

#Westgate Tweets: A Detailed Study in Information Forensics

By Brittany Card, Justine MacKinnon and Patrick Meier at QCRI¹

INTRODUCTION

Al Shabaab's horrific attack of the Westgate Mall in Nairobi generated over 730,000 tweets during the four-day siege in September 2013. The purpose of this study is to analyze the authors, content and frequency of these tweets in the hour leading up to the attacks and during the two hours after the onslaught began. The Qatar Computing Research Institute (QCRI) partnered with GNIP to collect the 730,000+ tweets within hours of the attack unfolding.² QCRI Research Assistants Ms. Brittany Card and Ms. Justine MacKinnon carried out the subsequent categorization and analysis of tweets under the guidance of QCRI's Director of Social Innovation, Dr. Patrick Meier.

We are actively soliciting feedback on this study to inform the next phase of our research. Our aim is to categorize and study the entire Westgate dataset. To do this, we are considering the use of machine learning for the automatic categorization of tweets and also exploring a potential partnership with iHub Research in Kenya.

RESEARCH QUESTIONS

The research questions that drove this study are listed below. These were formulated by QCRI and by readers who posted suggestions on iRevolution.³

1. Who authored the most tweets? Eyewitnesses, civilians, hostages, local journalists, international journalists, government, police, NGOs?
2. How did the frequency of tweets posted by eyewitnesses change over time and why?

¹ Published by Qatar Computing Research Institute (QCRI) on November 18, 2013. For media enquiries or other information, please contact: pmeier@qf.org.qa.

² Sincere thanks to our partners at GNIP for providing free and rapid access to their platform.

³ Meier, Patrick (2013). "Forensics Analysis of Westgate Tweets," October 3, 2013. Available online: <http://iRevolution.net/2013/10/03/forensics-analysis-of-westgate-tweets>. Ms. Anahi Ayala from Internews also provided input on the research questions.

3. Who were the tweets directed to? Police, Kenyan political figures or agencies, Red Cross, local media, international media, Kenyans in general?
4. What was the content of the tweets? Did they relay rumors/false information or dangerous info? Were there any suspicious tweets or tweets that could have served as early warnings, such as references to the electricity/power being cut? Were Twitter users posting pleas for help or offering help? Do any tweets document killings, injuries, threats? Do some refer to terrorism, religion or hostages? Were there any references to the perpetrators? Did any tweets convey anger and criticism? Did any call for calm and patience or express trust in the government or police (or the opposite)? Did some tweets refer to locations of friends or family? Was Twitter used to request more information? Were there any calls by Twitter users to be more critical of information provided with respect to credibility and veracity?
5. What terms were used to reference the attackers? Terrorists, Al Shabaab, thugs, robbers, gunmen, masked men, attackers, gangsters? And how did the frequency of these terms change over time?
6. What kind of multimedia content was shared during the first two hours of the siege?

DATA & METHODOLOGY

A total of 732,386 tweets were collected by QCRI's Social Computing Center using GNIP's platform. The keywords (hashtags) used to collect these crises tweets were: #Kenya, #Nairobi, #Westagte, #WestagateMall, #WestgateAttack, #WestgatemallAttack, #westgateshootout, #WeAreOne, #Westlands.⁴ The tweets are time-stamped from 12:11 am local time on September 21 to 8:31 am local time on September 25. The attacks are thought to have begun around 12:30 pm local time. This first study in information forensics analyzes tweets posted from 11:30 am on September 21 to 2:30 pm the same day. The total number of tweets posted during this time period was 13,250. We filtered out retweets (duplicates) and all tweets not directly relevant to the attacks. This reduced the total number of tweets to 902. Next, we developed a coding framework to categorize the tweets. This framework, displayed below, was

⁴ Note that the Westgate keyword and hashtag was only used to collect tweets on September 21. All other has keywords and hashtags were used for the full duration of the collection.

developed to answer the research questions outlined above.⁵ QCRI Research Assistants had access to the full tweet, author name, time of tweet and, if available, the location of the tweet.

Table 1: Coding Framework for Westgate Tweets

Code Source of Tweet by: Author of tweet (see bio, past tweets)	Code	Definition
Eyewitness	EW	Authored by eyewitness, bystander, someone in vicinity
Civilian	CV	Authored by a civilian
Hostage	HO	Authored by hostage, someone hiding, someone caught
Local Journalist	LJ	Authored by local/national journalist
International Journalist	IJ	Authored by international journalist
Government	GO	Authored by government official, institution
Police	PO	Authored by police officer, police organization
Civil Society Org	CS	Authored by civil society organization
Terrorist	TE	Authored by terrorist, terrorist organization
Pro-terrorist	PT	Authored by someone pro-terrorism, pro-attack
Other	OT	Other

Code Content of Tweet by: Text refers to	Code	Definition
Actionable/Tactical	AC	Info provides situational awareness to hostage, terrorist, police, etc...
Rumor / False Info	RF	Obvious or possibly false, non-credible info, rumors
Dangerous Info	DI	Info that puts hostage in danger, helps terrorists
Electricity/Power	EL	Messages about electricity/power outages

⁵ Dr. ChaTo Castillo (QCRI) and Anahi Ayala (Internews) kindly provided feedback on early version of the coding framework.

Suspicious/Early Warning	SU	Early indications of possible attack prior to attack
Killing	KI	Messages about killing(s)
Injury	IN	Messages about injury(ies)
Threat	TH	Message related to threats
Religion	RE	Info referring to religion
Hostages	HO	Info related to hostages
Criticism	CR	Messages criticizing police, media, gov, etc...
Terrorism	TE	General messages about terrorism
Need help	NH	Messages asking for help, resources, etc...
Offer of help	OF	Messages offering help, info, resources, etc...
Reference to perpetrators	RP	Messages refer to perpetrators in one way or ther
Call for calm/patience	CC	Tweets that ask people to remain calm and invite people to respect/wait for authorities to do their job
Location of friends/family	LF	Tweets asking where friends/family are or using twitter to responding to inquires
Contradicting terrorists	CT	Tweets that expressly contradict, deny or respond to tweets sent by terrorists organizations
Contradicting government/authorities	CA	Tweets that clearly contradict, deny official information provided by the authorities
Request for more information	RM	Request for more information from authorities or questions about the events
Express trust in the government/police	ET	Tweets that express trust, support and solidarity with the police and the authorities handling the emergency or to the government
Express mistrust in the government/police information	EM	Tweets that clearly express mistrust for the information provided by the police
Incite people to be more critical about the information provided	IC	Tweets that clearly challenge the official version/s
Tweets to (@) Police	TO	
Tweets to (@) Kenyan political figure/agency	TP	

Tweets to (@) Red Cross	TR	
Tweets to (@) Local media	TL	
Tweets to (@) International media	TI	
Tweets to (@) Kenyans	TK	
Tweets to (@) Internationals	TA	
Tweets to (@) Terrorists/terrorist organization	TT	
Other	OT	Other

Code Multi-Media of Tweet by: Tweet links to	Code	Definition
Relevant Picture	P	Picture relevant to attack
Relevant Video	V	Video relevant to attack
Other	OT	Other

Code GPS of Tweet by: Location of tweet author	Code	Definition
GPS Within 1 mile of mall	W	Tweet tagged as user located within one mile of mall but not in mall
GPS inside Mall	I	Tweet suggests user inside mall or very close
GPS Outside 1 mile of mall	O	Tweet tagged as user located outside one mile radius of mall

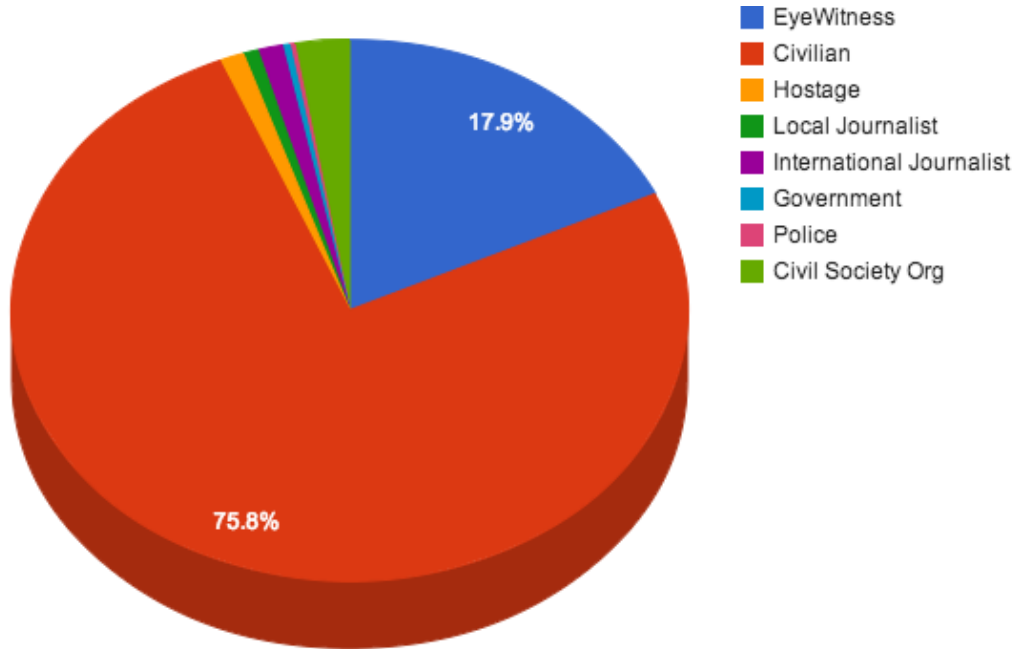
PRELIMINARY RESULTS

A total of 902 relevant tweets were posted between 10:04 am and 2:26 am on September 21. Of these, 6 tweets posted during the hour leading to the attack were categorized as “suspicious”.⁶ A total of 2,642 tweets were posted during the first hour of the attack. Of these, 187 were categorized as relevant. A total of

⁶ The 6 tweets categorized as “suspicious” referenced violence and insecurity in Nairobi and in other areas in Kenya.

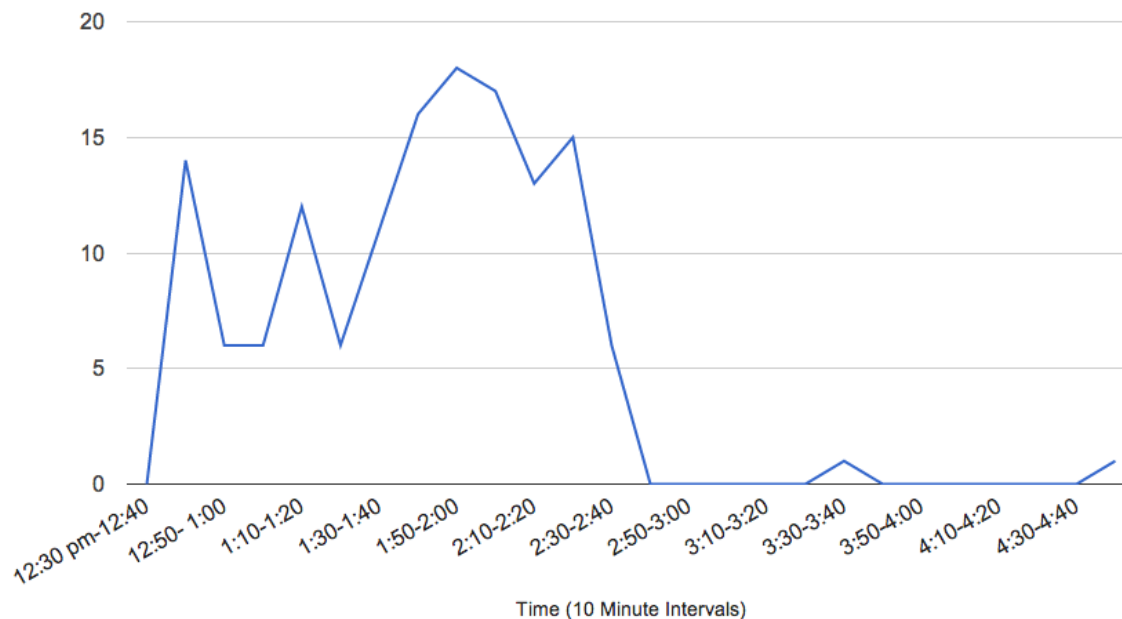
9,434 tweets were posted during the second hour and 420 of these were considered relevant.

Figure 1: Who Authored the Most Tweets?



Civilians posted the vast majority of tweets (75.8%). Surprisingly, eyewitnesses authored the next highest number of tweets (17.9%). Tweets authored by an eyewitness were defined as tweets authored by a bystander, someone in the vicinity or an eyewitness to events relating to the attacks. In total 151 tweets were considered to be authored by eyewitnesses, with 100 unique eyewitnesses responsible for the 151 tweets. 10 tweets were apparently authored by hostages. In these tweets, the messages contained statements about being trapped, in hiding or being rescued by the police. In a couple cases, people tweeted that they were hiding under mattresses in the mall.

Figure 2: Frequency of Tweets by Eyewitnesses Over Time?

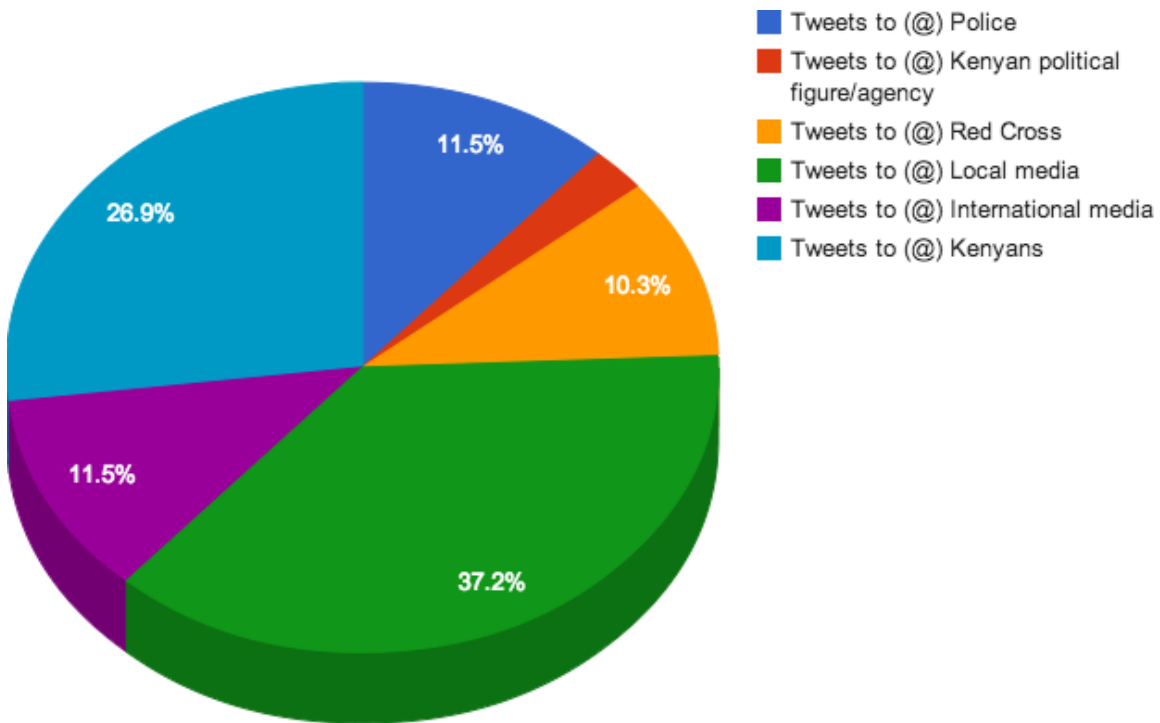


When examined over 10 minute intervals, the majority of eyewitness tweets occurred in the first two hours of the attack as expected. In this time frame, eyewitness tweets remained consistent between 1:40 pm- 2:10 with the highest number occurring between 1:50 pm- 2:00. The significant decline in eyewitness tweets at 3:00 pm and onward may be explained by evacuation and rescue efforts of civilians and hostages from the mall and its surrounding area. For example, the analysis of tweets revealed that police cordoned the roads surrounding the mall while conducting evacuations.

Additionally, the majority of eyewitness tweets revealed information that was considered to be actionable or tactical relating to the actions of Kenyan military, first responders and civilians in the mall. Examples of the types of information communicated are the presence of helicopters, continuing explosions and gunfight, the movement of police and ambulances, and the location of snipers. One tweeted even stated that shoppers were jumping from the 1st and 2nd floor of the mall.

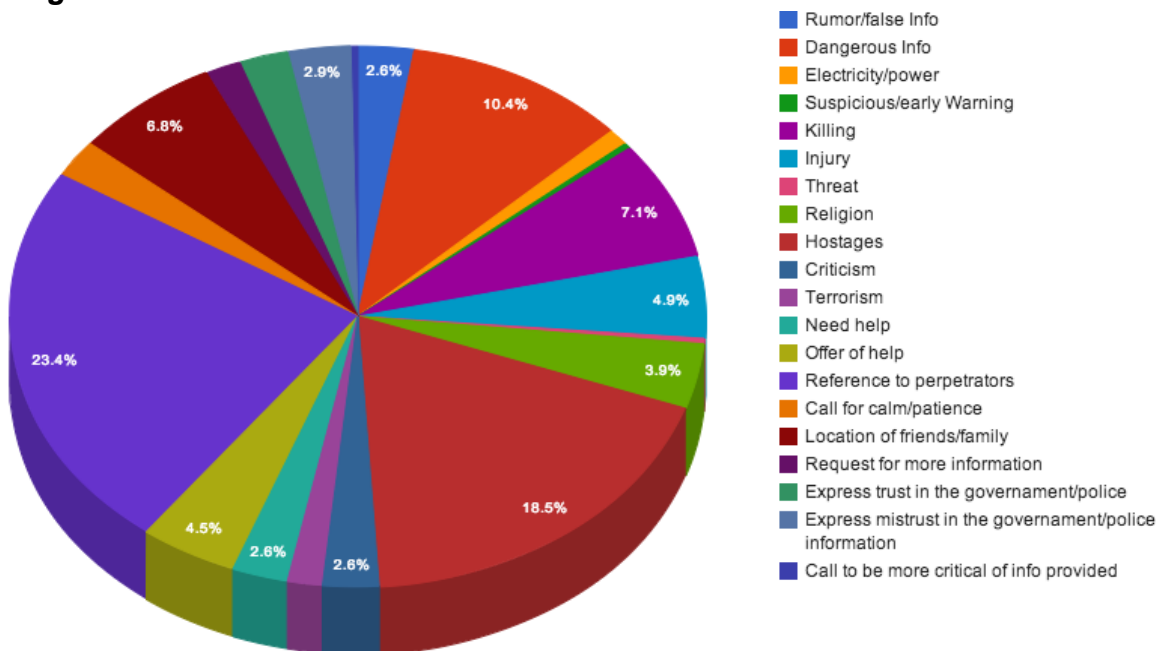
Examination of available coordinates for the location of the tweets revealed that none of the tweets were geo-located inside the mall. Also, only 2 tweets were geo-located within 1 mile of the mall. Both of these tweets were authored by civilians, one mentioned gunshots and the other reported helicopters above Westgate. Location information was not available for the majority of the tweets.

Figure 3: Who Were the Tweets Directed At?



The majority of tweets (853) analyzed were not directed at anyone in particular. Interestingly, the second largest number of tweets (29) were directed at local media followed by 21 tweets to Kenyans. Local media played an extremely fascinating role in the developing events of the attack. Analysis of the relevant tweets from the attack revealed that local media outlets offered live coverage of the attack and used twitter to communicate their coverage. As this coverage happened, people also used to tweets to direct tweets to local media. These tweets included information about the event but also called on local media outlets to be mindful of what they post and show on the television because the gunmen were also likely watching television coverage of the events.

Figure 4: What Content Did Tweets Contain?



The chart above summarizes the content analysis of 595 tweets that contained information relevant to our coding framework. Of these tweets, 64.5% (287) contained information considered to be “actionable” or “tactical”. Actionable or tactical information was defined as information that provides situational awareness to hostage, terrorist, police, etc... These 287 tweets, however, are not included in this chart to allow for a clearer display of categories of content in the remaining 308 tweets. The chart shows that the largest categories after actionable or tactical information are tweets that referenced the perpetrators, related to the hostages and contained dangerous information. The rest of the categories are distributed relatively evenly.

A more detailed analysis of the 287 “actionable” or “tactical” tweets revealed that specific information related to military or a police movement in response to the attack was shared openly on Twitter. Civilians and eyewitnesses shared the majority of actionable information. This information included specific units that were brought in to respond to the attack such as the General Service Unit (GSU) and Securex. GSU, the General Service Unit, is a paramilitary wing of the Kenyan Military and Police and Securex is a security provider in Kenya. Some tweets also referred to units’ locations. Table 2 below summarizes the “actionable” content shared about three different groups: 1) Security Forces; 2) Attackers; and 3) Civilians.

Table 2: Content Analysis of “Actionable” Tweets

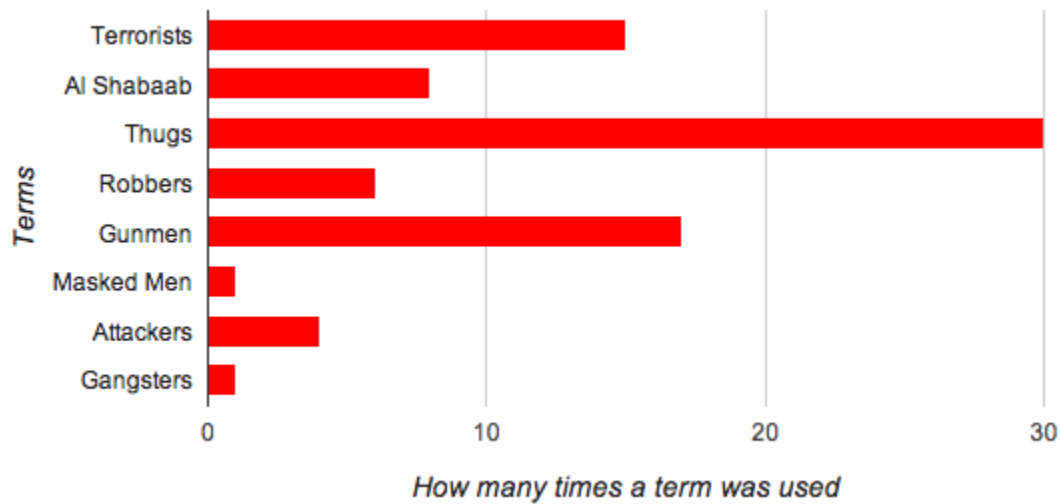
Kenyan Military, Police, Security Forces, First Responders
Police aerial surveillance over Westgate
Presence of military, Kenyan Defence Force (KDF), army and police helicopters
Police cordon of streets surrounding Westgate, specific streets mentioned in some cases
Sharp shooters and GSU forces trained in anti-terrorism moving in to secure the area
General Service Unit (GSU) arrived on scene
Fire brigade arriving
Ongoing security operations
Securex Agency escorting ambulances from 3rd Parklands Ave to Westgate
Ambulances arriving at Westgate
Police evacuating people from the mall
Police take control of ground floor
Kenya Red Cross first to respond, evacuate people
Snipers on the roof of Westgate
Police, military and people inside the mall using social media to communicate

Attackers
Cut electricity before attack
Several big explosions, smaller explosions followed by gunfire
Attackers heavily armed with sophisticated weapons, heavy automatic weapons
Grenade attacks
Control second floor of the mall
Security guard reports the attackers ran out of several cars and into into the mall while shooting
Take hostages, number of hostages inconsistent
Hostages held on top of parking lot

Shooting people inside the mall whose phones ring
Releasing hostages selectively
Have bombs

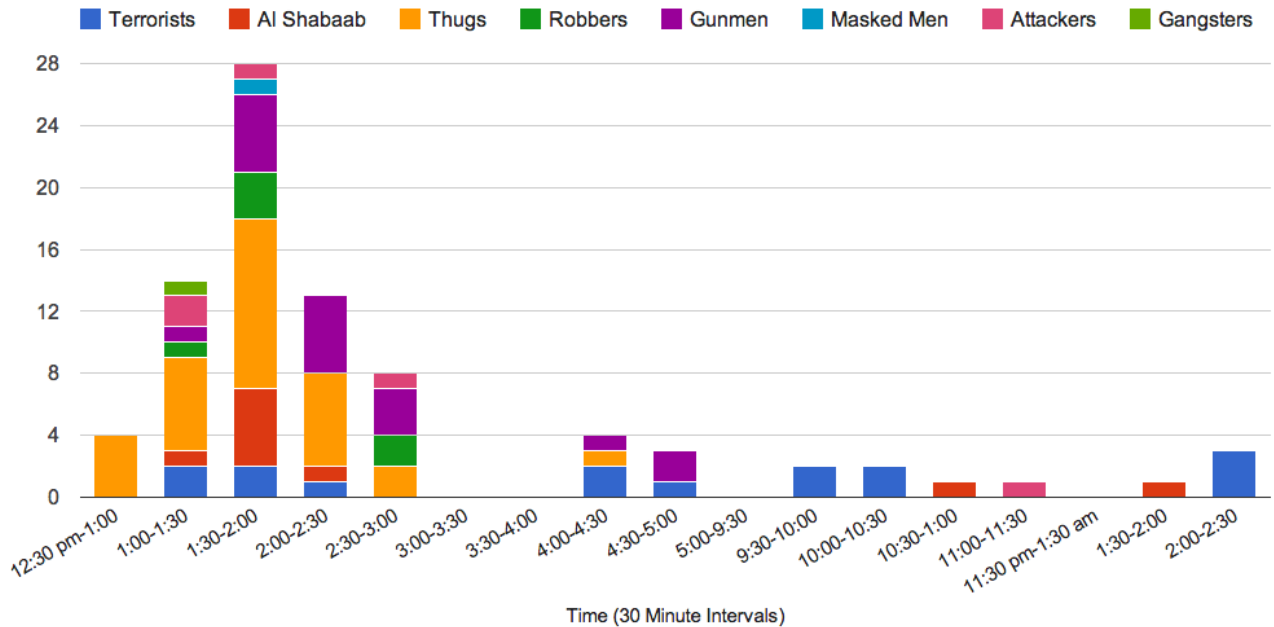
Civilians
Westgate is a "ghost town," abnormally low human traffic
Traffic jams
Roads are clear
Big crowds close to the mall

Figure 5: What Terms Were Used to Reference the Attackers?



The term most often used (30 times) to reference the attackers was “thugs.” The terms “gunmen” and “terrorists” were also frequently used to describe the attackers, with each used 17 and 15 times, respectively. It is fascinating that the term most used to describe the attackers was a more general term used to describe a criminal or violent person. However, terms that denote specific acts carried out by the attackers were used less frequently.

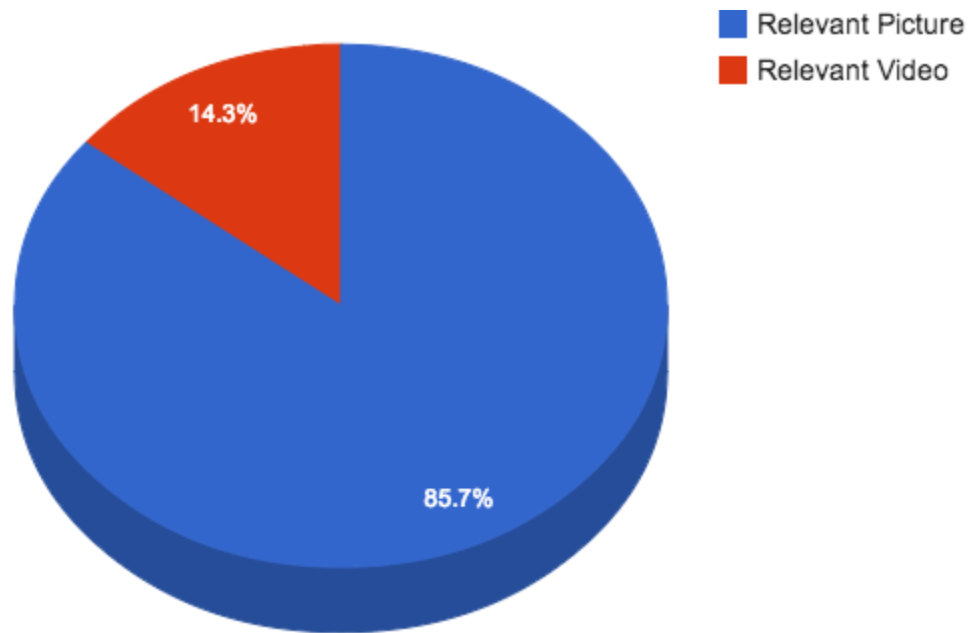
Figure 6: What Terms Were Used to Reference the Attackers Over Time?



The chart above examines the frequency at which terms were used to reference the attackers over 30 minute intervals. The term most used, thugs, increased consistently from when the attack started, approximately 12:40 pm, until 2:00. Within this timeframe, thugs was most frequent used between 1:30 pm- 2:00. The term terrorists was also consistently used throughout the first day of the attack.

It is surprising that the term Al Shabaab was most frequently used between 1:30 pm- 2:00 but severely declines over the rest of the day. The increased use of this term to describe the attackers may reveal the timeframe that it began to be reported that Al Shabaab was responsible for the attack. However, it is interesting that the term Al Shabaab was not consistently used to describe the attackers throughout the day despite their reported involvement. Instead, more general terms like gunmen or terrorists were used.

Figure 7: What Kind of Multimedia Content Was Shared?



The first picture was shared less than 10 minutes after the attack started. The caption states the picture is of Westgate employees fleeing from gunshots and grenades. In the first hour of the attack, a picture was also shared of an open back truck with the caption “GSU have just arrived in westlands.” Throughout the day following the attack, graphic pictures from inside the mall were also shared from the Daily Post and the New York Times.

CONCLUSION

This study reveals that twitter was used to share and seek information about the Westgate attack quickly after it began, particularly in the first two hours after the attack started. Our analysis reveals that mostly civilians, including eyewitnesses, used twitter to communicate information about the developing situation. The majority of these tweets contained actionable or tactical information that provided both general and specific situational awareness.

Twitter was also used to seek out information regarding the location of friends or family members. In some cases, people replied to these inquires about their safety using twitter. Additionally, the narrative of events displayed through this twitter analysis shows that local media outlets played an important role. While the local media also used twitter to share information about the attack, people also

used twitter to ask the media to be mindful and sensitive about what they were sharing through its live coverage because the attackers were also probably watching the coverage.

During this study, the tragic human suffering that occurred as a result of the Westgate attack was not forgotten. We hope that this analysis will help inform the responsible use of twitter during public emergencies and that this sheds light on the potential use of twitter for emergency response in such settings. We are actively soliciting feedback on this study to inform the next phase of our research in which we will code and analyze the remaining tweets. Our aim is to categorize and study the entire Westgate dataset.