



# Introducción a la Minería de Datos y Redes Sociales

**DRA. KARINA RUBY PÉREZ-DANIEL**

FACULTAD DE INGENIERÍA  
UNIVERSIDAD PANAMERICANA

ENERO - MARZO, 2020



UNIVERSIDAD  
PANAMERICANA

# CONTENIDO

- 1 INTRODUCCIÓN A LAS REDES SOCIALES
- 2 INTRODUCCIÓN A LOS GRAFOS
- 3 INTRODUCCIÓN A LAS BASES DE DATOS



UNIVERSIDAD  
PANAMERICANA

## SECCIÓN 1

# SOCIALES

## INTRODUCCIÓN A LAS REDES



UNIVERSIDAD  
PANAMERICANA

# LAS REDES SOCIALES Y LA INFORMACIÓN

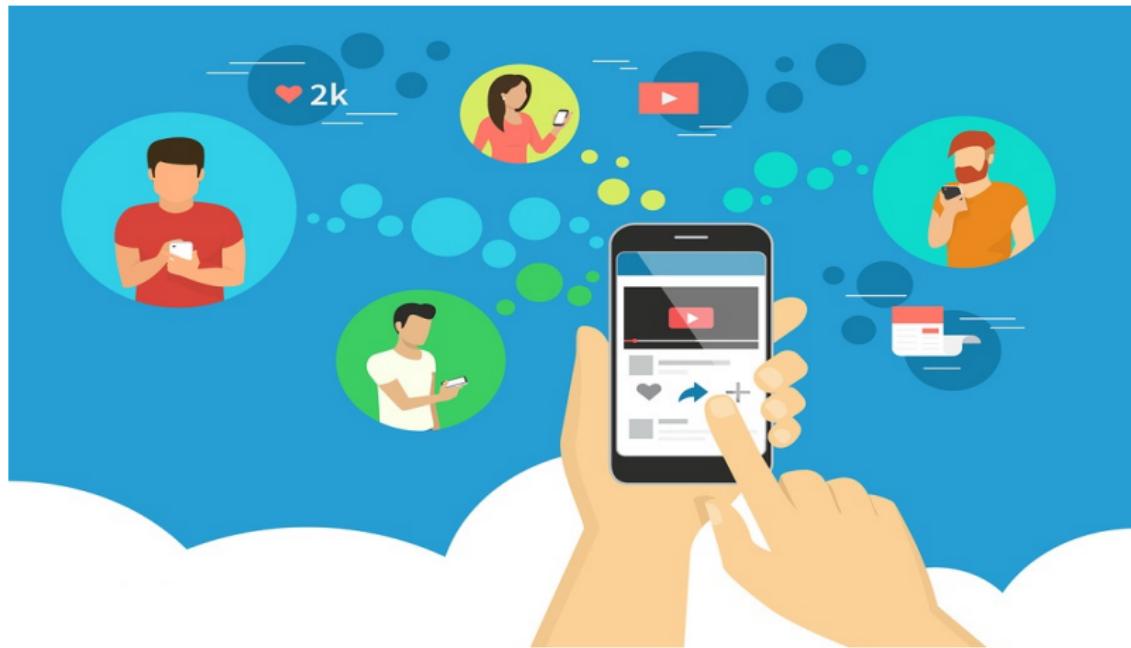


*Figura 1. Ejemplos de algunas Redes Sociales*



UNIVERSIDAD  
PANAMERICANA

# LAS REDES SOCIALES Y LA INFORMACIÓN



*Figura 2. Información compartida en Redes Sociales*



UNIVERSIDAD  
PANAMERICANA

# LAS REDES SOCIALES Y LA INFORMACIÓN

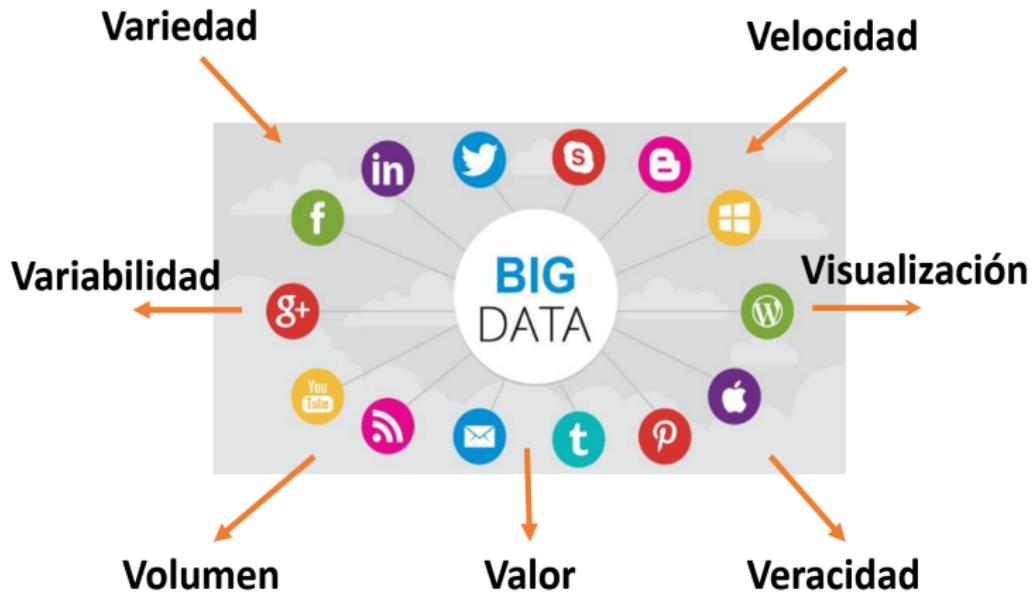


*Figura 3.* Cantidad de información generada en Internet en 60 segundos, 2019

Estadísticas en tiempo real: <https://www.internetlivestats.com/>



# LAS REDES SOCIALES Y LA INFORMACIÓN



*Figura 4. Esto es Big Data*



UNIVERSIDAD  
PANAMERICANA

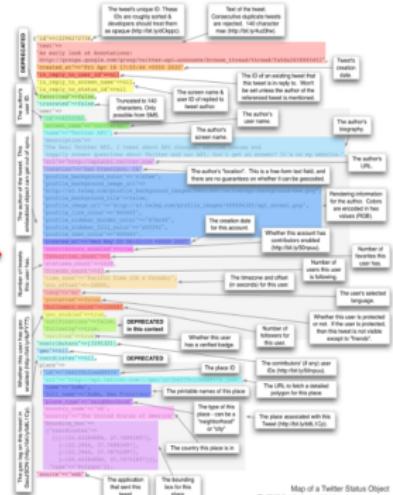
# LAS REDES SOCIALES Y LA INFORMACIÓN



Datos implícitos y explícitos de un tweet

- Metadatos
- Datos del usuario
- Texto

The figure consists of two side-by-side screenshots of the Twitter mobile application. The left screenshot shows the 'Your Profile' section with fields for Username (@myusername), Email (myemail@gmail.com), Phone (+1 415-555-0123), Account creation (Feb 15, 2007 at 4:33PM), Gender (Female), Age (26-34), Language (English, Spanish), and Profile location (San Francisco, CA). The right screenshot shows the 'Your Data' section with tabs for 'YOUR PROFILE' and 'YOUR DATA'. Under 'YOUR PROFILE', there are sections for Devices (4 logged-in devices, 7 total devices), Places you've been (3 locations), Interests from Twitter (14 interests), and Interests from partners (50 interests). Under 'YOUR DATA', there is a 'Request your data' button.



*Figura 5. Información explícita e implícita de un tweet*

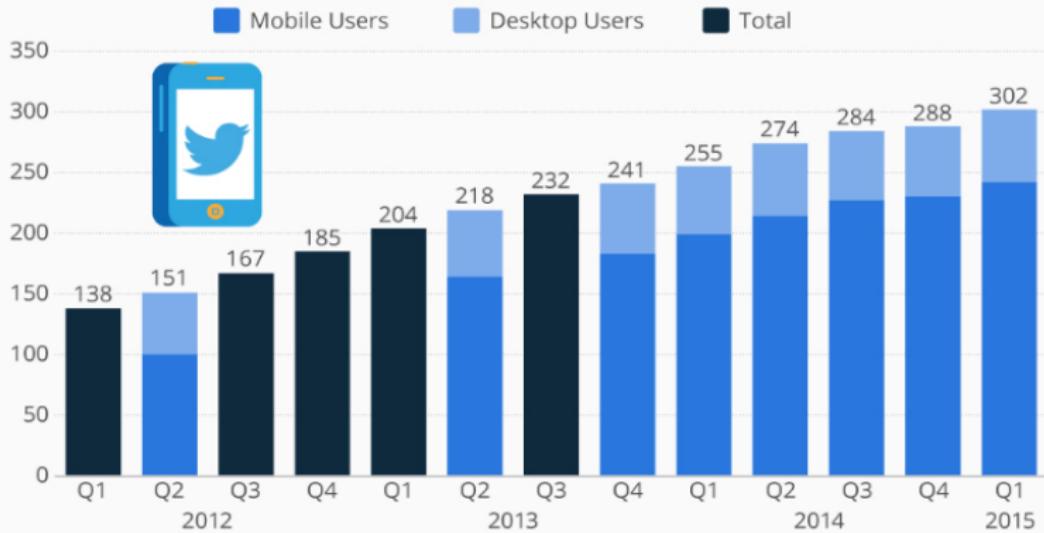


UNIVERSIDAD  
PANAMERICANA

# LAS REDES SOCIALES Y LA INFORMACIÓN

## 80% Of Twitter's Users Are Mobile

Number of monthly active Twitter users (in millions)\*



*Figura 6. Estadísticas a partir de datos obtenidos de tweeter*

**Los datos estan abiertos, por lo tanto, podemos analizarlos**

# LAS REDES SOCIALES Y LA INFORMACIÓN



*Figura 7. Los usuarios son lo importante*



UNIVERSIDAD  
PANAMERICANA

# LOS USUARIOS DE REDES SOCIALES



*Figura 8. Información Individual a partir de datos obtenidos de tweeter*

**Interesante pero los usuarios no estan solos**

# LOS USUARIOS Y SUS CONEXIONES



*Figura 9. Los usuarios y sus conexiones en redes sociales*

**Los usuarios publican para que la gente los vea**



UNIVERSIDAD  
PANAMERICANA

# GRAFO SOCIAL

## Grafo Social



*Figura 10. El grafo social: Los usuarios y sus conexiones en redes sociales*



UNIVERSIDAD  
PANAMERICANA

# LOS GRAFOS

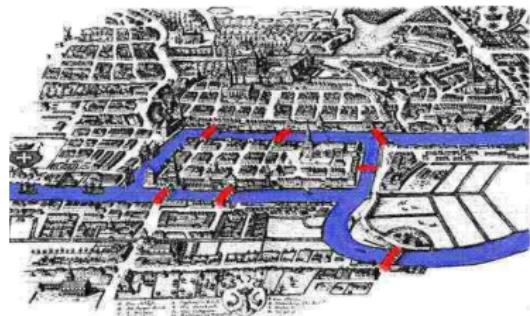


*Figura 11. Los grafos estan presentes en varios fenómenos*

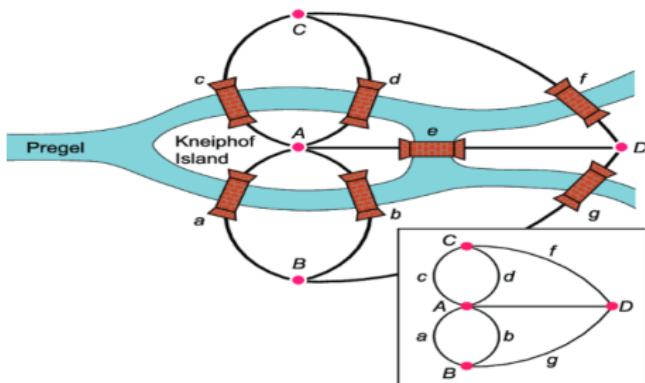


UNIVERSIDAD  
PANAMERICANA

# LOS PUENTES DE KÖNIGSBERG



*Figura 12. Los puentes de Königsberg*



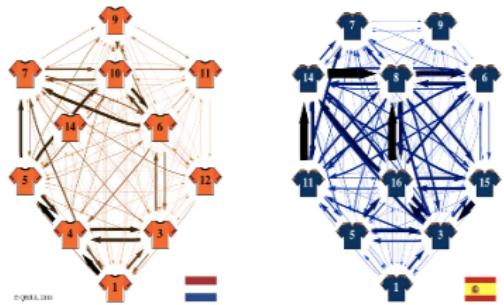
*Figura 13. Los puentes de Königsberg y la Teoría de Grafos*

**Euler Walk:** ¿Es posible recorrer toda la ciudad cruzando una y solo una vez cada uno de los puentes de Königsberg?



UNIVERSIDAD  
PANAMERICANA

# OTRAS APLICACIONES DE LOS GRAFOS



Spain and Netherlands pass maps from the 2010 World Cup Semifinals for cluster analysis

*Figura 14.* Análisis de partidos de Soccer



*Figura 15.* Descubrimiento de vínculos en redes de triangulación

¿Otros ejemplos?

## SECCIÓN 2

## INTRODUCCIÓN A LOS GRAFOS



UNIVERSIDAD  
PANAMERICANA

# ¿QUÉ ES UN GRAFO?

## Características de un grafo

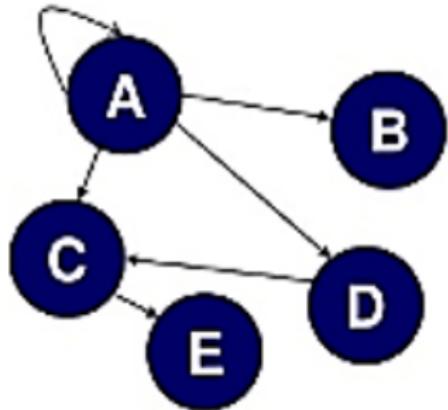
Un grafo esta compuesto por:

- Un conjunto de **nodos** ó **vértices**.
- Un conjunto de **enlaces** ó **aristas** que unen a los nodos en diversas formas.
- Si dos nodos son los extremos de un enlace entonces se dice que los **nodos son adyacentes**.



UNIVERSIDAD  
PANAMERICANA

# LOS GRAFOS



## Adyacencia del grafo

- B es adyacente a A,
- D es adyacente a A,
- C es adyacente a A,
- A es adyacente a A,
- C es adyacente a D,
- E es adyacente a C

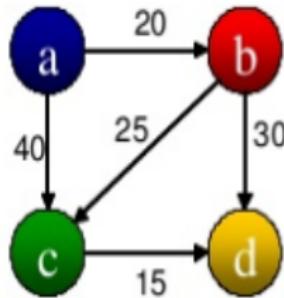
*Figura 16. Grafo*



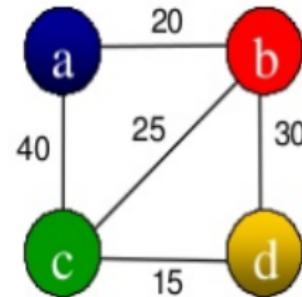
UNIVERSIDAD  
PANAMERICANA

# EL GRAFO ETIQUETADO

- El grafo etiquetado es aquel en el que cada **enlace** tiene asociada una **etiqueta**, la cual puede ser un *símbolo*, un *número*, o *texto*.



(a) Etiquetado y  
Dirigido



(b) Etiquetado y NO  
Dirigido

*Figura 17. Tipos de grafos etiquetados*

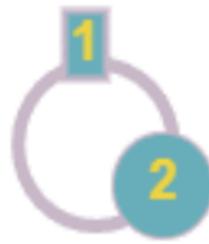


# VARIANTES DE UN GRAFO

- ① **Multigrafo:** Grafo que tiene varios enlaces entre 2 nodos adyacentes.
- ② **Pseudografo:** Grafo que posee enlaces cuyos extremos son el mismo nodo (lazos).
- ③ **Digrado:** Grafo dirigido (posee flechas) donde los extremos de los enlaces poseen un orden.



(a) Multigrafo



(b) Pseudografo



(c) Digrado

Figura 18. Tipos de grafos

**NOTA: Todas estas variantes se pueden combinar.**

# GRADO DE UN NODO

- 1 El **grado de un nodo**: Es el número de enlaces que tiene un nodo.
- 2 Un **Grafo regular**: Es aquel en el que todos sus nodos tienen el mismo grado.

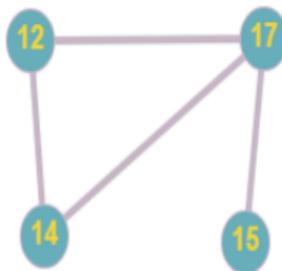


Figura 19. Grafo

## Grado del Nodo

- 12: 2,
- 14: 2,
- 17: 3,
- 15: 1



# CAMINOS EN UN GRAFO

Un camino es una sucesión finita de nodos y enlaces.

- ① **Grafo conexo:** Si para cada par de nodos existe un camino entre ellos (El grado de todos los nodos es  $> 0$ ).
- ② **Longitud:** Número de enlaces de un camino.
- ③ **Ciclo:** Camino cerrado en el que coinciden el primer y el último nodo.
- ④ **Círculo:** Ciclo en el que no se repiten enlaces.



UNIVERSIDAD  
PANAMERICANA

# ÁRBOL

Un árbol es un digrafo conexo sin ciclos y entre cada par de nodos existe un único camino.

- ① **Nodo raíz:** Nodo del que parten todos los caminos.
- ② **Nodo interno:** Nodo intermedio de un camino.
- ③ **Nodo hoja:** Nodo donde termina un camino.
- ④ **Frontera:** Conjunto de todos los nodos hoja de un árbol.

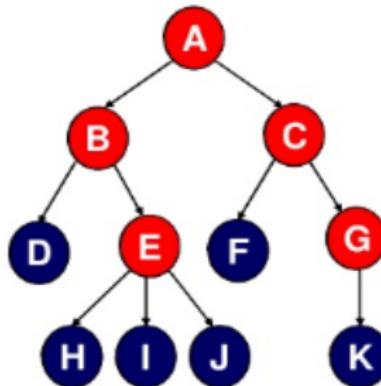
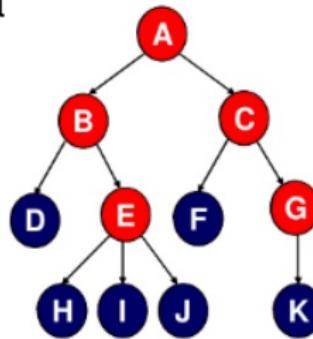


Figura 20. El árbol



# JERARQUÍA DENTRO DE UN ÁRBOL

- ① **Nodos hijos**: Nodo del que parten todos los caminos.
  - **Ramificación**: Número de hijos de un nodo.
  - **Expandir Nodo**: Crear y asignar nodos hijos.
- ② **Nodo padre**: Nodo origen del enlace que llega hasta un nodo.
- ③ **Nodos ascendentes**: Todos los nodos que hay desde un nodo hasta el nodo raíz.
- ④ **Nodos descendentes**: Todos los nodos que hay desde un nodo hasta sus nodos hoja



*Figura 21. El árbol*



## SECCIÓN 3

# DATOS

## INTRODUCCIÓN A LAS BASES DE



UNIVERSIDAD  
PANAMERICANA

# BASES DE DATOS RELACIONALES VS BASES DE DATOS NO RELACIONALES

- ① **Relacionales:** SQL Server, MariaDB, Oracle, PostgreSQL, MySQL, etc.
- ② **NO Relacionales:** Cassandra, Elasticsearch, Dynamodb, Neo4j, MongoDB, Firestore, bigQuery, etc.

## Bases de Datos NO relacionales

Es la familia que engloba a todas aquellas bases de datos que no son basadas en SQL.

Existen varios tipos/clases de bases de datos No relacionales. Las más populares son:

- Clave-valor,
- Documentos (tipo JSON),
- Grafos,
- En memoria,
- Optimizadas para búsquedas.

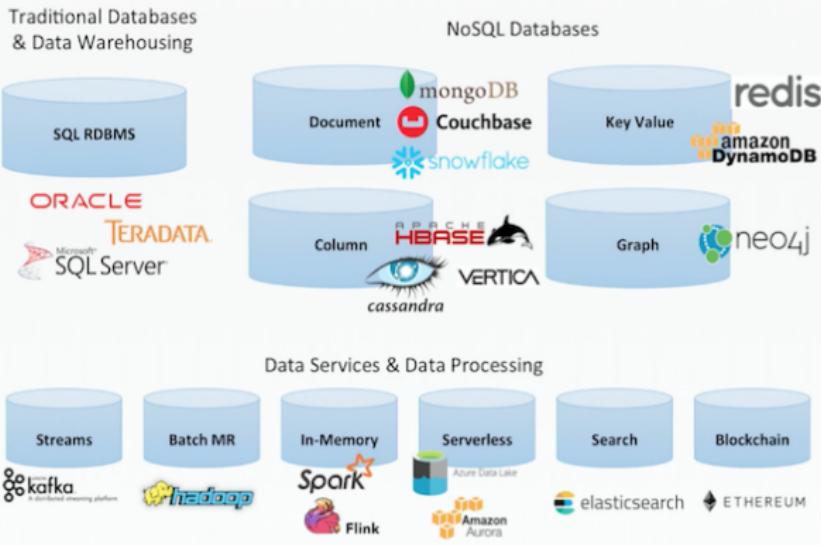
# TIPOS DE BASES DE DATOS NO RELACIONALES

- ① **Clave-Valor:** Cassandra, DynamoDB. Son ideales para almacenar y extraer grandes cantidades de datos con una clave única de forma rápida.
- ② **Documentos:** MongoDB, Firestore. Se les llama documentos a objetos de tipo JSON.
- ③ **Grafos:** Neo4j, TITAN, OrientDB, InfiniteGraph, InfoGrid, HyperGraphDB. Son basadas en grafos. Ideales para almacenar relaciones complejas y son muy usadas en IA.
- ④ **En Memoria:** Memcached, Redis. Son muy rápidas, pero son limitadas.
- ⑤ **Optimizadas para búsquedas:** BigQuery, Elasticsearch. Son optimizadas para búsquedas. Ideales para búsquedas complejas de manera sencilla.



# VISIÓN GENERAL DE LAS BASES DE DATOS

## The Database Landscape is Changing



*Figura 22. Visión general de las bases de datos*



UNIVERSIDAD  
PANAMERICANA

# Gracias por su atención

*Fin*



UNIVERSIDAD  
PANAMERICANA