

# **Introducción al Análisis de Redes Complejas**

Guillermo de Anda Jáuregui (INMEGEN/CONACYT/C3-UNAM)

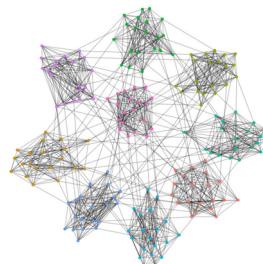
**II Escuela de Verano Modelación y Herramientas para el Análisis de Datos**

# Programa

- Sesión 1, parte 1:
  - Complejidad, ciencia de redes, y biología...
- Sesión 1, parte 2:
  - Tipos de redes
- Sesión 1, parte 3:
  - Análisis de redes: propiedades y métricas
- Sesión 2:
  - Comparaciones topológicas de redes (ejercicios).

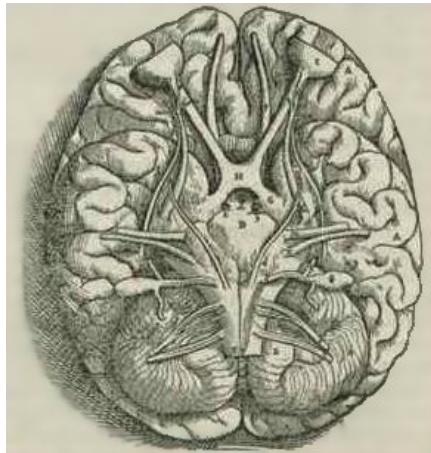
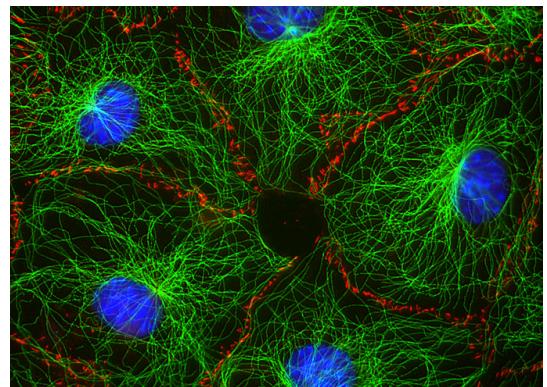
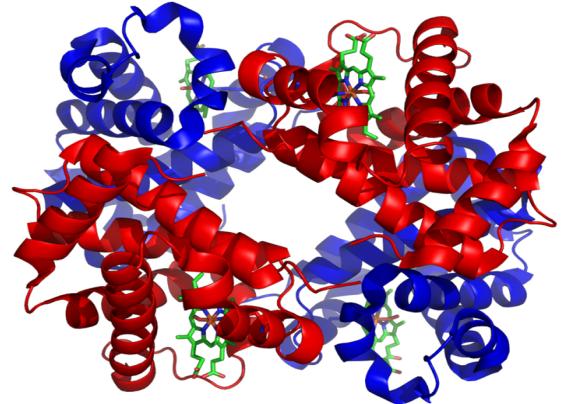
# Recursos

- <https://github.com/guillermodeandajaregui/EVHMAD2022>
  - Ahí encuentran estas láminas, guías de instalación, ejercicios, y los datos para los ejercicios.



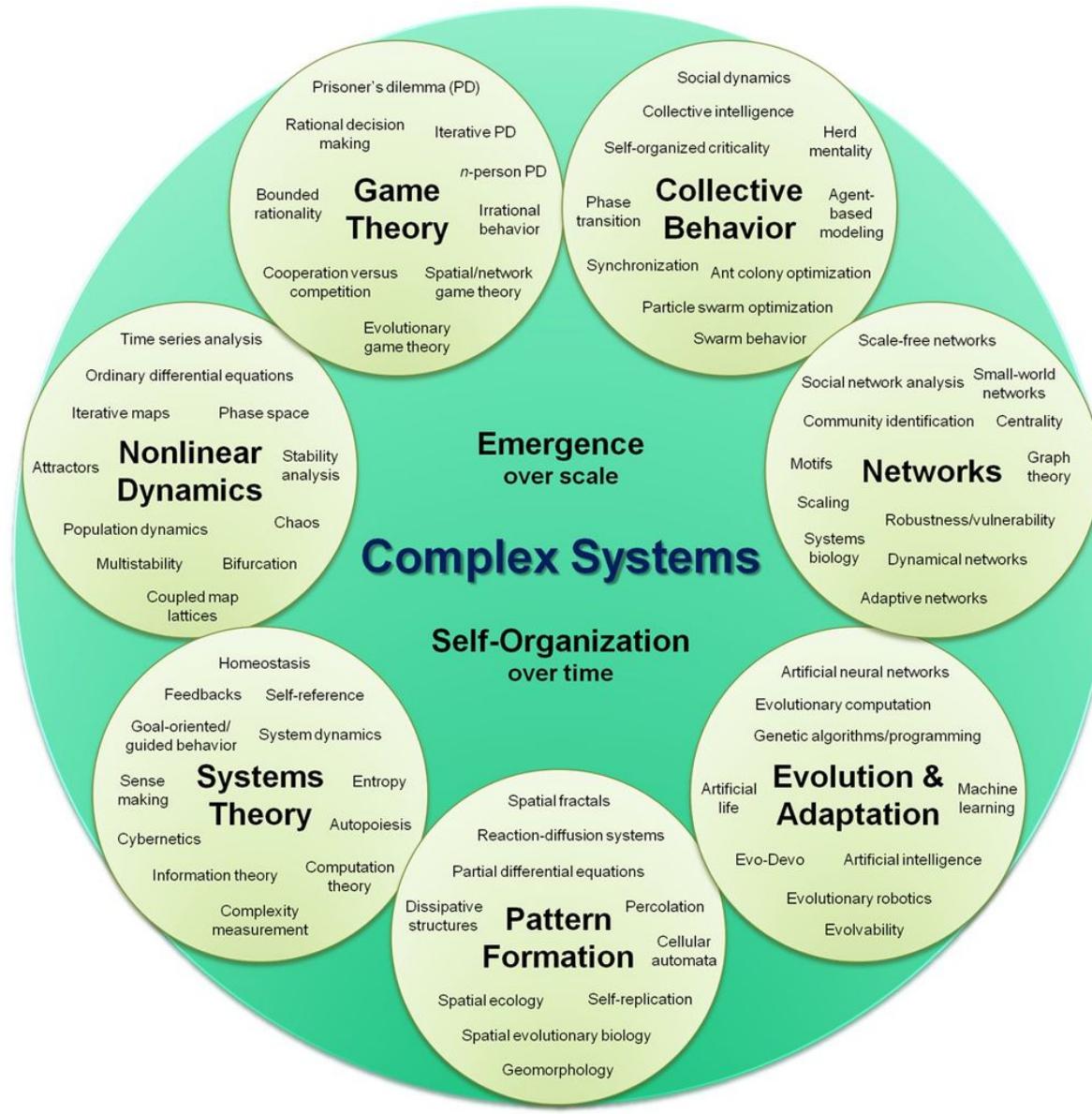
# Parte 1: Complejidad, Redes, y Biología ...

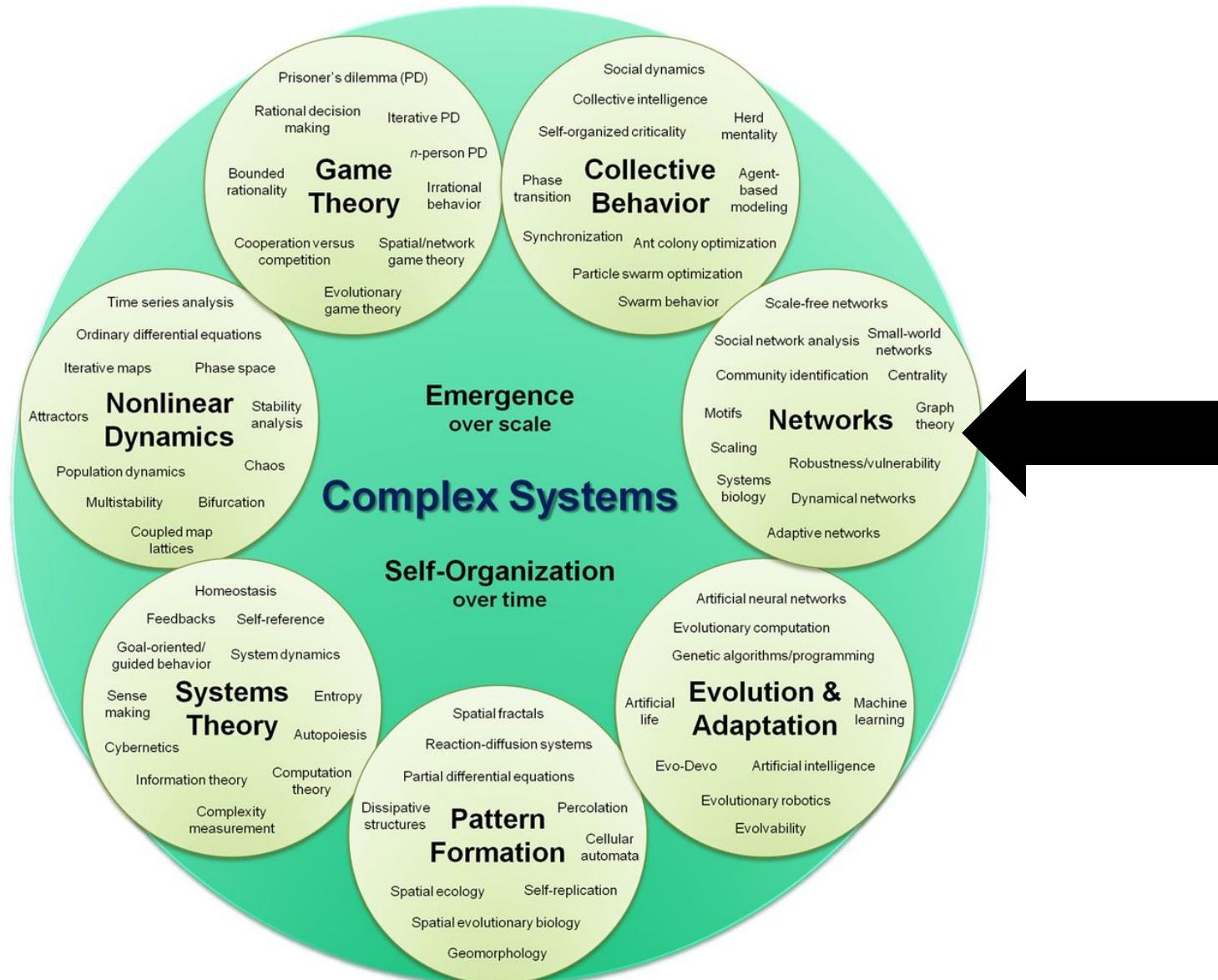
# La vida es compleja...



“No hay amor en un átomo de carbono,  
Ni huracán en una molécula de agua,  
Ni colapso financiero en un billete de dólar”

*Peter Dodds*



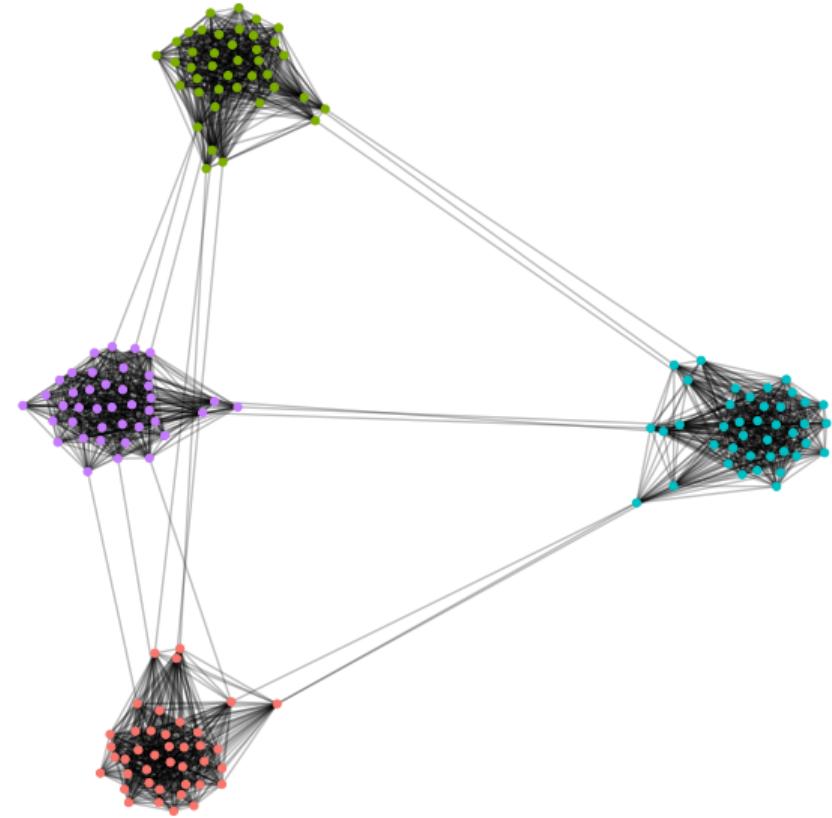


# Las redes están en todas partes

El concepto de redes es amplio y general.

Permite describir como se conectan las cosas.

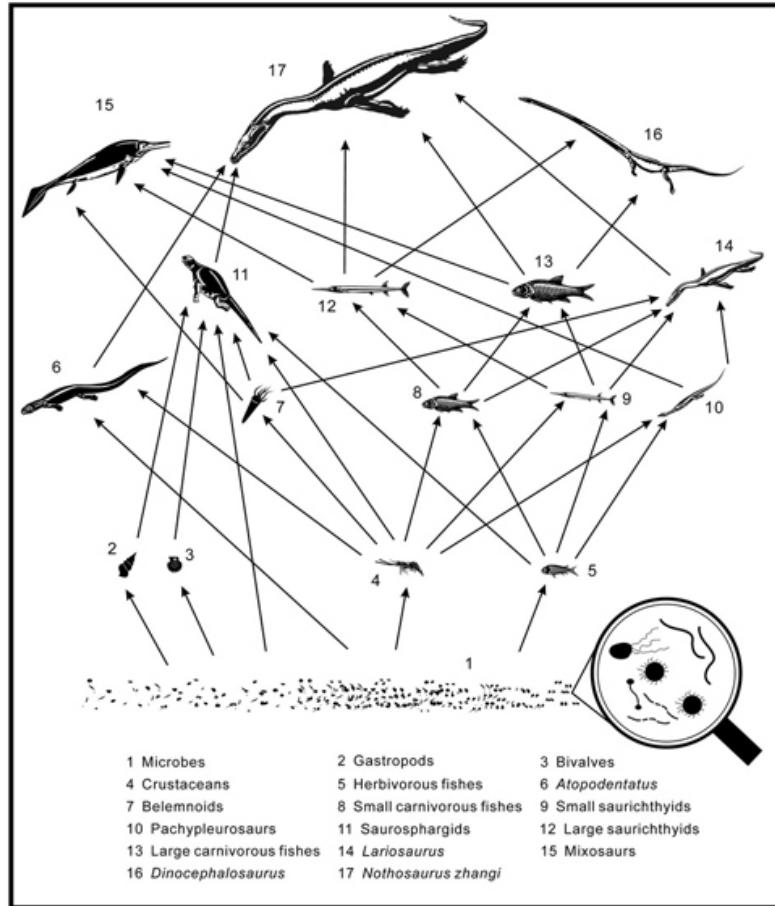
Muchos sistemas (complejos) pueden representarse como redes



Las redes están en todas partes

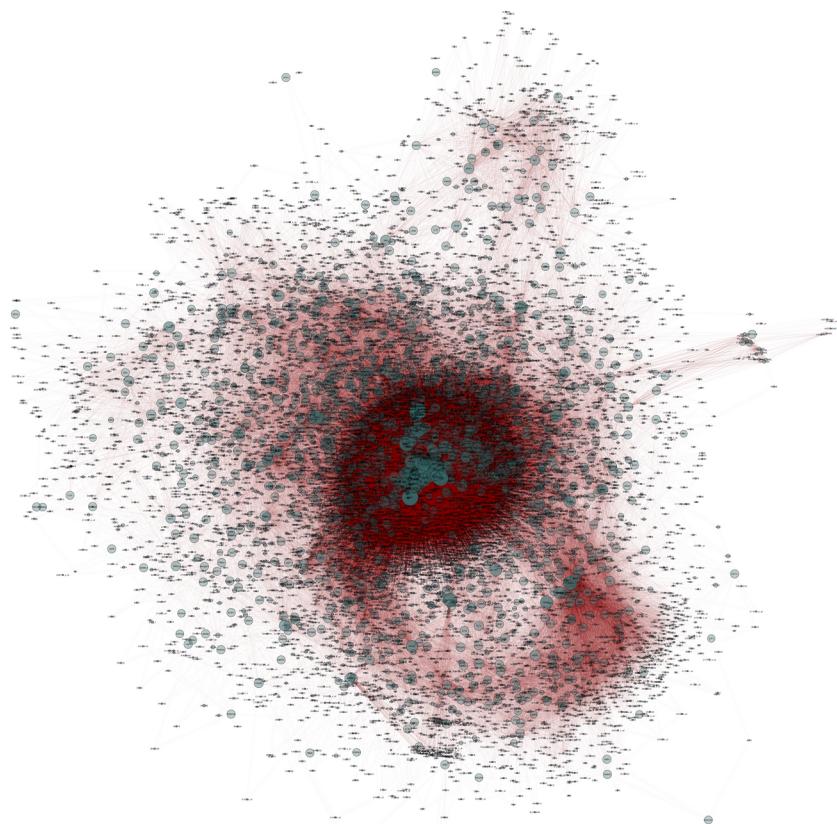
**Los sistemas  
biológicos  
forman redes**

# Los sistemas biológicos forman redes



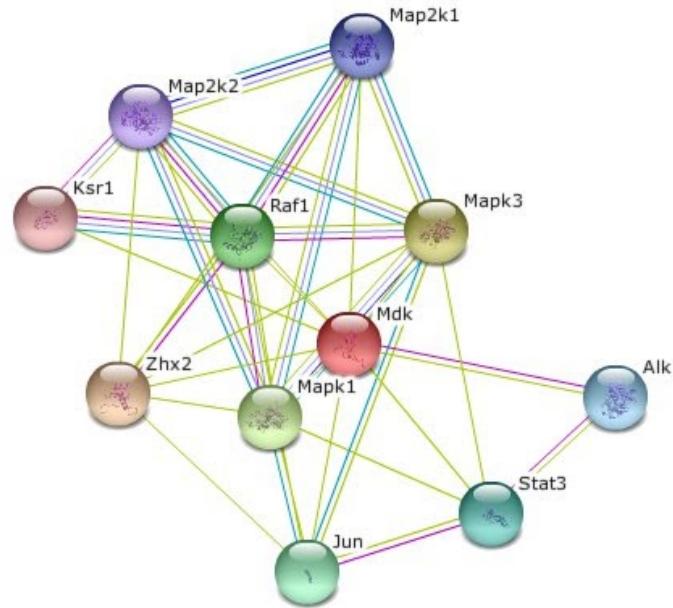
## Redes tróficas

# Los sistemas biológicos forman redes



Regulación genética

# Los sistemas biológicos forman redes



Interacciones de  
proteínas

# Los sistemas biológicos forman redes



neuronas

# Los sistemas biológicos forman redes

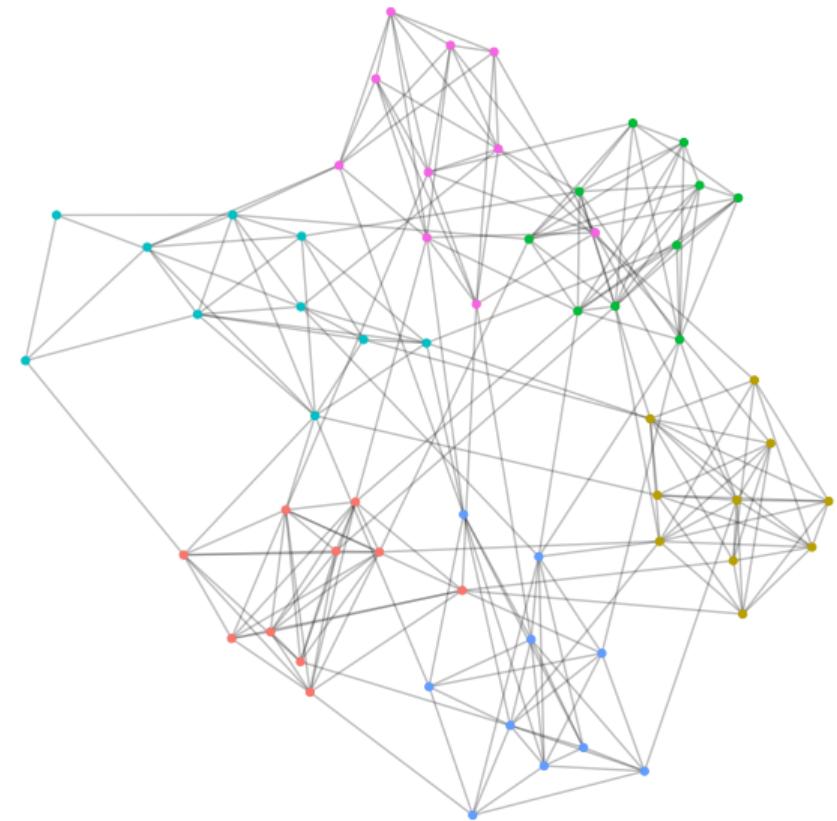
epidemias



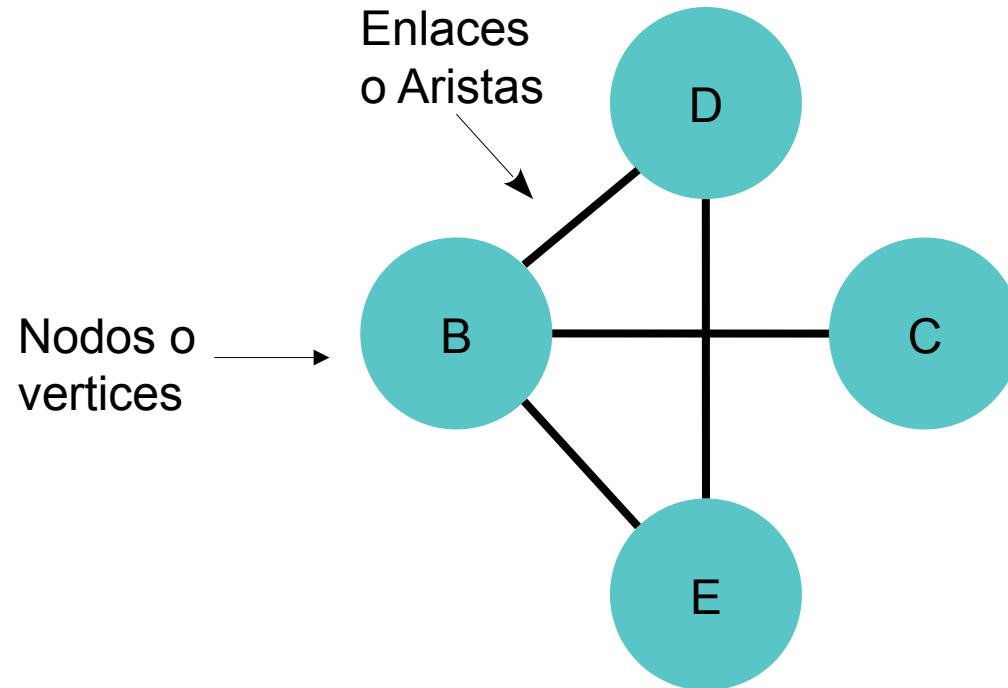


# Las redes explican como los sistemas se conectan e interactuan

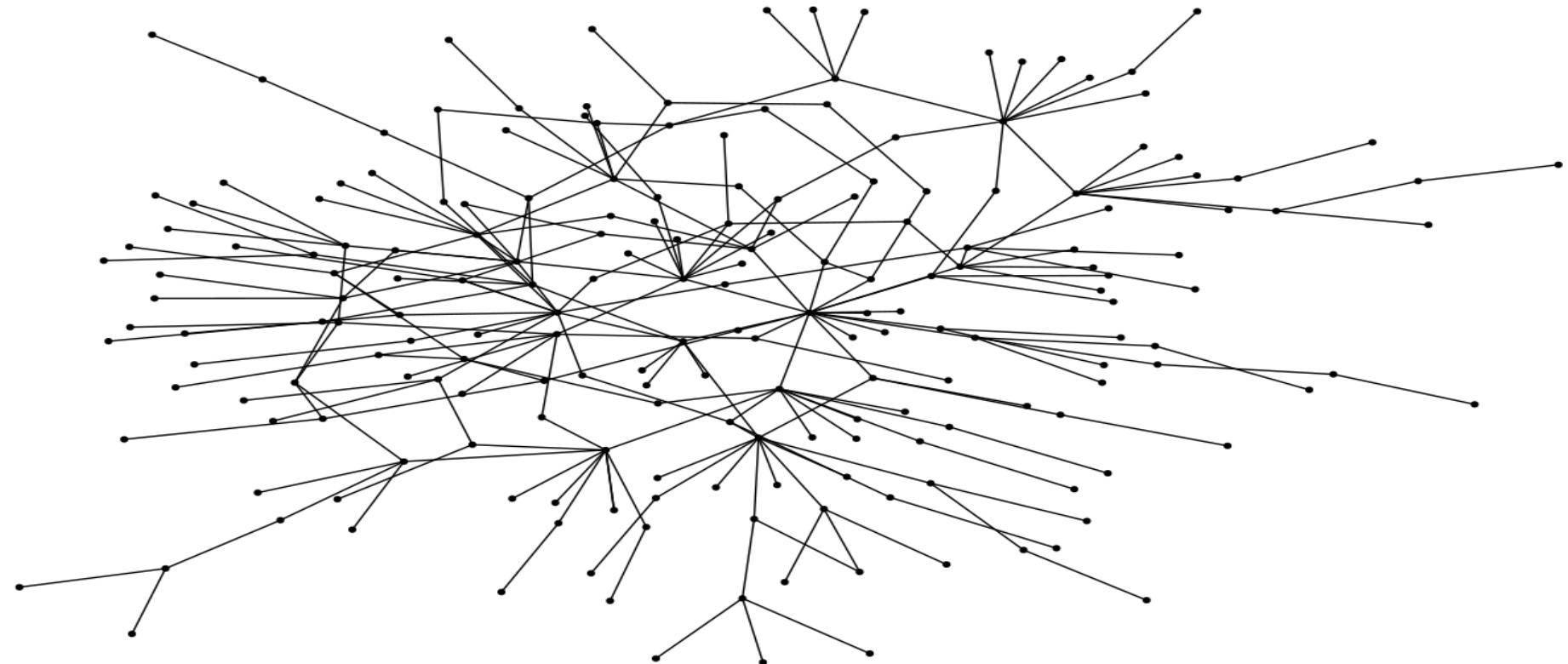
La ciencia de redes  
evoluciona del campo de las  
matemáticas conocido como  
“Teoría de Grafos”



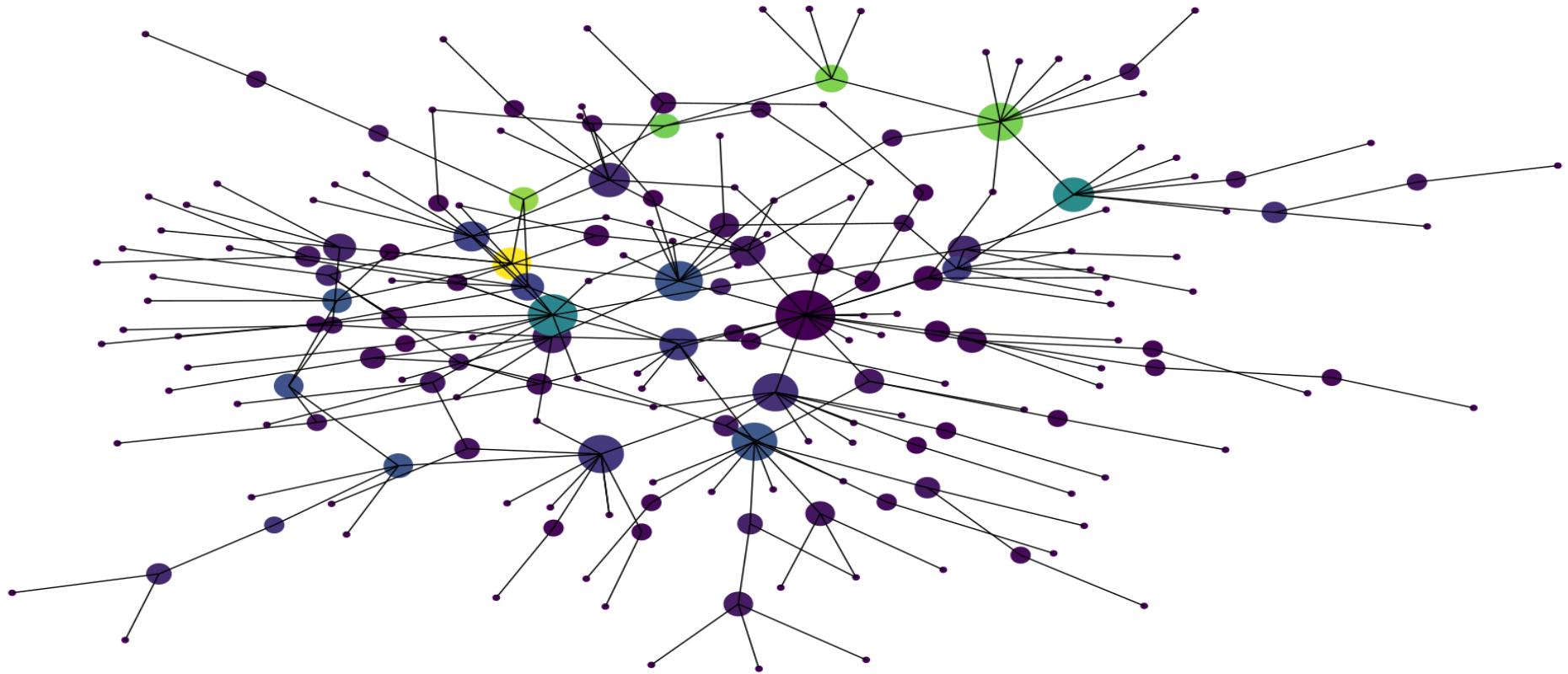
# Las redes explican como los sistemas se conectan e interactuan



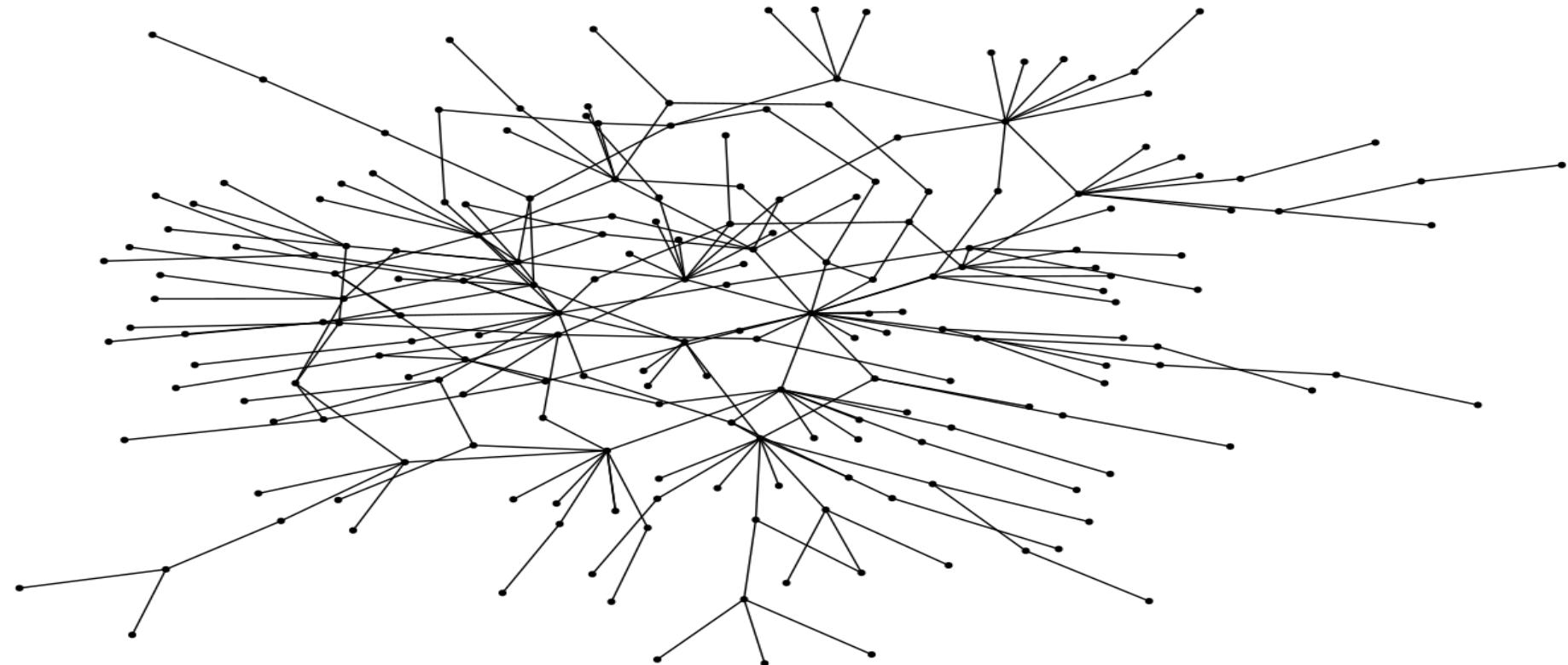
# Aprender cosas sobre los elementos...



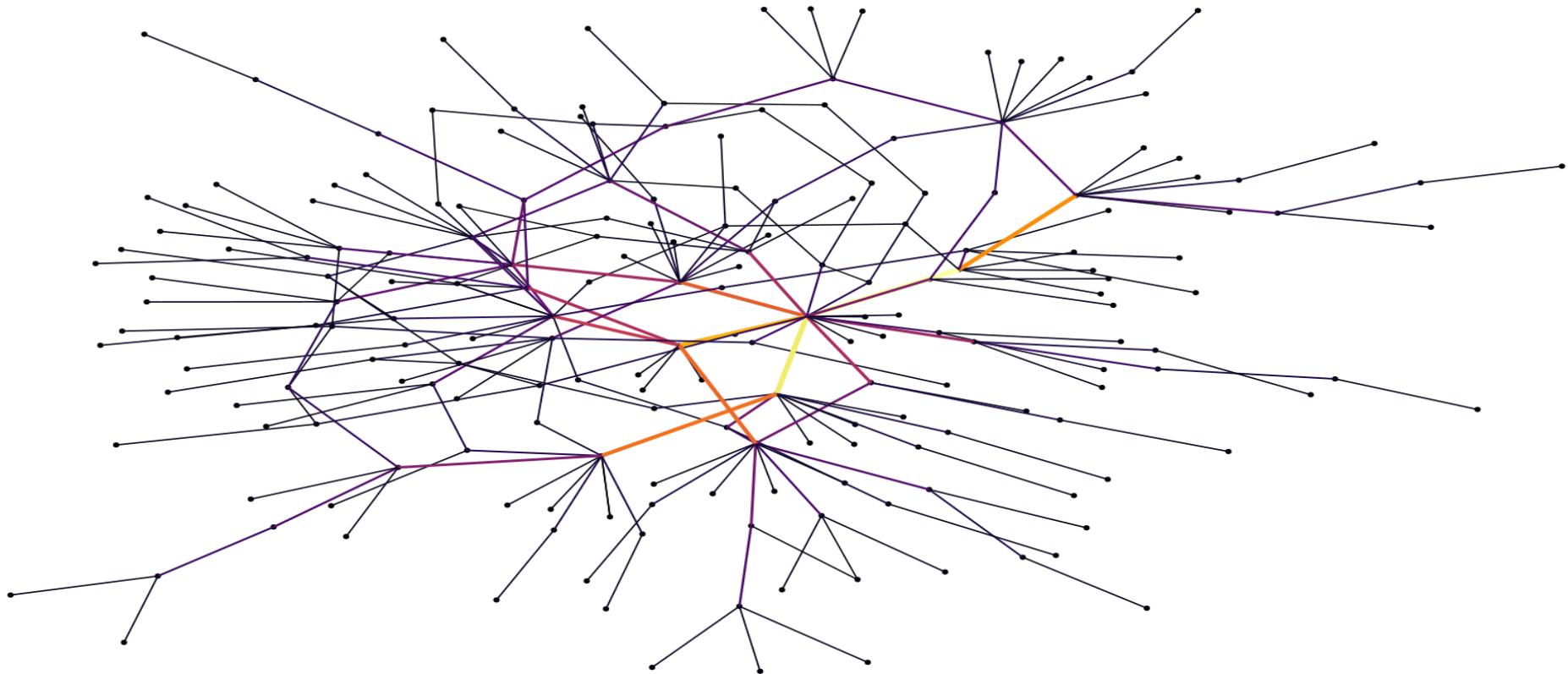
# Aprender cosas sobre los elementos...



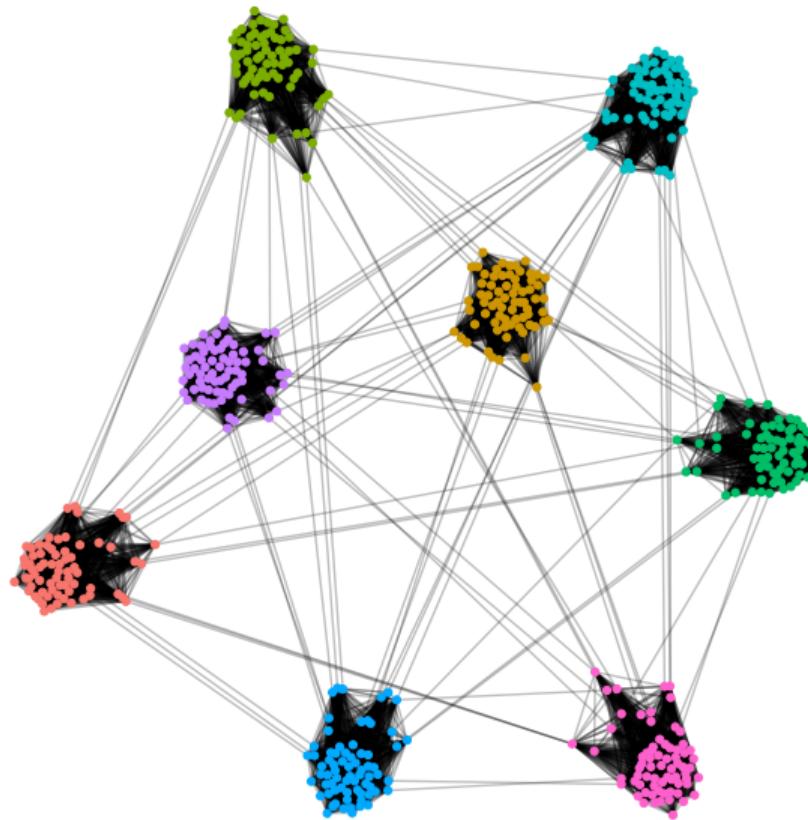
# ...y de como se conectan



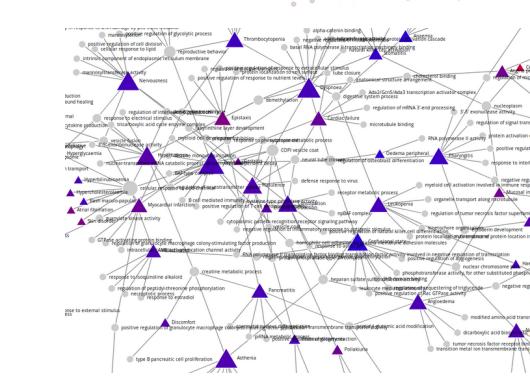
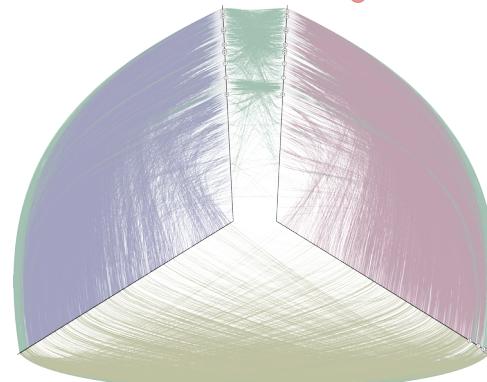
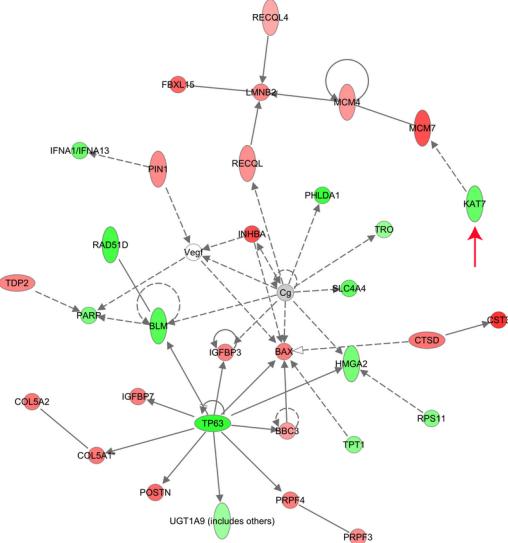
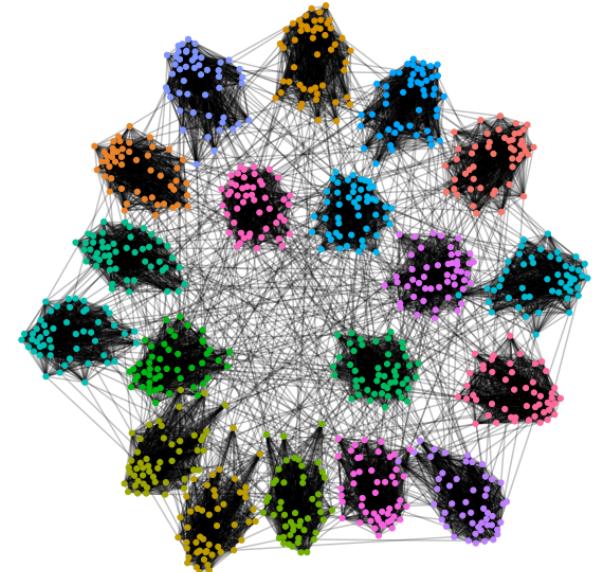
# ...y de como se conectan



# Y del sistema completo



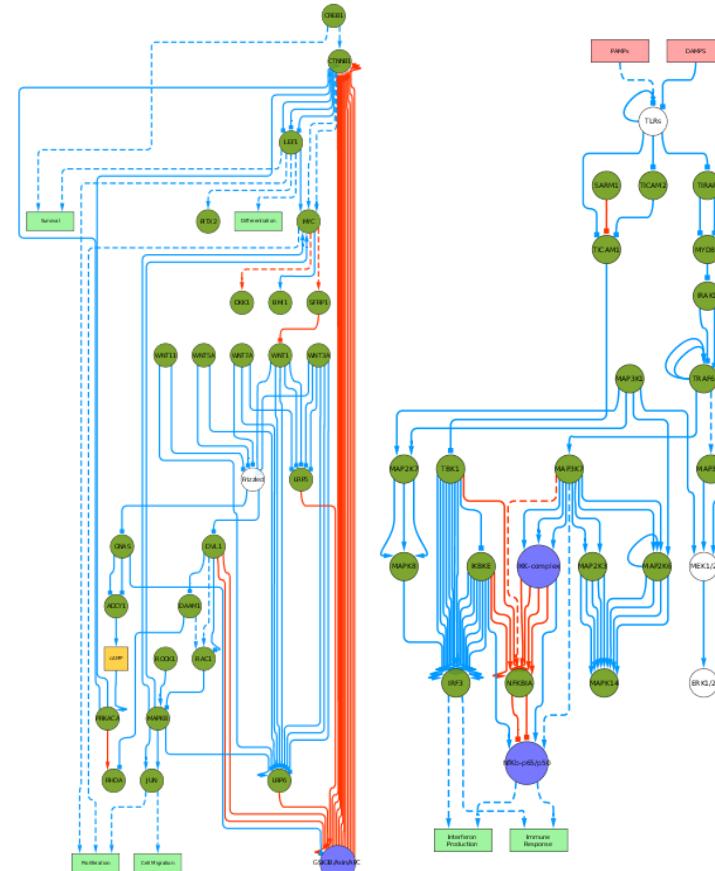
Con redes podemos visualizar datos...



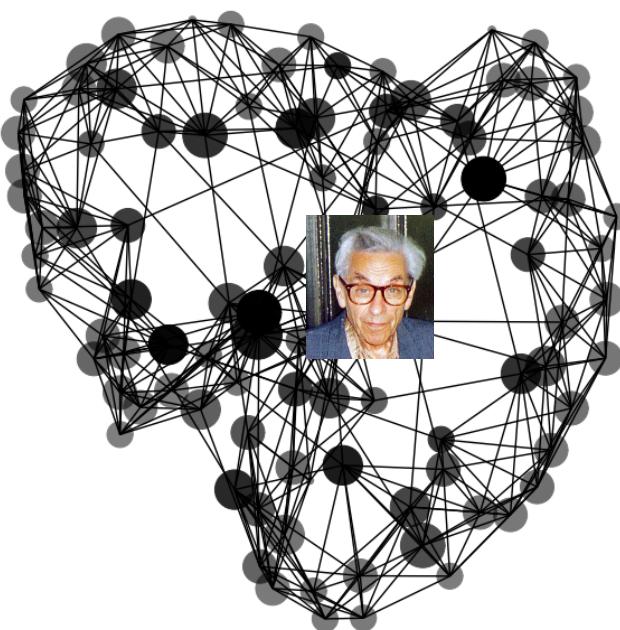
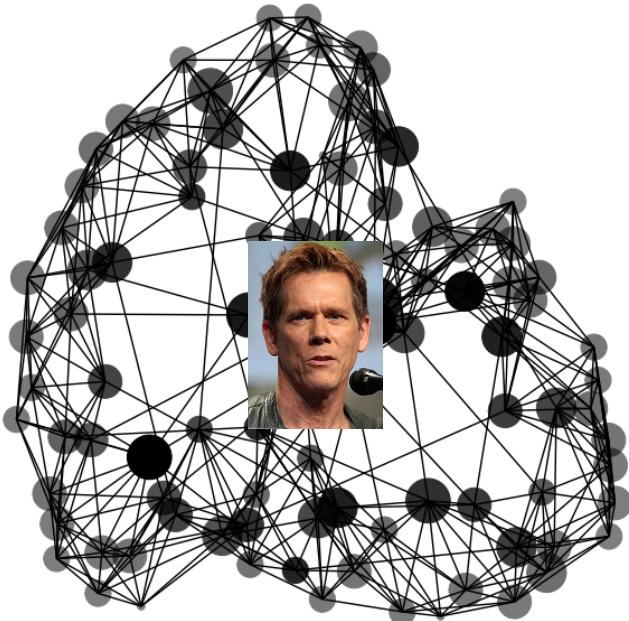
# Las redes permiten revelar patrones

Con las redes podemos:

- Identificar fortalezas y debilidades del sistema
  - Modelar procesos dinámicos dentro del sistema
  - Inferir y predecir la evolución del sistema



# Con las redes podemos comparar sistemas

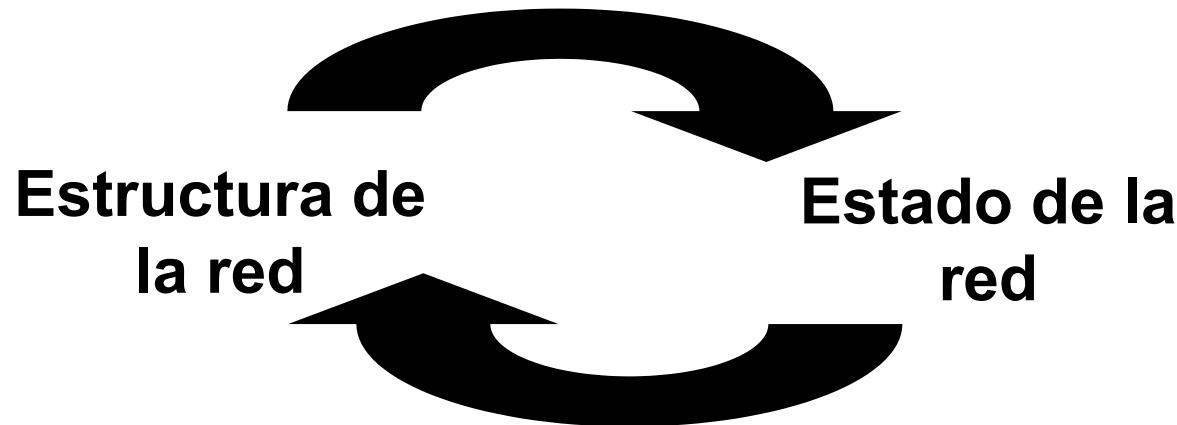


- ¿Cómo se parecen...
- ...redes metabólicas y neuronales?
- ...redes tróficas y vías de señalización?
- ...regulación genética y redes de transporte?

Ciencia de redes → Interdisciplinariedad

# Estructura y estado de la red

Como se conectan los elementos entre ellos

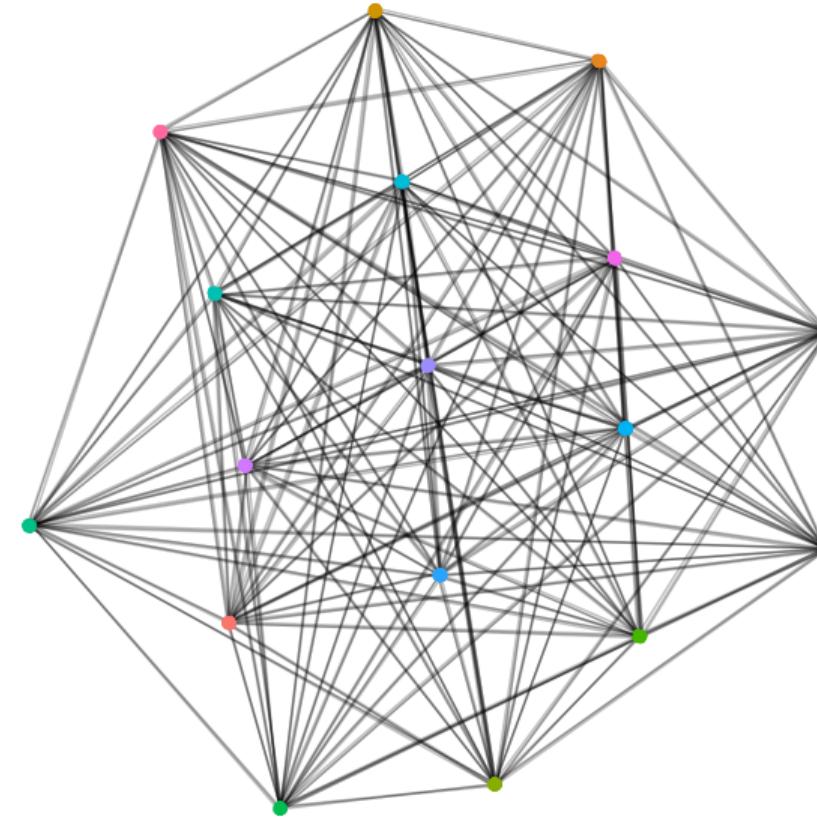


Como se comportan los elementos individuales

Propiedades de cada enlace

# Podemos estudiar redes del mundo real el día de hoy ...

- Tenemos la teoría
- Tenemos los datos
- Tenemos la tecnología



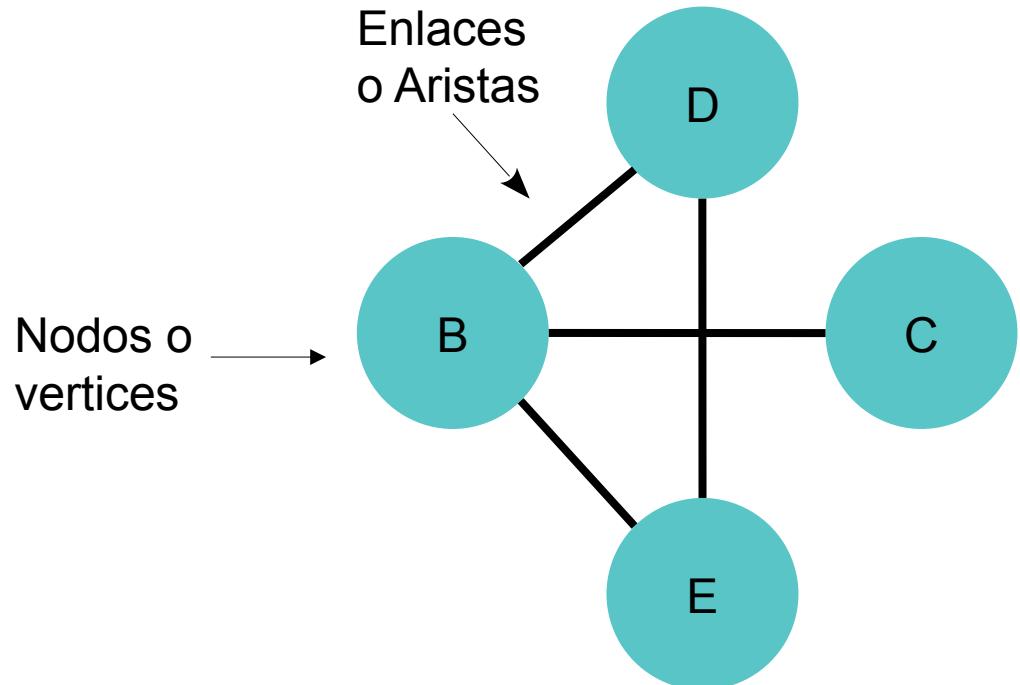
## Parte 2: Tipos de redes

# Definición de grafo

Conjunto de nodos (vértices)

$$G(V, E)$$

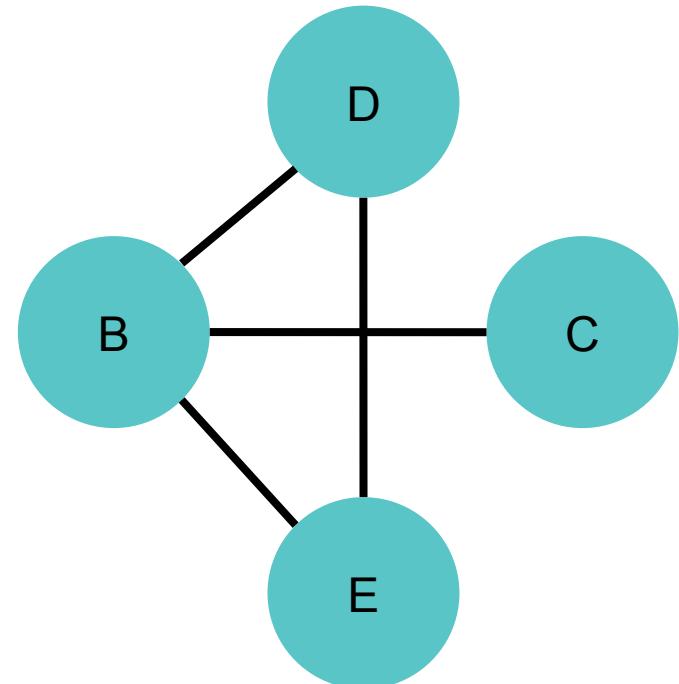
Conjunto de enlaces (aristas)  
Parejas de elementos en V



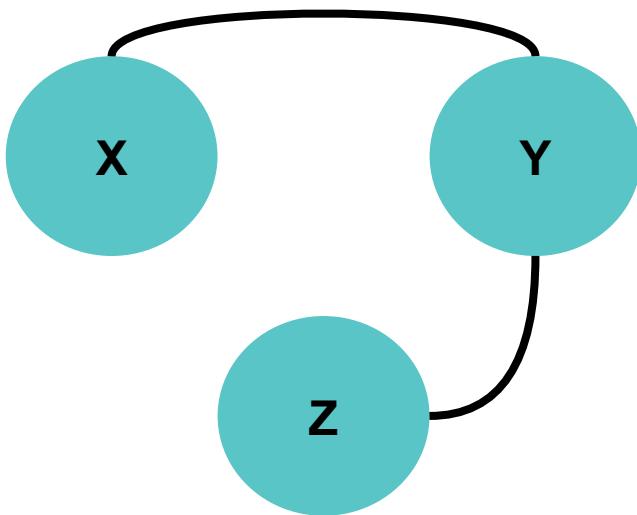
# Matriz de adyacencia

NODOS ▶	B	C	D	E
↓	0	1	1	1
B	1	0	0	1
C	1	0	0	0
D	1	0	0	0
E	1	0	1	0

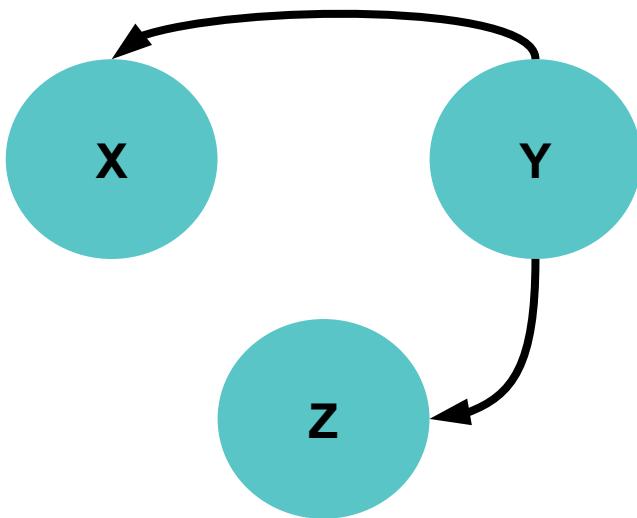
Enlaces



# Diferentes clases de enlace

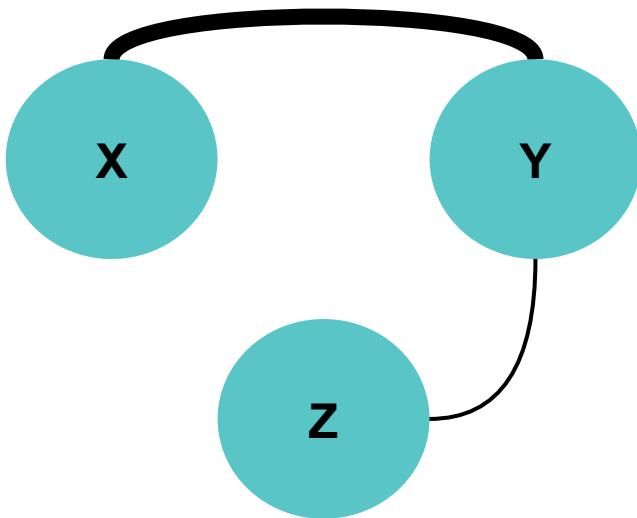


# Diferentes clases de enlace



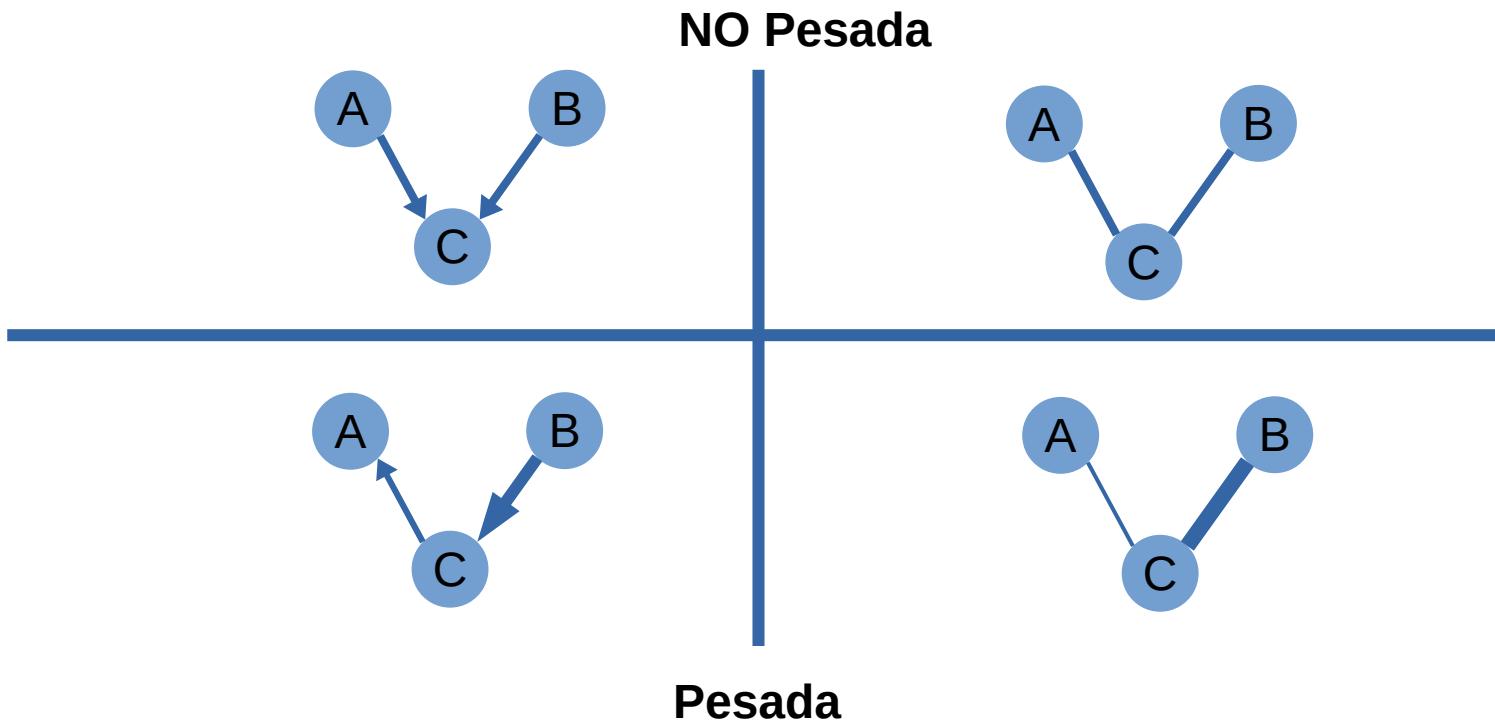
Dirección

# Diferentes clases de enlace

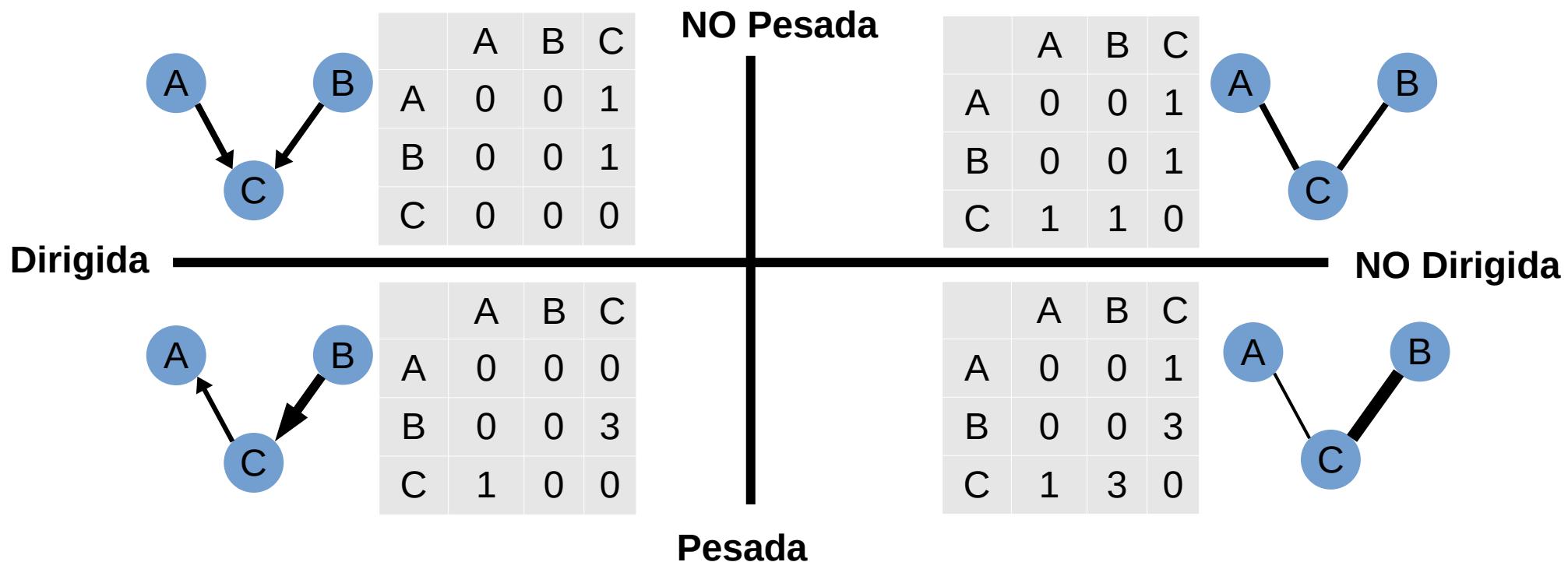


Peso

# Diferentes tipos de redes



# Podemos representarlas todas con matrices...



# Pregunta para pensar

¿Cómo sería la matriz de una red con signos?

*Por ejemplo, un circuito regulatorio.*

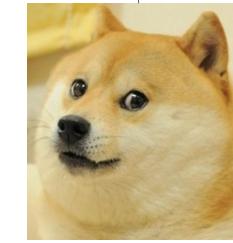
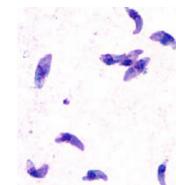
¿Qué tipo de red sería?

# ¿Todas mis redes van a ser una matriz?

A veces, para representar mejor un sistema, puedo decidir usar otro formalismo para representar redes...

# Redes bipartitas

Nodos  
TOP



Enlaces  
entre top  
y bottom

Nodos  
BOTTOM

Top y bottom son  
mutuamente excluyentes

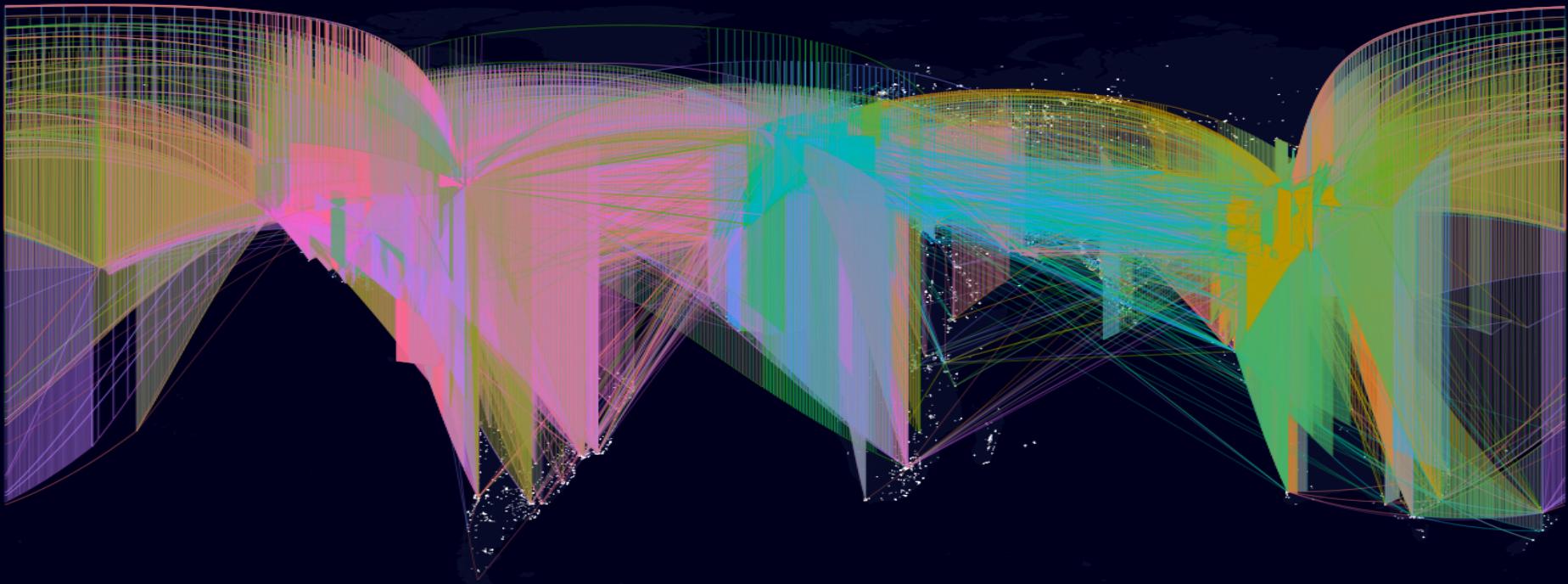
# Matriz de incidencia

	1	2	3	4	5
a	1	1	1	0	0
b	0	1	1	0	0
c	0	0	0	1	1

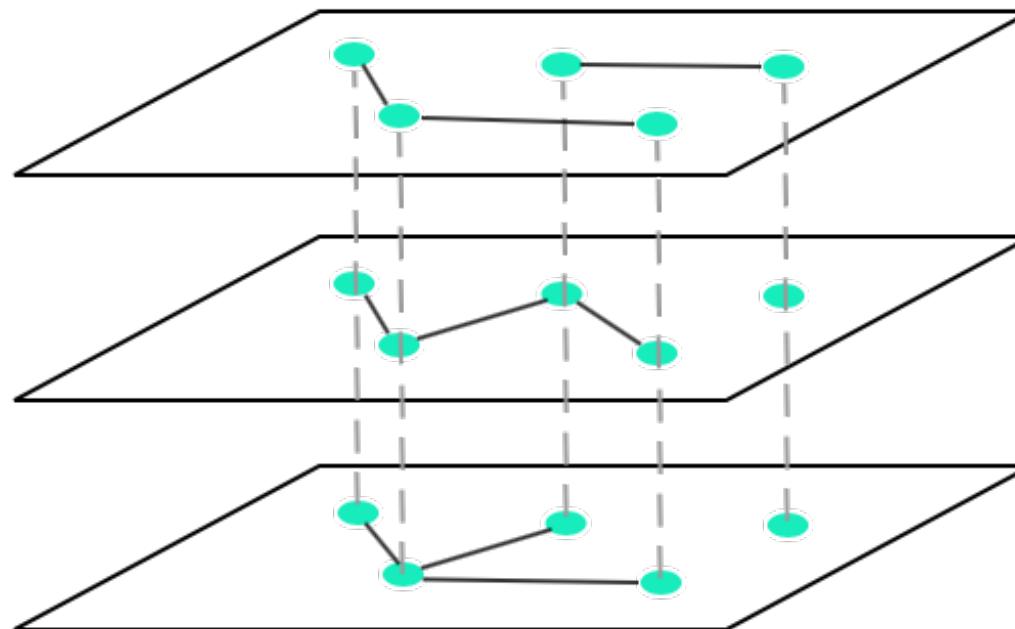
# Redes multicapa



# Redes multicapa



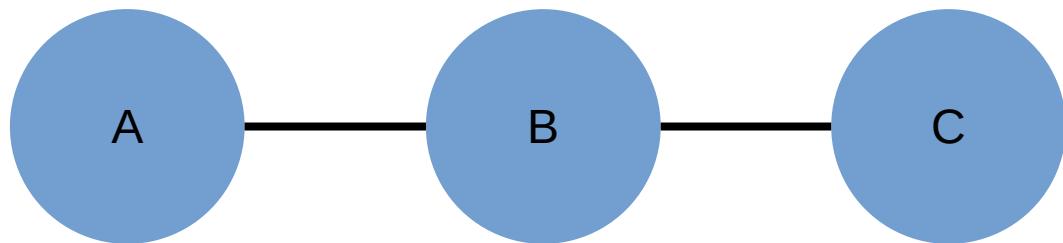
# Redes multicapa



# Parte 3: Análisis de redes: propiedades y métricas

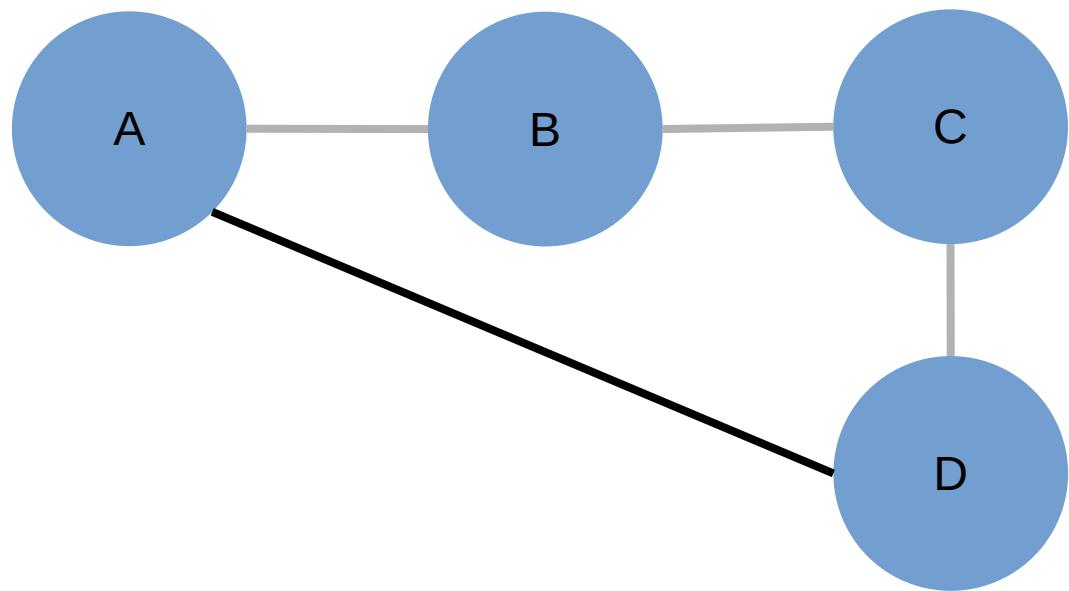
# Propiedades de la **RED**

# Caminos

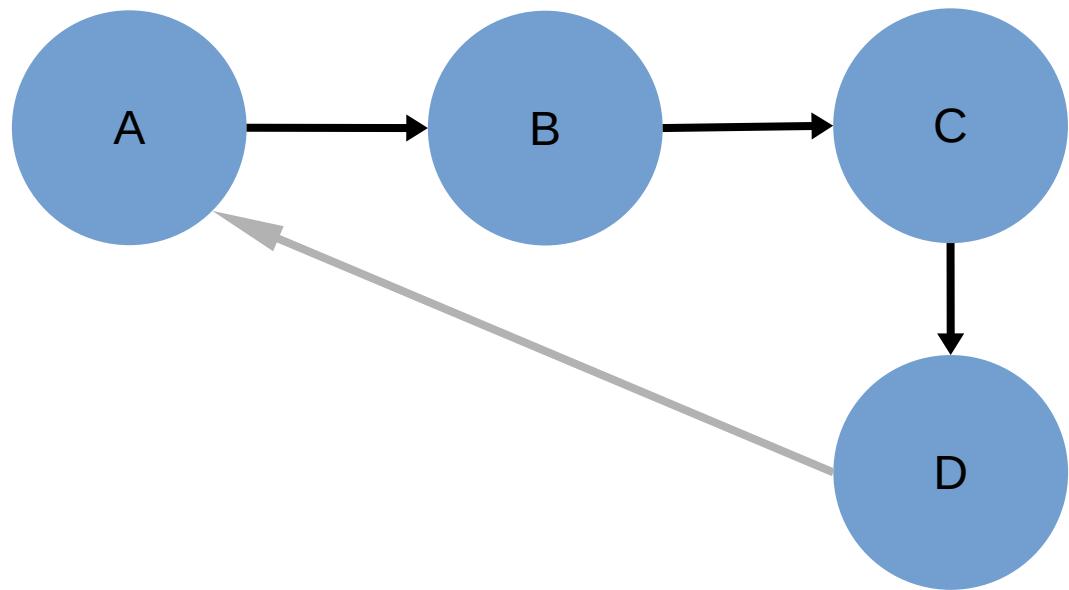


**Secuencia de nodos  
conectados por una  
secuencia de enlaces**

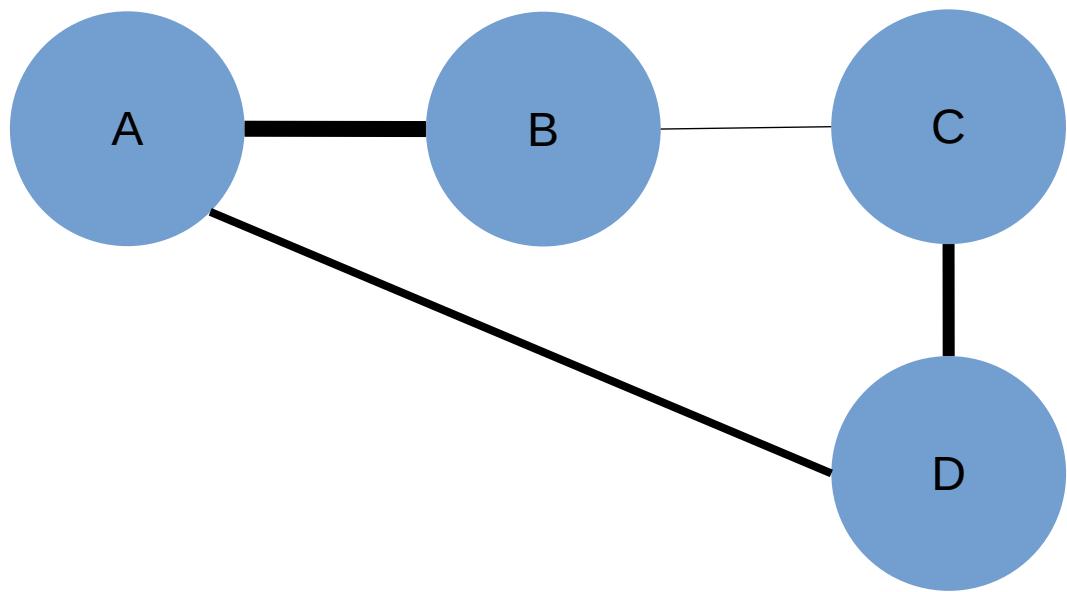
# Camino más corto (*Shortest Path*)



# Camino más corto (*Shortest Path*)



# Camino más corto (*Shortest Path*)



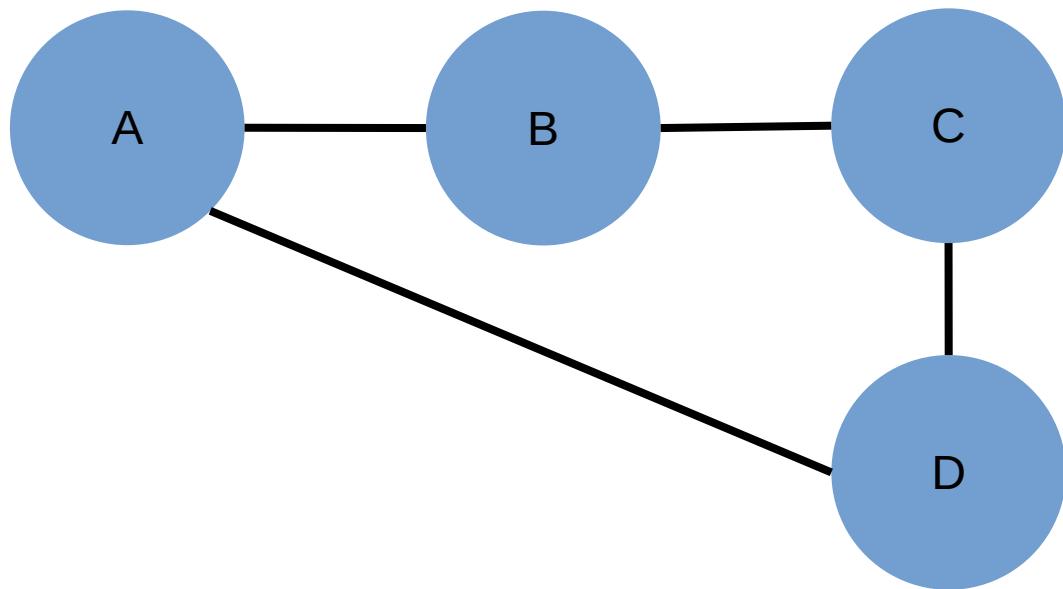
Los pesos representan mayor o menor “distancia” según nos convenga

Ejemplos:

- Distancia filogenética:  
Mayor peso ~ más lejano en la red  
→ **CASO ESTÁNDAR**

- Afinidad de sustrato:  
Mayor peso ~ más cercano en la red

# Camino más corto (*Shortest Path*)



Algoritmos:

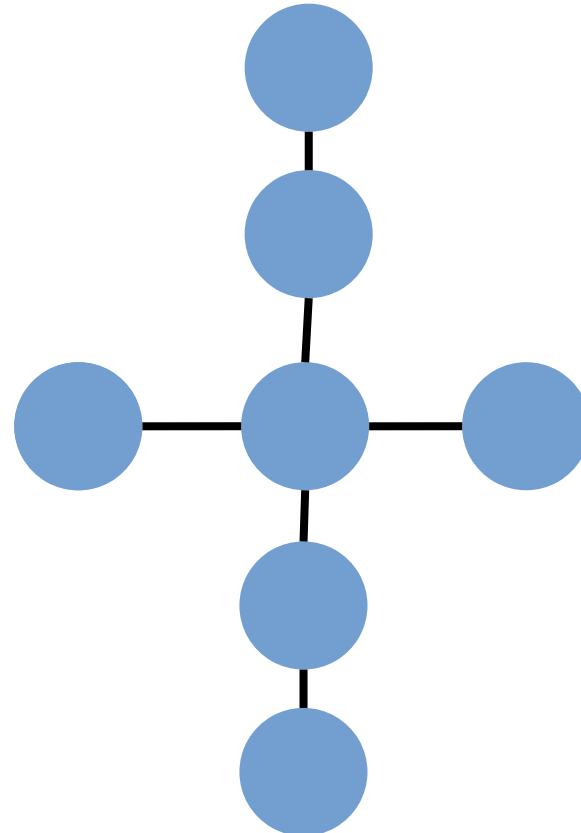
- Djikstra
- Bellman-Ford
- A\*
- Floyd – Warshall

# Camino más corto promedio

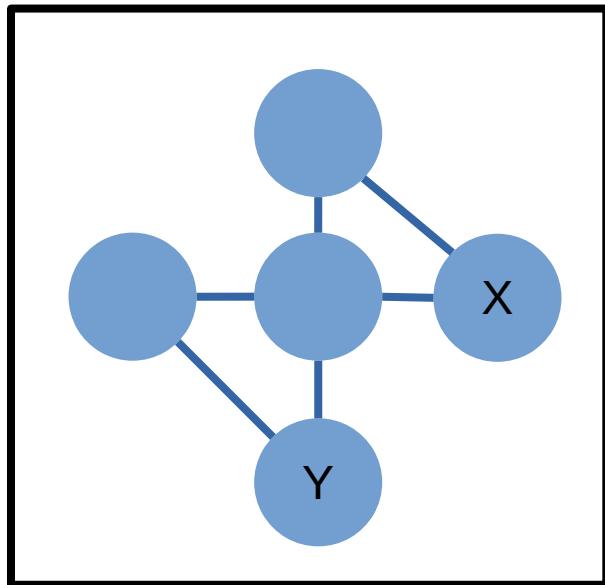
$$l_{prom} = \frac{1}{N(N-1)} \sum_{i=1}^N \sum_{j=1 \neq i}^N d_{ij}$$

# Diámetro

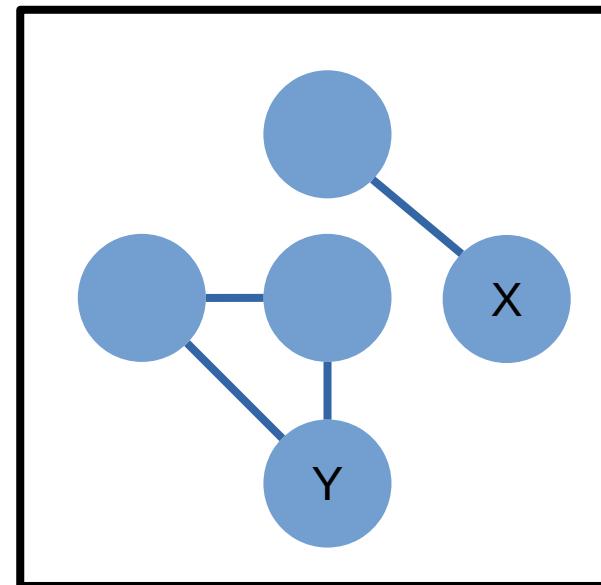
$$D = \max d_{ij}$$



# Componentes conectados

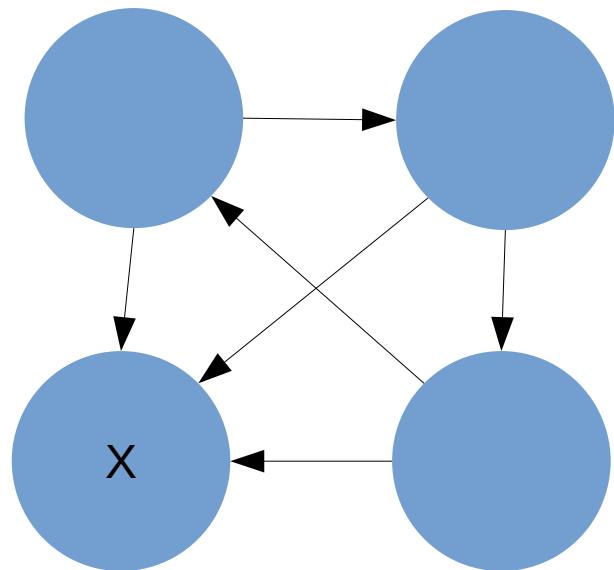


$$d_{xy} \neq \infty$$

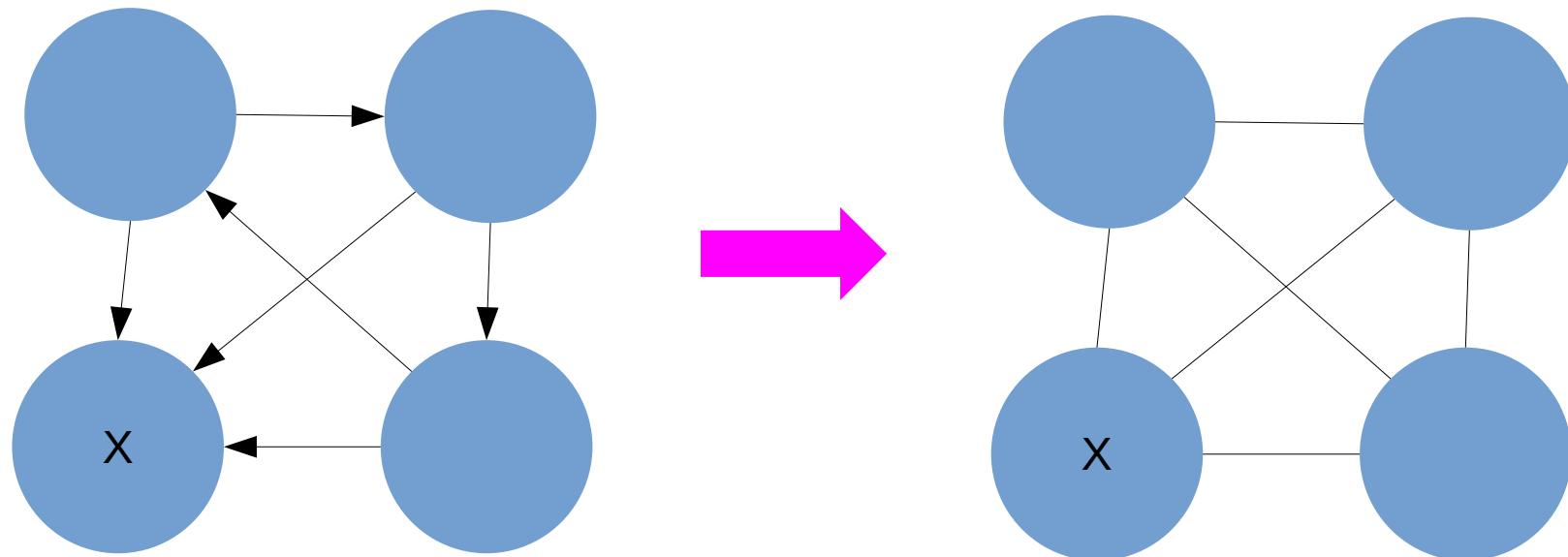


$$d_{xy} = \infty$$

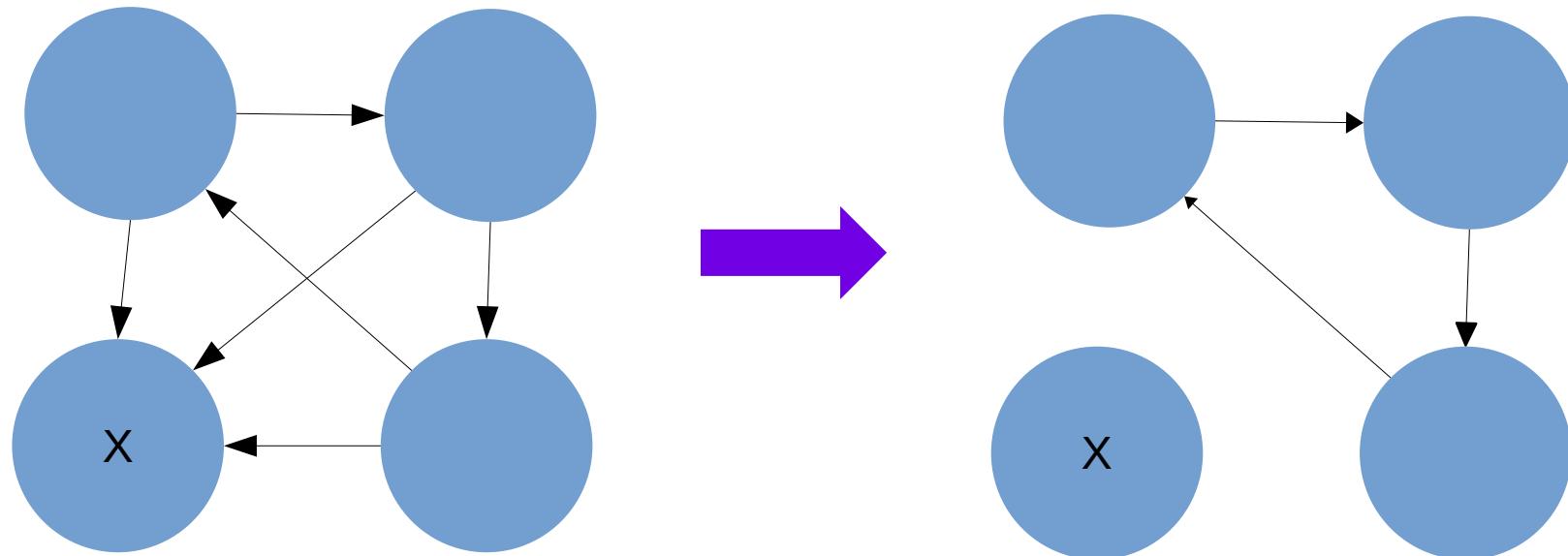
# Componentes conectados



# Componente débilmente conectado



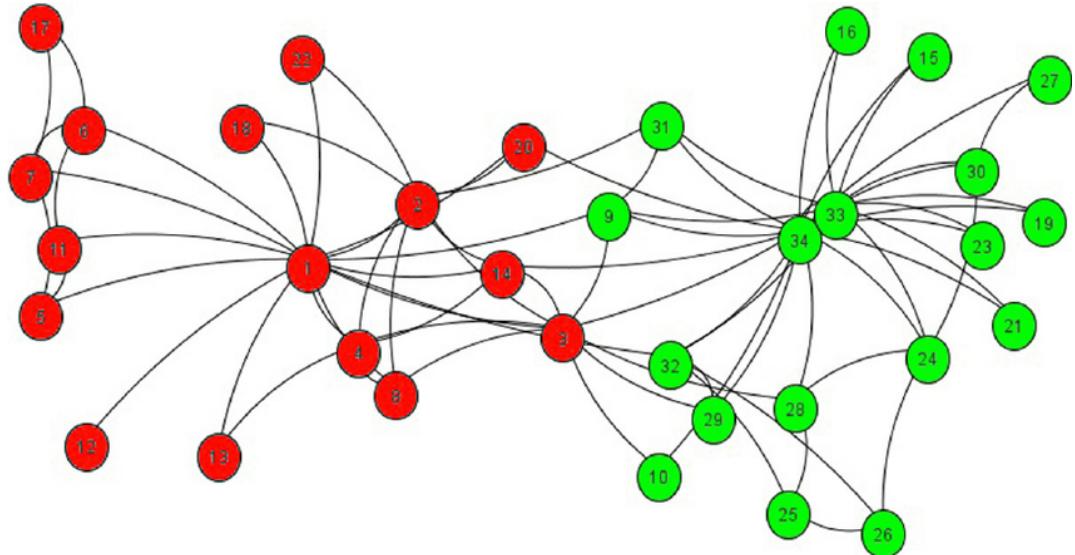
# Componente fuertemente conectado



# ¿Componente gigante?

- Siempre tendremos (al menos) un componente con un mayor número de nodos. Lo llamamos el **“componente más grande.”**
- Podemos tener un componente que contenga más de la mitad ( $50\% + 1$ ) de los nodos. Lo llamamos **“Componente gigante.”**
- Algunas personas utilizan los conceptos indistintamente. ¡Cuidado!

# Comunidades o módulos



## Partición de la red

Número de enlaces **DENTRO** de la comunidad

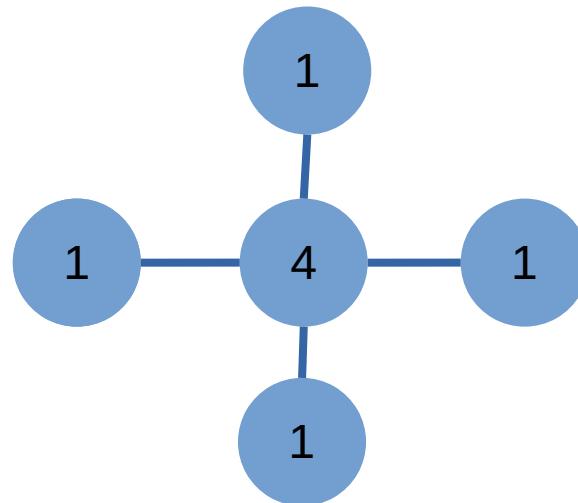
MAYOR que

Número de enlaces **FUERA** de la comunidad

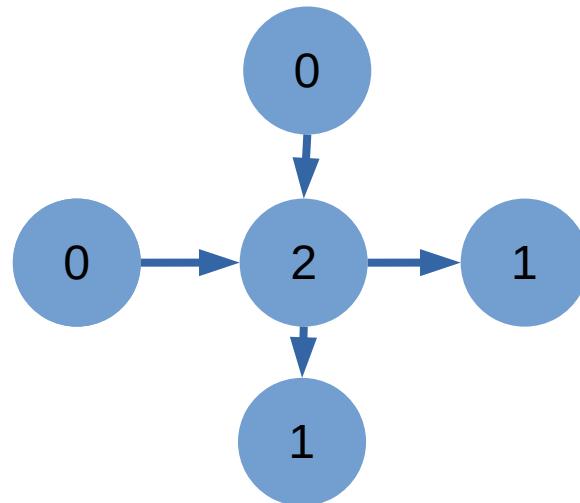
# Propiedades de los **NODOS**

# Grado

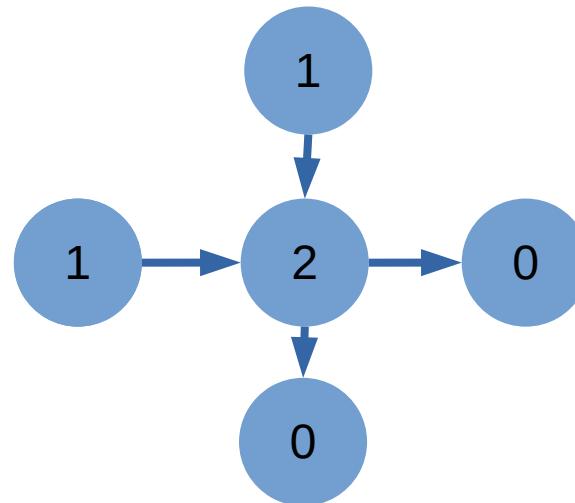
Número de vecinos



# Grados de entrada y salida

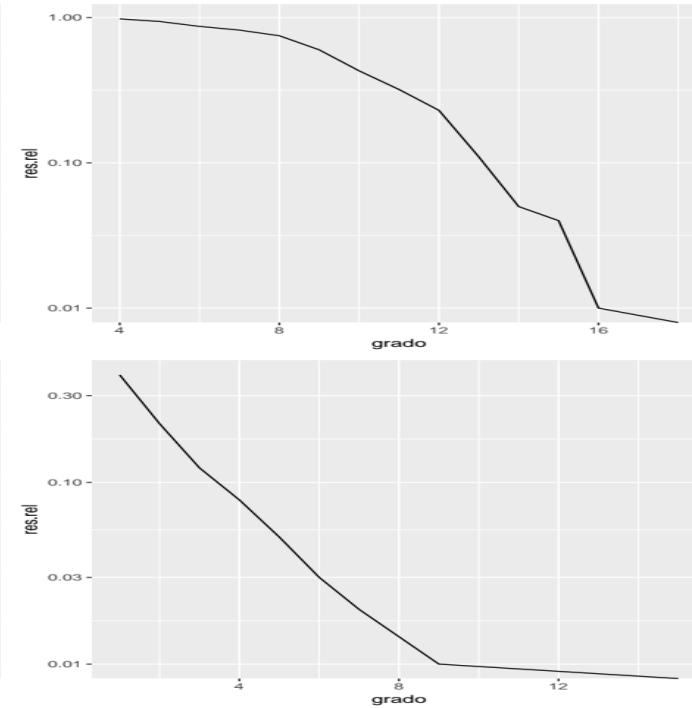
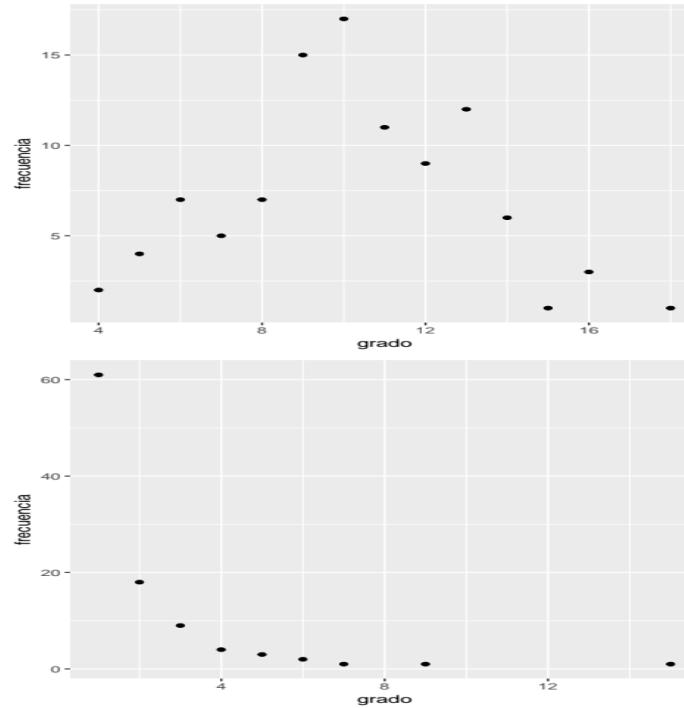
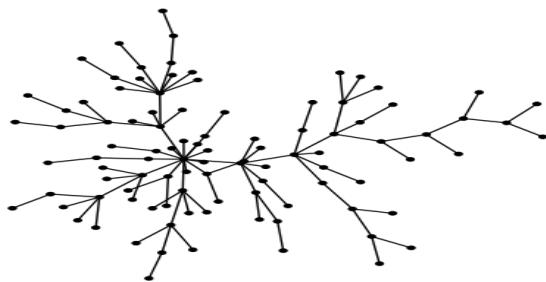
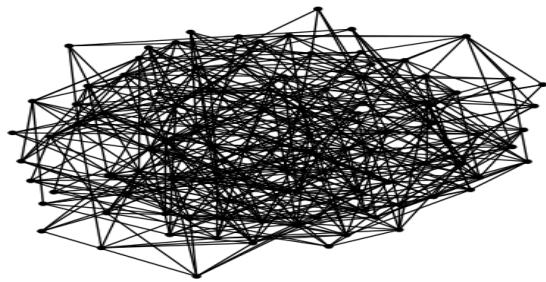


Entrada



Salida

# Distribución de grado



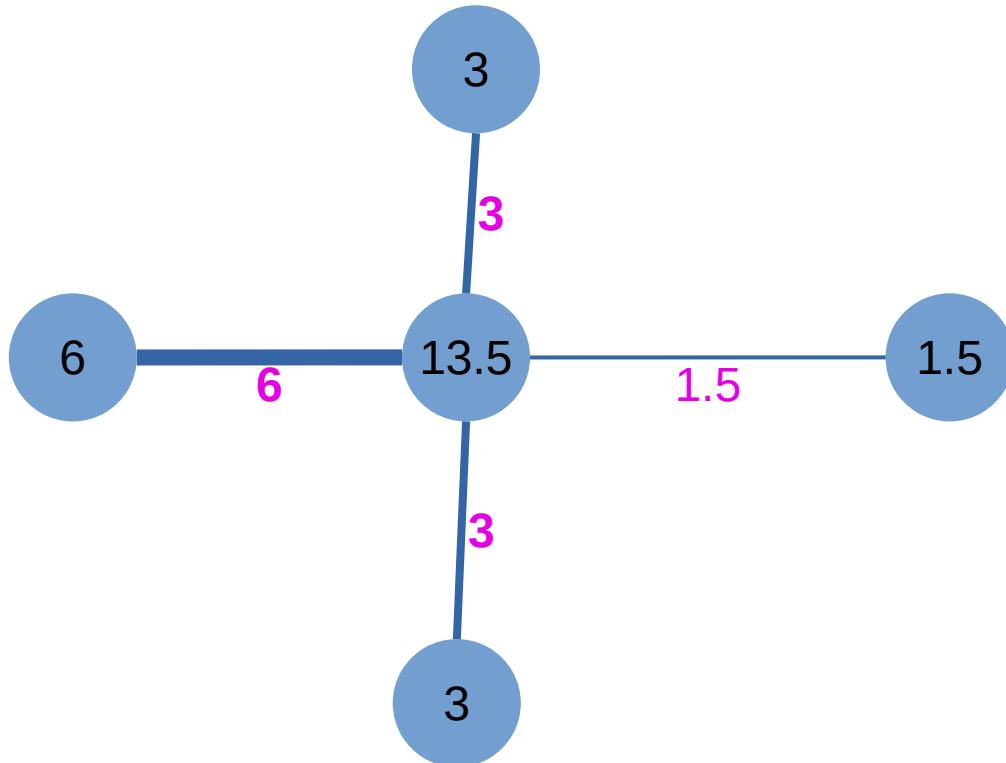
$$P(k) = \text{accum } p(k)$$

# Hubs

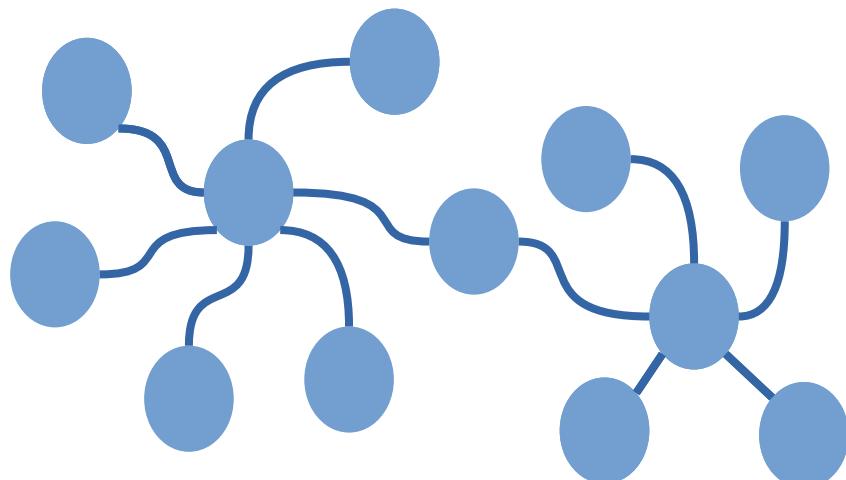


- Formalmente:
  - Llamamos hub a un nodo con grado significativamente mayor al grado promedio de la red
    - Requerimos un modelo para evaluarlo
- Informalmente:
  - Llamamos hub a los nodos con mayor grado en la red

# Fuerza



# Betweenness Centrality



$$b_i = \sum_{r, s} \frac{n_{rs}^i}{g_{rs}}$$

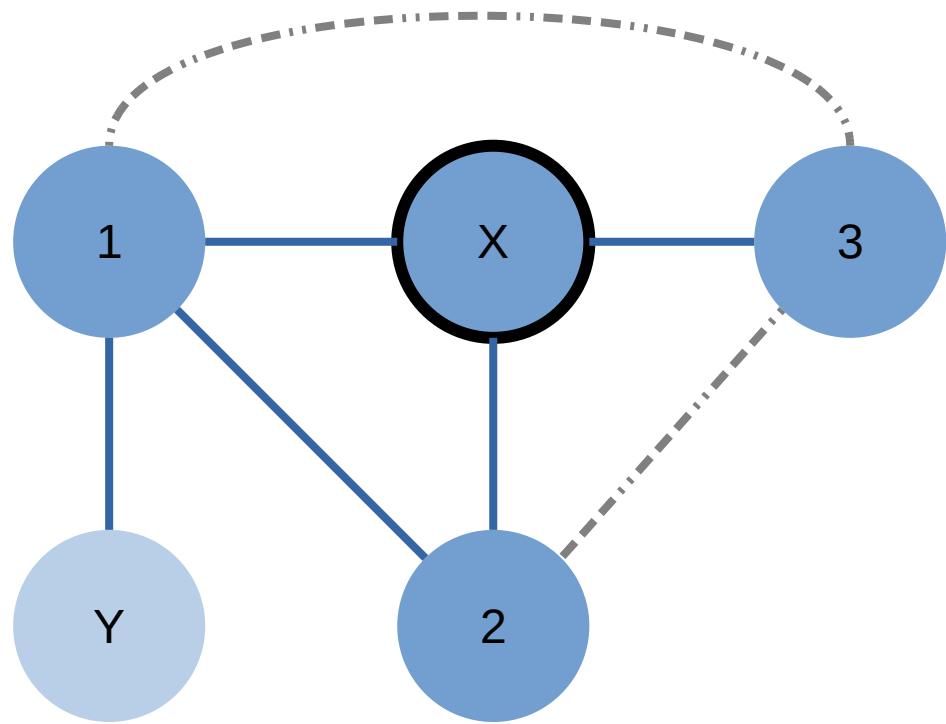
$n$  = Número de caminos más cortos entre  $r$  y  $s$  que pasan por  $i$

$G$  = número de caminos más cortos entre  $r$  y  $s$

# Otras centralidades

- Eigenvector
- Katz
- PageRank
- Closeness
- Comunicabilidad

# Clustering Coefficient



$$C_i = \frac{\text{Num. } \Delta_i}{k_i(K_i - 1)/2}$$

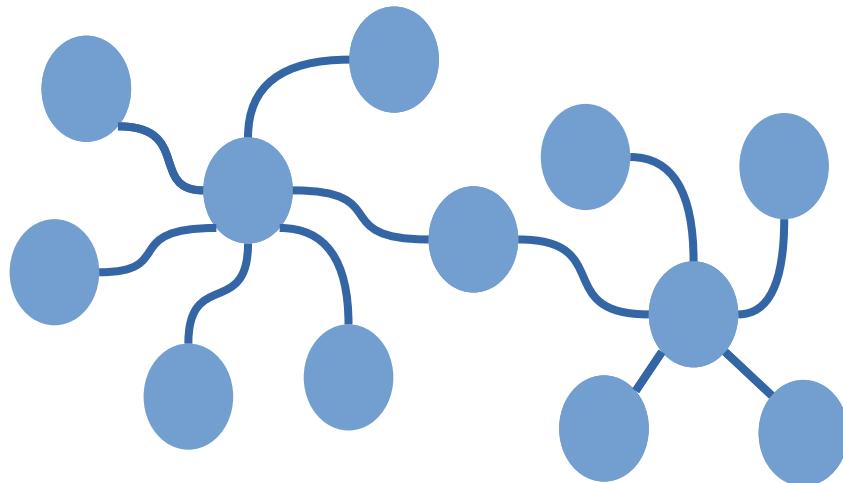
0 si  $k_i = 0$  o  $k_i = 1$

$$C = \frac{1}{N} \sum_{i=1}^N C_i$$

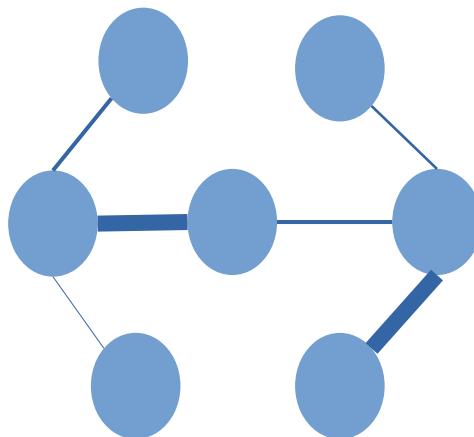
$$T = 3 \times \frac{\text{Num. } \Delta}{\text{Num. Paths de } d_{ij}=2}$$

# Propiedades de los **ENLACES**

# Edge Betweenness



# Distribución de pesos



# Visualización de redes

# Visualización de datos

- Mostrar los datos
- No distorsionar la información
- Presentar mucha información en poco espacio
- Hacer conjuntos de datos grandes comprensibles

Tufte, 1983: The Visual Display of Quantitative Information

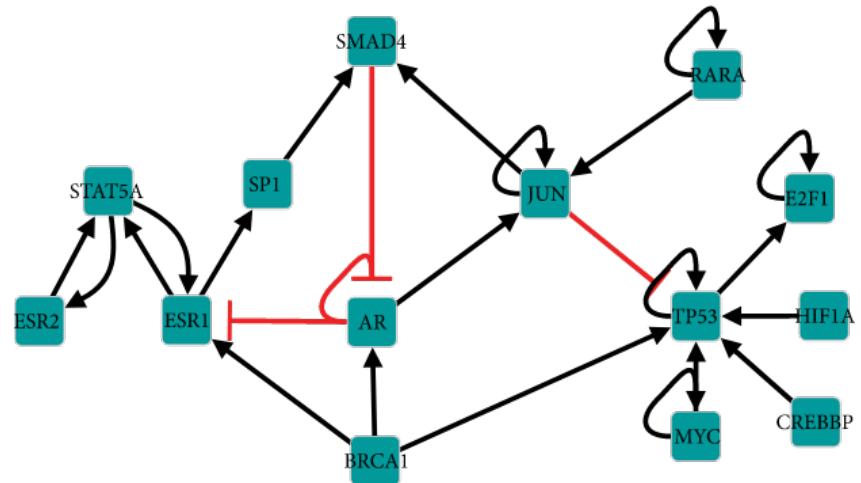
# Visualización de redes

- Convención tradicional: nodos son círculos, discos, cajas, etc... y enlaces son líneas
- ¿Existen otras convenciones? Mapas de calor de matriz de adyacencia, *circle packing* ...

# Algunos tipos de redes biológicas

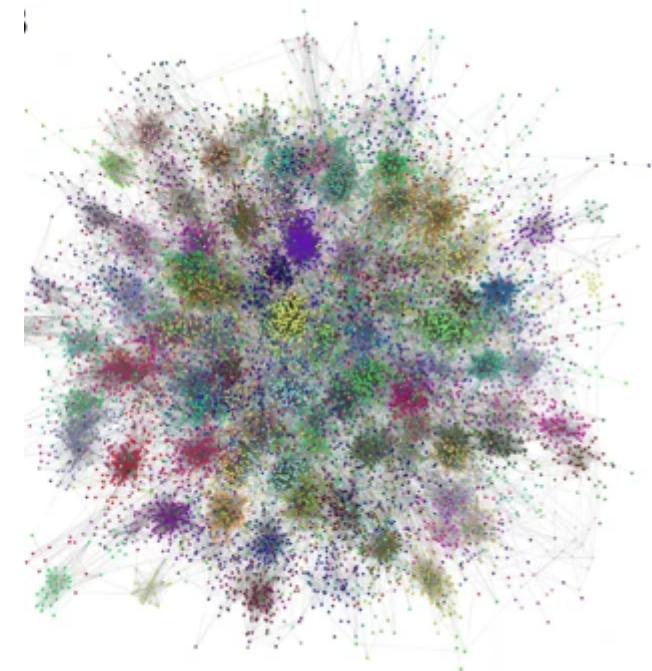
# Redes de regulación genética

- Nodos: genes\*
- En versión plana; proyección de una red bipartita gen - proteína
- Enlaces: regulación
- Tipo: dirigidas, signo/peso



# Redes de coexpresión

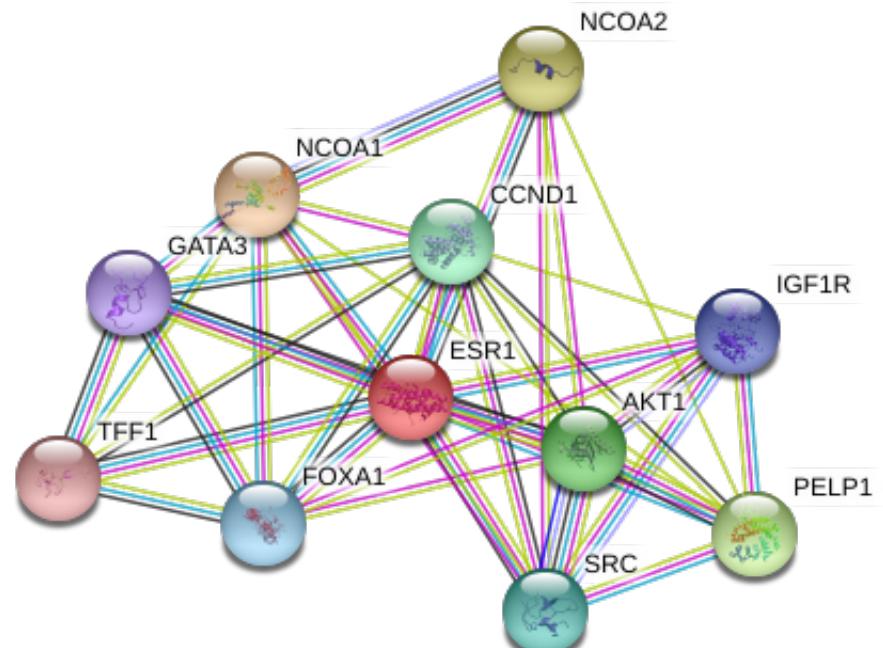
- Nodos: genes
- Enlaces: dependencia estadística en coexpresión
- Tipo: No dirigidas



<https://appliednetsci.springeropen.com/articles/10.1007/s41109-019-0129-0>

# Redes de interacción proteína-proteína

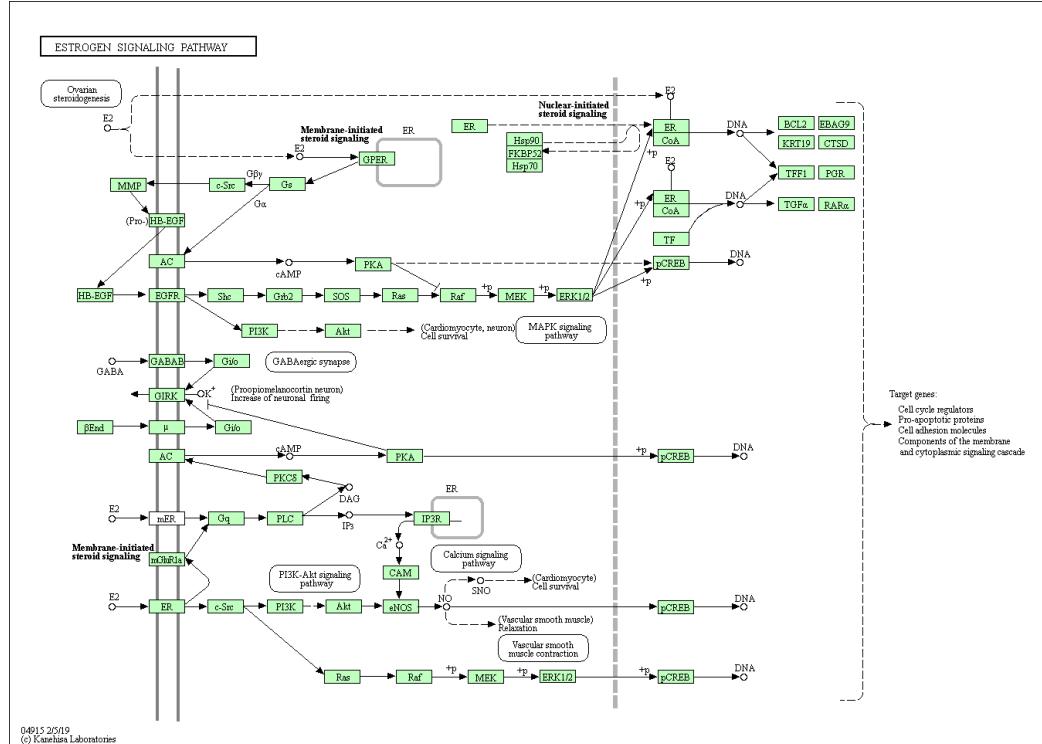
- Nodos: proteínas
- Enlaces: interacciones proteína-proteína (afinidad)
- Tipo: No dirigidas



<https://string-db.org/cgi/network.pl?taskId=yCCTcKj2uJMS>

# Vías de señalización

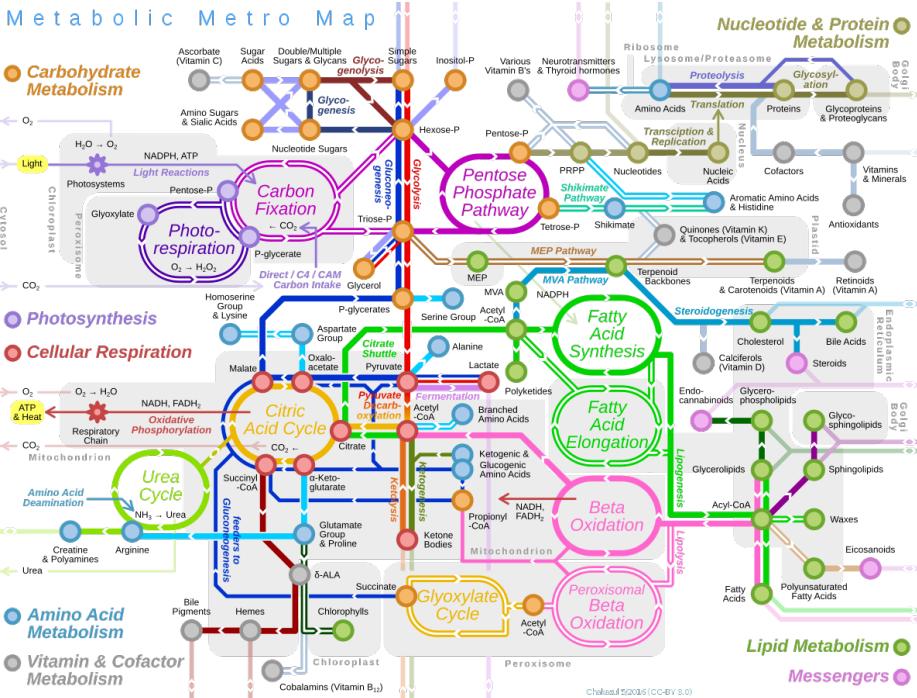
- Nodos: proteínas
- Enlaces: transducción de señales
- Tipo: Dirigidas, signo/peso



[https://www.genome.jp/kegg-bin/show\\_pathway?hsa04915](https://www.genome.jp/kegg-bin/show_pathway?hsa04915)

# Rutas metabólicas

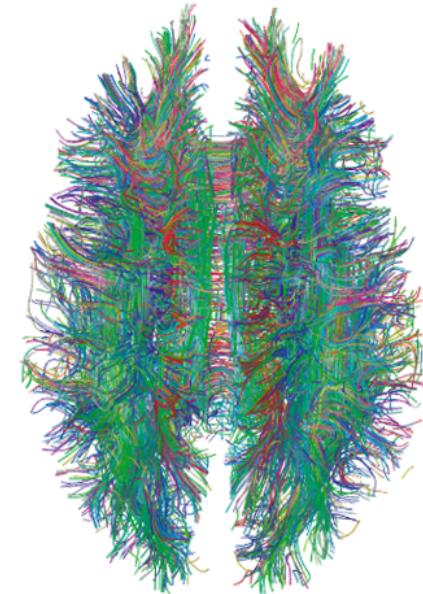
- Nodos: proteínas\*
  - En versión plana; proyección de redes bipartitas de proteína - metabolito
- Enlaces: flujo de materia
- Tipo: Dirigidas, signo/peso



[https://en.wikipedia.org/wiki/File:Metabolic\\_Metro\\_Map.svg](https://en.wikipedia.org/wiki/File:Metabolic_Metro_Map.svg)

# Conectomas

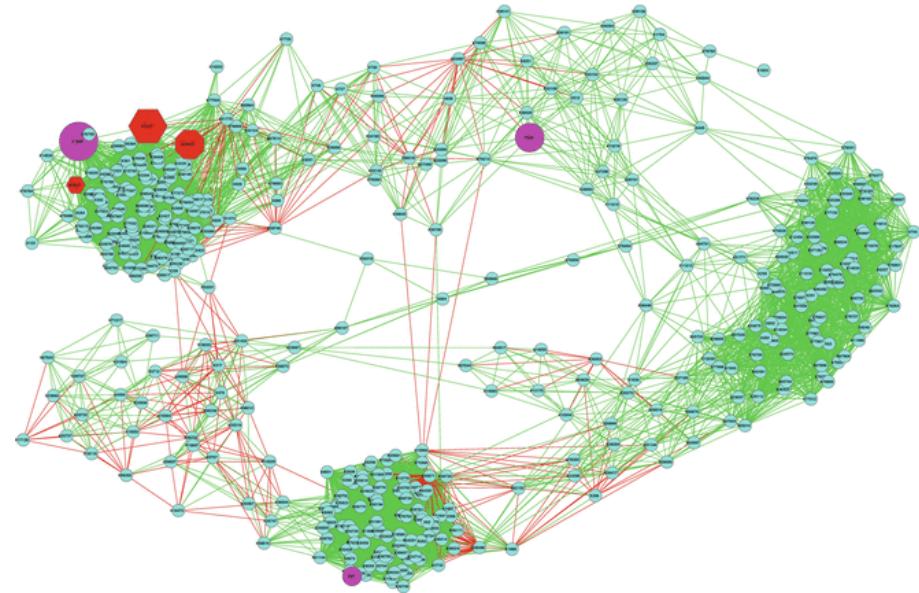
- Nodos: regiones cerebrales
- Enlaces: circuitos neuronales



[https://en.wikipedia.org/wiki/  
File:White\\_Matter\\_Connections\\_Obtained\\_with\\_MRI\\_Tractogr  
aphy.png](https://en.wikipedia.org/wiki/File:White_Matter_Connections_Obtained_with_MRI_Tractography.png)

# Redes ecológicas

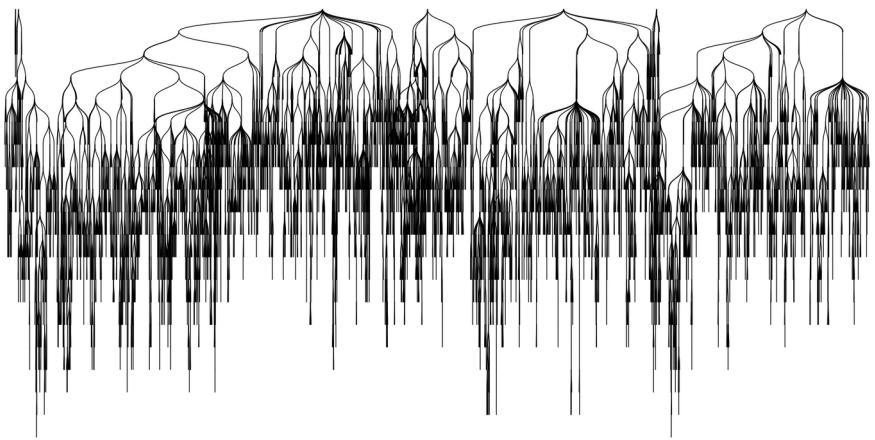
- Nodos: especies
- Enlaces: relaciones entre especies
- Tipo: usualmente multipartitas



[https://media.springernature.com/original/springer-static/  
image/chp%3A10.1007%2F8623\\_2016\\_204/MediaObjects/  
416256\\_1\\_En\\_204\\_Fig7\\_HTML.gif](https://media.springernature.com/original/springer-static/image/chp%3A10.1007%2F8623_2016_204/MediaObjects/416256_1_En_204_Fig7_HTML.gif)

# Árboles de contagio

- Nodos: personas
- Enlaces: Flujos de transmisión de patógenos
- Tipo: dirigidas, no pesadas



# Anexo: formatos de redes (y cuándo usarlo)

# Matriz de adyacencia

	[a]	[b]	[c]	[d]	[e]	[f]	[g]	[h]	[i]	[j]
[a]	0	0	0	0	0	0	0	0	0	0
[b]	0	0	0	1	0	0	0	0	1	1
[c]	0	0	0	0	0	1	0	0	0	0
[d]	0	1	0	0	0	0	1	0	0	0
[e]	0	0	0	0	0	0	0	1	0	0
[f]	0	0	1	0	0	0	0	0	0	0
[g]	0	0	0	1	0	0	0	0	0	0
[h]	0	0	0	0	1	0	0	0	0	1
[i]	0	1	0	0	0	0	0	0	0	0
[j]	0	1	0	0	0	0	0	1	0	0

# Edge list

b	d
c	f
d	g
e	h
b	i
b	j
h	j
d	l
h	m
k	m

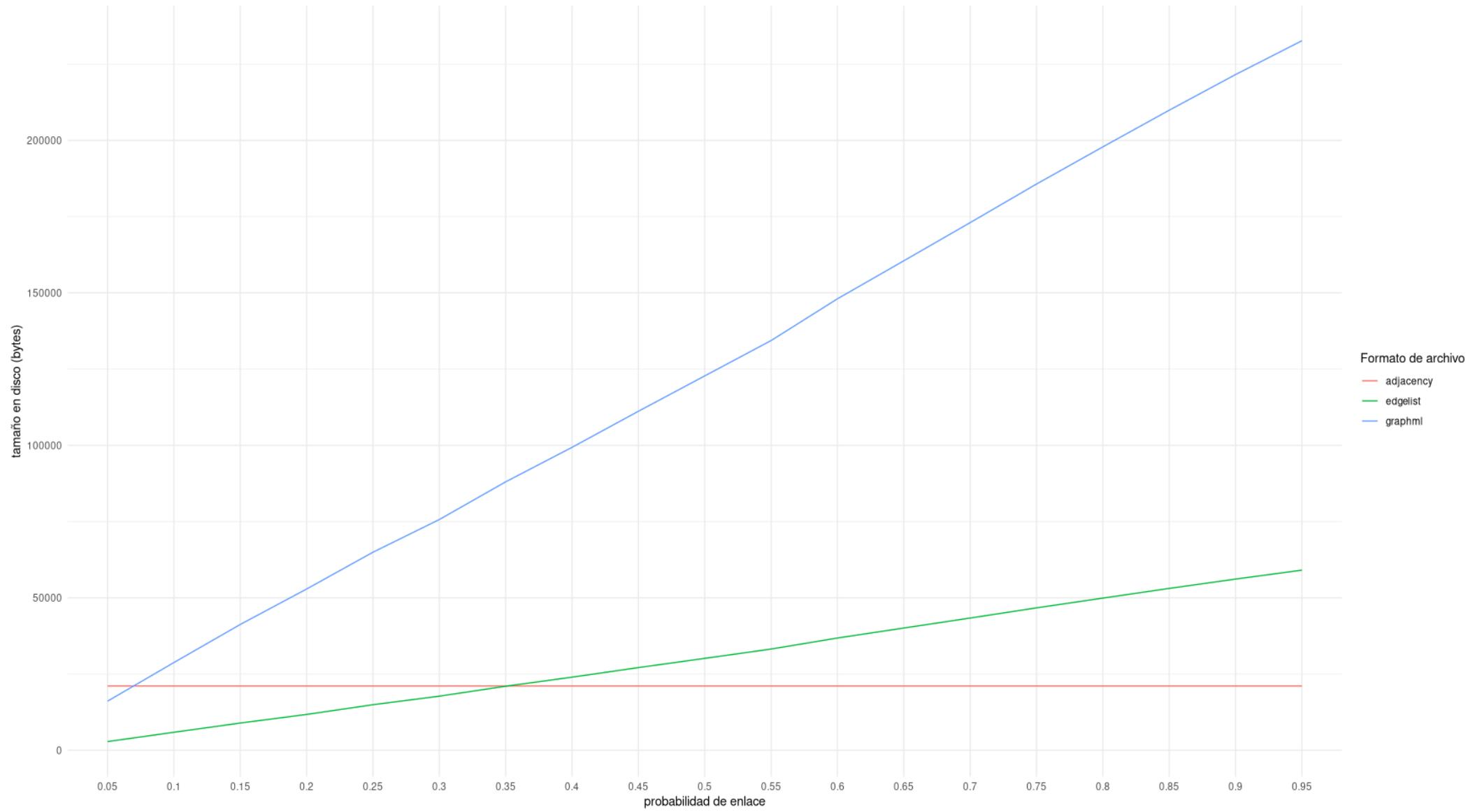


Puedo llevar (meta)datos de los enlaces, pero no de los nodos.  
Tampoco nodos sueltos

# graphml

```
<?xml version="1.0" encoding="UTF-8"?>
<graphml xmlns="http://graphml.graphdrawing.org/xmlns"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://graphml.graphdrawing.org/xmlns
    http://graphml.graphdrawing.org/xmlns/1.0/graphml.xsd">
  <!-- Created by igraph -->
  <key id="g_name" for="graph" attr.name="name" attr.type="string"/>
  <key id="g_type" for="graph" attr.name="type" attr.type="string"/>
  <key id="g_loops" for="graph" attr.name="loops" attr.type="boolean"/>
  <key id="g_p" for="graph" attr.name="p" attr.type="double"/>
  <key id="v_name" for="node" attr.name="name" attr.type="string"/>
  <graph id="G" edgedefault="undirected">
    <data key="g_name">Erdos renyi (gnp) graph</data>
    <data key="g_type">gnp</data>
    <data key="g_loops">false</data>
    <data key="g_p">0.1</data>
    <node id="n0">
      <data key="v_name">a</data>
    </node>
    <node id="n1">
      <data key="v_name">b</data>
    </node>
    <node id="n2">
      <data key="v_name">c</data>
    </node>
    <node id="n3">
      <data key="v_name">d</data>
    </node>
    <node id="n4">
      <data key="v_name">e</data>
    </node>
    <node id="n5">
      <data key="v_name">f</data>
    ...
    ...
    ...
    <edge source="n1" target="n9">
    </edge>
    <edge source="n7" target="n9">
    </edge>
    <edge source="n3" target="n11">
    </edge>
    <edge source="n7" target="n12">
    </edge>
    <edge source="n10" target="n12">
    ...
    ...
  </graph>
</graphml>
```

Otras opciones:  
gml  
json



# Comentarios finales

- Las redes son una herramienta muy poderosa para entender toda clase de fenomenos complejos.

# Comentarios finales

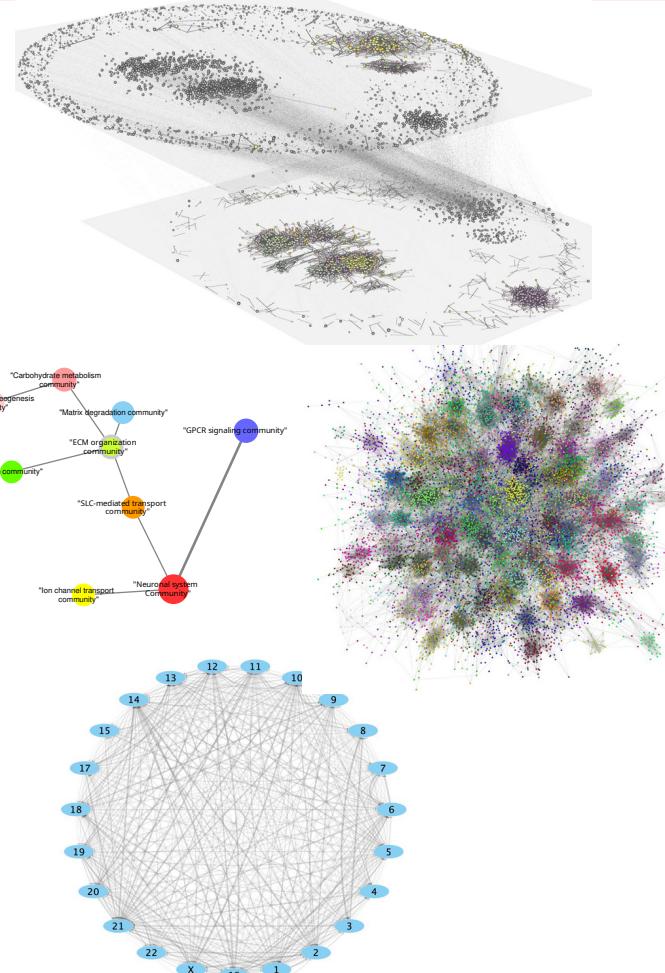
- Al analizar los fenómenos desde una perspectiva integrada, podemos identificar características que no pueden observarse en sus componentes individuales.

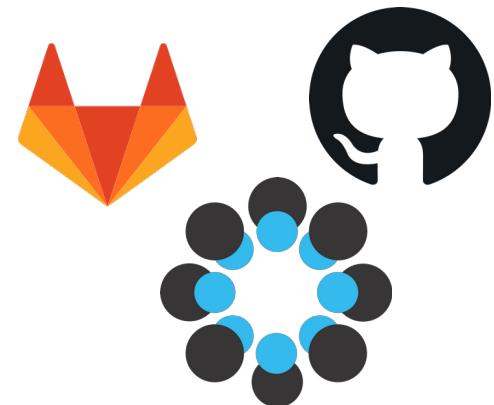
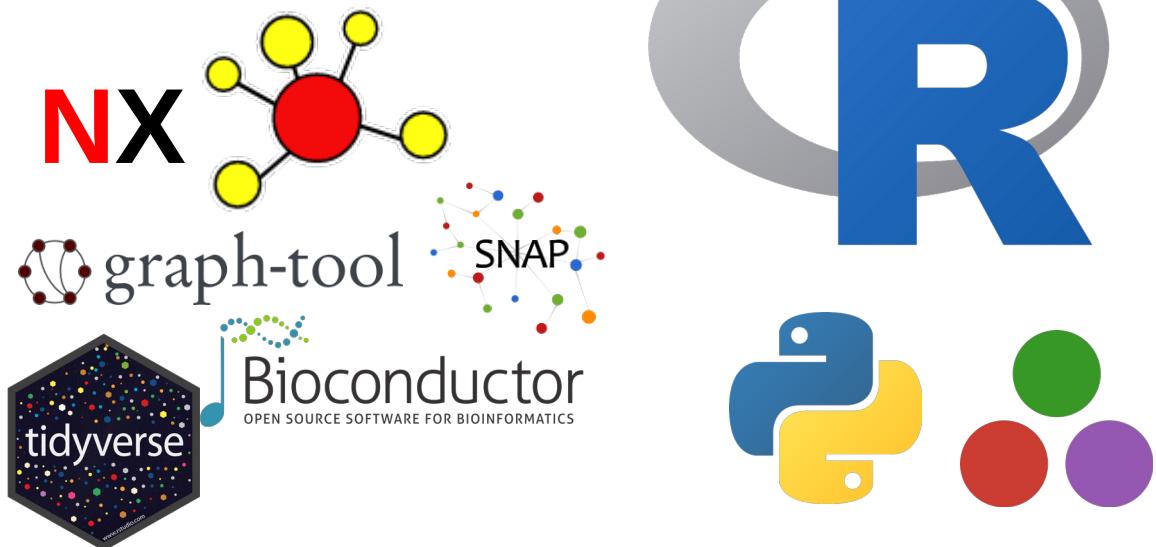
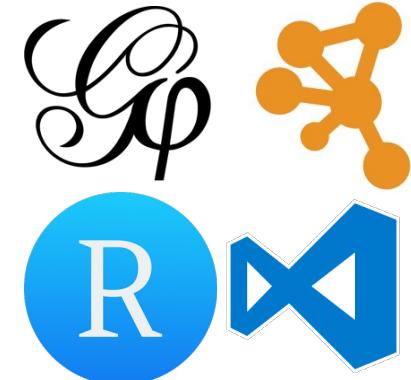
# Comentarios finales

- Podemos aprender mucho de los componentes individuales si entendemos el sistema en el que se encuentran

# Proyectos actuales...

- Redes **multiescala / multiómicas** para cáncer
- Perturbación funcional de fármacos
- RA de **antibióticos** por perturbación del **microbioma**
- Aprendizaje de máquina para descubrir **mecanismos de acción** de anticancerígenos.
- Bastante más cosas de COVID-19 que las que me gustaría





# ¡Gracias por su participación!

## Contacto:

- [gdeanda@inmegen.edu.mx](mailto:gdeanda@inmegen.edu.mx)
- [guillermodeandajauregui.github.io](https://guillermodeandajauregui.github.io)
- TW: @gdeandajauregui