

# Análisis de Datos en Instagram

## Análisis de tendencias personales y su relación con respecto a otras personas

Juan Esteban Ladino Mateus - Carlos Sebastian Martinez Vidal

September 7, 2020

### 1 Descripción del problema

El proyecto constituye dos problemas a resolver. El primero consiste en crear un esquema de análisis de las tendencias que sigue una persona para así obtener información implícita acerca del personaje en cuestión. El segundo problema consiste en almacenar toda la información(explicita e implícita) obtenida de dos o mas personas y analizar la relación entre los individuos en cuestión. Es decir, nivel de relación, gustos semejantes, likes recíprocos, etc... En otras palabras mostrar la compatibilidad de relación entre dos o mas usuarios.

### 2 Resumen

Mediante este proyecto aspiramos a crear un perfil para cada persona que es analizada. La idea es trabajar mediante porcentajes sobre cada categoría posible para así al final poder comparar estas categorías entre dos usuarios o mas usuarios y observar su relación.

### 3 Marco Teórico

#### CONCEPTOS:

- Nodos Principales

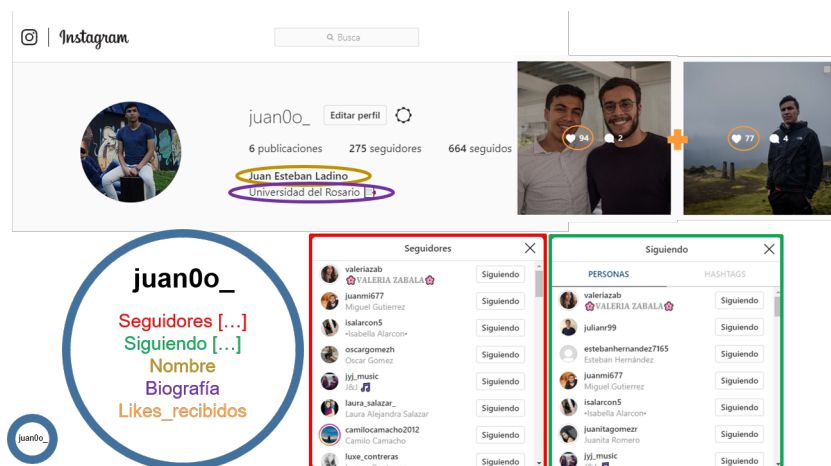


Figure 1: Caption

El **nodo principal** constituye al usuario(objetivo) para ser analizado, el nodo contiene la información publica del usuario proveída por el servido(Seguidores, seguidos, Nombre, Biografía, Likes\_recibidos).

- Nodos secundarios

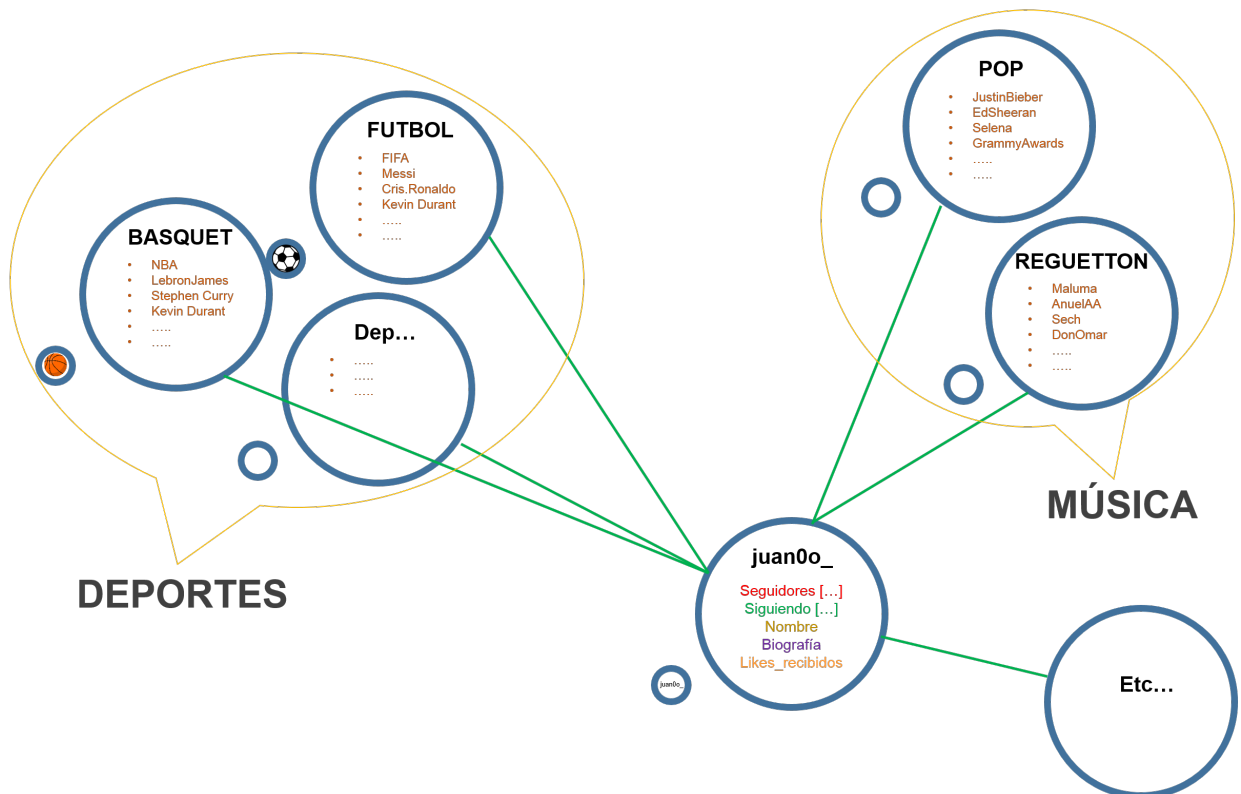


Figure 2: Caption

Los **nodos secundarios** constituyen listas divididas por temas, que contienen los usuarios mas representativos de cada temas correspondientemente.

**Ejemplos:** Deportes, Música, Cine, Entretenimiento, Arte, etc...

- Esquema de tendencia favorable

El **esquema de tendencia favorable** es la relación entre un nodo principal y los nodos secundarios, el constituye el porcentaje de tendencia favorable entre los temas de los nodos secundarios, es decir que tan afín eres acerca de los temas representados por los nodos secundarios

- **Grafos ponderados**

Cada arista de todos los grafos(Principal y secundarios) cargara con cierta información del individuo en cuestión. Esta información viajara primero de el grafo principal al secundario y viceversa, para así recopilar toda la información de vuelta en el grafo principal.

## 4 Solución del problema

### 4.1 1er problema

- Se elige a un objetivo que será el usuario elegido a analizar, luego se representa este usuario como el **nodo principal del grafo**, posteriormente se guarda en este nodo la información publica obtenida a través de Instagram, entonces se procederá a hacer el correcto análisis de la información del nodo para recibir datos implícitos(**esquema de tendencia favorable**) acerca del usuario principal obtenidos a partir de sus datos explícitos(seguidores, seguidos, likes\_recibidos, biografías).

### 4.2 2do problema

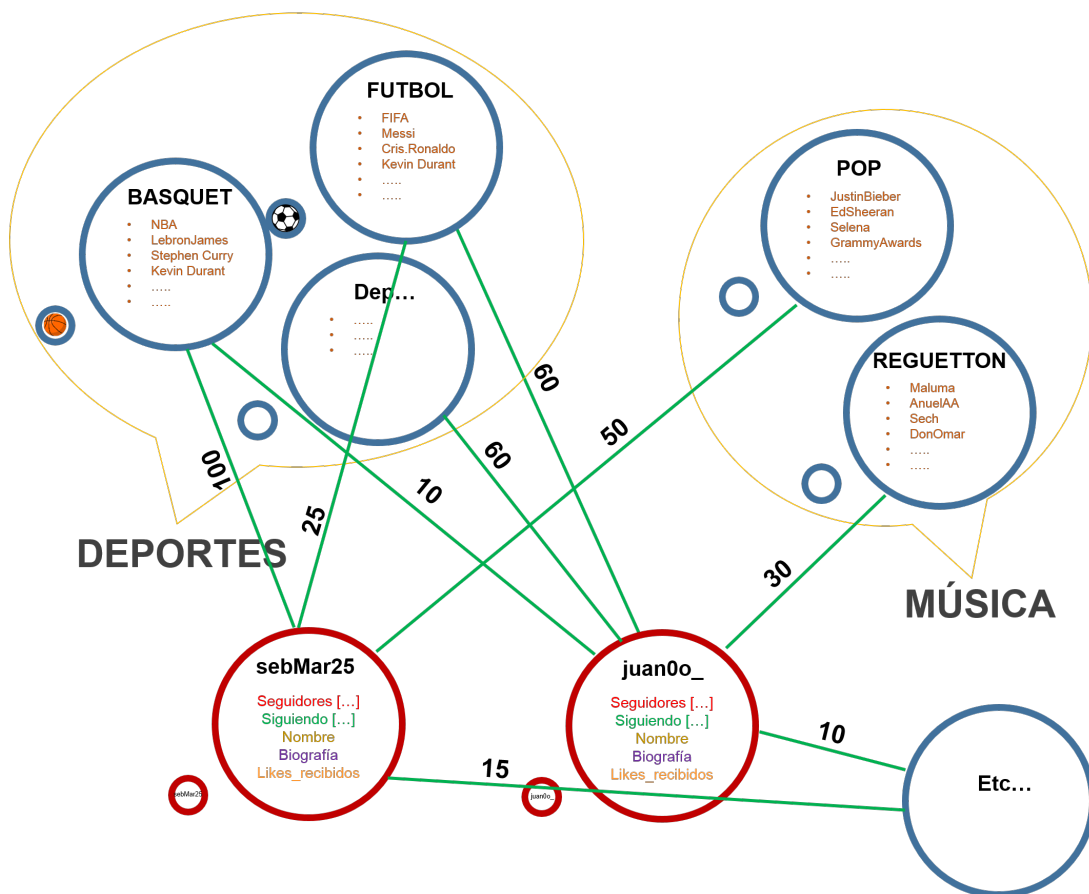


Figure 3: Caption

- Relación entre dos o mas usuarios: se selecciona el conjunto de **nodos principales** que se quiere relacionar y se ubican en el grafo, se obtiene el **esquema de tendencia favorable** + la info. Publica de cada nodo correspondiente, la relación entre los nodos se representara por medio de arista dirigidas cuya función será hacer el análisis de compatibilidad, es decir la interceptación de información entre los esquemas de tendencias favorables, mostrando un porcentaje de compatibilidad entre los nodos analizados acerca de temas de interés que compartan y su nivel de relacion.

## 5 Descripción de la implementación computacional

Se implementara la solución computacional en el lenguaje de programación Python, mas específicamente usando el Framework y librería de Selenium de Python, una herramienta de web scraping donde se conecta con el servidor de Instagram y se extrae la información publica almacenada por medio de solicitudes en HTML y se almacena la información en Python, donde se crea la clase nodo que contiene la información descrita en la solución, posteriormente se implementan las funciones de compatibilidad y se relacionan los nodos por medio de herramientas de probabilidad.

## 6 Implementación de la solución

Se creo una clase en python llamada nodo.py la cual implementa del nodo principal mencionado anteriormente. Esta contiene los métodos obtener: lista de seguidores, lista de seguidos, cantidad de likes, Descripción de la biografía, nombre de usuario por medio de una herramienta de Web scraping llamada selenium. Posteriormente se crearon las funciones de relación las cuales son : seguidores en común, seguidos en común y el esquema de tendencia favorable de cada nodo que retorna el porcentaje de compatibilidad entre las personas en un tema en específico.

## 7 Resultados obtenidos

- 15 diferentes porcentajes sobre la afinidad en cada aspecto antes mencionado
- Un perfil con caractericas unicas sobre la persona siendo analizada.
- Conclusion sobre la compatibilidad con otro usuario estudiado.

## 8 Conclusiones

Al usar diferentes personas para analizar observamos que el tiempo de trabajo del algoritmo es diferente. El numero de seguidores, seguidos, likes, etc.?Al mismo tiempo examinando a personas cercanas o incluso a nosotros mismos pudimos deducir que la realidad no es tal como en Instagram. Muchos gustos no se ven realmente reflejados en el test ya que en su Instagram no tenían ninguna relación con estos temas. Por lo tanto, el test de compatibilidad con otro usuario es bastante parcializado. Sin embargo, este es un proyecto que solo cuenta con datos recopilados en Instagram por lo cual los resultados serán muy precisos de acuerdo a sus datos.

Comparando cada lista con el usuario se determinó que una afinidad con un porcentaje de mas de 50 implicaba que era un fiel seguidor de esta categoría. Aunque parezca un porcentaje bajo, como las listas están definidas con algunos de los usuarios más representativos en Instagram -escogidos por nosotros- sobre ese género no se podía cubrir todo el territorio que implica cada gusto. Por ende, los porcentajes bajos son considerados una afinidad normal a esa clase de lista y porcentajes medianamente bajos se considera una afinidad alta.

Por ultimo comprobamos que esta herramienta funciona para conocer información implícita sobre cada persona. En unos días nos fue posible analizar datos a pequeña escala sobre cualquier individuo y crear un perfil sobre todo lo que le gusta en Instagram. Lo que lleva a pensar que la información esta disponible para todo el mundo y que la privacidad en las redes sociales no existe.