

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



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 2013-2014

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 alumno 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. Posteriormente, 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 alumno 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.

La página web <http://aquirin.ovh.org/research/mstpathfinder.html> 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 **Viernes 29 de Noviembre de 2013** 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.scimagolab.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. Todas están en formato *Pajek*, por lo que si el alumno desea manejarlas con *Gephi* tendrá que convertir los ficheros a un formato admitido por este último.

El alumno seleccionará una de las variantes de Pathfinder que permita escoger cualquier valor de q y r : Pathfinder original, Binary Pathfinder y Fast Pathfinder. Los códigos están disponibles en la página web <http://aquirin.ovh.org/research/mstpathfinder.html>. En cualquiera de los casos, **habrá que realizar una pequeña modificación en el código** para considerar la distancia de Minkowski genérica con cualquier valor de r (en la implementación actual, el código está particularizado para el caso de $r=\infty$).

Se ejecutará el algoritmo elegido sobre **10 de los 20 scienciogramas** (los que el alumno quiera) considerando distintos valores para r y q . En concreto, para r sólo se tomarán dos valores, 1 e ∞ . En cambio, para q se emplearán 10 valores distintos, equidistribuidos en el conjunto de valores posibles: $q=\{2, n/8, 2\cdot n/8, \dots, 7\cdot n/8, n-1\}$.

Para cada uno de los scienciogramas elegidos, se obtendrá una tabla con el siguiente formato (un total de 10 tablas):

² Quirin, A., Cordon, 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

Tabla 2.X: Resultados obtenidos en el cienciograma X

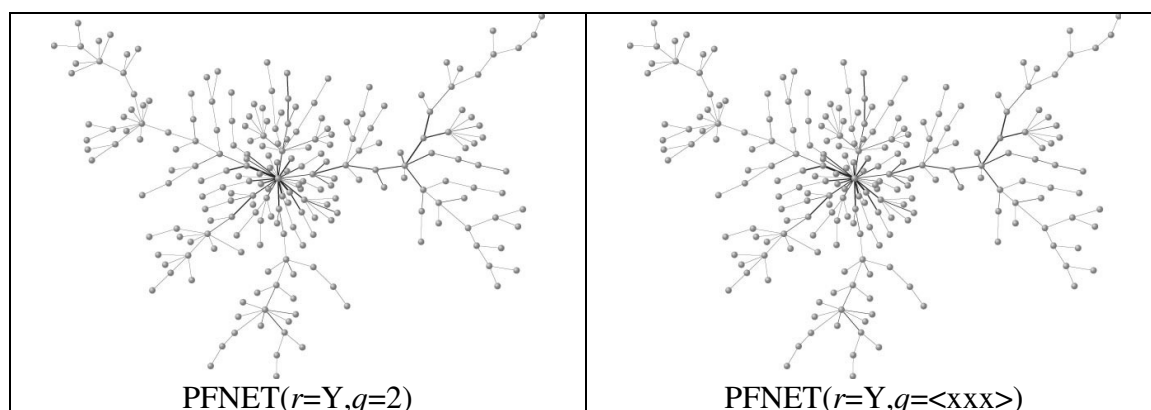
<Nombre cienciograma> ($n=<xxxx>$)	$r=1$ Número de enlaces / Densidad	$r=\infty$ Número de enlaces / Densidad
Red original	<xxx> / <xxx>	<xxx> / <xxx>
$q=2$	<xxx> / <xxx>	<xxx> / <xxx>
$q=n/8$	<xxx> / <xxx>	<xxx> / <xxx>
...	<xxx> / <xxx>	<xxx> / <xxx>
$q=7 \cdot n/8$	<xxx> / <xxx>	<xxx> / <xxx>
$q=n-1$	<xxx> / <xxx>	<xxx> / <xxx>

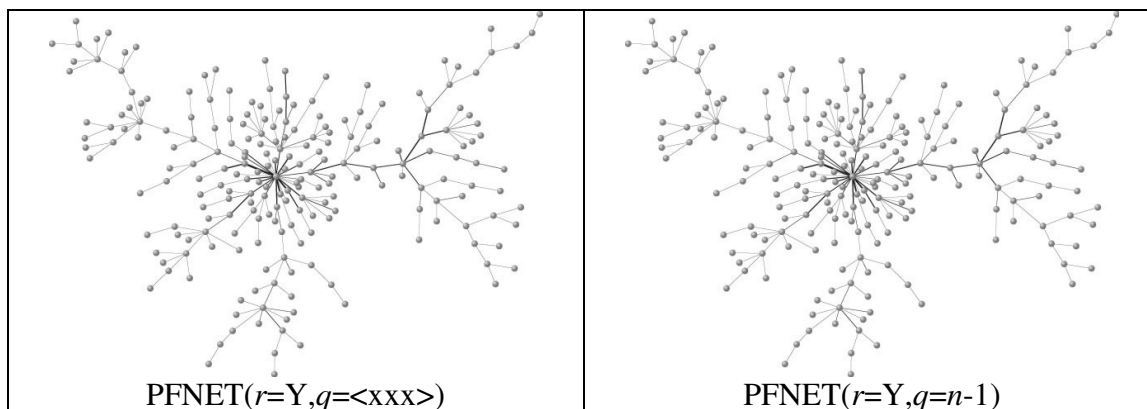
A continuación, el alumno realizará un análisis de los resultados obtenidos en las tablas, comentando el patrón de poda asociado a los valores de q en cada cienciograma y tratando de extraer conclusiones generales y específicas del tipo “¿qué correlación hay entre la intensidad de la poda y el valor de q ?”, “¿a partir de qué valor de q la PFNET(r,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 alumno escogerá **2 de los 10 cienciogramas** considerados y representará **8 de las PFNETs asociadas a cada una de ellas** (4 para cada valor de r) empleando dos algoritmos de distribución distintos, K-K y F-R. Esas 4 PFNETs serán la PFNET($r,q=2$), la PFNET($r,q=n-1$) y otras dos intermedias donde el alumno observe un cambio significativo en el número de enlaces.

De este modo, generará un total de 32 visualizaciones distintas (2 cienciogramas \times 2 valores de $r \times$ 4 valores de $q \times$ 2 métodos de distribución) y los presentará agrupados en bloques de 4 del siguiente modo:

Figura 2.XYZ: Visualizaciones de las PFNETs del cienciograma X con $r=Y$ realizadas con el método de distribución Z





Finalmente, el alumno realizará un estudio comparativo de la influencia de usar un valor de r u otro y , sobre todo, 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 (<http://aquirin.ovh.org/research/mstpathfinder.html>)— en redes aleatorias de 500 a 10000 nodos.

Para ello, el alumno 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 la plataforma de la asignatura³. Luego ejecutará las **5 variantes de Pathfinder con $r=\infty$ y $q=n-1$** sobre cada red 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

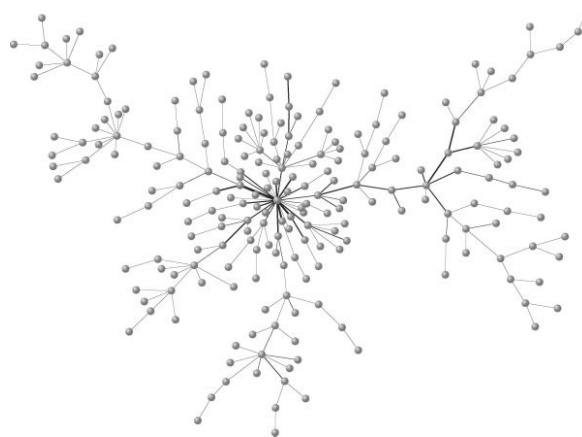
n	Media $ E $	Tiempo PF original	Tiempo Binary PF	Tiempo Fast PF	Tiempo MST-PF (baja empl.)	Tiempo MST-PF (práctico)
500	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>
1000	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>
2000	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>
5000	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>	<xxx>
10000	<xxx>	<xxx>	<xxx>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

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

El alumno 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 $O(n^3)$).

Por último, el alumno 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 alumno 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 alumno.

- 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 del algoritmo de distribución considerado en cada caso.
 - 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 alumno (Ej. Pérez Pérez Manuel.zip). Este fichero será entregado por internet a través del acceso identificado de la web del departamento de CCIA (<https://decsai.ugr.es>).