Empirical Evaluation of Social Network Sensors on Twitter During the Russia-Ukraine Conflict

by

Miranda Nicolle Ahlers

B.S. Operations Research
United States Air Force Academy, 2021

Submitted to the Institute for Data, Systems, and Society in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN TECHNOLOGY AND POLICY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2023

Authored by: Miranda Ahlers
Institute for Data, Systems, and Society
May 12, 2023

Certified by: Dean Eckles
Associate Professor of Marketing, Sloan
Thesis Supervisor

Accepted by: Noelle Eckley Selin
Professor, Institute for Data, Systems, and Society and
Department of Earth, Atmospheric and Planetary Sciences
Director, Technology and Policy Program

Empirical Evaluation of Social Network Sensors on Twitter During the Russia-Ukraine Conflict

by

Miranda Ahlers

Submitted to the Institute for Data, Systems, and Society
on May 12, 2023, in partial fulfillment of the
requirements for the degree of
Master of Science in Technology and Policy

Abstract

The immense magnitude of information sharing, paired with increased privacy considerations, has rendered global monitoring of social media platforms virtually infeasible. Heuristic algorithms grounded in the friendship paradox have provided simple, accessible methods for strategic sampling of information from platforms while only requiring knowledge of the local network structure. However, it still remains unclear how well such algorithms perform in contexts where the spread of information consists of exogenous and endogenous modes of propagation.

Herein, I evaluate the ability of randomly selected friends of random users to provide early awareness of discussions related to the Russia-Ukraine conflict on Twitter. I find that while selected sensors are more centrally located within the Twitter network, they fail to reliably provide early awareness of conflict-related hashtags. Lack of performance is exacerbated when only early adopters from each group are included in evaluations. Additionally, I find that the difference in time of adoption between control and sensor groups provides limited information about how popular a hashtag will become. Further, I propose a framework for using early participation in conflict discourse to condition the selection of sensors for future war-related trends – exploring both friendship and prior retweet connections as potential sensors. I then outline two systematic approaches for objectively quantifying the value of information acquired from selected sensor groups – a count-based approach and a predictive modeling framework. Ultimately, I find that both local and retweet sensors significantly reduce the noise of information produced by a random control group while effectively capturing over 80% of hashtags that become widely shared.

Thesis Supervisor: Dean Eckles
Title: Associate Professor of Marketing, Sloan
Mitsubishi Career Development Professor

Acknowledgements

First and foremost, I would like to thank my advisor, Dr. Dean Eckles. Dean, thank you for all of the guidance, insights, and patience that you offered throughout this process. It was an absolute privilege to get the chance to learn from you. I also would like to thank Lincoln Laboratory for this opportunity. Ed, thank you for welcoming me into Group 41.

Mom and Dad, I would be absolutely nowhere if it weren't for your constant support, love, and guidance. Thank you for everything. Ben and Kinzie, I couldn't have asked for better people to look up to!

Sydney, thank you for taking a leap of faith and moving to Boston with me – getting the chance to live with my best friend was truly the best part of the last two years.

And finally, to Jeff – I'm not sure what I would have done without your love and support over the last four years. Thank you for being the best sounding board and for showing me endless patience as I tried to navigate grad school. I cannot wait to start our life together.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

With nearly 4.7 billion active users worldwide – a global penetration rate of 59% – social media platforms have grown to dominate information sharing across the world (*Internet and Social Media Users in the World 2023*, 2023). As these platforms begin to supersede traditional forms of media, there is an increased interest in understanding the nature of information sharing on the platforms as well as associated consequences on physical occurrences.

It has been shown that activity on social media can serve as a predictor for offline events such as disease activity, box office revenues, and changes in stock prices (Asur & Huberman, 2013; Babichev & Lytvynenko, 2022; Signorini et al., 2011). Online discussions have also been linked to changes in offline behavior and perception of individuals, including increasing protest participation and heavily influencing public valuation of markets (Enikolopov et al., 2015; May et al., 2008). In recent humanitarian crises, such as the large-scale earthquake in Turkey, online activity was used to supplement relief and recovery efforts. In recent global conflicts, such as the ongoing Russo-Ukrainian War, social media has proven to be a "key instrument in reflecting the experience of war in the civilian population" (Zasiekin et al., 2022).

Diverse use-cases for online activity are continuously unveiled. This means finding mechanisms for efficient detection of information throughout OSNs is now more pivotal than ever. There have been several heuristic algorithms presented for strategically selecting nodes to be used for monitoring activity in social networks. However, there remain few empirical analyses of these algorithms to reinforce effectiveness and to

understand generalizability across diverse contexts and scenarios. Such proof of performance is key if approaches are to be responsibly implemented for policy-relevant applications – such as search and rescue, gauging public sentiment, etc. This work contributes to the task of understanding and evaluating mechanisms for sensor identification in social networks – specifically as they perform in conversations related to geo-political conflicts on Twitter. The empirical analyses conducted focus on discussions surrounding the Russia-Ukraine conflict.

In Chapter 2, I provide a brief summary of the on-going Russia-Ukraine conflict and discuss the various roles that social media has played throughout. I show that as the war on the ground has developed over the last year, online platforms have been consequential for Ukrainian resilience and fundamental in constructing the public's perception of the ongoing conflict. In Chapter 3, I discuss relevant literature in the field of sensing, and I also reinforce the value derived from identifying and implementing sensors in an applied setting.

In the fourth chapter, I outline the primary method for sensor identification that is used throughout the analysis and discuss principles underlying the mechanism. I then use two approaches to analyze the efficacy of the sensing mechanism when applied to discussions related to the Russia-Ukraine conflict. First, I look within specific hashtag networks to understand if randomly selected friends could have provided early awareness of hashtag sharing. Second, I use small samples to evaluate the sensing mechanism for a wider array of hashtags. Overall, I find that random one-hop connections were unable to reliably provide early awareness of conflict-related hashtags. Efficacy of sensors further declined when only the earliest adopters from either group are considered for lead time evaluations. Additionally, the success of sensors provides little information about how widespread a hashtag will become on Twitter.

In Chapter 5, I propose a framework for using early participation in conflict-related hashtags as an instrument for selecting sensors for future war-related trends. Both random friends and random prior retweet connections are explored as potential sensors. I find that retweet sensors are more centrally located in the network than a random control group and are also generally more active on Twitter than a group of local sensors. Finally, I propose two methods for objectively assessing the value provided by sensors –

a count-based approach and a predictive modeling framework. I find that local and retweet sensors successfully detect over 80% of relevant, widespread hashtags while sharing 36% fewer tweets than the control group. Additionally, I show that relative improvements in predictive performance can be used to directly quantify the gain from identifying sensors.

I conclude in Chapter 6 with a discussion of limitations and biases of the approaches taken and areas for future work.

Chapter 2

Empirical Context

2.1 Russia-Ukraine Relations

On Dec 1st, 1991, following the fall of the Soviet Union, Ukraine held a public referendum to confirm the "Act of Declaration of Independence of Ukraine" – formalizing Ukraine as an independent state (Futey, 1996). In efforts to protect their newfound independence, Ukraine forfeited warheads, missiles and other nuclear capabilities in exchange for guarantees that Russia, the US, and the UK would "respect the independence and sovereignty and the existing borders of Ukraine" (Sullivan, 2022; Treaty Series 3007, 2021).

Ukraine spent the following years continuing to forge its path as an independent entity while handling Russian interference, economic turmoil, and political corruption scandals. Parties seeking closer relations to NATO, the EU, and the West remained in constant tension with those in favor of tighter affiliation to Russia (*A Historical Timeline of Post-Independence Ukraine*, 2022). When Russian backed president Viktor Yanukovych announced that Ukraine would forgo signing an association agreement with the EU in late 2013, he fled to Russia, and protests erupted (Biersack & O'Lear, 2014). In response to growing support for the western favored government that replaced Yanukovych, Russian forces entered and illegally annexed the eastern peninsula of Crimea (*Timeline: Political Crisis in Ukraine and Russia's Occupation of Crimea | Reuters*, 2014). After Crimea, focus of pro-Russian separatists shifted to the Donbass region where violence continued (*Ukraine - The Crisis in Crimea and Eastern Ukraine | Britannica*, n.d.).

Nearly 14,000 lives had been lost due to fighting in Donbass by 2020 (Pifer, 2020). In November of 2021 the conflict began to escalate further as Russian troops and military equipment amassed at Ukrainian borders for the second time in eight months (Sullivan, 2022). Disregarding pleas from NATO and the West, on February 24, 2022, Putin publicly commenced a "special military operation" in Ukraine – marking the beginning of a full-scale invasion that continues on as this thesis is being written (Gill, 2022). Figure 2-1 shows snapshots of territorial control over the course of the conflict. Ukraine has far surpassed global expectations with their continued resistance against Russian forces. As of late 2022, Ukraine began to regain pieces of territory previously acquired by Russia.

Figure 2-1: Military Control of Ukraine Over Time. Image formatted by BBC from Institute for the Study of War (War, 2023)

Over the past year, the struggle for territory has not only resulted in a humanitarian crisis for those that remain in the country, but has also caused a large-scale refugee crisis, an energy crisis, and a global food shortage (Behnassi & El Haiba, 2022; Jaroszewicz et al., 2022). These dire consequences have drawn in the attention and resources of civilians and governments from all corners of the globe – many seeking to support Ukrainian efforts while searching for ways to de-escalate the conflict.

2.2 Role of Social Media in the Russo-Ukrainian War

Upon, and even before, Russia entered Ukraine, it became overtly evident that the online dimension of the conflict was going to play a pivotal role in the dynamics of the war. Fights for territory, hearts, and minds transcended the bomb-ridden battlefields and quickly made their way behind screens – with social media sitting at the forefront of these digital conversations. The extensive use and implications of social media platforms have earned the conflict the title of the "first social media war" (Ciuriak, 2022).

Leaders of both Russia and Ukraine turned to online platforms to share information about events unfolding throughout the conflict – each shaping accounts to fulfill the narratives they wished to portray (Ghasiya & Sasahara, 2022). Russian affiliated entities flooded social media streams with justification for their actions accompanied by denial and deflections of the atrocities taking place, such as the Bucha Massacre (Whalen & Dixon, 2022). On the other hand, Ukraine parties focused on using platforms to undermine the perceived success of the Russian military, to maintain patriotism throughout the population, and to garner support from western countries (Smart et al., 2022). Both Ukrainian and Russian activity online has reinforced the speed and scale with which information can be spread across online social networks – with messages from both parties reaching countries all over the globe (at least those whose internet use is not under sovereign control). Additionally, the activity has demonstrated the ability of social media platforms to be used as a forum for swaying public opinion. Capitalizing on the power of social media in the ongoing information war has proven critical for Ukraine's unanticipated resilience – aiding in recruitment of volunteer fighters and eliciting support from foreign entities.

Beyond use for official communication streams, individuals directly impacted by the conflict have turned to social media to share personal testimonies – ranging anywhere from written accounts of encounters with Russian soldiers to live-stream videos of ongoing attacks (Zasiekin et al., 2022). Refugees that have been displaced by the violence have used platforms to seek asylum in foreign countries and users within the Ukrainian population, as well as around the world, have extended a helping hand online –

establishing connections and coordinating refuge for those in need (Talabi et al., 2022). In addition to aiding the refugee crisis, others have taken to social media to share their opinions of the invasion and judgements of subsequent steps taken by government leaders in response to Russia's behavior. Overall, the ability for easy, informal global information exchange has amplified the narratives that are able to be told by those on both sides of the war. This has made platforms such as Telegram and Twitter a primary location for gauging public sentiment and activity during conflict.

While democratization of information sharing has changed the landscape of coverage during this war, it has also opened channels for disinformation and misinformation narratives to thrive. Russian information operations are not a novelty in conflict (Golovchenko, 2020). Even before the time of extensive social media use, Russian actors have exploited the power of mainstream media to manipulate public opinion and conceal intentions – the current conflict with Ukraine has been no exception. Russian propaganda justifying a "special military operation" has been pushed to the public since before the official invasion on February 24th, 2022. Continued narratives throughout the conflict have promoted hostility against the West and have diminished support for Ukraine. In this conflict (and in general), the uncontested scalability of disinformation narratives afforded by social media has fueled concerns about platforms being used "to increase political division and influence public opinion as a tool of modern warfare" (Geissler et al., 2023).

In addition to disinformation, unintentional spreading of false information has also been a point of concern. For example, videos from previous conflicts have been recycled and presented as "live footage" of the ongoing fighting and deep fakes have been created depicting Ukraine surrendering (Stănescu, 2022). Given the limited ability for individuals to gain first-hand evidence of events taking place, such instances of false information are easily propagated with their validity often going unquestioned. Companies including YouTube, Facebook, TikTok and Twitter have organized action groups dedicated to managing the spread of both disinformation and misinformation but recent work estimates that only 8-15% of the false content circulating about the conflict on Facebook and Twitter has actually been flagged or taken down (Note: these are estimates from a domain based analysis where reliability of information is determined by

website links contained in a post. A post is deemed “low-credibility” if the shared link is on the “Iffy Index of Unreliable Sources” and “high-credibility” if the link is considered reputable by the Media Bias/Fact Check website) (*Iffy.News*, n.d.; *Media Bias/Fact Check - Search and Learn the Bias of News Media*, n.d.; Pierri et al., 2023).

It is extremely difficult to pinpoint the exact relationship between online activity and real-world outcomes. However, it is abundantly clear that social media has become ingrained in many facets of this ongoing contest between Russia and Ukraine. In the interest of preserving the integrity of online conversations as well as proactively understanding prevailing beliefs among the public, effective engagement with social media platforms has become a necessity.

Chapter 3

Using Limited Network Structure to Identify Central Nodes in a Network

The accessibility of social media platforms has drawn in billions of users seeking to share news, personal stories, humorous anecdotes, and connect with others in separate geographic locations. Rapid influx of digital technologies and increases in global internet access have aided in the significant increase in membership and activity seen across OSNs.

As one of the longest standing social media platforms, Twitter is no stranger to this exponential increase in active users – with statistics from 2023 showing a near 1500% increase from the 30 million users that were active monthly in 2010 (*Twitter MAU Worldwide 2019*, n.d.). Recent work by Pfeffer et al. coordinated a large-scale collection of every tweet published on Twitter in the 24-hour period of September 21, 2022 (Pfeffer et al., 2023). The data collected contained 375 million tweets – equating to 15.6 million tweets being shared every hour, 260,000 tweets shared every minute, and 4,400 tweets shared every second. Other estimates report rates nearing 500 million tweets per day – 5,000 tweets per second – stemming from one of 450 million monthly active users (*22 Essential Twitter Statistics You Need to Know in 2023*, n.d.). We continue to increase our understanding of social media's ability to sway public perception in times of crises – but how do we go about finding a way to detect the activity that is driving these occurrences amongst the enormous volume of information being shared?

An apparent solution to tackle the immensity is to uniformly sample users in the network to serve as points of observation. Intuition tells us however, that such an

approach leaves much on the table regarding information about how ideas move throughout a network. Work over the last decade has focused on using only information about the local structure of networks to identify key nodes for strategically evaluating activity.

3.1 Related Work

Strategically targeting individuals in a network is not a novel problem. Across many disciplines such as "marketing, public health, [and] development", decision-makers are tasked with effectively carrying out interventions under strict budget and/or supply constraints (Eckles et al., 2022). To combat these limitations, policymakers and businesses seek to find "key informants to diffuse new information" to their surrounding communities (Banerjee et al., 2019). A wide-array of literature has explored optimal ways to select individuals expected to maximize total adoption in a network – often referred to as "influence maximization" (Hinz et al., 2011; Kempe et al., 2003).

Work in seeding has bred a contrasting but related field of research – "sensing" – that is geared towards problems such as outbreak detection and immunization (Chami et al., 2017; Christakis & Fowler, 2010). Over the last two decades, several methods have been proposed to select sensors in social networks. For example, Bagavathi & Krishnan (2019) implement a ranking scheme for users on Twitter based on "participation frequency" and "mean adoption time", where nodes with the highest rank are selected to be part of a sensor set. Xie et al. formulate sensor identification as a linear programming problem and use the subgradient method to identify individuals that efficiently detect cascades with "bursty" behavior (Xie et al., 2018). Batlle et al. present a framework for early detection of epidemic outbreaks, using "submodularity optimization techniques" to find an optimal subset of nodes to administer viral tests. (Batlle et al., 2020).

While mentioned methods provide theoretical performance guarantees, their complexity reduces general accessibility, and they overlook the value afforded by exploiting connections of the underlying network. There is an expansive collection of both theoretical and empirical literature demonstrating the "prevalence of peer effects in

adoption processes" – an indication that considering peer relationships would be beneficial for targeting strategies (Chin et al., 2022). However, given the extensive size of many social networks, obtaining complete information about the structure of a network can be costly (Eckles et al., 2022). One-hop targeting strategies grounded in the "Friendship Paradox" have provided an avenue for incorporating network information without needing a comprehensive picture of a network. The paradox chiefly states that "your friends have more friends than you do" (Feld, 1991).

Theoretical evaluations, such as that by Nettasinghe & Krishnamurthy, have exploited the friendship paradox to reach the tails of a degree-distribution – demonstrating that randomly sampling a single friend from a random pair of friends allows for efficient sampling of high-degree nodes in a network (Nettasinghe & Krishnamurthy, 2021). Initial proposals of friendship based targeting came from Cohen et al., who demonstrated that immunizing "random acquaintances" of random nodes could prevent epidemics with "a small finite immunization threshold" (Cohen et al., 2003). This pioneering application was shown to work for "any broad-degree distribution". Recent empirical studies have sought to build upon work by Cohen et al. and evaluate how useful such methods are in discipline-specific contexts.

Christakis and Fowler proposed monitoring the friends of randomly selected individuals as a method for detecting outbreaks of contagious epidemics such as the flu. Studying students at Harvard college, they found that the progression of the flu in a friend group was nearly two weeks earlier than that of the randomly chosen group of students (Christakis & Fowler, 2010). On a larger scale, Sun et al. used smart-card-based bus fare data from Singapore to simulate contagious outbreaks in "massive metropolitan encounter networks" (Sun et al., 2014). They show that a sensor group – created by sampling friends of random patrons – can provide detection of an outbreak up to 12 hours ahead of the random control group.

Garcia-Herranz et al. shifted prior work in sensing to the domain of online social networks – using followees of randomly selected Twitter users to construct groups of sensors. In evaluating hashtag use over 6 months of Twitter activity, they found that for widely used hashtags spreading endogenously through Twitter's network, sensor groups

provided early awareness (relative to a control group) of up to 7 days (Garcia-Herranz et al., 2014).

While high centrality nodes prove to be beneficial for early awareness in instances of viral contagion, information sharing in online social networks is not always contained to endogenous means. “External out-of-network sources” often play a role in what users decide to share on a platform (Myers et al., 2012). This component of online activity is likely to be exacerbated in circumstances where there is an influx of media coverage on a specific topic – i.e., geopolitical conflicts such as the ongoing war in Ukraine. There is limited literature evaluating the performance of the described heuristic algorithms for identifying sensors in scenarios where both endogenous and exogenous modes of propagation are present.

One effort comes from work by Kryvasheyeu et al., where they provide an empirical evaluation of sensors on Twitter during Hurricane Sandy – a relevant example of a complex environment of information sharing, where information about the disaster was “carried simultaneously by many other external channels” (Kryvasheyeu et al., 2015). Their work shows that using friends as sensors could have provided up to 25 hours advance warning of disaster-related discussions on Twitter and that lead time was maximized when users in the sensor group were physically located in the path of the storm while those in the control group were not. There currently exists a hole in the literature for proof of sensor performance in other contexts with mixed forms of information sharing – such as geo-political conflicts. The study of the Russia-Ukraine conflict on Twitter conducted in the following chapters seeks to add to the small array of empirical evaluations and further the understanding of how heuristic algorithms for sensor identification perform in diverse contexts.

Elaboration of the Friendship Paradox

In his 1991 paper, Scott Feld introduced the “Friendship Paradox” stating that “on average, the number of friends of a random friend is always greater than or equal to the number of friends of a random individual” (Feld, 1991)**.** The notion generalizes for “any arbitrary social network (with variance in the degree distribution)” (Garcia-Herranz et al., 2014)**.**

As an online social network, Twitter can be described as a directed network graph $G = (V, E)$, where nodes V are represented by the set of users on the platform and an edge $(i, j) \in E$ exists (connecting node i and node j) if user i follows user j. The friendship paradox can be restated in the context of this graph as follows:

For a uniformly chosen node $X \in V$ and a node Y from a uniformly chosen edge $(X, Y) \in E$:

$$E\{k_Y\} \geq E\{k_X\}$$

Where k_X denotes the out-degree of node X.

This formulation states that on average, the degree of node X will be less than the degree of its randomly chosen friend, node Y. The basic intuition behind the paradox is laid out succinctly by Kumar et al. in saying "there are few well-connected hubs in real networks, and since they are connected to many other nodes (by definition), obtaining a friend (or neighbor) of a random node is likely to result in a hub with greater likelihood, compared to the case of randomly selected nodes" (Kumar et al., 2021).

3.2 Value of Sensors

The virtual infeasibility of global monitoring renders sensing mechanisms necessary. And, while much of the literature seems to agree that using the local structure of a network to identify sensors is a functional approach, it is important to reiterate the value derived from implementing such techniques in policy relevant contexts. Why do sensors have the potential to be invaluable to decision-makers? By heuristically identifying highly connected nodes in social networks, we are able to mitigate four main concerns with global monitoring in the current landscape of social media:

Access to Information: Just because information is being publicly shared across a platform doesn't necessarily mean there is a feasible way for it to be collected for analysis on a large-scale. Rate limits on the Twitter API are a prime example of this

limitation. Under the version of the Twitter API available during data collection for this thesis, a maximum of 1% of all tweets could be collected in real-time using the Streaming API. If interested in evaluating activity that had already taken place, only 10 million historic tweets could be scraped per month by a single account (even with a high-level Academic Researcher API). This amounts to only 0.08% of total tweets published over the course of a month.

In early February 2023, Twitter announced that significant changes were being made to overall API access levels. Instead of offering free access to developers (with short applications for elevated access) the platform will be charging fees for even the most basic levels of developer use-cases. A basic tier will cost $100 monthly and will only provide access to 10,000 tweets per month (*Twitter API Documentation | Docs | Twitter Developer Platform*, n.d.). Elevated enterprise access is available but will likely come with a hefty monthly fee – with estimates ranging from $400 to multiple thousand dollars. These recent modifications to the Twitter API profoundly alter the landscape for understanding and utilizing Twitter activity for policy relevant applications – particularly for organizations that already face financial barriers. Looking beyond Twitter, some platforms such as Facebook do not even offer free streaming APIs. These drastic changes at Twitter and the limited data access for other online platforms amplify the necessity of finding a way to efficiently sample the information being shared online to minimize the magnitude of data that needs to be collected.

Processing Resources: While advanced computing resources are becoming increasingly accessible, there are still many entities that do not possess the capabilities necessary to quickly sort through the extremely large volumes of data being posted to social media platforms. If institutions wish to extract meaningful information of interest from the overwhelming sea of posts, they must find a way to strategically diminish the information that must be sorted through. This is directly intertwined with temporal considerations mentioned below. Even with extensive resources, the smaller the amount of information collected the faster a human or machine will be able to draw conclusions.

Privacy: Influx of interest into the activity of OSNs has brought forth many justified concerns about privacy of users. Data from social media "can provide insights into the lived experience … through unparalleled access to the everyday lives of individuals" (Nicholas et al., 2020). This means that collection and storage of large amounts of data from platforms such as Twitter, even with promises of anonymity, threatens to compromise user privacy – with risks such as "accidental and purposeful misidentification" and "unauthorized secondary use" of aggregated private information (Di Minin et al., 2021; Osatuyi, 2015).

Additionally, there still is not a clearly defined and globally understood line of consent for collection of social media data. Nicholas et. al discuss that while some users argue that "informed consent was implied due to the public nature of posts", others are of the opinion that the structure of the platforms provides them with the expectation that only family, friends and "friends of friends" will see and consume posts – so informed consent is needed for any further collection (Nicholas et al., 2020). Even if global monitoring were to be feasible, these considerations for user privacy provide significant motivation for minimizing the quantity of data and metadata collected. Identifying sensors helps to satisfy this minimization while still providing value to the populace interested in activity on online platforms.

Time: Whereas information spreading by word-of-mouth through a network of humans is held up by moments of physical interaction, virtual sharing can occur almost instantaneously, at all times of the day, regardless of physical location. This significantly expedites information sharing processes, which subsequently confines the window of time across which topics can be identified before they reach a wide audience.

In scenarios such as the refugee crises in the Russia-Ukraine War, where platforms are being utilized to supplement search and rescue efforts for humanitarian crises, natural disasters, etc. – time is undoubtedly important. Early detection of information is vital for maximizing the success of missions. However, in many other contexts – including disinformation and misinformation identification, marketing, public policy — gauging online activity is still a time sensitive matter. Many users online fashion their realities and preferences through the posts that they are exposed to. This

ability to “construct individual and collective memories” is where social media derives a large part of its power to impose a “social political impact” (Zasiekin et al., 2022). Waiting until topics have completely run their course through a social network maximizes the potential power of platforms and subsequently leaves vested parties with the task of correcting information or updating beliefs after-the-fact.

For institutions concerned with what individuals think, feel, and perceive, early awareness of online activity is irreplaceable. Finding a set of users that (on the aggregate) reliably share or interact with information ahead of the general adoption curve or a set of users whose interaction is indicative of the future sharing trajectory of information is vital for proactively managing the impact social media platforms are able to have.

Chapter 4

Empirical Evaluation: Early Awareness

The field of sensing lacks empirical evaluations to reinforce effectiveness – particularly in contexts concerning international conflict. Literature shows that by exploiting principles of the friendship paradox, random friends of random users should result in a group that is more connected on average than the random users themselves – helping to identify key nodes for understanding activity and providing early awareness online. However, with information sharing during the Russia-Ukraine War extending far beyond the boundaries of Twitter, can we still expect sensors to be more informative than their randomly selected counterparts?

The complex relationship between exogenous sharing – information shared by media outlets, on other platforms, or by offline connections – and endogenous sharing – information spreading via contagion through the Twitter network – makes it unclear if previously proposed sensing techniques would be reliable for war-related trends. For example, posit that CNN televises a broadcast during which they share details on the advancement of Russian Forces. An individual who has just watched this broadcast may take to Twitter to relay the information they just consumed. In such a scenario, engagement on the platform "cannot be attributed to network effects" and is unrelated to the position of the user in Twitter's network (Myers et al., 2012). Inversely, if a user sees a friend's post about the war and decides to reshare or contribute original commentary, network position becomes a potential factor in the adoption of the content.

This chapter seeks to contribute an evaluation of the performance of previously proposed heuristic algorithms for identifying sensors in discussions related to the

Russia-Ukraine conflict on Twitter. I will specifically aim to answer the following questions:

- How useful would sensors have been for providing early awareness of hashtags related to the conflict on Twitter?

- Is the method for evaluating lead time that has been proposed robust to late adopters?

- Does the efficacy of the sensing mechanism provide any indication of the future popularity of hashtags on Twitter?

4.1 Data Set

Tweets

The primary dataset I use for the following analyses was collected and made publicly available by Shevtsov et al. in their preliminary work titled "Twitter Dataset on the Russo-Ukrainian War". Beginning February 24, 2022 – the day Russia invaded the borders of Ukraine – Twitter's streaming API was used to query for tweets containing one of a set of pre-selected hashtags related to the Russo-Ukrainian War. These are complemented by tweets from February 22nd and 23rd that were retrospectively collected using the Search API on Twitter. Information collected also includes metadata pertaining to each tweet such as the Tweet ID, user ID, date and time of publishing, user friend and follower counts, and tweet favorite counts. Further details regarding the collection process of this data can be found in Shevtsov et al. (2022) and a list of the hashtags used to guide collection can be found in Appendix A (Shevtsov et al., 2022).

Due to restrictions enforced by Twitter for distributing content obtained via APIs to third party users, only TweetIDs were made available (*Use Cases, Tutorials, & Documentation*, n.d.). These IDs were "rehydrated" using Twitter APIs, a process through which only tweets that are still publicly available may be retrieved. Reasons for

unavailability include deletion by the user, deletion by Twitter due to inappropriate or harmful content, tweets made private by the user, or temporary disabling of an account. Approximately 75% of the TweetIDs contained in the data set were rehydrated from the platform. The final data set contains ~25 million tweets from February 22, 2022 to March 10, 2022 from ~4.8 million unique users.

Tweet Counts

In addition to tweets provided in the data set, aspects of the following analysis require understanding what overall participation in conversations on Twitter looked like over time. To capture holistic hashtag use, the Historic Tweet Count API on Twitter was used to collect daily hashtag counts from January 1, 2022 to August 31, 2022. Queries for hashtags using the Tweet Count API are not case sensitive.

While the tweet count endpoint is not subject to the same level of restrictions as the Streaming API, the endpoint can still only return counts of tweets that currently exist on Twitter. Therefore, if tweets containing a specific hashtag were deleted by the user, taken down by Twitter, etc. the use will not be included in the total count returned.

As a preliminary check, daily counts of tweets containing given hashtags found in the data set were compared to daily counts retrieved by the Count API. Figure 4-1 shows that for a majority of the hashtags used for collection by Shevtsov et. al (2022), the incidence curves for the data set and Count API are extremely similar.

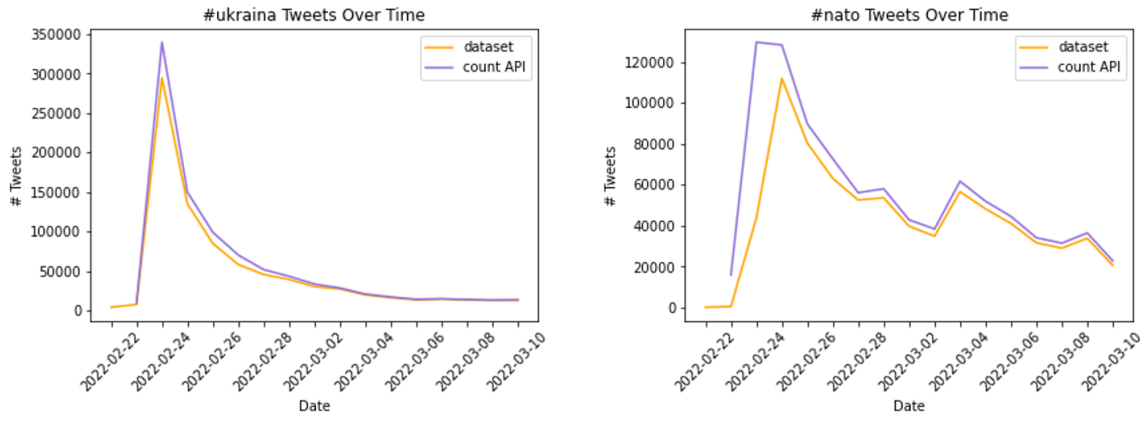


Figure 4-1: Hashtag Incidence Curves for Dataset and Count API

Slight differences can be observed for some hashtags. Where gaps exist between the curves, occurrences in the data set follow the general pattern of the ground truth count. For those hashtags for which the data set does not capture every occurrence, I make the assumption that tweets missing are missing at random – that there is no underlying bias created due to these missing observations. Implications of missing tweets and using hashtags to guide queries are discussed further in Chapter 6.

Friend Lists

Because information on friendship/follower connections was not scraped at the time of collection, this information was retrospectively scraped using the Friends List Twitter API. Retrieving friends lists from Twitter is computationally expensive and time-consuming, so therefore users from the data set were randomly sampled to have their information collected. Approximately 270,000 users in the data set had information on their friendship connections (the accounts that they follow) scraped. While collecting friends lists ex post facto introduces the potential for edges to be included in the analysis that did not exist at the time a tweet was published, this should not significantly impact results – the implications of this approach are further discussed in Chapter 6. Future efforts to assess sensing algorithms should strive to capture relationships between users at the time of collection to preserve true network structure to the maximum extent possible.

4.2 Sensing Framework

Below I outline the framework that will be used for selecting sensors. I also introduce the mechanism implemented in previous work that will be used to evaluate early awareness provided by a sensor set.

4.2.1 Selecting Sensors

For the first set of the evaluations in this thesis, a control group is constructed by randomly sampling users who had information collected on their friendship network. This control group can also be thought of as a "control sensor" group – where sensors are

chosen via the non-network signal of tweeting about the conflict. From this control group, two methods are used to select additional sensor sets. Following the approach taken by Kumar et. al (2021), I will refer to these sensors from this point forward as "local" and "global".

Local Sensors: For each user, *r*, in the control group, a node *s* is chosen with uniform probability from *r's* friends and added to the sensor set. After removing any duplicates from the set, we arrive at a sensor set of similar size to the original control group.

Global Sensors: For each user, *r*, in the control group, each of *r*'s friends is added to the sensor set with probability p, where $0 < p \leq 1$. Given the probabilistic nature of selecting global sensors, the size of the final group may be significantly larger than the control group. In this case, the sensor group is uniformly down sampled to equal the size of the control group.

Sampling in prior work using "friends as sensors" falls in line with the first method described, where the number of sensors selected is fixed (minus slight variations due to duplicate removal). Kumar et al. provide an in-depth appraisal of both approaches –

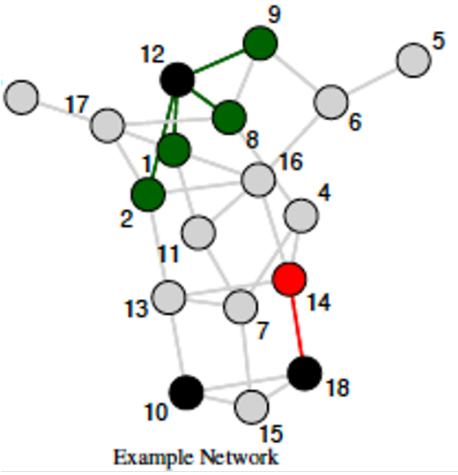


Figure 4-2: Example Network from Kumar et al. For control node 18, one of its friends, node 14, is chosen at random to be included in the local sensor group. For control node 12, each of its friends is probabilistically added to the global sensor set with probability *p* (Kumar et al., 2021).

showing that out of various different networks, Twitter has the highest ratio between the local and the global mean – that the centrality of locally sampled sensors is greater than those selected globally. For our purposes, the value of global sampling comes from the ability to easily compute the exact probability that some sensor set has been selected. This will not be explored in depth in this thesis; however, it is beneficial nonetheless to include global sensors in the analysis to understand how they perform relative to their local counterparts. Figure 4-2 provides an example network of nodes and illustrates connections between the randomly sampled control group (in black) and each of the sensor groups.

4.2.2 Sensor Methodology

To evaluate early awareness provided by the sensor groups, I focus on the first tweet within a hashtag for each user. Given we only observe tweets from February 22nd onward, I assume the first use of a hashtag by a user captured in this data set is their "entry" or "adoption" time. Using the framework proposed by Garcia-Herranz et al. and Kryvasheyeu et al., I define lead time as the difference in the average adoption time of the sensor group and the average adoption time of its counterpart control group (Garcia-Herranz et al., 2014; Kryvasheyeu et al., 2015). If t_i^h = the time at which sampled user *i* first mentions hashtag *h*, lead time is defined as the difference between the average adoption time of the sensor group and the average adoption time of the control group. It is formulated as follows:

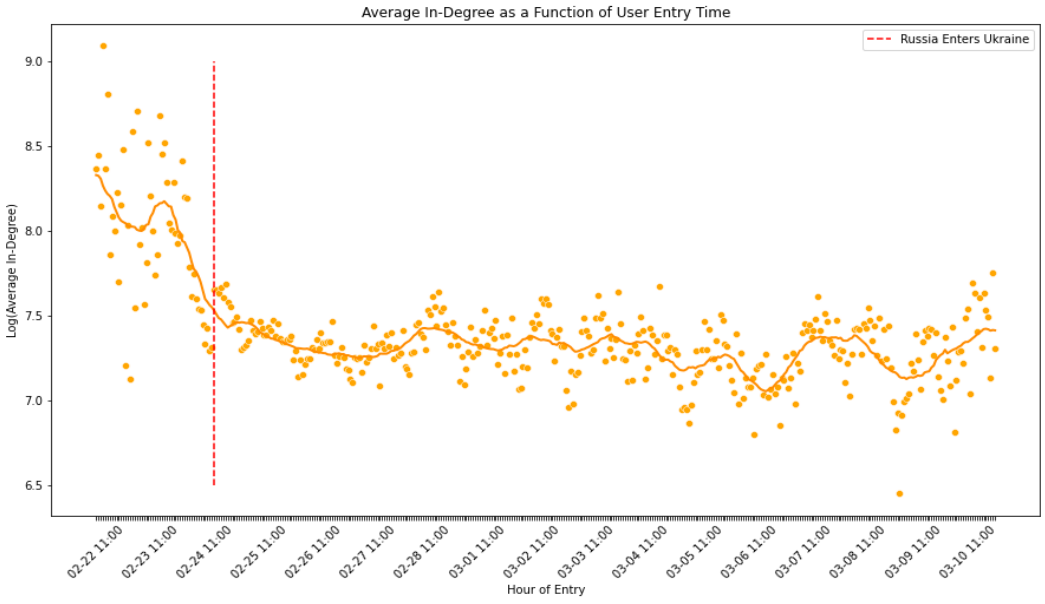
$$\Delta t^h = \langle t \rangle_{i \epsilon S} - \langle t \rangle_{i \epsilon C} \tag{4.1}$$

Where S represents the respective sensor set of users and C represents the control set of users. A negative Δt^h indicates early awareness provided by the sensor group.

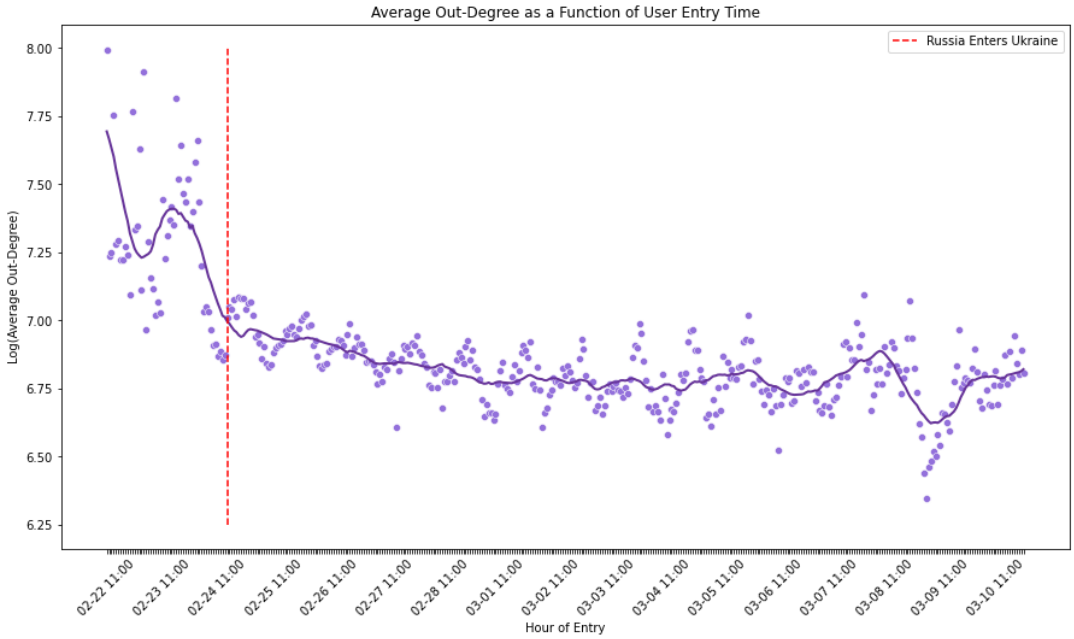
4.2.3 Hypothesis

For the proposed method to be successful, we should see some relationship between the topological (node degree) characteristic of users and their entry time – where those that are more centrally located in the network would generally participate earlier than those on the periphery (Kryvasheyeu et al., 2015).

Figure 4-3 clearly shows that on average, users with earlier entry times possess greater network centrality than those who engage with the hashtags later on – as characterized by their in-degree (number of followers) and out-degree (number of people they follow i.e., "friends" or followees). This increased level of centrality is particularly evident in the two days leading up to the invasion. In the days following, the disparities in network centrality become less accentuated. Average out-degree of users maintains a steady decrease as time continues while average in-degree levels out fairly quickly after the date of the initial invasion – indicating there is not much difference between the number of followers of users who engage shortly after the invasion and those who do not engage until a week or two later. Kryvasheyeu et al. noted similar behavior in



(a) Average In-Degree vs Hour of Entry



(b) Average Out-Degree vs Hour of Entry

Figure 4-3: Average degree of users as a function of entry time. Smooth line shows the line of best fit using second order polynomials and a window of size 40.

characteristics of users when evaluating participation in discussions related to Hurricane Sandy on Twitter (Kryvasheyeu et al., 2015).

While the aggregate network characteristics of users suggests the sensing mechanism should be successful, the role of exogenously transmitted information on the outcome is still unclear. In the analyses below, guided by work in previous literature, I use two approaches to evaluate the magnitude of the lead time provided by selected sensor groups. For the first approach, I look within specific hashtag networks, sampling control and sensor groups from only the population contained in isolated networks. For the second approach, I evaluate the sensor method using 'small' samples – sampling control groups from the entire population of users for which we have followee network information. I then discuss potential biases present amongst lead time metrics that have been implemented in the past and propose new metrics to evaluate sensors' ability to provide early awareness of conflict related discussions. Finally, I explore the relationship between lead time and a hashtag's subsequent level of popularity.

I find that for widespread hashtags used to condition collection, both local and global sensors successfully provide early awareness over 60% of the time. However, both sensors produce extremely large levels of variance for several hashtags, which brings into question their ability to provide consistent results.

For small samples, I find that the percentage of hashtags for which sensors provide early awareness is only slightly higher than the baseline 50%. When adoptions considered in lead time evaluations were restricted to earliest participants, the sensing mechanism fails to provide early awareness for over 80% of hashtags. Finally, I find that there is a weak positive correlation between the efficacy of the sensing mechanism and the future popularity of a hashtag.

4.3 Sensor Method with Hashtag Networks

In the sections below, I evaluate the performance of sensors among individual hashtag networks.

4.3.1 Methods

To start evaluations of the sensor method, I look within hashtag networks for each of the hashtags used in collection queries. Given the approach to data collection, a majority of each of these hashtag networks is captured in the data set.

First, hashtags are extracted from each tweet and lower-cased. This step was taken to account for the case-insensitive scraping parameters of the Twitter API (both Streaming API and Tweet Count API). Hashtags containing identical text are semantically equivalent so this preprocessing step also seeks to negate some of the minor differences in hashtag use that could result in two nearly identical hashtags being considered as different objects. I then construct the hashtag network for each "query hashtag" where each network is a directed graph $G = (V_h, E)$, in which nodes V_h are represented by the set of users that tweeted hashtag h at any time and an edge $(i, j) \in E$ exists if user i follows user j.

Following a similar approach to that used by Garcia-Herranz et al., I sample 5% of the hashtag network to construct a control group (Garcia-Herranz et al., 2014). The median size of control groups across all hashtags was ~3,000 users. The 25th, 75th and 90th percentiles of group size were 355, 8,283 and 17,630 respectively. I then use both the local and global approach to construct sensor groups from friends of control users within the network. For each set of samples, I calculate a local Δt^h and global Δt^h. I repeat this process 500 times for each hashtag to generate a distribution of lead times. When constructing sensor groups, if the randomly selected friend of a user was not present in the data set, I assume the user did not tweet using one of the query hashtags in the time frame of interest. While acknowledging that tweets may have been deleted before hydration which would result in occurrences being left out of the analysis, this should be a safe assumption that has minimal impact on results.

4.3.2 Results

Figure 4-4 shows average lead times for collection hashtags across 500 trials of control groups. Of the 35 hashtag networks evaluated, 24 (68.5%) had a negative average lead time – indicating that, on average, the local sensor group successfully provided early awareness. For 13 of those hashtags, sensors demonstrated early awareness across every one of the 500 samples. For 21 hashtags, sensors demonstrated early awareness more than 75% of the time. Average lead times fell within the range of -20 to 15 hours for 34 out of 35 of the hashtags. However, for some hashtags there was considerable variation in lead times across the 500 iterations. Nearly 20% of the hashtags demonstrated a variance greater than 20 – with “#stopnazism”, “#russian_ukrainian”, and “#запутина” showing a variance of over 100 (standard deviation of over 10 hours). Hashtags with smaller average lead times tended towards having lower levels of variation.

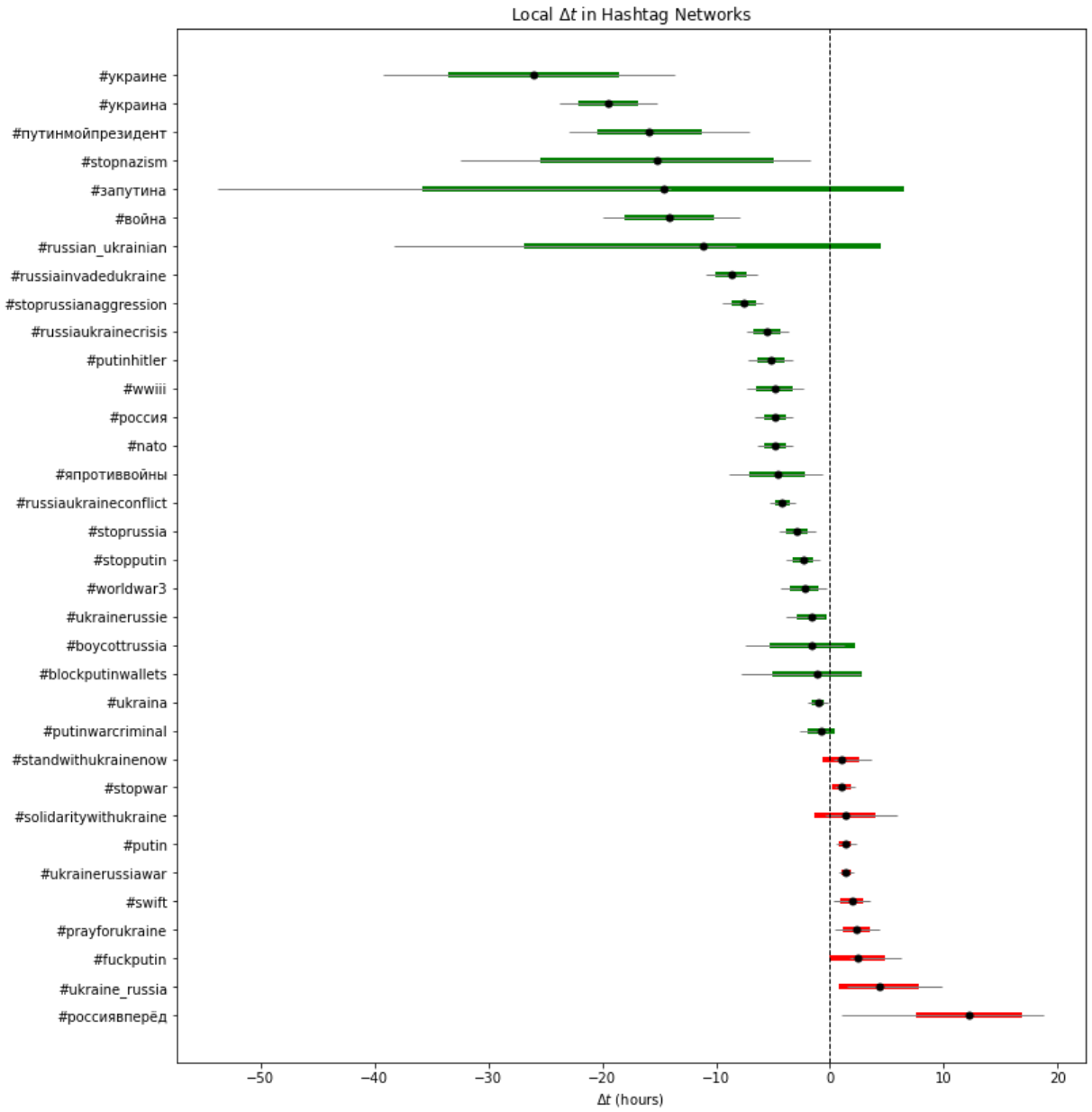


Figure 4-4: Local Δt^h for hashtag networks. Shown are the distribution of lead time values for each hashtag across 500 iterations of randomly selected control groups. Black points represent average lead time, colored bars indicate ±1 standard deviation of lead times, and gray bars indicate the 5th and 95th percentile of lead time values across the 500 samples.

Figure 4-5 outlines the performance of global sensors within individual hashtag networks. Compared to the local sensors, the number of hashtags with an average

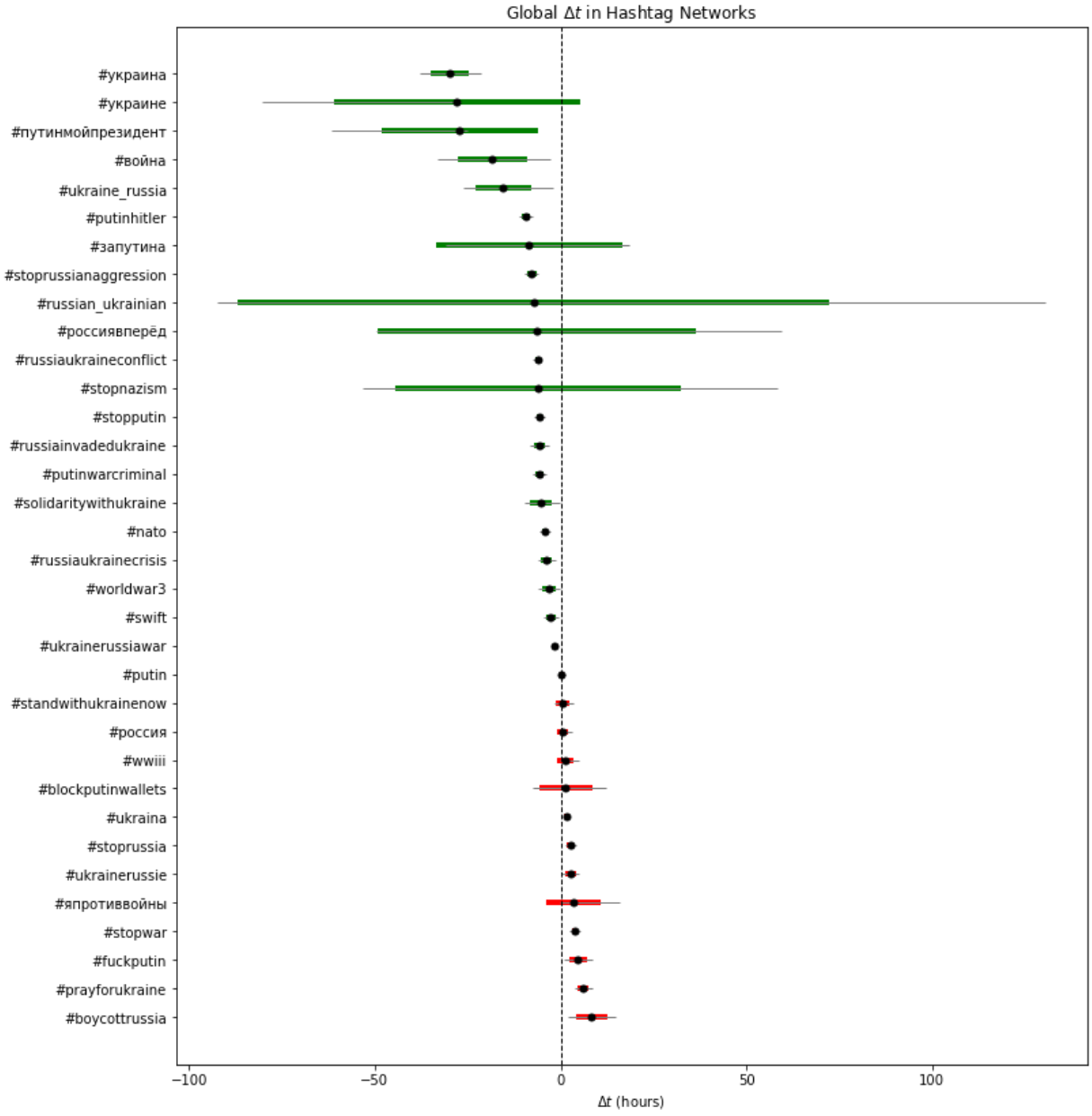


Figure 4-5: Global Δt^h for hashtag networks. Shown are the distribution of lead time values for each hashtag across 500 iterations of randomly selected control groups. Black points represent average lead time, colored bars indicate ±1 standard deviation of lead times, and gray bars indicate the 5th and 95th percentile of lead time values across the 500 samples.

negative lead time drops slightly, from 24 to 21. Of those 21 hashtags, sensors provided early awareness across every trial for only 9 of them – and across at least 75% of the trials for 18. There is one striking difference between the lead times of local and global sensors – the level of variance seen across the 500 trials. While, when using local sensors, 3 hashtags showed a variance of greater than 100, that number doubled for global sensors. In fact, 4 hashtags demonstrated a variance over 1,000. This translates into a standard deviation of over 30 hours, and Figure 4-5 shows that such a spread results in awareness as far as 95 hours in advance or lack thereof by nearly 130 hours.

Given the probabilistic nature of global sensor selection, this result is not particularly surprising – especially when looking within a specific hashtag network. The probability of being added to the global sensor set in this analysis was 0.01. For hashtags with a smaller network, this low probability could very well result in a single or very few sensors being selected, which would explain the extreme variation seen across trials (Note: For each trial, it is required that at least one user be in the global sensor set).

4.3.3 Discussion

While maximum lead times for many hashtags seen in Garcia-Herranz et al. (2014) are on the order of days, hourly lead times, like those that we see here, were also observed in work by Kryvasheyeu et al. (2015) when evaluating sensor performance during Hurricane Sandy. This shift from daily to hourly lead times suggests that significant events in the physical world drive increased rates of discussion on platforms (which is to be expected) and also suggests that the increase in information traveling through external channels may decrease correlation between centrality and time of engagement.

One notable difference of consequence from prior work, is the large variation of lead times seen across repeated samples. In contrast to Kryvasheyeu et al.'s findings where the maximum variance (across even the smallest sample size of 1K users) was 10, several query hashtags present a variance of greater than 100. This large variance for some hashtags brings into question the ability of the sensing mechanism to deliver consistent results – a desired characteristic if it is to be implemented by decision-makers.

This initial analysis presents mixed results on the effectiveness of the sensing mechanism for conflict-related hashtags. Over two-thirds of the query hashtags showing negative average lead times suggests that some pieces of information about the Russia-Ukraine conflict are traveling endogenously through Twitter's network and therefore the sensing mechanism may still be effective. However, lead times varied significantly for some hashtags of interest. Additionally, the hashtags evaluated above were all "popular" or general enough to be selected to guide initial data collection. Realistically we are interested in understanding how well the sensing mechanism performs beyond only widely used hashtags and also how well it performs for samples not contained to individual networks. I explore both of these questions in the next section.

4.4 Sensor Method with Wider Array of Topics

While performance of sensors in individual networks provides some insight into the overall effectiveness of this sensing strategy, for practical applications we are more interested in performance for a wide range of information. Do the initial results generalize to a wider array of war related topics? To investigate this, for the following analysis I randomly select control groups across all of the users in the data set with no restrictions on the friends that can be selected as sensors – i.e., they do not have to exist in the same hashtag network.

This approach allows us to evaluate participation in any Russia-Ukraine related hashtag conditioned on the use of more popular hashtags related to the crisis (those that were used for collection). I make the reasonable assumption that by conditioning collection on more widely used hashtags, a random sample of other war-related hashtags is captured.

4.4.1 Methods

I once again extract and lower-case hashtags from all tweets. I then restrict the population of hashtags to those that contain one of a set of keywords related to the Russia-Ukraine conflict. These keywords were: “ukrain”, “russia”, “ucrain”, “kiev”, “kyiv”, “donbas”, “luhansk”, “donetsk”, “zelensky”, “putin”, “nuclear”, “war”, “biolabs”. Final hashtags considered in the analysis include more general hashtags such as “#battleforkyiv” and “#russiansanctions” and also more pointed hashtags such as “#naziukraine” and “#putinhitlerfascism”. Whilst there are likely other hashtags utilized on Twitter in reference to the Russia-Ukraine conflict that are not captured by this set of keywords, this set of criteria for restriction should not significantly impact results.

Prior work shows that the size of the control sample plays a role in the expected efficacy of the sensing mechanism. This is due to the fact that for groups that are too large, the centrality characteristics of the control and sensor groups begin to overlap. Limited differences in the centrality of the control and sensor groups diminishes the value of the sensors. For samples that are too small, the participation captured may vary significantly. In this scenario, small samples may result in very few to no users in the sensor group having participated in the data set. Such variation can lead to large statistical errors when estimating lead time. Although this is not a comprehensive, theoretical analysis of appropriate sample sizes for the sensing mechanisms, I construct control samples of sizes ranging from 1K users to 50K users in increments of 5K, to get an idea of how results vary.

For each sample size, I uniformly sample several control groups and an equivalent number of local and global sensor groups from their followees. For smaller sample sizes, 20 groups were sampled. As sample size increased, a smaller number of groups were taken to prevent oversampling of the population who had their friendship edges collected. For Russia-Ukraine related hashtags that were used by at least 5 users in at least half of the random samples, I find the average lead time (both local and global) across all of the groups. Detailed information about group sizes and associated number of samples can be found in Appendix A.

4.4.2 Results

In a perfect world – i.e., if the sensor mechanism were always successful – we would see distributions partially bound at the upper limit by zero (or a value very close to zero). This would indicate that the sensor group consistently provides early awareness relative to its control counterpart. Figure 4-6 shows the empirical distribution of Δt for both local and global sensors using a sample size of 40K users. Looking at Figure 4-6, we see the empirical lead time distributions are slightly right skewed but nearly centered around zero.

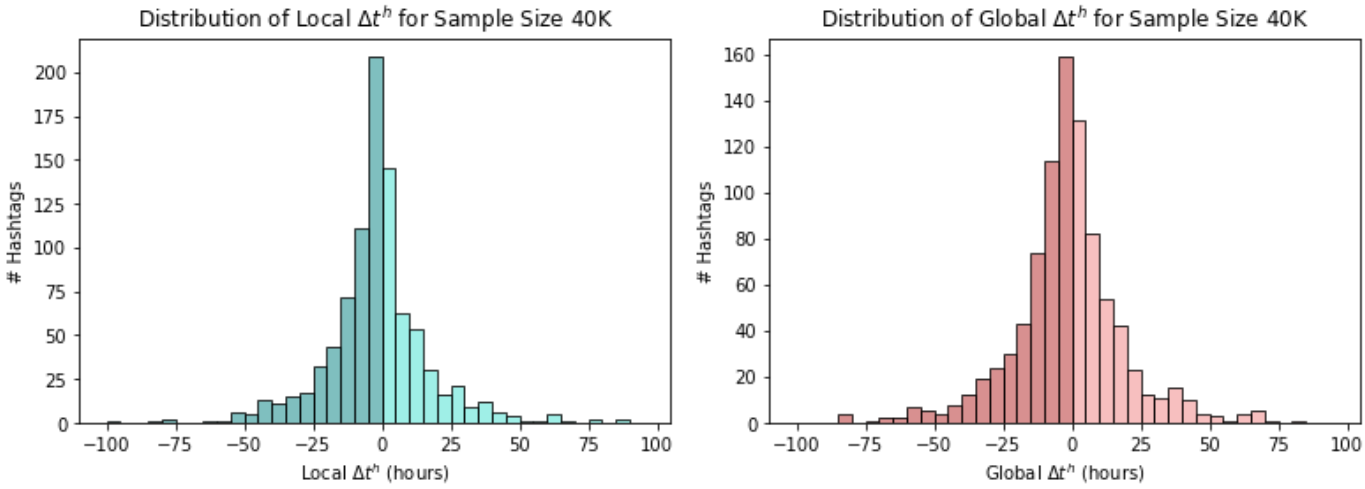


Figure 4-6: Empirical distributions for local and global Δt^h or all hashtags used by at least 5 users in at least 4 of the 5 random samples of 40K users.

Out of 915 conflict hashtags used by at least 5 users in at least 4 of the 5 random samples, local sensors resulted in a negative lead time for 59% of hashtags. Only about 55% of hashtags had a negative Δt using the global sensors. The mean lead time for the local sensors was only -1.57 hours (SEM 0.66 hours). For the global sensors, the mean lead time was slightly lower at -.87 hours (~52 minutes) (SEM .76 hours). The distribution of lead times in the global group appears to have slightly less kurtosis than that of the local distribution. Empirical cumulative distributions (ECDFs) for hashtags in both the local and global groups shown in Figure 4-7 further highlight that hashtag lead

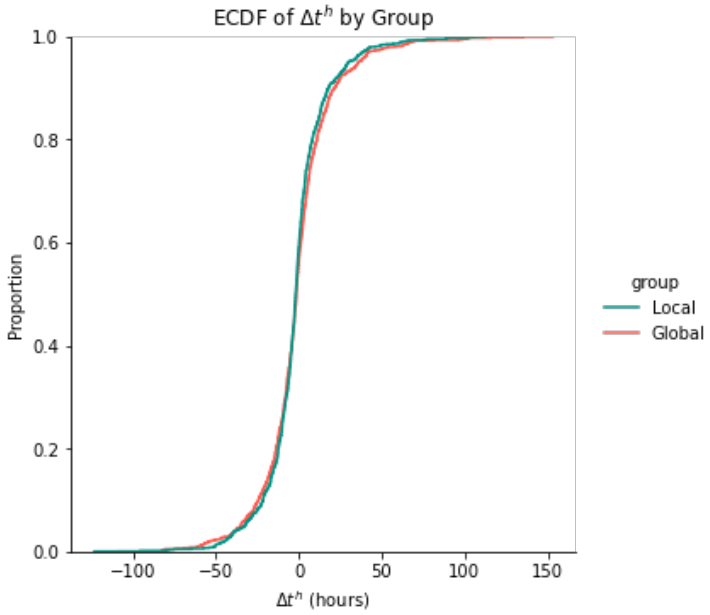


Figure 4-7: ECDF of Δt^h by group for all hashtags used by at least 5 users in at least 4 of the 5 random samples of 40K users.

times are centered around 0. Lead times from local sensors have a slightly longer left tail than the global group – indicating more instances of extreme early awareness – and the opposite is true on the other tail of the distribution. A sample size of 50K users exhibited very similar tendencies to those shown here.

In contrast to work done by Garcia-Herranz et al. where nearly 70% of hashtags had a lead time less than 0, this 10% decrease indicates that the presence of extensive sharing exogenous to Twitter may in-fact depreciate the efficacy of local sensors. Alternatively, this may suggest that the following network may not be the most relevant network to follow for sharing in this environment. These results were fairly consistent across sample sizes. Figure 4-8 shows how the proportion of hashtags with negative lead times varies with the size of the control group that is sampled. Peak proportion of negative lead time (for local sensors) was seen with a sample size of only five thousand users. However, because we are looking at specific hashtags instead of broad topic participation (as in Kryvasheyeu et al.), such a small sample size only captures 347 hashtags vs the 1105 captured by a sample size of 50K users. The larger sample size

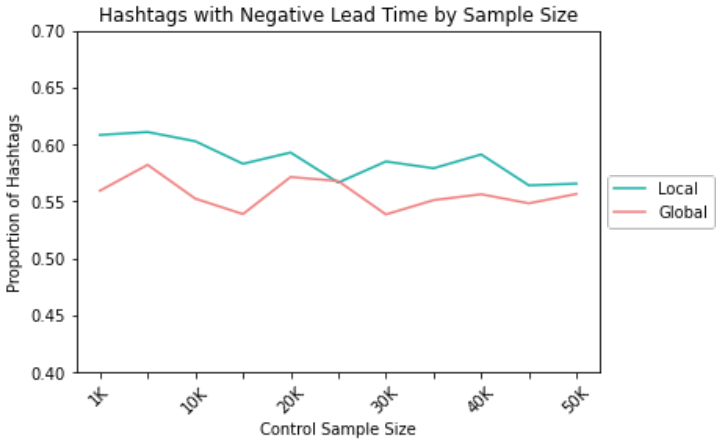


Figure 4-8: Proportion of hashtags with negative lead times for both local and global sensors across varying sample sizes.

gives us the greatest picture of current discussions about the conflict without making drastic sacrifices in performance.

Both evaluations of lead time in hashtag networks as well as evaluations with small samples show that on average, sensors provide early awareness over a control group ~60% of the time. While the magnitude and frequency of early awareness found is not as dramatic as has been seen in other contexts, having early awareness just over half of the time is better than no awareness at all and perhaps may still be useful when combined with other strategies. These outcomes provide initial hope that using the local structure of a network to guide sampling may still be an effective strategy, even during times when there is an influx of discussion on channels external to the platform of interest.

4.4.3 Alternate Metrics for Lead Time

So far, lead time from small samples show that sensors engage earlier than their control counterparts for just over half of the war-related hashtags, but are these results robust to the window of observation used to identify users? Evaluating adoption time between control and sensor groups using the previously proposed Δt^h metric runs the risk of the

average adoption time from either group being misleadingly shifted upwards by late adopters.

Figure 4-9 provides an example of this concern. Posit 5 users from a randomly sampled control group tweet using the hashtag *#stopputinwarcrimes* during the window of observation. The average adoption time across all of the users – time used in calculations for Δt^h – is shown by the dashed vertical black line. Even though over half

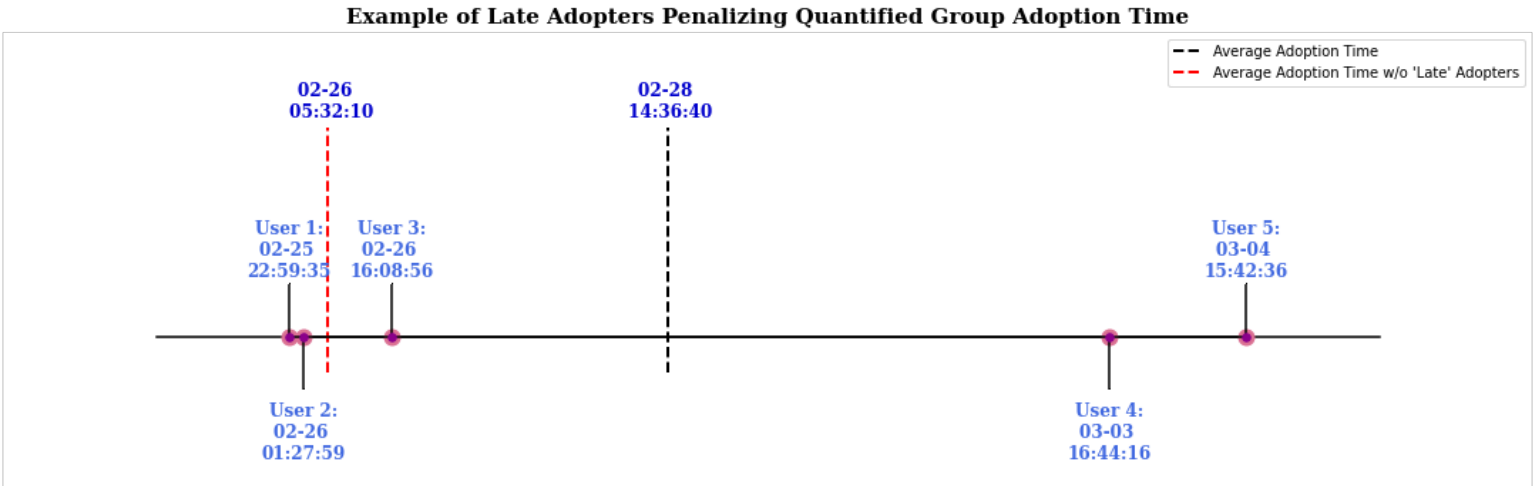


Figure 4-9: A hypothetical example of late use of a hashtag decreasing the overall adoption time of a group.

of the users in the group have tweeted the hashtag by 16:08 on February 26, the recorded average adoption time for the group is 14:36 on February 28. If we were to only consider the participation of the first three users however (the top 50th percentile), the average adoption time of the group shifts up by over two days – shown by the dashed vertical red line.

This is an important variable to consider for two reasons. First, the avenue used to select control users in this setting naturally lends itself to more users participating in the control group than in the sensor group – as control users are selected from a population of individuals that have already engaged in a conflict related hashtag. This means that there is a higher probability of average adoption time for the control group being penalized.

Consequently, estimated Δt^h values are likely to overestimate the ability of a randomly selected sensor group to provide early-awareness of conflict-related discussions.

Second, from an applied perspective, an indication of the average participation for a set of users is not beneficial for informing real-time decision making. For purposes of gaining awareness, we derive value from sensors through their capacity to engage with information *prior* to a random group of users. This means – for assessing lead time – we have virtually no interest in the right-hand side of the adoption curve. So, it is important to question whether prior characterizations of lead time that include these late adopters are truly sensible measures.

To further understand the impact of late entry on the perceived effectiveness of sensor groups, we propose several metrics whose implementation could help combat the bias inherent in averaging across all adoption times. Table 4.1 outlines a summary of the proposed metrics.

ΔtU_X^h	The difference in time between the Xth occurrence of hashtag *h* in the control group and Xth occurrence of hashtag *h* in the sensor group for X ϵ {1, 2, 3, 5, 10}
ΔtP_Y^h	The difference between the average adoption time for hashtag *h* of the top Yth percentile of the control group and the average adoption time of the top Yth percentile of the sensor group for Y ϵ {95, 90, 75, 50}

Table 4.1: Proposed metrics to evaluate early awareness.

ΔtU_X^h **:** Looking at the difference in adoption time between individual users in each group is an interesting metric from a pragmatic perspective – as it allows those analyzing activity to make declarations about lead time after a concrete number of occurrences. Additionally, evaluating the difference between ordered users could serve as a proxy for rate of diffusion among either group – which could be useful for understanding the sharing trajectory of a certain hashtag.

However, this approach to characterizing participation poses several limitations. First, given the procedure for selecting users into the control group for these evaluations, it is not unlikely that the number of occurrences in the control group would be greater than the number of occurrences seen in the sensor group. This means adoption time of higher order users in a group with increased participation will be compared to the adoption time of the first few users in the group with less participation. In such scenarios, this metric may continue to misrepresent the difference in adoption between two groups. Second, methods using single instances of participation to quantify lead time are extremely sensitive to slight changes in involvement from either group. The choice by a single user to engage with a hashtag could cause lead time to fluctuate by many hours or even days.

ΔtP_Y^h : Using the average adoption time of the top Yth percentile of adopters from each group to calculate lead time is also a promising approach because it helps directly diminish the bias created by including late adopters. Instead of considering every single user in a group that tweets a specific hashtag, we can shift our focus to a smaller group of users that participate earlier on.

Compared to evaluating differences between strict counts of participation, percentile averaging is much more robust to the number of occurrences in either group – reducing the overall sensitivity of the lead time metric. However, this approach presents one main drawback. Similar to full group averaging, to include this metric in a framework for active evaluation of online activity, there must be a delineated time frame across which percentiles can be computed. To know which users should be included in the top 75th percentile, we have to see the full spectrum of individuals that use the hashtag within the group – including those that are not in that percentile. Additionally, such a metric still has the ability to be influenced by additional participants that enter a conversation on the far right-hand side of the adoption curve. Future iterations of work may consider calculating percentiles relative to the total number of individuals in the groups of interest as opposed to the number of individuals in the group that share the

hashtag of interest. I argue that even though this metric has practical limitations, it is still very useful for evaluating robustness of previous outcomes.

4.4.4 Results for Additional Metrics

Results from initial Δt^h investigations indicate that there are greater instances of negative lead times than there are positive – that the sensing mechanism successfully provides early-awareness for more than half of the conflict related hashtags. However, how does the success of the mechanism change when using one of these newly proposed metrics? To answer this question, I use the same method for selecting control and sensor groups discussed above. As before, the proposed metrics are calculated for each group and then for hashtags used by at least 5 users in at least 4 of the groups, values were averaged across samples.

User Based Lead Time

When comparing single instances of participation between the control and sensor groups, users in the control participated ahead of those in the sensor group for a majority of the hashtags. Figure 4-10 highlights the significant decline in proportion of hashtags that

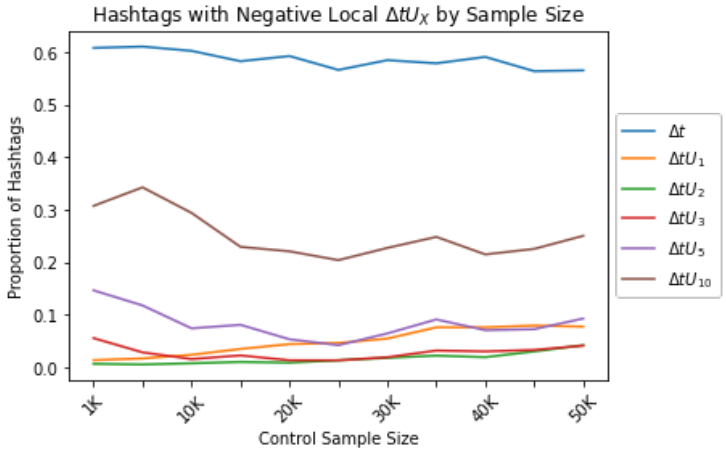


Figure 4-10: Proportion of hashtags with negative lead times for local sensors across varying sample sizes.

exhibit negative lead times under different variations of this adjusted metric. For every metric besides ΔtU_{10} , sensor groups provided early awareness for less than 10% of the conflict hashtags.

Further investigations of the underlying distributions of each of these metrics reveals that not only do ordered users in the control group participate earlier than their sensor group counterpart, but the magnitude of the time difference is relatively substantial. As an example, Figure 4-11 shows the binned distribution for ΔtU_1 and ΔtU_3 for a sample size of 50K users. It is apparent that these lead time distributions are heavily skewed. For ΔtU_1, nearly 40% of hashtags had lead times in the uppermost bin – indicating adoption by the sensor group was more than 30 hours behind that of the control.

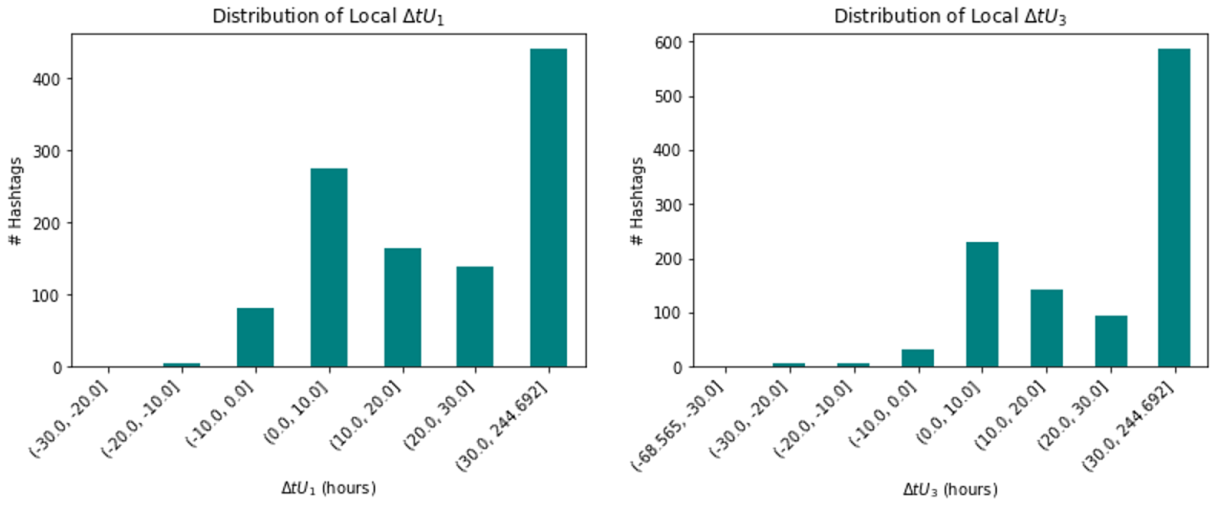


Figure 4-11: Binned distributions of lead times for ΔtU_1 and ΔtU_3

Percentile Based Lead Time

Figure 4-12 uses bins of equal width to display the frequency of hashtag lead times for the four 'percentile' based metrics. At first glance, it is clear that lead time distributions for each of the metrics are no longer positively skewed. Restricting the population of users for each hashtag to those that were in at least the first half of adopters, increases the mean lead time by over 15 hours – from -.98 hours (SE .67) for local Δt to 14.95 hours

(SE .86) for local ΔtP_{50}. The number of hashtags for which the sensor group provided early awareness decreased by almost 50% – dropping from 625 to 330.

This degraded performance supports the notion that by taking the average entry time of all users that tweeted a particular hashtag, those in the group that participate later on tend to drive down the overall adoption time. Again, with the increased likelihood that a control group contains a larger number of users than its counterpart sensor group (due to the selection of control groups being conditioned on initial use of a popular hashtag), this bias results in the sensing mechanism appearing more effective than it really is.

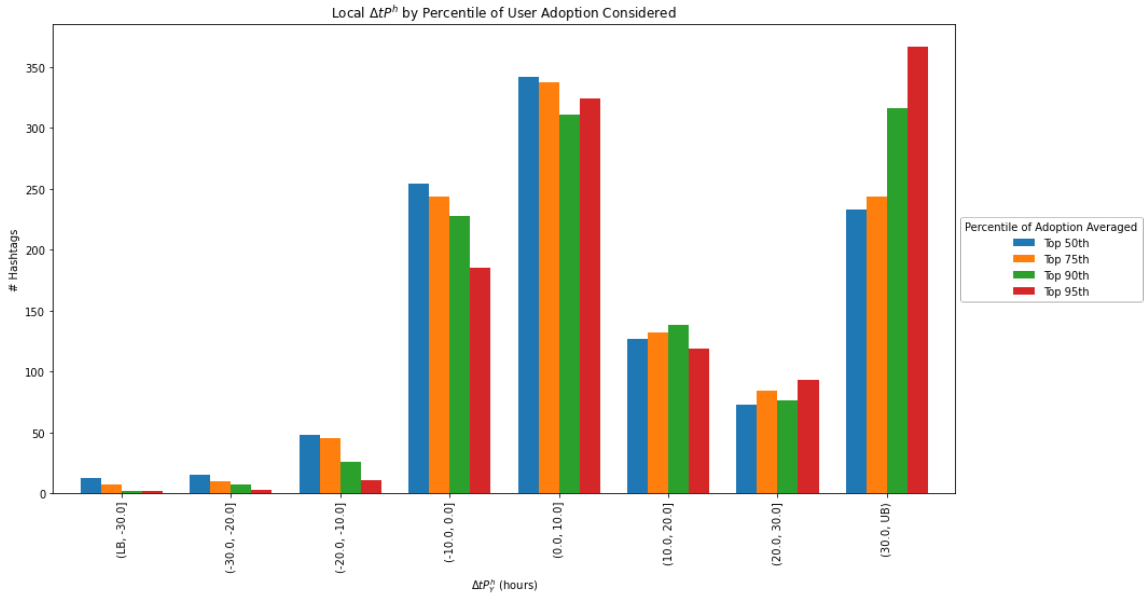


Figure 4-12: Lead times provided by local sensor groups by percentile of user adoption considered. Note: the tail bins contain all occurrences greater than or less than the stated bound.

The claim that bias created by late adopters inflates the performance of the sensing mechanism is further reinforced by the variation in lead time distributions seen across the different metrics in Figure 4-12. Lead times calculated using the average adoption of the top 50% of users in a group are depicted by the blue bars. It is clear that, when compared to the other three features, ΔtP_{50} has the greatest number of hashtags with a negative lead time. As smaller proportions of each group are considered for averaging, the distribution of lead times begins to shift to the right. For ΔtP_{75} the average

lead time increases to 17.4 hours (SEM .93 hours). For ΔtP_{90} the average lead time increases once again to 23.5 hours (1.08 hours). This shift continues for each marginal decrease in percentile considered until, when looking at the top 95th percentile of users, a majority of hashtags evaluated present a positive lead time.

This analysis shows that in times of increased sharing exogenous to Twitter, random friends are unable to reliably provide early awareness over a group of random users. So, can the difference in adoption between the two groups provide value beyond simple hashtag awareness?

4.5 Lead Time as a Potential Indicator of Popularity

While it is interesting to understand the ability of a sensor group to provide early-awareness of online discussions, it is not evident how beneficial early awareness itself would be for decision-makers – particularly given the hourly scale of differences that we see for conflict-geared sharing. It is then reasonable to ask, can lead time outcomes provide insights into other characteristics of interest for online discussions?

When leveraging social media during times of conflict, there is generally a heightened interest in items that are or are going to become widespread on a platform. I acknowledge, this may not always be the case. For instance, an extremely fervent chain of discussions encouraging dangerous collective behavior would be of interest to public leaders even if contained to several hundred users. However, in scenarios where time and/or resources are of the essence – attention of decision-makers will likely be directed towards messages that are going to reach thousands of users versus messages that are contained to a cascade of 10 individuals. With this in mind, a natural question to ask is 'What can observed lead time tell us about the future popularity of a particular hashtag?'. If, as an analyst proactively monitoring lead times of hashtags, I see that a group of sensors adopted "*#BiolabsFoundNow*" 20 hours prior to a control group – what does this tell me about how many other people are going to be talking about this topic in the future?

Figure 4-13 shows the composition of lead times (both ΔtP_{50} and ΔtP_{95}) by the overall number of times a hashtag was tweeted in the four weeks following the initial invasion of Ukraine. The bulk of hashtags that appeared in more than 18,000 tweets are concentrated around lead times from -10 to 10 hours (shown by the purple segment of the bars) – an indication that the sensing mechanism is generally effective for hashtags that are tweeted a greater number of times. This 20-hour range of lead times, however, is also heavily populated with hashtags that were shared a fewer number of times. In fact, for ΔtP_{50} specifically, over 75% of hashtags that display lead times between -10 and 10 hours were shared less than 18,000 times. Beyond widespread hashtags being focused near zero, there does not appear to be any clear visual trend for overall hashtag usage.

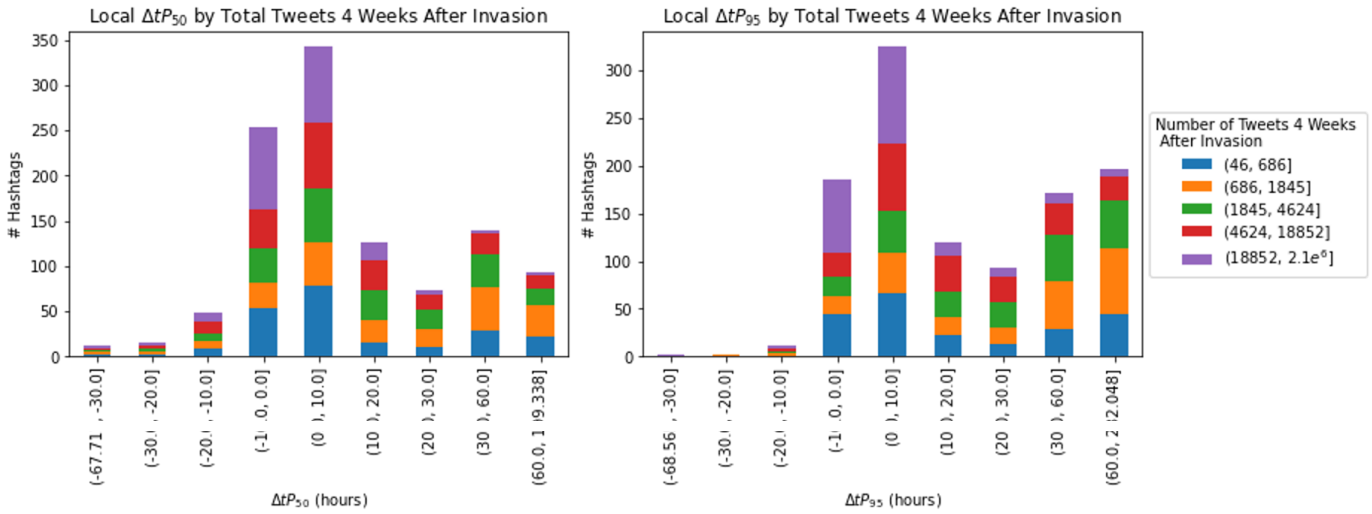


Figure 4-13: Lead time by total number of tweets containing hashtag

Work by Garcia-Herranz et al. found similar results surrounding the behavior exhibited in Figure 4-13, namely that hashtags that end up being used by a greater number of people were "more likely to exhibit smaller lead times" but this behavior "does not work the other way around" (Garcia-Herranz et al., 2014).

Given the excessive quantity of hashtags posted to social media platforms, the value of early awareness without any indication of "importance" or virality is debatable. So, we are left with the task of trying to identify the relationship between lead time and sharing trajectory on Twitter. In the section below, I attempt to quantify the underlying

relationship between outcomes of the proposed sensing mechanism and overall engagement with a hashtag on Twitter. I find that for nearly all metrics, larger lead time values correspond to a fewer number of shares. This suggests that the sensing mechanism is more useful for hashtags that are more widely shared. However, the magnitude of this relationship is fairly small and therefore would only be beneficial for ballpark projections of hashtag popularity.

4.5.1 Methods

To understand the relationship between lead time and overall use of a hashtag I fit a univariate model of the following form:

$$Y_{ph} = \beta L_h + \epsilon_h \tag{4.1}$$

where Y_{ht} is the number of tweets containing hashtag h shared p weeks after the date Russia initially invaded Ukraine and L_h is the lead time metric of interest. With this simple univariate approach, we are seeking to identify the incremental change in sharing of a hashtag that we should expect to see depending on the lead time produced by the sensing mechanism.

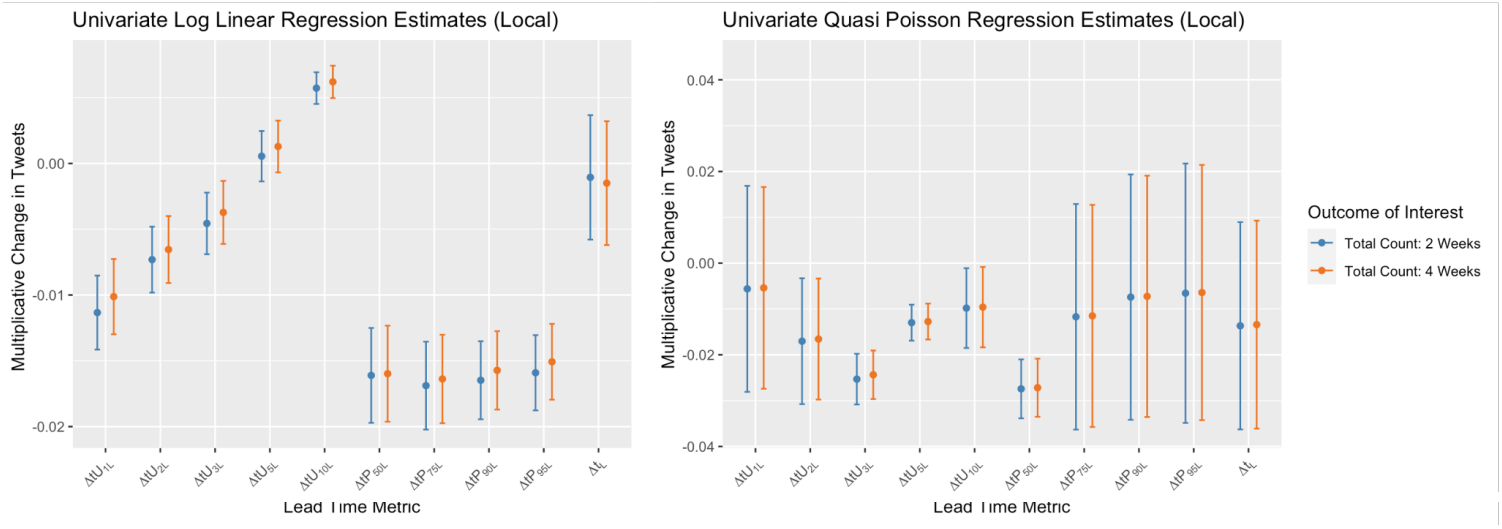
I once again use the lead time outcomes from the 5 samples of 50K users. For initial fitting, I hold out a random set – 20% of the 1105 total hashtags present – to be used for testing/validation.

4.5.2 Results

I estimate the model in Equation 4.1 separately for each lead time metric proposed in Section 3 – a total of 20 individual models. I use two main approaches for modeling count data as an outcome variable: log-linear model and quasi-Poisson generalized linear model (GLM). The quasi-Poisson GLM accounts for overdispersion present in total tweets counts. Figure 4-14 shows the coefficient estimates for each local lead time metric for total counts of both two and four weeks. Each model was fit with heteroskedasticity robust standard errors.

For the log-linear model, 80% of the local lead time metrics were significant at the 5% level. Of those that are significant, the only variable with a positive coefficient was that of ΔtU_{10}. Estimates for those with negative coefficients range from -.0025 to -.017. In this context, a negative coefficient signifies that the less successful the sensing mechanism (the larger the lead time metric) the less popular a hashtag is likely to become.

For the quasi-Poisson models, standard errors for a majority of the local metrics are too large to conclude that any relationship exists between lead time and the total popularity of a hashtag. Coefficient estimates for both ΔtU_5 and ΔtU_{10} flip from being positive in the log-linear model to now being negative – aligning with the hypothesis that negative lead times indicate trends to pay attention to. ΔtU_2, ΔtU_3, and ΔtP_{50} are the only three metrics that had significant, negative estimates across both of the models. Model outcomes for global sensors were extremely similar to those found using local sensors. Plots of coefficient estimates for global sensors can be found in figure A-1 in Appendix A.



(a) Coefficient estimates for log linear model

(b) Coefficient estimates for Poisson model

Figure 4-14: Univariate regression coefficient estimates under different modeling decisions.

While the coefficient estimates from both models are fairly small, it must be kept in mind that e^{β} is the multiplicative change in the total number of tweets containing the given hashtag after the time period of interest. For example, a coefficient of -.02 indicates we should account for a 2% decrease in total tweets (relative to the intercept) for every hour increase in lead time. For a majority of the models, the intercept hovered around 8, which equates to ~ 3,000 total shares.

The inconsistency in results seen across the two models indicates they are not robust to modeling decisions. And, when paired with the small coefficient values, it is unclear whether the magnitude of lead time displayed by the sensing mechanism provides a helpful indication of how widespread a hashtag will become. In the next section, I investigate how well the models above are able to predict the total shares a hashtag will receive.

4.5.3 Predicting Total Use

If we see a hashtag present a certain lead time, how well are we able to predict the magnitude of its future sharing? To understand how informative the estimated models would be in an applied setting, I use the log-linear model to make predictions on the set of hashtags that were held-out from initial fitting. I use ΔtP_{50} as the univariate predictor due to its significant and consistent coefficient estimates in both the log-linear and quasi-Poisson models. Results across each of the outcome variables were nearly identical so I use two-week totals as the response – near-term popularity is also likely to be of greater concern when addressing information spreading on a platform.

When applied to predict the log of total tweets two weeks after the invasion, performance by the log-linear model was not overwhelming. The adjusted R-squared value of the model was very low at .05. However, given all of the other factors that play a role in the spread of information through a network, we wouldn't necessarily expect a univariate model to be able to account for a massive portion of the variation in total shares (Cheng et al., 2014). The model produced a mean absolute error (MAE) of 1.60 and a mean squared error of 4.79. While these error values are well above zero, it is

important to consider if such a magnitude of error is still acceptable given the overarching goal.

As a decision-maker concerned with discourse about the Russia-Ukraine conflict on Twitter, the exact number of times that a piece of information is tweeted or retweeted is not of interest. Pragmatically, it is of much greater value to have information on the relative extent to which a discussion is going to be shared. Is the hashtag going to be used by 10 people or is it going to be used by 100,000 people? For lower estimates of tweet volume, an error of 2 could be the difference between projecting 8 total tweets of a hashtag (e^2) or ~ 55 total tweets (e^4). Taken in context, this error is relatively minor and has little potential to thwart any crucial decisions. For higher log estimates, the magnitude of the error becomes slightly more substantial – an error of 2 hours could now be the difference between predicting 22,000 total tweets (e^{10}) or 160,000 total tweets (e^{12}) . Arguably however, once total sharing surpasses a certain threshold – this may be ~5,000 tweets – the overall total becomes less valuable.

When used alone, the predictive power of lead time produced by the sensing mechanism is not overwhelming. Although, with a mean error in log-uses around 1.6, the model could be beneficial for prioritizing attention and resources in instances of compressed time. Even a strategic reduction of the volume of information by 50% would be an improvement over wading through every hashtag related to the conflict that is shared among a sampled group of users.

4.6 Discussion

To summarize, much of my analyses suggest that for discussions related to the Russo-Ukraine conflict, sensors were unable to reliably provide early awareness of discourse on Twitter. Using the Δt metric from prior literature, sensors provided early awareness for only 59% of hashtags related to the conflict. After accounting for late adopters in either group, newly proposed metrics show that the sensing mechanism was successful even less than half of the time. This strongly suggests that the influx of communication through external channels during the conflict decreases the correlation between degree

and activity of a user – which subsequently devalues friendship connections for sensing purposes on the platform.

Even for hashtags for which sensors were able to provide early awareness, the lead time was on the order of hours. It is important to question how much value early awareness is on such a small scale. For use cases where interventions are made directly on the platform, several hours could prove to be advantageous. For example, on Twitter's community-based fact-checking system – Birdwatch – early identification of content would provide the opportunity for labelers to effectively assess information and subsequently flag or recommend for removal before it has the opportunity to spread to extended parts of Twitter's network. Such lead time could help mitigate overall exposure to misleading or pernicious content on the platform – directly combatting the overall impact of misinformation, disinformation, and harmful speech. Several hours may also prove to be critical for search and rescue efforts.

Beyond direct interventions and instances of time-sensitive emergencies, the value of early awareness becomes unclear. Realistically, it takes time for bureaucratic agencies to compile and then subsequently incorporate information into decisions or any form of public interaction. If content on the platform is being used to gauge public sentiment (for example), the resources required to follow friendship connections may not justify attaining knowledge a few hours earlier.

Additionally, my analysis shows that lead time, when used as a univariate predictor, can give ballpark estimates of overall hashtag sharing. However, the model fails to explain a majority of the variation in total shares and decent sized errors may result in paying attention to hashtags that are inevitably shared only a few times or more importantly missing hashtags that eventually become widespread.

Overall, these results illuminate the need to continue empirical evaluations of these sensing mechanisms to expand our understanding of their capabilities in diverse contexts, to explore ways in which they can be effectively deployed in an applied setting, and to thoroughly think through the role that acquired information may have in decision-making processes.

Chapter 5

Active Selection & Value Quantification of Sensors

In the previous chapter, I evaluated the ability of sensors to provide early awareness of hashtag use and assessed the value of subsequent lead times for predicting popularity. However, there remain a few important unanswered questions. 1) In a practical setting where we are interested only in discussions about the Russia-Ukraine war, how do we go about selecting an initial control group from which to sample sensors? 2) Is the following edge truly the most informative for selecting useful sensors? 3) How can we systematically quantify the value that sensor groups add to our understanding of the information being shared on a platform?

In the following chapter, I propose and evaluate a framework for using early participation in widespread hashtags as a means for identifying sensors for future trends in conflict. I incorporate the multi-layer component of Twitter's network and explore the retweet edge for purposes of sensor selection. I show that sensors sampled from both explicit following connections and prior retweet connections result in more connected users than a random sample who employed a popular war-related hashtag. However, control and retweet sensor groups were overall more active and shared more tweets related to the conflict than users in a local sensor group.

Finally, I use two methods to quantify the value of sensors. First, a simple, count-based approach shows that local and retweet sensors (together) would have detected over 80% of 'widespread' conflict-related hashtags while sharing 36% less than a random control group. Second, a predictive modeling approach shows that basic indicators of

participation from sensor groups can help improve our ability to predict how widespread a hashtag from the control group will become. Predictive performance declines when models include all hashtags found across the three groups. Nonetheless, evaluating sensor value through a predictive modeling lens is much more robust than methods used in the past.

5.1 Framework and Data Collection

5.1.1 Early Participation to Identify Control Groups

For analyses in previous sections as well as for those in Kryvasheyeu et al., a control group is randomly sampled from a population that has already been refined to a topic of interest – i.e. a group of users that used a keyword related to Hurricane Sandy or a group of users who have used a Russia-Ukraine hashtag of interest (Kryvasheyeu et al., 2015). In an applied setting, we wouldn't have the ability to restrict the population in such a precise way prior to sampling. And, when concerned with specific topics like we are in these cases, randomly sampling from the entire Twitter population runs the risk of missing out on many relevant discussions.

As an initial effort to address this shortcoming, we propose that prior engagement in more widespread hashtags – in this case hashtags such as #Ukraine, #UkraineRussia, #Putin – could be used to condition selection of a control group from which we can then identify sensors for *future* war-related trends. While this framework does require some period of initial participation and is consequently directed towards settings of extended activity, I argue it is still a beneficial starting point for developing a robust framework for employing sensors during times of conflict. Further work may seek to understand if prior participation in discussions of similar genre or categories – as opposed to specific events – could be used in a similar fashion.

5.1.2 Retweet Connections as Sensors

While the nature of follower/followee connections create an explicit, publicly available social graph to describe relationships on Twitter, other interactions on the platform, such as retweets, may also be used to connect users. The retweet graph can be described as a directed network graph $G = (V, E)$, where nodes V are represented by the set of users on the platform and an edge $(i, j) \in E$ exists (connecting node i and node j) if user i has retweeted one of user j's tweets.

A notable component of literature studying the dynamics of information sharing on Twitter have used this layer of Twitter's graph as the pathway of information exchange (Barberá et al., 2015; Thomas et al., 2021). Work by Bild et al. shows that compared to features typically seen in follower graphs, both the clustering and assortativity of retweet networks are more aligned with characteristics of real-world relationships – demonstrating that "retweets more closely mirror real-world relationships and trust" than standard follower connections (Bild et al., 2015).

Considering these revelations and also the fact that platforms have shifted from chronological timelines of posts from "in-network" accounts to timelines dictated by content recommendation algorithms incorporating an amalgamation of factors from various sources – 50% in-network and 50% out-of-network – it is necessary to start exploring whether friendship connections are truly the best edge to follow for purposes of identifying sensors (*Twitter's Recommendation Algorithm*, n.d.).

5.1.3 Group Selection

I incorporate both retweet connections and prior participation into the sampling framework outlined below. To create a control group, I randomly sample 20K users that participated in the original Shevtsov et al. data set within the first seven days of Russia's initial invasion of Ukraine – February 24th to March 2nd. Next, following the same process as before, I randomly sample one friend from each user in the control group – removing any duplicates – to create a sensor group. In-line with the notation used in previous chapters, I will refer to this as the "local sensor group".

For retweet connections, I identify all retweets of users in the control group during the first week of the conflict – 82% of users in the control group had at least one retweet captured in the data set during this initial period. For each control user with at least one retweet, I randomly sample a retweet connection and add the original author to the sensor set – again removing any duplicate users. This group will be referred to as the "retweet sensor group". A total 18,603 users are present in the final local sensor group and 7,236 users are present in the final retweet sensor group. The dramatic reduction in users found in the retweet sensor group is an indication that many control users were retweeting the same accounts early on in the conflict. This could prove to be an effective way of reducing noise (from extraneous, irrelevant tweets) if the same source accounts were consistently retweeted as the conflict persisted.

5.1.4 Data Set

To understand the performance of the selected groups for detecting future trends in the Russia-Ukraine conflict, we are interested in the activity of each group in the months following the initial invasion. Therefore, I collected the entire timeline – from February 22, 2022 to March 31, 2022 – of each user contained in one of the three samples. Tweets from the week of the invasion were scraped to further understand initial activity levels surrounding the conflict. The full set of data consists of 53.3 million tweets – 32.6 million from the control group, 12.2 million from the local sensor group, and 8.5 million from the retweet sensor group.

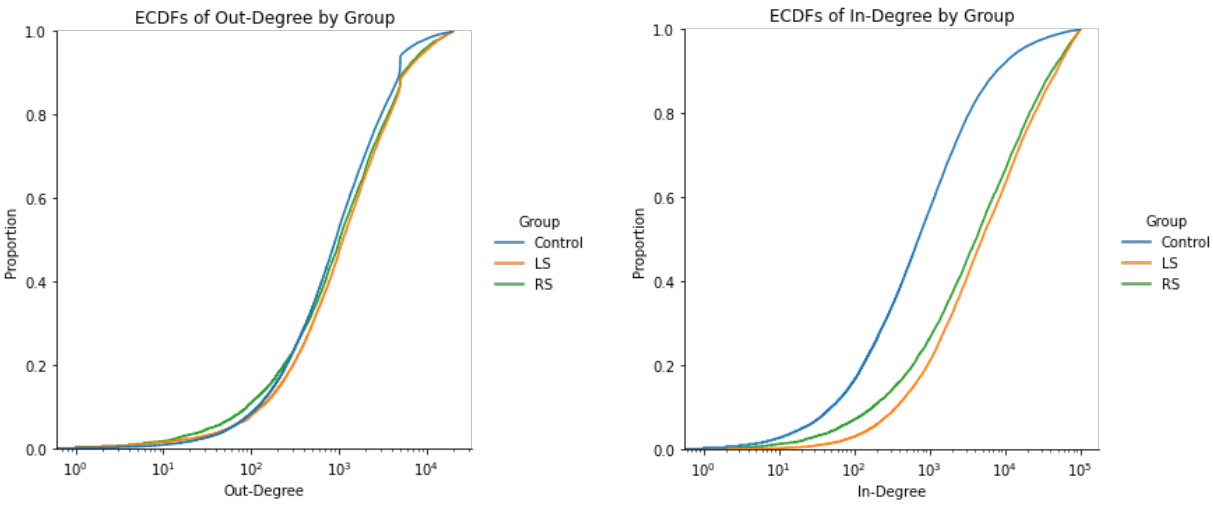
5.2 Group Characteristics

The primary reason that randomly sampling friends as sensors is relied upon is its ability to access groups of people that are 'on average' more central in a network. However, it's not a-priori clear how the connectivity of a retweet sensor group would compare to the control group and its local sensor equivalent. Control users could be retweeting anything from niche opinions from less popular accounts to general news broadcasts from

mainstream accounts with many followers. Work by Barberá et al. has shown that in instances of political protests, a majority of retweets about protests are "sourced" from a core group of participants (Barberá et al., 2015). Such evidence indicates that following the retweet network during times of conflict could help identify people that are highly connected or at least users that are sharing content that is successfully spreading across Twitter. The section below explores this further.

5.2.1 Group Centrality

If information related to the war is spreading endogenously via edges in the Twitter network, users with more friends (larger out-degree) have a greater opportunity to be exposed to information prior to those who are less connected. On the other hand, users with a larger in-degree are interesting because they have greater exposure capabilities – they can reach more users with the information that they post on their account. If high in-degree users captured in the sensor groups are responsible for generating content related to the conflict, the information they share may be indicative of topics that are going to reach and subsequently be adopted by a large number of users.



(a) Out-Degree of Users by Group (b) In-Degree of Users by Group

Figure 5-1: Degree Distributions by Group.

Figure 5-1 shows the empirical cumulative distribution function (ECDF) of the in-degree and out-degree of users in the control, local sensor, and retweet sensor groups. Local and retweet sensors are referred to as "LS" and "RS" respectively.

In line with results from recent literature, the leftmost plot shows that there are more users with a high out-degree in the local sensor group than there are in the control group. This is the outcome we would expect given the properties of the friendship paradox. However, the close proximity of the curves indicates that the differences in degree between the two groups are relatively minor. Interestingly, the ECDF for the retweet sensor group shows that it contains more users of low out-degree than the control group and also more users of extremely high out-degree. This reinforces that early in the conflict, individuals retweet both users with very few friends and users with many friends.

As for in-degree, the rightmost plot demonstrates that both the retweet sensor and local sensor groups contain more highly followed users than those that are captured in the control group. Retweet sensors are not quite as highly followed as local sensors. Still, this confirms that following retweet connections is a feasible way to sample users that are more connected on average than a randomly selected control group.

5.3.2 Group Activity Levels

Total Activity

How did the activity levels of each of the groups compare? Figure 5-2 below shows the total number of tweets shared per day by each of the groups. In the days following Russia's invasion of Ukraine, the control group had nearly three times the number of tweets per day than those produced by either the local or retweet sensor group. Later in the month, this disparity decreases slightly. While the difference in tweet quantity between the control and retweet sensor group is not surprising – due to the difference in number of users in the final groups – we would expect the number of tweets from the sensor group to be more closely aligned to that of the control.

Further evaluations reveal that the disparity in tweet volume between the control and local sensor group stems in part from the lack of active users among the local sensors. Figure 5-3 shows that 80%-90% of users in both the retweet sensor and control group were active each day across the month of interest. Adversely, the proportion of active users in the local sensor group hovered between 50%-60%. This highlights one of the downfalls of randomly selecting one-hop connections to serve as sensors – they may not be active on the platform. Even for those users that were active on the platform, the average number of tweets per active user in the local sensor group was ~30, while that for the retweet and control groups were ~35 and ~50 respectively. Figure A-2 in the Appendix provides further detail on the average number of tweets per active user over the month of collection.

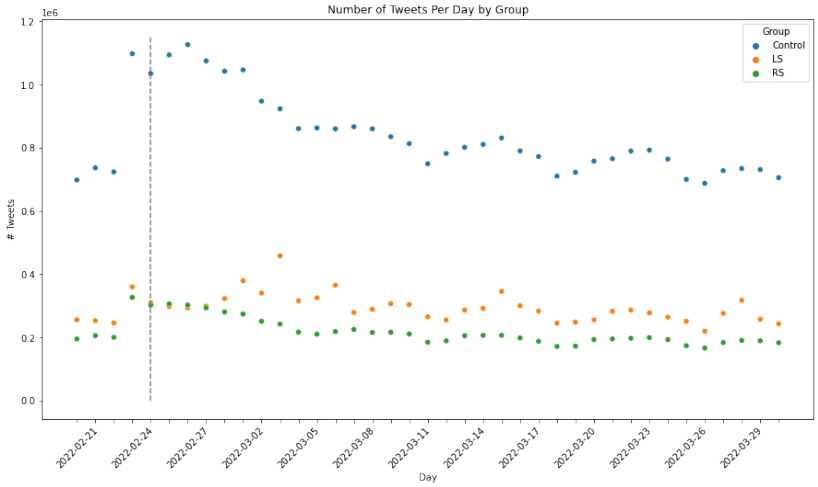


Figure 5-2: Total number of tweets per day by group

This stark contrast in level of activity seen between the control and local sensor group contradicts prior work by Garcia-Herranz et al. (2014) and Kryvasheyeu et al. (2015), who found that groups of one-hop connections displayed higher levels of activity than those found in a group that was randomly selected. This suggests that using the non-network signal of participation in early discussions may produce users who are generally more active on the platform – more so than users found through random sampling with no signal at all.

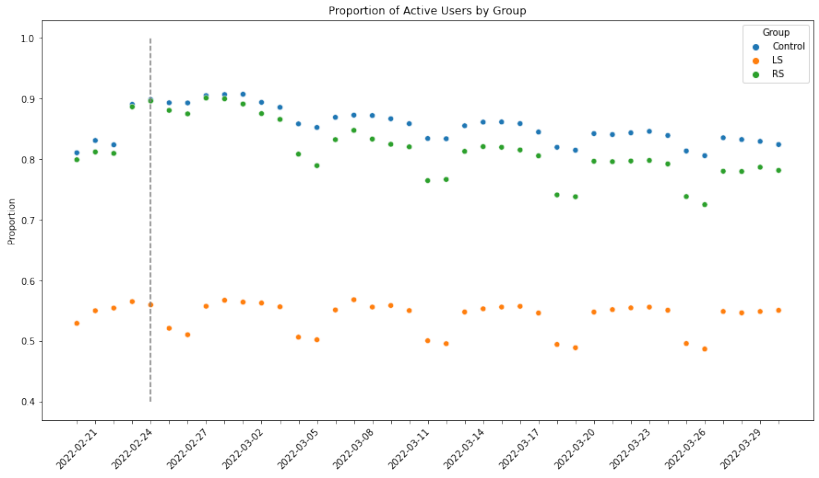
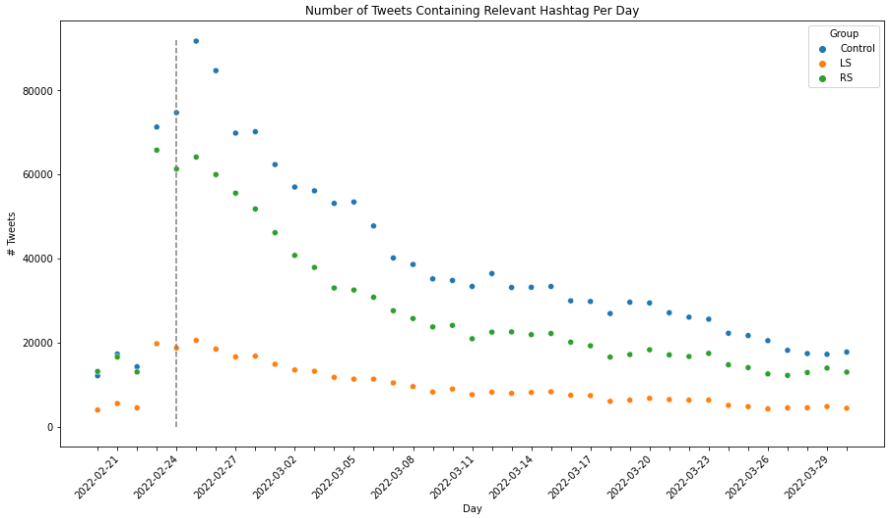


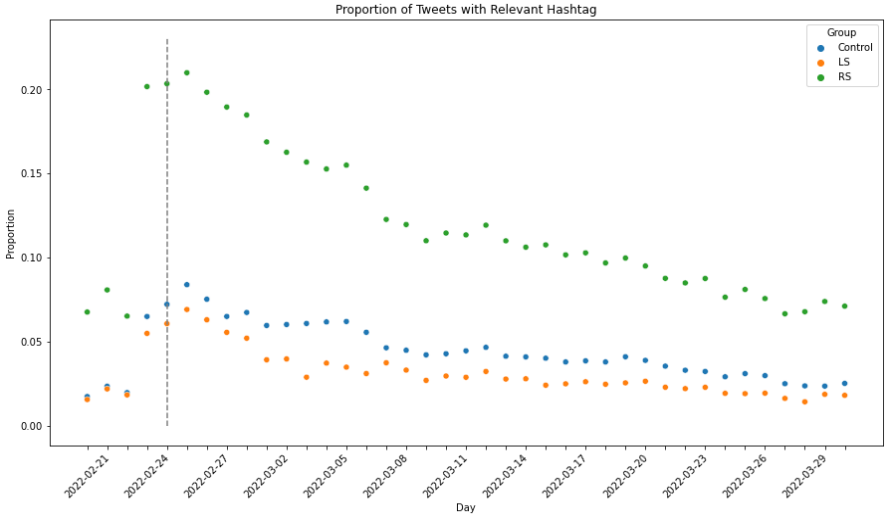
Figure 5-3: Proportion of active users per day by group

Relevant Activity

While total tweet volume is relevant for understanding the quantity of information produced by each sample, we are really only concerned with tweets discussing topics related to the Russia-Ukraine war. Just because the control group tweets more often, doesn't necessarily mean that the tweets are relevant. In fact, Figure 5-4a shows that although the retweet sensor group had lower levels of total activity, it rivaled the control group in number of tweets related to the conflict. From an efficiency standpoint, the retweet sensor group consistently had the largest proportion of tweets containing hashtags related to the Russia-Ukraine conflict – as shown in Figure 5-4b.



(a) Volume of relevant tweets per day



(b) Proportion of tweets with relevant hashtags per day

Figure 5-4: Participation in relevant hashtags per day by group.

This simple analysis of activity levels shows that a group of prior retweet connections shares more relevant content than a group of local sensors and rivals the amount shared by a control group – both of which are nearly three times the size. Having a group that generates fewer total tweets while still sharing comparable amounts of relevant information is a valuable feature for efficient information detection on a platform. In the

sections below I explore the range of information captured in these on-topic tweets and outline a framework that can be used to characterize the benefit of observing activity among these groups of sensors.

5.4 Value Quantification of Sensor Participation

Prior evaluations of sensing mechanisms do not provide clear guidance on how to quantify the gain derived from information found in sensor groups. In the section below, I start by using a simple, count-based approach to understand the prevalence of widespread hashtags captured by each of the groups. I then outline a predictive modeling framework that presents a novel, systematic approach to evaluating the value afforded by sensors. This framework seeks to answer the question of whether sensor participation truly increases our understanding of overall sharing on the platform.

5.4.1 Data Pre-Processing

I use the same approach from previous chapters to identify and extract hashtags related to the Russia-Ukraine conflict from tweets in all three groups. All hashtags are once again lower-cased to account for the case-insensitivity of Twitter's Count API (used to collect total tweet counts).

To eliminate hashtags that have been circulating on Twitter before the captured window of activity, I restrict the group of conflict hashtags to those that were "born" after February 28, 2022. Given we only have information on the total tweet counts starting January 1, 2022 – a hashtag was included in the final set if it was not used between January 1 and February 28. Additionally, I remove hashtags whose first occurrence on the platform was after March 23, 2022, to ensure that we are able to observe activity of all hashtags for at least one week. The final data set includes 5,232 unique conflict-related hashtags – 3,773 from the control group, 990 from the local sensor group and 1,953 from the retweet sensor group.

5.4.2 Hashtag Use by Group

One simple way to evaluate the value of these sensor groups is to look at how many hashtags were shared by each of the groups and investigate what proportion eventually become widely shared on Twitter.

Of the 3,773 hashtags found in the control group, 837 were shared by the retweet sensor group (43% of all hashtags in that group) and 571 were shared by the local sensor group (58% of hashtags in the group). Figure 5-5 demonstrates the relative proportion of unique hashtags that were shared by each group as well as the fraction of hashtags that were also shared in others. Clearly a large percentage of hashtags shared came only from the control group – meaning they were "missed" by both groups of sensors. The question is, how popular were these hashtags?

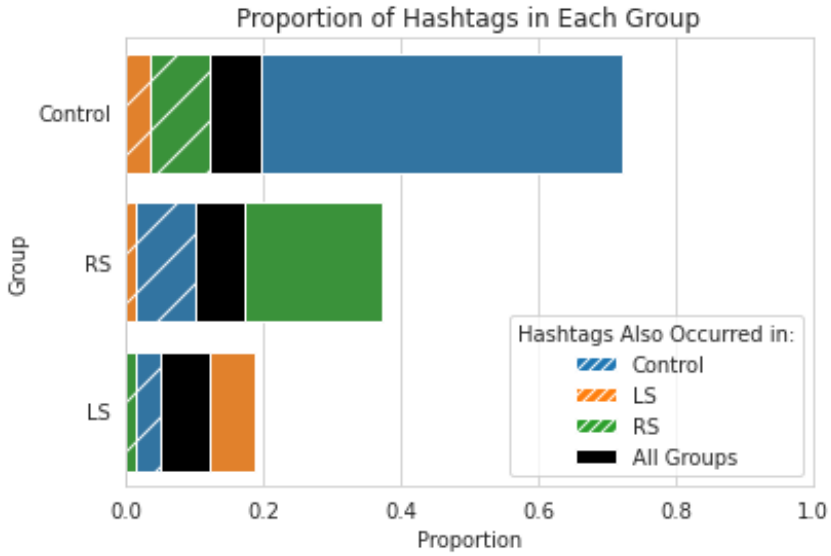


Figure 5-5: Proportion of total hashtags within each group and co-occurrence across groups.

Figure 5-6 shows that over 85% of hashtags shared by only the control group received less than 20 shares. Out of the 494 observed hashtags that eventually received more than 100 shares, only 91 (18%) of them were "missed" by the sensor groups. Of importance is

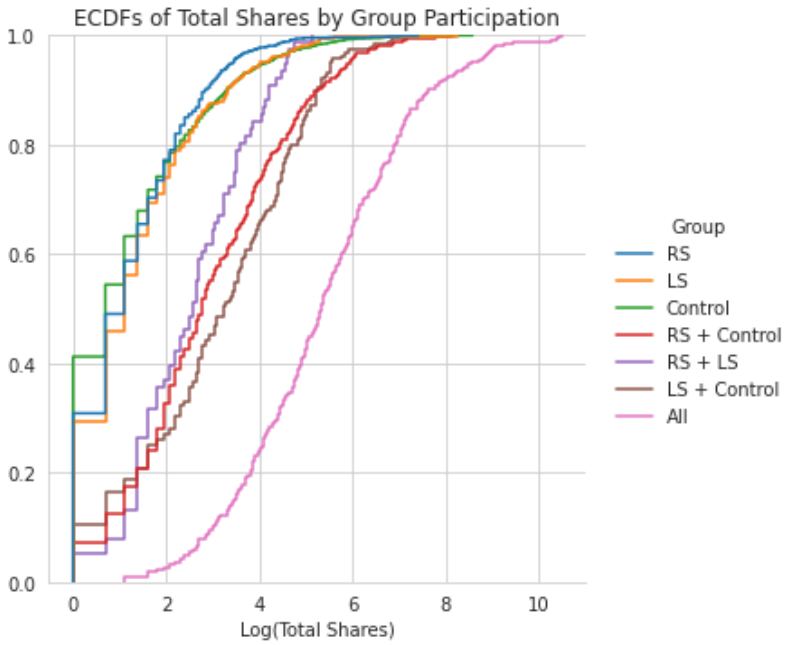


Figure 5-6: Log of total shares after 3 weeks by group participation.

that these 91 hashtags account for only 3.3% of the 2748 hashtags that were shared in the control group only. This suggests that of the hashtags that sensors fail to "detect", a majority of them are inconsequential. Additionally, it shows that activity in the control group is extremely noisy – there are a lot of hashtags present that are never shared more than 2 or 3 times. Hashtags contained to only the local or retweet sensor group also exhibited very similar behavior. Only 12 hashtags from each population were eventually shared more than 100 times – accounting for 3% and 1% of hashtags in each isolated group respectively.

Looking at the two-group interactions – the right shift of all three curves indicates that if a hashtag has been shared by at least two of the groups, there is an increased probability of it being shared a greater number of times. Intuitively, this makes sense. What is interesting however, is that the interaction between the control group and either sensor group signals an increased probability for larger sharing – more so than interaction between sensor groups. Only 3 of the 76 hashtags (4%) that occurred in both sensor groups were shared over 100 times. Whereas 41 of the 188 hashtags (22%) in both the control and local sensor group and 83 of the 454 hashtags (18%) in both the control and

retweet sensor group were shared over 100 times. This confirms that having knowledge of participation in groups of varying centralities may be an effective proxy of information diffusion across a network. It also suggests that the value of sensors may be amplified when accompanied by knowledge of activity in a control group.

What if a hashtag was shared by all three groups? The clear right shift in the ECDF for all groups illustrates that over 65% of hashtags were shared more than 100 times. Of the 494 conflict related hashtags that eventually received greater than 100 shares, 252 (51%) were used by all three groups. Taking it one step further – 117 of the 157 hashtags (75%) that were shared more than 500 times were used by all three groups.

There are two ways to view the outcome of this analysis. First, if we made the decision to only monitor activity in the sensor groups – which would be on par with insinuations from prior work – we could cut the total amount of data collected by nearly half, decrease the overall number of hashtags present from 5,232 to 2,484 (a 52% decrease), and only miss out on 18% of widespread hashtags. 403 of the 494 hashtags that received more than 100 shares would have been captured without the control group. With only a single sensor group, the number of widespread hashtags detected decreases – to 61% for local sensors and to 70% for retweet sensors.

This is validation that using both friend and retweet connections as points of monitoring could help strategically minimize the amount of data that must be collected while only sacrificing ~20% of widespread topics. If we were to select a single sensor group, retweet sensors capture a larger number of conflict-related hashtags, with around 4 million fewer total tweets than local sensors – making them an appealing choice. Additionally, there is much more overlap between hashtags in the control and retweet sensor group than there is between the control and local sensor group. This suggests that in this context, retweet connections may be more informative of behavior in a control group than randomly selected following edges. Note: Due to the scope of data collection for this analysis, we do not know the number of conflict-related hashtags that were not captured by any of the three groups. This means we have no way to quantify a "true" false negative rate of the sensor groups. However, these relative comparisons are still beneficial.

An alternate interpretation to these results is – maybe there is value in observing all three groups simultaneously, instead of simply forgoing information found in the control group. As shown above, hashtags used by multiple groups had an increased probability of becoming widespread. So, instead of disregarding the control group, what if we only look at hashtags that were present in at least two of the three groups? Granted, in doing so, we would not reduce the total amount of data collected – one of the primary objectives for identifying sensors. However, such an approach would decrease the set of hashtags among collected data that need to be evaluated from 5,232 to 1,101 (a 79% decrease), while still capturing 376 (76%) of all hashtags that become widespread.

Given we do not have knowledge of the entire sharing landscape on Twitter, the detection rate of sensors found here is likely an upper bound. Regardless, this simple approach to evaluating sensors shows that if they would have been pursued as sole points of monitoring on Twitter (instead of the sampled control group), we would have been able to decrease the total amount of data collected by 61% while still observing nearly 80% of relevant, widespread hashtags. If we would have selected only the retweet sensor group as our points of monitoring, we would have been able to decrease total data collected by over 84% while still observing 70% of relevant, widespread hashtags. Results also suggest that, if resources allow, there may be value in monitoring both the control and sensor groups, to take advantage of the signal sent by hashtags that occur in multiple groups.

5.4.3 Predictive Modeling Framework

Prior work has sought to predict cascade size in online networks after observing a complete picture of early behavior – showing that temporal and structural features of a cascade are key indicators for future growth (Cheng et al., 2014). The inherent challenge with being forced to sample information from a platform is that we are unable to obtain a complete picture of what is going on. Given methods for strategically sampling groups that have been outlined in this thesis, I seek to understand if simple indicators derived from participation by local and retweet sensor groups improve our ability to predict the popularity of a hashtag.

Similar to the approach taken in Cheng et al., I frame this problem as a classification task where the ultimate goal is to predict whether a hashtag will grow above a certain threshold of total shares within three weeks of its initial observation in one of the three groups.

I propose two distinct, yet beneficial approaches to this task, that can serve as a simple groundwork for future efforts to build upon. For the first approach I ask – given we see a set of hashtags in a randomly sampled control group – does having information about whether that hashtag also appeared in a sensor group increase our ability to predict its overall popularity? This echoes the methodology used in prior literature (and previous chapters in this thesis) to evaluate the lead time afforded by a group of sensors. Namely, we observe a set of hashtags in a random control group – for this set of hashtags, would we have benefited from information provided by a sensor group?

For the second approach, I take a more holistic route that evaluates hashtags present across all three groups. As shown above, not every relevant hashtag is captured by the control group. So, it is important to ask how expanding the window of information considered affects our ability to predict future popularity.

5.4.4 Methods

Control Centric Approach

I restrict the population of hashtags to the 3,773 that were shared among the control group and randomly assign hashtags to training and test sets using an 80-20 split. I use three separate thresholds for total number of shares received based on measures of position across all hashtags observed – 5 (just above the median of total shares received), 15 (the top quartile of total shares) and 90 (the 90th percentile of total shares). The median number of total shares across hashtags was 4, so to account for class imbalance present in each of the training sets, I down-sample the "below X shares" class to equal that of the "above X shares" class. Logistic regression is used to predict whether the total number of shares of a hashtag will grow above the predefined threshold. Table 5-1 outlines the features considered for this prediction task.

Holistic Approach

For the second approach, I expand the set of hashtags to include all 5,232 shared at least once by any of the three groups and then use random assignment to create training and test sets. I once again use three separate thresholds for the total number of shares received – 5, 15, and 90. Median number of total shares when all hashtags are included drops down to 3, so the "below X shares" class was again down-sampled to equal that of the "above X shares" class. The same predictor variables as listed in Table 5-1 are used. However, because all hashtags are included in this approach – not only those that occurred in the control – I also include a simple binary indicator for participation in the control. Namely: $control_h = 1$ if hashtag h appeared in the control group, 0 otherwise.

Participation Indicators

$local_h$	1 if hashtag h appeared in local sensor group, 0 otherwise
$rtwt_h$	1 if hashtag h appeared in retweet sensor group, 0 otherwise

Within Group Temporal Features

$K2r_{g,\,h}$	time between the first and second adoption of hashtag h in group g : If only 1 user participates driven to very large value

Between Group Temporal Features

$local\_\Delta tU_{1,h}$	Difference in time between first adoption of hashtag h in control group and first adoption in local sensor group: If no user participates in the local sensor group, driven to a very large value
$rtwt\_\Delta tU_{1,h}$	Difference in time between first adoption of hashtag h in control group and first adoption in retweet sensor group: If no user participates in the retweet sensor group, driven to a very large value

Table 5.1: List of features used for learning

For the first approach, the relative increase in predictive performance when basic indicators for participation in sensor groups are included will help directly quantify what we gain from having sensors. For the second approach, performance achieved can be used as a benchmark for sampling approaches being evaluated in the future.

The predictors included for each classification task are used to characterize early participation in each of the groups. Framing features in this manner makes it more conducive for use in an applied setting where we are interested in identifying topics of consequence as soon as possible. Such an approach is also fairly robust to the sample sizes used for group selection as well as to late adopters of any hashtag. We only use a simple binary indicator of participation – which does not rely on the magnitude of participation in any group, the difference in adoption between the first user in each group – which as described before is robust to late adopters in a given group, and the rate of adoption between the first two users in a given group – which helps describe how quickly a hashtag is spreading between users without requiring a picture of the full adoption curve.

5.4.5 Results

Control Centric Approach

Figure 5-7 shows the accuracy and F1 score for each logistic regression model when used to make predictions on a held-out test set. Given the class imbalance present, we are generally more interested in the F1 score than we are in overall accuracy because by simply classifying every hashtag as "below threshold", we could achieve an accuracy of over 50%, 75% or 90% for the three sharing thresholds respectively.

Results from the control-centric logistic regressions show that using only the time between adoption of the first two users in a control group, we can achieve an F1 score of nearly .62 (for a threshold of 5 shares) and .68 (for threshold of 15 and 90 shares). We can think of this as a baseline performance for this set of regressions or what we could achieve without having sensors. Using only indicators for presence in the retweet sensor group or local sensor group – the F1 score decreased by 8-12%. In an ideal world – sensor participation alone would be enough to signal that a hashtag was going to be

shared a large number of times. The various complexities of information sharing unfortunately make this a very unlikely outcome, so the decrease in performance relative to using only temporal features from the control group is not an unexpected result.

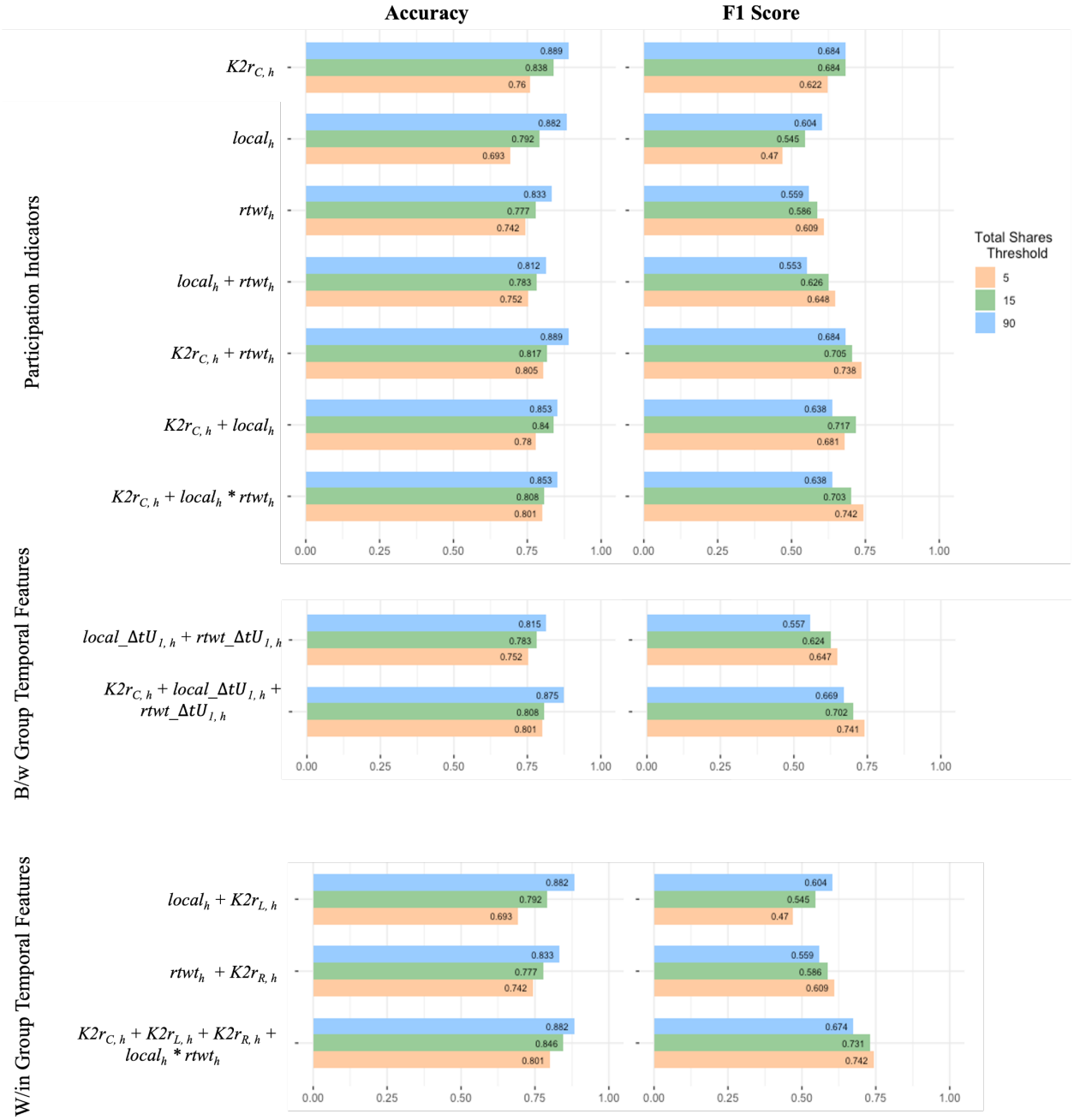


Figure 5-7: Accuracy and F1 score models including participation indicators, between group temporal features, and within group temporal features

Using $K2r_{C,h}$ as well as local and retweet indicators, the F1 score increases from the baseline by ~2% for a threshold of 15 shares and by almost 12% for a threshold of 5 shares. Performance decreased for the higher threshold of 90 shares, which suggests that sensor participation is beneficial for distinguishing between those hashtags that will be shared very few times and those that will receive a medium amount of attention, but not as useful for identifying which hashtags are going to become widely shared.

A very similar outcome was obtained when incorporating $local\_\Delta tU_{1,h}$ and $rtwt\_\Delta tU_{1,h}$ with $K2r_{C,h}$. The similarities in predictive performance between models that include simple participation indicators and those that include lead times further reinforces findings from Chapter 4, that the heavily touted lead time metric is not very useful for identifying which hashtags are going to become widespread (those that we would be more inclined to pay attention to).

The final model that incorporates the rate of sharing for all three groups outperforms all other combinations of predictors for thresholds of both 5 and 15 shares – obtaining F1 scores .12 and .07 higher than the baseline. Similar work done to predict whether a cascade will reach above a median 10 shares after directly observing the first 5 shares achieved an F1 score of .795 (Cheng et al., 2014). With simple information from samples of a population (not the complete picture) we were able to attain a comparable F1 of .73. This reinforces that drawing conclusions about activity across a platform is still feasible even with small samples of information and that sensors do in fact add value to our predictive capabilities.

Holistic Approach

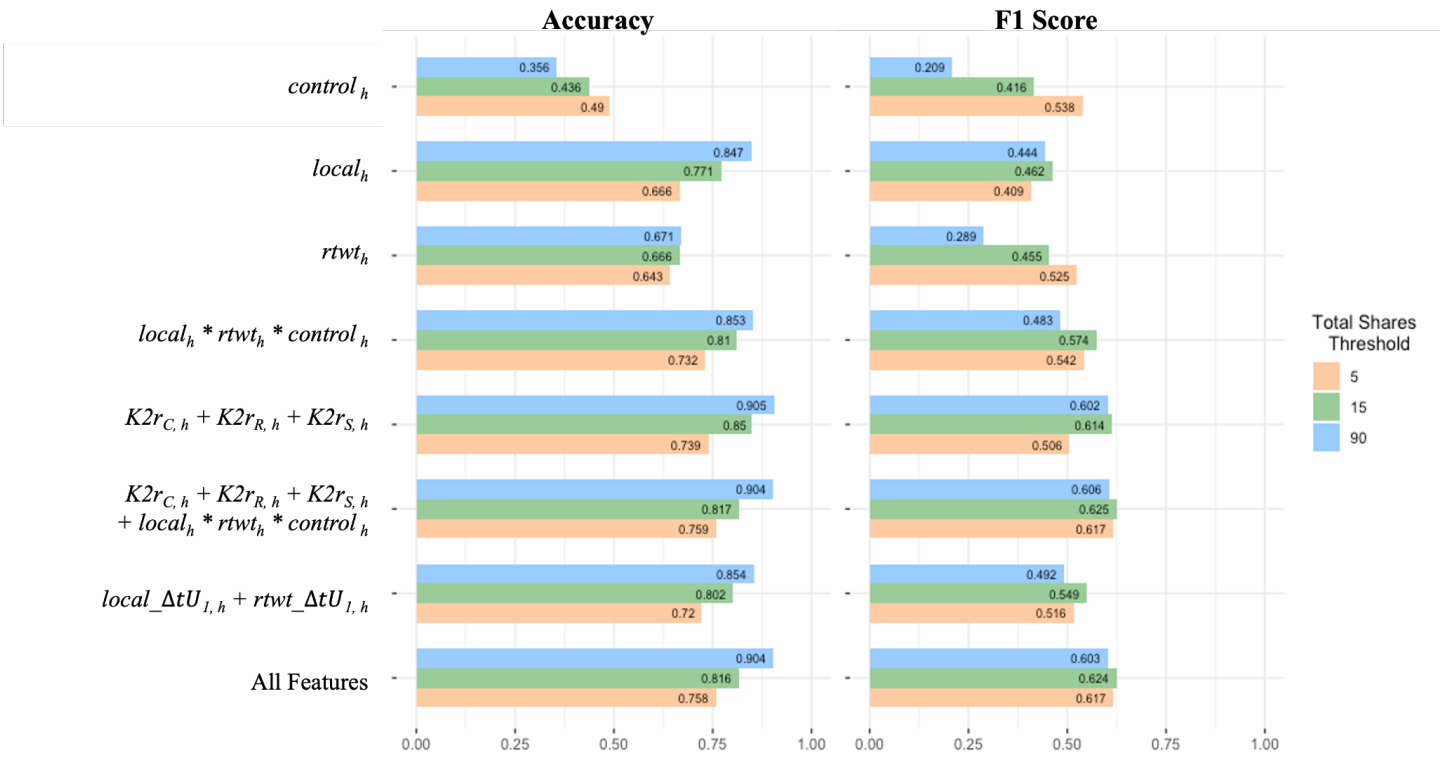


Figure 5-8: Accuracy and F1 score of logistic regression models fit on all hashtags.

Performance decreased across the board when all hashtags were included in training/testing. It is very clear by the low F1 scores for the first 3 models that participation alone by any single group is not a strong indicator of future growth. Given the number of hashtags that only occur in a single group (as discussed in section 5.4.2), it is not surprising that participation indicators do not perform well when used as univariate predictors.

The highest level of performance is achieved with a feature set that includes rate of sharing in all three groups as well as indicators of participation for all groups. F1 scores for thresholds of 5, 15, and 90 shares were .617, .625, and .606 respectively. These results are not groundbreaking and are likely limited in part by the smaller sample size that was used for collection. However, the approach provides a concrete evaluation of our ability to understand information sharing on the platform using strategically selected

samples. As work in this field evolves, I argue that using performance from predictive models – similar to those presented above – to compare viable sensor options is a much more robust process than others that have been used in recent years.

5.5 Discussion

While admittedly simple, the above analyses provide useful frameworks for objectively assessing sensors beyond only the lead time that they provide. Evaluating true positive and false negative rates associated with each sensor group allows us to gauge how confidently we can deploy sensors as our sole points of observation. Such metrics also highlight trade-offs between detected information and data collected. For this set of groups, by following only the local and retweet sensors, we would have been able to reduce data collected by over half while still capturing a majority of hashtags that are shared more than 100 times. Retweet sensors proved to be more effective than local sensors as they would have captured a larger number of widespread hashtags and also had a larger overlap with sharing in the control group.

Additionally, outcomes of the control-centric predictive models show that observing hashtag use in sensor groups improves our ability to predict the number of shares it will eventually receive. While performance declined when we consider all hashtags, the framework employed provides a robust approach for future efforts to expand upon as well as a benchmark level of performance to be used for comparison throughout future iterations.

Chapter 6

Discussion and Conclusion

6.1 Conclusion

Rapid proliferation of social media has permanently altered the landscape of global communication. Reliance on such platforms shows no sign of slowing down, which creates the unprecedented challenge of efficiently sampling the inordinate volume of information being shared in order to understand conversations circulating online.

In this thesis, I evaluated the efficacy of randomly sampled one-hop connections as sensors on Twitter during the Russia-Ukraine conflict. My findings show that the influx of information sharing via external channels decreased the correlation between activity and degree of users. Consequently, 'friends as sensors' would not have reliably provided early awareness of discussions related to the Russia-Ukraine conflict. When accounting for potential biases in prior framings of lead time calculations, following connections were found to be even less beneficial. Additionally, there is a weak, negative correlation between the lead time provided by sensors and the future popularity of a hashtag.

Further, results show that using the non-network signal of early participation in conflict discussions may be a viable approach to actively selecting control groups in future settings. Both local and retweet sensors effectively diminished the total volume of information shared, relative to a control group, while still capturing a majority of hashtags that became widely shared. Although, retweet sensors proved to be more beneficial than local sensors for efficiently detecting widespread information.

Finally, I demonstrated that by framing sensor evaluations as a predictive modeling problem, we can systematically quantify the value afforded by sensors and provide clear benchmarks of performance for future investigations.

While there are still many questions to be asked about what we can truly infer from users' posts on social media, awareness of online activity is still valuable for many policy-relevant contexts. Therefore, it is vital that we continue to study methods for efficient information detection on online platforms.

6.2 Limitations and Biases

While the above analysis provides one of few empirical evaluations of sensors in social networks, it is important to consider the biases and limitations that are present throughout.

Retrospective collection of information on edges of the network. Similar to many other types of networks – such as those seen among human relationships in the physical world – edges connecting users on Twitter fluctuate over time. The snapshot of the network used for this analysis was taken several months after tweets had been created – leaving room for new connections to be formed or old connections to be removed prior to collection taking place. In an ideal setting, all of this information would have been contemporaneously obtained with tweet metadata.

Tweet Deletion. It is important to consider the implications of the percentage of tweets that were able to be hydrated from the original data set. As previously mentioned, if tweets have been deleted from Twitter – either by the user or the platform – there is no way to access the content of the tweet, even with the original tweet ID. After hydrating the tweet IDs provided by Shevtsov et al., only 75% of the tweets were still accessible on the platform. While it would be advantageous for the external validity of these results if the missing tweets were deleted at random – it is highly unlikely that this was the case.

Realistically, the population of tweets remaining at the time of rehydration is biased towards information that is not explicitly harmful, hateful, or misleading.

This bias may impact the generalizability of results. Claims have been made that online information of different veracity diffuses at different paces through a network (Vosoughi et al., 2018). Recent work has challenged these initial assertions, finding that differences in spread can be attributed to the "infectiousness" of information rather than structural differences in diffusion (Juul & Ugander, 2021). Generalizing these results to all types of information assumes that they diffuse through a network in a similar fashion. However, the jury is still out on whether this is truly the case.

Overall, it is vital that those who wish to make future contributions to the empirical side of sensing seek to establish a population of interest at the outset and stream/live collect on information being shared to get the most accurate representation of the dynamics of online sharing.

Hashtags as unit of observation: Hashtags are "simply a keyword that assigns information to categories to help users retrieve it" which affords a relatively simple, objective way to track discourse on social media (Milan, 2015). However, in the comprehensive Twitter collection done in Garcia et al., only 14% of roughly 466 million tweets contained a hashtag (Garcia-Herranz et al., 2014). By conditioning collection on use of hashtags, the data set used for analyses is not a comprehensive set of discussions related to the Russia-Ukraine war on Twitter. Looking strictly at hashtag engagement is very likely to misrepresent overall engagement because after everyone is aware of a topic, a hashtag can be viewed as "superfluous and wasteful on the character-limited Twitter platform" and additionally hashtags are often seen as an avenue "only as useful for attracting attention to a particular topic, not for talking about it" (Tufekci, 2014).

An additional consideration for hashtag-based collection is that tweets within this set are "included because the user chose to use it" which is "a clear act of self-selection" (Tufekci, 2014). The sharing tendencies of those that use hashtags may be characteristically different than those that choose not to. While this does not render the work done in this thesis uninformative, it does mean that "safe inferences must be limited to those in the sample" (Keiding & Louis, 2016). Additional evaluations of discussed

sensing mechanisms need to be done – beyond strictly hashtag use – to gain a more holistic understanding of their capabilities and validity.

Twitter. The accessibility of Twitter has made it a haven for academic researchers trying to understand phenomena within social networks. However, publications using data from the platform have been met with criticisms of biased outcomes (Olteanu et al., 2019; Tufekci, 2014). Twitter offers (offered) multiple access points for researchers, but there is a lack of transparency in the exact sampling mechanisms used to produce outcome sets which can lead to unidentified biases in subsequent analyses. For example, Morstatter et al. show that while advertised as "random" samples, returns from the Twitter Streaming API are not always representative of the entire population of activity (Morstatter et al., 2013). General consensus now states that APIs should be "regarded as an unavoidable 'black box'" (Pfeffer et al., 2018).

Beyond potential unknown biases present due to collection methods and opaque sampling procedures, Twitter also differs structurally from other platforms in aspects that could influence the outcome of sensing approaches. For example, Twitter is a directed network with non-mutual ties created through the action of "following" whereas "friending" connections in a platform such as Facebook are mutual. Additionally, while Twitter is a micro-blogging platform dominated by short, text-based posts, a variety of other platforms are designed for sharing in non-text-based mediums such as videos (TikTok) and photos (Instagram). Other platforms, such as Telegram, contain direct messaging and both public and private groups and channels. Empirical analyses across these different types of platforms and across different mediums (beyond only text-based monitoring) are requisite before results found in this thesis (or other related works) can be confidently generalized to "social media platforms" as a whole. This is a clear pathway forward for future work in the field of sensing.

6.3 Directions for Future Work

As hinted at above, a natural extension to the work done in this thesis is to develop and evaluate peer-based sensing mechanisms on platforms beyond Twitter. Especially in light of all of the recent changes being made to Twitter's platform under the control of Elon Musk – there is no guarantee that Twitter will remain a prevalent (or accessible) social media platform in the coming years. Additionally, other platforms, such as Telegram, have become hotspots for harmful or misleading content due to notoriously lax moderation policies (Ghasiya & Sasahara, 2022). Methods for strategically collecting data to gain an understanding of current conversations on platforms beyond Twitter will be a key component to combating extremist, polarizing, and misleading activity in the future.

Another extension to this research would be to conduct a similar analysis using natural language processing techniques to evaluate the semantic content of posts (instead of conditioning all analyses on hashtags). A topic-based approach would require proper attention be paid to attaining a random, representative sample of activity – which as discussed above is difficult in and of itself. However, if achieved, such an analysis would allow us to better understand if friendship or retweet connections can provide early awareness of specific discussions during times of conflict – even though they proved to be ineffective for hashtag utilization.

Overall, this thesis reinforces the need for continued empirical evaluations of potential sensing mechanisms across diverse contexts.

Appendix A

Extra Tables and Figures

#Ukraine #Ukraina #ukraina #Украина #Украине #Россия #Russia #Putin #Война #ЯпротивВойны #WWIII #worldwar3 #Ukraine_Russia #Russian_Ukrainian #UkraineRussiaWar #UkraineRussie #RussiaInvadedUkraine	#RussiaUkraineConflict #RussiaUkraineCrisis #BoycottRussia #PrayForUkraine #solidarityWithUkraine #StandWithUkraine #BlockPutinWallets #StopPutin #StopRussianAggression #StopRussia #StandWithUkraineNOW #StopWar #SWIFT #NATO #FUCK_NATO #FuckPutin #PutinWarCriminal	#PutinHitler #with_russia #StopNazism #myfriendPutin #UnitedAgainstUkraine #ВпередРоссия #ЯМыРоссия #ВеликаяРоссия #Путинмойпрезидент #россиявперед #россиявперёд #ПутинНашПрезидент #ЗаПутина #Путинмойпрезидент #ПутинВведиВойска #СЛАВАРОССИИ #СЛАВАВДВ

Table A.1: Hashtags used as filters for the Twitter Streaming API during collection by Shevtsov et al. (2022)

Group Size (thousands of users)	**Number of Samples**
1	20
3	20
5	20
10	20
15	15
20	12
25	10
30	8
35	7
40	6
45	5
50	5

Table A.2: Table of sample size and associated number of samples for analysis of sensor method with small samples

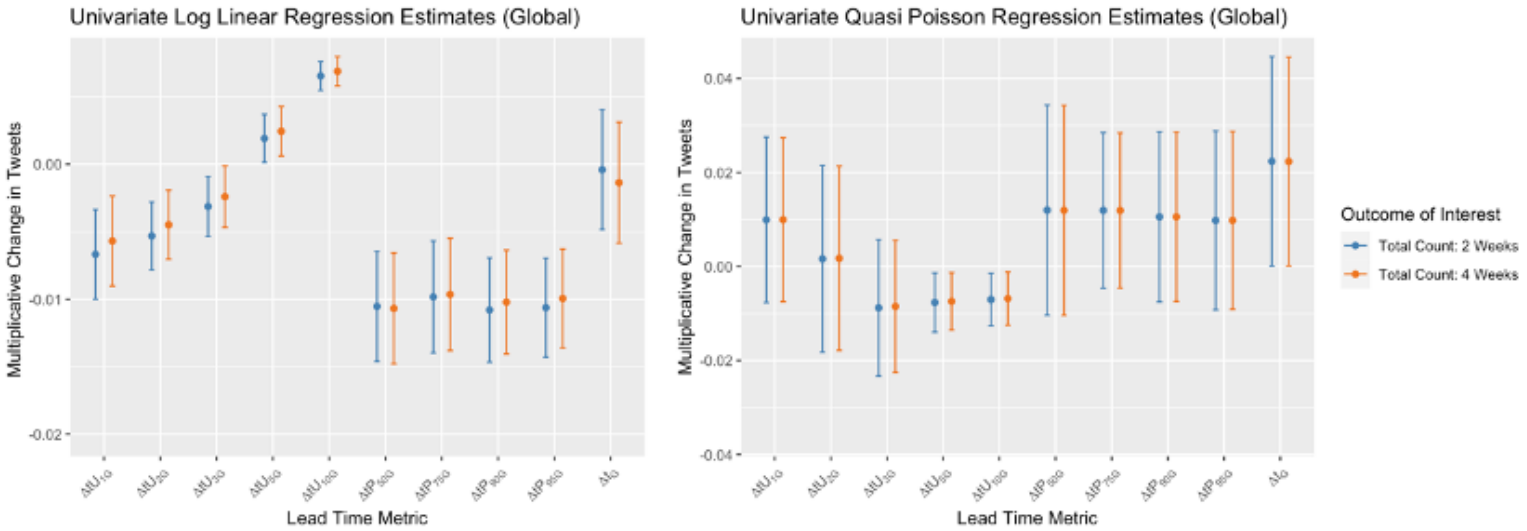


Figure A-1: Univariate regression coefficient estimates for global sensors.

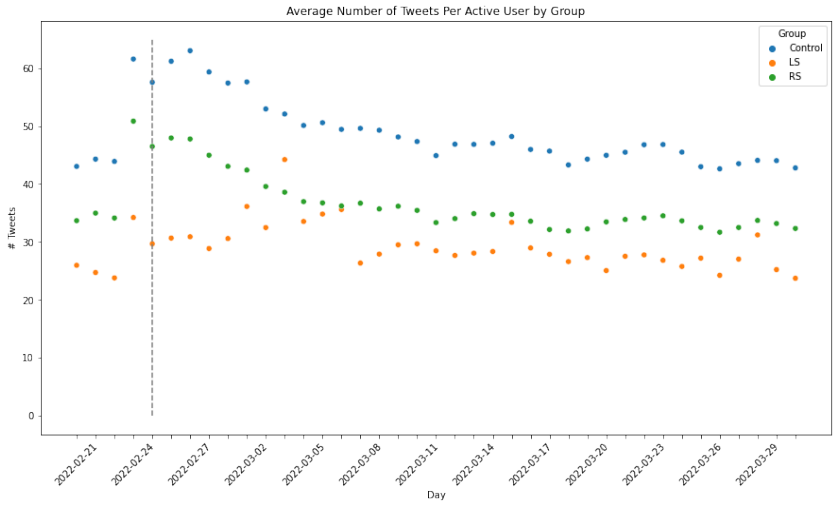


Figure A-2: Average number of tweets per active user by group

Bibliography

22 Essential Twitter Statistics You Need to Know in 2023. (n.d.). The Social Shepherd. Retrieved March 7, 2023, from https://thesocialshepherd.com/blog/twitter-statistics

A historical timeline of post-independence Ukraine. (2022, February 22). PBS NewsHour. https://www.pbs.org/newshour/world/a-historical-timeline-of-post-independence-ukraine

Asur, S., & Huberman, B. A. (2013). Predicting the Future with Social Media. *Applied Energy*, *112*, 1536–1543. https://doi.org/10.1016/j.apenergy.2013.03.027

Babichev, S., & Lytvynenko, V. (Eds.). (2022). *Lecture Notes in Computational Intelligence and Decision Making: 2021 International Scientific Conference "Intellectual Systems of Decision-making and Problems of Computational Intelligence", Proceedings* (Vol. 77). Springer International Publishing. https://doi.org/10.1007/978-3-030-82014-5

Bagavathi, A., & Krishnan, S. (2019). Social Sensors Early Detection of Contagious Outbreaks in Social Media. In T. Z. Ahram (Ed.), *Advances in Artificial Intelligence, Software and Systems Engineering* (pp. 400–407). Springer International Publishing. https://doi.org/10.1007/978-3-319-94229-2_39

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2019). Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials. *The Review of Economic Studies*, *86*(6), 2453–2490. https://doi.org/10.1093/restud/rdz008

Barberá, P., Wang, N., Bonneau, R., Jost, J. T., Nagler, J., Tucker, J., & González-Bailón, S. (2015). The Critical Periphery in the Growth of Social Protests. *PLOS ONE*, *10*(11), e0143611. https://doi.org/10.1371/journal.pone.0143611

Batlle, P., Bruna, J., Fernandez-Granda, C., & Preciado, V. M. (2020). *Adaptive Test Allocation for Outbreak Detection and Tracking in Social Contact Networks* (arXiv:2011.01998). arXiv. http://arxiv.org/abs/2011.01998

Behnassi, M., & El Haiba, M. (2022). Implications of the Russia–Ukraine war for global food security. *Nature Human Behaviour*, *6*(6), Article 6. https://doi.org/10.1038/s41562-022-01391-x

Biersack, J., & O'Lear, S. (2014). The geopolitics of Russia's annexation of Crimea: Narratives, identity, silences, and energy. *Eurasian Geography and Economics*, *55*(3), 247–269. https://doi.org/10.1080/15387216.2014.985241

Bild, D. R., Liu, Y., Dick, R. P., Mao, Z. M., & Wallach, D. S. (2015). Aggregate Characterization of User Behavior in Twitter and Analysis of the Retweet Graph. *ACM Transactions on Internet Technology*, *15*(1), 1–24. https://doi.org/10.1145/2700060

Chami, G. F., Ahnert, S. E., Kabatereine, N. B., & Tukahebwa, E. M. (2017). Social network fragmentation and community health. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(36), E7425–E7431. https://doi.org/10.1073/pnas.1700166114

Cheng, J., Adamic, L. A., Dow, P. A., Kleinberg, J., & Leskovec, J. (2014). Can Cascades be Predicted? *Proceedings of the 23rd International Conference on World Wide Web*, 925–936. https://doi.org/10.1145/2566486.2567997

Chin, A., Eckles, D., & Ugander, J. (2022). Evaluating Stochastic Seeding Strategies in Networks. *Management Science*, *68*(3), 1714–1736. https://doi.org/10.1287/mnsc.2021.3963

Christakis, N. A., & Fowler, J. H. (2010). Social Network Sensors for Early Detection of Contagious Outbreaks. *PLoS ONE*, *5*(9), e12948. https://doi.org/10.1371/journal.pone.0012948

Ciuriak, D. (2022). *The Role of Social Media in Russia's War on Ukraine* (SSRN Scholarly Paper No. 4078863). https://doi.org/10.2139/ssrn.4078863

Cohen, R., Havlin, S., & ben-Avraham, D. (2003). Efficient Immunization Strategies for Computer Networks and Populations. *Physical Review Letters*, *91*(24), 247901. https://doi.org/10.1103/PhysRevLett.91.247901

Di Minin, E., Fink, C., Hausmann, A., Kremer, J., & Kulkarni, R. (2021). How to address data privacy concerns when using social media data in conservation science. *Conservation Biology*, *35*(2), 437–446. https://doi.org/10.1111/cobi.13708

Eckles, D., Esfandiari, H., Mossel, E., & Rahimian, M. A. (2022). Seeding with Costly Network Information. *Operations Research*, *70*(4), 2318–2348. https://doi.org/10.1287/opre.2022.2290

Enikolopov, R., Makarin, A., & Petrova, M. (2015). Social Media and Protest Participation: Evidence from Russia. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2696236

Feld, S. L. (1991). Why Your Friends Have More Friends Than You Do. *American Journal of Sociology*, *96*(6), 1464–1477. https://doi.org/10.1086/229693

Futey, B. A. (1996). Comments on the Constitution of Ukraine Special Report. *East European Constitutional Review*, *5*(Issues 2 & 3), 29–34.

Garcia-Herranz, M., Moro, E., Cebrian, M., Christakis, N. A., & Fowler, J. H. (2014). Using Friends as Sensors to Detect Global-Scale Contagious Outbreaks. *PLoS ONE*, *9*(4), e92413. https://doi.org/10.1371/journal.pone.0092413

Geissler, D., Bär, D., Pröllochs, N., & Feuerriegel, S. (2023). *Russian propaganda on social media during the 2022 invasion of Ukraine* (arXiv:2211.04154). arXiv. http://arxiv.org/abs/2211.04154

Ghasiya, P., & Sasahara, K. (2022). *Messaging Strategies of Ukraine and Russia on Telegram during the 2022 Russian invasion of Ukraine* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-2288409/v1

Gill, T. D. (2022). The Jus ad Bellum and Russia's "Special Military Operation" in Ukraine. *Journal of International Peacekeeping*, *25*(2), 121–127. https://doi.org/10.1163/18754112-25020002

Golovchenko, Y. (2020). Measuring the scope of pro-Kremlin disinformation on Twitter. *Humanities and Social Sciences Communications*, *7*(1), Article 1. https://doi.org/10.1057/s41599-020-00659-9

Hinz, O., Skiera, B., Barrot, C., & Becker, J. U. (2011). Seeding Strategies for Viral Marketing: An Empirical Comparison. *Journal of Marketing*, *75*(6), 55–71. https://doi.org/10.1509/jm.10.0088

Iffy.news. (n.d.). Iffy.News. Retrieved April 21, 2023, from https://iffy.news/

Internet and social media users in the world 2023. (2023, February 24). Statista. https://www.statista.com/statistics/617136/digital-population-worldwide/

Jaroszewicz, M., Grzymski, J., & Krępa, M. (2022). The Ukrainian refugee crisis demands new solutions. *Nature Human Behaviour*, *6*(6), Article 6. https://doi.org/10.1038/s41562-022-01361-3

Ghasiya, P., & Sasahara, K. (2022). *Messaging Strategies of Ukraine and Russia on Telegram during the 2022 Russian invasion of Ukraine* [Preprint]. In Review. https://doi.org/10.21203/rs.3.rs-2288409/v1

Juul, J. L., & Ugander, J. (2021). Comparing information diffusion mechanisms by matching on cascade size. *Proceedings of the National Academy of Sciences*, *118*(46), e2100786118. https://doi.org/10.1073/pnas.2100786118

Keiding, N., & Louis, T. A. (2016). Perils and potentials of self-selected entry to epidemiological studies and surveys. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, *179*(2), 319–376.

Kempe, D., Kleinberg, J., & Tardos, É. (2003). Maximizing the spread of influence through a social network. *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 137–146. https://doi.org/10.1145/956750.956769

Kryvasheyeu, Y., Chen, H., Moro, E., Hentenryck, P. V., & Cebrian, M. (2015). Performance of Social Network Sensors during Hurricane Sandy. *PLOS ONE*, *10*(2), e0117288. https://doi.org/10.1371/journal.pone.0117288

Kumar, V., Krackhardt, D., & Feld, S. (2021). *Interventions with Inversity in Unknown Networks Can Help Regulate Contagion* (arXiv:2105.08758). arXiv. http://arxiv.org/abs/2105.08758

May, R. M., Levin, S. A., & Sugihara, G. (2008). Ecology for bankers. *Nature*, *451*(7181), Article 7181. https://doi.org/10.1038/451893a

Media Bias/Fact Check—Search and Learn the Bias of News Media. (n.d.). Retrieved April 21, 2023, from https://mediabiasfactcheck.com/

Milan, S. (2015). *Mobilizing in Times of Social Media. From a Politics of Identity to a Politics of Visibility* (SSRN Scholarly Paper No. 2880402). https://doi.org/10.2139/ssrn.2880402

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). *Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose* (arXiv:1306.5204). arXiv. http://arxiv.org/abs/1306.5204

Myers, S. A., Zhu, C., & Leskovec, J. (2012). Information diffusion and external influence in networks. *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '12*, 33. https://doi.org/10.1145/2339530.2339540

Nettasinghe, B., & Krishnamurthy, V. (2021). Maximum Likelihood Estimation of Power-law Degree Distributions via Friendship Paradox-based Sampling. *ACM Transactions on Knowledge Discovery from Data*, *15*(6), 106:1-106:28. https://doi.org/10.1145/3451166

Nicholas, J., Onie, S., & Larsen, M. E. (2020). Ethics and Privacy in Social Media Research for Mental Health. *Current Psychiatry Reports*, *22*(12), 84. https://doi.org/10.1007/s11920-020-01205-9

Olteanu, A., Castillo, C., Diaz, F., & Kıcıman, E. (2019). Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. *Frontiers in Big Data*, *2*. https://www.frontiersin.org/articles/10.3389/fdata.2019.00013

Osatuyi, B. (2015). Empirical Examination of Information Privacy Concerns Instrument in the Social Media Context. *AIS Transactions on Replication Research*, *1*(1). https://doi.org/10.17705/1atrr.00003

Pfeffer, J., Matter, D., Jaidka, K., Varol, O., Mashhadi, A., Lasser, J., Assenmacher, D., Wu, S., Yang, D., Brantner, C., Romero, D. M., Otterbacher, J., Schwemmer, C., Joseph, K., Garcia, D., & Morstatter, F. (2023). *Just Another Day on Twitter: A Complete 24 Hours of Twitter Data* (arXiv:2301.11429). arXiv. https://doi.org/10.48550/arXiv.2301.11429

Pfeffer, J., Mayer, K., & Morstatter, F. (2018). Tampering with Twitter's Sample API. *EPJ Data Science*, *7*(1), Article 1. https://doi.org/10.1140/epjds/s13688-018-0178-0

Pierri, F., Luceri, L., Jindal, N., & Ferrara, E. (2023). *Propaganda and Misinformation on Facebook and Twitter during the Russian Invasion of Ukraine* (arXiv:2212.00419). arXiv. https://doi.org/10.48550/arXiv.2212.00419

Pifer, S. (2020, March 17). Crimea: Six years after illegal annexation. *Brookings*. https://www.brookings.edu/blog/order-from-chaos/2020/03/17/crimea-six-years-after-illegal-annexation/

Shevtsov, A., Tzagkarakis, C., Antonakaki, D., Pratikakis, P., & Ioannidis, S. (2022). *Twitter Dataset on the Russo-Ukrainian War* (arXiv:2204.08530). arXiv. http://arxiv.org/abs/2204.08530

Signorini, A., Segre, A. M., & Polgreen, P. M. (2011). The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. *PloS One*, *6*(5), e19467. https://doi.org/10.1371/journal.pone.0019467

Smart, B., Watt, J., Benedetti, S., Mitchell, L., & Roughan, M. (2022). #IStandWithPutin Versus #IStandWithUkraine: The Interaction of Bots and Humans in Discussion of the Russia/Ukraine War. In F. Hopfgartner, K. Jaidka, P. Mayr, J. Jose, & J. Breitsohl (Eds.), *Social Informatics* (pp. 34–53). Springer International Publishing. https://doi.org/10.1007/978-3-031-19097-1_3

Stănescu, G. (2022). *Ukraine conflict: The challenge of informational war*. https://doi.org/10.5281/ZENODO.6795674

Sullivan, B. (2022, February 24). Russia's at war with Ukraine. Here's how we got here. *NPR*. https://www.npr.org/2022/02/12/1080205477/history-ukraine-russia

Sun, L., Axhausen, K. W., Lee, D.-H., & Cebrian, M. (2014). Efficient detection of contagious outbreaks in massive metropolitan encounter networks. *Scientific Reports*, *4*(1), Article 1. https://doi.org/10.1038/srep05099

Talabi, F. O., Aiyesimoju, A. B., Lamidi, I. K., Bello, S. A., Okunade, J. K., Ugwuoke, C. J., & Gever, V. C. (2022). The use of social media storytelling for help-seeking and help-receiving among Nigerian refugees of the Ukraine–Russia war. *Telematics and Informatics*, *71*, 101836. https://doi.org/10.1016/j.tele.2022.101836

Thomas, P. B., Saldanha, E., & Volkova, S. (2021). Studying information recurrence, gatekeeping, and the role of communities during internet outages in Venezuela. *Scientific Reports*, *11*, 8137. https://doi.org/10.1038/s41598-021-87473-8

Timeline: Political crisis in Ukraine and Russia's occupation of Crimea | Reuters. (2014, March 8). https://www.reuters.com/article/us-ukraine-crisis-timeline-idUSBREA270PO20140308

Tufekci, Z. (2014). *Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls* (arXiv:1403.7400). arXiv. http://arxiv.org/abs/1403.7400

Twitter API Documentation | Docs | Twitter Developer Platform. (n.d.). Retrieved April 9, 2023, from https://developer.twitter.com/en/docs/twitter-api

Twitter MAU worldwide 2019. (n.d.). Statista. Retrieved March 17, 2023, from https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/

Twitter's Recommendation Algorithm. (n.d.). Retrieved April 18, 2023, from https://blog.twitter.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm

Ukraine—The crisis in Crimea and eastern Ukraine | Britannica. (n.d.). Retrieved February 15, 2023, from https://www.britannica.com/place/Ukraine/The-crisis-in-Crimea-and-eastern-Ukraine

Treaty Series 3007, United Nations Office of Legal Affairs, November 11, 2021, https://doi.org/10.18356/9789214030966

Use Cases, Tutorials, & Documentation. (n.d.). Retrieved March 29, 2023, from https://developer.twitter.com/en

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

War, I. F. the S. of. (2023, March 3). *Interactive Time-lapse: Russia's War in Ukraine*. ArcGIS StoryMaps. https://storymaps.arcgis.com/stories/733fe90805894bfc8562d90b106aa895

Whalen, J., & Dixon, R. (2022, April 5). Russia denies and deflects in reaction to Bucha atrocities. *Washington Post*. https://www.washingtonpost.com/world/2022/04/04/russia-bucha-atrocities-war-crimes/

Xie, W., Zhu, F., Xiao, J., & Wang, J. (2018). Social Network Monitoring for Bursty Cascade Detection. *ACM Transactions on Knowledge Discovery from Data*, *12*(4), 1–24. https://doi.org/10.1145/3178048

Zasiekin, S., Kuperman, V., Hlova, I., & Zasiekina, L. (2022). War stories in social media: Personal experience of Russia-Ukraine war. *East European Journal of Psycholinguistics*, *9*(2), Article 2. https://doi.org/10.29038/eejpl.2022.9.2.zas