Thesis Proposal

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

Christine Sowa Lepird

February 26, 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.

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 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 impacts and 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. 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 find pink slime. Finally, I summarize the findings and relevant literature to make policy recommendations to counter this threat to local communities.

Acknowledgments

To my research collaborators who have brainstormed with and inspired me to find new solutions.
To my advisor for supporting me in researching a topic I believe is important and understudied.
To my parents for encouraging their daughter to be curious and look to science for answers.
To my husband, my greatest collaborator, for the thousands of dinner-time conversations about pink slime (an unappetizing phrase) and unending support.

Contents

Chapter 1

Introduction

1.1 Overarching Thesis Goals

While the digitization of the news industry has allowed more people to access information at the tip of their fingertips, 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 [36] 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 [47]. Furthermore, the publications have not analyzed the phenomenon past the 2020 U.S. Presidential election [27] [47] despite the ad spend and impressions on their largest social media sharer (Facebook) more than doubling from 2020 to the 2022 midterms and January 2024 ad spend reaching over 10x that of January 2022.

In this thesis I will apply social network analysis 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?

Chapter 2

Data and Tools

2.1 Data

This thesis largely analyzes how users interact and engage with pink slime sites as they are shared on various social media platforms. The following Fig 2.1 summarizes where each of these data sources are used throughout the chapters of the thesis, and they are explained in greater depth below.

		Chapter				
Dataset	**Size**	1	2	3	4	5
Facebook Group and Page Posts to Pink Slime Domains	1,249,643 posts					
Facebook Pink Slime Ad Data	4,281 ads					
Scraped Pink Slime Webpages						
Crowdtangle Facebook COVID 2020	600,000 posts					
Facebook/Reddit/Twitter All News Types Midterms 2022	1,347,917 posts					
Balikatan Facebook/Reddit/Twitter 2022						
March 2020 Twitter COVID Dataset	1.2 million tweets					
Trident Juncture 2018 Twitter Dataset	230,000 tweets					

Figure 2.1: Summary of data used in the chapters of the thesis

2.1.1 News Type Datasets

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 [6] lists many news sites and rates how factual and credible the reporting is for many; the George Washington University Dataverse [44] has a list of over 9600 Twitter accounts for media organizations, derived from over 160 million tweets between 2016 and 2020; there is also a Github repository [11] 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 [9]. 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 [32]. Additionally, a set of authentic local news sites owned by larger companies are compiled from [12].

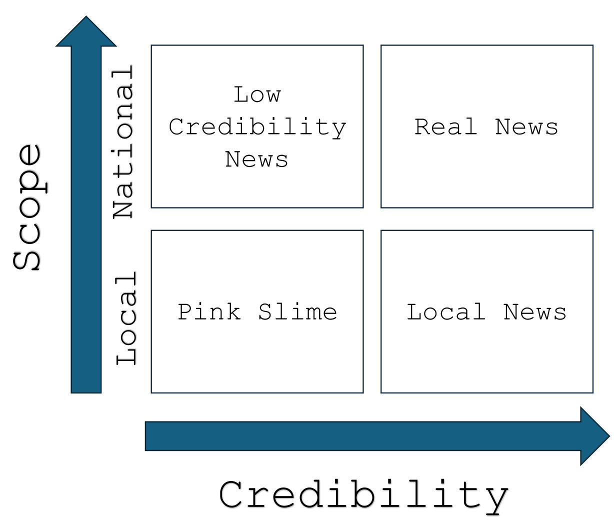

Figure 2.2: Relationships between news types

2.1.2 Facebook Datasets

Facebook Group and Page Posts to Pink Slime Domains Dataset consists of posts from the Python CrowdTangle API ([18]. For each pink slime domain listed in [9], 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 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 [2]. 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 ([18]. 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 [38]. For 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 [10], Reddit's Pushshift API [18], and Facebook's CrowdTangle API [3] were all used to pull the data for this research and was conducted with IRB approval in the Fall of 2022 Federalwide Assurance No: FWA00004206 IRB Registration No: IRB00000603.

2.1.3 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 [10] 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 [10] in March of 2020 with the following keywords: "coronaravirus", "coronavirus", "wuhan virus", "wuhanvirus", "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 [10] 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.

2.2 Tools Used

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

ORA is a dynamic network analysis and visualization tool with capabilities to import data from several social media sites [31]. It is used in this thesis visualize social media networks, calculate centrality metrics, and implement the BEND framework.

NetMapper is a text-based software [31] 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.

BotHunter [23] is a bot detection algorithm utilized to segment posts made on social media platforms as those from bots and humans. It outputs a probability ranging from 0 to 1, where a score closer to 0 is likely a human and a score closer to 1 is likely a bot. The cut-off for bot labeling in this thesis was 0.7 [51].

2.3 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 pending IRB approval in the Spring of 2024 Federalwide Assurance No: FP00010535.

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.

Chapter 3

Research Plan

This thesis is organized into six chapters. It begins by defining pink slime, what we know about it from previous research, 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 (Ch. 2). By including the text as well as network elements, it dives into which information operations are being conducted by the controlling pink slime organizations (Ch. 3). It establishes a new metric to find these sites and tests its performance (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 and case studies from similar phenomenon that has been observed internationally to make policy recommendations to address the issue in the United States (Ch. 6).

3.1 Background and Motivation (Ch. 1)

3.1.1 Research Questions

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 will describe 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.

- What is pink slime?
- 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?

3.1.2 Methods and Proposed Work

Defining 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, partisan groups
- 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

Furthermore, the concept of something being ‘local’ 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’

Background Hundreds of regional news sources that appear to be reliable local news have been spreading since 2019 [20]. These news websites consist of largely automated, low-quality partisan reporting and were nicknamed “Pink Slime” by journalist Ryan Smith in 2012 [55]. 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. 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 ChatGPT) [33].

American trust in local news organizations has remained higher than that of national news organizations [36]. To exploit the trust in local reporting, organizations like Metric Media LLC have created almost 1,000 local news sites [21]. 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 [15]. 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) [34].

The organizations like Media Metric that control vast swaths of pink slime sites do not appear to have foreign ties [21], 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. [48]”

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" [8]. 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 [4]. 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" [7].

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 [20]

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). Almost 70% of Americans get their news from the social media platform Facebook [37], but not all of this news is coming from quality news sources; 15% of referrals to fake news sites are coming from Facebook [39]. While 17.7% of visits to these sites are referred by Facebook 3.2% are through Twitter [47]. 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 [28].

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 autogenerated data stories, and those that made it to the front page pertained to state and national politics [54].

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 [47]. Furthermore, Biden supporters and people under 30 were more likely to visit these sites [47]. 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) [47]. Surprisingly, the consumption study also found that living in a news desert was not a significant predictor of visiting a pink slime site [47]. Finally, the researcher found that while a minority of pink slime sites are about politics, those are the ones most visited [47].

Parent Organizations Five parent organizations control over 1,000 pink slime news domains with the number of domains illustrated in 3.1. While these numbers are smaller than the Tow Center for Digital Journalism's published list [9], that is because the sites that have never been shared in the datasets in this thesis were excluded as well as any sites pertaining to outside of the United States. The largest organization, Metric Media, along with the Star News Network have conservative political leanings while the other three share liberal-leaning news.

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

Figure 3.1: Ownership of pink slime sites, colored by U.S. political leaning

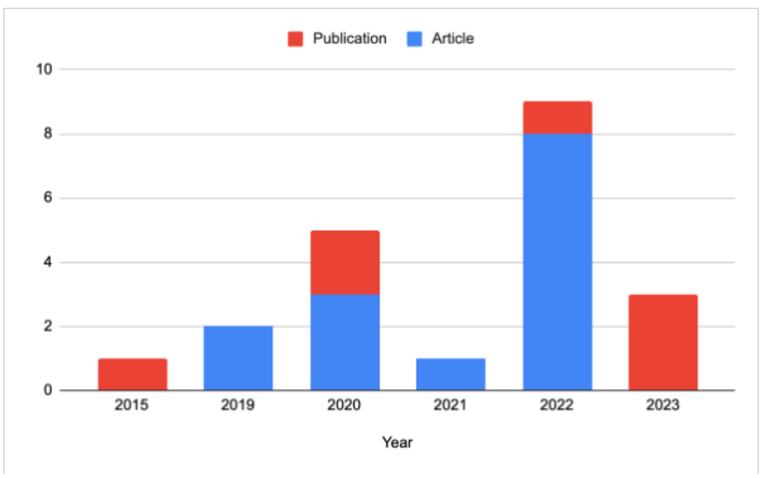

Figure 3.2: News Articles and Journal Publications Mentioning the Pink Slime Phenomenon

Published Research Per Fig 3.2, 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 [33]. 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 [47] [27] and who is funding it [4] [22], 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 [47] [19]; 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 [20]. 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 [53]. 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 [14] [46] [17]. In the final chapter (6), policy recommendations from the international incidents and based on the previous 5 chapters are recommended.

Research findings and gaps are referenced in Fig 3.3 as well as which chapter of this thesis will address those gaps.

Current Research Tells Us	Unanswered Questions	Where in this thesis is that question answered?
Who is consuming pink slime (exclusively during 2020 presidential election)	What are the network features of those sharing pink slime? What are they doing during midterms?	Chapter 2
Who is funding pink slime	Where are they spending that money, and what impact is it having?	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 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
Lessons 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
This phenomenon is happening internationally	What lessons can be applied to the situation in the United States? What policy should be enacted to counter it?	Chapter 6

Figure 3.3: Current Research Gaps

Future Work Future work for this chapter will involve actively following the research landscape of pink slime's academically published works and journalism's shared news on the topic.

Table 3.1: Facebook Ads Over Time

Year	Number of Ads	Total Impressions	Total Ad Spend
2018	364	1,301,818	$21,218
2019	1,130	3,313,935	$73,535
2020	1,960	25,785,521	$275,320
2021	1,735	14,787,133	$158,883
2022	3,640	42,881,181	$722,980
2023	1,396	18,617,304	$543,402

3.1.3 Challenges and Limitations

This chapter will need to continuously evolve until the dissertation is complete so that new articles published will be included and addressed in corresponding chapters.

3.2 Characteristics of Pink Slime (Ch. 2)

3.2.1 Research Questions

The key research questions for this chapter is:

- What has pink slime infiltrated communities the most?
- 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?

3.2.2 Methods and Proposed Work

Impact This chapter begins by describing the impact of pink slime via all of the Facebook ads the parent organizations have purchased and posts to public Facebook pages and groups that link to pink slime websites. The amount of ad spend and impressions - demonstrated in Table 3.1 - illustrates the importance of understanding the changes over time. While most academic literature focuses on pink slime during the most recent presidential election year (2020), less than 10 million impressions were garnered on ads in that year. The year of the midterm elections (2022) saw a strong increase in the ad spend and over 20 million impressions. These sites are only beginning to pick up traction, and general trends of these ads and posts are summarized by parent organization in Figure 3.4. Furthermore, ad spend by these organizations on Facebook was over 10 times higher in January 2024 than January 2022, which implies there will be a stronger effort during the 2024 election year.

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

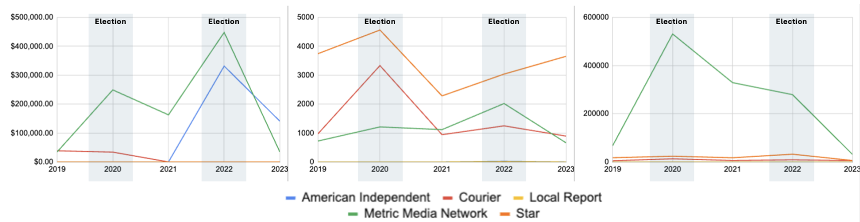

Figure 3.4: An over time plot of (i) Ad Spend (ii) Group Posts Linking to Pink Slime Site (iii) Page Posts Linking to Pink Slime Sites by Organization

using Python's wordcloud package[1] to formulate the word cloud including the top 100 words, sized by frequency of appearance. Figure 3.5 illustrates the changing focus of topics in election and off-election years. Throughout all of the years, the most targeted states ('Texas', 'Michigan', 'Georgia', 'Arizona', 'Florida', 'Wisconsin', and more) 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 presidential candidates' names were not at the forefront of the 2020 conversation, there was a focus on the phrase *Catholic*. The ads containing references to 'Catholic' 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.

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

[1]https://pypi.org/project/wordcloud/

Table 3.2: State variables and their Pearson correlation to 2022 pink slime ad spend

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

active and visible to the Facebook audience even during the down time.

Figure 3.5: Wordclouds of the top 100 words used in Facebook ads by Pink Slime Organizations over Time

With the advertising expenditure segregated by states in Figure 3.6 , I analyze the expenditure per state. 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 American Independent purchased ads in 2022 and 2023. All of the organizations who purchased ads targeted Pennsylvania - a key swing state that determined Biden's 2020 victory over Trump. Other swing states, like Michigan, Georgia, and Wisconsin were targeted by two of the three organizations.

The top ten states by ad spend (from highest to lowest) are Wisconsin, Georgia, Texas, Arizona, Pennsylvania, Iowa, California, Missouri, Michigan, and North Carolina.

While swing states are a likely place to target during election years, other states' significance was seen as a result of isolated events. For example, Courier spent $33,998 in ads to Iowa in 2020.

When analyzing the relationship between a state's attributes and the amount pink slime organizations spent to influence that state via ads in 2022, I find in Table 3.2 that 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.

Finally, to see how the ad spend influenced organic conversation, Figure 3.7 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.

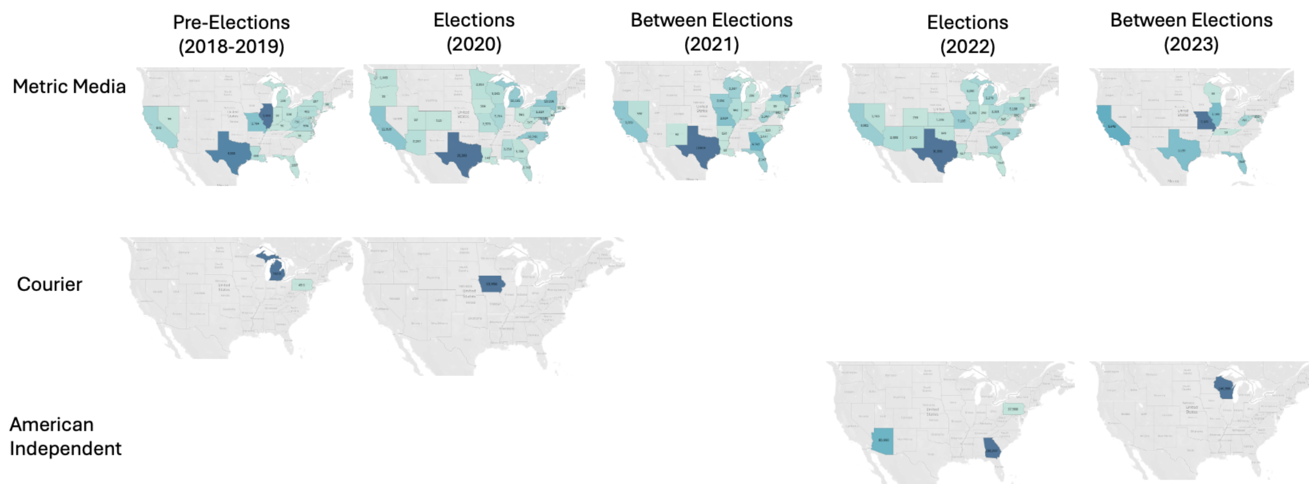

Figure 3.6: Advertising Expenditure by State

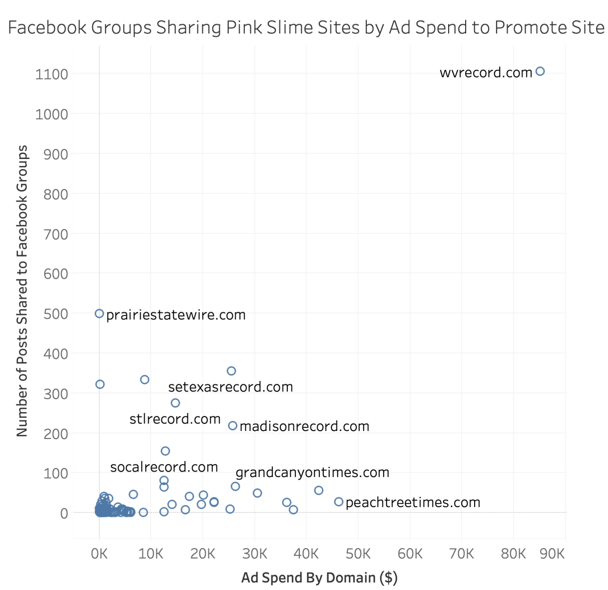

Figure 3.7: Number of Instances a Pink Slime Domain Appears in a Facebook Group by Ad Spend for those Domains

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

Table 3.3: Number of links with labels in the midterms dataset by platform

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 3.4: 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.

How pink slime differs from 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.

During the 2022 U.S. Midterm elections, posts pertaining to contentious elections and linking to external URLs were collected from Twitter, Facebook, and Reddit; the number of links from each platform can be seen in Table 3.3.

In analyzing the breakdown of news type shared by platform (see Fig. 3.4), it was discovered that all of the platforms studied had some pink slime news shared to them.

While pink slime sites make up a very small percentage of news shared on these platforms, they are disproportionately shared to smaller Facebook pages than other news types. Table 3.5 shows how 85% of pink slime site shares are to the smallest 50% of the Facebook pages.

Due to the differences in size of the audience it was shared to, their relative engagement should be considered; 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's 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. When the relative engagement of posts are separated by news type, an alarming trend emerges in Fig. 3.8 - pink slime receives a greater relative engagement than any other news type. This news is resonating with its audience, and real news is not. Further description of this

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

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 3.6: Relative Engagement Metrics by News Type

variable is broken down in Table 3.6

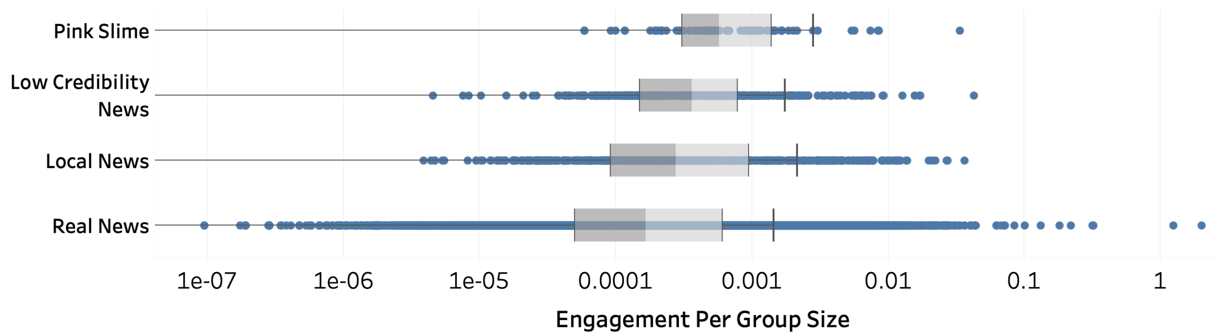

Figure 3.8: Facebook engagement per group size by news type

When trying to understand the agents sharing content on these platforms, Figure 3.9, made using visualization software ORA [30] [16], illustrates the changes in what news type a user on a platform will post after previously posting from a designated news type. Across all platforms, the most likely outcome is for a user to continue to share content within the same news type category.

However, there are some differences between the platforms. Outside of sharing the same content as previously shared, subreddits have a higher chance of going between sharing real news and local news. None of the subreddits who shared local news in this dataset also shared pink slime, and vice versa.

On Facebook, pink slime posters would also share local news and real news sites, which may give the pink slime more credibility in the eyes of the viewers.

Given that Twitter had the most posts in this dataset, its users were split into two categories before news spread was analyzed - bot accounts and accounts that were not bots under the delineation of [50]. Surprisingly, the behavior between bots and non-bots remained consistent. Pink slime was co-shared with all news types on Twitter, with a high likelihood of users sharing pink slime and low credibility news one after the other.

I filtered down the dataset to exclusively include agents on all platforms (Facebook, Reddit, and Twitter) who shared posts to pink slime sites to better understand the online behavior of these individuals. In Figure 3.11, using visualization software ORA [30] [16], the network of such accounts to the links they shared is visualized. The agents (gray nodes) link to the different news types. 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.

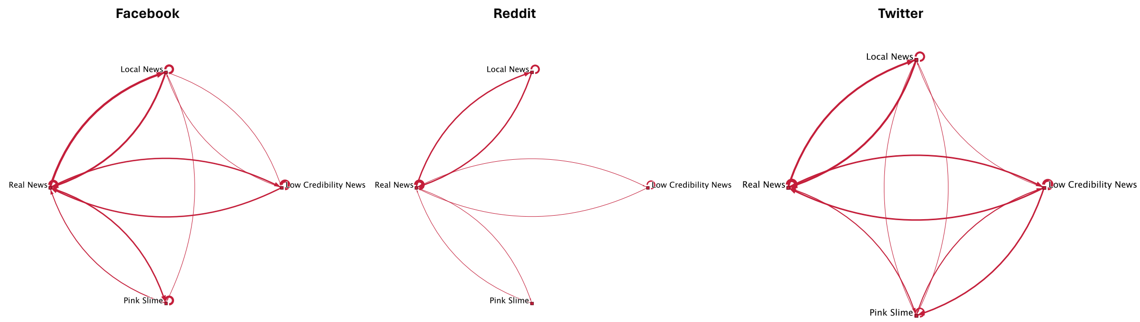

Figure 3.9: User likelihood (represented by line thickness) of sharing one news type based on previous news type shared by platform.

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%

Figure 3.10: Percentage of users who by platform and news type of continue to post within the same news type (i.e., self-loops).

The pink slime domains in the right component are all under the control of parent organizations pushing politically right-leaning news. Grandcanyontimes.com and keystonetoday.com are controlled by Metric Media. Georgiastarnews.com, theohiostar.com, and tennesseestar.com are all owned by the Star News Network.

The pink slime domains in the left component are all under the control of parent organizations pushing politically left-leaning news. Coppercourier.com, keystonenewsroom.com, cardinalpine.com, and upnorthnewswi.com are all controlled by the Courier Newsroom. 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.

To further assess this phenomenon, the agents in the subset of users sharing pink slime were labeled as "Left" for those sharing left-leaning pink slime sites, "Right" for those sharing right-leaning pink slime sites, and "Both" for the agents who shared pink slime sites from parent organizations on both sides of the political spectrum. For each of the three agent types, the distribution of news types that they shared as a total of all of the links they posted was calculated and can be seen in Figure 3.12.

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, we used a Chi-Squared test to compare the three distributions.

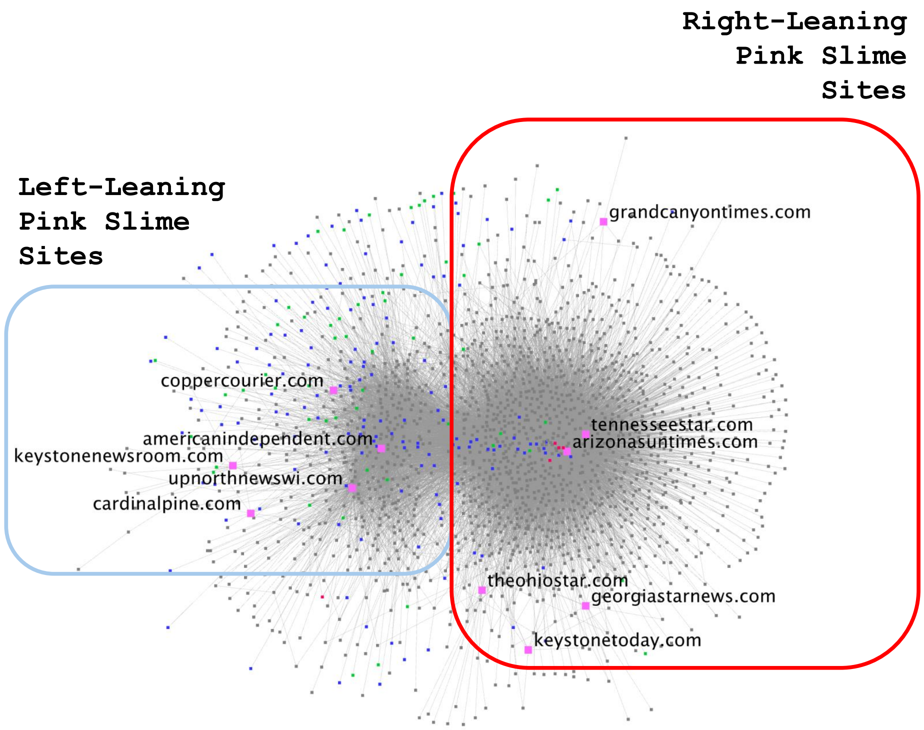

Figure 3.11: News sources shared by users (including all platforms) who shared pink slime domains. Pink slime sites are labeled and given a pink node coloring, local news sites are green nodes, real news sites are blue nodes, and low credibility news sites are red nodes.

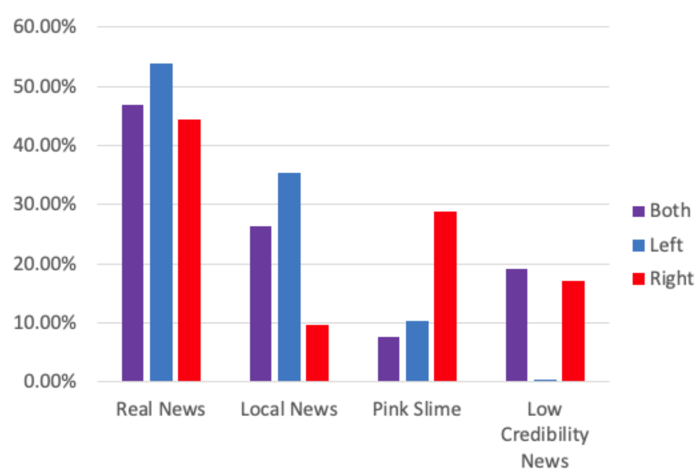

Figure 3.12: Distribution of the news types shared by agents who shared pink slime, grouped by whether the agent shared pink slime from a right-leaning pink slime organization, a left-leaning pink slime organization, or both

$$\chi^2 = \sum \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where:

χ^2 = Chi-squared statistic

O_{ij} = Observed frequency in cell (i, j) of a group of agents sharing a news type

E_{ij} = Expected frequency in cell (i, j) of a group of agents sharing a news type

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.

Next Steps Future work will focus on understanding how pink slime is spread via URL backlinks and how they differ from other news types as well as analyzing engagement on Twitter and Reddit.

3.2.3 Challenges and Limitations

This section is limited to the news types that are in the 32,000 labeled news sites in the CASOS news thesaurus. While most of the work has been completed, there are some challenges in finding differences in the backlink behavior between pink slime and authentic local news sites.

3.3 BEND Maneuvers of Pink Slime (Ch. 3)

3.3.1 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?
- Do maneuvers differ between parent organizations or topic?
- What linguistic cues can be included to identify the proper maneuvers?
- How do the maneuvers compare to those of local news organizations?

3.3.2 Methods and Proposed Work

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 [29]. 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

	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 3.13: Definitions of the 16 BEND Maneuvers, adapted from [24], [25], and discussions with the authors.

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 3.13. 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 [26], the Chinese balloon incidents [49], and events in Indonesia [35]), 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**.

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 [41] as well as the list of authentic local news sites owned by companies [12]. Using the CrowdTangle API [3], for each

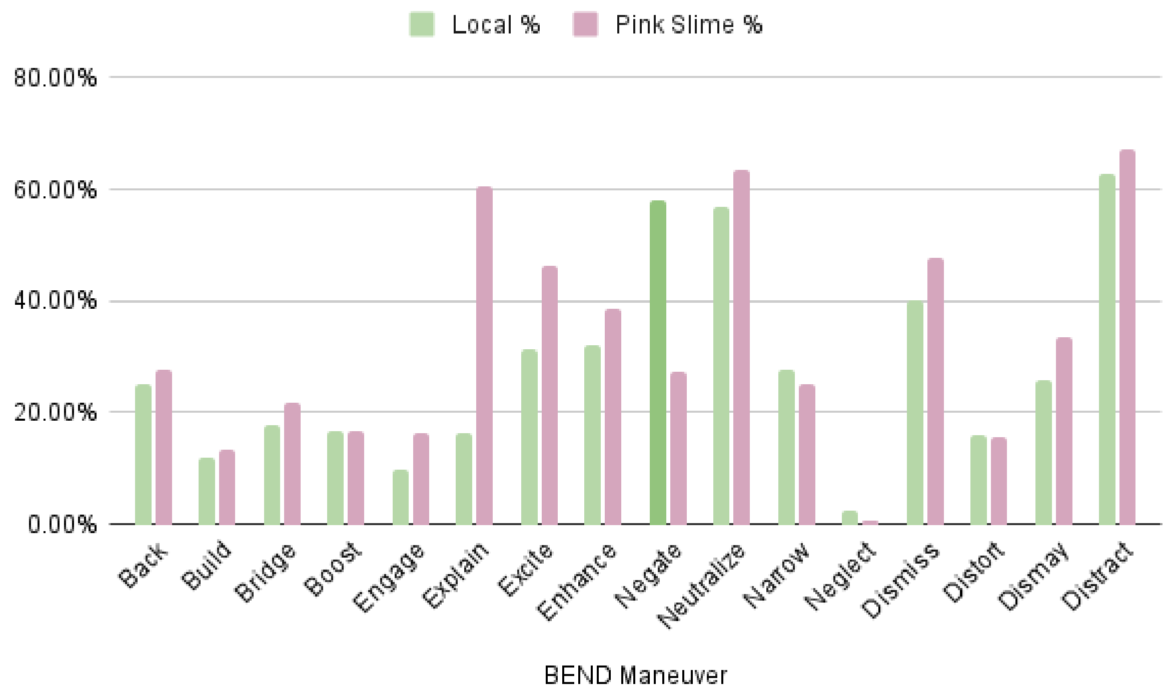

Figure 3.14: Percentage of Posts Using BEND Maneuvers by News Type

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 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 [47] 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 3.14 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 3.15 takes the values from Table 3.14 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.

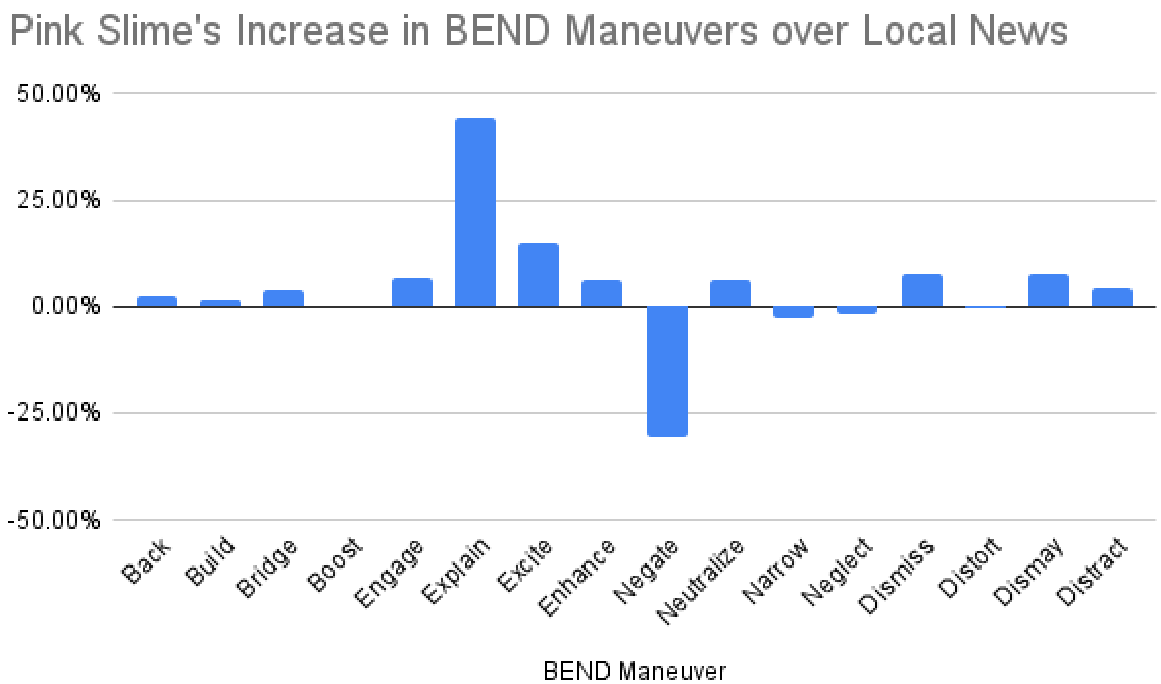

Figure 3.15: Increase in pink slime posts using BEND maneuvers over local news posts

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 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.

Future work After finding that there *are* differences in BEND maneuvers that pink sime sites and local news sites use to discuss elections, I chose to expand this research to include more variables. Future work for this chapter will involve performing a BEND analysis on the midterms dataset - comparing BEND maneuvers used on Twitter, Reddit, and Facebook and how it varies across the four news types discussed in Chapter 2 - low credibility news, real news, pink slime, and local news.

3.3.3 Challenges and Limitations

Challenges for this chapter involve include the appropriate network features from both Facebook and Reddit to create a community maneuver network that rivals Twitter's mentioning network. This will involve several attempts at trying different networks to find the one that shows coordination. Furthermore, new linguistic cues will be need to be generated as the research evolves. Limitations are primarily within the datasets I possess. Specifically, the Facebook and Twitter datasets are different because while Twitter has user-level posts, Facebook has page and group level posts. This may lead to differences in how BEND can be measured when they are compared to one another.

3.4 Finding New Sources of Pink Slime (Ch. 4)

3.4.1 Research Questions

The key research question for this chapter is:

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

3.4.2 Methods and Proposed Work

Academic research into site credibility algorithms can serve as an inspiration to understanding methods to find relevant websites of interest. In a precursor to PageRank, Kleinberg proposed a model that found authoritative webpages for search topics by analyzing the relationship between "authorities for a topic and those pages that link to many related authorities ... [called] hubs" [43]. Using the reasoning that "hubs and authorities exhibit what could be called a mutually reinforcing relationship: a good hub is a page that points to many good authorities; a good authority is a page that is pointed to by many good hubs" [43], they created a hub score that incorporated the authority scores of the nodes pointing to it. Likewise, they gave nodes a high hub score if it linked to nodes that are authorities on the subject. While others have taken these methods and applied them to citation networks and ranking academic journals, this research sets about viewing certain Facebook pages as hubs of disinformation and building out the network of authorities that they share.

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. 3.16. Additionally, Chapter 2 shows us that agents sharing pink slime are most likely to follow up by posting *more* pink slime.

The methodology for this research 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 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.

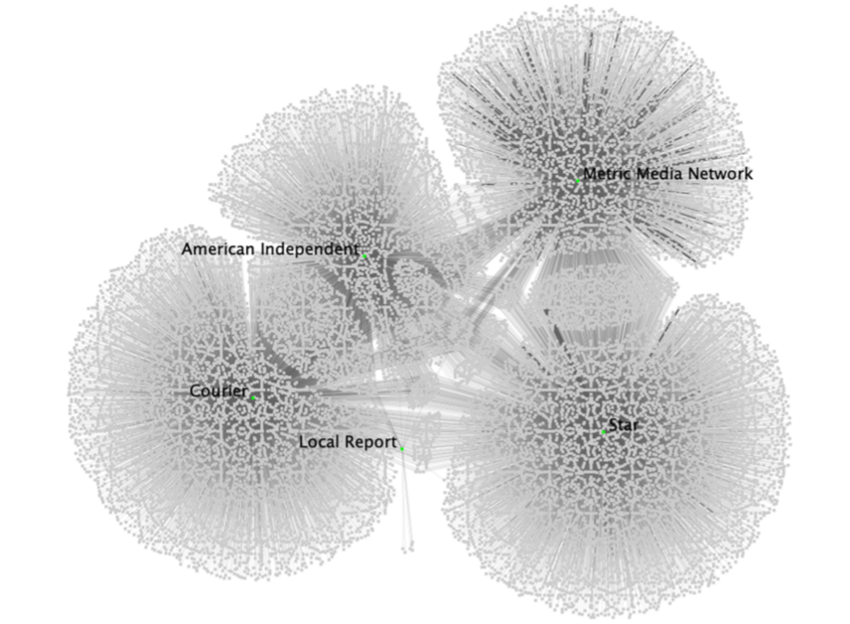

Figure 3.16: Network visual of Facebook Pages linking to Parent Organizations of Pink Slime Sites

From there, the larger, unlabeled domain dataset was analyzed using the set of features to find new source of pink slime.

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 (Noncredible Sharer Score). The higher the Noncredible Sharer Score, the lower the credibility of the Facebook Page. Then, the domains could be scored by averaging the Noncredible Sharer Score of the pages that shared the domain.

A common use case for this scoring algorithm would be to upload a Facebook dataset for topic of interest, run the algorithm and find the Noncredibility Scores, and then sort by descending Noncredibility Scores. This ranks sites from lowest credibility and is a place where the analyst can then use visual inspection and visit some of the highest scoring domains.

$$\text{Noncredible Sharer Score}_k = \frac{1}{n}\sum_{j=1}^{n} I_j \tag{3.1}$$

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

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

$$\text{Noncredibility Score}_z = \frac{1}{|K|}\sum_{k=1} \text{Noncredible Sharer Score}_k \tag{3.2}$$

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

In meta-network terms, in a network that linked Domains to Domains by the Facebook Pages that shared them (Domain x Facebook Page x Domain), the Noncredibility Score for any given domain would be the percentage of first degree neighbors that were known sources of pink slime or fake news.

For example, a page that shared 5 news articles, none of which link to domains that were labeled as fake news was assigned a score of zero. However, if a page shared 10 news articles and 5 of the articles were to domains labeled as fake news, that Facebook Page would be assigned a Noncredible Sharer Score of 0.50. Each Facebook Page that shared a domain had its score averaged to assign the domain a Noncredibility Score.

These network features, as well as those described in the Data Description, 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 fake 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 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 [13].

To see if the labels of "pink slime", "real news", and "low credibility news" could be predicted using the features extracted in the Data Description and Methods sections, a Naive estimator that assigned classifications to the dominant class (real news) was used as a baseline to compare to another estimator, the XGBoost model. In comparison, the XGBoost model demonstrated an increase in differentiating classification between real news and fake news. Low credibility news and real news saw an uptick in precision and recall when the actual estimator was used. However, class imbalance played a role in the classification of pink slime. Due to the extremely low support for pink slime instances in the dataset, the model struggled to find a threshold for pink slime that increased the model's accuracy.

While the accuracy value of the pink slime classification was low, the overall intent of the features are to use them as a starting point of analysis to find higher likelihood domains faster. Since the goal of this research isn't to do human-out-of-the-loop, the accuracy scores don't paint the entire picture of the model's performance. The below Figure 3.17 illustrates the strength of the model in predicting pink slime (despite the low occurrence) through the ROC curve. Overall its area under the curve (AUC) of 0.77 provides an acceptable value to use the network features of the model as a ranking heuristic in large scale data analysis [40]. When looking at how well the model could predict the presence of fake news, the ROC curve is slightly stronger with an AUC of 0.78.The presence of real news was similarly well predicted to that of pink slime. The ROC curve, with its AUC of 0.77 shows an ability for this approach to be generalizable to these three difference news types.

When analyzing the feature importance of the variables included in the model, only two had an effect greater than there. The strongest one was the domain's Noncredibility Score. The other attribute, with 25 times less importance, was the average number of likes a post from the domain received. These results validate that using a network-based approach to assessing credibility can be used to predict news labels.

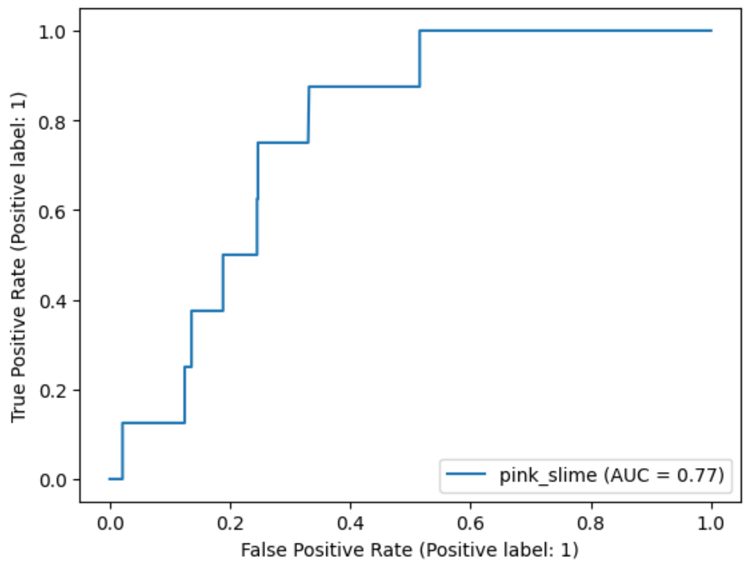

Figure 3.17: ROC Curve of Predicting Pink Slime

The primary function of the attributes discussed was to apply these measures to new datasets to quickly rank and assess probable new pink slime news agencies.Labeled pink slime sites are a small minority of the labeled news sources in a given dataset. Due to the issue of class imbalance, the general results for predicting pink slime show low accuracy but promising AUC curves. Given these results, using the model itself is not to output a list of two possible pink slime sites in a dataset of tens of thousands of domains. Rather, the intent is to apply the new attributes to new datasets and sort by the Noncredibility Score for a shorter, more targeted list of potential new pink slime sites. While it did not show measurable effect on the model, the attribute describing whether a domain contains a location can still be used as a filter for the quickest analysis.

Future work Future work for this chapter would be to apply this methodology to the Twitter and Reddit midterm datasets and develop platform-specific alterations depending on the accuracy. Additionally, more features of the dataset will be tested to see if they can improve the AUC.

3.4.3 Challenges and Limitations

The primary challenge in this chapter is finding datasets with a representative amount of pink slime (as seen in Chapter 2, this is less than 1% of news sites shared during elections) such that a 30% holdout will allow for enough class balance to have strong results.

3.5 Training Humans to Detect Pink Slime (Ch. 5)

3.5.1 Research Questions

The existence of pink slime alone is not a threat. However, the human consumption and spread of these sites is a danger. This chapter looks to see how humans process encountering pink slime on social media through a field study as part of Project OMEN. They will take 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.

While 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 [53], 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. Additionally, 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.

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?

3.5.2 Methods and Proposed Work

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. The game's objectives and setup can be found in [42]. As part of the measurement of effectiveness of training, a pre and post test of media literacy has been conducted. 23 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: covid and vaccination, climate change, and the Ohio train derailment. For stories with external links, participants are allowed to visit the websites. For each posts, the participants must answer a series of questions including:

- How trustworthy do you consider the poster of this message to be?
  - Trustworthy
  - Somewhat trustworthy
  - Neither trustworthy nor untrustworthy
  - Somewhat untrustworthy
  - Untrustworthy
- Do you believe the post was written by a local reporter?
  - Yes

- No
- Unsure

After the pretest, the participants were given a 31-minute training adapted from PBS [5] to include defining pink slime journalism, 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 3.18.

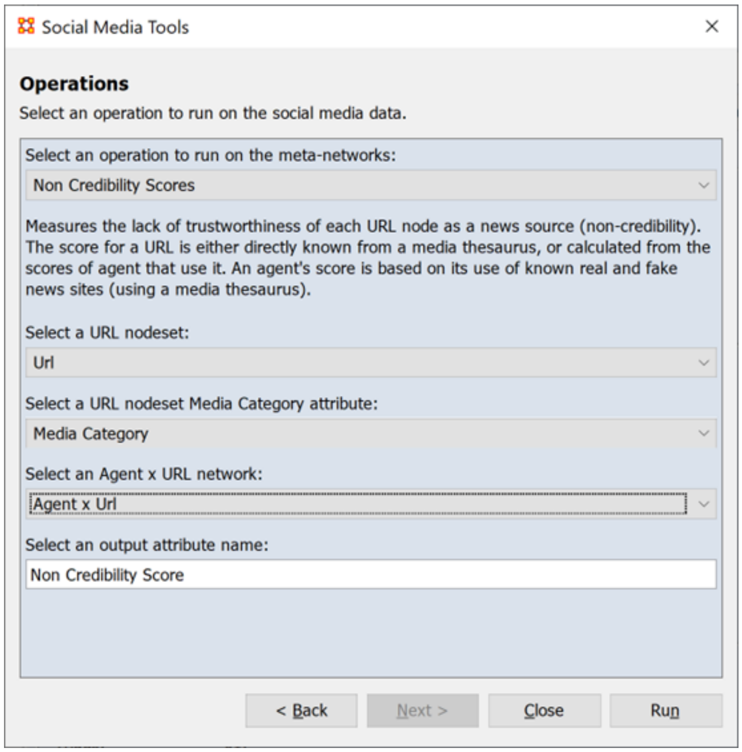

Figure 3.18: ORA interface for running the network features described in Chapter 4 as part of the lesson plan

After the training, the participants were given a post test with 16 new social media posts (with the same news type distribution). 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 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 3.19. Furthermore, since research shows that exposure to malicious news types can lead to a corroded trust of credible news sources [52], 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 2.5 on the trustworthy scale. After the training, trust of local news actually improved to a value of 2.1, while pink slime was rated a 4.2, indicating that the training had the intended affect of keeping trust in authentic local news high while increasing awareness of pink slime journalism's un-trustworthiness.

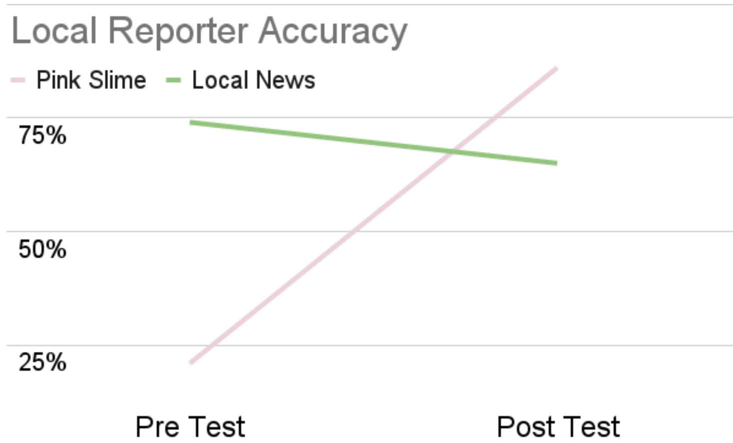

Figure 3.19: Participants' ability to correctly identify pink slime and local news before and after training

Future work Further analysis will focus on whether participants' responses for trust levels in the pink slime increased or decreased depending on if they clicked the embedded news story URLs and whether trust in local news decreased after the training. Statistical analyses will be conducted to determine whether the training resulted in significant changes in the ability for participants to identify pink slime.

3.5.3 Challenges and Limitations

The main challenge will be having enough participants to have strong results.

3.6 Conclusions and Recommendations (Ch. 6)

3.6.1 Research Questions

This chapter will highlight key findings from the previous 5 chapters as well as look 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 other countries done to control pink slime?
- What policies could decrease the threat of pink slime?

3.6.2 Methods and Proposed Work

While pink slime has not been an internationally-adopted phrase, the tenants of purporting to be local news but controlled by a larger national entity to sway public opinion is not a uniquely American phenomenon. In 2019, the European External Action Service's East Stratcom (an EU Disinformation Task Force) disclosed that 265 fake local news sites operating in 65 countries were under the control of an Indian influence campaign to undermine international support for Pakistan [14]. Much like American pink slime, these sites had domains with local areas embedded, published stories that were largely copy and pasted as local news to different areas, and had social media presence. Some even acquired previous *authentic* local news domains after the newspapers went out of business. The task force made policy recommendations and urged for reform to the domain name industry.

In early 2024, it was discovered that China had created over 100 'local' news sites in 30 countries in Europe, Asia, and Latin America to spread pro-Beijing content as part of an information operations campaign[1]. Efforts by Russia to influence local communities in Europe [46] and a campaign to hijack German local news [17] have been seen in the past, and will be analyzed (along with resulting policy changes) in this chapter.

Additionally, Some outlets have labeled pink slime as misinformation [45], 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.

3.6.3 Challenges and Limitations

Not all of the international incidents seen have generalizable nor applicable policy recommendations to the United States' pink slime phenomenon, and these limitatations will be discussed in this chapter.

Chapter 4

Contributions

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.

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 will contribute the methodology to apply the BEND framework to Facebook posts and categorize pink slime sites (as well as local news ones) into narrative and network maneuvers.

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.

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 2019-2023. 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.

4.5 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 will be 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.

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* interest. The information is not analyzed for its factual validity but rather for its marketed duplicity as local news.

Chapter 5

Proposed Timeline

Figure 5.1 shows my proposed timeline from Spring 2024 through my projected thesis defense in December 2024.

	Completed	Spring 2024	Summer 2024	Fall 2024
Ch 1: Background & Motivation	Data Collection, Literature Review, Analysis			
Ch 2: Network Features	Data Collection, Social Media Analysis	Finish SEO Analysis		
Ch 3: BEND	Data Collection, First Analysis	Analyze full dataset	Write Up Results	
Ch 4: Source Detection	Testing on FB Data, Analysis, Writing		Implement on Twitter and Reddit	
Ch 5: User Testing	Game Design, IRB, Deploy Game, Collect Data	Analyze and Write Up Results		
Ch 6: Conclusion and Recommendations			Conduct review of international examples	√
Finalize Thesis Document				√
Thesis Defense				√

Figure 5.1: Planned timeline for work to be completed

Bibliography

[1] PAPERWALL: Chinese Websites Posing as Local News Outlets Target Global Audiences with Pro-Beijing Content - The Citizen Lab — citizenlab.ca. https://citizenlab.ca/2024/02/paperwall-chinese-websites-posing-as-local-news-outlets-with-pro-beijing-content/. 3.6.2

[2] Ad library, . URL https://www.facebook.com/ads/library/?active_status=all&ad_type=political_and_issue_ads&country=US&media_type=all. 2.1.2

[3] CrowdTangle, . URL https://www.crowdtangle.com/. 2.1.2, 3.3.2

[4] Dark money political ads proliferate on facebook and instagram, . URL https://www.newsguardtech.com/misinformation-monitor/october-2022. 3.1.2, 3.1.2

[5] Lesson plan: How to spot 'pink slime' journalism — misinformation in long-trusted local news - PBS NewsHour classroom, . URL https://www.pbs.org/newshour/classroom/2022/10/lesson-plan-how-to-spot-pink-slime-journalism-misinformation-in-once-trusted-local-news/. 3.1.2, 3.5.1, 3.5.2

[6] Media bias/fact check news, . URL https://mediabiasfactcheck.com/. 2.1.1

[7] Nutrition label, . URL https://www.newsguardtech.com/metric-media-network/. 3.1.2

[8] Secretly partisan-funded websites posing as independent local news sites on verge of outnumbering daily newspapers in the u.s., . URL https://www.newsguardtech.com/press/partisan-funded-websites-nearly-outnumber-daily-newspapers-in-us/. 3.1.2

[9] TowCenter/partisan-local-news, . URL https://github.com/TowCenter/partisan-local-news. original-date: 2022-10-29T23:23:00Z. 2.1.1, 2.1.2, 3.1.2

[10] Twitter data for academic research, . URL https://developer.twitter.com/en/use-cases/do-research/academic-research. 2.1.2, 2.1.3, 2.1.3, 2.1.3

[11] unreliable-news/data at master · hearvox/unreliable-news, . URL https://github.com/hearvox/unreliable-news. 2.1.1

[12] Who owns the media?, . URL https://www.freepress.net/issues/media-control/media-consolidation/who-owns-media. 2.1.1, 3.3.2

[13] XGBoost documentation — xgboost 1.7.4 documentation, . URL https://xgboost.readthedocs.io/en/stable/. 3.4.2

[14] Alexandre Alaphilippe, Gary Machado, Roman Adamczyk, and Antoine Grégoire. Uncovered: 265 coordinated fake local media outlets serving indian interests. URL https://www.disinfo.eu/publications/uncovered-265-coordinated-fake-local-media-outlets-serving-indian-interests/. 3.1.2, 3.6.2

[15] Davey Alba and Jack Nicas. As local news dies, a pay-for-play network rises in its place - the new york times. URL https://www.nytimes.com/2020/10/18/technology/timpone-local-news-metric-media.html. 3.1.2

[16] Neal Altmanl and Kathleen M. Carley. Ora user's guide 2022, 2022. 3.2.2, 3.2.2

[17] K. Assmann. ise of the zombie papers: Infecting germany's local and regional public media ecosystem. 11, 2023. ISSN 3. doi: 10.17645/mac.v11i3.6816. 3.1.2, 3.6.2

[18] Jason Baumgartner, Savvas Zannettou, Brian Keegan, Megan Squire, and Jeremy Blackburn. The pushshift reddit dataset. In *Proceedings of the international AAAI conference on web and social media*, volume 14, pages 830–839. 2.1.2, 2.1.2, 2.1.2

[19] Priyanjana Bengani. Advocacy groups and metric media collaborate on local 'community news', . URL https://www.cjr.org/tow_center_reports/community-newsmaker-metric-media-local-news.php/. 3.1.2

[20] Priyanjana Bengani. Hundreds of 'pink slime' local news outlets are distributing algorithmic stories and conservative talking points, . 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. 3.1.2, 3.1.2

[21] PRIYANJANA BENGANI. The metric media network runs more than 1,200 local news sites. here are some of the non-profits funding them. URL https://www.cjr.org/tow_center_reports/metric-media-lobbyists-funding.php/. 3.1.2

[22] Priyanjana Bengani. 'pink slime' network gets $1.6m election boost from PACs backed by oil-and-gas, shipping magnates. 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. 3.1.2

[23] David M Beskow and Kathleen M Carley. Bot-hunter: a tiered approach to detecting & characterizing automated activity on twitter. In *Conference paper. SBP-BRiMS: International conference on social computing, behavioral-cultural modeling and prediction and behavior representation in modeling and simulation*, volume 3, page 3. 2.2

[24] Kathleen M. Beskow, David M.; Carley. Social Cybersecurity An Emerging National Security Requirement — armyupress.army.mil. https://www.armyupress.army.mil/Journals/Military-Review/English-Edition-Archives/Mar-Apr-2019/117-Cybersecurity/b/. [Accessed 01-12-2023]. 3.13

[25] Janice T. Blane. Social-cyber maneuvers for analyzing online influence operations. 3.13

[26] Janice T. Blane, Daniele Bellutta, and Kathleen M. Carley. Social-cyber maneuvers during the COVID-19 vaccine initial rollout: Content analysis of tweets. 24(3):e34040. doi: 10.2196/34040. URL https://www.jmir.org/2022/3/e34040. Company: Journal of Medical Internet Research Distributor: Journal of Medical Internet Research Institution: Journal of Medical Internet Research Label: Journal of Medical Internet Research Publisher: JMIR Publications Inc., Toronto, Canada. 3.3.2

[27] Jack Brandy and Nicholas Diakopouloshttps. Facebook's news feed algorithm and the 2020 us election. 9, 2023. ISSN 3. doi: 10.1177/20563051231196898. URL https://journals.sagepub.com/doi/epub/10.1177/20563051231196898. 1.1, 3.1.2

[28] Anthony G. Burton and Dimitri Koehorst. Research note: The spread of political misinformation on online subcultural platforms. doi: 10.37016/mr-2020-40. URL https://misinforeview.hks.harvard.edu/article/research-note-the-spread-of-political-misinformation-on-online-subcultural-platforms/. 3.1.2

[29] KM Carley. BEND: a framework for social cybersecurity. 6(2):22–27. 3.3.2

[30] K.M. Carley. Ora: A toolkit for dynamic network analysis and visualization. 2017. doi: 10.1007/978-1-4614-7163-9_309-1. 3.2.2, 3.2.2

[31] 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;. 2.2, 2.2

[32] 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. URL https://github.com/ercexpo/us-news-domains. original-date: 2021-11-04T19:58:16Z. 2.1.1

[33] Nicole S. Cohen. From pink slips to pink slime: Transforming media labor in a digital age. 18(2):98–122. ISSN 1071-4421. doi: 10.1080/10714421.2015.1031996. URL https://doi.org/10.1080/10714421.2015.1031996. Publisher: Routledge _eprint: https://doi.org/10.1080/10714421.2015.1031996. 3.1.2, 3.1.2

[34] NYU Web Communications. Russian trolls relied on local news more than fake news in 2016 presidential election, new analysis finds. 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-. 3.1.2

[35] Ng L.H.X. Carley K.M. Danaditya, A. From curious hashtags to polarized effect: profiling coordinated actions in indonesian twitter discourse. 12, 2022. ISSN 105. doi: 10.1007/s13278-022-00936-2. 3.3.2

[36] JEFFREY GOTTFRIED and JACOB LIEDKE. Partisan divides in media trust widen, driven by a decline among republicans. URL https://www.pewresearch.org/fact-tank/2021/08/30/partisan-divides-in-media-trust-widen-driven-by-a-decline-among-republicans/. 1.1, 3.1.2

[37] John Gramlich. 10 facts about americans and facebook. URL https://www.pewresearch.org/fact-tank/2021/06/01/facts-about-americans-and-facebook/. 3.1.2

[38] Christopher Groskopf. Who's ahead in competitive districts: 50 most competitive races, 2022. URL https://projects.fivethirtyeight.com/2022-election-forecast/house/. 2.1.2

[39] Andrew M. Guess, Brendan Nyhan, and Jason 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. 3.1.2

[40] David W Hosmer and S Lemeshow. Area under the ROC curve. 2:160–164. 3.4.2

[41] TOW CENTER FOR DIGITAL JOURNALISM. Domains as of august 3, 2020. URL https://datawrapper.dwcdn.net/TqILa/2/. 3.3.2

[42] Catherine King, Christine Sowa Lepird, and Dr Kathleen M Carley. Project OMEN: Designing a training game to fight misinformation on social media. 3.5.2

[43] Jon M Kleinberg. Authoritative sources in a hyperlinked environment. 3.4.2

[44] Justin Littman, Laura Wrubel, Daniel Kerchner, and Yonah Bromberg Gaber. News outlet tweet ids. URL https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/2FIFLH. 2.1.1

[45] Megan Lynch. 2022 midterms - pink slime, misleading ads, and more. URL https://www.audacy.com/kmox/news/local/2022-midterms-pink-slime-misleading-ads-and-more. 3.6.2

[46] Meta. Updates on our security work in ukraine, 2022. URL about.fb.com/news/2022/02/security-updates-ukraine/. 3.1.2, 3.6.2

[47] Ryan Moore, Ross Dahlke, Priyanjana Bengani, and Jeffrey Hancock. The consumption of pink slime journalism: Who, what, when, where, and why? Publisher: OSF Preprints. 1.1, 3.1.2, 3.1.2, 3.3.2

[48] H. Murphy and S. Venkataramakrishnan. Local news is drowning in 'pink slime' ahead of US election. 3.1.2

[49] Carley K. M. Ng, L. H. X. 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. 3.3.2

[50] Lynnette Hui Xian Ng and Kathleen M Carley. Botbuster: Multi-platform bot detection using a mixture of experts. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, pages 686–697, 2023. 3.2.2

[51] Lynnette Hui Xian Ng, Dawn C Robertson, and Kathleen M Carley. Stabilizing a supervised bot detection algorithm: How much data is needed for consistent predictions? 28:100198. Publisher: Elsevier. 2.2

[52] Katherine Ognyanova, David Lazer, Ronald E. Robertson, and Christo Wilson. Misinfor-

mation 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*, Jun 2020. doi: 10.37016/mr-2020-024. 3.5.2

[53] Sara Rafsky. Tow center audience study: Reader perspectives on partisan local news sites. URL https://www.cjr.org/tow_center_reports/tow-center-audience-study-reader-perspectives-on-local-partisan-news-sites.php/. 3.1.2, 3.5.1

[54] Asa Royal and Philip M. Napoli. Local journalism without journalists? metric media and the future of local news. 8:119–147. 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/. 3.1.2

[55] 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/. 3.1.2