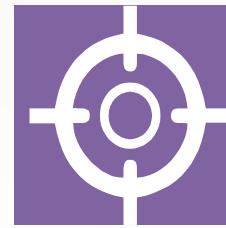
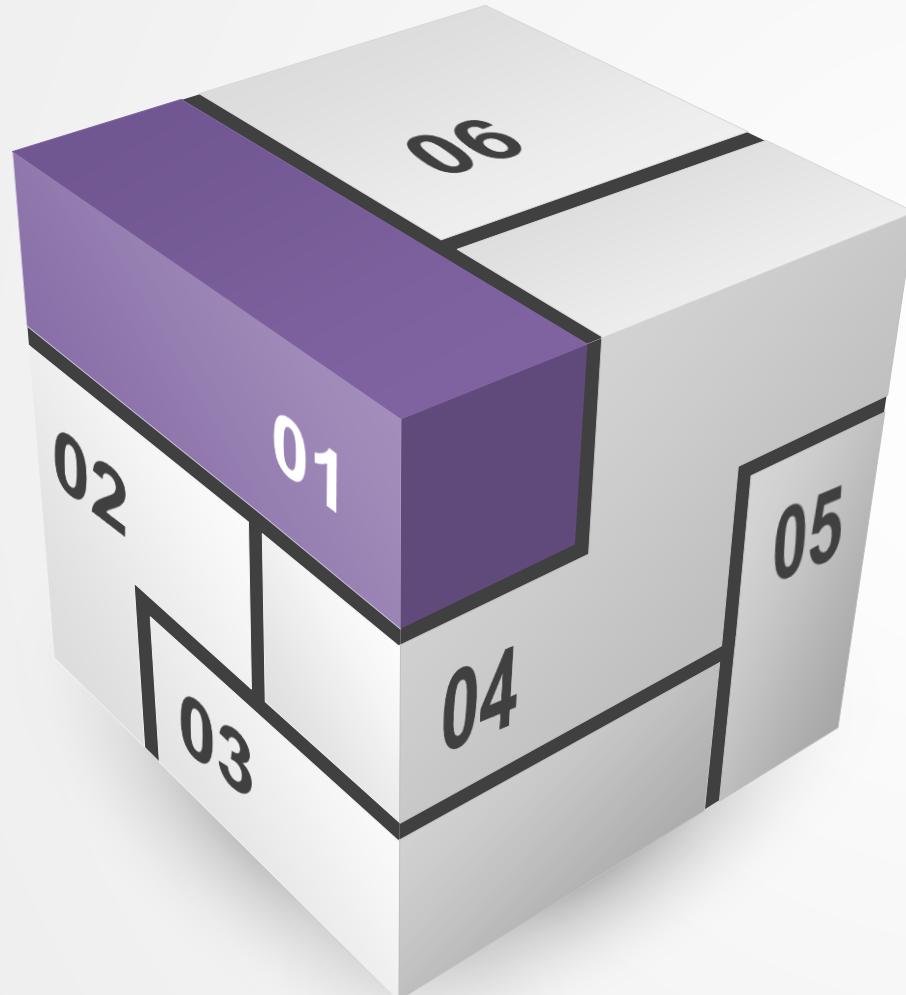


Analítica con redes sociales - Twitter

César A. Huertas Kaleda
26 de Enero de 2019

Analítica con redes sociales - Twitter



Introducción

- Presentación y acuerdos
- Expectativas
- Análitica
- Contexto
- Herramientas
- Plan de trabajo

Manos a la obra

Analítica con redes sociales - Twitter

El mundo de la analítica es muy **extenso** y **variado**. Mediante aplicaciones sencillas aprenderemos varios conceptos, su implementación y uso potencial en información obtenida de una fuente de contenidos conocida por todos como es Twitter.

Este taller ofrece un acercamiento teórico-práctico a conceptos como web scraping, análisis de redes, grafos, NLP, Análisis de sentimiento y predicción mediante el uso de información obtenida de la red social Twitter.

César Augusto Huertas Kaleda

- Ingeniero Eléctrico y MSc in Analytics
- Enel – Codensa S.A. 10 años.
- 8 años realizando estudios operativos.
- IoT Medición Inteligente – Smart Cities
- Primer puesto en el Global I&Nnovation Challenge Categoría: Social Media for Network Commercial Operations. Roma, Italia

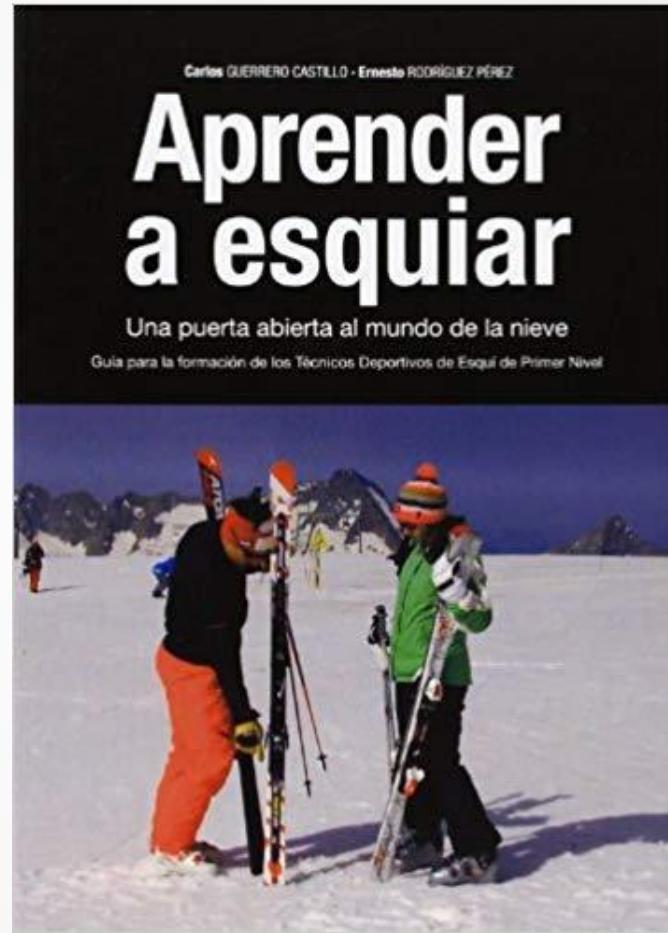
Motivación

- 3% trabaja con Big Data, un 2% con Inteligencia Artificial y un 1% utiliza la robótica, la realidad virtual y 'blockchain'.
- 'Analítica de datos' y 'seguridad digital' tienen un 95 y un 100 por ciento de empleabilidad respectivamente.. Pero no hay perfiles... dilema.
- No hay un método, y si lo hay, es de largo plazo.

Formación...



VIA 9GAG.COM



Consideré que...

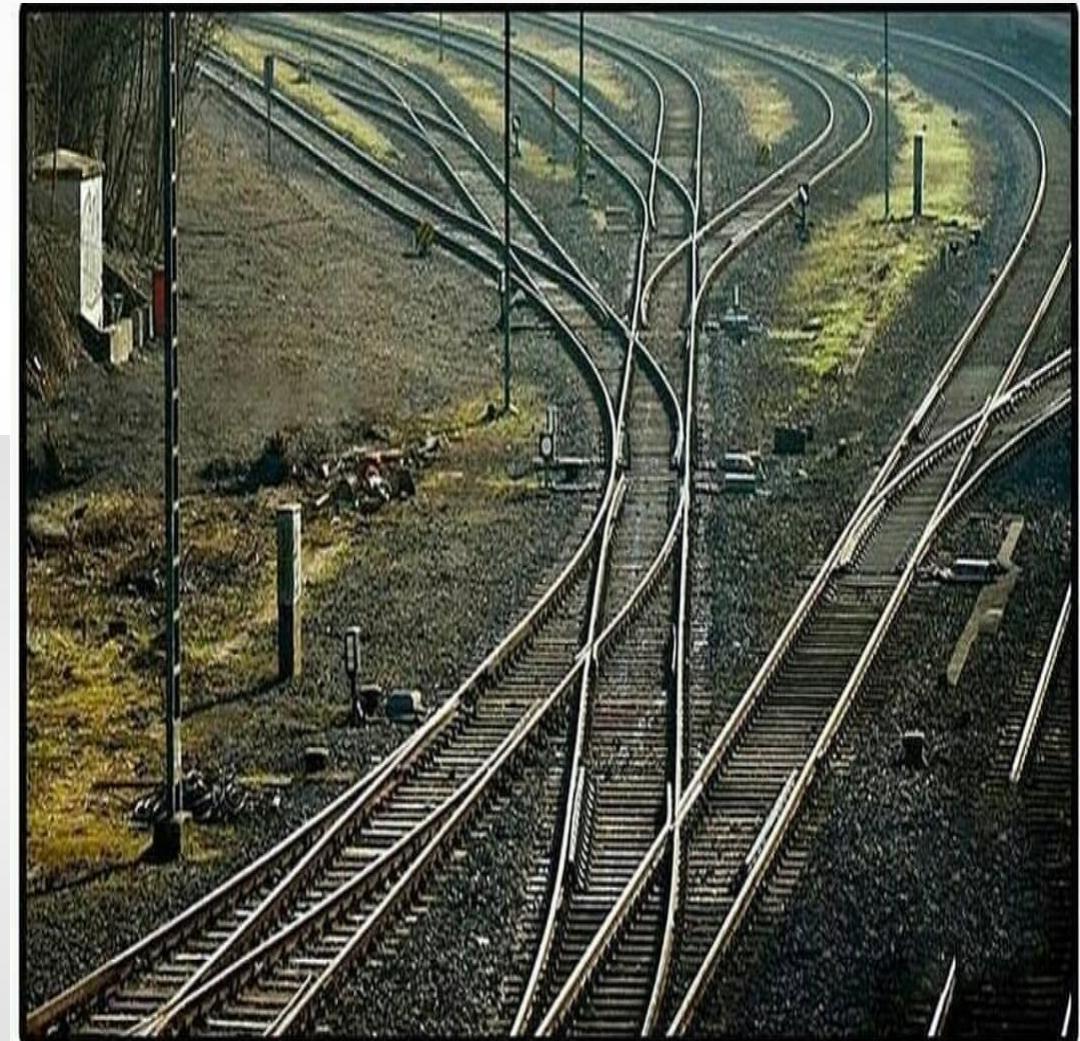
- No es un taller (curso) completo de Análisis de datos a profundidad.
 - Va a salir con más dudas(diferentes) que con las que entró... Yo también.
1. Es una construcción para entender conceptos, ejemplos de implementación rápida e identificación de usos potenciales.
 2. Se basa en las redes sociales como fuente de información para formación.
 3. Objetivo: despertar interés en usted para profundizar y avanzar.
- El conocimiento se construye entre todos
 - Hoy en día el conocimiento está en la red.
 - La gran mayoría de cosas acá presentadas, las encontrarán en internet.
 - En el análisis de datos no hay recetas, metodologías si, recetas mágicas no.

Una verdad universal:

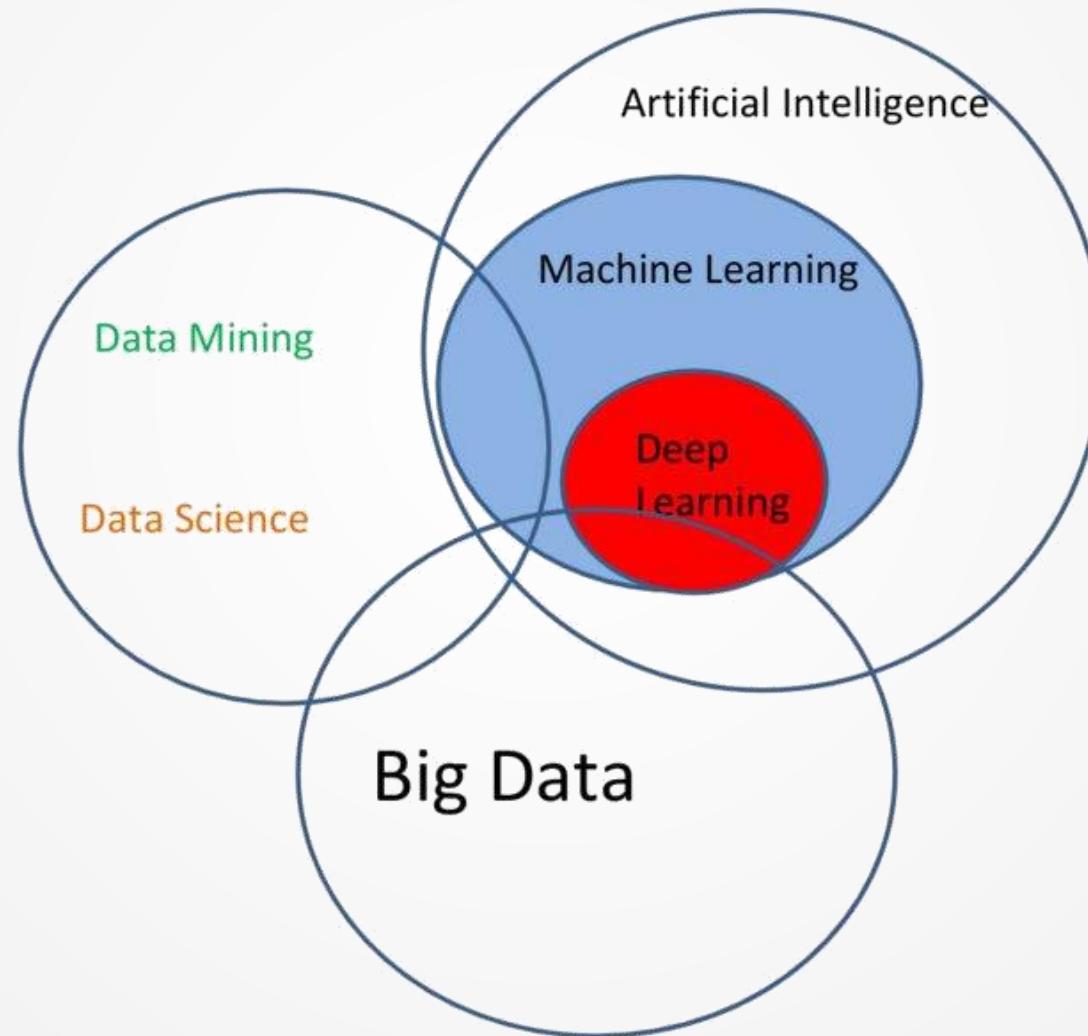
La teoría



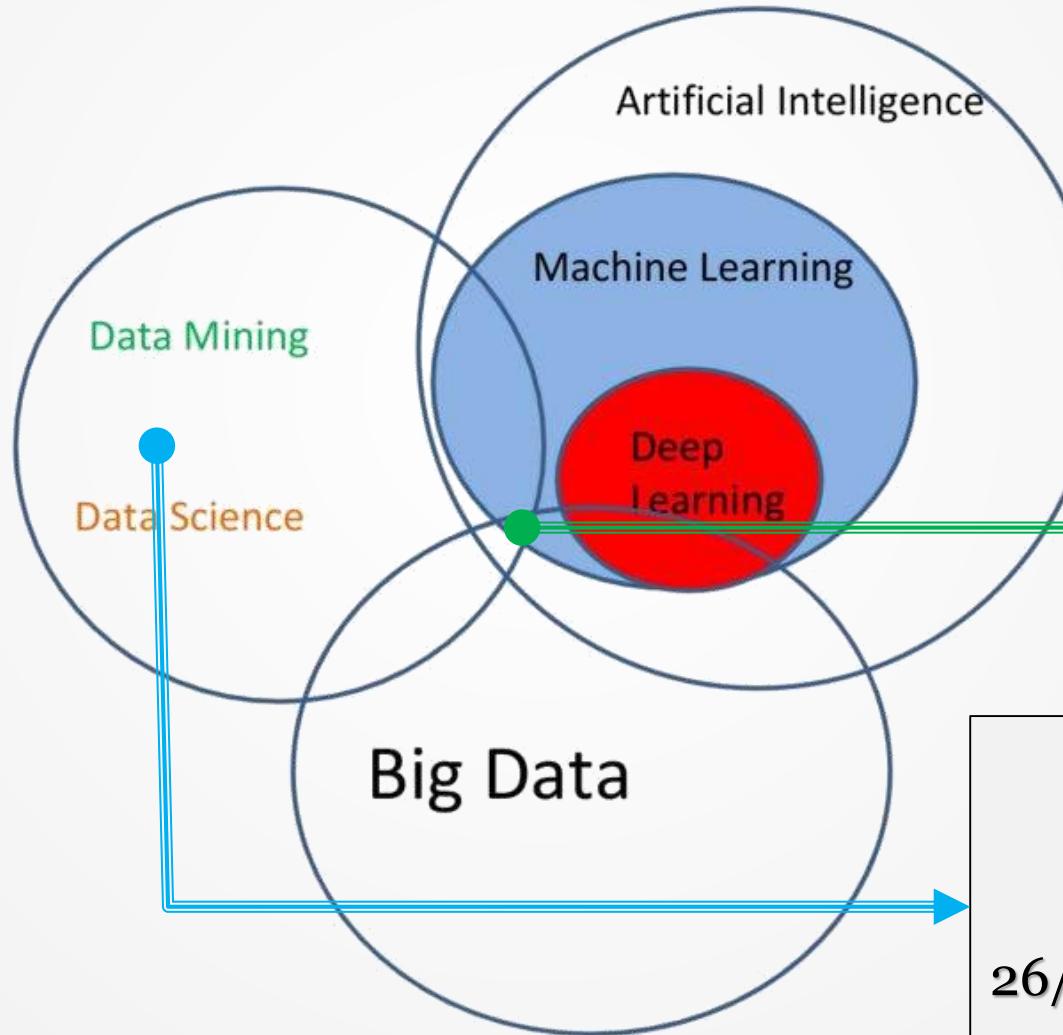
La práctica



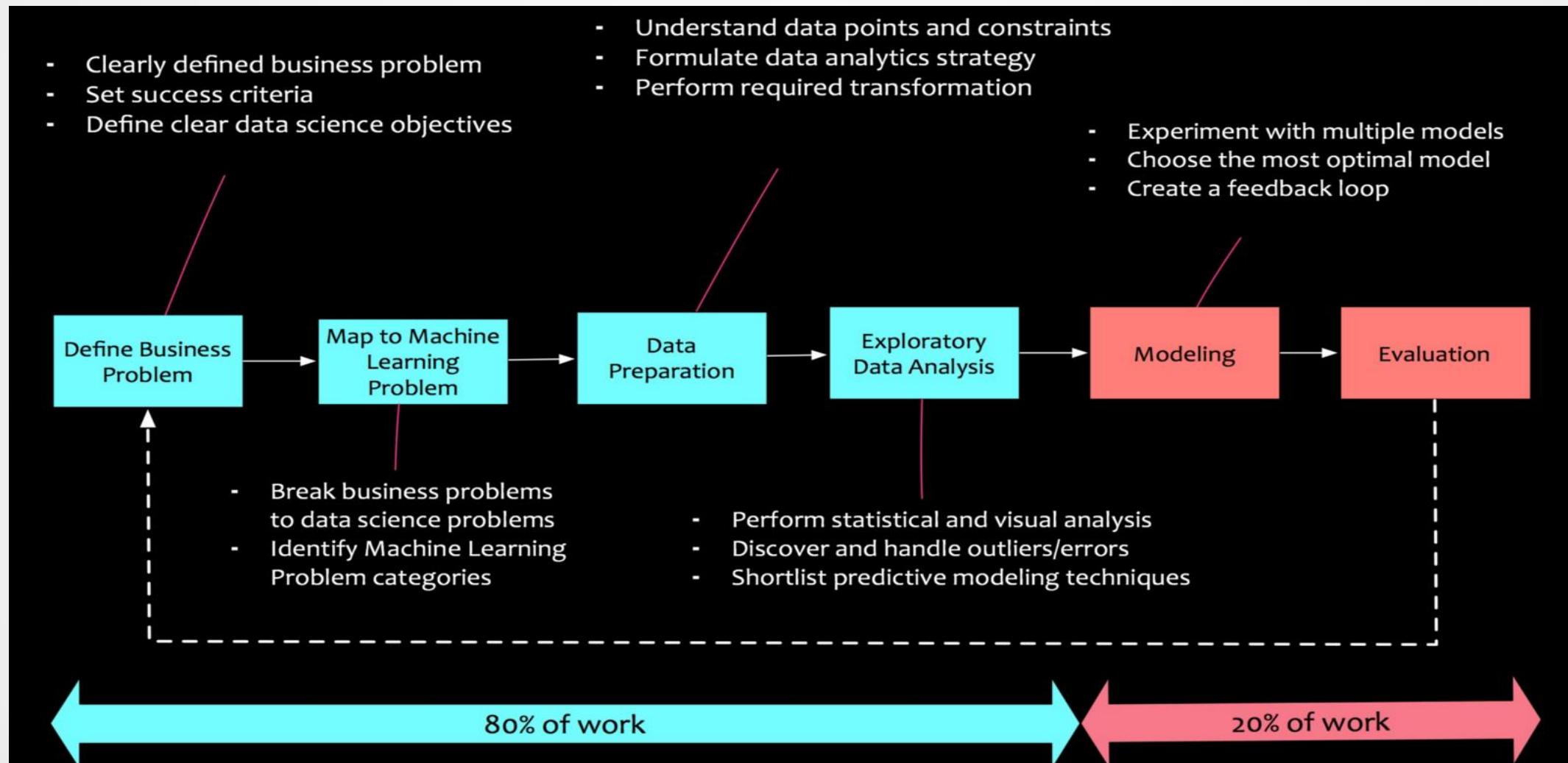
Entorno Big Data – IA – Data Science



Entorno Big Data – IA – Data Science



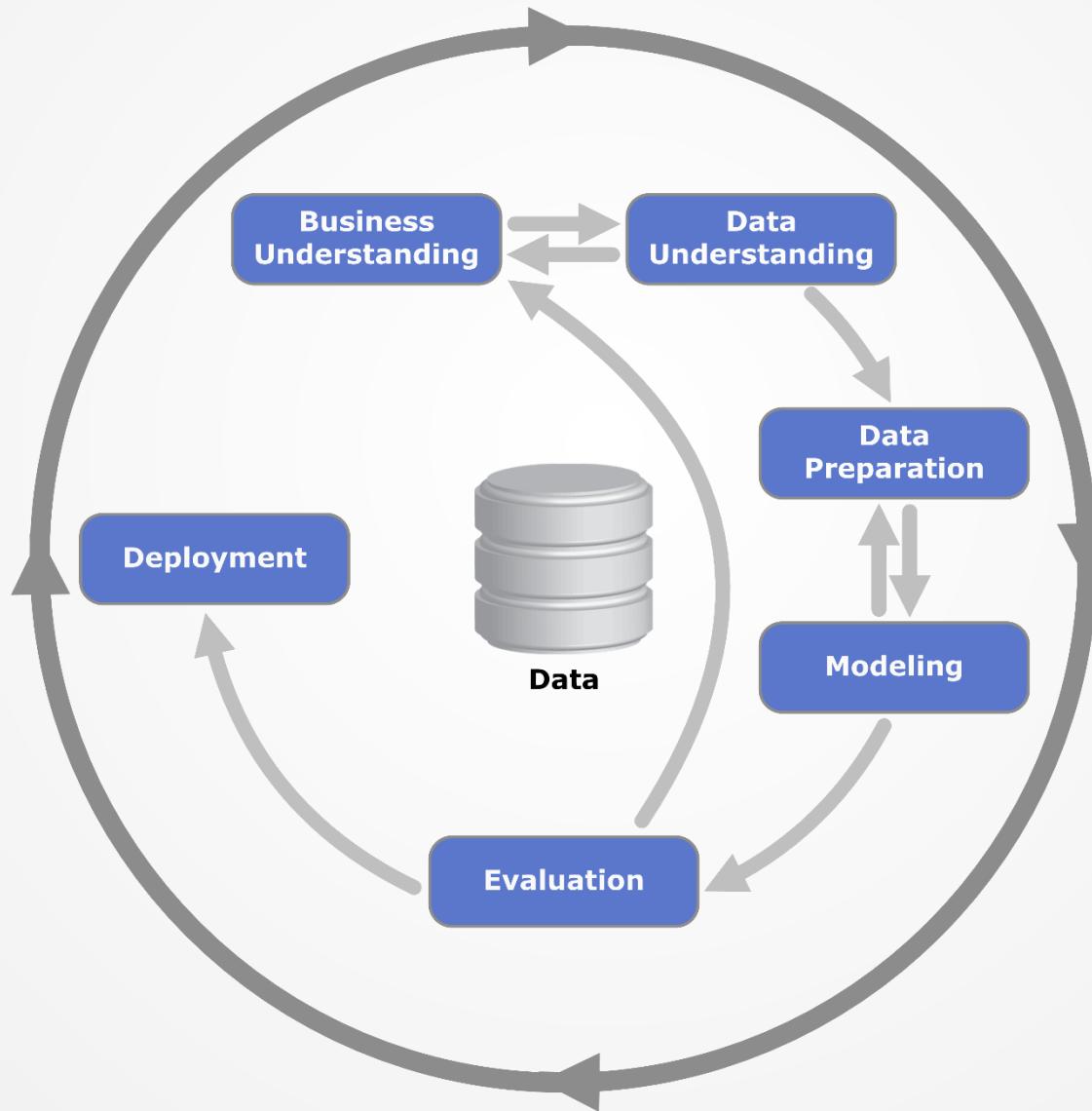
Si hay ciencia, hay método



Process of Data Science

Fuente: Data Science Simplified. Pradeep Menon

The Cross Industry Standard Process for Data Mining ([CRISP-DM](#))



¿POR QUÉ ANALÍTICA CON REDES SOCIALES?

UN JUEGO... Hagamos una lista de redes sociales conocidas, sin repetir...

Redes Sociales

- Las redes sociales no son las aplicaciones, las redes sociales son **grupos de personas** relacionadas entre sí.
- Aplicaciones = medio facilitador.
- **Su ventaja:** No hay barreras de distancia y tiempo
- El término es social media, un medio de comunicación 
- **El poder:** En la herramienta los usuarios crean una red social, se comunican y comparten información, **muuuucha información**



Redes Sociales.

Social Big Data



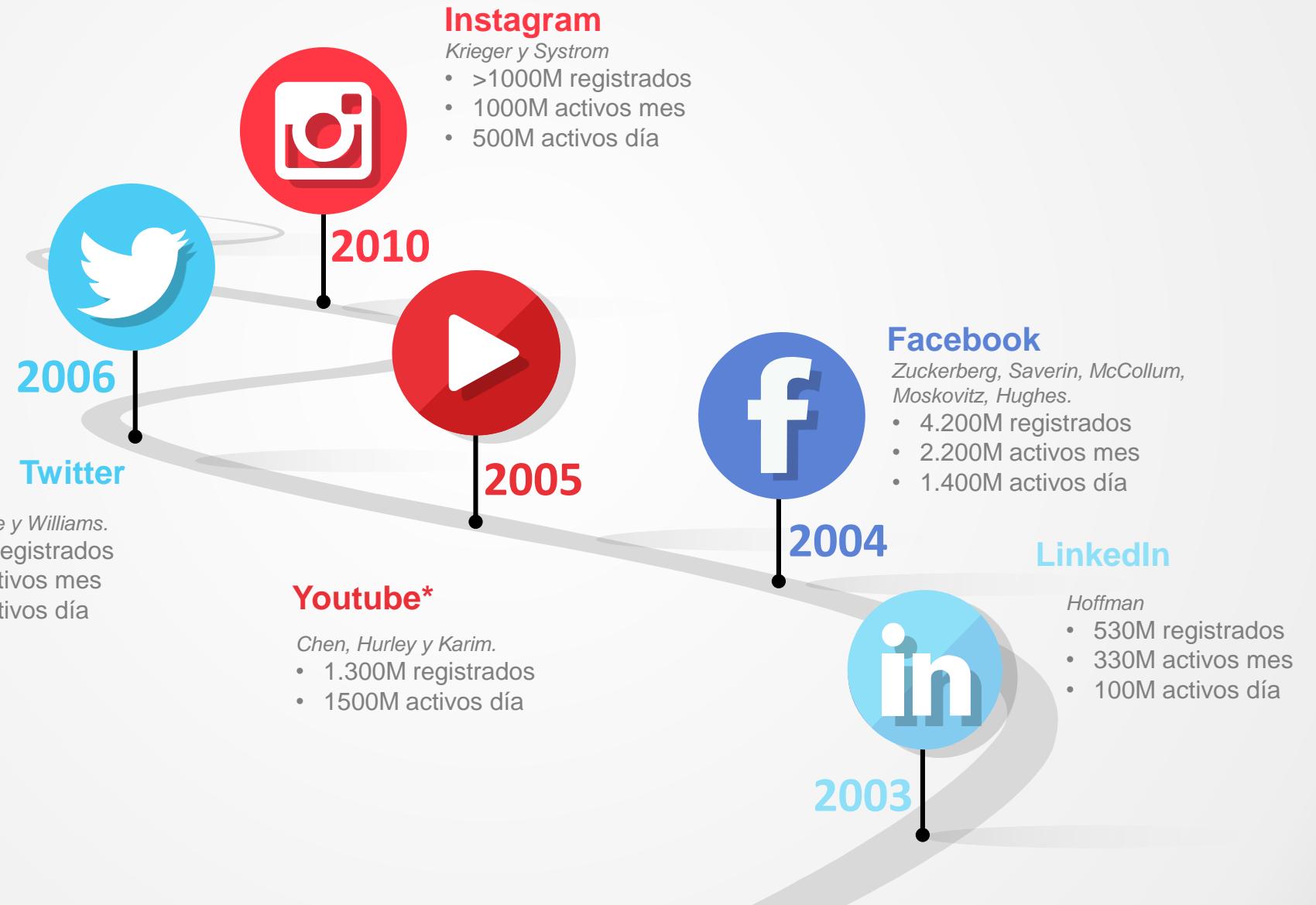
Perspectiva en tiempo real de procesos sociales, eventos actuales, procesos naturales y más

Reflejan información que los usuarios deciden compartir de forma pública.

Social Big Data

Redes Sociales

Fuente poderosa de información



2018 This Is What Happens In An Internet Minute



Redes Sociales

Gran volumen de datos

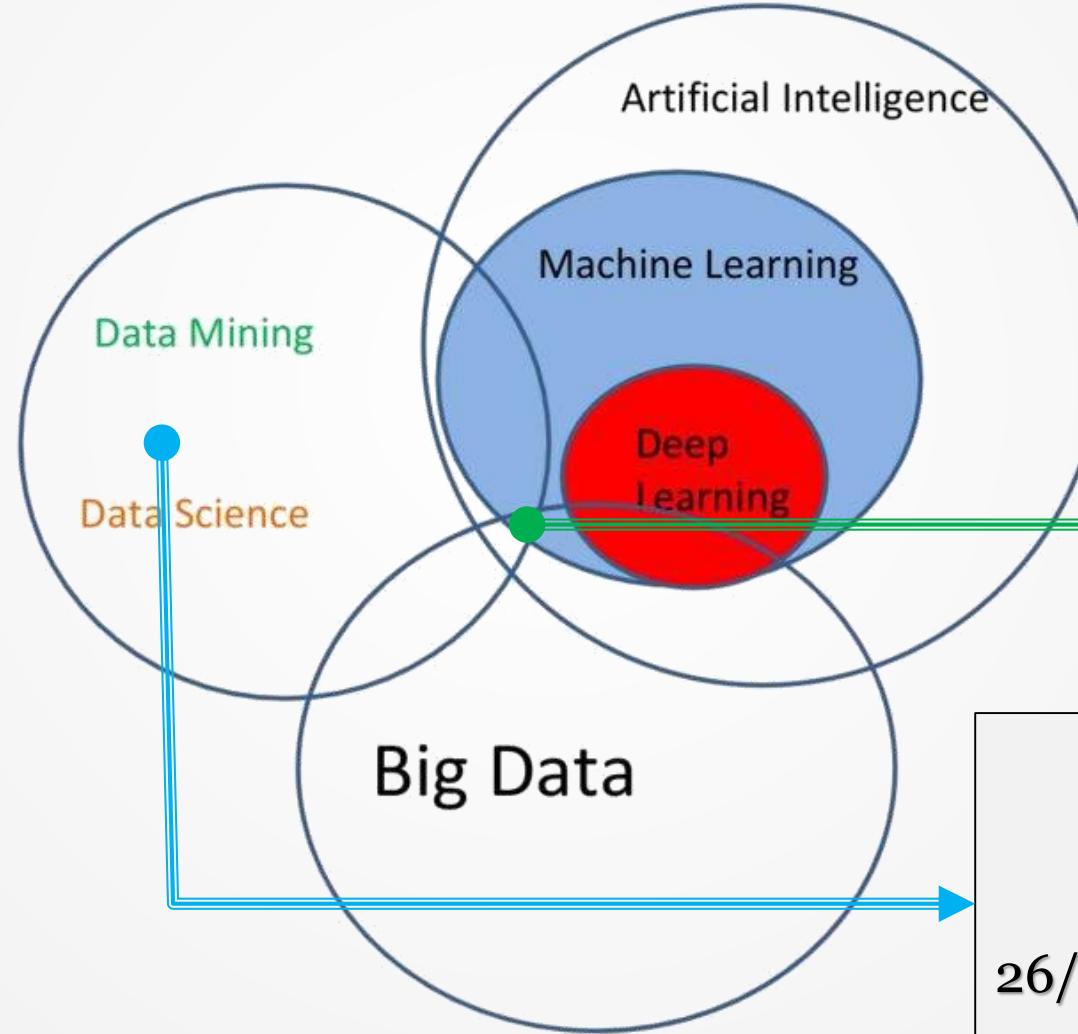
Acceso a través de API's*

Reflejan información que los usuarios deciden compartir de forma pública.

Hoy, es más fácil estudiar redes sociales, las actividades quedan registradas en millones de datos que no desaparecen con el tiempo.

Entorno Big Data – IA – Data Science

"Data Science is a broad field. It is an exciting field. It is an art. It is a science." Pradeep Menon.



Analítica con redes
sociales – Twitter

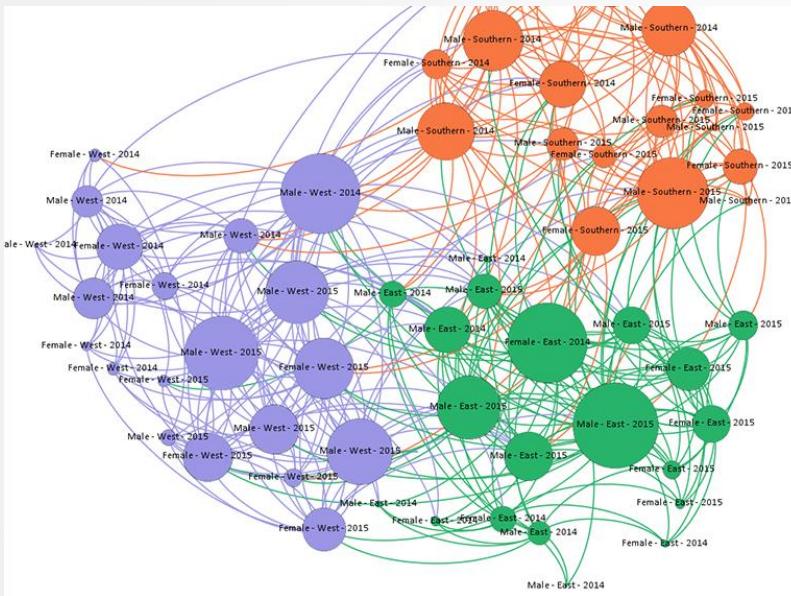
26/01/2019

¿Por qué analítica **con** redes sociales?

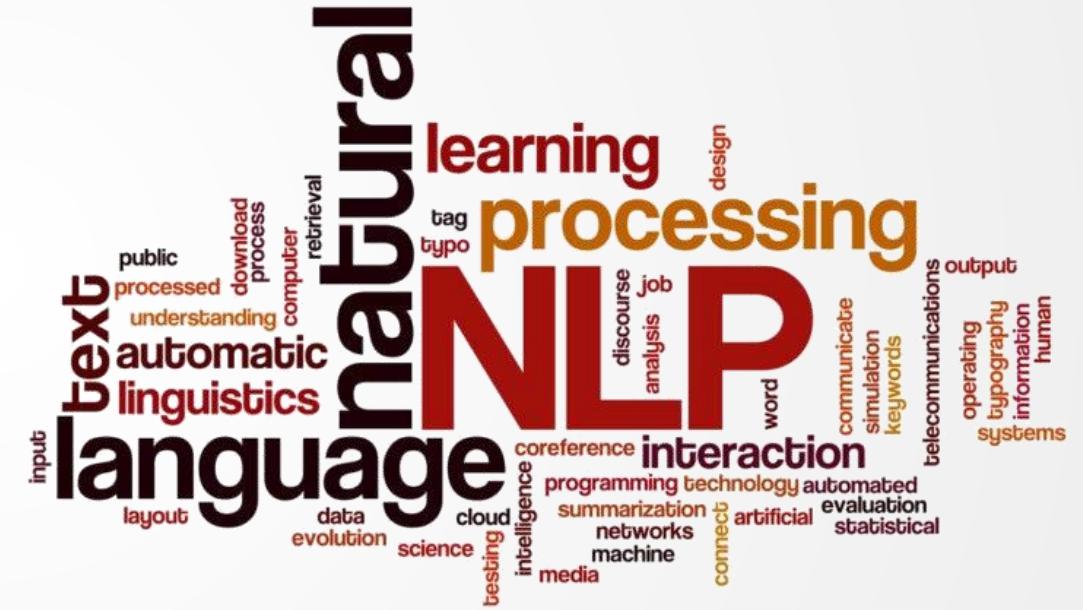
Analizar la red

y/o

Analizar el contenido



Uso de redes y teoría de grafos.
Complejidad

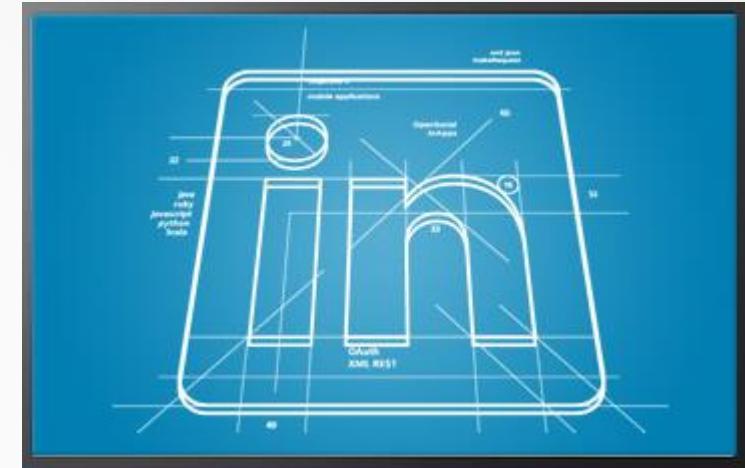


Data Mining – Machine Learning

DESAFIOS DEL MUNDO REAL

- Texto no estándar
- Minería de textos
- Falta de información
- Sesgo/Sobreajuste/Dimensionalidad
- Gran volumen de datos
- Limitaciones de la API - Censura
- Objetivo claro

HERRAMIENTAS - API Application Programming Interface



- Reglas
- Términos y condiciones
- Limitaciones
- Costo



HERRAMIENTAS - Lenguajes

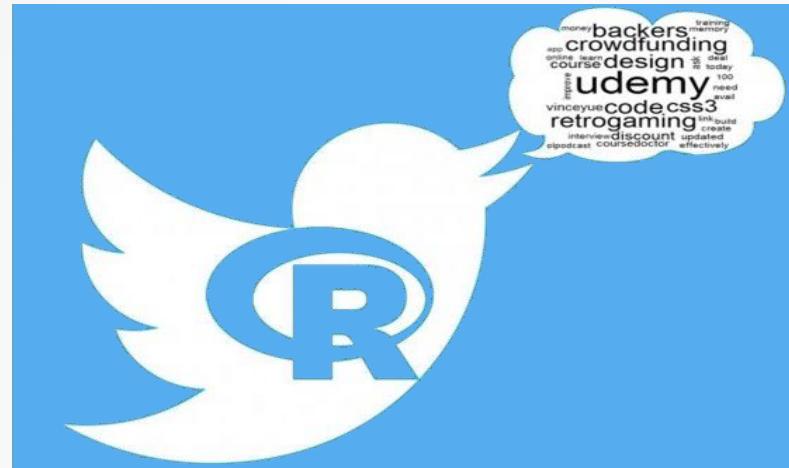


- Librerías
- Blogs
- Documentación



JavaScript

NUESTRAS HERRAMIENTAS



Analítica con redes sociales - Twitter



Introducción

Presentación, Contexto, Herramientas.



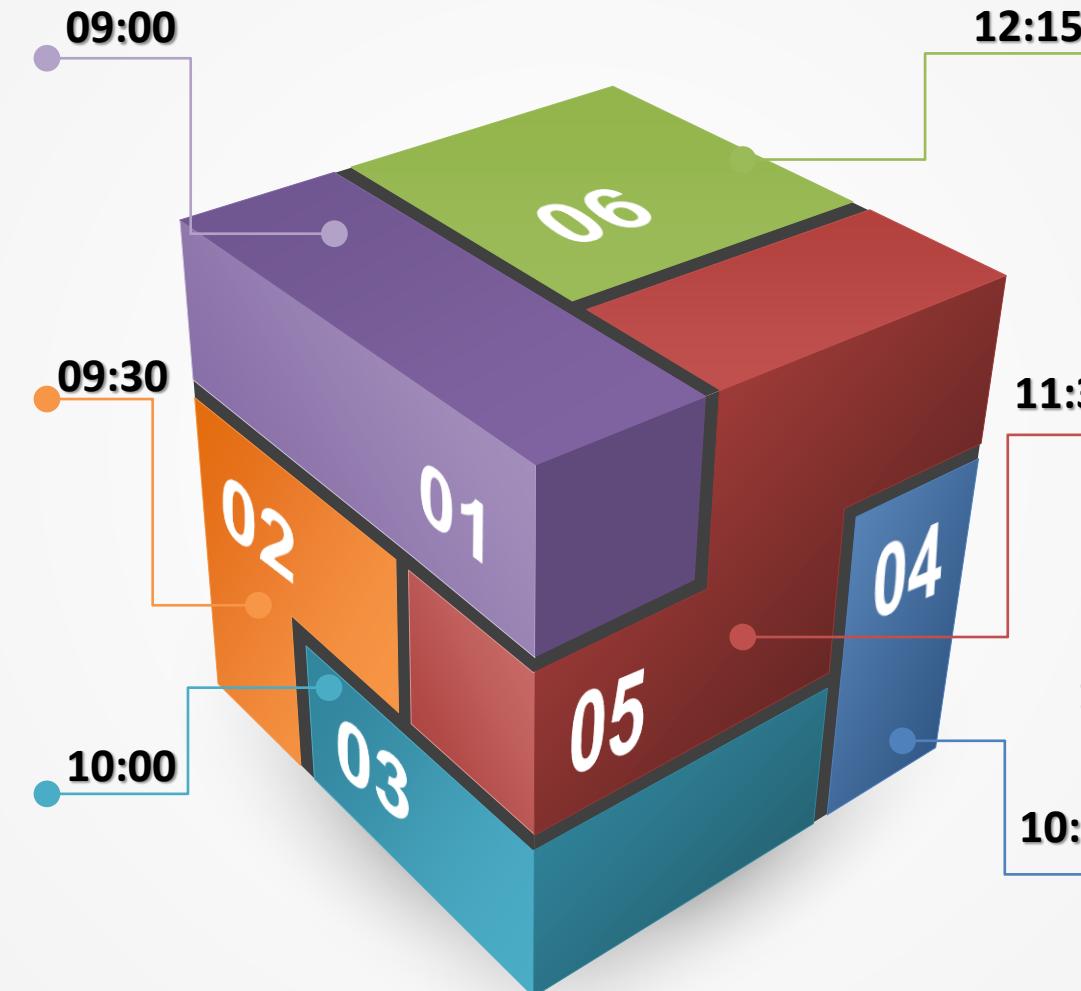
Extracción - API

Twitter API - Rtweet
NodeXL - LostCircles



Análisis de redes

¿Qué es? Conceptos
Ej: Facebook LC
Ej: NodeXL



10:00-10:15: Video Break - Opcional



Modelos predictivos



Modelos descriptivos



11:15-11:30: Break

Minería de texto

NLP, limpieza, retos, limpieza de la base

Analítica con redes sociales - Twitter



Extracción

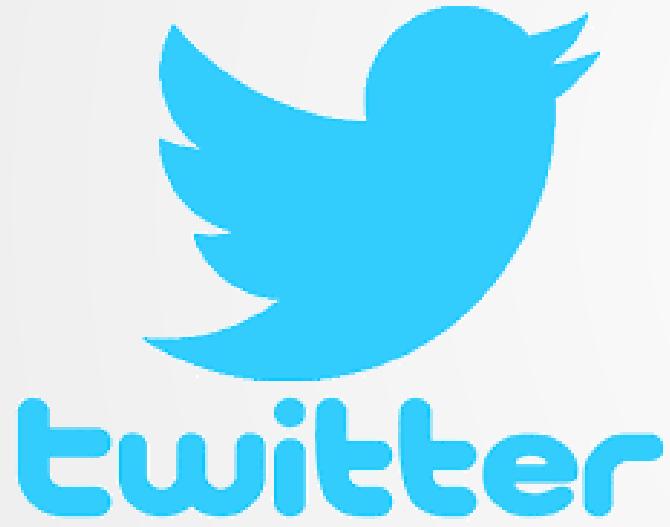
Twitter – API

Rtweet

Node XL

Opcional: LostCircles

Es un servicio de noticias y redes sociales en línea donde los usuarios publican e interactúan con mensajes, "tweets", restringidos a 280 caracteres. **Los usuarios registrados pueden publicar tweets, pero aquellos que no están registrados solo pueden leerlos.** - Fuente: Wikipedia



21 de marzo del 2006



"La posibilidad de generar y compartir ideas e información al instante, sin ningún tipo de obstáculos"

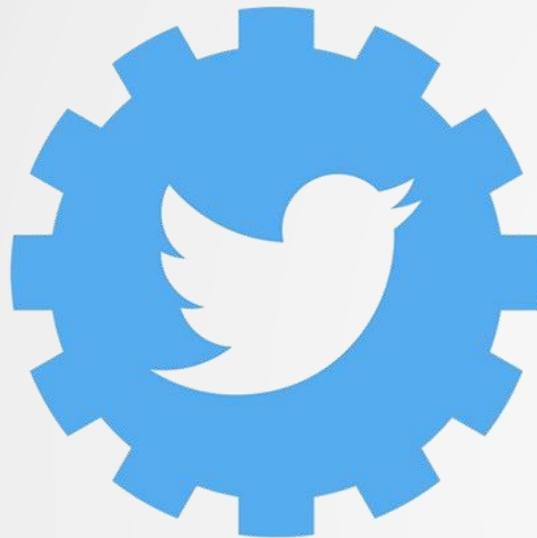
Límites en cuanto al tipo de contenido y los comportamientos tolerados

Twitter



- 500M trinos diarios
- Solo 20 de las “Fortune 500” se vinculan a sus clientes a través de Facebook, 83% tiene presencia en Twitter.
- El 96% de la gente habla sobre marcas online pero no las sigue
- El 78% de la gente que se queja en Twitter espera una respuesta en un tiempo máximo de una hora.
- 23m de usuarios son Bots (201803)

No es una muestra representativa del mundo



Puedes:

Desarrollar soluciones que retransmiten de forma automática información útil en los Tweets.

Publicar campañas de creatividades que responden automáticamente a los usuarios que interactúan con tu contenido.

Desarrollar soluciones que responden a los usuarios de forma automática en los Mensajes Directos.

Probar cosas nuevas que ayudan a la gente (y que cumplen con nuestras reglas).

NO puedes*:

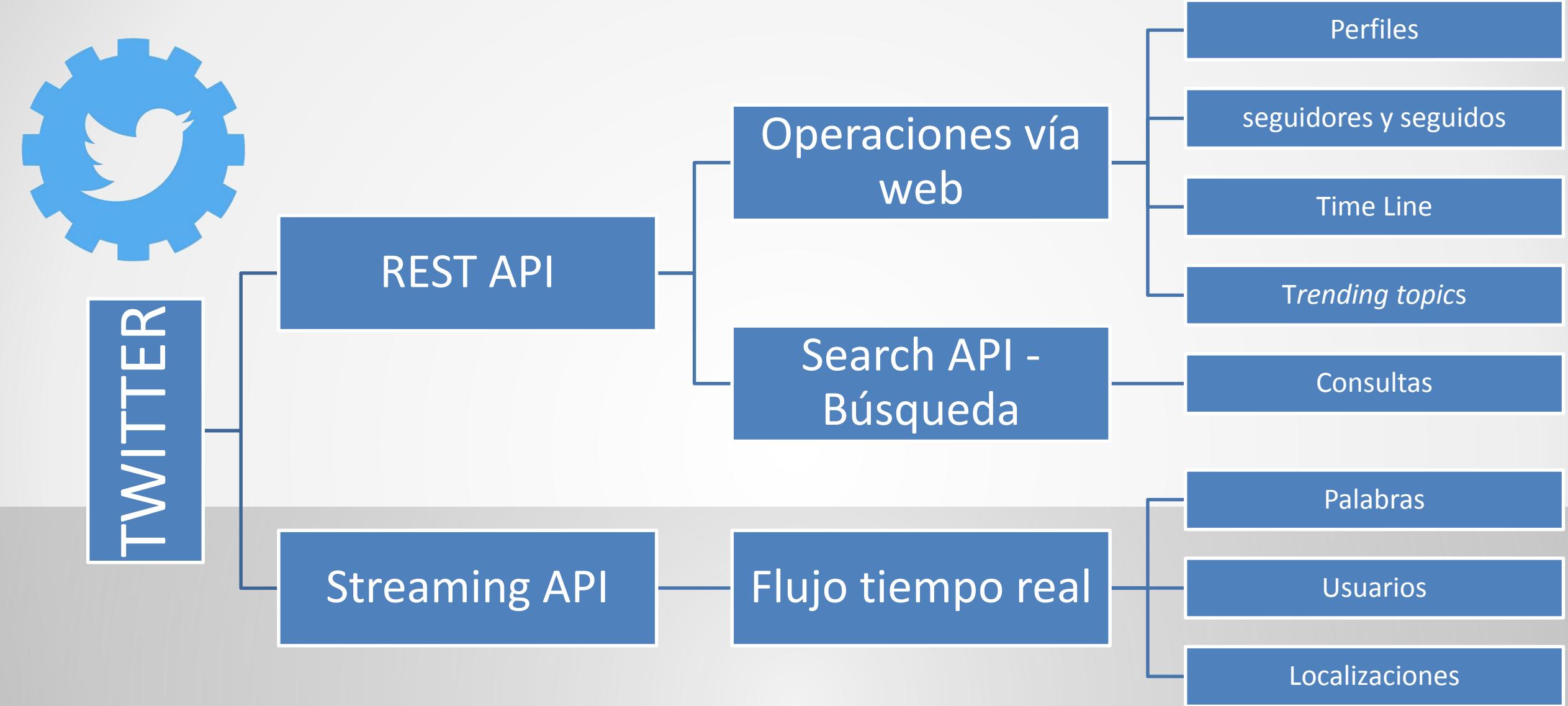
Registrar múltiples aplicaciones para un caso de uso único, o casos de uso sustancialmente similares o que se superpongan.

Abusar de la API de Twitter o intentar eludir los **límites de velocidad**.

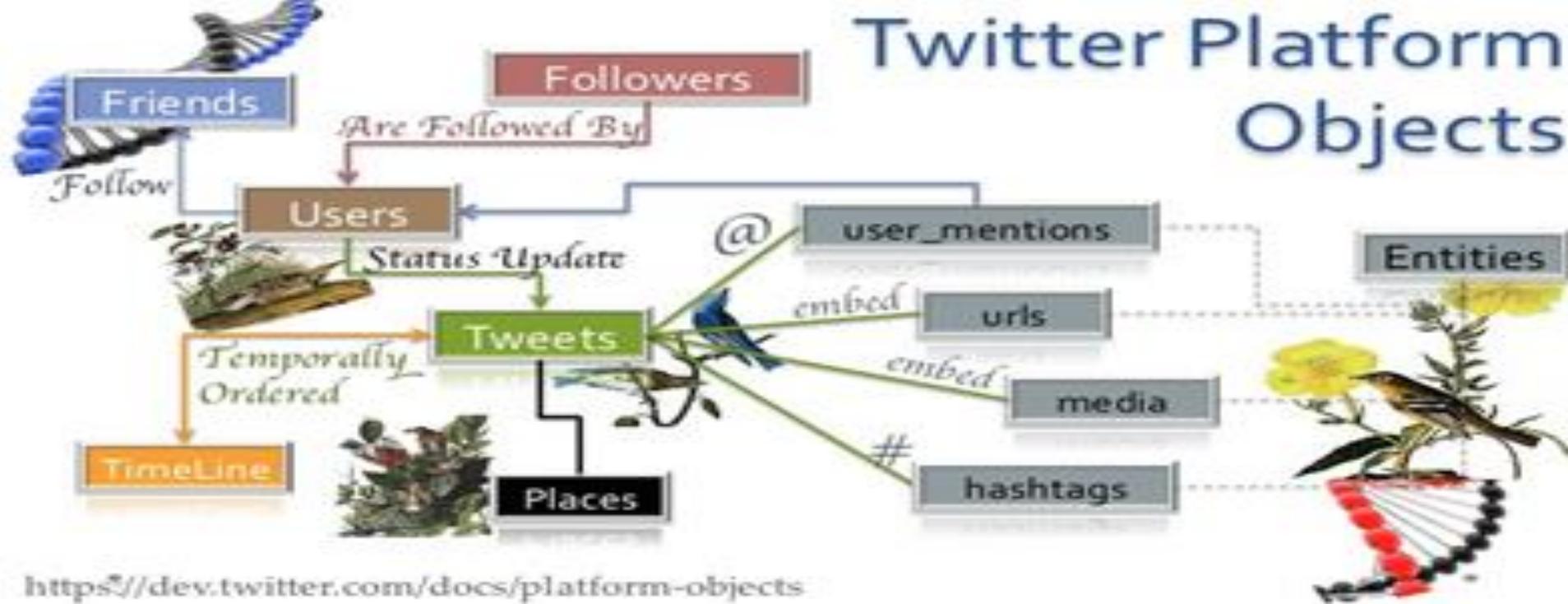
Enviar spam o importunar a los usuarios.

Automatización para publicar contenido idéntico, sustancialmente similar, o para realizar acciones como Me gusta o Retweets

* Excepto: Anuncios de servicio público de amplio interés comunitario: Meteorológico, (alertas de terremotos o tsunamis) o de emergencia



Twitter Platform Objects



<https://dev.twitter.com/docs/platform-objects>

Tweets persistence

REST API
Statuses/
user_timeline

Search API

Streaming
API

-x(*)

-7 days

Now

(*) Depending on user's published frequency, maximum last 3200 tweets

START



0. Crear tu Aplicación en Twitter

Ingresar a <https://developer.twitter.com/en/apps>

Crear una nueva App

- Nombre
- Descripción
- URL(LinkedIn, Twitter, GitHub, etc)
- CallBack Url: <http://127.0.0.1:1410>
- Detalles

Definir los permisos

- Read – Write – Direct Messages

App details

The following app details will be visible to app users and are required to generate the API keys needed to authenticate Twitter developer products.

App name (required) ?

Maximum characters: 32

Application description (required)

Share a description of your app. This description will be visible to users so this is a good place to tell them what your app does.

Please be detailed.

Between 10 and 200 characters

Website URL (required) ?

https://

0. Crear tu Aplicación en Twitter

Obtener las claves de la App - Keys and Access Tokens

The screenshot shows the 'Keys and tokens' tab of a Twitter developer application. It displays Consumer API keys and Access token & access token secret.

Consumer API keys:
efw...FvPEB (API key)
rhu...0bd (API secret key)

Access token & access token secret:
943...8337-Du8Gv...q59hP (Access token)
hdgll...XKIgIlc (Access token secret)

Access level: Read, write, and direct messages

Actions: Revoke, Regenerate

0. Crear tu Aplicación en Twitter

Obtener las claves de la App - Keys and Access Tokens

The screenshot shows the 'Keys and tokens' tab of a Twitter developer application. It displays Consumer API keys and Access token & access token secret.

Keys and tokens
Keys, secret keys and access tokens management.

Consumer API keys
efw... FvPEB (API key)
rhu... 0bd (API secret key)

[Regenerate](#)

Access token & access token secret
943... 8337-Du8Gv... q59hP (Access token)
hdgll... XKIgIlc (Access token secret)

Read, write, and direct messages (Access level)

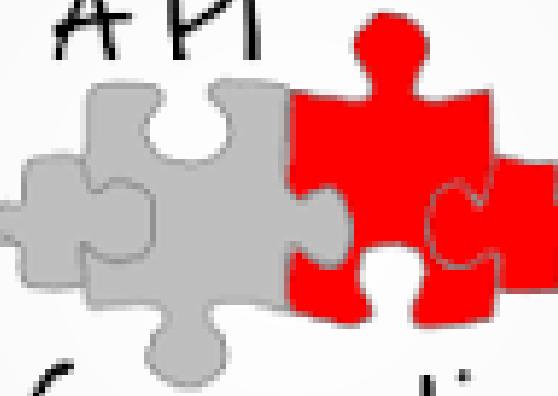
[Revoke](#) [Regenerate](#)



Script 1 - Collecting Twitter Data



API
Connection



BRAVO!



Analítica con redes sociales - Twitter



Introducción

Presentación, Contexto, Herramientas.



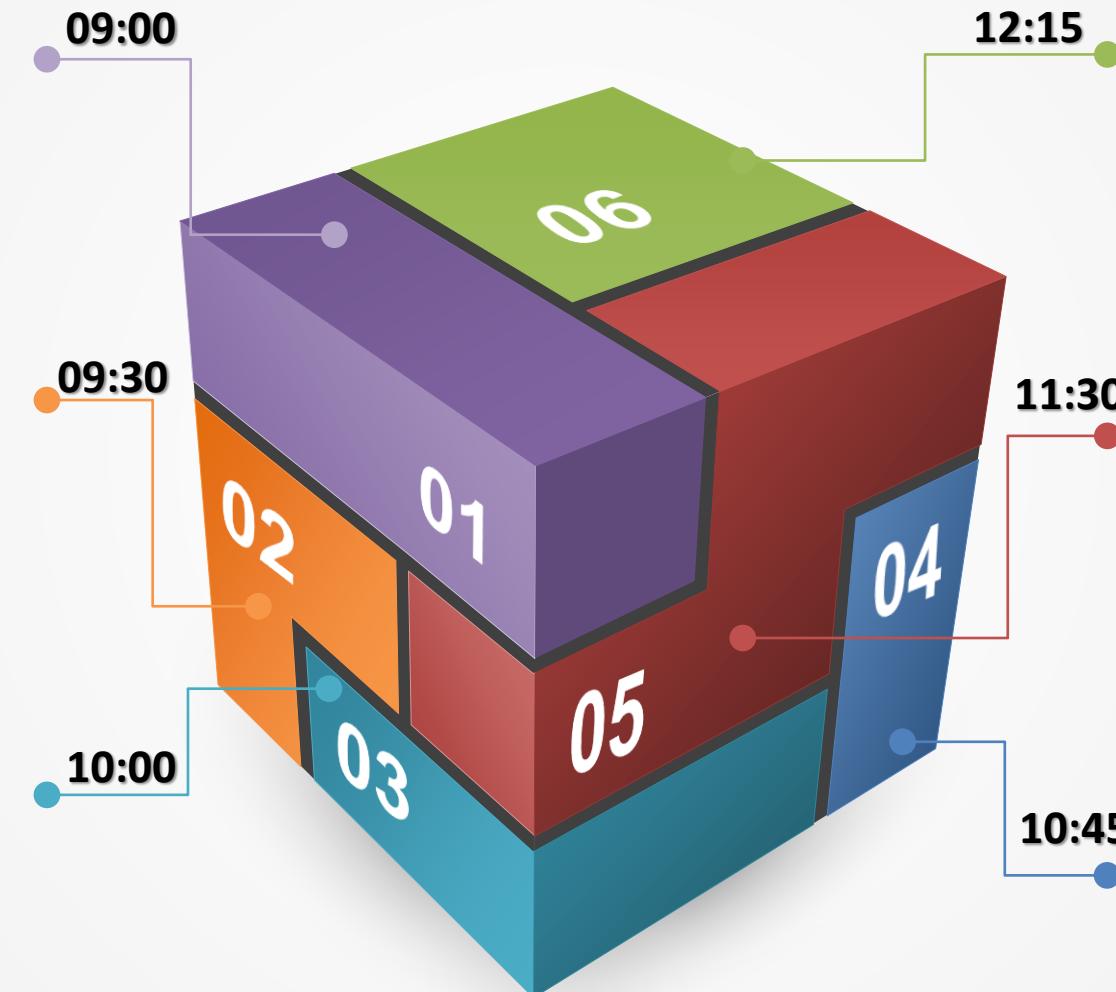
Extracción - API

Twitter API - Rtweet
NodeXL - LostCircles



Análisis de redes

¿Qué es? Conceptos
Ej: Facebook LC
Ej: NodeXL



10:00-10:15: Video Break - Opcional

11:15-11:30: Break



Modelos predictivos

This is a sample text. Insert your desired text here.



Modelos descriptivos

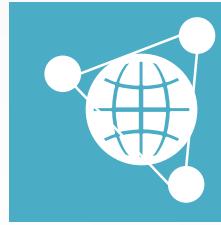
This is a sample text. Insert your desired text here.



Minería de texto

This is a sample text. Insert your desired text here.

Analítica con redes sociales - Twitter



Análisis de redes

Complejidad
Análisis de redes
Conceptos

Análisis de tu red de Facebook
LostCircles
Gephi

Redes aplicadas a Twitter

Analisis de red con NodeXL

Comentarios

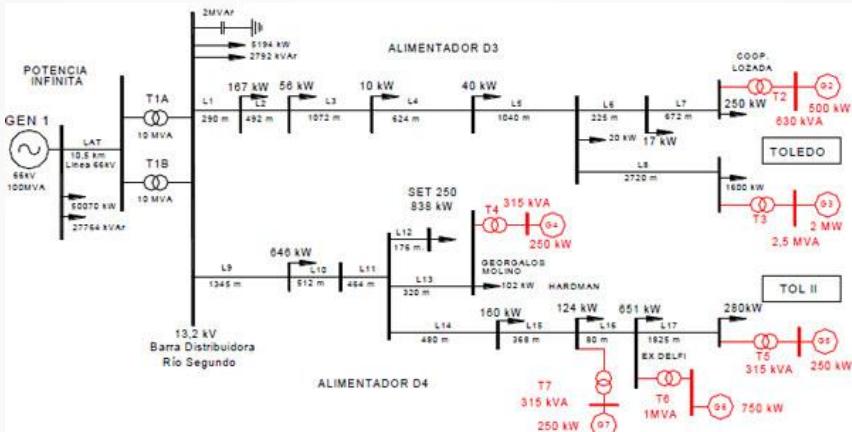
Análisis de redes

- Análisis: Teoría de redes - Teoría de grafos.

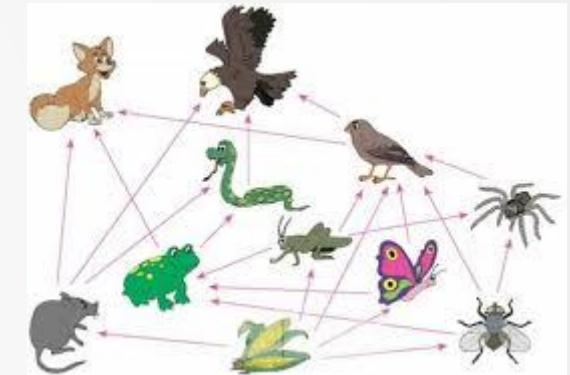


Transporte

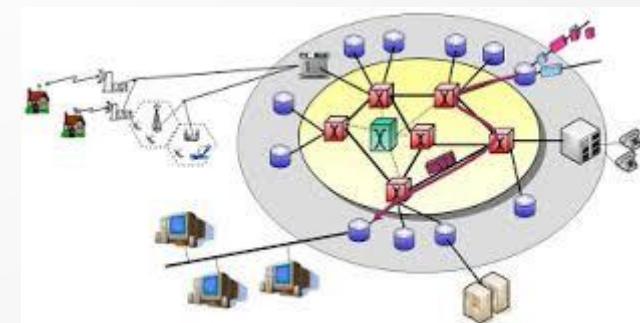
Social



Eléctrica



Biológica



Comunicación

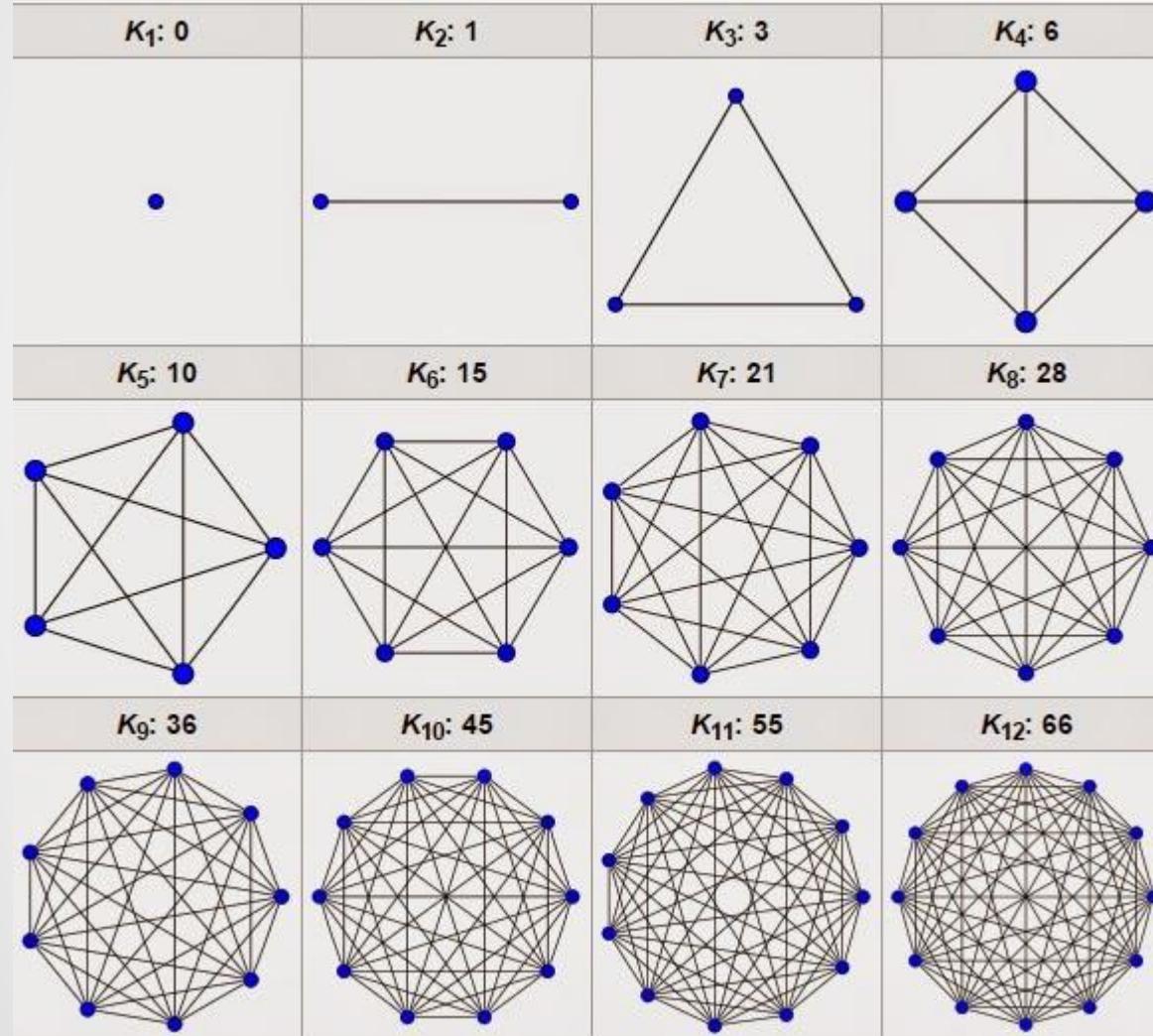
Grafos – Una red

- Dos vértices o nodos



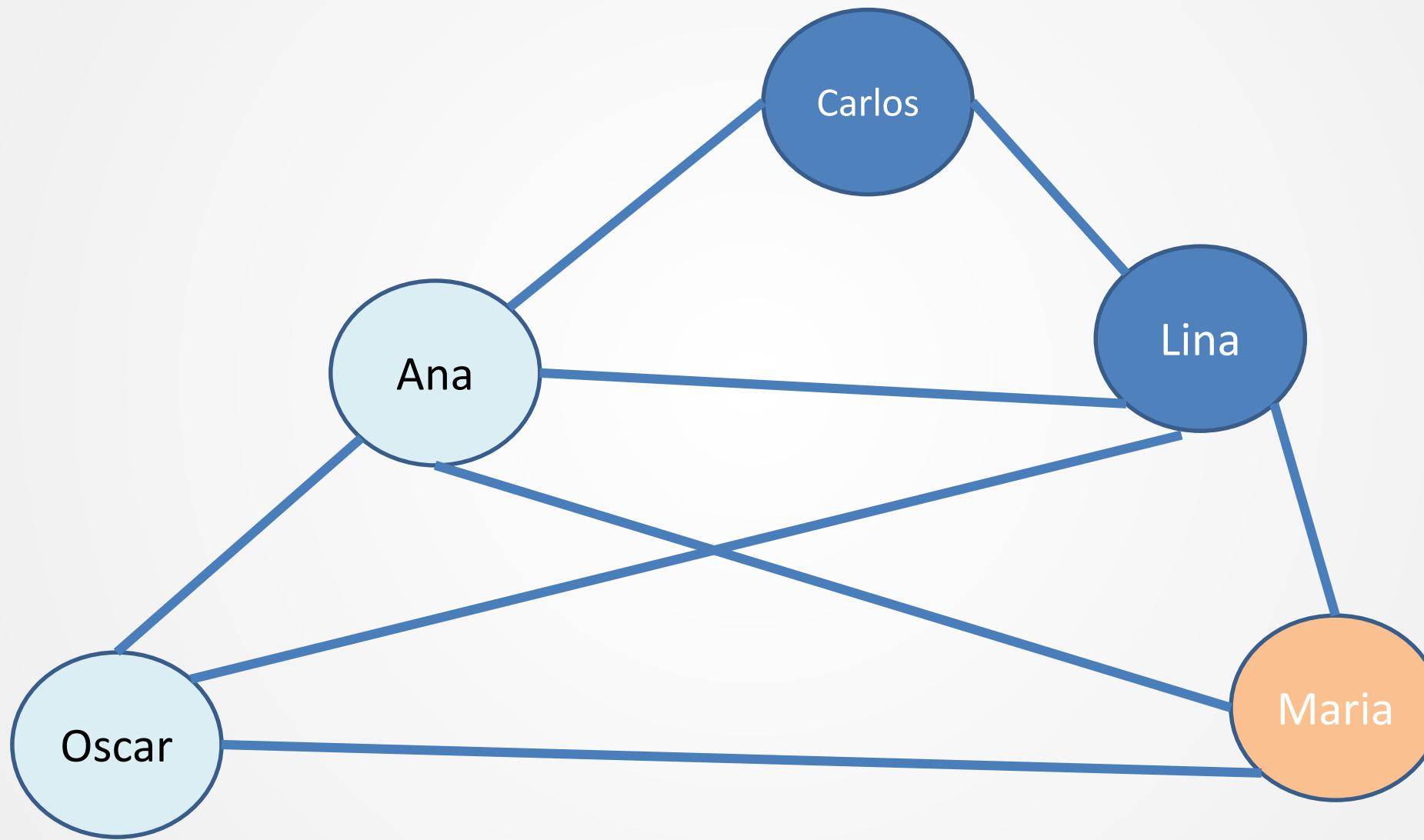
- La unión es una arista(Edge)
 - V: 2
 - E: 1

Grafos != Gráficos



- $G=(V,E)$
- Pareja ordenada
- V es un conjunto **no vacío** de vértices
- E es un conjunto de aristas
- E consta de pares **no ordenados** de vértices
- Si el grafo es dirigido se le llama dígrafo

Grafos – Una historia



COMPLEJIDAD

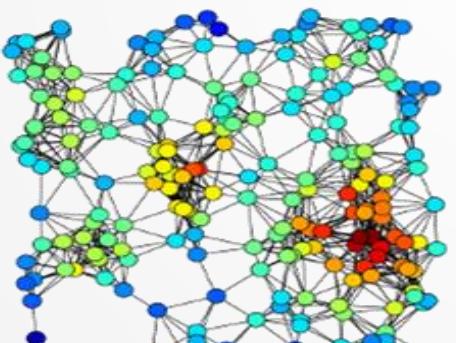
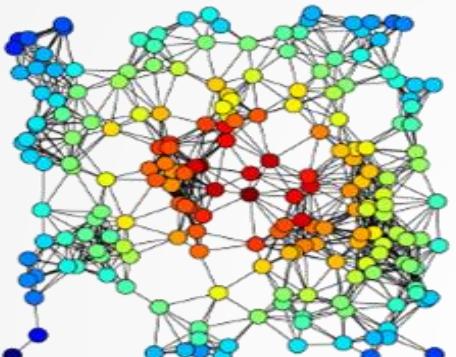
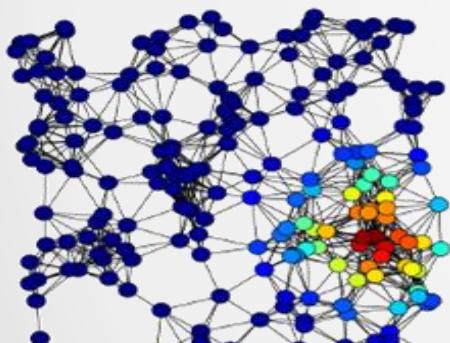
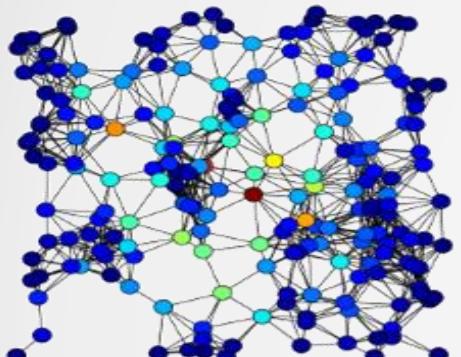
- **Paradigma convencional= reduccionista:** las partes son tomadas en cuenta de manera aislada, importan más sus propiedades que sus relaciones.
- **Paradigma complejo:** las conexiones toman la misma y a veces mayor importancia que la parte misma.

VIDEO BREAK

[https://www.ted.com/talks/eric berlow how complexity leads to simplicity?language=en](https://www.ted.com/talks/eric_berlow_how_complexity_leads_to_simplicity?language=en)

[https://www.ted.com/talks/nicholas christakis how social networks predict epidemics](https://www.ted.com/talks/nicholas_christakis_how_social_networks_predict_epidemics)

Medias de centralidad



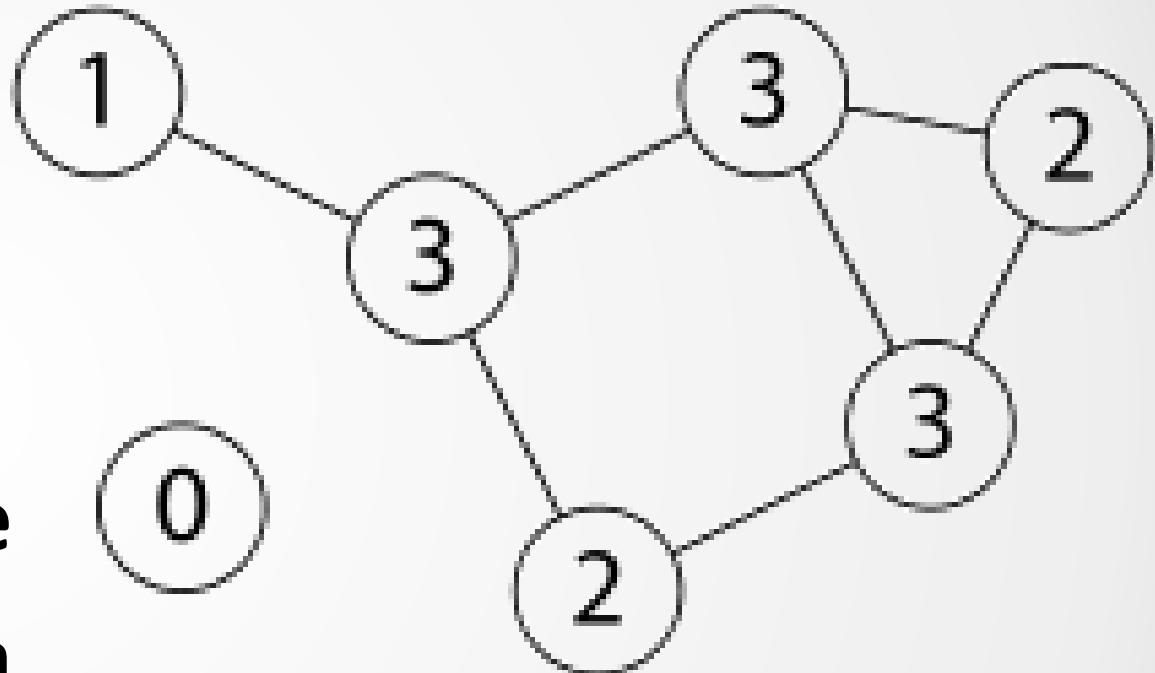
No es un atributo intrínseco de los nodos, es un atributo estructural.

Depende estrictamente de su localización en la red:

- A) intermediación
- B) cercanía
- C) valor propio
- D) grado

GRADO

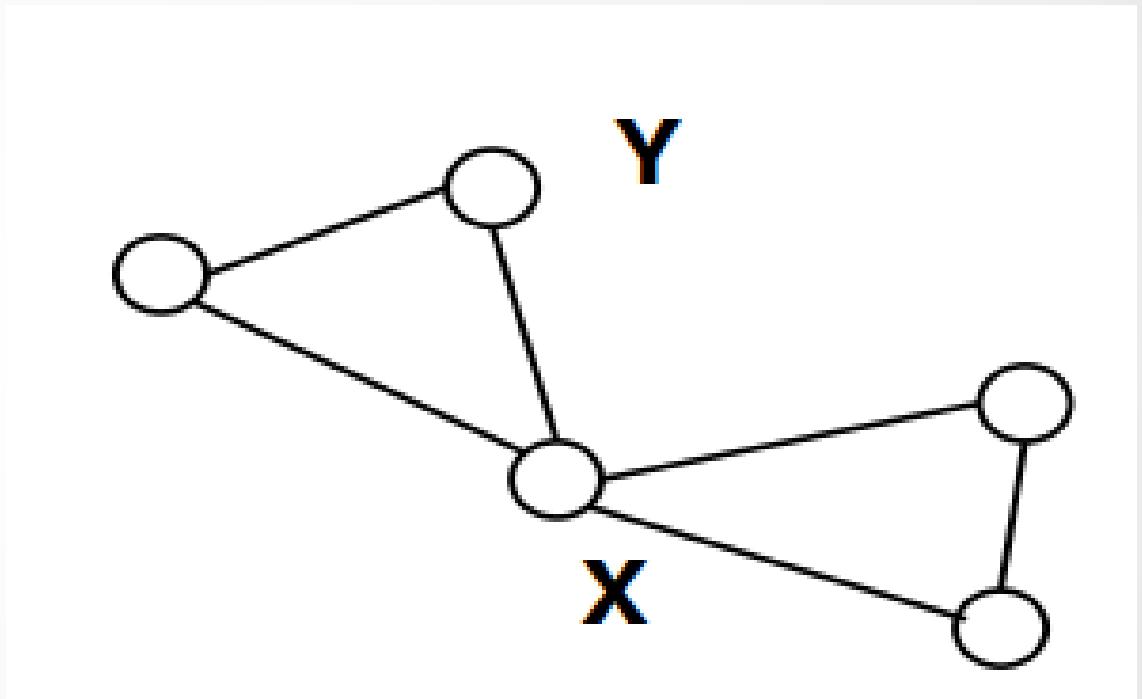
- Número de conexiones adyacentes
- En redes dirigidas:
- Prestigio de entrada **Soporte**
- Prestigio de salida **Influencia**

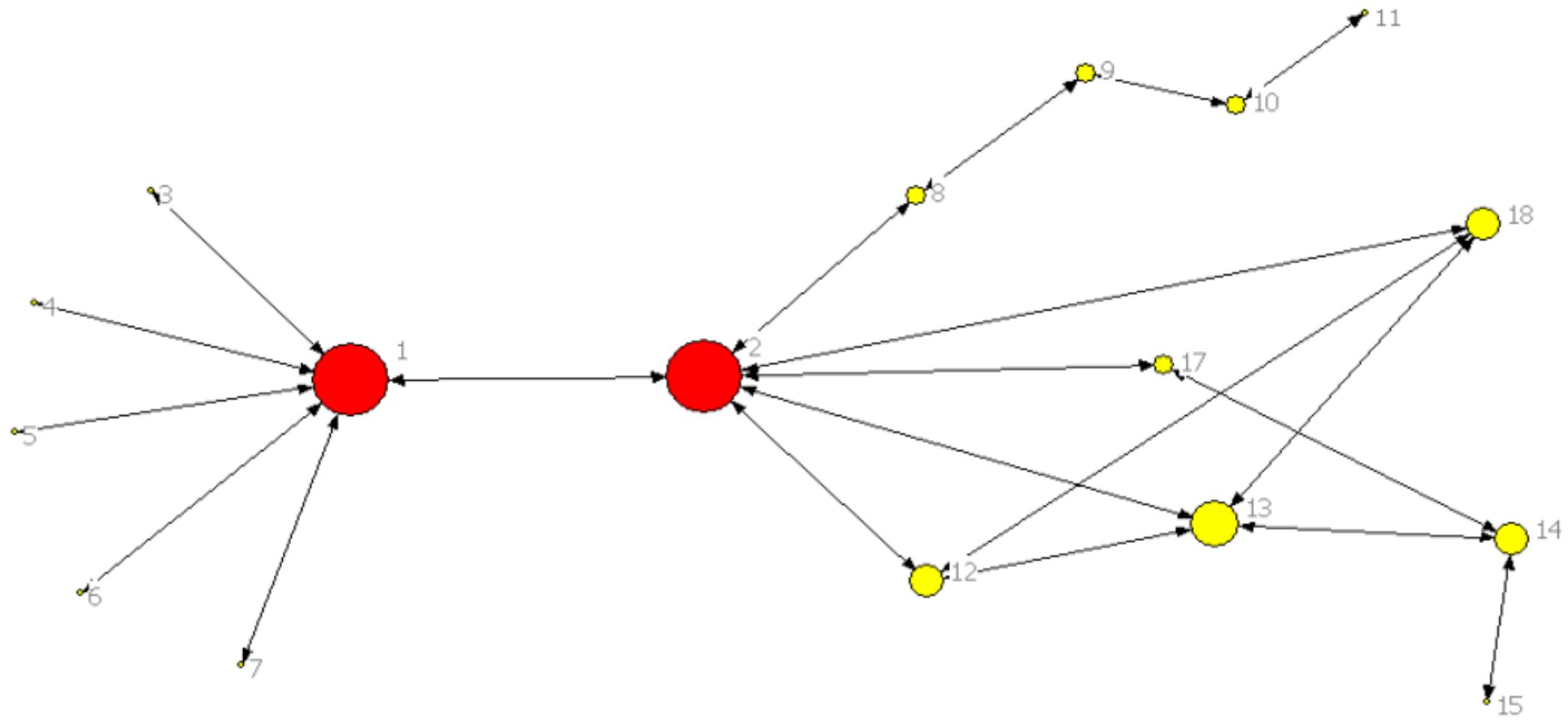


Sólo mide la importancia con respecto a los vecinos más cercanos. Se asume que las conexiones de los amigos no importan

Intermediación

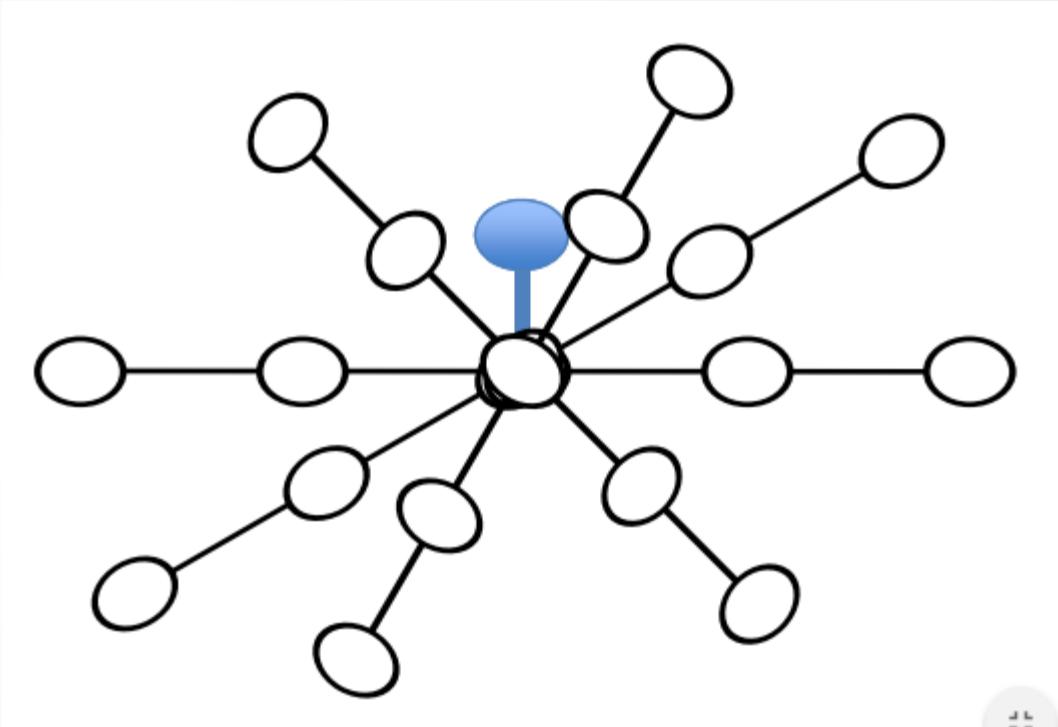
- “Cuantos más nodos que necesiten pasar por mi para hacer sus conexiones indirectas por los caminos más cortos, más central seré yo.”
- Gatekeepers





Cercanía/Lejanía

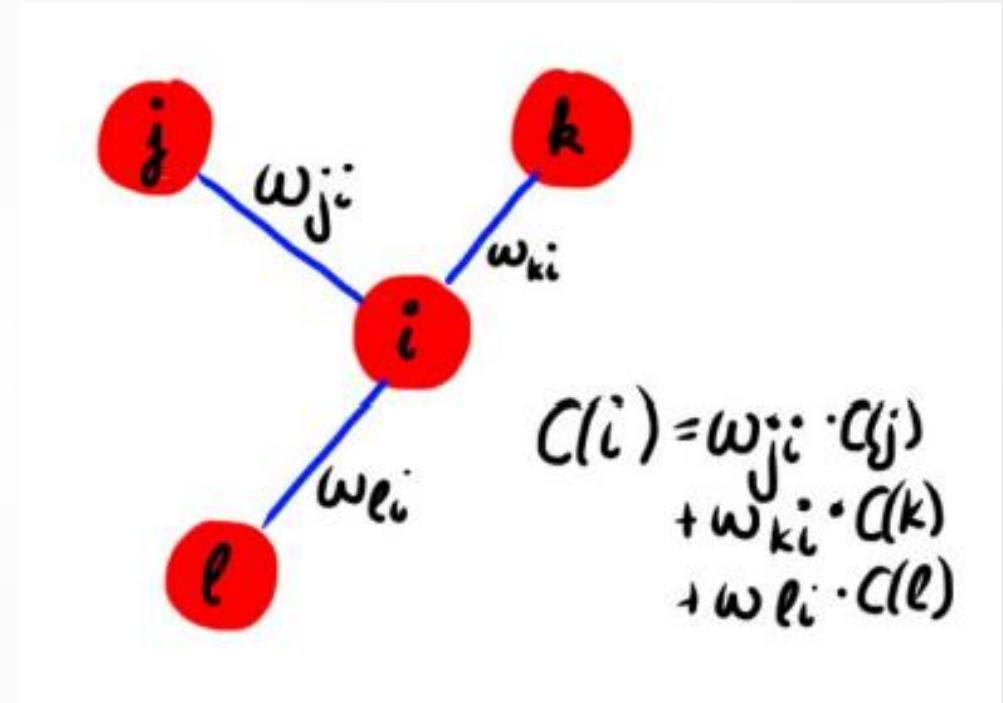
- “estar en medio de las cosas”, no demasiado lejos del centro, para lo cual no es necesario estar en una posición de correduría
- La suma de distancias de los caminos mínimos para cada actor es la lejanía. La inversa de dicha suma es la medida de cercanía



Centralidad de vector propio

Pagerank

- el poder y el status de un actor (ego) se define recursivamente a partir del poder y el status de sus vecinos (alters)

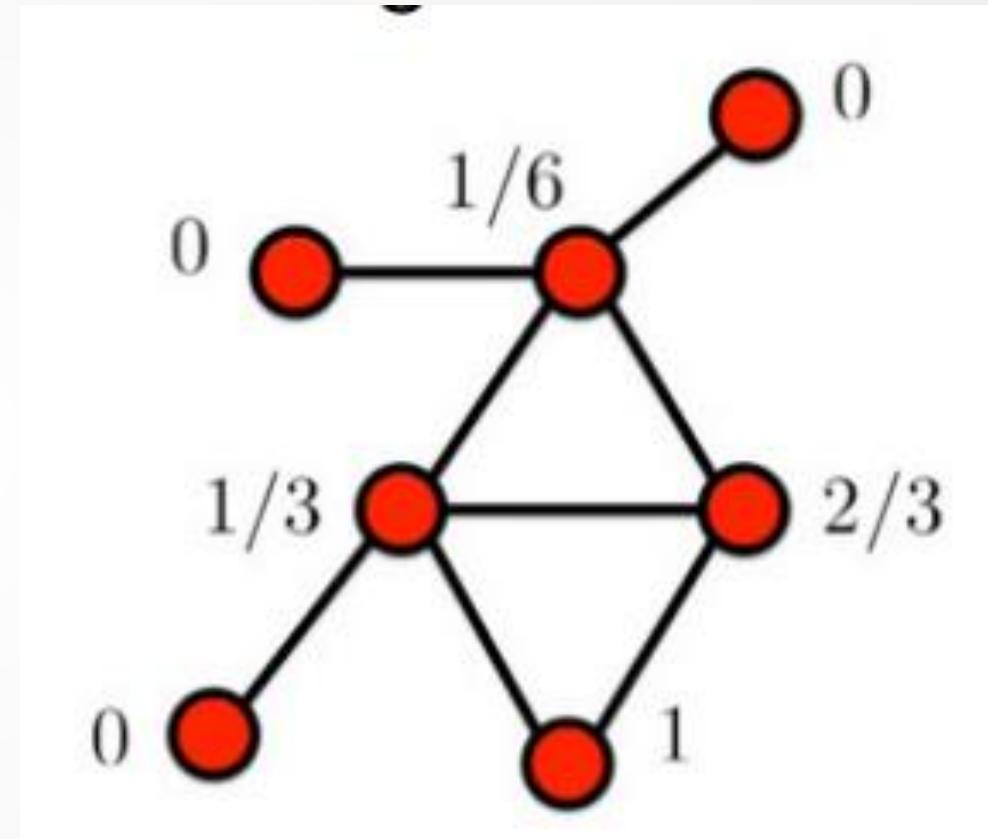


No todas las conexiones tienen la misma importancia. No se tiene en cuenta la cantidad sino la calidad de las mismas

Clustering

- Transitividad: los amigos de un actor dado también suelen ser amigos entre sí.
- Coeficiente de Clustering
- Densidad local

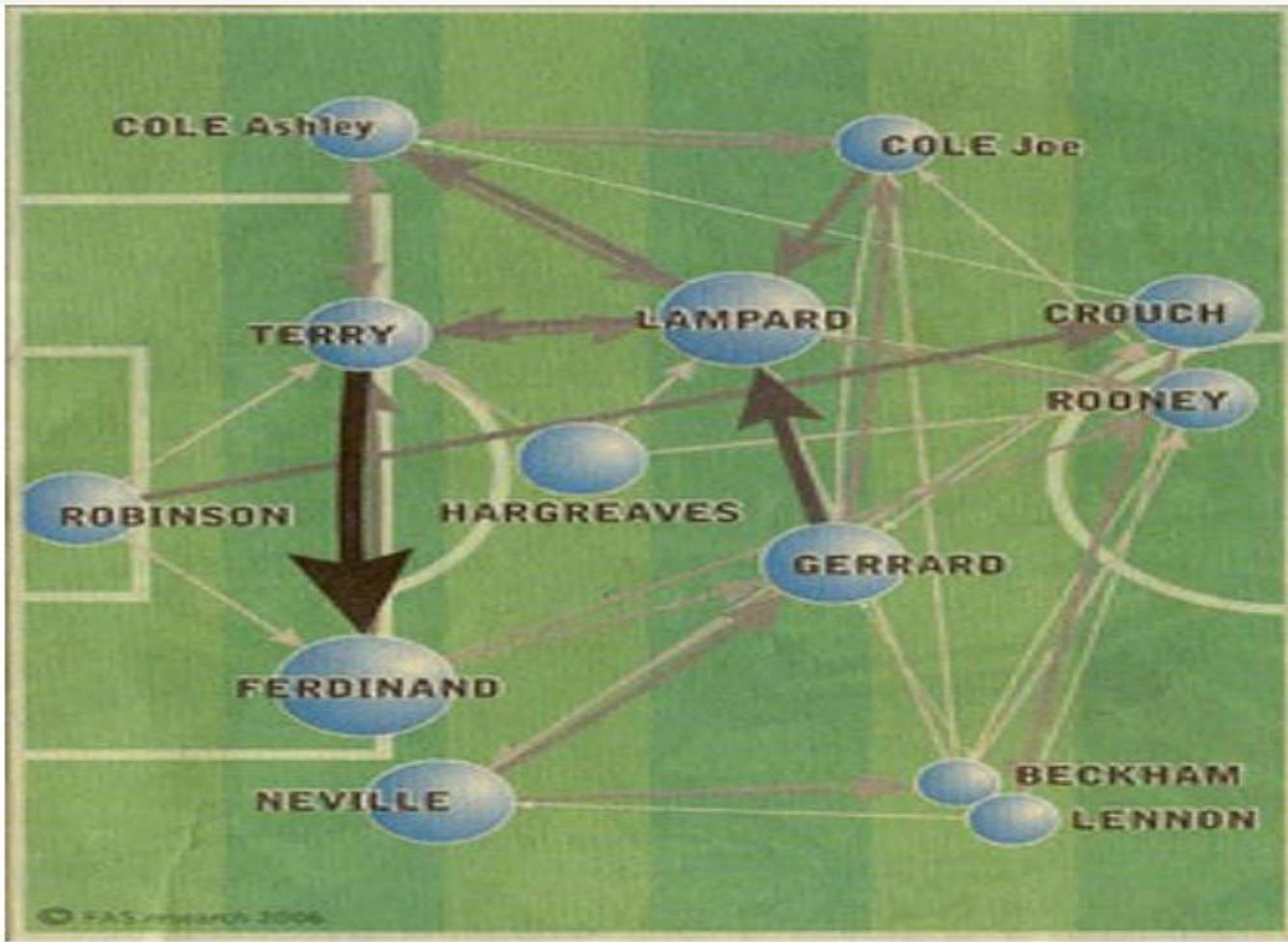
$$C_{CLC}(i) = \frac{2L_i}{k_i(k_i - 1)}$$



MEDIDAS DE LA RED

- **Distancia:** Es el camino entre un par de nodos.
- **Distancia Geodésica:** Es el camino más corto entre dos nodos.
- **Diámetro:** Es la longitud de la distancia geodésica más larga entre cualquier par de nodos en la red.
- **Densidad:** Proporción del número de lazos presentes en la red sobre el número máximo posible.

Otras aplicaciones:



Chelsea FC:

| | |
|-------------|----|
| Robinson | 1 |
| Cole Ashley | 2 |
| Terry | 3 |
| Ferdinand | 4 |
| Neville | 5 |
| Cole Joe | 6 |
| Lampard | 7 |
| Heargraves | 8 |
| Gerrard | 9 |
| Beckham | 10 |
| Lennon | 11 |
| Crouch | 12 |
| Rooney | 13 |

2. Vamos a Facebook y Gephi



Lost Circles^{Beta 2}



2. Salgamos de R – Vamos a Facebook



Lost Circles Beta 2

Visualize the hidden structure of your facebook network.

[Install Chrome Extension](#)

Like Share Be the first of your friends to like this.

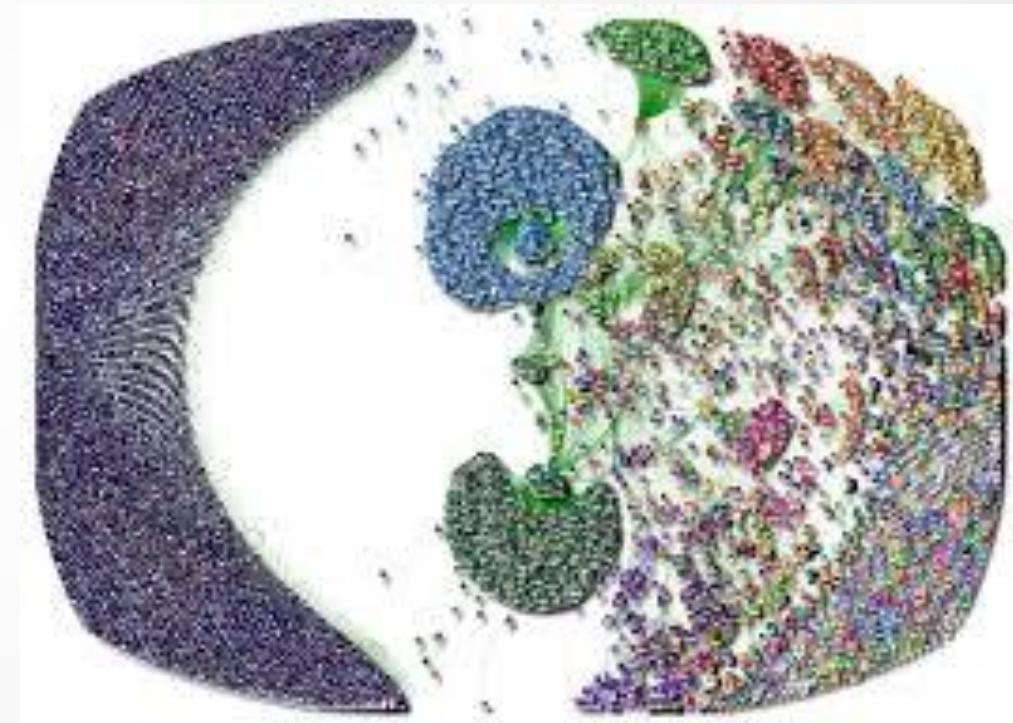
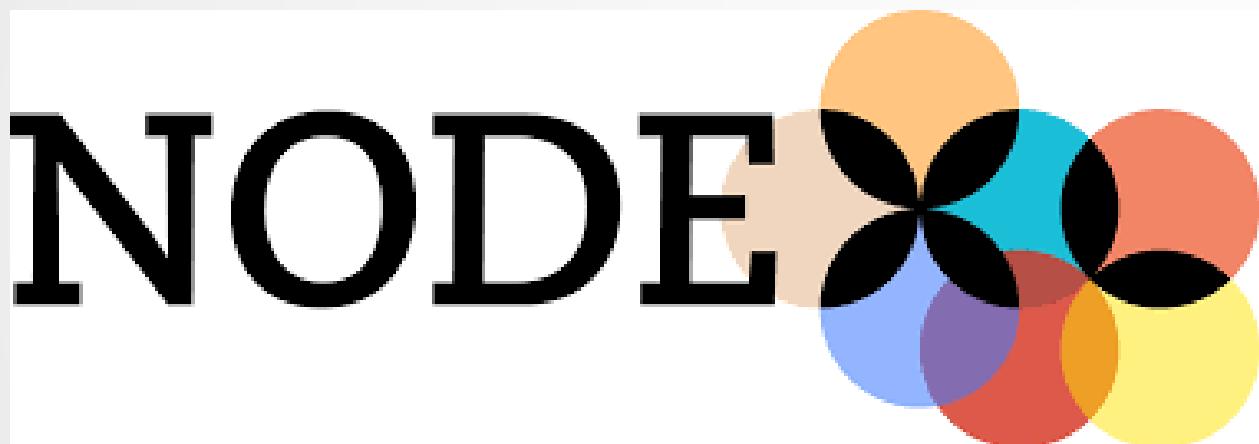
A screenshot of the Lost Circles Beta 2 website. It features two circular network graphs. The left graph shows a dense cluster of small colored dots connected by a web of thin gray lines. The right graph shows several distinct clusters of dots in various colors (blue, red, green) connected by a more complex network of lines. At the top, the text "Lost Circles Beta 2" is displayed. Below the graphs, the text "Visualize the hidden structure of your facebook network." is followed by a blue button with the text "Install Chrome Extension". At the bottom, there are "Like" and "Share" buttons with the subtext "Be the first of your friends to like this."

- Si tienes cuenta en [Facebook](#) configura el lenguaje de tu perfil a English(US)
- Instala el complemento de chrome [LostCircles](#) . Instala en tu equipo.

Tutorial LostCircles y Gephi

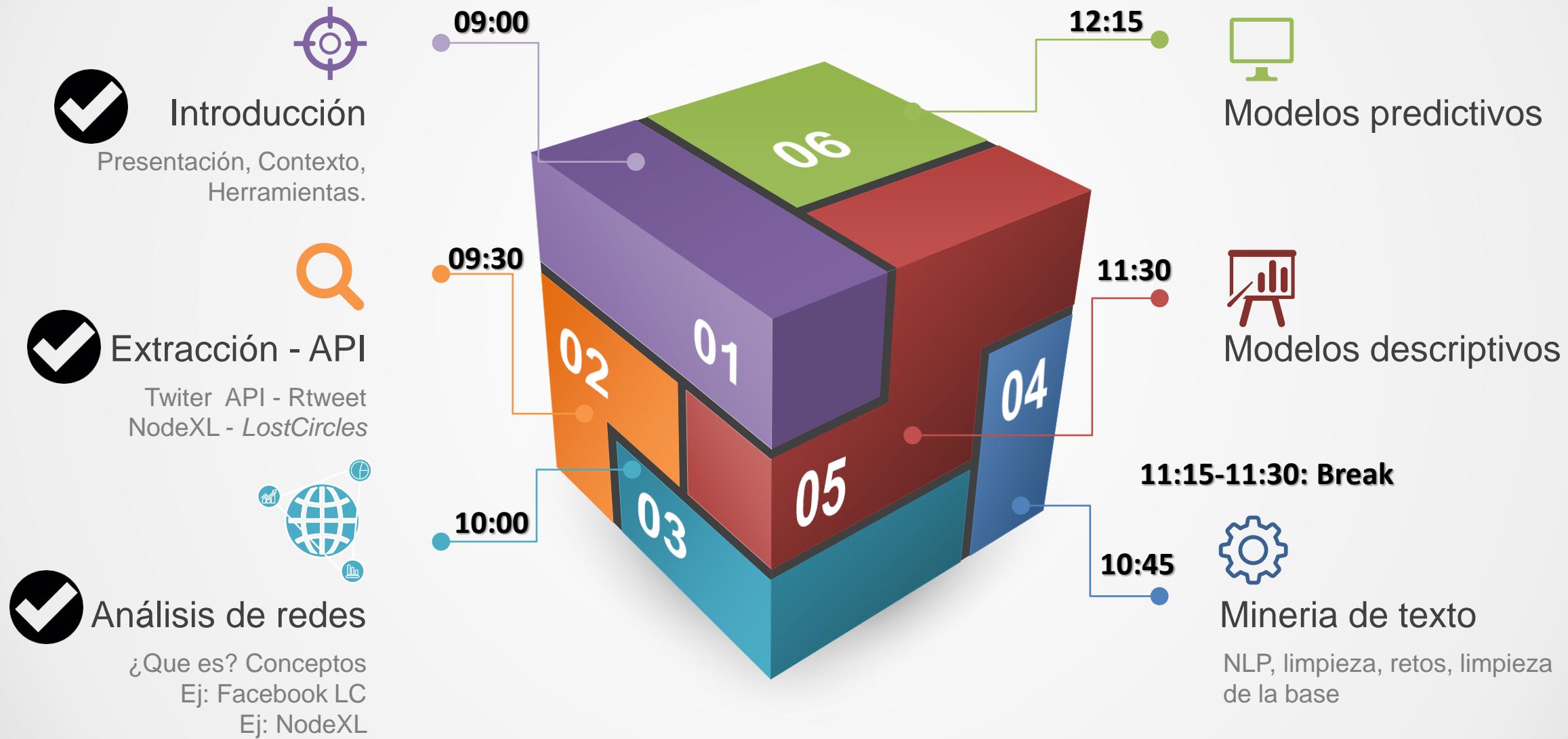
- EJERCICIO CON SU RED DE FACEBOOK
- Tamaño de la red
 - ¿Cuantos nodos y arcos componen su red?
 - ¿Cuál es el valor de la densidad de su red?
- Análisis descriptivo y medidas de la red
 - ¿Cuál es el grado medio de su red? Que significa?
 - Cuál es la persona con mayor grado ? Que concluye de su relación con ella
 - Identifique las personas con mayor grado de intermediación y cercanía ¿Es una misma persona en alguna de estas medidas? ¿Es la misma persona con mayor grado? ¿Si son diferentes concluya sobre sus relaciones con estas personas.

3. NODEXL – Ejemplo de extracción

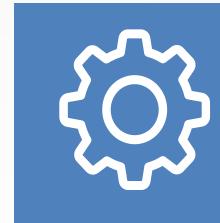
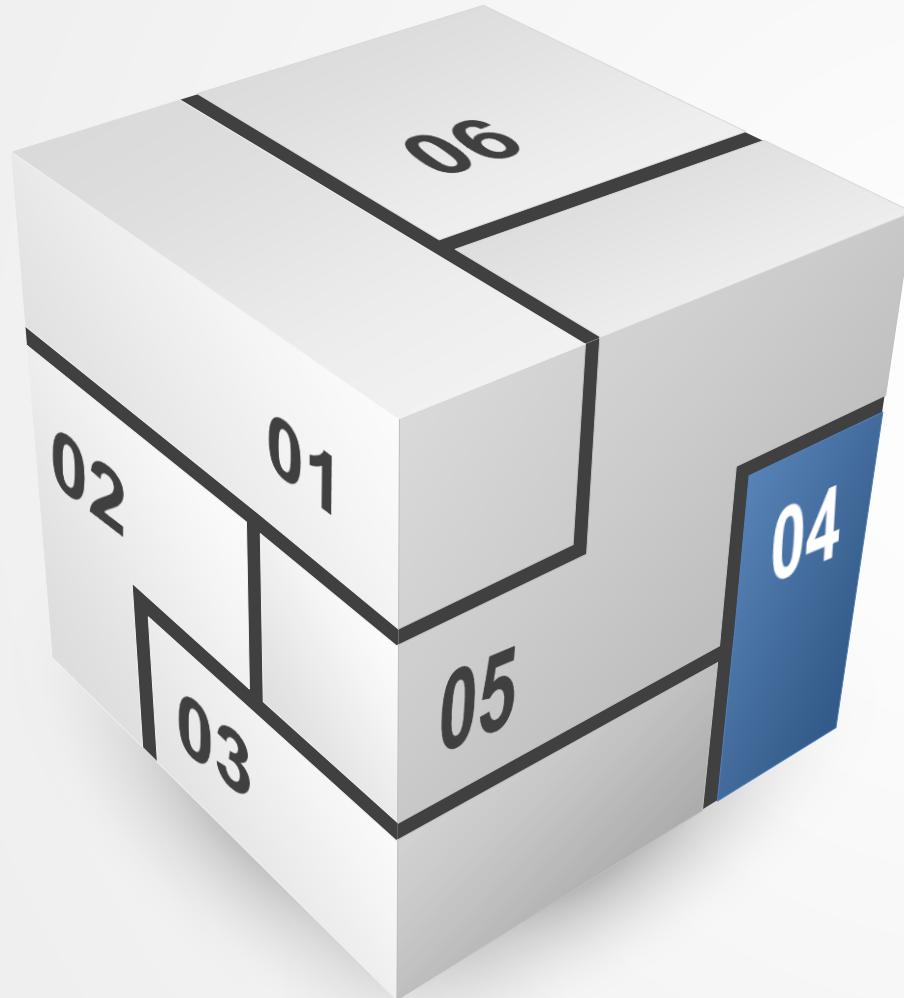


Tutorial NodeXL

Analítica con redes sociales - Twitter



Analítica con redes sociales - Twitter



Minería de texto

Conceptos

Limpieza

Estructuración

StopWords

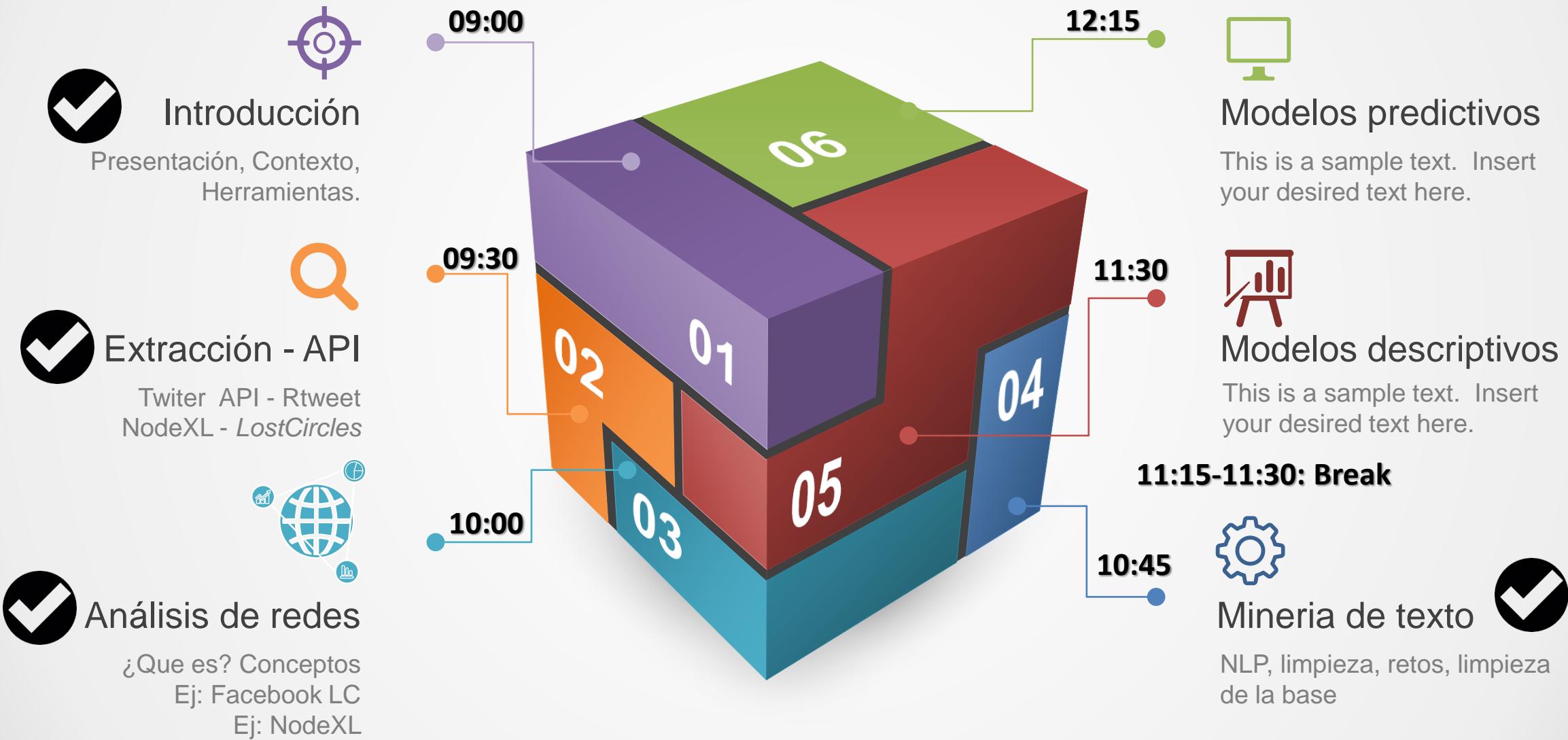
WordClouds

TEXT MINING

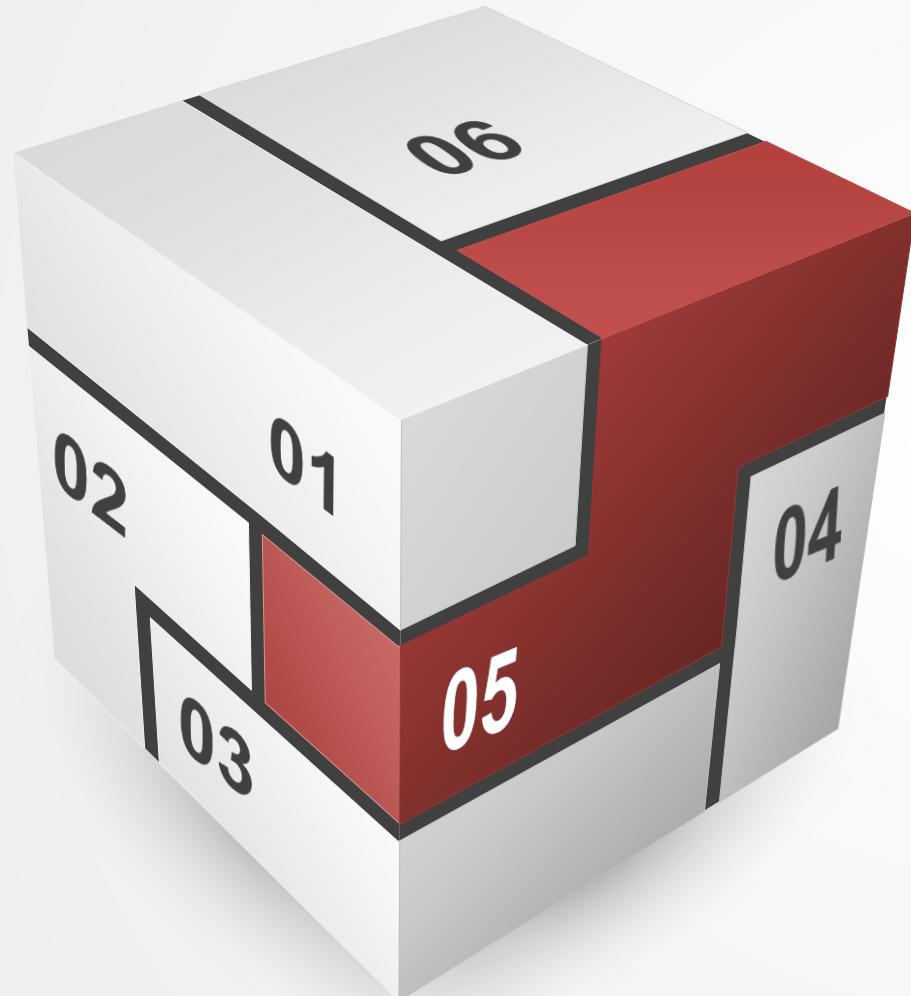


Script 2 - TextMining

Analítica con redes sociales - Twitter



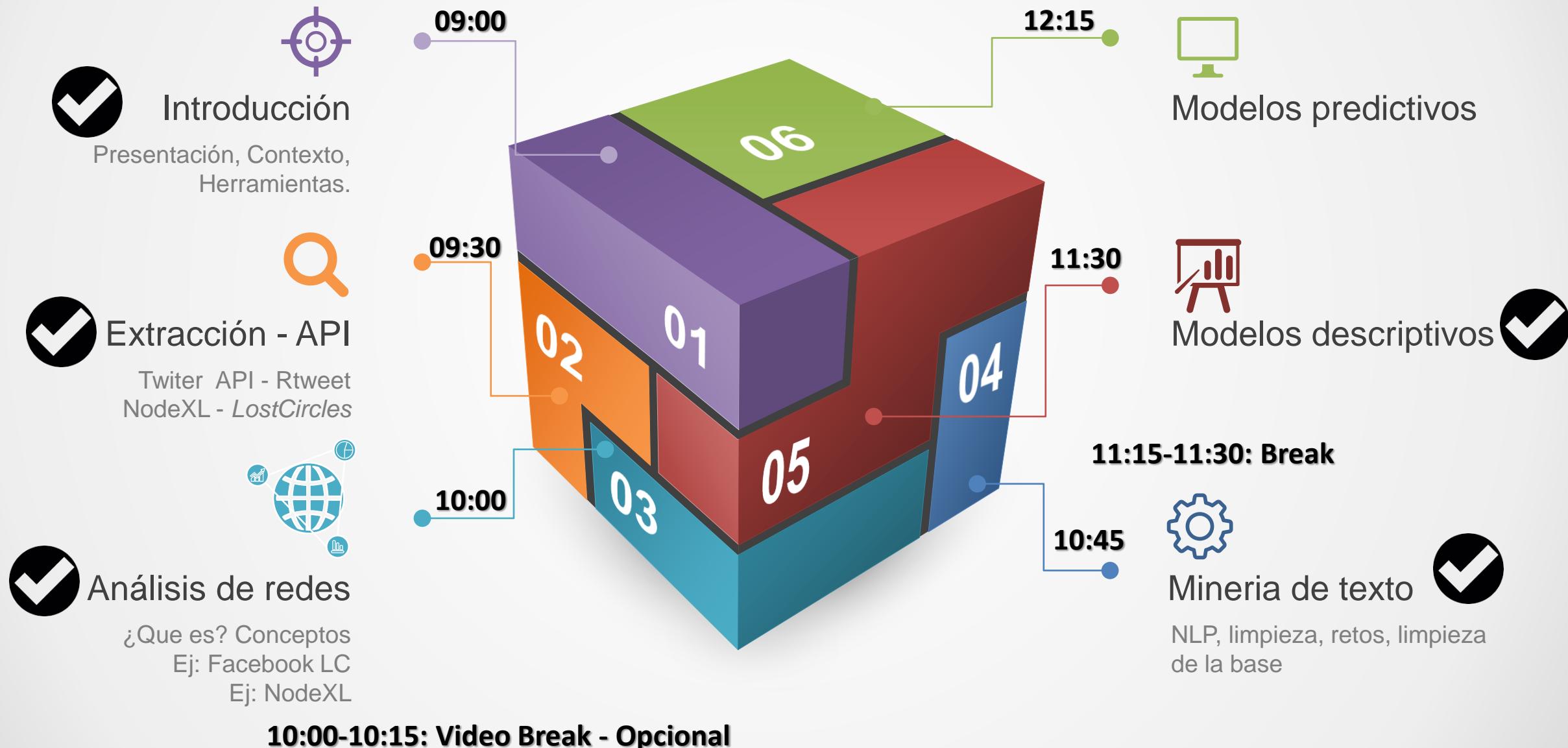
Analítica con redes sociales - Twitter



Modelos descriptivos

Script 3 - Descriptiva

Analítica con redes sociales - Twitter



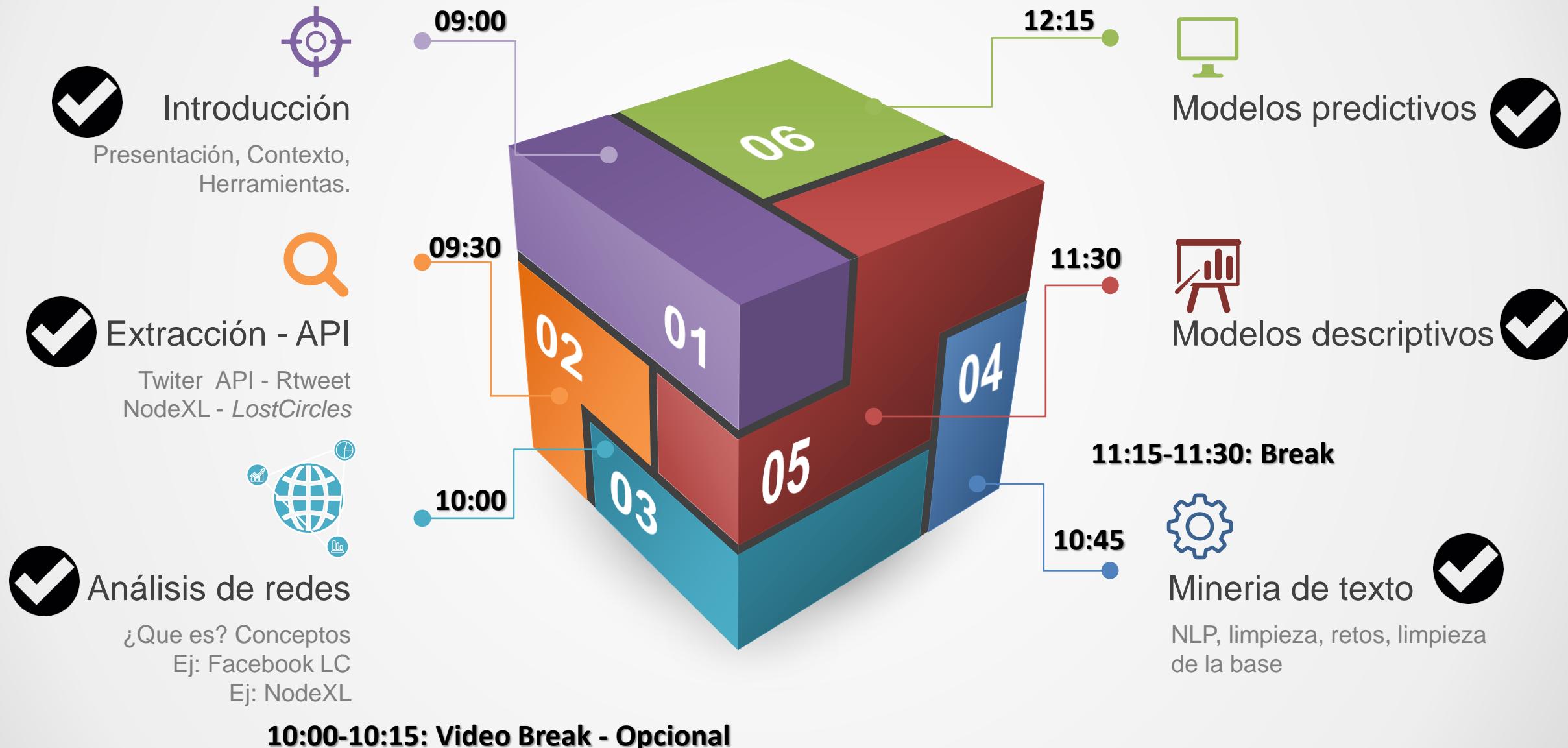
Analítica con redes sociales - Twitter



Modelos predictivos

Script 4 - Predictiva

Analítica con redes sociales - Twitter



Muchas gracias!!!



CeHuKa/AnaliticaRS_COL



cesar.huertask@gmail.com



César Augusto Huertas Kaleda



comunidad.colombia.ai

