



Reading In-Between the Lines: An Analysis of Dissenter

Erik Rye (CMAND) Jeremy Blackburn (Binghamton University) Robert Beverly (NPS)

Internet Measurement Conference 2020







Warning!

This talk contains content that some individuals will find offensive, including racist, misogynistic, homophobic, and anti-Semitic language.

Individuals that are sensitive to this type of content may instead wish to read the paper, as it characterizes this language in the aggregate without specifically reproducing it.



Motivation

- Online censorship a hot topic
 - Twitter warning labels
 - Facebook, Twitter banning link sharing
- Why not just study Twitter, YouTube, Facebook?
 - Mainstream social networks ban individuals
 - Sites turn off comment sections
- Dissenter is a platform for free speech...
 - (mostly) free from content moderation
- ...but what's really there?





What you see



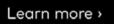


at qualifying restaurants, transit, grocery & drug stores* Tap & Go

to help support cancer research



*Online Mastercard transactions also apply.



Røtten omates

Search movies, TV, actors, more...

Q

MOVIES & DVDS ▼ TV ▼ NEWS ▼ TICKETS & SHOWTIMES

What's the Tomatometer®? Critics

Our New Podcast Scare Central RT Watch List Haunting of Bly Manor



SIGN UP | LOG IN

THEATERS MOVIES

TV SHOWS

MOST POPULAR

IN

92% The Trial of the Chicago...

STREAMING

- **86%** The More You Ignore Me
- **100**% Bellingcat: Truth in a P...
- *** 44%** The Lie
- *** 57%** Ms. White Light
- 80% Charm City Kings

No Score Yet Daddy's Girl

- 100% Buddies
- 75% Console Wars
- ***48**% Lost Girls & Love Hotels

View All



from qualifying restaurants, transit, grocery & drug stores*





CAPTAIN MARVEL

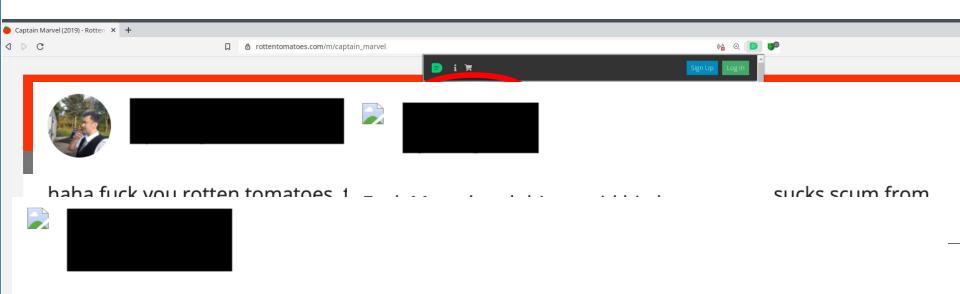
Critics Consensus



What Dissenter sees



Reading In-Between the Lines: An Analysis of Dissenter



Don't forget to watch this movie or you're a racist sexist Islamophob homophob nazi right wing nut that deserves to be put in gas chambers! This comment is sponsored by Disney









Jews will not re

"This movie is so good it silences all critics! 100% fresh!" - RT niggerfaggots



Packed with action, humor, and visual thrills, *Captain Marvel* introduces the



How did we get here?

- Reaction to content moderation policies
 - Anti-hate speech, norm enforcement
- Feb 2019
 - YouTube disabled comments for millions of videos
 - Rotten Tomatoes received deluge of negative reviews before Captain Marvel release



Gab Dissenter's Parent Social Network

- Founded 2016
- Alternative social network
 - Like Twitter, for people banned from Twitter
- Zannettou et al. "What is Gab?", WWW 2018
 - 22M posts by 336k users
 - Significant hate speech (~5% of all posts)
 - On the border between mainstream social networks and "fringe" web communities





Dissenter Overview

- Created by Gab in February 2019
- Subvert Big Tech censorship
 - Comment on any URL
 - Comments only visible to others using service
 - A "comment overlay" system
- Earlier attempts at web annotation:
 - Google Sidewiki, 2009-2011





Dissenter Overview

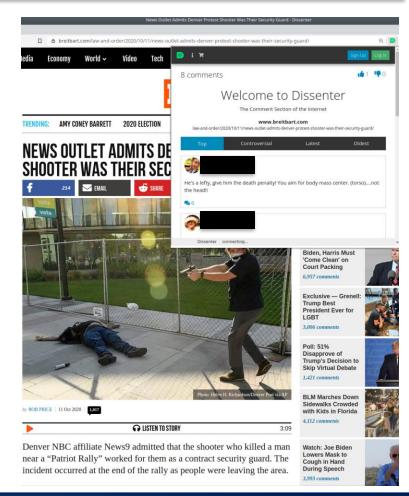
- First: Chrome/Firefox plugin
 - But, banned by Google/Mozilla for ToS violations within 2 months
- Current:
 - Dissenter Browser
 - Gab Trends





Dissenter Browser

- "Free Speech Web Browser"
 - Brave fork
- Dissenter plugin integrated
- Windows, macOS, and Linux versions





Gab Trends

- News aggregation site
- Displays trending, usersubmitted news and videos
- Provides brief description and comment board for each URL
 - Comment system is Dissenter





Reading In-Between the Lines: An Analysis of Dissenter



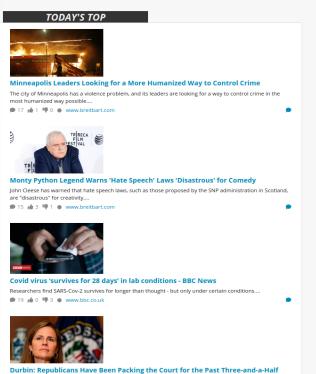


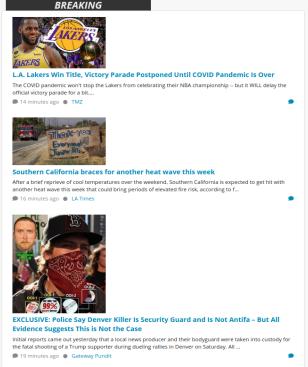
REVEALED: Charged Denver Shooter Matthew Dolloff's Left-Wing Facebook Posts

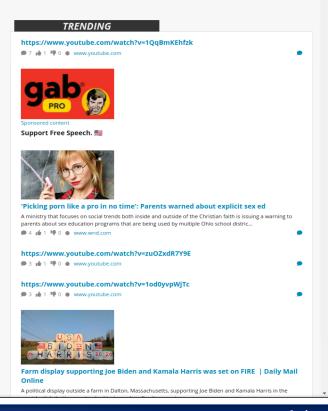
GAB TRENDS

Paste a URL to start a Trend...

START TREND





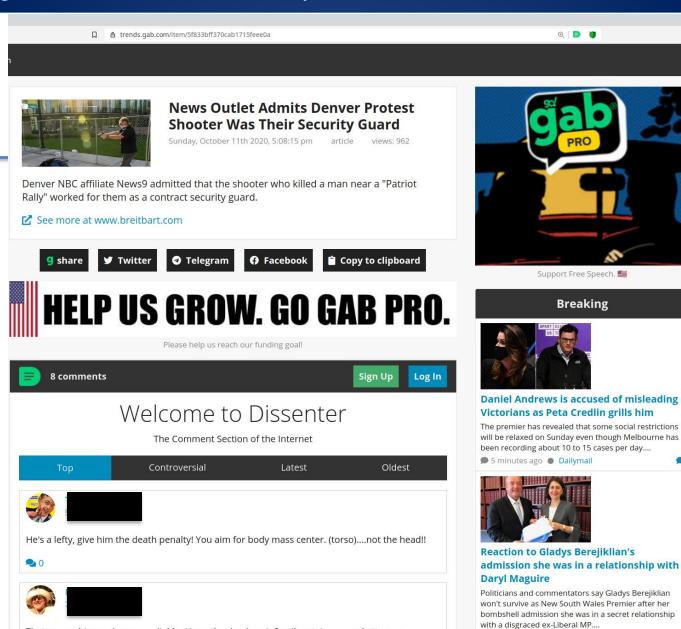


IMC 2020

Years -- 'They Brag About It'

Reading In-Between the Lines: An Analysis of Dissenter





IMC 2020

8 minutes ago Dailymail

That means his employers are liable. Hope the dead guy's family retains a good attorney.





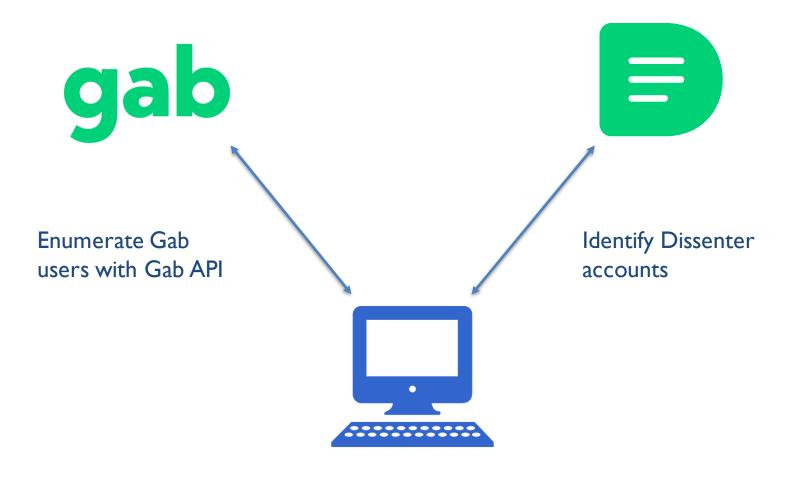
Our work

- 1. Enumerate Dissenter
- 2. Analyze Content
- 3. Case Study



Dissenter Enumeration Username Discovery

Username Discovery





Dissenter Enumeration Username Discovery

Username Discovery

Gab API

Dissenter account created

erik@rocinante:~ » curl https://dissenter.com/user/a 2>/dev/null | wc -c 72202 erik@rocinante:~ » curl https://dissenter.com/user/e 2>/dev/null | wc -c 145

Dissenter

No Dissenter account



Dissenter Enumeration Comment Harvesting

Comment Harvesting

- For each Dissenter user
 - Visit home page
 - Gather URLs they've commented on
- For each URL
 - Visit "URL page": Dissenter's aggregation of comments for a given URL
 - Harvest comments



Dissenter Characterization

- Discover 101k Dissenter users
 - 79k joined in first full month
 - 47k active (commented on >= 1 URL)
- 1.68M comments/replies
- 588k unique URLs



Dissenter Characterization Commented URLs

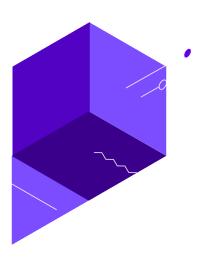
# Commented URLs by Domain			
youtube.com	121,928 (20.75%)	foxnews.com	12,196 (2.08%)
twitter.com	40,392 (6.87%)	bitchute.com	12,124 (2.06%)
breitbart.com	23,705 (4.03%)	zerohedge.com	8,634 (1.47%)
bbc.co.uk	16,213 (2.76%)	theguardian.com	8,010 (1.36%)
dailymail.co.uk	15,752 (2.68%)	youtu.be	7,819 (1.33%)
Other	320,962 (54.61%)	Total	587,735 (100%)

Video sharing, social media, and news websites are the most commonly commented-on domains. Mix of mainstream (YouTube, Twitter, BBC) and more fringe (Bitchute, Zero Hedge) content



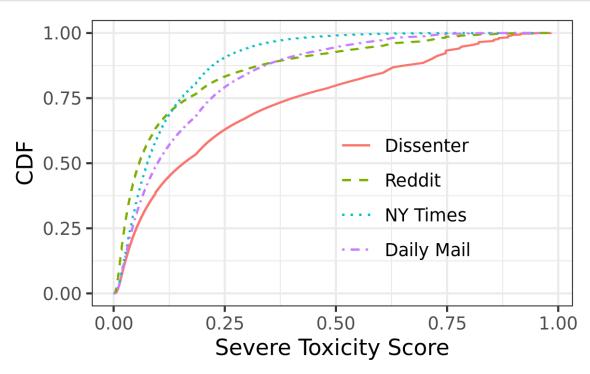
Content Analysis

- Use Perspective API to classify toxic and offensive content
 - Toxicity, hate speech detection a hard problem
- 17 unique "attributes" that measure toxicity, obscenity, other characteristics
- Validate Perspective API results consistent with
 - Hate word dictionary (Hatebase)
 - NLP classifier
 - Use Perspective API for remaining analysis





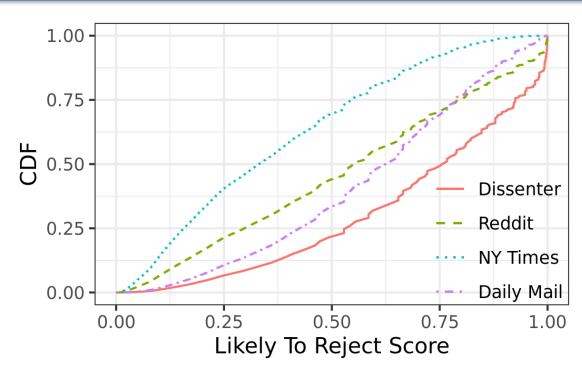
Dissenter Comments vs Related Datasets



- "Severe Toxicity" Perspective model measures very hateful speech; less sensitive to profanity than other toxicity models.
- Dissenter comments have substantially higher levels of toxicity than other platforms



Dissenter Comments vs Related Datasets



- "Likely to Reject" trained on NYT comments; how likely moderators are to reject comments
- Dissenter comments much more likely to be rejected than other datasets
- Validates reason for existence



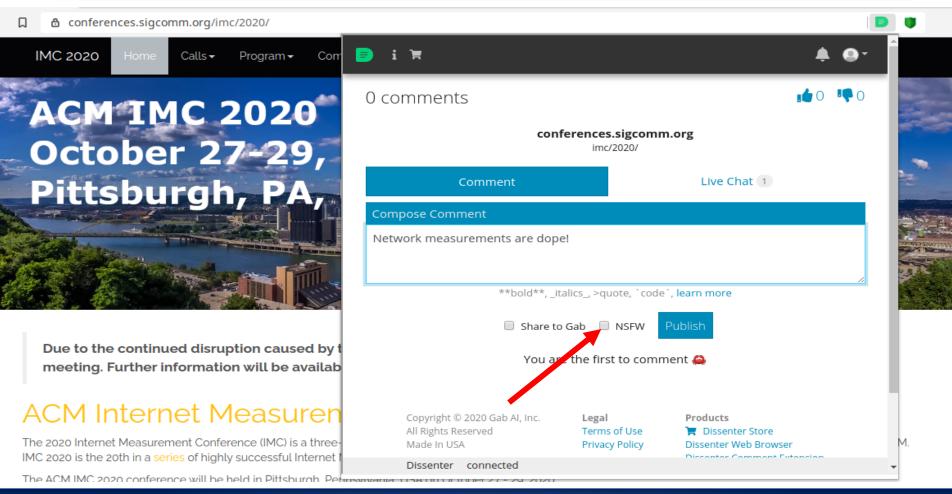
NSFW and Offensive Content

- Some Dissenter comments have additional labels
- Must explicitly opt-in to view
 - NSFW
 - User-tagged
 - Offensive
 - Dissenter-tagged
 - Tagging mechanism unknown
- Leaked metadata reveal numbers opted-in
 - 7,000 opted-in NSFW
 - 3,500 opted-in Offensive





NSFW-Tagged Content

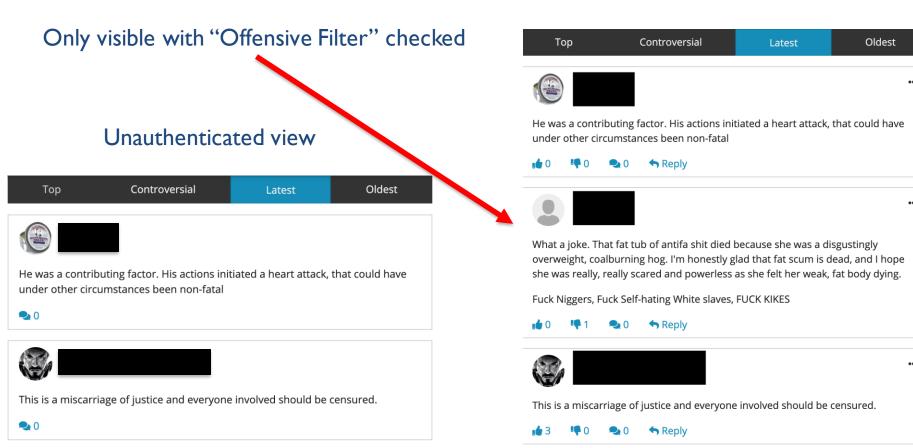






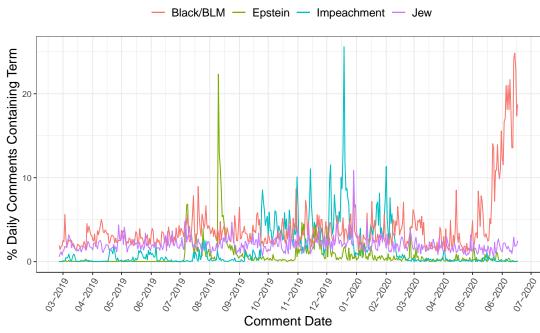
Offensive-Tagged Content

Authenticated with "Offensive" filter checked





Case Studies

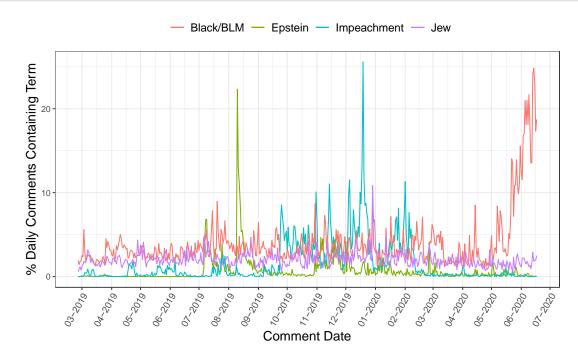


- Reflect both
 - Current events
 - Underlying biases, interests of user base

- Examine % of comments with term each day
 - In comment text
 - In URL string



Case Studies – Current Events



Epstein

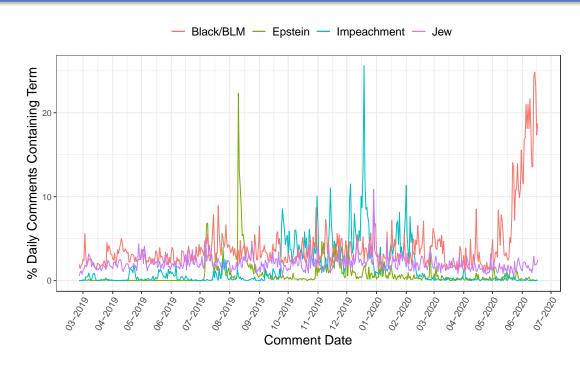
Peaks when arrested and after suicide

Impeachment

 Significant interest during hearings and trial



Case Studies – Community Interest

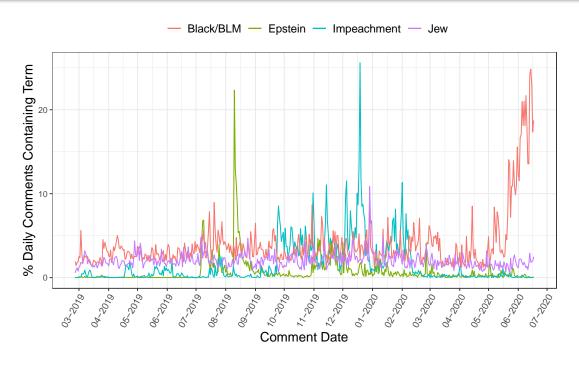


Jew/Jewish

- Sustained significant daily interest (1-5%)
- Small spike following mass stabbing during Hanukkah
- Prior work indicates high level of anti-Semitism in Gab



Case Studies – Community Interest + Current Events



Black/BLM

- Like Jew/Jewish, exhibits sustained interest
- Death of George Floyd and accompanying BLM protests



Conclusions

- Characterize Dissenter, the "Comment Section of the Internet"
 - Find 101k users, 1.68M comments, 588k URLs
- Comments significantly more toxic than related datasets
- Comments likely to be rejected by moderators
 - Underscoring Dissenter's niche as free speech bastion
- Case studies show current events as well as underlying community interests reflected in Dissenter comments

Thanks for listening!