INTRODUCTION

On the morning of February 18th, 2014, Kiev awoke to 20,000 protestors marching on Ukraine’s Parliament building. Three days later, hundreds of demonstrators and police were dead or missing, and well over 1,000 people had been seriously injured (Interfax-Ukraine, 2014). But when the dust settled on February 22nd, Euromaidan protestors had won. They had removed President Viktor Yanokovych from office.

The immediate conditions which led to the Ukrainian revolution began months earlier in November of 2013. Amid increasing domestic tensions over Ukraine’s future relationship with the EU and Russia, a comprehensive free trade deal between the EU and Ukraine suffered months of delays and postponements. Seeing this as a threat to Ukraine’s future relationship with Western Europe, Pro-EU demonstrators began their months-long occupation of Independence Square in an attempt to push their government into ratification. Months of clashes between police and protestors ensued, and the weeks preceding the final march on parliament were just as tense. But the uneasy stalemate between demonstrators and police eventually broke on February 18th, 2014, as 20,000 protestors marched on Parliament. Over the next two days, 75 people were killed and 571 were injured (BBC 2014). The conflict ultimately claimed the lives of 780 people and injured over 1,800 (Interfax-Ukraine, 2014).
Within a year following Ukraine’s political revolution, Russia had annexed Crimea, and separatists near Ukraine’s Eastern border declared their independence, starting the still ongoing War in Donbass (Fischer, 2019).

During this time of political upheaval, Ukrainian citizens documented their experiences on online platforms like Twitter, creating a large collection of time capsules, each containing individual narratives from the period. Recognizing the importance of this discourse, Dr. Marina Kogan, professor at the University of Utah’s School of Computing, began collecting data from Twitter on November 19th, 2013, gathering valuable conversations taking place around the crisis. With 96 million tweets recounting Ukrainian citizens’ experiences between 2013 and 2015, this data set can provide useful insights into the ways ideas and political discourse change over time during political crises.

Of interest to me in the case of the Ukrainian revolution is the use of English by native Ukrainian speakers. Although English isn’t the most prevalent language in Ukraine (Education First, 2014), it is extremely prevalent in the data, completely overwhelming the discourse 3-4 months after data collection began. It’s been suggested the widespread use of English by Ukrainian and Russian native speakers during the crisis represented a conscious, strategic communication effort to shape the narrative of their struggle in a particular way for a particular audience (Metzger et. al., 2015; Tucker et. al., 2014), but study into the particulars of what that discursive strategy did or looked like has been thin (Metzger et. al., 2015).

**LITERATURE REVIEW**

Twitter is a social media application which allows users to share messages up to 280 characters long with each other. These messages are called *tweets*, and can include text, images, or video.
Social media often finds a unique role to play in times of crisis, whether that crisis presents itself as a natural disaster like a hurricane (Starbird and Palen, 2011) or a political movement (Arif et. al., 2018). Twitter’s role as an effective platform for information dissemination and political organization has been well documented in numerous studies over the last decade (Jost et. al., 2018; Starbird & Palen, 2012; Sutton et. al., 2008; Hughes & Palen, 2009). Just three years after Twitter’s 2006 launch, Hughes and Palen (2009) found that users experiencing emergency events tend to utilize Twitter for information dissemination and brokerage far more than normal. Kogan and Palen (2018) found that users experiencing external crises tend to expand their conversations with others, both in the number of conversational partners they have, and in the content they share as they search for relevant and timely information. Ukraine has been a particularly salient example of this ad-hoc information network in action, as social media has been widely used for political purposes by organizers since the Orange Revolution in 2004 (Sikorska, 2014; Bohdanova, 2014).

In this specific uprising, the Euromaidan, social media heavily influenced the strategies deployed by Ukrainian protestors (Tucker et. al. 2014). Tracing back what exactly took place before, during, and after the crisis on these social networks requires an understanding of what function the networks had to begin with. Jost et al. (2018) claim in their meta-study of numerous articles on protest movements in the US, Spain, Turkey, and Ukraine, that social media platforms appear to be particularly useful for coordination, collective motivation, and political exposure. The authors cite Ukraine’s Euromaidan as an interesting example of this. They observed that the number of Ukrainian Twitter users increased in sync with waves of protest activity, with a particularly high set of spikes directly following the February 18th march on parliament, suggesting “many Ukrainians may have joined Twitter in order to follow Euromaidan events”. Bohdanova (2014) goes even further in their study of social media’s role in the Euromaidan,
finding social media to be an effective vector for activity organization and message amplification. This phenomenon is not unique to Ukraine. In their paper examining information diffusion during the 2011 Egyptian political uprisings, Starbird and Palen (2012) demonstrated the power Twitter had as a platform for crowd-based information processing as users recommended and boosted high priority information coming from original, local sources. An event as locally impactful and geopolitically significant (Zelinska, 2017) as the Euromaidan is worthy of study, and Twitter has proved itself as a medium of political communication worth studying it on.

**DATA & METHODS**

The data was collected by Dr. Kogan between November 2013 and August 2015. Collection began within days of the first protests on November 20th, 2013. To ensure only tweets related to the event were gathered, tweets were filtered by keywords and hashtags related to Kiev and the Euromaidan protests.

The research presented here applies two lenses of analysis to that data: content analysis and temporal analysis, modelled off research presented in Kogan and Palen (2018). In their paper, the authors analyzed conversation structure within Twitter threads during the natural disaster Hurricane Sandy. By analyzing the content of conversations, as well as how that content changed over time, the authors were able better understand the way individuals search for critical information online during times of crisis.

In this study, the content analysis centered around topic modelling, a machine learning technique for discovering hidden semantic structure within text. Topic modeling clusters words which co-occur in similar contexts to create clusters of words called “topics.” It can then be inferred which documents (in our case, Tweets) correspond most to which topic. This technique is considered
“unsupervised,” because there is no ground truth or statistical prior presented beforehand. We simply observe which words appear together or are used in similar contexts, and cluster those words accordingly. For example, if the phrases “bucket of water” and “bucket of sand” appear often in a text, “bucket”, “sand”, and “water” may appear in the same topic. Further extending this notion of “context surrounding a word” can give us complex topic clusters like the ones in Figure 1 [Figure 1]. State of the art topic modeling typically uses variants of the Latent Dirichlet Allocation (LDA) model described in Blei et. al. (2003). The particular model I used was the Biterm Topic Model (BTM), a derivative of LDA, as it was developed specifically with short text, like tweets, in mind (Yan et. al., 2013).

Next, in order to appreciate the full scope of discourse on Twitter, one must look beyond a single post as a unit of analysis. The aforementioned Kogan and Palen (2018) offer one such approach to temporal analysis which satisfies this goal. They highlight two temporal features of the data which must be analyzed alongside content to generate a fuller picture of the discursive strategies at play: the temporality of conversations, and how content affects temporal features.
Before analysis, however, some data cleaning had to be done. The tweets were collected via keyword matching, so when users spoke about things like “Kiev,” their conversations were taken into the database. However, not all discussions around those keywords were necessarily about the Euromaidan, so using topic models like LDA (Blei et. al., 2003) or BTM (Yan et. al., 2013) alongside the measure of information similarity KL Divergence (Kullback & Leibler, 1951) across varying timelines was useful in determining when conversations shifted away from topics I was interested in.

RESULTS & DISCUSSION

Here, I’ll present the findings generated by the preceding methods in the following order. First, I will highlight noteworthy statistical attributes of English tweets within the data. Then, I’ll go over the findings of the topic modeling proper. Finally, we’ll investigate the specific case of topics relating to the separatist cities Luhansk and Donetsk, as well as their relationship to the ongoing War in Donbass.

**Characteristics of English Tweets**

As mentioned before, the presence of English in the data set wasn’t particularly surprising, but the sheer volume of English tweets compared to tweets in Ukrainian and Russian were unexpected given the strict filtering rules. Nearly a third of the 96 million tweets collected were found to be English later on. Although we don’t have exact statistics of English use in the country, Education First’s (2014) annual English Proficiency Index survey ranked Ukraine as 44th out of the 63 countries surveyed, placing it at the bottom of the “low proficiency” bin. This is to say, in 2014 at least, although English was not a popular language in Ukraine, it was a popular language among Ukrainian Twitter users.
The average density of English tweets remains high throughout the data set. The distribution of those tweets, however, is not uniform. Initially, no more than 10% of tweets in a week are listed as being in English. As time goes on, there are multiple spikes where English usage overwhelms the data before falling within a month or so. Three of the most noticeable spikes began around the following dates: November 30th, 2013, January 16th, 2014, and February 18nd, 2014. Those dates correspond to three extremely important events: November 30th, the first instance of mass police violence against protestors; January 16th, the passing of President Yanukovych’s anti-protest laws; and February 18th, the final march on parliament leading to Yanukovych’s removal from office. These spikes in English tweet frequency may indicate some motivating force for discourse outside the network, and the dates corresponding to important political events is likely no coincidence either, though we’ll investigate that further when we reach content analysis.

For now, it should be noted which events did not appear to lead to spikes in English Twitter usage. Over the month of December, 2013, police and protestors clashed almost every night. On December 1st, the protests were declared a riot, and police fired crowd control munitions into the crowd. By the 11th, non-lethal rounds, as well as live rounds, were being fired by police, while protestors responded with crude weapons and explosives. These spikes in violence were significant events to the broader Euromaidan movement, but they didn’t seem to stir up as much conversation as later news did. The implications here are unexpected. Initially, we know social media lends itself particularly well to the spectacle of violence, and images of protestors fighting police brutality should have no problem going viral, as we saw this last Summer during the George Floyd protests. Although the initial spike in violence sparked some English language discourse, what really caught English users’ eyes seemed to be news about civil liberties and democratic institutions coming under fire.
Topic Modeling Results

To get a better understanding of what could have caused this, let’s look at some of the content these Tweets actually contained. I ran the BTM algorithm on the subset of English tweets separated into monthly and weekly intervals. This choice of time granularity was motivated by a desire to see both macro and micro changes in the discourse as time went on, and was an attempt to account for potential noise in the model interfering with topic resolution at time spans smaller than a week. The choice for week-based granularity ended up being far more interpretable and useful for analysis. Recall these topics are merely clusters of frequently co-occurring words or words occurring in similar context, so giving them names or titles can’t be justified by this technique. As such, in the following discussion, I’ll be referring to topics by a unique, ordinal number between 0 and 40. The values of these numbers are not empirically relevant to topic interpretation, and are only to be used for distinguishing them from each other.

The frequencies of each topic over the data collection period [Figure 2] indicate clear user preferences for topics 4, 12, and 21. For reference, the top five words for each topic are listed in Figure 3 by column [Figure 3]. Topic 4 contains words largely relating to the annual Eurovision Song Contest, an international song competition featuring participants representing primarily European countries. Eurovision 2014 began in May of 2014, directly in the aftermath of the Euromaidan while Ukraine had just begun fighting separatists on its Eastern border. Ukraine’s representative in the competition, Mariya Yaremchuk, placed 6th in the final round (EBU, 2014), and became a popular conversation piece. Conversations like these were supposed to have been filtered out in the preprocessing stages, but this topic slipped through. Given the specificity of this topic to Eurovision in particular, however, it’s unlikely to have had much of an effect on the interpretability of other topics.
Topics 12 and 21 are much more related to the political crisis. Topic 21 appears to relate to procedural and bureaucratic conversations regarding Ukraine’s political future. Words like “council”, “agreement”, “summit”, and “discuss” are among the top ten words within the topic, indicating a clear focus on generic diplomacy and politics. The top word in topic 21, however, is “lavrov”. Further investigation revealed these references to be pointing to Sergey Lavrov, Russia’s Minister of Foreign Affairs. Lavrov became a common antagonist among the Euromaidan protests and the new Ukrainian government that followed them for his advocacy of
a more Russia-aligned Ukraine (Euromaidan Press, 2014). Looking at the timeline of top topics by week [Figure 4], you’ll notice topic 21 becomes a main point of discussion in March of 2014, and from November of 2014 onward. The first spike in topic 21’s prominence is best explained by Russia’s incursion into Crimea on February 20th, 2014, marking the start of the ongoing Russo-Ukrainian War (Snyder, 2018). Early on in the conflict, Russia’s military buildup of nearly 40,000 troops near Ukraine’s eastern border was seen as a clear threat, but many held out hope that a diplomatic resolution would be possible. That did not end up being the case, and as of February 2020, the OHCHR estimates over 13,000 have been killed in the conflict, 3,000 of which were civilians (UN OHCHR, 2020). As of early 2014, however, this was not a scenario English speaking users dwelled on in detail, and instead focused on identifying political mechanisms which may have been used to avoid such a disaster.

Figure 4
[The most frequent topic to occur each week]
[Colors and height do not indicate any values, and are only meant to assist in distinguishing individual topics.]
Topic 12 stands in sharp contrast to the more reserved terms present in topic 21. Words like “nazis”, “militaries”, and terms relating to violent political seizures of power like “junta” and “coup” were all present in the top 10 words. The period in which these words emerged from the data, March of 2015, offer some explanation as to the great divide between the rhetorical strategies implied by topics 21 and 12. From USAToday (2015): “A volunteer brigade with self-proclaimed Nazis fighting alongside government troops against Russian-backed separatists is proving to be a mixed blessing to its cause.” The Azov Battalion was a volunteer regiment of the Ukrainian National Guard based on Mariupol, a city in south eastern Ukraine. Almost immediately after being incorporated into the National Guard, allegations of torture and war crimes emerged (UN OHCHR, 2016), creating a bit of a stir in the English speaking world, as US military aid had been reaching the group since 2014 (Kheel, 2018). What really set topic 12 loose, however, was the revelation of neo-Nazi sympathizers, as well as open and proud Nazis, within the group. In 2014, a spokesman unapologetically declared that 10-20% of the unit were neo-Nazis (Pugliese, 2015). That, alongside their repeated clashes with pro-Russian separatist groups in the eastern Ukrainian cities of Luhansk and Donetsk shifted the center of conversation in over 1.2 million tweets to Nazism, fascism, and military juntas.

This hard shift in conversation being driven by the emergence of armed neo-Nazis clashing with Ukrainian separatists caught my attention, and encouraged deeper exploration of the relationship between topic 12 and other topics in the data. Other topics that frequently occurred within the top 5 topics between March 2015 and July 2015 were topics 0, 20, and 25, the top 10 words of which can be found in Figure 5 [Figure 5]. These three topics were divergent in many ways, but there was one word which managed to appear in the top 50 for each: “luhansk”.

Luhansk and Donetsk

Luhansk and Donetsk are cities in Eastern Ukraine with a high population of Russian speakers and ethnic Russians. This isn’t uncommon for the region, and while the Crimean peninsula maintained higher demographic concentrations of Russo-Ukrainians, the concentrations in Luhansk and Donetsk, along with the annexation of Crimea in April of 2014 by Russia, were enough to set tensions particularly high between pro-EU and pro-Russian Ukrainians. This all came to a head on April 7th, 2014, when pro-Russian activists seized state security buildings and declared Donetsk and Lugansk “independent people’s republics” (BBC, 2014). Within hours, the Ukrainian interim president Oleksandr Turchynov vowed to launch a major “counter-terrorism” operation against the separatists. By April 15th, the War in Donbass had started, with several hundred pro-Russian separatists and Ukrainian troops fighting on the outskirts of Luhansk (Kramer, 2014). This protracted conflict, and the aforementioned Russo-Ukrainian war, are difficult to distinguish from each other today, as the war fronts, combatants, and strategic goals of both sides have largely melded into a broad, singular conflict. Russia’s repeated incursions into Crimea now involve providing military support to separatists in Eastern Ukraine, leading to the line between “pro-Russian Ukrainian separatist” and “Russian combatant” to be blurred. But in 2014, these conflicts were known to be related, but distinctly separate. As such, we can justify investigating the rhetoric surrounding separatist movements on Twitter by analyzing topics 0, 20, and 25.
Again, each of the topics 0, 20, and 25 contained the word “luhansk”, and appeared to relate to Luhansk and Donetsk in some way. But each did so in extremely divergent ways. Topic 0 contained words like “missions”, “entering”, “crossing”, and “defense”, seemingly bureaucratic discussions of the brewing War in Donbass. Topic 20 contained words like “civilian”, “front line”, “bomb”, “child”, and “red cross”, clearly indicating discussions of the victims of Ukraine’s offensive strikes, as well as local responses to that violence. Topic 25 appears to be similar to topic 20 at first glance, with words like “dead”, “blast”, “mine”, “explosion”, “dozen”, “wound”, and “bus”. But while topic 25 is certainly centered on violence in Lugansk and Donetsk, it’s notably different in how it represents that violence. Comparing the list of top 50 words of each topic, topic 25 contains noticeably fewer words relating to victims or emergency responses, and many more words describing the means by which the violent incidents themselves occurred. This alone isn’t terribly unexpected, people talk about violence in different ways based on the discursive contexts they find themselves speaking in. But further investigation reveals notable differences in the relative frequencies in which these topics appear over time.

By plotting time series of the relative ranks of topics 0, 20, and 25, a clear diverging pattern is revealed [Figure 6]. Topics 0 and 20 rise and fall in similar and predictable ways in tandem with spikes of violence in Eastern Ukraine, while topic 25 remains at a relatively low ranking until early 2015. In February 2015, topic 0 maintained its trajectory as expected, as neither the violence nor the conversations around Luhansk and Donetsk had waned. But topic 20 suddenly found its rank drop to the lowest it had been since April of 2014 nearly a year prior. At the exact same time, topic 25 experienced a massive jump in popularity, placing it even higher than topic 20 was at any time in the near past. Topic 25’s unexpected boost in relevance and topic 20’s
rapid decline became the norm for just a few months before reversing again right before our data set ended in August of 2015.

![Topic Rank Time Series (0, 20, 25)](image)

**Figure 6**
*Relative rankings of topics 0, 20, and 25 over time*

The inference we can gather from these numbers is this: when people were tweeting about violence in Luhansk and Donetsk between February and June of 2015, they were far more likely to use words in topic 25, and far less likely to use words in topic 20, than they had been at any point before or after. In other words, the frequency of conversations about victims dramatically fell at precisely the same time the frequency of explicit discussions surrounding specific acts of violence dramatically rose.

**CONCLUSION**
Social media communications change rapidly during times of crisis, and the Euromaidan is no exception. Twitter users in Ukraine generated an enormous 350 gigabyte dataset of 96 million tweets in which they distributed information, organized political movements, and described their feelings about the political events of the time. In this project, a quantitative analysis of Twitter users in and around Ukraine who chose to communicate in English on the platform revealed a few insights into the ways English may be deployed in political crises involving speakers that don’t use English as their primary language.

First, spikes in violence were heavily correlated with spikes in violence-related topics like topics 12 and 25. I do not attempt to demonstrate whether the methods presented in this paper are capable of uniquely identifying these spikes better than any other method, but instead point to the specific forms of violence which seem to trigger spikes in English conversation. Violence perpetrated by neo-Nazis (topic 12) and government (topic 25) against separatist forces appeared to engage English language speakers the most, while the initial violence perpetuated by police at the start of the Euromaidan didn’t attract as much sustained attention from the same users.

Next, large changes in the political landscape preceding the Russo-Ukrainian war corresponded with the emergence of political figures like Sergey Lavrov within popular topics like topic 21, which broadly demonstrated the capacity for particular topics to correspond to a particular geopolitical environment.

Finally, diverging topics illuminated diverging rhetorical tendencies taken by individuals sharing similar bits of information. As the War in Donbass was beginning, English users discussing Luhansk and Donetsk tended to use terms heavily correlated with topic 20, but briefly made a collective shift to topic 25 between February and June of 2015.
The methods employed by this study did not provide rich enough information to make broad statements about national identity formation or general political rhetorical strategy. The choice to focus solely on English enabled a broader survey of English user’s strategies, but also deprived us of a proper control group with which to make more holistic descriptive claims. The focus on English did, however, establish multiple trends in the ways English users tended to tweet about Ukraine’s political revolution. This information could serve as the foundation for a deeper analysis into the political strategies deployed by both the protestors and the government during the initial Euromaidan, as well as the ways the newly established government navigated public relations in the aftermath of the revolution. Using that information in tandem with the information uncovered by this study could reveal interesting similarities or differences between official communications, and the ways English users tended to relay that information to their followers.

The goal in understanding the rhetorical and political strategies employed by revolutionary or oppositional political movements should always be to improve mutual intelligibility and shorten the path to resolution. An understanding of why a particular rhetorical strategy is deployed must first be preceded by an understanding of what that rhetorical strategy looked like to begin with. This research contributes to our understanding of Ukrainian social media users’ collective strategies and tendencies during and after the Ukrainian political revolution of 2014. There is no reason to believe political struggle in the future will be identical to struggle in the past, but our communication strategies change far slower than our technologies do, making insights like these valuable.

REFERENCES


Kheel, Rebecca (2018-03-27). "Congress bans arms to Ukraine militia linked to neo-Nazis". The Hill.


Pugliese, David (26 June 2015). "Ukrainian unit accused of Neo-Nazi links wants Canada's help". Ottawa Citizen


