



Redes y Sistemas Complejos
Cuarto Curso del Grado en Ingeniería Informática
Tema 4: Algoritmos de Poda y Visualización de Redes

Oscar Cordon García

Dpto. Ciencias de la Computación e Inteligencia Artificial
ocordon@decsai.ugr.es

VISUALIZACIÓN DE REDES

INTRODUCCIÓN

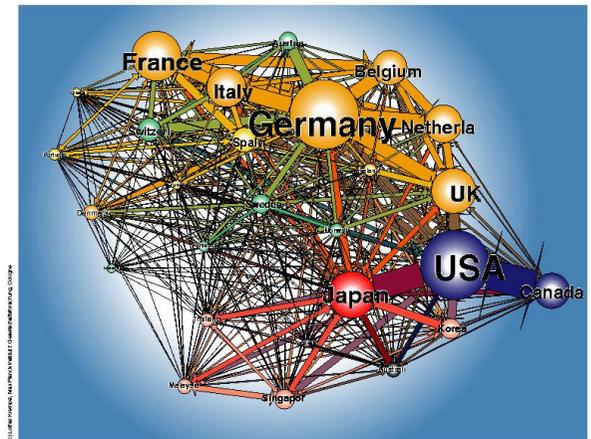
Necesidad de la simplificación y visualización de redes

La **visualización de información** tiene el potencial necesario para ayudar a las personas a acceder a la información necesaria de una forma más eficaz e intuitiva

La visualización de información comprende dos aspectos directamente relacionados: **Modelado estructural** y **Representación gráfica**

El propósito de la **representación gráfica** es transformar un modelo previo de una estructura en un modelo gráfico que permita examinar visualmente la estructura original e interactuar con ella

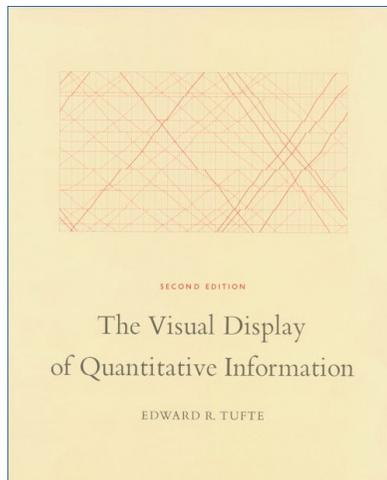
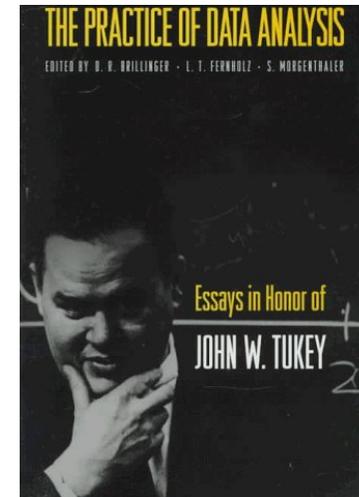
En el dominio de la **visualización de redes complejas**, la gran dimensión de las redes genera dificultades para obtener representaciones gráficas útiles para el análisis



Una misma red puede representarse gráficamente de diversas formas → uso de distintos algoritmos de distribución (*layout*), con distintas filosofías

John W. Tukey:

- Fue el que acuñó los términos *software* y *análisis exploratorio de datos*, así como la expresión: “*es mejor tener una respuesta aproximada a la pregunta correcta que una respuesta precisa a la pregunta equivocada*”



Edward F. Tufte:

- Su libro de 1983 popularizó las ideas de Tukey: “el éxito de una visualización se basa en el conocimiento profundo y el cuidado de la sustancia, así como en la calidad, relevancia e integridad del contenido”
- En nuestro ámbito, **¡conoce la red!**

Tufte enunció los cinco principios básicos de la visualización de datos:

- **Lo principal, mostrar los datos**
- Maximizar el *ratio data-ink*, dentro de lo razonable
(*data-ink* es la “tinta” usada para la presentación de los datos. Si se eliminara, el gráfico perdería contenido. *Non-Data-Ink* es la “tinta” que no contiene información y se usa para escalas, etiquetas y enlaces)
- Eliminar el contenido que no sea *data-ink*, dentro de lo razonable
- Eliminar el contenido *data-ink* redundante
- Revisar y editar

<http://www.edwardtufte.com/tufte/>



Tamaño de la red:

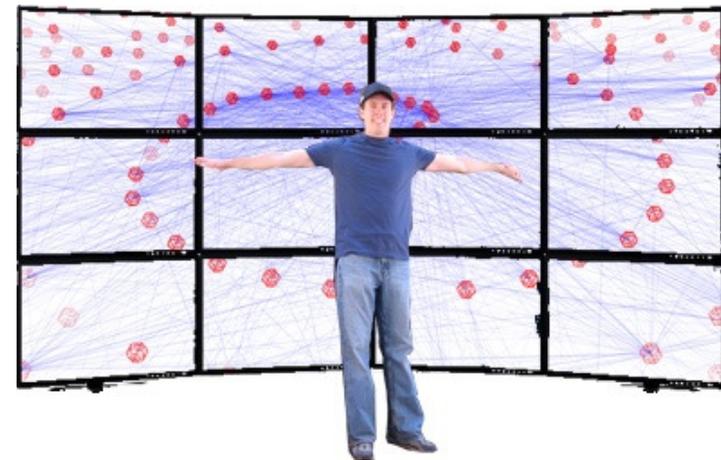
A mayor número de nodos, mayor número de píxeles serán necesarios para visualizar la red en la pantalla:

$800 \times 600 = 480,000$ pixels

$1024 \times 768 = 786,432$ pixels

$1920 \times 1200 = 2,304,000$ pixels

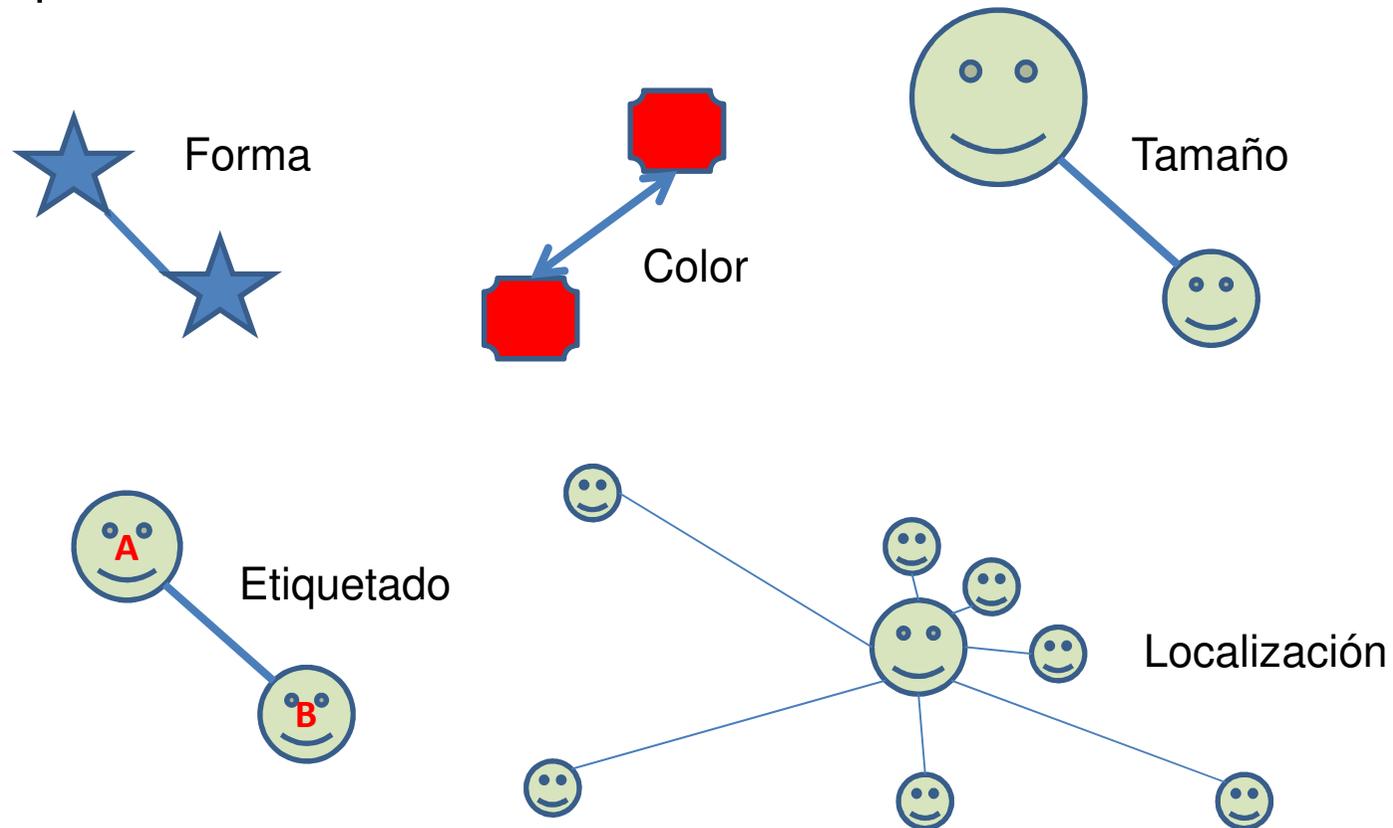
¡Puede ocurrir que no tengamos suficientes píxeles para visualizar la red!



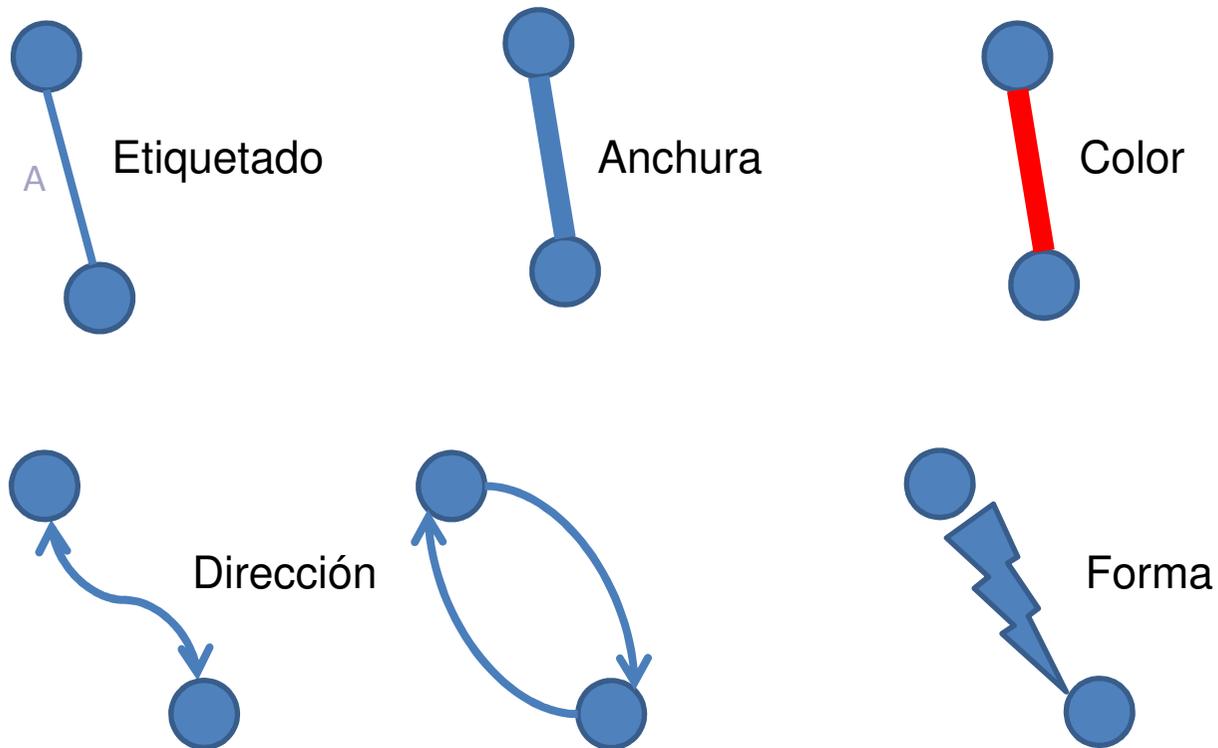
Cuanta más información tengamos que presentar en pantalla, más importancia tendrá el diseño y la gestión de la ventana de visualización

Aspectos de distribución (*layout*):

¿Cómo representar un nodo?



Representación de los enlaces:

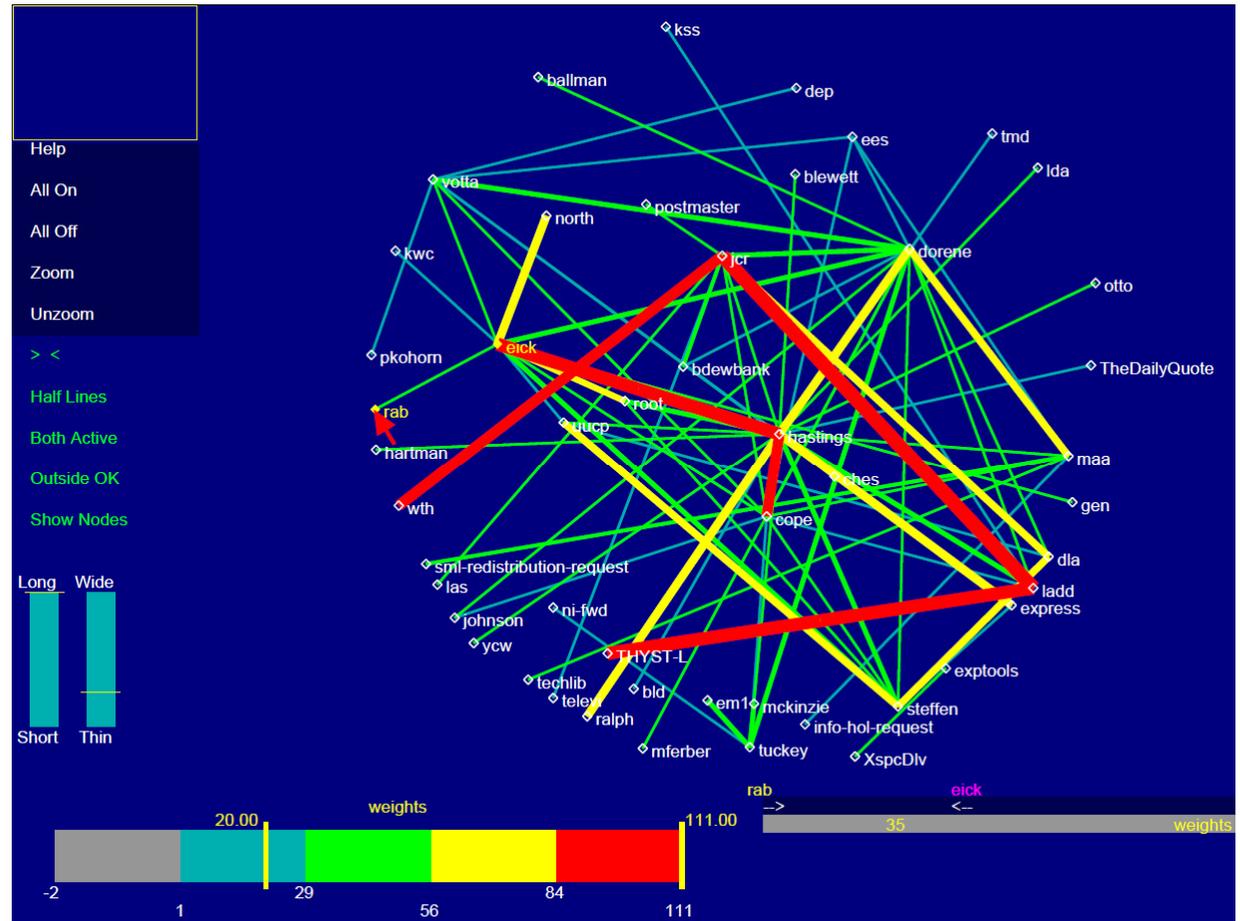


Ejemplo: anchura de los enlaces:

E-mails en un departamento de los AT&T Bell Labs:

- 1 mes de datos
- 1500 usuarios
- Umbral: usuarios con más de 20 mensajes

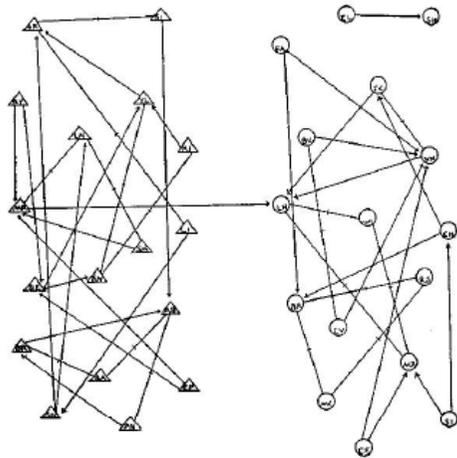
R.A. Becker, S.G. Eick, A.R. Wilks. 1995. [Visualizing Network Data](#). IEEE Trans. Visualization and Computer Graphics 1(1): 16-21



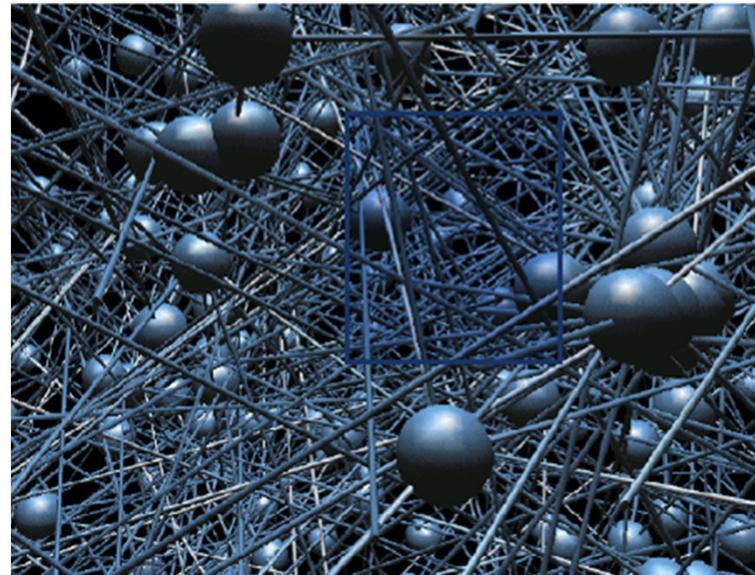
INTRODUCCIÓN

Aspectos básicos de la visualización de redes (4)

La visualización de un subconjunto de una red resaltando los atributos de los nodos mediante forma y color mejora su comprensión:



Una red de atracción en una clase de cuarto curso (**Moreno**, 'Who shall survive?', 1934).



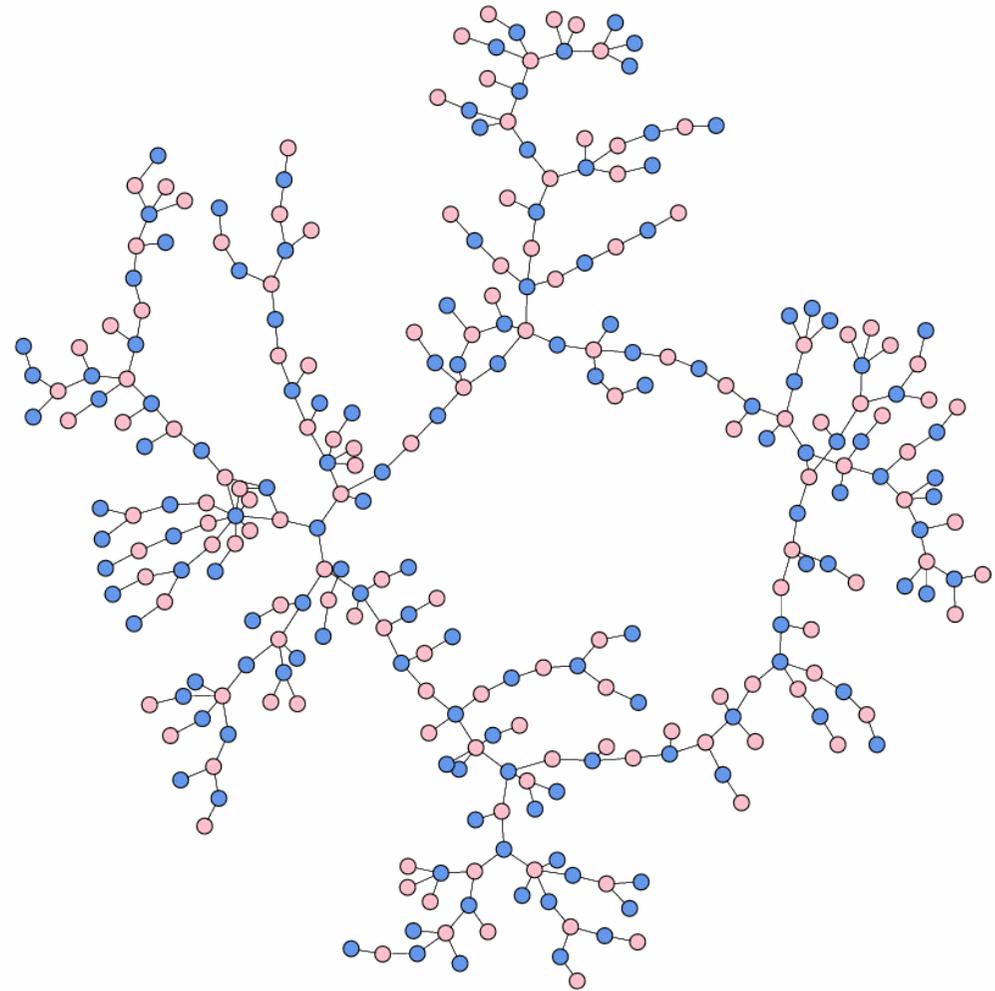
Alden Klovdahl: Parte central ($n \sim 450$) de una red social de unos 5000 residentes en Canberra, Australia

Ejemplo: color = género:

Citas en el instituto:

P.S. Bearman, J. Moody, K. Stovel.
2004. [Chains of affection: The structure of adolescent romantic and sexual networks](#). *American Journal of Sociology* 110: 44-91

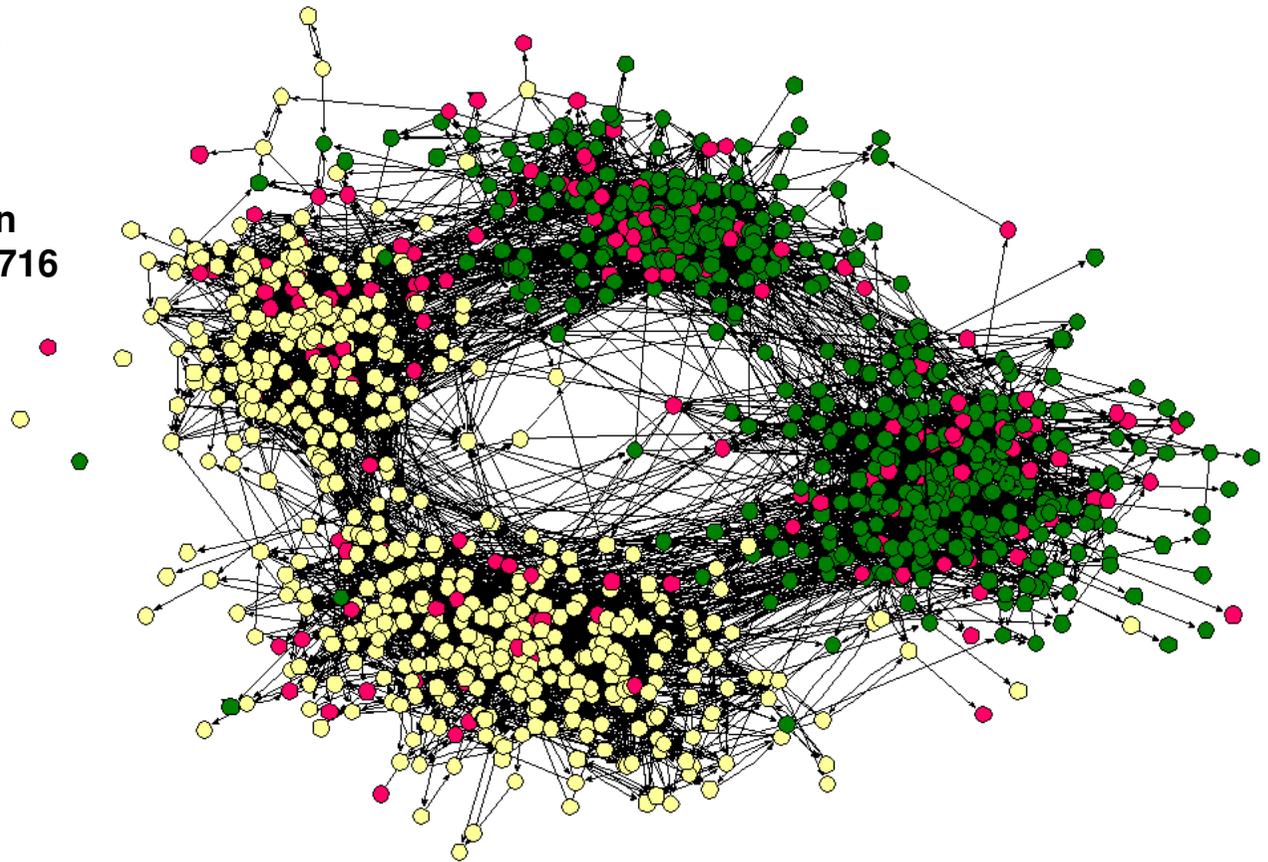
(imagen de Mark Newman)



Ejemplo: forma de los nodos:

Integración racial en el colegio:

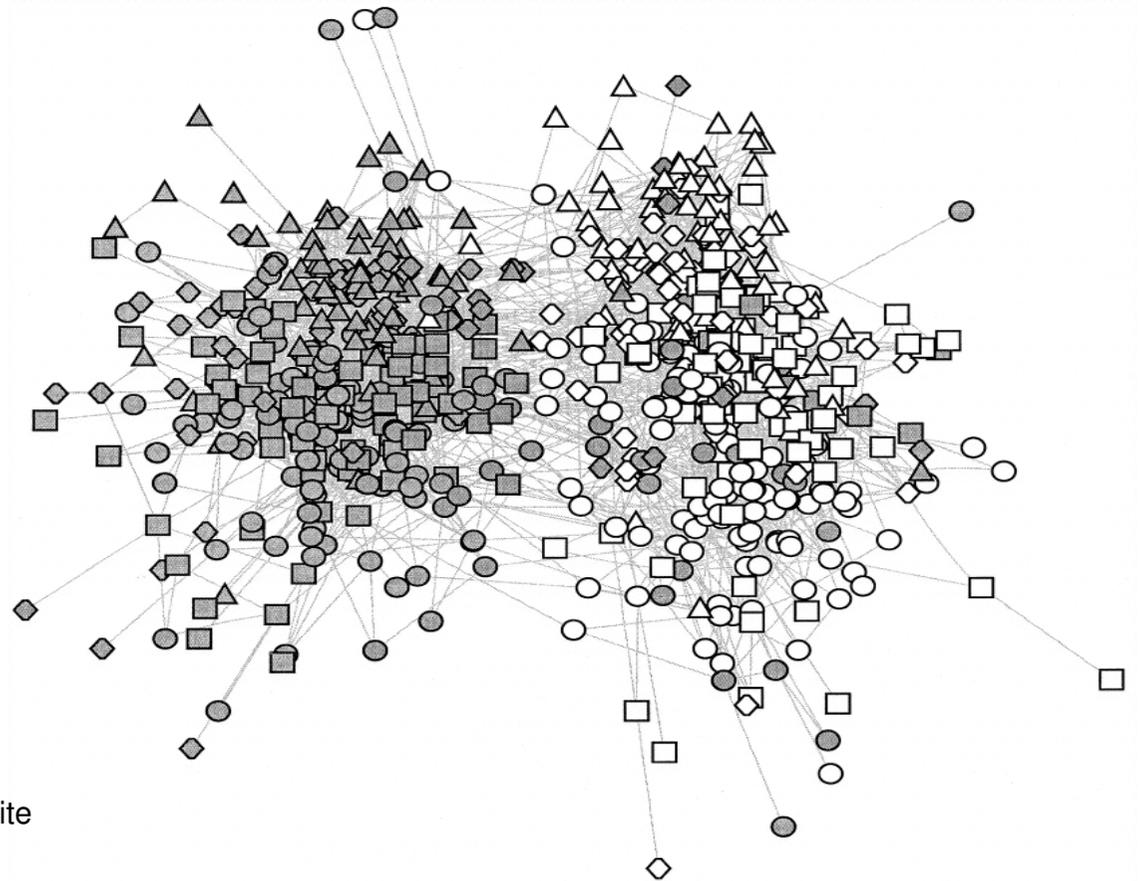
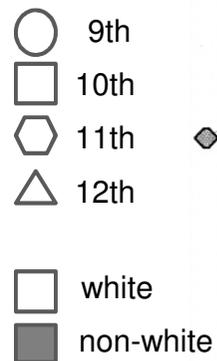
**J. Moody. 2001. Race, School
Integration, and Friendship
Segregation in America. American
Journal of Sociology 107(3): 679-716**



Ejemplo: forma de los nodos:

Integración racial en el colegio:

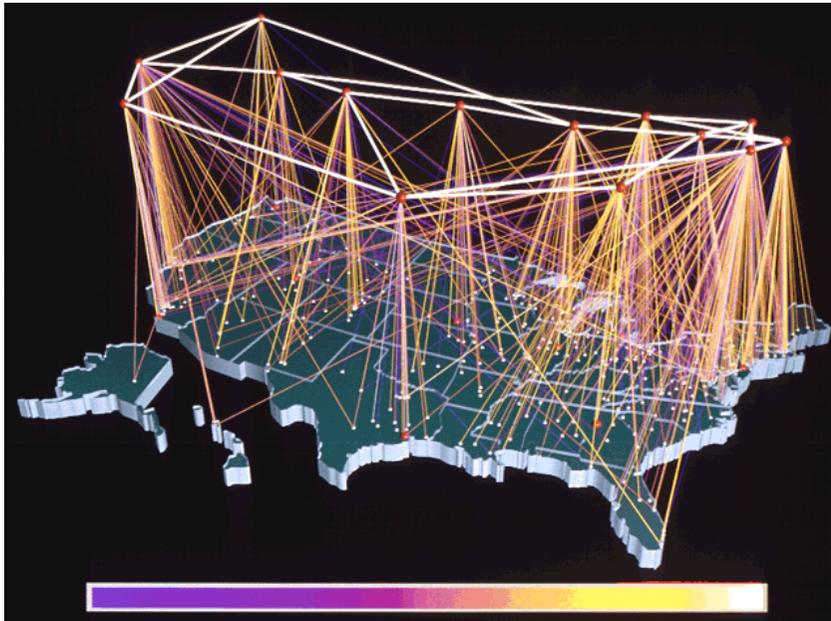
**J. Moody. 2001. Race, School
Integration, and Friendship
Segregation in America. American
Journal of Sociology 107(3): 679-716**



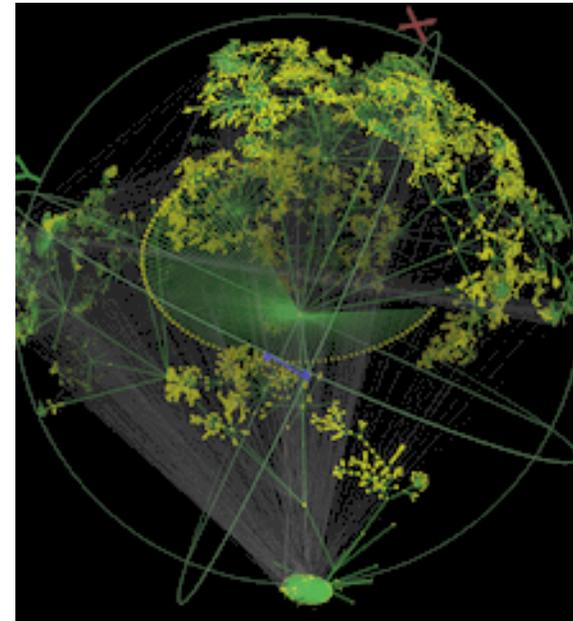
INTRODUCCIÓN

Aspectos básicos de la visualización de redes (8)

Posibilidad de combinar la visualización de la red con información geográfica:



Tráfico de datos en el backbone de la red
ANS/NSFnet T3 en Noviembre de 1993
(**Cox & Patterson**)



Imágenes Walrus de datos *Skitter* de Internet
(54893 nodos y 79409 enlaces)

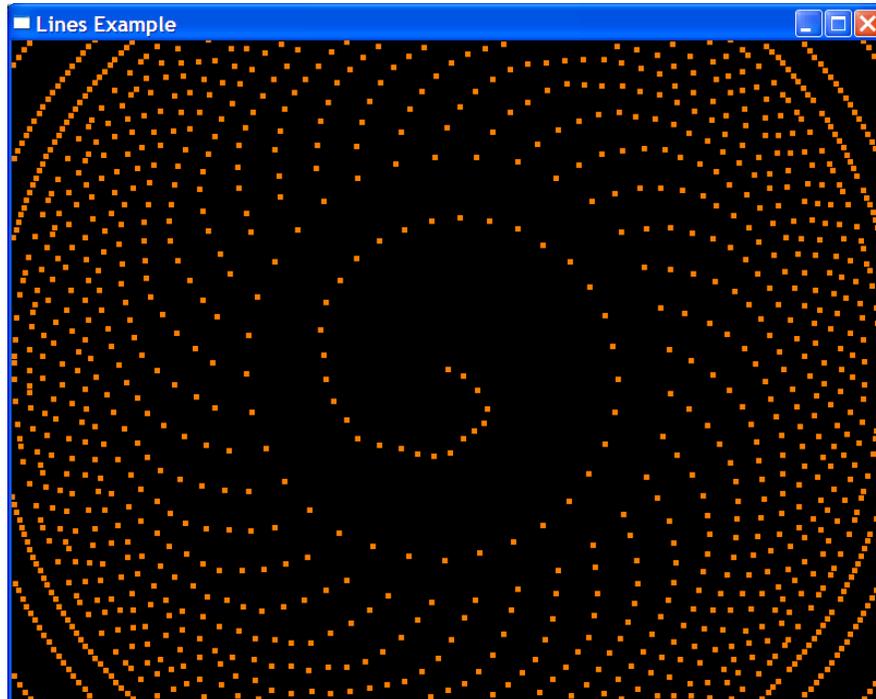
Walrus está disponible con licencia GPL:
<http://www.caida.org/tools/visualization/walrus/gallery1/>

DISTRIBUCIÓN (*LAYOUT*) DE REDES

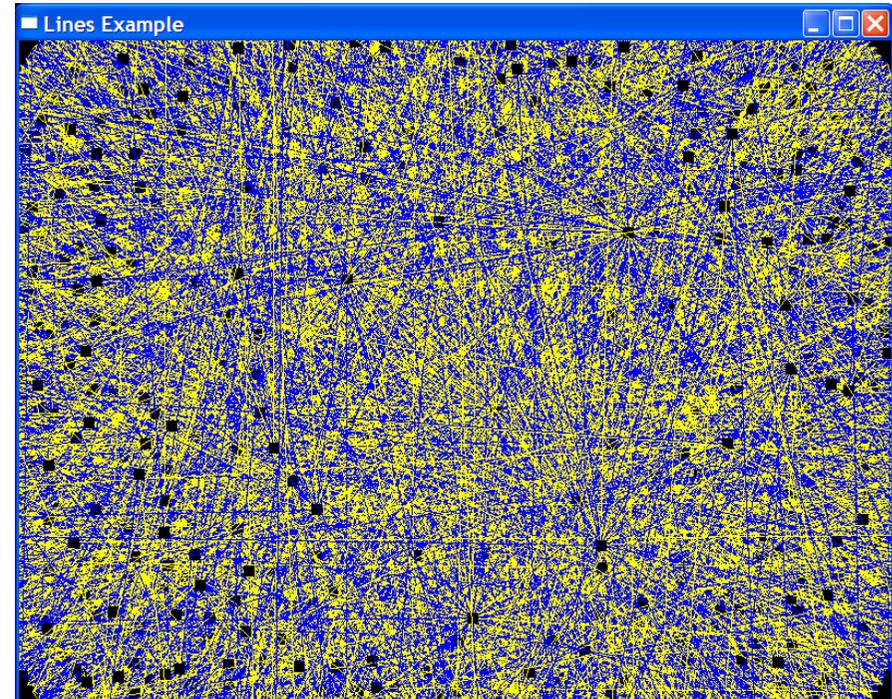
PROBLEMAS DE LA DISTRIBUCIÓN (*LAYOUT*) DE REDES

¡¡El problema de los enlaces es que pueden tapar partes del grafo!!

Antes de pintar los enlaces

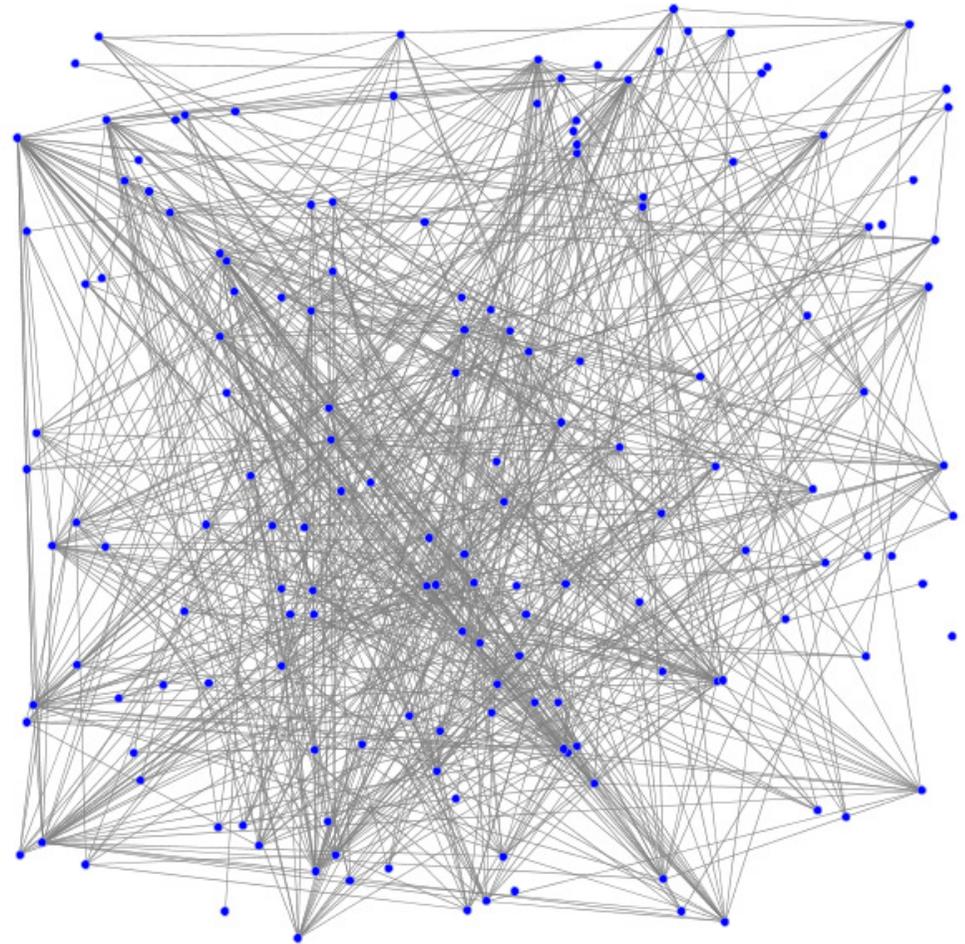


Después de pintar los enlaces



Se escogen aleatoriamente las coordenadas (x,y) para cada nodo

- **Ventaja:** muy rápida
- **Inconveniente:** imposible de interpretar

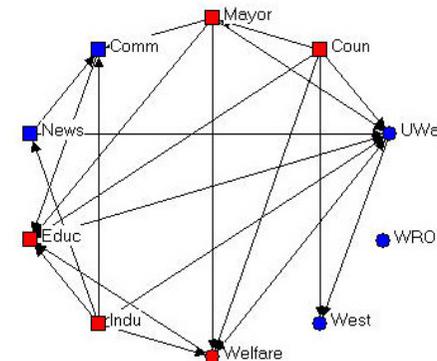
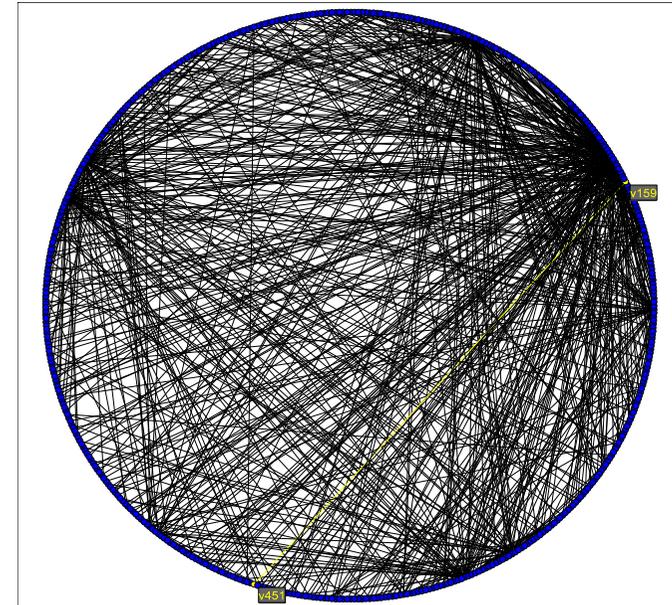


DISTRIBUCIONES BÁSICAS

Circular

Se distribuyen los nodos en círculo y se dibujan los enlaces entre ellos

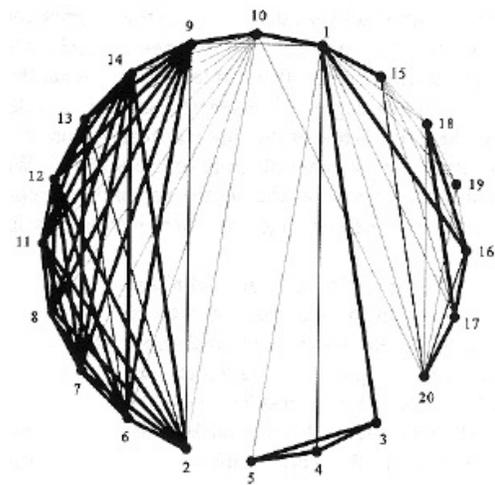
- **Ventajas:**
 - Muy rápida
 - Las coordenadas circulares pueden representar una propiedad de los datos (p.ej. latitudes o edades)
- **Inconvenientes:** difícil de interpretar en redes grandes y densas:
 - mucho solapamiento de enlaces
 - muchos enlaces largos (los nodos conectados no tienen por qué estar cerca)
 - Es difícil identificar clusters



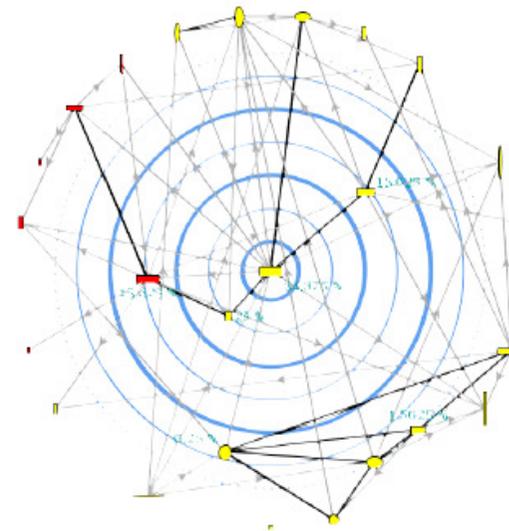
DISTRIBUCIONES BÁSICAS

Radial (1)

- Se distribuyen los nodos en círculo
- Se localiza el **nodo central** en el centro de la ventana de visualización
- Se distribuyen sus vecinos en círculos concéntricos de radio equivalente al peso de su enlace



(a) circular [Mit94]



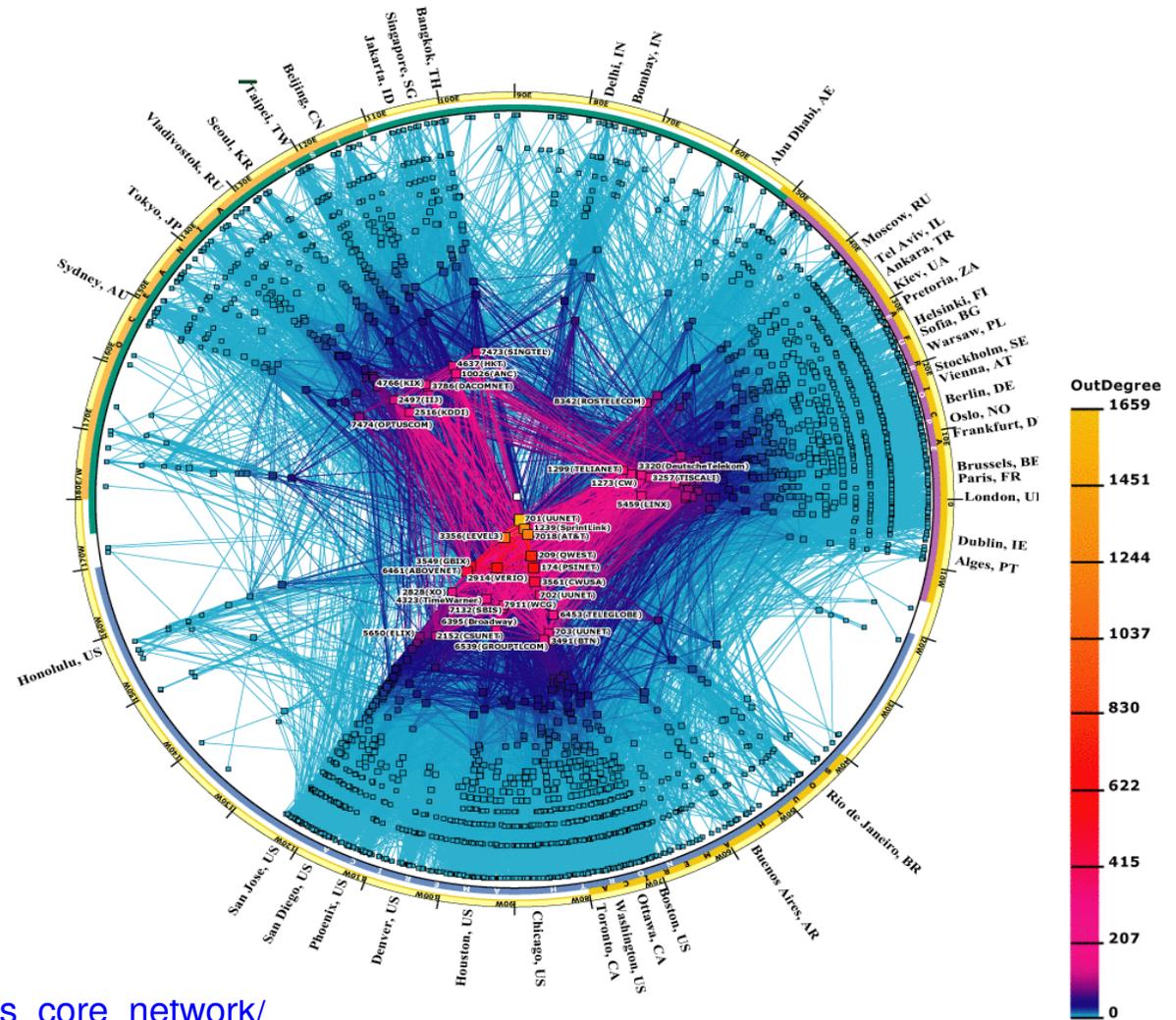
(b) radial [BW04]

DISTRIBUCIONES BÁSICAS

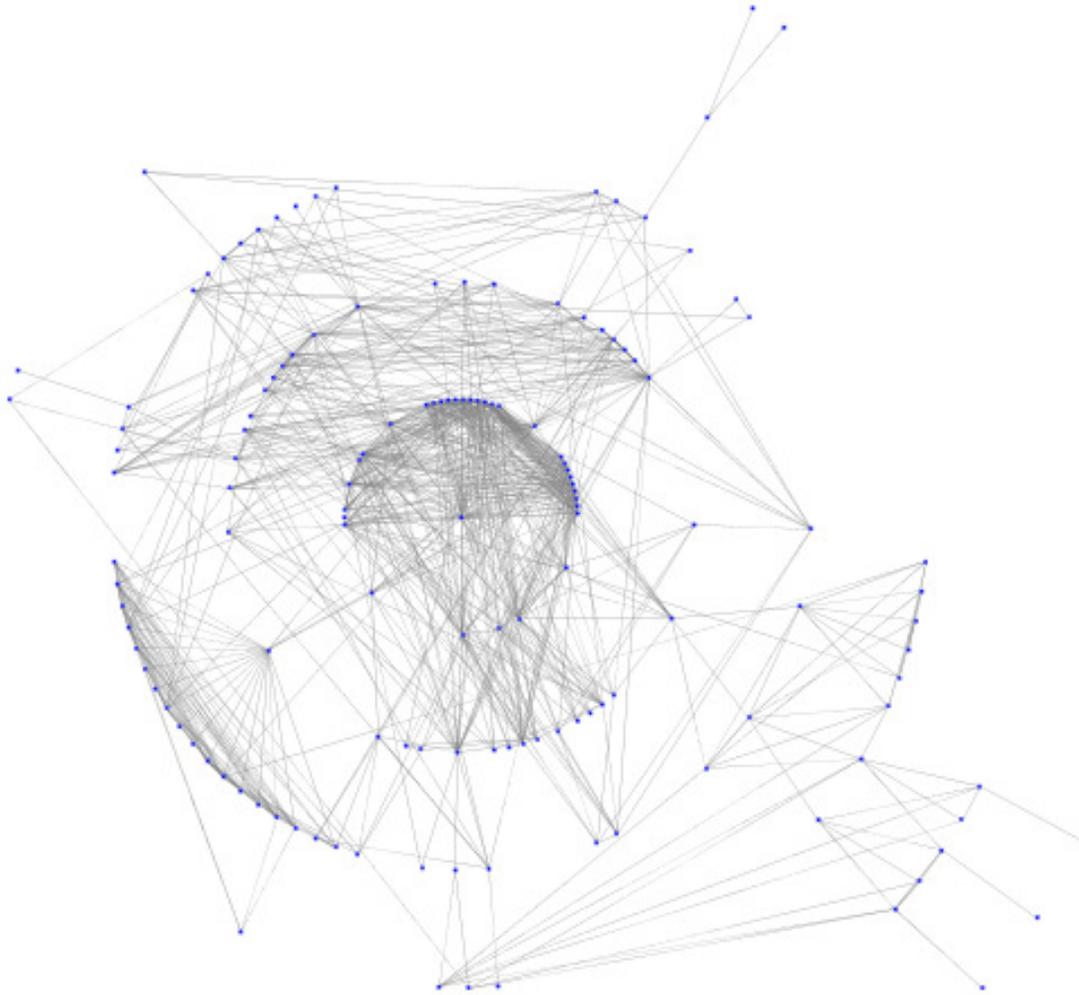
Radial (2)

Ejemplo:

- Grafo de Internet IPv4
- Mapa de Internet a nivel de sistemas autónomos (*AS-level*)



http://www.caida.org/research/topology/as_core_network/

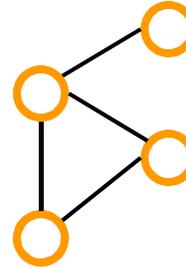


Se comienza con un nodo y se dibujan todos los demás nodos en capas semicirculares de acuerdo a cuántos saltos hay que dar para alcanzar a cada uno de ellos desde el inicial

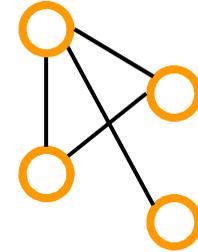
- **Ventajas:** Muy rápida
- **Inconvenientes:** No se tiene en cuenta que los nodos que estén conectados entre sí deberían estar cercanos en cada capa

CRITERIOS ESTÉTICOS PARA VISUALIZACIÓN DE REDES (1)

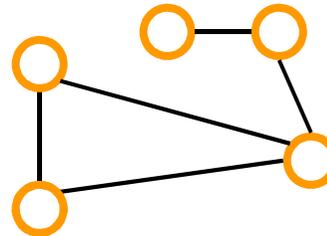
- Minimizar los cruces entre enlaces:



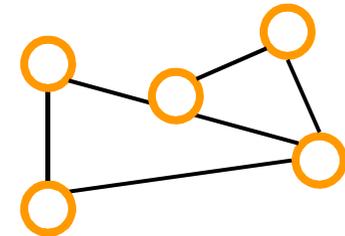
mejor que



- No permitir que los nodos se superpongan a enlaces que no incidan en ellos:

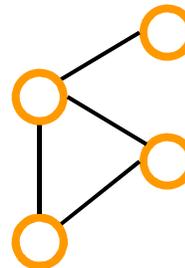


mejor que

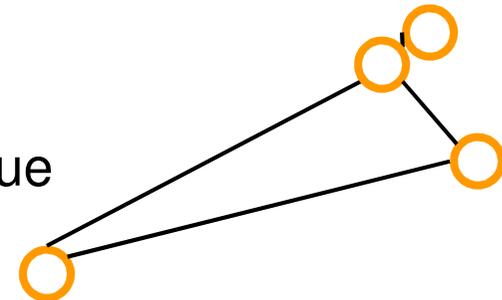


- En redes no ponderadas, mantener una longitud de enlace uniforme:

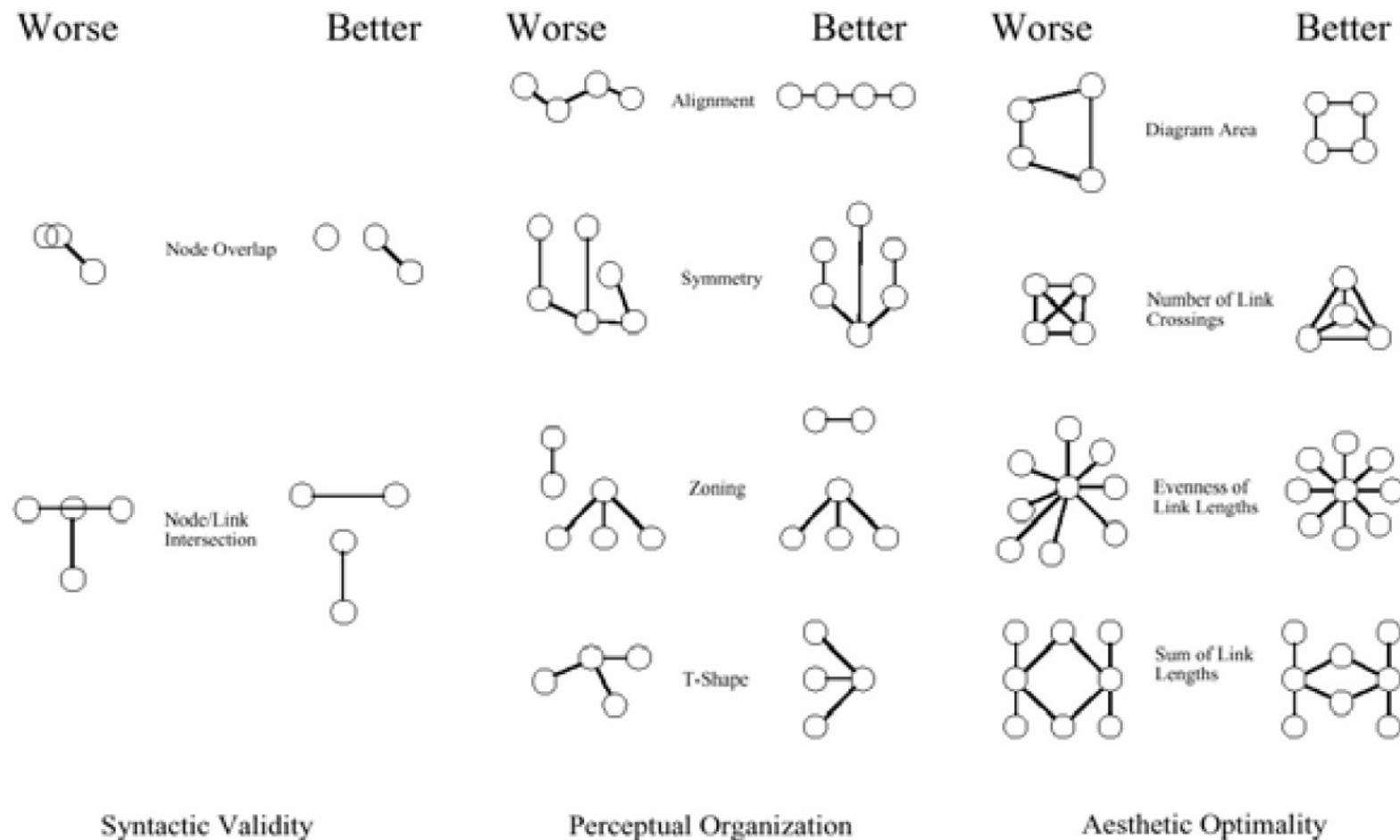
En redes ponderadas, la longitud de cada enlace debe ser proporcional a su peso



mejor que



CRITERIOS ESTÉTICOS PARA VISUALIZACIÓN DE REDES (2)



Preferred graph layout heuristics (Kosak et al., 1994). © 1994 IEEE.

MÉTODOS DE DISTRIBUCIÓN
GUIADOS POR FUERZAS
(*FORCE-DIRECTED PLACEMENT*)

INTRODUCCIÓN

Los **métodos de distribución guiados por fuerzas** son algoritmos de dibujo de grafos generales no dirigidos con enlaces de líneas rectas en el plano ([demo](#))

En general, fueron propuestos para verificar distintos criterios estéticos:

1. Distribución uniforme de los nodos
2. Longitud uniforme de los enlaces
3. Minimización de los cruces (superposiciones) entre enlaces
4. Simetría

Cada método pone un énfasis especial en algunos de estos criterios

Los criterios pueden ser mutuamente excluyentes. P.ej. un grafo simétrico puede requerir un cierto número de cruces entre enlaces, incluso si se pueden evitar

Un enfoque pragmático consiste en darle a los algoritmos la suficiente flexibilidad para adaptarse a aplicaciones concretas

http://en.wikipedia.org/wiki/Force-directed_graph_drawing

EL PROBLEMA

Problema: Dado un conjunto de nodos y enlaces, calcular las posiciones de los nodos en el plano

El objetivo es doble, pintar el grafo *bien* (uso de heurísticas) y pintarlo *rápido* (necesidad de escalado para grafos grandes)

Es un problema muy complicado porque presenta muy pocas restricciones

Los *métodos de distribución guiados por fuerzas* se propusieron inicialmente para el diseño VLSI para optimizar el diseño de un circuito con el menor número posible de cruces entre líneas

Existen muchas variantes distintas. La primera, basada en representaciones barimétricas, fue propuesta por **Tutte en 1963**

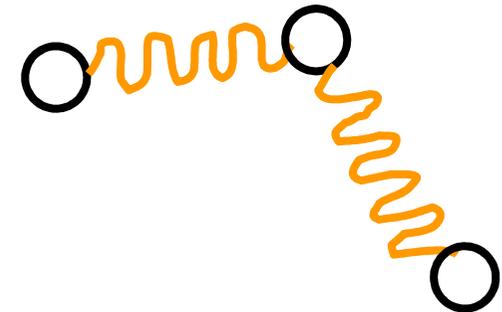
W.T. Tutte. 1963. How to draw a graph. Proc. London Math. Society 13(52): 743-768

LA METÁFORA DEL SISTEMA DE MUELLES (1)

P. Eades. 1984. A heuristic for graph drawing. *Congressus Numerantium*, 42: 149-160

Modelo *spring embedder*: Analogía física para dibujar grafos, interpreta los grafos como un sistema de objetos con fuerzas:

- Los nodos son **anillos de metal con cargas eléctricas**
 - Todos los nodos ejercen una **fuerza repulsiva** (f_r) entre sí
- Los enlaces son **muelles**
 - Ejercen una **fuerza atractiva** entre los nodos conectados (f_a)



Los nodos no conectados tienden a alejarse y los conectados a acercarse

Se persigue que la distancia Euclídea entre los nodos en la visualización equivalga a una longitud natural de enlace prefijada

Se asume que un sistema equilibrado, aquel en el que todos los nodos se encuentran a esa distancia teórica, implica una buena visualización

LA METÁFORA DEL SISTEMA DE MUELLES (2)

El sistema empieza con una distribución inicial aleatoria

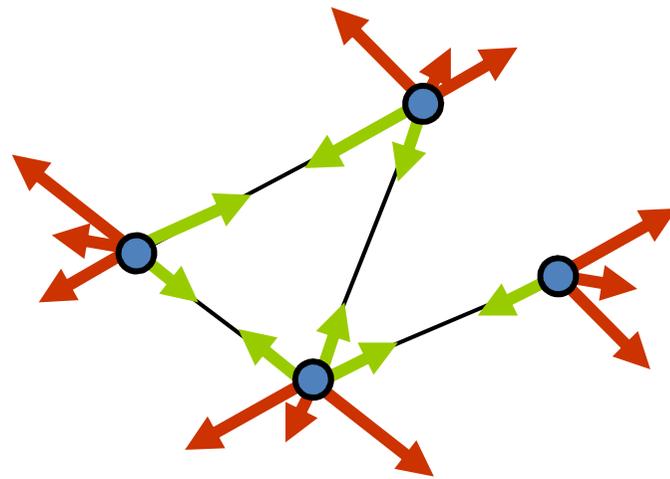
Los nodos se van moviendo de acuerdo a la acción de las fuerzas, cambiando la distribución

Cada distribución tiene una energía global

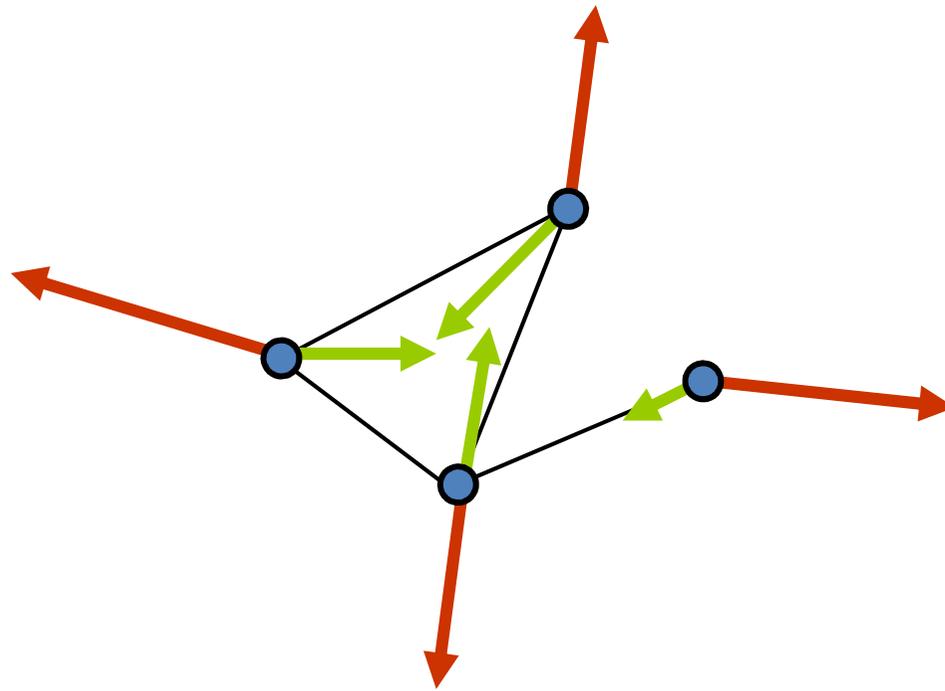
La distribución óptima corresponde a la distribución estable en la que **la suma de fuerzas de los nodos vale cero**

Existen otros algoritmos basados en la misma metáfora como el de **Fruchterman y Reingold de 1991**, que reduce la complejidad de las ecuaciones de fuerza del método de Eades

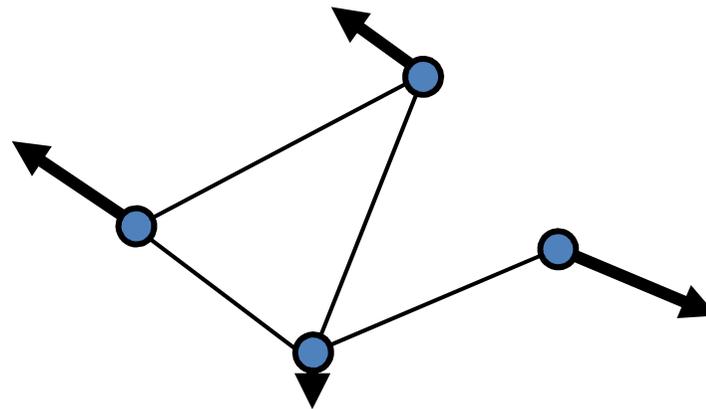
EJEMPLO ILUSTRATIVO (1)



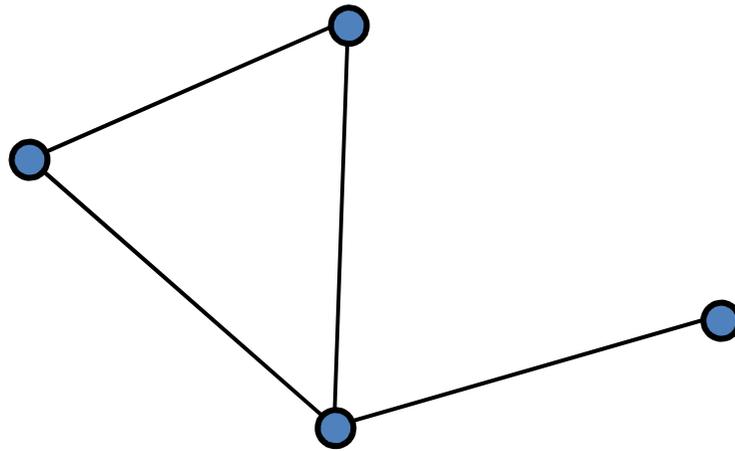
EJEMPLO ILUSTRATIVO (2)



EJEMPLO ILUSTRATIVO (3)



EJEMPLO ILUSTRATIVO (4)

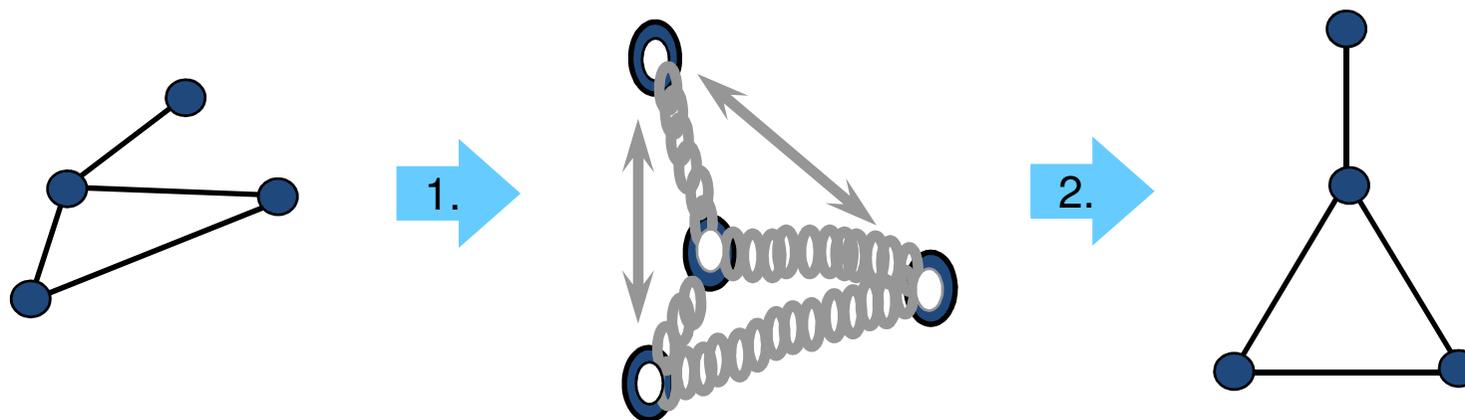


ALTERNATIVAS AL SISTEMA DE FUERZAS

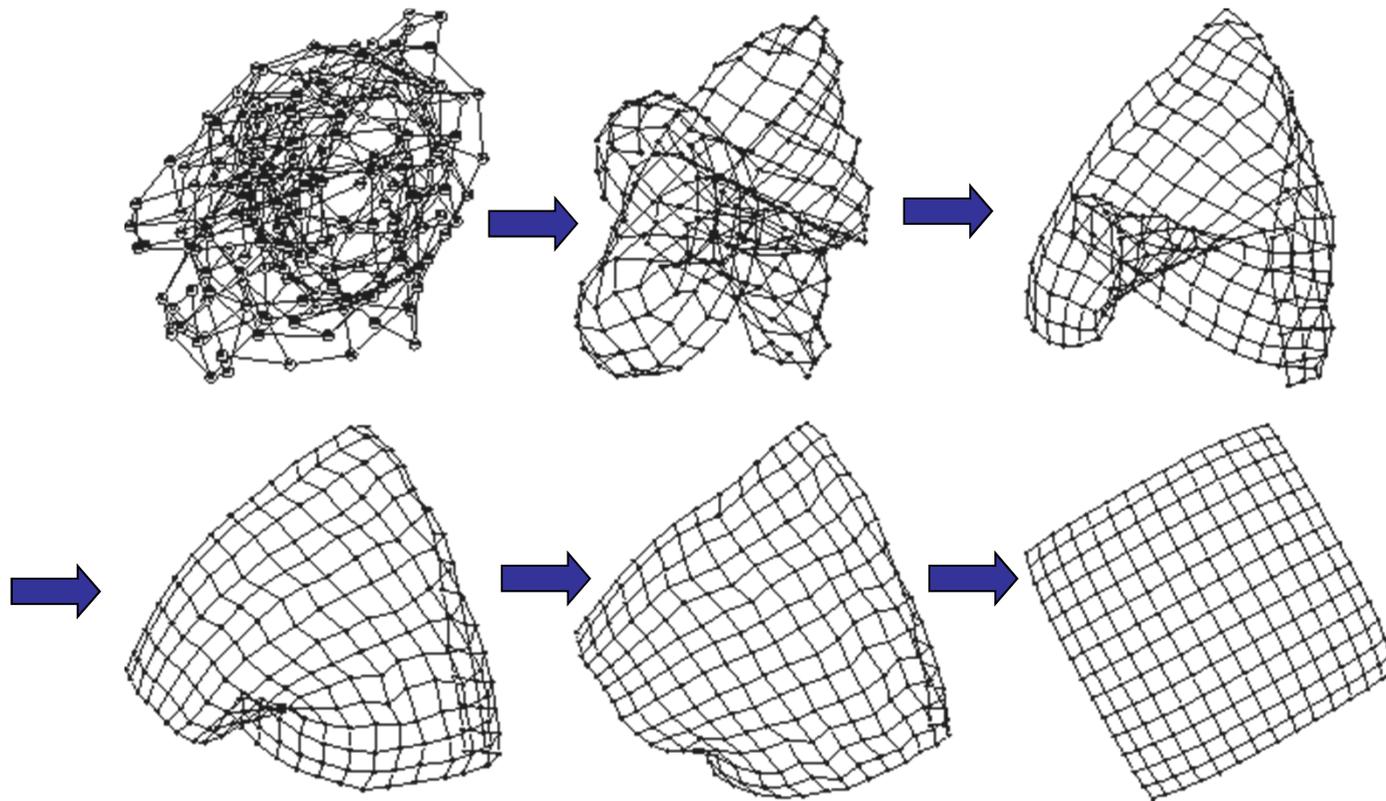
Alternativamente, se pueden considerar otros sistemas de fuerzas. P. ej., **las fuerzas entre los nodos u y v se calculan de acuerdo las distancias teóricas del grafo l_{uv}** , determinadas por los caminos geodésicos entre ellos (**Kamada y Kawai, 1989**)

Por tanto, **las fuerzas asociadas a los muelles son tanto atractivas como repulsivas** para conseguir que la localización de los nodos respete las distancias teóricas que existen en el grafo:

- Ejercen una **fuerza atractiva** si el muelle es más largo que su longitud natural en el grafo
- Ejercen una **fuerza repulsiva** si el muelle es más corto que su longitud natural en el grafo



OTRO EJEMPLO DE MAYOR TAMAÑO



© Sander

MÉTODOS DE DISTRIBUCIÓN GUIADOS POR FUERZAS

En general, todos los métodos pertenecientes a esta familia presentan dos componentes:

- 1. Modelo de fuerzas o de energía:** Sistema de fuerzas, definido por nodos y enlaces, que especifica la calidad de la distribución
- 2. Algoritmo de optimización:** Técnica para encontrar el estado de equilibrio, es decir, para calcular la distribución de nodos:
 - en la que la suma de fuerzas de los nodos es cero (equilibrio), o
 - que presenta la mínima energía local

MÉTODOS DE DISTRIBUCIÓN GUIADOS POR FUERZAS: Formulación inicial (1)

- Distancia ideal entre los nodos en la visualización (o longitud natural de enlace), k :

$$k = C \cdot \sqrt{\frac{\text{area_gráfico}}{n}} \quad ; \quad C=0.75 \quad (\mathbf{F\&R})$$

- Puede considerarse este valor o cualquier otro valor de referencia constante
- También se puede usar un valor de referencia distinto para la distancia ideal entre nodos (k_r) y la longitud natural del enlace (k_a), como hace el algoritmo de Eades

- Formulación cuadrática para las fuerzas repulsivas:

- La repulsión eléctrica producida por el nodo u sobre el nodo v sigue una ley cuadrática inversa:

$$f_r(u,v) = k / d^2$$

donde d es la distancia Euclídea existente entre los nodos u y v en el gráfico,

es decir, la magnitud del vector $\overrightarrow{(u,v)}$, $d = \|\overrightarrow{(u,v)}\|$

MÉTODOS DE DISTRIBUCIÓN GUIADOS POR FUERZAS: Formulación inicial (2)

- Formulación lineal para las fuerzas atractivas:

- Basada en la Ley de elasticidad de Hooke
- Proporcional a la diferencia de la distancia Euclidea y la distancia ideal k :

$$f_a(u,v) = d / k$$

- No se ejercen fuerzas cuando $d=k$ (el muelle tiene la longitud correcta)
- No da buenos resultados en la práctica: es demasiado fuerte cuando los nodos están muy lejanos

- **La distancia ideal se obtiene si $f_r(u,v) + f_a(u,v) = 0$** , es decir, cuando el nodo está en un estado de equilibrio en el que no actúan fuerzas sobre el

MÉTODOS DE DISTRIBUCIÓN GUIADOS POR FUERZAS: Algoritmos de optimización

Hay muchas técnicas para obtener la distribución óptima

Un algoritmo sencillo consiste en aplicar un proceso iterativo que:

- Inicializa los nodos en una posición aleatoria
- En cada paso:
 - Se calcula el vector de fuerza $\mathbf{F}(v)$ que actúa sobre cada nodo a partir de las fuerzas **atractivas** y **repulsivas**
 - Se **mueve** el nodo v **una cantidad pequeña** proporcional a la magnitud de $\mathbf{F}(v)$ en la dirección de $\mathbf{F}(v)$
- Se para cuando se alcanza el equilibrio o se cumple la condición de parada

No es muy rápido pero provoca una transición suave. La distribución final depende de las posiciones iniciales de los nodos

El coste de calcular las fuerzas atractivas es $O(|E|)$ (sólo los vecinos del nodo) pero el de las repulsivas es $O(n^2)$. **Es el cuello de botella del algoritmo**

MÉTODOS DE DISTRIBUCIÓN GUIADOS POR FUERZAS: Algoritmos de optimización (2)

Inicializar la distribución

Seleccionar la longitud natural de enlace (distancia entre nodos) k

Repetir

Para cada nodo v **hacer**

Para cada par de nodos u, v

calcular el vector de la **fuerza repulsiva**: $f_r(u, v) = \frac{k}{\|\overrightarrow{(u, v)}\|^2} \cdot \overrightarrow{(u, v)}$

Para cada enlace $e = (u, v)$

calcular el vector de la **fuerza atractiva**: $f_a(u, v) = \frac{\|\overrightarrow{(u, v)}\|}{k} \cdot \overrightarrow{(u, v)}$

Sumar todos los vectores de fuerzas: $F(v) = \sum_{(u, v) \in V \times V} f_r(u, v) + \sum_{(u, v) \in E} f_a(u, v)$

Mover la posición del nodo v una cantidad $\gamma \in [0, 1]$ de $\|F(v)\|$ en la dirección de $F(v)$

hasta FIN

MÉTODOS DE DISTRIBUCIÓN GUIADOS POR FUERZAS: Variantes (1)

- Formulaciones alternativas para las fuerzas:
- **(Eades, 1984):**
 - fuerzas repulsivas: $f_r(u, v) = \frac{k}{\sqrt{d}}$
 - fuerzas atractivas: Formulación logarítmica: $f_a(u, v) = \log\left(\frac{d}{k}\right)$
- Originalmente, la formulación propuesta por Eades incorporaba una serie de constantes y se aplicaba sobre grafos de tamaño pequeño (30)
- A pesar de que resuelve los problemas de la formulación inicial, sigue siendo muy costosa y demasiado estricta. No funciona bien en redes más grandes

MÉTODOS DE DISTRIBUCIÓN GUIADOS POR FUERZAS: Variantes (2)

- Formulaciones alternativas para las fuerzas:
 - Cuadrática, $p=2$ (**F&R, 1991**) estándar
 - Cúbica, $p=3$ (**Forster, 1999**) más rápida de calcular, no usa $\sqrt{\quad}$
 - Fórmulas ($p = 2,3$): $f_r(u, v) = -\frac{k^p}{d}$ $f_a(u, v) = \frac{d^p}{k}$
- Por otro lado, tenemos las formulaciones basadas en distancias teóricas del grafo, como (**K&K, 1989**)

MÉTODO DE FRUTCHERMAN Y REINGOLD (1)

T. Fruchterman, E. Reingold. 1991. Graph drawing by force-directed placement.
Software – Practice and Experience 21(11): 1129-1164

Acelera el cálculo de fuerzas del algoritmo de Eades

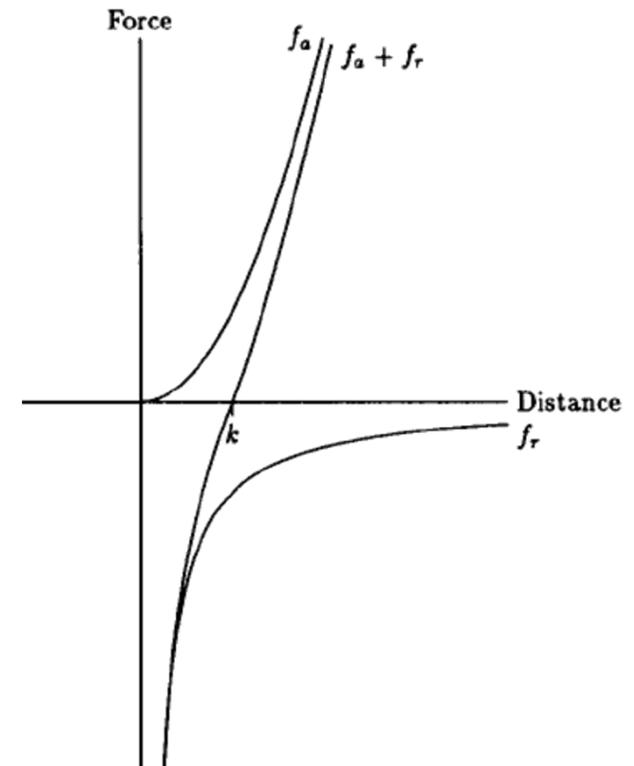
La fuerza de atracción es cuadrática y la de repulsión hiperbólica

Están referenciadas a la distancia ideal entre nodos deseada, k :

$$f_r(u, v) = -\frac{k^2}{d} \quad f_a(u, v) = \frac{d^2}{k}$$

k es el radio de la esfera que delimita la zona vacía en torno a un nodo

Al considerar una única referencia k , el algoritmo emplea directamente la matriz de adyacencia

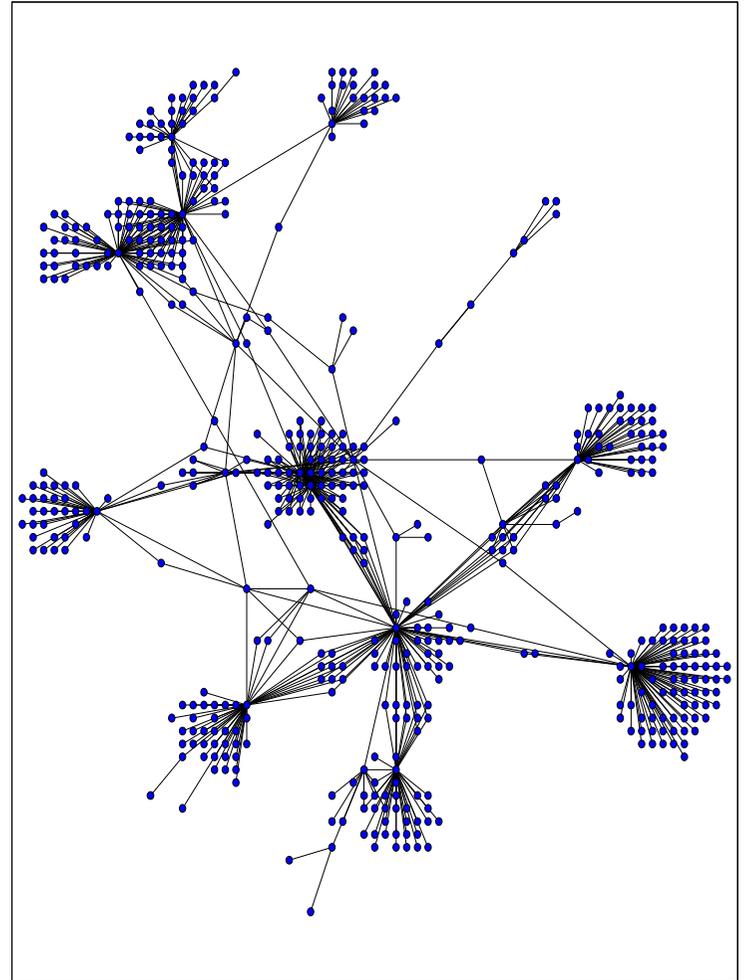


MÉTODO DE FRUTCHERMAN Y REINGOLD (2)

A grandes rasgos, el modelo corresponde a la atracción electrostática entre nodos conectados

Incorpora un parámetro de **temperatura t** que reduce el máximo desplazamiento posible para los nodos según avanza la ejecución (tipo *simulated annealing*)

t puede valer 1/10 de la anchura del gráfico y reducir linealmente su valor hasta llegar a 0



MÉTODO DE FRUTCHERMAN Y REINGOLD (3): Algoritmo

```
area:=  $W * L$ ; { $W$  and  $L$  are the width and length of the frame}
 $G := (V, E)$ ; {the vertices are assigned random initial positions}
 $k := \sqrt{\text{area}/|V|}$ ;
function  $f_a(x) := \text{begin return } x^2/k \text{ end}$ ;
function  $f_r(x) := \text{begin return } k^2/x \text{ end}$ ;
for  $i := 1$  to  $\text{iterations}$  do begin
  {calculate repulsive forces}
  for  $v$  in  $V$  do begin
    {each vertex has two vectors:  $.pos$  and  $.disp$ }
     $v.disp := 0$ ;
    for  $u$  in  $V$  do
      if ( $u \neq v$ ) then begin
        { $\delta$  is the difference vector between the positions of the two vertices }
         $\delta := v.pos - u.pos$ ;
         $v.disp := v.disp + (\delta/|\delta|) * f_r(|\delta|)$ 
      end
    end
  end
  {calculate attractive forces}
  for  $e$  in  $E$  do begin
    {each edges is an ordered pair of vertices  $.v$  and  $.u$ }
     $\delta := e.v.pos - e.u.pos$ ;
     $e.v.disp := e.v.disp - (\delta/|\delta|) * f_a(|\delta|)$ ;
     $e.u.disp := e.u.disp + (\delta/|\delta|) * f_a(|\delta|)$ 
  end
  {limit max displacement to temperature  $t$  and prevent from displacement outside frame}
  for  $v$  in  $V$  do begin
     $v.pos := v.pos + (v.disp/|v.disp|) * \min(v.disp, t)$ ;
     $v.pos.x := \min(W/2, \max(-W/2, v.pos.x))$ ;
     $v.pos.y := \min(L/2, \max(-L/2, v.pos.y))$ 
  end
  {reduce the temperature as the layout approaches a better configuration}
   $t := \text{cool}(t)$ 
end
```

MÉTODO DE FRUTCHERMAN Y REINGOLD (4)

A pesar de la optimización en el cálculo de las fuerzas, **el algoritmo puede ser lento si el tamaño de la red es grande**

Cada iteración del algoritmo básico calcula $O(|E|)$ fuerzas atractivas y $O(n^2)$ fuerzas repulsivas

Para reducir el tiempo de cálculo, los autores proponen una variante donde se ignoran las fuerzas repulsivas entre nodos lejanos

En grafos dispersos y con una distribución uniforme de nodos, esta **variante basada en grid** tiene $O(n)$ en el cálculo de las fuerzas repulsivas

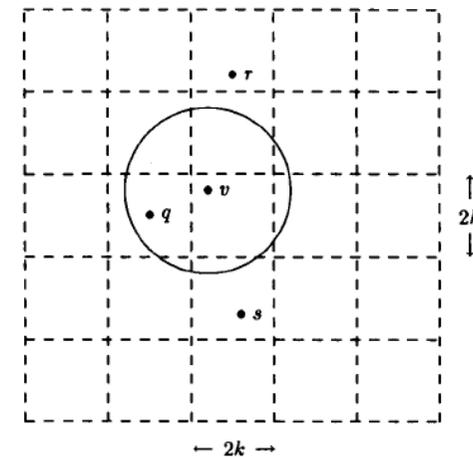
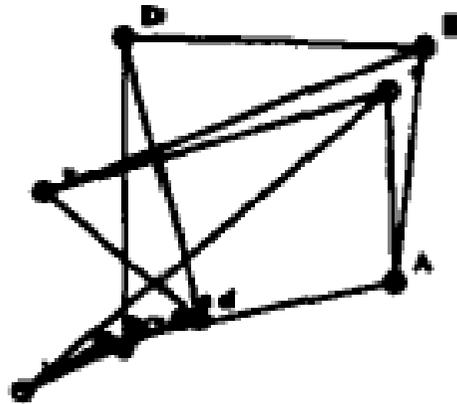
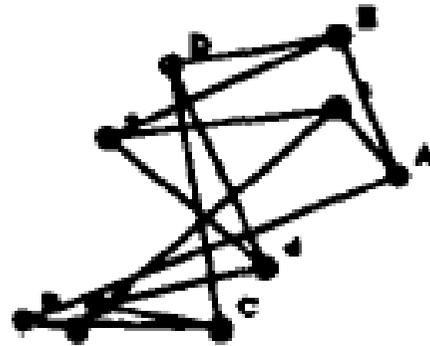


Figure 7. Calculating repulsive forces using the grid-square algorithm

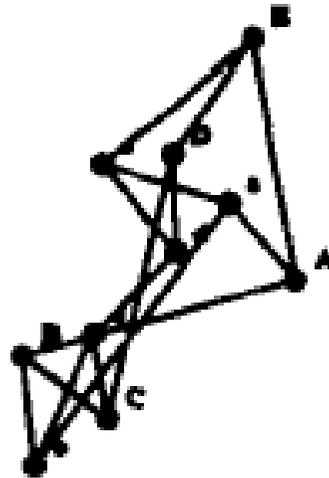
EJEMPLO ILUSTRATIVO: F&R (1)



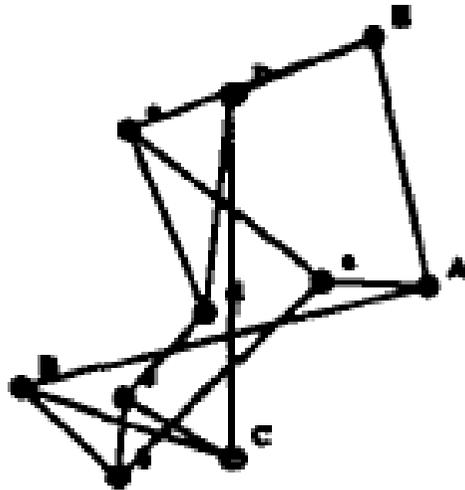
EJEMPLO ILUSTRATIVO: F&R (2)



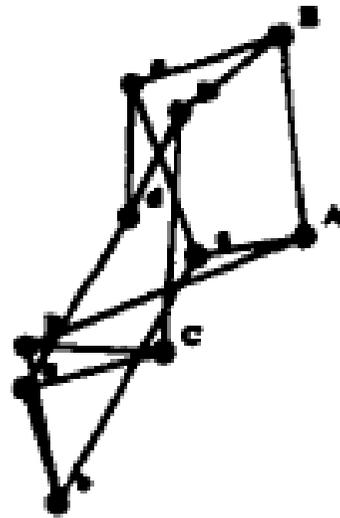
EJEMPLO ILUSTRATIVO: F&R (3)



EJEMPLO ILUSTRATIVO: F&R (4)



EJEMPLO ILUSTRATIVO: F&R (5)



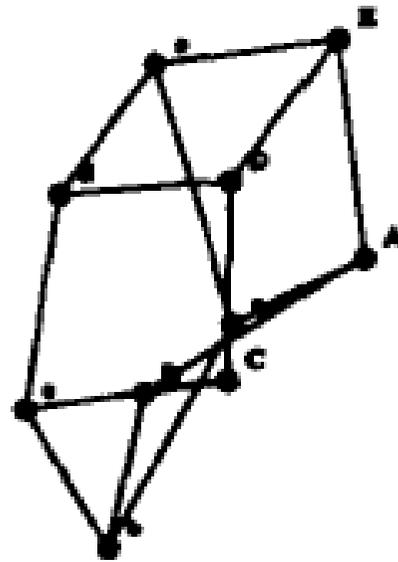
EJEMPLO ILUSTRATIVO: F&R (6)



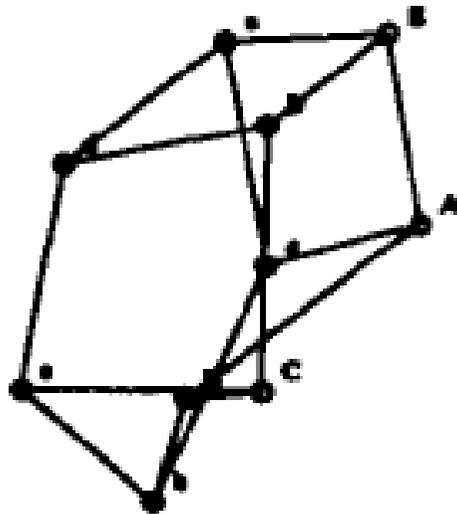
EJEMPLO ILUSTRATIVO: F&R (7)



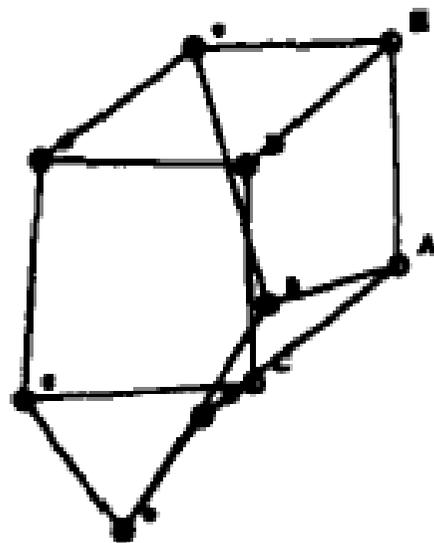
EJEMPLO ILUSTRATIVO: F&R (8)



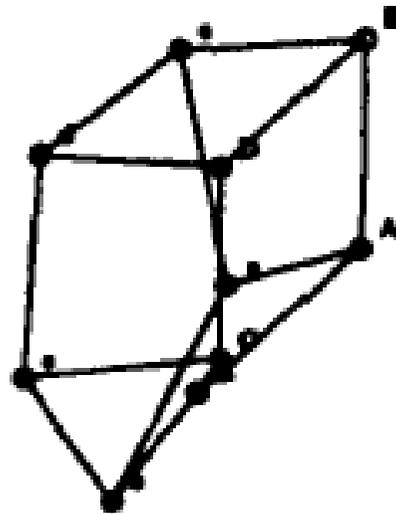
EJEMPLO ILUSTRATIVO: F&R (9)



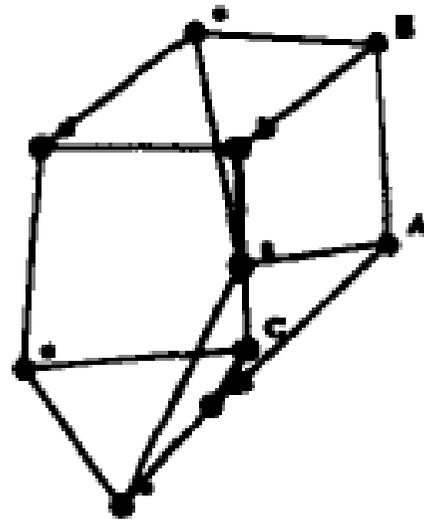
EJEMPLO ILUSTRATIVO: F&R (10)



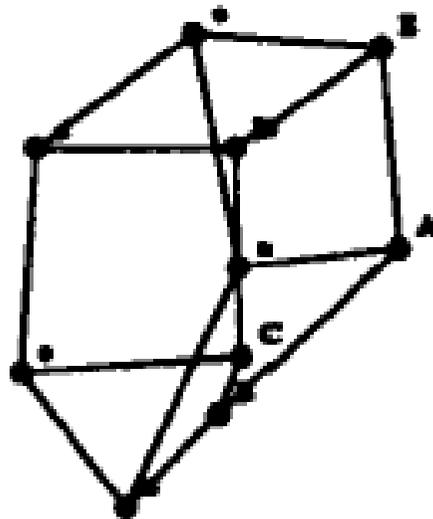
EJEMPLO ILUSTRATIVO: F&R (11)



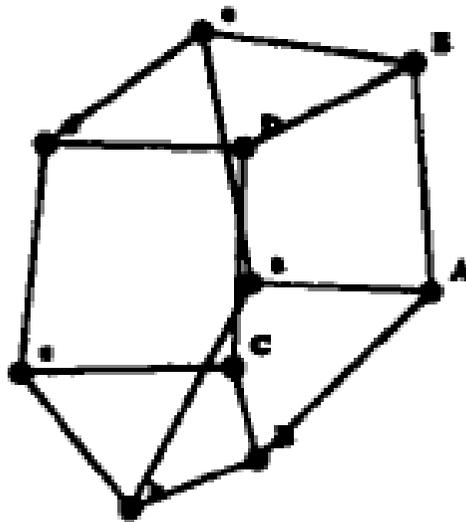
EJEMPLO ILUSTRATIVO: F&R (12)



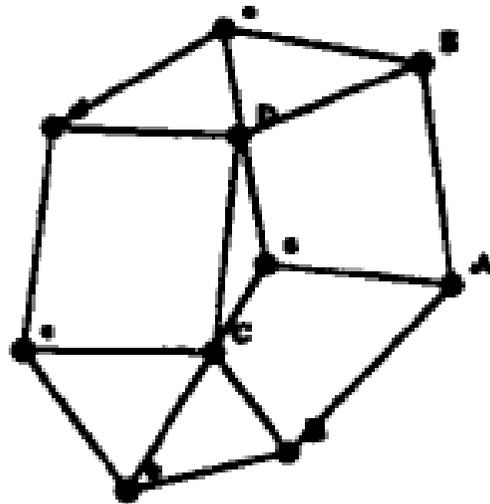
EJEMPLO ILUSTRATIVO: F&R (13)



EJEMPLO ILUSTRATIVO: F&R (14)

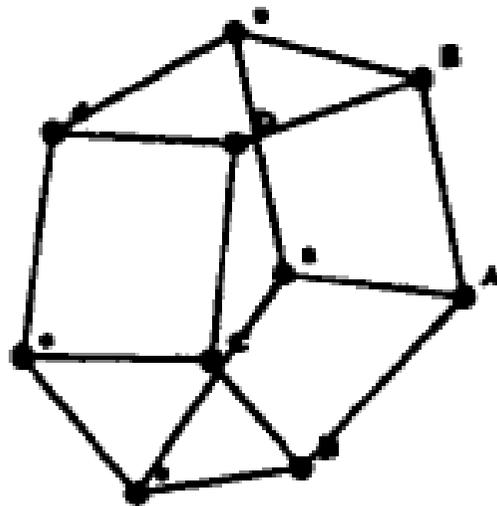


EJEMPLO ILUSTRATIVO: F&R (15)



EJEMPLO ILUSTRATIVO: F&R (16)

“Final”



MÉTODO DE KAMADA Y KAWAI (1)

T. Kamada, S. Kawai. 1989. An algorithm for drawing general undirected graphs.
Information Processing Letters 31: 7-15

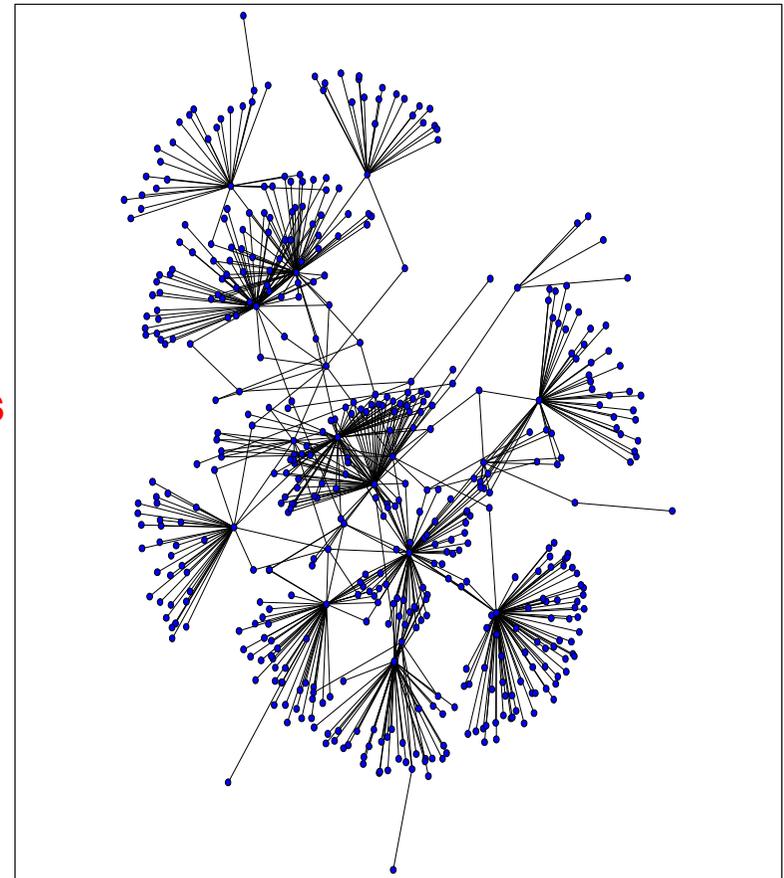
Todos los nodos i, j están conectados por muelles con una longitud (**distancia Euclidea en el gráfico**) ideal $l_{i,j}$ equivalente a la distancia del camino geodésico que los conecta en el grafo

Hay fuerzas de atracción y de repulsión entre cada par de nodos **determinadas por los muelles**

Como paso previo, es necesario calcular todos los caminos mínimos de la red

Ventaja: puede usarse en grafos ponderados

Para mejorar la eficiencia, en cada iteración el algoritmo **mueve solo el nodo peor situado**



MÉTODO DE KAMADA Y KAWAI (2)

Modela la distancia teórica del grafo con la distancia Euclídea en el plano:

- Las fuerzas tratan de localizar los nodos de forma que su distancia geométrica en el gráfico sea proporcional a su distancia teórica en el grafo

Para cada par de nodos (i,j) , $d_{i,j}$ es la distancia teórica en el grafo entre ellos:

- Número de enlaces o suma de los pesos en el camino geodésico entre i y j

La longitud ideal del muelle entre i y j se define como $l_{i,j} = L \cdot d_{i,j}$, donde L es la longitud deseable para una distancia de 1 en el gráfico

L se obtiene a partir del diámetro de la red y la longitud en pixels del lado más grande del área de visualización L_0 :

$$L = \frac{L_0}{\max_{i < j} (d_{ij})}$$

MÉTODO DE KAMADA Y KAWAI (3)

- $p_i=(x_i, y_i)$ y $p_j=(x_j, y_j)$ son las posiciones de los nodos i y j en el gráfico
- $d(p_i, p_j)=|p_i - p_j|$ es la distancia Euclidea entre dichas posiciones
- $l_{i,j}$ es la distancia teórica en el grafo entre dichos nodos

Objetivo: Encontrar una distribución tal que para cada par de nodos i y j , $d(p_i, p_j)$ sea aproximadamente igual a $l_{i,j}$

- Es decir, el sistema tiene una fuerza proporcional a $d(p_i, p_j) - l_{i,j}$

con la siguiente función de energía global, que se debe minimizar:

$$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{1}{2} k_{ij} \left(|p_i - p_j| - l_{ij} \right)^2$$

MÉTODO DE KAMADA Y KAWAI (4)

- La energía potencial en el muelle que conecta i y j es:

$$\frac{1}{2} \cdot k_{ij} \cdot (d(p_i, p_j) - l_{i,j})^2$$

- k_{ij} es un **parámetro de rigidez**, que indica la fuerza del muelle en cuestión. Los muelles que unen nodos con una distancia teórica pequeña son más fuertes:

$$k_{ij} = K / l_{i,j}^2 \quad (K \text{ es una constante})$$

- Considerando las coordenadas 2D de los nodos en el gráfico, la función de energía E se reescribe como:

$$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{1}{2} k_{i,j} \left((x_i - x_j)^2 + (y_i - y_j)^2 + l_{i,j}^2 - 2l_{i,j} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \right)$$

- El mínimo de esta ecuación se alcanza cuando las derivadas parciales son cero y es muy difícil de obtener. Hay que resolver $2 \cdot n$ ecuaciones no lineales**

MÉTODO DE KAMADA Y KAWAI (5): Algoritmo

- Alternativamente, el método usa un **enfoque iterativo** que obtiene un mínimo local (y no el mínimo global) de la ecuación anterior
- En cada paso, **sólo se mueve un nodo** a una posición estable que minimice la energía global. El resto de nodos se quedan fijos
- **El nodo escogido es aquel sobre el que tiene que actuar la mayor fuerza (el peor posicionado)**
- Se obtiene como el nodo m que maximiza la siguiente fórmula basada en las derivadas parciales:
$$\Delta_m = \sqrt{\left(\frac{\partial E}{\partial x_m}\right)^2 + \left(\frac{\partial E}{\partial y_m}\right)^2}$$
- El algoritmo finaliza cuando la mejora es menor que un umbral

MÉTODO DE KAMADA Y KAWAI (6): Algoritmo

compute $d_{i,j}$ for $1 \leq i \neq j \leq n$;

compute $l_{i,j}$ for $1 \leq i \neq j \leq n$;

compute $k_{i,j}$ for $1 \leq i \neq j \leq n$;

initialize p_1, p_2, \dots, p_n ;

while ($\max_i \Delta_i > \epsilon$) {

 let p_m be the particle satisfying $\Delta_m = \max_i \Delta_i$;

while ($\Delta_m > \epsilon$) {

 compute δx and δy by solving the following system of equations:

$$\frac{\partial^2 E}{\partial x_m^2}(x_m^{(t)}, y_m^{(t)})\delta x + \frac{\partial^2 E}{\partial x_m \partial y_m}(x_m^{(t)}, y_m^{(t)})\delta y = -\frac{\partial E}{\partial x_m}(x_m^{(t)}, y_m^{(t)});$$

$$\frac{\partial^2 E}{\partial y_m \partial x_m}(x_m^{(t)}, y_m^{(t)})\delta x + \frac{\partial^2 E}{\partial y_m^2}(x_m^{(t)}, y_m^{(t)})\delta y = -\frac{\partial E}{\partial y_m}(x_m^{(t)}, y_m^{(t)})$$

$$x_m := x_m + \delta x;$$

$$y_m := y_m + \delta y;$$

 }

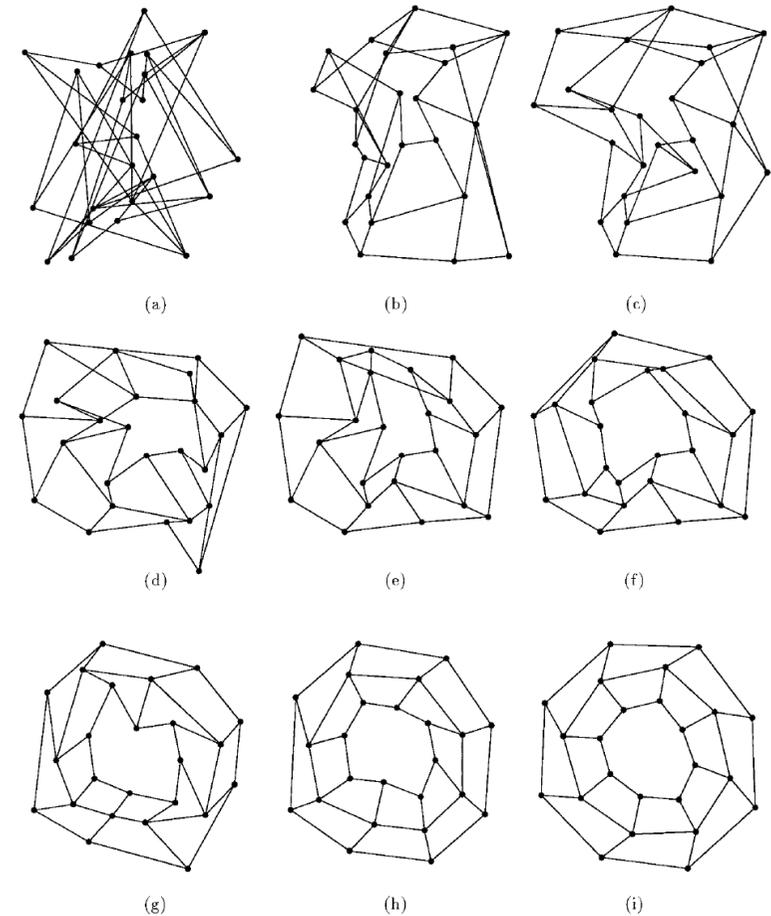
}

MÉTODO DE DAVIDSON Y HAREL

R. Davidson, D. Harel. Drawing graphs nicely using simulated annealing. 1996.
ACM Transactions on Graphics, 15(4):301-331

Este algoritmo incluye **restricciones adicionales al enfoque tradicional** basado en fuerzas con el objetivo explícito de minimizar el número de cruces entre enlaces y prevenir que los nodos se acerquen mucho a los enlaces no adyacentes

El algoritmo está basado en la metaheurística de **enfriamiento simulado** (simulated annealing) que permite movimientos grandes y de empeoramiento de la energía al principio de la ejecución



CALIDAD DE LAS VISUALIZACIONES EN MÉTODOS BASADOS EN FUERZAS (1)

- Cada método considera unos criterios estéticos distintos:

Table 3.1 Criteria for graph drawing algorithms

Criteria	Di Battista et al. (1994)	Eades (1984)	Kamada and Kawai (1989)	Fruchterman and Reingold (1991)	Davidson and Harel (1996)	NicheWorks (1997)
Symmetric	✓	✓	✓			
Evenly distributed nodes	✓		✓	✓	✓	Clustered
Uniform edge lengths	✓	✓	✓	✓	✓	Weights
Minimized edge crossings	✓		✓	✓	✓	

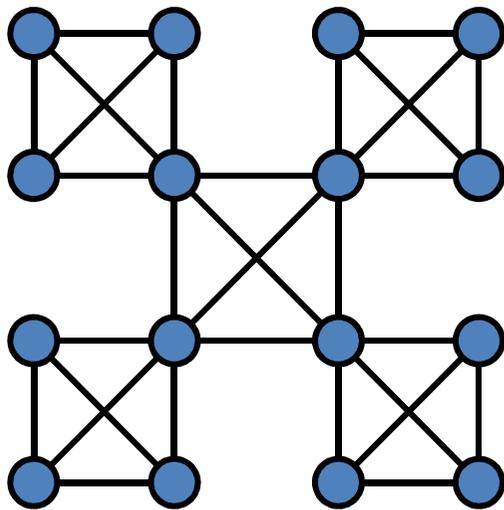
- (E) Eades ('84). *A heuristic for graph drawing*
- (F&R) Fruchterman & Reingold ('91). *Graph Drawing by Force-directed Placement*
- (K&K) Kamada & Kawai ('89). *An algorithm for drawing general undirected graphs*
- (D&H) Davidson & Harel ('96). *Drawing Graphs Nicely Using Simulated Annealing*

CALIDAD DE LAS VISUALIZACIONES EN MÉTODOS BASADOS EN FUERZAS (2)

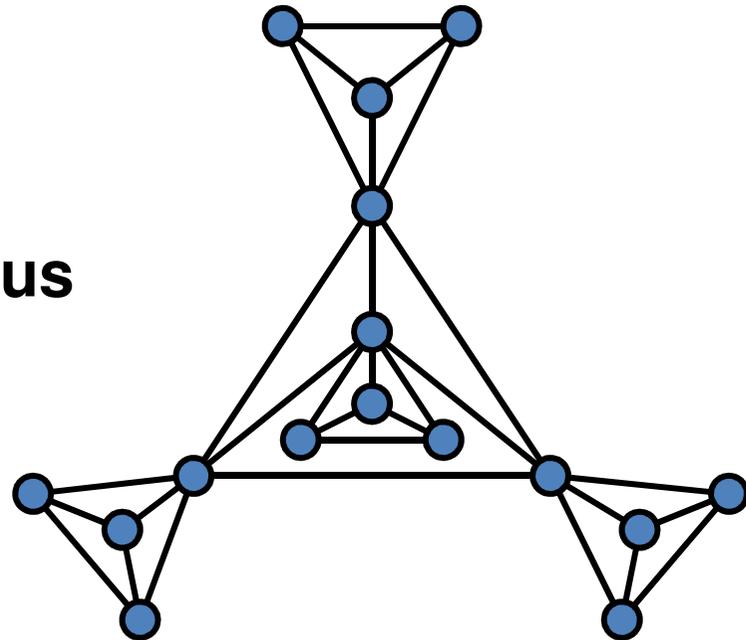
- Cada método considera unos criterios estéticos distintos:
 - **F&R, K&K, D&H:**
 - Distribuir los nodos y los enlaces de forma uniforme
 - Considerar la simetría siempre que sea posible
 - **F&R, D&H:**
 - Ajustarse a la forma de la ventana de visualización
 - Reducir el número de cruces entre enlaces
 - **F&R:**
 - Considerar una longitud de enlace uniforme
 - **D&H:**
 - Evitar que los nodos se acerquen demasiado a los enlaces
 - Permite que se ponderen los distintos criterios, por lo que se puede personalizar

CALIDAD DE LAS VISUALIZACIONES EN MÉTODOS BASADOS EN FUERZAS (4)

¿Qué se considera una buena visualización?



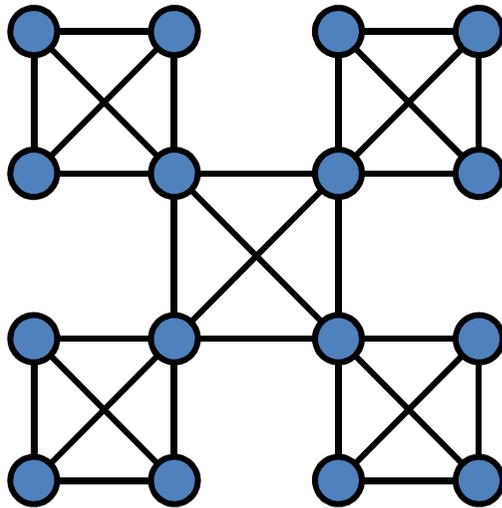
versus



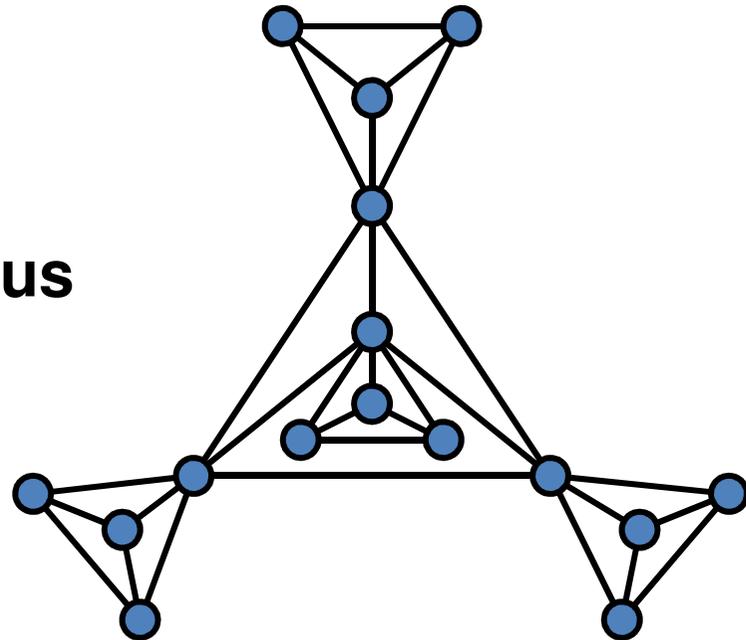
CALIDAD DE LAS VISUALIZACIONES EN MÉTODOS BASADOS EN FUERZAS (5)

¿Qué se considera una buena visualización?

- **K&K**: La minimización de los cruces entre enlaces no es un buen criterio por si solo
 - El “equilibrio” es más importante



versus



CALIDAD DE LAS VISUALIZACIONES EN MÉTODOS BASADOS EN FUERZAS (6)

Distribución inicial recomendada:

- **F&R**, **D&H**: aleatoria
- **K&K**: circular
- De todos modos, ambas funcionan “correctamente” en cualquier caso
- También se pueden especificar las posiciones iniciales de los nodos

Longitud de los enlaces:

- **F&R**: tratar de mantenerla uniforme
- **K&K**: los pares de los nodos deben tener una distancia en el plano (en la visualización) igual a su distancia teórica en el grafo
- **D&H**: tan corta como sea posible

CALIDAD DE LAS VISUALIZACIONES EN MÉTODOS BASADOS EN FUERZAS (7)

Movimiento de los nodos:

- Lo normal es mover un nodo en cada iteración (**K&K**, **D&H**)
- Es más eficiente pero puede caer en mínimos locales
- El mover todos los nodos a la vez puede proporcionar mejores resultados (**F&R**)
- El comportamiento más adecuado es restringir la amplitud del movimiento que se hace en cada paso
- Es mejor hacer movimientos grandes al principio de la ejecución y pequeños al final de la misma

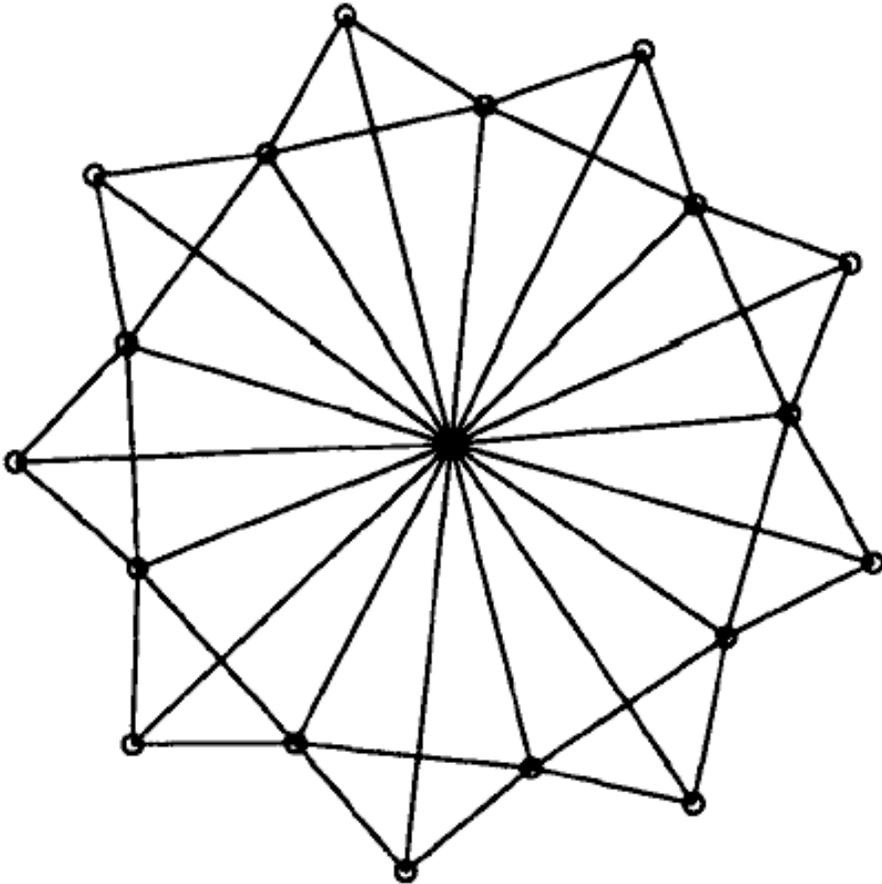
CALIDAD DE LAS VISUALIZACIONES EN MÉTODOS BASADOS EN FUERZAS (8)

¿Qué hacer con los cruces entre enlaces?

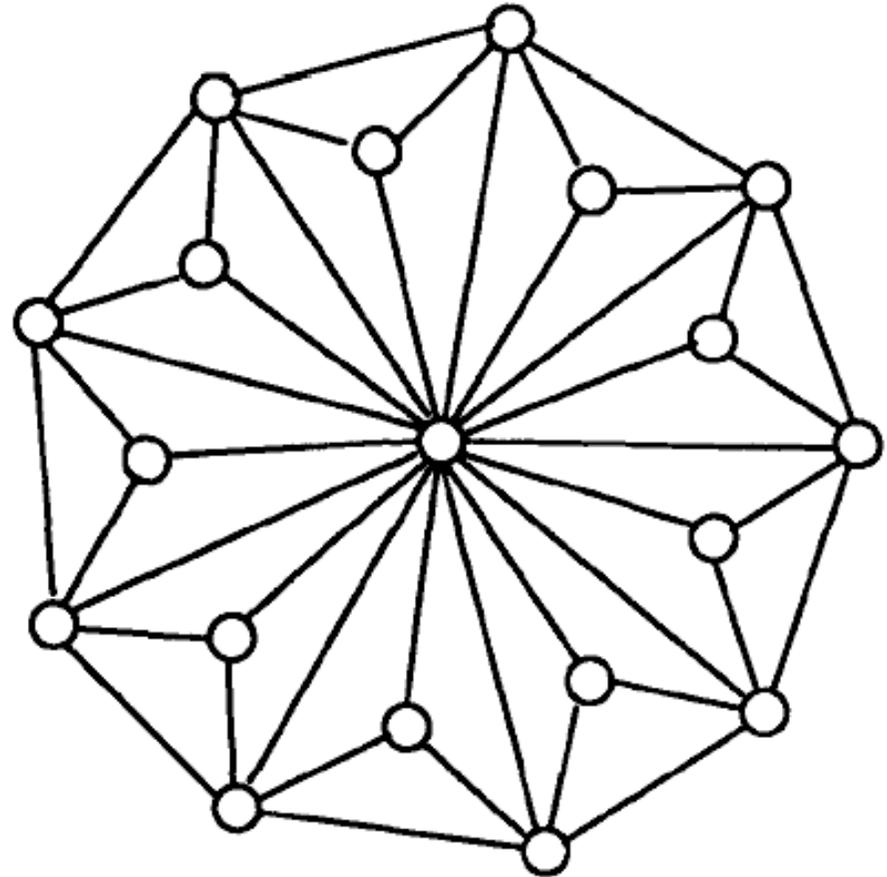
- **K&K**: Los cruces entre enlaces no son malos en absoluto
- **D&H**: Trata de reducir el tamaño de los enlaces, lo que tiende a evitar los cruces entre enlaces
- Existen algoritmos para dibujar grafos planos pero no siempre dan representaciones bonitas
- Es mejor tener unos pocos cruces en un buen gráfico que no tener cruces en una representación poco estética

CALIDAD DE LAS VISUALIZACIONES EN MÉTODOS BASADOS EN FUERZAS (9)

- **F&R:**



- **D&H:**



RESUMEN MÉTODOS DE DISTRIBUCIÓN GUIADOS POR FUERZAS (1)

Models	Attraction force	Repulsive force	Energy function
Eades (1984) Spring-embedder	$f_a(d) = k_a \log(d)$	$f_r(d) = k_r/d^2$	
Kamada and Kawai (1989) Local minimum			$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{1}{2} k_{ij} (p_i - p_j - l_{ij})^2$
Fruchterman and Reingold (1991) Spring-embedder	$f_a(d) = d^2/k$	$f_r(d) = -k^2/d$	
Davidson and Harel (1996) Simulated annealing			uneven-node-distribution-penalty + edge-crossing-penalty + long-edge- length-penalty

RESUMEN MÉTODOS DE DISTRIBUCIÓN GUIADOS POR FUERZAS (2)

Ventajas:

- Concepto intuitivo y relativamente fácil de implementar
- Buenos resultados en muchos grafos, buena representación de las simetrías
- Flexible: fácilmente adaptable, extensible y combinable:
 - más modelos de fuerzas:
 - Centro de gravedad (atractiva)
 - Campos magnéticos subyacentes (concéntrico, radial, horizontal)
 - fuerzas angulares (entre enlaces adyacentes a nodos)
 - desde los límites (repulsiva, rebote)
 - esquemas de enfriamiento para la finalización (**F&R, 1991**)
 - restricción de las oscilaciones y rotaciones (**Frick, Ludwig y Mehlau, 1995**)
 - adaptaciones de *simulated annealing* (**D&H, 1996**), algoritmos genéticos, búsqueda tabú, ...
- Posibilidad de gestionar restricciones del dominio (minimizar cruces de enlaces, alejar nodos de otros nodos no adyacentes, etc.) (**D&H, 1996**) y de incluir mejoras heurísticas
- Extensible a 3D (**Bruß y Frick, 1996**)

RESUMEN MÉTODOS DE DISTRIBUCIÓN GUIADOS POR FUERZAS (3)

Inconvenientes:

- Costosos en tiempo de ejecución
 - Hay que calcular la acción de las fuerzas repulsivas para cada par de nodos ($O(n^2)$) y cada enlace
 - Se puede optimizar considerando sólo repulsión entre nodos cercanos (*grid*) (F&R, 1991)
- ¿Criterio de parada?
 - ¿número máximo de iteraciones?
 - ¿umbral de energía?
 - ¿se ha alcanzado un mínimo local?
 - ¿especificado por el usuario?
- ¿Qué fuerzas usar?
- Necesidad de ajustar muchos parámetros
- Falta de resultados teóricos sobre la calidad de las visualizaciones obtenidas

MÉTODOS BASADOS EN FUERZAS AVANZADOS (1)

R. Hadany, D. Harel. A multi-scale algorithm for drawing graphs nicely. 2001.
Discrete Applied Mathematics, 113(1):3-21

Primer algoritmo (**técnica multi-escala**) capaz de producir buenas representaciones con redes de más de 1000 nodos:

A natural strategy for drawing a graph nicely is to first consider an abstraction, disregarding some of the graph's fine details. This abstraction is then drawn, yielding a "rough" layout in which only the general structure is revealed. Then the details are added and the layout is corrected. To employ such a strategy is it crucial that the abstraction retain essential features of the graph. Thus, one has to define the notion of coarse-scale representations of a graph, in which the combinatorial structure is significantly simplified but features important for visualization are well preserved. The drawing process will then "travel" between these representations, and introduce multi-scale corrections. Assuming we have already defined the multiple levels of coarsening, the general structure of our strategy is as follows:

MÉTODOS BASADOS EN FUERZAS AVANZADOS (2)

1. *Perform fine-scale relocations of vertices that yield a locally organized configuration.*
2. *Perform coarse-scale relocations (through local relocations in the coarse representations), correcting global disorders not found in stage 1.*
3. *Perform fine-scale relocations that correct local disorders introduced by stage 2.*

La variante de (**Harel y Koren, 2002**) considera una relocalización de escala amplia basada en el método de los k -centros (se escoge un subconjunto V' de k nodos que minimice las distancias de todos los nodos a V')

En las relocalizaciones de escala fina se usa K&K. El algoritmo es capaz de manejar redes con 15000 nodos en un tiempo razonable

D. Harel, Y. Koren. 2002. A fast multi-scale method for drawing large graphs. Journal of Graph Algorithms and Applications, 6(3):179-202

MÉTODOS BASADOS EN FUERZAS AVANZADOS (3)

```
Layout( $G(V, E)$ )
% Goal: Find  $L$ , a nice layout of  $G$ 
% Constants:
% Rad[= 7] – determines radius of local neighborhoods
% Iterations[= 4] – determines number of iterations in local beautification
% Ratio[= 3] – ratio between number of vertices in two consecutive levels
% MinSize[= 10] – size of the coarsest graph
  Compute the all-pairs shortest path length:  $d_{VV}$ 
  Set up a random layout  $L$ 
   $k \leftarrow MinSize$ 
  while  $k \leq |V|$  do
     $centers \leftarrow \mathbf{K-Centers}(G(V, E), k)$ 
     $radius = \max_{v \in centers} \min_{u \in centers} \{d_{vu}\} * Rad$ 
    LocalLayout( $d_{centers \times centers}, L(centers), radius, Iterations$ )
    for every  $v \in V$  do
       $L(v) \in L(center(v)) + rand$ 
     $k \leftarrow kRatio$ 
  return  $L$ 

K-Centers( $G(V, E), k$ )
% Goal: Find a set  $S \subseteq V$  of size  $k$ , such that  $\max_{v \in V} \min_{s \in S} \{d_{sv}\}$  is minimized.
 $S \leftarrow \{v\}$  for some arbitrary  $v \in V$ 
for  $i = 2$  to  $k$  do
  1. Find the vertex  $u$  farthest away from  $S$ 
    (i.e., such that  $\min_{s \in S} \{d_{us}\} \geq \min_{s \in S} \{d_{ws}\}, \forall w \in V$ )
  2.  $S \leftarrow S \cup \{u\}$ 
return  $S$ 

LocalLayout( $d_{V \times V}, L, k, Iterations$ )
% Goal: Find a locally nice layout  $L$  by beautifying  $k$ -neighborhoods
%  $d_{V \times V}$  : all-pairs shortest path length
%  $L$ : initialized layout
%  $k$ : radius of neighborhoods
for  $i = 1$  to  $Iterations * |V|$  do
  1. Choose the vertex  $v$  with the maximal  $\Delta_v^k$ 
  2. Compute  $\delta_v^k$  as in Kamada-Kawai
  3.  $L(v) \leftarrow L(v) + (\delta_v^k(x), \delta_v^k(y))$ 
end
```

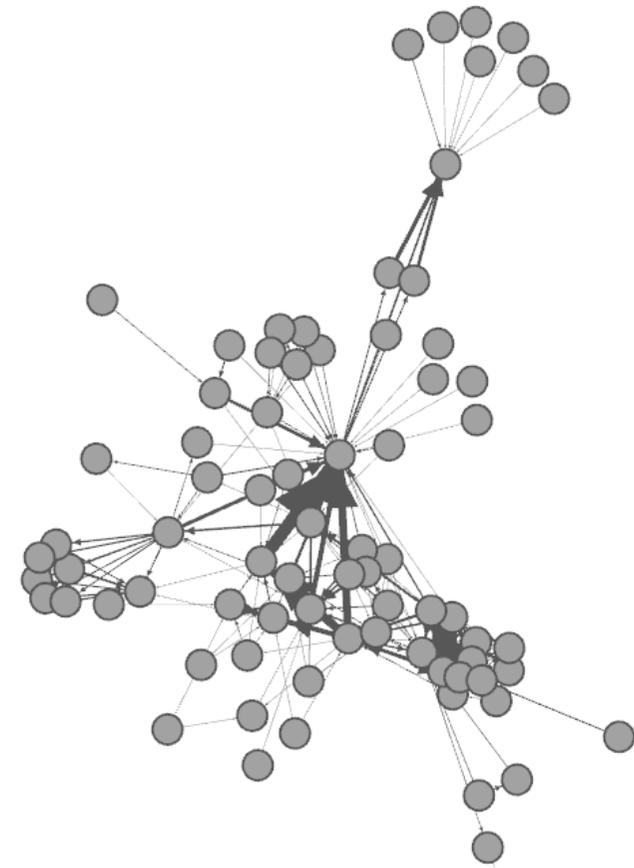
MÉTODOS BASADOS EN FUERZAS AVANZADOS (4)

Y. F. Hu. 2005. Efficient and high quality force-directed graph drawing.
The Mathematica Journal, 10:37-71

Combina un algoritmo basado en fuerzas con una **técnica multi-escala** para reducir la complejidad

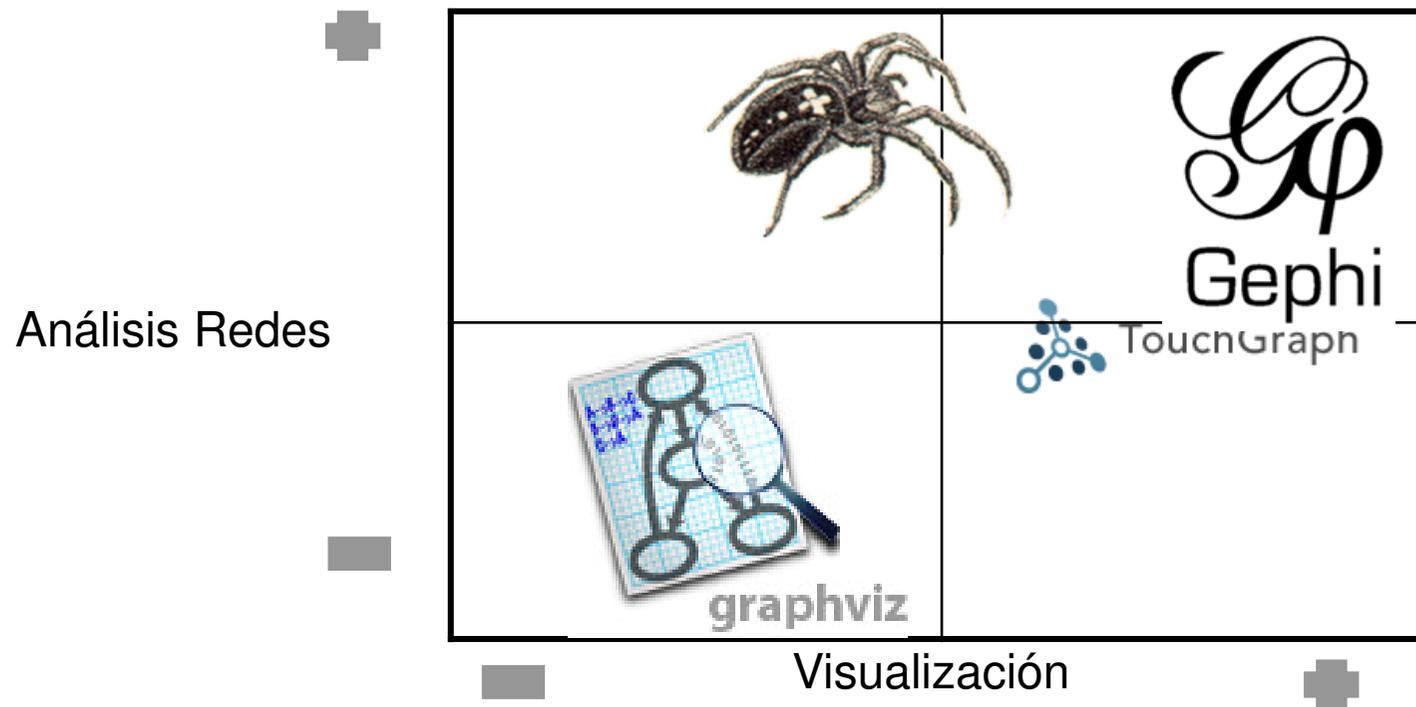
Las fuerzas repulsivas ejercidas sobre un nodo por un cluster de nodos (tratados como un **super-nodo**) lejanos se aproximan por un cálculo de Barnes-Hut

Tiene orden $N \cdot \log(N)$ y es capaz de manejar redes de hasta 100000 nodos



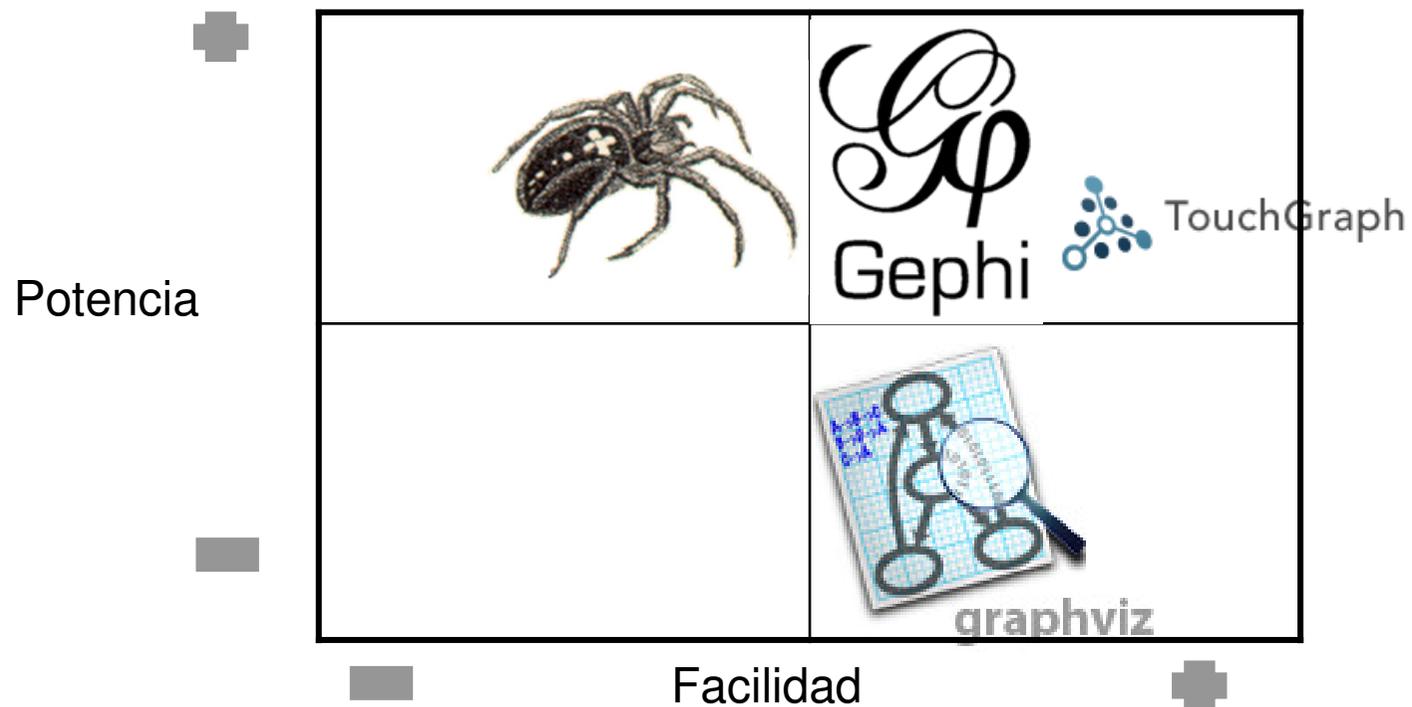
BIBLIOTECAS Y PROGRAMAS DE VISUALIZACIÓN DE REDES

PROGRAMAS DE VISUALIZACIÓN DE REDES (1)



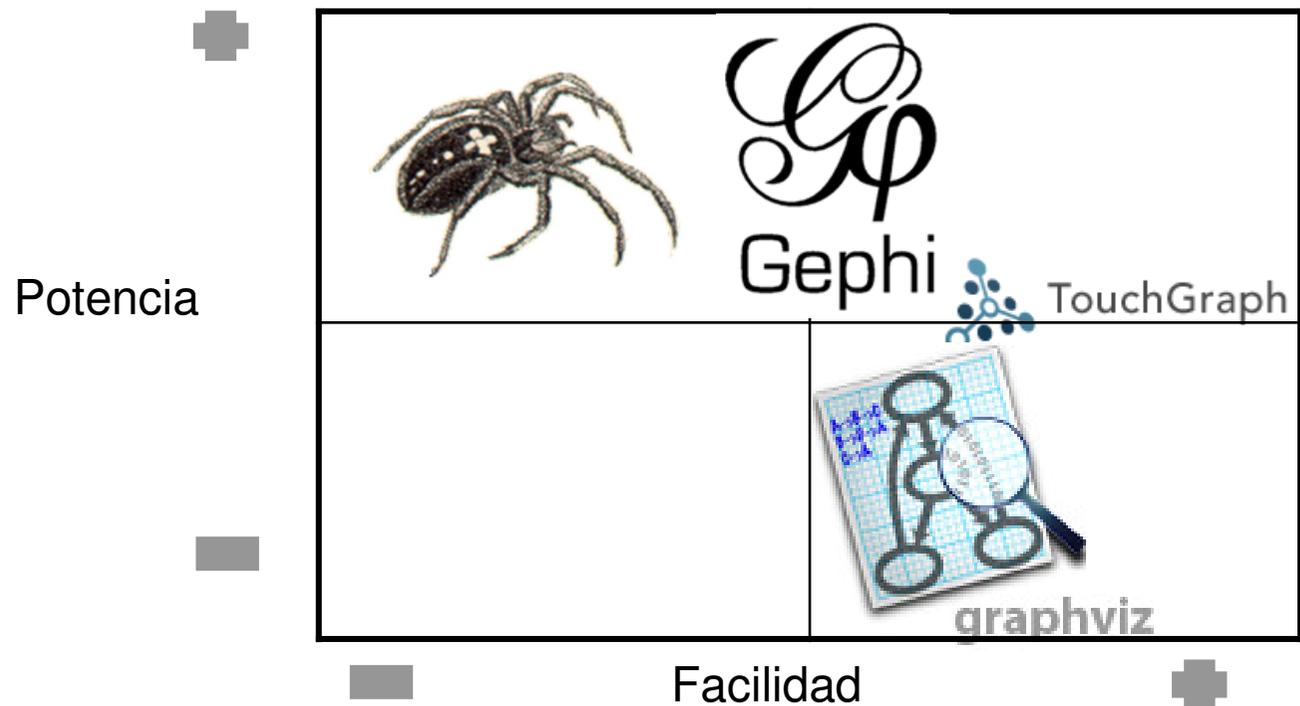
Relación Análisis Redes vs. Visualización

PROGRAMAS DE VISUALIZACIÓN DE REDES (2)



Formatos de entrada: Potencia vs. Facilidad

PROGRAMAS DE VISUALIZACIÓN DE REDES (3)



Interfaz Usuario: Potencia vs. Facilidad

PROGRAMAS DE VISUALIZACIÓN DE REDES: GraphViz (1)

- Toma como entrada descripciones de grafos en formatos de texto simples
- Devuelve salidas en formatos de imagen habituales
- Opciones para formas y colores
- Se puede usar independientemente o como una biblioteca

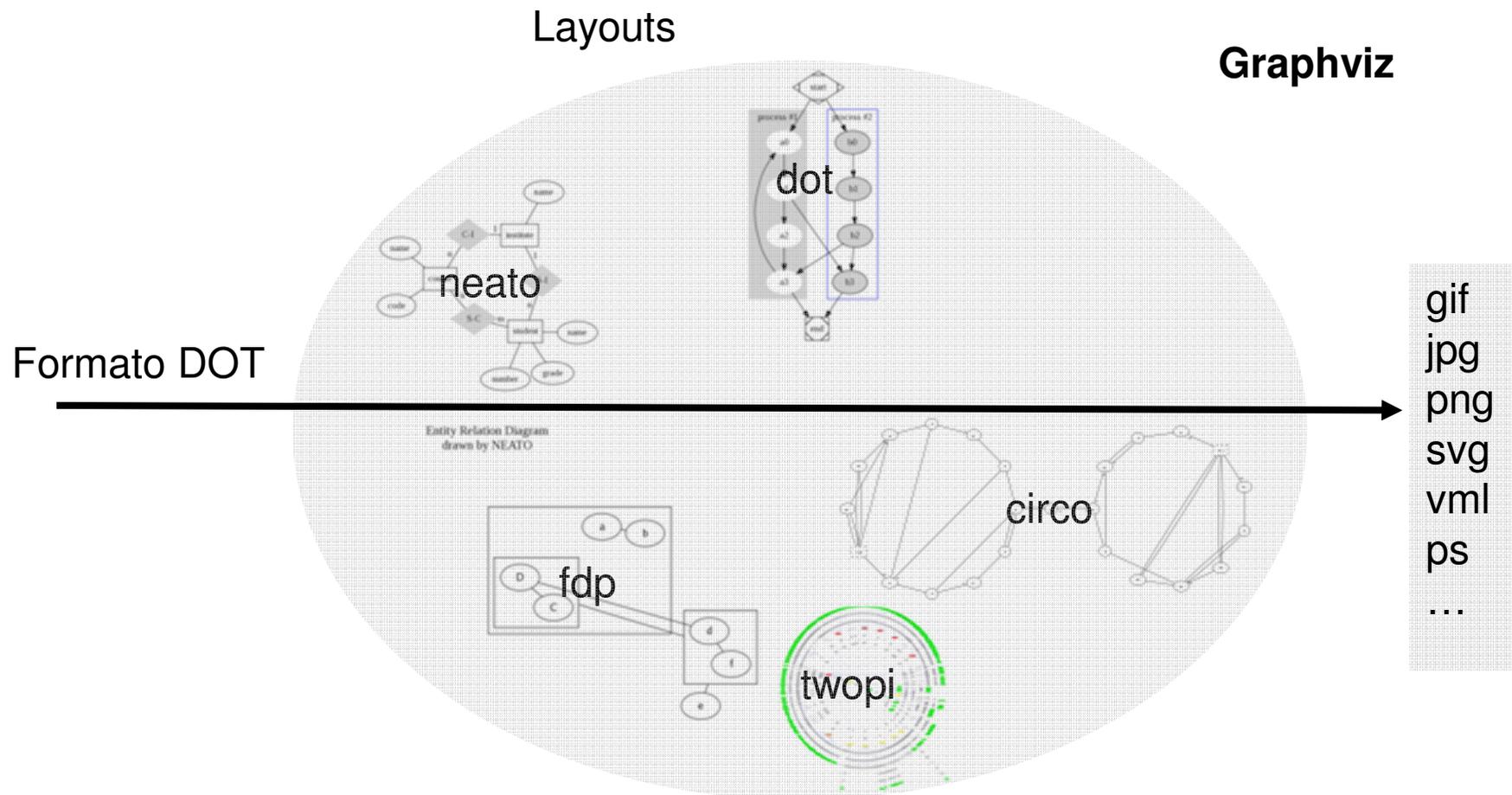


- **dot**: herramienta básica para dibujar grafos dirigidos, jerárquicos o por niveles, evitando cruces entre enlaces y reduciendo el tamaño de enlace
- **fdp**: F&R, **sfdp**: F&R multi-escala para manejar redes grandes
- **neato**: K&K para grafos de 100 nodos
- **twopi**: distribución radial
- **circo**: distribución circular



<http://www.graphviz.org/>

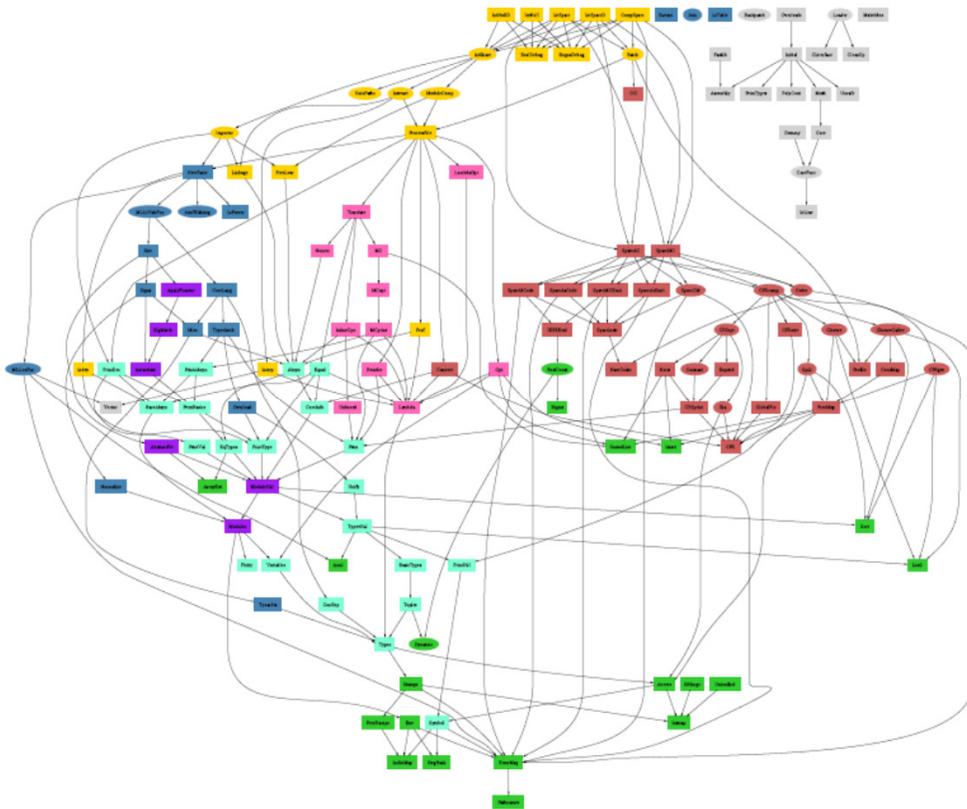
PROGRAMAS DE VISUALIZACIÓN DE REDES: GraphViz (2)



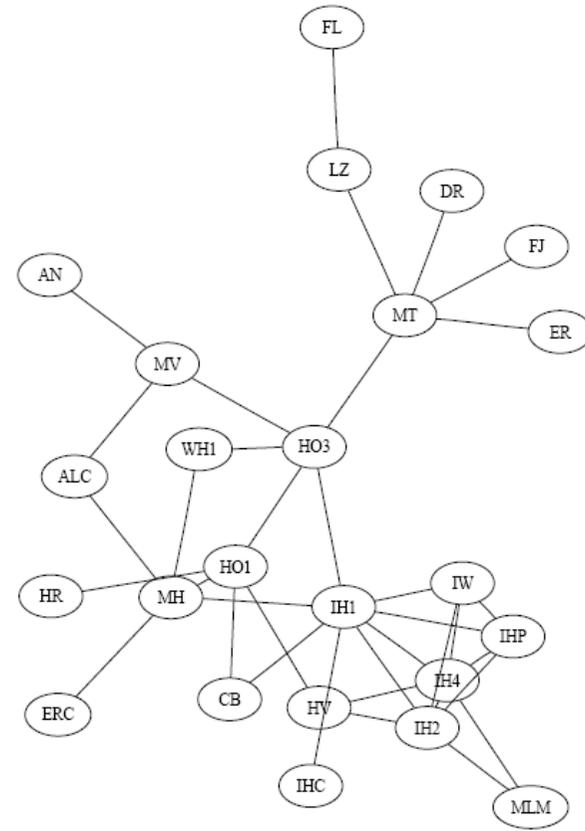
Más información: [Taller Graphviz](#)

PROGRAMAS DE VISUALIZACIÓN DE REDES: GraphViz (3)

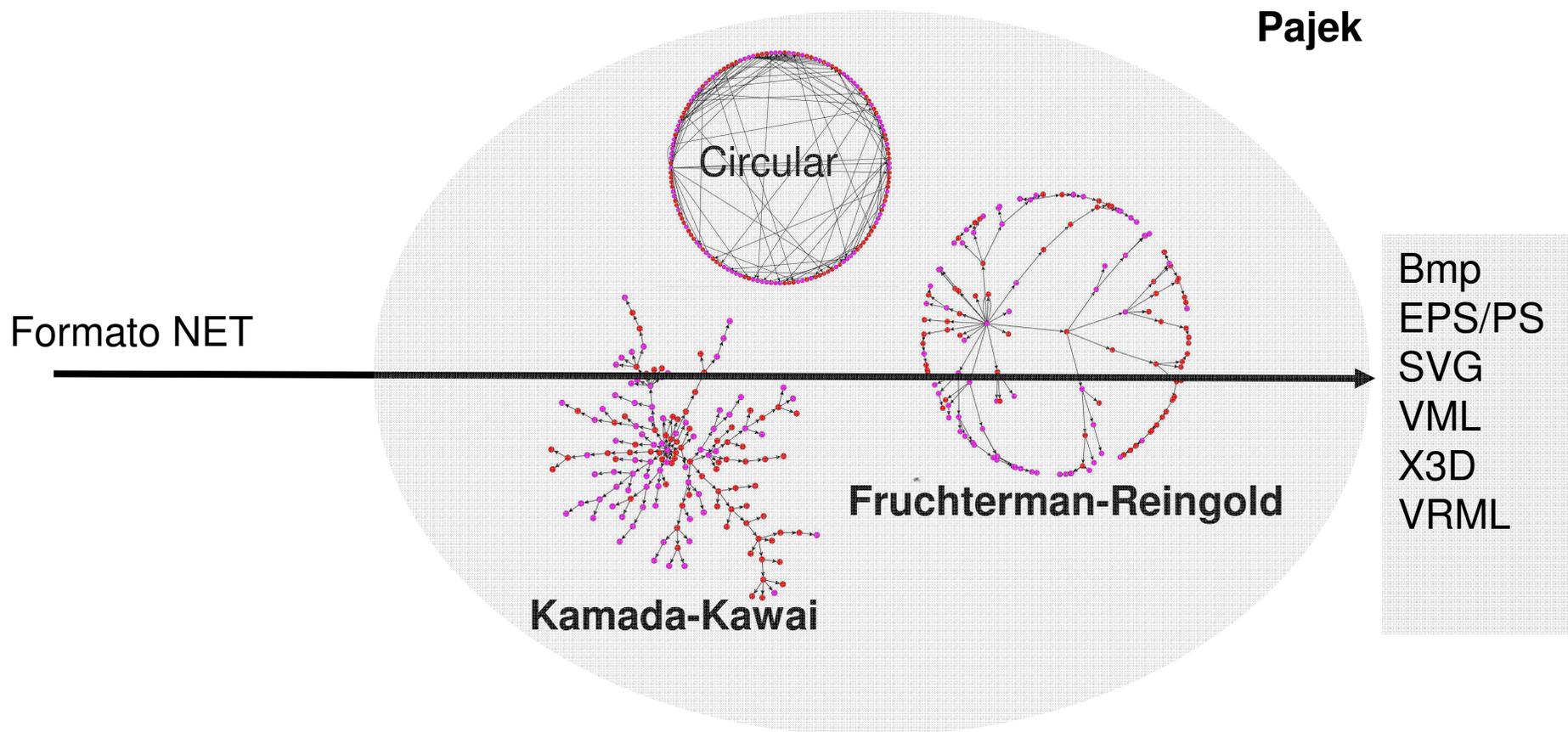
- **Dot:**



- **Neato:**



PROGRAMAS DE VISUALIZACIÓN DE REDES: Pajek (1)



Más información: [Taller Pajek](#)

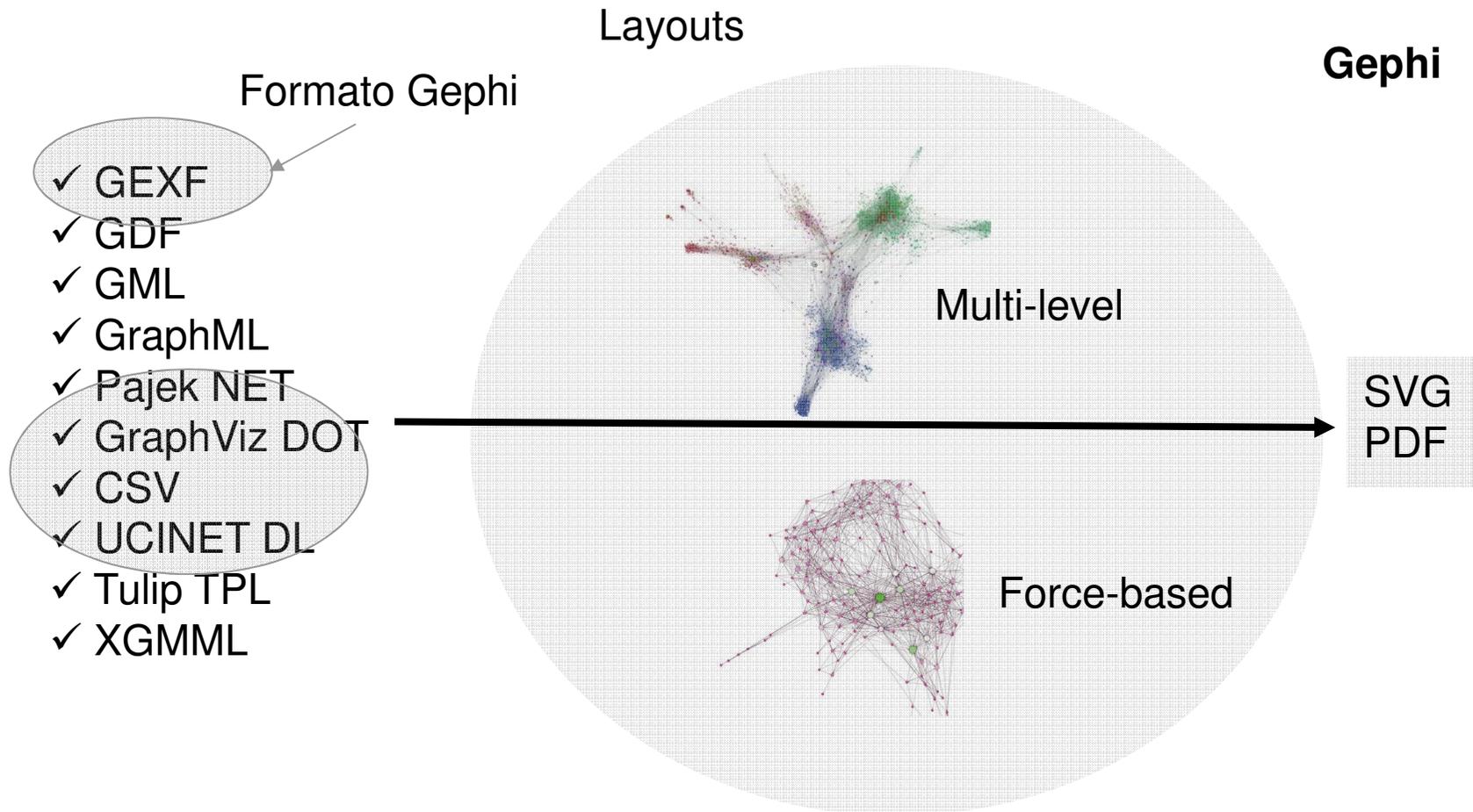
PROGRAMAS DE VISUALIZACIÓN DE REDES: Pajek (2)



Ejemplo: Red de metro de Madrid

<http://blog.vectart.com/2010/10/29/una-nueva-forma-de-ver-el-metro-de-madrid/>

PROGRAMAS DE VISUALIZACIÓN DE REDES: Gephi



Más información: [Taller Gephi](#)

PROGRAMAS DE VISUALIZACIÓN DE REDES: Touchgraph (1)

Características:

- Interfaz intuitiva
- Admite múltiples formatos de entrada, entre ellos Excel con varias pestañas
- Es muy sencillo de utilizar. Maneja redes grandes e identifica comunidades
- Se pueden asociar imágenes a los nodos y representa las propiedades mediante un halo semitransparente

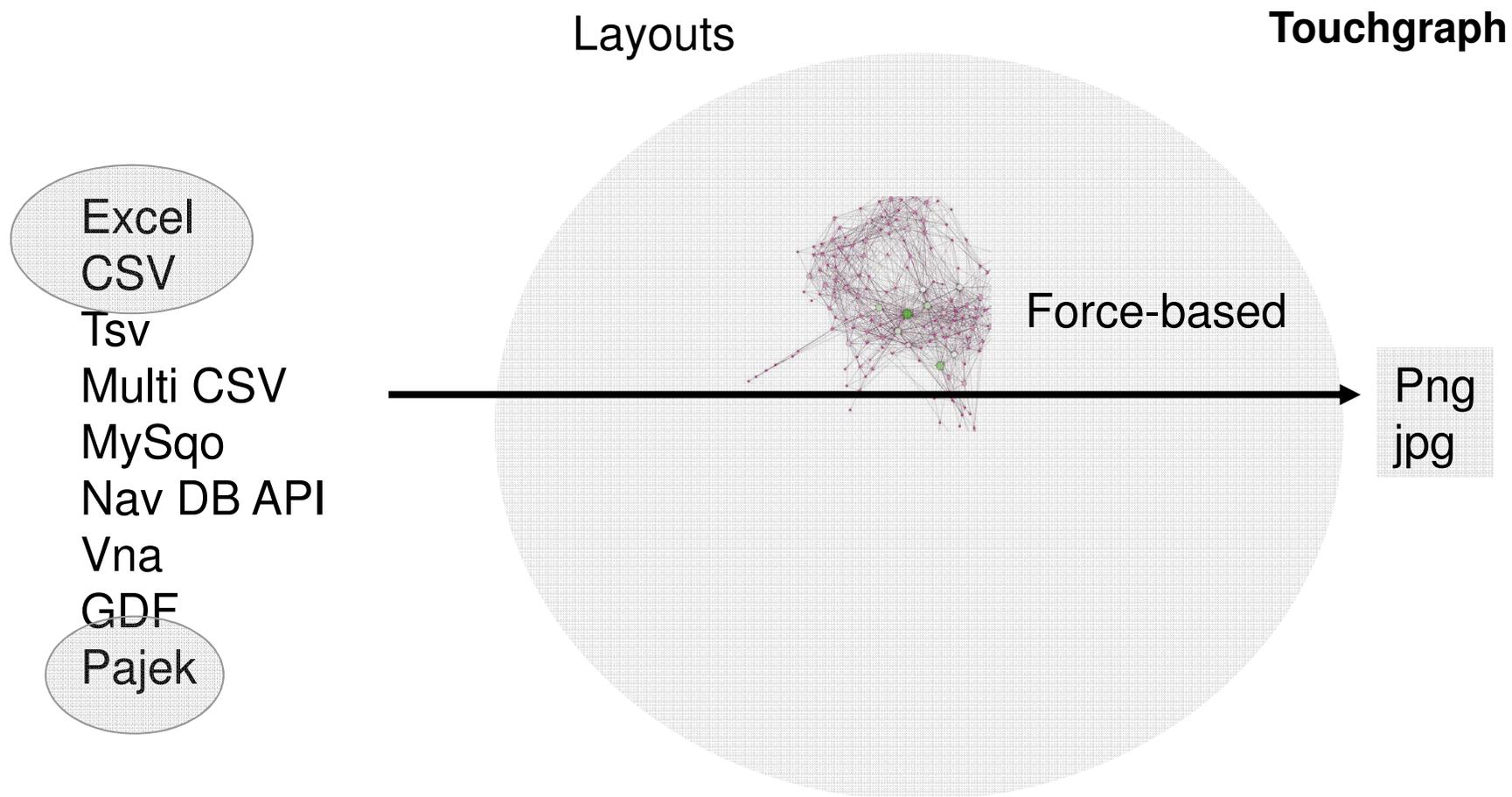


Funcionalidades:

- Interactividad: Permite interactuar con el grafo y ver la conectividad de cada nodo de forma visual
- layout: algoritmos de distribución basados en fuerzas, con parámetros ajustables
- Color: El color de los nodos se puede aplicar por atributos
- Tamaño nodos/enlaces: El tamaño de los nodos y los enlaces se pueden representar según su importancia mediante halos

<http://sourceforge.net/projects/touchgraph/>

PROGRAMAS DE VISUALIZACIÓN DE REDES: Touchgraph (2)



Más información: [Taller Touchgraph](#)

Referencias y Agradecimientos

Para diseñar los materiales de este tema, he hecho uso de material desarrollado por expertos en el área disponible en Internet:

- Curso “SI 614 Networks: Theory and Application”, Lecture 3: Network Visualization Tools, **Lada Adamic**, University of Michigan:

<http://www-personal.umich.edu/~ladamic/courses/networks/si614w06/index.html>

- Curso “CS 765 Complex Networks”, Lecture 27: Network Visualization, **Mehmet Hadi Günes**, University of Nevada, Reno:

<http://www.cse.unr.edu/~mgunes/cs765/cs765fa11/>

- Seminario “Graph Drawing”, Lecture Force directed methods, **Thomas van Dijk**, Universiteit Utrecht:

<http://www.cs.uu.nl/docs/vakken/gdr/>

- Curso “COMP4048: Information Visualisation”, **Seokhee Hong**, Lecture Force Directed Methods and Spring Algorithm, The University of Sydney:

<http://sydney.edu.au/engineering/it/~visual/comp4048/index07.html>

- Seminario “Un paseo por la Visualización en siete días: REDES”, **M^a Luz Congosto**:

<http://www.slideshare.net/mlcm/camon-visualizar20101014>