

Práctica 3:

Estudio comparativo de Métodos para Poda y
Visualización de Redes

Linqi Zhu
X6300759R
zhulinqi@correo.ugr.es

Curso 2022-23

Cuarto Curso del Grado en Ingeniería Informática



**UNIVERSIDAD
DE GRANADA**



1. Análisis de cienciograma

Las tablas de resultados de los 10 cienciogramas de distintos países y años, podado con el algoritmo de Original Pathfinder.

Argentina-2005(n=266)	Número de enlaces / Densidad	Distancia media $\langle d \rangle$
Red original	17938 / 0.509	1.4966
q=2	324 / 0,009	4.7002
q=3	277 / 0,008	5.7263
q=4	269 / 0,008	6.08
q=5	268 / 0,008	6.1138
q=n-1	267 / 0,008	6.2111

Canada-2005(n=288)	Número de enlaces / Densidad	Distancia media $\langle d \rangle$
Red original	30574 / 0.74	1.2604
q=2	340 / 0,008	4.9933
q=3	301 / 0,007	5.9317
q=4	290 / 0,007	6.7343
q=5	287 / 0,007	7.7893
q=n-1	287 / 0,007	7.7893

Chile-2004(n=256)	Número de enlaces / Densidad	Distancia media $\langle d \rangle$
Red original	14778 / 0,453	1.5526
q=2	318 / 0,01	4.7803
q=3	265 / 0,008	6.2451
q=4	258 / 0,008	7.1219
q=5	256 / 0,008	7.4112
q=n-1	256 / 0,008	7.4112

China-2002(n=259)	Número de enlaces / Densidad	Distancia media $\langle d \rangle$

Red original	20661 / 0,618	1.3831
q=2	306 / 0,009	4.9833
q=3	268 / 0,008	5.9861
q=4	262 / 0,008	6.8231
q=5	260 / 0,008	7.5935
q=n-1	260 / 0,008	7.5935

Cuba-2005(n=244)	Número de enlaces / Densidad	Distancia media $\langle d \rangle$
Red original	11306/ 0,381	1.6476
q=2	306 / 0,01	4.4773
q=3	266 / 0,009	5.0316
q=4	266/ 0,009	5.0316
q=5	258 / 0,009	5.5419
q=n-1	258 / 0,009	5.5419

France-2002(n=267)	Número de enlaces / Densidad	Distancia media $\langle d \rangle$
Red original	23986/ 0,675	1.3252
q=2	312 / 0,009	4.9703
q=3	275 / 0,008	5.8775
q=4	271/ 0,008	6.5192
q=5	270 / 0,008	6.7184
q=n-1	268 / 0,008	7.5218

Germany-2002(n=269)	Número de enlaces / Densidad	Distancia media $\langle d \rangle$
Red original	25395/ 0,705	1.2959
q=2	313 / 0,009	5.1805
q=3	277 / 0,008	6.0087
q=4	272/ 0,008	6.8469
q=5	270 / 0,007	7.6983
q=n-1	269 / 0,007	7.9078

Japan-2002(n=265)	Número de enlaces / Densidad	Distancia media $\langle d \rangle$
Red original	21754 / 0,622	1.3793
q=2	316 / 0,009	4.9724
q=3	279 / 0,008	5.6891
q=4	269 / 0,008	6.2717
q=5	267 / 0,008	7.1342
q=n-1	267 / 0,008	7.1342

Mexico-2005(n=270)	Número de enlaces / Densidad	Distancia media $\langle d \rangle$
Red original	20110 / 0,554	1.4504
q=2	336 / 0,009	4.8887
q=3	284 / 0,008	6.0236
q=4	275 / 0,008	6.8280
q=5	274 / 0,008	7.3910
q=n-1	273 / 0,008	7.4173

Portugal-2005(n=265)	Número de enlaces / Densidad	Distancia media $\langle d \rangle$
Red original	20699 / 0,592	1.4102
q=2	318 / 0,009	5.0591
q=3	280 / 0,008	6.1083
q=4	272 / 0,008	6.8657
q=5	270 / 0,008	7.3391
q=n-1	268 / 0,008	7.7503

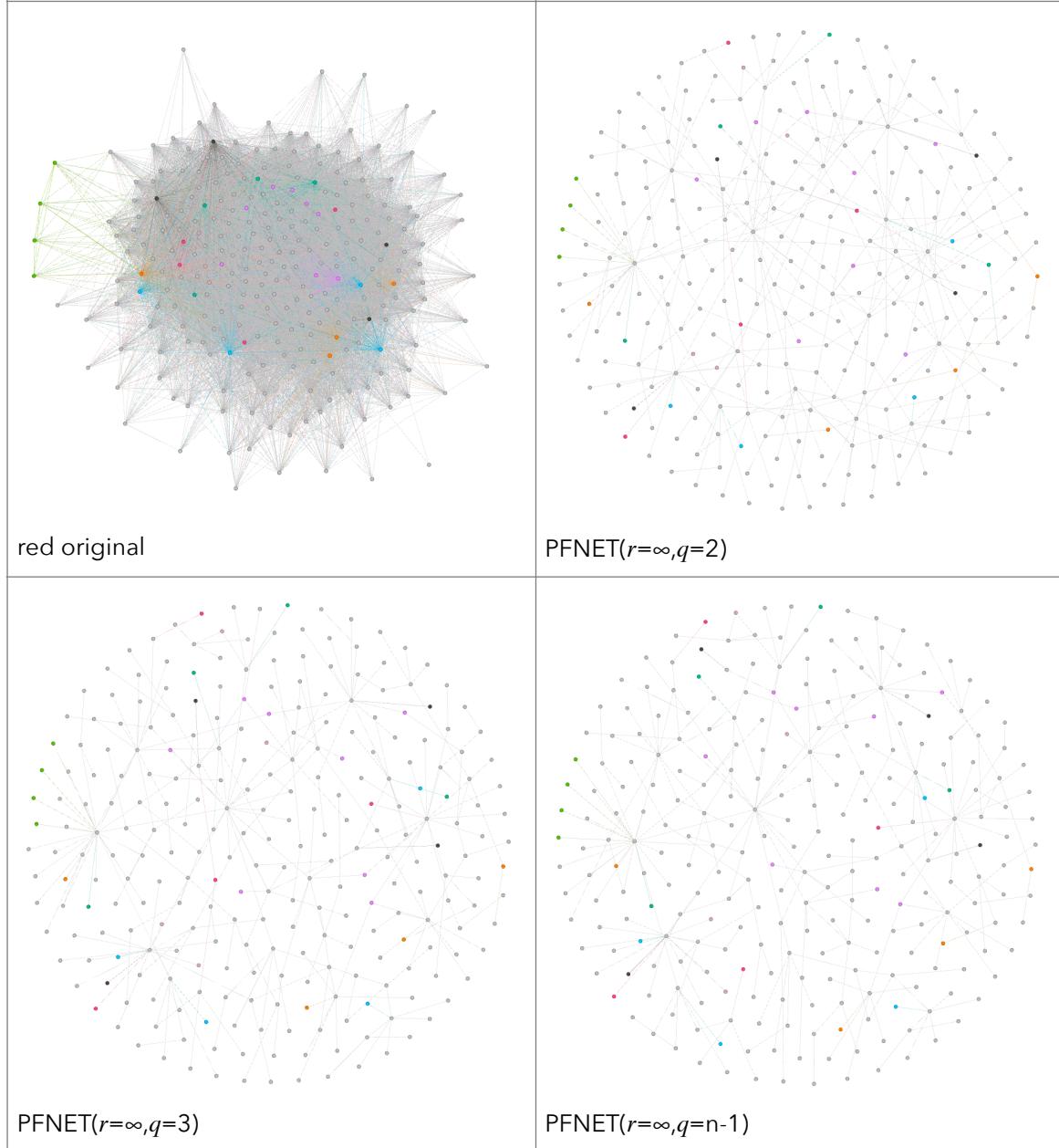
Podemos observar que en general hay una gran diferencia entre numero de enlaces de las redes original y las resultantes con q.

Con mayores valores de q se va rebajando un poco mas la densidad de la red y la distancia media va subiendo(al eliminar muchas conexiones directas, el numero de enlaces necesarios para conectar cada par de nodos de la red es mayor).

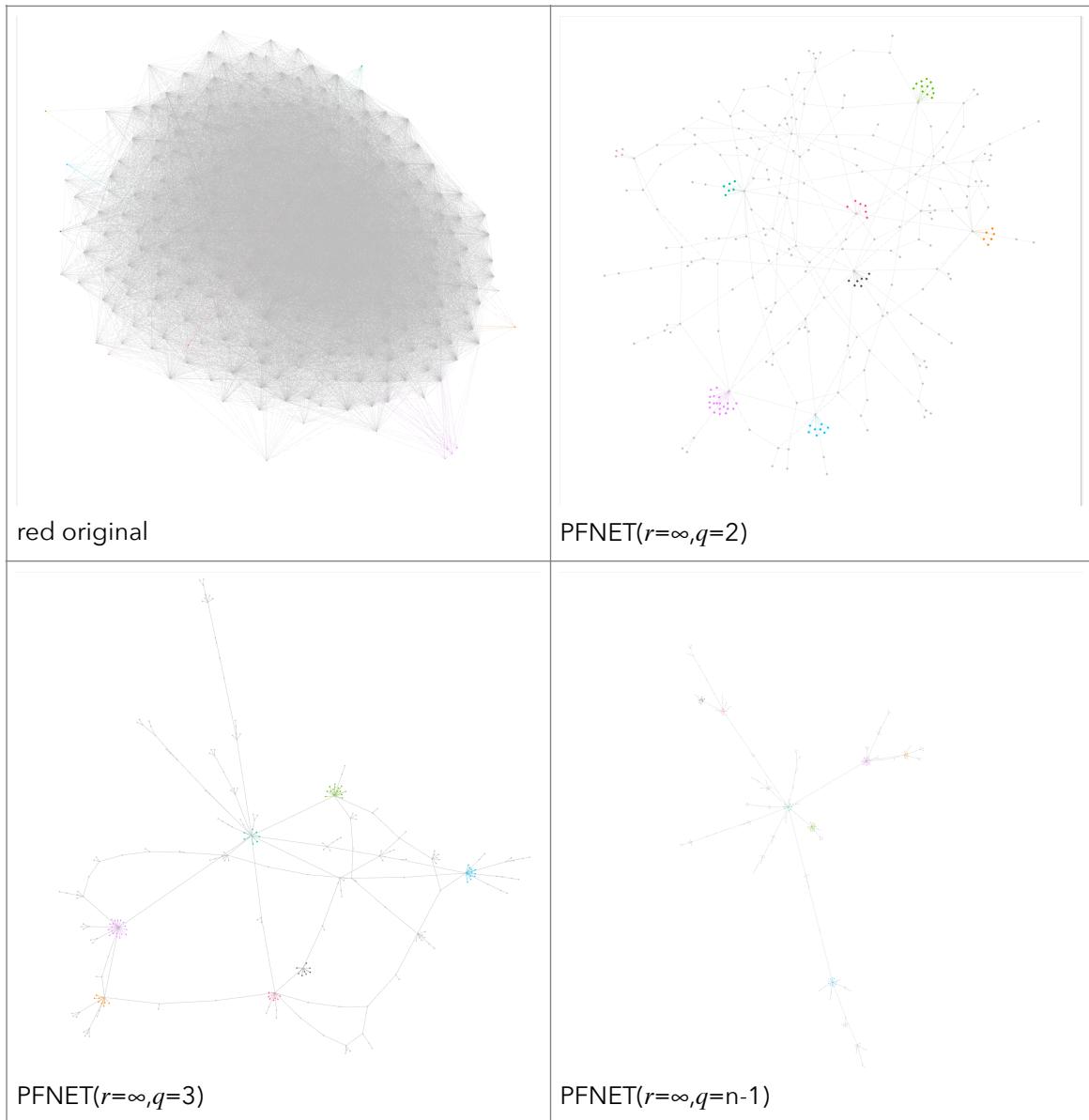
Representación de 4 de PENETs del cienciograma:

Argentina-2005:

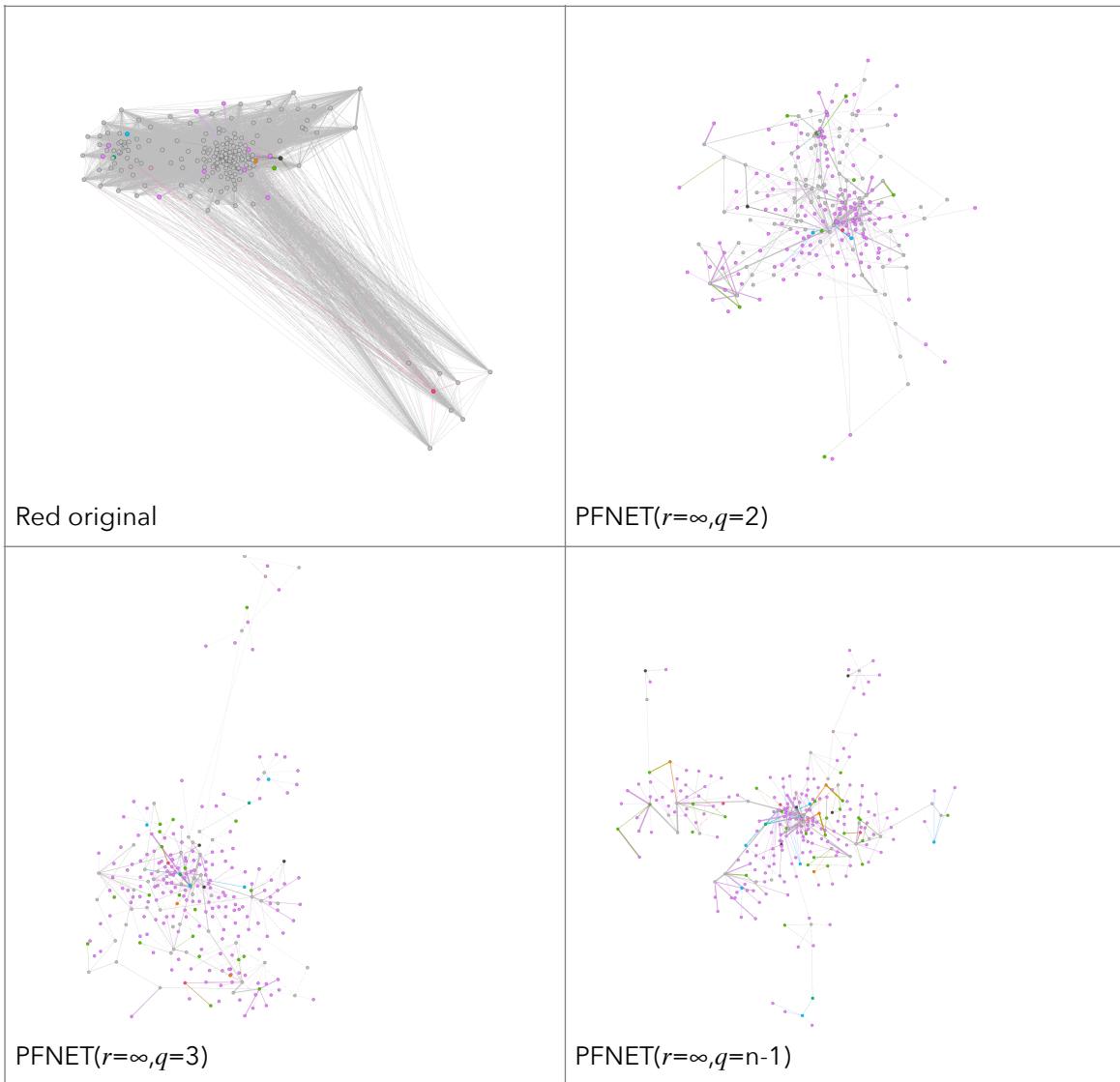
Fruchterman Reingold:



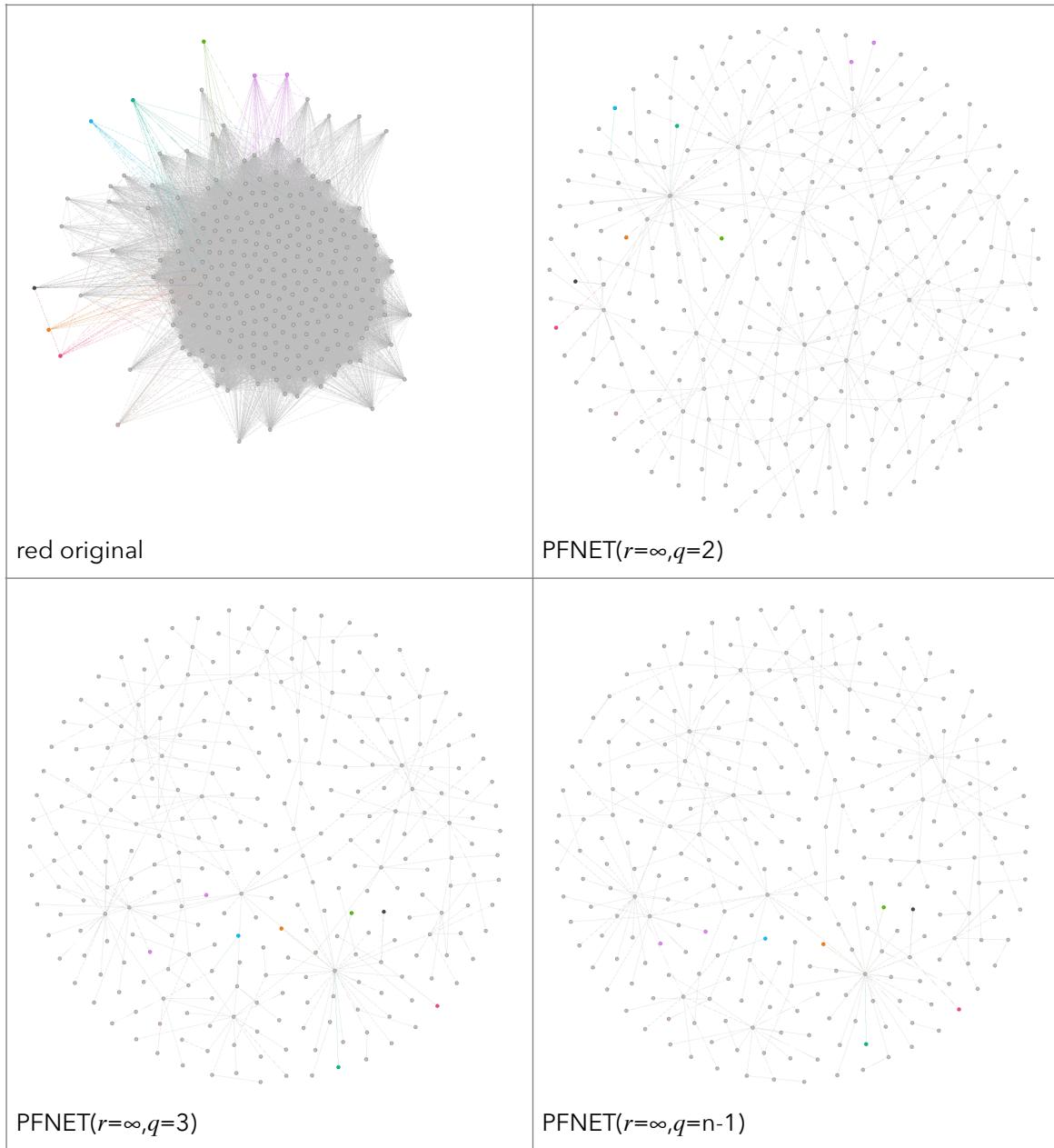
ForceAtlas 2:



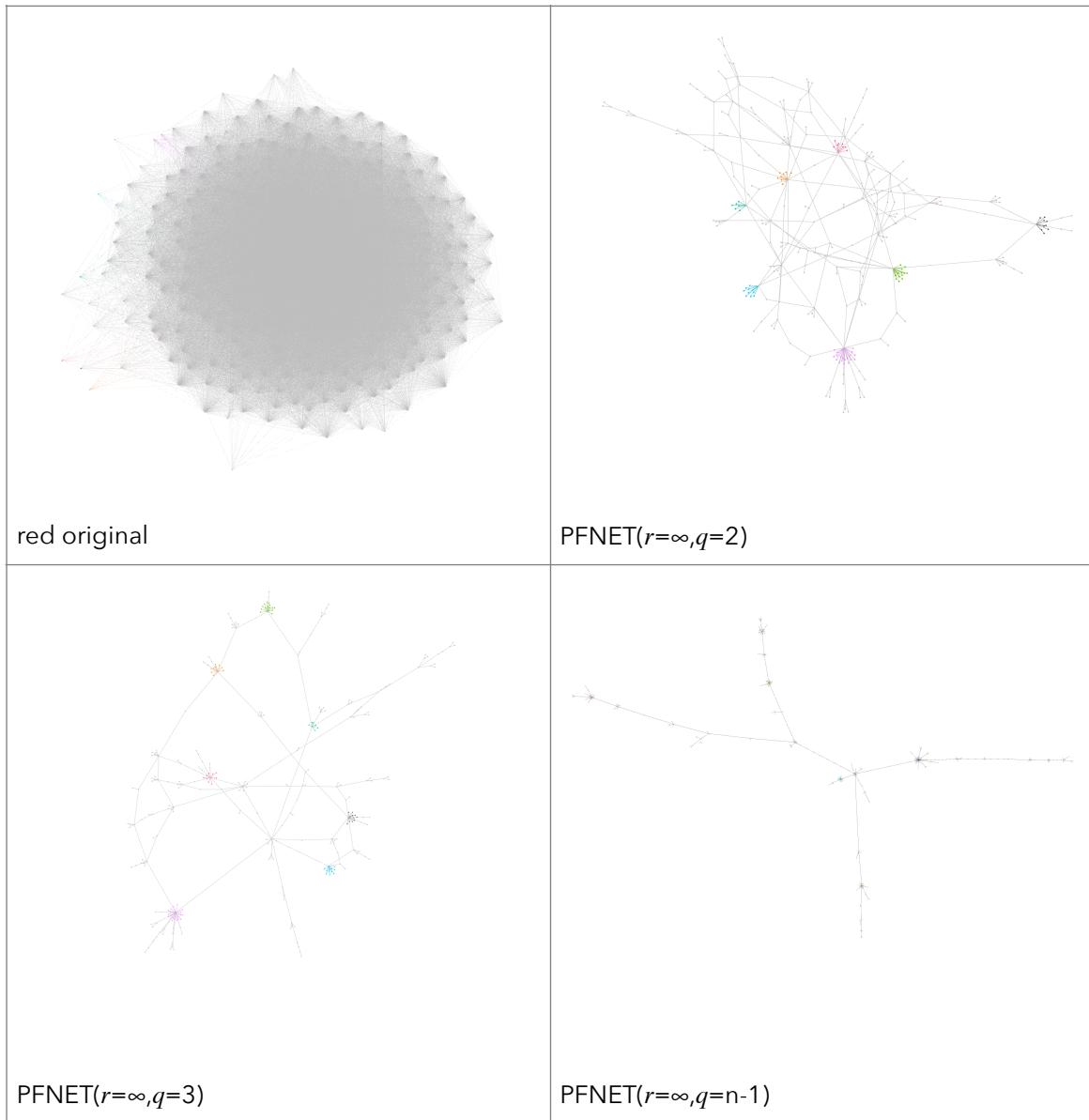
OpenOrd:



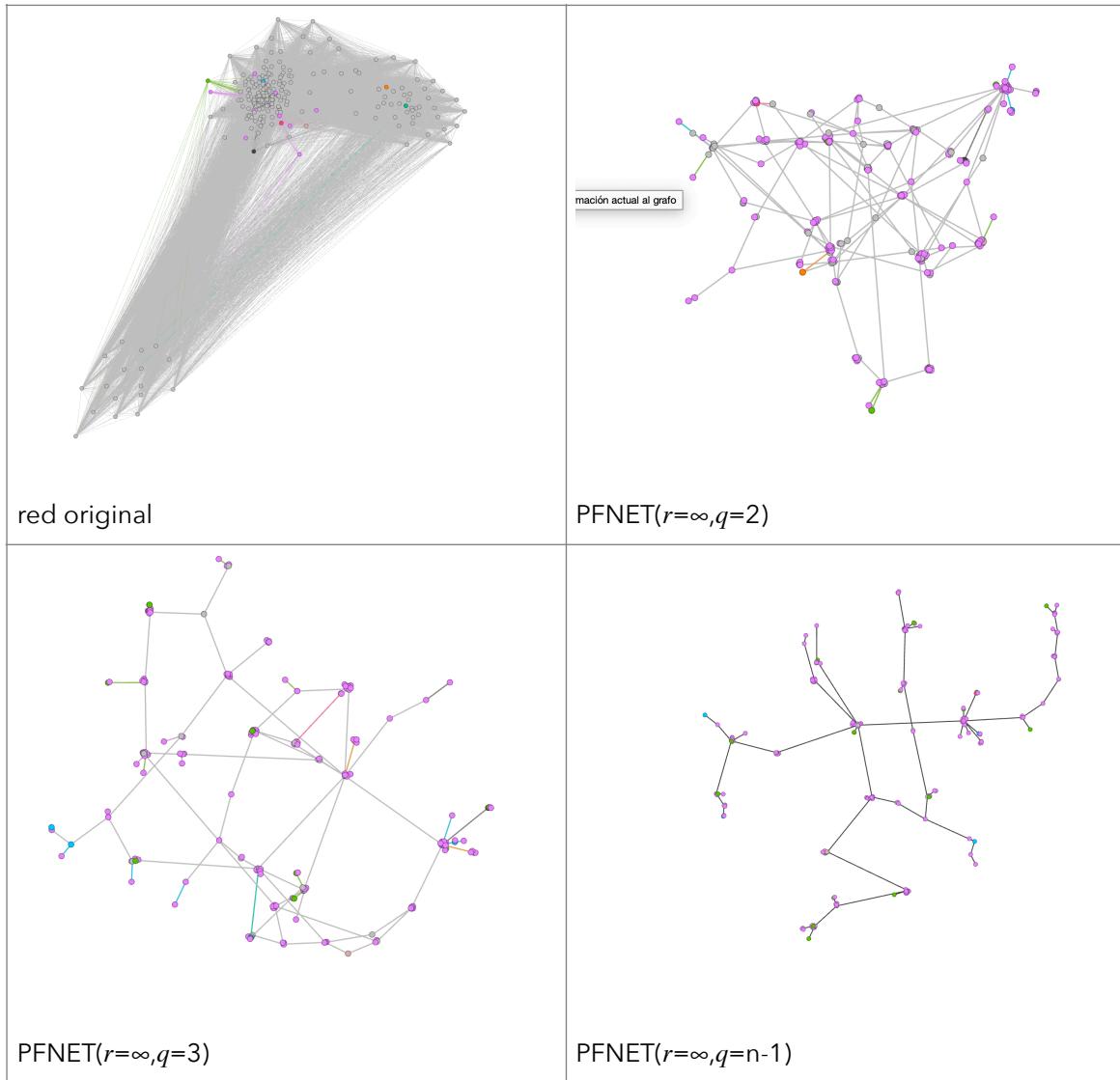
Canada-2005:
Fruchterman Reingold:



ForceAtlas 2:



OpenOrd:



En las visualizaciones de KK se separan más los diferentes grupos de nodo que están fuertemente entre sí, con los diferentes podas de la red podemos ver que va reduciendo número de nodo y arista de la red y podemos tener una visión cada vez más clara de los diferentes grupos.

En las visualizaciones de FR, las redes se dibujan en forma circular, no podemos ver con claridad los grupos de redes (lo que supone difícil para detección de comunidades). Pero podemos ver todos los nodos de la red de forma más unida.

En las visualizaciones de OpenOrd, con las podas se ve cada vez más con claridad como las dos anteriores. Incluso mejor que KK ya que al poner escalado a 100 no muestra de forma unida los nodos se extiende demasiado.

En todas las visualizaciones observamos que hay un cambio muy grande entre red original y red después del algoritmo de poda.

2. Análisis de la eficiencia

Comparación de los tiempos de ejecución de las cinco variantes del método Pathfinder —original, Binary Pathfinder, Fast Pathfinder, MST-Pathfinder de baja complejidad y MST-Pathfinder práctico en redes aleatorias de 500 a 10000 nodos (generado con el código C RandomNets del espacio de la asignatura del Prado).

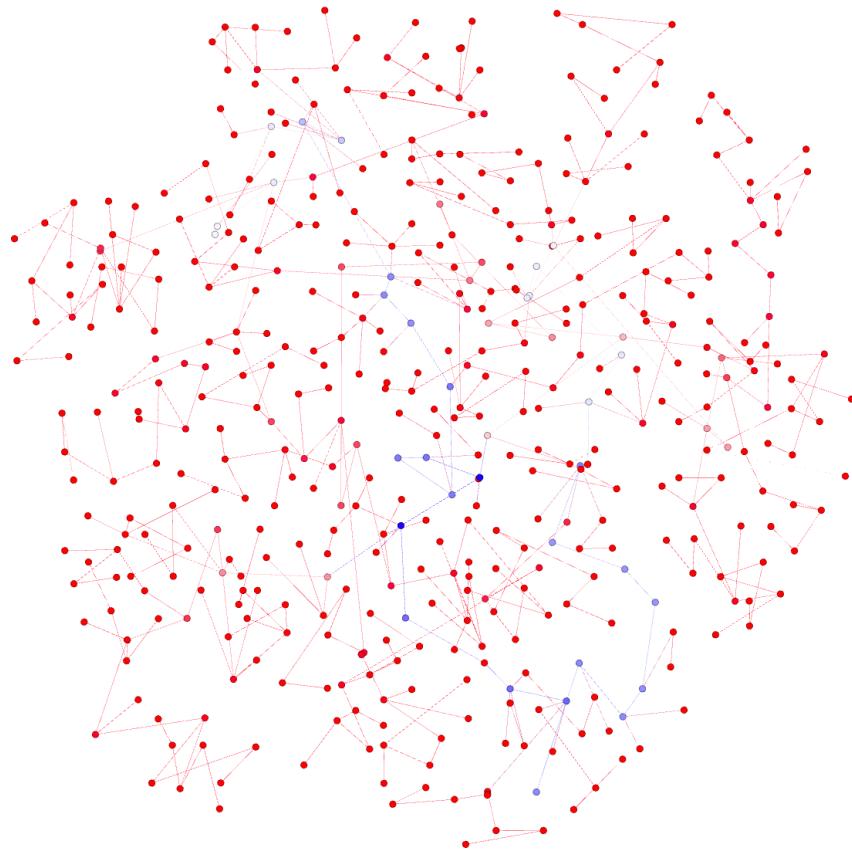
Tabla de comparación:

N	Media E	Media D	Tiempo PF original	Tiempo Binary PF	Tiempo Fast PF	Tiempo MST-PF (baja compl.)	Tiempo MST-PF (Práctico)
500	12415	0.0996	46s	5s	Aprox. 0s	0s	0s
1000	49913.4	0.1	1127s	308.6s	2s	0s	0s
2000	199335.2	0.0998	+1800s	+1800s	15s	0s	0s
5000	1249703.4	0.1	+1800s	+1800s	224s	1s	1s
10000	5000290.2	0.1	+1800s	+1800s	+1800s	6.5s	6.5s

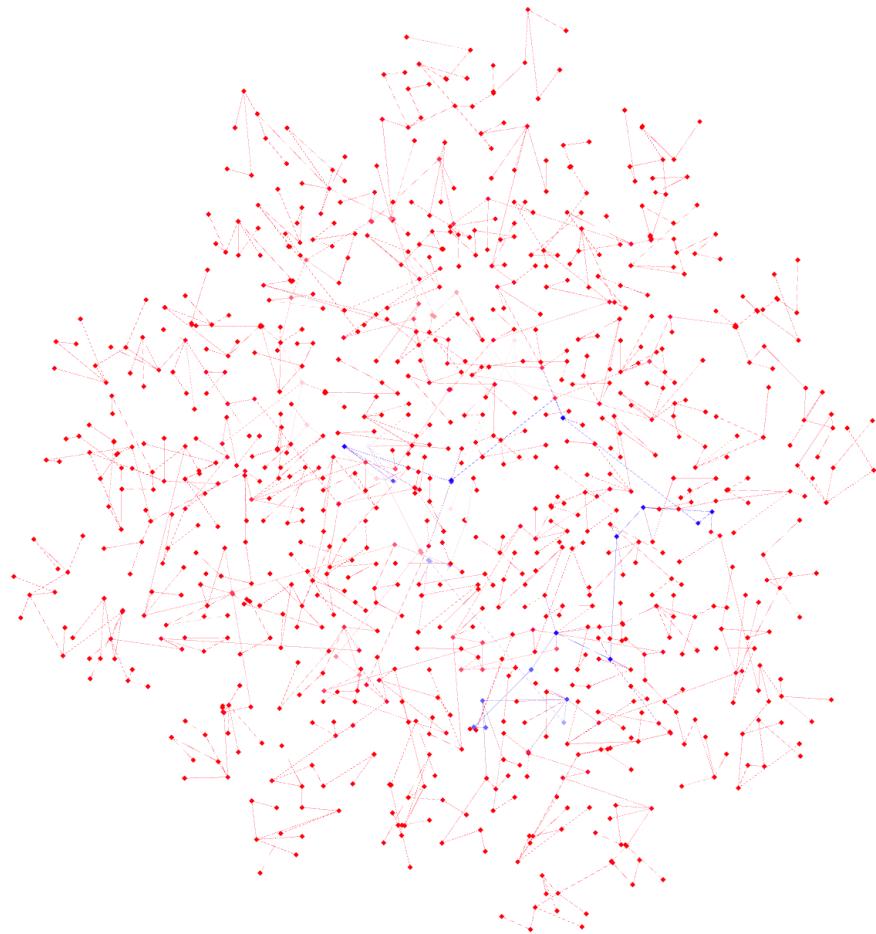
En primer lugar podemos observar que la densidad media de todas las redes es aproximadamente 0.1, y en cuanto tiempo de ejecución:

- El algoritmo de PF original se dispara para tamaño de red pequeño. Para red con nodo más de 2000 tardan más de media hora. No es eficiente para red de grade tamaño. Complejidad teórico es de $O(n^4)$
- El algoritmo de Binary PF es un casi 10 veces más rápido que PF original en red de pequeño tamaño, y 4 veces más rápida en red de 1000 nodos, pero para redes de gran tamaño pasa igual que el algoritmo original. Complejidad es $O(n^3 \log n)$
 - El algoritmo de Fast PF podemos ver que está muy clara la mejora, aunque en la red de 10000 nodos, un tamaño de red bastante grande, se colapsan como los anteriores. Tiene Complejidad de $O(n^3)$
 - El algoritmo de MST-PF de baja complejidad y práctico cuyo ordenes de eficiencia es de $O(n^2 \log n)$ y $O(n^3)$ podemos ver que hay mucha diferencia en tiempo de ejecución, hay una clara mejora en todos redes de los diferentes nodos. La ganancia en velocidad es lo bastante relevante para tamaño de redes muy grandes, al menos en término de velocidad.

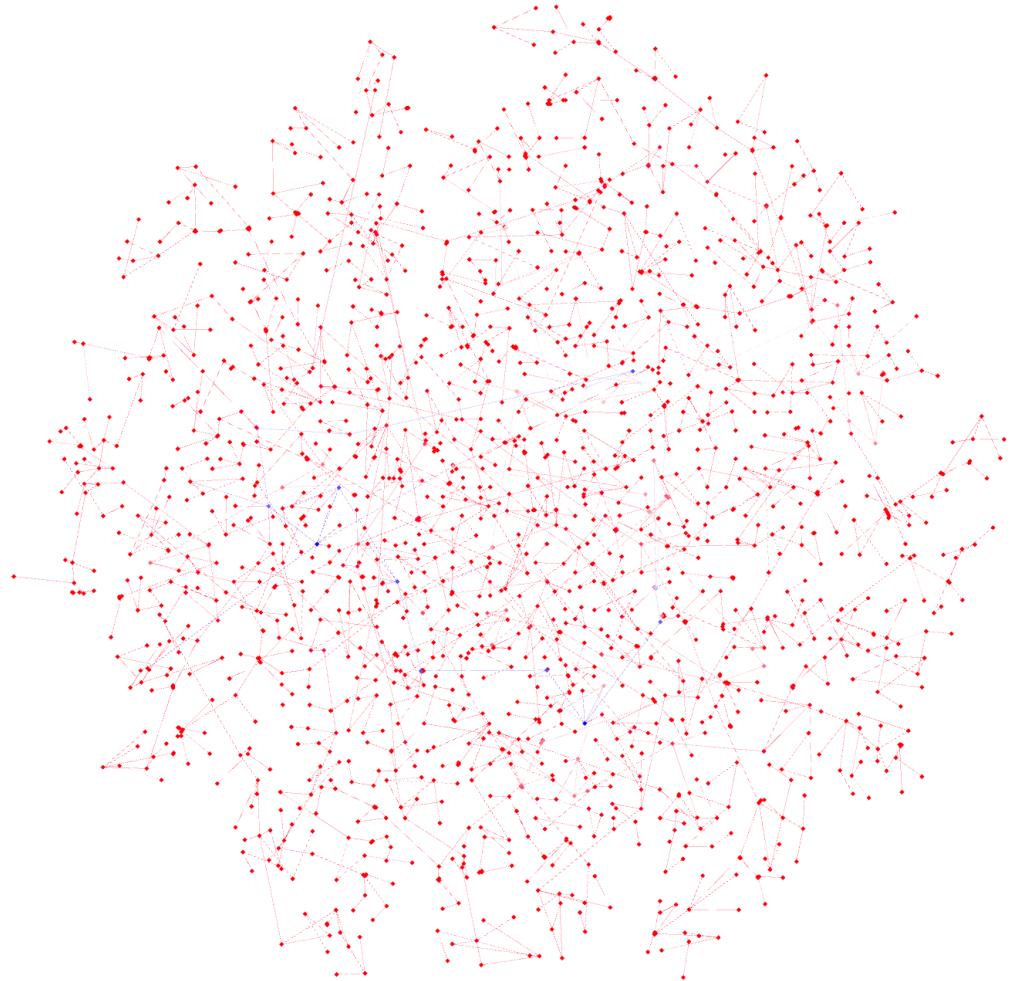
Visualización de redes anteriores, con $r=\infty$ y $q=n-1$, visualización con el algoritmo de distribución OpenOrd (los nodo rojo con pequeño centralidad y azul con mayor centralidad):



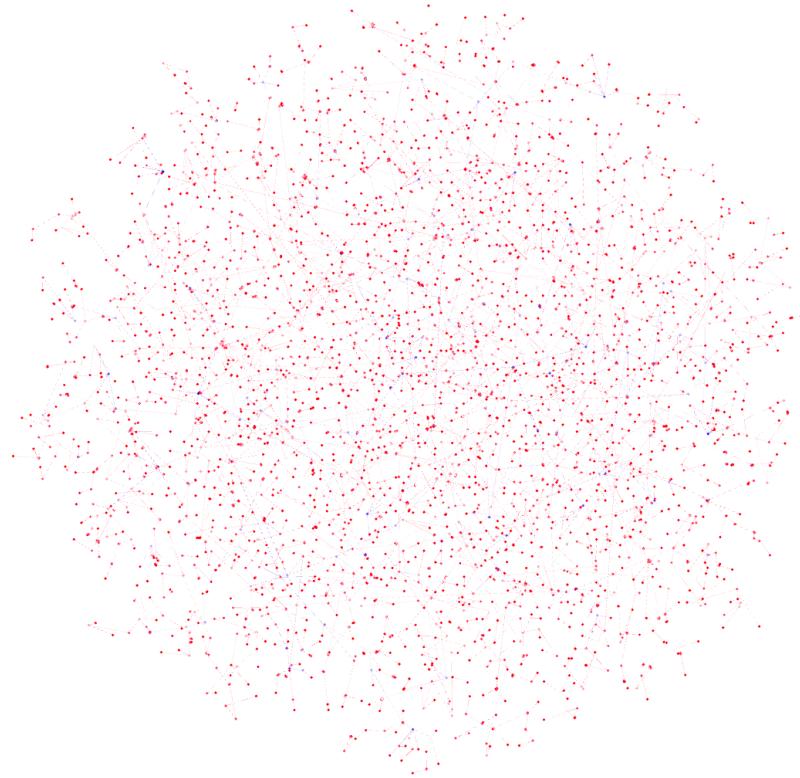
Red aleatoria de Nodo = 500 Aristas = 499



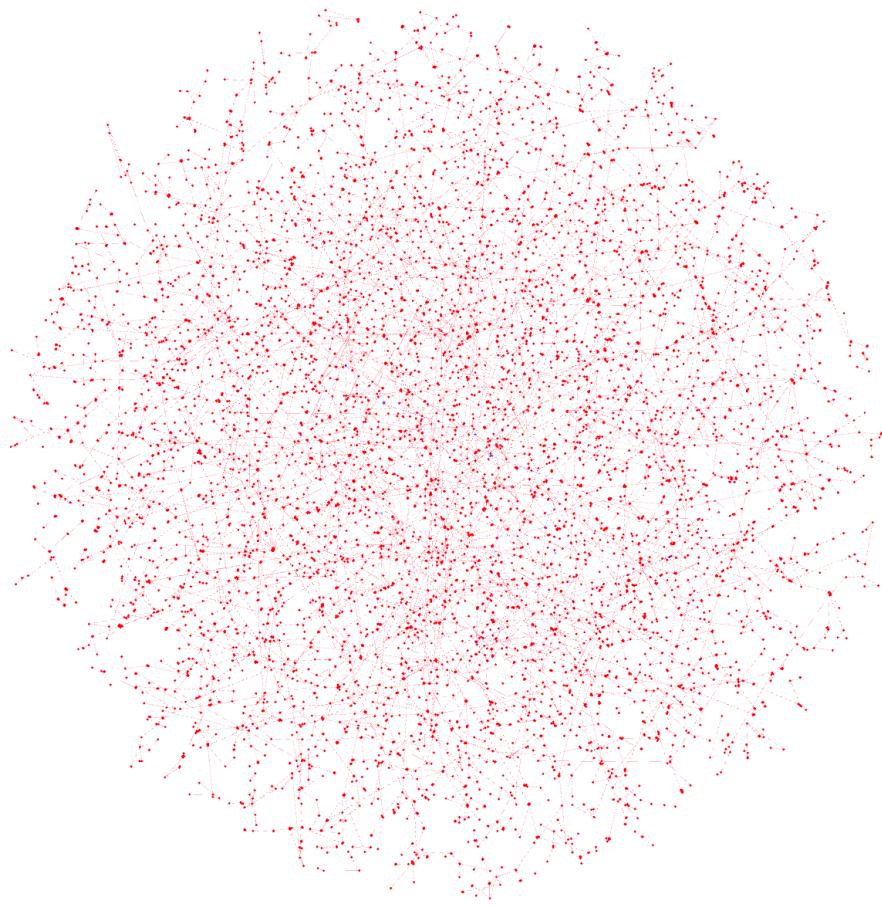
Red aleatoria de Nodo = 1000 Aristas = 999



Red aleatoria de Nodo = 2000 Aristas = 1999



Red aleatoria de Nodo = 5000 Aristas = 5001



Red aleatoria de Nodo = 10000 Aristas = 10001

Podemos ver las cinco redes tiene un aspecto parecido como las visualización F-R.

3. Extras opcionales

- Considerar algoritmo de visualización adicionales en la primera parte de la práctica. Nueva visualización: OpenOrd.