# Using Social Media Data to Understand Changes in Gender Norms

USAID
FROM THE AMERICAN PEOPLE

PEPFAR

MEASURE Evaluation

# Using Social Media Data to Understand Changes in Gender Norms

**Brittany Iskarpatyoti,** MPH
**Heather Biehl,** BS
**John Spencer,** MS

**MEASURE** Evaluation
University of North Carolina at Chapel Hill
123 West Franklin Street, Suite 330
Chapel Hill, North Carolina 27516
Phone: +1-919-445-9350
**measure@unc.edu**
**www.measureevaluation.org**

This publication was produced with the support of the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation cooperative agreement AID-OAA-L-14-00004. MEASURE Evaluation is implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of USAID or the United States government. TR-18-295
ISBN: 978-1-64232-092-3

# ACKNOWLEDGMENTS

Thanks to the United States Agency for International Development (USAID) and the United States President's Emergency Plan for AIDS Relief (PEPFAR) for funding this research.

A special thank you to John Spencer and Becky Wilkes, of MEASURE Evaluation, for their efforts to identify data sources and appropriate abstraction methods and for working with third-party vendors and obtaining and cleaning the data. Brittany Iskarpatyoti and Heather Biehl, both also of MEASURE Evaluation, were responsible for qualitative coding. Qualitative and quantitative analysis was conducted by Ms. Iskarpatyoti. Thanks to Jessica Fehringer and Carolina Mejia, of MEASURE Evaluation, for their formative contributions and review.

Thanks to MEASURE Evaluation's knowledge management team for editorial, design, and production services.

**Suggested citation:**

Iskarpatyoti, B., Biehl, H., & Spencer, J. (2018). Using Social Media Data to Understand Changes in Gender Norms. Chapel Hill, NC, USA: MEASURE Evaluation.

# CONTENTS

## FIGURES



## TABLES

# ABBREVIATIONS

| | |
|---|---|
| ART | antiretroviral therapy |
| GBV | gender-based violence |
| HTC | HIV testing and counseling |
| PEPFAR | United States President's Emergency Plan for AIDS Relief |
| STI | sexually transmitted infection |
| USAID | United States Agency for International Development |

# EXECUTIVE SUMMARY

As access to digital devices grows around the globe, social media platforms have emerged as tools to better understand a broad array of health issues. Social media have been harnessed to surveil infectious disease; forecast asthma exacerbations; monitor mental health following traumatic events; and gauge attitudes and beliefs about medical conditions, products, and treatments. (Gruebner, 2017; Bhattacharya, 2012; Nsoesie, 2016; Kuehn, 2015) Gender norms—the socialized expectations about women, men, boys, and girls and the power dynamics between them—are a frontier for social media research. Although many national and international programs seek to change harmful gender norms to achieve equities in health, few collect data on social attitudes. Because social media allow people to share details about their lives, these platforms may provide insights into attitudinal and behavioral aspects of gender inequality and capture information that is difficult and costly to obtain through surveys.

We explored the feasibility of using large social media data sets to track changes in attitudes toward and gender norms related to sexual relationships between younger women and older men ("blessers" or "sugar daddies") and gender-based violence (GBV) against women and girls in 10 sub-Saharan African counties (Kenya, Lesotho, Malawi, Mozambique, South Africa, Swaziland, Tanzania, Uganda, Zambia, and Zimbabwe). This study assessed the challenges and limitations of using social media data to track changes over time in attitudes and gender norms, the type of information that can be gathered from Twitter to track changes in gender norms, and the emerging themes and thematic variations from tweets about age-discordant relationships or GBV.

After reviewing several possible social media platforms including Facebook, Snapchat, and Instagram, we selected Twitter for its data availability and ease of access. We used and assessed different Twitter abstraction platforms and methods including accessing data through vendors (Crimson Hexagon and Union Metrics) and workflow automation tools (Microsoft Flow). While Microsoft Flow is an effective method of obtaining data, we encountered challenges owing to shifting privacy policies that created data gaps. Data from Crimson Hexagon and Union Metrics were used in this study.

We conducted validity assessments to understand how relevant the data were to the intended topic. Of the 1,777 subsampled blesser tweets from Crimson Hexagon, 11 (0.6%) were found to be irrelevant to the research topic. Of the 1,100 subsampled GBV tweets from Union Metrics, 65 (5.9%) were found to be irrelevant to the research topic.

We assessed the type of information that can be gathered on gender norms from social media to track changes in gender norms through overall tweet activity, tweet metrics, engagement metrics and tweet language metrics. Overall tweet activity was recorded as the number of tweets and users, which were related to specific time periods. The tweet metric analysis was performed by retrieving statistics about frequency of tweets with links, tweets with replies, and tweets where Twitter users were mentioned. Engagement metrics were retrieved by obtaining the number of users who tweeted over a set time period, grouped by the number of tweets sent per user, the number of followers they have, and so forth. Language metrics, such as most common words, were visualized using an online word cloud generator.

To quantify attitudes and emotions, sentiment analysis was conducted. We compared computer algorithm-generated sentiment provided by Crimson Hexagon to human-coded sentiment. Overall, less than half (n=723, 41%) of the coding matched. Hand-coded sentiment was used in analysis.

Data collected on blessers from Crimson Hexagon and Union Metrics and on GBV from Union Metrics primarily came from South Africa (97.6%, 75.6%, 77.8% respectively). This coincides with number of people online and using social media across the countries sampled (Portland, 2016). Timing patterns and frequency followed popular new reports and media events. These could be compared to net sentiment to find, for example, that an uptick in Twitter activity on blessers corresponded to a news report and that conversation was primarily negative in sentiment.

Key findings from data on "blessers" and GBV are shown on the following pages.

**Figure 1. "Blesser" tweets from Crimson Hexagon, January 2016 to January 2017**

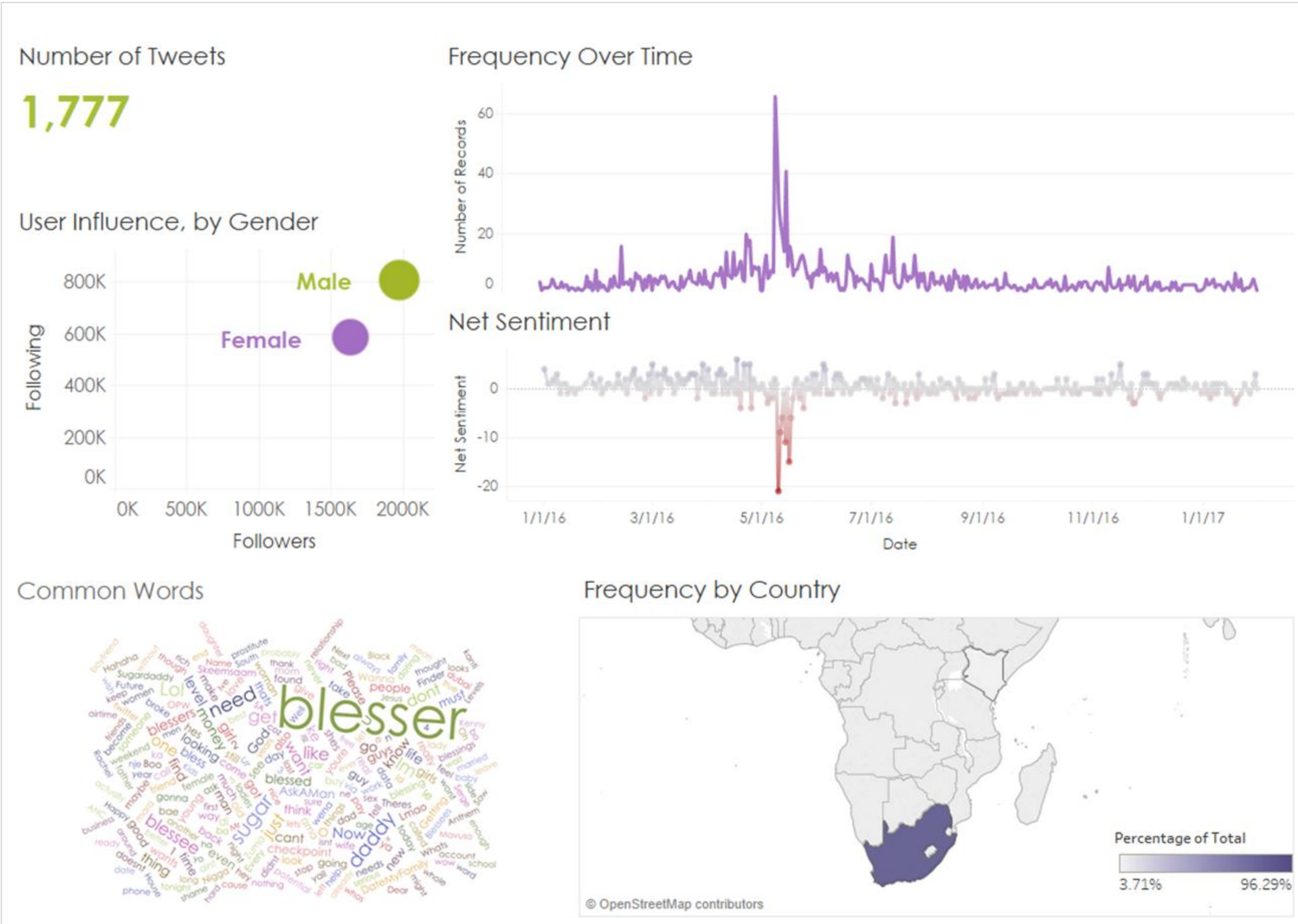


**Figure 2. "Blesser" tweets from Union Metrics, November 2017 to January 2018**

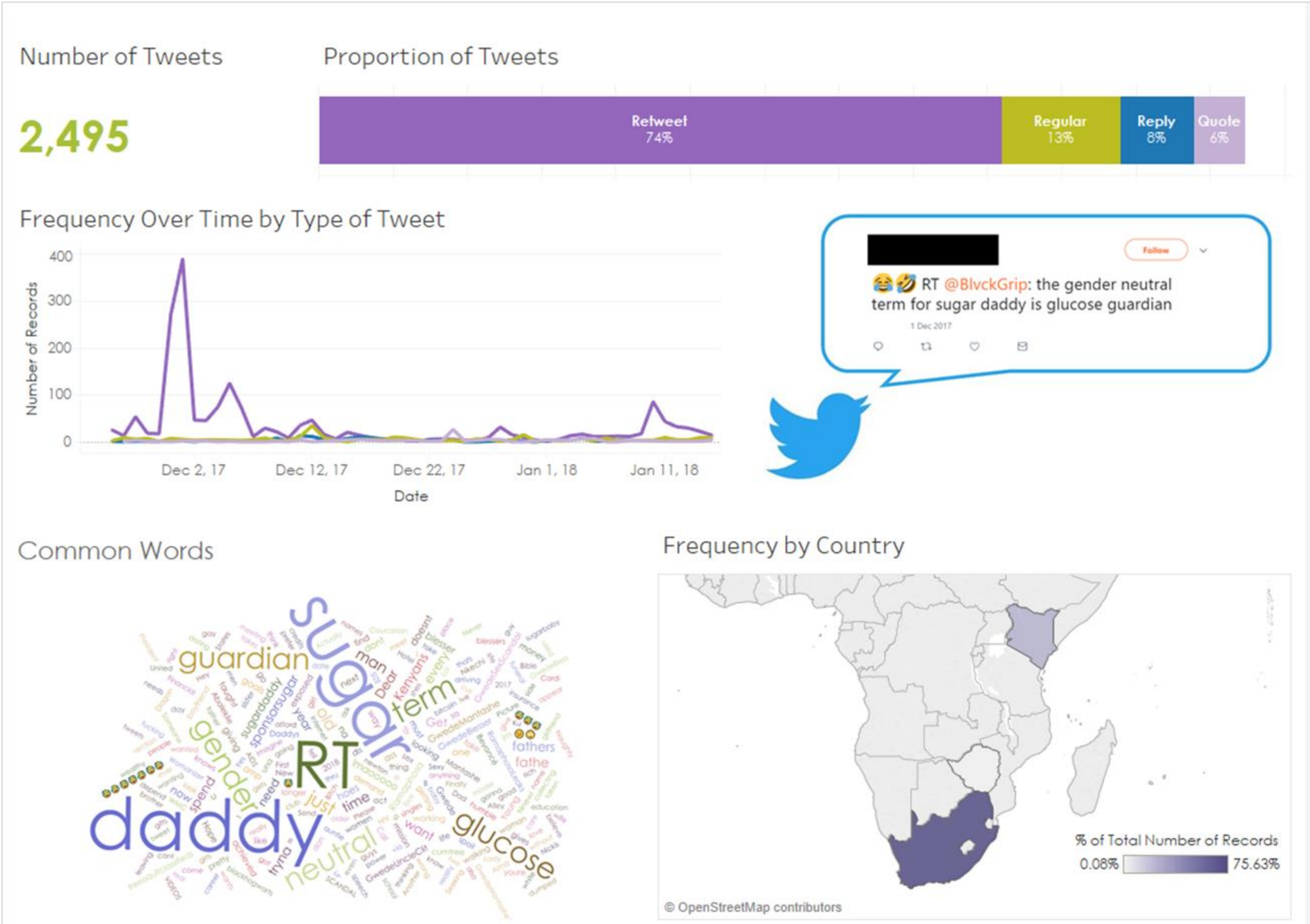
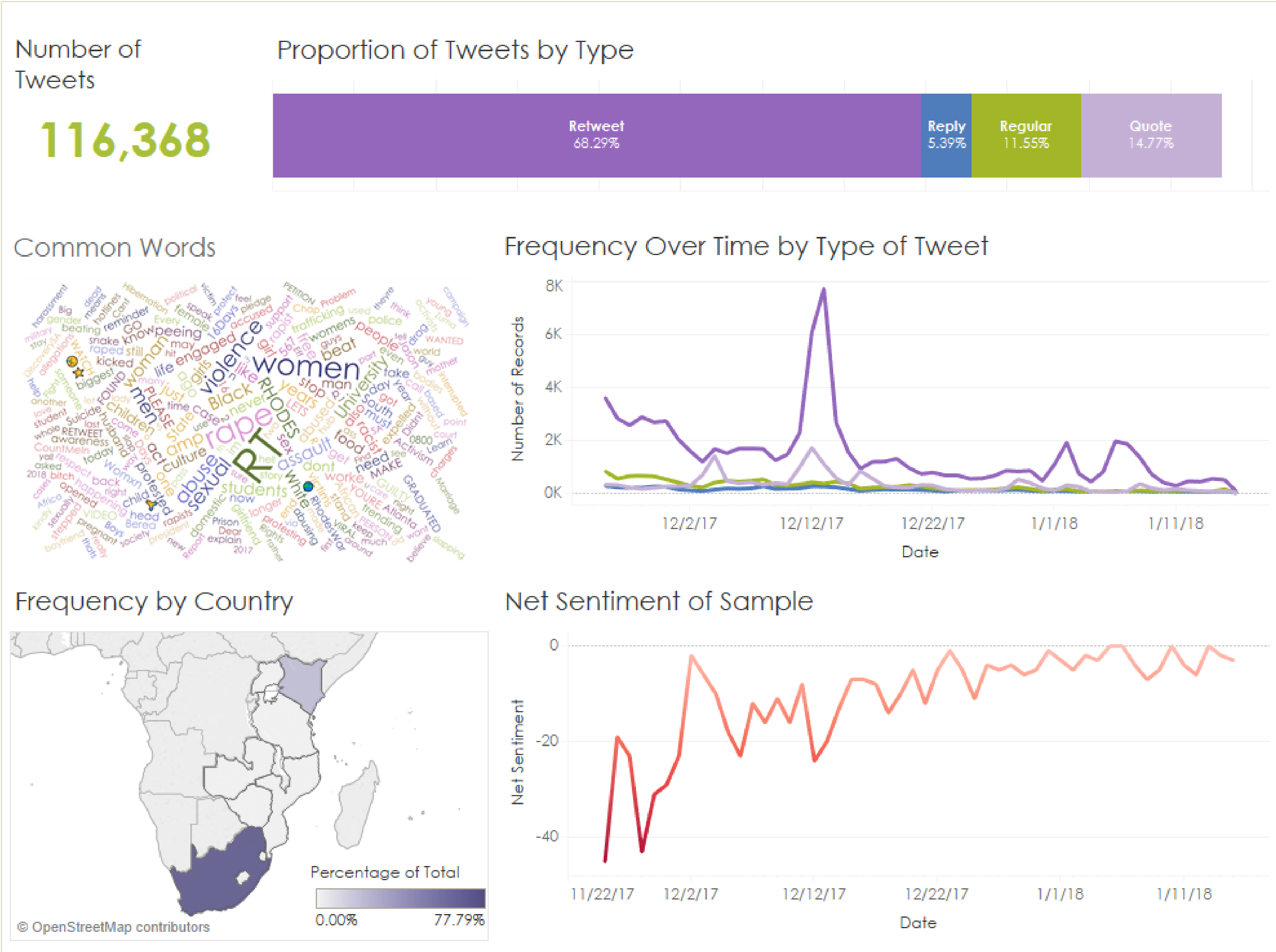


**Figure 3. GBV tweets from Union Metrics, November 2017 to January 2018**



Analysis of social media trends in timing, sentiment, and conversation themes can uncover user attitudes toward gender norms that drive health behaviors. Understanding how, when, and under what circumstances people talk about gender norms and behaviors can help programs tailor messaging and interventions more precisely, to reduce negative social and health outcomes. However, challenges remain in using social media data, including ethical concerns, shifting policies and procedures, and generalizability of findings.

# BACKGROUND

As access to digital devices grows around the globe, social media platforms have emerged as tools to better understand a broad array of health issues. Social media have been harnessed to surveil infectious disease; forecast asthma exacerbations; monitor mental health following traumatic events; and gauge attitudes and beliefs about medical conditions, products, and treatments (Gruebner, 2017; Bhattacharya, 2012; Nsoesie, 2016; Kuehn, 2015). In contexts where traditional information sources such as surveys, interviews, or direct observations are not feasible owing to timing, cost, or geography, social media may constitute a rapid, affordable, and accessible data source (Gruebner, 2017).

Gender norms—the socialized expectations about women, men, boys, and girls and the power dynamics between them—are a frontier for social media research (WHO, 2015). The United States President's Emergency Plan for AIDS Relief (PEPFAR) states that gender norms increase the vulnerability of women and girls to HIV when they "restrict women's access to HIV/AIDS information and services; severely limit women's control over their sexual lives, leaving them vulnerable to sexual violence and abuse and putting them at increased risk of HIV transmission; and deprive them of economic resources and legal rights necessary to protect themselves from HIV/AIDS" (PEPFAR, 2007). Fifty-nine percent of people living with HIV/AIDS in eastern and southern Africa are female, and young women (ages 14–24) constitute 74 percent of new infections in eastern Africa and 91 percent of new infections in southern Africa. (UNAIDS, 2017; UN WOMEN, 2017) Although gender norms are central to our understanding of sexual behavior, GBV, and HIV transmission, data collection on gender norms remains limited.

Prevention of GBV is of interest to researchers because of the known associations between gender norms, GBV, and HIV transmission. It is estimated that 28 percent–60 percent of women ages 15–49 years in sub-Saharan Africa have experienced physical violence; Estimates increase to 30 percent–68 percent with the inclusion of sexual and emotional violence (Borwankar, 2017). However, reliable and valid data on GBV are hard to find—particularly in low-resource settings—owing to the lack of a unified definition, methodological challenges to measuring frequency and duration, and under- or nonreporting because of stigmatization (Kilonzo, 2009; UN Women, 2005). Social media presents one alternative strategy for data collection. Preliminary studies indicate that this platform contains substantial GBV content, and that such content often tracks social movements and transient political or celebrity-related events (Elsherief, 2017; Purohit, 2016).

Literature on GBV in sub-Saharan Africa has frequently examined age-discordant relationships between young women and older men ("blessers" or "sugar daddies"), which are often characterized by material transaction: an exchange of gifts or money from the older man to the younger woman. Three principal paradigms have been used to conceptualize age-discordant relationships, which typically involve material transactions (Stoebenau, 2016). The first is the "sex for basic needs" paradigm, in which women are constructed as victims of predatory older men. The second paradigm is "sex for improved social status," in which women agentively seek age-discordant relationships as a means to material goods and social ascension. Finally, the "sex and material expressions of love" paradigm complicates the relationships between love, money, and gendered power to position material exchange as central to socialized concepts of romantic love (Purohit, 2016).

Age-discordant and transactional sex has been positively associated with HIV infection, adolescent pregnancy, and intimate partner violence (IPV) (Fox, 2007; Toska, 2015; Shefer, 2012; Dunkle, 2004). Thus, programs have historically targeted young women as a vulnerable population seeking "sex for basic needs," intervening at the behavioral level to shift sexual practices or to promote condom negotiation.[15, 16] However, qualitative evidence suggests a need to complicate traditional understandings of age-discordant relationships.

Participants are constructed not as "sex workers" and "clients," but as "girlfriends" and "boyfriends" (Hunter, 2002). Studies indicate that young women seek blessers not for subsistence, but for "fashionable clothes and prestige" (Fox, 2007). Blessers, for their part, are considered *isoka*, or "successful man with girls" idealizing young women as beautiful and pure (Hunter, 2002; Bhana, 2011). Thus, transactional sex in sub-Saharan Africa is intertwined with gendered understandings of love. Current HIV prevention programs, therefore, may not align with the cultural realities of young women and men. Interventions responsive to the complex dynamics of gender, power, and sexuality distal to HIV transmission are needed.

The analysis of gender norms through data is critical to pattern measurement, community-wide engagement, public opinion, and expression sensing, as well as to designing data-driven policies for raising awareness (Esty, 2007), PEPFAR and the associated DREAMS Initiative (Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe women) have collected program-level data using the GEND_NORM and GEND_GBV indicators,[1] a process requiring extensive resources, time, and the cooperation of partner programs (PEPFAR, 2009). However, the 2015 PEPFAR Monitoring, Evaluation, and Reporting Indicator Reference Guide dropped the GEND_NORM indicator (PEPFAR, 2017).

Such challenges prompt investigation into alternative sources of data. Existing literature suggests that social media may constitute a more rapid and affordable source of gender-related data (Mueller, 2017). This study addresses three major questions regarding the feasibility of using social media data to measure changes in gender norms. First, what are the challenges and limitations of using social media data to track changes over time in attitudes and gender norms? Second, what type of information can be gathered from social media to track changes in gender norms? Last, what are the emerging themes and thematic variations from social media posts about age-discordant relationships or GBV?

Companion guidance to this report that describes the process of and considerations for collecting, analyzing, and reporting Twitter data is available at [https://www.measureevaluation.org/resources/publications/ms-18-147/](https://www.measureevaluation.org/resources/publications/ms-18-147/).

[1] GEND_NORM and GEND_GBV were PEPFAR program indicators that measured the number of people completing an intervention pertaining to gender norms that meets minimum criteria and number of people receiving post-GBV care, respectively.

# METHODS

## Social Media Platform

We explored a variety of social media platforms including Twitter, Facebook, Instagram, and Snapchat to assess which had available, accessible data on gender norms from 10 sub-Saharan African counties (Kenya, Lesotho, Malawi, Mozambique, South Africa, Swaziland, Tanzania, Uganda, Zambia, and Zimbabwe). Certain platforms did not have a strong user base (e.g., Snapchat) or applicable content for analysis (e.g., Instagram). Facebook user protections prevented the abstraction of data from large numbers. Because of the reach, applicability, and accessibility of data, we selected Twitter as our social media source.

## Data Sources and Study Design

Data can be extracted from Twitter through online vendors such as Crimson Hexagon and Union Metrics; through automation tools such as Microsoft Flow; or by writing a script in R, Python or another language. Different methods or sources lend themselves to different study designs. We chose not to use written script methods, as they are time-consuming and require technical expertise beyond those available to most programs and projects. We assessed the feasibility of using Microsoft Flow, and ran several abstractions. However, any changes to Twitter privacy settings or policies interrupted data collection and led to several time gaps in collection. These issues are described more in our companion guidance (MEASURE Evaluation, 2018). To analyze attitudes towards GBV and blessers, we used two different data extraction methods or sources. Due to cost and time limitations, GBV data were not collected through retrospective-only means (e.g., Crimson Hexagon). Table 1 summarizes this process.

**Table 1. Summary of study design and sources**

| | "Blessers" | GBV |
|---|---|---|
| Study design | Data source | Data source |
| Retrospective | Crimson Hexagon | n/a |
| Retrospective + prospective | Union Metrics | Union Metrics |

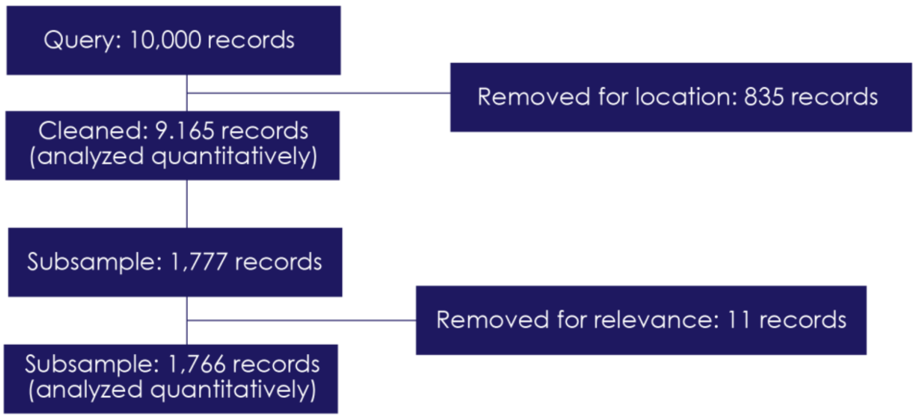
### Crimson Hexagon

Crimson Hexagon is a commercial analytics software platform for social media data. The company's online data library consists of over 1 trillion posts and includes documents from social networks such as Twitter and Facebook as well as blogs, forums, and news sites. Because the data is stored within Crimson Hexagon's library, we could extract data retrospectively over a long period of time; however, they limit the number of tweets to 10,000 per query.

We extracted 10,000 geolocated Twitter messages or "tweets" using a set of selection criteria to extract relevant tweets on the topic of age-discordant relationships or blessers (Appendix A). We included retrospective data from January 1, 2016 to January 31, 2017, in 10 DREAMS countries. Only search terms in English were included. Retweets were excluded. Tweets containing relevant language along with other analytics (user ID, date, time, location, sentiment, emotion, user gender) were captured and stored in an Excel spreadsheet. We removed 835 records for not meeting geographic inclusion criteria; 9,165 records were used for quantitative analysis.

Of these 9,165 Crimson Hexagon-provided records, 1,777 had specified "gender." Using OpenEpi (https://www.openepi.com/), we conducted a Pearson's chi-square test to check for any differences between records that had specified gender and those that had not, and found no statistical difference in other characteristics (sentiment, country of origin). Upon further examination, an additional 11 records were removed for relevance. A total of 1,766 records were therefore used for qualitative analysis for blessers (Figure 4).

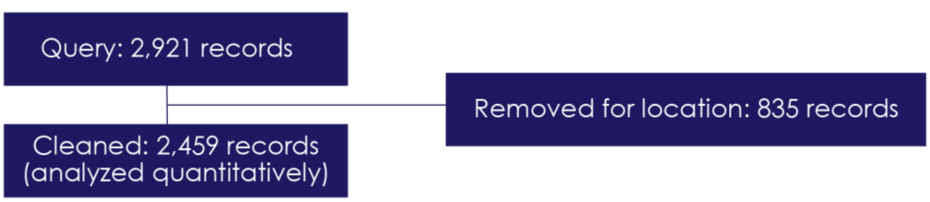
**Figure 4. Summary of review process of Crimson Hexagon blesser tweet records**




## Union Metrics

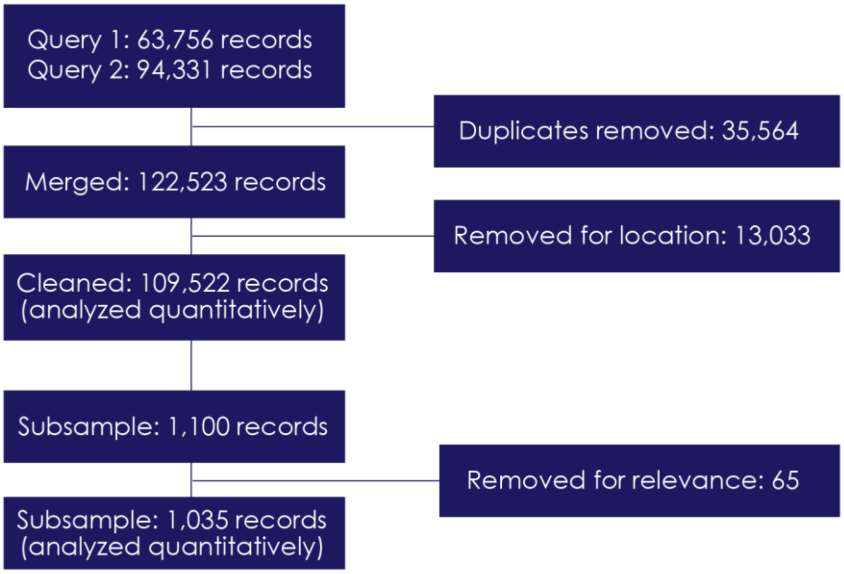
Like Crimson Hexagon, Union Metrics is a commercial software platform that extracts social media data. There are fewer built-in analytics, but the data can be collected either retrospectively or prospectively. We extracted data from November 25, 2017 to January 16, 2018 on both topics under study in the 10 DREAMS countries. We requested tweets on blessers using the same search criteria used for the Crimson Hexagon extraction. A total of 2,921 records were collected from Twitter; 835 were removed during cleaning for not meeting location criteria. The remaining 2,459 records were used for quantitative analysis (Figure 5).

**Figure 5. Summary of review process of Union Metrics blesser tweet records**




To extract tweets on GBV, we requested tweets with language related to GBV, IPV, and violence against women (Appendix A). Owing to the length of the search strings, two queries were conducted and merged after data collection. Tweets containing relevant language along with other metadata (user ID, date, time, location, type of tweet, replies, and retweets) were captured and stored in two Excel spreadsheets: one for GBV and one for blessers. In all, 109,522 records were used for quantitative analysis. We used simple random sampling to subsample approximately one percent of the Union Metrics-provided records, excluding retweets and quotes (n=1100), for the qualitative analysis of sentiment and GBV attitudes (Figure 6).

**Figure 6. Summary of review process of Union Metrics GBV tweet records**




## Data Analysis

This study assessed the challenges and limitations of using social media data to track changes over time in attitudes and gender norms; the type of information that can be gathered from social media to track changes in gender norms; and the emerging themes and thematic variations from social media posts about age-discordant relationships or GBV. Certain analyses were limited by data source (Table 2). All quantitative analyses were descriptive and conducted using Excel and Tableau.

**Table 2. Summary data analysis**

| Questions | | Data Sources | Type of Analysis |
|---|---|---|---|
| What are the challenges and limitations of using social media data to track changes over time in attitudes and gender norms? | | | |
| a) | What are the challenges and limitations of using social media data to track changes over time in attitudes and gender norms? | (review of the process) | Descriptive: *Cost, time, availability of data, sampling, generalizability, ethics, data security* |
| b) | Validity check: How relevant are the tweets to the intended topic? (subsample of blesser and GBV tweets) | Crimson Hexagon<br>Union Metrics | Quantitative and qualitative |
| c) | Validity check: How accurate is machine- versus human-driven sentiment analysis? (subsample of blesser tweets) | Crimson Hexagon | Quantitative and qualitative |
| What type of information can be gathered on gender norms from social media to track changes in gender norms? | | | |
| a) | What is the frequency of posts in a specified time period? | Crimson Hexagon<br>Union Metrics | Quantitative |
| b) | What are the timing patterns of posts (by month, day, or time of day)? | Crimson Hexagon<br>Union Metrics | Quantitative |
| a) | What are the most frequently used words in the online discussion? | Crimson Hexagon<br>Union Metrics | Quantitative |
| b) | What is the frequency of positive, negative, or neutral sentiments? | Crimson Hexagon<br>Union Metrics | Quantitative |
| a) | What sociodemographic characteristics of the people (i.e., users) posting are available? | Crimson Hexagon | Quantitative |
| c) | How influential are the people posting on the topic? | Crimson Hexagon | Quantitative |
| d) | What patterns exist related to each category (e.g., user's gender, timing, influence, location)? | Crimson Hexagon<br>Union Metrics | Quantitative |
| e) | What are the changes in patterns during a specific period? | Crimson Hexagon<br>Union Metrics | Quantitative |
| What are the emerging themes? | | | |
| a) | What are the thematic patterns of posts? | Crimson Hexagon<br>Union Metrics | Qualitative |

## Analysis of Process

We assessed the processes for extracting data from Twitter and analyzing metrics to understand the challenges and limitations of using social media data to track changes over time in attitudes and gender norms. Using the blesser search string, we assessed different methods for collecting and analyzing Twitter data, and compared the cost, time, availability, and types of data abstracted. We included extraction of GBV data to assess method replicability across different gender topics and issues. We also reviewed ethics and data security considerations with each method by conducting a review of rules and regulations, along with peer-reviewed and grey literature.

## Analysis of Gender Norms

We assessed the type of information that can be gathered on gender norms from social media to track changes in gender norms through overall tweet activity, tweet metrics, engagement metrics, and tweet language metrics. We also conducted validity assessments to understand how relevant the data were to the intended topic.

Overall tweet activity was recorded as the number of tweets and users, and these were related to specific time periods. The tweet metric analysis was performed by retrieving statistics about the frequency of tweets with links, tweets with replies, and tweets where Twitter users were mentioned. Engagement metrics were retrieved by obtaining the number of users who tweeted over a set time period, grouped by the number of tweets sent per user, the number of followers they have, and so forth. Language metrics, such as most common words, were visualized using an online word cloud generator.[2]

We performed a sentiment analysis to quantify attitudes and emotions. Computer algorithm-generated sentiment was provided by Crimson Hexagon. We sought to compare computer-generated with human-coded sentiment. To code for sentiment, a subsample of tweets was uploaded into Excel software and coded independently by two researchers using predetermined sentiment categories (positive, negative, or neutral). For the blesser analysis, we categorized the sentiment of the tweet as:

- Positive: request or advertisement for a blesser, pros of having a blesser, pros of being a blesser, positive attitude toward concept of blesser
- Negative: cons of having a blesser, cons of being a blesser, negative attitude toward concept of blesser
- Neutral: neither positive nor negative attitude toward blesser, colloquial use of term blesser
- Unknown: partial or unclear statements with links out, non-English, link only
- NA: not applicable to the topic

For the GBV analysis, we categorized the sentiment of the tweet as:

- Positive: pros of GBV, desire to perpetrate or receive, excusing GBV or the perpetrator, positive attitude toward concept of GBV, positive metaphorical use
- Negative: cons of GBV, negative attitude toward GBV or the perpetrator, negative metaphorical use
- Neutral: neither positive nor negative attitude toward GBV, partial thought, no link
- Unknown: partial or unclear statements with links out, non-English, link only
- NA: not applicable to the topic

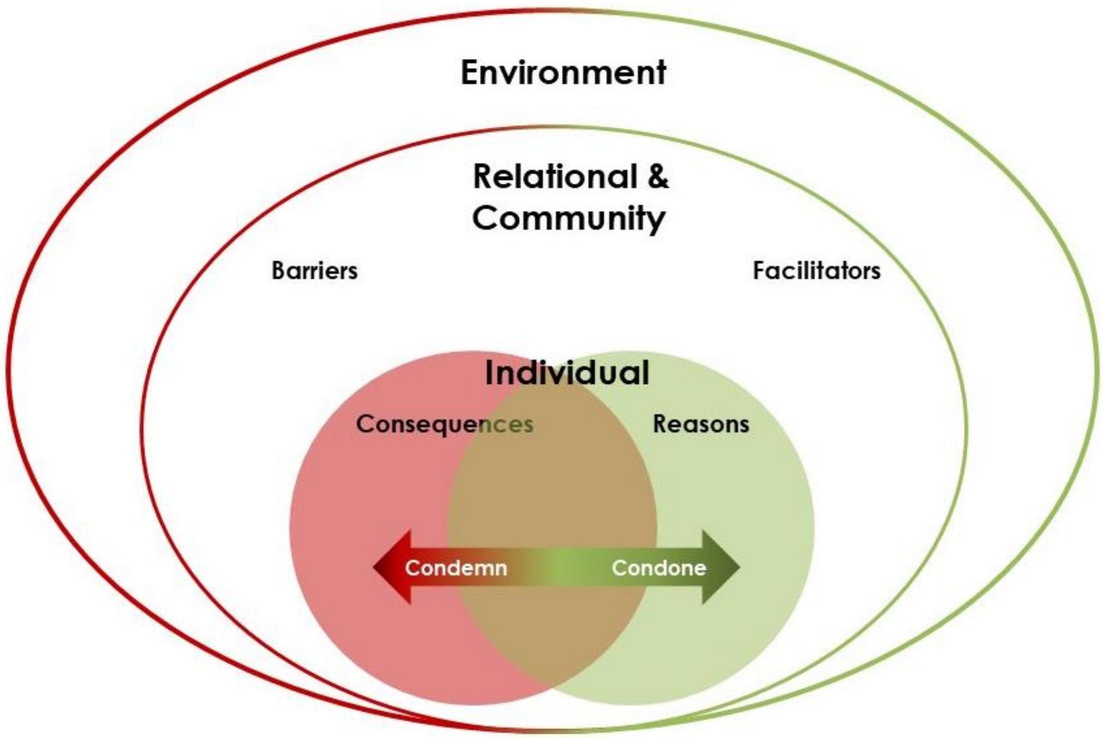
To maximize inter-coder reliability, we conducted quality assurance checks after the first 50 consecutive tweets, the second 50 consecutive tweets, and on the full set of tweets. Discordant tweets were filtered, discussed, and resolved. If a resolution could not be obtained, a final code was applied by a third independent coder. We compared the accuracy between computer- and human-applied sentiment by cross-checking codes from the research team against those automatically assigned by Crimson Hexagon.

[2] https://www.jasondavies.com/wordcloud/

Sentiment was hand-coded for blessers (Crimson Hexagon data only) and GBV data, and net sentiment was calculated as the number of positive tweets minus the number of negative tweets.

To answer the qualitative questions, we used this same subsample to further explore emerging themes. Inductive thematic coding was conducted using the subsample sets (blesser n=1,766, GBV n=1,035). We followed the ecological model, understanding how individual attitudes were affected and shaped by relational, community, and environmental (societal) factors (Figure 7). For blesser data, we also analyzed thematic variations by gender.

**Figure 7. Conceptual framework for qualitative analysis**




## Ethical Review

The Office of Human Research Ethics at the University of North Carolina reviewed the study and determined that it did not constitute human subjects research as defined under federal regulations—45 CFR 46.102 (d or f) and 21 CFR 56.102(c)(e)(l)—and did not require IRB approval. However, the Office of Human Research Ethics required additional security measures to ensure that data were adequately protected from inadvertent disclosure. Owing to the nature of the data, the investigators were required to implement Level II Data Security Requirements: [https://research.unc.edu/files/2017/05/Updated-DSL-Notification.pdf](https://research.unc.edu/files/2017/05/Updated-DSL-Notification.pdf).

# RESULTS

In this section we first describe the results of our analysis of the different abstraction platforms and methods, including the types of information provided by each platform, the relevancy of the data we collected, and the validity of sentiment generated by a computer versus coded by hand.

We then present the results of those data collected and analyzed on blessers: first the retrospective data collected through Crimson Hexagon then the mixed retrospective and prospective data collected through Union Metrics. Finally, we present the results of those data collected and analyzed on GBV from Union Metrics.

## Analysis of Methods

We used and assessed different abstraction platforms and methods including accessing data through vendors (Crimson Hexagon and Union Metrics) and workflow automation tools (Microsoft Flow). Cost, data, and available analytics vary depending on the method of abstraction (Table 3).

**Table 3. Types of information provided by each platform**

| | Crimson Hexagon | Union Metrics | Microsoft Flow |
|---|---|---|---|
| Cost of platform | $$$$ | $ | Free |
| Cost of analysis | $$ | $$$ | $$$$ |
| **Meta Data** | | | |
| Tweet ID | ✓ | ✓ | ✓ |
| Tweet URL | ✓ | ✓ | ✓ |
| User ID | ✓ | ✓ | ✓ |
| User location | ✓ | ✓ | ✓ |
| Country | ✓ | ✓ | ✓ |
| Tweet text | ✓ | ✓ | ✓ |
| Time | ✓ | ✓ | ✓ |
| Retweet ID | | ✓ | ✓ |
| Reply to ID | | ✓ | ✓ |
| Impressions | | ✓ | |
| Likes | | ✓ | ✓ |
| User number posts | ✓ | | ✓ |
| User followers | ✓ | | ✓ |
| User following | ✓ | | ✓ |
| **Quantitative Analytics** | | | |
| Average tweets/day | | ✓ | |
| Type of tweet breakdown | | ✓ | |

| | | | |
|---|---|---|---|
| Retweet rate | | ✓ | |
| User gender | ✓ | | |
| Tweets per user (contributors) | | ✓ | |
| Top contributors | | ✓ | |
| Klout score | ✓ | | |
| Location (user entered) | ✓ | ✓ | ✓ |
| Geography | ✓ | With additional subscription | |
| **Qualitative Analytics** | | | |
| Sentiment* | ✓ | With additional subscription | |
| Emotion | ✓ | | |
| Top hashtags | | ✓ | |

* Note: Sentiment generated by computer algorithm

While Microsoft Flow is an effective method of obtaining data, we encountered challenges due to shifting privacy policies that created data gaps. These data were also the most "raw" and would require significant human resources to clean and analyze. At the time of the study, Flow did not have the capacity to filter location; this would have to be manually filtered after collection.

Union Metrics data were fairly clean but needed additional location cleanings. Union Metrics data also came with additional analytics through their online platform. Data provided by Crimson Hexagon were the most "clean" and came with advanced analytics, which reduced the amount of resources needed for analysis. Data from Crimson Hexagon and Union Metrics were used in this study.

## Relevance

Of the 1.777 subsampled blesser tweets from Crimson Hexagon, 1 (0.6%) were found to be irrelevant to the research topic. All 11 were related to religion or faith, not using the term "blesser" as a metaphor or colloquialism.

Of the 1,100 subsampled GBV tweets from Union Metrics, 65 (5.9%) were found to be irrelevant to the research topic. These were primarily tweets where homonyms were used, but not metaphorically. For example, a tweet about a song's beat would have been included in the data collection search for the word "beat" but does not refer to violence. Metaphoric use of terms such as "rape" were retained and included in analysis.

## Sentiment Validity

While sentiment was provided by Crimson Hexagon, we sought to compare computer-generated with human-coded sentiment. We used the subsample blesser dataset of 1,766 tweets for this analysis. Overall, less than half (n=723, 41%) of the coding matched, although this varied across sentiments:

- Neutral (n=231, 28%)
- Positive (n=159, 41%)
- Negative (n=221, 54%)
- Unknown or blank (n=112, 84%)

Coding for "unknown" or "blank" matched the majority of the time, with greater discordance among what was coded as positive, negative, and neutral. We found greater variation in sentiment when a computer coded the tweets than when coding was done by hand. Crimson Hexagon found two times more positive tweets than negative ones and coded nearly five times more tweets as neutral; hand-coded tweets were more equally distributed.

For the purpose of this study, we used human-coded sentiment during qualitative analysis.

**Table 4. Comparison of human- and computer-generated sentiment codes**

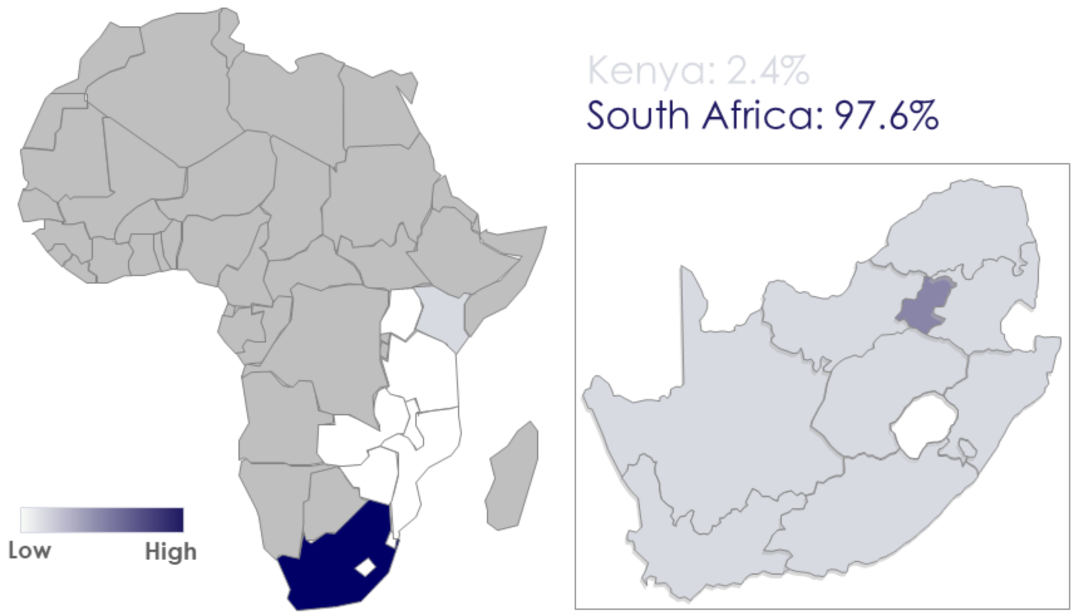
| | | Crimson Hexagon Computer-Coded | | | | |
|---|---|---|---|---|---|---|
| | | Positive | Negative | Neutral | Unknown/ blank | TOTAL |
| Human Coded | Positive | 231 | 52 | 203 | 7 | **493** |
| | Negative | 108 | 159 | 139 | 4 | **410** |
| | Neutral | 139 | 56 | 221 | 11 | **427** |
| | Unknown/blank | 77 | 26 | 221 | 112 | **436** |
| | TOTAL | **563** | **294** | **784** | **134** | **1766** |

## Analysis of Gender Norms: Age-Discordant Relationships or "Blessers"

### Geographic Distribution

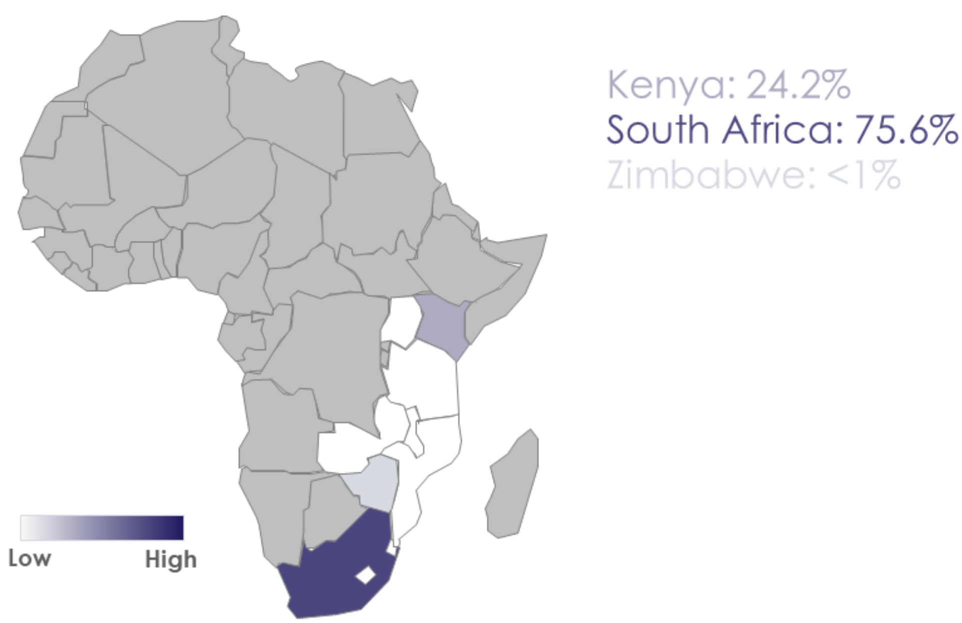
Data collected by Crimson Hexagon from January 1, 2016 to January 31, 2017, primarily came from South Africa (97.6%), with a few data points from Kenya (2.4%). No other DREAMS countries were represented. Closer examination of South Africa's distribution showed while all provinces of South Africa provided tweets on the subject of "blessers" or "sugar daddies," Gauteng produced the most tweets (42.3%), heavily influenced by Johannesburg and Pretoria (Figure 8). This coincides with number of people online and using social media across the countries sampled (Portland, 2016).

**Figure 8. Geographic distribution of tweets on "blessers" and "sugar daddies," January 1, 2016 to January 31, 2017 (Crimson Hexagon)**




One year later and using data from a different platform, we saw similar geographic patterns, with South Africa producing the majority of tweets about "blessers" and "sugar daddies" (75.6%). A greater proportion of tweets were identified as from Kenya (24.2%) than the previous year. Few tweets were identified as from Zimbabwe (<1%); no other DREAMS countries were represented (Figure 9).

**Figure 9. Geographic distribution of tweets on "blessers" and "sugar daddies," November 25, 2017 to January 16, 2018 (Union Metrics)**

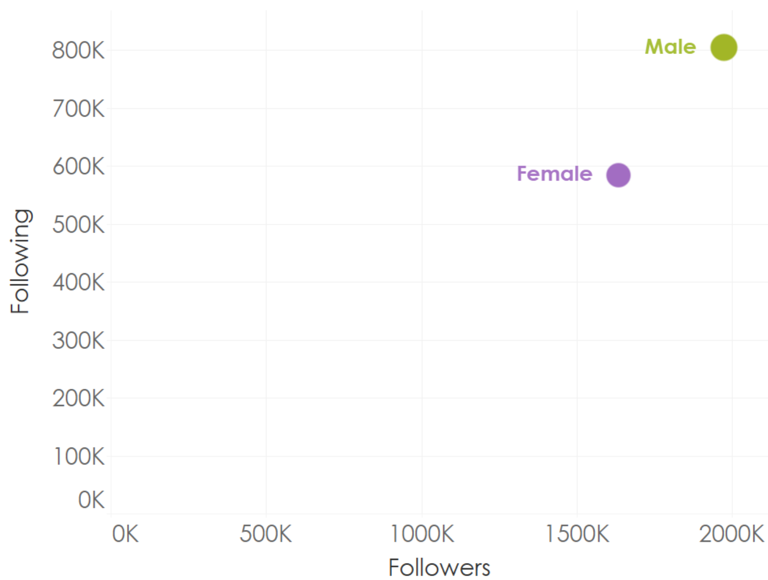


## User Characteristics

Slightly more men tweeted about "blessers" than did women (55% men, 45% women; Figure 12). This correlates with the distribution of Twitter users overall (53% men, 47% women) (Portland Communications, 2016). Note that gender data is only available in the Crimson Hexagon dataset, so comparison of data sources is not available.

Men's tweets were more influential in that they had on average higher numbers of followers than women did (Figure 10). A higher number of followers correlates with a larger digital influence or reach, in that a single tweet has higher potential to be seen by more people. Males also followed more people.

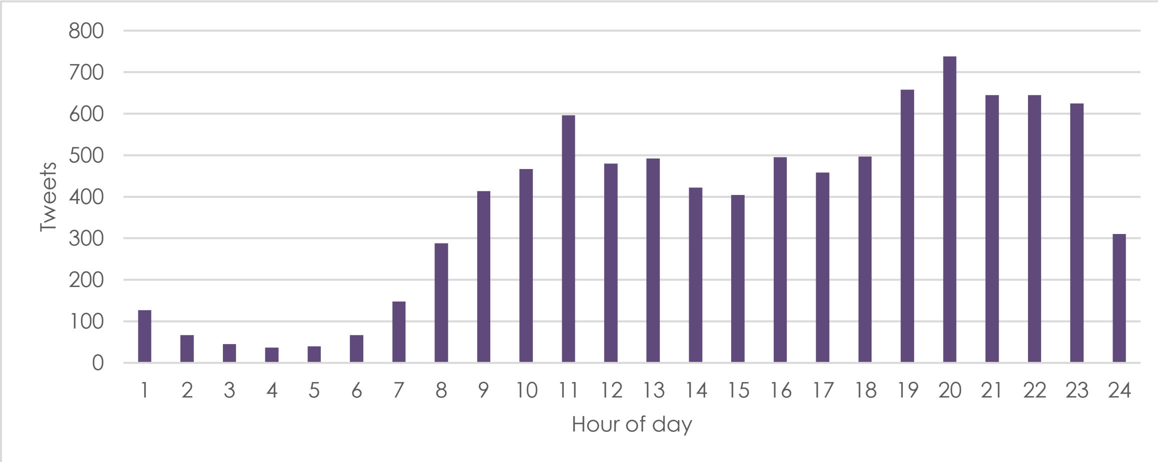
**Figure 10. Influence of those tweeting about blessers, by gender**




## Timing Patterns

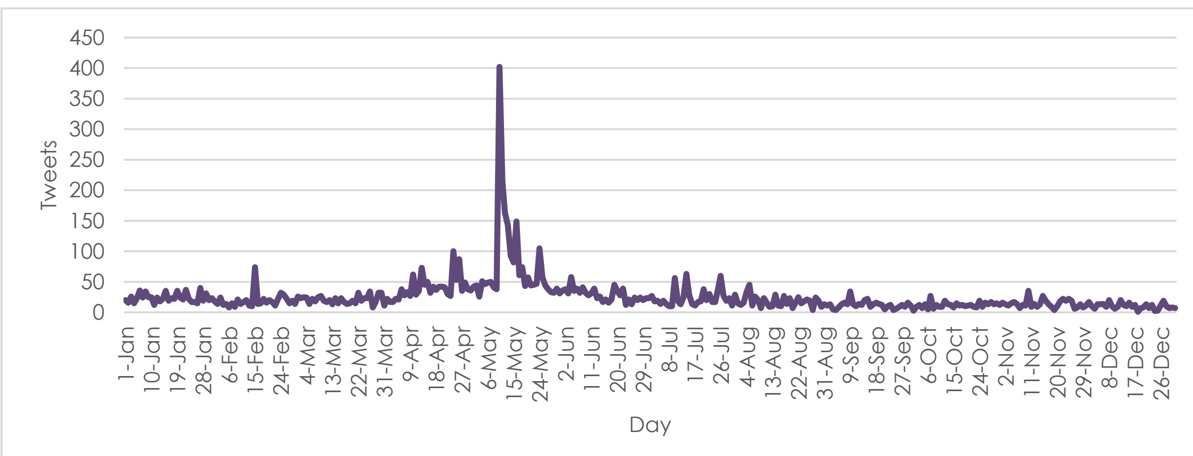
Twitter provides date and time metadata. Tweets about "blessers" saw increases in the late morning and late evening. Similar trends were found using data from both Union Metrics (Figure 11) and Crimson Hexagon.

**Figure 11. Frequency of blesser tweets by hour (Crimson Hexagon)**




When assessing the frequency of tweets over a 12-month period, we see a significant uptick of activity in May 2016 (Figure 12). This uptick corresponds with a South African news report on "blessers" Serge Cabonge and Kenny Kunene. Analysis of frequently used words and hashtags indicate this uptick was driven primarily by this report.

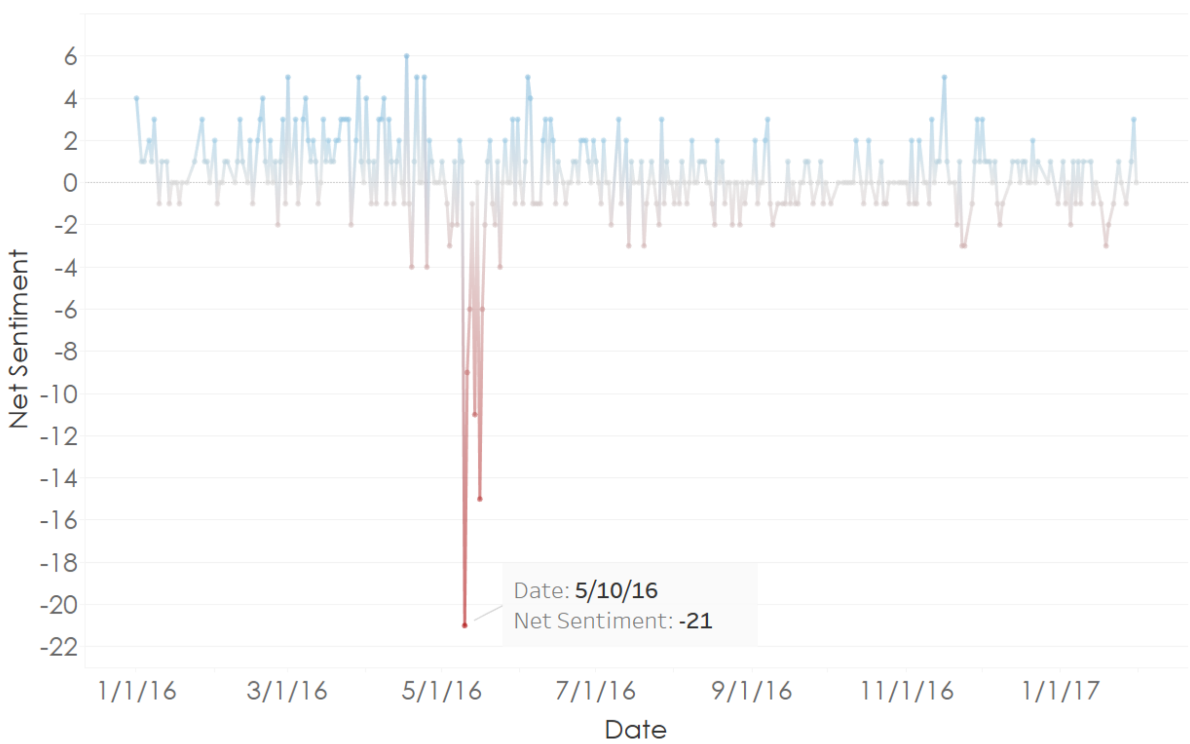
**Figure 12. Frequency of blesser tweets over 12 months (Crimson Hexagon)**




We saw no significant differences in frequency patterns between men and women.

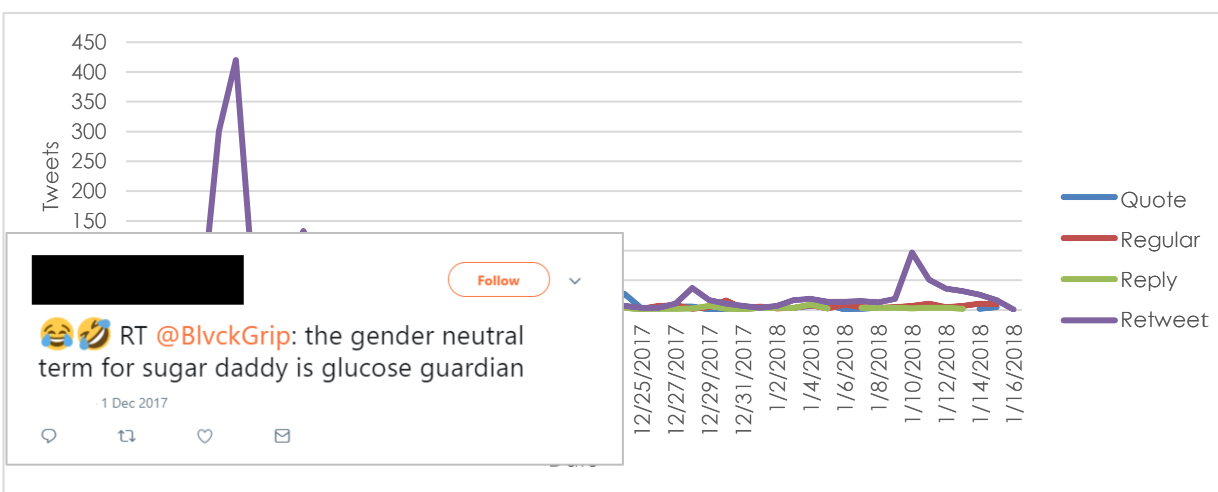
Sentiment analyses found that while the net sentiment fluctuated rather steadily between +4 and -4, the uptick in tweet activity in May 2016 was primarily negative, indicating a general negative reaction to the news articles around "blessers" Serge Cabonge and Kenny Kunene (Figure 13).

**Figure 13. Net sentiment of blesser tweets over 12 months (Crimson Hexagon)**



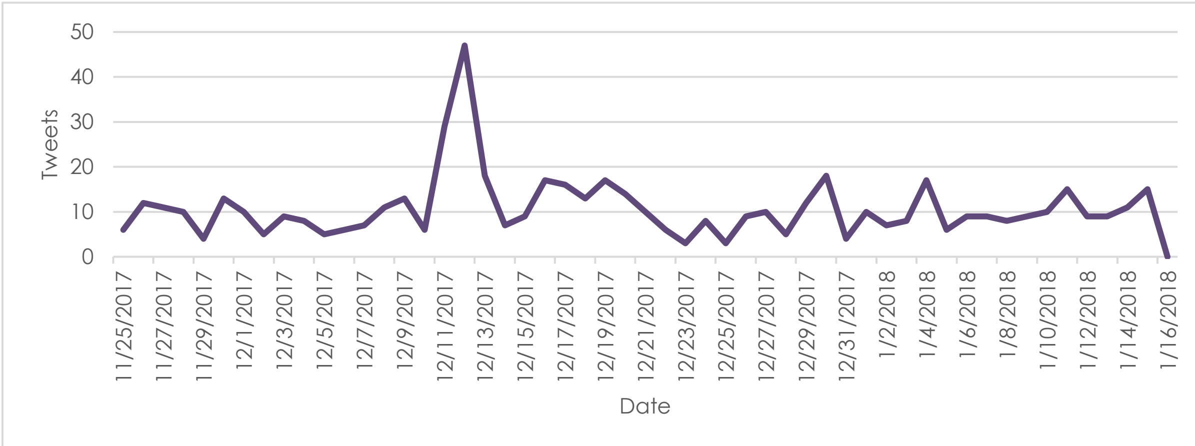

Because we only requested original tweets (excluding retweets) from Crimson Hexagon, we could not assess timing patterns for all types of tweets. Contributions to a conversation often involve retweeting and quoting existing tweets. With Union Metrics data we explored the timing patterns for all tweets, disaggregating data by type of tweet (Figure 14). Regular tweets are messages containing text, photos, a GIF, or video. A reply is when one user responds to another user. A quote tweet allows a user to tweet another person's tweet with their own comment added. Retweet is a re-posting of another user's tweet.

**Figure 14. Frequency of blesser tweets by date, disaggregated by type (Union Metrics)**



We removed retweets and quotes to explore the timing patterns of new ideas (Figure 16). A surge of original tweets was seen on December 11. After investigating, these tweets primarily revolved around a news story on African National Congress Secretary General, Gwede Mantashe, who was accused of being a blesser.

**Figure 16. Frequency of blesser tweets by date, without retweets or quotes (Union Metrics)**




## Common Words

Because we searched on certain terms, a high frequency of words like "blesser" and "sugar daddy" was expected. Word clouds help visualize the frequency of terms to better understand the common words that were used beyond search terms. For example, Figure 17 shows a word cloud created from data collected from January 1, 2016 to January 31, 2017. "Blesser" is clearly the most common word, but other words such as need, money, Checkpoint and AskAMan (both news shows) shed light quickly on the characteristics of the conversation. We discuss these qualitative themes later in the report.

**Figure 17. Word cloud of blesser tweets from January 2016 to January 2017 (Crimson Hexagon)**

Visualizing a word cloud from data collected a year later (Figure 18) shows how the conversation has evolved and the influence of retweets. The majority of the terms highlighted are components of the popular retweet mentioned earlier. Also noticeable are references to different celebrities and scandals.

**Figure 18. Word cloud of blesser tweets from November 2017 to January 2018 (Union Metrics)**


This type of analysis allows for a quick qualitative inventory of themes.

## Emerging Themes

We also conducted a more thorough and rigorous analysis of emerging qualitative themes to better understand user attitudes, personal considerations, and societal factors. The themes are organized here according to the ecological model.

### *Individual Roles and Attitudes*

The roles and attitudes towards blessers were divided along gender lines.[3] When speaking as participants, very few women mentioned being blessers themselves but rather as a blessee. Men, on the other hand, spoke generally from the point of view of the blesser. This follows traditional gender norms, whereby men take on the role of the provider in relationships. There were exceptions to this, with a few men discussing having blessers. This may be genuine, or it may be a result of poor assignment of gender or, at least in one case, of non-heterosexual relationships. The dearth of men as blessees may also reflect that the term "blesser" is infrequently used to describe a relationship where women provide money or gifts in exchange for sex with men.

Several users spoke about blessers either as observers or about the phenomenon more generally. In this context, attitudes tended to be more negative than positive, expressing attitudes of pointlessness,

[3] As determined by Crimson-Hexagon's algorithm and may not reflect actual gender of Twitter user.

awkwardness, and judgement. For example, some tweets described blessees as lazy or unmotivated to find a job to pay for their needs or desires. Others equated blessees with prostitutes.

> *There is no difference between a blessee and a prostitute, just the word used.*[4] *(Men)*

While both genders expressed positive and negative attitudes towards blessers, more condemning tweets came from men than women. In general, women celebrated the benefits a transactional relationship could offer while men criticized the individual expectations, associations, and outcomes of the blesser lifestyle.

> *This blesser thing isn't for me. . . . unnecessary pressure to spend. (M)*

An individual's attitude often coincided with gender roles and was balanced against personal considerations: reasons for and consequences of blesser/blessee relationships, as well as the enabling environment.

**Personal Considerations**

Individual attitudes, either condoning or condemning, were closely linked to personal considerations. In general, tweets coded as having negative attitudes towards blessers were more likely to cite consequences, while positive attitudes were more likely to cite reasons for having or being a blesser.

**Reasons**

Twitter users discussed many reasons for wanting or being in transactional relationships. Both men and women mentioned transactional sex as a way of covering certain living costs including rent, child support, and school fees.

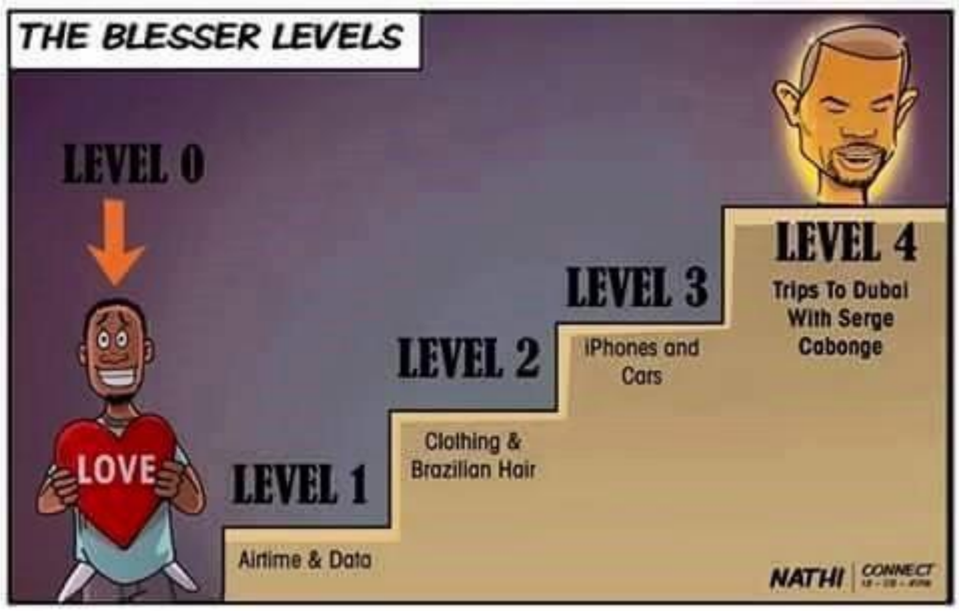
> *I need a blesser for my school fees, but not sure what level that would be. (Women)*

The tweet above refers to a common idea that there are "levels" of blessers, whereby each level is positively associated with the value of exchanged good. This concept originated online and has been perpetuated in the media (Figure 16) and through online social networks like Twitter.

---

[4] To ensure confidentiality, "aggregated quotations" have been used where several quotations around a topic are brought together to maintain the essence of meaning but rendering it impossible to identify any individual poster. (Bond, et al., 2013)

Figure 19. A commonly tweeted comic illustrating the stages of being #blessed




These levels focus mostly on sex to obtain social status and material goods. This paradigm was the most commonly discussed motivation for engaging in transactional sex, particularly by women, who are more likely to be the recipient of these gifts. Material goods mentioned included but were not limited to, electronics (e.g., cell phones), designer shoes and clothes, accessories, nice cars, and travel. The value of a material good was tied to the "blesser level" and resulting sexual expectation.

> *I want a blesser but only for data; I'm sure he won't be expecting sugar for just data. (W)*

The desire for love and commitment were less commonly discussed by participants; however, a few users mentioned knowing of people who have this expectation. Those users often reported skepticism around the blesser relationship becoming more than a transaction and judgement towards those who believed it.

> *Ha! This girl catching feelings for a blesser like he's going to stay with her. (M)*

Lastly, sex as a motivation for a blesser relationship was seen predominantly among men. This is expected, given the transactional aspect of these relationships and men's predominant role as blesser (provider of goods, receiver of sex) and women's role as blessee (receiver of goods, provider of sex). Some women, conversely, wondered if they could have a transactional relationship without providing sex. For example:

> *When you want a sugar daddy but giving them sugar is the problem (W)*

These differing attitudes expressed by men and women show how one person's reason is another's consequence in transactional relationships.

### Consequences

Positive consequences of blesser relationships include obtaining those things that were listed as reasons for participating in these relationships, such as material goods and sex. However, while these relationships may result in positive consequences for one gender or side of the relationship, the same may not be true for the other. While some women did mention the negative consequence of providing sex in these transactional relationships, men were more likely to discuss negative consequences of blesser relationships than women.

Tweets that focused on consequences were closely associated with negative attitudes. For example, as discussed, some men had negative attitudes towards blessers because of the time and resources required to uphold the relationship. Women also talked about this aspect, but with more neutral tone, possibly indicating that upkeep of one's self for a blesser not seen as a significant detriment when balanced against positive consequences.

Users also noted that the existence and acceptance of blesser relationships had a detrimental effect on non-blesser relationships. Because of the high levels of secrecy required for the former, there was higher suspicion and reduced trust within non-blesser relationships. Users also commented on how relationships that offered financial benefit may create a sense of dependency or reduce the need for non-blesser relationships.

> *Boyfriends are becoming extinct, replaced by blessers. (M)*

While non-blesser relationships offer love rather than goods or services, some users did mention love or commitment within blesser relationships. It was stated, however, that this rarely occurs, and Twitter users judged those who believed it was a possibility. Users who spoke as observers rather than participants of blesser relationships tweeted about the gossip, judgment, or social awkwardness that could come with being a blesser or blessee.

We specifically coded for the mention of increased violence or poor health (STIs, HIV) as a consequence of blesser relationships. Few tweets mentioned these as possibilities, often prompted by a specific event such as a news article. These consequences were mentioned much less frequently than other consequences discussed.

### *Relational, Community, and Environmental Levels*

Women and men discussed how larger social pressures, norms, and structures prevented or facilitated engagement in blesser-blessee relationships.

**Barriers**

Twitter users who both condoned and condemned blesser relationships discussed possible social barriers to engaging in these types of relationships. A common hashtag used among blesser tweets was #MoralsMustFall. This hashtag, made popular by an online "blesser-finder" website, hints at morals being a barrier to engaging in transactional sex. Some users explicitly discussed morals, as well as religion or belief in God, as reasons not to participate in blesser relationships.

There also simply may not be a need for blessers according to some users. For example, if a person is employed and educated, they can take care of their material and social needs themselves and do not have a need for a "blesser."

> *Walk in the kitchen to find leftover KFC . . . I am my own blesser! (W)*

While some users joked about "being their own blesser," most discussed the need for the social and financial security as a reason for or facilitator of blesser relationships.

Users who did want to be a blesser or blessee but found they could not do so often cited not meeting certain criteria based on money, looks, or age. A few users remarked that the blesser phenomenon created a micro-society from which you could be included or excluded based on these criteria. Users not meeting criteria were thus excluded even if they desired to be a blesser or blessee. Additionally, individuals may be excluded from having or being a blesser simply because they lacked information about where and how to engage in these transactions. Twitter has itself emerged as a platform for obtaining information and acting as a facilitator of these relationships.

### Facilitators

Several users expressed the belief that transactional sex is entrenched in society, just with different names and more modern approaches. Social networks and online platforms have facilitated the growth of blesser culture. For example, a blesser-finder website allows older men to find and contact women who are interested in being supported financially. Some individuals used Twitter to direct to this website or even used Twitter to advertise for blessers or blessees.

> *Hey @BlesserFinder, looking for a man to take me to Durban for the weekend. (W)*

Twitter was also used by people to advertise parties and events that would cater specifically to blessers and blessees.

At the individual level, users discussed the desire for financial or material goods as reasons for engaging in transactional sex. This extended into the community level, whereby poverty was cited as an enabler for blessers.

> *As long as there is poverty in our society, there will be blessers. (M)*

These types of larger social issues were commonly blamed for the existence of blessers by men and women who condemned the practice. These users viewed blessers as a consequence of larger, more entrenched social issues, such as a reduction in social ethical standards and a lack of strong role models.

### Mass Media

Many tweets were written around or in reference to people or events in the media. Because the term “blesser” was born out of social media, it is no surprise there is frequent conversation around blessers and media. These conversations heavily influenced the frequency of tweets. For example, quantitative measures revealed an uptick in discussions around May 2016. This coincided with news reports featuring Kenny Kunene and Serge Cabonge. Tweets also served as platforms for commentary against popular media events and shows, using hashtags like #SkeemSaam and #7deLaan to discuss plotlines from television programs featuring blesser characters. Two other common shows discussed were “Our Perfect Wedding” and “Date My Family,” reality TV shows that featured contestants Twitter users believed to be “blessees.” Radio shows and talk shows also were mentioned, including a popular segment on local radio entitled “Ask a Man,” which featured callers asking for advice about their blesser relationships. Tweets about mass media were primarily neutral in sentiment and came from both men and women.

### Linguistic Mainstreaming

How trends become incorporated into language through jokes and metaphor can speak to how pervasive an idea is in society. Tweets may use the term “blesser” figuratively, as in:

> *When Rhianna drops a new album, . . . she is a music blesser.*

They are not indicating that Rhianna herself is a blesser in the normative sense, but rather that in a figurative sense she provides gifts to others in the form of her music.

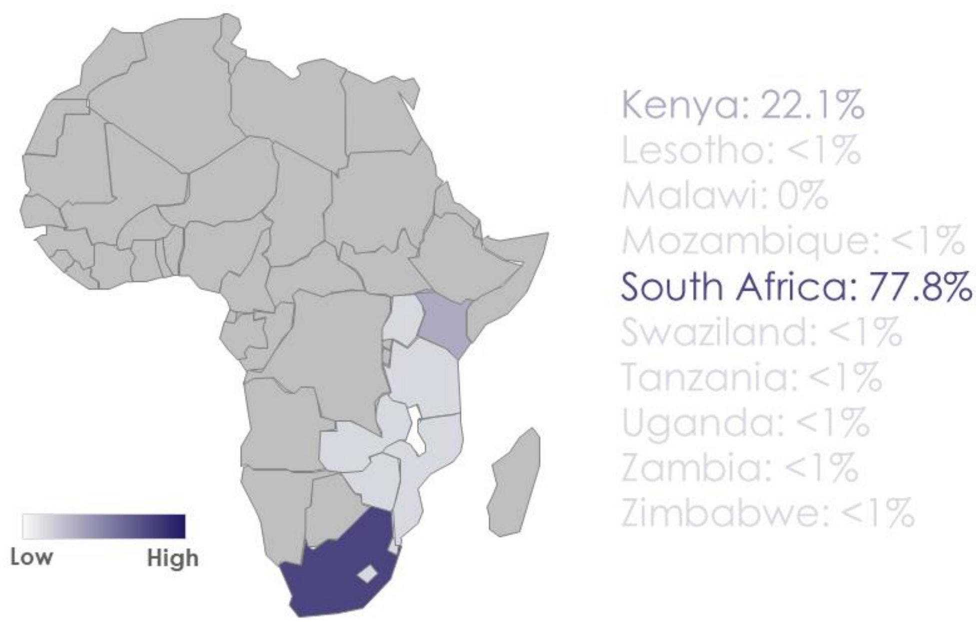
We found that tweets early in our search period focused primarily on the term “blesser” in the normative sense, referring to an intergenerational transactional relationship. However, tweets from later in the year became more figurative or liberal in meaning, either referring to simply transactional (not necessarily intergenerational) relationships or using the term metaphorically as described above.

## Analysis of Gender Norms: GBV

### Geographic Distribution

Data collected on GBV primarily came from South Africa (77.8%), followed by Kenya (22.1%). All other DREAMS countries were represented at much lower levels (<1%) except for Malawi, which produced no tweets matching our GBV search criteria during the data collection period (Figure 20).

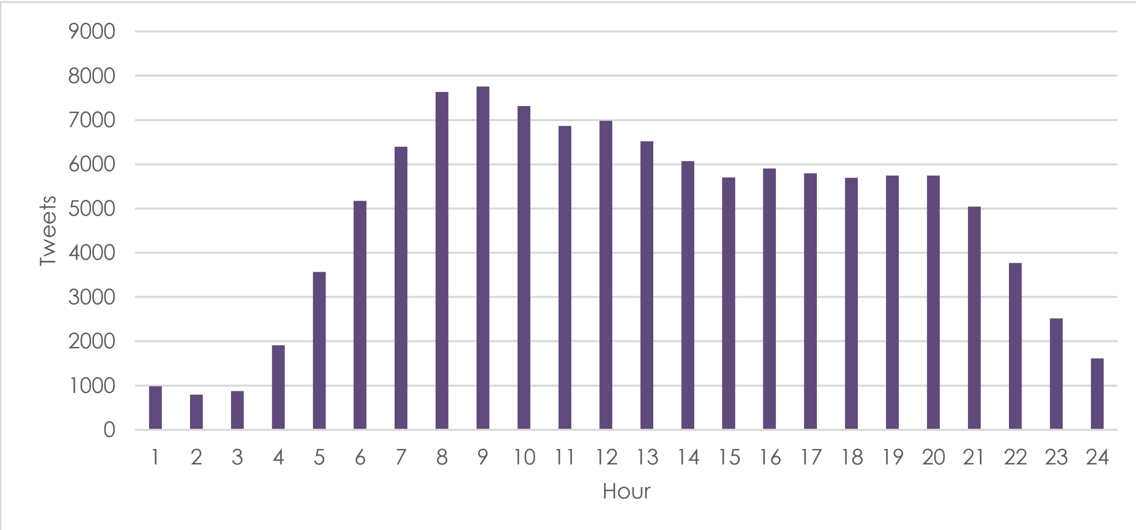
**Figure 20. Geographic distribution of tweets on GBV, November 25, 2017 to January 16, 2018 (Union Metrics)**




### Timing Patterns

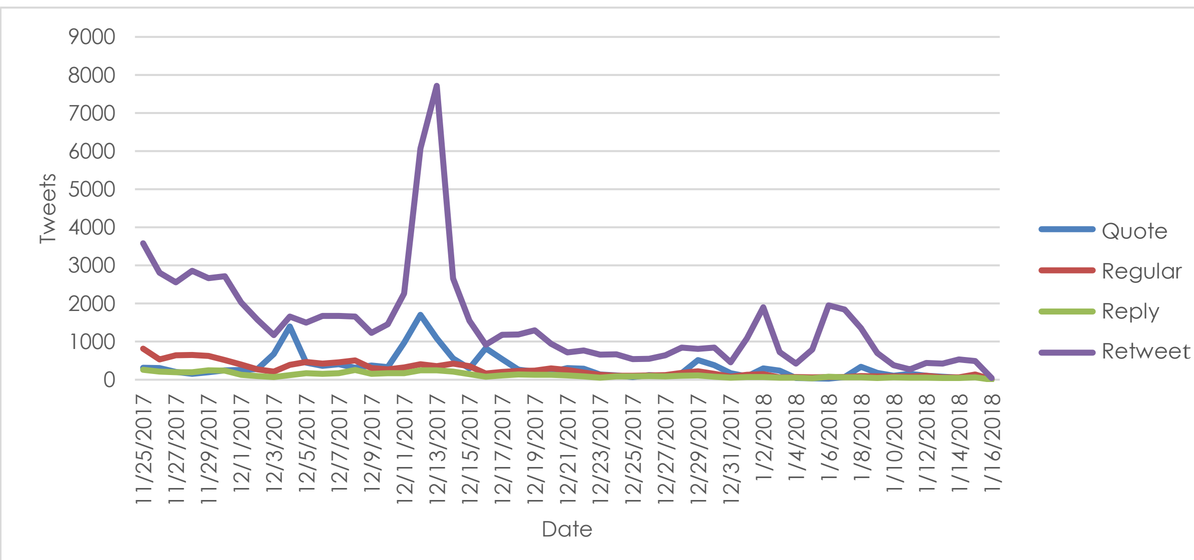
Twitter provides date and time metadata. Tweets about GBV saw increases in the morning, decreasing throughout the day (Figure 21).

**Figure 21. Frequency of GBV tweets by hour (Union Metrics)**



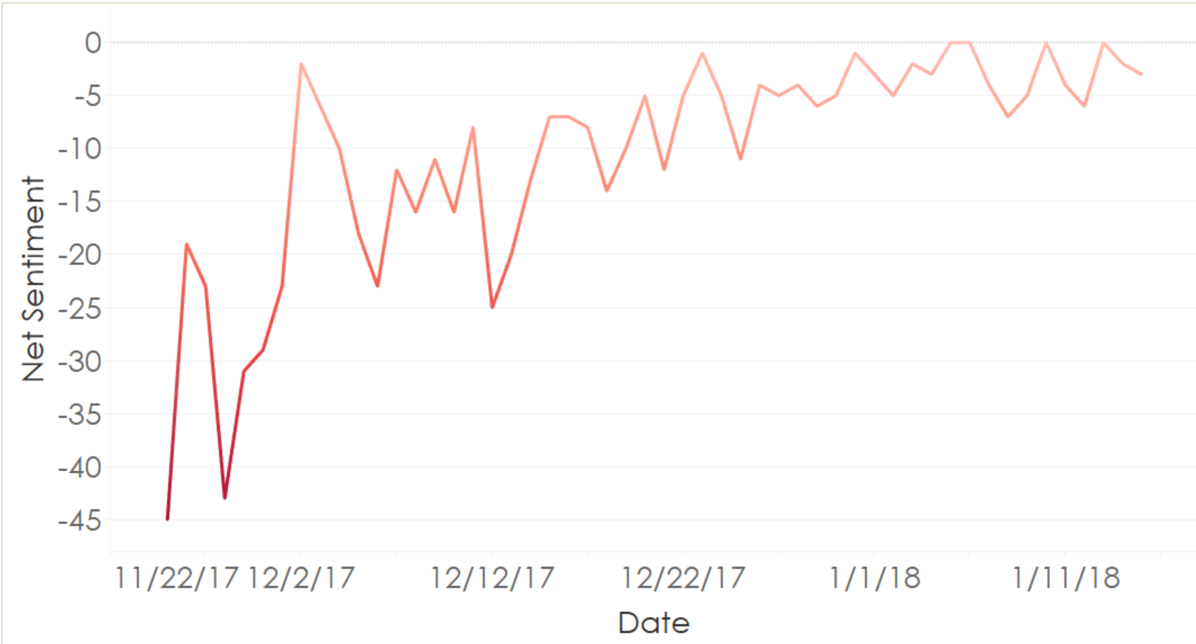

A large uptick in activity in mid-December was primarily driven by reaction to and discussion of the expulsion of two student activists from Rhodes University (Figure 22).

**Figure 22. Frequency of GBV tweets by date, disaggregated by type (Union Metrics)**




The full sample of tweets were not coded for sentiment; those that were did not include retweets. When the subsample (quotes, replies, and regular tweets only) were coded for sentiment and plotted over time, we did find that the net sentiment for all days was negative. More net negative tweets occurred during the first part of the sampling time frame when the 16 Days of Activism was occurring (November 25-December 10). Over 10 percent of the subsampled tweets (n=122) were directly linked through hashtags to the 16 Days of Activism.

**Figure 23. Net sentiment of GBV tweets by date (Union Metrics)**




## Keywords

Figure 24 shows a word cloud from data collected from November 25, 2017 to January 16, 2018. Because we searched on certain terms, a high frequency of words like "sexual abuse," "violence," and "women" was expected. The prominence of "RT" indicates a high proportion of tweets are retweets. Other themes that arose were Rhodes University, 16 days (of Activism against Gender-Based Violence Campaign), and black (a nod to themes of intersectionality). We dove deeper into these emerging themes through qualitative analysis.

**Figure 24. Word cloud of GBV tweets from November 2017 to January 2018 (Union Metrics)**

## Emerging Themes

We conducted a more thorough and rigorous analysis of emerging qualitative themes of the GBV data to better understand user attitudes, personal considerations, and societal factors. The themes are organized here according to the ecological model.

### *Individual Roles and Attitudes*

While individual user gender is unknown from this dataset, there were many conversations about the roles of perpetrators and victims of GBV that followed gender lines, in which perpetrators were often characterized as men and victims as women. However, there were also tweets that challenged the hetero-normative and male perpetrator narrative, exploring women as perpetrators and violence within LGBTQ+ communities.

> *While the majority of victims of domestic violence are women, men may also be victims of relationship violence.*
>
> *Let's also talk about the violence that LBQT Women and gender non-binary people face for pushing back on problematic ideas around gender in our society.*

As we noted, net sentiment of tweets about GBV skewed negative. Many of the tweets that condemned GBV denounced the concept as a social phenomenon rather than a personal experience. This was primarily driven by response to social movements such as the 16 days of Activism against Gender-Based Violence Campaign or specific events such as the Rhodes University rape case.

Few tweets explicitly encouraged or supported GBV; however, there were a number of tweets that excused perpetrators or blamed victims.

### *Personal Considerations*

In general, most tweets did not discuss GBV as a personal issue through direct experience, but rather as commentary on someone else's actions or experiences. From this commentary, however, it is possible to glean certain perceived reasons for and consequences of GBV.

**Reasons**

Behaviors of both women and men were described as reasons for GBV. How women acted towards men, how they were dressed, their perceived promiscuity, the establishments they frequented, and other behavior were cited as reasons violence against women occurs, particularly among those who engaged in victim-blaming or excusing the perpetrator.

> *Girls these days- eish! They just asking for it with those tight tops and loose morals!*
>
> *You remove your top to go march against rape. . . . Are you trying to make the job easy for the rapist?*

Others, more supportive of the woman as the victim, noted the lack of power to fight back or inequality as reasons GBV could occur in a relationship. Power dynamics drove many of the reasons for GBV. A struggle to exert power and establish dominance was often described or referenced as reasons men primarily engaged in GBV. This feeds into a belief that when power is challenged, a proportional response is necessary.

> *Men are expected to control their women, and violence is a common way to punish transgressions.*

At a societal scale, movements around feminism could challenge notions of male superiority, so violence was mentioned as a way of rebalancing the status quo.

Other male behaviors or characteristics—such as alcohol use, exposure to violent video games, machismo attitudes, and even race or nationality—were noted as drivers of GBV. Additionally, some Twitter users saw violence as a response to violence, be it reacting to or taking revenge on violence they had experienced previously inside or outside of the relationship itself and taking those feelings out on the partner, or intergenerationally through a cycle of violence.

Lastly, there were tweets that espoused intolerance for GBV in any circumstance.

> *I don't care how mad you are. There is never a reason to beat a woman; there is never an excuse to violence.*

**Consequences**

Consequences were discussed primarily in terms of those experienced by the individual victim or the perpetrator. Traditional consequences of violence for the victim included negative health outcomes (e.g., STIs, pregnancy) injury, and death. We also found themes of shame, embarrassment, and stigma tied closely to negative consequences for the victim. Themes around the difficulty in seeking and obtaining help after instances of GBV—such as appropriate and affordable care, social support, and sensitive caregivers—arose either by those offering alternatives or those seeking further information through social media.

> *Where does one go for assistance? This is in relation to an unreported rape allegation of a girl.*

Consequences of violence for the perpetrator included police involvement, jail time, and criminal records. Those excusing the perpetrator often described an accusation and the resulting social ostracization as an undeserved negative outcome.

> *There are too many jailed while innocent. Why are these girls coming out now? They say rape; I say holding a grudge.*

Few Twitter users noted the larger social consequences of GBV, ranging from perpetuating a cycle of violence to galvanizing harmful gender norms to limiting the potential for growth as a society. It was discussed that GBV can negatively influence the educational opportunities and social welfare of victims and create an economic burden on societies in terms of support, care, and treatment.

> *Child marriage will cost developing countries trillions of dollars by 2030 #EndChildMarriage*

### *Relational, Community, and Environmental Levels*

Because tweets on GBV focused primarily on commenting on the experiences of others, stories, or the concept of GBV itself, they can illustrate community perceptions and the larger social norms and structures that facilitate or prevent GBV.

**Facilitators**

GBV was discussed in terms of a physical manifestation of deeply engrained norms and expectations supported by social structures. Both religious and cultural practices were described as facilitators for GBV. These included traditional practices of bride price, female genital cutting or mutilation, and child marriage. Other facilitators of GBV that were discussed included exposure during firewood collection and community property laws that prevent women from leaving violent relationships. The intersection of migration and violence was also discussed, both in terms of immigrants in general and refugees of war. In particular, users cited news stories about rapes of Rohingya women in Myanmar.

> *The rape of Rohingya women by Myanmar's security forces has been sweeping and methodical. AP reports.*

While these cultural practices and phenomena put women at risk, Twitter users described a culture around rape itself that excuses men and perpetuates violence through victim silence and blame.

*So, we can't call out the rapist if he is powerful? We shame the victim? Shame thrives in silence. We will not be silent #Speakup #16days*

Blame is tied closely to users' perceptions of reasons for GBV, as discussed above, including women's promiscuity, frequented establishments, and behaviors. Since the study period corresponded with 16 days of Activism, many of the tweets that discussed this theme were advocating for grassroots movements to destigmatize and end GBV.

There was also conversation around how focusing on male violence against women overshadows other forms of GBV and creates an opportunity for these to occur.

*Feminism won't solve the problem. This focus on women it aggravates it causes men to feel unprotected from various forms of abuse. Women abuse too.*

It was argued that this focus creates stigma for reporting violence perpetrated by a woman against a man. Other identities such as race, sexual orientation, and disability status that influence power dynamics and intersect with gender were also discussed.

Few users discussed weak legal and political support such as police response, access to legal aid, and prosecution by court systems as facilitators to GBV. For the most part, however, these were seen as barriers rather than facilitators.

**Barriers**

Twitter is often used as a platform for social activism; we found several tweets promoting women's empowerment, economic empowerment, and education as barriers to GBV.

*I believe economic empowerment for the girl child is key in ending violence as they will be bold enough to speak, walk out, and support themselves.*

There were also those who supported male engagement in these efforts, both as agents of change to harmful gender norms that underpin GBV and as perpetrators of GBV.

*A real man is one who does not subject a woman to abuse. It's our duty as men not to rape.*

Some tweets were examples of online advocacy, pushing for improved policies and laws against GBV. However, as at least one user noted, laws do exist, but enforcement is low.

While some users argued a larger social movement against GBV is necessary for change, others focused on individual accountability and responsibility.

We noted earlier that silence was described as a facilitator for violence. In the same vein, voice was described as a barrier, allowing survivors to speak up and out against GBV, but also discouraging indifference among family, friends, and communities.

*The key to the end of sexual harassment and violence against women is men calling out their friends who bad-mouth, harass, assault, and victimize women.*

**Mass Media**

In regard to GBV, we found Twitter was primarily used to share information and news reports; politics and activism-related tweets dominated the conversations. The data collection timing overlapped with the 16 Days of Activism against Gender-Based Violence, an international campaign to challenge violence against women and girls (UN Women, 2017).

*1 in 3 women worldwide are subject to violence over the course of their lives. Speak up!!! #16DaysofActivism2017*

*This 16 Days of Activism we say no to violence against women and children. #CountMeIn #SayNoToViolence*

Additionally, Twitter reflected trending stories around current local events such as the Rhodes University rape case, celebrity, and important political figures. Some stories originated outside the countries searched, such as reactions to accusations against U.S. celebrities and political figures.

Quantitative timing analysis showed an uptick in activity around that time, as well as deeper negative net sentiment towards GBV. Qualitative analysis showed discussions of mass media intersected with previously described themes of reasons for, consequences for, barriers to, and facilitators of GBV.

**Linguistic Mainstreaming**

We found a high number of metaphorical uses of different violence-related words including rape, dragging, and b--- slap. For GBV data, there was a frequent and steady use of metaphor over time. One term that increased in use was "dragging," which is slang for insult or joke with bad intent. In some ways, this type of behavior qualifies as emotional violence, but this analysis focused primarily on physical forms of GBV.

# DISCUSSION

## Emerging Themes

### Blessers

Visualization of key words through word clouds and deeper analysis of themes through qualitative analysis provide some insight into conversations around gender norms. These can be viewed and compared over time, as we found by examining word clouds from blesser data at two time points.

This type of analysis can show both how the conversation has evolved and the influence of retweets. Key words and language used can change and be influenced by social media itself. The first set of data collected from Crimson Hexagon (January 1, 2016 to January 31, 2017) showed a lot of discussion around blessers as a concept, as well as conversation related to timely media events. A year later, we found in data from Union Metrics (November 2017 to January 2018) the events had changed, but the influence of timely media and celebrities driving the conversation remained.

Qualitative analysis found an increase in the metaphoric use of search terms over time. Metaphor may indicate a trivialization of an issue or an expression of linguistic mainstreaming that mirrors social normalization (Purohit, 2016). Metaphors are a common device in conversational language; however, it may be important to note how these common devices change in use and meaning over time. In the blesser data, we found that early tweets used the term to specifically refer to an inter-generational transactional relationship. However, tweets from later in the year became more figurative and liberal in meaning. This evolution underscores how social media not only created a term but adapted it and diffused it into greater conversation. This greater acceptance acknowledges larger social normalization.

### Gender-Based Violence

While we did not look at themes over time for GBV, we can see how time-based constructs such as news reports, local events, and media can influence online conversation around GBV. Word clouds and deeper analysis highlight the influence of newsworthy events (e.g., the Rhodes University trial) and international advocacy days (e.g., 16 days of Advocacy) on online conversations. While these larger societal movements dominated the conversation, a few tweets did share more intimate personal experiences or sought out or offered resources for care and counseling. More work should be done on understanding appropriate ways to engage at a program level with those who are using social media in this way to offer services, resources, or support.

Across the data collection time, we found net sentiment to be negative. However, a small percentage of tweets did support GBV, either directly through statements of violence or indirectly through perpetrator excusal and victim-blaming. In engaging in victim-blaming attitudes, society allows the abuser to perpetrate relationship abuse or sexual assault while avoiding accountability for those actions. By also removing accountability, or saying "it wasn't really rape," instances of GBV are downplayed in society as normative. This normalization was mirrored in common metaphoric or non-literal use of GBV-related words, as discussed above.

Additionally, while most of the conversation centered on a hetero-normative cis-gender relationship between men and women, emerging themes of intersectionality and LGBTQ+ challenged this common narrative. Social media may allow for traditionally marginalized groups to express themselves and their challenges more freely (James, 2009).

## Benefits of Using Social Media

For social media users, online spaces allow for connections; identity exploration; a space to express ideas, sexual identities, feelings, and problems; and a place to receive feedback from others (James, 2009). For researchers, social media contributes substantially to the "big data revolution," with its volume, variety, and velocity of available data (Data2x, nd). Effective use of this data in development policymaking and advocacy could improve the lives of women and men, boys and girls by resulting in more efficient services and programs.

Quantitatively, social media data are useful for understanding the timing patterns and geographic concentrations of Twitter activity around a topic. For example, we can see the influence of a specific news story or hashtag on the frequency of tweets. We can also understand where these tweets are occurring, particularly at the country level, even given limitations of location data.

In our dataset we were able to see how events like the 16 days of Activism against Gender-Based Violence Campaign affected the frequency and net sentiment of tweets. We can also see how news articles influenced the discussion, such as the case of blessers Serge Cabonge and Kenny Kunene. Understanding how social media is used to describe and communicate social experiences can help program managers by tapping into these trends.

The majority of these shares were in the form of links to external articles, with little or no additional personal attitudes expressed towards the article. However, hashtags may provide insight into intended sentiment. Hashtags also allow for quick and efficient searching of tweets with higher rates of relevancy.

Additionally, social media data provide an opportunity to collect user-created data that have not been filtered through a research lens. Word clouds can quickly identify popular themes or topics. These data clarify how users are discussing a topic: the popular words and phrases that are used and the events or sub-topics that are driving tweets. Programs can leverage Twitter to develop appropriate, tailored communications by using the language used by the target population.

## Challenges and Limitations

Although Twitter can be useful in identifying the language used around a topic, social media has its own language that can differ by place and topic. Some basic knowledge of the correct acronyms and terminology is necessary to craft an appropriate search string. Because the data is user-created, search strings may include slang or expletives. When conducting the GBV search, we found that users don't speak in research jargon. The same was true for intergenerational, transactional sex, which is how we came to the term "blessers." While "sugar-daddy" produced some results, that term has gone out of style among younger generations. "Blesser" was a much more popular and trendy term, particularly in South Africa. Language, especially online, is quick to evolve. Even during the analysis of this data, we discovered we had unknowingly excluded much data on transactional relationships between older women and younger men, because these are associated with the term "Ben 10," not "blesser." The search terms used can greatly affect the data collected.

The data available vary depending on the social media site and the method of abstraction used. The data available are also subject to the rules and privacy regulations of the social media site at the time of abstraction. These rules and regulations are subject to change at any time and may interrupt your data collection process if not closely monitored. Working through a third-party site like Crimson Hexagon or Union Metrics mitigates this challenge, as they are monitoring and responding to such changes. When scraping data through Twitter API or Microsoft Flow, the onus is on the researcher to monitor and account for any changes.

The data may also be biased and not representative. For instance, income and education are known to be positively associated with the likelihood of tweeting about health-related issues, while age and male gender are negatively associated with this (Gruebner, 2017; Nsoesie, 2016). Uneven access to the Internet across public and private spaces similarly may impact who is able to tweet and when, limiting the representativeness of certain geosocial data (Bhattacharya, 2012). Furthermore, social media behavior is complex and contextually dependent. Behaviors such as retweeting, screen captures, or "hate-linking" may be algorithmically invisible, and thus subject to incomplete or incorrect representation in studies utilizing computational methods. Finally, social media is produced for public consumption; users are self-aware and may curate content to project an image not representative of their actual health-related beliefs, attitudes, or behaviors (Tufekci, 2014).

It is known that significant gender-related content exists in social media, and that statistical models can be used to determine the gender of uncharacterized users (Burger, 2011). However, because a "gender" variable is not collected directly from Twitter users and is only available from Crimson Hexagon through an algorithm, only a small subsample of users could be identified as male or female with confidence. There may be error in the accuracy of these determinations (e.g., how a computer assigns gender may be different than how one self-identifies). Recent studies have utilized automatic classifiers such as Support Vector Machines and Naive Bayes to predict gender with up to 92 percent accuracy using lexical variation, exploring relationships between gender and health-related content, quality and quantity of social media posts, cultural venue preferences, and social networks (Burger, 2011; Purohit, 2016; Mueller, 2017). Interestingly, it has been found that individuals whose gender is not correctly predicted by machine-learning algorithms have fewer same-gender social connections in their online network (Bamman, 2014).

Similarly, the sentiment generated by a computer algorithm may not reflect the sentiment intended by the user in regard to the research topic. We found discordance in 59 percent of tweets coded by computer versus human for sentiment. (MEASURE Evaluation, 2018) This is in part to the driving purpose of the coding: pure language analysis versus language analysis regarding a topic. These algorithms and technologies are rapidly improving and evolving and may in the near future be able to account for this distinction. Currently, costly human coding would be necessary to ensure quality results if the goal was to understand prevailing sentiment on a topic.

Additionally, it may be difficult to interpret the intent of a user's retweet without context. For example, a user may retweet the following:

*RT @BBCWorld: Emmanuel Macron: Domestic violence is "France's shame" https://t.co/IETtk1smo4*

This may indicate the user supports Emmanuel Macron's statement. However, if the user is outspoken against Mr. Macron in other tweets, it may be viewed through a more negative lens. Without more context, it is difficult to understand true intention and sentiment. For this reason, we excluded retweets and simple quotations from our qualitative analysis, focusing on regular tweets and replies.

Finally, we found that not all issues discussed were reflective of local events, ideas, or people. This illustrates an important challenge in the interpretation of social media data in an increasingly globalized world.

## Ethical Considerations

The issue of how ethical principles are applied to social media research proved a challenge. A full discussion of these issues is beyond the scope of this document, but users of this type of data should fully consider ethical and privacy implications around the data. A guidance document based on this research includes lessons learned and ethical considerations. (MEASURE Evaluation, 2018)

Ethics around social media and privacy is an evolving field, and individual social media platforms can change their policies at any time. This may influence not only the data that is collected as discussed earlier, but also an IRB's inclusion of social media data under review.

Because tweets themselves can be identifying, this report chose to use aggregated quotations. Several tweets around the same topic were brought "together to maintain the essence of meaning but rendering it impossible to identify any individual poster." Aggregate quotations may lose some nuance and introduce bias in analysis and reporting, but we chose to use them to maintain confidentiality, particularly around sensitive subjects such as GBV.

# CONCLUSIONS

Analysis of social media trends in timing, sentiment, and conversation themes can uncover user attitudes toward gender norms that drive health behaviors. Understanding how, when, and under what circumstances people talk about gender norms and behaviors can help programs tailor messaging and interventions more precisely, to reduce negative social and health outcomes. However, challenges remain in using social media data, including ethical concerns, shifting policies and procedures, and generalizability of findings.

# REFERENCES


Bamman, D., Eisenstein, J., & Schnoebelen, T. (2014). Gender identity and lexical variation in social media. *Journal of Sociolinguistics, 18*(2), 135–160. Retrieved from https://onlinelibrary.wiley.com/doi/abs/10.1111/josl.12080

Bhana, D., & Pattman, R. (2011). Girls want money, boys want virgins: The materiality of love amongst South African township youth in the context of HIV and AIDS. *Culture, Health, and Sexuality, 13*(8), 961–972. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/21656409

Bhattacharya S., Tran H., & Srinivasan, P. (2012). Discovering health beliefs in Twitter. *AAAI Fall Symposium - Technical Report*, FS-12-05 (January), 2–7. Retrieved from https://www.aaai.org/ocs/index.php/FSS/FSS12/paper/viewFile/5591/5842

Borwankar, R., Diallo, R., Goings, S., Sommerfelt, E., & Oluwole, D. (2009). *Gender-Based Violence in Sub-Saharan Africa: A Review of Demographic and Health Survey Findings and Their Use in National Planning*. Washington, D.C. Retrieved from https://www.wilsoncenter.org/sites/default/files/GBV in SSA.pdf

Burger, J.D., Henderson, J., Kim, G., & Zarrella, G. (2011). Discriminating Gender on Twitter. *Association for Computational Linguistics*, *146*, 1301–1309. Retrieved from http://www.aclweb.org/anthology/D11-1120

Data2x. n.d. Retrieved from http://www.data2x.org/partnerships/big-data-gender/

Dunkle, K.L., Jewkes, R.K., Brown, H.C., Gray, G.E., McIntryre, J.A., & Harlow, S.D. (2004). Transactional sex among women in Soweto, South Africa: Prevalence, risk factors and association with HIV infection. *Social Science & Medicine, 59*(8), 1581–1592. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/15279917

Elsherief, M., Belding, E., & Nguyen, D. (2017). # NotOkay: Understanding gender-based violence in social media. *Eleventh International AAAI Conference on Web and Social Media*, 52–61. Retrieved from https://aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15602

Esty, D., & Rushing, R. (2007). The promise of data-driven policymaking. *Issues in Science and Technology*, 23(4), 67. Retrieved from https://issues.org/esty-2/

Fox, A.M., Jackson, S.S., Hansen, N.B., Crewe, M., & Sikkema, K.J. (2007). A qualitative study of women's risk for intimate partner violence and HIV. *Violence Against Women, 13*(6), 583–602. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/17515407

Gruebner, O., Sykora, M., Lowe, S.R., Shankardass, K., Galea, S., & Subramanian, S.V. (2017). Big data opportunities for social behavioral and mental health research. *Social Science & Medicine*, 189, 167–169. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/28755794

Hunter, M. (2002). The materiality of everyday sex: Thinking beyond "prostitution." *African Studies, 61*(1), 99–120. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/00020180220140091

James, C., Davis, K., Flores, A., Francis, J.M., Pettingill, L., Rundle, M., et al. (2009). *Young People, Ethics and the New Digital Media: A Synthesis from the Good-Play Project*. Cambridge, MA: MIT Press.

Kilonzo, N., Ndung'u, N., Nthamburi, N., et al. (2009). Sexual violence legislation in sub-Saharan Africa: The need for strengthened medico-legal linkages. *Reproductive Health Matters, 17*(34), 10–19.

Kuehn, B.M. (2015). Twitter streams fuel big data approaches to health forecasting. *Journal of the American Medical Association, 314*(19), 2014–2016. Retrieved from https://jamanetwork.com/journals/jama/fullarticle/2470443

MEASURE Evaluation. (2018). Using social media data to understand changes in gender norms: Guidance. Chapel Hill, NC, USA: MEASURE Evaluation. Retrieved from https://www.measureevaluation.org/resources/publications/ms-18-147/

Mueller, W., Silva, T.H., Almeida, J.M., & Loureiro, A.A. (2017). Gender matters! Analyzing global cultural gender preferences for venues using social sensing. *EPJ Data Science, 6*(1), 5. Retrieved from https://link.springer.com/article/10.1140/epjds/s13688-017-0101-0

Nsoesie, E.O., Flor, L., Hawkins, J., et al. (2016.) Social media as a sentinel for disease surveillance: What does sociodemographic status have to do with it? *PLoS Currents*, 8, 1–26. Retrieved from http://currents.plos.org/outbreaks/article/social-media-as-a-sentinel-for-disease-surveillance-what-does-sociodemographic-status-have-to-do-with-it/

PEPFAR (President's Emergency Plan for AIDS Relief). (2007). *The Power of Partnerships: The President's Emergency Plan for AIDS Relief Third Annual Report to Congress*. Retrieved from https://www.pepfar.gov/press/c21604.htm

PEPFAR (President's Emergency Plan for AIDS Relief). (2017). Monitoring, Evaluation, and Reporting Indicator Reference Guide (MER 2.0). Retrieved from https://www.pepfar.gov/documents/organization/263233.pdf

Portland Communications. (2016.) How Africa tweets. Retrieved from http://www.howafricatweets.com/

Purohit, H., Banerjee, T., Hampton, A., Shalin, V.L., Bhandutia, N., & Sheth, A. (2016). Gender-based violence in 140 characters or fewer: A #BigData case study of Twitter. *First Monday, 21*(1). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/6148

Shefer, T., & Strebel, A. Deconstructing the "sugar daddy": A critical review of the constructions of men in intergenerational sexual relationships in South Africa. (2012). *Agenda,* 26(4):57-63. Retrieved from http://dx.doi.org/10.1080/10130950.2012.760837

Stoebenau, K., Heise, L., Wamoyi, J., & Bobrova, N. (2016). Revisiting the understanding of "transactional sex" in sub-Saharan Africa: A review and synthesis of the literature. *Social Science & Medicine*, 168, 186–197. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/27665064

Toska, E., Cluver, L.D., Boyes, M., Pantelic, & M., Kuo, C. (2015.) From "sugar daddies" to "sugar babies": Exploring a pathway among age-disparate sexual relationships, condom use and adolescent pregnancy in South Africa. *Sexual Health.*, 12(1), 59–66. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4546918/

Tufekci, Z. Big questions for social media big data: Representativeness, validity and other methodological pitfalls. (2014). Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewFile/8062/8151

UNAIDS. (2017). *UNAIDS Fact Sheet: Latest Statistics on the Status of the AIDS Epidemic*. Retrieved from http://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf

United Nations Division for the Advancement of Women (UN WOMEN). (2005). *Violence against Women: A Statistical Overview, and Challenges and Gaps in Data Collection and Methodology and Approaches for Overcoming Them Needs of Users of Data Collection*. Vol 4. Retreived from http://www.un.org/womenwatch/daw/egm/vaw-stat-2005/docs/final-report-vaw-stats.pdf

United Nations Division for the Advancement of Women (UN WOMEN). (2016). World AIDS Day Statement: For young women, inequality is deadly. Retrieved from http://www.unwomen.org/en/news/stories/2016/11/un-women-statement-aids-day

World Health Organization (WHO). (2015). Glossary of terms and tools. Geneva, Switzerland: WHO. Retrieved from http://www.who.int/gender-equity-rights/knowledge/glossary/en/

# APPENDIX A. SEARCH TERMS

## Search Query for Age-Discordant Transactional Relationships:

- ("blesser" OR "blessee" OR "Sugar Daddy" OR "Sugardaddy") (place_country:KE OR place_country:LS OR place_country:SZ OR place_country:UG OR place_country:ZA OR place_country:MW OR place_country:MZ OR place_country:TZ OR place_country:ZM OR place_country:ZW)

## Search Query for GBV:

- [woman OR women OR girl OR girls OR female OR boyfriend OR girlfriend OR boy-friend OR girl-friend OR bitch OR domestic] **AND** [drag OR dragged OR dragging OR kick OR kicked OR kicking OR beat OR beating OR beaten OR slap OR slapped OR slapping OR punch OR punched OR punching OR violence OR abuse OR abused OR abusing OR assault OR assaulted OR assaulting OR harass OR harassed OR attack OR attacked OR attacking OR grope OR groped OR groping OR stalk OR stalking OR stalked OR trafficking OR trafficked]

- “rape” OR “sexual assault” OR “sexual violence” OR “sexual abuse” OR “force sex” OR “child marriage” OR “children marriage” OR “underage marriage” OR “forced marriage” OR “sex trafficking” OR “child trafficking” OR “children trafficking” OR “female genital mutilation” OR “female genital cutting OR “forced prostitution”

**MEASURE** Evaluation
University of North Carolina at Chapel Hill
123 West Franklin Street, Suite 330
Chapel Hill, North Carolina 27516
Phone: +1-919-445-9350
**measure@unc.edu**
**www.measureevaluation.org**

This publication was produced with the support of the United States Agency for International Development (USAID) under the terms of MEASURE Evaluation cooperative agreement AID-OAA-L-14-00004. MEASURE Evaluation is implemented by the Carolina Population Center, University of North Carolina at Chapel Hill in partnership with ICF International; John Snow, Inc.; Management Sciences for Health; Palladium; and Tulane University. Views expressed are not necessarily those of USAID or the United States government. TR-18-295
ISBN: 978-1-64232-092-3