

NATO Trident Juncture on Twitter: Public Discussion

**Will Frankenstein
Binxuan Huang
Kathleen M. Carley**

January 13, 2016
CMU-ISR-16-100

Institute for Software Research
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213



Center for the Computational Analysis of Social and Organizational Systems
CASOS technical report.

This work was supported in part by the Office of Naval Research under N00014140737 and N000140811186, and by NATO as part of TJ15. Additional support was provided by the center for Computational Analysis of Social and Organizational Systems and the Institute for Software Research at Carnegie Mellon University. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of NATO or the U.S. government. We are grateful to the Data Mining and Machine Learning Lab at Arizona State University for their assistance. The CASOS team worked directly with the NATO MISO cell under the auspices of the ONR. We would also like to thank Dr. Rebecca Goolsby (ONR), Dr. Brian Steckler (NPS), Major Rick Galeano and Jennifer Franco (BBN) for feedback on the products as they were being developed.

Keywords: NATO, North Atlantic Treaty Organization, Twitter, Latent Dirichlet Analysis, Network Analysis, Social Media, LDA, Media, Geotagged Tweets

Abstract

NATO conducted one of its largest exercises over the past decade, Trident Juncture, from October 21st to November 6th 2015. The exercise included over 36 thousand troops from over 28 allies, 9 partner nations, and engaged 18 observations and 12 international and non-governmental organizations and aid agencies. As part of this exercise, the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon was asked to assess, in partnership with the Data Mining and Machine Learning Lab at Arizona State University, the social media response to Trident Juncture. We collected data from Twitter and VK to provide daily updates and intelligence reports on the social media discussion surrounding Trident Juncture and NATO. We focus on three distinct collections of Twitter data: geotagged tweets, tweets by Ministries of Foreign Affairs (MFAs), and news media tweets. This is our summary report for the time period from October 1st through November 15th. The report includes as appendices our briefings made during the exercise.

Table of Contents

1	Introduction	1
2	Data Description	1
2.1	Geo-tagged tweets.....	1
2.1.1	NATO Tweets.....	2
2.2	MFAs and Media	3
3	Key Actors.....	4
4	Tweet Analysis	7
4.1	Tweet Volume.....	7
4.2	Sentiment Trends towards NATO	9
4.3	Topic Analysis	10
4.3.1	Geotagged Tweets LDA Clusters	11
4.3.2	MFA Tweets LDA Clusters	12
4.3.3	Media Tweets LDA Clusters	14
5	Conclusion.....	15
6	References	15
7	Appendix A: Complete list of MFA Twitter handles	16
8	Appendix B: Lessons Learned on Twitter Analytics.....	18
8.1	Effective Collection	18
8.2	Feature of the Twitter Discussion Space	18
8.3	Targeting Audiences in Social Media.....	19
8.3.1	Issues in Targeting an Audience in Twitter	19
8.3.2	Conversation Method.....	19
8.3.3	Communicative Reach Approach	20
9	Appendix C: October 16 2015 TJ Briefing	21
10	Appendix D: October 18 2015 TJ Briefing	23
11	Appendix E: October 19 2015 TJ Briefing.....	25
12	Appendix F: October 21 2015 TJ Briefing.....	36
13	Appendix G: October 22 2015 TJ Briefing	38

14	Appendix H: October 23 2015 TJ Briefing	40
15	Appendix I: October 24 2015 TJ Briefing.....	42
16	Appendix J: October 25 2015 TJ Briefing.....	44
17	Appendix K: October 26 2015 TJ Briefing	46
18	Appendix L: October 27 2015 TJ Briefing - Country	49
19	Appendix M: October 27 2015 TJ Briefing – General.....	52
20	Appendix N: October 28 2015 TJ Briefing	57
21	Appendix O: October 29 2015 TJ Briefing	59
22	Appendix P: October 30 2015 TJ Briefing.....	60

Table 1: Basic statistics of our geotagged tweets collection.....	1
Table 2: keywords used to extract tweets related to NATO in 11 languages.....	2
Table 3: Basic statistics of geotagged tweets related to NATO.....	3
Table 4. Country-level MFA Handle Positions	3
Table 5. MFA Tweet Statistics	4
Table 6. Media Tweet Statistics.....	4
Table 7 .Most active tweeters in Geofenced data	5
Table 8. Accounts with highest mentions in NATO Geofenced dataset	5
Table 9. Most active MFA twitter accounts by sender	6
Table 10. Most active Media twitter accounts by sender.....	6
Table 11: Number of pro/anti-NATO tweets in each country	10
Table 12. Time 1: October 1st through 15th LDA Clusters for Geotagged Tweets.....	11
Table 13. Time 2: October 15th through 31st LDA Clusters for Geotagged Tweets.....	11
Table 14. Time 3: November 1st through 12th LDA Clusters for Geotagged Tweets.....	12
Table 15. Time 4: November 13th through 15th LDA Clusters for Geotagged Tweets ..	12
Table 16. Time 1: October 1st through 15 th LDA Clusters for MFA Tweets.....	12
Table 17. Time 2: October 16th through 31st LDA Clusters for MFA Tweets.....	13
Table 18. Time 3: November 1st through 12th LDA Clusters for MFA Tweets.....	13
Table 19. Time 4: November 13th through 15th LDA Clusters for MFA Tweets	13
Table 20. Time 1: October 1st through 15th LDA Clusters for Media Tweets	14
Table 21. Time 2: October 16th through 31st LDA Clusters for Media Tweets	14
Table 22. Time 3: November 1st through 12th LDA Clusters for Media Tweets	14
Table 23. Time 4: November 13th through 15th LDA Clusters for Media Tweets.....	15
Figure 1: Trends of Russia and Paris topic in NATO-related tweets by day.....	8
Figure 2. Total count of Media and MFA Tweets over time	8
Figure 3: Sentiment trends towards NATO by day	9

1 Introduction

NATO conducted one of its largest exercises over the past decade, Trident Juncture, from October 21st to November 6th 2015. The exercise included over 36 thousand troops from over 28 allies, 9 partner nations, and engaged 18 observations and 12 international and non-governmental organizations and aid agencies. The exercise took place over 16 locations around the Mediterranean in Portugal, Spain, and Italy.

As part of this exercise, the Center for Computational Analysis of Social and Organizational Systems (CASOS) at Carnegie Mellon was asked to assess, in partnership with the Data Mining and Machine Learning Lab at Arizona State, the social media response to Trident Juncture. We collected data from Twitter and VK to provide daily updates and intelligence reports on the social media discussion surrounding Trident Juncture and NATO. We focus on three distinct collections of Twitter data: geotagged tweets, tweets by Ministries of Foreign Affairs (MFAs), and news media tweets. This is our summary report for the time period from October 1st through November 15th.

2 Data Description

We collected data from Twitter in two ways: using a geofence to collect geotagged data across Europe, as well as utilizing the Twitter API to identify tweets with keywords or tweeted by key users. Media and MFA tweets were collected via keywords. Data was collected from October 1st through November 16th.

2.1 Geo-tagged tweets

For geotagged tweets, we collected all tweets from a large area of Europe utilizing Twitter’s Streaming API[1]. We used the geographic boundary box $[13.971488 W, 33.807153 N] \times [41.399602 E, 58.728686 N]$, which covers countries including France, United Kingdom, Spain, Germany, Poland, Ukraine, Italy, Turkey, part of Russia and so on. The geo-tagged tweets we collected are either tagged with a longitude/latitude coordinates or associated with a place defined as a small geographic box like ”Roma”: $[12.234427 E, 41.655874 N] \times [12.855864 E, 42.140959 N]$.

Each day, approximately 5 Gigabytes of data, over 1.5 million tweets were collected. Table 1 shows the average number of tweets, average number of tweeters, hashtags, URL, and tweet mentions as calculated on a daily basis.

Table 1: Basic statistics of our geotagged tweets collection

	# of tweets	# of tweeters	# of hashtags per tweet	# of URLs per tweet	# of mentions per tweet
Average	1,547,044	387,614	0.34	0.24	0.60

Maximum	1,860,426	436,476	0.38	0.27	0.54
Minimum	1,202,798	343,568	0.30	0.21	0.66
Standard Deviation	121,232	16,871	0.018	0.011	0.027

2.1.1 NATO Tweets

Geotagged tweets were filtered to identify tweets related to NATO and Trident Juncture. We identified 12 English keywords that related to NATO and Trident Juncture and translated them into 10 additional languages based on the volume of tweets collected in those languages: Estonian, French, German, Greek, Italian, Lithuanian, Portuguese, Russian, Spanish, and Ukrainian. On average, we found approximately 290 tweets related to NATO on a daily basis from the geotagged dataset.

Table 2: keywords used to extract tweets related to NATO in 11 languages

Estonian	French	German	Greek	Italian	Lithuanian	Portuguese	Russian	Spanish	Ukrainian	English
NATO	OTAN	NATO	Tou NATO	NATO	NATO	NATO	HATO	OTAN	HATO	NATO
Trident pöördpunktis	trident moment	Dreizeck Stelle	τρίαίνα συγκυρία	tridente congiuntura	Trident konjunktūra	tridente conjuntura	трэзубець соединение	tridente conjuntura	тризуб з'єднання	trident juncture
TRJ4	TRJ4	TRJ4	TRJ4	TRJ4	TRJ4	TRJ4	TRJ4	TRJ4	TRJ4	TRJ4
NATO	OTAN	NATO-	Tou NATO	NATO	NATO	OTAN	HATO	OTAN	HATO	NATO's
NATOS	Natos	Natos	Tou NATO	NATOs	Natos	Natos	NATOS	NATOS	NATOS	NATOS
natos	natos	natos	του NATO	Natos	Natos	natos	natos	natos	natos	natos
Põhja-Atlandi Lepingu Organisatsioon	Organisation du Trait é de l'Atlantique Nord	North Atlantic Treaty Organisation	Οργανισμός Βορειοατλαντικού Συμφώνου	North Atlantic Treaty Organisation	Šiaurės Atlanto sutarties organizacija	Organização do Tratado do Atlântico Norte	Организация Северо-Атлантического Договора	Organización del Atlántico Norte	Організація Північноатлантичного договору	north atlantic treaty organization

TJ15	TJ15	TJ15	TJ15	TJ15	TJ15	TJ15	TJ15	TJ15	TJ15	TJ15
NATO_JFCBS	NATO_JFCBS	NATO_JFCBS	NATO_JFCBS	NATO_JFCBS	NATO_JFCBS	NATO_JFCBS	NATO_JFCBS	NATO_JFCBS	NATO_JFCBS	NATO_JFCBS
Trident pöördpunktis 2015	Trident Junctionure 2,015	Trident Junctionure 2015	Trident συγκυρία του 2015	Trident Junctionure 2015	Trident sand ūroje 2015	Trident Junctionure 2015	Тризубец на стыхах 2 015	Trident Соунтура 2015	Тризуб на стихах 2015	Trident Junctionure 2015
TJ 15	TJ 15	TJ 15	TJ 15	TJ 15	TJ 15	TJ 15	TJ 15	TJ 15	TJ 15	TJ 15
tj15	tj15	tj15	tj15	tj15	tj15	tj15	tj15	tj15	tj15	tj15

Table 3: Basic statistics of geotagged tweets related to NATO

	# of tweets (daily)	# of tweeters (daily)	# of hashtags per tweet	# tweets with URLs	# of mentions per tweet
Average	293	197	0.63	0.28	1.1
Maximum	687	463	2.7	0.42	1.8
Minimum	163	132	0.3	0.14	0.49
Standard Deviation	133	76	0.37	0.061	0.30

2.2 MFAs and Media

To identify relevant Twitter discussion of NATO we first identified all official Twitter accounts by identifying NATO and non-NATO member country twitter accounts associated with government legislative and executive functions, as identified in Table 4 below.

Table 4. Country-level MFA Handle Positions

Domestic Government	Ministries	Embassy
<ul style="list-style-type: none"> - Executive leader - Legislative body 	<ul style="list-style-type: none"> - Ministry of Foreign Affairs - Foreign Minister - Ministry of Defense - Defense Minister 	<ul style="list-style-type: none"> - US Embassies in the country - Country embassies in the US - Missions to NATO

Many countries have multiple Foreign Ministry twitter accounts: sometimes one account is used for tweeting in the native language of the country, and the second account is used for English-language press releases. Where we found multiple accounts for a country, we included both accounts in our analysis.

In addition to these country-level official Twitter accounts, we identified the Twitter accounts related to international organizations that also participated in the Trident Juncture exercise. In this report, we collectively refer to all tweets made by these 205 identified twitter handles as “MFA Tweets”. A full list of Twitter handles used can be found in Appendix A.

The table below highlights some summary statistics about the MFA tweets over the time period.

Table 5. MFA Tweet Statistics

	# of tweets (daily)	# of tweeters (daily)	# of hashtags per tweet	% tweets with URLs	# of mentions per tweet
Average	435	42	1.7	96%	0.53
Maximum	1329	76	16	n/a	6
Minimum	5	0	0	n/a	0
Standard Deviation	314	20	1.8	n/a	0.66

We also utilized a list of known official Twitter media handles based on the CASOS Universal Thesaurus. Summary statistics for this set of accounts are listed below. There are 5,934 media handles in the set and they are available upon request.

Table 6. Media Tweet Statistics

	# of tweets (daily)	# of tweeters (daily)	# of hashtags per tweet	% tweets with URLs	# of mentions per tweet
Average	466	20	0.38	97%	0.60
Maximum	4446	76	14	n/a	8
Minimum	0	0	0	n/a	0
Standard Deviation	814	16.5	0.98	n/a	0.54

3 Key Actors

To facilitate comparison across all three datasets, we divided our analysis of the data into three time periods: October 1st through October 15th, October 16th through October 31st, and November 1st through November 16th. In analyzing the geofenced data, we identified certain Twitter accounts as bots.

A major policy concern is how competing groups utilize bots and automated message generators in social media to promote viewpoints, spread malware, and promote commercial activity [6][7][8]. We focused on ongoing conversations about NATO and did not analyze messages clearly generated and produced by bots. We focus on bot structure and information propagation in a separate paper.

In the aggregate geofenced data, “ZhannaPfaiffer” is the most active user across the three periods. However, the user’s profile indicates that this user is a bot with very high probability, so we excluded this user in analyzing the messages

Table 7 .Most active tweeters in Geofenced data

Oct.1-Oct.15		Oct.16-Oct.31		Nov.1-Nov.16	
Screen name	# of tweets	Screen name	# of tweets	Screen name	# of tweets
ZhannaPfaiffer	731	ZhannaPfaiffer	295	ZhannaPfaiffer	153
MsadeceM	147	MsadeceM	104	MsadeceM	74
A_Rockas	143	PegioBelgrade	101	A_Rockas	68
AslnurBarlal	51	A_Rockas	63	HazteEco6	43
mabelle128	45	MarForEUR_AF	56	timelivenews	38

This is the table for most mentioned users in the NATO dataset. “Steven2077”, “twitter” and “YuliaTymoshenko” in period 1 and period 2 are all mentioned by the bot “ZhannaPfaiffer”.

Table 8. Accounts with highest mentions in NATO Geofenced dataset

Oct.1-Oct.15		Oct.16-Oct.31		Nov.1-Nov.16	
Screen name	# of tweets	Screen name	# of tweets	Screen name	# of tweets
NATO	1141	NATO	510	NATO	738
Steven2077	731	Steven2077	295	twitter	306
twitter	731	twitter	295	YuliaTymoshenko	302
YuliaTymoshenko	731	YuliaTymoshenko	295	Steven2077	296
jensstoltenberg	51	UN	38	UN	45
UN	44	NATO_JFCBS	18	ONU_es	40

gazetesozcu	28	odatv	18	RT_com	31
TBMMresmi	20	argirogr	17	NASA	30
Reuters	19	Europarl_EN	15	lorenzo99	26
BritishMonarchy	19	USNATO	13	thinktankurjc	26

Among MFA twitter accounts, Ukraine and Russia are the two most active MFA accounts – followed closely by the French (*francediplo*), German (*AuswaertigesAmt*, *GermanyDiplo*), and UK (*foreignoffice*) Foreign Ministries. Belarus and Kosovo were also quite active.

Table 9. Most active MFA twitter accounts by sender

Oct.1-Oct.15		Oct.16-Oct.31		Nov.1-Nov.16	
Screen name	# of tweets	Screen name	# of tweets	Screen name	# of tweets
MFA_Ukraine	57	MFA_Ukraine	46	MFA_Ukraine	54
Mfa_russia	27	Mfa_russia	23	Mfa_russia	22
francediplo	14	MFAKOSOVO	13	francediplo	16
BelarusMID	11	minbuza	13	minbuza	16
AuswaertigesAmt	11	MAECgob	12	AuswaertigesAmt	15
MSZ_RP	11	foreignoffice	10	foreignoffice	15
GermanyDiplo	11	francediplo	10	GermanyDiplo	13
minbuza	8	BelarusMID, DutchMFA, AuswaertigesAmt, MSZ_RP	9	MFA_KZ	12
foreignoffice	8			MSZ_RP	11
MAECgob	8			BelarusMID, MFAKOSOVO	10

Among the media accounts, there was considerably more variance. The data was filtered to specifically identify tweets related to NATO or Trident Juncture, so there are comparatively fewer tweets captured in our dataset made by media accounts.

Table 10. Most active Media twitter accounts by sender

Oct.1-Oct.15		Oct.16-Oct.31		Nov.1-Nov.16	
Screen name	# of tweets	Screen name	# of tweets	Screen name	# of tweets
RT_com	14	elisabettaly	4	Reuters	2

AlArabiya_Eng	12	lanuovasardegna	3	DailyTimes_DT	2
La_stampa	7	BrookingsFP	3	prensacom	2
Reuters	6	US_EUCOM	3	japantimes	2
USArmyEurope	5	UN	2	Corriereit, thenation, BILD, USArmEurope, OttawaCitizen, la_stampa, WashTimes, FT, MarcellFriesse, Il_Centro	1
GDS_it	5	TheTruthIsViral	2		
CBCNews, SKyNews, abcnews, MoscowTimes, MsAmyMacPherson, XHNews, GulfTimes_QATAR	4	Reuters, MoscowTimes, ReutersIndia, abcnews, manila_bulletin, FT, GDS_it, USAandEurope, AlrtNet, la_stampa, Avvenire_NEI, RT_com, DailyTimes_DT, NPR, TorontoStar, naiz_info, ylenews, alto_adige	1		

4 Tweet Analysis

We first examine the volume of tweets to identify potential biases in the data. We find that Russian involvement in Syria as well as the terror attacks in Paris significantly impact the volume of tweets being analyzed from the public. Then, we analyze the sentiment trends at a country level and analyze topics using LDA.

4.1 Tweet Volume

In the geofenced data, two peaks of tweets appeared around Oct. 5 and Nov. 14 when two big events happened---Russia launched its attack in Syria and terrorists attacked Paris. We excluded the tweets about Russia and Paris from the NATO-related tweets when analyzing tweets for content topics, as there are two obvious topics change happened when these two events happened.

While we see a similar spike in the number of tweets surrounding the October 5th Russian attacks in Syria, we do not see a surge in the number of tweets related to the Paris

terrorist attack on the 13th of November. This is due to the fact that tweets were already filtered to identify NATO topics.

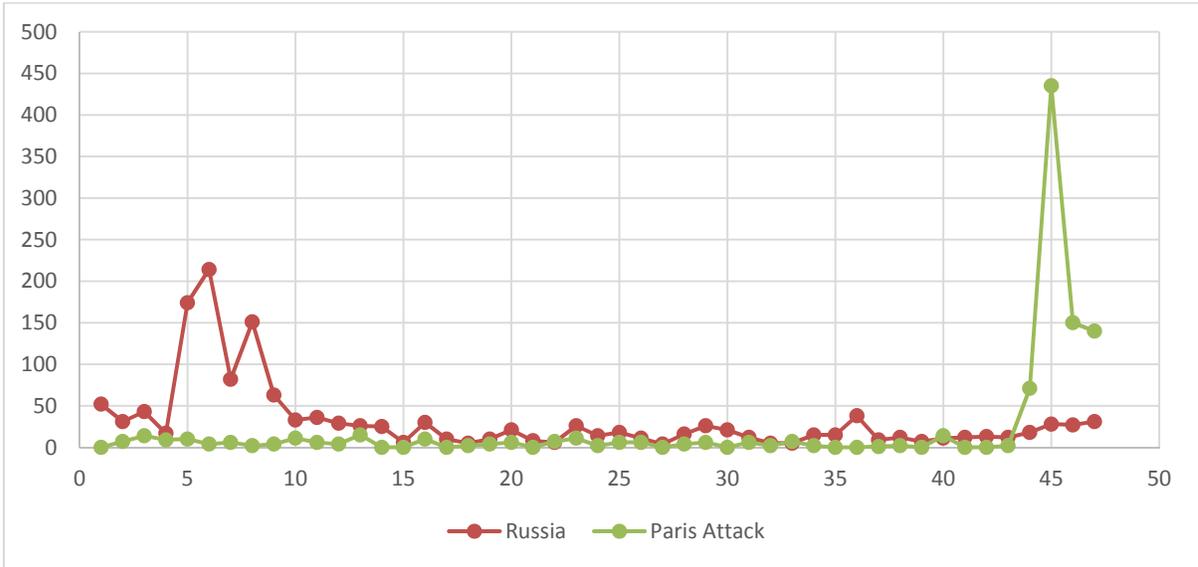


Figure 1: Trends of Russia and Paris topic in NATO-related tweets by day

We find there is a significant spike in media and MFA tweets at the beginning of the collection period, followed by a relatively constant volume of tweets, even during the Paris terror attacks. This is likely due to the fact that tweets were filtered to identify NATO-relevant tweets; while Paris attack tweets dominate topics in the last time period analyzed, we only captured tweets that also mentioned NATO keywords.

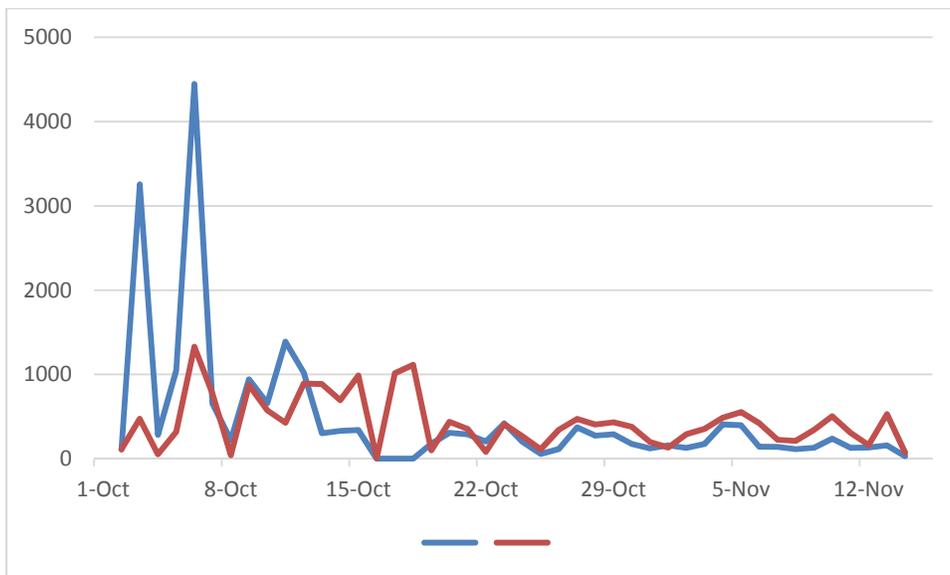


Figure 2. Total count of Media and MFA Tweets over time

4.2 Sentiment Trends towards NATO

To determine the attitudes behind each tweet(positive or negative), we developed a sentiment classifier based on VADER [2] which is a rule-based sentiment analysis model for social media text. We extended it to deal with multiple languages in our dataset. Specifically, we first choose lexicons in English from several sources. The words in the our lexicon were created from a merger of terms found in several different sentiment lexicons, and then were augmented with related terms found in wordnet[3], wikipedia, and various on-line corpus of terms related to sentiment. Each term was classified as positive or negative based on a) the consensus of at least three sources, or when there were not three sources - the consensus opinion of multiple reviewers. There are in total 16,487 terms in our original English lexicon. Then we translated the English terms into 12 languages: Arabic, Estonian, French, German, Greek, Italian, Lithuanian, Portuguese, Russian, Spanish, Turkish, and Ukrainian using Google translate. For terms corresponding to same word in English, they were assigned the same sentiment polarity: negative, positive or neutral.

In this part, we evaluated the polarity of tweets in the geofenced data related to NATO. As is shown in the figure below, the total number negative tweets about NATO is slightly larger than the number of positive one(4576 versus 4002).

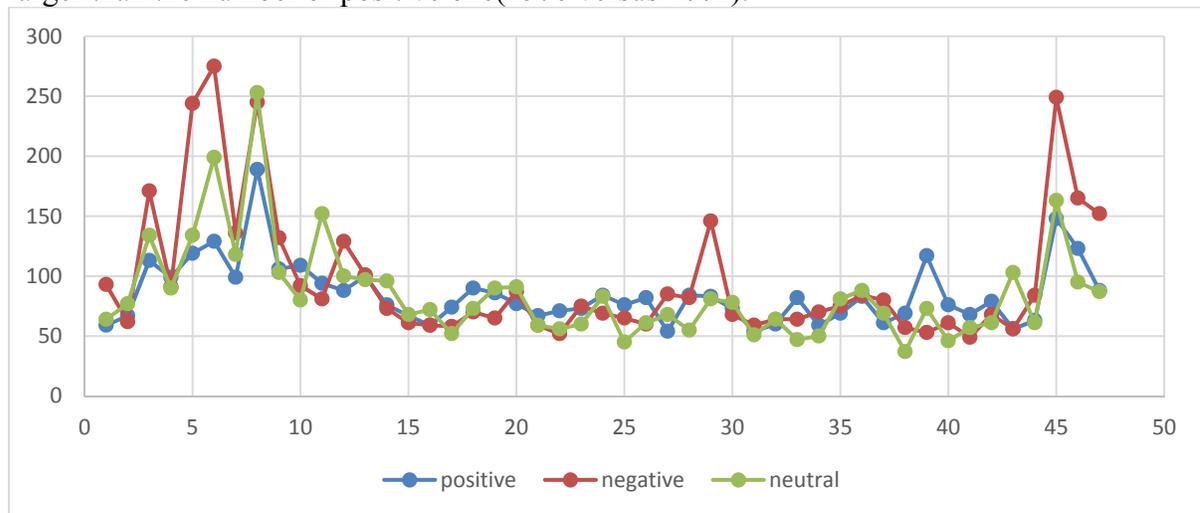


Figure 3: Sentiment trends towards NATO by day

91.2% tweets related to NATO are sent from those 15 countries we listed in Table 11: Number of pro/anti-NATO tweets in each country. Most tweets about NATO come from Spain, Italy, Turkey, Ukraine and UK. Consider the collection area only cover part of Russia, we cannot make a conclusion that tweeters in Russia pay less attention to NATO than those countries. Countries like United Kingdom, Serbia, Germany, Greece

and Montenegro are typical countries where positive tweets about NATO are far less than negative ones.

Table 11: Number of pro/anti-NATO tweets in each country

Country	# of pro-NATO	# of anti-NATO	# of Neutral	# of Tweeters	# of Tweets
Spain	1325	1124	835	1836	3284
Italy	725	620	496	1252	1841
Turkey	502	505	774	812	1781
United Kingdom	326	705	325	565	1356
Russia	154	125	182	217	461
Ukraine	122	89	176	107	387
Serbia	61	204	121	114	386
France	130	158	91	211	379
Germany	66	151	100	153	317
Poland	79	77	137	109	293
Greece	58	107	125	97	290
Portugal	97	89	89	164	275
Netherlands	63	90	79	77	232
Belgium	34	65	62	75	161
Montenegro	13	62	73	35	148
Sweden	36	40	71	81	147

4.3 Topic Analysis

To compare tweet topics we used LDA to find three primary clusters of tweet topics. LDA is a well-known methodology that allows us to find clusters of topics in text datasets over large numbers of documents [4][5]. For geotagged tweets, we used Google Translate to translate hashtags, using the tweet-level language identifier in the Twitter JSON to indicate the Google Translate “from” language, and manually checked that the acronym NATO mapped to the correct topic – i.e. OTAN and HATO mapped to NATO. Then, we imported tweets into ORA and merged Topic nodes based on the translated hashtags. This ensured that tweets about Syria, for example, all mapped to the same #Syria hashtag – regardless of the original tweet’s language, whether French (Syrie), Arabic (سوريا), or Russian (Сирия). These hashtags were used to identify tweet topics.

In our tweets, we initially split topic analysis using LDA into three 15-day time periods to compare and contrast topics over time. However, we found in the topic analysis, the third period (November 1-16) topics became overwhelmed with information on Twitter related to the ISIS attack in Paris on November 13th. To separate out the impact of the Paris attacks, we split the third time period into two sections: November 1-12, and November 13-16.

To highlight changes in topic clusters, we ran LDA for each time period to find three clusters. Each cluster has unique words, shared topics, and common topics. By unique words, we mean hashtags that only appeared in that cluster. By shared topics, we refer specifically to hashtags that appeared in only two out of the three clusters; for common topics, we refer to hashtags that appeared in all three clusters.

Broadly speaking, MFA tweets focused on high-level meetings of multilateral organizations, while tweets from the media focused on much more popular current events topics. For example, MFA tweets would discuss and track the movements of UN Secretary General Ban Ki-Moon and the climate negotiations in Paris, while media tweets would focus on recent catastrophes and breaking news about strikes in Syria.

Currently, in the Twitter data collected from the MFA's we find that Russia is only discussed in the context of ISIS. If the MFA's want to continue to exert pressure in Twitter regarding Russia and Ukraine then they need to continue to tweet about Russia and Ukraine. Without such continued discourse, attention shifts to the topic of the day, and the the potential to use Twitter to continue to exert diplomatic pressure is not being exercised.

4.3.1 Geotagged Tweets LDA Clusters

Table 12. Time 1: October 1st through 15th LDA Clusters for Geotagged Tweets

	Cluster A	Cluster B	Cluster C
Unique	EU		Putin, Kunduz, Afghanistan
Shared	Syria, news, Luiv, TJ15		
		ISIS, MSF	
	Russia		
Common	Trndl, USA, NATO, Turkey		

Clusters B, C, focused on bombing of Doctors Without Borders hospital in Kunduz, Afghanistan; Clusters A and C focused on Russian response. Clusters A and B co-mention TJ15 – Trident Juncture 2015. All clusters discuss NATO, the US, and Turkey.

Table 13. Time 2: October 15th through 31st LDA Clusters for Geotagged Tweets

	Cluster A	Cluster B	Cluster C
Unique	TJ15, notridentjuncture,	Putin, TridentJuncture,	Turkey, Ukraine, Syria, Georgia

	TJ2015, ATA_TJ2015, TRJE15	Serbia, Kosovo, UNESCO, NoKosovoUnesc, Trael,Serbian,USA	
Shared		EU	
	Russia, NATO, Lviv, trndnl, dkpol		

Cluster A more focused on Trident Juncture actors, clusters A and C address Russian response. Cluster B seems more focused on leftovers from NATO involvement in Kosovo.

Table 14. Time 3: November 1st through 12th LDA Clusters for Geotagged Tweets

	Cluster A	Cluster B	Cluster C
Unique	Palm, Syria	News, Tridentjuncture2015	RigaConf
Shared	General		
		Trndnl, Lviv	
	USA, EU		
Common	NATO		

All clusters address Trident Juncture (TJ15). Cluster C more focused on the Riga Conference; Cluster A on nexus of EU involvement with Syria.

Table 15. Time 4: November 13th through 15th LDA Clusters for Geotagged Tweets

	Cluster A	Cluster B	Cluster C
Unique	USA, EU	PrayForPeace, trndnl, BigOil	Daesh
		Turkey	
	France, Hollande		
Common	NATO, Paris, ParisAttack, Russia		

All clusters discussed the ParisAttack ; clusters B and C identified ISIS and Turkey's key role. Cluster A focused more on the role of the US.

4.3.2 MFA Tweets LDA Clusters

Table 16. Time 1: October 1st through 15th LDA Clusters for MFA Tweets

	Cluster A	Cluster B	Cluster C
Unique	natopastavanger1	HMCS, Lavrov,	TwitterKurds, TJ15, Russia,

	5, EU, Kunduz, MH17, Turkey, Estonia, Afghanistan, United For Thanksgiving, allies Ukraine, US, DefMin, Ukraine, Syria	Poland, MH17, Latvia, Turkey, Estonia, SilvanAndNusaybinUnderA ttack,Baltic, BismilUnderAttack,Ankara
Shared	Montenegro	
Common	NATO	

Far more distinct clusters; Trident Juncture was mostly discussed in relation to Turkey.

Table 17. Time 2: October 16th through 31st LDA Clusters for MFA Tweets

	Cluster A	Cluster B	Cluster C
Unique		NATO, Russia, Poland, Montenegro, UNSC, OSCE, BankimoonInSpain	Syria
Shared	TJ15		
	EU, Kosovo, COP21, UN70, UNBlue, Lavrov, Steinmeier		EU, Kosovo, COP21, UN70, UNBlue, Lavrov, Steinmeier
Common	Ukraine, KosovoinUNESCO		

Table 18. Time 3: November 1st through 12th LDA Clusters for MFA Tweets

	Cluster A	Cluster B	Cluster C
Unique	Syria, Ukraine, Steinmeier, ISIS, Georgia, Stage, AlliedStrong, Zakharova	????????? (Belarus), ASEMFMM12	gov, kamerstuk, RigaConf,BerlinFPF
Shared	Russia, EU		
		NATO, TJ15, News, COP21, KosovoinUNESCO, WSF2015	
Common	NATO		

Table 19. Time 4: November 13th through 15th LDA Clusters for MFA Tweets

	Cluster A	Cluster B	Cluster C
Unique	Femforeignpolicy, 24HoursofReality,ModiInUK,	Beirut, EU, Baltic	Syria, kamerstuk, FAC

	Terrorisme		
Shared	Migration		
		Steinmeier, 13novembre	
	Ukraine		Ukraine
Common	UDiSkolan, France, Paris, ParisAttacks		

4.3.3 Media Tweets LDA Clusters

Table 20. Time 1: October 1st through 15th LDA Clusters for Media Tweets

	Cluster A	Cluster B	Cluster C
Unique	Ukraine, Drones, Spain, UAV	US, Russian, BreakingNews	Turkey, News, ISIS
Shared	Kunduz, Rusya		
		Syria, Russia, BREAKING	
	MSF, Syrian		MSF, Syrian
Common	NATO, Afghanistan		

Table 21. Time 2: October 16th through 31st LDA Clusters for Media Tweets

	Cluster A	Cluster B	Cluster C
Unique	MSF, topstories, Contacto ConMaduroNro44	Video, Ukraine, UNBlue	Afghanistan, isCBCoccupied
Shared	NATO, ONU70años		
		Syria, kenfm, ContactoConMaduroNro43	
	News, cdnpoli, ahora		News, cdnpoli, ahora
Common	Russia, Kunduz		

Table 22. Time 3: November 1st through 12th LDA Clusters for Media Tweets

	Cluster A	Cluster B	Cluster C
Unique	Turkey, News, TwitterKurds	Venezuela, EnsayoParaLaVictoria	Fluchtlingsskriese, ContactoConMaduroNro46
Shared	Syria, Russia, Sanctions		
		FreeSavchenko, kenfm, ISIL, ASPA2015	
	NATO, Video, ISIS		NATO, Video, ISIS

Common	Merkel
--------	--------

Table 23. Time 4: November 13th through 15th LDA Clusters for Media Tweets

	Cluster A	Cluster B	Cluster C
Unique	Fotos, ISIL, charliehebdo	Merkel, Mnoal	USA, ClimateChange, Paris
Shared	iostoconVale, Libya		
		ISIS, climatehope	
	Video		Video
Common	24HoursofReality, COP21, ArrancoElHuracanBolivariano, ParisAttacks		

5 Conclusion

While the Trident Juncture exercise only took place over a short 17 days, from this analysis we can see that the online discussion surrounding Trident Juncture started well before the actual exercise and continued afterwards. We have shown that three very distinct collections of data lead to distinct inferences being drawn: MFA discussion is distinct from Media Twitter discussion relating to NATO. We have demonstrated a new approach to handling multilingual tweets in our geotagged data that allows us to do topic clustering on these tweets.

6 References

- [1] <https://dev.twitter.com/streaming/overview>
- [2] Hutto, C. J., and Eric Gilbert. "Vader: A parsimonious rule-based model for sentiment analysis of social media text." *Eighth International AAI Conference on Weblogs and Social Media*. 2014.
- [3] Miller, George A. "WordNet: a lexical database for English." *Communications of the ACM* 38.11 (1995): 39-41.
- [4] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993–1022.
- [5] Wei, W., Joseph, K., Lo, W., & Carley, K. M. (2015). A Bayesian Graphical Model to Discover Latent Events from Twitter. *Ninth International AAI Conference on Web and Social Media*. 2015.
- [6] Grier, C., Thomas, K., Paxson, V., & Zhang, M. (2010). @ spam: the underground on 140 characters or less, 27–37.
- [7] Miller, Z., Dickinson, B., Deitrick, W., Hu, W., & Wang, A. H. (2014). Twitter spammer detection using data stream clustering. *Information Sciences*, 260(0), 64–73.

[8] Berger, J. M., and Jonathon Morgan. "The ISIS Twitter Census." *The Brookings Project on US Relations with the Islamic World, Analysis Paper 20* (2015).

7 Appendix A: Complete list of MFA Twitter handles

VladaRH	MNEGOVPT	NATOHR	NATOSocial
fhollande	MAERomania	ESTNATO	1GENLCorps
Rigas_pils	SlovakiaMFA	franceonu	hqmncne
DGrybauskaite	MZZRS	ITALY_NATO	NATO_MARCOM
premierRP	MAECgob	LV_NATO	NATOCanadaWest
presidencia	TC_Disisleri	LitdelNATO	NATODsg
desdelamoncloa	foreignoffice	MdeKwaasteniet	NATODepSpox
david_cameron	statedept	NORWAYNATO	HQEuroCorps
potus	valismin	PLinNATO	NDC_Research
gouvernementFR	mon_pl	slovenianato	NATO_DefCollege
kormany_hu	MiroslavLajcak	USNATO	SHAPE_NATO
Palazzo_Chigi	PHammondMP	DFATD	eu2015lu
gouv_lu	johnkerry	mzvcr	eu_commission
rijksoverheid	CFOperations	francediplo	epoorg
Polska	Defense_gouv	germanydiplo	ESA
govpt	Defensiemissies	minbuza	eu_eeas
guv_ro	mod_pl	Utenriksdept	EUatUN
Congreso_Es	fap_emfa	MSZ_RP	ArcticCouncil
UKParliament	MO_RS	cabinetofficeuk	coe
SpeakerBoehner	Defensagob	cabinet	commonwealthsec
AlbanianMFA	TC_NATO	NSCPress	dfat
BelgiumMFA	defencehq	FranceOTAN	ecb
MFABulgaria	deptofdefense	NATO	Federation
CanadaFP	GonulVecdi	ANDERSFOGHR	IAEAorg
MVEP_hr	PennyMordauntMP	ccdcoe	interpol_hq
czechmfp	pentagonPressSec	e3acomponent	intlcrimcourt
denmarkdotdk	denmarkinusa	NATO_ACT	OEA_Oficial
estonia_eu	franceintheus	NATO_MARCOM	OSCE
francediplo_en	germanyinusa	NATODEPSPOX	donaltdusk
auswaertigesAmt	greeceinUSA	NATODSG	refugees
greeceMFA	LatEmbInDC	NATOLIBRARY	UN
MFAIceland	LTEmbassyUS	NATOPress	undp
FarnesinaPress	NorwayUS	jensstoltenberg	un_women

Latvian_MFA	PolisyEmbassyUS	NATO_JFCBS	un_spokesperson
LithuaniaMFA	PortugallnUSA	JFC_Naples	ebrd
dutchmfa	USUN	LANDCMD	oecd
NorwayMFA	UNinBrussels		
PolandMFA	canadaNATO		
imohq	embassyli		
AzerbaijanPA	NATOmoscwR		
GovernAndorra	MID_RU		
<u>DrZvizdic</u>	SwedenNato		
cyprusPIO	swiss_un		
valtioneuvosto	post2015_CH		
GovCyprus	SwissHumAidUnit		
FinGovernment	SwissOSCE2014		
governmentGeo	travel_edadfae		
MerrionStreet	georgianembassy		
guvernulRMD	irelandembUSA		
gvtmonaco	kazakhembassy		
MeGovernment	AZEMBASSYUS		
GovernmentRF	AZMISSIONNATO		
diplomacy_RM	E3AComponent		
serbiangov	NATO_SPS		
MFA_Russia	iea		
KremlinRussia_E	imfnews		
SweMFA	FinnEmbassyDC		
post2015_CH			
MFA_Ukraine			
BR_Sprecher			
Utrikesdep			
belarusmid			
MFA_Austria			
AzerbaijanMFA			
blearusmfa			
sebastiankurz			
cyprusmfa			
Ulkoministerio			
MFAgovge			
dfatirl			
MFA_KZ			
MFAKOSOVO			

pressslujba
MFA_LI
NikolaPoposki

8 Appendix B: Lessons Learned on Twitter Analytics

These lessons fall into three categories:

1. How to effectively collect data
2. Features of the Twitter discussion space

Targeting audiences in social media

8.1 Effective Collection

Without access to the entire data stream, care must be taken in how to pull the tweets to identify the trends in what the public is discussing.

If using the streaming API – it is best to run multiple collections.

Political Actors: One stream might be focused on the official and political actor accounts. These actors tweet so little that it is possible to follow and capture all of their tweets. One stream might use key words to capture relevant data.

General Public: The problem here is that as the conversation evolves, the key words of relevance may change. Another problem is that some keywords can generate too much data and your sample will go beyond the allowable 1%.

Geo-tagged: In many cases it is possible within the 1% to capture all geo-tagged tweets. However, these may or may not be representative of the general tweets. They often will have a similar temporal signature and address similar topics.

Most twitter collection tools do not support exporting the tweets for analysis – so all you can do is read them.

8.2 Feature of the Twitter Discussion Space

- Twitter is not the Universe –
 - You need to capture and simultaneously assess many different media feeds
 - You should analyze the impact on the GENERAL PUBLIC separately from the impact on the POLITICAL ACTORS
- Sentiment in tweets is expressed by words, emoticons, emoji, and images. No sentiment miner could handle all the languages, the emoji and the images.

- All sentiment tools are measuring whether the message as a whole is positive or negative and NOT what the messenger is positive or negative ABOUT
- Twitter is dominated by discussions related to sports and entertainment celebrities. This meant for NATO that there were very few tweets about NATO. In general, we found similar number of tweets concerning cyber attacks and NATO.
- Most twitter users do not use geo-tagging so it is difficult to tell where they are, even at the country level. This meant for NATO that most of the tweets concerning NATO could not be attributed as coming from a specific country.
- Twitter data is multi-lingual.
- Tweets are very local event driven. Thus the number of tweets concerning NATO went up prior to the Ukrainian elections and down immediately after.
- Twitter is over-run by bots. Estimates suggest that as many as a quarter of tweeters at any time may be bots. Bots can cause huge spikes in the number of tweets and the apparent sentiment. For example, the jaanpfeffer bot created such spikes for tweets concerning NATO.

8.3 Targeting Audiences in Social Media

8.3.1 Issues in Targeting an Audience in Twitter

- The audience you want may not be on Twitter
- You first need to find the group you want to target
- You cannot guarantee that only that group will attend to your tweets
- Simple targeting can be achieved using the conversation method
- More directed targeting requires crafting messages using communicative reach

8.3.2 Conversation Method

1. Identify the group
 1. What kind of people or issue do you want to address
 2. Collect tweets from this group
2. Identify who they key actors in the group are
 1. Those who are central in the co-mention network

2. Those who are frequent tweeters
3. Identify what the key hashtags are
 1. Those that are used by many group members
 2. Those that are frequently used
4. Send out tweets in which you
 1. Mention one or more of the key actors
 2. Use one of more of the key hashtags

Note: ORA is designed to help you do steps 2 and 3

Note: this approach was effective in looking at the discussion by the MFA's and political actors.

8.3.3 *Communicative Reach Approach*

1. Identify a small number of members of the group
 2. Do a pruned snowball search to identify members
 1. Like was done by CMU CASOS to identify ELF and ISIS
 2. Requires large storage space
 3. Conduct spectral clustering and develop machine learning algorithm to remove actors not of interest
 3. Locate key actors, hashtags and topics
 1. Utilize co-mention network, shared hashtag, and sentiment-LDA
 4. Construct a message
 1. using words that score high in communicate reach vis this group
 2. Mention key actors, hashtags
- Not for the faint of heart!
 - ORA supports step 3. We have been developing techniques to make this approach simple and easy. Prior studies suggest that this will enable the broadest reach.

9 Appendix C: October 16 2015 TJ Briefing

Carnegie Mellon

Trident Juncture 2015 Twitter Report

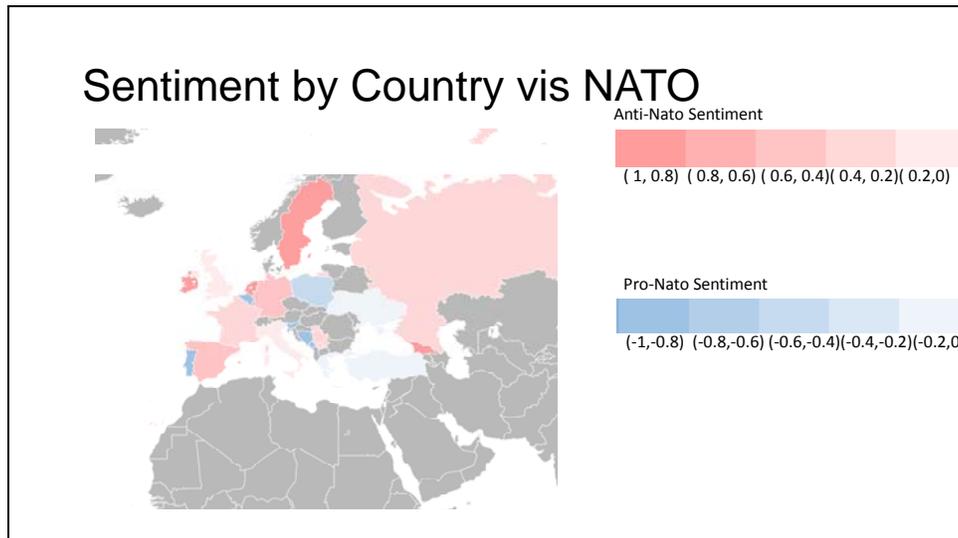
Prepared by Carnegie Mellon University Team at the center for Computational Analysis
of Social and Organizational Systems – CASOS

POC: Dr. Kathleen M. Carley
kathleen.carley@cs.cmu.edu
412-268-6016

Team: Binxuan Huang, William Frankenstein, Matt Benigni



Slide 2

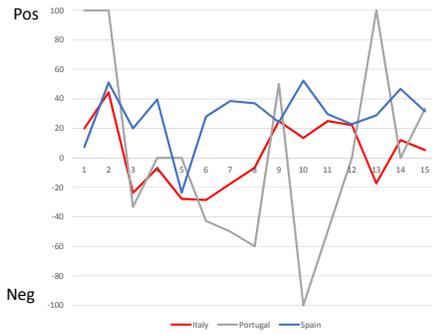


There may be pro and anti-sentiment in each country – but this indicates the difference. So if Blue there is more pro than anti-sentiment, and if red there is more anti- than pro. Gray means we are not tracking those countries or – there is no data from them in our sample.

This is a heatmap of sentiment value of each country.

Slide 3

Sentiment as ratio from Oct 1 to Oct 15



- Generally Spain is Positive
- Italy is less positive
- Wild swings in Portugal

11 Appendix E: October 19 2015 TJ Briefing

Carnegie Mellon

Trident Juncture 2015 Twitter Report 10-18-2015

Prepared by Carnegie Mellon University Team at the center for Computational Analysis
of Social and Organizational Systems – CASOS

POC: Dr. Kathleen M. Carley
kathleen.carley@cs.cmu.edu
412-268-6016

Team: Binxuan Huang, William Frankenstein, Matt Benigni



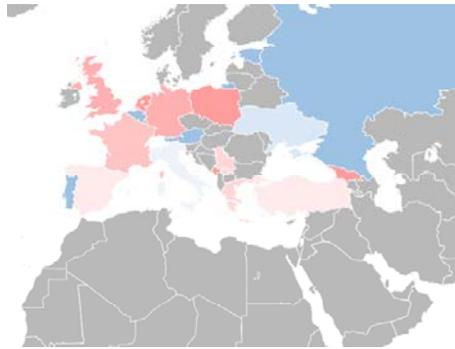
Slide 2

Mapping the “sentiment” regarding NATO in twitter data

- Sentiment: the expression of a favorable (pro or positive) or unfavorable (con or negative) attitude toward a person, place, thing, etc.
- These maps, show the extent to which the tweets from a country that refer to NATO or trident juncture (regardless of the language) contained positive or negative sentiment regarding NATO
- The more saturated the blue, the more positive
- The more saturated the red, the more negative
- Only tweets with geo-tags (latitudes and longitudes) are used.
- All geo-tagged tweets in the region are used
- Sentiment is extracted in the most dominant languages of the region

Slide 3

October 1 2015

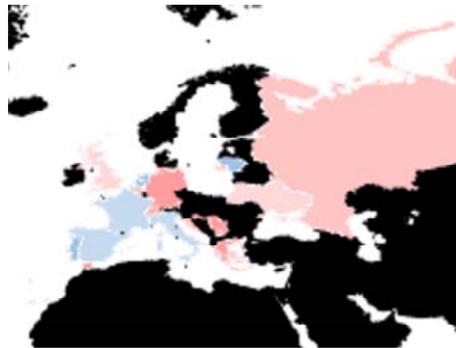


Much of Europe is expressing anti

The darker the color the more sentiment is present
Blue is NATO - pro
Red is NATO - con
Gray is - no pro or con tweets were extracted

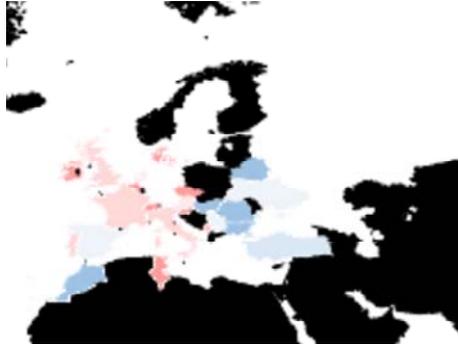
Slide 4

October 2, 2015



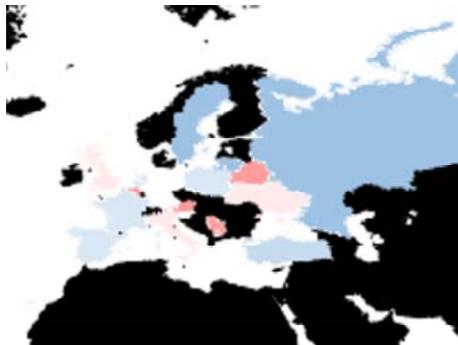
Slide 5

October 3, 2015



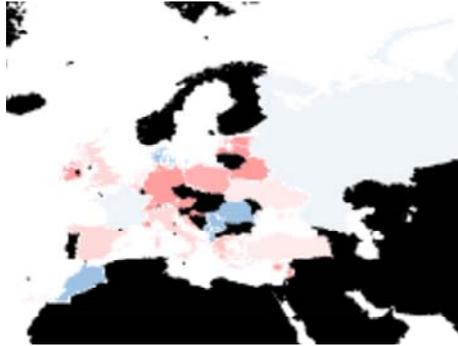
Slide 6

October 4, 2015



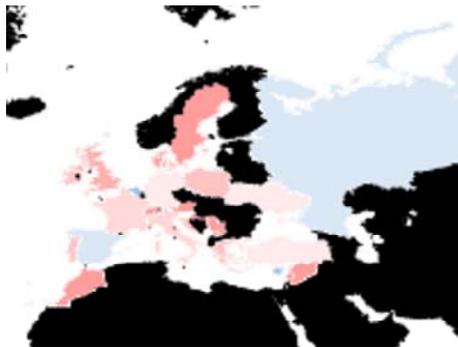
Slide 7

October 5, 2015



Slide 8

October 6, 2015



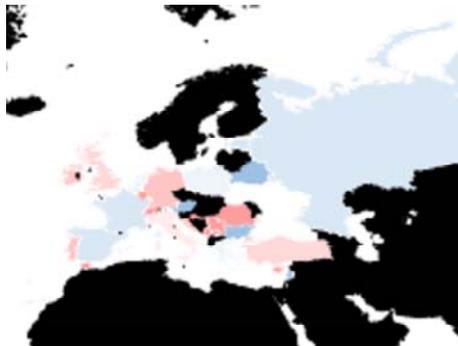
Slide 9

October 7, 2015



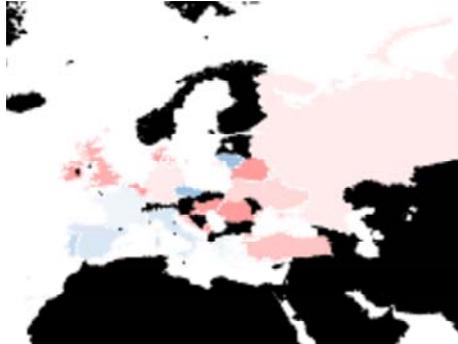
Slide 10

October 8, 2015



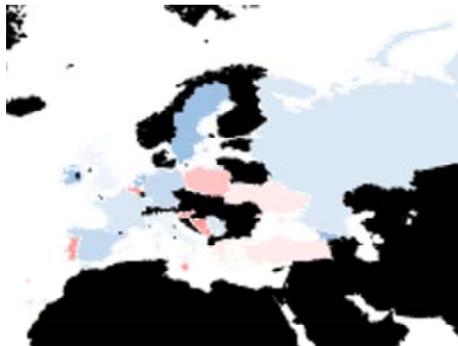
Slide 11

October 9, 2015

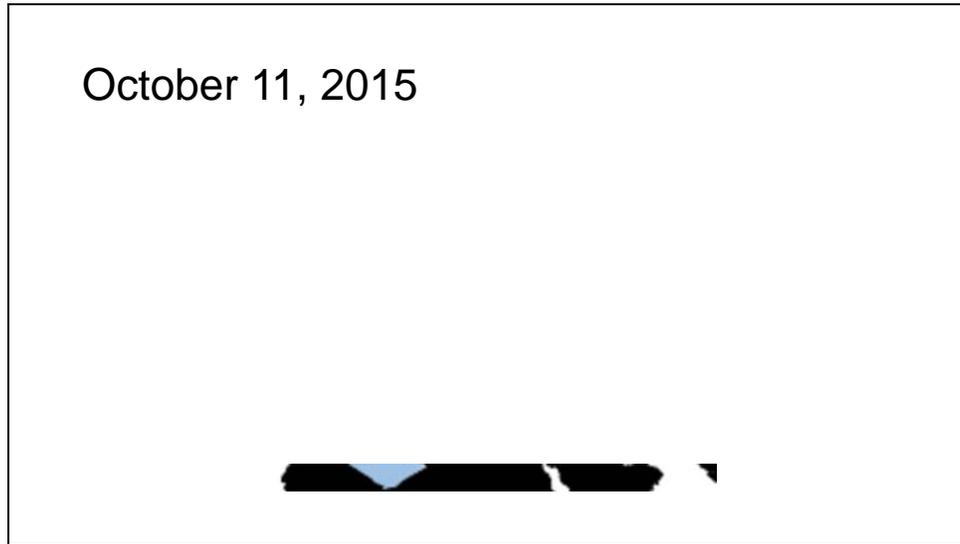


Slide 12

October 10, 2015



Slide 13



Slide 14



Slide 15

