

# Redes Complejas

Inteligencia Colectiva y Sistemas de Recomendación

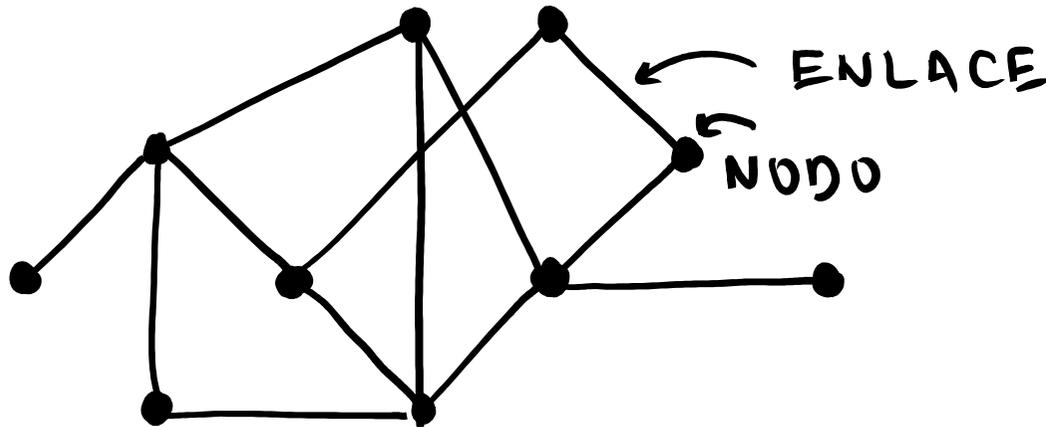
Master Propio en Data Science y Big Data

# ¿Qué es una red?

Conjunto de puntos, algunos de ellos conectados mediante enlaces.

Los **puntos** se llaman también nodos o vértices.

Los **enlaces** se llaman lados o aristas.



# Redes y sistemas complejos

- Muchos sistemas complejos pueden representarse de forma esquemática mediante una red:
  - Componentes del sistema = nodos.
  - Interacciones = conexiones.
- Los sistemas complejos suelen evolucionar en el tiempo, son dinámicos: no siempre vienen representados mediante la misma red, hay cambios en número de nodos y en conexiones a lo largo del tiempo.

# Redes y sistemas complejos

- Se puede estudiar:
  - Naturaleza de los nodos.
  - Naturaleza de las conexiones entre nodos.
  - **Patrón de conexiones.**
- Resulta tan crucial como los otros dos puntos para entender el funcionamiento del sistema complejo.
- Aunque el sistema complejo evolucione, y con él la red que lo representa, el patrón de conexiones suele ser estable.

# Redes y sistemas complejos: antecedentes

- Los sociólogos fueron los primeros en considerar el estudio de redes para aplicaciones reales:
  - Personas = nodos, interacciones entre ellas = conexiones.
  - Estudio de patrones de formación de conexiones para comprender la sociedad.
  - Utilización de encuestas para recoger datos y representar la red.
  - Estudios limitados a pequeñas redes.

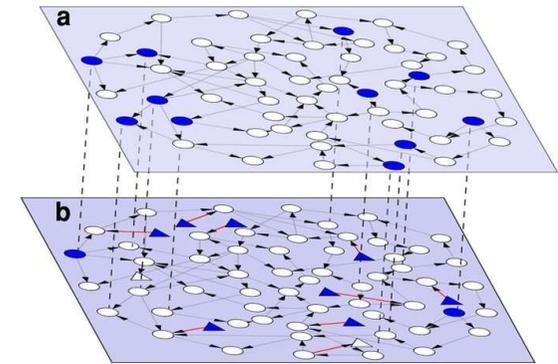
# Redes y sistemas complejos: nueva visión

- **Redes enormes** con millones de nodos.
- Preguntas tradicionales sin interés:
  - Antes: ¿Qué ocurre si quito un nodo?
  - Ahora: ¿Qué porcentaje de nodos debo quitar para que afecte a la funcionalidad de la red?
- El interés pasa de estudiar nodos fijos a considerar **propiedades estadísticas y condiciones globales.**
- Imposibilidad de representar la red.



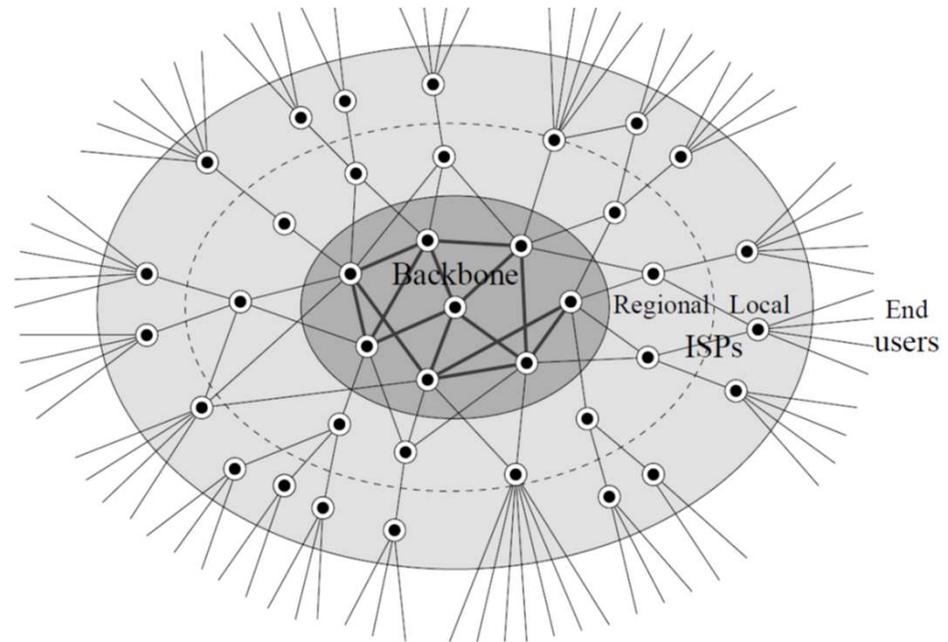
# Redes y sistemas complejos: nueva visión

- ¿Qué “forma” puede tener la red aunque no podamos verla gráficamente?
- ¿Cómo clasificar y analizar distintos tipos de redes?
- Buscamos predicciones de comportamiento de sistemas complejos basados en la medición de propiedades estructurales de la red y de reglas locales que afectan al comportamiento de los nodos.
- Modelización matemática y computacional.



# Ejemplos de redes en el mundo real

- Redes de información:
  - World Wide Web: hyperlinks
  - Redes de citación
  - Redes de Noticias y Blogs
- Redes tecnológicas:
  - Energéticas
  - Transporte (aéreo, carreteras, fluviales,...)
  - Telefónicas
  - Internet
  - Sistemas Autónomos

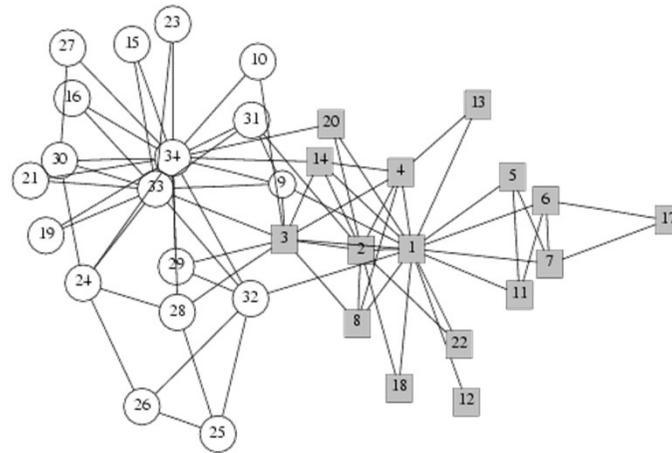


Esquema de la red de internet



# Ejemplos de redes en el mundo real

- Redes sociales



Karate club network

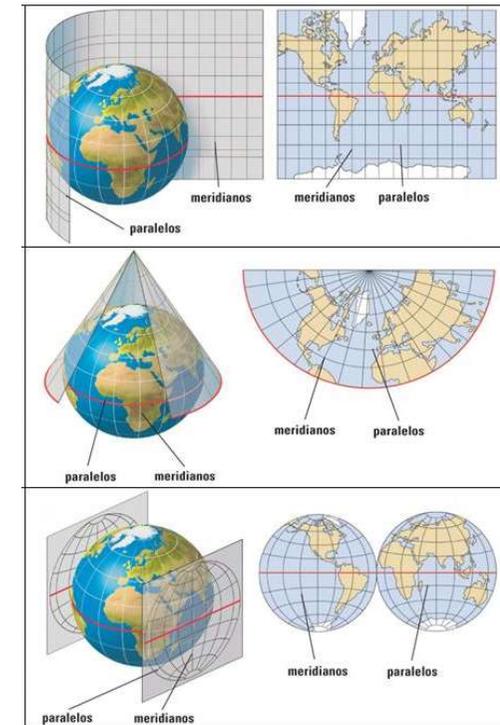


Redes de amistad

# Modelo de Representación unificado: Teoría de Grafos

**Grafo:** marco general de representación abstracta.

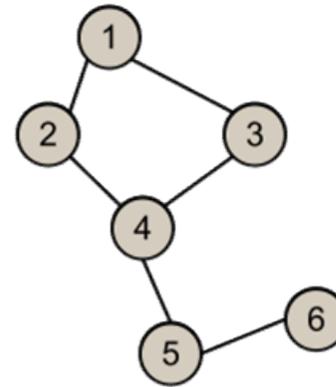
- Flexibilidad en la representación: permite proyectar el fenómeno real en multitud de posibles grafos, cada uno pudiendo resaltar una “visión” particular de la realidad.
- Robustez del modelo matemático subyacente:
  - Posee resultados de gran potencia.
  - Permite combinarlos con otras teorías matemáticas de gran potencia: probabilidad, computación.



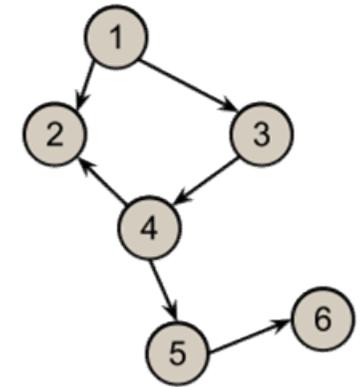
# Introducción a la Teoría de Grafos

# Fundamentos de Teoría de Grafos

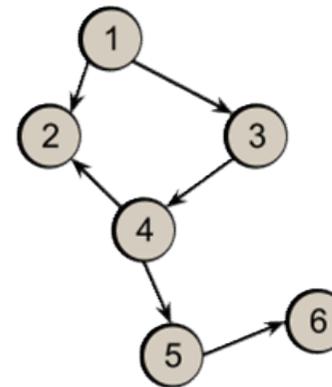
- Elementos sustanciales:
  - Nodos
  - Aristas.
- Aristas dirigidas o no dirigidas.
- Información (o no) en los nodos y/o aristas: pesos.



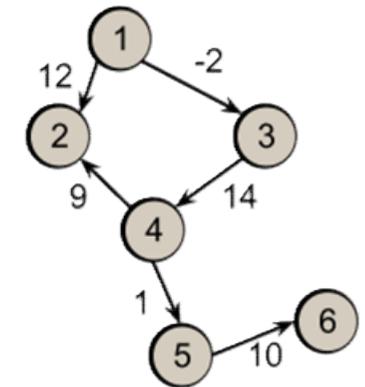
Undirected



Directed



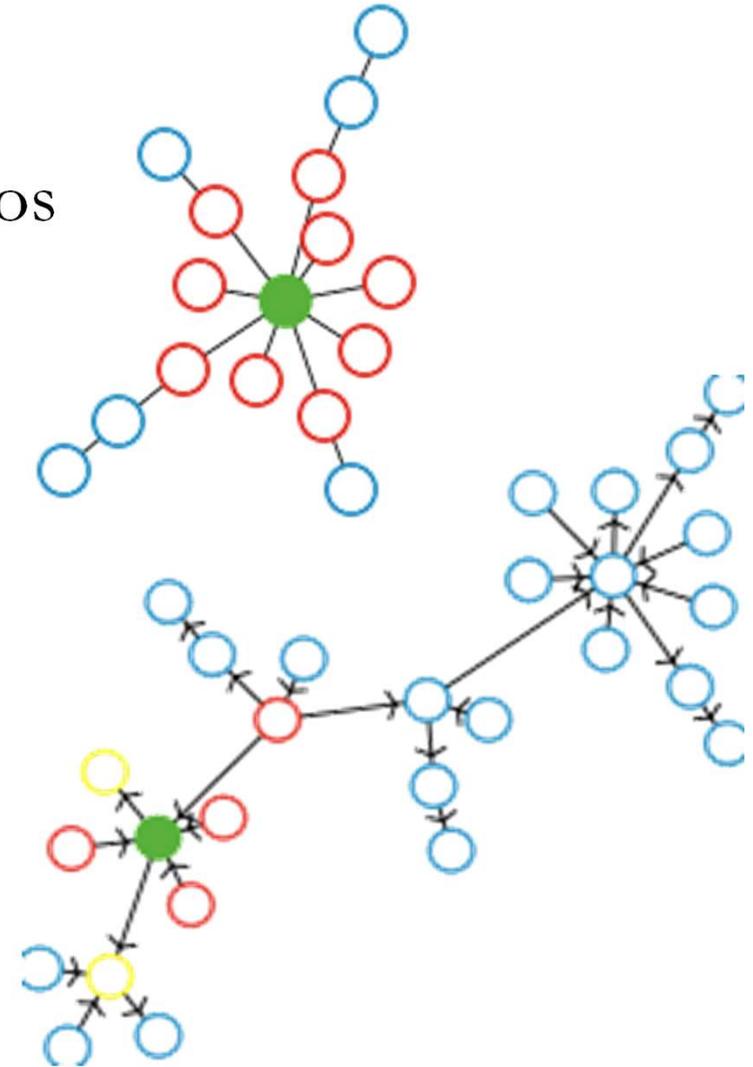
Unweighted



Weighted

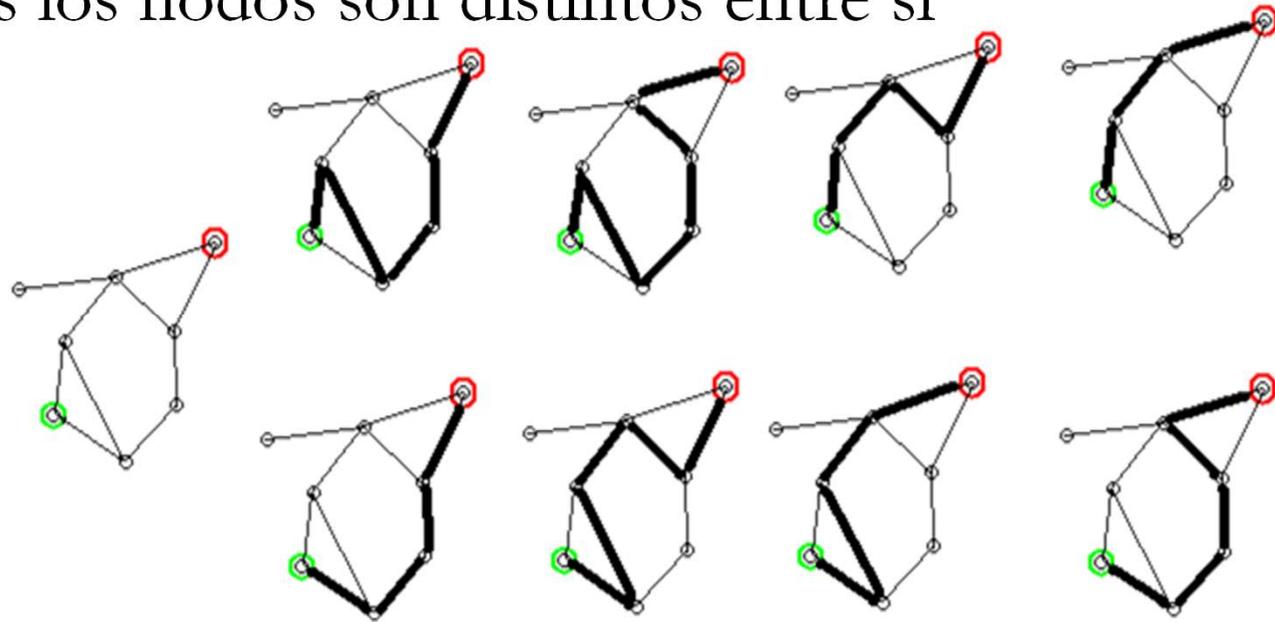
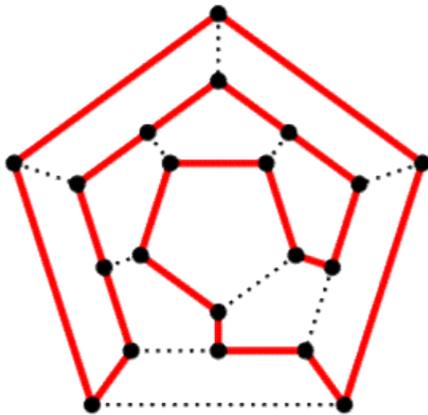
# Vecindarios

- Vecindario de un nodo: conjunto de nodos conectados con él:
- Si el grafo es dirigido, se puede hablar de vecindarios de entrada/salida:



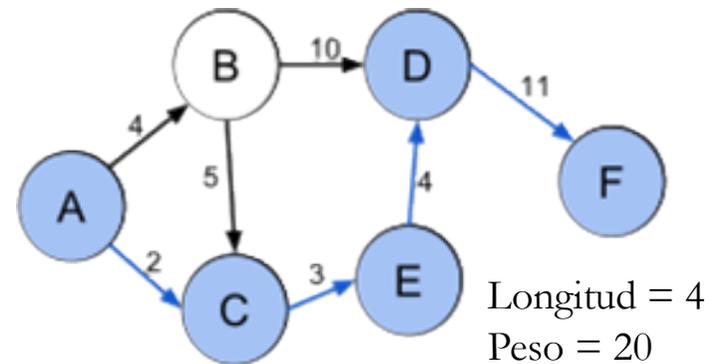
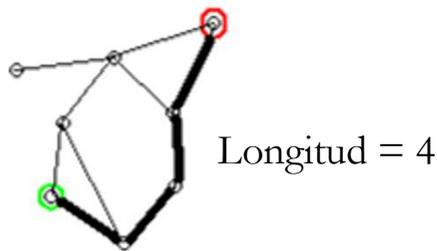
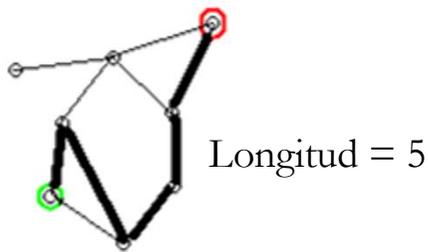
# Camino y Ciclos

- **Camino** conectando nodo A con nodo B:
  - Sucesión de nodos que comienza en A y acaba en B, y donde cada par de nodos consecutivos está conectado por una arista
- **Camino Simple**, si todos los nodos son distintos entre sí
- **Ciclo**, si  $A = B$



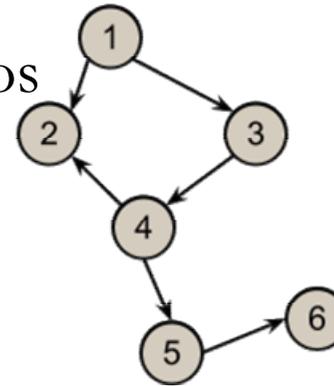
# Caminos y Ciclos

- **Longitud:** número de aristas del camino
- **Peso** (en grafos con pesos): suma de los pesos de las aristas del camino

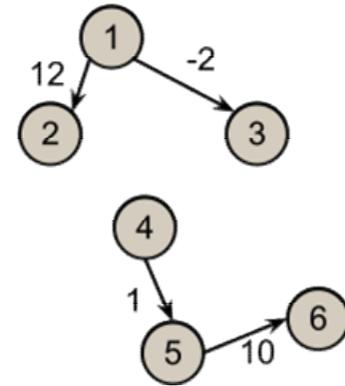


# Conectividad

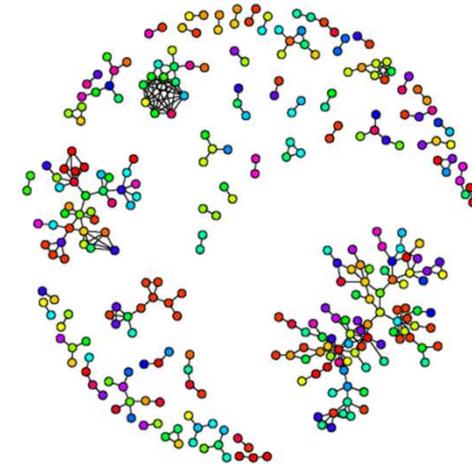
- **Conexo:** Si todo par de nodos están conectados
- **Componente Conexo:** conjunto maximal conexo de nodos
- En grafos dirigidos: fuertemente conexa.
  - **Componente de entrada:** nodos que pueden alcanzar la componente conexa, pero no pueden ser alcanzados desde ella.
  - **Componente de salida:** el recíproco.



Connected

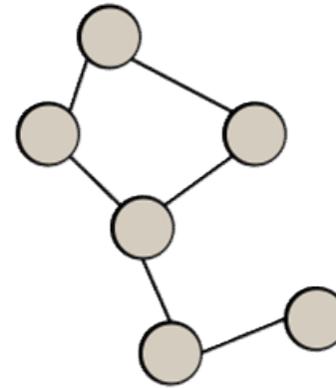


Disconnected

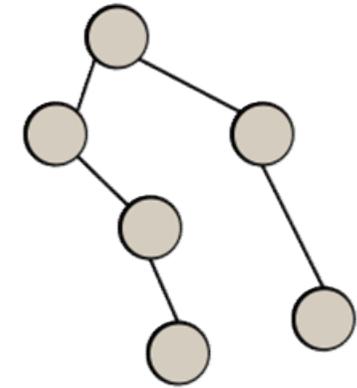


# Árboles y Bosques

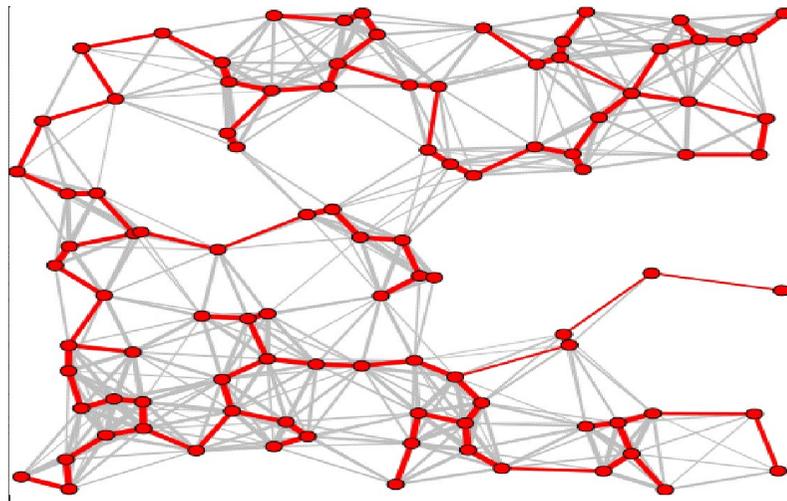
- Grafos sin ciclos
- Conexo : **Árbol**
- Colección de árboles: **Bosque**



Graph



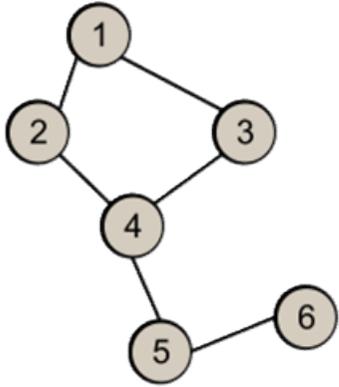
Tree



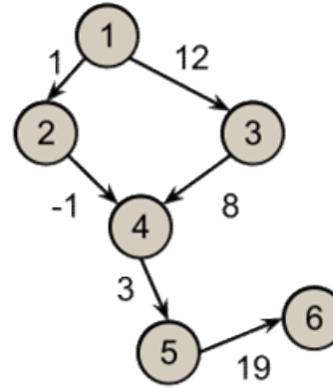
Árbol Maximal de un grafo: árbol que contiene todos los nodos del grafo original.

# Representación de Grafos

Matriz Adyacencias

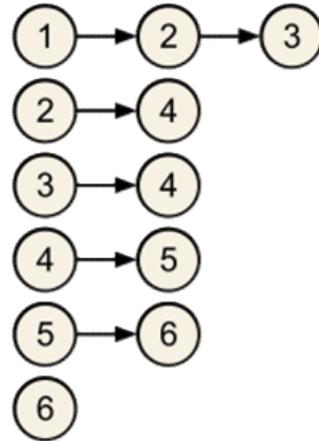
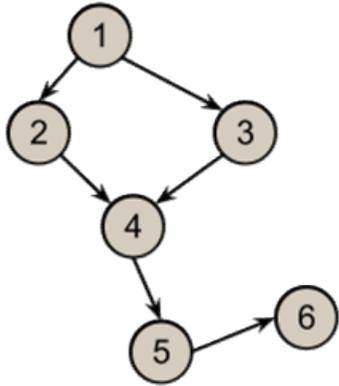


	1	2	3	4	5	6
1	0	1	1	0	0	0
2	1	0	0	1	0	0
3	1	0	0	1	0	0
4	0	1	1	0	1	0
5	0	0	0	1	0	1
6	0	0	0	0	1	0



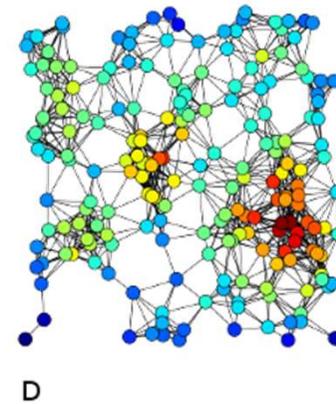
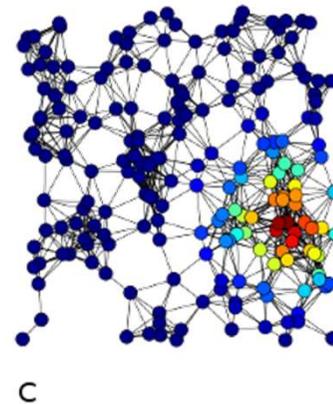
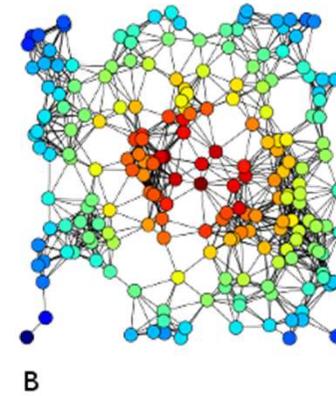
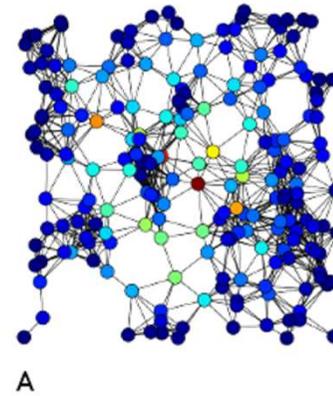
	1	2	3	4	5	6
1	0	1	12	0	0	0
2	-1	0	0	-1	0	0
3	-12	0	0	8	0	0
4	0	1	-8	0	3	0
5	0	0	0	-3	0	19
6	0	0	0	0	-19	0

Listas Adyacencias



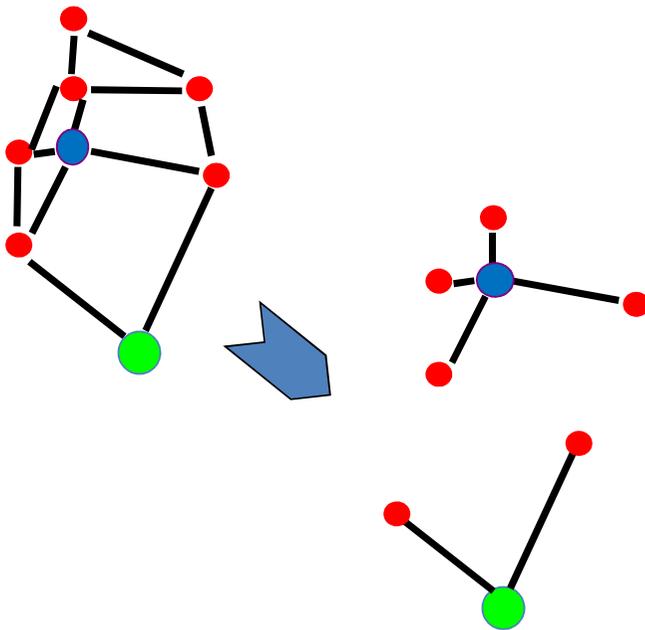
# Medidas usuales en Teoría de Grafos

- A. Centralidad Betweenness (intermediación) por nodos o aristas.
- B. Centralidad Closeness (cercanía)
- C. Centralidad Eigenvector (Autovalor)
- D. Grado y Distribución de Grados.
- E. Coeficiente de Clustering (agrupación).
- F. Conectividad: caminos aleatorios y distancia.



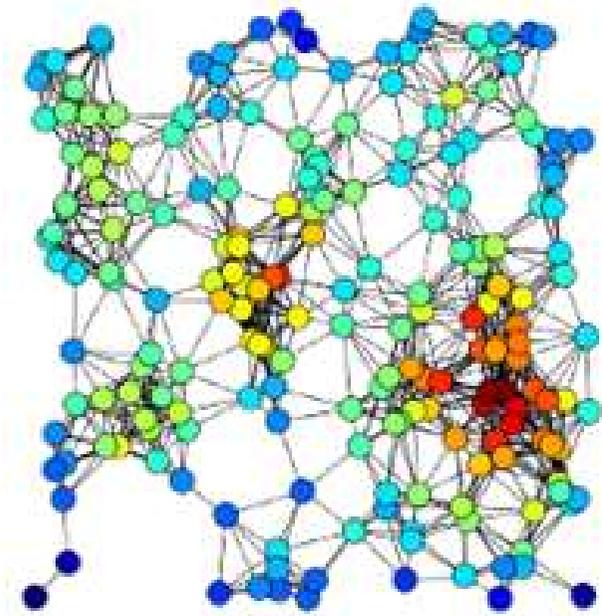
# Medidas: Grado y Distribuciones de Grados

- **Grado de un nodo:** número de nodos conectados a él. Lo denotaremos por  $k$



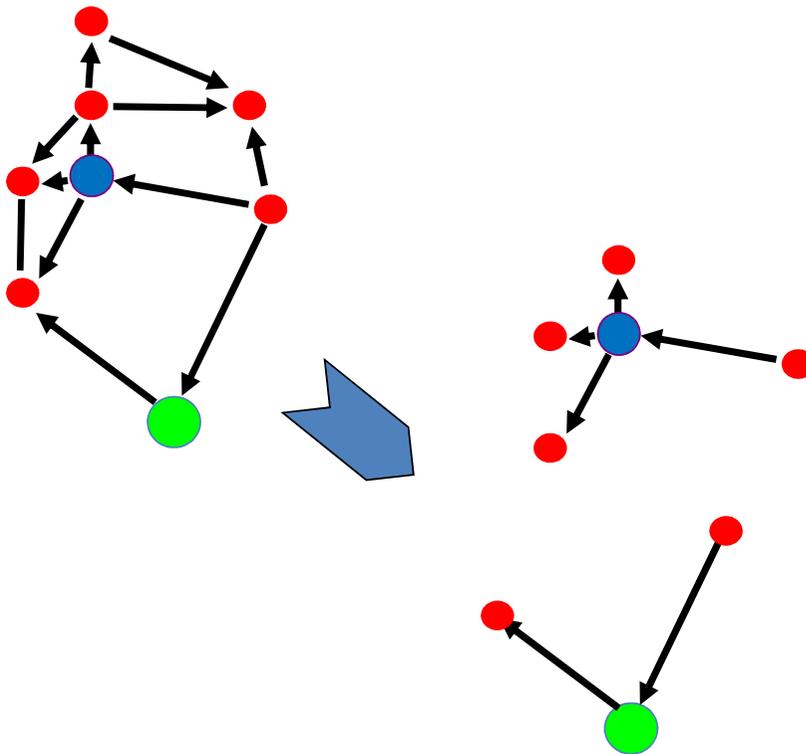
$$k(\bullet) = 4$$

$$k(\bullet) = 2$$



# Medidas: Grado y Distribuciones de Grados

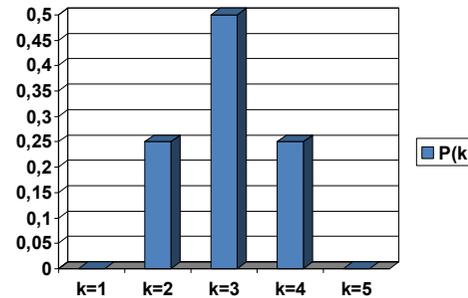
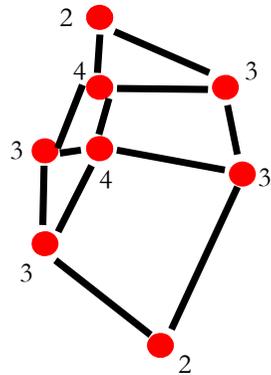
- En el caso dirigido se distingue entre **grado entrante** y **grado saliente**:  $k_{in}$ ,  $k_{out}$



$$\begin{aligned}k_{in}(\text{blue}) &= 3 \\k_{out}(\text{blue}) &= 3 \\k_{in}(\text{green}) &= 1 \\k_{out}(\text{green}) &= 1\end{aligned}$$

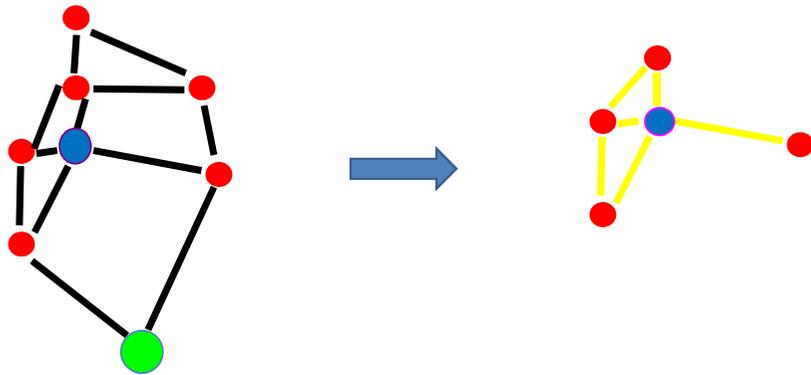
# Medidas: Grado y Distribuciones de Grados

- **Distribución de Grados:**  $P(k)$ , probabilidad de que un nodo tenga grado  $k$ .



# Medidas: Coeficiente de Clustering o de Transitividad

- Probabilidad de que dos nodos vecinos a uno dado, sean vecinos entre sí.



$$E_i = 2$$

$$k_i = 4$$

$$C_i = \frac{2}{\frac{1}{2}4 \cdot 3} = \frac{1}{3}$$

$$C_i = \frac{E_i}{\frac{1}{2}k_i(k_i - 1)}$$

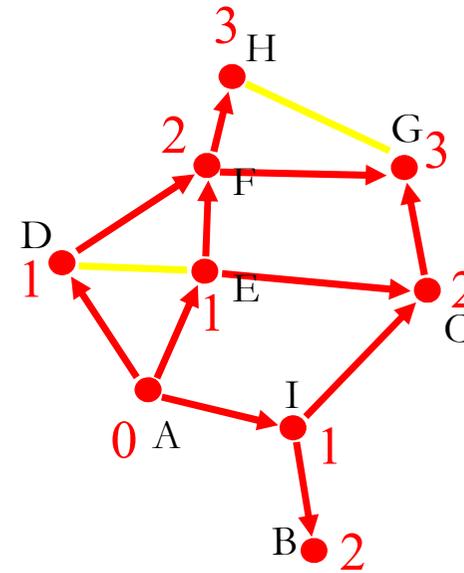
donde  $E_i$  es el número de aristas que conectan entre sí los nodos adyacentes al nodo  $i$ .

- De igual forma, se trabaja con la distribución de clustering,  $C(k)$ .

# Medidas: distancia

- **Distancia entre dos nodos:** menor longitud de los caminos que los unen.
- **Diámetro:** mayor distancia.

		nodo $i$								
		A	B	C	D	E	F	G	H	I
nodo $j$	A	0	2	2	1	1	2	3	3	1
	B	2	0	2	3	3	4	3	4	1
	C	2	2	0	2	1	2	1	2	1
	D	1	3	2	0	1	1	2	2	2
	E	1	3	1	1	0	1	2	2	2
	F	2	4	2	1	1	0	1	1	3
	G	3	3	1	2	1	1	0	1	2
	H	3	4	2	2	2	1	1	0	3
	I	1	1	1	2	2	3	2	3	0

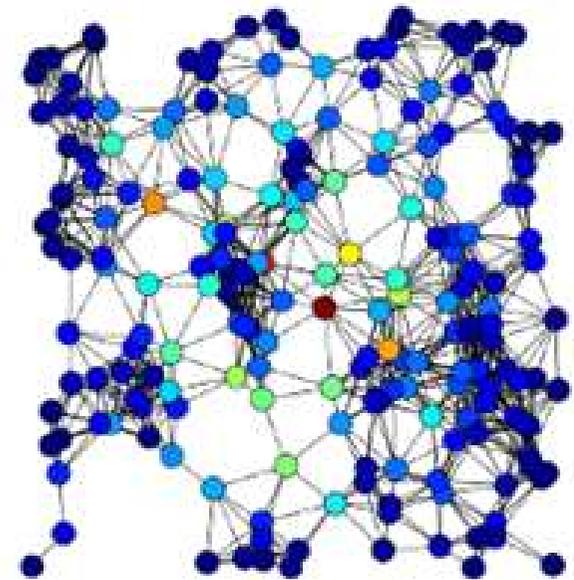


Distancia media:  $d = 1.94$

Diámetro: 4

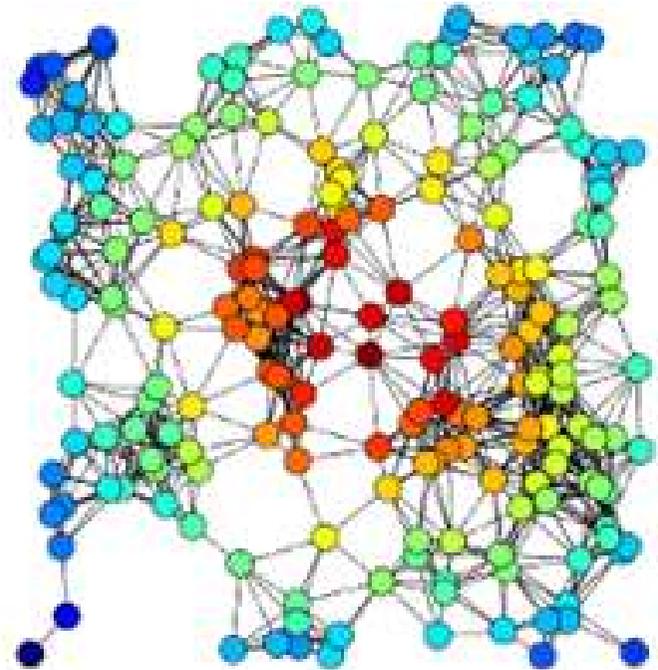
# Medidas: Betweenness, Carga o Intermediación

- **Carga del nodo  $u$ :** proporción de caminos (más cortos) que van de un nodo  $s$  a un nodo  $t$  pasando por el nodo  $u$ .
  - Dan idea de la conectividad relativa y de la capacidad de dirección de tráfico de un nodo.
  - Proporción de veces que el nodo actúa de puente.
- **Carga de una arista  $e$ :** proporción de caminos (más cortos) que van de un nodo  $s$  a un nodo  $t$  pasando por la arista  $e$ .



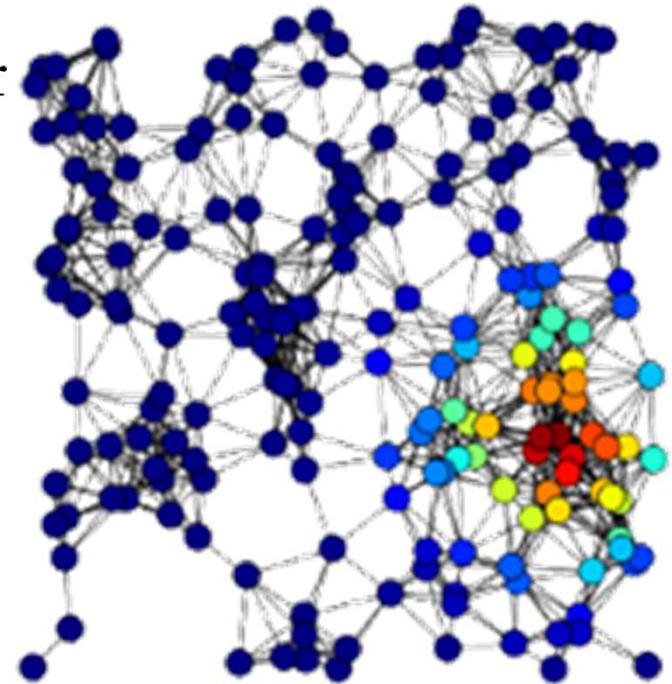
# Medidas: Closeness

- Se basa en calcular la suma o bien el promedio de las distancias más cortas desde un nodo hacia todos los demás.
- A veces se define como el recíproco, para evitar el problema de las distancia en los grafos no conexos.



# Medidas: Autovalor

- Mide la influencia de un nodo en el grafo, y corresponde con el **autovalor principal** de la matriz de adyacencia del grafo analizado.
- Intuitivamente, los nodos que poseen un valor alto de esta medida de centralidad están conectados a muchos nodos que a su vez están bien conectados (en este sentido)
- Por ejemplo, son buenos candidatos para difundir información, divulgar rumores o enfermedades, etc.



# Medidas: Modularidad

- Es una medida de la red completa, mide la fuerza que tiene la red para dividirse en **módulos (comunidades)**:
  - Redes con modularidad alta tienen gran densidad de conexiones entre nodos de los mismos módulos, pero baja entre nodos de diferentes comunidades.
- Se suele usar como **medida de optimización** para detectar comunidades en redes.
- Se ha comprobado que estas medidas pueden funcionar para detectar grandes comunidades, pero son incapaces de detectar comunidades pequeñas.

