

USING BIG DATA FOR UN WOMEN EVALUATIONS

A FEASIBILITY STUDY ON USING BIG DATA WITH A CASE FROM MEXICO

**SEMINARIO INTERNACIONAL SOBRE EVALUACIÓN CON
PERSPECTIVA DE GÉNERO**

by Claudia Abreu Lopes

1 Junio 2018

Outline

1

¿Por qué Big Data para la evaluación?

2

Método y casos de estudio

3

Análisis y resultados

4

Conclusiones y recomendaciones

¿Por qué Big Data para la evaluación?

Cómo se cruzan big data y la evaluación

Evaluation:

métodos tradicionales,
dentro de los plazos y
recursos disponibles

Big Data

Big Data surge como una fuente para identificar tendencias a gran escala, entendiendo si las intervenciones han funcionado y prediciendo los resultados de las intervenciones

¿Cómo pueden los big data fortalecer las evaluaciones de UNW?

- Complementa los métodos tradicionales y permite la expansión del alcance de las evaluaciones
- Proporciona una base de evidencia adicional para la triangulación y aumenta la credibilidad de las evaluaciones
- Revela tendencias a gran escala que los métodos tradicionales no pueden capturar (por ejemplo, geográfica o longitudinal)
- Puede elucidar en la atribución de resultados
- Permite un uso eficiente de los recursos

Método y casos de estudio

Study Objectives

- Estudio exploratorio de factibilidad basado en casos relacionados con la Evaluación WLPP en México
- Determinar si es posible mejorar la evaluación UNW utilizando otras formas de evidencia para complementar, triangular o ampliar la base de evidencia
- Proporcionar feedback para el uso y el refinamiento de los nuevos Principios del UNDG relacionados con el uso ético de Big Data para la evaluación
- Comprender cómo UNW y otras organizaciones pueden usar big data

Método

- Como estudio exploratorio, se necesitaba aprender cada paso para determinar cómo obtener información de las redes sociales para el uso de la evaluación:
 - Revisión de posibles fuentes de datos (pros y contras)
 - Selección de fuente de big data y criterios
 - Protocolo para extracción y análisis de datos
 - Adecuación y limitaciones de cada fuente de big data

Método: fuentes de datos

Fuentes de redes sociales seleccionadas – Twitter and Facebook

- De fácil acceso y posible dentro del marco de tiempo de tres meses del estudio para acceder a los conjuntos de datos (por ejemplo, no se seleccionaron los datos de radio, pero fueron prometedores)
- naturaleza de las campañas intervenciones de ONU Mujeres en Mexico
 - Twitter: mejor para compartir información rápida y públicamente, con opiniones cortas, links y encabezados de noticias
 - Facebook: más adecuado para interacciones más duraderas, discusiones más profundas y un intercambio más personal dentro del contexto de las redes de amistad

Análisis y resultados

Twitter analysis in Mexico

- **Contexto:** alta aceptación de las redes sociales; un buen nivel de recursos (español) con bibliotecas accesibles y herramientas de análisis textual; los 3 mejores países para usuarios de Twitter; varias campañas de ONU Mujeres México en las redes sociales (#DemosElPaso y # Mexico5050).
- **Dataset:** subconjunto de tweets de México en español de 2012 a 2017 - 65.512 usuarios con tweets que contienen al menos uno de los 12 hashtags utilizados por la cuenta de Twitter de ONU Mujeres México
- **Población:** usuarios de hashtags, seguidores de usuarios y seguidores seguidores en conversaciones y retweets.
- **Herramientas:** Hexágono Crimson, paquetes Python / R, Stompol para análisis de sentimiento

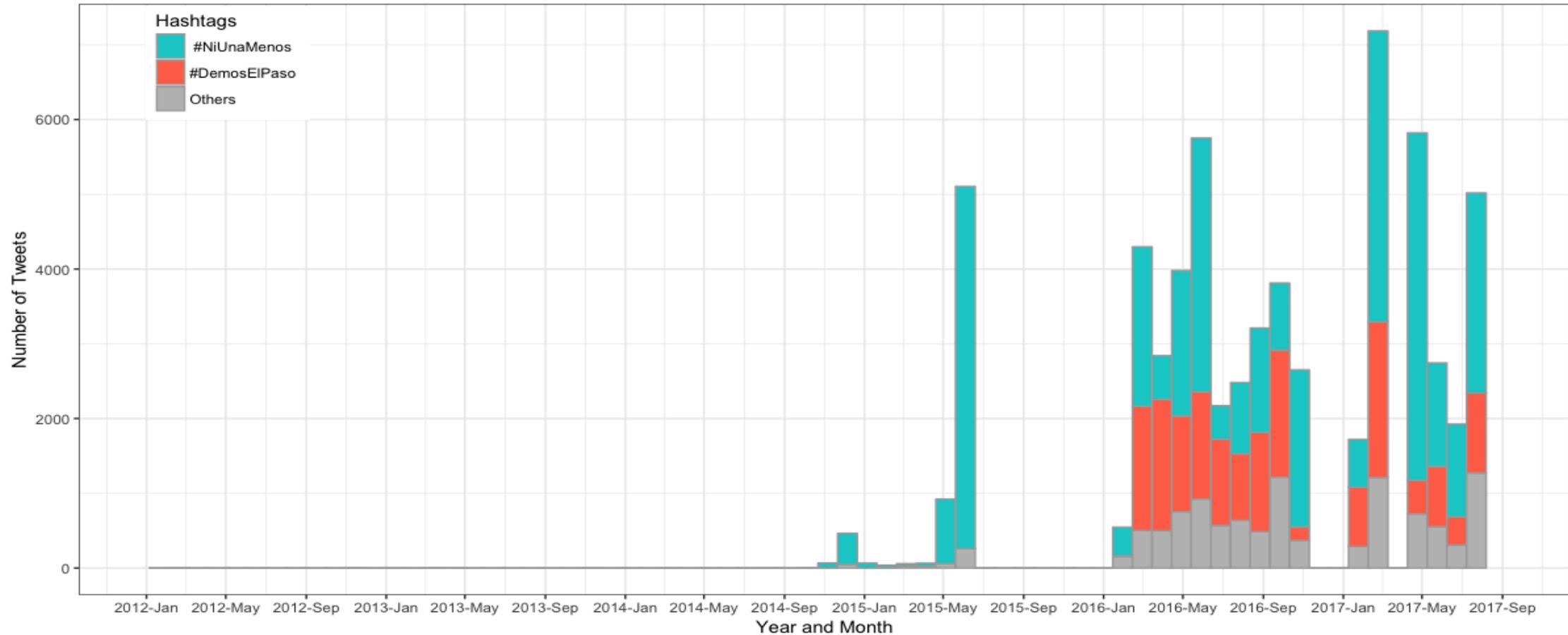
Pasos para el análisis de Twitter en México

1. Defina un modelo de medición para seleccionar los mejores indicadores (por ejemplo, hashtags) que se correlacionen con los datos tradicionales (encuestas, datos de registro de votantes, KII, FGD)
2. Describir el universo de usuarios y evaluar exclusiones de usuarios relacionadas con el idioma y la demografía
3. Análisis Hashtag para descubrir tendencias geográficas y longitudinales
4. Análisis del sentimiento y co-ocurrencias estadísticas para correlacionar el sentimiento con las regiones
5. Triangular fuentes tradicionales y grandes de datos para complementar los resultados de la evaluación

Característica demográfica de los usuarios

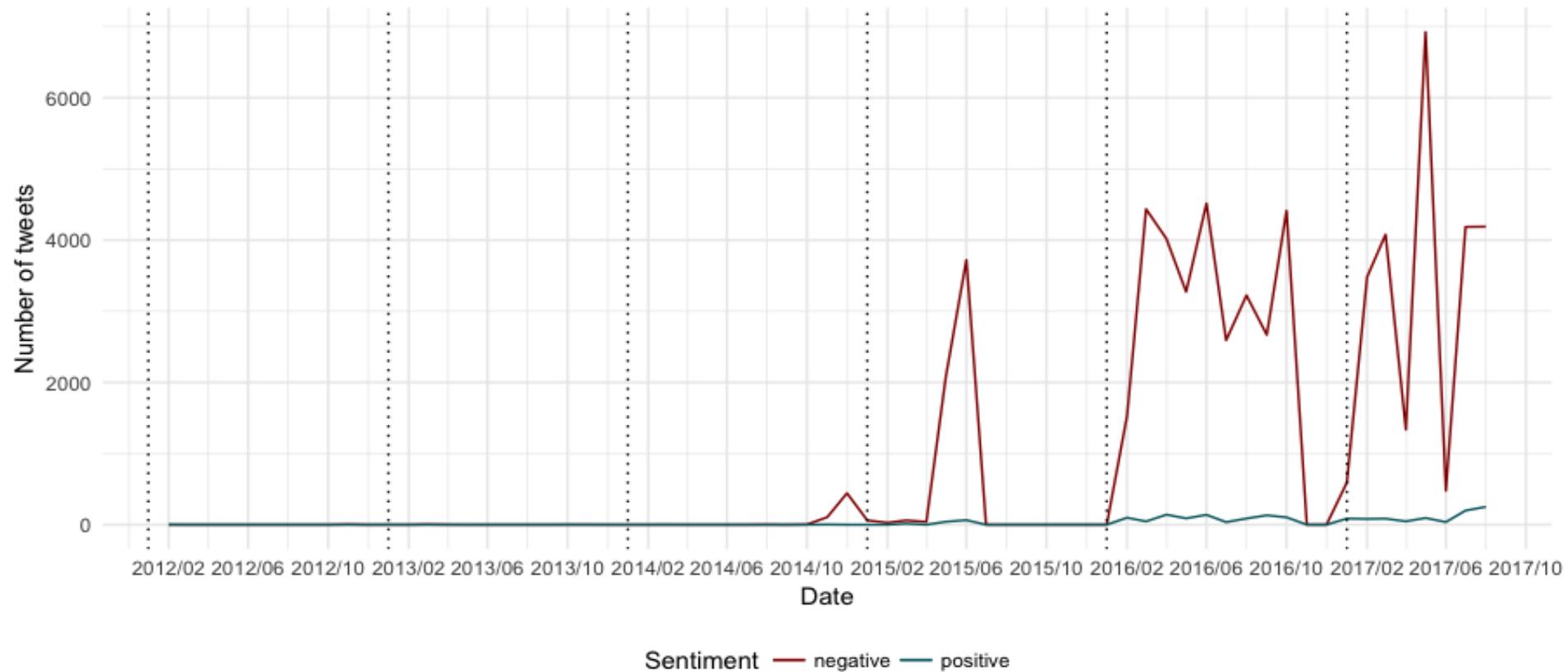
- 65,512 usuarios diferentes
- 55.1% de los usuarios son mujeres; 44.9% son hombres
- Distrito Federal con 59.7% de tweets
- Mexico City con 62.8% de tweets
- Idioma de los tweets clasificados como español

Análisis de hashtags: tendencias longitudinales



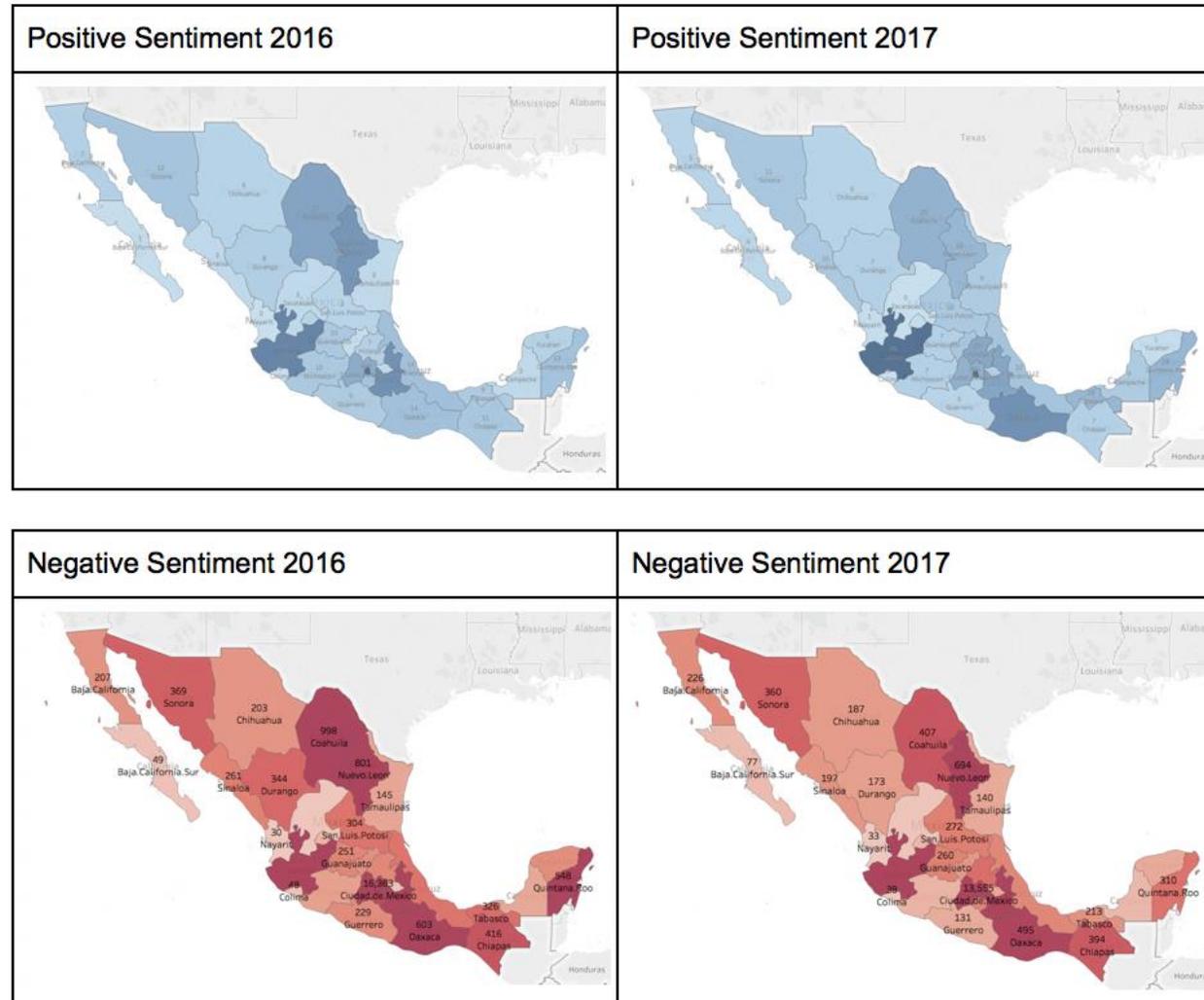
Tweets mexicanos de enero de 2012 a agosto de 2017 que contienen al menos uno de los hashtags

Análisis de sentimiento: tendencias longitudinales



Análisis de sentimiento de tweets subdivididos por año

Análisis de sentimiento: tendencias geográficas



Heatmaps para tweets con sentimiento positivo y negativo

Aprendizajes del análisis

- Desafíos para obtener datos históricos (que consumen mucho tiempo, múltiples consultas)
- Dominación de ciertos hashtags
- Alto volumen de re-tweets (75%)
- El análisis del sentimiento refleja la naturaleza del tema (vs a la actitud del usuario)
- Necesidad de una interpretación contextual de los hallazgos

Conclusiones y recomendaciones

Twitter analysis

- Twitter parece más apropiado para evaluar las intervenciones de ONU Mujeres basadas en opiniones y actitudes hacia la igualdad de género y la participación política de las mujeres.
- El análisis de redes sociales puede ayudar a revelar la red en línea de usuarios y su grado de influencia dentro de su red. Este tipo de análisis puede responder preguntas relacionadas con el alcance y la diseminación de información a través de Twitter.
- Dada la corta vida de los hashtags, el análisis longitudinal basado en los mismos hashtags no es significativo.

Twitter analysis

- El análisis y la interpretación de las conversaciones dentro del contexto cultural pueden mejorarse mediante grupos de enfoque con usuarios de Twitter y / o validados por expertos en medios y comunicación del país
- Otras fuentes también son prometedoras: datos de radio, respuestas a campañas de SMS; respuestas a artículos online.

Recomendaciones

- 1. El análisis de Big Data debe incorporarse en el diseño de la evaluación desde el principio**

Recommendations

- 1. Understand the bigger picture of big data in a country before considering it as a source for evaluation**
- 2. Comprender el panorama general de los grandes datos en un país antes de considerarlo como una fuente de evaluación**
- 3. Los grandes datos se pueden configurar de manera que mejoren su valor**
- 4. Los datos grandes deben preceder a los datos tradicionales cuando se secuencian y evalúan**

MUCHAS GRACIAS