# El complejo de los grafos

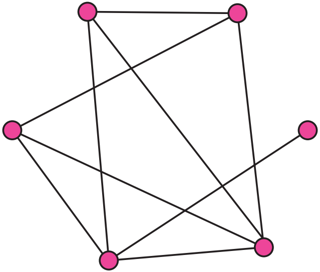
Flavia Bonomo

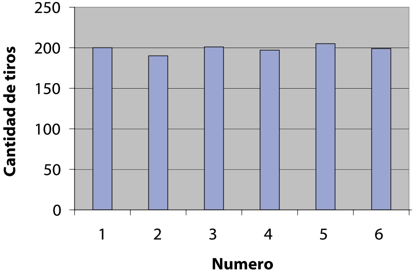
Charla de borrachos, cosecha 2011

# Generando grafo para testear algoritmos

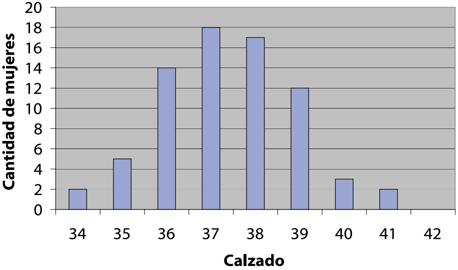
# Generando grafo para testear algoritmos

Modelo de Erdös-Renyi $G(n, 0.5)$: Para definir si una arista existe o no, tiro una moneda.





Cantidad de tiros
250
200
150
100
50
0
1
2
3
4
5
6
Numero

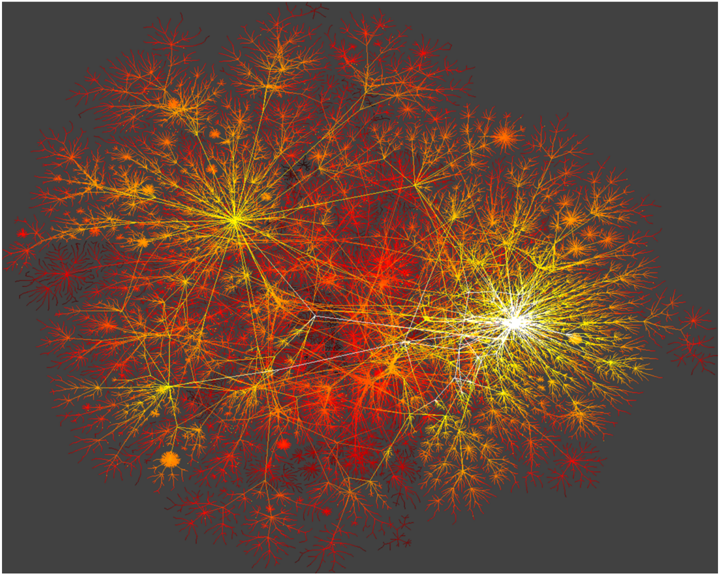

Cantidad de mujeres
20
18
16
14
12
10
8
6
4
2
0
34
35
36
37
38
39
40
41
42
Calzado

# Redes complejas

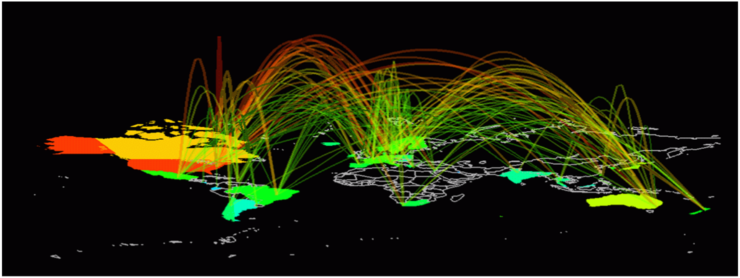
**Grafos muy grandes** de la vida real. Algunas de las principales preguntas sobre ellos son:

- Cuáles son las estadísticas de este tipo de redes?
- Podemos explicar su proceso de formación?

# Ejemplo: el grafo de Internet



# Internet



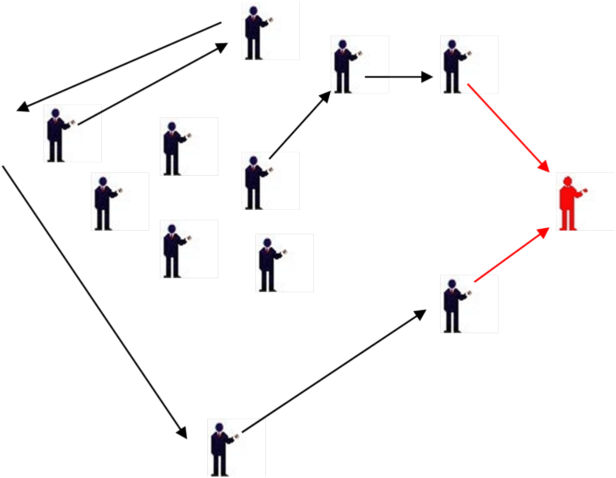
# Más ejemplos

- Redes sociales
  - grafo de amigos (ahora modelado en parte por facebook)
  - redes de colaboración científica
  - redes de mensajes de texto o llamadas telefónicas
  - tratados bilaterales de inversión
- Redes tecnológicas
  - Internet
  - redes de teléfono
  - redes de transporte
- Redes biológicas/psicológicas
  - interacción proteina-proteina
  - regulación de genes
  - cadena alimentaria
  - interacción de neuronas
  - asociación de palabras

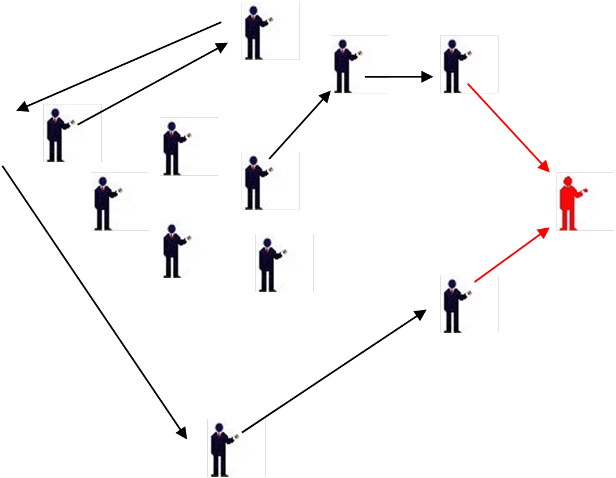
# Bibliografía fundacional


- Watts and Strogatz, “Dynamics and small-world phenomenon”
- Faloutsos, Faloutsos and Faloutsos, “On power-law relationships of the Internet Topology”
- Kleinberg et al., “The Web as a graph”
- Barabasi and Albert, “The emergence of scaling in real networks”

# El experimento de Milgram: Small World

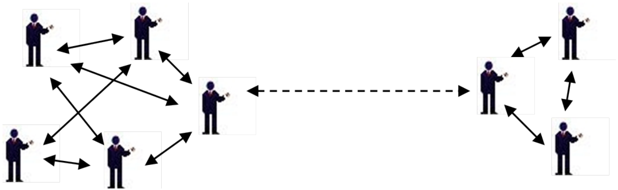


# El experimento de Milgram: Small World



Unas 65 cartas de las casi 300 enviadas llegaron a destino, en 5.5 pasos promedio.

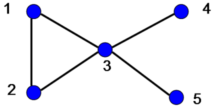
# El experimento de Milgram: Small World



**Conjetura:** Una parte importante del mundo está en una gran componente conexa de diámtro a lo sumo 6.

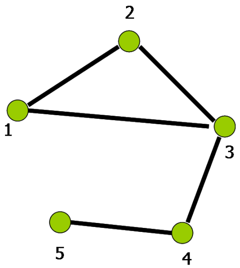
# Coeficiente de clustering

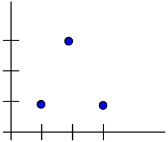
- coeficiente de clustering del vértice $i$:
  - si $d(i) > 1$, el número de aristas entre vecinos de $i$ dividido por $d(i)(d(i)-1)/2$
  - si $d(i) \leq 1$ puede ser definido como 0 o 1
- coeficiente de clustering del vértice 3: $1/6$
- coeficiente de clustering del vértice 1: 1



# Distribución de grados

- secuencia de grados: $[d(1), d(2), d(3), d(4), d(5)] = [2, 2, 3, 2, 1]$
- distribución de grados: $[(1, 1), (2, 3), (3, 1)]$

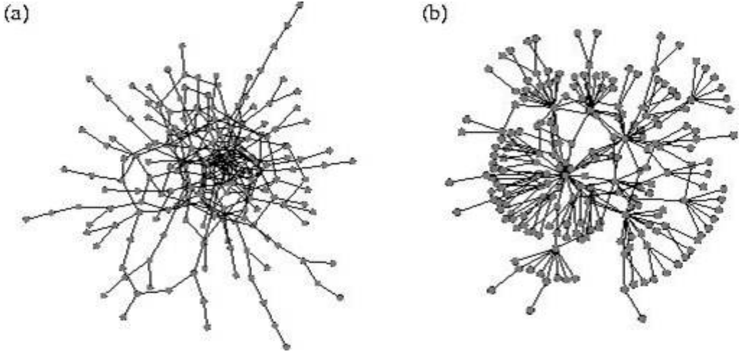


# Caracterización de redes complejas

- Diámetro, coeficiente de clustering, distribución de grados.
- Centralidad: Número de caminos mínimos que pasan por un vértice (cómo afecto la conectividad o el diámetro sacando ciertos vértices).
- Comunidades: estructura de las cliques
- Subestructuras locales: a veces la red es una red de bloquecitos con una estructura interna distinta que la de la red general.
- "Assortativity": los vértices de grado alto se conectan preferentemente entre si?

# Assortativity



# Características típicas de una red real

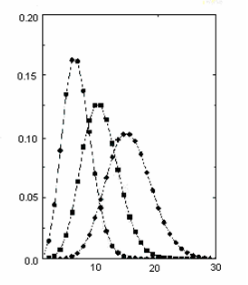
- Muchos vértices tienen grado bajo pero existe una cantidad no despreciable de vértices de grado muy alto (distribución de grados siguiendo una power-law)
  - redes scale-free
- Si un vértice $x$ es vecino de $y$ y $z$, entonces $y$ y $z$ están conectados con una alta probabilidad.
  - coeficiente de clustering alto
- La mayoría de los vértices están cerca entre si.
  - redes small world

Redes de áreas diversas (internet, redes biológicas, etc.) tienen propiedades similares

- Es posible describir un proceso común de generación?

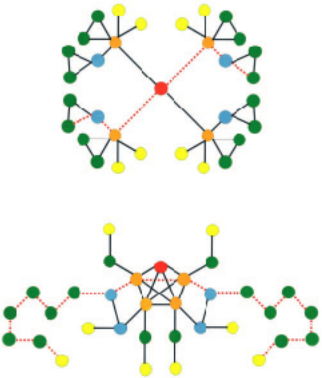
# Modelo de Erdős-Renyi

- La distribución de grados es una Gaussiana alrededor de la media (la probabilidad con la que determino las aristas)




- la probabilidad de dos vértices de estar conectados es independiente de que tengan o no vecinos comunes.
- las distancias promedio son cortas, pero no constantes

# Distribuciones de grados



Estos grafos tienen la misma distribución de grados pero muy distinto fiámetro, modularidad, simetría, robustez...

# Distribuciones power-law

- Muchas de las redes reales siguen una distribución power law

$$P(k) = Ck^{-\alpha}$$

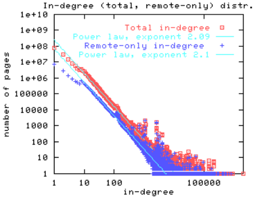
  - tienen una cantidad no despreciable de vértices de grado alto (hubs)
  - scale-free: el promedio de los grados no es de gran información.
- En contraste con el modelo random...
  - Distribución de Poisson

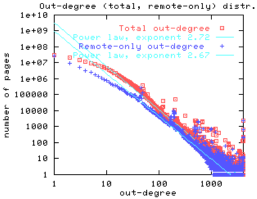
$$P(k) = \frac{(np)^k}{k!} e^{-np}$$

  - muy concentrados en la media
  - la probabilidad de vértices de grado alto es exponencialmente baja

# Ejemplo: el grafo de la WWW

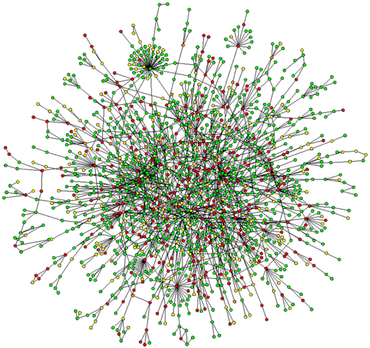
- Distribución de grados de entrada: Power-law con exponente 2.1
- Distribución de grados de salida: Power-law con exponente 2.7
- El hecho de que el exponente sea mayor que 2 implica que el valor esperado del grado es constante (no crece con $n$).
- Entonces, la cantidad de links crece lineal con la cantidad de páginas $n$.



# Grados esperados

- Promedio:
  - Grafos random: $np$.
  - Grafos scale-free: constante si $\alpha \geq 2$, y diverge si $\alpha < 2$.
- Máximo:
  - Grafos random: típicamente, no mucho mayor que la media.
  - Grafos scale-free: $k_{\mathrm{max}} \approx n^{1/(\alpha-1)}$.



# Modelos de generación

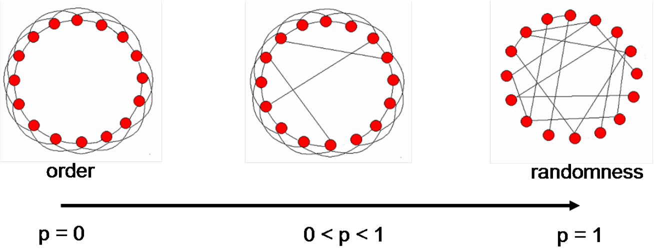
- Grafos Random (Erdős-Renyi)
- Preferential attachment (Price, Albert, Barabasi, Jeong)
- Modelos de copia (Kumar, Raghavan)
- Small World (Watts, Strogatz)
- Fitness (Caldarelli, Capocci, De Los Rios, Muñoz)

# Preferential Attachment

- Introducido por Price (1965) como modelo para redes de citas (científicas! ;))
  - cada paper nuevo es generado con $m$ citas (media)
  - un paper nuevo cita a otro con una probabilidad proporcional a la cantidad de citas previas más uno (para darle alguna chance a los nuevos, sino no anda).
- Power law con exponente $\alpha = 2 + 1/m$.
- El modelo de Barabasi-Albert es similar y se obtienen grafos con distribución power law con exponente $\alpha = 3$.

# Grafos Small World (Watts and Strogatz, 1998)

- Se empieza con un ciclo donde cada vértices está conectado con los siguientes $z$ vértices.
- Con probabilidad $p$, intercambio dos aristas (o, agrego una arista a un destino elegido uniformemente).



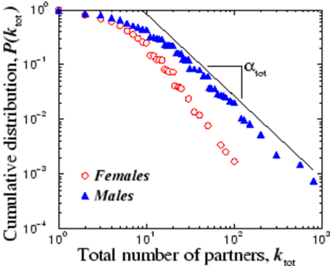

- Para $0 < p < 1$, tenemos coeficiente de clustering alto y diámetro chico.

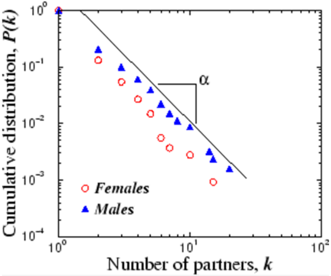
# Difusión en redes o epidemias

Redes sociales y difusión de rumores (Ej: The Big Bang Theory, “The Herb Garden Germination”, 4ta temporada).

# Difusión en redes o epidemias

Redes sexuales (Lilijeros et al., Nature (2001))




Encuesta en Suecia en 1996 a 4.781 personas sobre la cantidad de parejas. Esta red es scale-free, a diferencia de las redes de amistad.