



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Análisis de frecuencia de hashtags en Twitter

TRABAJO FINAL DE MÁSTER

José Alberto Pérez Melián
Valencia, 26 de septiembre de 2016

Índice

1 Introducción

2 ETL

3 Metodología y resultados

4 Conclusiones

1 **Introducción**

1 Introducción

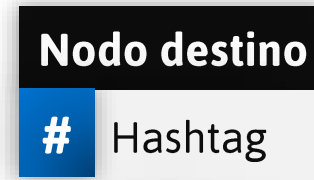
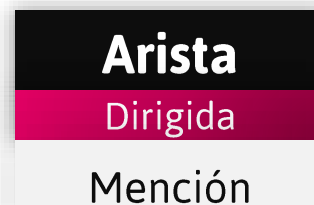
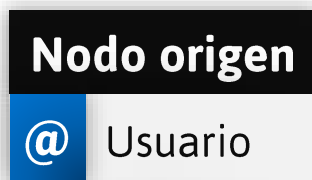
Redes sociales

Servicios que permiten a los usuarios establecer conexiones y compartir información entre ellos

Twitter



En este trabajo, Twitter se puede representar como



1

Introducción

Objetivo

Conocer el comportamiento de los hashtags estudiando la distribución de su frecuencia de acuerdo a su popularidad

Una vez conocida la distribución de frecuencias, nos permitiría detectar patrones que indiquen comportamientos anómalos

2

ETL

2

ETL

Hispatweets

11,9 GB

Elecciones

1,35 GB

Bárcenas

52,6 MB

{JSON}

26 campos

usuario

hashtag

Hashtag

Estandarización

MapReduce

CSV

hashtag

Nº #

Hispatweets

6,33 MB

Elecciones

130 KB

Bárcenas

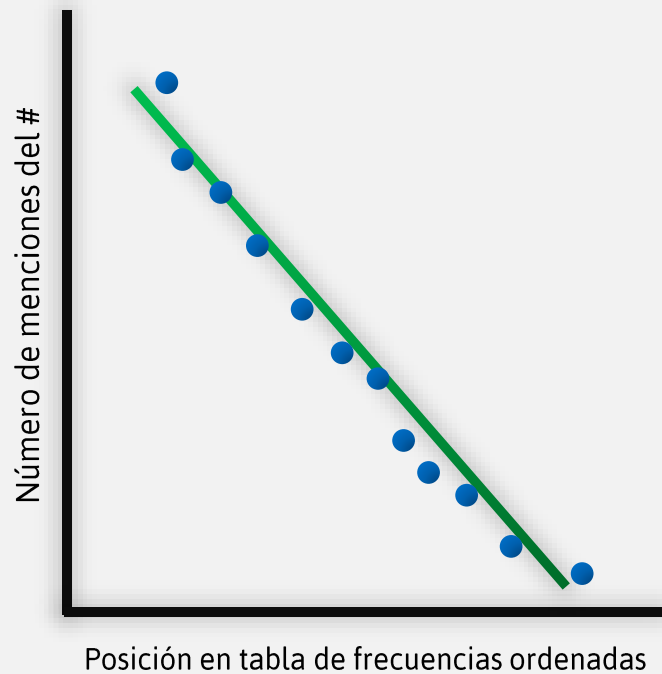
60 KB

3 Metodología y resultados

3

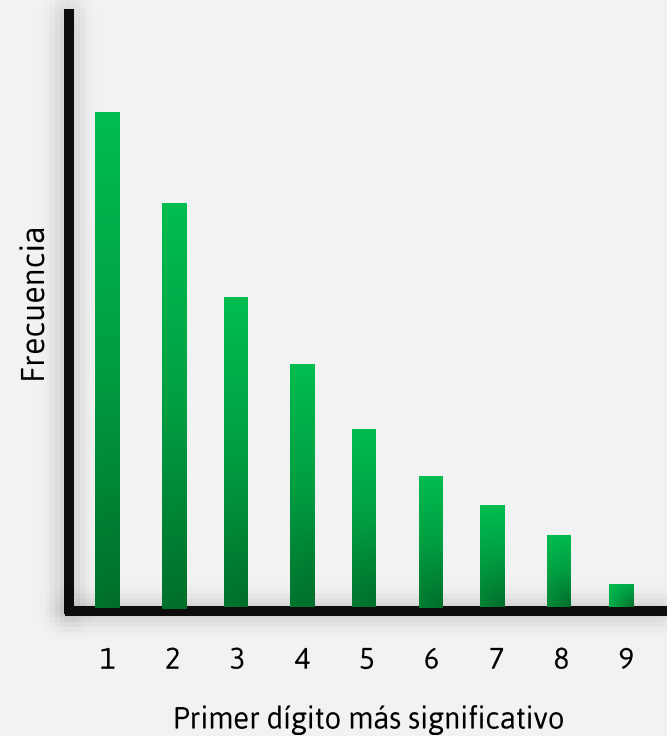
Metodología y resultados

Ley de Zipf



Análisis de la calidad de los datos

Ley de Benford

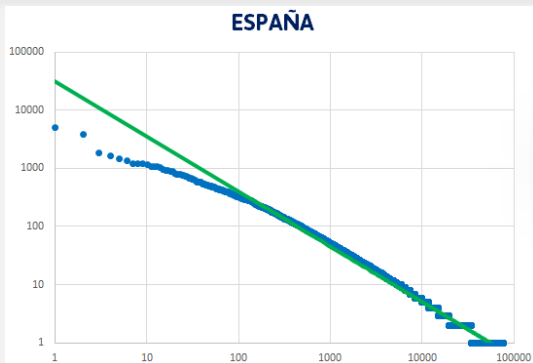


Detección de anomalías en la distribución de los hashtags

3 Metodología y resultados

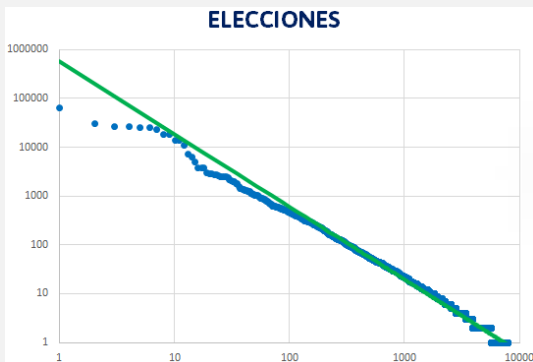
Ley de Zipf

AJUSTE



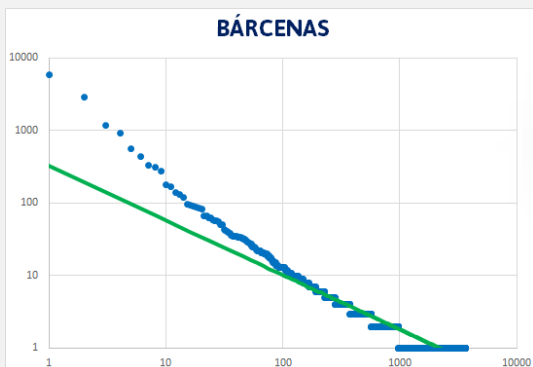
0,9628

Si



0,9879

Si

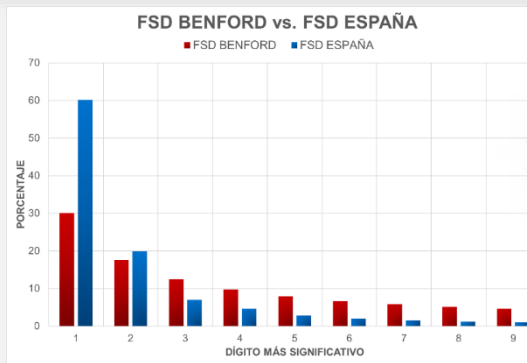


0,8785

No

Ley de Benford

TESTS



CHI

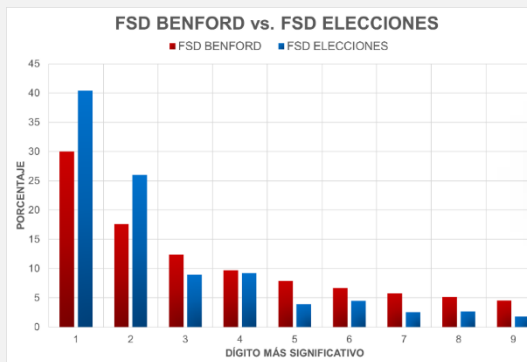
No

MAD

No

PEA

Si



CHI

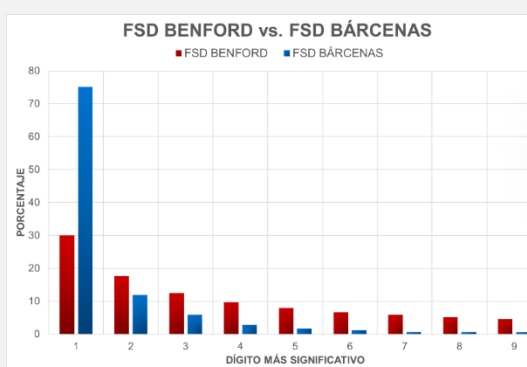
No

MAD

No

PEA

Si



CHI

No

MAD

No

PEA

Si

4

Conclusiones

4 Conclusiones

Ley de Zipf

HISPATWEETS

ELECCIONES

BÁRCENAS

- No presenta anomalías en la distribución de #
- Presenta anomalías en la distribución de # (Acumulación de hashtags con poca repercusión)

Ley de Benford

HISPATWEETS

ELECCIONES

BÁRCENAS

- Los tests estadísticos se han comportado de acuerdo a estudios previos
- Las distribuciones no se ajustan porque presentan anomalías en el comportamiento del dígito 1

Propuesta a futuro

- Distribución temporal
- Ajuste de una curva a los datos



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Análisis de frecuencia de hashtags en Twitter

TRABAJO FINAL DE MÁSTER

José Alberto Pérez Melián
Valencia, 26 de septiembre de 2016