense. Chi gestisce più siti spesso usa lo stesso account Analytics o AdSense. Di conseguenza, se nei codici sorgente di siti apparentemente irrelati tra loro trovate lo stesso ID, vuol dire che in realtà una relazione c'è. Per fortuna, trovare l'ID nel codice sorgente è piuttosto semplice.

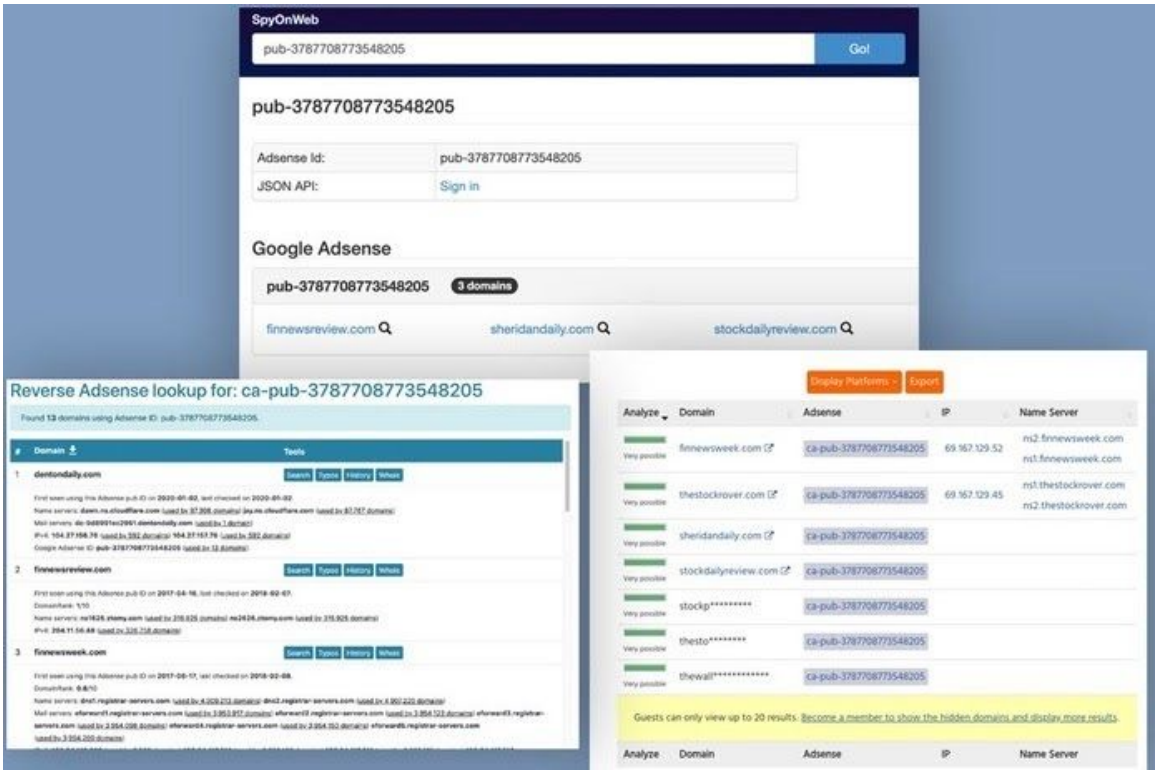
Per prima cosa, vai sul sito che vuoi analizzare. Usiamo dentondaily.com. Su Chrome per Mac, seleziona il menù "Visualizza" e poi "Opzioni per sviluppatori" e "Visualizza sorgente" (se usi Chrome in inglese, il percorso è: "View" > "Developer" > "View Source"). Si apre così una nuova tab con il codice sorgente della pagina (se usi Chrome per PC, usa la combinazione Ctrl-U).



Tutti gli ID di Google Analytics iniziano con le lettere “ua-” seguite da una stringa di numeri, quelli di AdSense con “pub-” seguito da una stringa di numeri. Puoi rintracciarli nel codice sorgente del sito tramite una semplice ricerca sulla pagina con lo strumento “Trova”. Per farlo su un Mac premi Command-F, su PC invece Ctrl-F. Ti apparirà una piccola casella di ricerca. Inserisci “ua-” o “pub-” e scoprirai così se ci sono ID all’interno della pagina.



Se trovi un ID, copialo e incollalo nella casella di ricerca di servizi come SpyOnWeb, DNSlytics, NerdyData o AnalyzeID. Tieni presente che spesso i risultati ottenuti cambiano da un servizio all'altro, per questo è importante cercare l'ID tramite più servizi e poi confrontare i risultati. Nell'immagine sottostante puoi vedere che SpyOnWeb ha trovato tre domini con lo stesso AdSense ID, ma DNSlytics e AnalyzeID ne hanno trovati molti di più.



Può capitare che un sito che in passato aveva un ID ora non ce l'abbia più. Per questo è essenziale usare il metodo descritto sul codice sorgente di tutti i siti che presumiamo condividano lo stesso ID per confermare che questo sia effettivamente presente. Ricorda anche che gli ID di AdSense e Analytics rimangono presenti nelle versioni di un sito archiviate sulla Wayback Machine. Quindi, se non trovi l'ID su un sito attualmente presente online, controlla anche sulla Wayback Machine.
I servizi che abbiamo nominato offrono alcuni risultati in forma gratuita. Spesso, tuttavia, per ottenere risultati completi occorre pagare, soprattutto se l'ID che stai cercando compare in un alto numero di altri siti.

Una nota finale sull'analisi dei codici sorgente: vale la pena scorrere l'intera pagina anche se non conosci i linguaggi HTML, JavaScript, PHP o altri comuni linguaggi di programmazione. Ad esempio, quando si utilizza lo stesso template di progettazione per più siti, spesso ci si dimentica di cambiare il titolo della pagina o del sito. A te questo semplice errore può fornire un collegamento.

Mentre indagavo sulla frode pubblicitaria che coinvolgeva società di copertura come Atoses, mi interessai a una società chiamata FLY Apps. Andai a guardare il codice sorgente del suo sito web a pagina unica e notai che accanto alla parte superiore del codice del sito era riportata una parola "normale", non scritta in codice, “Loocrum” (evidenziazione mia):

Cercando quella parola su Google trovai una società chiamata Loocrum , che nel suo sito usava esattamente lo stesso design di FLY Apps e proponeva alcuni dei suoi stessi contenuti. Una ricerca Whois rivelò che l'indirizzo email usato per registrare loocrum.com era lo stesso usato per registrare altre società di comodo che avevo già identificato nello stesso piano fraudolento. Questa connessione tra FLY Apps e Loocrum fornì un'importante, ulteriore prova che i quattro gestori di FLY Apps erano coinvolti nel piano generale. Lo scoprii semplicemente scorrendo il codice sorgente alla ricerca di parole in chiaro che sembravano fuori luogo.

Conclusione

Pur avendo nella tua cassetta degli attrezzi tutti i metodi e gli strumenti descritti qui sopra, può comunque capitarti di sentirti arrivato a un vicolo cieco. Ma spesso esiste comunque un altro modo per trovare collegamenti o vie che ti permettono di proseguire le indagini. Clicca su ogni link, studia i contenuti, leggi il codice sorgente, controlla chi appare nei crediti del sito, guarda chi lo condivide ed esamina qualsiasi altra cosa che ti venga in mente per svelare cosa sta realmente accadendo.

9. Analizzare annunci pubblicitari sui social network

Scritto da: Johanna Wild

***Johanna Wild** è investigatrice open source a Bellingcat, dove si occupa anche di sviluppo di tecnologie e strumenti per indagini digitali. Viene dal mondo del giornalismo online, e precedentemente ha lavorato con altri giornalisti in zone di conflitto e post-conflitto. Tra i ruoli che ha ricoperto c'è stato quello di supporto ai giornalisti impegnati in Africa Orientale nel produrre trasmissioni e contenuti per The Voice of America.*

Gli annunci pubblicitari che compaiono sulla tua timeline nei social network non sono gli stessi che vedono le persone sedute a fianco a te in autobus o in metro. Ad esempio, mentre tu visualizzi annunci di lussuose suite per trascorrere le vacanze a Malaga, il tuo vicino potrebbe vedere annunci di giochi giapponesi per mobile: questo in base a fattori come la località, il genere, l'età, i like che hai messo o i contenuti che hai condiviso sui tuoi social network.

Il microtargeting, ovvero la categorizzazione degli utenti in gruppi target al fine di mostrargli annunci in linea con i loro interessi e circostanze di vita, è diventato un problema centrale durante le elezioni. A suscitare preoccupazione è il fatto che le campagne possano rivolgersi a fette molto piccole di popolazione con pubblicità che provocano paura o disprezzo, o che diffondono informazioni false. Di norma gli annunci pubblicitari dei politici sui social network non sono sottoposti a fact-checking. A gennaio 2020, ad esempio, Facebook ha riconfermato che continuerà a permettere ogni pubblicità di natura politica a patto che rispetti gli standard della community di Facebook. Ciò significa che determinati gruppi di utenti potranno convertirsi dopo essere stati bersaglio di annunci disinformativi su argomenti cruciali di natura politica e sociale.

Fino a poco tempo fa per giornalisti e ricercatori era praticamente impossibile avere accesso a dati utili sulle pubblicità targhettizzate per utenti diversi. In risposta alle critiche pubbliche sulla mancanza di trasparenza, molti social network hanno creato librerie (cataloghi) di annunci pubblicitari che consentono a chiunque di vedere informazioni riguardanti le inserzioni pubblicate sulle loro piattaforme.

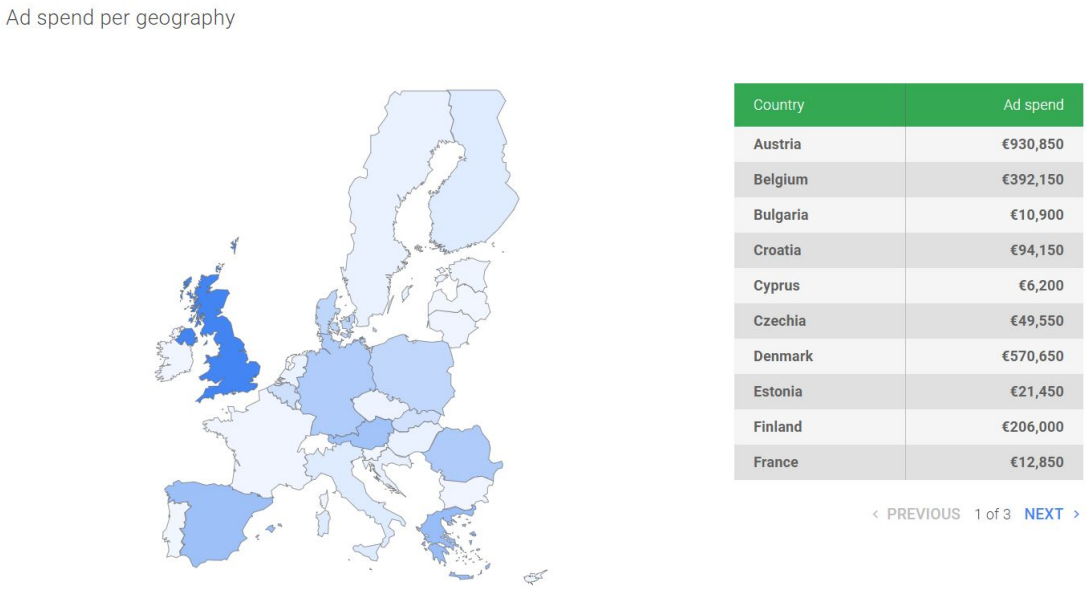
Nello specifico, la libreria delle inserzioni di Facebook è stata accusata di non mostrare in maniera affidabile tutti gli annunci disponibili. Quindi, quando consulti questi cataloghi, prenditi un po' di tempo per verificare se tutte le pubblicità che hai visto sulla tua timeline possono essere ritrovate al loro interno.

Le librerie sono nondimeno un importante passo avanti verso una maggiore trasparenza, e forniscono ai giornalisti e ad altre figure nuovi e interessanti modi di investigare sulle pubblicità digitali. Le seguenti tecniche ti aiuteranno a iniziare ad analizzare le pubblicità che si trovano sulle principali piattaforme, come Google, Twitter e Facebook.

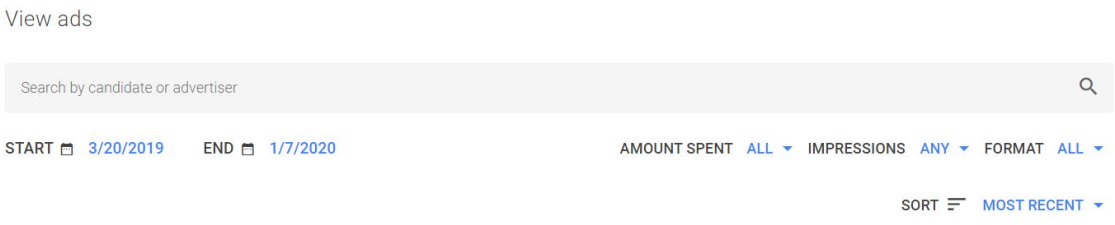
Google

Il centro annunci di Google è ben nascosto nel suo Rapporto sulla Trasparenza. Usa questo link per accedere alla sezione della pubblicità politica, che fornisce informazioni sugli annunci Google e YouTube in Unione Europea e Regno Unito, India, Nuova Zelanda e negli Stati Uniti.

La pagina di ciascuna regione mostra una lista di stati e la spesa totale in pubblicità dal momento del lancio del report.



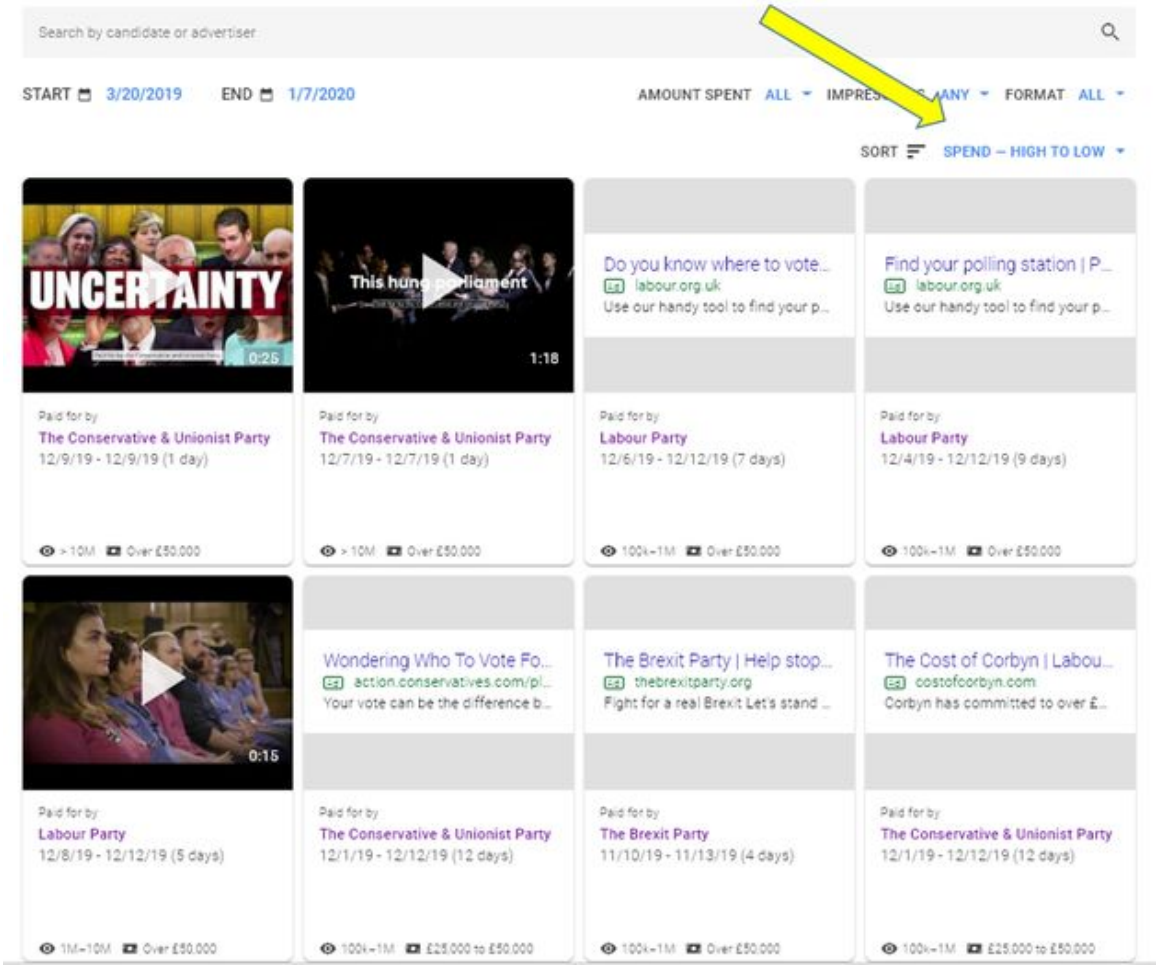
Clicca su uno stato e verrai reindirizzato ad una pagina contenente il suo database di pubblicità:



Puoi filtrare i risultati per data, somma di denaro spesa e numero di volte in cui un annuncio viene mostrato agli utenti (impressioni). Se desideri visualizzare i risultati

per annunci incentrati solo su video, immagini o testi, puoi filtrare i risultati in base al formato dell'annuncio.

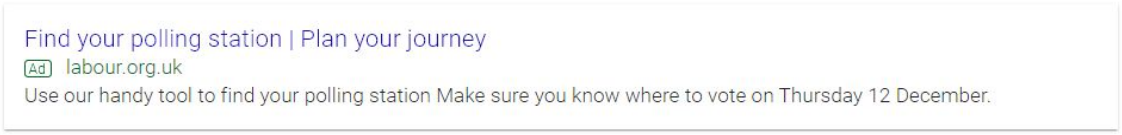
Si può anche risalire in maniera semplice a chi spende di più. Ad esempio, se vuoi visualizzare le più grandi campagne pubblicitarie politiche condotte nel Regno Unito dal lancio del rapporto a gennaio 2020, clicca su "Ordina" e seleziona "Spesa - decrescente", come mostrato di seguito (l'immagine sotto e quelle successive sono tratte dalla pagina in inglese, che rispecchia quella italiana nella collocazione delle categorie, delle sezioni e dei pulsanti).



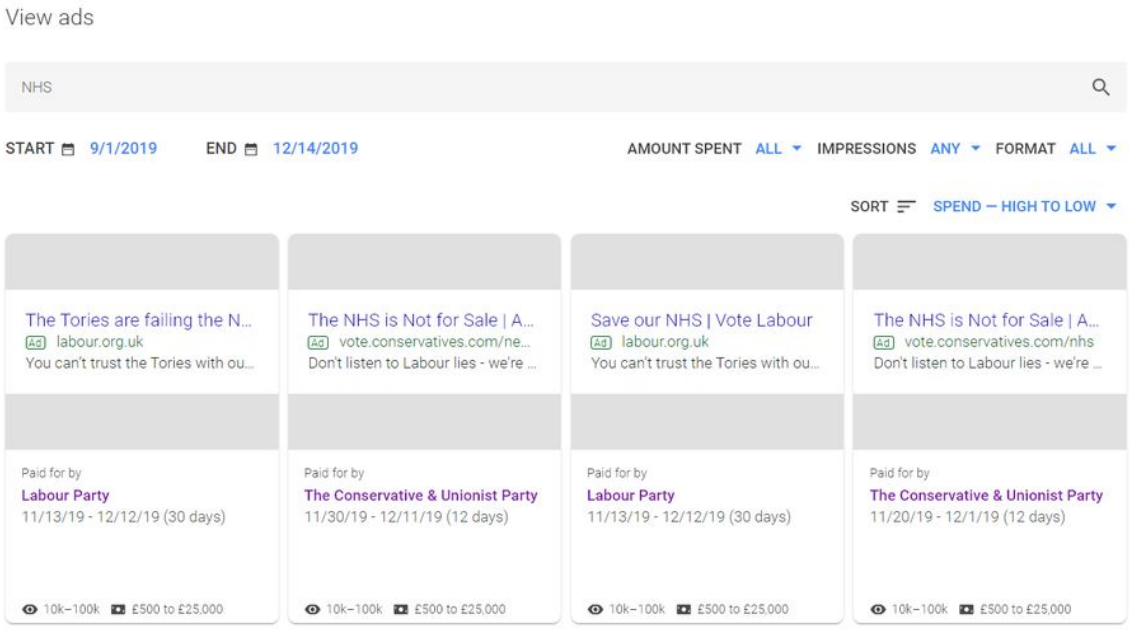
Non sorprende che i maggiori acquisti di annunci si siano verificati poco prima delle elezioni generali e nel giorno stesso di queste ultime, il 12 dicembre 2019.

Puoi vedere che il Partito Conservatore e Unionista ha investito più di 50.000 sterline su ciascuno di due annunci YouTube fatti circolare per un solo giorno.

Il Partito Laburista, invece, ha speso più di 50.000 sterline per posizionare sulle pagine dei risultati di ricerca di Google un annuncio riguardo uno strumento che, a sua detta, poteva aiutare gli elettori a trovare il proprio seggio elettorale.



Puoi cercare anche per parole chiave. Se scrivi NHS (acronimo di National Health Service, il Servizio Sanitario Nazionale) scoprirai che nel novembre e nel dicembre 2019 il Partito Laburista e quello Conservatore hanno comprato inserzioni su Google per criticarsi a vicenda i rispettivi piani per l'NHS.



Cliccando sul nome dell'inserzionista puoi anche visualizzare la somma totale che ha speso in inserzioni su Google da quando è stato lanciato il Transparency Report. Ecco i risultati per i due partiti più importanti del Regno Unito nel gennaio del 2020:

Viene fornita anche una timeline delle spese dell'inserzionista. Il report sulla sinistra mostra le tendenze di spesa per il Partito Conservatore e Unionista, mentre quello a destra riguarda il Partito Laburista:

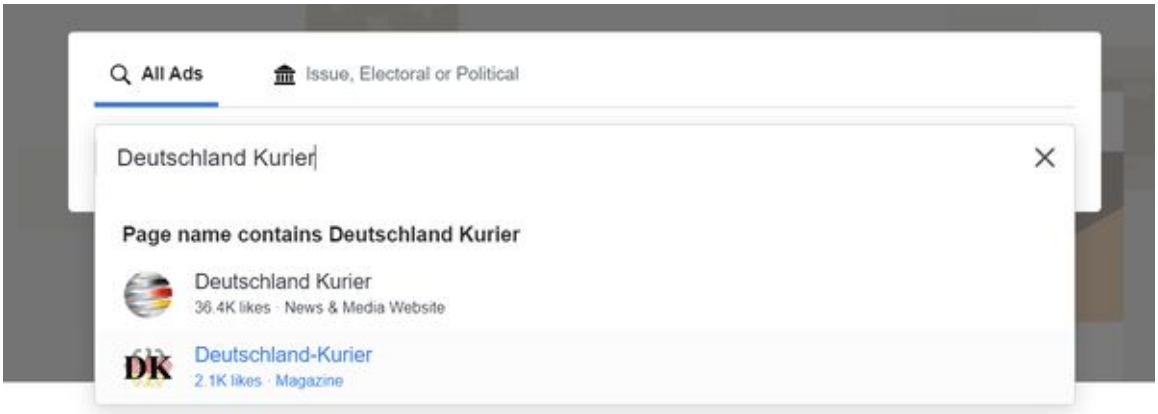
Se vuoi analizzare ulteriormente il database delle inserzioni pubblicitarie, scorri la pagina fino a trovare un bottone con scritto "Scarica dati (CSV)", che ti permette di scaricare i dati in formato CSV.

Così facendo potrai importare i dati in fogli di calcolo, usando ad esempio Google Sheets o Excel, filtrarli ulteriormente e proseguire nella tua analisi.

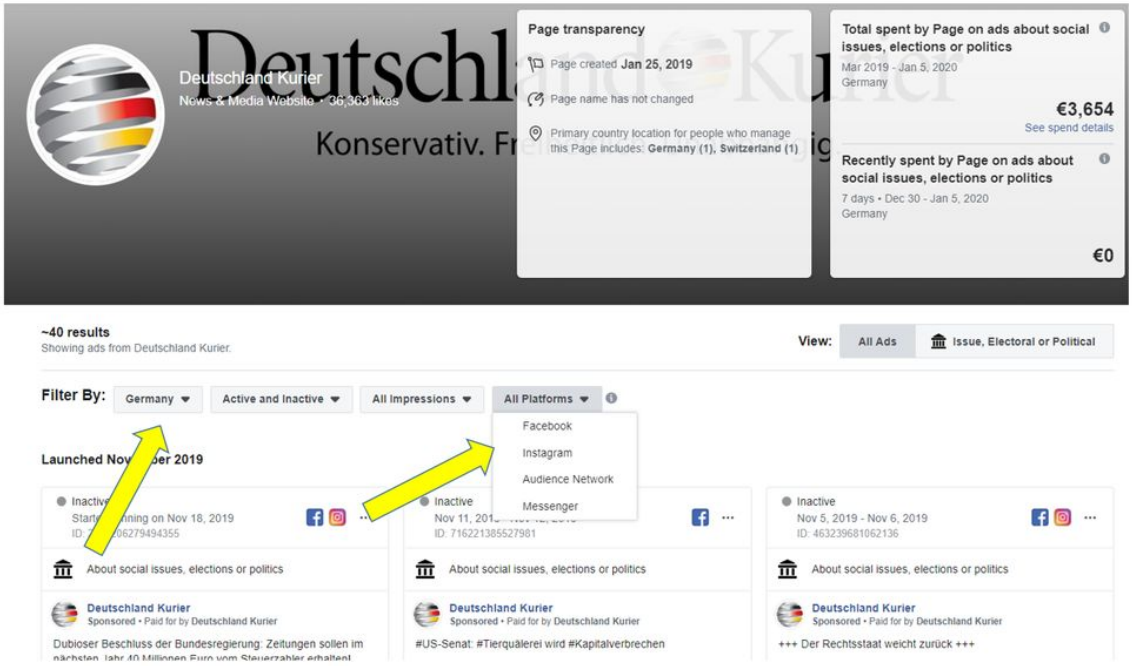
Facebook

La libreria delle inserzioni di Facebook è divisa in due parti: "Temi sociali, elezioni o politica" e "Qualsiasi". Se fai clic su "Qualsiasi" potrai cercare gli inserzionisti solo per nome, non potrai usare parole chiave.

Se ad esempio vuoi vedere le inserzioni del Deutschland Kurier, una testata che pubblica spesso contenuti a sostegno del partito di estrema destra tedesco AfD, puoi scrivere il suo nome e Facebook ti suggerirà pagine che contengono il testo digitato:

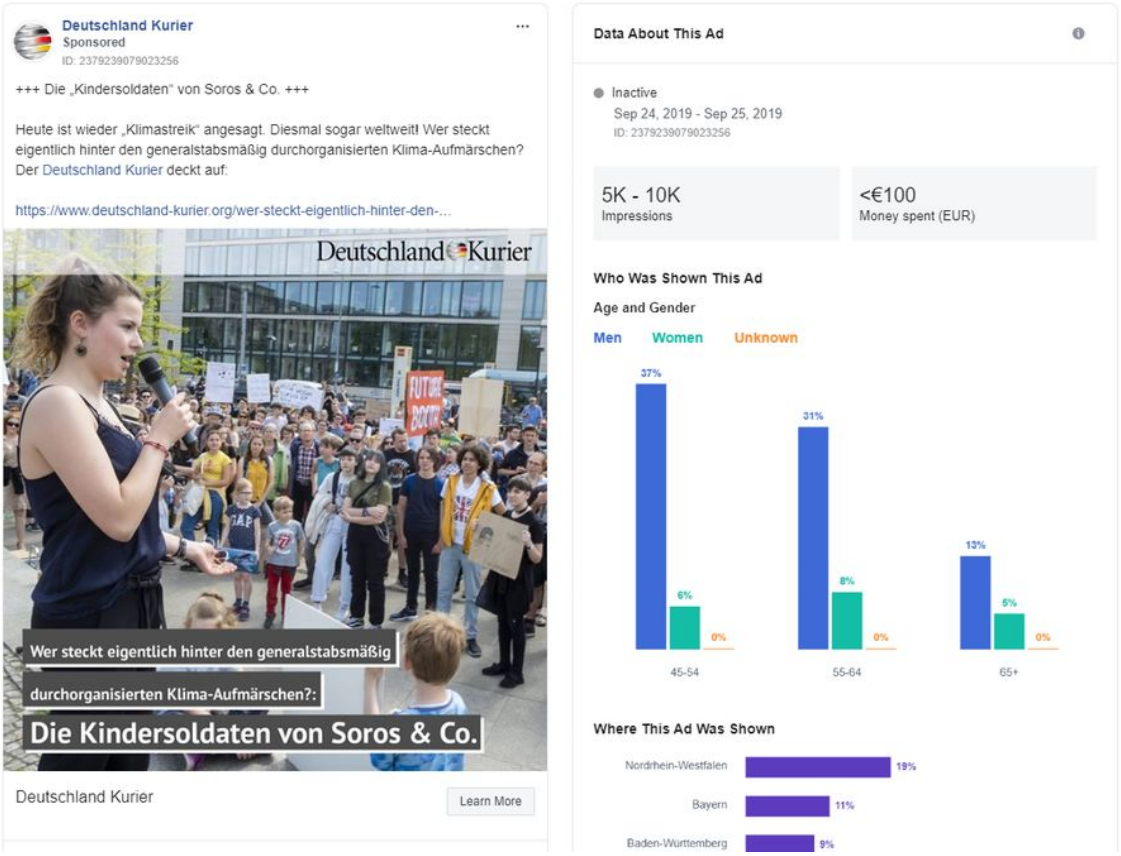


La pagina dei risultati mostra che tra il marzo del 2019 e il gennaio del 2020 il Deutschland Kurier ha speso 3.654 euro per pubblicare inserzioni in Germania.



Quando ti trovi sulla pagina dei risultati, assicurati di selezionare correttamente il Paese in cui eseguire la ricerca (o l'opzione per tutti i paesi) e di scegliere se vuoi vedere le inserzioni da Facebook, Instagram, Messenger o Facebook Audience Network. L'Audience Network è una rete pubblicitaria gestita da Facebook che piazza inserzioni su app mobile e siti web che non sono di proprietà di Facebook. Nella maggior parte dei casi la scelta migliore è eseguire la ricerca su tutte le piattaforme, così da ottenere una fotografia completa delle inserzioni di un'organizzazione.

Su ogni singola inserzione puoi cliccare sul pulsante “Vedi tutti i dettagli” per avere informazioni aggiuntive.



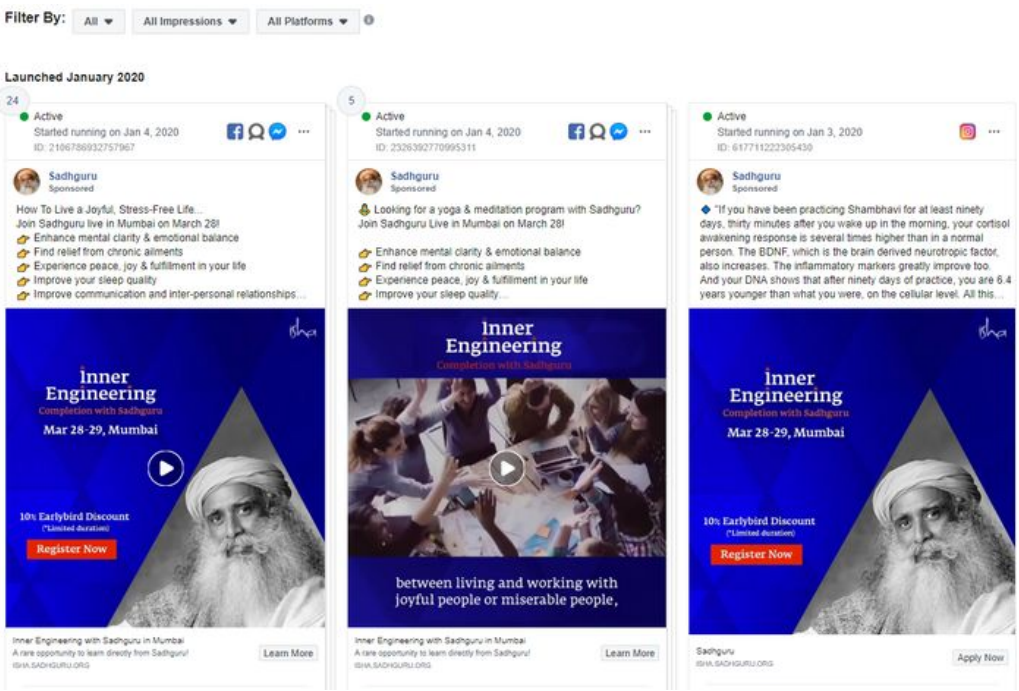
In questo esempio, il Deutschland Kurier ha speso meno di 100 € per questa inserzione che definisce i partecipanti alle manifestazioni sui cambiamenti climatici “bambini soldato di Soros & Co.”, e ha ricevuto tra le 5.000 e le 10.000 impressioni, la maggior parte da uomini dai 45 anni in su.

La seconda opzione per utilizzare la libreria delle inserzioni è scegliere il database “Temi sociali, elezioni o politica”, un archivio di inserzioni su temi di natura politica, elettorale o sociale. I due grandi vantaggi di questa opzione sono che puoi cercare usando qualsiasi parola chiave, e che questi tipi di inserzioni sono archiviati da Facebook.

Guardiamo un esempio.

Sadhguru è il nome di una famoso guru spirituale indiano che dichiara di non essere legato a nessun partito politico. Ha detto che considera un suo dovere supportare qualsiasi governo in carica nel “fare del proprio meglio”. Se inserisci il suo nome nella sezione “Qualsiasi” della libreria, Facebook ti mostrerà la pagina Facebook personale di Sadhguru.

I risultati ci propongono una selezione di inserzioni apolitiche pubblicate da Sadhguru per promuovere i suoi corsi di meditazione e yoga.



Ora proviamo a scrivere il suo nome nella barra di ricerca della sezione “Temi sociali, elezioni o politica”, senza accettare le pagine Facebook suggerite che appaiono:

Il risultato della ricerca cambia radicalmente. Ora siamo davanti a una serie di inserzioni pubblicate da altri account che menzionano il nome di Sadhguru.

Una pubblicità pubblicata dal BJP, il partito nazionalista indiano attualmente al governo, mostra un video in cui Sadhguru esprime il suo supporto alla controversa legge sulla cittadinanza. Il disegno di legge consente agli immigrati non registrati provenienti da alcuni paesi confinanti con l'India di ottenere più facilmente la cittadinanza indiana, ma non concede la stessa opportunità ai musulmani. L'inserzione suggerisce una possibile relazione tra Sadhguru e il BJP, argomento molto discusso in India.

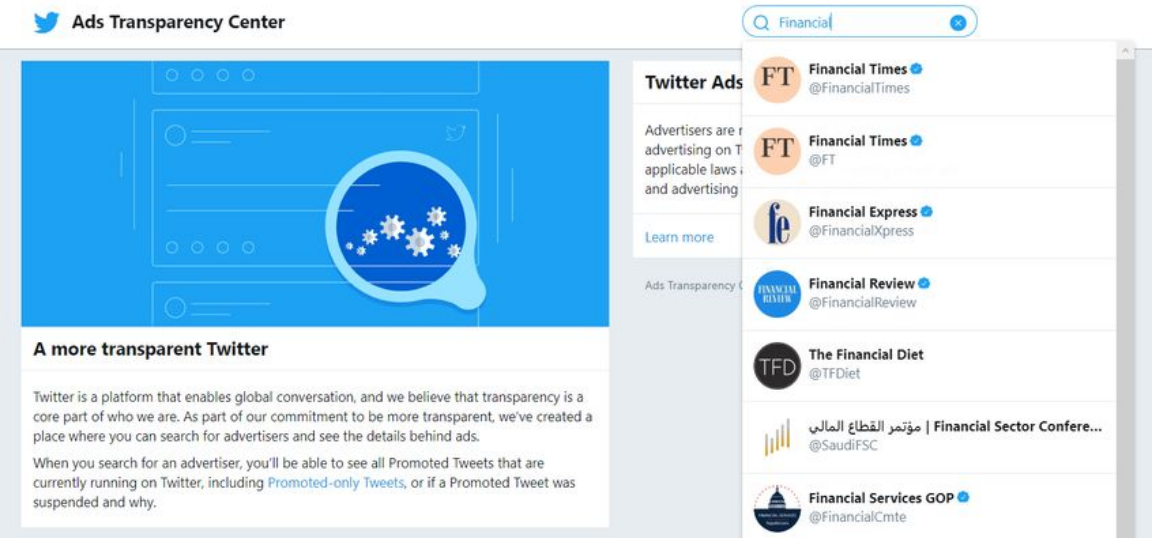
Questo esempio mostra come si può usare la libreria delle inserzioni di Facebook per ottenere informazioni chiave per le proprie indagini. Potresti trovare utile anche

dare un'occhiata al report della Libreria inserzioni di Facebook, dove sono raccolte informazioni chiave ricavate dalle inserzioni politiche di paesi diversi.

Twitter

Alla fine del 2019 Twitter decise di proibire la pubblicità politica sulla sua piattaforma. Ad ogni modo è ancora possibile utilizzare il centro trasparenza sugli annunci del social network per ottenere informazioni riguardo le inserzioni non politiche degli ultimi sette giorni.

Trovare annunci tramite questo strumento è piuttosto scomodo, perché non esiste la funzione di ricerca per parole chiave. Per avviare una ricerca, vai alla casella nell'angolo in alto a destra della pagina e digita il nome utente o lo pseudonimo che ti interessa.



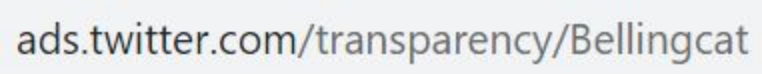
Se negli ultimi sette giorni sono stati pubblicati annunci, verranno visualizzati in una lista.

Se facciamo una ricerca sul Financial Times, scopriamo che ha pagato per generare più interesse attorno alla sua storia intitolata “How native speakers can stop confusing everyone else” ("Come le persone di madrelingua inglese possono smettere di mandare in confusione gli altri"). Il tweet fu pubblicato il 3 dicembre del 2019, ma le informazioni fornite da Twitter non ci dicono quando esattamente la promozione è stata attiva.

Per velocizzare le tue indagini, puoi usare questo piccolo trucco. Dopo aver eseguito una prima ricerca, guarda la URL nel tuo browser:

ads.twitter.com/transparency/FinancialTimes

La URL ha sempre la stessa struttura, che termina con uno pseudonimo di Twitter. Cancella solo questa ultima parte e sostituiscila con un altro pseudonimo:



Ricarica la pagina: ora puoi visualizzare i risultati riguardanti Bellingcat. Se negli ultimi sette giorni l'account non ha pubblicato pubblicità, comparirà il messaggio: "Questo account non ha sponsorizzato annunci negli ultimi sette giorni". Dal momento che si possono vedere solo gli annunci degli ultimi sette giorni, fai dei controlli frequenti per vedere se account per te rilevanti hanno pubblicato degli annunci pubblicitari. Fai degli screenshot ogni volta che vedi nuovi annunci.

Snapchat

La "Snap political ads library" offre informazioni approfondite su pubblicità politiche o relative a determinate questioni rilevanti o di attivismo. Gli annunci dell'ultima categoria sono definiti come "annunci riguardanti questioni od organizzazioni che sono oggetto di dibattito a livello locale, nazionale o globale o che sono di rilevanza pubblica". Argomenti come, ad esempio, l'immigrazione, l'istruzione o le armi.

Andando alla libreria vedrai una lista divisa per anni.

Clicca su uno degli anni e potrai scaricare un foglio di calcolo con tutte le informazioni disponibili riguardo le pubblicità di quell'anno. Il contenuto del foglio di calcolo non sembra particolarmente interessante a una prima occhiata, ma in realtà lo è! Ogni linea rappresenta un annuncio e mostra chi lo ha pubblicato, quanto denaro ci ha investito, e persino che caratteristiche ha scelto per micro-targhettizzare gli utenti.

Nell’esempio riportato qui sotto, l’investitore voleva puntare su “Amanti dell’avventura, esperti di arte e cultura, bagnanti e surfisti, esperti di bellezza, topi di biblioteca e lettori accaniti, studenti universitari, appassionati di cibo, hipster e trendsetter, osservatori di notizie politiche, appassionati di outdoor e natura, amanti di animali e animali domestici, filantropi, viaggiatori e lifestyle femminile”.

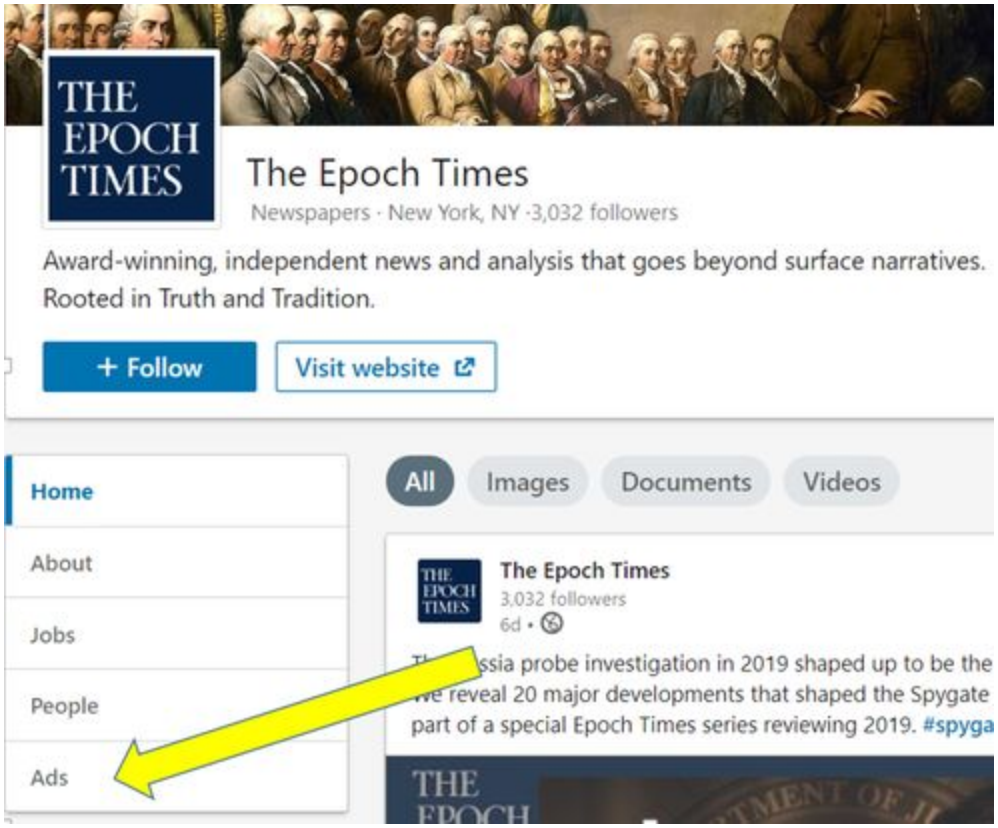
Le librerie di altre piattaforme non forniscono questo genere di informazioni sui target degli annunci.

Nel file trovi anche una URL per visualizzare l'annuncio vero e proprio. In questo esempio, il messaggio dell'annuncio incoraggiava le persone a ordinare bandiere arcobaleno gratuite per sostenere l'imminente voto in Svizzera sulla difesa dalle discriminazioni delle persone LGBT.

LinkedIn

Linkedin non consente pubblicità di matrice politica sulla sua piattaforma, e non possiede una libreria di annunci. Fortunatamente, c’è un altro modo per ottenere informazioni riguardo l’advertising di una compagnia specifica all’interno della piattaforma.

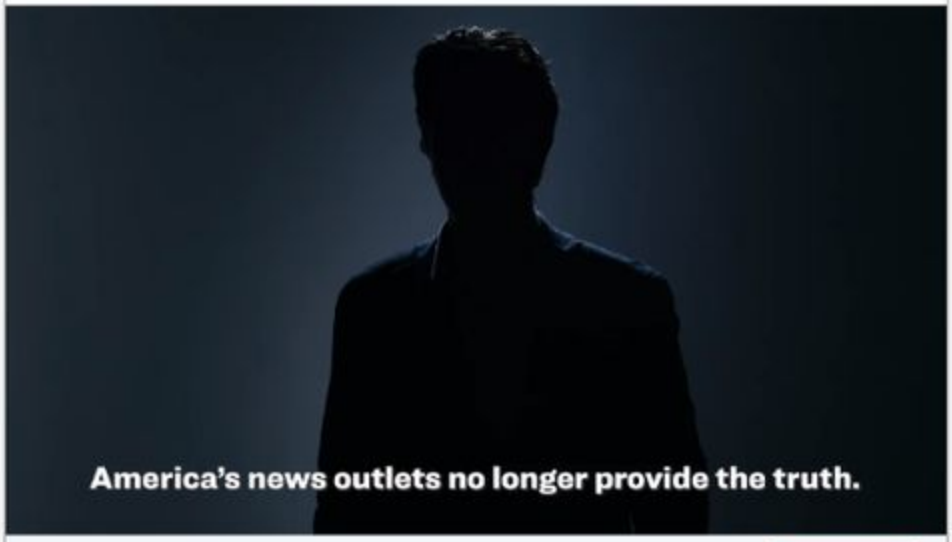
Se vai sulla pagina LinkedIn dell'azienda che ti interessa, in fondo alla colonna a sinistra vedrai una sezione intitolata “Ads”.



Clicca su questa sezione e LinkedIn ti mostrerà una lista di tutti gli annunci pubblicati dall'azienda negli ultimi sei mesi. Questa funzione ha permesso ad esempio di scoprire che l'Epoch Times continuava a pubblicare annunci su LinkedIn anche dopo che gli era stato vietato di farlo su Facebook. Le due pubblicità sponsorizzate dall'azienda sostenevano che "Gli organi di stampa americani non dicono più la verità", presentando di contro l'Epoch Times come "media indipendente" e "non di parte".

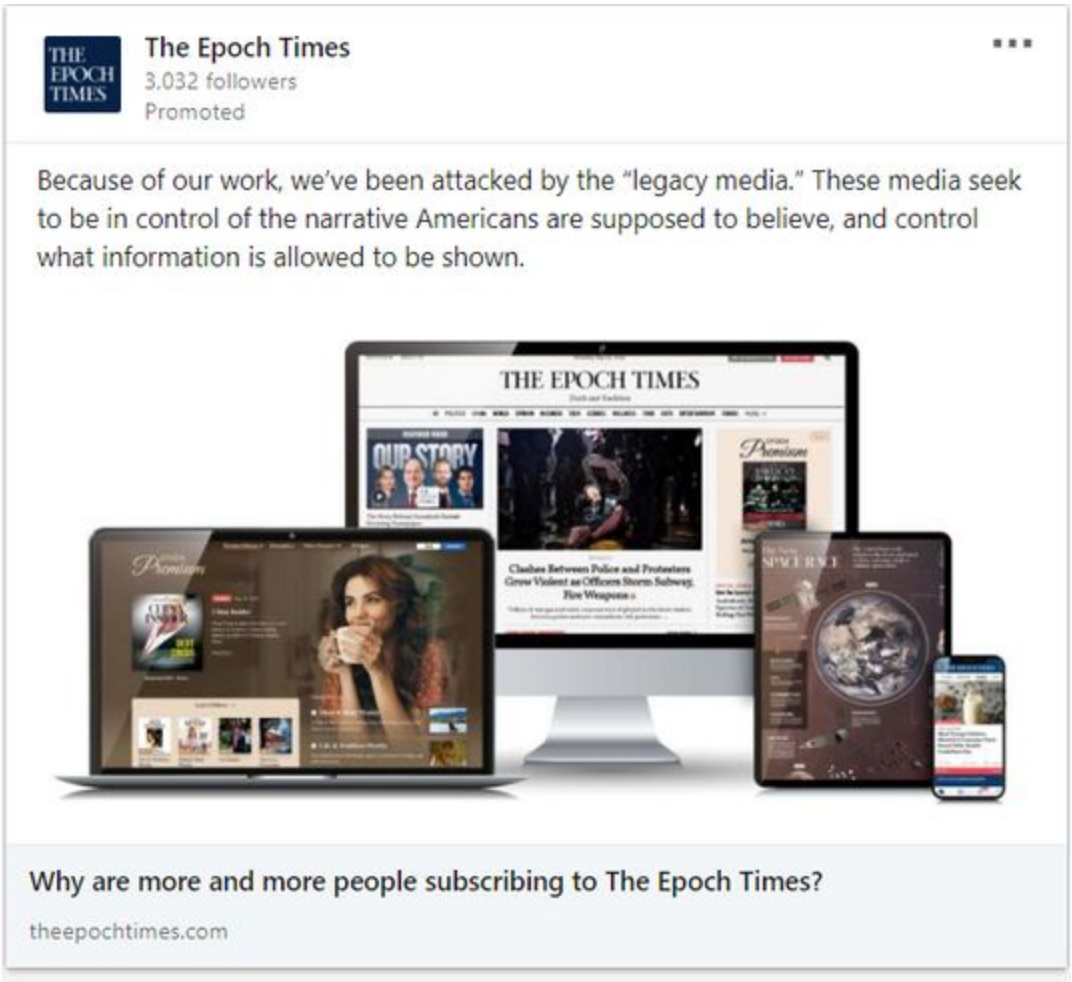
The Epoch Times
3,032 followers
Promoted

90% of news outlets in the US are controlled by 6 corporations. Where can you find real news without false narratives?

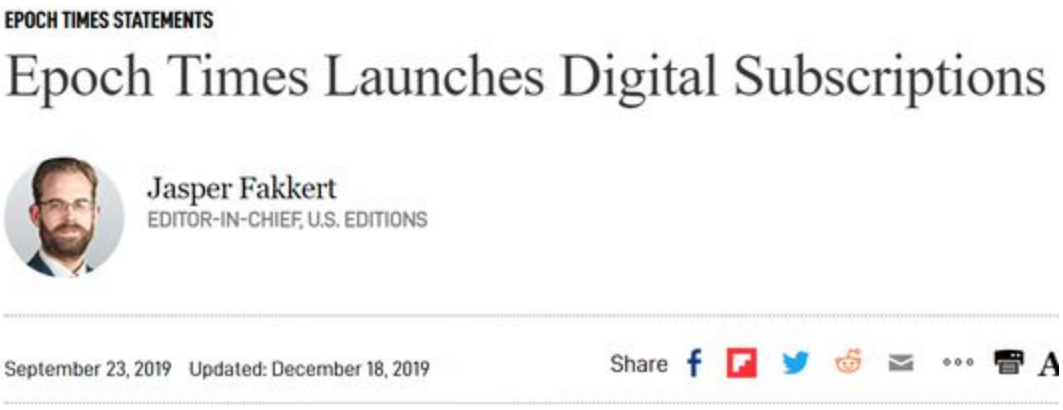


Get Real News + Your Free Poster

Subscribe



Non vengono fornite le date esatte di pubblicazione, ma in alcuni casi se si clicca sull'annuncio (lo si può fare anche se l'annuncio non è più attivo su LinkedIn) il sito a cui si approda può fornire dati più concreti. Ad esempio, il primo annuncio dell'Epoch Times portava a un testo datato "23 settembre 2019" e "aggiornato: 18 dicembre 2019". Queste informazioni ci aiutano a stimare quando l'annuncio è andato online.


EPOCH TIMES STATEMENTS

Epoch Times Launches Digital Subscriptions

Jasper Fakkert
EDITOR-IN-CHIEF, U.S. EDITIONS

September 23, 2019 Updated: December 18, 2019

Share

Se conosci le loro funzionalità nascoste, le librerie di annunci pubblicitari sono un ulteriore strumento, potente e semplice da usare, da aggiungere al tuo arsenale per le indagini digitali; nonché un'importante risorsa da controllare quando si indaga su una persona o su un'azienda presente sui social media.

10. Monitorare soggetti attraverso più piattaforme

Scritto da Ben Collins

Ben Collins *lavora come giornalista alla NBC, occupandosi di disinformazione, estremismi e internet. Negli ultimi cinque anni si è dedicato all'aumento delle teorie del complotto, alle comunità di hater, alle campagne di manipolazione estere e alle falle delle piattaforme. In precedenza ha lavorato per The Daily Beast, dove con il suo gruppo di lavoro ha scoperto account, gruppi ed eventi offline creati dalla fabbrica di troll della russa Internet Research Agency durante le elezioni del 2016.*

Il 3 agosto del 2019 Patrick Crusius entrò nel Walmart di El Paso e uccise 22 persone. La sparatoria era motivata da ragioni di nazionalismo bianco. Prima di entrare nel negozio, Crusius aveva pubblicato un manifesto nella sezione di discussione politica /pol/ su 8chan.net, una piattaforma di messaggi anonimi che negli ultimi anni è diventata luogo di aggregazione di nazionalisti bianchi. Le sezioni /pol/ su 4chan e 8chan sono quasi del tutto prive di moderazione, e nell'estate del 2019 8chan era diventato uno spazio che raccoglieva contenuti e discussioni riconducibili al nazionalismo bianco violento.

In parte per questo, gli utenti di 8chan a volte segnalavano la pubblicazione di un nuovo manifesto violento. Le segnalazioni venivano fatta tramite commenti sotto al manifesto stesso, oppure inviando segnalazioni online ai media o alle forze dell'ordine. La prima volta che il tiratore di El Paso pubblicò il proprio manifesto (inizialmente con l'allegato sbagliato), un utente commentò sotto al post "Hello FBI". Il manifesto corretto fu in seguito pubblicato proprio sotto questo commento rivolto all'FBI.

Nei primi momenti in cui si seguono tragedie come quella di El Paso, auto-denunce di questo genere possono fornire ai giornalisti informazioni essenziali. In alcuni casi, prima delle sparatorie i manifesti o post sospetti vengono segnalati su piattaforme web più aperte, popolari e civili come Reddit o Twitter da utenti più volenterosi di altri. Dato che perdersi un post o un commento importanti su 4chan o 8chan è molto facile, queste iniziative sono fondamentali.

Piattaforme anonime come 4chan o 8chan giocano un ruolo importante nella diffusione online di disinformazione e di ecosistemi di troll, perché spesso è proprio in questi spazi che le persone si organizzano per lanciare e coordinare le campagne. Reddit, altro popolare spazio web dove gli utenti sono in gran parte anonimi, ospita una grande varietà di community. Alcune di queste si raccolgono in subreddit (ovvero sottocanali) rigorosamente moderati dove gli utenti possono confrontarsi sui propri hobby o discutere riguardo notizie o eventi; altre invece sono spazi in cui vige il principio del "liberi tutti" e in cui l'odio può diffondersi senza freni. È

essenziale che i giornalisti sappiano monitorare e raccontare queste community, e che conoscano le complessità del loro funzionamento.

Tenendo presente tutto ciò, ecco cinque regole da rispettare quando, per raccogliere informazioni sulla tua storia, si rende necessario ricorrere a 4chan o 8chan (o la sua nuova versione, 8kun):

1. Non fidarti di niente di quello che trovi su 4chan/8chan.
2. Non fidarti di niente di quello che trovi su 4chan/8chan.
3. Non fidarti di niente di quello che trovi su 4chan/8chan.
4. Può darsi che su 4chan/8chan si possa trovare qualche informazione utile (o persino qualche prova) riguardo un crimine o una campagna di troll o di disinformazione.
5. Non fidarti di niente di quello che trovi su 4chan/8chan.

Non sottolineerò mai abbastanza quanto sia importante per i giornalisti seguire le regole 1, 2, 3 e 5, anche se dovesse impedire loro di scovare materiale succulento che potrebbe aiutarli al punto 4. Questi siti web sono costruiti apposta per trollare, diffondere insinuazioni e falsità sui nemici percepiti e pompare bugie su persone emarginate. Di tanto in tanto, pubblicano anche qualche contenuto semi-divertente su cosa significhi essere un adolescente, montato come se fosse una storia vera.

A dimostrazione di quanto detto c'è il fatto che questi spazi sono stati usati da nazionalisti bianchi, incel (abbreviazione di involuntary celibate, "celibe non per propria volontà": secondo la definizione dell'Oxford Dictionary sono uomini giovani che si reputano incapaci di risultare sessualmente attraenti agli occhi delle donne, e che mostrano visioni ostili nei confronti di queste e degli uomini sessualmente attivi) e giovani attentatori bianchi o disturbati di qualche altro tipo, come discarica dove scaricare i loro manifesti.

Ripetiamolo ancora una volta: se stai guardando un contenuto su 4chan o 8chan (che da qui in poi continueremo a chiamare 8chan, anche se il suo nome è cambiato in 8kun) c’è un'altissima probabilità che sia una menzogna volta a seminare il caos e a seccare i giornalisti. Non entrare in una conversazione per chiedere maggiori dettagli. Anzi, non pubblicare proprio nulla. Altrimenti verrai preso di mira da gente che ha fin troppo tempo a disposizione.

Verificare il manifesto

In virtù di tutto ciò, è molto utile che membri di queste community facciano uno sforzo per denunciare pubblicamente manifesti o altri contenuti di rilievo per i giornalisti. Il commento “Hello FBI” su 8chan è ciò che mi ha permesso di scoprire l’esistenza del manifesto di El Paso. Poco dopo aver pubblicato la notizia della sparatoria, ho cercato su Twitter le parole chiave “El Paso 4chan” e “El Paso 8chan.” Quando si verificano eventi simili è utile fare una ricerca impostando la query con questa formula: [nome della città] + [8chan, 4chan, incels.co o altri siti estremisti].

Grazie alla ricerca su Twitter scoprii che alcuni utenti avevano condiviso su 8chan gli screenshot dei post pubblicati dal colpevole della sparatoria, sebbene molti avessero erroneamente attribuito il post a un utente di 4chan. Dovevo quindi trovare quel post.

Qual è il metodo più veloce per cercare un post di 8chan? Google. Subito dopo l'attentato cercai su Google digitando "site:8ch.net", aggiungendo una parte della frase presa dal presunto post dell'attentatore su 8chan (una nota: 4chan cancella automaticamente i post dai suoi server dopo un certo periodo di tempo, ma esistono siti di archivio automatico di 4chan. Il più completo si chiama 4plebs.org. I post archiviati di 4chan possono essere trovati semplicemente rimpiazzando nella URL 4chan con 4plebs e rimuovendo il prefisso "boards". Per esempio: boards.4chan.org/pol/13561062.html potrebbe essere trovato alla URL 4plebs.org/pol/13561062.html).

Durante alcune sparatorie può rivelarsi utile provare a cercare "site:4chan.net + 'manifesto' o 'fbi'" e usare le opzioni di ricerca di Google per restringere la ricerca alle ultime 24 ore. Gli utenti di Chan potrebbero aver già tentato di denunciare l'attentatore commentando il suo post.

La mia strategia di ricerca iniziale non fece emergere nessun post rilevante su 8chan, il che mi portò a credere di trovarmi davanti a una bufala montata in fretta. Ma qualcosa non tornava. Il post mostrato nello screenshot su Twitter mostrava, di fatto, uno user ID e un numero di post. Questi dettagli mi portarono a pensare che fosse un post vero e non un fake. Su 8chan ogni post è infatti associato a uno user ID unico, generato algoritmicamente, che viene mostrato accanto alla data di pubblicazione. Questo sistema permette agli utenti di avere degli ID statici con cui identificarsi all'interno di una conversazione.

Questo sistema di user ID, tra le altre cose, è ciò che permette alle persone di riconoscere i post dell'utente "Q", la figura chiave della teoria complottista QAnon. Di fatto, gli utenti possono creare nomi utente e password permanenti inserendo durante la creazione di un post un nome utente nel campo ID, seguito da un # e da una password.

Tramite lo user ID scoprii che l'utente che aveva pubblicato per sbaglio il PDF con il nome del responsabile della sparatoria era lo stesso che due minuti dopo aveva pubblicato il manifesto vero. Entrambi i post avevano il medesimo user ID, generato in modo casuale: 58820b.

Accanto allo user ID c'è un numero relativo al post, una sorta di attributo permanente che crea una URL unica per ogni post. Nello screenshot condiviso su Twitter del manifesto di El Paso era incluso anche il post ID: 13561062. Con questo numero ho ricreato la URL 8ch.net/pol/res/13561062.html. Puoi usare questa regola di composizione della URL sia su 4chan che su 8chan.

In questo caso, tuttavia, il post non esisteva. Pensai che forse era stato cancellato (in seguito scoprii che il proprietario di 8chan, Jim Watkins, lo aveva rimosso dopo essere stato allertato sul suo contenuto).

Con il post sparito, la mia ultima e maggiore speranza era che qualcuno ne avesse riconosciuto l'importanza e lo avesse archiviato. Fortunatamente, un utente sveglio di 8chan aveva salvato il post sul sito di archiviazione archive.is. Incollando la URL nella casella di ricerca di archive.is, che si chiama “I want to search the archive for saved snapshots” (nella versione italiana il campo di ricerca si chiama semplicemente "Cercare nell'archivio"), scoprii che il post del manifesto era reale e ora potevo vederlo.

Ma c’era un altro problema: quando era stato pubblicato per la prima volta su 8chan? Avevo bisogno di una marca temporale (timestamp) affidabile per confermare che il manifesto era stato pubblicato prima che l’attentatore di El Paso scatenasse la sua furia.

Sia 4chan che 8chan localizzano i timestamp, rendendo difficile ricavare l'orario effettivo corrispondente dai siti di archiviazione. Per fortuna, c’è una metodo infallibile per aggirare il problema: cliccando con il tasto destro sul timestamp e selezionando “Ispeziona”, viene mostrato il codice sorgente del sito e viene evidenziata una sezione che inizia con “‘<time unixtime=‘[number].’”

Copiando e incollando questo numero in un convertitore di timestamp Epoch/Unix, ad esempio unixtimestamp.com, si ottiene una marca temporale precisa al secondo in orario UTC. Così facendo e convertendo l'orario a cui ero risalito al fuso orario di El Paso scoprii che il manifesto era stato pubblicato alle 10.15 del mattino del fuso orario CT (Central Time), ovvero pochi minuti prima dell'inizio della sparatoria.

Questo lavoro mi aiutò a confermare che il manifesto pubblicato su 8chan era, in effetti, una prova legittima in un caso di terrorismo interno a sfondo razzista.

Monitorare soggetti attraverso più piattaforme

Nel 2017, Lane Davis, ex "Gamergate researcher" (dicitura da leggersi "stalker di internet professionista": il Gamergate è stata una campagna virale a sfondo sessista e anti-progressista nata e sviluppatasi nel mondo dei videogiochi) per il personaggio di estrema destra ormai caduto in disgrazia Milo Yiannopoulos, uccise suo padre nella propria abitazione.

Davis aveva avuto una discussione con i suoi genitori. Una chiamata al 911 attesta che poco prima dell’omicidio si fosse messo a urlare cose usando il tipico linguaggio degli estremisti di destra su Internet. Prima che suo padre chiamasse la polizia per farsi aiutare a cacciare Davis dalla loro casa, dove il figlio viveva ancora, questi apostrofava i suoi genitori come "pedofili di sinistra".

Davis era conosciuto online come “Seattle4Truth” e nei suoi video su YouTube parlava spesso di immaginari circoli segreti di pedofili che secondo lui erano le forze che stavano dietro al liberalismo. Su YouTube c'era un video a suo nome intitolato “Progressive ideology’s deep ties to pedophilia” (“I profondi legami dell’ideologia progressista con la pedofilia”).

Per un giornalista, lo scenario migliore che può presentarsi quando indaga sull’estremismo online è quello in cui l'autore di un crimine usa uno stesso username statico in tutte le piattaforme. Fu questo il caso di Davis, che si identificava come Seattle4Truth su YouTube e su Reddit, dove i suoi post rivelavano una mente ancora più imbottita di teorie cospirazioniste.

Come fu possibile scoprirlo? Semplicemente inserendo seattle4truth nella URL convenzionale dei nomi utente di Reddit: reddit.com/u/[username]. Dopo averlo fatto, puoi ordinare per post più recenti, più popolari e per quelli più controversi, sfruttando un filtro che ordina i risultati secondo un rapporto tra quante volte i post sono stati apprezzati e quante disprezzati.

Un modo veloce di fare una ricerca su un nome utente è usare Namechk, che ricerca un nome utente su quasi 100 servizi Internet. Come spiego più nel dettaglio in seguito, ciò non significa necessariamente che tutti gli account sono gestiti dalla stessa persona, ma è un modo efficace per vedere dove viene usato il nome utente, e quindi dove andare a scavare più a fondo con le ricerche. Puoi anche cercare su Google tutti i nome utente che ti interessano.

È importante anche conoscere le community internet estremamente di nicchia in cui il tuo sospetto potrebbe essere attivo. Il responsabile di una sparatoria in una scuola del New Mexico nel 2017, William Edward Atchison, venne identificato dagli utenti di KiwiFarms, un sito dedito soprattutto al bullismo anti-trans, come l'utente con lo username @satanicdruggie. Gli utenti dissero che era attivo su Encyclopedia Dramatica, un sito di meme senza alcuna regola che talvolta ospita retorica estremista.

Non soltanto Atchison era attivo su Encyclopedia Dramatica, ma era anche un SysOp (system operator), vale a dire un amministratore e utente con poteri particolari (abbiamo confermato con utenti del sito che hanno sviluppato relazioni non solo virtuali con Atchinson, soprattutto su Skype, che gli account erano suoi. Se bannato, Atchison indirizzava volontariamente gli utenti ad altri suoi account). Una ricerca su Google sul suo nome utente usando la stringa “site:encyclopediadramatica.rs + [username]” ha fatto emergere che si faceva chiamare Satanic Druggie (drogato satanico), ma anche "Future School Shooter” ("futuro tiratore in una scuola") e “Adam Lanza,” il nome del responsabile della sparatoria di Sandy Hook.

La storia delle cose da lui pubblicate sul web metteva in luce un'ossessione per le sparatorie nelle scuole che nemmeno la polizia aveva scoperto nelle ore successive alla sparatoria.

Ancora una volta, è importante sottolineare che il fatto che un nome utente appaia su più piattaforme non significa necessariamente che gli account siano stati tutti creati dalla stessa persona. Un caso esemplare a questo proposito si ebbe quando i famigerati diffusori di disinformazione di estrema destra Ian Miles Cheong, Mike Cernovich, InfoWars e GatewayPundit affermarono che un uomo che aveva ucciso due persone e feritene altre 10 a un torneo di videogiochi a Jacksonville era anti-Trump.

Le loro ragioni? L'autore della sparatoria, David Katz, partecipava a tornei di videogiochi online con il nome utente “Ravens2012Champs”, nome simile a quello di un utente su Reddit, “RavenChamps”, che però era contrario a Trump.
La diffusione della notizia fu tanto rapida quanto falsa. InfoWars titolava "L'impazzito autore della sparatoria di Jacksonville aveva criticato i Trumptards (sostenitori di Trump) su Reddit", e nell’articolo si affermava che egli “odiava i sostenitori di Trump”.

Alla fine venne fuori che RavenChamps era un’altra persona: un operaio del Minnesota di nome Pavel.
“Sono vivo, sapete?”, scrisse su Reddit qualche ora dopo la sparatoria (il vero responsabile si suicidò dopo aver commesso il massacro).

Per fare indagini serve molto più di un semplice username; tuttavia, ciò non toglie che questo potrebbe essere un punto chiave da cui partire per portare avanti le ricerche e, quando poi contatterai le forze dell’ordine, scaverai nei documenti pubblici e farai delle telefonate.

Ricostruire campagne in tempo reale (o quasi)

Le campagne di disinformazione e di manipolazione dei media vengono spesso diffuse attraverso Reddit e 4chan, e alcune possono essere tracciate in tempo reale.

Per esempio, per molti anni 4chan è stato usato nel business della manipolazione dei sondaggi online per sostenere certi candidati. Nel 2016 su 4chan vennero ripetutamente pubblicati link a siti di news, sia nazionali che estremamente locali, che trasmettevano i sondaggi fatti subito dopo i dibattiti presidenziali e che mostravano che il candidato preferito dagli utenti era Donald Trump.

Cambiando i parametri di ricerca di Google per filtrare i post relativi all’“ultima ora” e digitando nella barra di ricerca “site:4chan.org ‘polls’” avrai una visione abbastanza chiara dei sondaggi che gli utenti di 4chan tentano di manipolare in tempo reale.

La stessa cosa accadde anche nella successiva tornata elettorale. I sondaggi di 4chan spinsero Tulsi Gabbard, che veniva chiamata “Mommy”, nei sondaggi di The Drudge Report e NJ.com. Usando la semplice ricerca su Google menzionata sopra, chiunque poteva vedere il cambiamento in tempo reale dei sondaggi dopo che uno dei channer (gli utenti di 4chan) aveva detto agli altri “DATE A LEI IL VOSTRO POTERE”.

L'utile filtro di Reddit "Rising" permette di notare con facilità quali sono le operazioni di trolling in azione all'interno di spazi come la community di Reddit r/The_Donald.

Per vedere i risultati che in un dato subreddit a una data ora stanno guadagnando visibilità a una velocità insolita, usa la URL convenzionale “reddit.com/r/[subreddit-name]/rising” .

Guardando su reddit.com/r/all/rising puoi inoltre vedere i cosiddetti "overperforming post", post che in generale stanno avendo un successo inconsueto. Per i risultati di questa pagina vengono indicizzati tutti i post nella maggior parte delle community di Reddit. Non vengono considerati i subreddit in quarantena, che provengono da community tossiche con l’abitudine a condividere contenuti profondamente offensivi, e che prendono di mira altre community con campagne di trolling. I subreddit in quarantena non sono indicizzati nemmeno su Google, ma potete includerli nella ricerca con la query “reddit.com/r/[subreddit-name]/rising”. Mettere in quarantena i post è un'azione molto efficace per limitare il raggio d'azione di campagne di trolling al di fuori di un pubblico molto selezionato, ma rende molto più difficile capire come si stanno organizzando in un dato momento i soggetti mossi da intenti malevoli.

In generale, durante momenti salienti dell'informazione politica, eventi tragici o elezioni, è buona cosa tenere aperte delle schede sulle sezioni in ascesa di community conosciute per le loro campagne di trolling, come r/the_donald.

Non si può negare che talvolta le misure messe in atto dalle piattaforme per contrastare i manipolatori e chi diffonde disinformazione possano ostacolare i giornalisti nello svolgere indagini importanti. Ci sono strumenti che possono essere d'aiuto, ma gran parte del lavoro è manuale e richiede metodi di verifica che algoritmi e computer non sono in grado di riprodurre.

A conti fatti, questo tipo di lavoro non può essere sostituito dai computer. Il compito di farlo spetta a noi.

11. Analisi dei network e attribuzione

Scritto da: Ben Nimmo

***Ben Nimmo** è direttore delle indagini presso Graphika e membro senior non residente del Digital Forensic Research Lab dell'Atlantic Council. È specializzato nello studio di **operazioni di informazioni** e di ingerenza su larga scala e attraverso più piattaforme. Passa il suo tempo libero sott'acqua, dove non può essere raggiunto al telefono.*

Quando si ha a che fare con una sospetta operazione informativa, una delle questioni chiave per un ricercatore è quanto grande sia e quanto ampiamente si stia diffondendo. Sono questioni diverse rispetto alla misurazione dell'impatto dell'operazione, a sua volta importante: si tratta soprattutto di trovare gli account e i siti gestiti dall'operazione stessa.

L'obiettivo di un investigatore è scoprire quanto più possibile riguardo l'operazione prima di scriverne, perché dopo che un'operazione viene portata alla luce ci si può aspettare che i suoi responsabili si nascondano, ad esempio eliminando o abbandonando le risorse di cui si servivano.

Il primo anello della catena

In ogni indagine il primo indizio è il più difficile da trovare. Spesso le indagini cominciano da una soffiata da parte di un utente allarmato o, più raramente, da parte di una piattaforma di social media. Il lavoro del Digital Forensic Research Lab (il Laboratorio di Ricerca Forense Digitale) per portare a galla la sospetta operazione russa di intelligence chiamata "Secondary Infektion" cominciò con una soffiata di Facebook, che aveva rilevato 21 account sospetti sulla sua piattaforma. L'indagine raggiunse il suo culmine sei mesi dopo, quando Graphika, Reuters e Reddit svelarono un tentativo di condurre un'operazione simile per interferire nelle elezioni britanniche. In un altro caso fu la scoperta, da parte di un dipendente dell'organizzazione Vietnam Veterans of America, che su Facebook c'era una pagina che imitava il suo gruppo e che aveva il doppio dei follower rispetto a quelli realmente presenti sulla piattaforma a far partire un'indagine su un'operazione di disinformazione che aveva come obiettivo i veterani americani.

Facendo leva sulle tue risorse, non c'è una regola unica per trovare il primo anello della catena. La strategia più efficace è *cercare le incongruenze*. Potrebbe trattarsi di un account in apparenza localizzato in Tennessee, ma registrato con un numero di cellulare russo, oppure di una pagina che dichiara di operare dal Niger, ma che invece è gestita dal Senegal e dal Portogallo. Potrebbe essere un account di YouTube con milioni di visualizzazioni che nel 2019 pubblica una gran quantità di contenuti a

sostegno della Cina, ma in cui le visualizzazioni provengono quasi totalmente da alcuni episodi di sitcom britanniche caricati nel 2016.

Ancora, potresti trovarti davanti a un sito web anonimo che si occupa di politica estera americana, ma che è registrato dal Dipartimento delle Finanze di un distretto militare orientale della Federazione Russa. O avere a che fare con una presunta intervista a un "agente dell'MI6" formulata in un inglese ampolloso, quasi shakespeariano, o con un account Twitter che intervalla inviti a visitare un sito porno e citazioni incomplete tratte da "Ragione e sentimento" di Jane Austen.

Di fronte a tutti questi segnali, il segreto è prendersi il tempo di analizzarli a fondo. Investigatori e giornalisti spesso si trovano sotto pressione per via della mancanza di tempo, quindi può capitargli abbastanza facilmente di ignorare alcuni indizi pensando "è solo una stranezza" e passando oltre. Spesso le cose strane sono strane per un motivo. Prendersi il tempo per domandarsi "Perché è strano?" può essere il primo passo per far emergere una nuova operazione.

Elementi, comportamenti, contenuti

Una volta individuato il primo elemento — ad esempio un account o un sito —, la sfida è scoprire *a cosa* porti. A questo proposito ci sono tre domande fondamentali da porsi, formulate sulla base del Disinformation ABC di Camille François:

- Quali sono le informazioni disponibili riguardo l'elemento iniziale?
- Qual è stato il suo comportamento?
- Che contenuti ha pubblicato?

Il primo passo da fare è raccogliere quante più informazioni possibile riguardo l'elemento iniziale. Se si tratta di un sito internet, quando è stato registrato, e da chi? Presenta dati identificabili, ad esempio un codice Google Analytics, un numero di AdSense, una mail o un numero di telefono di registrazione? Questi aspetti possono essere verificati tramite gli archivi WhoIs, forniti da servizi come lookup.icann.com, domaintools.com, domainbigdata.com o spyonweb.com (dal nome orribile).

Domain Information

Name: nbenegroup.com

Registry Domain ID: 1558058690_DOMAIN_COM-VRSN

Domain Status:
clientTransferProhibited

Nameservers:
dns1.netbreeze.net
dns2.netbreeze.net

Dates

Registry Expiration: 2020-06-04 06:17:42 UTC

Registrar Expiration: 2020-06-04 06:17:42 UTC

Created: 2009-06-04 06:17:42 UTC

Contact Information

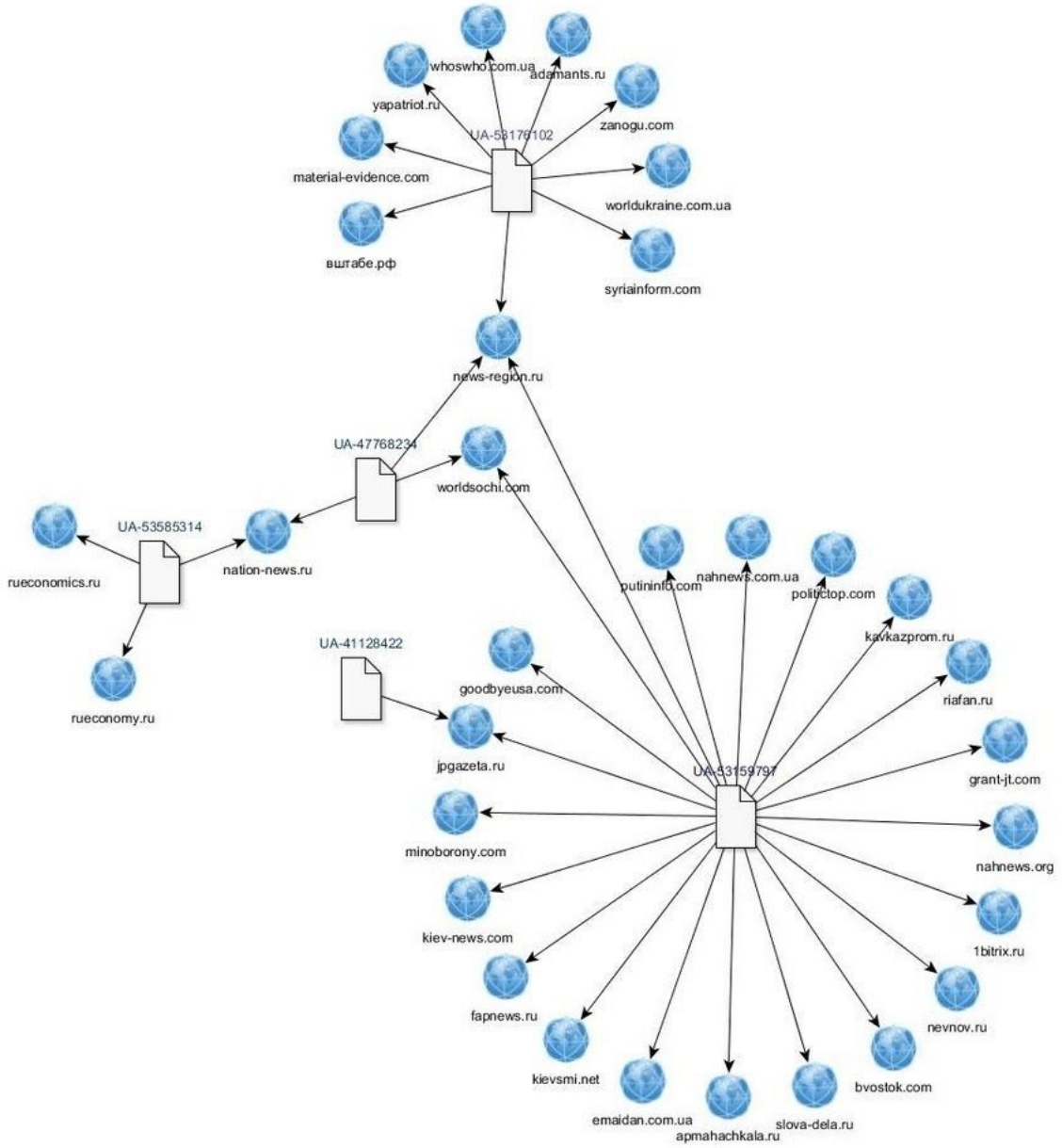
Registrant:

Name: Finance Department of the Far Eastern Military district

Dettagli della registrazione del sito web NBeneGroup.com, che si presenta come "gruppo di analisi giovanile", che mostrano la registrazione del sito da parte del dipartimento delle finanze del distretto militare dell'Estremo Oriente della Federazione Russa. Da lookup.icann.org.

Le informazioni sul sito possono essere utilizzate per arrivare ad altri elementi. Sia domaintools.com che spyonweb.com permettono agli utenti di fare ricerche usando dati come l'indirizzo IP e il codice di Google Analytics, operazioni che possono potenzialmente portare a siti associati, anche se oggi le *operazioni di informazione* più scaltre nascondono i dati di registrazione dietro soggetti commerciali o servizi di privacy, rendendo le cose più difficili.

In una prima fase della sua indagine, il ricercatore britannico Lawrence Alexander identificò 19 siti gestiti dalla Russian Internet Research Agency seguendo i codici di Google Analytics. Nell'agosto del 2018 la società di sicurezza FirmEye smascherò un'operazione di ingerenza iraniana su larga scala collegando tra loro siti in apparenza scollegati proprio attraverso i dati di registrazione dei siti, inclusi gli indirizzi email.



Network di siti web correlati collegati tra loro grazie ai codici di Google Analytics (sequenze di otto cifre preceduti dalle lettere UA) individuati dal ricercatore britannico Lawrence Alexander

Se l'elemento di partenza è un account sui social media, valgono le linee guida proposte nei precedenti due capitoli dedicati ai bot e alle attività non autentiche e alle indagini sugli account social. Quando è stato creato l'account? Il nome utente coincide con lo pseudonimo? (Se lo pseudonimo è @moniquegrieze e il nome utente è “Simmons Abigayle”, è possibile che l’account sia stato hackerato, o che faccia parte di una operazione di creazione massiva di account)

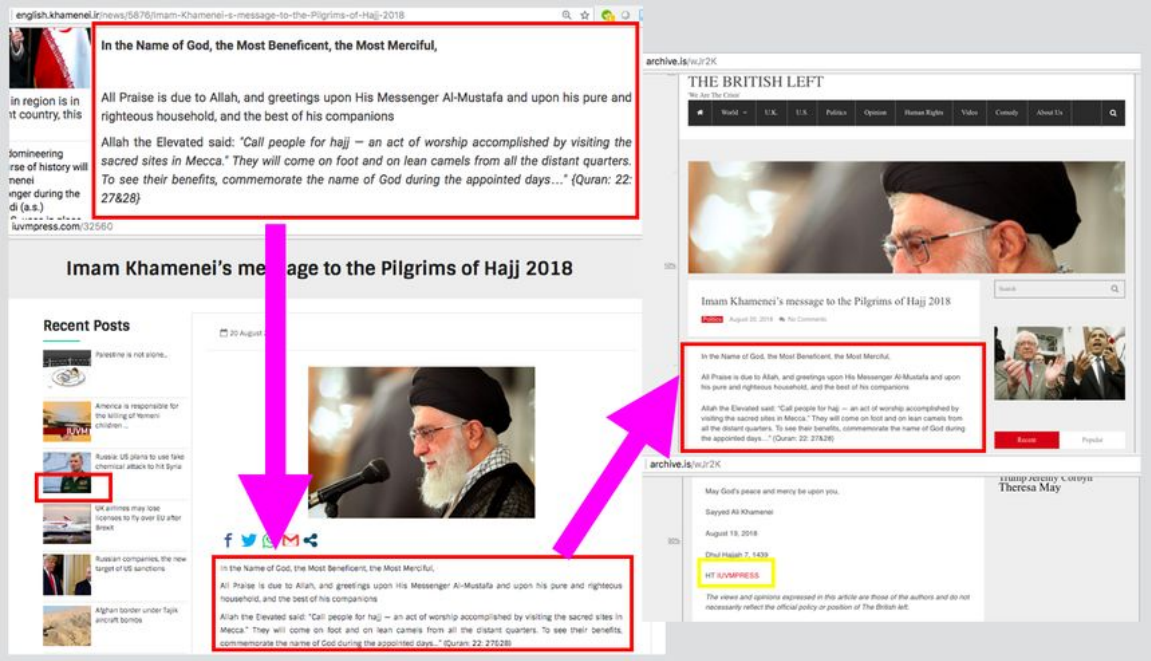
Tre account Twitter coinvolti in una grande operazione tramite bot nell'agosto del 2017. Confrontando i nomi utente e gli pseudonimi ci si accorgeva che molto probabilmente gli account erano stati hackerati, rinominati e riconvertiti per i propri scopi dal gestore dei bot.

L'account mostra informazioni biografiche verificabili o link ad altri elementi sulla stessa piattaforma o su altre? Se si tratta di una pagina o di un gruppo su Facebook, chi lo gestisce e dove si trova? Di chi è follower, e chi lo segue? Le informazioni mostrate nella sezione "Trasparenza della pagina" e in quella "Membri" di Facebook forniscono spesso indizi preziosi, così come come le informazioni dei profili Twitter, ad esempio la data di iscrizione e il numero totale di tweet e di like (su Facebook e Instagram non è possibile vedere la data di creazione di un account, ma la data di caricamento della prima foto profilo può essere ragionevolmente considerata un utile riferimento).

Sito web e Trasparenza della pagina Facebook del presunto sito di fact-checking "C'est faux — Les fake news du Mali" ("È falso — Fake news in Mali): il sito afferma di essere gestito da un gruppo di studenti in Mali, ma in realtà le attività sono condotte dal Portogallo e dal Senegal. Immagine da DFRLab.

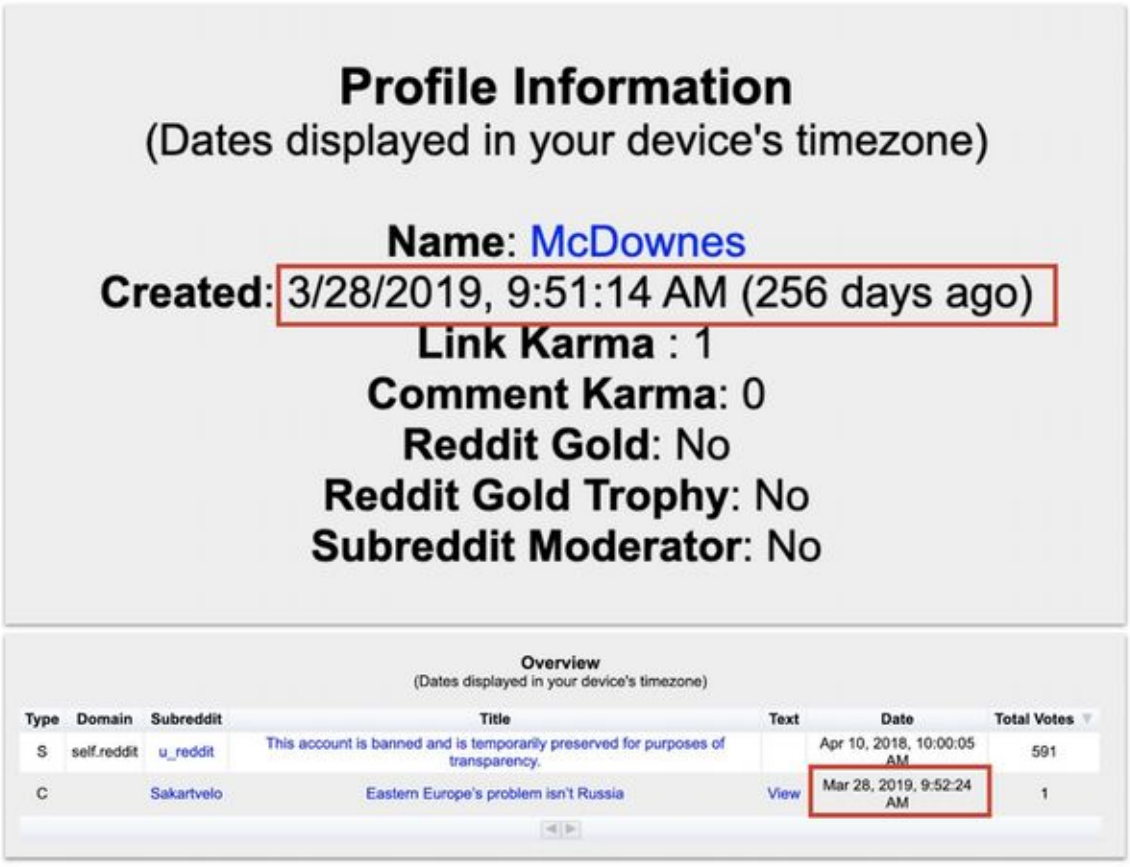
Dopo aver salvato e archiviato i dettagli riguardanti l'elemento di cui ci stiamo occupando, il passo successivo è determinarne il comportamento. La domanda chiave da porsi in questo caso è "Quali sono i tratti comportamentali più tipici di questo elemento e che potrebbero essere utili per identificare altri elementi coinvolti nella stessa operazione?"

Si tratta di una domanda ad ampio raggio che può avere molte risposte, alcune delle quali potrebbero arrivare solo negli stadi più maturi dell'indagine. I comportamenti in questione potrebbero essere, ad esempio, quelli di canali YouTube con nome e foto profilo occidentali, ma che pubblicano video politici in cinese inframmezzati a una gran quantità di brevi video presi da Tiktok. Oppure reti di account Facebook e Twitter che condividono link sempre dallo stesso sito o dalla stessa serie di siti. O ancora, account che nelle loro bio usano le stesse frasi, o loro variazioni; account di "giornalisti" che non riportano informazioni biografiche verificabili o che riportano informazioni che si sono scoperte false; siti che copiano la maggior parte dei propri contenuti da altri siti e che solo occasionalmente pubblicano articoli di parte, polemici o ingannevoli. Oppure, si potrebbe avere a che fare con molti di questi comportamenti insieme: per chi sta indagando, la sfida è individuare una combinazione di caratteristiche che permetta di dire: "questo elemento fa parte di questa operazione".



Modelli di comportamento: un articolo originariamente pubblicato sul sito web dell'ayatollah iraniano Khamenei e poi riprodotto senza attribuzione da IUVMpress.com e britishleft.com, due siti web facenti parte di una rete di propaganda iraniana. Immagine da DFRLab.

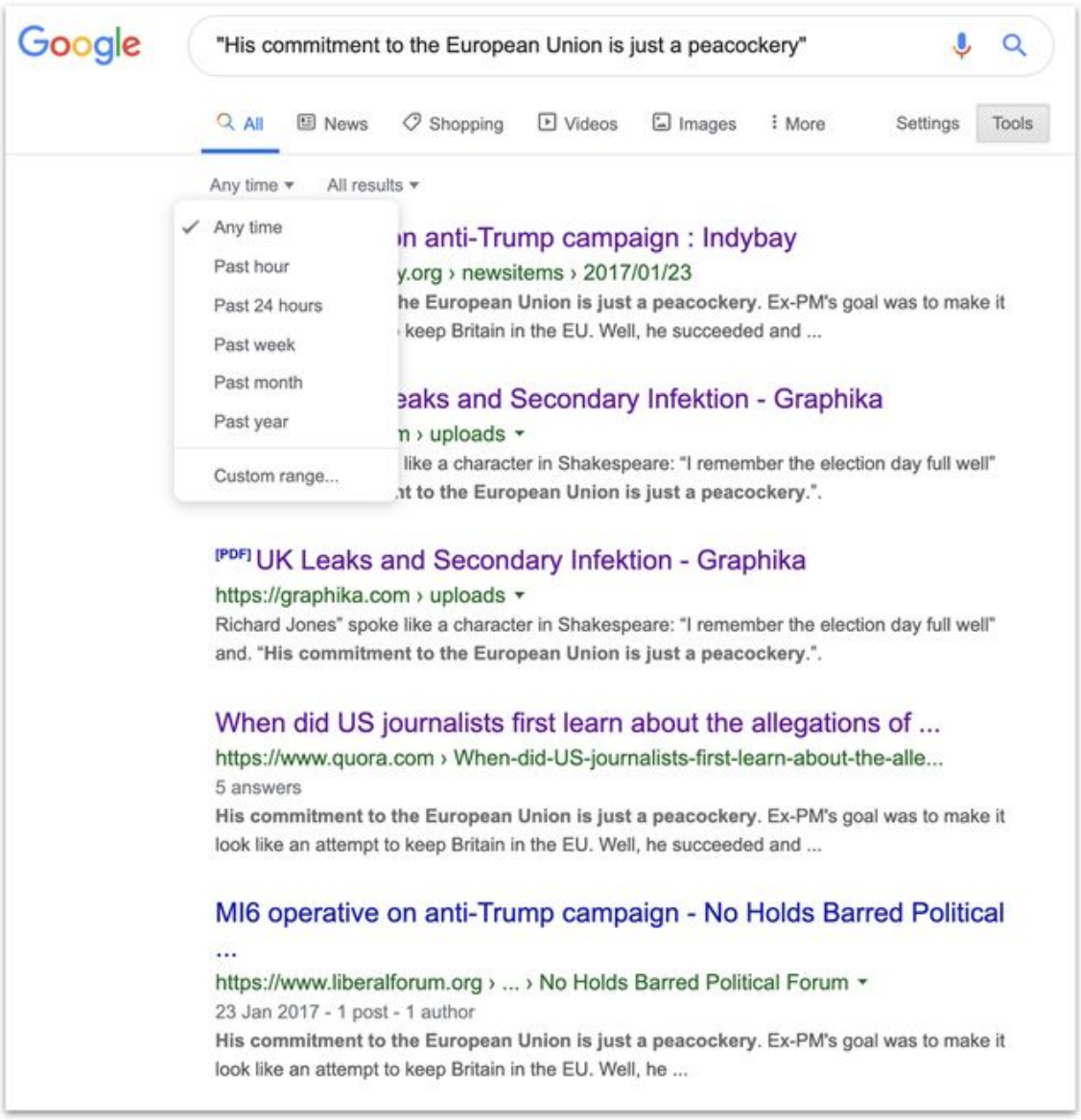
A volte la mancanza di caratteristiche identificative è essa stessa una caratteristica identificativa. È il caso della campagna "Secondary Infektion" condotta dalla Russia. La campagna usava centinaia di account su diverse piattaforme di blogging: ciascun account includeva informazioni biografiche minime, aveva pubblicato un articolo nel giorno in cui era stato creato e poi era stato abbandonato per non essere mai più usato. Questo modello di comportamento ricorreva in così tanti account che nel corso dell'indagine si profilò chiaramente come firma dell'intera operazione. Quando, poco prima delle elezioni generali britanniche del dicembre 2019, alcuni account anonimi iniziarono a far circolare documenti trapelati riguardanti il commercio tra Stati Uniti e Regno Unito, Graphika e Reuters dimostrarono che l'operazione si contraddistingueva proprio per quel tipo di comportamento. Reddit confermò l'analisi.



Il profilo Reddit di un account chiamato "McDownes," ricondotto da Reddit all'operazione russa "Secondary Infektion". L'account fu creato il 28 marzo 2019, pubblicò un articolo pochi minuti dopo la sua creazione e poi piombò nel silenzio. Immagine da Graphika, dati da redective.com.

Indizi utili riguardo l'appartenenza a uno stesso network possono essere rintracciati anche nei contenuti. Se un elemento conosciuto condivide una foto o un meme, vale la pena fare una ricerca inversa dell'immagine per verificare in quali altre occasioni è stata usata. Il plug-in per browser web di RevEye è particolarmente utile in questo senso, dal momento che permette di effettuare la ricerca inversa parallelamente su Google, Yandex, TinEye, Baidu e Bing. È sempre bene considerare diversi motori di ricerca, perché spesso portano a risultati diversi.

Se un elemento che abbiamo individuato condivide un testo, vale la pena andare a vedere in quali altri spazi è apparso quel testo. È consigliabile, specialmente con testi lunghi, scegliere una frase o due dal terzo o quarto paragrafo, o anche più in basso, dal momento che gli artefici delle operazioni di disinformazione modificano titoli e lead degli articoli copiati, ma è più difficile che si prendano il tempo che serve a modificare il corpo del testo. Per trovare le esatte corrispondenze del testo, inserisci la porzione di testo scelta tra virgolette nella barra di ricerca di Google. Dal menù "Strumenti" puoi ordinare i risultati per data.



I risultati di una ricerca su Google riferiti a una frase pubblicata nel contesto di una sospetta operazione di disinformazione russa, che mostrano la funzionalità degli strumenti di Google per limitare la finestra temporale della ricerca.

Dato che i refusi, per loro stessa natura, sono più rari delle parole scritte correttamente, i testi che contengono errori sono di particolare valore. Ad esempio, in un articolo che faceva parte di una sospetta operazione di intelligence russa ci si riferiva a Salisbury, la città britannica dove fu avvelenato l'ex agente segreto russo Sergei Skripal, come a "Solsbury". L'errore venne sfruttato per fare una ricerca molto più raffinata su Google, ottenendo molti meno risultati, ma molto più significativi, di quanti non ne avrebbe prodotti una ricerca con "Skripal" e "Salisbury".

Quando si guarda agli indizi di contenuto, per confermare o meno l'appartenenza di un elemento a un'operazione è particolarmente importante considerare anche altri fattori, ad esempio i modelli di comportamento. Gli utenti possono infatti essere spinti a condividere inconsapevolmente alcuni contenuti che fanno parte di un'operazione informativa, per molti motivi sensati. Ciò significa che l'atto di condividere un contenuto facente parte di un'operazione informativa è, di per sé, un indizio debole. Per esempio, i meme della Russian Internet Reasearch Agency avevano molte caratteristiche per diventare virali, e per questo sono stati condivisi da molti utenti. Il semplice fatto di aver condiviso un contenuto non è sufficiente a stabilire che un certo elemento faccia parte di un'operazione.

Raccogliere le prove

Le *operazioni di informazione* e di manipolazione sono complesse e si muovono in fretta. Una delle esperienze più frustranti per un ricercatore open source è vedersi mettere offline tutto un insieme di elementi nel bel mezzo di un'indagine. Vale sempre la regola d'oro "quando trovi una cosa, archiviala": potresti non avere una seconda possibilità di farlo.

Ricercatori diversi hanno preferenze diverse in merito all'archiviazione degli elementi che individuano, e le esigenze cambiano da operazione a operazione. I fogli di calcolo sono utili per archiviare informazioni di base su un gran numero di elementi; cartelle condivise in cloud sono utili per raccogliere grandi quantità di screenshot (se gli screenshot dovessero rendersi necessari, è fondamentale dare ai file nomi immediatamente identificabili: ci sono poche cose più seccanti di cercare di capire quale dei 100 file chiamati "Screenshot" è quello che ti serve). I file di testo vanno bene per registrare informazioni di varia natura, ma se l'operazione è estesa si rivelano ben presto scomodi e caotici.

Qualsiasi sia il formato, ci sono alcune informazioni che dovrebbero essere sempre registrate. Tra queste ci sono, ad esempio, il modo in cui l'elemento è stato trovato (è un punto essenziale), il suo nome e la sua URL, la data in cui è stato creato (se nota), il numero di follower, gli utenti seguiti, like e/o visite. Va inclusa anche una descrizione minima delle risorse (per esempio, "account in lingua araba filo-saudita con Emma Watson come immagine di profilo") che permetta di ricordarsi di che cosa si tratta dopo aver passato in rassegna altri 500 elementi. Se lavori con un team, vale la pena tenere traccia anche di chi ha trovato cosa.

I link possono essere salvati usando un servizio di archiviazione come Wayback Machine o archive.it, ma abbi cura che gli archivi non espongano utenti autentici che potrebbero aver interagito inconsapevolmente con un elemento sospetto, e assicurati che il link archiviato mantenga le parti visive; in caso contrario, fai uno screenshot di backup. Assicurati inoltre che tutte le risorse siano archiviate in luoghi sicuri, come file protetti da password o una cartella criptata. Tieni traccia di chi può accedere, e controlla regolarmente gli accessi.

Infine, vale la pena assegnare agli elementi un punteggio che ne indichi il livello di affidabilità. Spesso gli artefici di operazioni di manipolazione cercano utenti inconsapevoli per amplificare i loro contenuti. Di fatto, in molti casi è proprio questo il loro scopo. Quanto sei sicuro che l'ultimo elemento che hai trovato faccia effettivamente parte dell'operazione, e perché lo dici? È consigliabile prevedere una voce separata per indicare il livello di affidabilità (alto, moderato o basso), e aggiungere alle note le motivazioni (ne parliamo adesso).

Attribuzione e fiducia

La più grande sfida quando si ricostruisce un'operazione informativa sta nell'attribuirla a uno specifico attore. In molti casi, farlo con precisione è fuori dalla portata degli investigatori open source. Il massimo che si può ottenere è un certo grado di fiducia in merito al fatto che un'operazione sia stata probabilmente condotta da un determinato attore, o che vari elementi appartengano a una data operazione, ma raramente è possibile stabilire con certezza chi c'è dietro facendo affidamento solo sull'open source.

Informazioni come dati di registrazioni, indirizzi IP e numeri di telefono possono condurre ad una attribuzione solida, ma spesso sono oscurati a tutti tranne che alle piattaforme dei social media. Ecco perché contattare le piattaforme di riferimento è una parte fondamentale del lavoro investigativo. Dato che le piattaforme hanno ampliato i loro team investigativi interni, sono diventate più disponibili a condividere pubblicamente i risultati delle loro indagini in merito all'attribuzione delle **operazioni di informazione**. In casi recenti, l'attribuzione più solida è stata fornita proprio dalle piattaforme, come nel caso della denuncia da parte di Twitter dell'operazione informativa sostenuta dallo stato cinese che aveva come bersaglio Hong Kong, o nel caso della denuncia da parte di Facebook delle operazioni collegate al governo saudita.

Gli indizi di contenuto possono avere un ruolo. Per esempio, in un'operazione su Instagram smascherata nell'ottobre del 2019 erano stati postati meme praticamente identici a quelli pubblicati dalla Russian Internet Research Agency, a cui erano solo stati tolti i watermark dell'agenzia russa. L'unico modo in cui quei meme potevano essere stati prodotti era procurandosi le immagini originali dei post dell'agenzia russa per poi ricostruire i meme su di esse. Ironicamente, il tentativo di mascherare le origini dei post dell'IRA suggeriva che l'attore dietro l'operazione fosse proprio l'IRA.

Un caso simile è quello di un ampio network di siti apparentemente indipendenti che pubblicava ripetutamente articoli copiati, senza attribuzione, da fonti del governo iraniano. Questa attività era tanto frequente che risultò essere la principale del sito. Fu così possibile attribuire l'operazione ad attori filo-iraniani, ma non fu possibile arrivare ad attribuirla al governo iraniano in sé.

In fondo, l’attribuzione è una questione di autodisciplina. Gli investigatori devono chiedersi “come puoi provare che questa operazione sia stata condotta dalla persona che stai accusando?”. Se nel rispondere non sono assolutamente certi di quel che dicono, dovrebbero evitare di fare accuse. Individuare e portare in superficie un'operazione informativa è un lavoro difficile e importante: attribuire l'operazione a qualcuno senza prove o in maniera non accurata può compromettere tutto quello che è stato fatto prima di quella fase.

11a. Caso di studio: trovare gli autori dell'Endless Mayfly

Scritto da: Gabrielle Lim

***Gabrielle Lim** è una ricercatrice del Technology and Social Change Research Project allo Shorenstein Center dell'Harvard Kennedy School, e membro del Citizen Lab. Studia le conseguenze della censura e della manipolazione dei media sulla sicurezza e sui diritti umani.*

Nell'aprile del 2017 venne pubblicato su Reddit un finto articolo che faceva il verso al quotidiano britannico The Independent. L'articolo conteneva una falsa citazione dell'ex vice primo ministro britannico Nick Clegg, secondo cui l'allora primo ministro Theresa May stava "leccando i piedi ai regimi arabi". Il post venne subito segnalato dagli utenti esperti di Reddit come sospetto e falso. Non solo l'articolo era stato pubblicato sul sito "indep**n**edent.**co"** invece che su www.independent.co.uk, ma l'utente che l'aveva pubblicato per primo era un individuo superficiale che su Reddit aveva già pubblicato molti altri articoli non autentici.

Partendo da quel primo articolo falso, dal quel dominio e da quell'utente, nei successivi 22 mesi i ricercatori del Citizen Lab ricostruirono e indagarono il network che si sviluppava dietro la sfaccettata operazione informativa online in cui si inscriveva il caso in questione. L'operazione, chiamata Endless Mayfly (Effimera Immortale), prendeva di mira giornalisti e attivisti con siti non autentici, imitando siti di testate autorevoli e disseminando informazioni false e divisive.

In linea generale, il network funzionava così: veniva prodotto un articolo falso che imitava quelli di testate autorevoli, se ne amplificava la diffusione sfruttando una rete di finti siti web e account falsi su Twitter e, alla fine, dopo aver creato un po' di scompiglio online, l'articolo falso veniva cancellato o reindirizzato. Qui sotto, un esempio di un articolo falso che si fingeva di Bloomberg.com, ma che era stato invece pubblicato su bloomber**q**.com:

Questa immagine mostra due falsi personaggi online facenti parte dell'operazione Endless Mayfly che twittano un link che porta a una versione falsa del Daily Sabah,

una testata turca. La persona sulla destra, "jolie prevoit", ha come immagine di profilo una foto dell'attrice Elisha Cuthbert.

Quando pubblicammo il nostro report, nel maggio del 2019, i dati che avevamo raccolto comprendevano 135 articoli falsi, 72 domini, 11 persone, un'organizzazione fasulla e un network di informazioni filo-iraniane che amplificava le menzogne trovate negli articoli falsi. Alla fine, concludemmo con un moderato livello di fiducia che Endless Mayfly era un'operazione informativa filo-iraniana.

L'operazione Endless Mayfly ben illustra come, combinando l'analisi dei network e delle narrative cui ricorrono i suoi artefici tramite report esterni, sia possibile arrivare a identificare gli autori di un'operazione informativa.

Endless Mayfly è anche utile per mettere in luce le difficoltà che comporta attribuire l'operazione a uno specifico responsabile, i motivi per cui occorra raccogliere una molteplicità di prove, e come indicare il livello di fiducia dell'attribuzione per segnalarne il grado di affidabilità.

In ultima analisi, a meno che tu non sia in grado di ottenere una confessione o una prova sicura e definitiva, l'attribuzione è una missione difficile e spesso ostacolata da informazioni imprecise. Questo spiega perché, in molti casi di manipolazione dei media, l'indicazione dei responsabili sia spesso accompagnata da una stima probabilistica dell'affidabilità della scoperta.

Combinare molteplici dati e analisi

Data la natura clandestina delle *operazioni di informazione*, l'abilità degli attori in gioco di fare campagne sotto falsa bandiera e la natura effimera delle prove, l'attribuzione dovrebbe essere il risultato di una combinazione di prove e analisi. Nel caso di Endless Mayfly concludemmo con moderata certezza che si trattava di una campagna filo-iraniana grazie ai risultati di tre tipi di analisi:

1. Analisi delle narrative
2. Analisi del network
3. Report e analisi esterne

1. Analisi delle narrative

Tramite l'analisi dei contenuti e l'analisi del discorso condotte su 135 articoli falsi raccolti durante la nostra indagine, determinammo che la narrativa diffusa dall'operazione era in linea con gli interessi dell'Iran. Dopo aver letto preliminarmente tutti gli articoli in questione, determinammo una serie di categorie e assegnammo ciascun articolo a una categoria. Organizzammo due tornate di classificazione: la prima venne condotta da due ricercatori che lavorarono in maniera indipendente l'uno dall'altro; la seconda fu condotta dai due ricercatori insieme per risolvere ogni eventuale discrepanza. Questa tabella rappresenta il risultato del loro lavoro.

Category	Article count	Category description
Geopolitical discord	63 (46.7%)	The article describes events, actions or statements made by government officials toward a foreign state that may be construed as provocative, hostile or counter to the foreign state's interests.
Domestic discord	16 (11.9%)	The article describes events, actions or statements made by political actors that may sow discord between political parties or actors within the same state.
Cooperating with Israel	14 (10.4%)	The article describes events, actions or statements made by political actors or government officials that show cooperation between Israel and another state.
Saudi Arabia supports terrorism	9 (6.7%)	The article describes events, actions or statements that either link Saudi Arabia to terrorist activity or allege that Saudi Arabia supports terrorism.
Other	5 (3.7%)	The article does not fit into any of the categories.
No archive	31 (23%)	The article cannot be coded because it no longer exists and there is no cache, screenshot or copy of the text to perform any meaningful analysis.
Copy of existing article	5 (3.7%)	The article is a direct copy/paste of an already existing real article.

Dopo aver classificato tutti gli articoli, fummo in grado di determinare le narrazioni più comuni diffuse dall'operazione Endless Mayfly. Le confrontammo poi con le nostre ricerche preliminari riguardo il paese. Ciò comportò una ricerca estesa per capire le alleanze e le rivalità del territorio, gli interessi e i pericoli geopolitici e la storia del controllo delle informazione. Questo ci fu necessario per contestualizzare le prove e situare le narrazioni in un più vasto scenario politico. Con in mano i risultati della classificazione in categorie, stabilimmo che queste narrazioni erano molto probabilmente al servizio degli interessi dell'Iran.

2. Analisi del network

Per stabilire attraverso quali domini o piattaforme venivano amplificati i contenuti dell'operazione, fu condotta un'analisi del network. I network coinvolti dall'operazione Endless Mayfly nella diffusione di articoli falsi e delle loro menzogne nell'operazione erano due: un network di siti filo-iraniani e un cluster di utenti di Twitter filo-iraniani. Tutti e due vennero inclusi tra i responsabili dell'operazione sulla base del fatto che entrambi pompavano sistematicamente storie in linea con le

politiche, le dichiarazioni pubbliche e le posizioni dell'Iran nei riguardi di Arabia Saudita, Israele e Stati Uniti.

Il network di pubblicazione — Il network di pubblicazione era formato da un grande numero di siti in apparenza filo-iraniani, che si presentavano come organi di stampa indipendente. In totale trovammo 353 pagine web (su 132 domini) che facevano riferimento agli articoli falsi dell'operazione Endless Mayfly, oppure li linkavano. Per arrivare a tale conclusione fu necessario cercare su Google tutte le URL degli articoli falsi e i loro titoli. Inoltre, analizzammo i link twittati dagli utenti del network, trovando così pagine web che contenevano riferimenti o link a quegli articoli.

Seguendo questo procedimento, individuammo i 10 domini principali che rimandavano più di frequente agli articoli falsi. Di questi 10 domini, otto condividevano lo stesso indirizzo IP o dati di registrazione, il che significava che forse erano gestiti dallo stesso soggetto. Anche il contenuto di questi siti era volto a promuovere gli interessi iraniani. IUVM Press, ad esempio, che aveva linkato o citato articoli falsi di Endless Mayfly per 57 volte, conteneva al suo interno un documento PDF intitolato "Statute" ("Statuto") dove dichiarava esplicitamente di essere contro "le attività e i progetti degli stati dell'arroganza globale, contro l'imperialismo e contro il sionismo" e che "la sede dell'Unione si trova a Teheran, capitale della Repubblica Islamica dell'Iran".

Il network di utenti — Così come le posizioni degli articoli falsi e del network di pubblicazione, anche gli utenti Twitter che facevano parte della rete di Endless Mayfly erano decisamente critici nei confronti dell'Arabia Saudita, di Israele e, in generale, degli stati occidentali. Analizzando le loro attività su Twitter, fu possibile scoprire che questi account promuovevano alternativamente articoli credibili e articoli falsi molto critici verso i rivali politici dell'Iran. Prendiamo, ad esempio, l'account Twitter di "Peace, Security, Justice Community," una finta organizzazione che smascherammo durante le indagini e che mostriamo qui di seguito. L'organizzazione diffondeva contenuti contro l'Arabia Saudita, Israele e gli Stati Uniti, e usava persino la sua foto profilo e l'immagine di copertina per prendere di mira l'Arabia Saudita: nota il mirino sopra l'Arabia Saudita nella foto del profilo e la mappa usata nell'immagine di copertina. Come se non bastasse, nella bio dell'account si dice esplicitamente che l'Arabia Saudita e l'ideologia Wahhabi sono le cause dell'estremismo.

Impostazione simile ha il tweet che mostriamo di seguito, pubblicato da un altro account affiliato a Endless Mayfly, "Mona A. Rahman", il quale nel post critica il principe saudita Mohammad bin Salman menzionando, contestualmente, il giornalista e critico dell'Arabia saudita Ali al-Ahmed.

3. Report e analisi esterne

Abbiamo inoltre confrontato le nostre scoperte e i nostri dati con report esterni. Ad esempio, a seguito di un'indicazione ricevuta da FireEye, nell'agosto 2018 Facebook disattivò alcuni account e alcune pagine collegate al network di pubblicazione utilizzato da Endless Mayfly. Nelle sue analisi, FireEye aveva individuato molti domini che facevano parte del network da noi ricostruito, ad esempio institutomanquehue.org e RPFront.com. Come noi, anche FireEye concluse con un livello di fiducia moderato che "la sospetta operazione di manipolazione" sembrava partire dall'Iran. Anche Facebook, nella sua dichiarazione, scrisse che le operazioni in questione sembravano proprio provenire dall'Iran.

A ciò si aggiunse Twitter, che rilasciò una serie di dati riguardanti account legati all'Iran che erano stati sospesi a causa di una "manipolazione coordinata". Sebbene all'epoca della sospensione gli account con meno di 5000 follower erano stati resi anonimi, fummo in grado di identificare tra i dati di Twitter uno degli utenti (@Shammari_Tariq) coinvolti nell'operazione Endless Mayfly.

Le indagini di Twitter, Facebook e FireEye furono utili per corroborare le nostre ipotesi, in quanto fecero affiorare prove che i nostri sforzi non erano riusciti a far emergere e che allo stesso tempo coincidevano con gli elementi dell'operazione che avevamo già identificato. Ad esempio, i numeri di telefono e dati di registrazione raccolti dalle indagini di FireEye, e che noi non avevamo tra le informazioni da noi raccolte, erano collegati ad account Twitter e domini associati a Endless Mayfly. Probabilmente Facebook e Twitter avevano accesso ad alcuni dati di registrazione degli account, ad esempio indirizzi IP, a cui noi non avevamo accesso. I dati aggiuntivi forniti da questi report esterni ci aiutarono ad ampliare l'insieme delle prove.

Arrivare a un livello di fiducia moderato

Nel caso di Endless Mayfly, in base alle prove che avevamo raccolto — le narrative filo-iraniane, gli utenti e il network di pubblicazione — l'origine dell'operazione era da rintracciare nell'Iran. Le prove da noi raccolte vennero confrontate con report esterni affidabili prodotti da FireEye, Facebook e Twitter, i quali confermarono le nostre scoperte. Ogni singola prova era, da sola, insufficiente a procedere con l'attribuzione; ma considerata assieme alle altre in una visione integrata e messa in rapporto con la totalità delle prove portate alla luce, contribuiva a confermare e rafforzare la nostra ipotesi.

Tuttavia, malgrado molteplici elementi conducessero all'Iran, non avevamo ancora una prova definitiva. Per questo abbiamo utilizzato un quadro di riferimento per l'attribuzione di contenuti informatici comune tra chi si occupa di intelligence. Il quadro si serve di molteplici indicatori di certezza probabilistica (bassa, moderata,

alta), permettendo ai ricercatori di esprimere le proprie scoperte e allo stesso tempo classificarne il livello di affidabilità.

Alla fine, assegnammo ai risultati delle nostre indagini un livello di fiducia moderato, il che significa, come definisce lo U.S. Office of the Director of National Intelligence che "le informazioni provengono da fonti credibili e sono plausibili, ma non di qualità sufficiente o sufficientemente corroborata da garantire un livello di fiducia più elevato".

Scegliemmo di non optare per un livello di certezza più elevato, perché sentivamo di non avere le prove sufficienti per escludere completamente che si trattasse di una operazione di "false flag", ovvero sotto falsa bandiera, espressione utilizzata per indicare qualcuno che agisce facendo sembrare che dietro l'operazione ci sia qualcun altro: in questo caso l'Iran, o una terza parte che simpatizza con gli interessi dell'Iran.

L'attribuzione di *operazioni di informazione* come Endless Mayfly poggerà quasi sempre su informazioni incomplete o imperfette. Per questo ai fini dell'attribuzione è importante assegnare livelli di fiducia alle scoperte, per venire incontro al principio di massima cautela. Un'attribuzione errata o un livello di fiducia esagerato possono comportare conseguenze catastrofiche, specialmente se da queste valutazioni errate dipendono politiche governative o la decisione di applicare delle sanzioni. Per evitare di procedere all'attribuzione in maniera avventata o debole, è importante considerare molteplici indicatori, molteplici tipi di prove e analisi, e avere un livello di fiducia che tenga in considerazione anche ipotesi alternative e fatti mancanti.

11b. Caso di studio: Indagare su una operazione informativa in Papua Occidentale

Scritto da: Elise Thomas, Benjamin Strick

***Benjamin Strick** è ricercatore open source per la BBC, collaboratore di Bellingcat e docente di tecniche open source, intelligence geospaziale e analisi dei network. Ha un background nel mondo militare e del diritto. Si concentra sull'uso di OSINT/GEOINT, geolocalizzazione e metodi di intelligence a fin di bene, spaziando tra le aree dei diritti umani, dei conflitti e della privacy.*

***Elise Thomas** è una giornalista freelance e ricercatrice che lavora con l'International Cyber Policy Centre all'Australian Strategic Policy Institute. I suoi contributi sono stati pubblicati su Wired, Foreign Policy, The Daily Beast, The Guardian e altre testate. Precedentemente ha lavorato come assistente editoriale per l'Office for the Coordination of Humanitarian Affairs dell'ONU, come autrice di podcast e ricercatrice.*

Nell'agosto del 2019 tensioni separatiste tornarono a divampare in Papua Occidentale, una provincia dell'Indonesia che a seguito di una decisione piuttosto controversa venne annessa al Paese negli anni Sessanta. Da allora, nella regione sono piovute numerose accuse di violazioni dei diritti umani per reprimere il dissenso da parte delle autorità indonesiane.

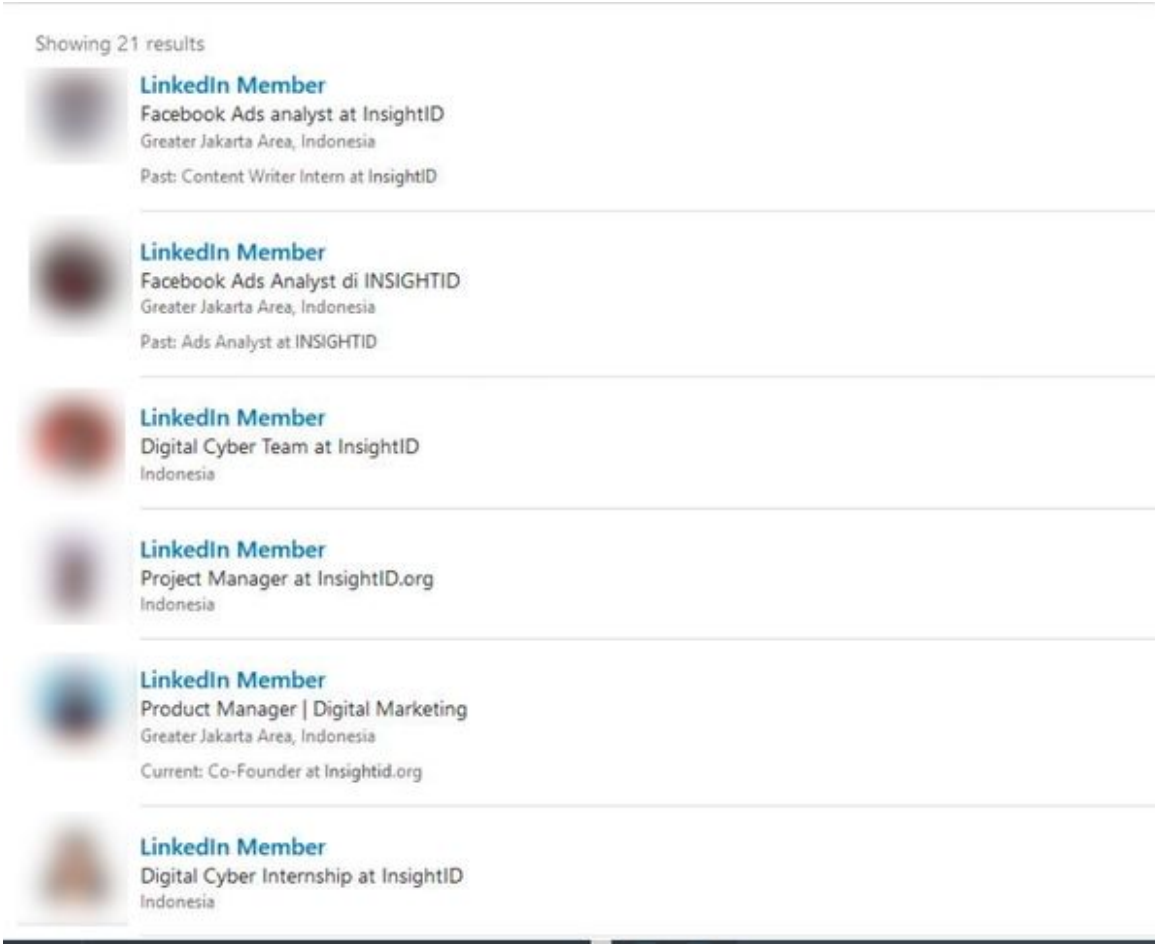
L'accesso alla regione è fortemente limitato, e ai giornalisti stranieri è vietato lavorare nell'intera provincia. Tutto ciò rende i social media una risorsa cruciale per monitorare e raccontare quel che accade in Papua Occidentale.
Mentre provava a geolocalizzare alcuni dei video che stavano saltando fuori sulle violenze commesse nella città di FakFak, uno di noi identificò due hashtag che si stavano diffondendo su Twitter: #WestPapua e #FreeWestPapua.

Conducendo delle ricerche tramite questi hashtag, fu possibile rivelare un'ondata di account fake che, utilizzandoli, pubblicavano automaticamente gli stessi video e lo stesso testo. Gli account si retwittavano e si mettevano like a vicenda, contribuendo ad amplificare e aumentare le interazioni attorno agli hashtag.

Il procedimento seguito per analizzare questi account automatici è stato descritto nel dettaglio al capitolo 3 di questo manuale. Partendo da quel lavoro, allargammo la nostra indagine con l'obiettivo di identificare le persone o i gruppi che stavano dietro l'operazione. Nel corso delle indagini scoprimmo una campagna simile a questa, ma più piccola e apparentemente non correlata, e riuscimmo anche a individuarne il responsabile. Quando la BBC contattò gli autori di entrambe le campagne, questi ammisero il loro coinvolgimento.

La portata della prima campagna, e il fatto che operasse su più piattaforme, ci mise di fronte a una serie di possibili strade che avremmo potuto percorrere per trovare indizi su cui far leva al fine di ottenere più informazioni riguardo i suoi autori.

I siti internet che venivano condivisi dagli account Twitter e Facebook del network furono il nostro primo indizio utile. Ricerche Whois rivelarono che quattro di quei domini erano registrati sotto falso nome e con un indirizzo mail fantoccio, ma con un numero di telefono vero. Inserimmo il numero su WhatsApp per vedere se era collegato a un account. Lo era, e aveva anche una foto profilo. Cercando quella foto con la ricerca inversa delle immagini di Yandex, riuscimmo a metterla in relazione con alcuni account su Facebook, LinkedIn e Freelancer.com. Grazie all'account LinkedIn scoprimmo persino il luogo di lavoro di quella persona, e potemmo vedere chi erano i suoi colleghi.



La persona in questione era un impiegato di un'azienda chiamata InsightID, con sede a Giacarta. Secondo il sito internet l'azienda si occupava di "programmi integrati di pubbliche relazioni e digital marketing".

Raccogliemmo ulteriori dati a riprova del fatto che InsightId era responsabile dell'operazione informativa. Sul suo sito internet, InsightID parlava del suo lavoro per la “Papua Program Development Initiative" (Iniziativa per un Programma di Sviluppo di Papua) che “esamina[va] il rapido sviluppo socio economico di Papua ed esplora[va] le sue sfide”. Ex dipendenti e stagisti di InsightID dissero che parte del loro lavoro sul Programma di Sviluppo di Papua consisteva in produrre contenuti video e scrivere e tradurre contenuti.

C'era poi un ex dipendente che dichiarava sul proprio profilo LinkedIn che il suo lavoro poteva essere trovato su “West Papuan (Instagram, Facebook, Website).” West Papuan era uno dei cinque siti di news coinvolti nella campagna. Inoltre, un altro dipendente di InsightID aveva creato un account Youtube a suo nome per pubblicare un video come parte della campagna. Questo video fu poi embeddato su westpapuan.org.

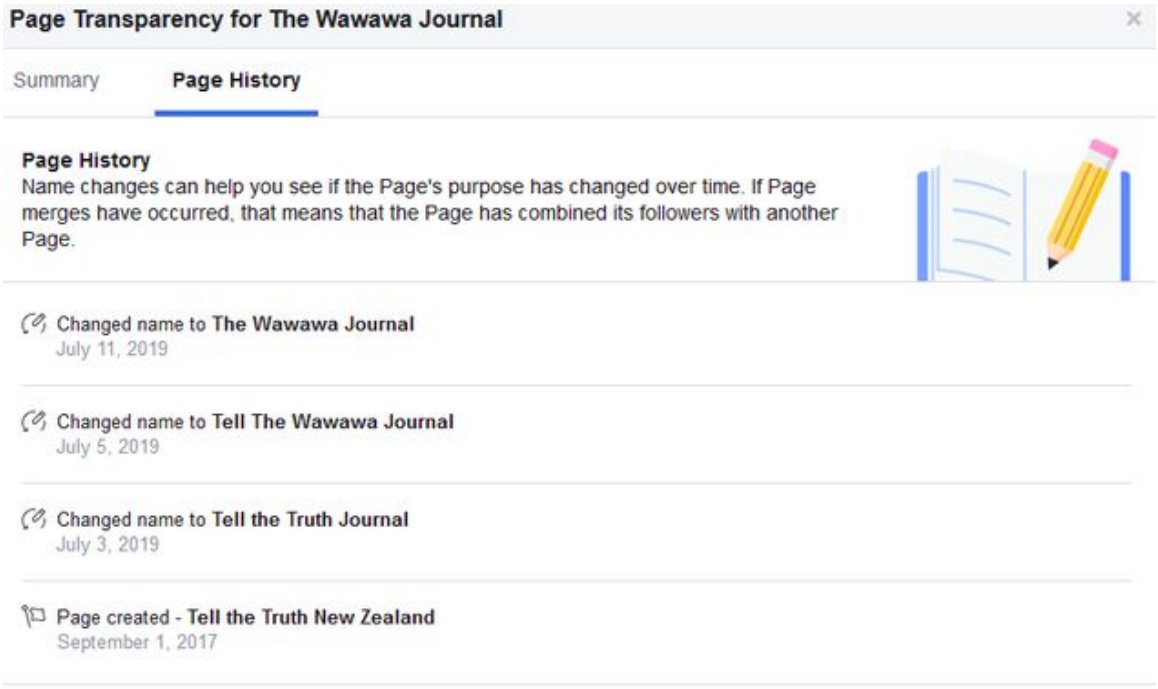
Ulteriori ricerche sui dati di registrazione dei domini rivelarono che il co-fondatore di InsightID aveva utilizzato il suo indirizzo email aziendale per registrare 14 domini nello stesso giorno, la maggior parte dei quali chiaramente e direttamente collegati alla Papua Occidentale. Tra questi c'erano westpapuafreedom.com, westpapuagenocide.com e westpapuafact.com. Ogni informazione in più che ottenevamo si aggiungeva alle prove che il responsabile dell'operazione fosse InsightID.

A quel punto, i giornalisti della BBC provarono a contattare InsightID per una dichiarazione in merito alla questione. L'azienda non rispose. Ciononostante, scrisse un post sui social dichiarando che “i nostri contenuti difendono l’Indonesia contro la falsa narrativa dei gruppi separatisti Free Papua”, ammettendo in sostanza la sua responsabilità.

Non riuscimmo a identificare il committente che ingaggiò InsightID per portare avanti la campagna di disinformazione.

Mentre portavamo allo scoperto questa operazione più ampia, indagavamo anche su un piccolo network formato da tre siti, camuffati da fonti di notizie indipendenti e associati a dei profili social. Nonostante questi siti non fossero in apparenza connessi alla prima campagna, miravano alla percezione internazionale della situazione in Papua Occidentale, concentrandosi sul pubblico della Nuova Zelanda e dell'Australia.

La chiave per risalire al responsabile di questo piccolo network fu il fatto che la pagina Facebook di uno dei siti, quello del Wawawa Journal, si chiamava in origine Tell the Truth NZ. Lo scoprimmo osservando lo storico dei nomi della pagina. Ciò ci permise di collegare il sito al dominio tellthetruthnz.com, che era stato registrato a nome di Muhamad Rosyid Jazuli.



Contattato dai giornalisti della BBC, Jazuli ammise di essere l'autore della campagna. Jazuli lavora con il Jenggala Center, un'organizzazione creata dall'allora vice presidente dell'Indonesia, Jusuf Kalla. Venne creata nel 2014 per promuovere la sua rielezione e per supportare l'amministrazione del presidente Jokowi.

Questa indagine ha dimostrato che per scoprire le campagne di disinformazione e trovarne i responsabili (singoli individui o gruppi), non servono necessariamente tecniche o strumenti sofisticati. Quel che è certo, tuttavia, è che servono pazienza e una certa dose di fortuna. Per questa indagine abbiamo fatto ricorso a strumenti open source, come i registri Whois, la ricerca inversa delle immagini, l'analisi dei profili sui social e dei codici sorgente dei siti internet. Il fatto che la campagna fosse attiva su più piattaforme e in combinazione con i social media e i profili LinkedIn dei dipendenti di InsightID è stato fondamentale per consentirci di mettere insieme molti piccoli indizi per costruire il quadro generale.

Se c'è una lezione chiave da portarsi a casa alla luce di questo esempio, è quella di soffermarsi sempre a riflettere su come sfruttare i dettagli o gli indizi ottenuti da una piattaforma per spostarsi su un'altra.

Credits

Editor: Craig Silverman

Contributing Editor: Claire Wardle

Copy Editor: Merrill Perlman

Contributors: Ben Collins, Ben Nimmo, Benjamin Strick, Brandy Zadrozny, Charlotte Godart, Claire Wardle, Craig Silverman, Donie O'Sullivan, Elise Thomas, Farida Vis, Gabrielle Lim, Gemma Bagayaua-Mendoza, Hannah Guy, Henk van Ess, Jane Lytvynenko, Joan Donovan, Johanna Wild, Sam Gregory, Sérgio Lüdtke, Simon Faulkner, Vernise Tantuco

Production Manager: Arne Grauls

Questo manuale è stato pubblicato dall'European Journalism Centre e è stato possibile grazie al finanziamento di Craig Newmark Philanthropies.

La traduzione italiana è stata curata da Slow News e resa possibile grazie al sostegno di Pagella Politica e Facta News.

Traduzione in italiano: Andrea Coccia

Revisione: Elena Brilli

Correzione bozze: Francesca Menta