

Analítica de datos en Twitter

Proyecto Fin de Carrera realizado por:
Dani Mir Montserrat
Director:
Jose López Vicario

Febrero 2015

Índice

1. Introducción y objetivos
2. Desarrollo
3. Resultados experimentales
4. Conclusiones y trabajo futuro

Índice

1. Introducción y objetivos
2. Desarrollo
3. Resultados experimentales
4. Conclusiones y trabajo futuro

1. Introducción y objetivos

◆ Las redes sociales

- Facebook, Twitter, Google+, LinkedIn,...
- Canal comunicación → Fidelizar clientes
- Ing. Telecom. → Comp. mercados
 - ◆ Comunicación eficiente
 - ◆ Mejora continua



1. Introduucción y objetivos

◆ Objetivos iniciales

- @adidas_ES VS @Nike_Spain
- Twitter analytics → Patrones e info. interés → Com. eficiente
 - ◆ Técnicas ML
 - ◆ Métodos NLP
 - ◆ Python



*Campos de la
Ing. Telecomunicaciones*

Índice

1. Introducción y objetivos
2. **Desarrollo**
3. Resultados experimentales
4. Conclusiones y trabajo futuro

2. Desarrollo

◆ Herramientas

- Python

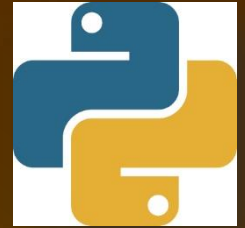
- ◆ Spider
- ◆ IPython
- ◆ Twython → API REST

- ML

- ◆ Unsupervised Learning → K-means

- NLP

- ◆ Bag of Words → tf-idf



2. Desarrollo

◆ Propuesta a resolver

1. Descartar seguidores no deseados

- ◆ fakes
- ◆ poca actividad
- ◆ baja influencia
- ◆ otros mercados

2. Localizar seguidores más influyentes

2. Desarrollo

◆ Propuesta a resolver

3. Conocer a fondo los seguidores más influyentes

- ◆ Volumen en la cuenta y contribución
- ◆ Relación con la marca → Premium, basic,...

4. Temas más comunes comentados

- ◆ Temas de interés general
- ◆ Relaciones entre temas y seguidores

2. Desarrollo

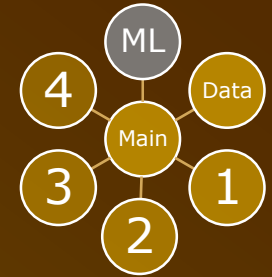
◆ Implementación

1. Filtro de followers
2. Detección celebrities
3. Análisis celebrities
4. Topics

◆ Estructura

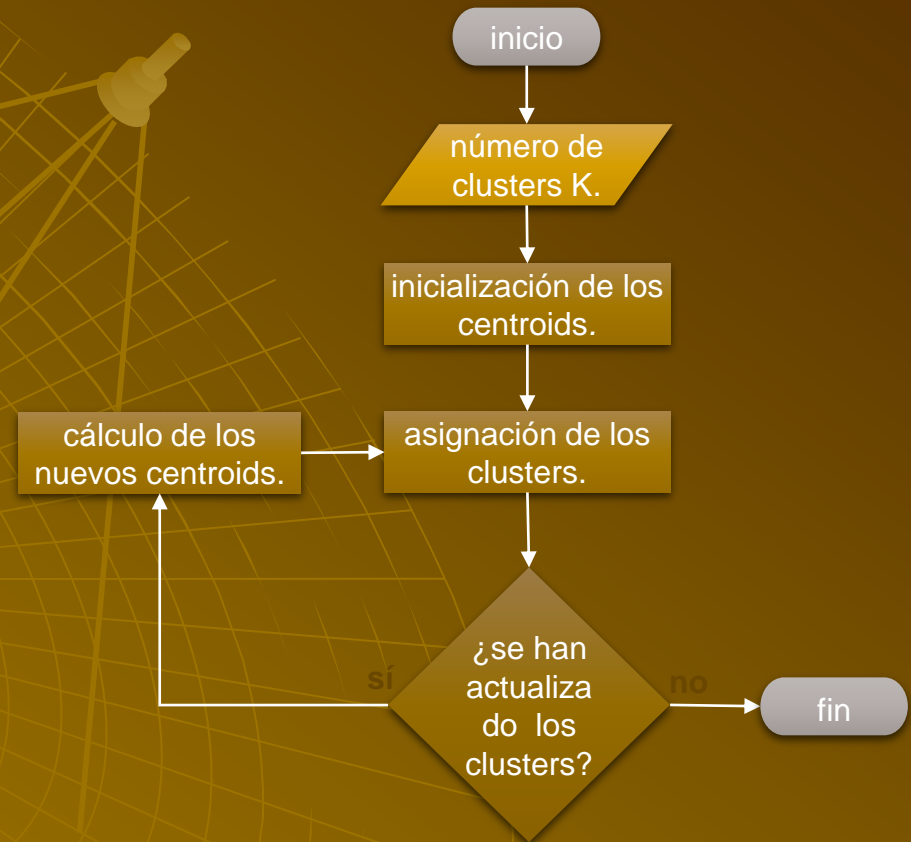
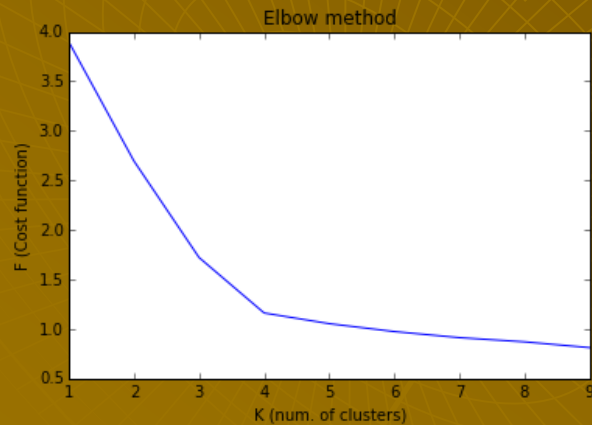


2. Desarrollo



◆ Algoritmos ML

- PCA
- K-means
 - ◆ Elbow



2. Desarrollo



◆ Obtención de los datos

	Núm. friends	Núm. followers
adidas	188	728504
Nike	203	56289

Título <i>método</i>	Pet. 15-min	Usuarios / pet.	T. espera / pet.	T. adidas	T. Nike
GET friends/ids <i>get_friends_ids</i>	15	5000	60s	2.42h	0.18h
GET followers/ids <i>get_followers_ids</i>	15	5000	60s	2.42h	0.18h
GET users/lookup <i>lookup_user</i>	60	100	15s	30.35h	2.34h
GET statuses/user_timeline <i>get_user_timeline</i>	300	1	3s	606.96h	46.8h
Total marca:				642.15h	49.5h
TOTAL:				28.81 días	

2. Desarrollo



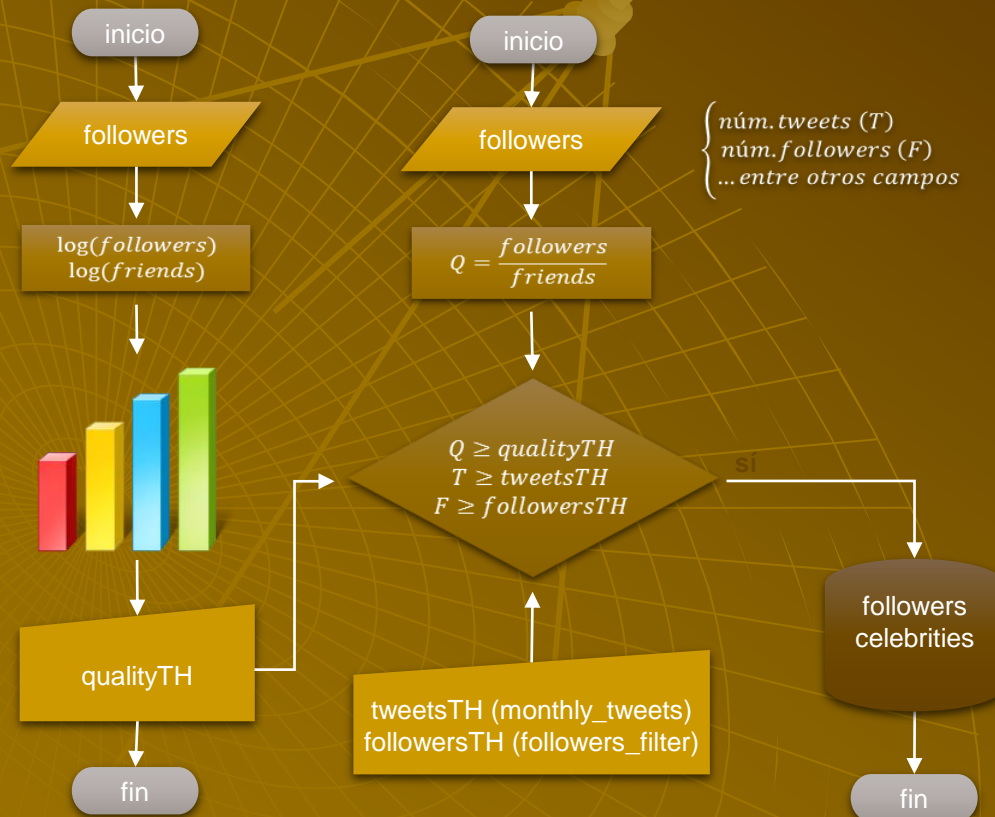
1. Filtro followers

- Núm. de followers mínimo
- Media de tuits por mes
- Idioma

2. Desarrollo



2. Detección celebrities



2. Desarrollo



3. Análisis celebrities

- Densidad, actividad, seguidores,...
- Historial menciones
- Top10 celebrities & Potential friends
- Segmentación k-means (Elbow)

2. Desarrollo



4. Topics

- tf-idf



$$W_{t,d} = tf_{t,d} \cdot \log \frac{N}{df_t}$$



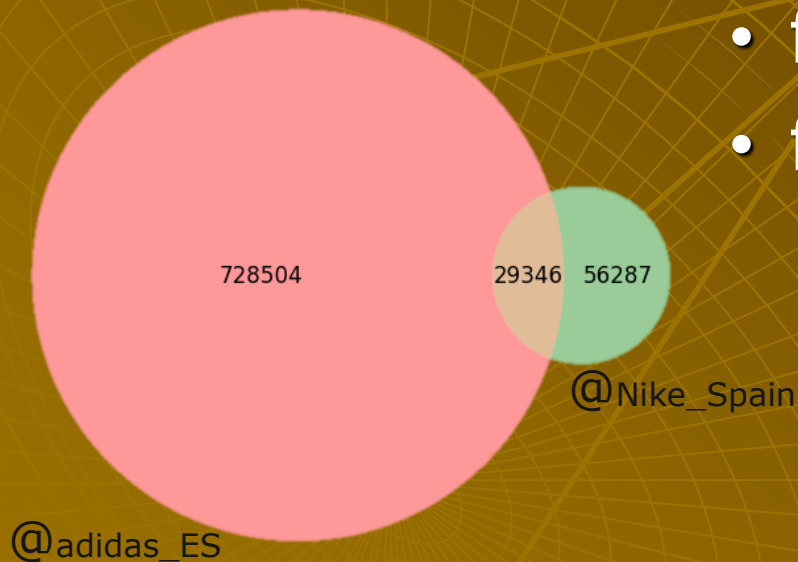
Índice

1. Introducción y objetivos
2. Desarrollo
3. Resultados experimentales
4. Conclusiones y trabajo futuro

3. Resultados experimentales

◆ Exp. 1: Filtro de followers

- flw. adidas >> flw. Nike
- flw. irrelevantes^{↑↑}

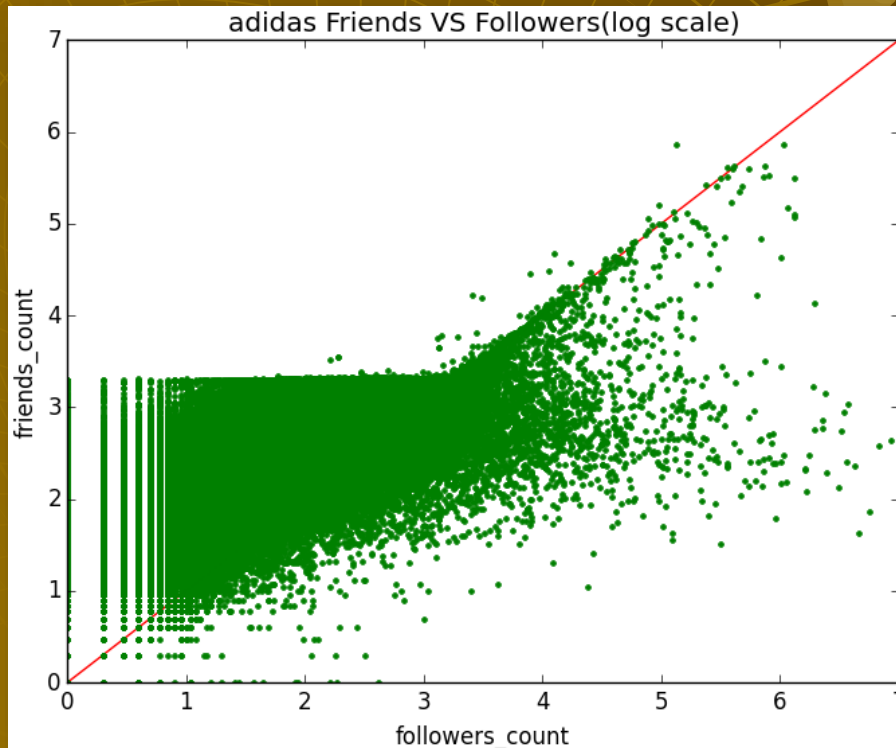


followers_filter=10
monthly_tweets=4
lang_filter=es

Brand	Flw.	Flw. filter	Flw. Filter %	P. reach
adidas	727223	182639	25,11	245 M
Nike	56283	23832	42,34	71 M

3. Resultados experimentales

◆ Exp. 2: Detección de celebrities



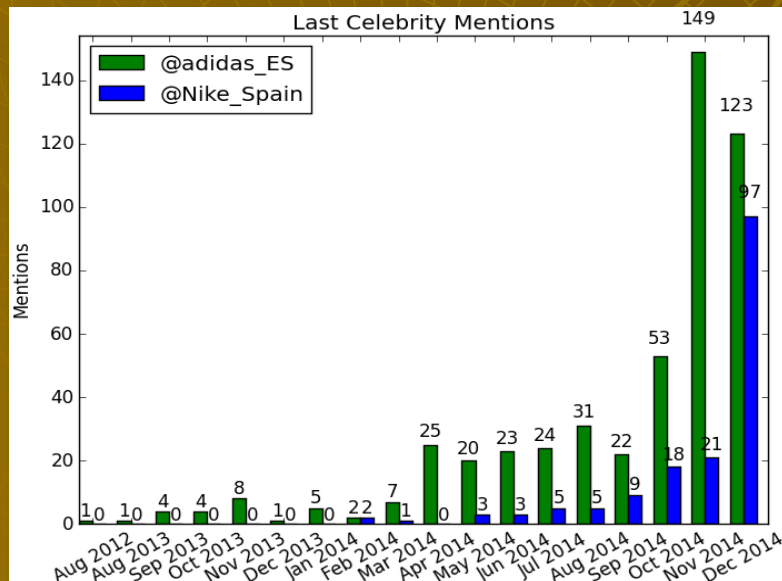
- Celebrities

- ◆ followers_count^{↑↑}
- ◆ friends_count^{↓↓}

3. Resultados experimentales

◆ Exp. 3: Análisis de celebrities

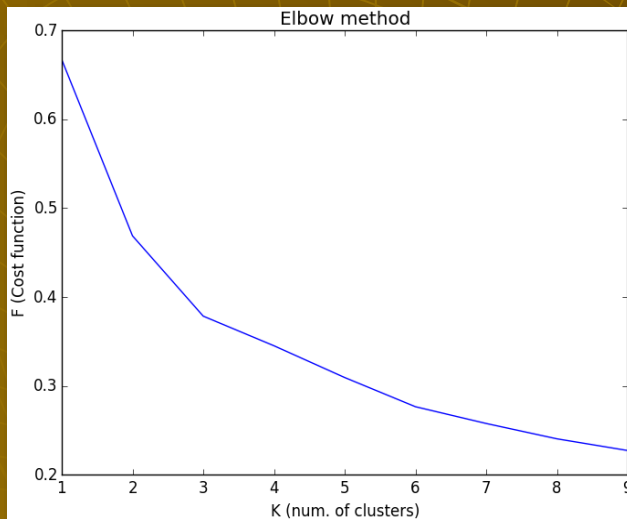
Brand	Cel. (%)	Av. activity	Av. Flw.	Ment. (Av.)
adidas	650 (0.35)	13402	125894	503 (0.77)
Nike	260 (1.09)	11787	114301	164 (0.64)



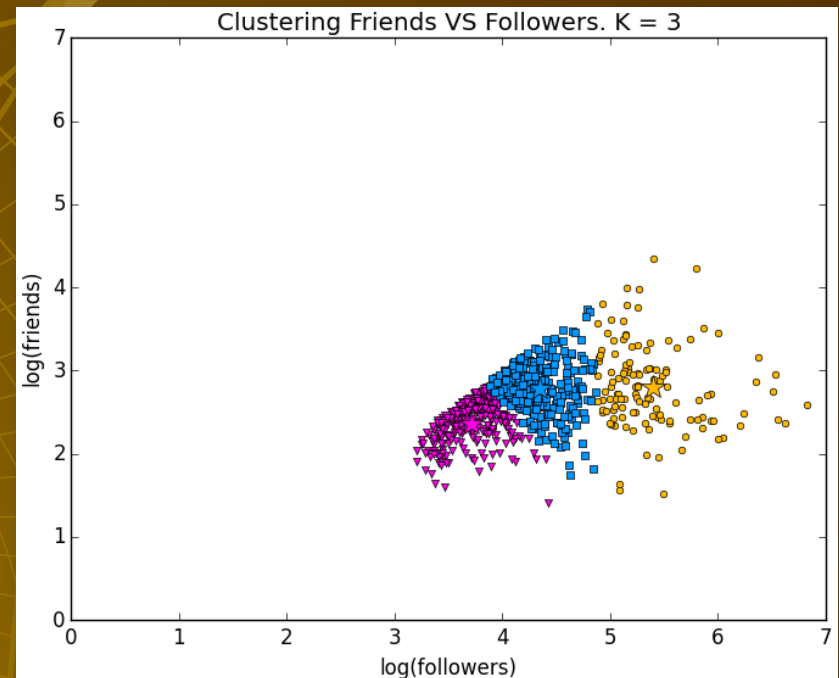
- número cel. adidas^{↑↑}
- relevancia cel. adidas^{↑↑}
- menciones adidas^{↑↑}

3. Resultados experimentales

◆ Exp. 3: Análisis de celebrities

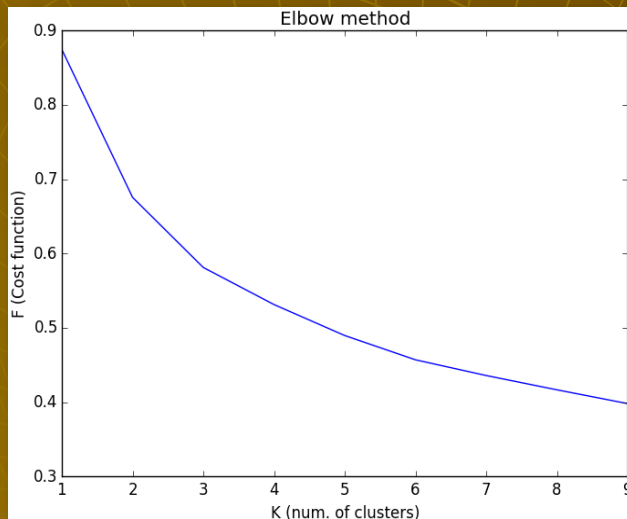


- Ej. seg. 3 grupos
 - ◆ embajadores (producto)
 - ◆ famosos (sueldo)
 - ◆ mediáticos (sueldo más elevado)



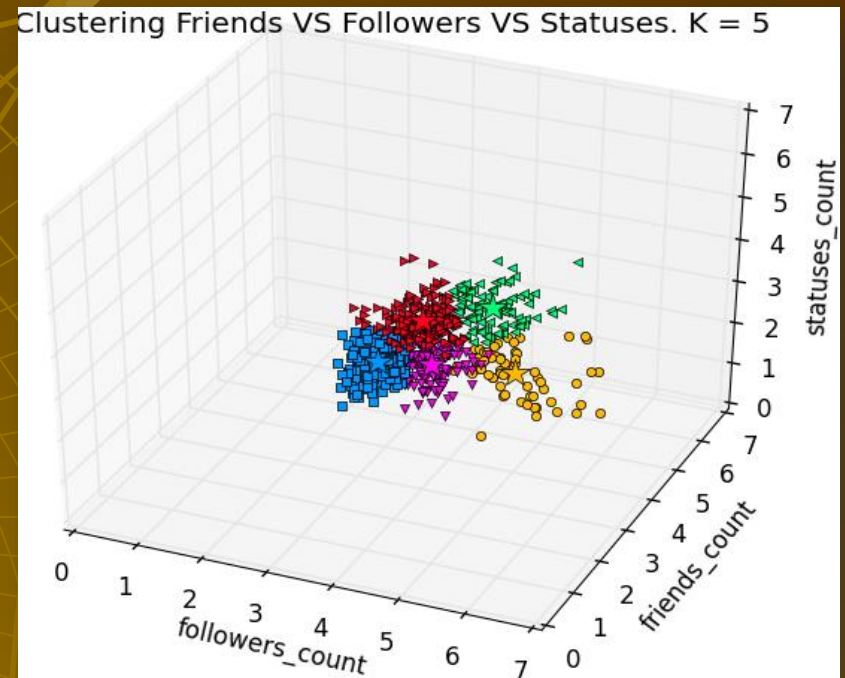
3. Resultados experimentales

◆ Exp. 3: Análisis de celebrities



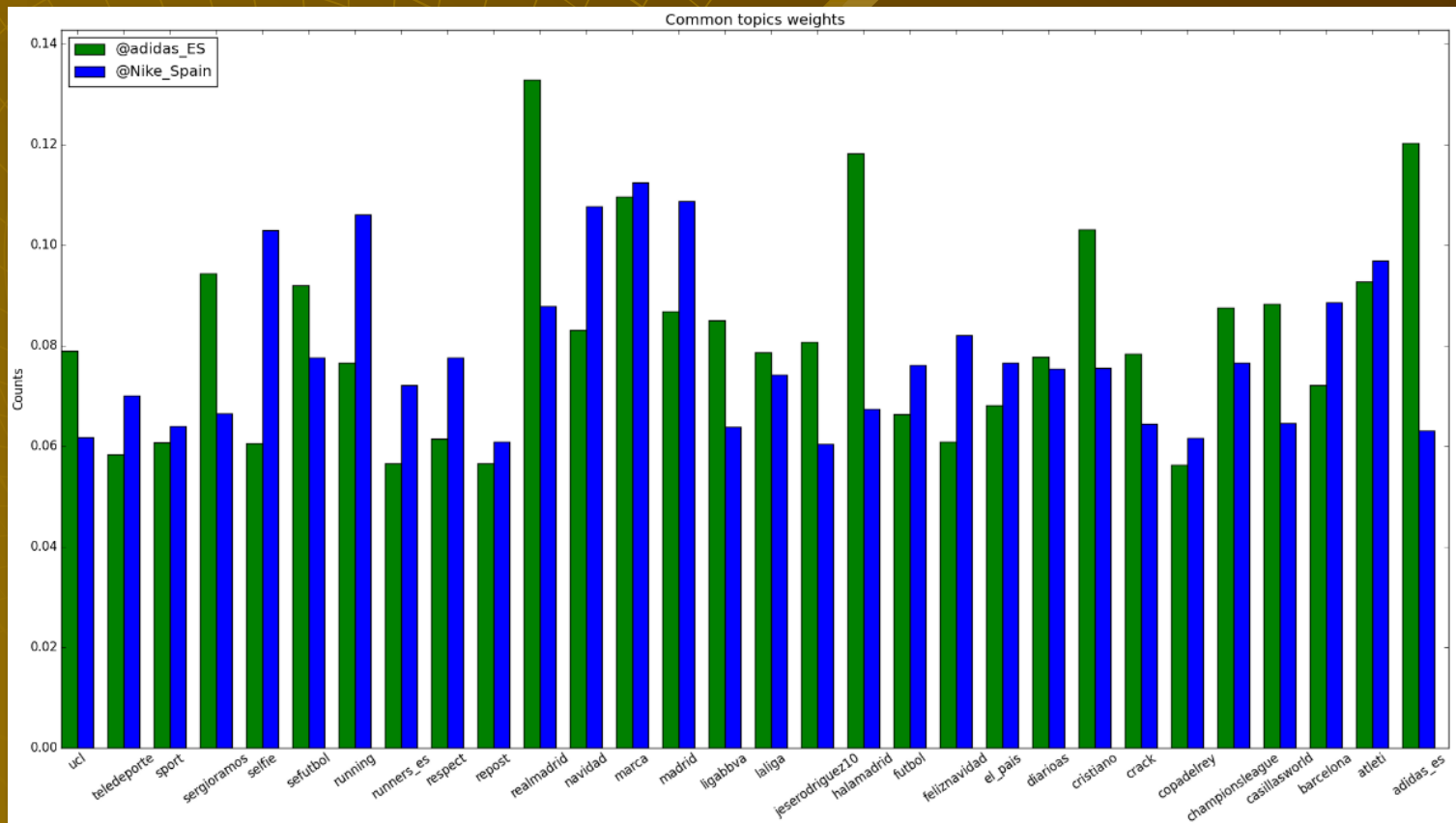
- Ej. seg. 4 grupos

- ◆ conocidos (producto esporádicamente)
- ◆ embajadores (reciben producto)
- ◆ famosos (sueldo)
- ◆ mediáticos (sueldo y campañas)

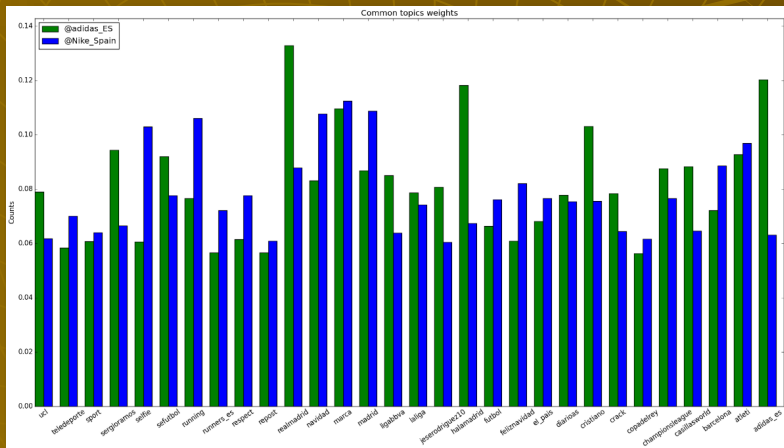


3. Resultados experimentales

◆ Exp. 4: Topics



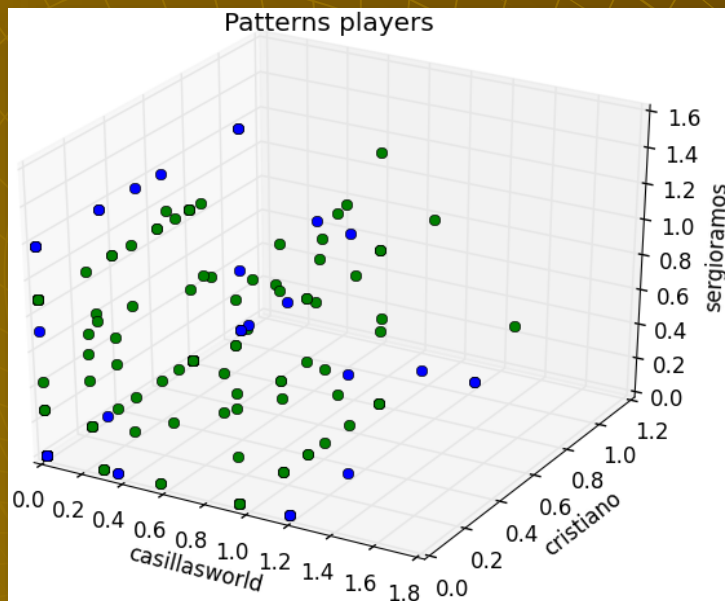
◆ Exp. 4: Topics



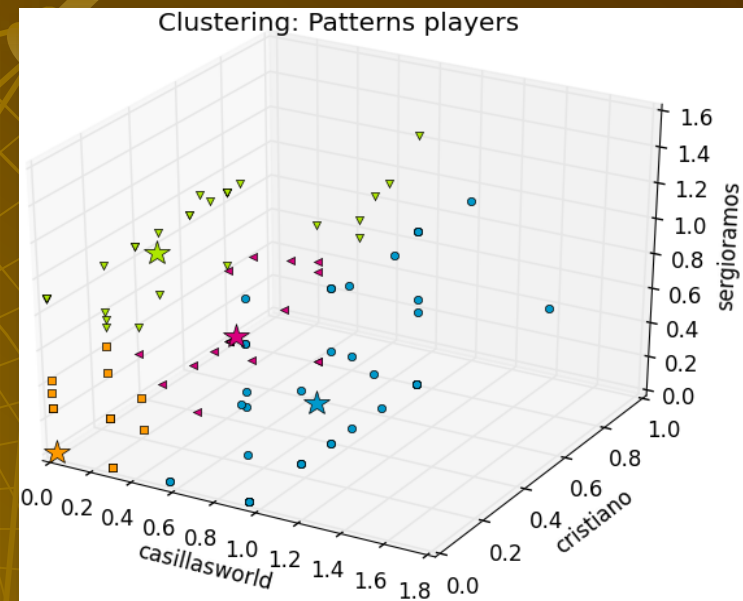
- adidas
 - ◆ sergioramos
 - ◆ cristiano
 - ◆ realmadrid
 - ◆ halamadrid
 - ◆ adidas_ES
- Nike
 - ◆ running
 - ◆ selfie
 - ◆ navidad
 - ◆ madrid

3. Resultados experimentales

◆ Exp. 4: Topics



- Relación entre topics y celebrities



- Segmentación celebrities según topics

Índice

1. Introducción y objetivos
2. Desarrollo
3. Resultados experimentales
4. Conclusiones y trabajo futuro

4. Conclusiones y trabajo futuro

◆ Objetivos

- Analytics en social media
- Patrones e información de interés
- Experimentado: ML, NLP, Python

4. Conclusiones y trabajo futuro

◆ Problemas

- Gran número de followers
- Limitaciones
 - ◆ Peticiones de API
 - ◆ Proceso open-sabe-close

4. Conclusiones y trabajo futuro

◆ Trabajo futuro

- Big Data
- Crear una aplicación
- Nuevos datos
 - ◆ Geolocalizar tuits y usuarios
 - ◆ API de Streaming
 - ◆ Análisis de sentimiento

Gracias por su atención

¿Preguntas?