CAN BIG DATA BE USED FOR EVALUATION?

A UN Women feasibility study

by **Claudia Abreu Lopes,
Savita Bailur** and
Giles Barton-Owen
13 September 2018

OUTLINE

- What is big data and why use it?

- What are the sources of big data?

- What are the risks and challenges in using big data?

- Findings: Can big data be used for evaluations?

- Conclusions and Recommendations

STUDY OBJECTIVES

- **Exploratory feasibility study** based on cases related to a global WPP Evaluation with site visits in Mexico and Pakistan

- **Determine if it is possible to improve UN Women evaluation** using additional evidence streams to complement, triangulate or widen evidence base

- Provide feedback for the use and refinement of **the new UNDG Principles related to ethical use of big data for evaluation**, specifically for GEWE

- **Understand how UN Women and partners can use big data**

What is big data and why use it?

What is big data and why use it?

- **By-product of people's digital behavior.**

- **Requires interpretation after the event.**

- **Imperfectly matches the entire universe of cases.**

- **Non-coverage is often a concern when assessing data quality.**

- **Is often accessible in near real time (at the time the data are produced).**

- **Requires combining different data sources.**

- **Can be harnessed to improve decision making.**

Evaluation: traditional methods, within timeframe & resources available

Big Data

Big Data is emerging as a source for identifying large-scale trends, understanding whether interventions have worked, and predicting outcomes of interventions.

What are the sources of big data?

- **Twitter**

- **Facebook**

- **WhatsApp**

- **Radio data**

- **News data**

What are the sources of big data?

Social media sources selected: Twitter & Facebook

- **Easily accessible, nature of UN Women's campaigns/intervention in selected case countries, interest by UN Global Pulse**

- **Twitter (Mexico): better for sharing information quickly and publicly, with short opinions, links and news headings**

- **Facebook (Pakistan): more suited for longer lasting interactions, more in-depth discussion and more personal sharing within context of friendships networks**

What are the risks and challenges in using big data?

- **Elite capture: major risk of 'black holes' of data where entire demographics can be missed because of restricted access and use**

- **Restricted use: factors that constrain women's ability to participate in social media or other interactive platforms**

- **Ambient sexism: perceived misogyny on social media**

- **Platform drifts: population, usage and system**

- **Ethical and privacy issues**

METHOD:

Can big data be used for evaluations?

Pilot cases

MEXICO

🐦 25.7 M Twitter users in 2017

3️⃣ One of the top three countries of Twitter users

📶 65.3% of the population has internet access

PAKISTAN

🐦 3.1 M Twitter users in 2016

f More than 31 M Facebook users by end of 2017

📶 22% of the population has internet access

Method

As an exploratory study, learning was needed for each step towards determining how to derive insights from social media for evaluation use:

- Selection of case countries (Mexico and Pakistan)
- Review of potential data sources (pros and cons)
- Big data source selection and criteria
- Protocol for data extraction and analysis
- Adequacy and limitations of each data source

Four stages for social media analysis

1 MEASUREMENT MODEL TO SELECT THE BEST BIG DATA INDICATORS

Proposed indicator to measure SDG 5.5

Percentage of seats held by women in elected offices

Traditional data sources to measure political participation...

Valid and trusted international survey results e.g. World Values Survey

... are used to validate indicators from social media

Explore concepts from Theory of Change and map them onto possible social media indicators (e.g, hashtags and keywords)

Hybrid approach traditional + big data

Validate the big data indicators with traditional data i.e. measuring the same construct

2 DESCRIBING THE UNIVERSE OF RELEVANT TWEETS (& FACEBOOK POSTS)
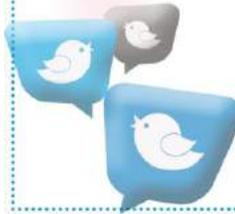
Analysis of tweets

All tweets/posts with relevant hashtags and keywords
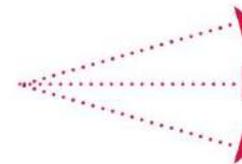
Engagement : likes, shares and comments

Demographics of users

Sentiment of tweets/posts/comments

3 ANALYSIS OF RESULTS ACROSS REGIONS & OVER TIME

Regional comparison of engagement

For example, comparing engagement in Mexican districts where UN Women invested more resources with other districts

Longitudinal analysis of engagement, demographics and sentiment

Mapping a timeline of UN Women events in the target regions onto the patterns of social media results over time

4 TRIANGULATE & COMPLEMENT EVALUATION FINDINGS WITH BIG DATA

The results of the big data analysis can be triangulated with the results of the corporate evaluation and interpreted alongside traditional data, such as interviews that provide a deeper understanding of the contribution of UN Women to observed changes.

Pilot study in Mexico

Social media campaigns captured by a series of hashtags used by the Country Office in Mexico for different purposes over a long period of time.

The hashtags were from different initiatives not necessarily linked to the UN Women strategy on WPP.

Pilot study in Pakistan

Twitter use remained generally low and remained within upper echelons of society in Pakistan, as recognized in the risks above, Facebook use increased dramatically since the 2013 elections, particularly within youth demographics

#IgualdadeDeGenero

#NiUnaMenos

#Planeta5050

#México5050

#DemosElPaso

#NinasNoEsposas

#ODS

#ODS5

#MujeresPoderosas

#Agenda2030

#ÚNETE

#ATENEA

FINDINGS:

Can big data be used for evaluations?

Achievements

Study allowed for identification of big data analysis steps for each source and associated challenges for accessing and analyzing social media data

1. Protocol now available for other UN Women evaluations using big data (and partners).

   - **Twitter:** steps for data analysis and results for each step
   - **Facebook:** stages for data analysis; no results given time restrictions to develop language models for Urdu
   - **Radio data:** process discussed, but not implemented due to timeframe

2. Learnings associated with challenges in methods, data access and analysis and solutions discussed

Steps for analysis of Twitter in Mexico

1. **Define a measurement model to select the best indicators (e.g., hashtags) that correlate with traditional data (surveys, voters' registration data, KII, FGDs)**

2. **Describe the universe of users and evaluate exclusions of users related to language and demographics**

3. **Hashtag analysis to uncover geographic and longitudinal trends**

4. **Sentiment analysis and statistical co-occurrences to correlate sentiment with regions**

5. **Triangulate traditional and big data sources to complement findings from evaluation**

Hashtag analysis: Longitudinal trends
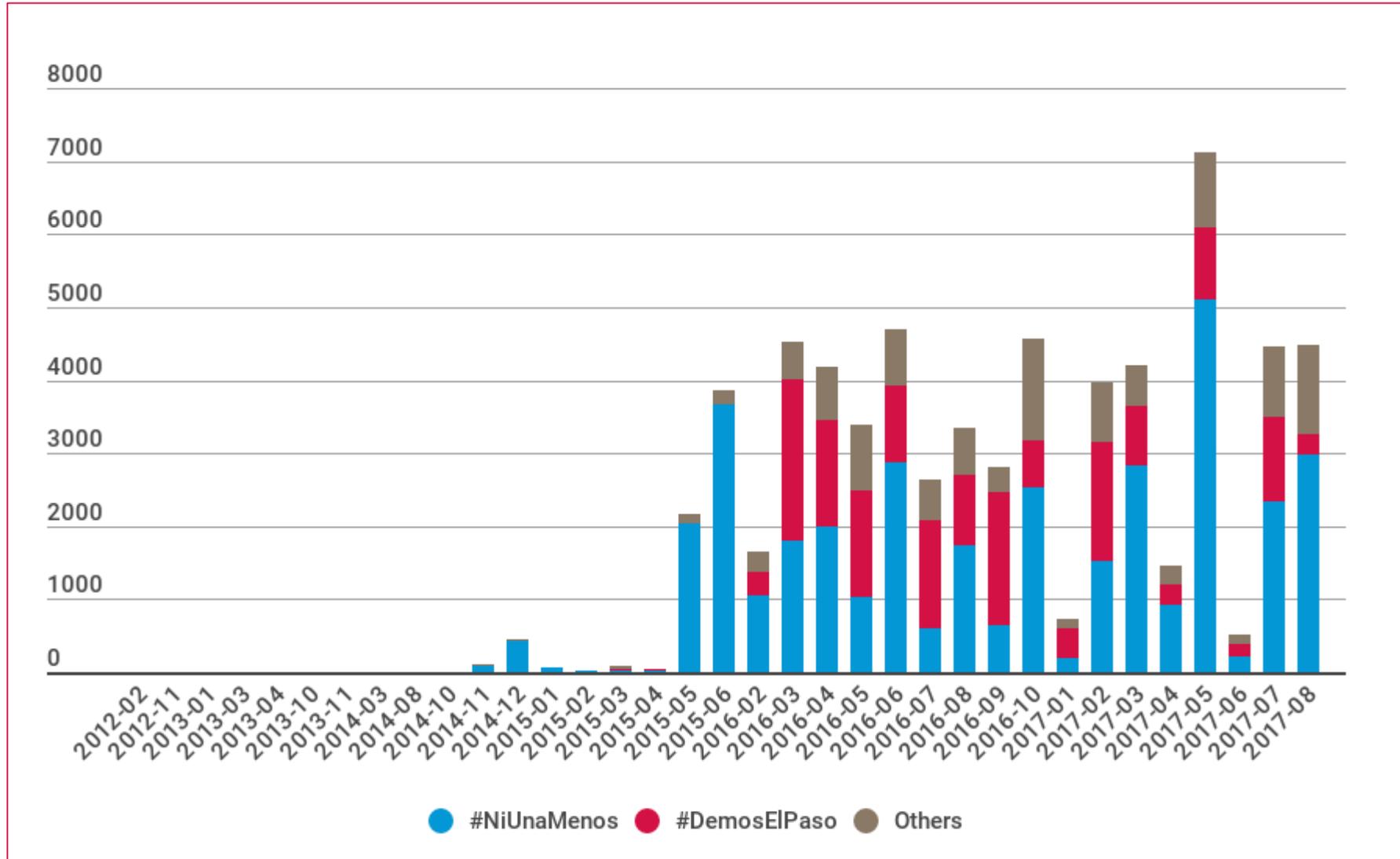
Legend: ● #NiUnaMenos ● #DemosElPaso ● Others

Positive sentiment 2016

Positive sentiment 2017

Negative sentiment 2016
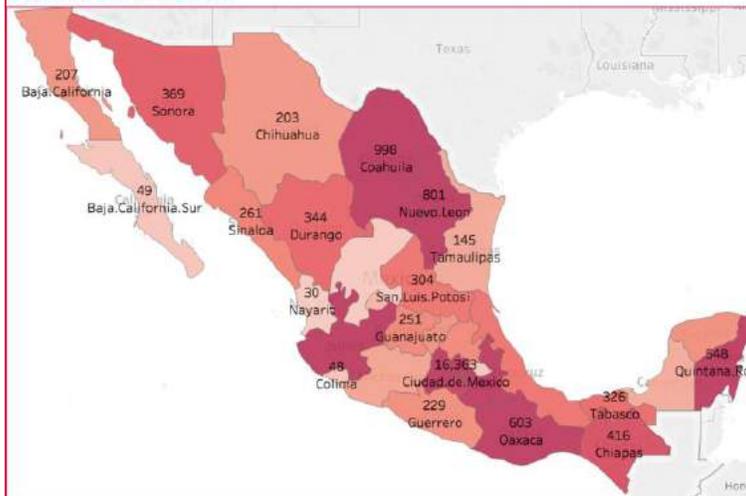
Negative sentiment 2017

Learning from the analysis

- **Challenges in getting historical data (time-consuming, multiple queries)**

- **Dominance of certain hashtags**

- **High volume of re-tweets (75%)**

- **Sentiment analysis reflects nature of topic (vs attitude of user)**

- **Automatic content analysis limiting (vs. crowd-coding)**

- **Need for contextual interpretation of findings**

Conclusions and Recommendations

- **Twitter appears more appropriate for evaluating UN Women's interventions** aimed at fostering political participation and attitudes towards gender equality.

- **Social network analysis can help to reveal the online network of users and their degree of influence within their network.** This type of analysis may be able to answer questions related to the reach and spread of information through Twitter.

- Analysis and interpretation of conversations within a cultural context can be enhanced **by focus groups with Twitter users and/or validated by media and domain experts from the country.**

- **Private or semi-private discussions may pose ethical issues** because they can reveal sensitive personal details that could place users at risk.

- **Many pages from organizations do not contain much discussion;** pages associated with political or social issues have biased samples, as people self-select strongly based on their views on those issues.

- Other sources hold more promise, such **as radio data, responses to SMS campaigns and responses to newspaper articles online.**

Findings on Radio data

- Radio can be a **significant social venue.**

- **Historical streaming of radio data is not always present.**

- Radio programmes can be **designed to gather useful information for evaluation** through voice or SMS.

- **Requires careful recording and coordination** to ensure large volume of data is available for analysis, but can be highly relevant and rich (e.g., documenting community conversations).

Recommendations

1. **Understand the bigger picture of big data in a country** before considering it as a source for evaluation.

2. Big data should be incorporated **in the design of the evaluation from the outset.**

3. **Big data should precede traditional data** when sequencing and evaluating.

4. Big data can be **shaped in ways that enhance its value.**

THANK YOU