Digital Pink Slime: Measuring, Finding, and Countering Online Threats to Local News

Christine Sowa Lepird

November 7, 2024

School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Kathleen M. Carley, Chair, Carnegie Mellon University
Ananya Sen, Carnegie Mellon University
Hong Shen, Carnegie Mellon University
Jeff Hemsley, Syracuse University

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

To all of the local news reporters who work tirelessly to bring accurate and unbiased reporting to their communities.

Abstract

In the past twenty years, journalism has had to evolve to keep up with the digital era that publishes stories not on print mediums but on online websites that are shared via social media in order to be consumed by the public. Unfortunately the smallest of journalistic outlets - those serving local communities - have been hit the hardest. Many local newspapers have closed due to budget cuts, but the trust in local news reporting remains high. In their place, some unsavory actors have decided to exploit this trust to share national messaging under the guise of local news. They have created hundreds of websites designed to appear as part of small American communities - particularly communities in swing states and those of national electoral importance. With little to no actual reporters, these sites are largely filled with automated reporting on community budgets, weather, and sports. The primary pull for these websites are their partisan political articles which are shared on Facebook, Twitter, and Reddit.

This thesis is a comprehensive study into these sites that are masquerading as local news while pushing a national agenda and spending significant money to do so. Each element of this research aims to answer the questions: how is the behavior of those creating and sharing pink slime sites different from that other news sites (be it local, real, or low credibility news)? Furthermore, how can I leverage their defining characteristics to train others to find and be wary of these sites? I start by concretely defining this phenomenon of ‘pink slime’, how it gained footing in the online news ecosystem, and what gaps in current literature inform the research I conducted. I then utilize computational social science and social network analysis methods to quantify the characteristics of these sites (in comparison to real news, local news, and low credibility news sites) in the first large scale empirical assessment of pink slime. From an information operations perspective, I categorize the network and narrative BEND maneuvers utilized by pink slime across multiple social media platforms and compare it to three other news types across social media platforms. Applying natural language processing, machine learning, and network analysis, I propose a network feature that can find new sources of these sites and prove its effectiveness. In a study on human subjects, I learn how a reader’s trust in pink slime and local news differs and how training impacts their ability to recognize pink slime. Finally, I summarize the findings and relevant literature to make policy recommendations to counter this threat to local communities.

Acknowledgments

In writing this thesis, I'd be remiss if I didn't note all of the people who have helped me to get here today. The journey to completing a PhD is an arduous one under the most straightforward of conditions; I could have never made it to my defense while enduring a global pandemic and becoming a parent these past six years if it weren't for the individuals mentioned below.

I've been very fortunate to have worked with wonderful researchers and am grateful for the guidance from my advisor from my first day on campus. Kathleen M. Carley has helped me to forge a path in my research, explore new areas, and become a better teacher. I am also grateful to my committee members, Ananya Sen, Hong Shen, and Jeff Hemsley, for making my thesis better by providing their suggestions based on their extensive areas of expertise.

When I started the PhD journey, I did not know how many wonderful collaborators I would meet along the way. Everyone in the CASOS research group gave phenomenal feedback during the many stages of my research. Catherine King worked tirelessly over the past five years on making the OMEN project and our media literacy testing a success. Isabel Murdock was always available to discuss the ways in which multi-platform work could be studied. Lynnette Ng helped me to grow in my scientific writing and served as a sounding board whenever I'd have fresh results. Joshua Uyheng was a fantastic co-teacher who inspired me to persevere through any setback. My good friend, Luke Osterritter, led me to prioritize and view this program as a marathon, not a sprint. I'm grateful to Matt Hicks who always made time to work my research into his advanced scenario generator. As one of the first collaborators interested in advancing pink slime research, I was lucky to learn about backlinks and SEO exploitation of news from Peter Carragher. My fellow OMEN team members - Janice Blane, Geoff Dobson, Charity King, Ian Kloo, and Jonathan Morgan - showed me how much one team could accomplish. My summer interns, Anna Wu and Mugilan Nambi, were important to my growth in understanding how to manage others and break down the tasks of our many projects.

I was also lucky enough to receive generous funding to pursue novel research. For that, I'd like to thank the Knight Foundation for their IDeaS Center Knight Fellowship and the Office of Naval research's Project Scalable Tools for Social Media Assessment (N000142112229).

Finally, I'd like to thank my family for their enduring support. To my parents, Chris and Nancy Sowa, who have always encouraged me to work hard and pursue education. Fostering my love of local news, my mom would drive me into Philadelphia every weekend for my first internship at KYW News Radio in high school. To my grandfather, Professor Walter Sowa, who eagerly asked for research updates and regaled me with tales of how the computer science field has evolved since his grad school days in the 1940s during every weekly call. To my husband, Jack Lepird, who would brainstorm research ideas with me over dinner and keep me positive with every setback I faced. Finally, to my daughter, Diane - you are my inspiration to take on challenges; I love you endlessly.

Contents

List of Figures

List of Tables

Chapter 1

Introduction, Background, and Motivation

1.1 Overarching Thesis Goals

While the digitization of the news industry has allowed more people to rapidly access information on any smartphone, there has been a dark downside to the ease at which companies can register domain names and populate a website with automated filler content. Pink slime journalism has oozed its way into the media diet of unsuspecting Americans who still place a higher trust in local news institutions [51] to keep them informed on issues that matter to them.

Academic research on the topic has been largely limited to consumption and answering the question of who is likely to click on these websites [81]. Furthermore, the publications have not analyzed the phenomenon past the 2020 U.S. Presidential election [29] [81] despite the ad spend on their largest social media sharer (Facebook) more than doubling from 2020 to the 2024.

In this thesis I apply social network analysis, statistical analysis, machine learning, and user studies to answer the following questions: how is pink slime fundamentally different from other news types vying for attention on social media? Furthermore, by answering the first question, can I use the characteristics to inform algorithms to more quickly discover new sources of these pink slime websites? Finally, with the first two questions answered, can I take the results and effectively teach others to recognize pink slime when it is encountered?

This thesis is organized into six chapters. It begins by defining pink slime, what we know about it from previous research, similar phenomenon that has been observed internationally, and motivating why it is worth studying (Ch. 1). From there, it measures the impact that these sites are having and defining the network characteristics they present in comparison to other news types (Ch. 2). By including the narrative as well as network elements, it dives into which information operations are being conducted by the controlling pink slime organizations across various social media platforms (Ch. 3). It establishes a new metric to find these sites and tests its performance across social media platforms and in differing regions of the world (Ch. 4). It then reports on experimental findings to assess human trust of pink slime and the effectiveness of training on detection (Ch. 5). Finally, it draws on findings from the earlier chapters to make policy recommendations to address the issue in the United States (Ch. 6).

1.2 Background and Motivation

This chapter serves as an introduction to pink slime journalism. In defining the phrase pink slime, it establishes what will be analyzed and researched throughout the thesis. It describes the online landscape of these sites and how casual viewers may come into contact with them. It surveys the current landscape of academic and journalistic research surrounding the topic and identifies where the gaps in research exist. By identifying the gaps, it explains how the subsequent chapters of this thesis will address them and what datasets and software will be utilized to do so. Finally, it analyzes the other countries which have dealt with pink slime campaigns.

1.2.1 Research Questions

The following questions are addressed within Chapter 1:

- What is pink slime?
- What conditions allowed pink slime to enter the news ecosystem?
- What do we know from previous research about pink slime?
    - Who owns these sites and with what goal?
    - Where are these sites shared?
    - What gaps exist in current research? How will they be addressed in this thesis?
- Which other countries have dealt with attacks to local news? What do they have in common?
- What data and tools will be used to answer these questions?

1.2.2 Evolution of Journalism Over Time

In order to understand how pink slime gained a footing in the American news diet, it's important to understand the history and evolution of the more unsavory news practices that led us here.

Yellow Journalism In the 1890s, New York City saw the rise of sensational journalism from publishers like William Randolph Hearst and Joseph Pulitzer who aimed to attract mass readership from previously un-marketed, lower-class audiences [66]. Their daily newspapers, printed cheaply and focusing on scandal and crime, were dubbed "yellow journalism" in homage to the cartoon of a yellow-colored child ("the Yellow Kid") seen in some of the publications and Figure 1.1[1]. Historian Frank Luther Mott defined yellow journalism has containing the following features: large font headlines proclaiming false excitement over irrelevant news, excessive use of photos that are often faked, interviews and stories that never took place or include pseudoscience and misinformation, Sunday comics, and a framing narrative of rooting for an underdog. [82].

These publications reached over a million New York City readers for daily publication [66], and succeeded at swaying public sentiment on international affairs. When a US battleship sunk

[1] toonopedia.com/yellow.htm

Figure 1.1: A comic of the Yellow Kid

in the Havana harbor in 1898, Hearst and Pulitzer printed sensational rumors and pushed anti-Spanish rhetoric to call for the Spanish-American War, which started weeks later [88].

Click Bait While yellow journalism was limited to its daily publication in the early 20th century, 100 years later the Internet would allow for a new application of the sensational news form - click bait. Click bait uses yellow journalism's sensational headlines in order to garner more "clicks" from social media platforms, like Facebook, where news stories are shared [10]. The headlines are designed to include key words that would intrigue the reader and leave an element of doubt - thus, prompting the reader to click on the article.

Scholars define the construction of these headlines as those containing discourse deixis and cataphora [25]. Discourse deixis previews a part of the story relative to the reader's current understanding, i.e. "*This* story will leave you speechless." Meanwhile, the cataphora element uses a word in the headline to forward-refer to something that is later explained in the headline, i.e. "When *he* returned from war, this veteran was in for the shock of his life." This type of sentence construction is designed to foster 'suspense and anticipation' [25].

Others have claimed to have cracked the code on click bait headlines with formulas such as "X things Y can teach you about Z" [67] wherein X is a number, Y is a cultural touchpoint, and Z is something worth knowing.

Social Media News Sharing In the digital era, news like click bait is being crafted for and shared via social media as opposed to traditional print medium. A 2021 survey by Pew Research Center shows that almost half (48%) of the adults in the US agree that they regularly get news from social media platforms, with the largest proportion of news consumption coming from Twitter, Facebook and Reddit[113]. Of these platforms vying for users' clicks and time, Facebook emerges as the dominating platform for news sharing in a study across multiple social media platforms [65]. Almost 70% of Americans get their news from the social media platform Facebook [52], but not all of this news is coming from quality news sources; 15% of referrals to fake news sites are coming from Facebook [56]. However, when it comes to news dissemination, research finds that smaller communities, like the subreddit dedicated to President Donald Trump have an outsized influence on larger, external communities like Twitter and Gab [114].

When it comes to local communities, Americans find posts made from local news Facebook Pages to be more trustworthy and relevant compared to posts from non-regional control Facebook Pages [101]. However, many owners of community weekly newspapers view the rise of news on Facebook as a direct threat to both ad revenue and the lack of community content sharing to their newspapers in favor of the digital medium [75].

Automated News Reporting While the phrase "automated reporting" may conjure up images of ChatGPT creating stories, the phenomenon has been around since the 1960s when meteorologists used punch cards and the weather science data processing service to generate written weather forecasts: "The more routine duties can be handled by computer, thereby freeing the meteorologist for the more challenging roles of meteorological consultant and specialist on high-impact weather situations." [49]. Later research relied on databases of sports, financial, or traffic data to utilize natural language generation to create news stories [42]. Some local newsrooms have used AI to automate tasks like natural disaster prediction alerts[100].

Interviews with journalists found that the automated reporting software felt too template-based and formulaic [72]. Researchers found that automated news reporting and human-written news have some similar elements - like timely news - but more profound differences. Human-written news articles include more negativity and impact as well as interpretation; meanwhile, the automated news articles are shorter and do not include human sources [107].

In 2015, the article "From Pink Slips to Pink Slime: Transforming Media Labor in a Digital Age" was published highlighting the dangers of news aggregation and "robot reporters" (a term used before the invention of large language models like ChatGPT).

The Creation of Pink Slime By the late 2010s, news had once again evolved, this time using elements of the previous news types. Hundreds of regional news sources that appear to be reliable local news have been spreading since 2019 [15]. These news websites consist of largely automated, low-quality partisan reporting and were nicknamed "Pink Slime" by journalist Ryan Smith in 2012 [108]. The term was coined as a comparison to the cheap fillers added to beef, here with cheap reporting being added to a self-reported news outlet. [37]. Smith recounted the story of how he was hired by Journatic and asked to write local news stories for towns across the USA [70]. He was editing stories originating from outsourced reporters in the Philippines who were not given credit for their writing and paid 40 cents per story. Journatic owner, Brian Timpone, came under fire when the story of his company broke on the popular podcast, This American Life [7]. Years later, Timpone would create Metric Media, a network of conservative-leaning pink slime sites that have more local news sites than Gannett and claimed to present data-driven news free of political bias; however, in the New York Times reporting, they learned that Metric

Media reporters were emailed that "clients want a politically conservative focus on their stories, so avoid writing stories that only focus on a Democrat lawmaker, bill, etc."[7].

Pink slime utilizes large datasets to create automated "local" reporting, includes numeric reporting like portions of click bait, and attempts to create a political undercurrent like yellow journalism through sharing of their partisan news stories on social media platforms.

1.2.3 The Significance of Pink Slime

Exploited Trust in Local Reporting American trust in local news organizations has remained higher than that of national news organizations [51]. To exploit the trust in local reporting, organizations like Metric Media LLC have created almost 1,000 local news sites [17]. While there is a dearth of authentic local reporting by local reports, it remains highly trusted, and creators of these networks are taking advantage of this trust. A New York Times investigation focusing on sites under Metric Media's control highlighted that while pink slime sites may seem insignificant on a national scale with tens of thousands of shares on social media, the focus on small towns require less readership for the impact to be felt [7]. Which is perhaps why 30% of the links pushed by the Russian troll farm, the Internet Research Agency (IRA), during the 2016 U.S. Presidential Election were to stories on local news websites (occasionally fake local news sites created by the Russians) [116]. Researchers conducting experiments into trust of news source find that individuals assessing pink slime news (particularly from Metric Media) rate it, along with authentic local news sources, as higher trustworthiness [92].

Financial Power to Influence Elections The organizations like Media Metric that control vast swaths of pink slime sites do not appear to have foreign ties [17], but they are currently financed by political candidates and political action committees with the hope of swaying election results. When speaking of threats to election integrity, Alex Stamos, director of the Stanford Internet Observatory, remarked "The issue ... is not going to be foreign interference. It's much more likely that legitimate domestic actors possibly operating under their own name — with LLCs or corporations with very shady funding that are not required to disclose what that funding is — are going to dominate the online conversation about the outcome of the election. [84]" Beyond elections, Metric Media mobilized its sites to run hundreds of articles pertaining to the reopen movement in Spring 2020 pertaining to regional covid lockdowns [19].

In discussing the 2022 U.S. Midterm elections, the co-CEO of NewsGuard (an organization dedicated to countering misinformation using online tools) Gordon Crovitz stated that: "Partisan sites masquerading as independent local news publishers are designed to fool readers into trusting untrustworthy sources of information, which has the result of reducing trust in all local news as people realize they've been targeted for biased reporting...The partisan groups secretly solicit millions of dollars from donors who are willing to undermine trust in news. The social media companies take advertising money designed to spread false and one-sided news coverage, in many cases microtargeting swing voters. These partisan donors and irresponsible social media companies have helped undermine trust in news. The resulting uncertain 'local news' environment cuts readership and advertising support for the legitimate news sites that need both now more than ever" [86]. Research out of NewsGuard went on to criticize the almost $4 million spent on ads run over 115 million times on Meta platforms in 2022 [12]. It comes as no surprise that NewsGuard created a nutrition label for the Metric Media sites claiming "A network of websites that falsely present themselves as locally based news sites. The sites do not disclose their conservative agenda, and much of the content is created by algorithms" [76].

Metric Media is not the only parent organization of pink slime, but it is the largest. Metric Media has several subnetworks all associated with and sharing IP space with the Metric Media sites [15]. Leaders at Metric Media affiliates have financial stakes in political action committees and non-profits. Timothy Dunn, the secretary of one of Metric Media affiliates is affiliated with groups focused on lowering taxes and limiting government (such as Defend Texas Liberty PAC, Texas Pubic Policy Foundation, Empower Texans, and Citizens for Self-Governance) [19]. John Tillman, secretary to another Metric Media affiliate, is involved in Illinois think tanks and non-profits like Illinois Policy Institute, Franklin News Foundation, Think Freely Media, and American Culture Project [19]. Finally, Dan Proft, founder of yet another Metric Media affiliate, runs political action committees including Liberty Principles PAC and People Who Play By the Rules PAC [19].

Reaching Large Audiences via Social Media Pink slime news sources do not exist in silos. Many of the known sources of pink slime have their own associated social media accounts on platforms like Facebook to amplify the spread of the messaging to the community (as the names of these sites frequently have the targeted community in the domain name). While 17.7% of visits to these sites are referred by Facebook, 3.2% are through Twitter [81]. Over 300,000 Twitter posts contain links to pink slime URLs [8]. A lack of pink slime on Reddit and 4chan has been documented by researchers who believe it is due to the communities on these platforms having higher media literacy [30].

1.2.4 Defining Pink Slime

Characteristics from Academic Research Those who have scraped news articles and analyzed the content of Metric Media sites (in 2020) have found that front-page stories have a median age of 81 days, 97% of the articles are auto-generated data stories, and those that made it to the front page pertained to state and national politics [104]. When analyzing the origin of the content, [59] found that pink slime sites largely copied their news articles from authentic local news sites or the Associated Press.

Research out of Stanford was the first to analyze news consumption of pink slime and found that during the 2020 U.S. Presidential election, 3.7% of American adults visited at least one pink slime site [81]. Furthermore, Biden supporters and people under 30 were more likely to visit these sites [81]. While living in a news desert was not a significant predictor of visiting a pink slime site, the distance from a visitor's self-reported location to that of the pink slime site was smaller (506 miles) than the distance from the visitor to the authentic local news sites he visited (598.1 miles) [81]. Surprisingly, the consumption study also found that living in a news desert was not a significant predictor of visiting a pink slime site [81]. Finally, the researcher found that while a minority of pink slime sites are about politics, those are the ones most visited [81].

When pink slime news articles are shared to social media platforms like Twitter, the text of the tweet contains the first sentence of the news article it links to *57%* of the time (compared to 27% for local news and less than 1% for national news tweets) [8].

Thesis Definitions of Pink Slime and 'Local' For the purposes of this research, pink slime is defined as media outlet websites that include the following:

- are run by national organizations
- have a local term in their name (i.e. "East Michigan News")
- are shared on social media platforms
- include aggregated and automated news reporting
- have a majority of their articles written by non-local reporters
- have a partisan leaning
- each website in a given organization is built via the same HTML template
- do not have a paywall

While the following incidents have taken place, they do *not* quality as pink slime per the definition set forth in this paper:

- generating fake social media accounts for local news organizations with no website presence
- hacking into known local news websites and running misleading stories

Furthermore, while we have seen pink slime occur in international instances, the focus of this paper is on the domestic pink slime. Therefore, the concept of something qualifying as 'local' in this thesis is defined as:

- a region in the *United States* that is either a state or a sub-community within the state
- sub-communities within the state are referred to as hyper-local but are also considered 'local'

Researchers' Differences in Definition Not all researchers who have used the term "pink slime" in research have utilized the same definition as outlined in this thesis. As automated news reporting was gaining footing in the early 2010s, Ryan Smith's 2012 podcast interview in which he coined the unsavory term was referring to inauthentic local news created cheaply, often overseas, to generate ad revenue, not political influence [109]. In response to this podcast, Canadian professor Nichole Cohen wrote *From Pink Slips to Pink Slime: Transforming Media Labor in a Digital Age* in 2015; while she did not set an explicit definition of what is or is not pink slime, her focus of pink slime was the dangers of automated, outsourced reporting (not exclusively mentioned as those targeting local news) [37]. It wasn't until 2019 when Tow Center for Digital Journalism fellow, Priyanjana Bengani, published her report on "partisan outlets masquerading as local news organizations" that researchers started defining pink slime as inauthentic local news websites set up for political gain, not financial profit [15]. Bengani followed up her report by publishing a list of news domains and their parent organization which serve as a seed dataset to the research in this thesis [45].

Organization	Sub-Specialty	Number of Websites
Metric Media	Metric Media	977
	Metro Business	56
	LGIS	35
	Record	11
	Franklin Archer	11
Local Report		49
Star		11
Courier		8
American Independent		5

Table 1.1: Ownership of pink slime sites, colored by U.S. political leaning

Parent Organizations Five parent organizations control over 1,000 pink slime news domains with the number of domains illustrated in 1.1. While these numbers are smaller than the Tow Center for Digital Journalism's published list [45], that is because the sites that target locals outside of the United States are excluded from this research. The largest organization, Metric Media, along with the Star News Network have conservative political leanings while the other three share liberal-leaning news.

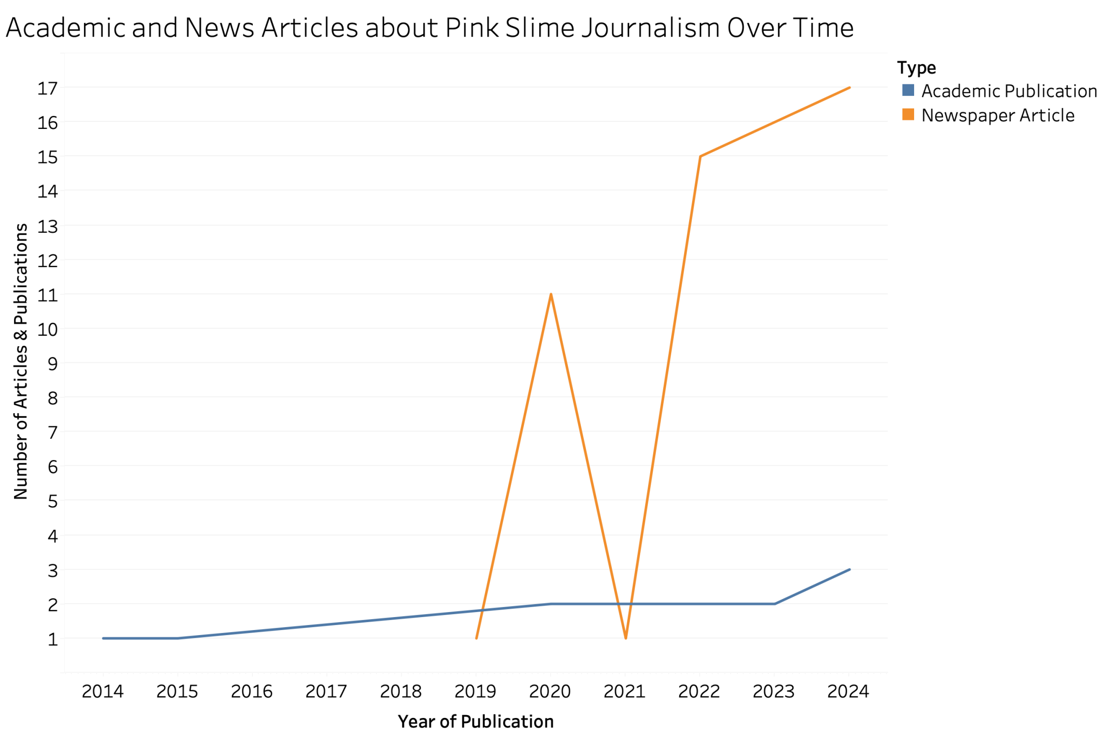


Figure 1.2: News Articles and Journal Publications Mentioning the Pink Slime Phenomenon

Published Research Pink slime has been largely under-studied in academic settings, but it has gained traction in newsrooms. To further illustrate this dichotomy, I plotted the number of academic publications and journal articles pertaining to this topic over time. Using the GNews Python Package [2], all of the news articles mentioning "pink slime journalism" were collected from Google News. After removing the few referring to the meat byproduct instead of the journalism phenomenon, 61 articles remained from 2019 through July of 2024. When looking at academic publications, the phrase "pink slime journalism" was used to search research articles on Semantic Scholar[3]. While there were 129 results, the vast majority of them were from the Biology and Agriculture research areas and had to be removed for pertaining to the pink slime meat byproduct. After removing the meat-related ones, only 9 publications remained (2 of which were first-authored by the author of this thesis). Per Fig 1.2, very few peer reviewed articles have been published on the topic of pink slime. Publications prior to 2019 focused on the auto-generation of news and conditions that allowed pink slime to gain a footing in the American news media ecosystem [37]. Most of the articles published have been by news outlets who are appalled by the emergence of this new threat; however, some peer reviewed publications on the topic exist.

While there have been a few publications on how pink slime news is consumed by individuals surrounding an election [81] [29] and who is funding it [12] [18], little remains known about the impact these news sites have during non-election years and where the funding from these parent organizations is going. Furthermore, no research has been published documenting the spread of these news sites on different social media platforms. Chapter 2 of this thesis addresses all of those gaps.

Some research has analyzed what movements the parent organizations are supporting [81] [16]; however, we are left unaware of how to quantitatively compare this support and the maneuvers used to authentic local news. Chapter 3 of this thesis answers this question.

One researcher has found that new pink slime sites can be uncovered through an expensive and tedious process of searching NewRelic IDs and Quantcast IDs [15]. A free and less intensive method of finding these sites is proposed and tested in Chapter 4.

PBS published lessons plans aimed at school-aged children to teach them media literacy on the subject of pink slime [5]. The effectiveness of those plans have not yet been tested, but Chapter 5 performs this testing. Additionally, user studies looked at humans viewing pink slime sites to conclude that there are negative impressions of these sites after repeated exposure [102]. What remains absent from this study is how the pink slime sites rate with regards to trustworthiness in comparison to authentic local news. Chapter 5 also addresses this concern.

Finally, pink slime is a term for the hijacking of local news in the United States, but this phenomenon has been seen in other countries and regions around the world [6] [80] [13]. In the final chapter (6), policy recommendations from the international incidents explored in Chapter 1 and based on the previous 5 chapters are recommended.

Research findings and gaps are referenced in Table 1.2 as well as which chapter of this thesis will address those gaps.

[2] pypi.org/project/gnews/
[3] semanticscholar.org

Current Research Tells Us	Unanswered Questions	Where this Thesis Answers Those Questions
This phenomenon is happening internationally	How is this similar to the situation in the United States?	Chapter 1
Who is consuming pink slime (exclusively during 2020 U.S. Presidential Election)	What are the network features of those sharing pink slime? What are they doing during the U.S. Midterms Election?	Chapter 2
Who is funding pink slime	Where are they spending that money? How do they decide where to spend the money?	Chapter 2
What movements some organizations are supporting, common topics of articles	What maneuvers are they utilizing to show support for candidates, movements, and topics, specifically around elections?	Chapter 3
Pink slime sites can be found through a tedious process of searching NewRelic IDs and Quantcast IDs.	How can we sift through websites to quickly find new sources of pink slime for free?	Chapter 4
Lesson plans have been crafted to teach humans what pink slime is.	Are these lesson plans effective?	Chapter 5
User studies have found that individuals viewing the sites repeatedly form negative impressions of pink slime.	How does human trust of pink slime compare to that of authentic local news?	Chapter 5
International groups are looking into ways to counter pink slime.	What policy should be enacted to counter it?	Chapter 6

Table 1.2: Current Research Gaps

1.3 International Local News Hijacking

While the phenomena of pink slime is new in the United States, the concept of infiltrating local news websites to share political propaganda is not new nor limited to the USA. In this section, I've compile all other known cases of this untoward form of local journalism as an information operation campaign. By understanding the ways in which bad actors are able to infiltrate the local news ecosystem for political gain, we can learn what to look for when the attacks are domestic and consider future policy action to combat them.

1.3.1 Methodology

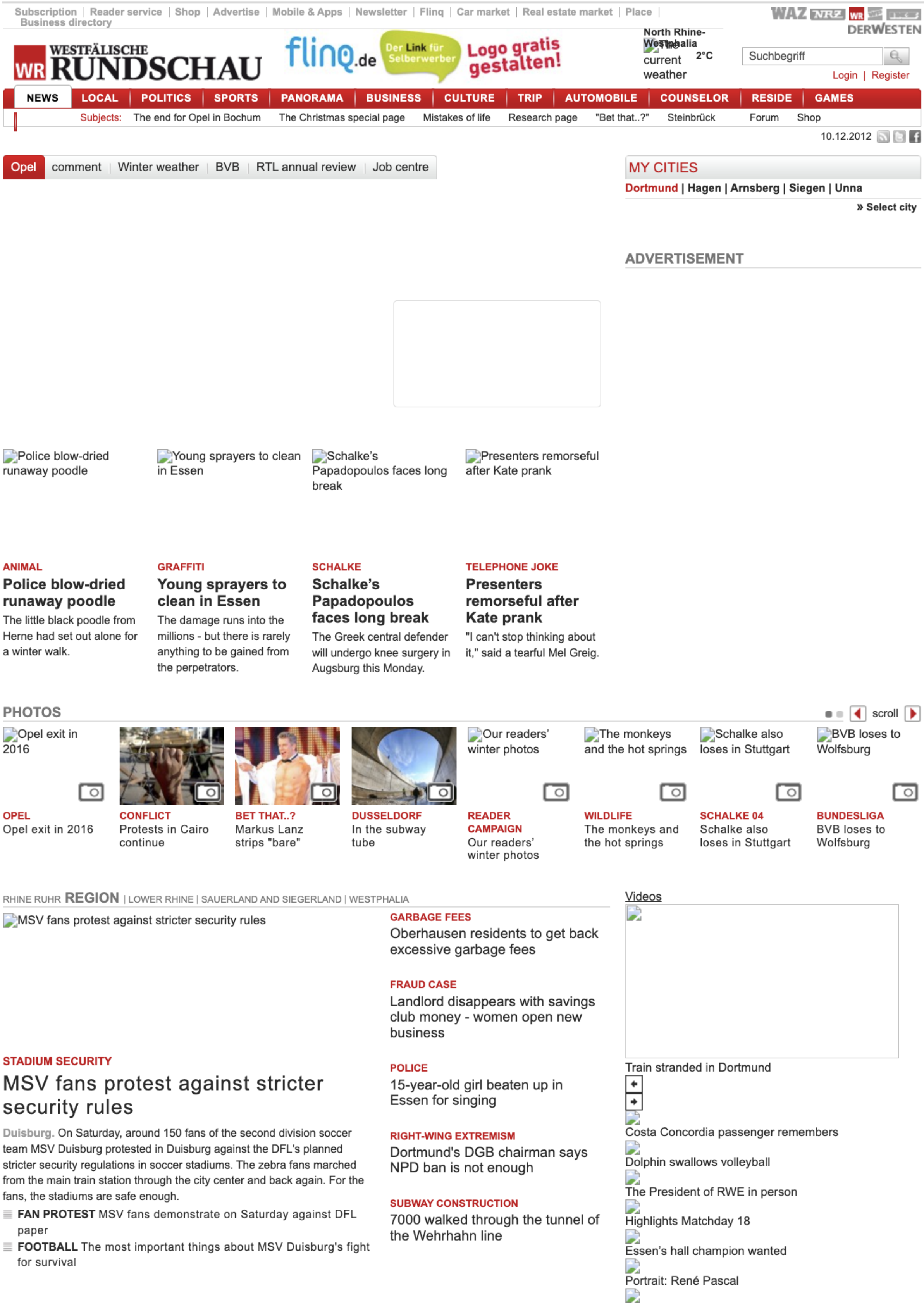
In order to find examples of pink slime abroad, I chose to focus on keywords other than "pink slime" since it is an American-ized term. Instead, I searched for terms that captured the essence of the creation of an information operation campaign going after trusted local news sources with the phrases: "fake local news", "hijacked local news" and "infiltrating local news". These phrases were searched on Google Scholar, Google News, and the social media platform X; it was important to expand beyond academic publications, as many of the articles about these campaigns were done by fellow news reporters. This selection criterion was to find instances of deceitful local news websites by non-local reporters that were trying to influence a specific community. Overall 7 such instances were found and are described in the sections below.

1.3.2 International Instances of Local News Hijacking

In the paragraphs below, each campaign is described in greater detail in chronological order.

2007: Germany's 'Zombie' Papers The earliest known example of inauthentic local news dates back to 2007 in Germany. Unlike our other examples, this one has a less-sinister and, unfortunately, more practical origin story. Due to financial pressures, local newsrooms began laying off their reporters, writers, and staff [13]. Eventually, many of these newsrooms had no physical presence in a region but maintained their websites which largely contained stories copied by competitors and money-generating ads [13]. Citizens who were aware of these "zombie" newsrooms grew distrustful of local news, a painful consequence [13]. Figure 1.3 shows what one of these sites looked like before it was taken over by non-local reporters and Figure 1.4 shows what the same news site looks like a few years after the takeover. These sites still exist throughout Germany and may have served as a template for later cases of pink slime, showing organizations that local reporters and a physical presence are not required to run a local news agency [13].

2010: A Pro-India Campaign in the European Union In 2019, the European Union Disinfo Lab uncovered an Indian campaign to influence the European Union by creating 265 "local" news sites within 65 countries, dating back to 2010 [3]. The websites (most of which were named after extinct, real local newspapers, the German "zombie" approach) shared anti-Pakistan content on their websites as well as associated Twitter accounts [3], as seen in Figures 1.5 and 1.6.



Subscription | Reader service | Shop | Advertise | Mobile & Apps | Newsletter | Flinq | Car market | Real estate market | Place | Business directory

WAZ NRZ WR DERWESTEN

WESTFÄLISCHE RUNDSCHAU

flinq.de Der Link für Selberwerber Logo gratis gestalten!

North Rhine-Westphalia current weather 2°C

Suchbegriff

Login | Register

NEWS | LOCAL | POLITICS | SPORTS | PANORAMA | BUSINESS | CULTURE | TRIP | AUTOMOBILE | COUNSELOR | RESIDE | GAMES

Subjects: The end for Opel in Bochum | The Christmas special page | Mistakes of life | Research page | "Bet that..?" | Steinbrück | Forum | Shop

10.12.2012

Opel | comment | Winter weather | BVB | RTL annual review | Job centre

MY CITIES

Dortmund | Hagen | Arnsberg | Siegen | Unna

» Select city

ADVERTISEMENT

Police blow-dried runaway poodle

Young sprayers to clean in Essen

Schalke's Papadopoulos faces long break

Presenters remorseful after Kate prank

ANIMAL

Police blow-dried runaway poodle

The little black poodle from Herne had set out alone for a winter walk.

GRAFFITI

Young sprayers to clean in Essen

The damage runs into the millions - but there is rarely anything to be gained from the perpetrators.

SCHALKE

Schalke's Papadopoulos faces long break

The Greek central defender will undergo knee surgery in Augsburg this Monday.

TELEPHONE JOKE

Presenters remorseful after Kate prank

"I can't stop thinking about it," said a tearful Mel Greig.

PHOTOS

scroll

Opel exit in 2016

Our readers' winter photos

The monkeys and the hot springs

Schalke also loses in Stuttgart

BVB loses to Wolfsburg

OPEL
Opel exit in 2016

CONFLICT
Protests in Cairo continue

BET THAT..?
Markus Lanz strips "bare"

DUSSELDORF
In the subway tube

READER CAMPAIGN
Our readers' winter photos

WILDLIFE
The monkeys and the hot springs

SCHALKE 04
Schalke also loses in Stuttgart

BUNDESLIGA
BVB loses to Wolfsburg

RHINE RUHR **REGION** | LOWER RHINE | SAUERLAND AND SIEGERLAND | WESTPHALIA

MSV fans protest against stricter security rules

STADIUM SECURITY

MSV fans protest against stricter security rules

Duisburg. On Saturday, around 150 fans of the second division soccer team MSV Duisburg protested in Duisburg against the DFL's planned stricter security regulations in soccer stadiums. The zebra fans marched from the main train station through the city center and back again. For the fans, the stadiums are safe enough.

FAN PROTEST MSV fans demonstrate on Saturday against DFL paper

FOOTBALL The most important things about MSV Duisburg's fight for survival

GARBAGE FEES
Oberhausen residents to get back excessive garbage fees

FRAUD CASE
Landlord disappears with savings club money - women open new business

POLICE
15-year-old girl beaten up in Essen for singing

RIGHT-WING EXTREMISM
Dortmund's DGB chairman says NPD ban is not enough

SUBWAY CONSTRUCTION
7000 walked through the tunnel of the Wehrhahn line

Videos

Train stranded in Dortmund

Costa Concordia passenger remembers

Dolphin swallows volleyball

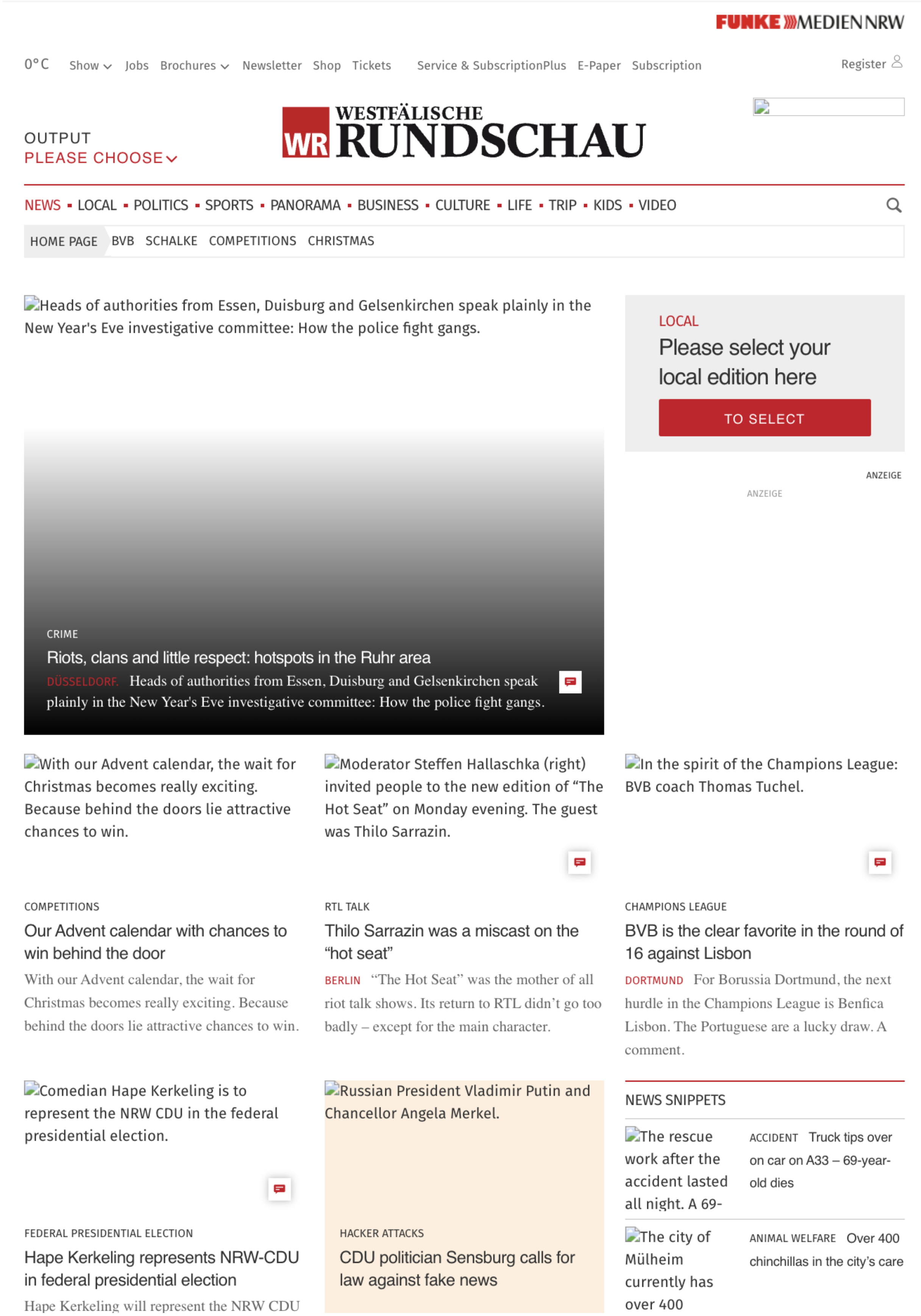
The President of RWE in person

Highlights Matchday 18

Essen's hall champion wanted

Portrait: René Pascal

Figure 1.3: A 2012 translated screenshot of one of the German "zombie" newspapers before it was taken over by non-locals.


FUNKE MEDIENNRW

0°C Show Jobs Brochures Newsletter Shop Tickets Service & SubscriptionPlus E-Paper Subscription Register

OUTPUT
PLEASE CHOOSE

WR WESTFÄLISCHE RUNDSCHAU

NEWS · LOCAL · POLITICS · SPORTS · PANORAMA · BUSINESS · CULTURE · LIFE · TRIP · KIDS · VIDEO

HOME PAGE BVB SCHALKE COMPETITIONS CHRISTMAS

Heads of authorities from Essen, Duisburg and Gelsenkirchen speak plainly in the New Year's Eve investigative committee: How the police fight gangs.

CRIME
Riots, clans and little respect: hotspots in the Ruhr area
DÜSSELDORF. Heads of authorities from Essen, Duisburg and Gelsenkirchen speak plainly in the New Year's Eve investigative committee: How the police fight gangs.

LOCAL
Please select your local edition here
TO SELECT

ANZEIGE

ANZEIGE

With our Advent calendar, the wait for Christmas becomes really exciting. Because behind the doors lie attractive chances to win.

COMPETITIONS
Our Advent calendar with chances to win behind the door
With our Advent calendar, the wait for Christmas becomes really exciting. Because behind the doors lie attractive chances to win.

Moderator Steffen Hallaschka (right) invited people to the new edition of "The Hot Seat" on Monday evening. The guest was Thilo Sarrazin.

RTL TALK
Thilo Sarrazin was a miscast on the "hot seat"
BERLIN "The Hot Seat" was the mother of all riot talk shows. Its return to RTL didn't go too badly – except for the main character.

In the spirit of the Champions League: BVB coach Thomas Tuchel.

CHAMPIONS LEAGUE
BVB is the clear favorite in the round of 16 against Lisbon
DORTMUND For Borussia Dortmund, the next hurdle in the Champions League is Benfica Lisbon. The Portuguese are a lucky draw. A comment.

Comedian Hape Kerkeling is to represent the NRW CDU in the federal presidential election.

FEDERAL PRESIDENTIAL ELECTION
Hape Kerkeling represents NRW-CDU in federal presidential election
Hape Kerkeling will represent the NRW CDU

Russian President Vladimir Putin and Chancellor Angela Merkel.

HACKER ATTACKS
CDU politician Sensburg calls for law against fake news

NEWS SNIPPETS
The rescue work after the accident lasted all night. A 69-
ACCIDENT Truck tips over on car on A33 – 69-year-old dies
The city of Mülheim currently has over 400
ANIMAL WELFARE Over 400 chinchillas in the city's care

Figure 1.4: A 2016 translated screenshot of one of the German "zombie" newspapers after it was taken over by non-locals.

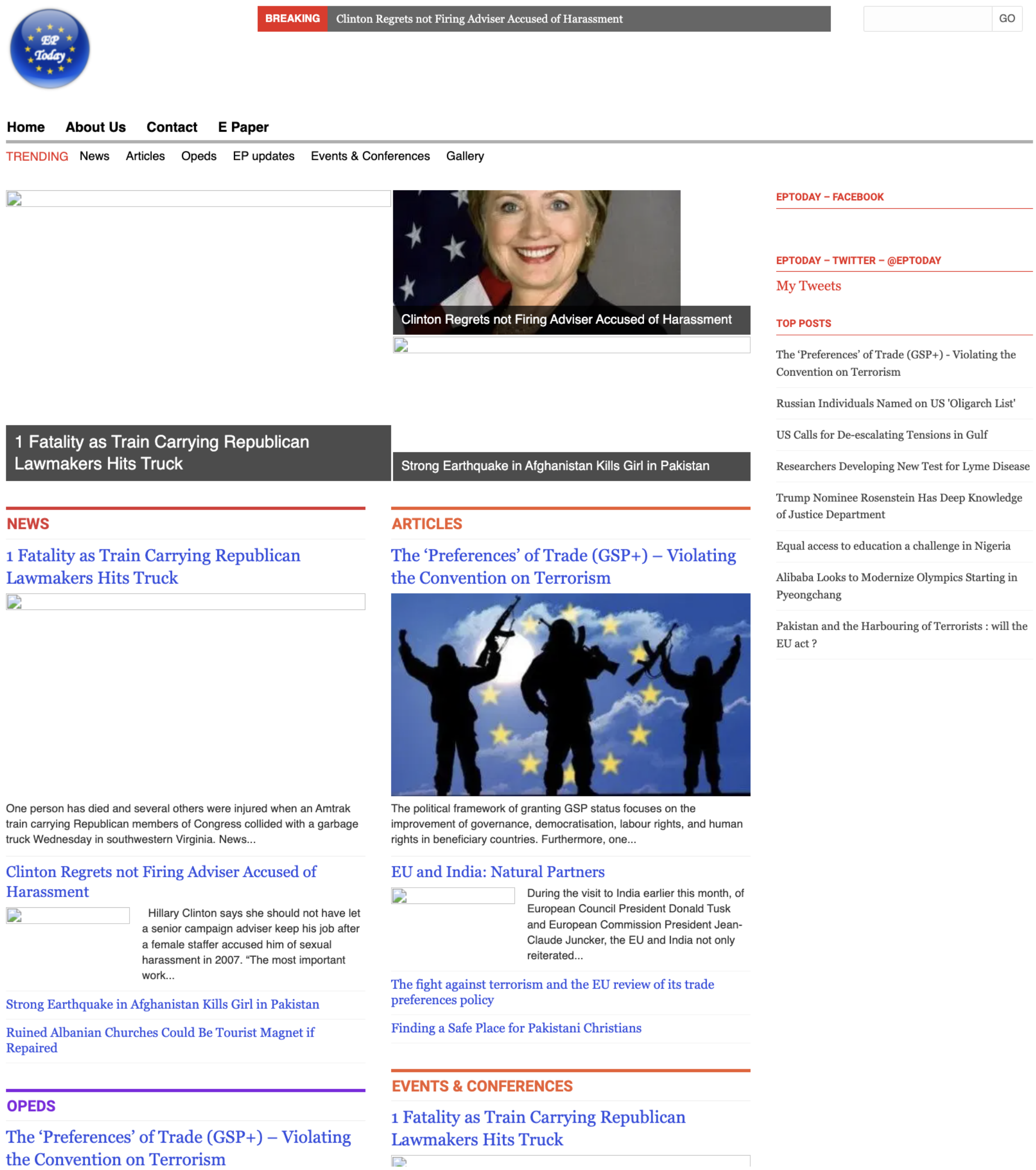


Figure 1.5: A 2018 screenshot from EP Today, a fake local news website established by India to influence the European Union. One headline states "EU and India: Natural Partners."

Figure 1.6: EP Today's "About" section, highlighting that its audience is the European Parliament and insisting that it is operating out of Brussels, Belgium (one of the two locations where the European Parliament convenes).

2018: Romania Driving Canadian Misinformation Initially, Canadians believed that there was a new local news site about updated driver laws in the country [31]; however, the scheme unearthed was much more sinister and involved actors from Romania creating twelve "local" news websites in Canada, as shown in Figure 1.7 [105]. These websites used WordPress templates to share misinformation about recalls, immigration, and driving laws [31] [105]. The articles garnered much interaction on social media platforms like Facebook [31]. While the sites did not venture into promoting political agendas, those researching the phenomena believe that was a next step [105].

2019: The Great Chinese Paperwall Against the World A 2024 discovery by The Citizen Lab in Toronto found a pro-Beijing campaign of 123 "local" news websites in 30 countries operated by a PR Firm in China [44]. These sites were largely created using WordPress and contained both local and national news stories aggregated from other news sources so as to not draw suspicion with its own original content which contained targeted attacks and conspiracy theories [44]. An example of one of their sites targeting residents of Venice, Italy and the information it shared about Chinese President Xi Jiping can be found in Figures 1.8 and 1.9.

2023: Russia Infiltrates Israel During the Russian-Ukraine conflict, the Russian government worked to change the narrative in the Middle East by creating three "local" news websites in Israel [64]. These news sites mimic more well known Israeli news sites but include anti-Ukrainian propaganda [64]. Furthermore, the Russians websites ran fake stories accusing U.S. President

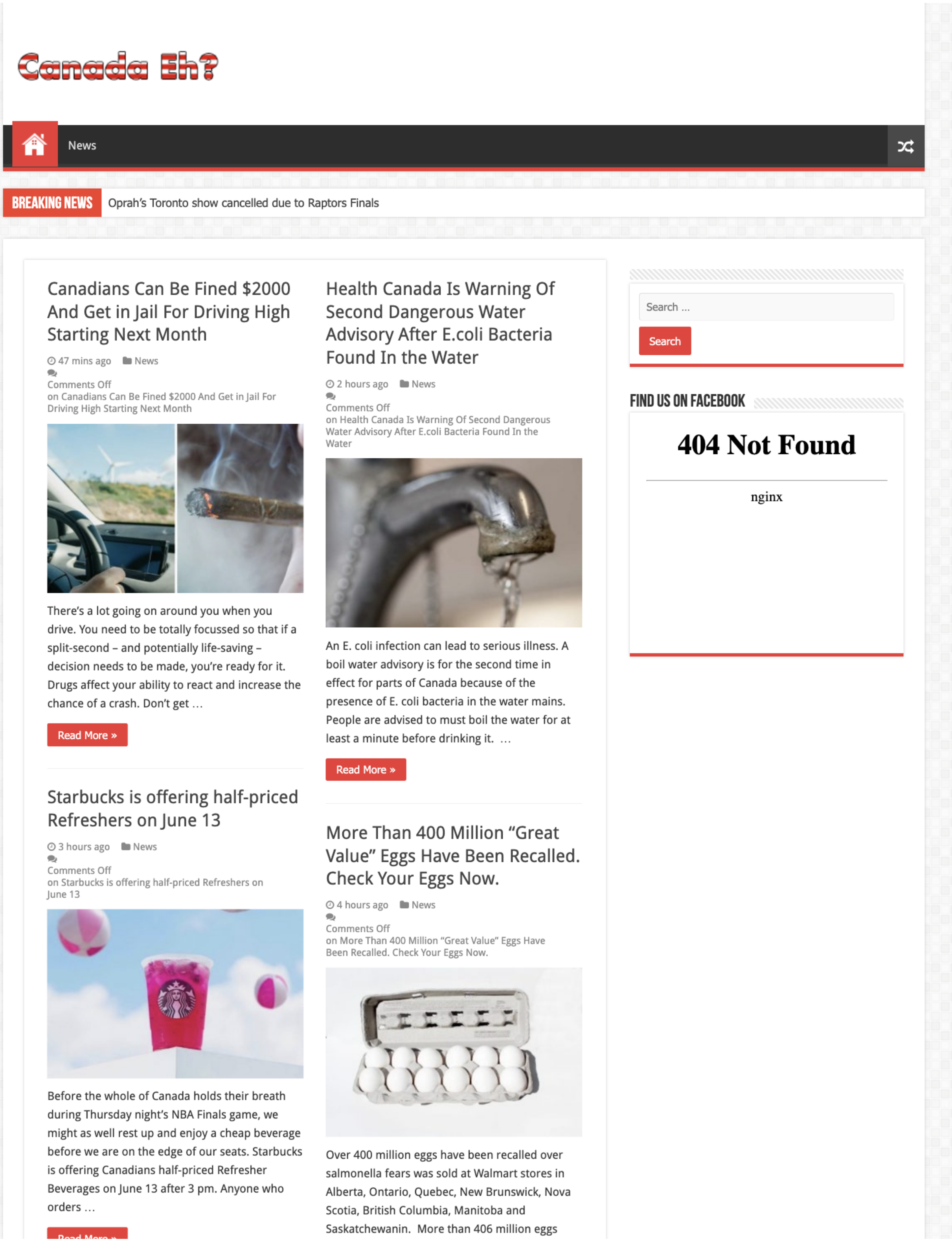


Figure 1.7: A screenshot of the "Canada Eh" news site created by Romanians.

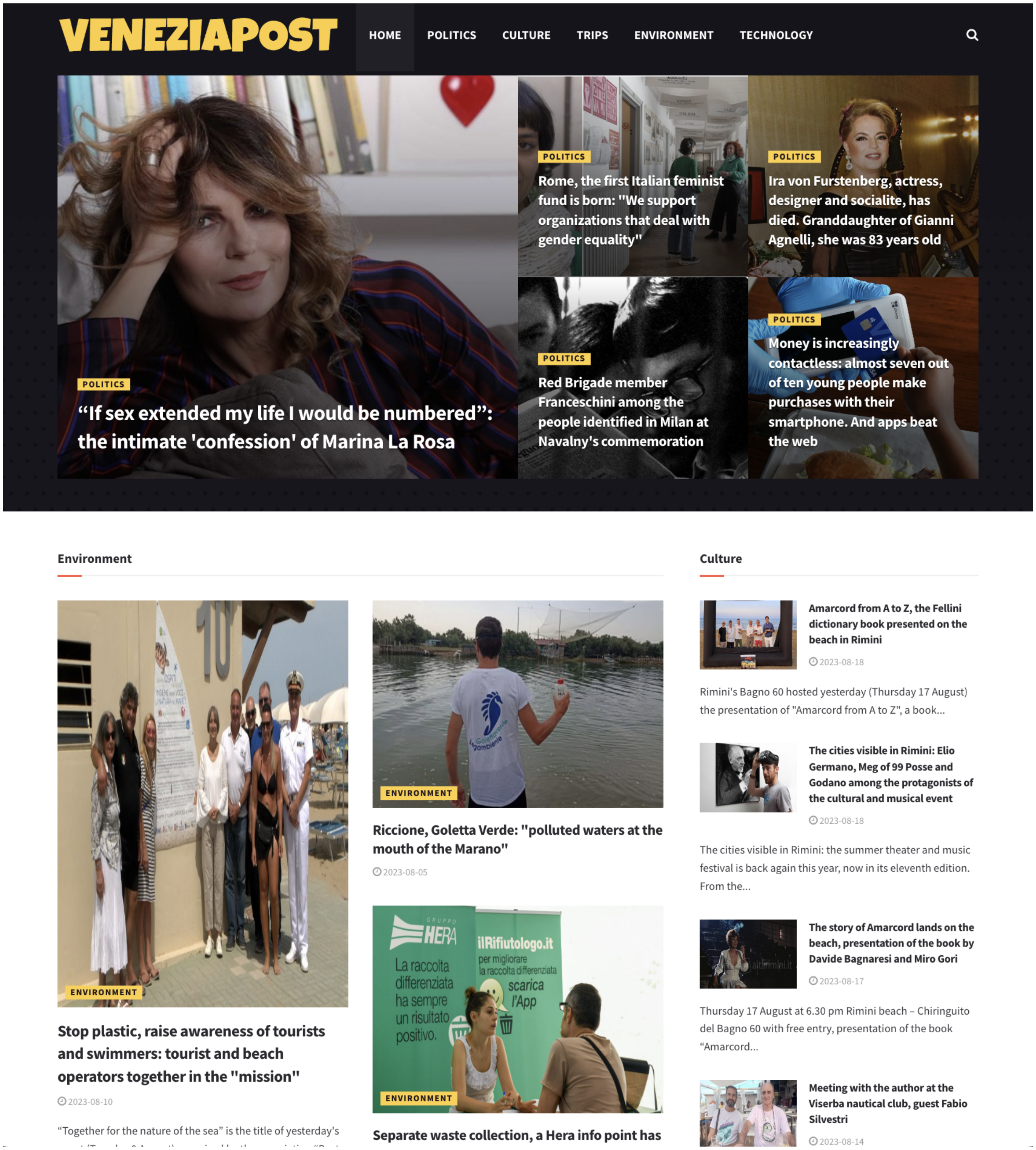


Figure 1.8: A translated screenshot of the "Venezia Post" news site created by Chinese to target Venice, Italy.

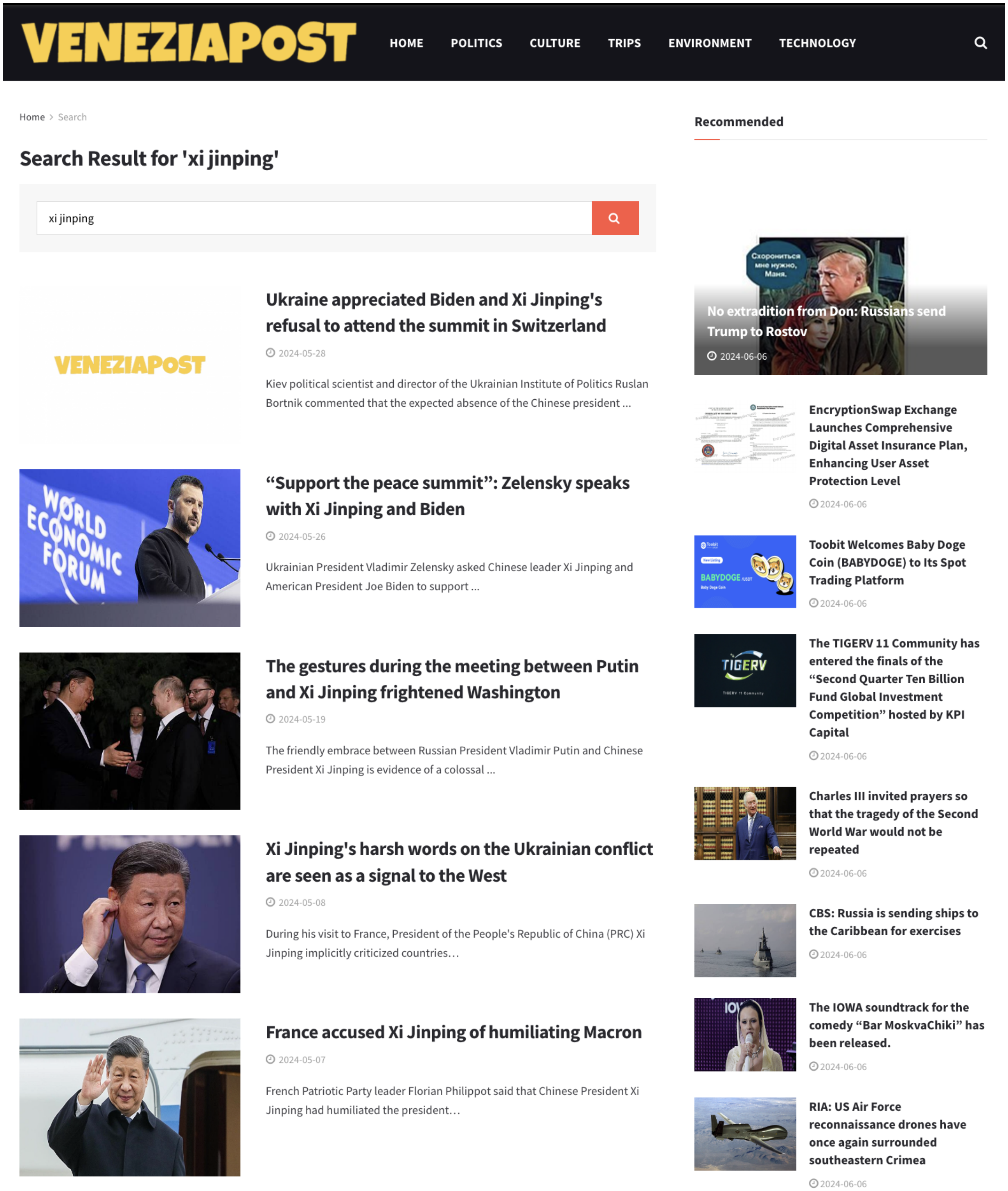


Figure 1.9: A screenshot of the "Venezia Post"'s responses when searching Chinese President Xi Jiping.

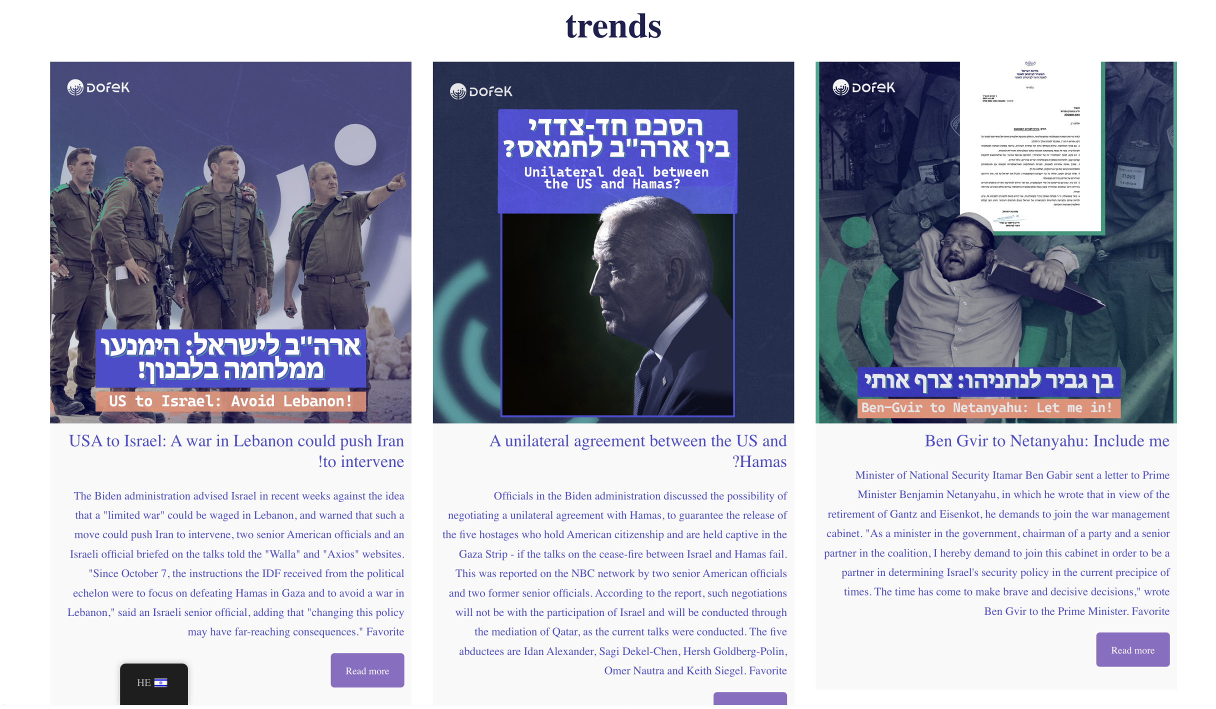


Figure 1.10: An auto English-translated version of Dofek.TV, the Lebanon-backed ‘Israeli’ News Site.

Biden of trying to “topple the Israeli government” [64], further attempting to create a wedge between Israel and the United States.

2024: Lebanon *Also* Infiltrates Israel As tension in Israel heated up, so did the influence campaigns. Lebanon created “Dofrek TV”, a website designed to share news with Israelis, as seen in Figure 1.10 [58]. In only a few days, news from the site was shared on many social media platforms [58]. Despite claiming to be a voice for Israelis, the messaging is anti-Israel, heavily critical of Prime Minister Netanyahu, and pro-Palestinian [58]. The majority of the news content is lifted directly from other Israeli news outlets [58].

2024: Russia Comes for America In the most recent instance, Russia created 4 news sites that appeared to be local news for 4 major U.S. cities - D.C., New York, Chicago, and Miami - in an attempt to influence the upcoming 2024 U.S. Presidential Election [85]. These sites used WordPress templates, and the Chicago site was akin to another instance of a “zombie” paper as the Chicago Chronicle was a reputable local newspaper from 1895-1907 (sadly too early to register a news domain on the World Wide Web) that shuttered due to low profits before becoming a ploy in Russian propaganda [85]. The group who discovered these sites wisely surmised that “The purpose is not to fool a discerning reader into diving deeper into the website, let alone subscribing. The goal instead is to lend an aura of credibility to posts on social media spreading the disinformation”[85], a goal the group accomplished. While many of the articles were lifted from other national news sources, some of the original content still included ChatGPT prompts within the text. An example of one of their sites can be seen in Figure 1.11.

Commonalities of the International Pink Slime Examples When analyzing all of the international news hijacking instances together, a few themes emerge. First, nearly all of the campaigns involve zombie papers. These zombies are actual local news websites that go out of business, but their news domain is "resurrected" from the dead for nefarious purposes. Second, news articles from these sites are all shared on social media. This drives home the point that the objectives are that the viewers see these headlines in their social media feeds, note that the linked domain sounds familiar or trustworthy, and believe the reporting. Third, authentic news stories from various sources are copied onto the front page to lend credibility to the other reporting. Finally, most of these sites are generated using web templates from WordPress. In order to scale their operations, the invading group will copy and paste templates throughout all of the "local" news sites they generate.

The map in Figure 1.12 shows all of the creating actors and regions attacked in the seven campaigns discovered. Many of these campaigns, like the ones from China and India, targeted dozens of countries, so a more detailed network was created to illustrate all of the countries who fell victim to fake local news campaigns. The network in Figure 1.13 shows all of the creating actors and individual countries attacked in the seven campaigns discovered.

To understand the differences between those countries creating fake local news and those who were victims of these attacks, I analyzed the differences in these countries' freedom of the press, democracy, and cyber-security. For freedom of the press, I used the Reporters Without Borders World Press Freedom Index 2024 who bases the index upon "a score ranging from 0 to 100 that is assigned to each country or territory, with 100 being the best possible score (the highest possible level of press freedom) and 0 the worst" [26]. To measure democracy, an index by the Economist Intelligence Unit (2006-2023) was used; it scores countries based on their ability to fairly elect their political leaders and enjoy civil liberties, ranging from 0 to 10 (most democratic) [61]. Finally, as a measure of cyber-security, the Cybersecurity Exposure Index (CEI) was analyzed to compare the countries' exposure to cyber crime; this index ranges from 0 to 1 (higher exposure) [78]. The results can be found in Table 1.13, with a final column added to specifically analyze those countries who have been repeatedly targeted in these attacks. Overall, countries creating fake local news websites enjoy less freedom of the press and democracy and are more subject to cyber crime than their victims. Furthermore, those countries that have been victims of multiple fake local news campaigns enjoy the highest freedom of the press and democracy and lowest exposure to cyber crime. This may suggest that countries with the most freedom of the press and ability to criticize their democracies are more prone to attacks of harmful actors claiming to be "press" to enjoy the many liberties that those countries afford to such members.

Comparing International News Hijacking to Domestic Pink Slime The main focus of this thesis is on the *domestic* pink slime attacking the Untied States by American organizations. However, the creators of these organizations may have taken a few pages out of the playbook from international local news invaders. Some also include actual news stories from real reporters at organizations like the Associated Press to fill out the remainder of content on their home pages. Just like the international groups streamline the operations, each of the pink slime parent

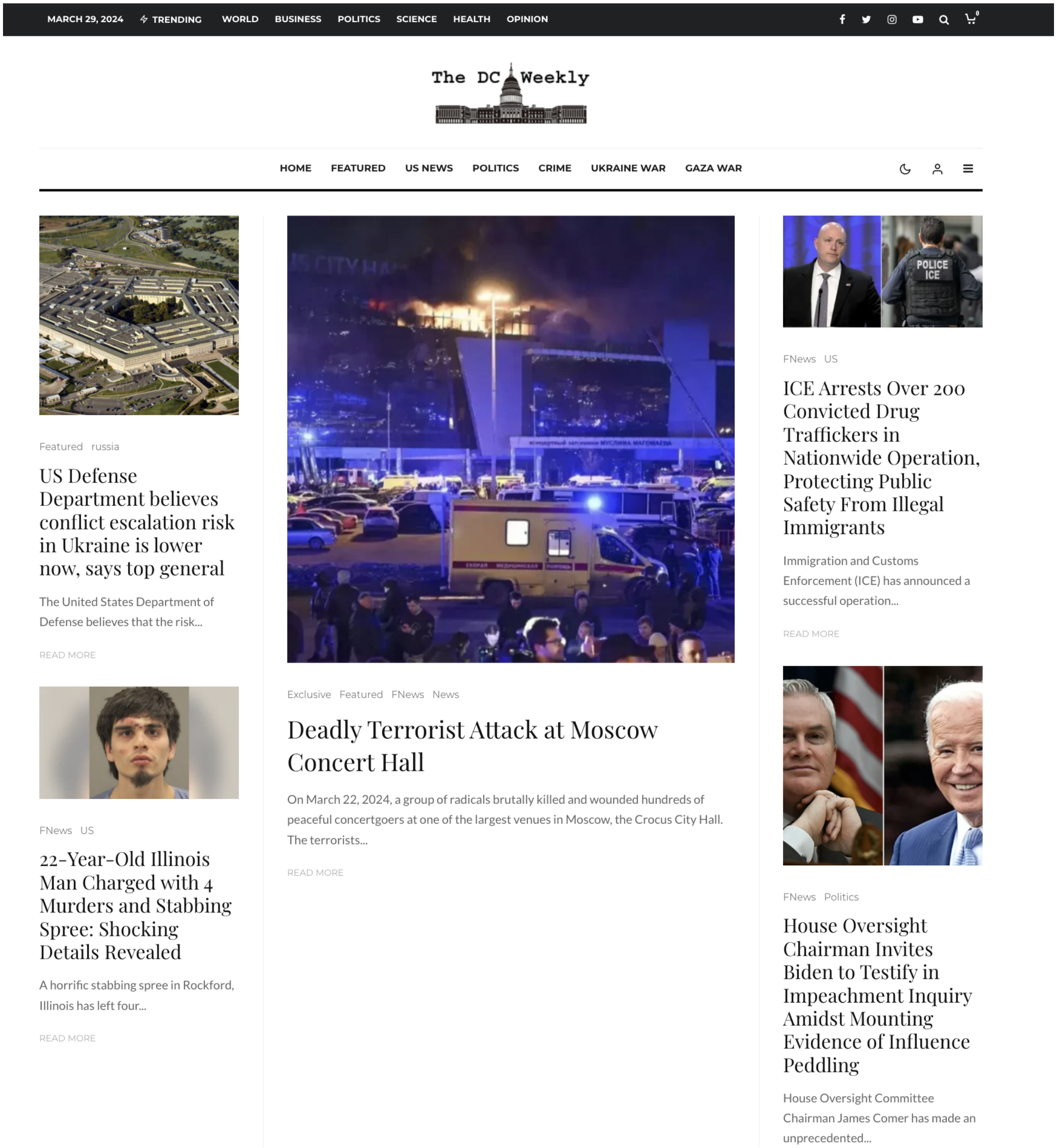


Figure 1.11: A screenshot of the D.C. Weekly website run by Russia

Figure 1.12: A map representing the sources of inauthentic local news and the regions where they created the "local" news. The geographic regions the arrows point towards are the ones that were infiltrated by the source nodes. Image generated using the ORA network visualization software [34].

	Fake Local News Creating Countries	All Fake Local News Victim Countries	Repeat Fake Local News Victim Countries
Press Freedom Index	46.5	58.2	72.4
Democracy Index	5.06	6.38	8.06
Cybersecurity Exposure Index	0.481	0.412	0.262

Table 1.3: Average index values for creators and victims of fake local news attacks.

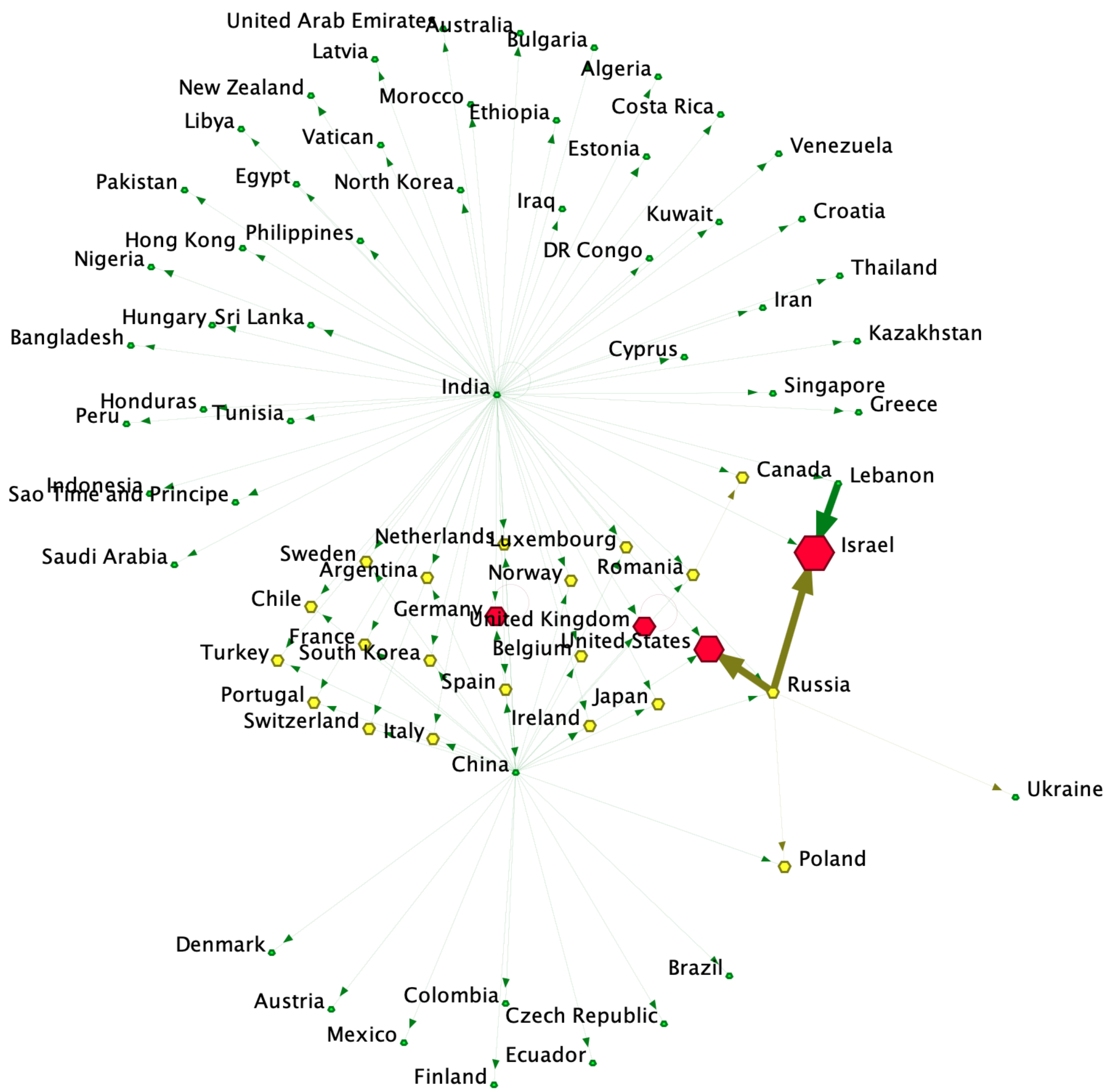


Figure 1.13: A network visualizing all of the individual countries creating and being attacked by local news hijacking campaigns. Red nodes were attacked at least 3 times, yellow nodes were attacked 2 times, and the green nodes were only attacked 1 time. Image generated using the ORA network visualization software [34].

organizations uses the same website HTML template throughout its multi-state operations. Much like the international campaigns, domestic pink slime has focused their attacks on spreading their news articles and political ads to social media platforms like Facebook.

There are two primary differences between the international campaigns and the domestic ones presented in this thesis. First, the U.S. pink slime organizations have yet to create zombie websites of previously legitimate local news sites. While they are registering news domains that *sound* like local news, they have not attempted to fool their readers into thinking they are the owners of some of their previously favorite local news sources (a rare moral win for these companies). The second difference is the reliance on APIs to generate content. In the American spirit of innovation, creators of U.S. pink slime have been able to further streamline the process of owning hundreds of news sites by finding datasets that other organizations are responsible for updating (such as unemployment data in a region, budget reports, etc.) and auto-reporting these numbers as new articles for their homepage without wasting journalistic manpower.

Given that one of the primary objectives (both internationally and domestically) of going through the effort to create hundreds or thousands of news websites is to have their headlines shared on social media, it's important to understand how these articles are being shared on social media, by whom, and what impact they are having. The remaining chapters of this thesis rely on social media news sharing data so that we can answer these questions.

1.4 Data

This thesis largely analyzes how users share, interact and engage with pink slime sites on various social media platforms. Table 1.4 summarizes where each of these data sources are used throughout the chapters of the thesis, and they are explained in greater depth in the subsections below.

1.4.1 News Type Datasets

Throughout this thesis, the "Four News Types" are used to describe the different types of news as defined in [71]. These news types are categorized based on their credibility and scope. Pink slime, the primary news type discussed in this thesis, is low in credibility and local in scope. The list of sites that qualify as pink slime are discussed in the **Pink Slime Dataset**. Low Credibility News is low in credibility and national in scope; these sites are listed in the **CASOS Thesaurus Dataset**. Real News, which is high in credibility and national in scope is also labeled from sources within the **CASOS Thesaurus Dataset**. Finally, local news is high in credibility and local in scope; a list of these sites is found through the **Local News Dataset**. A visual representation of this can be found in Figure 1.14.

CASOS Thesaurus Dataset consists of posts from the media thesaurus compiled by the CASOS University Center at Carnegie Mellon University. The media thesaurus has been compiled from multiple publicly available lists of news media URLs and media organizations' Twitter accounts: Media Bias/Fact Check [1] lists many news sites and rates how factual and credible the reporting is for many; the George Washington University Dataverse [73] has a list of over 9600

		Chapter				
Dataset	**Approximate Size**	**1**	**2**	**3**	**4**	**5**
Facebook Group and Page Posts to Pink Slime Domains	1.3 million posts					
Facebook Pink Slime Ad Data	5,000 ads					
Scraped Pink Slime Webpages	35,697 articles					
Facebook/Reddit/Twitter All News Types Midterms 2022	1.4 million posts					
CrowdTangle Facebook COVID 2020	600,000 posts					
CrowdTangle United Kingdom 2024 Election	1.2 million posts					
March 2020 Twitter COVID Dataset	1.2 million Tweets					
Trident Juncture 2018 Twitter Dataset	230,000 Tweets					

Table 1.4: Summary of data used in the chapters of the thesis

Twitter accounts for media organizations, derived from over 160 million tweets between 2016 and 2020; there is also a Github repository [50] of unreliable, misleading, and/or low credibility news sources that includes lists from Snopes Field Guide, Melissa Zimdars' OpenSources, Wikipedia, and others. There is often overlap between these sources, particularly for the less factual news outlets; to resolve any conflicts that emerge between the sources, the thesaurus errs on the side of not labeling a news source in question as low credibility news. From this thesaurus, the labels of low credbility and real news domains are utilized.

Pink Slime Dataset consists of domains from the Tow Center of Digital Journalism's study of pink slime and published on Github [45]. While not all of these sites ended up meeting the definition of pink slime set forth by this thesis, this was used as a baseline for data acquisition from APIs. The targeted state column was also utilized for geospatial analysis.

Local News Dataset contains a list of domains that are classified as local news in the United States. They can be found on Github [112]. Additionally, a set of authentic local news sites owned by larger companies are compiled from [97].

1.4.2 Facebook Datasets

Facebook Group and Page Posts to Pink Slime Domains Dataset consists of posts from the Python CrowdTangle API ([14]. For each pink slime domain listed in [45], the API was called to return all of the posts linking to them from public Facebook groups, pages, and profiles from 2019-2023. It should be noted that the API limits the response to 1,000 posts per call, so the

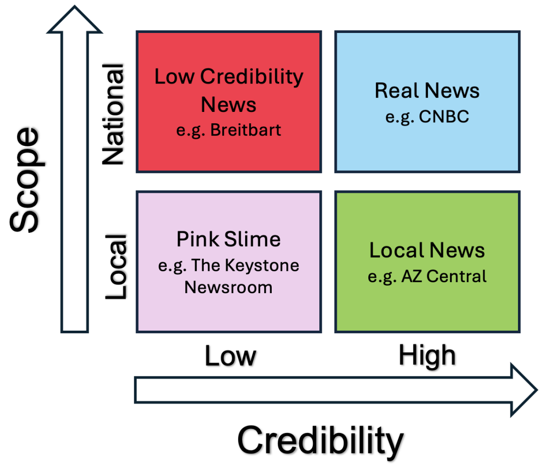


Figure 1.14: Relationships between news types

data was dynamically and recursively pulled; if a date range had more than 1,000 posts to a pink slime domain, then the date range was halved until the response would be under 1,000 posts. The minimum date range that the API would allow is one day, so if a domain had more than 1,000 posts from these sources on a given day, the additional posts were not included. Overall this yielded 1.2 million posts.

Facebook Pink Slime Ad Dataset consists of posts that pink slime parent organizations paid to promote on Facebook Pages. It was acquired through the Facebook Ad Library [62]. The amount paid for these posts and impressions garnered was taken as the average of the minimum and maximum values listed. When a majority of the ad impressions were in one state, this was listed in this research as the targeted state.

Facebook COVID 2020 Dataset consists of posts made in 2020 to Facebook pages and groups pertaining to the "reopen" movement and general "elections" from the Python Crowdtangle API ([14]. All of the posts contain external links. It is used as a training set in Chapter 4.

Midterms 2022 Dataset consists of posts from Facebook, Reddit, and Twitter pertaining to the United States 2022 Midterm Elections in regions with the most contention elections. The posts pulled from each of the platforms contain URLs to external sites for further analysis. The elections took place on November 8, 2022, and the data was collected from October 1, 2022 to December 1, 2022. Elections selected for this analysis included the most competitive districts and regions in Arizona, Georgia, Pennsylvania, Nevada, and Wisconsin [54]. The full set of keywords includes: (Kelly OR Blake OR AZSen OR Lake OR Hobbs OR AZGov OR Crane OR Halleran OR AZ02 OR Hodge OR Schweikert OR AZ01 OR Engel OR Ciscomani OR AZ06 OR Warnock OR Walker GASen OR Kemp OR Abrams OR GAGov OR McBath OR Handel OR GA06 OR Oz OR Fetterman OR PASen OR Shapiro OR Mastriano OR PAGov OR Scheller OR Wild OR PA07 OR Bognet OR Cartwright OR PA08 OR Shaffer OR Deluzio OR PA17 Mastro

OR Laxalt OR NVSen OR Sisolak OR Lombardo OR NVGov OR Becker OR Lee OR NV03 OR Peters OR Hosford OR NV04 OR Robertson OR Titus OR NV01 OR Johnson OR Barnes OR WISen OR Evers OR Michels OR WIGov OR Van Orden OR Pfaff OR WI03 OR Vance OR Ryan OR OHSen OR DeWine OR Whaley OR OHGov OR Chabot OR Landsman OR OH01 OR Sykes OR Gilbert OR OH13 OR Kaptur OR Majewski OR OH09 OR Beasley OR Budd OR NCSen OR Nickel OR Hines OR NC13) AND (vote OR election OR elect OR race OR democrat OR republican OR AZ OR Arizona ORGA OR Georgia OR PA OR Pennsylvania OR NV OR Nevada OR WI OR Wisconsin OR OH OR Ohio OR NC OR North Carolina)

The Twitter researcher API [111], Reddit's Pushshift API [14], and Facebook's CrowdTangle API [110] were all used to pull the data for this research.

United Kingdom 2024 Election Dataset consists of posts from Facebook groups and pages pertaining to the United Kingdom 2024. The posts pulled from each of the platforms contain URLs to external sites for further analysis. The elections took place on July 7, 2024, and the data was collected from March 16, 2024 to August 12, 2024. Using the House of Commons Library to create a set of keywords of all of the candidates in the election[4], the full set of keywords can be found in Appendix A.

1.4.3 Twitter Datasets

Trident Juncture 2018 Dataset contains tweets pertaining to the 2018 NATO-led military exercise in Norway. The tweets were collected from October 22, 2018 to November 13, 2018 via the Twitter API [111] and included the hashtags: #tridentjuncture, #nato, and their non-English variants. It represents 81,555 unique Twitter users and is used as a training dataset in the OMEN game experiment detailed in Chapter 5.

March 2020 Covid Dataset consists of 1.2 million tweets from the Twitter API [111] in March of 2020 with the following keywords: "coronaravirus", "coronavirus", "wuhan virus", "wuhan-virus", "2019nCoV","NCoV","NCoV2019". It is used as a training dataset in the OMEN game experiment detailed in Chapter 5.

Balikatan 2022 Dataset consists of tweets pulled from the Twitter API [111] in April 2022 related to the annual military exercise between the Philippines and the United States. It is used as a training dataset in the OMEN game experiment detailed in Chapter 5.

1.4.4 Other Datasets

The Scraped Pink Slime Webpages contains news articles listed on the homepage of the pink slime sites listed in the Pink Slime Dataset as of July 30, 2024. This dataset includes 35,697 news articles from 7 Courier news domains, 43 Local Report news domains, 1,107 Metric Media news domains, and 10 Star news domains. Overall there were 342 articles scarped from the Courier

[4] https://researchbriefings.files.parliament.uk/documents/CBP-10009/CBP-10009.pdf

news domains, 573 from Local Report, 34,380 from Metric Media, and 402 from Star. It was collected using Python's Beautiful Soup [5] and Newspaper3k [6] packages.

1.5 Tools Used

A series of computational tools are used throughout this thesis to identify networks, characterize their activity and their interactions.

ORA is a dynamic network analysis and visualization tool with capabilities to import data from several social media platforms [34]. It is used in this thesis to visualize social media networks, calculate centrality metrics, and implement the BEND framework.

NetMapper is a text-based software [34] used to extract linguistic cues pertaining to emotion, pronouns, and icons in a set of input text files. It appends this metadata to the original text files so that it can be imported into ORA to help classify which BEND maneuvers are taking place.

1.6 Internal Review Board (IRB) Approval

The collection of datasets and human subjects research were performed under the following IRBs.

The Midterms 2022 dataset was collected with IRB approval in the Fall of 2022 Federalwide Assurance No: FWA00004206 IRB Registration No: IRB00000603

The Facebook Pink Slime dataset was collected with IRB approval in the Spring of 2024 Federalwide Assurance No: FWA00004206 IRB Registration No: IRB00000338.

The Media Literacy Test was conducted with IRB approval in the Spring of 2024 and determined to be Exempt under the 2018 Common Rule 45 CFR 46.104.d.

[5] crummy.com/software/BeautifulSoup/bs4/doc/
[6] newspaper.readthedocs.io/en/latest/

Chapter 2

Characteristics of Pink Slime

This chapter seeks to characterize the behavior of pink slime sites as they exist on the Internet as well as how they entice people to visit their articles via social media. I start by analyzing the specific content being shared on these websites and note how different parent organizations deploy different strategies to continually republish the same "local" information across a wider network. To further motivate this research, I report on how much web traffic these websites garner via search engines before pivoting to social media. Through analysis of the Facebook ads the parent organizations have purchased as well as how these sites are shared on Facebook pages and groups, I dive into the evolving social media strategy by organization from 2018-2024. Finally, I take a step back from analyzing exclusively pink slime social media sharing to analyze how these sites are shared in comparison to the three other news types across Facebook, Reddit, and Twitter.

2.1 Research Questions

The key research questions for this chapter are:

- How similar is the content on pink slime sites?
- How many people visit these sites?
- What has pink slime's social media strategy been?
- Who is pink slime targeting?
- What are the network characteristics of pink slime?
- How are these sites shared differently on different platforms?
- How do their network features look different from other news types?

2.2 Pink Slime Website Similarity

To understand how much of the content from pink slime sites is copy and pasted, content similarity was analyzed looking at the articles on the pink slime sites as well as the articles shared on each site across networks. Using the scraped pink slime webpage dataset (which was able to collect news articles from all of the pink slime parent organization with the exception of the American Independent network), I compared the text of each news article shared on the homepage of these webpages with the text of all of the other articles shared on the homepage. This comparison was performed using cosine similarity with the TfidfVectorizer function of the scikitlearn package [1] and returned a numeric value between 0 and 1 for each pair of news articles. For each distinct url, the maximum cosine similarity value was returned to indicate which article had a similar article already on the same homepage. The results of this distribution can be found in Figure 2.1.

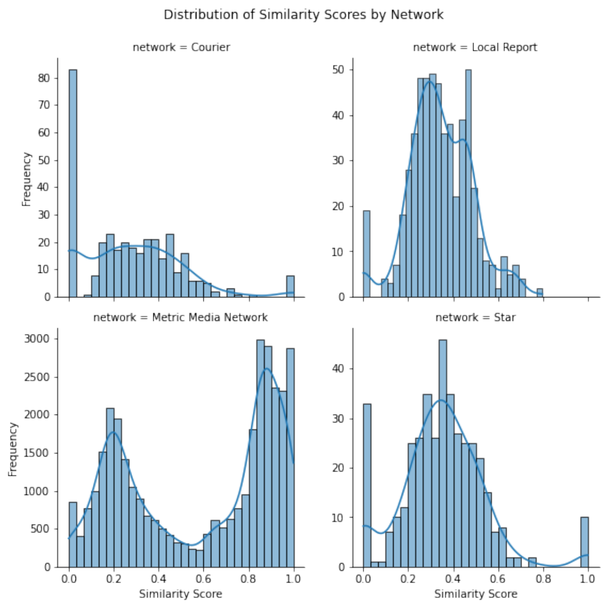


Figure 2.1: The distribution of the similarity score of each article on a pink slime homepage with the most similar news article on the same homepage, by network

Researchers analyzing similarity of news content online have identified a tf-idf threshold of 0.6, above which two articles are sharing the same content [27]. Using this cutoff, we see that Metric Media has the majority of its content with a similarity score above 0.6, indicating that the majority of its homepage is copy and pasted *within* the homepage. This phenomenon is not as pronounced for the other pink slime networks. This may be due to the extremely large number of websites whose content Metric Media must fill - over 1,000 websites targeting hyperlocal communities. However, the three other networks have fewer websites (49, 11, and 8) so they may not need to automate away the production of homepage news articles relevant to their audience as much as Metric Media does.

In order to understand how much of these articles are copy and pasted *across* the pink slime networks, I filtered down to the articles that did not have a matching news article (maximum similarity score below 0.6) on the same homepage and compared the similarity of these articles to the other articles on other websites within the network. Again, for each distinct url that did not have a matching article on its same homepage, the maximum cosine similarity value was

[1] scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.TfidfVectorizer.html

returned to indicate which article had a similar article on another homepage within the same news network. The results of this distribution can be found in Figure 2.2.

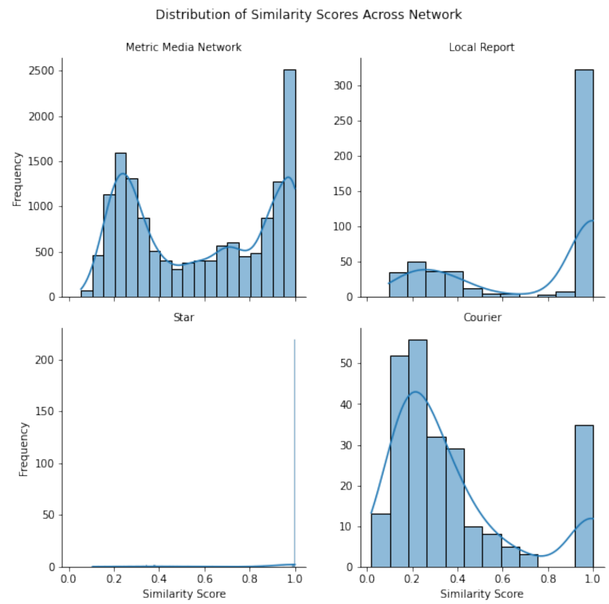
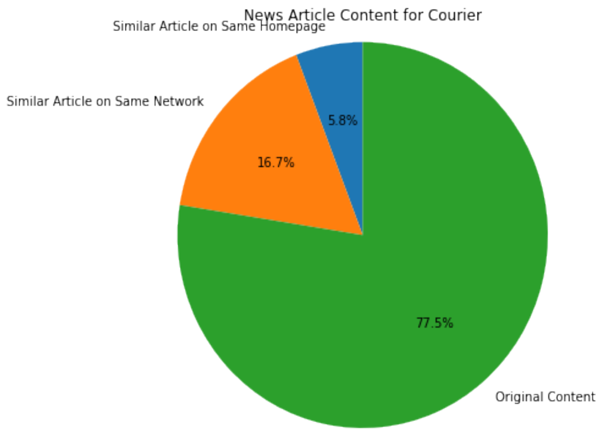


Figure 2.2: The distribution of the similarity score of each article on a pink slime homepage with the most similar news article on different homepages owned by the same parent organization, by network

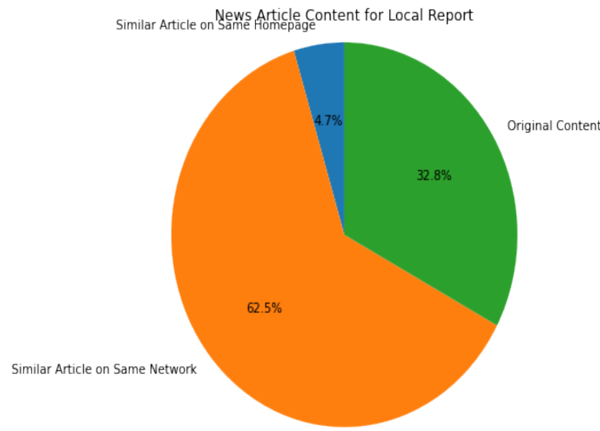
We see that the parent organization Star has the bulk of its content with over 0.95 cosine similarity across the platform, meaning it's sharing the same content on The Tennessee Star that it's sharing on The Minnesota Sun and the other news sites they own. Metric Media again has a substantial number of articles above the 0.6 similarity cutoff across their network. While Courier has some articles shared across their news network, it is not nearly as pronounced as it is for the other news sites.

Using the 0.6 cutoff, all of the homepage news articles were divided into three categories - having a matching article on the same homepage, having a matching article on another homepage of the same news network, and original content. For each news network, the distribution of the news content shared can be found in Figure 2.3.

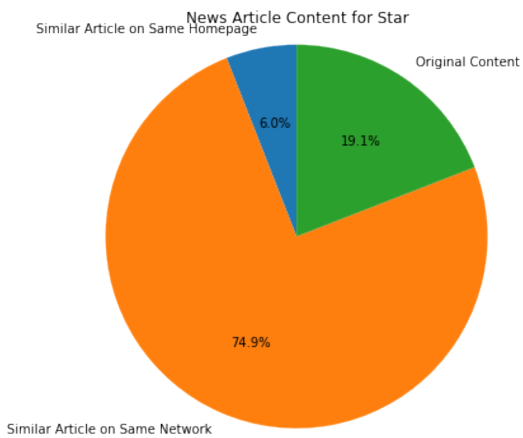
The vast majority of content (77.5%) of news articles appearing on the Courier Newsroom network is original content. While some organizations, like the Tow Center for Digital Journalism, have viewed a change in the Courier business strategy to be less "pink slime"-like and more authentic (as this visual would support), I will continue to include this network in the remainder of the thesis since their original intent at creation was modeled off the pink slime framework. Networks like Metric Media see their strategy of creating over 1,000 separate websites tested when they need enough fresh content for each site - ultimately 55.7% of articles on every webpage have a matching article on the *same* homepage. Meanwhile, organizations like Star News have 74.5% of their articles shared *across* the network. While each organization takes a different approach to creating and populating content on their websites, we will see this shaping the number of interactions news articles get on social media as well as website traffic in later sections of



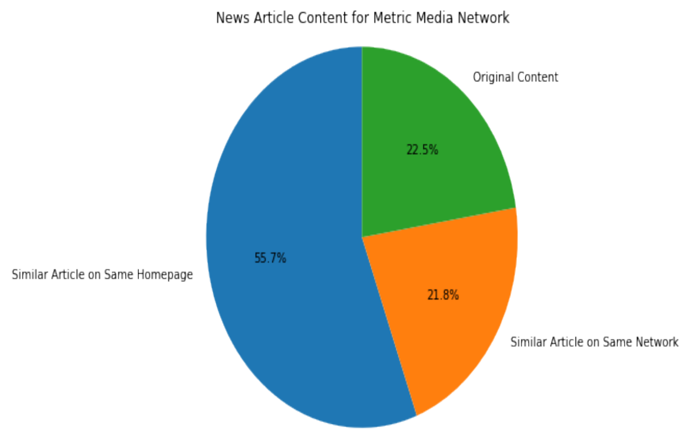
(a) Courier News Content Distribution



(b) Local Report News Content Distribution



(c) Star News Content Distribution



(d) Metric Media News Content Distribution

Figure 2.3: Each pink slime network's content distribution.

this thesis.

2.3 Pink Slime Web Traffic

In order to understand the amount of web traffic these pink slime sites are acquiring, the ahrefs API[2] was used to pull an estimation of the total monthly search engine traffic to these websites leading up to the 2024 U.S. Presidential Election (the data was pulled on September 27, 2024). The results, summarized by network in Table 2.1. While the average Metric Media news site only receives 301 visitors per month, it's important to note that over 1,000 of these sites exist, allowing the monthly traffic to surpass a quarter of a million clicks to sites in their control. While Local Report has few monthly visitors, the other three news organizations, which focus their efforts

[2]ahrefs.com/api/documentation/positions-metrics

Network	Sum	Average	Median
Metric Media Network	345,865	301	34
Courier Newsroom	132,802	14,755.8	17,736
Star News Network	25,451	2,313.7	94
American Independent	5,031	1,006.2	1,490
Local Report	237	4.8	1

Table 2.1: Statistics on the monthly search-engine-generated traffic to the pink slime sites, grouped by their parent organization.

on having news sites targeting swing states, receive thousands of visits via search engines. The target audience for these sites live in counties and states where a few thousand votes could sway a national election. While Google is the largest referrer of pink slime, it only accounts for 23.4% of the visits [81], meaning clicked traffic to these sites is around four times higher.

To extend this analysis, the pink slime sites targeting the following states received the most visits (in order): Illinois, Pennsylvania, Michigan, Arizona, Wisconsin, and North Carolina. With the exception of Illinois (where a disproportionately high number of pink slime sites exist), these states are swing states in the upcoming 2024 Presidential Election. When taking into consideration the population of the states, the states that received the most visits relative to the size of the state are: Illinois (where 0.6 visits occurred per 100 residents), West Virginia, Iowa, Wisconsin, Arizona, Tennessee, Pennsylvania, Michigan, and North Carolina (with 0.22 visits occurring per 100 residents). While these numbers may seem small, the New York Times considers Wisconsin, Arizona, Pennsylvania, Michigan, and North Carolina to be swing states that play an outsized role in determining the election with previous elections coming down to as few as 40,000 votes to determine which presidential candidate gets all of the state's electoral college votes [103].

2.4 Pink Slime's Social Media Strategy

2.4.1 Facebook Ads

This section describes the impact of pink slime via all of the Facebook ads the parent organizations have purchased. Meta publicizes the political ads purchased to run on their platforms in their Ad Library [3]. I gathered all of the ads purchased by the following accounts: Courier's (”Cardinal & Pine”, ”Courier Newsroom”, ”Courier Newsroom, Inc.”, ”Floricuas”, ”Granite Post”, ”Iowa Starting Line”, ”The 'Gander Newsroom”, ”The Copper Courier”, ”The Keystone”, ”The Nevadan”, ”UpNorthNews”), Metric Media's (”Metric Media LLC”, ”Franklin Archer”, ”Local Government Information Services”, ”The Record”), and American Independent's (”American Independent Media”). No ads were found from the Star and Local Report parent organizations. For each ad, the dates it was active on Meta platforms as well as a minimum and maximum number of views and ad spend are listed. For this research, those two numbers were averaged to determine the impressions and ad spend. It should be noted that for ads with over a million

[3] facebook.com/ads/library

Year	Number of Ads	Total Impressions	Total Ad Spend
2018	374	1,328,813	$21,713
2019	1,367	16,091,317	$200,067
2020	2,774	105,847,118	$1,118,763
2021	1,735	14,787,133	$158,883
2022	3,696	43,776,653	$731,752
2023	1,390	11,726,806	$281,905
2024	1,970	149,043,538	$3,135,265

Table 2.2: Pink Slime Facebook Ads Over Time (Through September 2024)

impressions, only a lower bound (of one million) is publicized, and this research defaults to labeling the number of impressions as one million even though that number is a very conservative estimate that can be much higher. Some demographic information (age, gender, and location) about who viewed the ads is also given and is analyzed later in this section. The amount of ad spend and impressions - demonstrated in Table 2.2 - illustrates the importance of understanding the changes over time. While most academic literature focuses on pink slime during the last presidential election year (2020), only 105 million impressions were garnered on ads in that calendar year; however, during our current presidential election year (2024), almost 150 million impressions have been garnered on these ads through September (weeks ahead of the election). While there is a drop off of ad spend and impressions during non-election years, there are still tens of millions of impressions during this time, including over 43 million views during the year of the midterm elections (2022). These sites are only beginning to pick up traction, and general trends of these ads are summarized by parent organization in Figure 2.6.

In order to analyze the messaging of these impactful ads, I broke down the content within Facebook ad messaging across the years and populated the frequency of messaging into a word cloud. I first pre-processed the text in the messaging to remove stopwords and URLs, before using Python's wordcloud package[4] to formulate the word cloud including the top 100 words, sized by frequency of appearance. Figure 2.4 illustrates the changing focus of topics in election and off-election years (word clouds for all of the years can be found in the Appendix in Figure B.1) ; furthermore, Figure B.2 shows how these words changed between the two elections generated using [47]. Throughout all of the years, the most targeted states ('Texas', 'Michigan', and 'Arizona') remain as top terms. One key tactic used is to write ads with the same message but switching out the name of the state for the one that is being targeted in the ad. For example, some of the ads run in 2022 had the following titles: "'Inflation has shot up a staggering 13.2%' since Biden took office, Arizona's CPI at 13%" , "'Inflation has shot up a staggering 13.2%' since Biden took office, Michigan's CPI at 8.1%", "3 in 5 Americans concerned about housing affordability, North Carolina's average rent up 30%", and "3 in 5 Americans concerned about housing affordability, Wisconsin's average rent up 17%."

During election years, ad spending increased drastically, and the conversations naturally turn political. While the then-candidate Biden's name was not at the forefront of the 2020 conversation, there was a focus on the phrase *Catholic*. The ads containing references to 'Catholic'

[4] `https://pypi.org/project/wordcloud/`

were mostly run in September and October 2020 by the Metric Media organization, leading into the appointment of Catholic Supreme Court Justice Amy Coney Barrett. The top two of these ads, garnering 125,000 and 50,000 impressions, respectively, were titled ‘President Trump addresses Catholics directly’ and ‘Catholic Vote: Biden’s anti-school choice stance should worry WI Catholic school parents’; the ‘Catholic’ phrase was indirectly used to support President Trump’s re-election.

During the midterms in 2022, President Biden was the top phrase, with secondary attention paid to key economic issues like ‘inflation’, ‘gas’ , and ‘prices’. The top two of these ads (ran by Metric Media) garnered 300,000 total impressions with the title ‘As Pennsylvanians receive fourth stimulus check, Pigott points out negative real wage growth: ‘Joe Biden is the pay cut president”. Much like the 2020 efforts to use the Catholic narrative into praise for President Trump, the 2022 tactic was to use negative economic news to undermine President Biden (and the Democratic party) ahead of the midterm election.

Leading up to the 2024 Presidential Election, there is renewed conversation around both candidates as well as the current president, Joe Biden. Primary conversation is around phrases focused on immigration from the southern border with terms like like “illegal”, “aliens”, and “border” as well as “abortion” and “inflation.” Liberal-leaning Courier ran ads with millions of impressions supporting Harris by critiquing Trump’s stances on abortion with text like: “WATCH as North Carolina Auditor Jessica Holmes (@jessicafornc) explains how Donald Trump is responsible for abortions bans across the country, including North Carolina, “Right now, I have fewer rights than my mom or grandma had fifty years ago” she said at a rally in support of the Harris-Walz ticket on Monday.” Metric Media also used abortion to tear down Ohio GOP primary candidate Bernie Moreno: “On non-profit board, GOP Senate hopeful Moreno approved $2.1 million in grants to Planned Parenthood, pro-abortion expansion groups”. Instead, many of Metric Media’s ads were in support of another GOP primary candidate Frank LaRose: “Sec. of State Frank LaRose (R-Ohio), who is running for U.S. Senate, announced a border security plan today that includes sending three U.S. military divisions to the U.S.-Mexico border.”

Ad expenditure during the years between elections drastically diminished, and the discourse focuses more on “court”. These ads highlighted state supreme and high courts, and are not necessarily political or partisan in nature. For example, ‘Appeals court vacates ruling against Parkways Authority over Turnpike toll fees’ received 10,000 impressions in 2023. This strategy may be to establish the Facebook pages sharing the news as a nonpartisan, trustworthy local news outlet when they aren’t actively trying to push a political message, and to keep the organizations active and visible to the Facebook audience even during the down time.

With the advertising expenditure segregated by states over election years in Figure 2.5, I analyze the expenditure per state when the Facebook Ad Library included data on where the impressions were viewed (which was included in approximately 27% of reported ads). Additional map coverage of the maps for all years and categorized by parent organization can be found in the Appendix Figure B.3. I observe that three of the five organizations purchased Facebook ads, and Metric Media was the only organization to consistently purchase ads across all time periods. Courier purchased ads through 2020 and then again in 2024, and American Independent purchased ads in 2022-2024. All of the organizations who purchased ads targeted Pennsylvania - a key swing state that determined Biden’s 2020 victory over Trump as well as Michigan, Wisconsin, and Arizona - fellow swing states with some of the closest voting spreads in the 2020

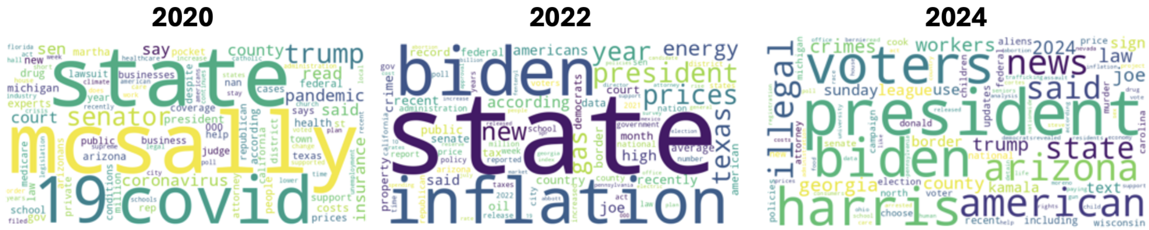


Figure 2.4: Word clouds of the top 100 words used in Facebook ads by Pink Slime Organizations during election years over time

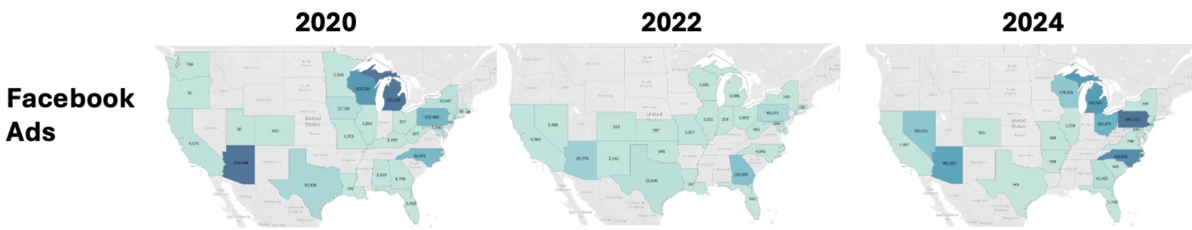


Figure 2.5: Advertising Expenditure by State

Presidential Election.

The top ten states by ad spend (from highest to lowest) are Pennsylvania, Arizona, Michigan, Wisconsin, North Carolina, Ohio, Nevada, Georgia, Texas, and California. All of these states (with the exception of Texas and California) were among those with the closest 2020 Presidential Elections voting spread. Iowa had the 11th highest ad spend due to its unique position as the first state in the country to hold an electoral event with its caucus every four years. While swing states are an important target during election years, other states' significance was seen as a result of isolated events. For example, Courier spent $16,999 in ads to Iowa in 2020. During the 2020 elections, Trump won Iowa against Biden by 8.2 percentage points. Given the small difference, it is likely that Courier, a left-leaning party, poured more resources to boost chances in Iowa.

Furthermore, the different pink slime parent organizations target different populations by age and gender, per Table 2.3. American Independent and Courier Newsroom, two of the politically liberal-leaning organizations focus more on women than men when selecting who should receive their ads, and the majority of their impressions are coming from Facebook users who are under 55. Meanwhile, Metric Media shows their ads to a more male audience that skews older; their largest target age demographic is for social media users over the age of 65, and the majority of their ads are viewed by people over the age of 55. The left-leaning pink slime parent organizations generally target younger women, while the right-leaning parent organizations are going after the older male demographic.

In order to understand whether attributes of given states had an effect on how much ad money was spent to target that state, we performed Pearson Correlations of state-specific variables to

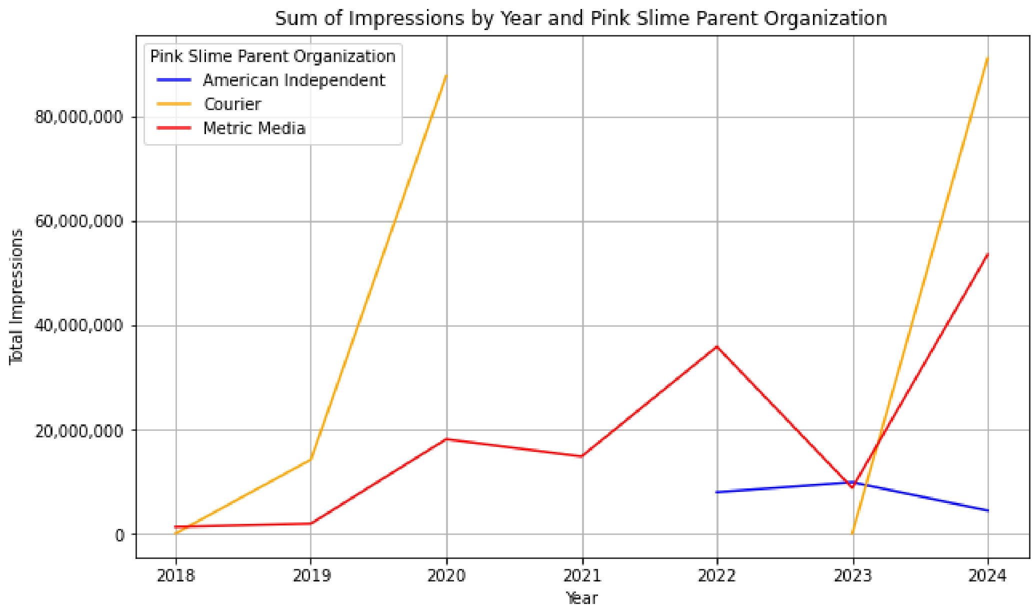


Figure 2.6: Advertising Expenditure by Parent Organization Over Time

	Female	Male	18-24	25-34	35-44	45-54	55-64	65+
American Independent	67.8%	31.5%	4.5%	18.6%	22.3%	20.2%	17.9%	16.4%
Courier Newsroom	59.9%	36.6%	16.1%	29.5%	21.0%	14.6%	8.6%	7.4%
Metric Media	43.3%	50.0%	5.5%	11.5%	10.4%	14.4%	22.5%	29.7%

Table 2.3: Breakdown of targeted ad demographics by gender and age for pink slime organizations

Variable	Correlation	Significance
2020 Voter Spread	**-0.42**	**0.004**
Cities Over 100k Population	0.09	0.54
Percent of Population Living in Rural Areas	-0.12	0.43
Percent of Population with a Bachelors Degree	-0.008	0.96
2020 GDP	0.13	0.39
Median Age of State	-0.08	0.61
March 2020 Governor's Party	-0.17	0.27
Electoral College Votes	0.17	0.25

Table 2.4: State variables and their Pearson correlation to 2022 pink slime ad spend

the 2022 pink slime ad spend in each state. This includes the 2020 voter spread[5], 2020 GDP[6], percentage of population living in rural areas[7], percentage of population with a bachelors degree[8] and so forth. All of the state-specific datasets were from 2020, as that would be the most recent datasets advertising organizations would have access to when determining how much money to spend during the 2022 midterm elections. In Table 2.4 I found the Pearson correlation coefficient between the state variables and the ad spend as well as the corresponding p-values. The only statistically significant (p-value < 0.05) relationship is the 2020 presidential voter spread of a state. This indicates that pink slime organizations are attempting to exert influence in the closest of races that they believe could flip.

Finally, to see how the ad spend influenced organic conversation, Figure 2.7 was generated to compare the amount of money spent promoting a given pink slime domain via Facebook ads and the number of times that domain appeared in a Facebook group. A linear trend line ($R^2 = 0.35$) indicates a weak positive correlation between the two measures, ad spend and organic group conversation.

2.4.2 Posts to Facebook Pages and Groups

Ads aren't the only place where people will experience pink slime on their social media feeds. Many Facebook Page and group posts contain links to pink slime sites. Plots breaking out how many posts were made linking to the various parent organization sites by year can be found in Figure B.4 and Figure B.5. This data is also visualized using geospatial metadata in Figure 2.9 to understand which pink slime sites affiliated with the various states are shared more during election years. The remainder of the years (and broken down by parent organization) can be found in the Appendix Figure B.4 and Figure B.5.

Metric Media is the most dominant parent organization through their many Facebook pages posts, with posts on pages relating to all 50 states. Metric Media owns hundreds of these pink slime sites whereas the other four organizations own fewer than 15 of these sites each. Their goal

[5] https://www.presidency.ucsb.edu/statistics/elections/2020

[6] https://www.bea.gov/

[7] https://www.census.gov/en.html

[8] https://fred.stlouisfed.org/release/tables?rid=330&eid=391444&od=2020-01-01

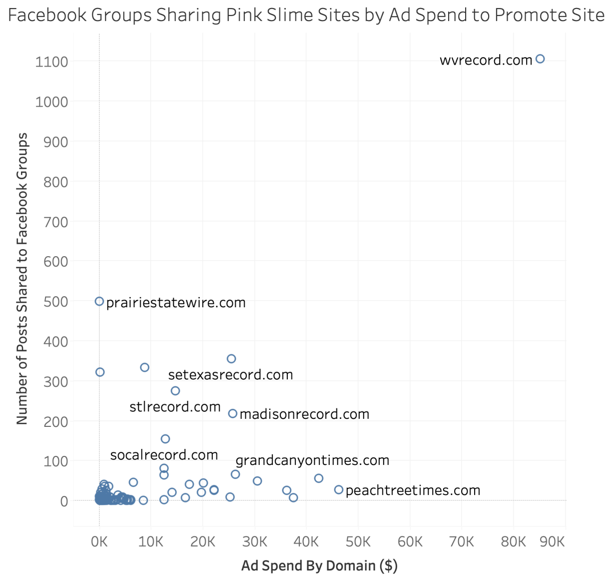
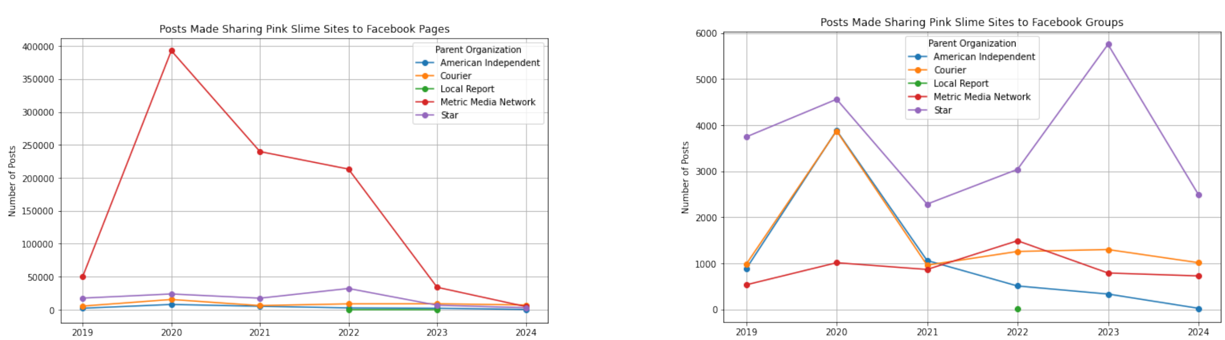


Figure 2.7: Number of Instances a Pink Slime Domain Appears in a Facebook Group by Ad Spend for those Domains



(a) An over time plot of Page Posts Linking to Pink Slime Sites by Organization (through August 2024)

(b) An over time plot of Group Posts Linking to Pink Slime Site by Organization (through August 2024)

Figure 2.8: Facebook Pages and Group Posts Over Time

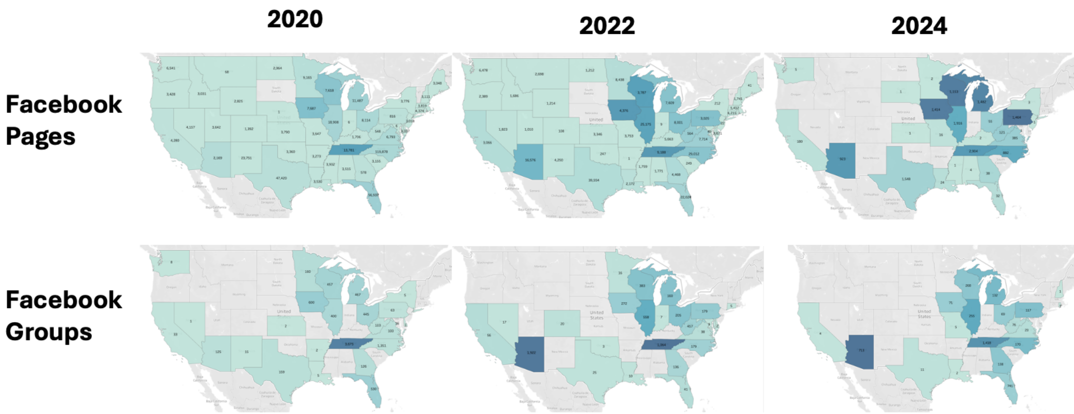


Figure 2.9: Facebook Pages and Groups Posts by State

seems to be to reach *every* corner of the United States with strong focus on Illinois, where the founder of Metric Media, Brian Timpone, resides. The other pink slime organizations have more concentrated sharing in Facebook groups and pages to states of higher electoral importance.

Furthermore, Metric Media's prolific posting to Facebook pages (usually to the formal pages promoting each of their many websites) peaked in 2020 and decline drastically during the 2022 Midterms. At the same time, they pivoted towards doubling their ad spend, signifying a business decision to slow down on the spamming of their Facebook pages and focus on reaching wider audiences via paying for Facebook ads. The other parent organizations all have fewer than 35,000 posts per year to Facebook pages, and the Facebook pages with the most shares are the official Facebook Page accounts for the websites.

Despite not purchasing any ads, the Star news network consistently has more shares of its news articles to Facebook groups than any other parent organization. This could be due to intense spamming of posts to specific groups. Five Facebook groups in this dataset had more than 300 Star network posts shared from 2018-2024, with the top group having 1,088 posts linking to these sites. Only three other Facebook groups have over 300 posts linking to sites from a single parent organization - one sharing Metric Media and two sharing Courier news sites.

2.5 Network Differences Between Pink Slime and Other News Types

To understand how pink slime spreads differently from other news types on different platforms, a case study was performed on the midterms dataset to compare patterns of how users share news articles from the big four news types across three social media platforms - Twitter, Reddit, and Facebook - via statistical and network-based analysis.

	Facebook	**Twitter**	**Reddit**
Posts	28,178	1,383,896	16,375
Posts with News URLs	17,268	851,828	7,811

Table 2.5: Number of links with labels in the midterms dataset by platform

2.5.1 Multi-platform Midterms Dataset

For context, the 2022 United States Midterm elections were held on November 8, 2022 with over 107 million Americans electing 36 governors, 35 senators, and all 435 voting seats in the House of Representatives [40]. Since the office of the President was not on the ballot, smaller, regional elections were the focus of the election news coverage. Posts containing links to external URLs as well as keywords pertaining to contentious elections in battleground states were collected from Twitter, Reddit, and Facebook (keywords can be found in Chapter 1's Data section). Elections in Arizona, Georgia, Pennsylvania, Nevada, Wisconsin, and North Carolina were included in this data collection for one month prior to the election (starting October 1, 2022) and one month after the election took place (through December 1, 2022) to include conversation points before and after the election.

The data collection yielded 1,383,896, 28,178, and 16,375 posts from Twitter, Facebook Pages, and Reddit, respectively. These posts were then cleaned and their external URLs were compared against the CASOS Media Thesaurus to assign the corresponding news type label (Real News, Local News, Pink Slime, or Low Credibility News as defined in the Datasets portion of Chapter 1) to the post. Not all of the posts had a URL with a news domain in the CASOS Media Thesaurus; this resulted in 851,828 tweets, 17,268 Facebook Page posts, and 7,811 Reddit posts and comments linking to URLs with a designated news type rating. The number of posts with these news links from each platform can be seen in Table 2.5.

2.5.2 Data Analysis

This analysis was guided by three key research questions:

1. **Platform-Based Analysis**: How do different social media platforms differ by the types of news they share?
2. **News Type-Based Analysis**: How does the engagement on posts differ based on the type of news they are sharing?
3. **User-Based Analysis**: For users sharing pink slime news, what other news are they sharing? Furthermore, is this sharing done because they care about sharing news close to their local community or to share news that aligns with their political ideology?

Platform-Based Analysis To answer the question of "How do different social media platforms differ by the types of news they share?", each of the datasets of social media posts corresponding to the different platforms was segregated into which news type the posts shared and analyzed. In Table 2.6 we can see that pink slime is present in all of the social media platforms; however, it varies greatly by platform. This is notable since previous researchers [30] did not find pink

News Type	Facebook	Twitter	Reddit
Real News	65.6%	78.7%	89.5%
Local News	29.3%	13.8%	9.7%
Pink Slime	1.1%	2.3%	0.1%
Low Credibility News	4.1%	5.3%	0.6%

Table 2.6: Breakdown of news types shared on the three platforms, as percentages of the amount of each news type site as a total of the number of sites shared within each platform.

slime on political subreddits and concluded that these sites were not present on the platform; the midterms research suggests that pink slime *is* present on Reddit, but it is shared more frequently to smaller subreddits that are tailored to regional communities.

Reddit, leading with the highest percentage of real news and the lowest percentage of low credibility news only saw 0.6% of its posts and comments pertaining to these elections linking to pink slime sites. Reddit's subreddits must have moderators, who we have seen delete almost 3 million comments in less than a year due to hateful speech and poor quality news [36]. A built-in system that encourages moderation may be what is keeping Reddit full of high credibility news.

Facebook pages have the most news shared that is local in scope; nearly 30% of posts in this dataset link to local news and 1.1% to pink slime. This suggests that news pertaining to local communities is shared more extensively on Facebook pages than the other platforms studied in this research.

Finally, Twitter has the highest percentage of news that is low in credibility; it leads in the proportion of low credibility news (5.3%) and pink slime (2.3%) sites shared. While other researchers have shown that low credibility news is most prevalent on Twitter [99] and especially so surrounding political elections [28, 53], this may be more pronounced due to a change of ownership of the Twitter platform during the midterms that resulted in the layoffs of the content moderation teams.

The next analysis performed was a network analysis which explored the sequential transition of news type sharing of users on each platform. Directed network graphs were created to represent the transition of a user's news sharing from one type to another. Each node represents a news type, and the link between them shows a user who shared multiple posts linking to the four news types. For example, if a user shares a local news URL in the first post then shares a real news URL in the second post, a link on the graph will be drawn in the direction from the local news node to the real news node. Finally, the proportion of each type of link is calculated, and it is represented through the thickness of the link width in the network graph, as illustrated in Figure 2.10.

We observe that the most common transition pattern among all news types and among all platforms is that of self-loops, which shows that the most likely behavior for users is to continue sharing the same type of news. When expanding our analysis to consecutive news sharing of *different* news types, we observe that all of the platforms have a high likelihood of sharing real news and then local news (and vice versa), showing that those users sharing high credibility news generally continue to do so regardless of the scope. However, Twitter and Facebook see high overlap of low credibility news sharing with real news. It's possible that these users are not

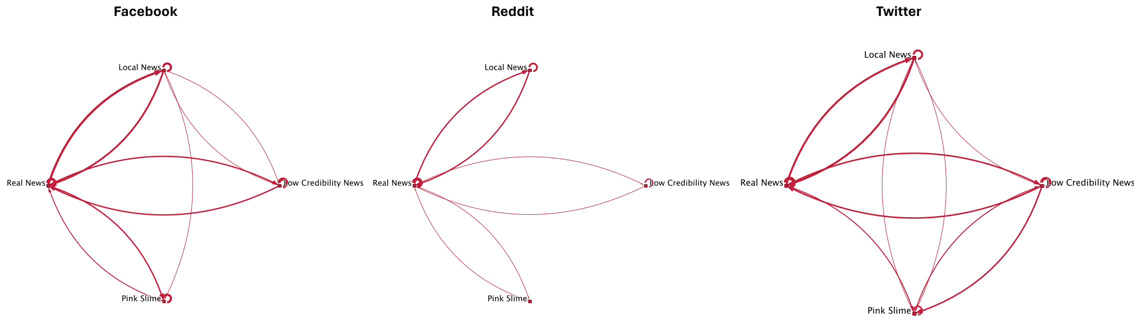


Figure 2.10: User likelihood (represented by line thickness) of sharing one news type based on previous news type shared by platform.

News Type	Facebook	Twitter	Reddit
Real News	68%	83%	91%
Local News	40%	34%	27%
Pink Slime	46%	78%	
Low Credibility News	83%	52%	41%

Table 2.7: Percentage of users who by platform and news type of continue to post within the same news type (i.e., self-loops)

aware that the news they are sharing is low credibility, but it's important to note that one can still share both of these news types.

No Reddit users that shared local news also shared pink slime news, but the sharing between local news and pink slime news exists on Facebook and Twitter. Much like the sharing of real news and low credibility news, it is again possible that those sharing local news and pink slime have a hard time differentiating between these news sources that are local in scope.

When we analyze pink slime specifically, these network visualizations show that the most likely subsequent post for a user sharing this news type is to continue sharing pink slime (what we refer to as a "self-loop." However, on Twitter, where we have our richest dataset with the most pink slime posts, there is a strong likelihood that someone sharing pink slime is also sharing low credibility news. This insight will be useful in Chapter 4 when we are using these characteristics to find new sources of pink slime based on news sharing behavior.

Filtering down to only self-loops by platform and news type in Table 2.7, we observe that the highest number of self-loops are seen in Twitter and Reddit users who are sharing real news. However, for Facebook, self-loops are highest among pages that are sharing low credibility news. We don't observe any self loops of pink slime on Reddit since no single user shared two instances of pink slime on the platform; however, on Twitter 78% of the times when a user shared pink slime and then shared a subsequent tweet, that subsequent tweet contained a link to a pink slime news site.

News Type-Based Analysis In this analysis we are interested in answering the question, "How does the engagement on posts differ based on the type of news they are sharing?" We first de-

Quartile of Facebook Group Size	**Q1**	**Q2**	**Q3**	**Q4**
Real News	12%	14%	18%	56%
Local News	9%	20%	43%	28%
Pink Slime	21%	7%	36%	36%
Low Credibility News	22%	20%	43%	28%

Table 2.8: Distribution of news types shared to Facebook across the quartiles of the Facebook Pages group sizes.

scribe a metric that can broadly compare engagement of the posts regardless of the size of the group that was reached.

Through the collection of Facebook data through the CrowdTangle API, we were given information about the size of the group to which the posts were shared. However, due to changes in the Twitter algorithm that allow users to see posts from accounts they do not follow, it is not clear that the relationship between the number of followers of a Twitter account and the views of its posts are linear. On Reddit, the PushShift API provided the number of subscribers to a given subreddit but only for the posts collected and not the comments (which constitutes 79% of the Reddit dataset). Due to these data limitations, only the Facebook dataset was utilized to understand the relationship between news type and engagement.

The median group size of the Facebook pages is 8,928 subscribers, and the smallest 25% of groups have fewer than 1,653 subscribers. To better understand where these news types are shared, we observe the distribution of these news types by quartiles of Facebook page subscribers in Table 2.8. This information reveals that the two news types shared most frequently to the smallest quartile of Facebook pages are pink slime (21% of its posts are shared to groups of this size) and low credibility news (22%). Meanwhile, the majority of the time that real news posts are shared, they are shared to the largest quartile of Facebook pages. As it has a national scope and high credibility, it may appeal to larger communities of people.

Since the posts of different news types are shared to groups of different sizes, comparing the absolute engagement of the posts would show real news, with a whopping 56% of its posts shared to the largest quartile of Facebook pages, with the most engagement due to sheer number of impressions made. However, I propose normalizing the engagement of these posts by the size of the page to which it was shared. If a post gets 100 likes in a group of 100 people, it's a hit, but if a post gets 100 likes in a group of 1,000,000 people it should be considered a flop. On Facebook, I define the relative engagement as the number of likes a post received divided by the size of the group the post was shared to. This relative engagement metric measures engagement for Facebook posts on pages as a proportion of the total number of followers or likes to the page. Specifically, the relative engagement is defined as per Equation 2.1.

$$\text{Relative Engagement} = \frac{\text{\# of Likes a Post Receives}}{\text{\# of Followers or Likes the Page Has}} \tag{2.1}$$

When the relative engagement of posts are separated by news type with its logarithmic distribution presented, an alarming trend emerges (as seen in Figure 2.11) - pink slime receives a

News Type	Average Relative Engagement	Median Relative Engagement	Std Dev of Relative Engagement
Real News	0.0015	0.00016	0.026
Local News	0.0014	0.00022	0.24
Pink Slime	0.0020	0.00039	0.005
Low Credibility News	0.0010	0.00036	0.003

Table 2.9: Relative Engagement Metrics by News Type

greater relative engagement than any other news type. This news is resonating with its audience, and real news is not as the next highest relative engagements are seen in posts sharing low credibility news, local news, and finally real news.

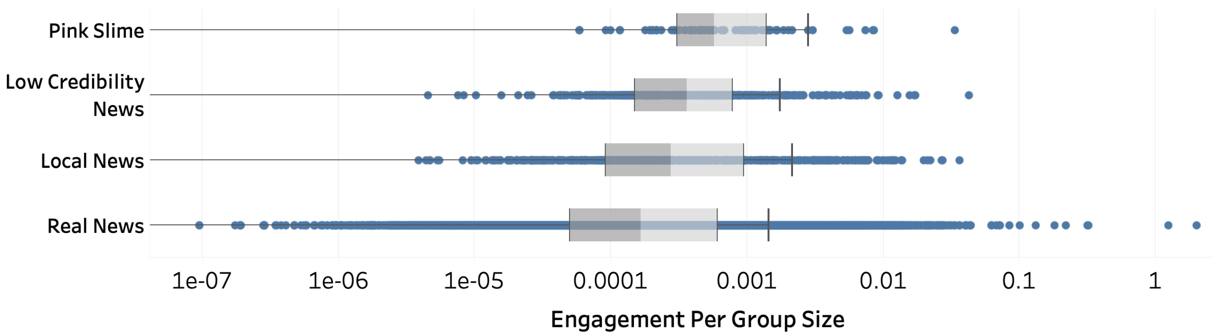


Figure 2.11: Facebook engagement per group size by news type

Beyond looking at the average Relative Engagement, further description of this variable is broken down in Table 2.9. The standard deviation of Relative Engagement for local news is the highest, suggesting that the reception of local news varies widely. The standard deviation of low credibility news is the lowest, meaning the audience with which this news resonates is more consistent with their engagement of these posts.

User-Based Analysis In this analysis we are interested in answering the questions, "For users sharing pink slime news, what other news are they sharing? Furthermore, is this sharing done because they care about news close to their local community or to share news that aligns with their political ideology?" The dataset is filtered down to exclusively users that shared pink slime news articles regardless of platform to see what other news sources they are sharing.

First, an analysis was done to understand the pink slime news sharing based on whether the users were sharing pink slime based on their political ideology. Users were then place in one of three categories: Left, which included users who only shared news from left-leaning pink slime organizations; Right, which consisted of users who shared only news from right-leaning pink slime organizations; and Both, which was comprised of users who shared news from both left- and right-leaning pink slime organizations. For each of these three user buckets, I analyzed the distribution of the news types shared by these users as a proportion of the total links the users shared which can be seen in Figure 2.12.

By inspection, agents sharing left-leaning pink slime sites shared more local news and less low credibility news than those sharing right-leaning pink slime sites. To understand if there was a significant difference in this distribution, I used a Chi-Squared test to compare the three

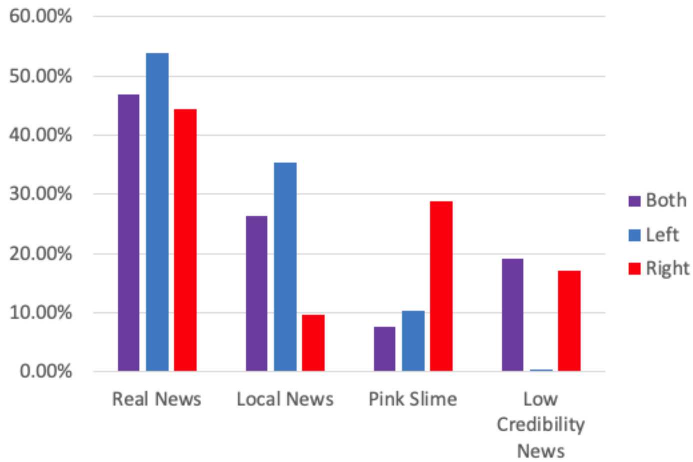


Figure 2.12: Distribution of the news types shared by agents who shared pink slime, grouped by whether the agent shared pink slime from a right-leaning pink slime organization, a left-leaning pink slime organization, or both

distributions and see if there are differences with news sharing based on the pink slime political ideologies. The null hypothesis H_0 is that there is no difference in the distribution of the types of news shared among different political leanings, i.e. each political leaning shares the same proportion of each type of news. The Chi-squared formula we used is expressed in Equation 2.2.

$$\begin{aligned} \chi^2 &= \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \\ &\text{where:} \\ \chi^2 &= \text{Chi-squared statistic} \\ O_{ij} &= \text{Observed frequency in cell } (i, j) \\ &\quad \text{of a group of users sharing a news type} \\ E_{ij} &= \text{Expected frequency in cell } (i, j) \\ &\quad \text{of a group of users sharing a news type} \end{aligned} \tag{2.2}$$

The chi-square statistic (with 6 degrees of freedom) was 9234.5219 with a p-value < 0.00001. Using a significance level of p $<$ .05, we conclude that the distribution in news type sharing among those sharing pink slime from the different politically leaning organizations is significantly different. Therefore, I conclude that there is a difference in the distribution of the types of news shared among different political leaning.

As a next step in the User-Based analysis, I construct a User x News Domain network diagram consisting of two types of nodes - users and news domains. The user nodes (colored gray) are linked to the domains they share, and the domain nodes are colored by the type of news. The blue nodes represent real news, green represents local news, red represents low credibility news, and pink represents pink slime domains. The pink slime domains are labeled, and they are split between two inner-connected components, as visualized in Figure 2.13 using the ORA software's force-directed continuous graph layout algorithm [33] [9].

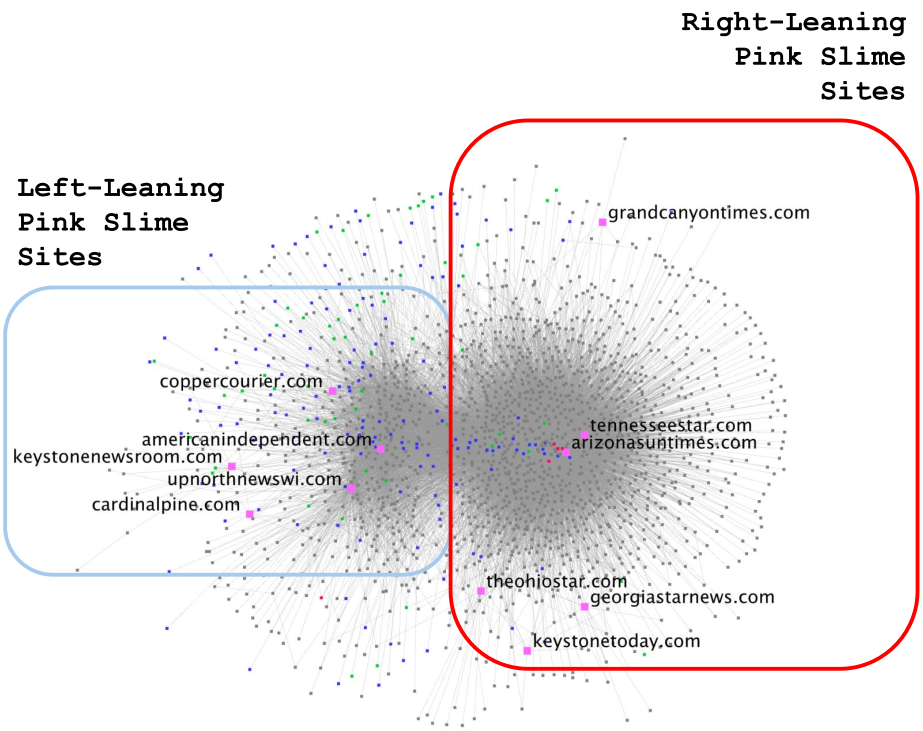


Figure 2.13: News sources shared by users (including all platforms) who shared pink slime domains. Pink slime sites are labeled and given a pink node coloring, local news sites are green nodes, real news sites are blue nodes, and low credibility news sites are red nodes.

The pink slime domains in the right component (captured in the red box) are all under the control of parent organizations pushing politically right-leaning news. This includes `grandcanyontimes.com` and `keystonetoday.com` which are controlled by Metric Media; and `georgiastarnews.com`, `theohiostar.com` and `tennesseestar.com` which are owned by the Star News Network.

The pink slime domains in the left component (captured in the blue box) are under the control of parent organizations that push politically left-leaning news. `coppercourier.com`, `keystonenewsroom.com`, `cardinalpine.com` and `upnorthnewswi.com` are controlled by the Courier Newsroom, and `americanindependent.com` is under the control of The American Independent which has many more state-specific sites.

This visual division suggests that the news spread is not done along regional lines (a component along each of the six states in the dataset) but rather along political lines.

2.6 Limitations

This section is limited to the news types that are in the 32,000 labeled news sites in the CASOS news thesaurus.

2.7 Conclusions

This chapter answered many key questions for those who ask *"Well, what are these pink slime sites doing in digital spaces?"*.

First, the content of these sites was analyzed to determine how the different parent organizations have different strategies for writing news articles based on how many sites they control. While organizations like Metric Media, with over 1,000 websites to write content for, copy and paste most of the content on the same homepage, organizations like Star News (with only a handful of sites) repeat their articles *across* their network. Finally, Courier Newsroom appears to have made strides to produce more original content on their websites and may now be classified more as pink slime-adjacent. However, since they have previously exhibited characteristics of pink slime, they are still studied throughout this thesis for their previous contributions.

By analyzing recent web traffic to these sites, we see around 500,000 monthly visitors to the sites from search engines like Google. While these numbers are small, they only account for 23.4% of visits to the sites, and visits are concentrated around pink slime sites targeting swing states where few votes can sway the outcome of an election.

In assessing the Facebook advertising expenditure of these sites, we observe cyclical ad spend by three of the parent organizations, with peaks during presidential election years and a focus on spending in swing states with a low voter spread. While the left-leaning organizations focused on a more female and younger-skewing demographic to receive the ads, Metric Media (a right-leaning organization) targeted an older male demographic.

When comparing how pink slime is spread on social media platforms in relation to the other major news types, we find that while pink slime makes up a minority of the posts on the different news platforms, the posts sharing pink slime receive the highest relative engagement of the news types. Furthermore, the individuals sharing pink slime tend to continue to share pink slime or share low credibility news. Finally, users sharing pink slime on social media are doing so along political lines as opposed to geographic.

Chapter 3

BEND Maneuvers of Pink Slime

3.1 Research Questions

The key research questions for this chapter is:

- What BEND maneuvers are pink slime sites utilizing?
- How do these maneuvers vary from platform to platform?
- How do the maneuvers compare to those of local news organizations?

3.2 Comparing Pink Slime to Local News Maneuvers

3.2.1 Background on BEND

The CASOS Center at Carnegie Mellon University has produced substantial research in the field of categorizing online influence operations; they have published a set of 16 defined maneuvers utilized in influence operations, referred to as the BEND framework [32]. The 16 categories can be broken into narrative (based on the text messaging and the way in which it is presented) and network (based on the way in which the messaging is spread and communities are formed around the key actors) maneuvers. Each of the letters includes four maneuvers of the same starting initial. The B maneuvers (Back, Build, Bridge, and Boost) are positive network maneuvers. The E maneuvers (Engage, Explain, Excite, and Enhance) represent positive narrative maneuvers. The N maneuvers (Neutralize, Negate, Narrow, and Neglect) are utilized via negative network means. Finally, the D maneuvers (Dismiss, Distort, Dismay, and Distract) are negative narrative maneuvers. Their individual definitions can be found in Figure 3.1. This framework allows for a more defined, measured, and analytical way to compare ways in which influence tactics are employed in information operations.

While BEND has largely been utilized for analyzing behavior on Twitter (such as narratives around vaccines [24], the Chinese balloon incidents [87], and events in Indonesia [38]), this research will implement the methodology to categorize the maneuvers of sharers of the four news types used throughout this thesis on Twitter, Facebook, and Reddit.

	Community Maneuvers		Narrative Maneuvers	
	Affects who is talking/listening to who		Affects what is being discussed	
Positive	**Back**	Discussion or actions that increase the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic	**Excite**	Discussion or actions related to a community or topic that cause the reader to experience a positive emotion such as joy, happiness, liking, or excitement
	Build	Discussion or actions that create a group, or the appearance of a group, where there was none before	**Explain**	Discussion or actions that clarify a topic to the targeted community or actor often by providing details on, or elaborations on, the topic
	Bridge	Discussion or actions that build a connection between two or more groups or create the appearance of such a connection	**Engage**	Discussion or actions that increase the relevance of the topic to the reader often by providing anecdotes or enabling direct participation and so suggesting that the reader can impact the topic or will be impacted by it
	Boost	Discussion or actions that increase the size of a group and/or the connections among group members, or the appearance of such	**Enhance**	Discussion or actions that provide material that expands the scope of the topic for the targeted community or actor often by making the topic the master topic to which other topics are linked
Negative	**Negate**	Discussion or actions that decrease the actual, or the appearance of, an actor's importance or effectiveness relative to a community or topic	**Dismay**	Discussion or actions related to a community or topic that cause the reader to experience a negative emotion such as worry, sadness, disliking, anger, despair, or fear
	Neutralize	Discussion or actions that cause a group to be, or appear to be, no longer of relevance, e.g., because it was dismantled	**Distort**	Discussion or actions that obscure a topic to the targeted community or actor often by supporting a particular point of view or calling details into question
	Narrow	Discussion or actions that lead a group to be, or appear to be, more specialized, and possibly to fission, or appear to fission, into two or more distinct groups	**Dismiss**	Discussion or actions that decrease the relevance of the topic to the reader often by providing stories or information that suggest that the reader cannot impact a topic or be impacted by it
	Neglect	Discussion or actions that decrease the size of a group and/or the connections among group members, or the appearance of such	**Distract**	Discussion or actions that redirect the targeted community or actor to a different topic often by bring up unrelated topics, and making the original topic just one of many

Figure 3.1: Definitions of the 16 BEND Maneuvers, adapted from [20], [23], and discussions with the authors.

3.2.2 Applying BEND to Facebook Data

When BEND is applied to Twitter data, the networks that the maneuvers were built on were for User x User by shared hashtag, retweet, or reply. Due to the way in which Meta shares Facebook data via CrowdTangle, information about direct relationships between Facebook Pages was unavailable. Instead, the network that was used for this study (Facebook Page x Domain x Facebook Page) is more limited because it does not imply a direct interaction between the two users.

To ascertain whether differences exist between pink slime and local news shared on Facebook, a proof of concept was devised for the below analysis. To find the news site domains I was interested in studying, I consolidated a list of known pink slime sites [45] as well as the list of authentic local news sites owned by companies [97]. Using the CrowdTangle API [110], for each of the domains on the list, the 1,000 most recent instances of a link to the domain being shared on a Facebook Page was collected. In total 335,609 posts were collected from 12 pink slime organizations and sub-organizations and 8 local news organizations. Of the 12 pink slime organizations, there were 1,238 domains linked to from 285,640 posts. Of the 8 local news organizations, there were 50 domains linked to by 49.969 posts.

After performing topic modeling on the titles of the shared links, the largest common topic found pertained to elections. Since research shows that the most consumed pink slime sites are those pertaining to politics [81] and in order to analyze how these two groups discussed the same topic, the posts were filtered down to ones mentioning elections, judicial selections, and voting. This left 385 posts linking to 47 local news domains and 465 posts linking to 76 different pink slime domains. The local news posts ranged from November 17, 2022 to January 12, 2022. The pink slime posts ranged from January 27, 2020 to May 12, 2023. The posts linking to local news sites averaged a higher number of likes (27.2) than that of pink slime (21.7).

Table 3.2 illustrates the percentage of Facebook posts that contain each of the BEND Maneuvers (a note that a post can contain multiple BEND Maneuvers).

Both groups had over half of their messages falling in the Distract category. While less than 20% of documents had each of the B maneuvers, the percentages utilized by local news and pink slime are fairly equal.

Table 3.3 takes the values from Table 3.2 and subtracts the local news values from the pink slime values. This shows how much more the pink slime posts are utilizing each BEND maneuver more than the local news posts.

Most interestingly, many more pink slime posts utilize the Explain, Excite, Nuke, and Dismiss maneuvers than local news. Local news posts, however, were more heavily involved in the Neutralize maneuver. Both groups had over half of their messages falling in the Distract category.

For those sharing pink slime sites, the Explain maneuver can be seen in titles like "Ninety-three percent of Arizona Catholics say religion should not play a factor in judicial selection" and "Townsend: Audit of secretary of state's use of private funds in elections necessary 'to feel good about yes vote' on budget." The text of these posts convey statistics or quotes that provide insight into the topic. Meanwhile the messaging around Excite can be seen in posts like "Allen: 'We must restore our trust in the election process'" and "Coyne: 'We are thrilled with this year's local election results and are very proud of whatever impact we had in producing them'" Much like

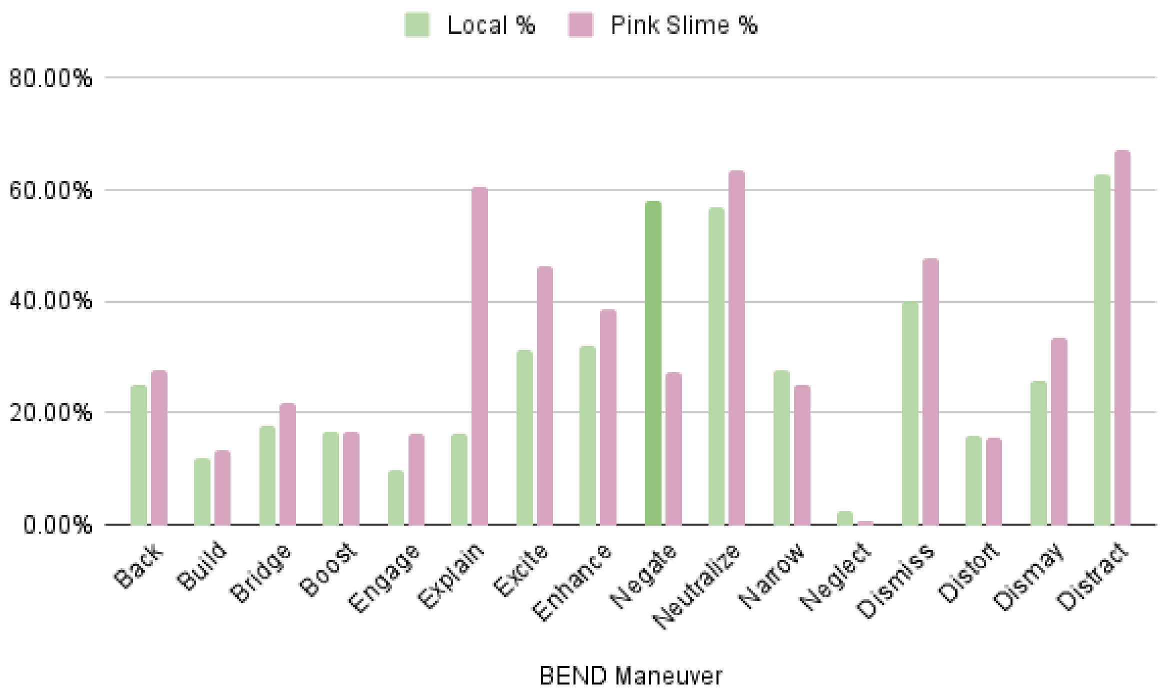


Figure 3.2: Percentage of Posts Using BEND Maneuvers by News Type

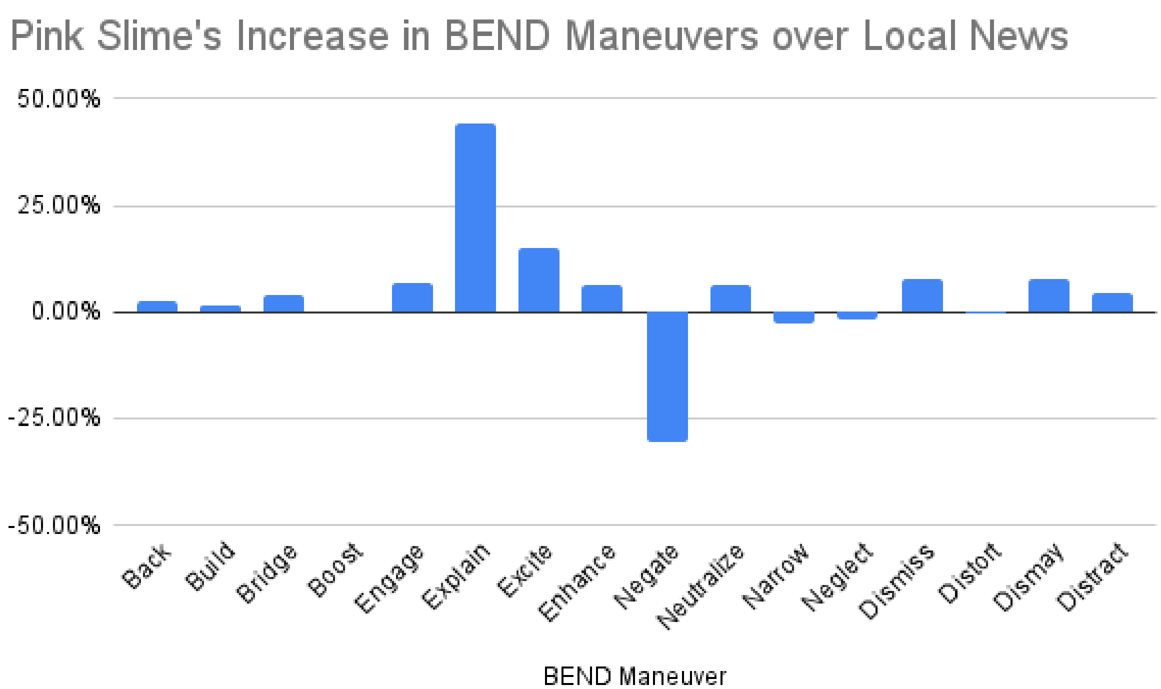


Figure 3.3: Increase in pink slime posts using BEND maneuvers over local news posts

with the Explain posts, the titles for Excite rely heavily on quotes. The narrative is one meant to bring positive emotion towards the audience. More than half of the messages fall into the Explain and Excite categories, keeping a majority of the messaging *positive* in sentiment. The remainder of maneuvers analyzed fall into the categorization of *negative* in their influence.

Examples of pink slime sites being shared with a Nuke message include "Arizona legislators protest election results, request decertification" and "Kansas legislature overrides Kelly's veto of election integrity bill." When the Dismiss maneuver is analyzed for the pink slime sites, examples include "Harbin: Georgia is experiencing 'more election irregularities because our Secretary of State could not get the job done'" and "Nagel: 'Democrats in Springfield are offering temporary election year gimmicks that attempt to trick voters instead of truly help them'", the later of which links to an article owned by the LGIS pink slime organization targeting a small city in Illinois. By referring to the state's capitol (Springfield), it gives the appearance of local news coverage; however, the same author also wrote articles for a different pink slime organization, Media Metric, targeting Grand Haven, Michigan. These Dismiss campaigns are aimed at minimizing the efforts of individuals or groups.

When the maneuvers for local news are analyzed, Neutralize (the largest increase over pink slime) is seen in messages like "Trump: People who think 2020 election was fair are 'very stupid'", "Donald Trump's response to criminal charges revives election lies" and "School elections are now political: NYC Community and Education Council voting is getting too nasty." Broadly, these messages are designed to reduce positive messaging on a topic or individual.

Both of the groups utilized the Distract maneuver heavily, a narrative maneuver that attempts to make other topics seem more important through misdirection. For pink slime this was seen in messaging like "Rats and needles hot election issue in Rogers Park Aldermanic race" and "Kansas challenger for secretary of state: Opponent's refusal to sign election integrity pledge 'should be a red flag for any Republican voter'". In local news, Distract looks like "Biden launches 2024 campaign; jury selection to start in Trump rape lawsuit; N. Dakota's near total abortion ban; and more morning headlines" (linking to an Idaho-based local news site) and "Did they vote twice in the 2022 election? RI investigating 5 cases of potential double voting."

For both sites controlled by pink slime organizations and sites controlled by organizations owning multiple local news domains, the top-ranking BEND maneuver utilized was Distract - a negative narrative maneuver. However, pink slime sites used distraction in messaging pertaining to local and state elections while the local news sites had a greater focus on national elections and events in other regions. Surprisingly, mentions of former President Trump were see in 3.2% of posts linking to pink slime sites, but he appeared in 8.1% of local news headlines; current President Biden was mentioned in only 1.5% of pink slime sites but in 8.6% of pink slime text.

Interestingly, sites controlled by pink slime organizations were shared on Facebook with more positive messaging than posts from local news organizations. Explaining and excite were utilized to highlight facts and nuance from both hyper-local and national political topics. When they used negative messaging through Dismiss, not-local reporters highlighted reasons of local concern to dismiss efforts by political parties.

Facebook Pages sharing local news sites heavily utilized the Neutralize maneuver to dismiss positive stories about national politicians and local organizations.

This current analysis only includes a few hundred Facebook posts and is limited to comparing pink slime and local news. In the next section, the BEND framework is applied to the midterms

	Real News	**Local News**	**Low Credibility News**	**Pink Slime**	**Total**
Twitter	7,010	1,803	1,186	60	10,059
Facebook	1,862	1,083	160	43	3,148
Reddit Posts	989	157	112	1	1,259
Reddit Comments	866	238	10	3	1,117

Table 3.1: Total number of posts from each dataset and news type mentioning the Fetterman v. Oz senate race.

dataset across three social media platforms.

3.3 Applying the BEND Framework to the Multi Platform Midterms Dataset

In order to get a comparison of the maneuvers used across the different platforms, I took a subset of the Midterms dataset that was discussing the same election across the three platforms. The Pennsylvania senate seat election between John Fetterman (D) and Mehmet Oz (R) was the most discussed on all of the platforms, so the Midterms dataset was filtered to posts mentioning either candidate for a more equal comparison of the nuances of how election news is shared on Twitter, Facebook, and Reddit. While John Fetterman ultimately won the election, there was heated debates that brought in some non-policy issues - namely, whether Fetterman was fit to serve following the stroke that he survived in May of the election year and the recent residency change Oz made from New Jersey to Pennsylvania to allow him to run for office in the keystone state. The number of posts sharing the various news types by platform is summarized in Table 3.1. Furthermore, the length of the text of the posts differs by platform - the average tweet length is 21.6 words, the average Reddit post is 14.6 words, the average Reddit comment is 145.5 words, and the average Facebook post is 37.7 words.

When analyzing how BEND maneuvers differ by platform, it's important to consider the ways in which users choose to share news on these platforms. In a meta-study on datasets including this midterms dataset, researchers discovered that multi-platform news links appear on Twitter and Facebook before they are shared on Reddit [83]. While the news was generally shared first on Twitter, the same articles were shared a median of 2.5 hours later on Facebook and 18-22 hours later on Reddit [83].

We observe that Twitter has extremely high proportions of posts falling into the Explain, Enhance, and Excite maneuvers. As Twitter is referred to as a "global town square" and we know that it is usually the *first* place a news story is shared, it's where we see news being broken (Excite), described (Enhance), and responded to (Enhancing). The most liked tweet in this dataset, with over 31,000 likes illustrates all three of these maneuvers happening by saying "NEW: Pennsylvania US Senate candidate Mehmet Oz staged an event for media in Philadelphia at which he consoled a woman whose family members were killed in a shooting. He didn't tell media she was actually a paid staffer. https://t.co/Bie3nq9L9M". Twitter is the only platform where post employ the *Backing* maneuver due to its unique use of the mentions feature that is not directly

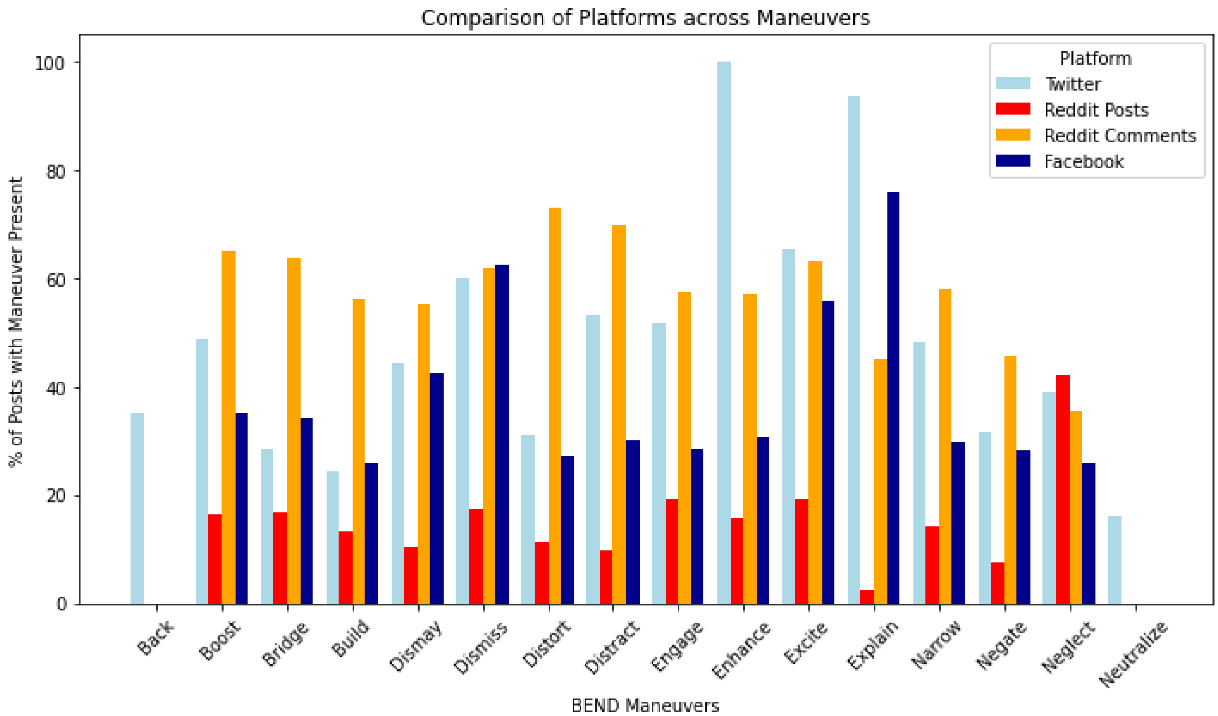


Figure 3.4: Percentage of posts from each dataset that fall into the 16 BEND maneuvers.

comparable on the Facebook and Reddit platforms; users are backing other users via their mentions, and the differences of how this is done my news type is mentioned in the Twitter section below. Like the Backing maneuver, Twitter is the only platform that sees posts utilizing the Neutralize maneuver. This is seen in tweets wherein the user is trying to expose wrongdoings of the candidates and reduce support for them. The two most liked Neutralize tweets each go after different candidates: "Did you know that Oz has had so many ethical lapses that his medical colleagues at Columbia wrote a letter protesting his affiliation? And that Columbia eventually dropped him? Check it out: https://t.co/ffvE7NNZeI" and "@JohnFetterman Fetterman chased down an unarmed black jogger with a shotgun: https://t.co/j4xchTPe0r"

We observe a contrast between the maneuvers present in Reddit posts and comments. Reddit posts show a unique phenomenon wherein the text of the post is exclusively the headline of the URL shared with the exception of the following headlines which appended an opinion as the second sentence to the post text after the news headline: "The decisive vote: Fetterman and Oz bet big on women in the Philadelphia suburbs, Vote Blue. Keep this puppy killer away from power" and "Rasmussen poll shows Mastriano and Oz closing the gap! Strong disapproval of Biden in PA. Make sure to vote PA!" Furthermore, the two largest sharers of real news and local news links on Reddit, accounting for 29% of the Reddit posts in this dataset, were explicit bot accounts that are set to post news articles. Effectively, this means the narrative maneuvers of BEND are measuring the BEND maneuvers present within *article headlines.* This is reflected in the word counts of these article titles; the average word count of all of the posts is shorter than any other platform. Despite this way of sharing posts, Reddit posts had the highest percentage of the Neglect maneuver. In particular these appear in posts that try to decrease Oz supporters' numbers by referencing gaffes he and his party made: “John Fetterman Campaign Serves Crudités at Election Night Party” a reference to a viral video of his opponent shopping for the posh snack; “Pennsylvania by the sea: Mehmet Oz implies state has Atlantic coastline”; “Head of Republican Party mocks speaking abilities of Fetterman, Biden”

Reddit comments paint a different story. Users, not bots, deliver more passionate rebukes and include news links as sources to back up their claims. This can be seen in the average number of words of each comment - almost four times higher than the next most verbose platform. The comment with the highest score shows a more typical example of how news links are used as citations within Reddit comments and shows how news that was broken on Twitter gets shared as a reference to commentary a day later in Reddit comments: “[This map](https://www.washingtonpost.com/politics/2022/11/09/fetterman-rural/) from WaPo of his margins compared to Biden just gave me so much joy. His strategy was basically to say “Hey PA. I know I won’t win every county, but I give a sh*t about the people in every county” and it paid off like hell.” This way of sharing news results in Reddit comments having the highest Boost, Bridge, Build, Distort, Distract, Engage, Narrow, and Negate maneuvers of all the platforms studied. Other example of comments containing all of these maneuvers include: “If they knew Philadelphia was good they’d be in Pittsburgh instead.Looks like Obama is hitting both sides of the state on the same day:https://triblive.com/local/obama-will-rally-with-fetterman-in-pittsburgh/; Former President Barack Obama is coming to Pittsburgh this weekend to hold a rally for Democratic Senate candidate John Fetterman. According to the Fetterman campaign, Obama will rally voters, encouraging them to turn out and vote for Fetterman and Democrats up and down the ballot.Details about specifics had not yet been released as of Tuesday, except that it will be Saturday. More details will be announced in the coming days. During the same day, Obama, Fetterman, President Joe Biden and Democratic gubernatorial candidate Josh Shapiro will also be rallying in Philadelphia.” and “Muhlenberg College released a poll showing that only about 3% of those polled are considering changing their vote after the debate. Like this poll, it still showed Fetterman in the lead.I’m sure if Emerson had asked if the debate negatively impacted voter’s opinions of Oz you would also find that half of all voters would answer yes, because we live in an extremely polarizing time.John Delano did a piece on [KDKA](https://www.cbsnews.com/pittsburgh/video/poll-finds-oz-fetterman-debate-didnt-change-voters-minds/) about it last night. The guy has been following PA politics for the better part of thirty years. I trust his opinion on the matter above that of armchair pundits be they Oz or Fetterman supports. The race will be close, so go vote.” This matter of using news links more than a day after the story broke as a citation is more unique to the Reddit comment ecosystem.

Facebook page posts are an interesting variation of Reddit posts but with slightly more context. As a function of how Facebook shares links, when an individual does not provide any additional commentary beyond a link, the default text when retrieved from the CrowdTangle API is the title of the news article as well as the subheading. This is reflective in the word counts of the text - Facebook posts containing news articles are longer than tweets and Reddit posts but shorter than Reddit comments. Facebook, where news ends up a few hours after it is broken on Twitter features posts high in Explain and Dismiss maneuvers, although not as high as Twitter and Reddit comments are for these maneuvers, respectively.

In the subsections below, the BEND maneuvers each *news type* are utilizing are analyzed by platform.

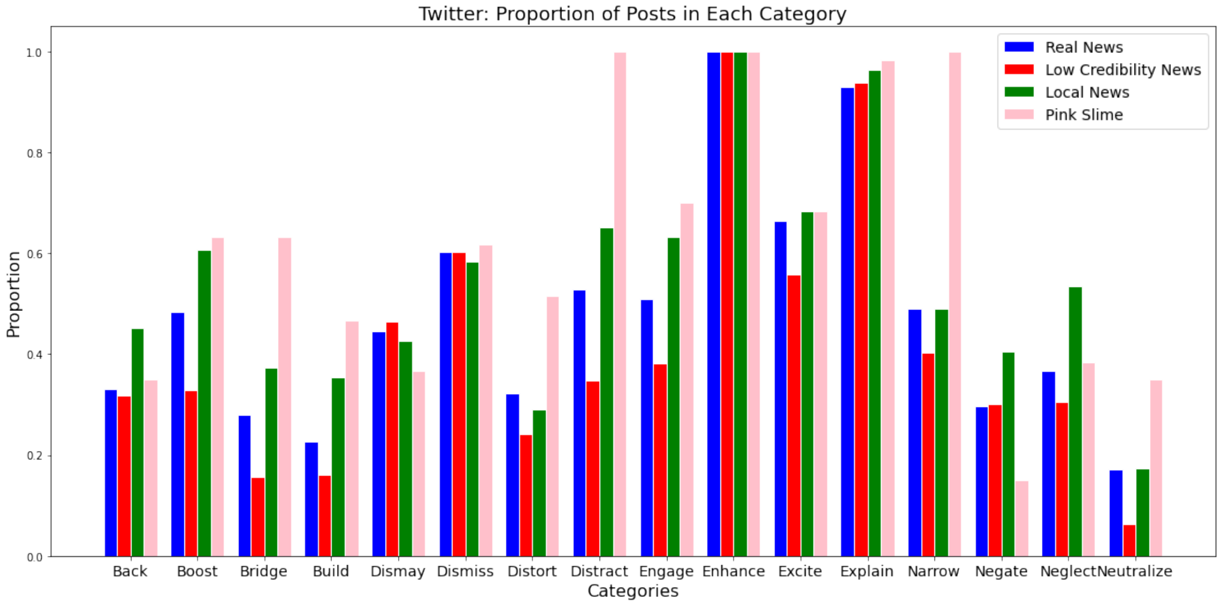


Figure 3.5: Proportion of posts by news type from the Twitter dataset that fall into the 16 BEND maneuvers.

3.3.1 Twitter

The proportion of each news type whose tweets are categorized as the BEND maneuvers is visualized in Figure 3.5. Furthermore, we observe that the length of the tweets varied by news type shared - local news (19.9 words), low credibility news (17.9 words), pink slime (24.3 words), and real news (22.7 words). Previous research [8] shows that when pink slime news articles are shared on Twitter, the text of the tweet contains the first sentence of the news article it links to 57% of the time (compared to 27% for local news and less than 1% for national news tweets). These statistics help us to interpret the differences in maneuvers utilized by these news types.

While the analysis across platforms showed Twitter to be the only platform using the backing maneuver, the ways in which these users are backing varies tremendously based on the news type shared. Tweets containing real news mentioned @DrOz, @marklevinshow, and @JohnFetterman, the official accounts of the two candidates and a Republican-leaning broadcast news show, most frequently. However, tweets containing low credibility news most frequently mentioned @BreitbartNews and @gatewaypundit, the official Twitter accounts of the low credibility news source outlets. Tweets containing local news, the news type with the highest proportion of the back maneuver, mentioned @JohnFetterman, @Will_Bunch, and @DrOz most frequently. In addition to the candidates, the most mentioned account is that of Will Bunch, a national columnist with the Philadelphia Inquirer. While there were few tweets containing links to pink slime, the following accounts were mentioned the most: @DrOz and @CheriJacobus, a national political strategist. While the high credibility news sources focused more on the candidates, the scope of the news type impacted who else was mentioned - the real news focusing on a national broadcast show while the local news referenced a local reporter following the election. Tweets containing low credibility news focused their efforts on *their news sources*. While local news focused on *local* reporters, pink slime mentioning a *national* political strategist most frequently lends further credence to the fact that these news sources have national agendas.

When looking at which tweets shared pink slime news, all but *one* of the tweets shared pink slime from left-leaning organizations. The one tweet linking to the right-leaning organization was critical of the concept of pink slime: "@SteveSchmidtSES @JohnFetterman When garbage like this comes to your home mailbox disguised as a legitimate NEWSPAPER, https://t.co/cZMyBesxts it breaks my heart that some of my neighbors will be fooled... https://t.co/fTDFwxOIVK" Due to this divide, the messaging within these tweets are largely critical of Dr. Oz for various reasons, with a particularly extreme use of the Distract and Narrow maneuvers. These strategies were deployed to counter Dr. Oz and his supporters with messaging such as: "RT @CheriJacobus: Oz's statement about meeting with Erdo$f$üan about Turkish politics contradicts past claims https://t.co/9DQDOXEdnm" and ""RT @dabbs346: Mehmet Oz claims to be tough on crime while opposing steps to actually address it #pasen https://t.co/2eOPe8eg72" Furthermore, pink slime dominates in the Boost, Bridge, and Build maneuvers. An example of a tweet using these maneuvers is "@DrOz Meanwhile, Oz has no clue has usual. The @PAGOP including @dougmastriano have cut mental health and addiction funding and voted down ever gun safety bill. PA has been under GOP rule for 30 years with the min wage at 725 since 2009. https://t.co/nqzmanrcqw" as it creates a group of Republican politicians within Pennsylvania and bridges Dr. Oz with that group.

Tweets sharing low credibility news have the highest proportion of the Dismay maneuver, with much of these negative emotions targeting Fetterman: "It's Happening: Dementia Joe will Go to Pennsylvania to Stump For Stroke Victim John Fetterman https://t.co/36HTthAiAT via @gatewaypundit", "This RADICAL MARXIST PRO-CRIME hack @JohnFetterman is UNFIT for of[fice] & a danger to PA @PAGOP @DrOz @PhillyInquirer @WNEP @PittsburghPG @TuckerCarlson https://t.co/6sX3s6270t via @gatewaypundit", and "Folks please recognize what Fettermans hometown newspaper AND Police have - HE IS NOT fit for office & will sow more chaos if elected- please vote Republican!! John Fetterman's Hometown Newspaper Endorses Dr. Oz In Pennsylvania Senate Race https://t.co/Sbl8FlAkvc" In this messaging, we see more impassioned, capitalized messages tearing down the former Lieutenant Governor of Pennsylvania.

Finally, the messaging from those sharing local news via tweets have the highest proportion of the Negate and Neglect maneuvers. Examples are seen in tweets attacking both candidates. Here, an account is promoting Oz by noting that Fetterman was deemed not worth of endorsement by the Pittsburgh Post-Gazette "@JohnFetterman Funny, your hometown paper just endorsed Oz!!! https://t.co/Gsg67Zo4z8". Meanwhile, another Twitter user shared a local Pennsylvania article about Oz's dual citizenship and added in hateful commentary to justify voting for Fetterman: "Why doesn't Dr. Oz give up his Turkish citizenship now? — PennLive letters https://t.co/hnDJYTjjSs dude he's a terrorist . A sleeper cell. Abandonment Emmett! #vote for @JohnFetterman . He's an #American . #Pennsylvania".

3.3.2 Facebook

The proportion of each news type whose Facebook posts are categorized as the BEND maneuvers is visualized in Figure 3.6. Again, we observe that the length of the tweets varied by news type

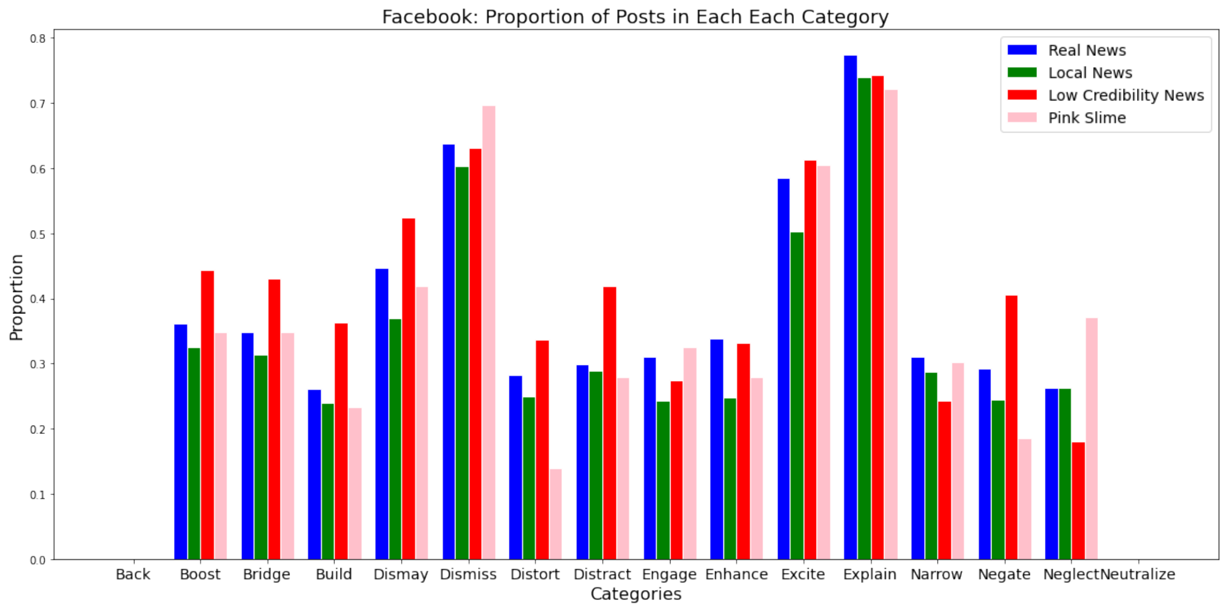


Figure 3.6: Proportion of posts by news type from the Facebook dataset that fall into the 16 BEND maneuvers.

shared - local news (26.9 words), low credibility news (29.1 words), pink slime (71.0 words), and real news (44.0 words). Much like tweets, Facebook posts containing links to pink slime have the most verbose text. However, *unlike* tweets, all of the Facebook posts containing links to pink slime sites link to websites owned by the conservative-backed Star News Network.

The posts containing pink slime also score higher than other news types for containing the Dismiss and Neglect maneuvers. Much like how Twitter users deployed the Neglect maneuver when discussing how the Pittsburgh Post-Gazetter's endorsed Oz, the official Facebook Page for the Tennessee Star shared this news via the Tennessee Star pink slime outlet with the following message seen in the post in Figure 3.7:"Pennsylvania GOP Senate nominee Dr. Mehmet Oz is touted an endorsement from the Pittsburgh Post-Gazette, a major newspaper in the state whose readership is largely in Democratic opponent John Fetterman's home county. The editorial board of the Post-Gazette, Pennsylvania's second-largest paper, questioned Fetterman's capabilities in a Sunday opinion piece. The board said Fetterman's "lack of transparency" following a serious stroke he suffered in May "suggests an impulse to conceal and a mistrust of the people." The paper also said Fetterman's "life experience and maturity are also concerns" as he has "lived off his family's money for much of his life." "

The plurality of low credibility news shares linked to Breitbart, a Republican-leaning news outlet, and all of the most engaged with posts linked to these sites from the official Breitbart Facebook Page. The posts sharing low credibility news articles have the highest proportion of the Negate, Excite, Distract, Distort, Boost, Bridge, Build, and Dismay maneuvers of any news type. The most liked of these posts contains the Build, Excite, Negate, Neglect, and Distract maneuvers with the text quoting musician Kid Rock's negative sentiment towards endorsements of Fetterman from Breitbart: "Kid Rock called Oprah Winfrey a fraud after she endorsed Pennsylvania Democrat U.S. Senate candidate John Fetterman over Republican Mehmet Oz, who got

Figure 3.7: A post from the pink slime news network discussing the Post-Gazette's endorsement of Oz.

his start working under Winfrey's wing." Another Breitbart post shows an example like the pink slime one mentioned above where a lower credibility news outlet uses a headline from a higher credibility news outlet to further sink a candidate with the Build, Bridge, Boost, Excite, and Dismay maneuvers present: "Democrat Pennsylvania Senate candidate John Fetterman's campaign spiraled into crisis Wednesday after an NBC report by Dasha Burns said it "wasn't clear" if he "was understanding our conversation" when unaided by closed captioning. Now, NBC is backpedaling about its own reporter's reporting."

The majority of the posts on Facebook linked to real news headlines, which had the highest percentage of the Enhance and Narrow maneuvers. The most interacted-with of these posts linked to NBCNews and the NYTimes (both with a left-center bias per Media Bias-Fact Check). Due to this bias, we observe that the Enhance maneuver is used to unearth headlines that are more positive towards Fetterman: "Pennsylvania Gov.-elect Josh Shapiro's more than 14-point win helped boost Sen.-elect John Fetterman to a key victory, marking the first time since the 1940s that Pennsylvania will have two elected Democrats representing the state in the Senate." Meanwhile, the use of the Narrow maneuver is seen in posts showing Oz's splintering from the Republican elected officials in Pennsylvania: "Dr. Oz declines to say if he would have voted for the recent bipartisan gun bill, which retiring Republican Sen. Pat Toomey supported." as well as posts minimizing Fetterman's potential disability post-stroke: "BREAKING: John Fetterman, the Democratic nominee for Senate in Pennsylvania, says that his stroke recovery changes everything but that he's fit to serve as senator. More on the exclusive interview: https://nbcnews.to/3fU7dCp:=:https://www.nbcnews.com/politics/ 2022-election/fetterman-says-stroke-recovery-changes-everything-s-fit-serve-senator-rcna51498".

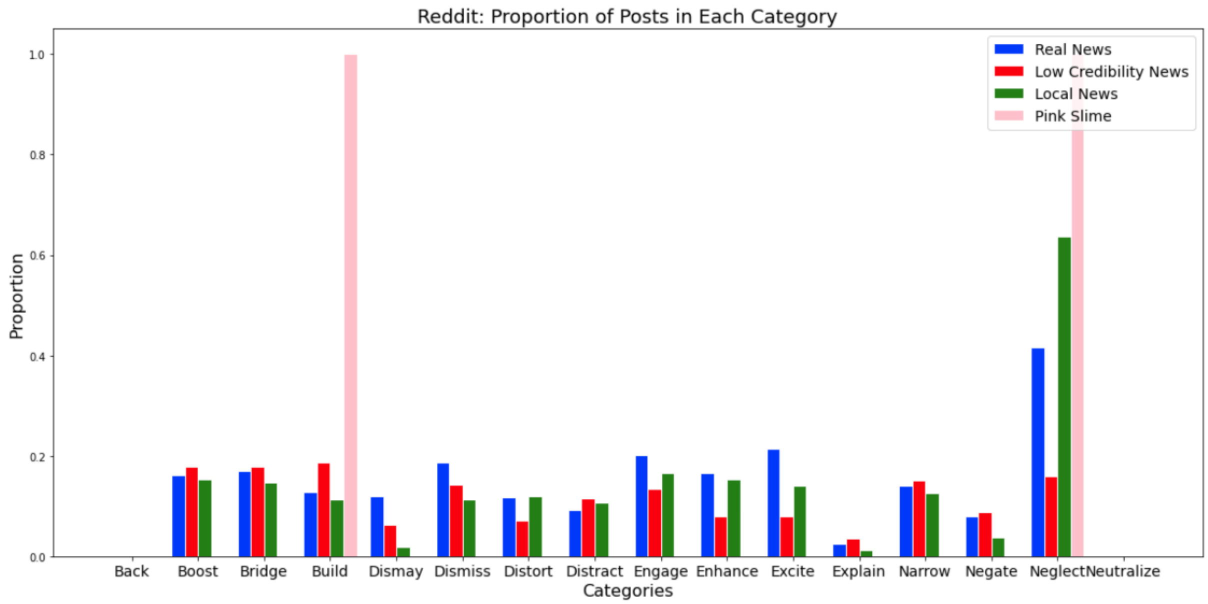


Figure 3.8: Proportion of posts by news type from the Reddit post dataset that fall into the 16 BEND maneuvers.

3.3.3 Reddit Posts

The proportion of each news type whose Reddit posts are categorized as the BEND maneuvers is visualized in Figure 3.8. Each of the news types have few average words per post - local news (14.6 words), low credibility news (13.4 words), pink slime (9 words), and real news (14.7 words). For posts, the top subreddits by news type are local news (r/AutoNewspaper, r/Triblive, r/TWTauto, r/politics, and r/Pennsylvania), low credibility news (r/BreitbartNews, r/NewsWhatever, r/Conservative, r/conservatives, and r/Republican), pink slime (r/Pennsylvania), and real news (r/AutoNewspaper, r/politics, r/FOXauto, r/TrendingQuickTVnews, and r/RedditSample). As a reminder, 29% of the posts in this dataset were created by Reddit news bots sharing articles, and all posts average just 14.6 words. While we can extract *some* maneuvers from these limited messages, it follows that these posts have fewer maneuvers present in them than their longer counterparts on Twitter, Facebook, or Reddit comments.

There is only one example of a pink slime article being shared by a Reddit post, which would explain why 100% of the pink slime posts in the plot show that the Build and Neglect maneuvers are present, as that post shares a Republican-backed link to poll results showing “Oz Leads Fetterman in Pennsylvania Senate Race.”

Outside of the one post to a pink slime link outlier, local news has the highest percentage of posts falling under the Neglect maneuver. This is seen in posts like the following that ice out Oz from his previous relationship with television host, Oprah: “Oprah backs John Fetterman over Mehmet Oz in Pa. Senate race.” In other rare shows of Reddit users adding their own personal feelings into news headlines from the links shared, the following poster inserted his categorization of Oz in the quotations to ostricize the candidate and remove potential support for the candidate: “Midterm Election Results: Fetterman wins Pennsylvania Senate race, “Big

Scambag/Forced Birth/Animal Abuse" Oz concedes."

The overwhelming majority of posts in this dataset link to the real news type. As we've observed in Chapter 2, Reddit has a higher percentage of real news posts and a lower percentage of low credibility news than other platforms, possibly due to its moderation efforts. The real news posts then have the highest percentage of posts categorized in the Dismiss, Engage, Enhance, and Excite maneuvers. We see posts supporting both sides of the political spectrum engaged in the Dismiss maneuver against the opposing candidates with posts such as "[National] - At Fetterman Rally, Obama Mocks Oz and Tells Crowd to Vote for Democracy — NY Times" and "Fetterman attempts to wrangle support from GOP voters after he said Republican base is xenophobic, homophobic". Posts with the three positive narrative maneuvers are seen in some of the victory messaging about Fetterman: "John Fetterman wins Pennsylvania Senate race, defeating TV doctor Mehmet Oz and flipping key state for Democrats." Meanwhile, Oz supporters shared a message designed to excite those not pleased with the results: "Tucker Carlson says it would be 'absurd' for voters to accept Pennsylvania election as legitimate if John Fetterman wins."

All but four of the posts linking to low credibility news shared news from the Breitbart news outlet, leading towards a right-leaning bias in the messaging. These posts contained higher percentages of the Boost, Bridge, Distract, and Narrow maneuvers than the other news types shared via Reddit posts. The Narrow maneuver is seen in posts trying to detach Fetterman from important voting blocks, like gun owners: "Exclusive Video: Democrat John Fetterman Wants to Ban 'Ownership' of Rifles, Not Just Sale" as well as those critizing Oz for his New Jersey residence: "'Jersey Shore Gisele': PA Democrat Fetterman Attacks Oz over NJ Home, but His Wife Lived There." The Distract maneuver was deployed to criticize Fetterman's stance on drugs: ("Democrat John Fetterman Refuses to Commit to Legislation to More Easily Lock Up Fentanyl Dealers", "Democrat John Fetterman Applauded Oregon's Decriminalization of Heroin, Hard Drugs"). The posts engaging in Boost are also the same posts categorized here as Bridge, by accomplishing both maneuvers in posts bridging the candidates with previous presidents that endorsed them, such as: "Donald Trump: John Fetterman Is the 'Single Most Dangerous Democrat Seeking to Join Congress'" and "Joe Biden Rallies with Barack Obama to Prop Up John Fetterman in Pennsylvania."

3.3.4 Reddit Comments

The proportion of each news type whose Reddit comments are categorized as the BEND maneuvers is visualized in Figure 3.9. Again, we observe that the length of the comments varied by news type shared, but for Reddit comments they are substantially longer than the other platforms and types - local news (62.9 words), low credibility news (528.7 words), pink slime (13.7 words), and real news (164.2 words). For comments, the top subreddits by news type are local news (r/Politics, r/Pennsylvania, r/philadelphia, r/pics, and r/pittsburgh), low credibility news (r/Conservative, r/Firearms, r/GeopoliticsIndia, r/IndiaSpeaks, and r/Yankee_Clickers), pink slime (r/Pennsylvania, r/philadelphia, and r/worldnews), and real news (r/politics, r/Pennsylvania, r/moderatepolitics, r/philadelphia, and r/Pennsylvania_Politics).

There are only 3 Reddit comments containing links to the pink slime sites, but we see all of them fall into the Neglect maneuver. All of these comments linked to the left-leaning pink slime site, The American Independent, and they were used as a source to a statement bashing Oz in fa-

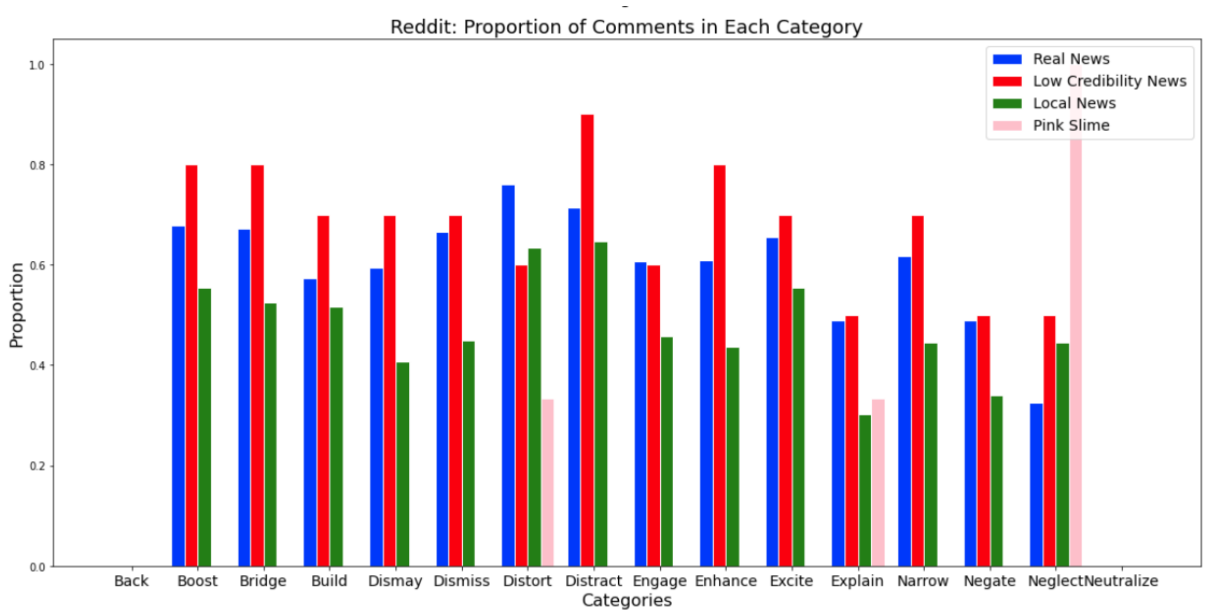


Figure 3.9: Proportion of posts by news type from the Reddit comment dataset that fall into the 16 BEND maneuvers.

vor of Fetterman: “https://americanindependent.com/mehmet-oz-erdogan-turkey-pennsylvania-senate/ He’s bragged about it. Educate yourself.People recover from strokes. In a year Fetterman is gonna be a lot better. In a year Oz is still going to be the worst possible choice.”

There are only 10 Reddit comments containing links to low credibility news, but they lead the way in their usage of the following maneuvers - Boost, Bridge, Build, Dismay, Dismiss, Distract, Enhance, Excite, Explain, Narrow, and Negate. These posts tend to be long tirades against political issues that arose during the election, such as the use of mail in ballots which one user opposed since it kept voters from viewing the debates prior to voting, using low credibility news articles to bolster his claims: “It’s called a ’though experiment.’ Consider, if you will, that there are people out there who will blindly pull the D (or R) lever, without much thought. But, they didn’t know just how bad of a shape that Fetterman was in - yet, they sent their votes in early. How many would have changed their minds, if they saw that debate prior to voting? Also, this only took about a second to find (Bing search - I trust it to be apolitical a lot more than I trust Google): [https://www.breitbart.com/politics/2022/10/26/report-half-of-pennsylvania-vote-by-mail-ballots-cast-before-fettermans-disastrous-debate/](https://www.breitbart.com/politics/ 2022/10/26/report-half-of-pennsylvania-vote-by-mail-ballots-cast-before-fettermans-disastrous-debate/) 48% of requested mail-in votes were received prior to that debate. Add in the mail being the mail, and it’s quite likely that well north of 50% of those mail-in early votes were cast before the debate.”

Again, comments containing links to real news make up the majority of this dataset, and they have the highest proportion of their messaging falling into the Distort maneuver. Like other comments, we see more thoughtful messages responding to other Reddit users, with detailed stances backed up by news sources. The Distort maneuver was used to call into question hypocrisy from the Republican platform in the following comment with the New York Times cited: “* Nationwide, most Republicans rail against liberal elites and then block a $15 an hour minimum wage, paid leave laws and workplace safety protections. * They stymie bills to help workers unionize, and top it off by starving the National Labor Relations Board of funding, even as it faces a surge of union election requests. * Several Republican attorneys general have sued to stop wage hikes for nearly 400,000 people working for federal contractors. * Republicans also opposed extending the popular monthly child tax credit that helped so many working families afford basic necessities. * The issues section on the campaign websites of Mr. Vance and Dr. Oz contain virtually no labor policy. Howling about China, as they do, isn't a comprehensive labor plan. ... from the article https://www.nytimes.com/2022/10/29/opinion/election-workers-republican-oz-vance.html" Others used Distort, citing a CNN polling article, to question a user's interpretation that "crackpots" would only vote for Fetterman: "Do you have a source for that? The main exit poll I'm seeing is CNN's:https://www.cnn.com/election/2022/exit-polls/pennsylvania/senate/0 They cited that in articles as being important, but the exit poll doesn't actually ask if that's important to how people decided to vote. And more to the point, they didn't ask if being a crackpot matters. So this poll, if they had asked me, would not have captured why I voted for Fetterman instead of Oz."

3.4 Conclusions

Many of the differences in BEND maneuvers present in different social media platforms are functions of how users more broadly share news on the platform. While owners of Facebook pages and posters to Reddit tend to post a story and its headline as a starting point for conversation, Twitter users take pride in *breaking* the news and responding to it with their commentary. A day later, Reddit commenters use those same news links to serve as citations to their opinions and political rebukes. When specifically looking at how pink slime is shared across the platforms, it is important to note that all but one of the tweets containing pink slime linked to *left leaning* pink slime sites while all of the Facebook posts linked to *right leaning* pink slime sites. Across the platforms, we observe that users sharing pink slime in their posts do so with more negative maneuvers than other news types. On Twitter, this is seen in their higher prevalence for the Distract and Narrow maneuvers; on Facebook, with Dismiss and Neglect. For Reddit, we see both posts and comments fitting into the Neglect maneuver.

Chapter 4

Finding New Sources of Pink Slime

4.1 Research Questions

The key research question for this chapter is:

- How can we detect new sources of pink slime sites?

4.2 Related Work

The current expert in discovering new sources of pink slime was established by Priyanjana Bengani, a senior research fellow at Columbia Journalism School's Tow Center for Digital Journalism. She discovers these new sites by collecting identifiers like IP addresses of the sites and servers as well as tracking numbers [17] of domains that were related or sharing IP space. Her discoveries were that pink slime sites seldom shared Google Analytics IDs, but some of the Metric Media sites shared three of the identifiers. Meanwhile, Local Government Information Services (LGIS), Franklin Archer, and LocalityLabs (all losely related to Metric Media) shared three other unique IDs. Other pink slime networks had additional identifiers (Quantcast and NewRelic IDs) in common [15]. Her compilation of these sites is publicly available [63] and serves as a reference label for this thesis' research.

While not explicitly searching for pink slime sites, other researchers have recognized clusters of websites using the same third party analytics trackers to identify malicious online campaigns [106]. These sites, while seemingly unrelated, were acting together and new sites were able to be discovered by finding other domains using the same identifiers.

Both methods utilized by [15] and [106] require substantial manual input of domains and collection of third party identifiers which is time and labor intensive. Furthermore, these methods are only capable of finding new domains within an existing network of known pink slime organizations.

Looking outside research strictly identifying malicious news domains owned by larger parent organizations, PageRank is an established algorithm that researchers have modified to find misinformation domains using text commonalities between sites of unknown validity with those known to be spreading false news [115]. Other adaptions used PageRank to rank Twitter users who act as authorities and "give" credibility to the events they Tweet [57].

In a precursor to PageRank, Kleinberg proposed a model that found authoritative webpages for search topics by analyzing the linkage to other sites and assigning values that designate the site as an authority or a hubs (a site that links to many related authorities) [69]. In the search engine optimization space, researchers have also used known labels of news sources to discover potential new sources of misinformation [35].

4.3 Pink Slime Network Spread

We begin our analysis with an exploratory network analysis of the news domain sharing structure to better understand how these sites are shared on Facebook. First, we extract the Facebook pages that are sharing known sources of pink slime and construct a network diagram, where there are parent organizations represented as green nodes, and pink slime domains represented as grey nodes. The grey nodes are linked to the green nodes if the pink slime domains are originated from the parent organization. The grey nodes representing pink slime domains are then linked to each other if a Facebook page shares both domains. When looking at networks of Facebook pages to the parent organization of the pink slime sites they shared, there are many Facebook pages that are posting links to pink slime sites owned by multiple parent organizations. One such Facebook page, "Democrats of the Alachua County Area", linked to 5 different parent organizations of pink slime sites. Most of the Facebook pages sharing these sites were smaller (under 1,000 followers) and targeted a hyper-local area. See Fig. 4.1.

Additionally, Chapter 2 shows us that agents sharing pink slime are most likely to follow up by posting *more* pink slime. When analyzing this network, 20,065 Facebook Pages shared over 1.4 million posts that linked to these five known parent organizations. 13.4% of these Facebook Pages shared news linking to more than one pink slime parent organization.

This insights from the network analysis inform a new strategy of finding local community Facebook pages sharing news from multiple pink slime parent organizations. Using this strategy, we can find new pink slime websites belonging to emerging pink slime parent organizations as they are shared on Facebook pages containing topics pertaining to local communities. In order to solidify this phenomenon, a network-based score was established and named the Non-Credibility Score. We then put the score into a machine learning model to determine if it is a successful method of categorizing news domains. While some of the results of this research was previously published [71], this chapter includes newer additions to the model, specifically the natural language features such as when location names are in the domain as well as a local news label and applications to multiple platforms and countries.

4.4 Non-Credibility Score (NCS)

The Non-Credibility Score is a feature of each unlabeled news domain in a dataset, ranging from [0,1] and provides information regarding the credibility of the news site, based on the known credibility of other sites shared by the same Facebook Pages that the news site is shared on.

The Non-Credibility Score involves creating network features of the news domains to signal their credibility based on which hubs are sharing the authorities and what other authorities those

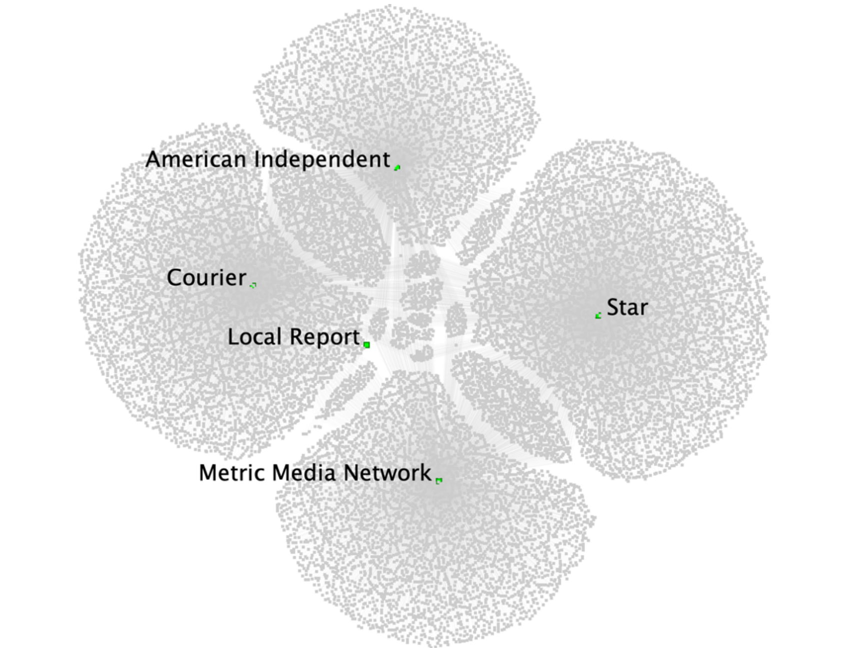


Figure 4.1: Network visual of Facebook Pages linking to Parent Organizations of Pink Slime Sites

hubs were sharing. In this research, Facebook pages act as the hubs, and the domains they share are the authorities. The labeled news domains are utilized to test the validity of the features. From there, the larger, unlabeled domain dataset was analyzed using the set of features to find new source of pink slime.

Figure 4.2 illustrates the process of assigning Non-Credibility Scores for news sites. It uses concepts from network science as well as the HITS algorithm, creating a network of two node types - Facebook Pages and News Sites. If a Facebook Page shares a News Site, then there is a link between the two nodes. In network terms, if we create the network of News Sites x Facebook Pages x News Sites, then the NCS is the percentage of first-degree neighbors that are known sources of pink slime or low-credibility news. In HITS terms, we are looking at which hubs (Facebook Pages) are sharing which authorities (News Sites).

In order to determine the credibility of domains, the following equation was implemented. First, each Facebook Page present in the dataset was given a score to indicate the proportion of content it shared that originated from a known source of low credibility news or pink slime, referred to as the Noncredible Sharer Score (NCSS). This value ranges from [0,1] and relays the percentage of content a given Facebook Page shares that is of known low credibility or pink slime. The higher the Noncredible Sharer Score, the lower the credibility of the Facebook Page. The equation is elaborated below:

$$\text{Noncredible Sharer Score}_k = \frac{1}{n}\sum_{j=1}^{n} I_j \tag{4.1}$$

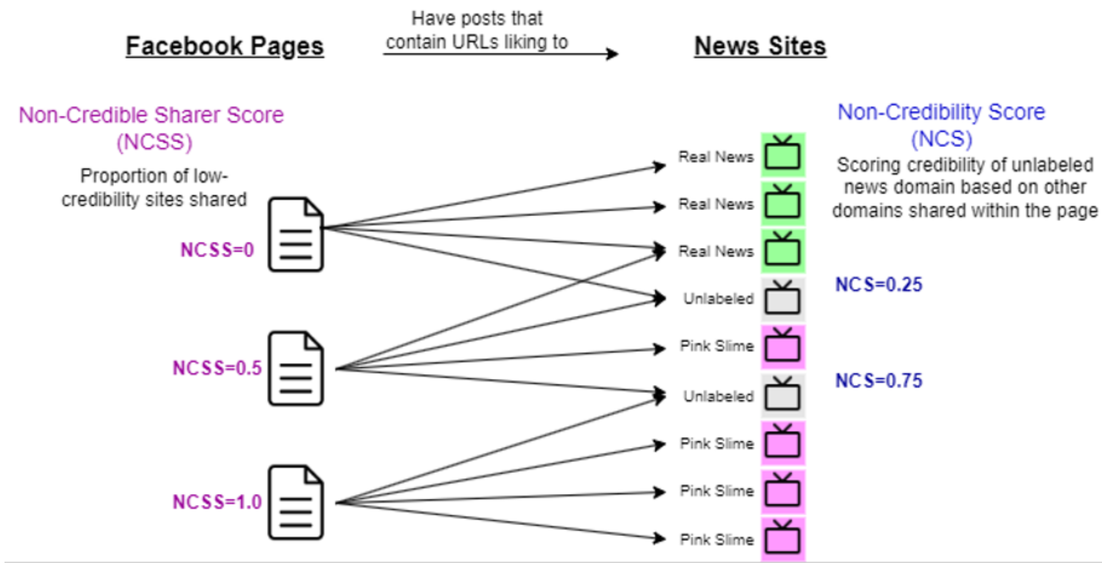


Figure 4.2: Visualization of the Noncredibility Score

$$I_j = \begin{cases} 1 & if \text{domain j is a labeled pink slime or low credibility news site} \\ 0 & otherwise \end{cases}$$

where k is Facebook Page linking to domains and n is the number of domains shared by Facebook $Page_k$

Then, the domains could be scored by averaging the Noncredible Sharer Score of the pages that shared the domain:

$$\text{Noncredibility Score}_z = \frac{1}{|K|}\sum_{k=1} \text{Noncredible Sharer Score}_k \tag{4.2}$$

where K is the set of Facebook Pages sharing a given unlabeled domain and z number of unlabeled domain shared by a Facebook Page

4.5 Validation: Predicting News Category

In order to verify the validity of using the NCS to find pink slime, we constructed a machine learning model that differentiates between pink slime, local news, low-credibility news, and real news; based off our analysis of how pink slime spreads, we expect that the NCS will help assign correct news labels to unknown news sites. To do so, we first perform feature engineering to extract input features. Table 4.1 below outlines which features are selected as inputs for the machine learning model. These features were designed to provide further insight into whether a news site is particularly credible and popular among social media sharers and viewers. Since most of the pink slime sites in our dataset have the name of an area in the United States in their domain, we also included whether the domain name of the link shared contains a location. We

Feature	Description	Type
Credibility-Based Features		
Non-Credibility Score (NCS)	Measure of credibility of news site	Float [0,1]
Domain Contains Location	Whether the domain name corresponds to a value within a US-based gazetteer	Boolean [0,1]
Popularity-Based Features		
Average Number of Likes	Average number of likes posts sharing the domain received	Float [0,inf)
Unique Pages	Number of Facebook Pages that share the domain	Integer [0,inf)
Number of Occurrences	Number of times the domain appears in the dataset	Integer [0,inf)

Table 4.1: List of features included in the model for each domain.

parsed the domain name to see if it contained a US-based gazetteer [1]. The names of each county, city, region, and town in the country were analyzed to see if that specific string was included in the domain name.

The features designed to capture the popularity of a news domain include the average number of likes that the posts sharing the domain received, the number of unique Facebook Pages that share the domain, and the number of occurrences the domain appears within the dataset.

These features were utilized as inputs in a machine learning model to predict whether or not a domain in the dataset was pink slime. 70% of the known sources of low credibility news, local news, pink slime, and real news are utilized in training the model to see if they can accurately predict the legitimacy of the 30% of withheld domains. The 30% of withheld domains are treated as unlabeled news sources during the calculation of the network measures. The output is the news category label that the model believes best captures what type of news the unlabeled domain belongs in. The XGBoost model, an open-source implementation of the gradient boosted trees algorithm, is used to perform the training and validation due to its high efficiency and accuracy [41]. After learning of the performance of the model, I analyze the feature importances using sklearn's feature_importance function to see which of the input features played the largest role in assigning the label to the domains' classifications.

[1] census.gov/geographies/reference-files/time-series/geo/gazetteer-files.html

News Domains	Number of Domains	Average Number of Occurrences	Average Number of Likes	Average Number of Pages	Domain Contains Location
Pink Slime	71	34	172	8	61%
Low Credibility News	254	193	144	42	18%
Real News	1774	464	57	132	38%
Local News	913	80	18	11	58%

Table 4.2: Statistics for Facebook 2020 dataset

4.6 Results

For the initial proof of concept, I analyzed over 2.9 million Facebook posts related to 2020 elections and the ReOpen movement with news domains within the messaging. After matching the shared domain names with the Media Thesaurus we constructed to label the known sources of news, we found the statistics in Table 4.2 pertaining to the collected dataset. A vast majority (96%) of the domains shared in this dataset are unlabeled by the Media Thesaurus. This underscores the challenges of finding and identifying pink slime domains (or even low-credibility news) with a vast swath of unlabeled domains. An analyst would have to spend weeks, manually reviewing tens of thousands of URLs.

Per Table 4.2, of the (minority of) news domains that are labeled, only a handful are of known pink slime or low credibility news labels. While this means that Facebook may be doing a good job of removing any particularly egregious fake news sites posted to their platform, the majority of domains shared that are not well-known enough to be included in our Media Thesaurus indicates that an automated labeling of credibility of these news sites would benefit the platform (and analysts).

Looking at some of the other input features of our model, we see that real news domains are shared the most frequently (1774 domains in this dataset, shared an average of 464 times by an average of 132 Facebook Pages) while the 71 pink slime domains are only shared an average of 34 times per domain by 8 Facebook Pages. Since pink slime targets a smaller demographic, it follows that it would be shared with smaller, presumably more geo-specific community groups than the broadly-targeting real news sites. Unsurprisingly, local news and pink slime have a majority of their news domains containing U.S.-based gazetteers while real news and low credibility news have a minority of domains including such phrases. Finally, the lower credibility news types receive far more likes per domain (172 and 144 for pink slime and low credibility news, respectively) than their higher credibility news domain counterparts (18 and 57 for local news and real news, respectively).

In order to assess the credibility of news domains, I introduced the new network-based feature called Non-Credibility Score (NCS). As mentioned in Section 4.4, the NCS is calculated based on the other news domains that are shared within the Facebook Page that shared the news domain in focus. In Table 4.3, I list the average NCS by labeled news types in this dataset. Unsurprisingly, I find that the average NCS for pink slime and low credibility news sites are close; as I defined

News Domains	Average Non-Credibility Score (NCS)	Standard Deviation
Pink Slime	.289	.292
Low Credibility News	.351	.285
Real News	.003	.012
Local News	.001	.004

Table 4.3: Average Non-Credibility Scores (NCS) of the Facebook 2020 training data by news type.

	Precision	Recall	F1 Score
Naïve	0.01	0.25	0.00
Logistic Regression	0.33	0.37	0.35
Random Forests	0.62	0.45	0.48
Gradient Boosted Classifier	0.65	0.45	0.49
XGBoost	0.62	0.43	0.45

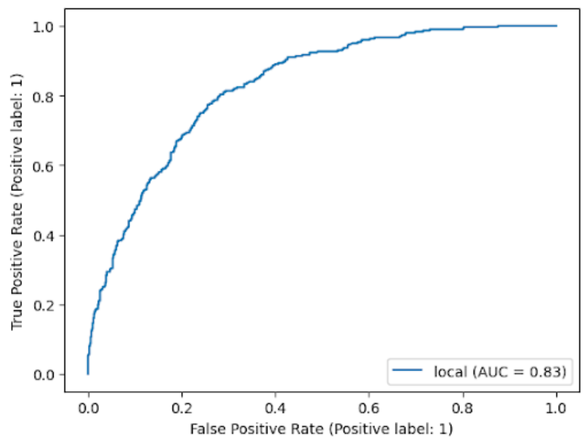
Table 4.4: Macro average accuracy results for news category prediction model

earlier, both news types are of similar credibility, just differing scope.

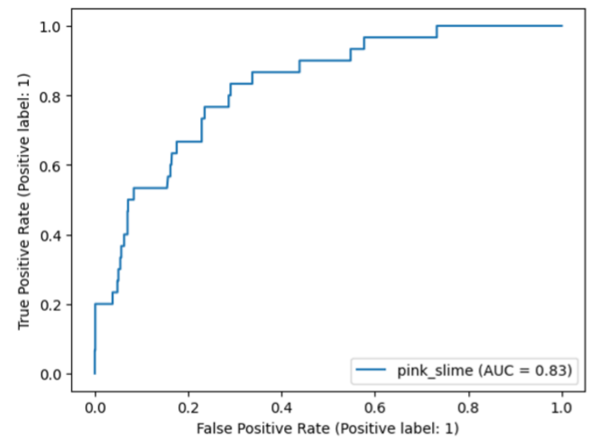
Table 4.4 shows the results of the macro accuracy using different machine learning models that take the input features of Table 4.2 and assigns the news domains to a given news label. The macro accuracy averages the performance metrics of each news label. Assigning all of the values to the dominant class (real news), as the Naïve classifier did, resulted in the lowest precision, recall, and F1 score. Since there is an issue of class imbalance as described earlier in this section, we used the macro score to take into account the results for the minority news labels. By including the input features, the precision more than doubles in classifiers like Random Forests, Gradient Boosted Trees, and XGBoost. Overall, there are only modest gains in the Recall and F1 scores, indicating that false negatives are a more common occurrence than false positives. While the overall accuracy value peaks at about 65%, the intent of this classifier is to use the classification as a start point of analysis to find higher likelihood domains faster, before performing deeper investigation.

One of the top-performing models, XGBoost, was then used to understand the classification performance via ROC curves. XGBoost does well in data science settings because it uses an efficient and optimized gradient boosting tree method to classify data. ROC curves plot model sensitivity by showing the performances of the models across varying classification thresholds, and curves with higher area under the curve (AUC) indicate better performances. These curves can be seen in Figure 4.3, with an overall AUC of 0.83 of classifying pink slime sites; per experts, this is an acceptable value, showing that this model has merit [60].

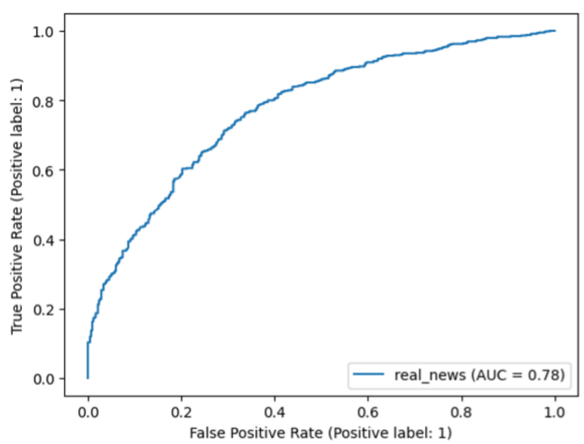
The below Figure 4.3 illustrates the strength of the model in predicting pink slime (despite the low occurrence) through the ROC curve. The similarity in the AUC for all four news categories to be above 0.78 shows the ability of the model to be applicable to at least these four different



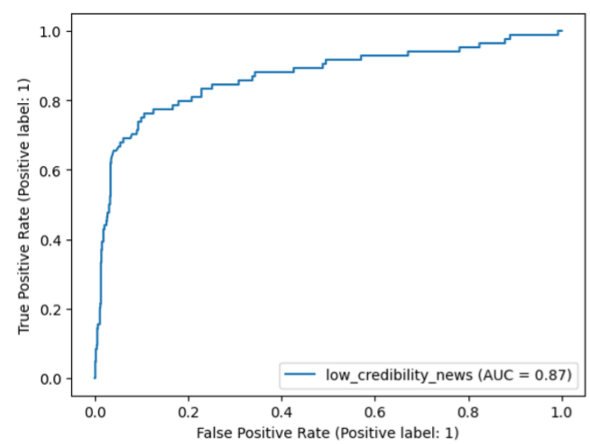
(a) Local News ROC Curve



(b) Pink Slime ROC Curve



(c) Real News ROC Curve



(d) Low Credibility News ROC Curve

Figure 4.3: ROC Curves for Predicting News Labels of the 2020 Facebook Dataset

news types. When looking at how well the model could predict the presence of low credibility news, the ROC is slightly stronger with an AUC of 0.87.

Using one of the top performing models, XGBoost, we look at the feature importance of the input features in Table 4.5. The strongest feature used for prediction of news label is the average number of likes a domain received when shared on Facebook with the NCS as the second-most important feature. Interestingly, the NCS was the most important feature in the model when initially trying to classify whether something was real news, low credibility news, or pink slime. By including the local news category, the average number of likes rose in importance, likely due to the differences in averages likes seen in Table 4.2. This top two features indicate how the combination of network and popularity features are important to predict news labels. With the NCS combined with other features extracted from the posts, we can quickly surmise whether a set of currently unlabeled news domains is likely to be pink slime based on the type of news shared by the Facebook pages sharing the site.

While these results are very promising for lending support to the NCS as a way to find pink

Feature	Importance
Average Number of Likes	0.469
Non-Credibility Score	0.248
Unique Pages	0.122
Domain Contains Location	0.082
Number of Occurrences	0.080

Table 4.5: Feature Importance of the XGBoost Model for the 2020 Facebook dataset.

	Midterms Twitter Data	Midterms Facebook Data	Midterms Reddit Data
Number of Posts	1,306,829	126,995	13,839
Number of Sharers	351,732	28,431	1,157

Table 4.6: Number of posts in each of the datasets and the number of people (or subreddits, in the case of Reddit) who shared them.

slime sites, it's important to mention the size of the dataset. This dataset had 2,914,911 posts coming from 277,601 Facebook pages.

4.7 Multi-Platform Validation on 2022 US Midterm Elections Dataset

Now that we have a working model, I decided to perform an an external validation to datasets from three platforms - Facebook, Twitter, and Reddit. For each of these platforms, I used the same date range and keywords, all pertaining to swing state elections during the 2022 United States Midterm Elections. The goal is to see if the NCS can be used in a platform-agnostic manner. While we know that the NCS can help us to find pink slime on a sufficiently large Facebook dataset, we will analyze its ability to find pink slime on other platforms in this section.

4.7.1 Comparison of Datasets

As a reminder, the above example of Facebook data from 2020 included almost 3 million posts shared by over a quarter of a million Facebook pages. In the following examples from the 2022 U.S. Midterm Election, each of our datasets is smaller than the Facebook 2020 example. Per Table 4.6, the only dataset with over a million posts is the Twitter dataset. The Facebook and Reddit datasets have much fewer posts (126,995 and 13,839, respectively) and fewer people sharing them.

Furthermore, part of the NCS revolves around creating Non-Credible Sharer Scores (NCSS) for the accounts that share a given news domain. In the case of the Facebook data above, we used Facebook pages as the "Sharer." Since a Facebook page isn't fully universal, we needed to determine what the sharers would be for the Twitter and Reddit data. For the Twitter dataset, I opted to use the individual user sharing the Tweets, as there are no groups Twitter users join. For

News Domains	Number of Domains	Average Number of Occurrences	Average Number of Likes	Average Number of Users	Domain Contains Location
Pink Slime	10	1,738	1.45	1570	51%
Low Credibility News	28	690	1.58	506	31%
Real News	429	1350	2.07	959	38%
Local News	146	24.4	0.86	15.5	50%

Table 4.7: Statistics for Twitter Midterms dataset

Reddit, I *could* have used the individual users in the dataset, but due to the nature of the way this dataset was collected (searching keywords as opposed to a given subreddit for posts), I seldom saw Reddit users making repeat posts in this dataset. Instead, I opted to make the sharers the subreddits (community groups through which Reddit is organized and news is posted to). As seen in Table 4.6, the only dataset with *more* sharers than the 2020 Facebook dataset was the Twitter dataset with 351,731 Twitter users sharing the Tweets. Again, the Facebook and Reddit Midterms sharers were much smaller, with only 28,431 and 1,157 sharers, respectively.

4.7.2 Twitter Midterms Dataset

The features of the news types within the Twitter Midterms dataset followed similar patterns to what was seen in the 2020 Facebook data. Per Table 4.7, the majority of the news domains were real news, with pink slime being the lowest occurring. Again, pink slime and local news (unlike the other two news types) had a majority of their domains contain a location name. The main differences between the Twitter Midterms dataset and the Facebook 2020 dataset are that the average number of likes is *highest* for real news on Twitter. Furthermore, the average number of users sharing pink slime is the highest (1570), while it was only 8 Facebook pages in the example above. This is due to more bot-like behavior on the Twitter platform than on Facebook. The pink slime domain, arizonasuntimes.com was shared by 14,375 users while the next highest domain, tennesseestar.com, was only shared 997 times. Given there are only 10 pink slime domains in this dataset, there is increased sensitivity to bot or anomoalous activity.

While the popularity and user stats may be subject to volatility, the NCS remains a constant. Per Table 4.8, pink slime and low credibility news have the highest average Non-Credibility Score in the Twitter Midterms dataset.

When the XGBoost model is applied to this dataset, we observe in Table 4.9 that the NCS rises to the highest feature importance. The number of unique users had the second-highest importance, but as mentioned previously, bot activity played a role in this feature.

When the XGBoost model is applied to this dataset, we observe a macro average precision and recall value of 0.33 with an F1 score of 0.32. While these numbers fall short of the 2020 Facebook dataset, the class imbalance is even stronger in the Twitter dataset. Looking at the ROC Curves in Figure 4.4, we see that pink slime actually had an AUC of 0.96 (higher than the 2020 Facebook dataset), with the lowest AUC being for real news (0.69).

News Domains	Average Non-Credibility Score (NCS)	Standard Deviation
Pink Slime	.351	.127
Low Credibility News	.432	.261
Real News	.002	.006
Local News	.001	.002

Table 4.8: Average Non-Credibility Scores (NCS) of the Twitter Midterms training data by news type.

Feature	Importance
Non-Credibility Score	0.336
Unique Users	0.257
Average Number of Likes	0.196
Number of Occurrences	0.119
Domain Contains Location	0.092

Table 4.9: Feature Importance of the XGBoost Model for the Twitter Midterms dataset.

The results for applying the NCS to the Twitter Midterms dataset are very promising. It shows the first example of using a non-Facebook dataset to show the importance of the NCS in finding pink slime.

4.7.3 Facebook Midterms Dataset

While we have seen that the NCS can predict pink slime through Facebook data, the Facebook Midterms Dataset is 22 times smaller than the 2020 Facebook data. Below we will perform the same analysis and model as the original 2020 Facebook data to see how the size differences change the outcome.

To start, the Facebook Midterms and the 2020 Facebook datasets have the similarity of the gazetteer presence and average number of likes per Table 4.10. However, the number of likes received by pink slime sites (29.6) is only *slightly* higher than that of real news (26.9) whereas it was much more profound of a difference in the 2020 Facebook dataset (172 to 57).

As expected, and seen in the 2020 Facebook dataset as well as the Twitter Midterms dataset, the NCS of the Facebook Midterms dataset is highest for the lower credibility domains, seen in Table 4.11. In fact, the NCS of the pink slime sites (0.557) is much higher than that of the previously mentioned two datasets (0.289 and 0.351), indicating a higher signal and presence of lower credibility news.

Once the features are input into the XGBoost model, similar to the Twitter Midterms dataset, the NCS proved to have the highest feature importance (per Table 4.12). The average number of likes was the second most indicative feature, dropping slightly from its prominence in the 2020 Facebook dataset.

News Domains	Number of Domains	Average Number of Occurrences	Average Number of Likes	Average Number of Pages	Domain Contains Location
Pink Slime	21	10.0	29.6	4.0	57%
Low Credibility News	73	28.9	75.7	6.9	11%
Real News	951	26.2	26.9	11.9	16%
Local News	425	10.5	7.81	2.0	51%

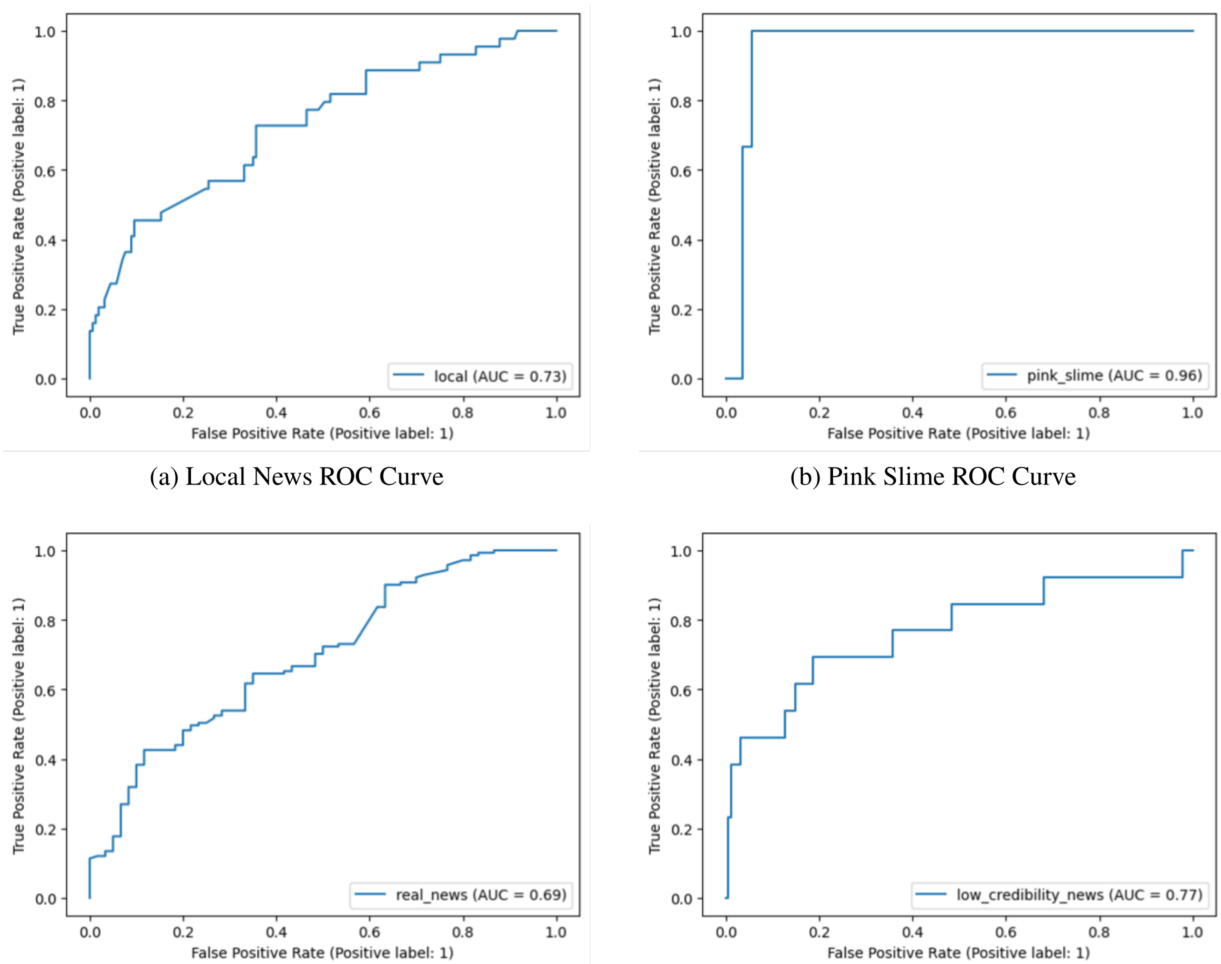
Table 4.10: Statistics for Facebook Midterms dataset

News Domains	Average Non-Credibility Score (NCS)	Standard Deviation
Pink Slime	.557	.366
Low Credibility News	.642	.330
Real News	.002	.010
Local News	.001	.006

Table 4.11: Average Non-Credibility Scores (NCS) of the Facebook Midterms training data by news type.

Feature	Importance
Non-Credibility Score	0.377
Average Number of Likes	0.357
Unique Pages	0.096
Domain Contains Location	0.094
Number of Occurrences	0.076

Table 4.12: Feature Importance of the XGBoost Model for the Facebook Midterms dataset.



(a) Local News ROC Curve

(b) Pink Slime ROC Curve

(c) Real News ROC Curve

(d) Low Credibility News ROC Curve

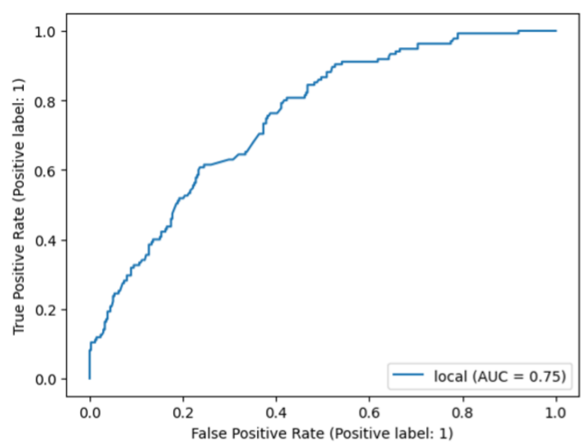
Figure 4.4: ROC Curves for Predicting News Labels of the Twitter Midterms Dataset

When the XGBoost model is applied to this dataset, we observe a macro average precision of 0.56 and recall value of 0.33 with an F1 score of 0.33. While these numbers are improvements over the Twitter Midterms dataset and more in line with the 2020 Facebook dataset, the ROC Curves in Figure 4.5 tell a less successful tale. Pink slime had an AUC of 0.65, with the lowest AUC being for low credibility news (0.50, no better than randomly guessing).

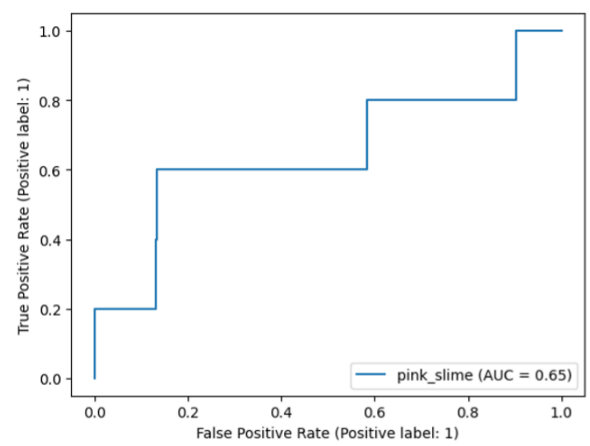
While these results are discouraging, it serves as an important reminder that the recommendations for using the NCS should come with some guidelines of minimum dataset size to achieve successful results.

4.7.4 Reddit Midterms Dataset

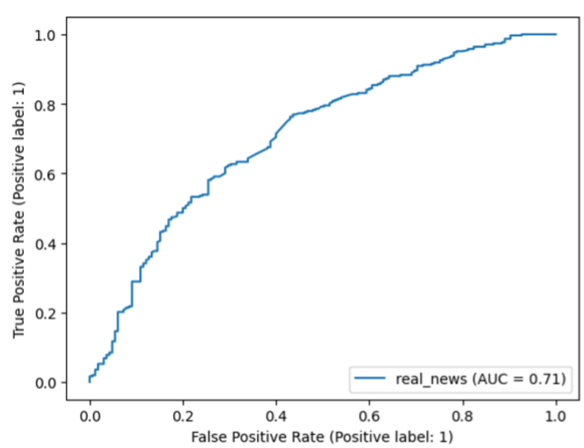
As mentioned earlier, the Reddit Midterms dataset is the smallest dataset we are running the model on. Unlike the previous datasets, we see some differences emerge in the feature analysis of the inputs in Table 4.13. While pink slime and low credibility news remain a minority and local



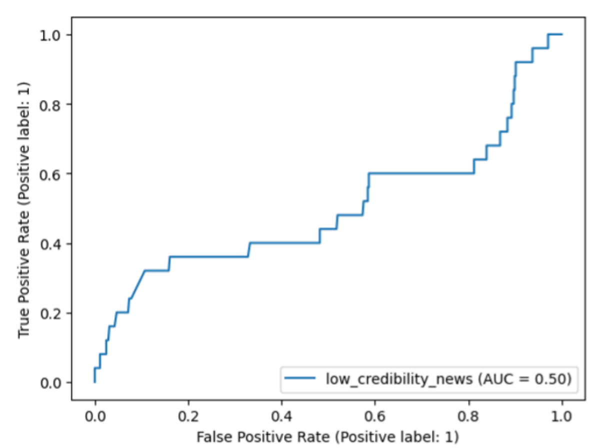
(a) Local News ROC Curve



(b) Pink Slime ROC Curve



(c) Real News ROC Curve



(d) Low Credibility News ROC Curve

Figure 4.5: ROC Curves for Predicting News Labels of the Facebook Midterms Dataset

news and pink slime remain the two news types with more than half of the domains containing a U.S.-based location, we see stark differences in the ‘score’ feature. Reddit’s ‘score’ value serves as a proxy to the number of likes a post or comment receives. Reddit posts and comments default to a score of 1, and other users can either upvote (resulting in a higher score) or downvote (resulting in a lower score) the post or comment. In the Reddit Midterms dataset, we see that all but one news type (real news, with an average score of 1.28) have average scores of 1.0, indicating a lack of popularity and interaction from other users on the platform. This could be for a variety of reasons, including being placed on less subscribed-to subreddits. It’s also important to note that there are only *3* pink slime domains in this dataset, providing us only a handful of data points.

Like all the previous datasets, the low credibility news types received the highest average non-credibility score; however, this score for pink slime (.009) is the lowest of all the datasets. This is likely due to the extreme dearth of pink slime and low credibility news instances in the dataset.

News Domains	Number of Domains	Average Number of Occurrences	Average Score	Average Number of Subreddits	Domain Contains Location
Pink Slime	3	3.3	1.0	2.0	58%
Low Credibility News	22	4.4	1.0	3.7	39%
Real News	388	21.7	1.28	8.8	35%
Local News	25	3.3	1.0	1.9	51%

Table 4.13: Statistics for Reddit Midterms dataset

News Domains	Average Non-Credibility Score (NCS)	Standard Deviation
Pink Slime	.009	
Low Credibility News	.176	.153
Real News	.002	.010
Local News	.001	.002

Table 4.14: Average Non-Credibility Scores (NCS) of the Reddit Midterms training data by news type.

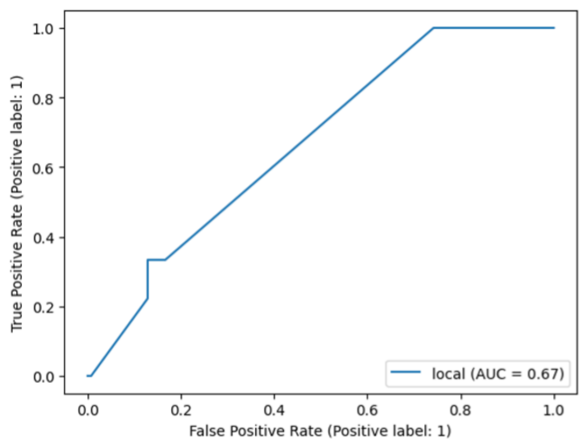
After running the XGBoost model on various iterations of the feature set, the best results came from only including two features - the score and the NCS. Ultimately, per Table 4.15, the score had the highest importance.

Ultimately, after running the XGBoost model, the average macro precision of prediction the news type was only 0.21, with a recall of 0.25 and an f1 score of 0.23. These results are the lowest of any sampled dataset, with relatively poor ROC Curves showing an AUC of predicting pink slime as 0.63 (Figure 4.6.

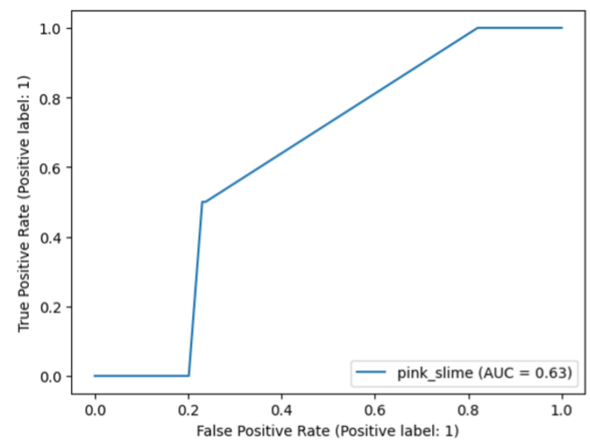
Overall, applying the NCS to the Reddit Midterms dataset was unsuccessful. Most likely this was due to the extremely small sample size, but as we've seen in Chapter 2, the presence of pink slime on Reddit is the most rare of any platform studied. It's likely that even a large Reddit dataset could not produce adequate results due to its limited occurrence of pink slime.

Feature	Importance
Score	0.606
Non-Credibility Score	0.394

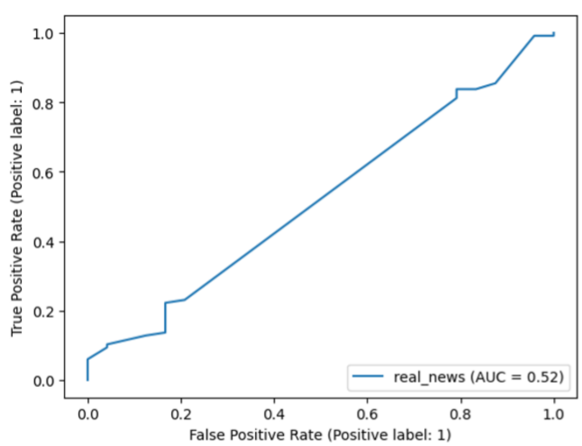
Table 4.15: Feature Importance of the XGBoost Model for the Reddit Midterms dataset.



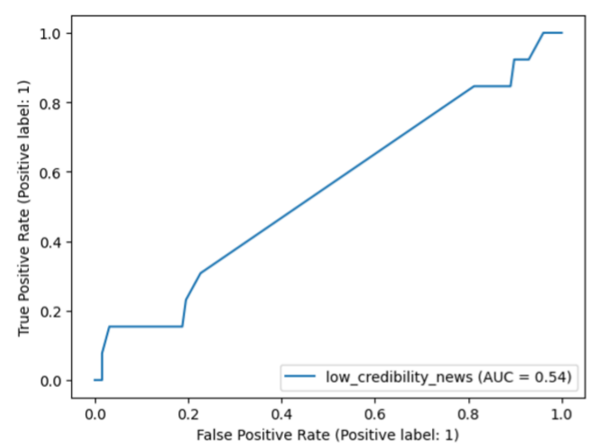
(a) Local News ROC Curve



(b) Pink Slime ROC Curve



(c) Real News ROC Curve



(d) Low Credibility News ROC Curve

Figure 4.6: ROC Curves for Predicting News Labels of the Reddit Midterms Dataset

4.7.5 Dataset Recommendations

In this chapter, we've seen that the success of the Non-Credibility Score in predicting the news type shared within a dataset depends largely on the size of the dataset and the amount of pink slime that is typically shared on the platform. The original 2020 Facebook dataset as well as the Twitter Midterms datasets showcase the strength of the NCS in sufficiently large (2.9 million and 1.3 million posts, respectively) datasets. It also shows that the NCS, as a network feature, can be applied to multiple platforms and is not merely limited to Facebook data. The lack of success in the Facebook Midterms and Reddit Midterms datasets (126k and 13k posts, respectively) show that there should be a recommend size cutoff for when to apply the NCS. While the drop from the successful Twitter Midterms dataset of 1.3 million posts to the unsuccessful Facebook Midterms dataset of 126k posts seems like a large range, I implore any future users of the NCS to understand that using this feature is best done on datasets with hundreds of thousands or millions of posts. While I am very critical of the model's success on the smaller datasets, I want to end

Discovered Domain	Targeting American Region	No Paywall	<50% Articles Has Authors Listed	Owned by Larger Entity	Has Sites in Other States
Georgiastarnews.com	✓	✓	✓	✓	✓
Colchestersun.com	✓	✓	✓		
Texasscorecard.com	✓	✓			
Baltimoresun.com	✓				
Californiajournal.com	✓				
Shelbycountyreporter.com	✓				
Coloradosun.com	✓				

Table 4.16: Most commonly shared sites and their characteristics

this section on a more positive note. When applying the NCS to all of the *unlabeled* Facebook Midterms posts, we are still able to find new sources of pink slime by sorting the results by NCS and reviewing the top 100. This analysis is in the section below.

4.7.6 Searching for Pink Slime Using NCS

By applying the machine learning model with the NCS to the Facebook Midterms dataset, we can quickly analyze it to find new and emerging sources of pink slime. This dataset included 10,223 domains which would be challenging to quickly manually analyze.

The top three labeled pink slime news articles in this dataset by number of likes included ones with the headlines, "Tammy Baldwin Gets Republicans to Back a Marriage Equality Bill, Making Final Approval Likely" (published on Up North News, a site owned by Courier Newsroom that targets Wisconsin residents), "Republicans Don't Want Black, Working-Class Voters To Turn Out" (published on Cardinal and Pine, a North Carolina-focused site owned by Courier Newsroom) , and "A Victory Over Extremism: Josh Shapiro Wins Pa. Governor's Race" (published on Keystone Newsroom, another Courier Newsroom site). While the number of pink slime sites with a conservative backing outnumber those with liberal backing, the liberal sites are receiving the most interaction when their articles are shared on Facebook.

After calculating and sorting the list of domains by a descending NCS value and visually inspecting the top 100 domains, I compiled Table 4.16 to illustrate some of the top examples of news domains shared in this dataset that contain many of the characteristics of pink slime domains; however, none of these sites are currently labeled as pink slime. For sites like the colchestersun.com, I would want to flag them to see if they eventually begin to own other "local" news websites in other states to meet the definition of a pink slime organization.

4.8 International Application

A final goal of this research is to understand whether the NCS can be used in international settings. While the network portion of the NCS is language-agnostic, it does depend on having certain culturally and regionally-relevant elements from the countries of interest. Specifically,

News Domains	Number of Domains	Average Number of Occurrences	Average Number of Likes	Average Number of Pages	Domain Contains Location
Low Credibility News	95	17.0	10.6	7.1	37%
Real News	1477	293.9	22.5	102.5	45%
Local News	295	95.2	6.3	6.4	40%

Table 4.17: Statistics for Facebook UK dataset

Gazetteers of the targeted country are necessary for the "Domain Contains Location" to remain accurate. Furthermore, having a set of news sites seeding the Media Thesaurus with authentic local news and any attempted pink slime attempts (like those referenced in Chapter 1 internationally).

4.8.1 United Kingdom Dataset

On July 7, 2024 the United Kingdom held a general election. We collected data related to this election from Facebook's CrowdTangle using all of the candidates' names to check for any potential instances of pink slime interfering with this election.

In order to run the NCS on the UK dataset, a set of UK-specific Gazetteers were compiled from the Association of British Counties[2]. Furthermore, in order to seed the Media Thesaurus with data relevant to the United Kingdom, a list of 1180 local news sites local to the UK from the Public Interest News Foundation[3] was added to the set of news labels.

Overall, this dataset had 1,249,741 posts shared by 293,453 Facebook Pages. Despite including some of the news domain examples of international news hijacking targeting the UK (mentioned in Chapter 1) as pink slime for the media thesaurus, none of these examples were present in the 2024 Elections dataset. Only the three other news types were present in this dataset, and we can see the breakdown of how the pages differ in Table 4.17 below. The vast majority of labeled news domains in this dataset are those belonging to the real news type. These sites appeared the most frequently, on the most pages, and received the most likes. In a confusing turn, these domains also contained UK location names a staggering 45% of the time, even more than we see for the local news domains.

Unsurprisingly, per Table 4.18, the low credibility news in the UK dataset have the highest NCS, over 100 times higher than that of local news or real news.

While initial results when running the XGBoost model on this dataset were adequate, they improved when adding a key feature - the domain suffix of the news domain. Interestingly, when these features were included in the US datasets, they did not improve the models and occasionally decreased the results. This may be due to a number of reasons, including that our Media Thesaurus, specifically in regard to real news and low credibility news is very US-specific so location-specific suffixes will appear more frequently in the local news websites collected

[2]gazetteer.org.uk/index
[3]https://www.publicinterestnews.org.uk/local-news-map-report-2024

News Domains	Average Non-Credibility Score (NCS)	Standard Deviation
Low Credibility News	0.514	.403
Real News	0.001	.006
Local News	0.005	0.014

Table 4.18: Average Non-Credibility Scores (NCS) of the Facebook UK training data by news type.

Feature	Importance
Domain Suffix: co.uk	0.674
Non-Credibility Score	0.157
Domain Contains Location	0.047
Number of Occurrences	0.022
Domain Suffix: .net	0.021
Domain Suffix: .org	0.018
Unique Pages	0.017
Domain Suffix: .de	0.016
Average Number of Likes	0.015
Domain Suffix: .com	0.015

Table 4.19: Feature Importance of the XGBoost Model for the Facebook UK dataset.

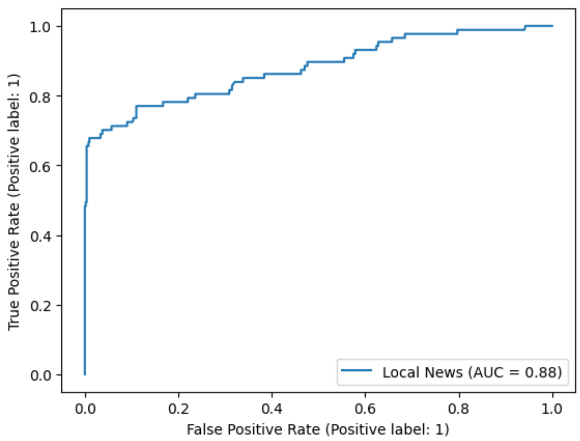
for this example. Overall, the importance of the domain suffix 'co.uk' had the highest feature importance per Table 4.19 (presumably to classify a news type as local news); following in importance were the NCS and whether the domain contains a location.

As speculated in the paragraph above, the distribution of domain suffixes is skewed by the inclusion of more specific local news websites in our Media Thersaurus for the UK but not as location-specific sites for real news and low credibility news. We observe that local news has the majority of its news domains ending with 'co.uk' while the remaining news types most frequently end in '.com' per Table 4.20.

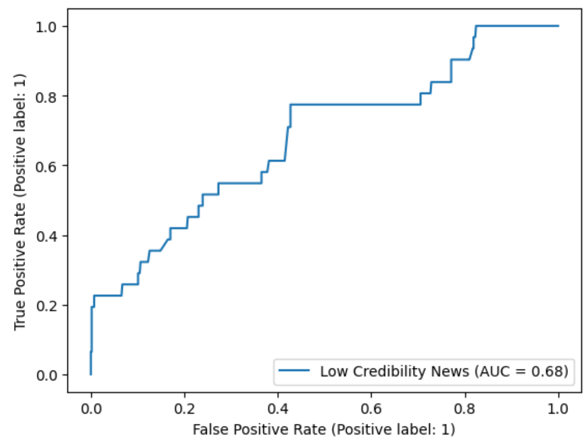
Finally, while there was no examples of pink slime in this dataset, the inclusion of the NCS as well as some of the linguistic features (whether the domain contains a location and the domain

	Domain Suffix				
News Type	.com	.org	co.uk	.net	.de
Real News	994	175	12	21	31
Local News	62	5	207	6	0
Low Credibility News	52	15	4	3	1

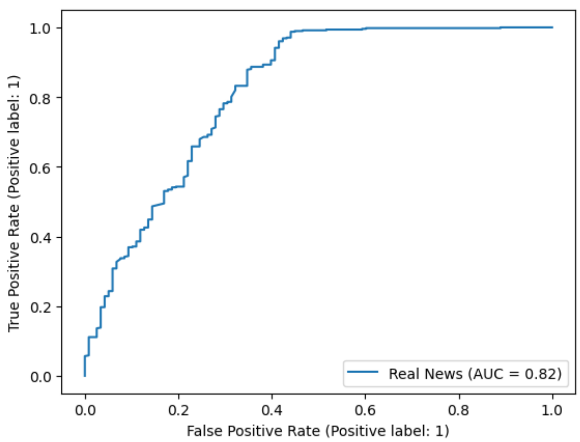
Table 4.20: Number of domain suffixes for each news type in the Facebook UK dataset.



(a) Local News ROC Curve



(b) Pink Slime ROC Curve



(c) Real News ROC Curve

Figure 4.7: ROC Curves for Predicting News Labels of the Facebook UK Dataset

suffix of the domain) provide a decent way to predict the remaining news labels in our international dataset. Below in Figure 4.7, we observe that the AUC for local news sites was high (0.88), similar to that of real news, and higher than that of low credibility news (AUC of 0.68).

When applying the NCS to the 2,975 unlabeled news domains in the dataset that were shared from a UK-admin Facebook Page, we discovered that only 63 of these sits had a NCS above zero. By ranking the sites by descending NCS values, we observed that the second, fourth, eleventh-highest ranking sites by NCS follow a similar domain naming pattern. The sites: bridgewaterburnhamconservatives.co.org, crawleyconservatives.co.org, and northumberlandconservatives.co.org, respectively, include a local UK region and the phrase "conservatives." While these sites appear to be legitimate sites created by the local conservative parties, it's interesting to note that they use the same website templates and the same Facebook avatars, indicating some coordination to influence local communities in multiple regions.

4.9 Limitations

There are two main limitations of the NCS and applying it to other datasets.

First, while we found that the vast majority of news domains in this dataset are "unlabeled", that is not the reality. Improvements to the Media Thesaurus to find other news sites through other labeled datasets would improve the starting point of finding the needle (low credibility and pink slime news sites) in the haystack of news sites shared on social media. In order to apply this research and the NCS, we need human labelers to "seed" the system with labels of low-credibility news, real news, or pink slime for different domains. In addition to including labels of new, credible sites that emerge, we must remain timely and purge the credible news labels from news sites like the nefarious "zombies" that take over news domains after legitimate a news organization closes, trying to steal its credible reputation with readers familiar with the name and domain [13].

The second limitation of applying the NCS is acquiring the large social media dataset required for the network measure to be based upon. In the past two years, the price for API pulls have called into question the future of the Twitter API. Furthermore, Meta made the bold decision to cancel the CrowdTangle API in August 2024. It remains to be seen how Meta's Content Library will work to maintain access to the type of Facebook data CrowdTangle researchers have grown accustomed to.

The more optimistic take on these limitations is that the NCS is an important metric that can be expanded by researchers with a larger media thesaurus to seed more news sites and continue to find these pink slime sites that infiltrate the news ecosystem. Furthermore, it can be applied professionally by social media platforms to downrank a news site that first appear on their platforms with a low NCS.

4.10 Conclusions

The Non-Credibility Score improves our ability to identify pink slime news sites by providing a quantifiable metric to evaluating the credibility of a site. Any individual with a large enough dataset of news sites shared by actors can apply this framework to quickly sift through millions of news domains and rank the likelihood that a given news site is low credibility or pink slime by sorting by descending NCS. For analysts overwhelmed by large datasets and time constraints, they can perform the network calculation to find the most concerning websites to investigate. Beyond analysts wanting to searching for fake local news, the NCS provides the ability to also find any sort of potential low credibility news that is a threat to an online ecosystem.

The primary function of developing a Non-Credibility Score and performing feature engineering on the datasets was to apply these measures to new datasets to quickly rank and assess probable new pink slime news sites. As presented in the above sections, labeled pink slime sites are a small minority of the labeled news sources in a given dataset. Not only are pink slime news sites obscure, they are also good at masquerading as local news. The general results for predicting pink slime news sites using our machine learning methods shows promising accuracy and AUC curves for multiple platforms of social media data (provided datasets are larger than a million posts). The model itself is intended to apply the derived network attribute of NCS to the dataset to be able to retrieve a more target list of potential new pink slime sites.

Chapter 5

Training Humans to Detect Pink Slime

5.1 Introduction

The existence of pink slime alone is not a threat. However, the human consumption and spread of these sites is a danger, particularly if they believe it to be trustworthy news from local reporters. Online misinformation has been shown to have an impact on matters of vital concern in areas like public health, where online exposure to misinformation has shown to increase vaccine hesitancy during the most recent covid-19 pandemic [95]. Even subtle misinformation with misleading headlines is shown to affect the memory, inferential reasoning, and impressions of people in the visuals provided of those who read it [96].

Furthermore, local news maintains relatively high trust regardless of political party, and Republicans and Independents who have much lower trust in national news outlets than Democrats still rate local news as more trustworthy [55].

This chapter looks to see how humans process encountering pink slime on social media through a field study as part of Project OMEN. The participants took a test to measure their trust of pink slime and local news before and after a training to see what impact the training has on their level of trust and awareness of these sites.

The key research questions in this chapter are:

- What is the difference in trust a human has for local news versus pink slime?
- Does trust in pink slime news change after a user visits the pink slime link?
- Can we train human users to identify pink slime campaigns?

5.2 Related Work

Previous research shows us that humans who visit pink slime sites directly have a negative impression of the sites with repeated exposure over 6 days going to the same site [102]. Other human user studies have found that pink slime sites gain a perceived trust advantage with its audiences due to having a local term in its name [92]. However, no further human subject testing has been performed to understand user impression of these sites in the format through which they are most frequently shared - social media posts.

Since some researchers have defined pink slime as a subset of misinformation [74], we can broaden our scope of related literature to human subject testing on the impression of more general misinformation. After simply subtly nudging participants to think about accuracy in the news they share online, researchers find that their sharing of false news decreases (relative to accurate news) [90].

While news literacy groups have published lessons plans on how to explain pink slime [5] no research has been done to assess the impact this type of training has on a user's ability to identify pink slime.

5.3 Data and Methods

Participants and Environment An ongoing project with the Office of Naval Research has been to simulate an information operation environment, teach analysts how to assess vast quantity of online data, and observe their ability to find bad actors and malicious information campaigns. All of the participants represented members of the defense community from Five Eyes (FVEY) alliance that consists of Australia, Canada, New Zealand, the United Kingdom, and the United States. The game's objectives and setup can be found in [68]. In the middle of the multi-day exercise, participants are asked to take a media literacy pre-test, are given a training on pink slime detection, and are then given a media literacy post-test to gauge the effectiveness of the training.

Social Media Post Selection As part of the measurement of effectiveness of training, a pre- and post-test of media literacy has been conducted. The participants viewed 16 generic social media posts including 4 low credibility news posts, 4 real news posts, 4 local news posts, and 4 pink slime posts. Topics were selected to be apolitical and include the following topics: the Coronavirus and vaccination, climate change, Narcan administration, and the Ohio train derailment. For stories with external links, participants are instructed that they are allowed to visit the websites.

In order to have equally challenging social media posts selected for the pre and post test, we had a group of media literacy experts review each of the social media posts in the survey and grade them for level of difficulty. Five or six experts reviewed each survey question, and the local news and pink slime posts were split into pre and post test designations based on ensuring an equal difficulty score. The experts were asked "How easy or difficulty do you think assessing the accuracy and trustworthiness of this post would be for an average American social media user?" They were prompted to then select either (1) Extremely easy, (2) Somewhat easy, (3) Neither easy nor difficult, (4) Somewhat difficult, or (5) Extremely difficult. The survey then saw local news posts with an average difficulty of 3.1 and 2.8 in the pre- and post- test, respectively. For pink slime posts, the split resulted in an average difficulty of 3.1 and 3.1 in the pre- and post-test, respectively.

Examples of pink slime and local news posts can be seen below in Figure 5.1 and Figure 5.2, and a complete set of all the posts used in this exercise can be found in Appendix D.

Figure 5.1: Pre-Test Local News Post #1

Survey Questions For each post, the participants must answer a series of questions including:

- What do you believe is the accuracy of the content in this post?
  - True
  - Somewhat true
  - Neither true nor false
  - Somewhat false
  - False
- How trustworthy do you consider the poster of this message to be?
  - Trustworthy
  - Somewhat trustworthy
  - Neither trustworthy nor untrustworthy
  - Somewhat untrustworthy
  - Untrustworthy
- How confident are you in your answers to the previous two questions? [0-10]
- Do you believe the post was written by a local reporter?
  - Yes
  - No
  - Unsure

Figure 5.2: Pre-Test Pink Slime Post #1

- Would you share this post online (for example, through Facebook or Twitter)?
  - Definitely yes
  - Probably yes
  - Might or might not
  - Probably not
  - Definitely not
- Do you believe the poster of this message is trying to influence you or the audience of this post?
  - Definitely yes
  - Probably yes
  - Might or might not
  - Probably not
  - Definitely not
- Please elaborate on the reasons for your answer to the previous question on influence

Training After the pretest, the participants were given a 31-minute training adapted from PBS [5] to include defining pink slime journalism as was outlined in Chapter 1 of this thesis, visiting multiple pink slime sites owned by different parent organizations, fact-checking a story mass-produced on Metric Media, and (per the research in Chapter 4) network features of the sites in ORA Figure 5.3. When assessing whether a post contained a pink slime article, participants were urged to use the following format which PBS developed in their training material:

1. Check the website's "About" page
2. Lateral reading – search keywords to see what other sources have to say about the topic
3. Reading upstream – click the links/sources within the news story to understand the evidence.
4. Check Fact-Checking sites

The same story suggested by PBS was used as an example for practicing these 4 steps. This story revolved around fictitious claims by a pink slime site that a school district was planning to implement race-based grading policies in its schools.

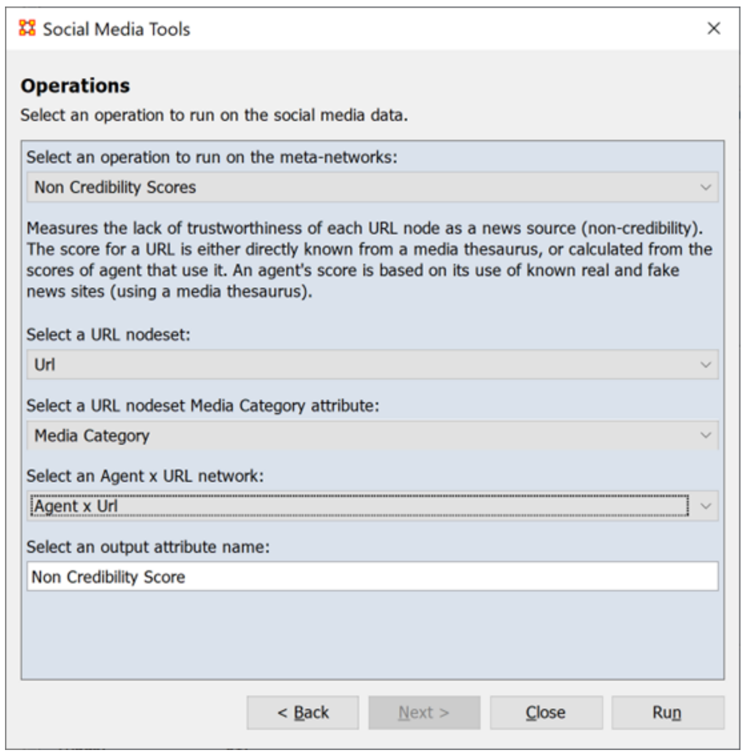


Figure 5.3: ORA interface for running the network features described in Chapter 4 as part of the lesson plan

After the training, the participants were given a post test with 16 new social media posts (with the same news type distribution) and the same follow up questions.

Survey Data In addition to each participant's responses to survey questions, the Qualtrics survey captured information regarding whether participants clicked the links embedded in the social media posts. This was of interest because research shows that 59% of links shared on Twitter were *never* clicked [46], indicating that the social media post and link preview are all that an audience will use to judge a post.

Participants also took an anonymous demographic survey which issued a participant identification number to tie to their results of the media literacy training surveys.

Methods Since the same participants take the pre-test and the post-test, their performance changes can be measured by a matched pairs t-test, provided there's a large enough sample size to meet the t-test's requirement of a normal distribution, for the following variables in Table 5.1.

5.4 Results

5.4.1 Participant Demographics

A total of 23 participants took both tests and attended the training. Their demographic distribution contains the following:

- Gender: 19 men and 4 women
- Race: 20 Caucasians, 1 Asian, 1 Native American, and 1 Latino or Hispanic
- Age: Average age of 35.6 with a standard deviation of 10.48
- Self-Described Political Leanings: 2 "Very Liberal", 6 "Liberal", 10 "Moderate", 4 "Conservative", and 1 "Very Conservative"
- Country of Residence: 16 USA, 3 Canada, 2, New Zealand, and 2 Australia
- Education: 5 Some College, 3 Associate's Degree, 9 Bachelor's Degree, 3 Master's Degree, 1 Professional Degree, and 2 Doctoral Degree
- 10 listed a form of local news as one of their primary news sources

5.4.2 Statistical Analysis of Variable Changes

For this study, a matched pairs t-test would be ideal to compare compare the pre test and post test results of all of the variables listed in Table 5.1. The results, illustrated in Table 5.2, are analyzed further below.

Local Reporter Identification During the pre-test, while participants accurately answered that local news posts were written by a local reporter 74% of the time, they only identified pink slime posts as not being written by a local reporter 21% of the time. In the post test, participants identified local news posts as written by a local reporter 65% of the time (a slight drop from the pretest); however, participants also correctly identified pink slime posts as *not* written by a local reporter 86% of the time, as illustrated in Figure 5.4. Per the statistical analysis, we find that this increase in identifying pink slime as not written by a local reporter has a statistically significant p-value of 1.58E-9 while the decrease in identifying local news as written by a local reporter does not have a statistically significant drop (p-value of 0.34). This lends credence to the increased ability to detect that pink slime is not written by a local reporter while not significantly decreasing the ability of a human to identify authentic local news as having been written by a local reporter.

Variable	Definition	Values
Pink Slime Trust	Response to "How trustworthy do you consider the poster of this message to be?" survey question on the pink slime posts.	1 = Trustworthy 2 = Somewhat trustworthy 3 = Neither trustworthy nor untrustworthy 4 = Somewhat untrustworthy 5 = Untrustworthy
Local News Trust	Response to "How trustworthy do you consider the poster of this message to be?" survey question on the local news posts.	1 = Trustworthy 2 = Somewhat trustworthy 3 = Neither trustworthy nor untrustworthy 4 = Somewhat untrustworthy 5 = Untrustworthy
Pink Slime Confidence in Trust	Response to "How confident are you in your answers to the trustworthiness question?" survey question on the pink slime posts.	0-10 where 0 = Very Unsure 10 = Very Confident
Local News Confidence in Trust	Response to "How confident are you in your answers to the trustworthiness question?" survey question on the local news posts.	0-10 where 0 = Very Unsure 10 = Very Confident
Pink Slime Reporter Correctness	Response to "Do you believe the post was written by a local reporter?" survey question on the pink slime posts.	0 = Yes or Unsure 1 = No
Local News Reporter Correctness	Response to "Do you believe the post was written by a local reporter?" survey question on the local news posts.	0 = No or Unsure 1 = Yes
Pink Slime Articles Clicked	Whether the pink slime article's embedded link was clicked in the survey.	0 = didn't click link 1 = clicked link
Local News Articles Clicked	Whether the local news article's embedded link was clicked in the survey.	0 = didn't click link 1 = clicked link

Table 5.1: Measured values from the surveys, their definitions, and the values that represent them.

	Pre Test		Post Test					
Variable	Average	Variance	Average	Variance	Pearsons Correlation	t Statistic	P(T<=t) one-tail	P(T<=t) two-tail
Pink Slime Trust	2.39	0.62	4.17	0.70	0.38	-8.96	4.30E-9**	8.61E-9**
Local News Trust	2.16	0.36	2.08	1.04	0.030	0.35	0.36	0.72
Pink Slime Confidence in Trust	7.24	3.38	8.24	3.10	0.25	-2.17	0.02*	0.04*
Local News Confidence in Trust	7.17	2.01	7.85	2.20	0.49	-2.21	0.02*	0.04*
Local News Reporter Correctness	0.74	0.08	0.65	0.12	0.11	0.98	0.17	0.34
Pink Slime Reporter Correctness	0.21	0.08	0.86	0.04	0.14	-9.85	7.91E-10**	1.58E-9**
Pink Slime Articles Clicked	0.68	0.13	0.70	0.14	0.76	-0.20	0.42	0.84
Local News Articles Clicked	0.70	0.15	0.67	0.14	0.56	0.29	0.39	0.78

Table 5.2: Matched pairs t-test results for the variables defined in 5.1 with a sample size of 23 participants (22 degrees of freedom). * represents significance <0.05 and ** represents significance <0.01.

Trust in News Types Since research shows that exposure to malicious news types can lead to a corroded trust of credible news sources [89], we were interested in seeing if an awareness of pink slime journalism would lessen the participants' trust in authentic local news. Using the trustworthy question where "Trustworthy" as assigned a value of 1 and "Untrustworthy" was assigned a value of 5, we found that prior to the training, participants rated local news posts as an average of 2.2 and pink slime posts as 2.4 on the trustworthy scale. This shows a similar level of trust placed in pink slime as is placed in local news prior to any training or awareness of its presence. After the training, trust of local news actually improved to a value of 2.1 with (a p-value of 0.72), while pink slime was rated a 4.2 (illustrated in Figure 5.5 with a statistically significant p-value of 8.61E-9, indicating that the training had the intended affect of keeping trust in authentic local news high while increasing awareness of pink slime journalism's untrustworthiness.

Confidence in Trust To better understand how sure the participants felt about the trust score they provided above, we asked them how confident they felt about their answer on a scale of 0 (Very Unsure) to 10 (Very Confident). Even though most of the participants were unfamiliar with pink slime and incorrectly stated that those articles were written by a local reporter during the pre-test, the participants still assessed their confidence in their high trust of pink slime as a 7.24, with their confidence in the local news trust score coming in slightly lower, at 7.17. Interestingly, there was more variance among the confidence in pink slime trust (3.38) than local news (2.01). After the training was conducted, both averages increased, with pink slime confidence in trust rising to 8.24 and local news confidence in trust showing a more modest increase to 7.85. While there was no statistically significant change for the confidence in the local news trust, the confidence in the pink slime trust rose enough to garner a p-value of 0.04, indicating that the training played a role in the participants' increased confidence in their ability to decipher pink slime as untrustworthy. The lack of a decrease in the confidence in trust of local news is also important for our findings. Just as it was important that trust in local news not falter once participants became aware of

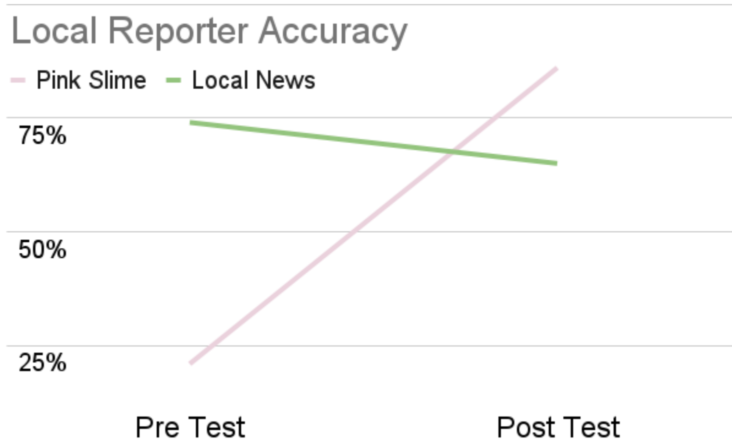


Figure 5.4: Participants' ability to correctly identify pink slime and local news before and after training

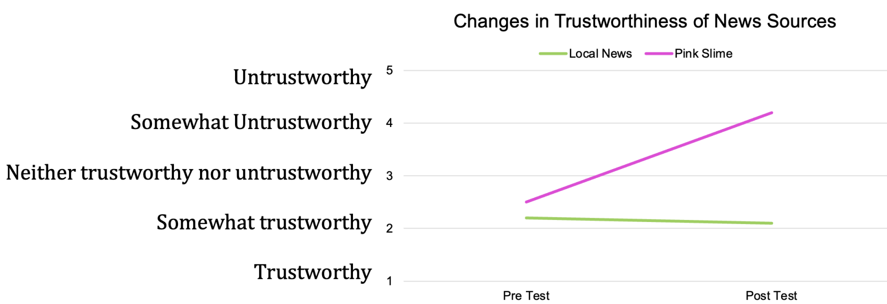


Figure 5.5: Participants' trust in news types before and after training

the pink slime phenomenon, we can see that their confidence in their ability to identify it as trustworthy is maintained.

Clicking Articles There was interest in understanding what role, if any, clicking on the embedded news story links from these social media posts would play in participants' ability to identify the news types. Literature tells us that 59% of links shared to Twitter were not clicked by viewers [46], and part of our training encouraged the participants to investigate the sources of the articles in the surveys. We were then surprised to learn that the majority of links in the pre-test were clicked for both pink slime (68%) and local news (70%). This indicates that our participants, possibly due to their professions, may not have been as casual as the typical social media users. After the training, their post-test click rates increased slightly for the pink slime articles (to 70%) and decreased slightly for local news articles (to 67%). Neither of these changes bared any statistical significance.

5.4.3 Detection of Pink Slime via Network Features

Additionally, nine made up pink slime news articles were embedded in the in the Balikatan training dataset in ORA for the participants to analyze. These news articles were designed to look like they were coming from an actual pink slime website targeting the Philippines, as seen in Figure 5.6. Using network features taught in the training, participants were asked to find the pink slime site whose articles were shared by the social media posters.

Figure 5.6: Example of pink slime site embedded into the OMEN exercise.

During the exercise, all of the groups were able to find and identify this malicious pink slime site. In previous iterations of this test where the pink slime training was not given, only one of the five teams was able to identify the embedded pink slime site as malicious and influencing activity.

5.4.4 Participant Feedback

In a survey given to the participants after the training and testing, we asked them which of the training elements they utilized to answer questions in the tests, and they responded as follows:

- 96% Clicking on the link and reading article
- 78% Checking a news website's About page
- 78% Looking up the author(s) of an article
- 78% Reading upstream - clicking links/sources in the article
- 61% Looking up the bias/accuracy rating of the news agency in question
- 57% Lateral reading - searching keywords or searching for similar stories

- 43% Checking fact-checking sites
- 13% Other (cited “common sense” and “previous knowledge”)

Finally, participants were asked how well they agreed or disagreed with the following statement: “This training has helped me become better at recognizing pink slime news.” The respondents overwhelmingly selected ‘Highly Agree’ (57%) or ‘Agree’ (39%) and only a single respondent selecting ‘Disagree’ (4%).

5.5 Discussions and Conclusions

The results from this study show that educating the public on pink slime and how to find it will not dampen the trust that authentic local news sources have worked hard to earn over decades. Additionally, the increased ability to detect that something was written by a pink slime site does not appear to have a correlation with increased scrutiny of the sites via clicking, as a similar portion of participants clicked the embedded URLs of the local news and pink slime sites in the pre test as they did in the post test.

A limitation of this study is that only 23 participants took the pre test, attended the training, and then took the post test. Generalizing to outside groups is a challenge, but this study represents a start.

Furthermore, a half hour training isn’t practical for mass education on the topic. In a condensed version of some of the attributes of the training mixed with the definition of pink slime from Chapter 1, I created an infographic to share as a media literacy resource with the public. The infographic can be seen below in Figure 5.7. While I have not conducted studies on the effectiveness of the infographic on the ability to detect pink slime, I present it as an option for educators who do not have a half hour to educate on the importance of this topic and awareness. Furthermore, in a meta-analysis on the effectiveness of infographics in educational environments, researchers found infographics to have a positive effect on academic achievement [43]

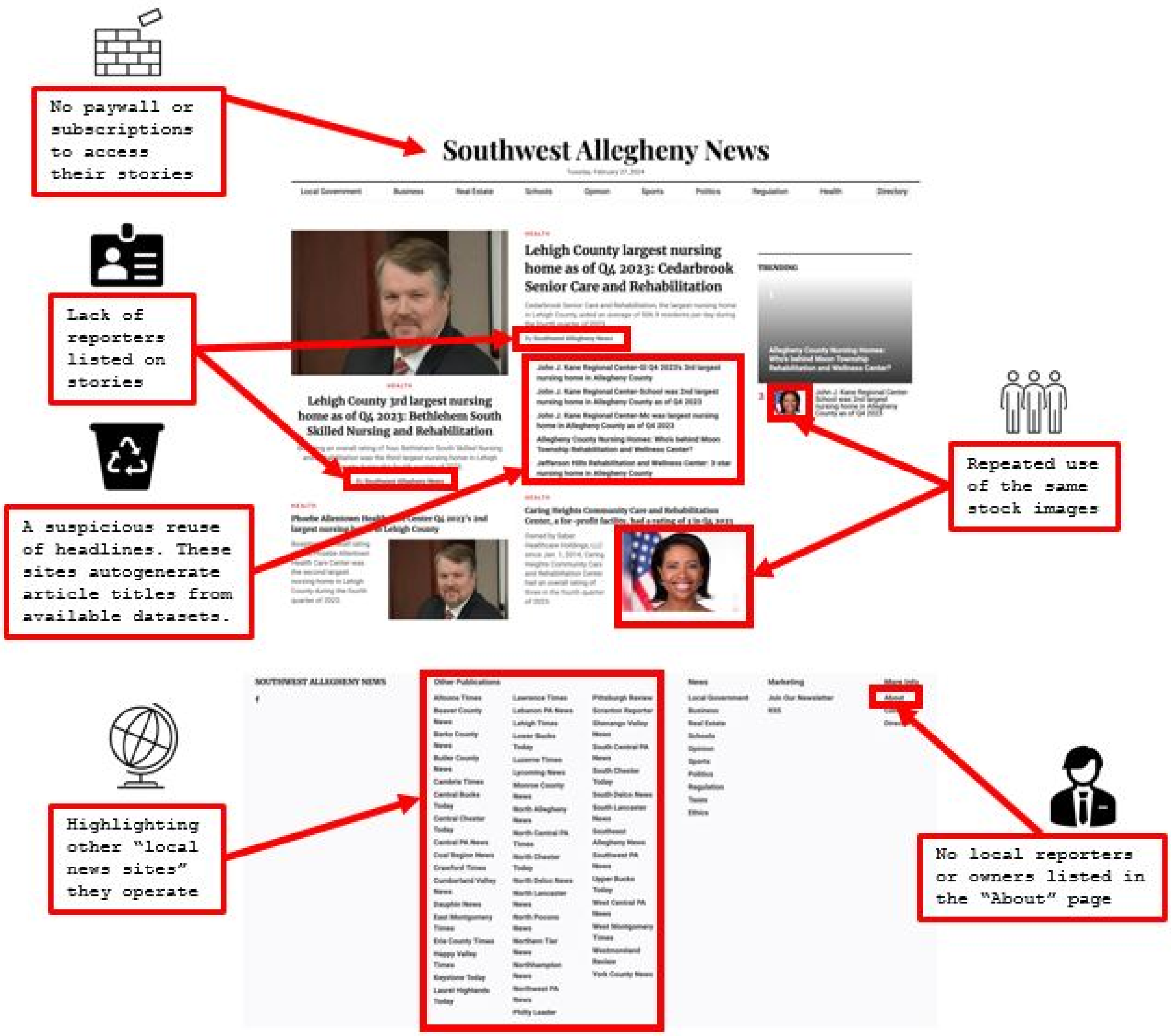


Figure 5.7: Infographic flyer to spread awareness of pink slime.

Chapter 6

Conclusions and Policy Recommendations

This chapter highlights key findings from the previous 5 chapters as well as looks outwardly to other countries facing issues similar to pink slime to generate policy recommendations for how to maintain the institution of authentic local news and discredit pink slime.

The key research question in this chapter is:

- What have we learned about pink slime?
- What are future concerns about pink slime?
- What policies could decrease the threat of pink slime?

6.1 Summary of Findings

This thesis provided a comprehensive look at one of the most dangerous threats to local news in the United States - digital pink slime journalism. From Chapter 1, I defined pink slime and discovered the evolution of journalism that allowed for this news type to gain a footing in the American news diet. Furthermore, I described how this phenomenon is not limited to the United States - 7 other international campaigns were analyzed and their commonalities were noted to understand potential paths for our domestic problem. In Chapter 2, I elaborated on the different strategies of the pink slime parent organizations for populating their web pages, advertising on Facebook, and posting to social media. In addition, I gained an understanding of how these news sites are shared differently from the other big three news types. I applied the BEND framework to analyze key differences in how pink slime differs from other news types across various platforms in Chapter 3. Taking the knowledge of how pink slime is shared from Chapter 2, I created the Non-Credibility Score in Chapter 4 and proved its effectiveness at identifying and classifying all of the news types. I performed user-studies to show the effectiveness of media literacy training on pink slime awareness and trust in Chapter 5. The remainder of this thesis is devoted to the future - what direction pink slime will travel in and how we can implement policy changes to combat the hijacking of local news.

6.2 The Future of Pink Slime

Organizations creating pink slime have benefited from the ease at which one can register a domain and establish a website template filled with API-driven content. Since the creation of many of these sites in 2018, further technological advances have been established that can streamline the process more. Generative AI can create not only online news articles through chatbots like ChatGPT given a few phrase prompt, but they can also generate article-relevant images to accompany the text on their website. Reputable newsrooms are already using ChatGPT as part of their news creation cycle, but journalists agree there are risks and not enough safeguards currently in place to keep the uses of it as ethical [11].

Furthermore, in the past two years, social media platforms have further restricted access to or raised prices on accessing data that researchers have used to identify pink slime. If this trend continues, the creators of pink slime may feel more emboldened to create more websites to spread influence on social media without fear of recourse.

6.3 Policy Recommendations

Upon reviewing the seven examples of international local news hijacking, several of the commonalities emerge which should be considered for policy action. Additionally, many of the following policy recommendations are based on the many contributions from the previous five chapters of this thesis. While each individual policy has its limitations in addressing all of the issues posed by pink slime journalism, I advocate for a combination of the policies to be enacted in order to increase the maximum effectiveness.

6.3.1 Government Policy Interventions

Strengthen Local News One counter-offense to this threat to local news would be a stronger defense - creating policies and funds that bolster America's local news ecosystem. The creation of the first international example of fake local news, as seen in Chapter 1, was prompted by the closing of local German newsrooms due to financial pressures. Researchers like Victor Pickard have argued that journalism is a public good [94] and pushed for the establishment of a publicly owned "and democratically governed media system" [93]. Despite how extreme this approach would be compared to the current business operation, smaller steps like greater public-private partnerships with funding for local newsrooms would be a step in the right direction. To start, I recommend focusing on maintaining local newsrooms in swing states, as residents in these regions are the subject to substantial pink slime advertisement campaigns, as evidenced in Chapter 2.

Combating Zombie Papers with Stricter IP Legislation The rise of "zombie papers," as termed by [13] to describe the happenings in Germany, was not limited to Germany. In Chapter 1, we saw other examples of foreign groups invading local news markets by using the names, logos, or likenesses of previously-active local news sites for the region. The European DisinfoLab suggested that we "urge the domain name industry to seriously reflect on this kind of fraudulent, disinforming behaviour as technical abuse of the domain name system" [48]. While we have not witnessed this happening domestically in the United States, other countries have used it to infiltrate American local news sites. The United States is in a unique position to counter these zombies through the use of legislation like the Anticybersquatting Consumer Protection Act [2]. While the law was passed in 1999, prior to the creation of these sites, it makes the action of creating a domain name that is in violation of a trademark illegal; however, the law fails to consider the international scope of these crimes. By passing legislation that classifies these international "local" news sites as cyber warfare, they can be removed from the online news ecosystem swiftly.

Increased Media Literacy Training in Schools In the United States, educational reform occurs at the state-level. While 75% of states agree that media literacy is important for students, only 19 states have passed legislation requiring such training in the classrooms and 9 more states have such pending legislation [77]. Including pink slime awareness as part of the news media literacy units and advocating for more states to include K-12 educational units on media literacy would help inoculate the next generation against this malicious threat to local news. As evidenced by Chapter 5's results, the use of pink slime lesson plans as originally designed by PBS [5] are effective at increasing awareness and lowering trust of pink slime while maintaining the ability to detect and trust authentic local news. The implementation performed in our user studies in tandem with the posts provided in Appendix D can serve as an additional training resource for teachers performing this training.

Mandate Increased Transparency and API Access for Researchers From Chapter 4, we learned that new sources of pink slime and low credibility news can be discovered with *sufficiently large* datasets of users sharing news sites on social media platforms. However, since that research was conducted, Meta closed access to its CrowdTangle API, Twitter/X substantially raised priced on its academic API, and Reddit blocked access to its academic-friendly API, PushShift. Fellow researchers have noted this problem and cited that APIs owned by platforms "hinder access, transparency and scientific knowledge" [39]. While the European Union passed legislation in 2022 that went into effect in 2024 [4] requiring access to data from the largest social media platforms, no such legislation exists in the United States. More robust legislation should be passed in the United States to allow free, expedited access to social media APIs for academic researchers.

6.3.2 Policy Recommendations for Companies

Removal or Flagging by Social Media Platforms As we have seen in Chapter 2, these sites are all shared on social media, and these social media platforms make up a plurality of the references to the sites. Many social media powerhouses have relied on Section 230 of the 1996 Communications Decency Act to protect them from being tried as a publisher of what their users share on their platforms. However, Facebook, Twitter, and Reddit have taken action against previous information operation campaigns that they are made aware of. Having a "tip" line for these instances that is shared with all of the social media platforms' content moderation teams could expedite the removal of their content or at the least, flagging it (or adding a "Community Note" in the case of Twitter/X). During the Covid-19 pandemic in 2020, Facebook announced that they would be providing ratings of "Altered", "Missing Context", "False", and "Partly False" via fact-checking partners to counter misinformation on their site, indicating that further labeling of pink slime would not be out of line with their news labeling efforts [98].

Website Templates' Terms of Service Violation As seen in Chapter 1's analysis of the international fake news campaigns, a majority of these sites are created using templates from the website-building software WordPress. While WordPress' Terms of Services does not hold them liable for the content posted on these sites, they have a streamlined process to report WordPress sites if they contain spam or infringe upon copyrights (many of which these sites do). Furthermore, if stricter IP legislation is passed (as recommended in the Government Policy Interventions section), it will be easier to file these copyright infringement reports to remove the sites.

6.3.3 User-Implemented Policy Recommendations

Treating Pink Slime Like Misinformation Some outlets have labeled pink slime as misinformation [74], and it's worth considering the approaches that misinformation and other low credibility news researchers have taken to combat fake news - nudges, fact checking, debunking, de-platforming - and seeing how well they would translate to the pink slime ecosystem based on the research in Chapters 1-5.

Public Media Thesaurus A recommendation utilizing the nudging countering approach from misinformation would be to have a governing news authority create a dynamic, public facing database of accredited local news organisations that meet certain requirements (such as having local news reporters, providing non-partisan reporting, etc.). This database would act as the CASOS Media Thesaurus has for the research in the research of Chapters 2-4. Residents can then cross-check their news sources credibility via browser extensions that highlight which news shared on their timeline is from accredited or non-accredited local news organizations. Researchers have tested this approach as it relates to broadly questionable and unreliable news sources by creating a browser extension rating Tweets for their content; it found that those exposed to such nudges were better able to distinguish the credibility of the information shared on their social media feed [22] [21]. As for news sharing, other research shows that having a credibility label on a Facebook news post would deter users from sharing that story [79]. A word of caution on this approach would be that the media thesaurus would need to be exhaustive. Research finds that, while labeled misinformation headlines result in viewers having a lower perception of the accuracy of the headlines, if a news misinformation news article is *not* labeled amongst a sea of labeled news sites, it is perceived as having higher accuracy than it does [91].

6.4 Contributions

6.4.1 Theoretical Contributions

The first theoretical contribution is a definition of pink slime and the conditions that allowed this phenomenon to gain success in online spaces. Secondly, through the human user studies, it contributes an understanding of human trust of pink slime sites as well as an evaluation of the impact of training on users' abilities to detect pink slime sites. The final theoretical contribution of the thesis is a set of policy recommendations for countering pink slime.

6.4.2 Methodological Contributions

The methodological contributions for this thesis are the creation of a hop-based method to discover low credibility news sites that is generalizable to not only pink slime but also assessing low credibility news and real news sites via the Noncredibility Score. Additionally, it contributes the methodology to apply the BEND framework to Facebook and Reddit posts and categorize pink slime sites (as well as the three other news types) into narrative and network maneuvers.

6.4.3 Empirical Contributions

This is the first large scale empirical assessment of the spread of pink slime sites on social media and the communities they targeted in online spaces during the 2020 U.S. Presidential Election and the 2022 U.S. Midterm election. It is also the first study showing pink slime spread on multiple platforms. Additionally, it contributes the quantitative measurement of impact of pink slime funding on organic community conversation.

6.4.4 Data Contributions

This thesis also offers several dataset contributions. It will be publishing the largest collection of Facebook posts sharing pink slime sites from 2019-2024. This includes over a million posts from every public Facebook account, page, and group that have ever shared a pink slime news article along with the engagement information and metadata representing which locale the pink slime site that was shared was targeting. The thesis also includes a dataset of over 4,000 ads purchased by pink slime organizations to promote their news articles including the targeted demographic, amount of money spent on the ads, and the number of impressions it received. This dataset will allow future researchers to join the two datasets to similarly understand the relationship between ad spend by these organizations and the organic conversations they generate in online spaces. Additionally, the thesis provides a collection of posts from Facebook, Reddit, and Twitter showing over 17,000 posts to pink slime sites during the 2022 U.S. midterm election.

6.5 Limitations

There are several important limitations to be addressed when proceeding with the scope of this thesis. The first is that the focus of pink slime sites is limited to those targeting the United States. While some of the research looks to similar cases in other countries for inspiration of how to address the issues, the United States is the focus and policy recommendations can be targeted to those capable of the U.S. government, companies, and citizens.

Second, the research is done largely on text in the English language since most of the social media platforms analyzed in the thesis contain posts written predominantly in English. Additionally, with the focus of the research being limited to the United States, the pink slime websites contain only English language articles. However, the methods proposed in Chapter 4 are designed to be language-agnostic, only focusing on the network features.

Third, this research is conducted using data from the social media platforms of Reddit, Facebook, and Twitter. From the conclusions drawn in Chapter 2 that Reddit contains minimal pink slime spread, the Facebook and Twitter datasets are a greater focus for analysis in later chapters. In recent years the APIs for these platforms have changed, and the methods utilized to acquire the data are referenced in the Data section of this proposal. While other platforms like Parler, Telegram, and NextDoor may contain posts linking to pink slime sites, the first two do not make up a significant amount of referral traffic per the SEO findings, and the third does not have a method to acquire data via an API.

Finally, this research is not focused on fact checking news articles that are shared by pink slime sites. The intent is to highlight that the stories shared by these platforms are those of a larger, *national and partisan* interest. The information is not analyzed for its factual validity but rather for its marketed duplicity as local news.

Bibliography

[1] Media bias/fact check news. URL https://mediabiasfactcheck.com/. 1.4.1

[2] Anticybersquatting consumer protection act. 15 U.S.C. § 1125(d), 1999. 6.3.1

[3] An investigation into a pro-indian influence network. Technical report, EU Disinfo Lab, 2019. URL https://www.disinfo.eu/wp-content/uploads/2020/01/20191213_InfluencingPolicymakers-with-Fake-media-outlets.pdf. 1.3.2

[4] Digital services act. OJ L 277, 27.10.2022, p. 1–102, 2022. 6.3.1

[5] Lesson plan: How to spot 'pink slime' journalism — misinformation in long-trusted local news - PBS NewsHour classroom, October 2022. URL https://www.pbs.org/newshour/classroom/2022/10/lesson-plan-how-to-spot-pink-slime-journalism-misinformation-in-once-trusted-local-news/. 1.2.4, 5.2, 5.3, 6.3.1

[6] A. Alaphilippe, G. Machado, R. Adamczyk, and A. Grégoire. Uncovered: 265 coordinated fake local media outlets serving Indian interests, November 2019. URL https://www.disinfo.eu/publications/uncovered-265-coordinated-fake-local-media-outlets-serving-indian-interests/. 1.2.4

[7] D. Alba and J. Nicas. As Local News Dies, a Pay-for-Play Network Rises in Its Place - The New York Times, October 2020. URL https://www.nytimes.com/2020/10/18/technology/timpone-local-news-metric-media.html. 1.2.2, 1.2.3

[8] A. Aljebreen, W. Meng, and EC Dragut. Analysis and detection of "pink slime" websites in social media posts. In *Proceedings of the ACM on Web Conference 2024*, New York, NY, USA, May 2024. ACM. 1.2.3, 1.2.4, 3.3.1

[9] N. Altmanl and K.M. Carley. Ora User's Guide 2022, Technical Report CMU-ISR-22-107, 2022. 2.5.2

[10] L. Alves, N. Antunes, O. Agrici, CMR. Sousa, and CMQ. Ramos. Click bait: You won't believe what happens next. *Fronteiras: Journal of Social, Technological and Environmental Science*, 5(2):196–213, 2016. 1.2.2

[11] A R Arguedas and F M Simon. *Automating democracy: Generative AI, journalism, and the future of democracy*. Balliol Interdisciplinary Institute, University of Oxford, 2023. 6.2

[12] L. Arvanitis and M. Sadeghi. Dark Money Political Ads Proliferate on Face-

book and Instagram, October 2022. URL https://www.newsguardtech.com/misinformation-monitor/october-2022. 1.2.3, 1.2.4

[13] Karin Assmann. Rise of the zombie papers: Infecting germany's local and regional public media ecosystem. *Media and Communication*, 11(3):360–370, 2023. ISSN 2183-2439. doi: 10.17645/mac.v11i3.6816. URL https://www.cogitatiopress.com/mediaandcommunication/article/view/6816. 1.2.4, 1.3.2, 4.9, 6.3.1

[14] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn. The Pushshift Reddit Dataset. In *Proceedings of the international AAAI conference on web and social media*, volume 14, pages 830–839. 1.4.2, 1.4.2, 1.4.2

[15] P. Bengani. Hundreds of 'pink slime' local news outlets are distributing algorithmic stories and conservative talking points, December 2019. URL https://www.cjr.org/tow_center_reports/hundreds-of-pink-slime-local-news-outlets-are-distributing-algorithmic-stories-conservative-talking-points.php/. Publication Title: Columbia Journalism Review. 1.2.2, 1.2.3, 1.2.4, 1.2.4, 4.2

[16] P. Bengani. Advocacy groups and Metric Media collaborate on local 'community news', October 2021. URL https://www.cjr.org/tow_center_reports/community-newsmaker-metric-media-local-news.php/. 1.2.4

[17] P. Bengani. The Metric Media network runs more than 1,200 local news sites. here are some of the non-profits funding them., October 2021. URL https://www.cjr.org/tow_center_reports/metric-media-lobbyists-funding.php/. 1.2.3, 1.2.3, 4.2

[18] P. Bengani. 'Pink slime' network gets $1.6m election boost from PACs backed by oil-and-gas, shipping magnates, October 2022. URL https://www.cjr.org/tow_center/pink-slime-network-gets-1-6m-election-boost-from-pacs-backed-by-oil-and-gas-shipping-magnates.php. 1.2.4

[19] P. Bengani, P. Brown, J. Bartholomew, S.G. Gotfredsen, and S. Rafsky. "pink slime": Partisan journalism and the future of local news. Technical report, Tow Center for Digital Journalism, 2024. 1.2.3

[20] DM. Beskow and KM. Carley. Social Cybersecurity An Emerging National Security Requirement. https://www.armyupress.army.mil/Journals/Military-Review/English-Edition-Archives/Mar-Apr-2019/117-Cybersecurity/b/, 2019. (document), 3.1

[21] Md Momen Bhuiyan, Kexin Zhang, Kelsey Vick, Michael A. Horning, and Tanushree Mitra. Feedreflect: A tool for nudging users to assess news credibility on twitter. In *Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing*, CSCW '18. ACM, October 2018. doi: 10.1145/3272973.3274056. URL http://dx.doi.org/10.1145/3272973.3274056. 6.3.3

[22] Md Momen Bhuiyan, Michael Horning, Sang Won Lee, and Tanushree Mitra. Nudgecred: Supporting news credibility assessment on social media through nudges. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2):1–30, October 2021. ISSN 2573-

0142. doi: 10.1145/3479571. URL `http://dx.doi.org/10.1145/3479571`. 6.3.3

[23] JT. Blane. Social-Cyber Maneuvers for Analyzing Online Influence Operations, May 2023. (document), 3.1

[24] JT. Blane, D. Bellutta, and KM. Carley. Social-Cyber Maneuvers During the COVID-19 Vaccine Initial Rollout: Content Analysis of Tweets. 24(3):e34040, March 2022. doi: 10.2196/34040. URL `https://www.jmir.org/2022/3/e34040`. 3.2.1

[25] JN. Blom and KR. Hansen. Click bait: Forward-reference as lure in online news headlines. *Journal of Pragmatics*, 76:87–100, 2015. ISSN 0378-2166. doi: https://doi.org/10.1016/j.pragma.2014.11.010. URL `https://www.sciencedirect.com/science/article/pii/S0378216614002410`. 1.2.2

[26] Reporters Without Borders. World Press Freedom Index 2024, 2024. URL `https://rsf.org/en/index`. 1.3.2

[27] Dimitrios Bountouridis, Monica Marrero, Nava Tintarev, and Claudia Hauff. Explaining credibility in news articles using cross-referencing. In *SIGIR workshop on ExplainAble Recommendation and Search (EARS)*, 2018. 2.2

[28] Alexandre Bovet and Hernán A Makse. Influence of fake news in twitter during the 2016 us presidential election. *Nature communications*, 10(1):7, 2019. 2.5.2

[29] J. Brandy and N. Diakopouloshttps. Facebook's News Feed Algorithm and the 2020 US Election. 9, September 2023. ISSN 3. doi: 10.1177/20563051231196898. URL `https://journals.sagepub.com/doi/epub/10.1177/20563051231196898`. 1.1, 1.2.4

[30] A.G. Burton and D. Koehorst. Research note: The spread of political misinformation on online subcultural platforms. *Harvard Kennedy School Misinformation Review*, September 2020. doi: 10.37016/mr-2020-40. URL `https://misinforeview.hks.harvard.edu/article/research-note-the-spread-of-political-misinformation-on-online-subcultural-platforms/`. 1.2.3, 2.5.2

[31] Alex Cadier. Imposter local news sites spread false claims about penalties for passengers using phones in canada, Aug 2020. URL `https://factcheck.afp.com/imposter-local-news-sites-spread-false-claims-about-penalties-passengers-using-phones-canada`. 1.3.2

[32] KM Carley. BEND: a framework for social cybersecurity. *Future Force*, 6(2):22–27. 3.2.1

[33] K.M. Carley. ORA: A toolkit for dynamic network analysis and visualization. 2017. doi: 10.1007/978-1-4614-7163-9_309-1. 2.5.2

[34] LR Carley, J Reminga, and KM Carley. ORA & NetMapper. 11th International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation;, July 2018. (document), 1.12, 1.13, 1.5, 1.5

[35] Peter Carragher, Evan M. Williams, and Kathleen M. Carley. Detection and discovery of misinformation sources using attributed webgraphs. *Proceedings of the International AAAI Conference on Web and Social Media*, 18:214–226, May 2024. ISSN

2162-3449. doi: 10.1609/icwsm.v18i1.31309. URL http://dx.doi.org/10.1609/icwsm.v18i1.31309. 4.2

[36] Eshwar Chandrasekharan, Mattia Samory, Shagun Jhaver, Hunter Charvat, Amy Bruckman, Cliff Lampe, Jacob Eisenstein, and Eric Gilbert. The internet's hidden rules: An empirical study of reddit norm violations at micro, meso, and macro scales. *Proc. ACM Hum.-Comput. Interact.*, 2(CSCW), nov 2018. doi: 10.1145/3274301. URL https://doi.org/10.1145/3274301. 2.5.2

[37] N.S. Cohen. From Pink Slips to Pink Slime: Transforming Media Labor in a Digital Age. 18(2):98–122, April 2015. ISSN 1071-4421. doi: 10.1080/10714421.2015.1031996. URL https://doi.org/10.1080/10714421.2015.1031996. 1.2.2, 1.2.4, 1.2.4

[38] A. Danaditya, L.H.X. Ng, and K.M. Carley. From curious hashtags to polarized effect: profiling coordinated actions in indonesian twitter discourse. 12, 2022. ISSN 105. doi: 10.1007/s13278-022-00936-2. 3.2.1

[39] Brittany I. Davidson, Darja Wischerath, Daniel Racek, Douglas A. Parry, Emily Godwin, Joanne Hinds, Dirk van der Linden, Jonathan F. Roscoe, Laura Ayravainen, and Alicia G. Cork. Platform-controlled social media apis threaten open science. *Nature Human Behaviour*, 7(12):2054–2057, November 2023. ISSN 2397-3374. doi: 10.1038/s41562-023-01750-2. URL http://dx.doi.org/10.1038/s41562-023-01750-2. 6.3.1

[40] Drew Desilver. Turnout in 2022 house midterms declined from 2018 high, final official returns show, 2023. URL https://www.pewresearch.org/short-reads/2023/03/10/turnout-in-2022-house-midterms-declined-from-2018-high-final-official-returns-show/. 2.5.1

[41] XGBoost Developers. XGBoost Documentation 1.7.4, 2022. URL https://xgboost.readthedocs.io/en/stable/. 4.5

[42] K.N. Dörr. Mapping the field of algorithmic journalism. *Digital Journalism*, 4(6):700–722, 2016. 1.2.2

[43] Şenel Elaldı and Taner undefinedifçi. The effectiveness of using infographics on academic achievement: A meta-analysis and a meta-thematic analysis. *Journal of Pedagogical Research*, 5(4):92–118, December 2021. ISSN 2602-3717. doi: 10.33902/jpr.2021473498. URL http://dx.doi.org/10.33902/JPR.2021473498. 5.5

[44] Alberto Fittarelli. Chinese websites posing as local news outlets target global audiences with pro-beijing content, Feb 2024. URL https://citizenlab.ca/2024/02/paperwall-chinese-websites-posing-as-local-news-outlets-with-pro-beijing-content/. 1.3.2

[45] Tow Center for Digital Journalism at Columbia University. Partisan Local News. URL https://github.com/TowCenter/partisan-local-news. 1.2.4, 1.2.4, 1.4.1, 1.4.2, 3.2.2

[46] M. Gabielkov, A. Ramachandran, A. Chaintreau, and A. Legout. Social Clicks: What and Who Gets Read on Twitter? In *ACM SIGMETRICS / IFIP Performance 2016*, Antibes Juan-les-Pins, France, June 2016. URL https://inria.hal.science/hal-

01281190. 5.3, 5.4.2

[47] Ryan J. Gallagher, Morgan R. Frank, Lewis Mitchell, Aaron J. Schwartz, Andrew J. Reagan, Christopher M. Danforth, and Peter Sheridan Dodds. Generalized word shift graphs: a method for visualizing and explaining pairwise comparisons between texts. *EPJ Data Science*, 10(1), January 2021. ISSN 2193-1127. doi: 10.1140/epjds/s13688-021-00260-3. URL http://dx.doi.org/10.1140/epjds/s13688-021-00260-3. 2.4.1

[48] Roman Adamczyk Gary Machado, Alexandre Alaphilippe and Antoine Grégoire. Indian chronicles: Subsequent investigation: Deep dive into a 15-year operation targeting the eu and un to serve indian interests. Technical report, EU Disinfo Lab, 2020. URL https://www.disinfo.eu/wp-content/uploads/2020/01/20191213_InfluencingPolicymakers-with-Fake-media-outlets.pdf. 6.3.1

[49] H.R. Glahn. Computer-produced worded forecasts. *Bulletin of the American Meteorological Society*, 51(12):1126–1132, 1970. 1.2.2

[50] B. Golding. Unreliable News Sites. URL https://github.com/hearvox/unreliable-news. 1.4.1

[51] J. Gottfried and J. Liedke. Partisan divides in media trust widen, driven by a decline among Republicans. August 2021. URL https://www.pewresearch.org/fact-tank/2021/08/30/partisan-divides-in-media-trust-widen-driven-by-a-decline-among-republicans/. 1.1, 1.2.3

[52] J. Gramlich. 10 facts about Americans and Facebook, February 2024. URL https://www.pewresearch.org/fact-tank/2021/06/01/facts-about-americans-and-facebook/. 1.2.2

[53] Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. Fake news on twitter during the 2016 us presidential election. *Science*, 363(6425):374–378, 2019. 2.5.2

[54] C. Groskopf. Who's ahead in competitive districts: 50 most competitive races, 2022. URL https://projects.fivethirtyeight.com/2022-election-forecast/house/. 1.4.2

[55] A. Guess, B. Nyhan, and J. Reifler. All Media Trust is Local? *Findings from the 2018 Poynter Media Trust Survey*, 2018. 5.1

[56] A.M. Guess, B. Nyhan, and J. Reifler. Exposure to untrustworthy websites in the 2016 US election. 4(5):472–480. ISSN 2397-3374. doi: 10.1038/s41562-020-0833-x. URL https://www.nature.com/articles/s41562-020-0833-x. Number: 5 Publisher: Nature Publishing Group. 1.2.2

[57] Manish Gupta, Peixiang Zhao, and Jiawei Han. Evaluating event credibility on twitter. In *Proceedings of the 2012 SIAM International Conference on Data Mining*, pages 153–164. Society for Industrial and Applied Mathematics, Philadelphia, PA, April 2012. 4.2

[58] Shlomi Heller and Bini Ashkenazi. Lebanon-linked operators run fake israeli 'news site', spread anti-israel sentiments - exclusive, Apr 2024. URL https://www.jpost.com/

israel-news/article-797725. 1.3.2

[59] BD Horne and M Gruppi. NELA-PS: A dataset of pink slime news articles for the study of local news ecosystems. *Proceedings of the International AAAI Conference on Web and Social Media*, 18:1958–1966, May 2024. 1.2.4

[60] D.W. Hosmer and S. Lemeshow. Area under the ROC curve. 2:160–164, 2000. 4.6

[61] Our World in Data. Economist Intelligence Unit (2006-2023) "Democracy index" [dataset], 2023. URL ttps://ourworldindata.org/grapher/democracy-index-eiu. 1.3.2

[62] Meta Platforms, Inc. Meta content library api version v2.0. doi: https://doi.org/10.48680/meta.metacontentlibraryapi.2.0. URL https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=US&media_type=all. 1.4.2

[63] TOW CENTER FOR DIGITAL JOURNALISM. Domains as of august 3, 2020. URL https://datawrapper.dwcdn.net/TqILa/2/. 4.2

[64] Raphael Kahan. Russian influence campaign shares bogus articles by fake israeli news websites, Jun 2023. URL https://www.ynetnews.com/business/article/rknlzep93. 1.3.2

[65] B. Kalsnes and A.O. Larsson. Understanding news sharing across social media: Detailing distribution on Facebook and Twitter. *Journalism studies*, 19(11):1669–1688, 2018. 1.2.2

[66] R.L. Kaplan. Yellow journalism. *The International Encyclopedia of Communication*, 11: 5360–5371, 2008. 1.2.2

[67] J. Kerswell. 6 Reasons You Always Fall For Click-Bait (and the Secret Formulas Publishers Won't Want You To See). https://messageandmuse.wordpress.com/2013/12/19/6-reasons-you-always-fall-for-click-bait-and-the-secret-formulas-publishers-wont-want-you-to-see/, 2013. [Accessed 27-03-2024]. 1.2.2

[68] C. King, C.S. Lepird, and K.M. Carley. Project OMEN: Designing a Training Game to Fight Misinformation on Social Media. 2021. 5.3

[69] J.M. Kleinberg. Authoritative Sources in a Hyperlinked Environment. *ACM*, 1999. doi: 0004-5411/99/0900-0604. 4.2

[70] S. Koenig. Switcheroo - This American Life. https://www.thisamericanlife.org/468/switcheroo, 2012. This American Life. 1.2.2

[71] Christine Sowa Lepird, Lynnette Hui Xian Ng, and Kathleen M. Carley. Non-credibility scores: Relative ranking of news sites shared on social media to identify new pink slime sites. *First Monday*, September 2024. ISSN 1396-0466. doi: 10.5210/fm.v29i9.13544. URL http://dx.doi.org/10.5210/fm.v29i9.13544. 1.4.1, 4.3

[72] C-G. Linden. Decades of automation in the newsroom. *Digital Journalism*, 5(2):123–140, 2017. doi: 10.1080/21670811.2016.1160791. URL https://doi.org/10.1080/21670811.2016.1160791. 1.2.2

[73] J. Littman, L. Wrubel, D. Kerchner, and Y. Bromberg Gaber. News Outlet Tweet Ids, 2020. URL https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/2FIFLH. Harvard Dataverse. 1.4.1

[74] M. Lynch. 2022 Midterms - Pink Slime, Misleading Ads, and more, October 2022. URL https://www.audacy.com/kmox/news/local/2022-midterms-pink-slime-misleading-ads-and-more. 5.2, 6.3.3

[75] Nick Mathews and Benjamin Toff. "we were facebook before facebook": The existential (not only economic) threat to community weekly newspapers in the us. *Digital Journalism*, 12(4):476–493, December 2023. ISSN 2167-082X. doi: 10.1080/21670811.2023.2293936. URL http://dx.doi.org/10.1080/21670811.2023.2293936. 1.2.2

[76] K. McDonald. NewsGuard - Metric Media Network Nutrition Label. URL https://www.newsguardtech.com/metric-media-network/. 1.2.3

[77] Erin McNeill. The u.s. media literacy policy report, 2024. URL https://medialiteracynow.org/wp-content/uploads/2024/02/MediaLiteracyNowPolicyReport2023_publishedFeb2024b.pdf. 6.3.1

[78] Kateryna Meleshenko. Cyber security indexes, 2023. URL https://www.kaggle.com/ds/3135173. 1.3.2

[79] Paul Mena. Cleaning up social media: The effect of warning labels on likelihood of sharing false news on facebook. *Policy amp; Internet*, 12(2):165–183, July 2019. ISSN 1944-2866. doi: 10.1002/poi3.214. URL http://dx.doi.org/10.1002/poi3.214. 6.3.3

[80] Meta. Updates on Our Security Work in Ukraine, 2022. URL about.fb.com/news/2022/02/security-updates-ukraine/. 1.2.4

[81] R. Moore, R. Dahlke, P. Bengani, and J. Hancock. The Consumption of Pink Slime Journalism: Who, What, When, Where, and Why? 2023. Publisher: OSF Preprints. 1.1, 1.2.3, 1.2.4, 1.2.4, 2.3, 3.2.2

[82] F.L. Mott. American Journalism. *Journalism Quarterly*, 27(3):368–368, 1950. doi: 10.1177/107769905002700333. URL https://doi.org/10.1177/107769905002700333. 1.2.2

[83] IE. Murdock. Information diffusion over diverse social media platforms and the simulated cross-platform impacts of interventions, May 2024. 3.3

[84] H. Murphy and S. Venkataramakrishnan. Local news is drowning in 'pink slime' ahead of US election. 2020. URL https://www.ft.com/content/f36c3e2e-bd62-4d93-853f-0d09cd8d9079. 1.2.3

[85] Steven Lee Myers. Spate of mock news sites with russian ties pop up in u.s., Mar 2024. URL https://www.nytimes.com/2024/03/07/business/media/russia-us-news-sites.html. 1.3.2

[86] NewsGuard. Secretly Partisan-Funded Websites Posing as Independent Local News Sites On Verge of Outnumbering Daily Newspapers in the U.S., December 2022. URL https://www.newsguardtech.com/press/partisan-funded-

websites-nearly-outnumber-daily-newspapers-in-us/. 1.2.3

[87] L.H.X. Ng and K.M. Carley. Popping the hood on Chinese balloons: Examining the discourse between U.S. and China-geotagged accounts. 28, 2023. ISSN 8. doi: 10.5210/fm.v28i8.13159. 3.2.1

[88] United States Department of State Office of the Historian, Foreign Service Institute. U.S. Diplomacy and Yellow Journalism, 1895–1898. https://history.state.gov/milestones/1866-1898/yellow-journalism. 1.2.2

[89] K. Ognyanova, D. Lazer, R.E. Robertson, and C. Wilson. Misinformation in action: Fake news exposure is linked to lower trust in Media, Higher Trust in government when your side is in power. *Harvard Kennedy School Misinformation Review*, Jue 2020. doi: 10.37016/mr-2020-024. 5.4.2

[90] Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio A Arechar, Dean Eckles, and David G Rand. Understanding and reducing the spread of misinformation online. *Unpublished manuscript: https://psyarxiv. com/3n9u8*, pages 1–84, 2019. 5.2

[91] Gordon Pennycook, Adam Bear, Evan T. Collins, and David G. Rand. The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, 66(11):4944–4957, November 2020. ISSN 1526-5501. doi: 10.1287/mnsc.2019.3478. URL http://dx.doi.org/10.1287/mnsc.2019.3478. 6.3.3

[92] Erik Peterson, Joshua P Darr, Maxwell B Allamong, and Michael Henderson. Can americans' trust in local news be trusted? the emergence, sources and implications of the local news trust advantage. *SocArXiv*, 2024. 1.2.3, 5.2

[93] Victor Pickard. *Democracy without Journalism?: Confronting the Misinformation Society*. Oxford University PressNew York, December 2019. ISBN 9780190946791. doi: 10.1093/oso/9780190946753.001.0001. URL http://dx.doi.org/10.1093/oso/9780190946753.001.0001. 6.3.1

[94] Victor Pickard, Josh Stearns, and Craig Aaron. Saving the news: Toward a national journalism strategy. 01 2009. 6.3.1

[95] Francesco Pierri, Brea L Perry, Matthew R DeVerna, Kai-Cheng Yang, Alessandro Flammini, Filippo Menczer, and John Bryden. Online misinformation is linked to early covid-19 vaccination hesitancy and refusal. *Scientific reports*, 12(1):5966, 2022. 5.1

[96] Lorenzo Prandi and Giuseppe Primiero. Effects of misinformation diffusion during a pandemic. *Applied Network Science*, 5:1–20, 2020. 5.1

[97] Free Press. Who Owns the Media? URL https://www.freepress.net/issues/media-control/media-consolidation/who-owns-media. 1.4.1, 3.2.2

[98] Meta Journalism Project. New Ratings for Fact-Checking Partners — facebook.com. https://www.facebook.com/journalismproject/programs/third-party-fact-checking/new-ratings, 2020. [Accessed 31-07-2024]. 6.3.2

[99] Cristina Pulido, Laura Ruiz-Eugenio, Gisela Redondo-Sama, and Beatriz Villarejo-

Carballido. A new application of social impact in social media for overcoming fake news in health. *International Journal of Environmental Research and Public Health*, 17(7):2430, Apr 2020. ISSN 1660-4601. doi: 10.3390/ijerph17072430. URL http://dx.doi.org/10.3390/ijerph17072430. 2.5.2

[100] M.A. Le Quéré and M. Jakesch. Trust in AI in Under-resourced Environments: Lessons from Local Journalism. In *CHI '22: Workshop on Trust and Reliance in AI-Human Teams*, New York, NY, 2022. ACM. 1.2.2

[101] M.A. Le Quéré, M. Naaman, and J. Fields. Local, Social, and Online: Comparing the Perceptions and Impact of Local Online Groups and Local Media Pages on Facebook. In *Proceedings of Computation+Journalism Symposium*, New York, NY, June 2022. C+J. 1.2.2

[102] S. Rafsky. Tow Center audience study: Reader perspectives on partisan local news sites, September 2022. URL https://www.cjr.org/tow_center_reports/tow-center-audience-study-reader-perspectives-on-local-partisan-news-sites.php/. 1.2.4, 5.2

[103] Allison McCann Robert Gebeloff and K.K. Rebecca Lai. Which Battleground State Voters Could Sway the Election? — nytimes.com. https://www.nytimes.com/interactive/2024/09/27/us/politics/battleground-state-voters.html, 2024. 2.3

[104] A. Royal and P.M. Napoli. Local Journalism without Journalists? Metric Media and the Future of Local News. 8:119–147, 2017. ISSN 21840466. doi: 10.56140/JOCIS-v8-2. URL https://jocis.org/2022/12/27/local-journalism-without-journalists-metric-media-and-the-future-of-local-news/. 1.2.4

[105] Craig Silverman and Jane Lytvynenko. These "canadian" websites and facebook pages are actually run from overseas, Aug 2020. URL https://www.buzzfeednews.com/article/craigsilverman/canadian-websites-and-facebook-pages-run-overseas. 1.3.2

[106] Oleksii Starov, Yuchen Zhou, Xiao Zhang, Najmeh Miramirkhani, and Nick Nikiforakis. Betrayed by your dashboard. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, New York, New York, USA, 2018. ACM Press. 4.2

[107] E.C. Tandoc, S. Wu, J. Tan, and S. Contreras-Yap. What is (automated) news? A content analysis of algorithm-written news articles. *Media e Jornalismo*, 2022. 1.2.2

[108] A. Tarkov. Journatic worker takes 'This American Life' inside outsourced journalism. Ju e 2012. URL https://www.poynter.org/reporting-editing/2012/journatic-staffer-takes-this-american-life-inside-outsourced-journalism/. 1.2.2

[109] Anna Tarkov. Journatic worker takes 'this american life' inside outsourced journalism. URL https://www.poynter.org/reporting-editing/2012/journatic-staffer-takes-this-american-life-inside-outsourced-journalism/. 1.2.4

[110] CrowdTangle Team. *CrowdTangle*. Facebook, Menlo Park, California, United States. URL https://www.crowdtangle.com/. 1.4.2, 3.2.2

[111] Twitter. Twitter Data for Academic Research. URL https://developer.twitter.com/en/use-cases/do-research/academic-research. 1.4.2, 1.4.3, 1.4.3, 1.4.3

[112] B. Clemm von Hohenberg, E. Menchen-Trevino, A. Casas, and M. Wojcieszak. A list of over 5000 US news domains and their social media accounts, 2021. URL https://github.com/ercexpo/us-news-domains. original-date: 2021-11-04T19:58:16Z. 1.4.1

[113] M. Walker and K.E. Masta. News Consumption Across Social Media in 2021. https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/, 2021. 1.2.2

[114] Y. Wang, S. Zannettou, J. Blackburn, B. Bradlyn, E. De Cristofaro, and G. Stringhini. A Multi-Platform Analysis of Political News Discussion and Sharing on Web Communities. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 1481–1492, 2021. doi: 10.1109/BigData52589.2021.9671843. 1.2.2

[115] Vinicius Woloszyn and Wolfgang Nejdl. DistrustRank. In *Proceedings of the 10th ACM Conference on Web Science*, New York, NY, USA, May 2018. ACM. 4.2

[116] L. Yin, F. Roscher, R. Bonneau, J. Nagler, and J.A. Tucker. Russian Trolls Relied on Local News More than Fake News in 2016 Presidential Election, New Analysis Finds, November 2018. URL http://www.nyu.edu/content/nyu/en/about/news-publications/news/2018/november/russian-trolls-relied-on-local-news-more-than-fake-news-in-2016-. 1.2.3

Appendix A

UK Election Keywords

“Geoff Cooper OR Caroline Nokes OR Tayab Ali OR Rowan Fitton OR Bob Bauld OR Daniel Matchett OR Jake Berry OR Andy MacNae OR Colin Taylor OR Paul Martin OR Tony Harrison OR Alexander Stafford OR Jake Richards OR Ishtiaq Ahmad OR David Atkinson OR Taukir Iqbal OR Tony Mabbot OR Adam Carter OR John Cronly OR Sarah Champion OR Anand Swayamprakasam OR Mark Townsend OR Becca Stevenson OR Richard Dickson OR Devenne Kedward OR Yousef Dahmash OR John Slinger OR Jess Lee OR Jonathan Banks OR Ian Price OR Tony Gill OR David Simmonds OR Paul Murphy OR Danny Clarke OR Chris Rowe OR Chris Copeman OR Jade Marsden OR Jason Moorcroft OR Mike Amesbury OR Nicholas Wood OR Michael Cressey OR Steven Ringham OR Stewart Mackay OR Robert King OR Ellen Nicholson OR Ben Spencer OR Harbant Sehra OR Lynn Irving OR Greg Webb OR Richard Mallender OR James Grice OR Ruth Edwards OR James Naish OR Andrew Daly OR John McArthur OR Jim Eadie OR Bill Bonnar OR Gloria Adebo OR Gary Burns OR David Stark OR Katy Loudon OR Michael Shanks OR Joanna Burrows OR Emma Baker OR James Moore OR Chris Clowes OR Joe Wood OR Alicia Kearns OR Stephen Lewthwaite OR Mustafa Abdullah OR Jake Austin OR Hilary Scott OR Wendy Olsen OR Keith Whalley OR Rebecca Long-Bailey OR Chris Harwood OR Arthur Pendragon OR Barney Norris OR Julian Malins OR Victoria Charleston OR Matt Aldridge OR John Glen OR Thomas Foster OR Asa Jones OR Lee Derrick OR Annette Hudspeth OR Robert Lockwood OR David Bowes OR Roberto Weeden-Sanz OR Alison Hume OR Scott Curtis OR Cahal Burke OR Abdul Butt OR Nick Cox OR Darren Haley OR Holly Mumby-Croft OR Nicholas Dakin OR Ralph James OR Gareth Lloyd-Johnson OR Kieran Dams OR Nagender Chindam OR Marcus Bleasdale OR Bill Esterson OR Christian Vassey OR Angela Oldershaw OR David Burns OR Charles Richardson OR Keir Mather OR Adam Hibbert OR Elwyn Jones OR Laura Manston OR Denise Scott-McDonald OR James Milmine OR Richard Streatfeild OR Laura Trott OR Jeremy Turner OR Mark Tyler OR Will Sapwell OR Maxine Bowler OR Aaron Jacob OR Christine Kubo OR Gill Furniss OR Annie Stoker OR Isabelle France OR Caitlin Hardy OR Alison Teal OR Sam Christmas OR Lucy Stephenson OR Angela Argenzio OR Abtisam Mohamed OR Mo Moui-Tabrizy OR Sam Chapman OR Andrew Cowell OR Jason Leman OR Issac Howarth OR Shaffaq Mohammed OR Olivia Blake OR Mick Suter OR Louise McDonald OR Steven Roy OR Helen Jackman OR Rebecca Atkinson OR Lorna Maginnis OR Alexi Dimond OR Louise Haigh OR Matthew Leese OR Muzafar Rahman OR Jack Carrington OR Hannah Nicklin OR Sophie Thornton OR Caroline Kampila OR Clive

Betts OR Jeremy Spry OR Lee Waters OR David Dobbie OR Sheila Greatrex-White OR Helen O'Hare OR Mark Spencer OR Michelle Welsh OR Darryl Morton-Wright OR Paul Shkurka OR Waqas Khan OR John Nagbea OR Will Grant OR Graham Reed OR Kevin Warnes OR Simon Dandy OR Philip Davies OR Anna Dixon OR James Gollins OR Chris Bovill OR Julian Dean OR Alex Wagner OR Victor Applegate OR Daniel Kawczynski OR Julia Buckley OR Mad Mike Young OR Matt Brown OR Frances Kneller OR Sam Banks OR Mike Baldock OR William Fotheringham-Bray OR Aisha Cuthbert OR Kevin McKenna OR Guy Phoenix OR Keith Tordoff OR Ryan Kett OR Andy Brown OR Andrew Murday OR Simon Garvey OR Malcolm Birks OR Julian Smith OR Matthew Winnington OR Martin Blake OR Robert Oates OR Benjamin Jackson OR Hanif Khan OR Caroline Johnson OR Nick Smith OR Jaswinder Singh OR Diana Coad OR Chandra Muvvala OR Adnan Shabbir OR Julian Edmonds OR Chelsea Whyte OR Robin Jackson OR Moni Kaur Nanda OR Azhard Chohan OR Tan Dhesi OR Ravaldeep Bath OR Christopher Graham OR Oliver Patrick OR Jay Anandou OR Nahim Rubani OR Roderick MacRorie OR Kate Fairhurst OR Pete Durnell OR Gurinder Josan OR Julian Knight OR Max McLoughlin OR Mary McKenna OR Ade Adeyemo OR Deirdre Fox OR Neil Shastri-Hurst OR Siobhan McErlean OR Lesley Veronica OR Roisin Lynch OR Mel Lucas OR John Blair OR Declan Kearney OR Paul Girvan OR Robin Swann OR Simon Breedon OR Steven Burnett OR Dave Thomas OR Elizabeth Grant OR Neil Speight OR Stephen Metcalfe OR Jack Ferguson OR James McMurdock OR James Gordon OR Miranda Fyfe OR Harrison Edwards OR Luke Viner OR Chris Carter-Chapman OR Pippa Heylings OR Owen Humphrys OR Martin Broomfield OR Sandy Steel OR Chris Twells OR Bob Eastoe OR Zoë Billingham OR Desi Latimer OR James Gray OR Roz Savage OR Paul Liversuch OR Amy Wheelton OR Aruhan Galieva OR Lucy Care OR Job West OR Heather Wheeler OR Samantha Niblett OR Becca Collings OR Robert Bagnall OR Daniel Steel OR Michael Bagley OR Anthony Mangnall OR Caroline Voaden OR Rosie Morrell OR Giovanna Lewis OR Joy Wilson OR Catherine Bennett OR Matt Bell OR Morgan Young OR Richard Drax OR Lloyd Hatton OR Hannah Westropp OR Declan Walsh OR Rosemary McGlone OR Michael O'Loan OR Jim Wells OR Andrew McMurray OR Diane Forsythe OR Colin McGrath OR Chris Hazzard OR Graham Cowdry OR Martin Corney OR Colin Martin OR Paul Wadley OR Sheryll Murray OR Anna Gelderd OR Rhys Baker OR Jack Braginton OR Mark Le Sage OR Paul Hilliar OR Matt Swainson OR John Hayes OR Mike Jelfs OR Paul Hartshorn OR Bill Piper OR Robert Parkinson OR Alberto Costa OR Jason Maguire OR Paco Davila OR Catherine Rowett OR Christopher Brown OR Chris Harrison OR Poppy Simister-Thomas OR Ben Goldsborough OR Stuart Robert OR Mick Stott OR Ian McCord OR Emmie Williamson OR Stewart Tolley OR Paul Hogan OR Rufia Ashraf OR Sarah Bool OR Stephani Mok OR Ange Turner OR Andy Hunter OR Katherine Fletcher OR Paul Foster OR Jonathan Aibi OR Ahmed Khan OR Craig Robinson OR David Francis OR Stephen Holt OR Emma Lewell-Buck OR Hilary Wendt OR Simon Thomson OR Charles Shackerley-Bennett OR Matthew Green OR Stuart Anderson OR Jessie Carter OR Tom Bartleet OR Beverley England OR Emma Bishton OR James Cartlidge OR Darryl Ingram OR Ben Davy OR Alan Spencer OR Lauren McLay OR Julian Brazil OR Stephen Horner OR Sarah Allen OR Rebecca Smith OR Ketankumar Pipaliya OR Michael McGetrick OR Victor Silkin OR Bernadette O'Malley OR Narinder Sian OR Keith Steers OR Alex Sufit OR Sally Symington OR Gagan Mohindra OR Lorraine Douglas OR Gary Conway OR Earl Elvis Of East Anglia OR Pallavi Devulapalli OR Josie Ratcliffe OR James Bagge OR Tobias McKenzie OR Liz Truss OR Terry Jermy OR Thomas

Culshaw OR James Ward OR Fay Whitfield OR Bret Palmer OR Garry Irvin OR Evelyn Akoto OR Andrew Murrison OR Declan Clune OR James Batho OR Neil Kelly OR Alex Culley OR Sidney Yankson OR Darren Paffey OR Maggie Fricker OR Wajahat Shaukat OR Thomas Gravatt OR Katherine Barbour OR John Edwards OR Ben Burcombe-Filer OR Satvir Kaur OR Bianca Isherwood OR Lee Clark OR James Allen OR Simon Cross OR Leslie Lilley OR Gavin Haran OR Bayo Alaba OR Robert Francis OR Lara Hurley OR Jason Pilley OR Tom Darwood OR James Miller OR Stephen Cummins OR Tilly Hogrebe OR Peter Little OR Anna Firth OR David Burton-Sampson OR Karl Vidol OR Geoff Moseley OR Lucy O'Sullivan OR Lauren Fulbright OR Charith Gunawardena OR Eric Sukumaran OR Bambos Charalambous OR Sean Halsall OR Edwin Black OR Erin Harvey OR Andrew Lynn OR Damien Moore OR Patrick Hurley OR Alistair Miller OR Manu Singh OR Rory O'Brien OR Harry Boparai OR Claire Tighe OR Lincoln Jopp OR Alison Brelsford OR Javed Bashir OR Martin Price OR Laura Evans OR Sarah Wood OR Kim Leadbeater OR Stewart Satterly OR Dafydd Morriss OR Simon Grover OR John Dowdle OR Sophia Bhatti OR James Spencer OR Daisy Cooper OR Angie Rayner OR Jay Latham OR Amanda Pennington OR Joanna Kenny OR Stephen Beal OR Steve Double OR Noah Law OR Joe Greenhalgh OR Pat Moloney OR Daniel Thomas OR Jayne Rear OR Malcolm Webster OR David Baines OR Brian Spencer OR Terence Price OR Emma Ellison OR James Tasker OR Raymond Peters OR Marie Rimmer OR John Harris OR Jason Saunders OR Paul Nicholson OR Dave Laity OR Ian Flindall OR Filson Ali OR Giane Mortimer OR Derek Thomas OR Andrew George OR Bev White OR Kathryn Fisher OR Stephen Ferguson OR Guy Lachlan OR Marianna Masters OR Anthony Browne OR Ian Sollom OR Craig Morton OR Titus Anything OR Peter Andras OR Scott Spencer OR Michael Riley OR Theo Clarke OR Leigh Ingham OR Graham Oakes OR Helen Stead OR Dave Poole OR Alastair Watson OR Karen Bradley OR Kamala Kugan OR Ian Owen OR Audel Shirin OR Robert Hodgetts-Haley OR Phil Chadwick OR Barbara Kaya OR Jonathan Reynolds OR Joshua Smith OR Paul Dawson OR Lisa Nash OR Peter Hopper OR Alex Clarkson OR Kevin Bonavia OR Andrew Adam OR Hamish Taylor OR Bill McDonald OR Neil Benny OR Alyn Smith OR Chris Kane OR Ashley Walker OR Ayesha Khan OR Wendy Meikle OR Helena Mellish OR Oliver Johnstone OR Lynn Schofield OR Navendu Mishra OR Jo Barton OR Samuel Bradford OR Niall Innes OR John McDermottroe OR Chris McDonald OR Monty Brack OR Vivek Chhabra OR Niko Omilana OR Mohammed Zaroof OR Nigel Boddy OR Anna-Maria Toms OR Steve Matthews OR Joe Dancey OR Matt Vickers OR AliRom Alirom OR Andy Poleshaw OR Laura McCarthy OR Adam Colclough OR Navid Kaleem OR Chandra Kanneganti OR Luke Shenton OR Gareth Snell OR Lucy Hurds OR Jag Boyapati OR Josh Harris OR Karl Beresford OR Jonathan Gullis OR David Williams OR Carla Parrish OR Peggy Wiseman OR Asif Mehmood OR Alec Sandiford OR Michael Baily OR Jack Brereton OR Allison Gardner OR Alexander Bramham OR Danni Braine OR Janice MacKay OR Sam Harper-Wallis OR Jacqueline Brown OR Gavin Williamson OR Mohammed Ramzan OR Christopher Bramall OR Stephen Price OR Richard Shaw OR Suzanne Webb OR Cat Eccles OR Barry Hetherington OR Gareth Burns OR Garreth Falls OR Alexandra Braidner OR Will Polland OR Noel Sands OR Ron McDowell OR Richard Smart OR Michelle Guy OR Jim Shannon OR Steve Hedley OR Fiona Lali OR Omar Faruk OR Janey Little OR Jeff Evans OR Nizam Ali OR Kane Blackwell OR Halima Khan OR Joe Hudson-Small OR Uma Kumaran OR Neil O'Neil OR Kevin Taylor OR Doug Rouxel OR Seyi Agboola OR James Crocker OR Chris Clarkson OR Manuela Perteghella OR Myles Owen OR Magdaline Nzekwue OR Waseem Sherwani OR Philip Watson OR Claire Bonham OR Anthony Boutall OR Scott Ainslie OR Steve Reed OR Jim Newell OR Mark Clayton OR Daniel Jerrome OR Khalila Chaudry OR Charlotte Faulkner OR Mark Cornes OR Andrew Western OR Jason Hughes OR Saskia Whitfield OR George James OR Pete Kennedy OR Chris Lester OR Siobhan Baillie OR Simon Opher OR Julian Cusack OR Julia Ewart OR Matthew Jackson OR Thérèse Coffey OR Jenny Riddell-Carpenter OR Rachel Featherstone OR Niall Hodson OR Greg Peacock OR Chris Enyon OR Lewis Atkinson OR Elizabeth Wallitt OR Jon Campbell OR Jessica Hammersley-Rich OR Sam Goggin OR Ed McGuinness OR Al Pinkerton OR Chris Magness OR Dominie Stemp OR Stephen Gander OR Shaun Bowler OR Austin Henderson OR Dipesh Patel OR David Morgan OR Danielle Newson OR Nus Ghani OR Hamilton Action-Man Kingsley OR Aasha Anam OR Ryan Powell OR Chrisni Reshekaron OR Tom Drummond OR Luke Taylor OR Wajad Burkey OR Ben Auton OR John Sweeney OR Mark Hoath OR Rob Pocock OR Andrew Mitchell OR Gareth Bromhall OR Peter Jones OR Tara-Jane Sutcliffe OR Gwyn Williams OR Michael O'Carroll OR Patrick Benham-Crosswell OR Torsten Bell OR Scott Hunter OR Debbie Hicks OR Flo Clucas OR Andy Bentley OR Les Willis OR Justin Tomlinson OR Will Stone OR Martin Costello OR Matthew McCabe OR Roderick Hebden OR Catherine Kosidowski OR Robert Buckland OR Heidi Alexander OR Adam Goodfellow OR Robert Bilcliff OR Jed Marson OR Susan Howarth OR Ian Cooper OR Eddie Hughes OR Sarah Edwards OR Nigel Hennerley OR Jonathan Smith OR Oliver Speakman OR Ryan Jude OR Esther McVey OR Rochelle Russell OR Ryan Trower OR Brenda Weston OR Charles Hansard OR Rebecca Pow OR Gideon Amos OR Jo McKenna OR John Adams OR Hannah Campbell OR Alan Adams OR Shaun Davies OR David Edgar OR Cate Cody OR Damola Animashaun OR Byron Davis OR Laurence Robertson OR Cameron Thomas OR Chris Shipley OR Pat McCarthy OR Anthony Lowe OR Richard Leppington OR Roh Yakobi OR Mark Pritchard OR Luke Brownlee OR Richard McLane OR Steve Mason OR Mark Robinson OR Lisa Banes OR Kevin Hollinrake OR Alexandra Jenner-Fust OR Rob Logan OR Andrew Banwell OR Luke Hall OR Claire Young OR Nimal Raj OR Yousaff Khan OR Michael Bukola OR Eugene McCarthy OR Jacqueline Doyle-Price OR Sophie Preston-Hall OR Jen Craft OR Mark Rochell OR Abdul Husen OR Mohammed Hussain-Billa OR Mark Redding OR Jack Sabharwal OR Shaun Bailey OR Antonia Bance OR Laura Buchanan OR Jonathan Barter OR Fred Keen OR Ian Liddell-Grainger OR Rachel Gilmour OR Ian Grattidge OR Tim Shaw OR John Woollcombe OR Teresa Hansford OR Anna Cope OR Lewis Bailey OR Tom Tugendhat OR Davinder Jamus OR Jas Alduk OR Tarik Hussain OR Andrew Price OR Judith Trounson OR Nick Humberstone OR Ethan Brooks OR Rosena Allin-Khan OR Paul Moor OR Charlie West OR Chris Wongsosaputro OR Gordon Scott OR Kevin Foster OR Steve Darling OR Nikki Brooke OR Lee Dunning OR Brendan Roberts OR Philip Davies OR Matthew Jones OR Nathan Edmunds OR Ian Williams OR Nick Thomas-Symonds OR Alan Rayner OR Judy Maciejowska OR Andrew Jackson OR Isabel Saxby OR Phil Hutty OR Geoffrey Cox OR Pamela Holmes OR Amelia Allao OR Andrew Miles OR Jennifer Obaseki OR Roger Gravett OR Hari Prabu OR Attic Rahman OR Nandita Lal OR David Craig OR David Lammy OR Peter White OR Peter Lawrence OR Karen La Borde OR Steve Rubidge OR Ruth Gripper OR Cherilyn Mackrory OR Jayne Kirkham OR Hassan Kassem OR John Hurst OR Hugo Pound OR John Gager OR Neil Mahapatra OR Mike Martin OR Umair Malik OR Chantal Kerr-Sheppard OR Alex Starling OR Tom Bruce OR Jonathan Hulley OR Munira Wilson OR Adam Thewlis OR Christopher Greener OR Kelly Oliver Dougall OR Mustaque Rahman OR John Appleby OR Chloe-Louise Reilly OR Rosie Elliott OR

Lewis Bartoli OR Alan Campbell OR Malachy Quinn OR Kate Evans OR Eóin Tennyson OR Catherine Nelson OR Carla Lockhart OR Geoff Courtenay OR Steve Gardner OR Gary Harbord OR Ian Rex-Hawkes OR Sarah Green OR Tim Wheeler OR Steve Tuckwell OR Danny Beales OR Steven Sluman OR Stuart Field OR Steven Rajam OR Lynden Mack OR Ian Johnson OR Toby Rhodes-Matthews OR Alun Cairns OR Kanishka Narayan OR Andrew McRobbie OR Mike King OR Aarti Joshi OR Chris French OR Catherine Dawkins OR Florence Eshalomi OR Nicholas Sanders OR Brent Hawksley OR Keith Mason OR Ash Routh OR Stewart Golton OR Arnold Cravan OR David Dews OR Simon Lightwood OR Ian Pugh OR Philip Bimpson OR Vicky Downie OR Jane Turner OR Robbie Lammas OR David Burgess-Joyce OR Angela Eagle OR Patrick Stillman OR Sadat Hussain OR Shannon Lloyd OR Elaine Williams OR Aftab Nawaz OR Valerie Vaz OR Ruth Rawlins OR Dan Edelstyn OR Nancy Taaffe OR Mohammed Ashfaq OR Imran Arshad OR Rebecca Taylor OR Martin Lonergan OR Sanjana Karnani OR Rosalind Rowlands OR Stella Creasy OR Maddison Wheeldon OR Hannah Spencer OR David Crowther OR Yasmin Al-Atroshi OR Trevor Nicholls OR Charlotte Nichols OR Graeme Kelly OR Peter Willett OR Stephanie Davies OR Graham Gowland OR Janet Balfe OR Andy Carter OR Sarah Hall OR Laurie Steele OR Louis Adam OR Hema Yellapragada OR Nigel Clarke OR James Uffindell OR Matt Western OR Sharon McLafferty OR Ciaran Morrissey OR Michal Chantkowski OR Shaun Parsons OR Paul Donaghy OR Sharon Hodgson OR Sarah Knott OR Arran Bowen-la Grange OR Khalid Chohan OR Gary Ling OR Ian Stotesbury OR Dean Russell OR Matt Turmaine OR Maya Severyn OR John Shreeve OR Gurpreet Padda OR Scott Huggins OR Richard Rout OR Adrian Ramsay OR John Howson OR Kate Walder OR Daniel Kersten OR Lenny Rolles OR Katie Lam OR Jeremy Brittin OR Christopher Townsend OR Paul Mannion OR Ben Habib OR David Goss OR Gen Kitchen OR Craig Clarke OR Abi McGuire OR Peter Welsh OR Joe Joseph OR Helen Hims OR Meg Powell-Chandler OR Tessa Munt OR Sarah Butcher OR John Munro OR Jack Aaron OR Grant Shapps OR Andrew Lewin OR David Neill OR Iris Leask OR William Linegar OR Brandon Innes OR Michael Turvey OR Kate Blake OR Glen Reynolds OR Andrew Bowie OR Sam Harding OR Parmjit Gill OR Mohammed Yasin OR Gita Joshi OR Ray Nock OR Will Goodhand OR Sarah Coombes OR Marcus White OR Oliver Chisholm OR Kelvin Clayton OR Donna Lumsden OR Chris Loder OR Edward Morello OR Kelly Wilson OR Andrew Muir OR Paul Kennedy OR Maurice Corry OR Paula Baker OR David Smith OR Martin Docherty-Hughes OR Douglas McAllister OR Lois Austin OR Kayode Shedowo OR Emily Bigland OR Georgie David OR Holly Ramsey OR Rob Callender OR Sophia Naqvi OR James Asser OR Graham Smith OR Charlotte Houltram OR Simon Evans OR Mike Prendergast OR Ashley Dalton OR Ivan Kinsman OR Luke O'Brien OR Katie Parker OR Mark Ereira OR Henry Batchelor OR David Bull OR Rebecca Denness OR Nick Timothy OR Stephen Lynch OR Leza Houston OR Stephen Donnelly OR Stevan Patterson OR Matthew Bell OR Daniel McCrossan OR Tom Buchanan OR Órfhlaith Begley OR Seonaid Barber OR Natalie McVey OR Christopher Edmondson OR Kash Haroon OR Dan Boatright-Greene OR Harriett Baldwin OR Wendy Long OR Izzy Solabarrieta OR John Studholme OR Phil Clayton OR Pippa Smith OR James Townley OR Matty Jackman OR Tim Farron OR Thomas Daw OR Patrick Keating OR Richard Pearse OR John Penrose OR Daniel Aldridge OR John Hall OR James Monaghan OR Arnold Warneken OR Mike Jordan OR Ben Pickles OR Alec Shelbrooke OR Chris Wills OR Jill Perry OR Andrew Johnson OR David Surtees OR Josh MacAlister OR Michael Murphy OR David Coveney OR Nancy Mills OR Sean Houlston OR Jake Fraser OR

Derek Twigg OR The Zok OR Jan Cunliffe OR Jane Leicester OR Brian Crombie-Fisher OR Maureen O'Bern OR Henry Mitson OR Andy Dawber OR Lisa Nandy OR Michael Watson OR Amy Lynch OR Sarah Barber OR Aaron Mafi OR Rachel Brooks OR Ben Cronin OR Eleanor Stringer OR Danielle Dunfield-Prayero OR Paul Kohler OR Andy Liming OR Kevin D'Cruze OR Chris Barfoot OR Andrew Davis OR Lorraine Estelle OR Hannah Dawson OR Sean Whelan OR Flick Drummond OR Danny Chambers OR Simran Dhillon OR David Buckley OR Michael Boyle OR Harl Grewal OR Julian Tisi OR Pavitar Mann OR Jack Rankin OR Peter Reisdorf OR Gail Jenkinson OR Ken Ferguson OR Jenny Johnson OR Matthew Patrick OR Chelsey Jay OR Ashley Thompson OR James Abbott OR Timothy Blaxill OR Rumi Chowdhury OR Priti Patel OR David Cox OR Barry Ingleton OR Andrew Prosser OR Antonio Weiss OR Richard Langridge OR Robert Courts OR Charles Maynard OR Tim Read OR Nataly Anderson OR Ese Erheriene OR Richard Barker OR Jonathan Lord OR Will Forster OR Merv Boniface OR Monica Hamidi OR Colin Wright OR Lucy Demery OR Clive Jones OR Peter Thornton OR Kwaku Tano-Yeboah OR Paul Williams OR Jane Stevenson OR Sureena Brackenridge OR Bart Ricketts OR Athar Warraich OR Paul Darke OR Victoria Wilson OR Carl Hardwick OR Pat McFadden OR Vikas Chopra OR Zahid Shah OR Phillip Howells OR Celia Hibbert OR Andrea Cantrill OR Don Brookes OR Mike Newton OR Warinder Juss OR Duncan Murray OR Mark Davies OR Mel Allcott OR Tor Pingree OR Andy Peplow OR Marc Bayliss OR Tom Collins OR Sally Griffiths OR Danny Moloney OR Nas Barghouti OR Jemma De Vincenzo OR David Jones OR Bradley Mitchell OR Craig Birtwistle OR Michael Wheeler OR Kathryn Attwood OR Morag Chugg OR Sonya Mallin OR Edmund Rooke OR Peter Bottomley OR Beccy Cooper OR Paul Ashton OR Tim Morgan OR Timothy Sly OR Becca Martin OR Charles Dodman OR Sarah Atherton OR Andrew Ranger OR Mark Smallwood OR Ed Gemmell OR Ajaz Rehman OR Catherine Bunting OR Khalil Ahmed OR Toni Brodelle OR Richard Phoenix OR Steve Baker OR Emma Reynolds OR Nigel Geary OR Leigh Whitehouse OR John Davis OR Shazu Miah OR Bill Hopkins OR Vicki Smith OR Mark Garnier OR Hilary Salt OR John Barstow OR Simon Lepori OR Melanie Earp OR Sarah Beament OR Julie Fousert OR Mike Kane OR Steve Ashton OR Serena Wootton OR Rebecca Montacute OR Laura Bailhache OR Marcus Fysh OR Adam Dance OR Sam Wood OR Sir Grumpus L Shorticus OR Leena Farhat OR Martin Schwaller OR Emmett Jenner OR Ieuan Williams OR Virginia Crosbie OR Llinos Medi OR Leo Mayne OR Ruairi Kendall OR Roger James OR Alisdair Lord OR Alan Page OR Cliff Bond OR Lars Kramm OR Richard Hudson OR Rachael Maskell OR Darren Borrows OR Hal Mayne OR Keith Hayden OR David Eadington OR Michael Kearney OR Andrew Hollyer OR John Crispin-Bailey OR Julian Sturdy OR Luke Charters"

Appendix B

Facebook Ad and Post Visuals

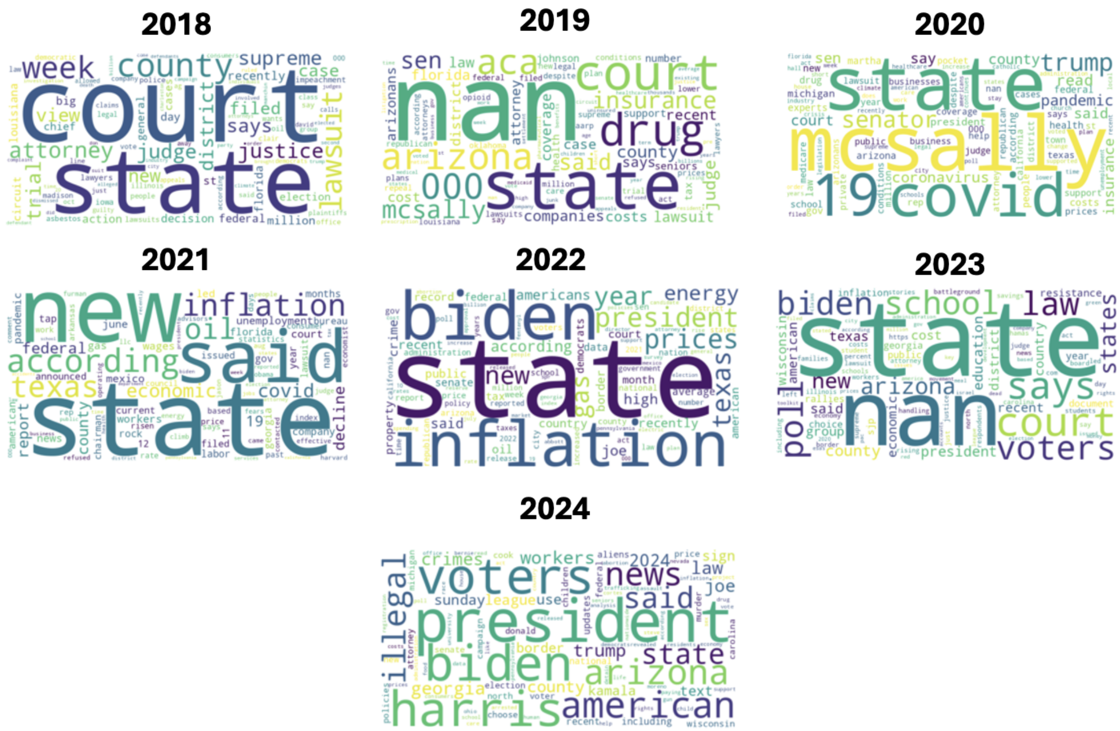


Figure B.1: Wordclouds of the Top 100 Words Appearing in Pink Slime Facebook Ads Over Time

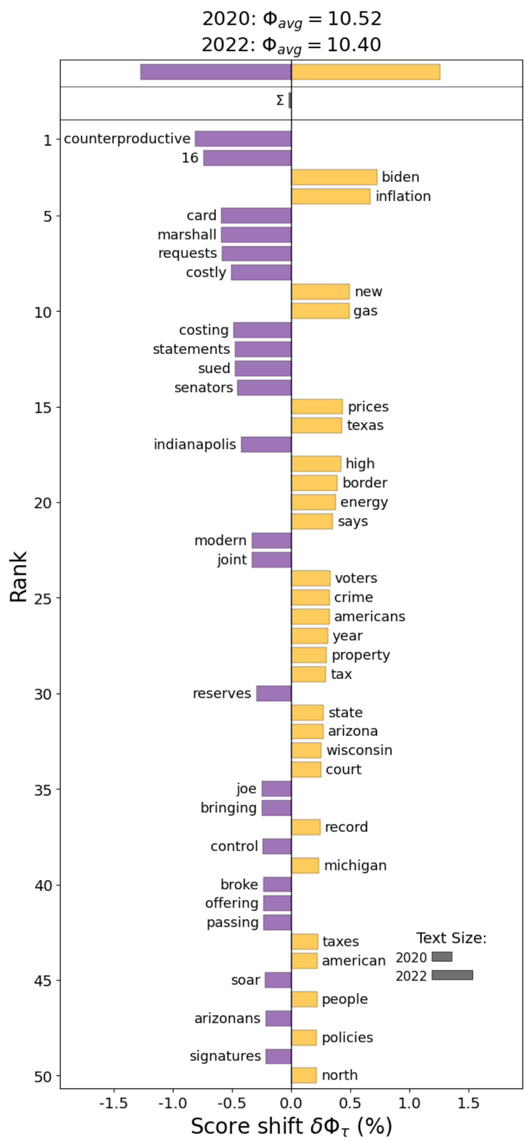


Figure B.2: Change in words used in Facebook ads by Pink Slime Organizations in 2020 (left) and 2022 (right)

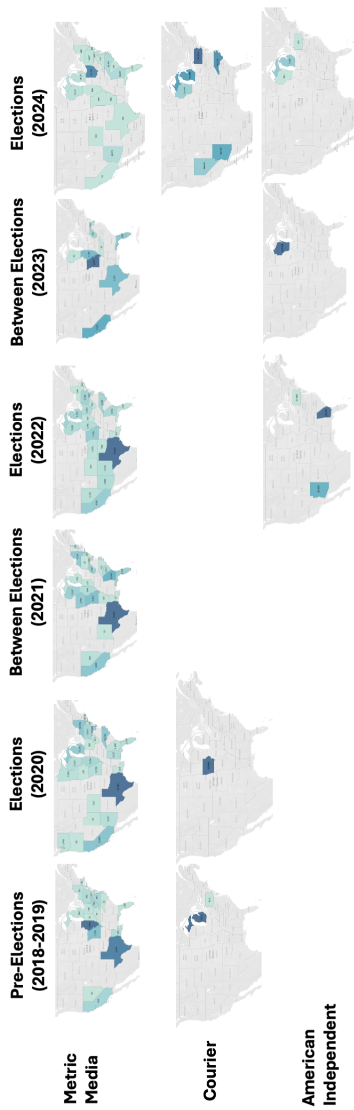


Figure B.3: Total Facebook ad expenditure by state over time by the various pink slime organizations.

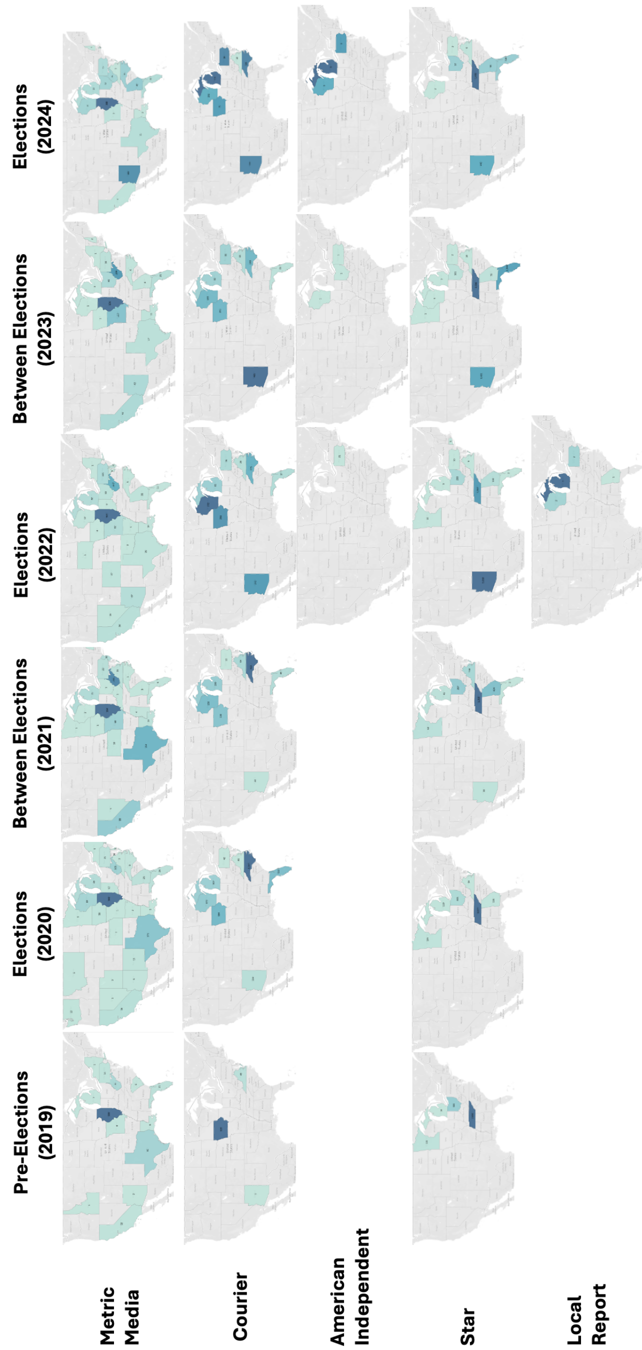


Figure B.4: Sum of all the posts linking from public Facebook groups to pink slime sites targeting different states by year through August 2024.

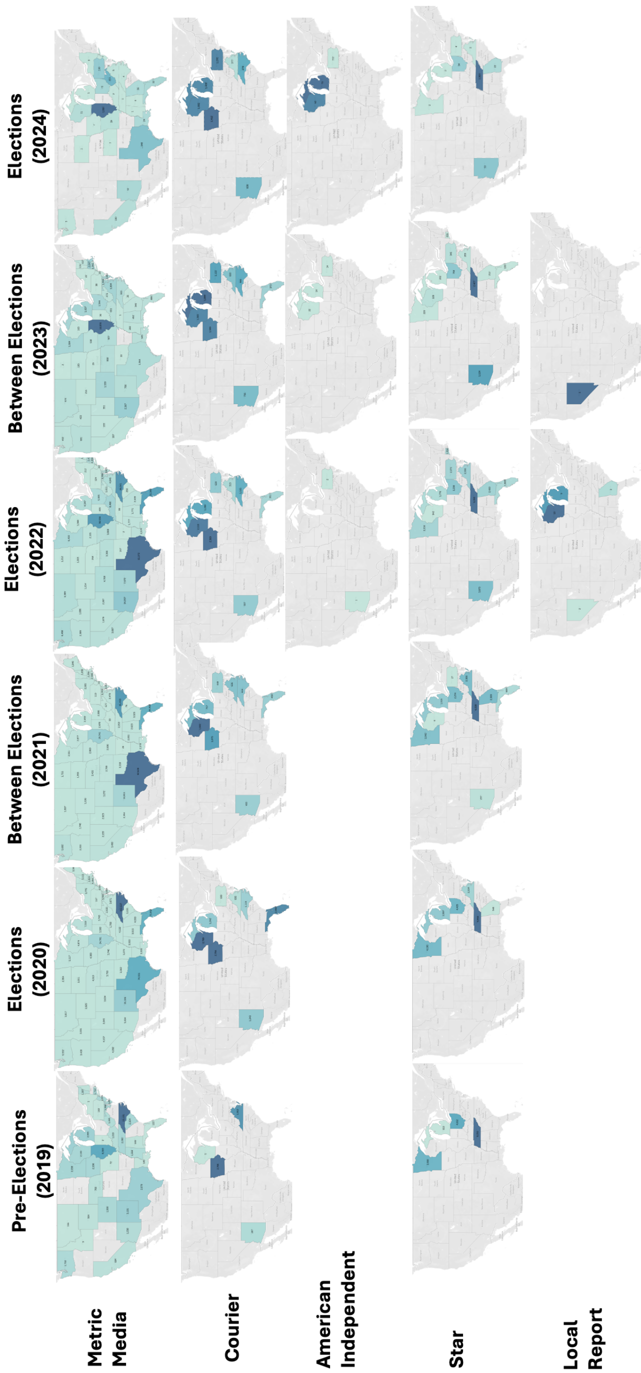


Figure B.5: Sum of all the posts linking from Facebook Pages to pink slime sites targeting different states by year through August 2024.

Appendix C

BEND Visuals

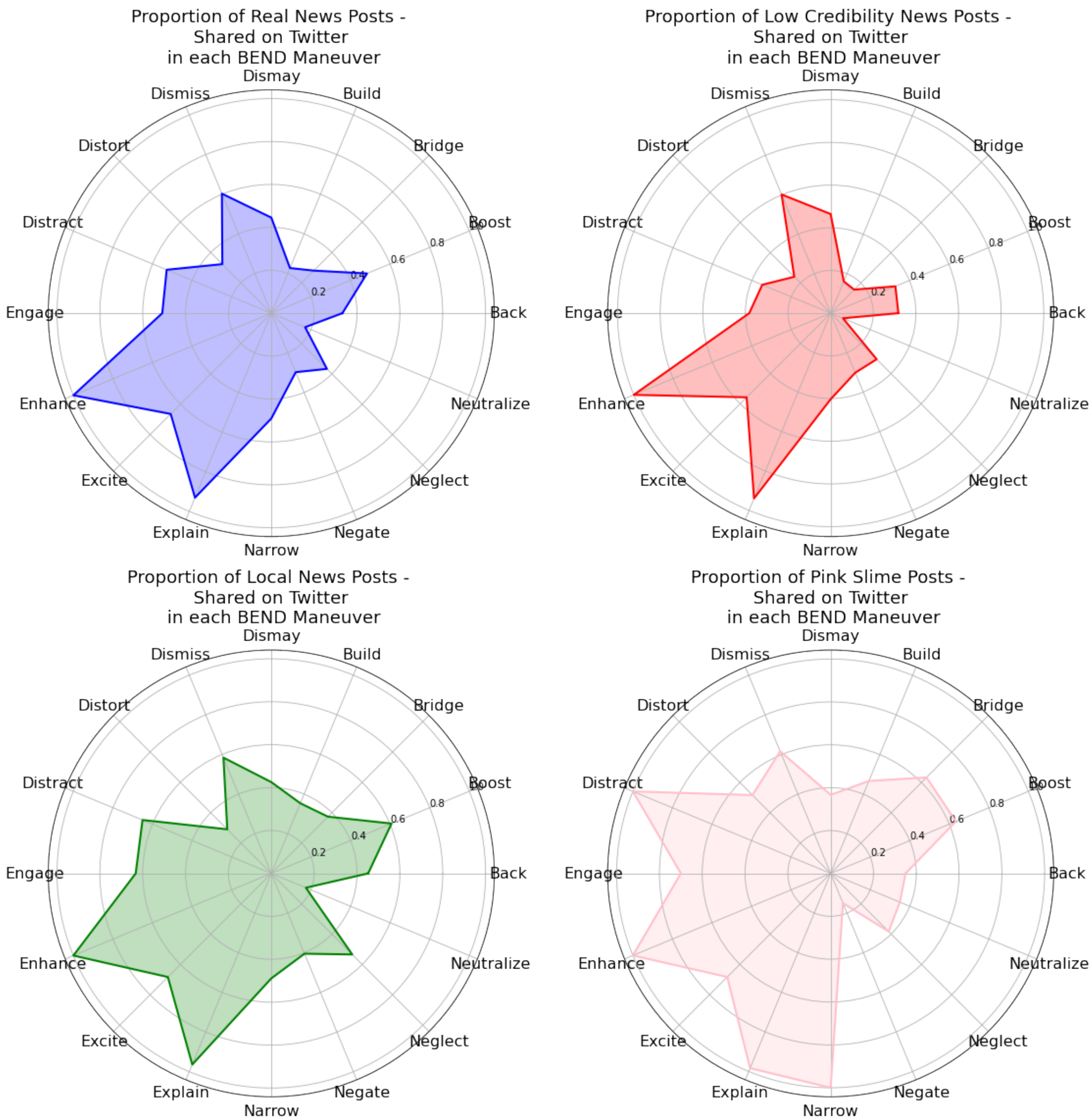


Figure C.1: Visualization of the proportion of posts for each news type fall into the BEND maneuvers on the Twitter midterms dataset for the Fetterman v. Oz senate race

Figure C.2: Visualization of the proportion of posts for each news type fall into the BEND maneuvers on the Facebook midterms dataset for the Fetterman v. Oz senate race

Figure C.3: Visualization of the proportion of posts for each news type fall into the BEND maneuvers on the Reddit midterms dataset for the Fetterman v. Oz senate race

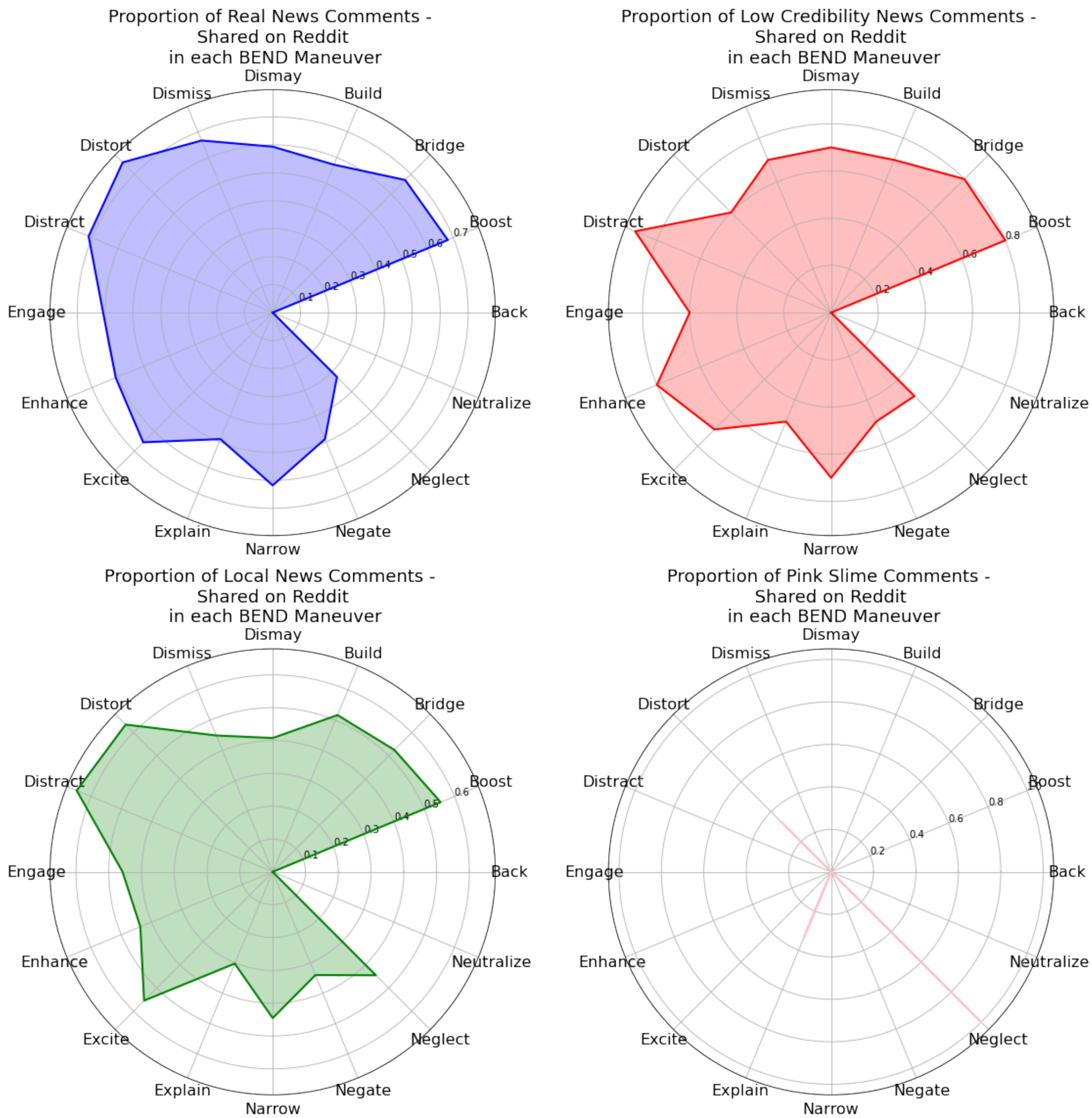


Figure C.4: Visualization of the proportion of comments for each news type fall into the BEND maneuvers on the Reddit midterms dataset for the Fetterman v. Oz senate race

Appendix D

Survey Posts

Figure D.1: Pre-Test Local News Post #1

27east.com
@27east

Stony Brook Southampton Hospital Suspended 16 Unvaccinated Staff On Tuesday.

27EAST.COM

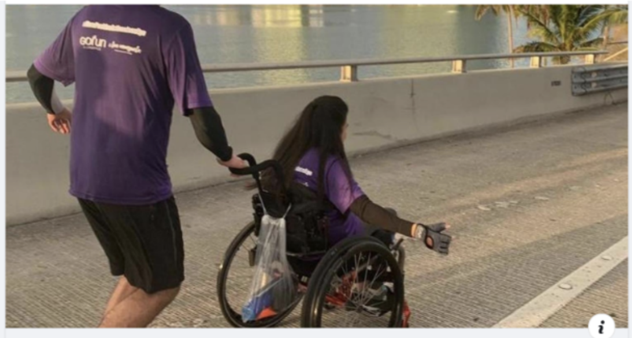
Stony Brook Southampton Hospital Suspended 16 Unvaccinated Staff On Tuesday - 27 East

5:11 PM Sept 27th, 2020

Figure D.2: Pre-Test Local News Post #2

Miami New Times
@miaminewtimes

For people in wheelchairs a couple of inches of water is enough to derail an entire day.



MIAMINEWTIMES.COM
Climate Change Is Already Affecting Miami's Disabled Residents
Residents with spinal-cord injuries struggle as rain and flooding become more frequent in ...

6:00 PM Oct 18th, 2021

Figure D.3: Pre-Test Local News Post #3

Cape May County News
@HeraldNews

Middle Township High School Senior Named 2024 Governor's STEM Scholar

CAPEMAYCOUNTYHERALD.COM
Middle Township High School Senior Named 2024 Governor's STEM Scholar - Cape May County Herald

12:22 PM Sept 15th, 2023

Figure D.4: Pre-Test Local News Post #4

Tennessee Star
@TheTNStar

More Than 1,600 Scientists, Nobel Laureates, Declare ‘Climate Emergency’ a Myth

TENNESSEESTAR.COM

More Than 1,600 Scientists, Nobel Laureates, Declare 'Climate Emergency' a Myth - Tennessee...

1:47 PM Aug 30th, 2023

Figure D.5: Pre-Test Pink Slime Post #1

The Pennsylvania Independent
@PennsylvaniaIndependent

Sweet news: Pennsylvania has the country's fourth-largest 'candy economy'

PENNSYLVANIAINDEPENDENT.COM
Sweet news: Pennsylvania has the country's fourth-largest 'candy economy' - Pennsylvania Independent

5:20 PM Oct 18th, 2023

Figure D.6: Pre-Test Pink Slime Post #2

Figure D.7: Pre-Test Pink Slime Post #3

Figure D.8: Pre-Test Pink Slime Post #4

Figure D.9: Post-Test Local News Post #1

The Des Moines Register
@DMRegister

It's allergy season in Iowa, and scientists say climate change might be making it worse.

DESMOINESREGISTER.COM

Sniffling more this allergy season? Scientists might have an explanation

While the study relies on compiled data from across the United States, Hartzler warns that ...

2:00 PM Sept 3rd, 2022

Figure D.10: Post-Test Local News Post #2

clevelanddotcom
@clevelanddotcom

East Palestine residents scream for emergency declaration, months after train derailment.

CLEVELAND.COM

East Palestine residents scream for emergency declaration, months after train derailment

4:15 PM July 17^{th}, 2023

Figure D.11: Post-Test Local News Post #3

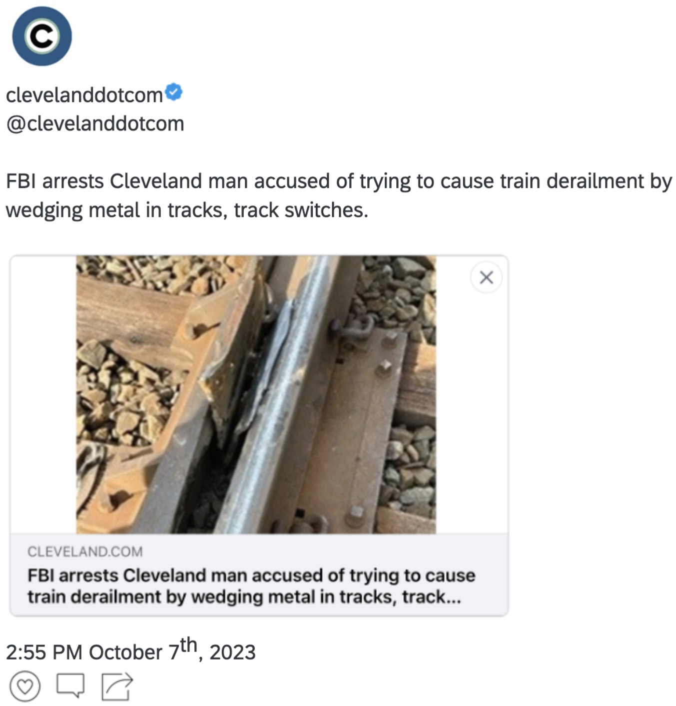


Figure D.12: Post-Test Local News Post #4

Figure D.13: Post-Test Pink Slime Post #1

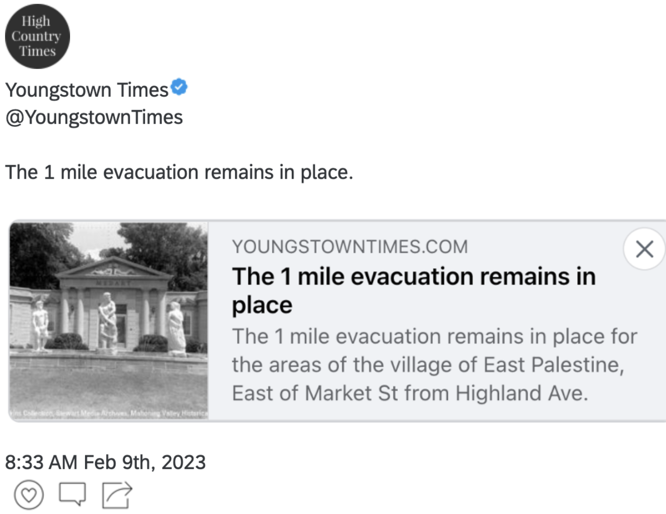


Figure D.14: Post-Test Pink Slime Post #2

Figure D.15: Post-Test Pink Slime Post #3

Figure D.16: Post-Test Pink Slime Post #4