

# **EXPLORANDO NUESTRA DATA**

**(A VER QUÉ HAY)**



# **CARACTERÍSTICAS DE LA DATA DE TW**

**(CARACTERÍSTICAS SIMILARES TENDRÁN SUS  
PROPIAS REDES DE TW)**

# ¿PRODUCCIÓN O REPRODUCCIÓN?

- ¿Qué hay más en la redes: producción de tuits o la reproducción de tuit?
- Veamos...

```
> table(is.na(duque_tw$retweet_name)) # Muchos RTs... esto es lo que esperamos.  
FALSE    TRUE  
252482  51882
```

- En nuestra data, 79.5% de los tuits son RTs. Esto es lo que esperamos: muchos más RTs que tuits. → (“scale-free world”)

# ¿A QUIÉNES RETUITEAMOS?

- Podemos ver quiénes son los usuarios más retuiteados (y luego compararlos con lo que más retuitean).
- Vamos a llamar a los creadores de tuits = **autoridades**; y vamos a llamar a los que retuitean un tuit = **hubs**. En la siguiente clase veremos la utilidad de pensarla de esta manera.
- *Nota: Sí, muchos usuarios son tanto autoridades como hubs. Pero en el trascurso de las clases veremos que suelen ser grupos con características distintas.*

```

> top_rt_auto <- duque_tw %>%
+   filter(!is.na(retweet_screen_name)) %>% # Eliminamos los que no fueron RTs
+   group_by(retweet_screen_name) %>% # Agrupamos por el nombre de la autoridad
+   mutate(frec_rt_name = n()) %>% # Calculamos la frecuencia
+   distinct(retweet_screen_name, retweet_name, frec_rt_name) # Mantenemos solo la info necesaria
a
>
> top_rt_auto <- top_rt_auto[order(top_rt_auto$frec_rt_name, decreasing=T), ] # Ordenamos
> names(top_rt_auto) <- c("Usario", "Nombre", "Tweets") # Damos nombres
> head(top_rt_auto, 10) # Vemos los más RT (y cuántas veces)
# A tibble: 10 x 3
# Groups:   Usario [10]
  Usario        Nombre      Tweets
  <chr>       <chr>       <int>
1 HELIODOPTERO Gonzalo Guillén     11871
2 FisicoImpuro  Físico Impuro      10264
3 petrogustavo Gustavo Petro      8681
4 Tolaymaruja Tola y Maruja      7775
5 tobonsanin    Gilberto Tobón Sanín 6241
6 IvanCepedaCast Iván Cepeda Castro 5599
7 GustavoBolivar Gustavo Bolívar    5304
8 carlosecaicedo Carlos Caicedo     4543
9 derlilopeza   Derli López        4114
10 piedadcordoba Piedad Córdoba  ⚡     4103

```

# ¿QUIÉNES RETUITEAN?

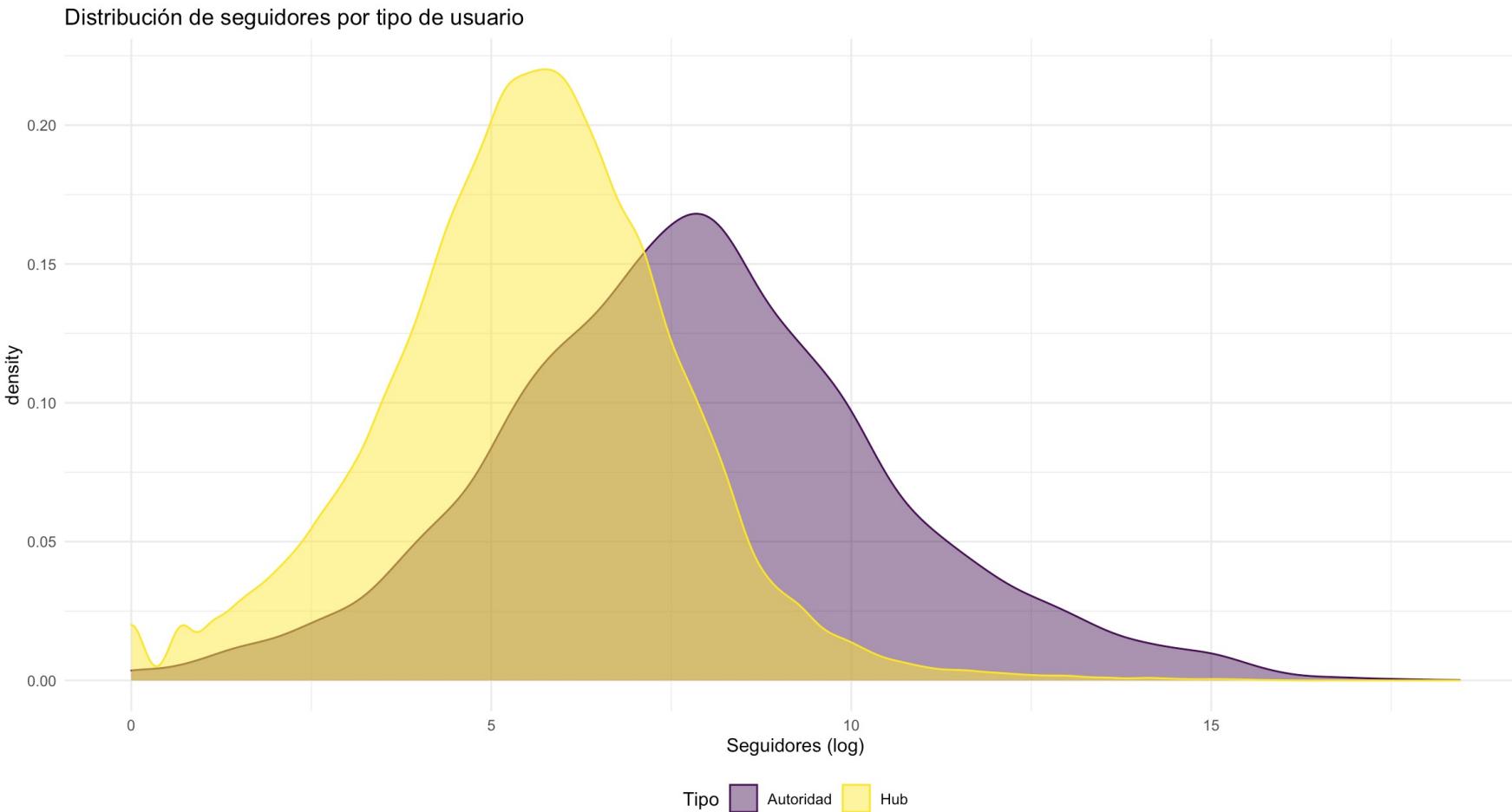
```
> top_rt_hub <- duque_tw %>%
+   group_by(screen_name) %>% # Agrupamos por el nombre del hub
+   mutate(frec_rt_name = n()) %>% # Calculamos la frecuencia
+   distinct(screen_name,frec_rt_name) # Mantenemos solo la info necesaria
>
> top_rt_hub <- top_rt_hub[order(top_rt_hub$frec_rt_name, decreasing=T), ] # Ordenamos
> names(top_rt_hub) <- c("Usuario","Tweets") # Damos nombres
> head(top_rt_hub, 10) # Vemos los más RT (y cuántas veces)
# A tibble: 10 x 2
# Groups:   Usuario [10]
  Usuario      Tweets
  <chr>        <int>
1 Duque_Francio    1358
2 duque_chamberi     918
3 Winchester_Musk    510
4 auribevoz        356
5 Arabelamay_0324    329
6 _rayen_duque       303
7 tyronramos_17       283
8 eltheyahliezeth     281
9 LuzStel17338771     270
10 Cazaratas11        270
```

- No son los mismo... y ahora vamos a ver si sus características son diferente.

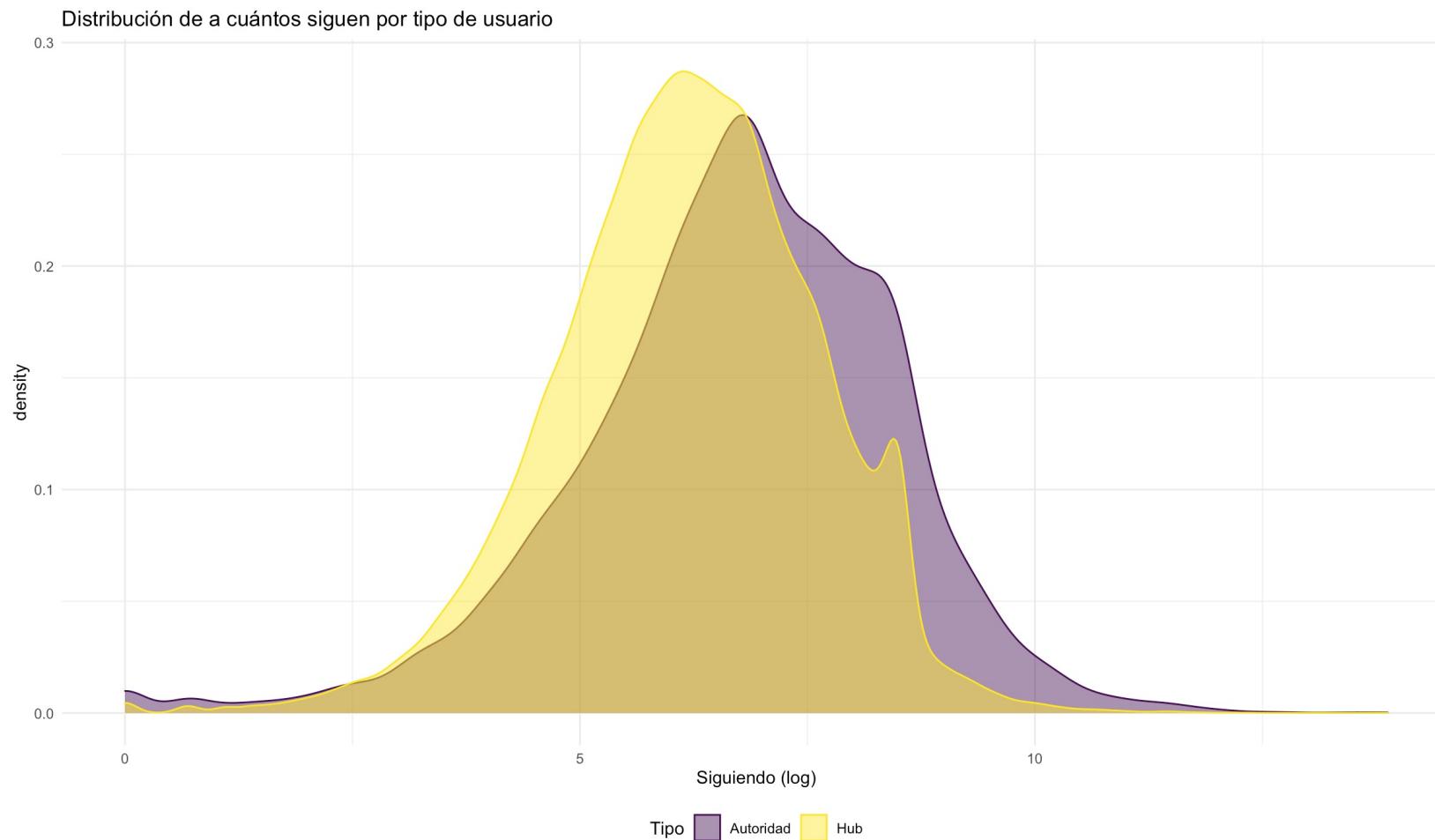
# CARACTERÍSTICAS DE *AUTORIDADES* Y *HUBS*

```
> table(duque_tw$retweet_verified[!is.na(duque_tw$retweet_screen_name)]) # autoridades  
  
 FALSE    TRUE  
175972  76510  
>  
> table(duque_tw$verified[!is.na(duque_tw$retweet_screen_name)]) # hubs q dieron RT  
  
 FALSE    TRUE  
251879     603
```

# CARACTERÍSTICAS DE *AUTORIDADES* Y *HUBS*



# CARACTERÍSTICAS DE *AUTORIDADES* Y *HUBS*



- Es evidente que las autoridades no son las mismas que los hubs.
- Eso significa que la producción de tuis (información) se lleva a cabo por un tipo de usuario, y su reproducción se está llevando a cabo por otro tipo de usuario.
- Para ver mejor quiénes son, dónde están, cómo afecta la ampliación de mensajes, etc.., debemos armar la red...

**¡ARMEMOS LA RED!**

## Pregunta de Investigación

- No lo vamos a cubrir
- ¡Importantísimo!
- Nuestra ventaja comparativa

## Data

- Obtener (TW)
- Limpieza de datos
- Organizar la data
- Explorar la data

## Crear la red

- Introducción a igraph
- V(net) y E(net) en igraph

## Características de la red

- Betweenness
- Centrality
- Paths

## Detección de comunidades

- Greedy algorithms / Randomwalk
- Características de las comunidades.

## Visualizar la red

- Layout
  - Force-directed, random.
- Centro / Periferia
- Baja y alta resolución

## Análisis de la red I

- Activación de la red
- La paradoja de la amistad
- Rising stars / rising tides

## Análisis de la red II

- Disonancias y resonancias cognitivas
- Viralidad
- Análisis de texto
- Experimentos