DIGITAL ACCESS TO SCHOLARSHIP AT HARVARD

Russia, Ukraine, and the West: Social Media Sentiment in the Euromaidan Protests

The Harvard community has made this article openly available.
Please share how this access benefits you. Your story matters.

(Article begins on next page)

RUSSIA, UKRAINE, AND THE WEST

Social Media Sentiment in the Euromaidan Protests

INTERNET MONITOR

INTERNET MONITOR is a research project to evaluate, describe, and summarize the means, mechanisms, and extent of Internet content controls and Internet activity around the world.

thenetmonitor.org

INTERNET MONITOR is a project of the Berkman Center for Internet & Society.
http://cyber.law.harvard.edu

Russia, Ukraine, and the West
Social Media Sentiment in the Euromaidan Protests

Bruce Etling

INTERNET MONITOR

ABSTRACT

This paper investigates sentiment in the online conversation about the Ukrainian Euromaidan protests across a range of English- and Russian-language social and traditional media sources. Results from this exploratory research show more support for the Euromaidan protests in Russian-language sources, including among sources and users based in Russia, than originally expected. Sentiment in English-language sources, including those located in the United States and United Kingdom, is more negative than anticipated given the rhetorical support among western governments for the Euromaidan protests. However, social media content in Ukraine, the US, and the UK is more positive than traditional media outlets in those countries.

ON NOVEMBER 21, 2013, Ukrainian President Victor Yanukovich decided to pull out of talks to more closely align Ukraine with the European Union. Instead, he chose to foster closer ties with Russia and a Kremlin-led Eurasian trade union. This move immediately sparked outrage among those Ukrainians hoping for greater integration into Europe and led to a series of massive protests centered in Kiev's Maidan Square. A permanent protest presence also emerged as a tent city, and barricades to protect protesters from the Berkut riot police were erected.

In short order, the protesters adopted the Twitter hashtag #euromaidan as their moniker. Protests continued at various levels of participation and intensity until late February, when violence between protesters and the Berkut riot police escalated. Hard core oppositionists stood their ground, and President Yanukovich eventually fled the country, turning up days later in the southern Russian city of Rostov. The president's party quickly abandoned Yanukovich, and the Parliament ushered in a provisional government and issued a warrant for Yanukovich's arrest. There was little time to reflect on the Euromaidan protests themselves, as they were quickly overcome by events such as Russia's annexation of Crimea, fighting between Ukraine and pro-Russian separatists in Ukraine's eastern provinces, and most recently the downing of Malaysia Airlines Flight 17. Yet the Euromaidan protests are worthy of study, given that they represent another successful bottom-up protest movement aided by the Internet that led to the fall of a government seen as corrupt and acting against the interests of its citizens.

The protests were in large part a debate within Ukraine about whether the country should integrate more closely into the EU, or instead turn east and back towards Russia. In addition to the offline protest actions, online debates on social media platforms such as Twitter, blogs, Facebook, and forums were a key part of how the world saw the protests, as well as how Ukrainians themselves talked about the events. To investigate these dynamics, this paper uses Crimson Hexagon to understand the sentiment in the online conversation about the Ukrainian protests across a range of English- and Russian-language social and traditional media sources. Results from this exploratory research show more support for the Euromaidan protests in Russian-language sources than originally expected, including in sources located within Russia. Sentiment in English-language sources, including those based in the US and UK, was more negative than expected given the rhetorical support among western governments for the Ukrainian protests.

RESEARCH QUESTIONS

This paper investigates the following research questions:

1. Given the interests of the EU, the US, and Russia in the outcome of the protests and the direction toward which Ukraine would eventually gravitate, are English-language social and traditional media more supportive of the protests than Russian-language sources?

2. How does this support or criticism of the protests and protesters vary over time?

3. Among Russian-language sources, are Ukraine-based sources more supportive of the protests than those based in Russia?

4. Are US- and UK-based sources more supportive of the protests than those based in Ukraine or Russia?

INTERNET MONITOR

METHODS AND DATA

This study uses data and software provided by Crimson Hexagon, which relies on an algorithm and method created by Hopkins and King (2010). This method consists of manually coding a small, nonrandom subset of stories into categories of interest: in this case, posts that are supportive, critical, or neutral with respect to the protests, or irrelevant. The algorithm then reports on the proportion of categories (again: positive, negative, neutral, or irrelevant) found in the entire text corpus. For a detailed description of the method see the Hopkins and King paper cited above. For an overview and some applications see the work with the platform by Pew's Project for Excellence in Journalism (Hitlin 2013). For an application to foreign language content see King, Pan, and Roberts (2013).

A series of what Crimson Hexagon calls monitors were created to assess the sentiment in online conversations and news stories about the protests. This involved the following steps:

1. Selecting a time period and media sources or platforms to include (i.e., Twitter, Facebook, forums, blogs, etc.). The date range for all monitors used in this paper is November 21, 2013, when Yanukovich decided to end talks with the EU, through February 26, 2014, the approximate date when Yanukovich left the country and other demands of the protesters were largely met, including the selection of a date for new elections.

2. Creating a Boolean search string with terms that represent the topic of interest. This string will pull relevant posts from Crimson's large data set of social media and news content for coding.

3. Training the algorithm by coding a random set of posts into positive, neutral, and negative sentiment categories, as well as an irrelevant category.

Approximately 30-45 posts were coded into each sentiment category, well above the 20-post minimum recommended by Crimson Hexagon. Posts coded as positive included those that were overtly supportive of the protests, pro-EU, critical of Yanukovich and violence by the Ukrainian police (including the Berkut riot police), and posts that were critical of Russia, Putin, and the Russian government's policy towards the protests. Posts coded as neutral included those that simply shared 'headline' type announcements about protests occurring in Ukraine as well as those that did not offer overt support or criticism of the protests or protesters. Negative posts included those that were overtly critical of the protests, that emphasized nationalist leadership in the protests or used language about fascism, Nazis, and Hitler to describe the protesters and protest leaders, were anti-EU, pro-Russian, and those that stated that protesters were paid for and organized by the West.

In summary, posts were coded for overt positive and negative sentiment, but an attempt was not made to distinguish fine shades of meaning. This conservative training of the algorithm was intended to identify sentiment that was clearly supportive or critical of the protests. Taking into account more nuance in the framing and discussion of protests would likely have led to more content falling into those sentiment categories and out of the neutral category. As recommended by Crimson Hexagon, the best examples of sentiment that could be found were selected for each category of interest, and categories needed to be mutually exclusive. As this is an exploratory study, only one bilingual researcher trained the monitor, and reliability tests were not conducted. Results could

INTERNET MONITOR

certainly vary depending on who trains the algorithm.

One issue that arose with Crimson Hexagon's Russian-language data is that many news websites pulled in irrelevant headlines, ads, and other text content from windows outside the main story text. An additional category for this type of 'mixed content' was coded (represented typically by a series of news headlines or story summaries, only one of which was related to the Ukraine protests.) To further prevent this type of content from biasing the results, the data from websites that most typically included this type of mixed or irrelevant content (the 'worst offenders') were also removed. This is unfortunate, since several of these were top Russian news sources such as NTV, which one would expect to have very pro-Russian frames that were opposed to the protests. The Russian news results are therefore possibly underestimating the amount of negative content.

DATA
The main sources of data used for the English- and Russian-language monitors are the text of Twitter (the full firehouse), blogs, news, forums, and Facebook. Filtering by country is currently only available for Twitter, forums, and Facebook, so the monitors for sources based in Ukraine and Russia were limited to those three platforms, and the training data was limited to those sources. Crimson Hexagon does not currently include Ukrainian-language sources, so a monitor could not be created for this language, but this could be an interesting area for future research. One would assume that content in Ukrainian, which is spoken more frequently in western Ukraine than elsewhere in the country, would be more supportive of the protests than Russian-language content.

FINDINGS

COMPARISON BETWEEN ENGLISH- AND RUSSIAN-LANGUAGE SOURCES
The first research question seeks to understand if English-language content is more supportive of the protests than Russian-language content, and to explore whether there are differences over time between the language groups. Although Ukrainians also speak Russian, Russia has a much larger population and has more Internet users. The hypothesis is that a majority of Russian-language sources will follow the Russian government's policy preference and be more critical of the protests. To drill down to sentiment by location, in the next section, Russian-language sources based in Ukraine are compared to those based in Russia, as well as to English-language sources based in the US and UK.

As all of the charts in this paper show, there are three major phases to the protests, which theoretically could be linked to what Tarrow (1998) calls 'cycles of contention.' The first phase occurs from late November and into late December and includes the abandonment of the EU negotiations, the first major protests, including one with an estimated crowd of 800,000 in early December, and the establishment of a tent city on Maidan Square. Online discussion and debate then cools off during the holidays and New Year celebrations, especially in English-language sources.

The second phase lasts from approximately January 20 to February 3. This phase included the passage of a restrictive protest law by the Ukrainian parliament, the first violence and deaths at the protests, the death of a high profile activist, Yuriy Verbytsky, the storming of regional government offices in Western Ukraine, the resignation of Prime Minister Mykola Azarov, and the passage of a broad amnesty law.

⋮⋮⋮ INTERNET MONITOR

The third, final, and most intense period of discussion online begins around February 17 and lasts through the end of the analysis in this paper on February 26. This final phase coincides with the greatest level of violence, including major clashes between protesters and the Berkut police force on February 18 that led to the death of 18 people, including 7 police, and attempts by the police to encircle and then clear Maidan Square. The worst spate of violence occurred around February 20, when 88 people were killed in a 48-hour period, and snipers also opened fire on protesters. On February 22, Yanukovich disappeared, protesters stormed the presidential administration, parliament removed the president from power and set a date for new elections, and opposition politician Yulia Tymoshenko was released from prison. During the period of February 23-26, parliament named an interim speaker and president and issued an arrest warrant for Yanukovich, and the Berkut police unit blamed for violence against protesters was disbanded (BBC 2014).

A comparison of the number of posts between the two languages shows over 100,000 more posts in English than in Russian (385,971 English; 273,648 Russian). This is not surprising given the significantly larger number of English-language Internet users. The first phase of the protests garners a higher and more sustained level of interest among Russian-language users. The highest peak in volume for both language groups is in the third and final phase. In general, English-language volume rises and drops off quickly, while the Russian-language sources tend to show a more sustained level of interest across the three phases. The second phase is also marked by a sustained level of consistent discussion of the protests in Russian-language sources, although the overall volume is lower compared to English.
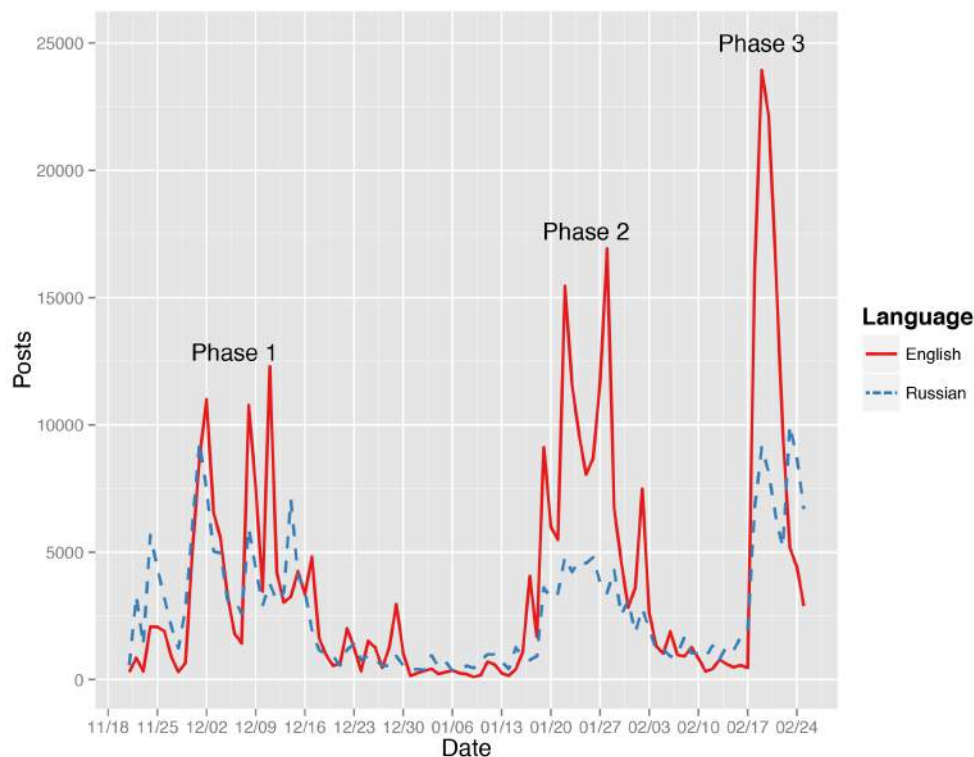

FIGURE 1: POST VOLUME BY LANGUAGE

Media sources

A comparison of sources across the two languages shows that blogs make up a much larger proportion of posts in Russian (23%) compared to English (just 5% of content). It appears that blogs, which are primarily from the LiveJournal platform, are still an important part of the networked public sphere on the Russian-speaking Internet. It is also clear that Twitter accounts for a majority of posts in both languages, but this is somewhat misleading, as a short tweet of 140 characters or fewer is not really comparable to a blog post, forum post, or newspaper article, which typically tend to be much longer and more time consuming to create. The amount of text generated on Facebook might also understate the importance of that platform for protest movements, especially coordination pages. Activity on Facebook, for example, might be better measured by 'likes' or the number of links pointing to a coordination page than by the number of posts.
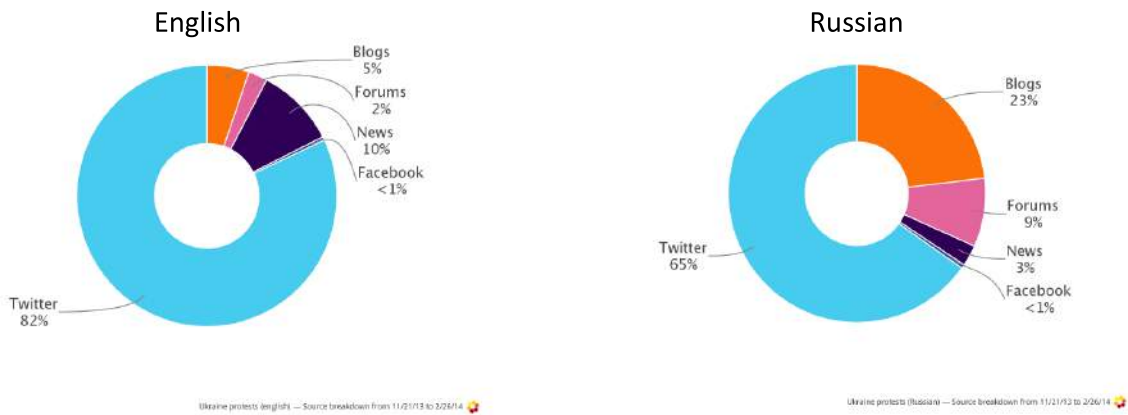

English

Russian

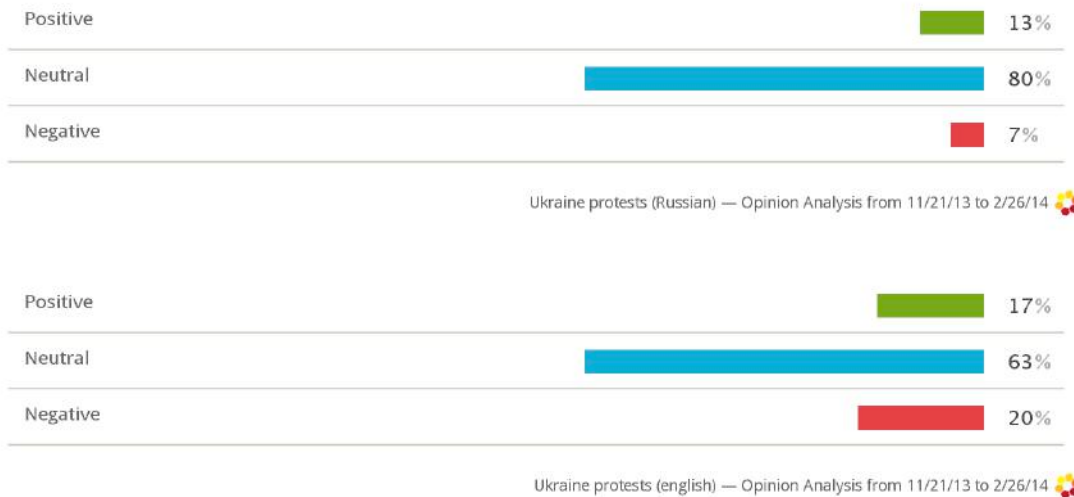FIGURE 2: ENGLISH- AND RUSSIAN-LANGUAGE SOURCES

Sentiment

FIGURE 3: RUSSIAN- AND ENGLISH-LANGUAGE SENTIMENT

INTERNET MONITOR

The first major takeaway regarding sentiment is that a majority of the content in both languages falls into the neutral category, which means reporting in a balanced or neutral way, or just in a 'headline' fashion that does not really express any opinion. For example the algorithm identified these posts as neutral, which fits well with the initial training posts:

> Ukraine protests grow as president responds
> **http://t.co/Z6RJxK08QN**

> Days of Protest in Ukraine In Focus - The Atlantic
> **http://t.co/SaRRF1W7mO**

Overall, English content is more negative than Russian content. Here is an example of negative content drawn from a day when protests turned violent in December:

> US, EU are 'neck deep' in plan to see pro-Western regime change in **#Ukraine**
> **http://t.co/uK3Ir71TDb #US #EU #NATO** behind **#Ukraine** protests.

> US, **#EU** ignore inconvenient Nazi, white power flags and symbols in **#Ukraine** 'protests'
> **http://t.co/VHq2ik62xI**

Here are two examples of posts with positive sentiment or framing in support of protests, drawn from the same day:

> Ukrainian protests turn violent, Yanukovych attacks instead of listening to pleas for democracy!
> **http://t.co/YuNDQygCnT #euromaidan**

> I went to Kiev last week for 8 days and took some pictures of a very peaceful Maidan. It is very sad to see all the violence, and how this beautiful place is burning. All I have personally seen was peaceful protests and a wide range of different Ukrainians concerned about the future of their country. I wish the Ukrainian people the very best. Let's hope they will succeed in bringing down this kleptocratic government.
> **http://imgur.com/a/fwUmZ#0**

Russian-language sources are more positive than negative, while English sources are overall a bit more negative than positive. Both these findings are at odds with this paper's original hypothesis. In the below images, the Russian- and English-language results are limited to only positive and negative sentiment across time to highlight differences and similarities. It appears that the Russian-language content is more positive in the first phase and grows more negative over time, with negative content becoming more frequent in the final phase of the protests. One hypothesis that might explain this shift is the amount of violence that took place in the last phase of the protests, which fit well with the Russian government's frame that the protests would lead to violence, chaos, and blood on the streets. English-language content does become more positive at the very end of the final phase of the protests, which is likely a reflection of the success of the protesters and the departure of Yanukovich.
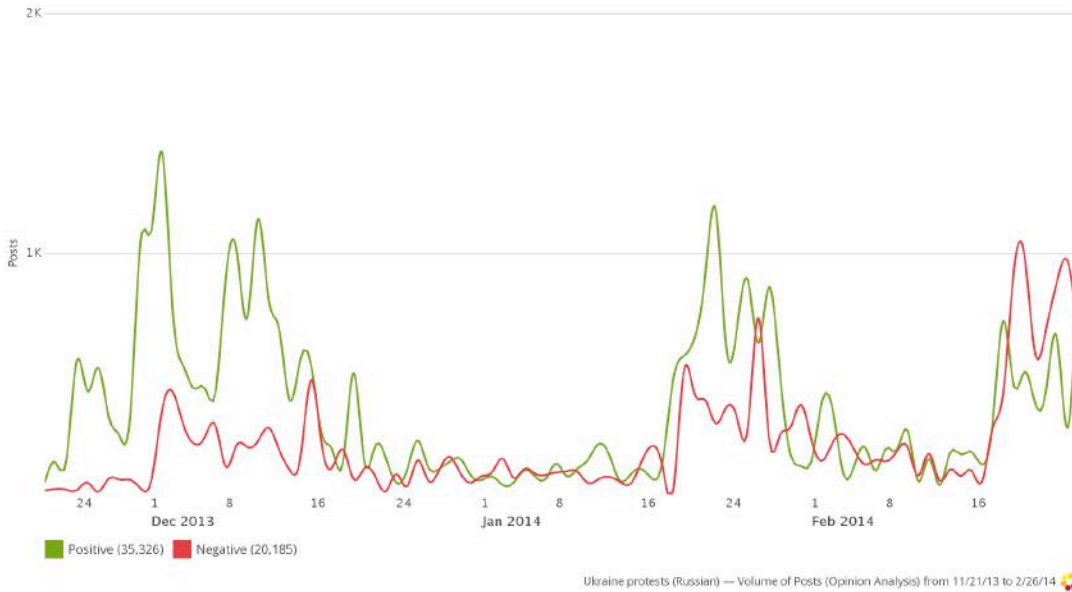
FIGURE 4A: RUSSIAN-LANGUAGE SENTIMENT OVER TIME

FIGURE 4B: RUSSIAN-LANGUAGE SENTIMENT VOLUME
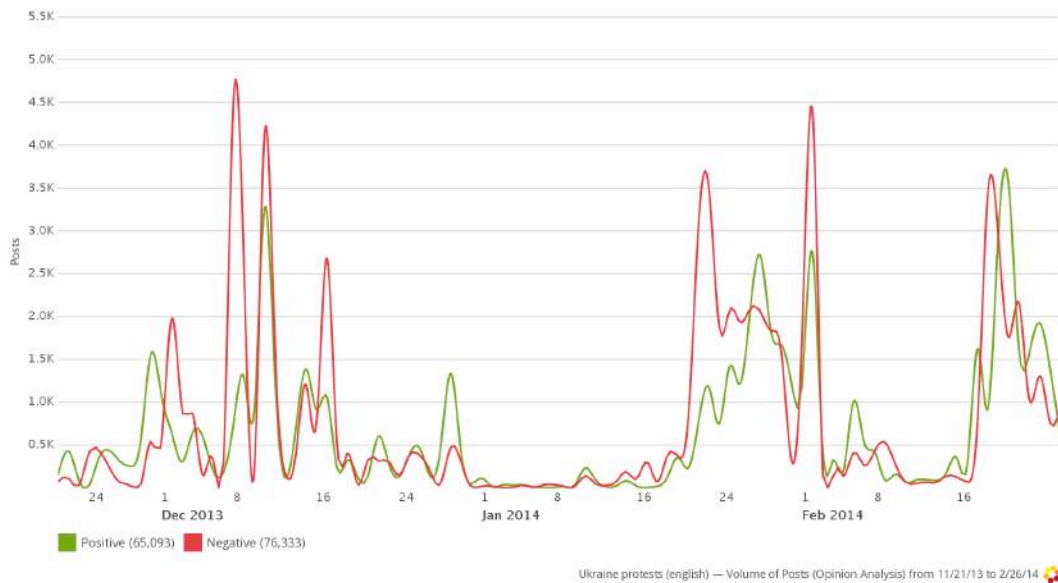
FIGURE 5A: ENGLISH-LANGUAGE SENTIMENT OVER TIME

INTERNET MONITOR

FIGURE 5B: ENGLISH-LANGUAGE SENTIMENT VOLUME

COUNTRY-BASED TWITTER, FACEBOOK, AND FORUMS

How did Russian-language users based in Russia talk about the protests compared to those based in Ukraine? How does this differ from sources based in the US and UK, which are the source of most English-language content in the English-language monitors referenced above? This paper's working hypothesis is that Ukraine-based users and sources will be more supportive than those based in Russia. As mentioned in the methods section, Crimson Hexagon only allows filtering by country for Twitter, forums, and Facebook, so news and blog content is not included in the below results. These results therefore are more representative of social media content than traditional media content.
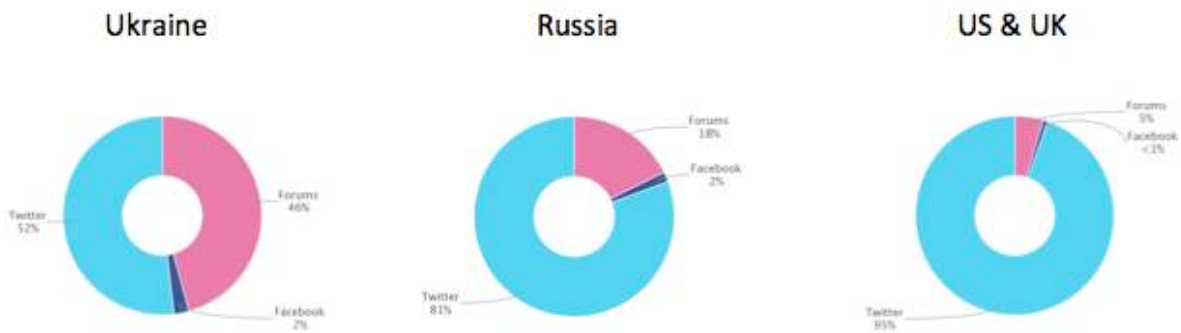

FIGURE 6: UKRAINE, RUSSIA, AND US & UK SOURCE VOLUME

Looking first at source volume across platforms and countries, Twitter remains the major source of content in all countries. However, forums account for 18% of content in Russia and almost half of the posts (46%) in Ukraine. For all countries, Facebook accounts for a small amount of posts (no greater than 2%).
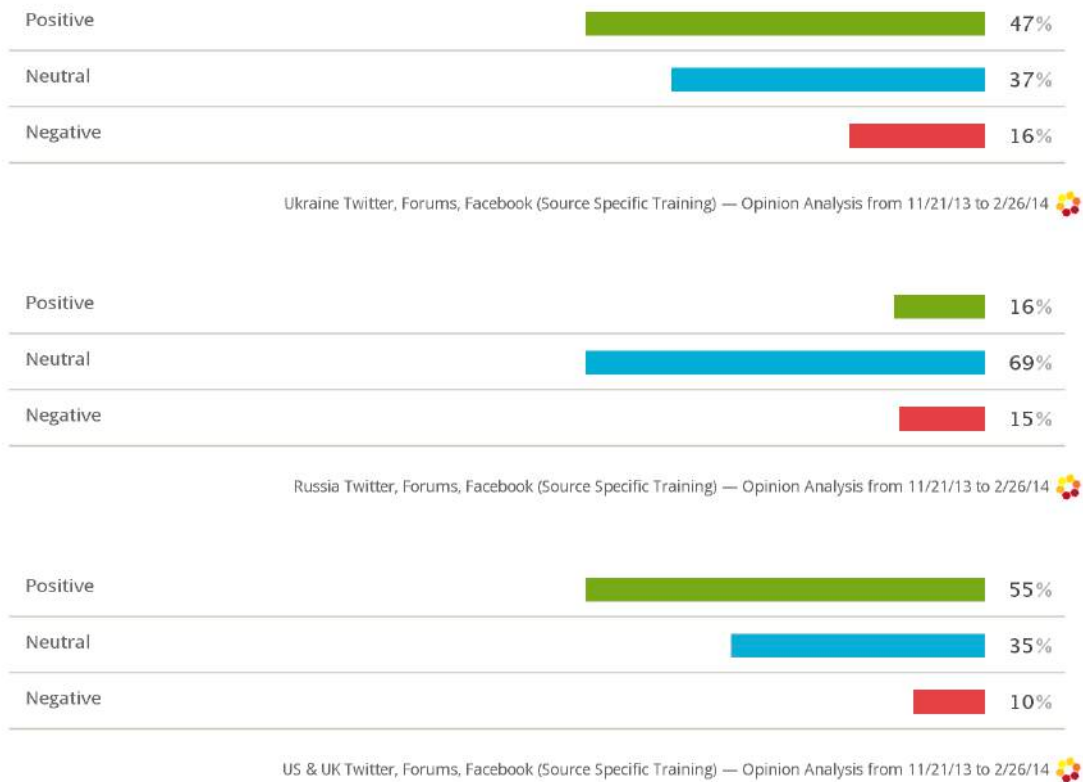
Positive	47%
Neutral	37%
Negative	16%

Ukraine Twitter, Forums, Facebook (Source Specific Training) — Opinion Analysis from 11/21/13 to 2/26/14

Positive	16%
Neutral	69%
Negative	15%

Russia Twitter, Forums, Facebook (Source Specific Training) — Opinion Analysis from 11/21/13 to 2/26/14

Positive	55%
Neutral	35%
Negative	10%

US & UK Twitter, Forums, Facebook (Source Specific Training) — Opinion Analysis from 11/21/13 to 2/26/14

FIGURE 7: SENTIMENT IN UKRAINE COMPARED TO RUSSIA AND THE US & UK

The results show that sentiment on social media in Ukraine is largely positive (47%), a substantial amount of content (37%) is neutral, and just 16% is negative. Social media content in the US and UK is even more positive than in Ukraine (55%), with about an equal amount of neutral content as in Ukraine (35%). For Russia-based social media content, most of the sentiment again falls into the neutral category, while the supportive content (16%) is about equal to that critical of the protests (15%). The social media findings for Ukraine and the US and UK, therefore, provide support for the hypothesis laid out at the beginning of the paper, while Russian social media is more neutral than expected. In the next charts, the neutral content is pulled out from all monitors to allow a focus on positive and negative sentiment from Ukraine and Russia.
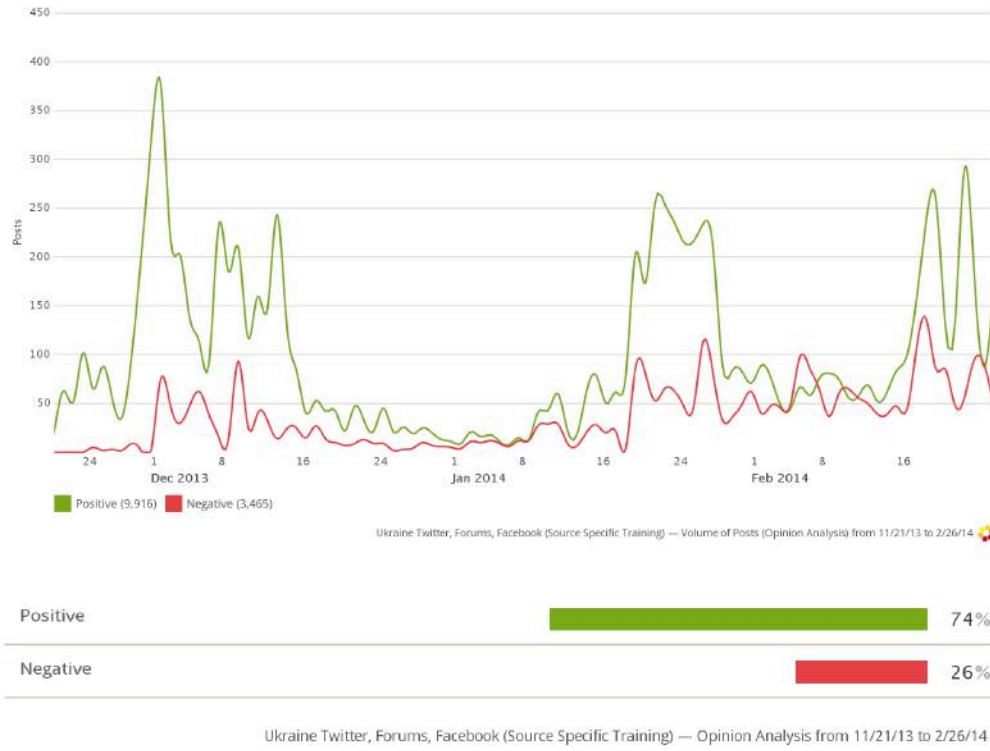
INTERNET MONITOR
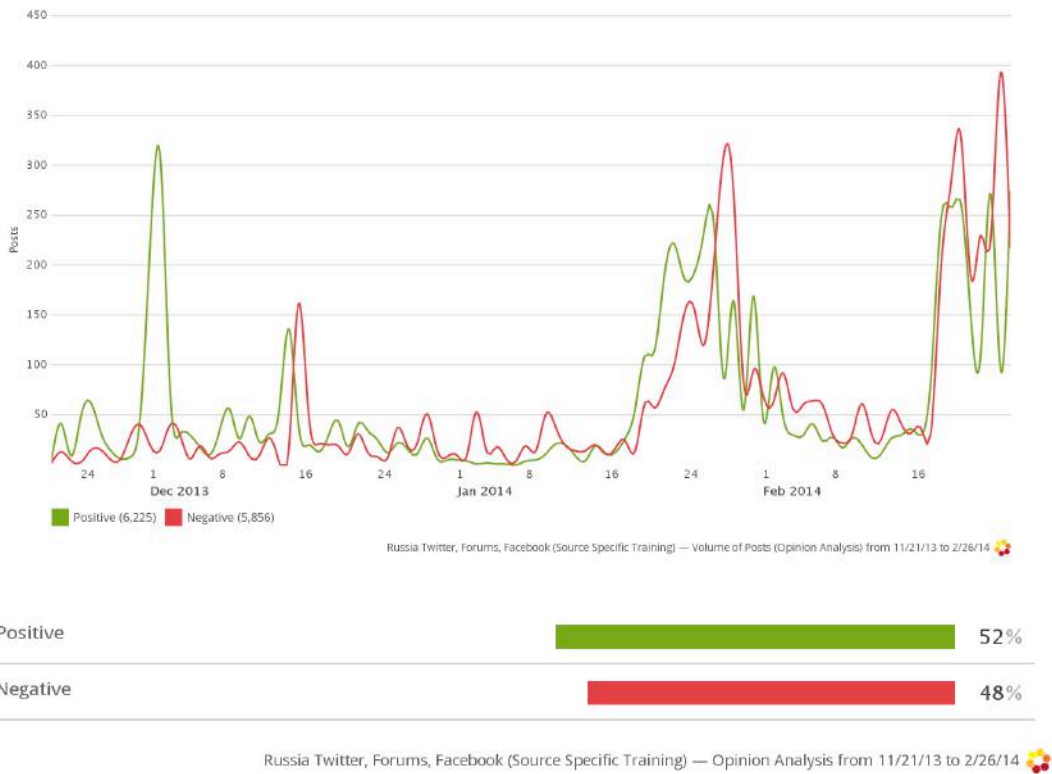
FIGURE 8: SENTIMENT OF UKRAINE-BASED SOCIAL MEDIA

FIGURE 9: SENTIMENT OF RUSSIA-BASED SOCIAL MEDIA

INTERNET MONITOR

Positive (51,545)   Negative (9,819)

US & UK Twitter, Forums, Facebook (Source Specific Training) — Volume of Posts (Opinion Analysis) from 11/21/13 to 2/26/14

| Positive | 84% |
| Negative | 16% |

US & UK Twitter, Forums, Facebook (Source Specific Training) — Opinion Analysis from 11/21/13 to 2/26/14
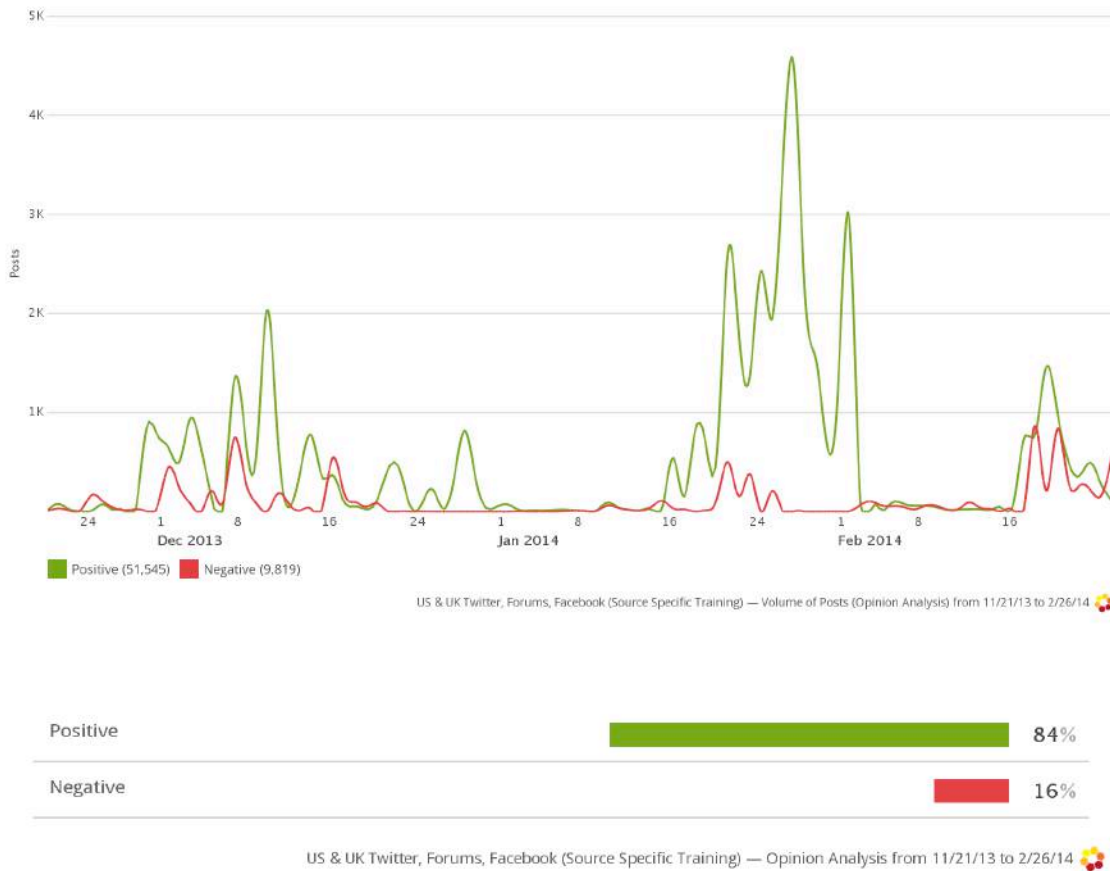
FIGURE 10: SENTIMENT OF US- & UK-BASED SOCIAL MEDIA

This data shows that discussion online in both Ukraine and Russia is more positive than negative, although Ukraine-based content is significantly more positive than content in Russia, and is supportive of the protests throughout all three major protest phases. The Russia-based sources show a peak of supportive sentiment early in the protests, while there are more negative peaks towards the later part of the second phase of the protests and during the third and final phase.

The US- and UK-based sources, similar to Ukraine-based sources, show a clear majority of posts falling into the positive category (55%). In fact, data from US and UK Twitter, forums, and Facebook has the largest share of positive sentiment of all the monitors. The highest volume of positive content for these two countries is found in the second phase of the Euromaidan protests.

It is also worth highlighting the differences between the sentiment on social media (Twitter, forums, and Facebook) compared to the overall English- and Russian-language results, which also include news and blog content. In Ukraine, the sentiment is much more positive on social media than in the overall results. Social media in Russia is somewhat more opinionated, but a little less positive. In the US and UK, social media on this topic is more opinionated and more positive in sentiment than in the overall

results. In the US and UK this finding might be explained by removing the news content, which in general is more neutral. However, it is also possible that this difference is an artifact of different training of the algorithm, since the social media monitors were trained with data exclusively from Twitter, forums, and Facebook.

CONCLUSION AND DISCUSSION

This exploratory research shows that there was more support online in Russia for the Ukraine protests than one might expect. This may be explained by the fact that previous research has shown that online political speech in Russia is more oppositional than supportive of the Kremlin, and that online sources used by bloggers that discuss politics tend to be more oppositional than supportive of the government (Etling et al.). This data also does not include discussion of Crimea, where the government's position may have enjoyed more support within Russia. It is also interesting that Russian-language sources in Ukraine are so supportive of the protests. This may be explained by the fact that many Ukrainians use both Russian and Ukrainian, and that language may not divide the country as clearly as many press accounts may have us believe. However, it is also possible that these early results point towards more support among Russian-speaking Ukrainians for the protests than the Russian government and mainstream media claimed. Much additional research would be needed to confirm this, but it is an interesting early finding. Finally, there was less support among English-language sources for the protests than originally assumed. This may be explained by the frequent finding in communication research that, at least in the US, news content overall tends to be negative (i.e., "If it bleeds, it leads"). Indeed, when the data relied upon for sentiment removes news sources and is purely based on social media content (Twitter, forums, and Facebook), US and UK sources are more supportive than critical of the Euromaidan protests.

There are many fruitful pathways for future research. Beyond confirming or refuting the early results from this study, interesting avenues of research include understanding differences in conversation and framing over Crimea, the role that violence plays in support or criticism online for protests, how closely social media sentiment tracks that of traditional news sources in the US compared to Ukraine and Russia, and a comparison of Ukrainian-language content and Russian-language content among sources and users within Ukraine proper. Finally, since knowing what communities are active online and how they engage in conversation over protests is essential, the need to combine social network analysis with automated text analysis is a critical next step for this type of research.

WORKS CITED

BBC. 2014. "Ukraine Crisis Timeline." Accessed July 21, 2014. http://www.bbc.com/news/world-middle-east-26248275.

Etling, Bruce, Karina Alexanyan, John Kelly, Robert Faris, John Palfrey, and Urs Gasser. 2010. "Public Discourse in the Russian Blogosphere: Mapping RuNet Politics and Mobilization." Berkman Center Research Publication 2010-11. http://cyber.law.harvard.edu/publications/2010/Public_Discourse_Russian_Blogosphere.

Hitlin, Paul. 2013. "Methodology: How Crimson Hexagon Works." Pew Project for Excellence in Journalism. http://www.journalism.org/2013/11/04/methodology-crimson-hexagon/.

Hopkins, Daniel, and Gary King. 2010. "A Method of Automated Nonparametric Content Analysis for Social Science." *American Journal of Political Science* 54 (1):229-247.

King, Gary, Jennifer Pan, and Margaret Roberts. 2013. "How Censorship in China Allows Government Criticism but Silences Collective Expression." *American Political Science Review* 107 (2):326-343. doi: 10.1017/S0003055413000014.

INTERNET MONITOR