

UNIVERSIDAD DE GRANADA
E.T.S.I. INFORMÁTICA Y TELECOMUNICACIÓN



**UNIVERSIDAD
DE GRANADA**



Departamento de Ciencias de la
Computación e Inteligencia Artificial

Redes y Sistemas Complejos

Guión de Prácticas

Práctica 3:
Estudio Comparativo de Métodos para Poda y
Visualización de Redes

Curso 2017-2018

Cuarto Curso del Grado en Ingeniería Informática

Práctica 3

Estudio Comparativo de Métodos para Poda y Visualización de Redes

1. Objetivos

Esta práctica tiene varios objetivos. Por un lado, comprender el funcionamiento del método *Pathfinder* para la poda de redes complejas y analizar la intensidad obtenida en la poda como resultado de su parametrización. Por otro, aplicar distintos métodos de distribución (*layout*) de redes basados en la filosofía de *force-layout* para determinar cuál es el que mejor se adapta a nuestros objetivos de visualización. Finalmente, realizar un estudio comparativo de la eficiencia de las distintas variantes de *Pathfinder* en su aplicación a redes de gran tamaño. Con estos objetivos globales en mente, dividiremos la práctica en dos partes:

- La primera estará dedicada a la poda y visualización de redes de tamaño pequeño. El estudiante aplicará el algoritmo *Pathfinder* sobre varias redes de este tipo considerando distintos valores para el parámetro q que determina la longitud de los caminos para la que se exige la verificación de la desigualdad triangular. Después analizará la densidad de las PFNETs obtenidas y visualizará algunas de ellas mediante los métodos *Kamada-Kawai* (K-K) y *Frucherman-Reingold* (F-R).
- La segunda parte se dedicará al análisis de la eficiencia de las distintas variantes de *Pathfinder* en redes de gran tamaño. Para ello, se requerirá que el estudiante genere varias redes aleatorias para cada tamaño considerado, ejecute las distintas variantes de *Pathfinder* con $r=\infty$ y $q=n-1$ sobre cada una de ellas y mida los tiempos de ejecución obtenidos en promedio.

El repositorio de GitHub <https://github.com/aquirin/pathfinder> contiene las implementaciones de las distintas variantes del algoritmo *Pathfinder*. Las visualizaciones se pueden realizar en *Gephi* (<http://www.slideshare.net/gephi/gephi-tutorial-layouts>) o *Pajek*. Alternativamente, se puede considerar cualquier otro programa o biblioteca de visualización de grafos¹.

La práctica se evalúa sobre un total de **1,5 puntos**. La fecha límite de entrega será el **Domingo 19 de Noviembre de 2017** antes de las 23:59 horas. La entrega de la práctica se realizará por Internet a través del acceso identificado de la web del departamento de CCIA (<https://decsai.ugr.es>).

¹ Una alternativa posible es la biblioteca de libre distribución *GraphViz* de representación gráfica para grafos de AT&T labs (<http://www.graphviz.org>), que implementa KK y FR en su utilidad *neato*.

2. Trabajo a Realizar en la Primera Parte

Consideraremos las 20 redes científicas (*cienciogramas*) siguientes creadas mediante la co-citación de categorías científicas a partir de datos reales disponibles en la gran base de datos científica *Scopus* de la editorial *Elsevier* (<http://www.scopus.com>):

Argentina (2005)	Mexico (2005)
Canada (2005)	Portugal (2005)
Chile (2004)	Spain (1996)
China (2002)	Spain (1998)
Cuba (2004)	Spain (2002)
Cuba (2005)	Spain (2004)
Europe (2002)	UK (2002)
France (2002)	USA (2002)
Germany (2002)	Venezuela (2005)
Japan (2002)	World (2002)

Estas redes han sido generadas por el grupo de investigación Scimago (<http://www.scimago.com>). La mayoría fueron empleadas en los artículos científicos en los que se propusieron las variantes avanzadas de Pathfinder². Dieciocho de las 20 corresponden a la producción científica de una serie de países concretos en un año concreto mientras que las otras dos representan la misma información para toda Europa y para todo el mundo en el año 2002. Las redes pueden cargarse directamente en Gephi, **simplemente marcando que son no dirigidas** e ignorando los *warnings* que indica la importación.

El estudiante seleccionará una de las variantes de Pathfinder que permita escoger cualquier valor de q y r : Pathfinder original o Binary Pathfinder. Los códigos están disponibles en el repositorio <https://github.com/aquirin/pathfinder>.

Se ejecutará el algoritmo elegido sobre **10 de los 20 cienciogramas** (los que el estudiante quiera). Para r sólo se tomará un valor, ∞ . En cambio, para q se emplearán **5 valores distintos**: $q=\{2, 3, 4, 5, n-1\}$. Para cada uno de los cienciogramas elegidos, se obtendrá una tabla con el siguiente formato (un total de 10 tablas):

Tabla 2.X: Resultados obtenidos en el cienciograma X

<Nombre cienciograma> ($n=<xxxx>$)	Número de enlaces / Densidad	Distancia media < d >
Red original	<xxx> / <xxx>	<xxx>
$q=2$	<xxx> / <xxx>	<xxx>
$q=3$	<xxx> / <xxx>	<xxx>
$q=4$	<xxx> / <xxx>	<xxx>
$q=5$	<xxx> / <xxx>	<xxx>
$q=n-1$	<xxx> / <xxx>	<xxx>

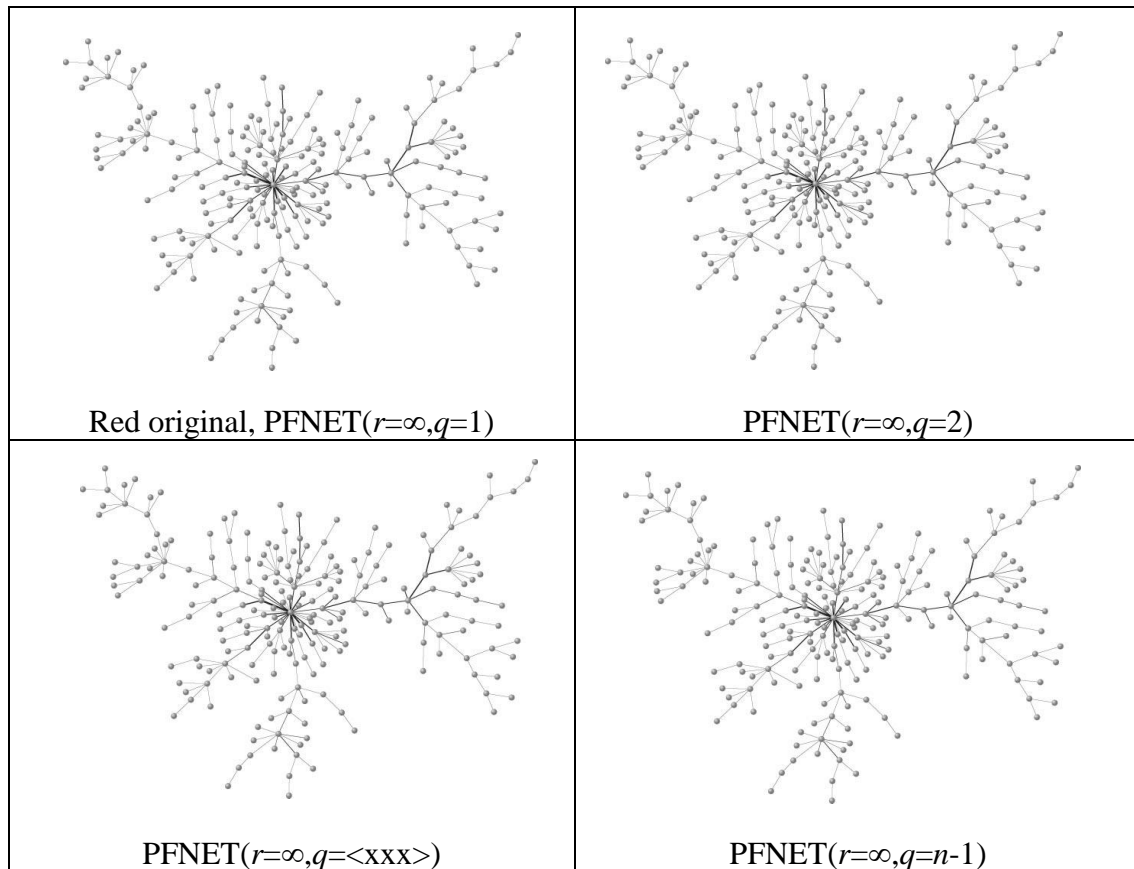
² Quirin, A., Cordón, O., et al. 2008. A Quick MST-Based Algorithm to Obtain Pathfinder Networks ($\infty, n-1$). Journal of the American Society for Information Science and Technology 59: 1912–1924

A continuación, el estudiante realizará un análisis de los resultados obtenidos en las tablas, comentando el patrón de poda asociado a los valores de q en cada cienciograma y tratando de extraer conclusiones generales y específicas del tipo “¿qué correlación hay entre la intensidad de la poda y el valor de q ?”, “¿qué correlación hay entre la distancia media de la red podada y el valor de q ?”, “¿a partir de qué valor de q la $PFNET(r=\infty, q)$ alcanza su máximo nivel de poda?”, “¿hay mucha dispersión en ese valor umbral de q dependiendo del cienciograma concreto?”, etc.

El siguiente paso será la visualización de los cienciogramas. El estudiante escogerá **2 de los 10 cienciogramas** considerados y representará **4 de las PFNETs asociadas a cada una de ellas** empleando dos algoritmos de distribución distintos, K-K y F-R. Esas 4 PFNETs serán la red original, es decir, la $PFNET(r=\infty, q=1)$, la $PFNET(r=\infty, q=2)$, la $PFNET(r=\infty, q=n-1)$ y otra intermedia a escoger por el estudiante.

De este modo, generará un total de 16 visualizaciones distintas (2 cienciogramas \times 4 valores de $q \times$ 2 métodos de distribución) y los presentará agrupados en bloques de 4 del siguiente modo:

Figura 2.XY: Visualizaciones de las PFNETs del cienciograma X realizadas con el método de distribución Y



Finalmente, el estudiante realizará un estudio comparativo de las capacidades de cada método de distribución para obtener una *visualización estética*.

3. Trabajo a Realizar en la Segunda Parte

El objetivo es comparar el tiempo de ejecución de las cinco variantes del método Pathfinder —original, Binary Pathfinder, Fast Pathfinder, MST-Pathfinder de baja complejidad y MST-Pathfinder práctico (<https://github.com/aquirin/pathfinder>)— en redes aleatorias de 500 a 10000 nodos.

Para ello, el estudiante generará **5 redes aleatorias distintas para cada uno de los 5 tamaños siguientes**: $n = \{500, 1000, 2000, 5000, 10000\}$. Para ello, puede utilizar el código *C RandomNets* disponible en el espacio de la asignatura en PRADO³ o cualquier otro de una biblioteca de libre distribución que permita generar grafos ponderados aleatorios. El estudiante podrá elegir la densidad deseada para las redes generadas y deberá indicarla en la memoria de la práctica. Luego ejecutará las **5 variantes de Pathfinder con $r=\infty$ y $q=n-1$** sobre cada una de las 5 redes de cada tamaño y medirá los tiempos de ejecución (en segundos) obtenidos **en promedio** para cada tamaño. **El tiempo máximo de ejecución será de 1800 segundos** (media hora) en cualquier caso.

Con los datos obtenidos, se construirá una tabla con el siguiente formato:

Tabla 3.X: Comparación de los tiempos de ejecución de las variantes de Pathfinder en redes aleatorias

n	Media $ E $	Media D	Tiempo PF original	Tiempo Binary PF	Tiempo Fast PF	Tiempo MST-PF (baja cmpl.)	Tiempo MST-PF (práctico)
500	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>
1000	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>
2000	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>
5000	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>
10000	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>

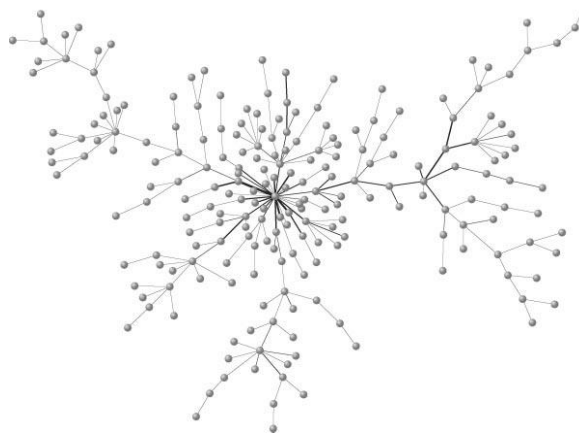
Nótese que cada una de las 5 filas corresponde a los resultados medios de las ejecuciones realizadas sobre las 5 redes generadas para ese tamaño. De este modo, la celda “Media $|E|$ ” de la fila $n=500$ debe recoger la media del número de enlaces de las 5 redes de ese tamaño, la celda “Media D” la media de las densidades de esas 5 redes, la celda “Tiempo PF original” la media del tiempo de ejecución empleado por el algoritmo Pathfinder original para realizar la poda de cada una de las 5 redes, etc.

El estudiante realizará un análisis de los resultados obtenidos. En particular, analizará detenidamente las diferencias de tiempo entre las dos variantes del MST-Pathfinder, la de baja complejidad (de orden $O(n^2 \cdot \log n)$) y la práctica (de orden $O(n^3)$).

³ Se recomienda compilar con la opción `-O3` para acelerar la ejecución de las variantes de Pathfinder.

Por último, el estudiante visualizará **3 de las PFNETs obtenidas**, una de 2000, otra de 5000 y otra de 10000 nodos. Para ello, empleará el algoritmo de distribución que desee, tratando de obtener una buena visualización para una red de gran tamaño desde un punto de vista estético. Presentará las visualizaciones obtenidas en 3 gráficos independientes, indicando el algoritmo de distribución considerado.

Figura 3.XY: Visualización de la PFNET($r=\infty$, $q=n-1$) de una red aleatoria realizada con el método de distribución Y. La PFNET tiene $n=X$ y $|E|=\langle xxx \rangle$



4. Extras Opcionales

El estudiante podrá realizar algún trabajo extra en la práctica que se valorará con hasta 1 punto adicional. Las opciones posibles son las siguientes:

- Considerar algoritmos de visualización adicionales en la primera parte de la práctica.
- Ampliar el tamaño de las redes aleatorias en la segunda parte de la práctica.
- Realizar un análisis de eficiencia híbrido para las distintas variantes de Pathfinder en la segunda parte (para ello, se puede usar el software *xmgrace* (<http://plasma-gate.weizmann.ac.il/Grace/>)).

5. Documentación y Ficheros a Entregar

La **documentación** de la práctica será un fichero *pdf* que deberá incluir, al menos, el siguiente contenido:

- a) Portada con el número y título de la práctica, el curso académico y el nombre, DNI y dirección e-mail del estudiante.
- b) Índice de contenidos.
- c) Una sección que incluya el material asociado a la primera parte de la práctica:
 - Indicación de la variante Pathfinder considerada y de los cienciogramas seleccionados.
 - Tablas de resultados y visualizaciones de las redes.
 - Los dos análisis de resultados, el correspondiente a la poda Pathfinder y el correspondiente a la visualización.
- d) Una sección que incluya el material asociado a la segunda parte de la práctica:
 - Tablas de resultados y visualizaciones de las redes, con indicación de los parámetros considerados para generar las redes aleatorias, el algoritmo de distribución considerado en cada caso, etc.
 - Análisis de resultados correspondiente al estudio comparativo de los tiempos de ejecución de las variantes de Pathfinder.
- e) Una sección que describa los extras realizados (en caso de haber hecho alguno).
- f) Referencias bibliográficas u otro tipo de material distinto del proporcionado en la asignatura que se haya consultado para realizar la práctica (en caso de haberlo hecho).

Aunque lo esencial es el contenido, también debe cuidarse la presentación y la redacción.

El fichero *pdf* de la documentación, los distintos ficheros de las redes podadas, los ficheros de las visualizaciones (proyectos de *Gephi*, ficheros de *Pajek*, etc.) y los gráficos de las mismas se comprimirán conjuntamente en un fichero *.zip* etiquetado con los apellidos y nombre del estudiante (Ej. Pérez Pérez Manuel.zip). Este fichero será entregado por internet a través del espacio de la asignatura en la plataforma PRADO2.