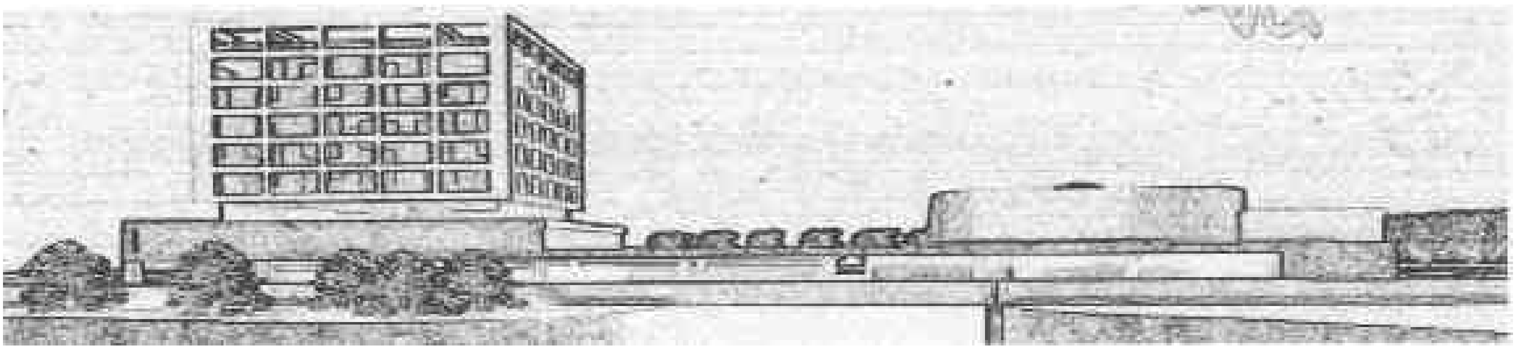
VOLUME VII, NUMBER 1



REPORTS ON SYSTEMS AND COMMUNICATIONS

Evaluación de la longitud media de caminos aleatorios en redes con ley de potencias

LUIS RODERO-MERINO
ANTONIO FERNÁNDEZ
LUIS LÓPEZ
VICENT CHOLVI

Móstoles (Madrid), Spain, Abril 2007
Depósito Legal: 50653-2004
ISSN: 1698-7489

Evaluación de la longitud media de caminos aleatorios en redes con ley de potencias

Luis Rodero-Merino[1], Antonio Fernández[1],
Luis López[1], Vicent Cholvi[2]
[1]Laboratorio de Algoritmia Distribuida y Redes, Universidad Rey Juan Carlos
Escuela Superior de Ciencias Experimentales y Tecnología
Campus de Móstoles (Madrid), C/ Tulipán S/N, 28933
E-mail: {lrodero,anto,llopez}@gsyc.es
[2]Departamento de Lenguajes y Sistemas Informáticos
Universitat Jaume I, Castellón, 12071
E-mail: vcholvi@lsi.uji.es

Abstract *In this paper we introduce a model to study random walks in power-law networks with one-hop replication. Basically, this model gives a set of expressions that captures how the* knowledge *about the network evolves as the random walk traverses the network: how many nodes have been known, either because they or their neighbors have been visited by the random walk. With this, we obtain an expression that gives a good estimation of the average number of hops needed to find some random peer from any other random peer. We denote this metric the average search length, and we deem it can be very useful to evaluate random walk based* resource location *solutions in P2P networks.*

1. Introducción

Las redes con *ley de potencias* (*power-law networks*) son aquellas en las que la distribución del número de vecinos de cada nodo (*i.e.*, su grado) cumple la siguiente propiedad:

$$p_k = k^{-\alpha} \tag{1}$$

donde p_k es la probabilidad de que el grado de un nodo cualquiera elegido al azar sea k. Este tipo de redes son especialmente interesantes porque se ha observado que varias redes del mundo real, como por ejemplo Internet [5] [6] o Gnutella [8] [9], son redes con ley de potencias.

Varios trabajos [1] [7] han intentado describir el comportamiento de las búsquedas en redes mediante *caminos aleatorios*[1], con especial atención a las redes con ley de potencias. Se dice que una búsqueda sigue un camino aleatorio si, en cada salto, el nodo que reenvía la búsqueda elige el destino de manera aleatoria, con probabilidad uniforme, entre sus vecinos. El interés en este mecanismo de búsqueda se debe a que, en escenarios reales, puede darse el caso de que se posea poca información acerca de dónde está localizado lo que se busca, y por lo tanto no se dispone de reglas o heurísticas que ayuden en la decisión de cómo encaminar la búsqueda.

Un ejemplo de escenario donde se ha propuesto la utilización de caminos aleatorios para realizar búsquedas son las redes entre iguales (*peer-to-peer*, P2P). En particular estas propuestas [13] [11] se basan en combinar caminos aleatorios con *topologías dinámicas*, y sus resultados parecen indicar que los caminos aleatorios pueden ser un mecanismo muy eficiente debido a que consumen muchos menos recursos que otras soluciones, como por ejemplo las búsquedas mediante *inundación*.

Este artículo intenta contribuir en el campo de la investigación de caminos aleatorios en redes con ley de potencias, proporcionando una serie de ecuaciones que modelan cómo evoluciona el *conocimiento* acerca de la red que se adquiere siguiendo un camino aleatorio. A partir de estas ecuaciones obtenemos otra expresión con la que estimar la *longitud media de las búsquedas*: el número de saltos que, como media, necesita una búsqueda desde un nodo cualquiera de la red para encontrar a otro nodo cualquiera de la red, siguiendo un camino aleatorio.

Nuestro estudio se centra en redes con una peculiaridad: cada nodo puede responder por cualquiera de sus vecinos. Es decir, basta con que el camino aleatorio visite un vecino del nodo a buscar para terminar la búsqueda. Esta característica añade dificultad al modelo, pero en muchos casos lo hace más cercano al mundo real. Por ejemplo, en una red social, para localizar a una persona basta con encontrar a cualquiera de sus amigos, que sabrán dónde se encuentra. Por esta razón esta misma propiedad también se supone en otros trabajos [1]. Esta particularidad hace al modelo especialmente interesante para aplicarlo al problema de las búsquedas en redes P2P, donde muchos

[1]También llamados *paseos aleatorios*.

trabajos [4] [10] suponen que cada miembro de la red conoce los recursos (*e.g.* un fichero, un servicio) de sus vecinos, con lo que es suficiente con visitar a un vecino del nodo que contiene el recurso buscado para saber dónde está el recurso.

Este artículo consta de las siguiente secciones. En la sección 2 hacemos un resumen de algunos trabajos de investigación acerca de caminos aleatorios, señalando las limitaciones que hemos encontrado en ellos. La sección 3 da algunas definiciones y supuestos que usamos como base de nuestra propuesta. La sección 4 presenta las métricas que forman el modelo, y explica y razona las ecuaciones con las que se calcula el valor de cada una. Después, la sección 5 muestra los resultados de algunos experimentos con los que hemos intentado validar el modelo. Finalmente, la sección 6 da las conclusiones y propone algunas líneas de trabajo futuro.

2. Estado del arte

El mayor problema a la hora de modelar caminos aleatorios es que no puede asumirse que funcionan como un proceso estocástico, al menos en muchas redes reales. Esto es especialmente cierto cuando la distribución del grado no es uniforme, sino que el grado de los nodos puede tomar valores muy diferentes, como por ejemplo en redes con ley de potencias.

En estas redes, unos pocos nodos tienen un grado mucho mayor que el resto, con lo que son visitados con más probabilidad por un camino aleatorio. Así, las búsquedas tienden a 'gravitar' alrededor de esos nodos de alto grado. Inicialmente, en redes en las que cada nodo responde por sus vecinos, esto es positivo. Al tener muchos vecinos, los nodos de mayor grado tienen un mayor conocimiento de la red y por lo tanto podrán responder con mayor probabilidad a las búsquedas que reciben. Sin embargo, según avanza el camino aleatorio, la probabilidad de revisitar una y otra vez los mismos nodos (que no pudieron responder anteriormente a la búsqueda) crece. Y por lo tanto, la probabilidad de éxito se reduce.

Además, hay otro factor a tener en cuenta. Aunque el nodo sea visitado por primera vez, puede que algunos de sus vecinos ya sean *conocidos* para el camino aleatorio (ellos o sus vecinos ya han sido visitados). Es decir, al llegar a un nodo de grado k en una red de N nodos, no se puede estimar la probabilidad de éxito como k/N. Dicho de otra manera, el problema de las búsquedas mediante caminos aleatorios no puede modelarse como el típico problema de *balls and bins*.

Adamic *et.al.* confirman esto en su trabajo [1]. Allí, proponen una serie de expresiones que intentan caracterizar caminos aleatorios en redes con ley de potencias, aplicando el formalismo de las funciones generatrices [14] ya introducido por Newman en [12] para el estudio de redes de grado arbitrario. Al intentar analizar la longitud del camino medio, Adamic *et.al.* encontraron una importante divergencia entre sus predicciones y los resultados obtenidos por simulación. Lo mismo ocurrió con sus análisis de la cobertura de la red (proporción de nodos conocidos por el camino aleatorio). Ellos mismos indicaron el motivo mediante otra serie de experimentos (cita traducida):

> "..En una red con ley de potencias, cuando el 50 % de la red ha sido visitada, el 80 % de los saltos son revisitas. Es decir, los nodos de grado alto son revisitados una y otra vez antes de que nodos de grado bajo sean visitados por primera vez"

En un trabajo previo, Newman *et.al.* [12] proponen otra serie de expresiones para calcular la longitud del camino medio *más corto* entre dos nodos cualesquiera de una red con distribución de grado arbitraria. Sin embargo, en su trabajo no incluyen simulaciones que avalen su modelo, que no tiene en cuenta los problemas descritos previamente.

Finalmente, Gkantsidis *et.al.* [7] afirman que un camino aleatorio puede simular un muestreo aleatorio uniforme de los nodos de una red. Tomando este trabajo como base, Bisnik *et.al.* [3] proponen una expresión para estimar la longitud del camino medio de búsqueda de un recursos en una red, en este caso sin que los nodos respondan por sus vecinos. Para ello, asumen que si un recurso tiene popularidad $p = n/N$ (donde n es el número de nodos que tienen el recurso y N el tamaño de la red), la probabilidad de no haber encontrado el recurso en T saldos es $(1-p)^T$. Es decir, se aproximan al problema asumiendo que es similar al de *balls and bins*. Sin embargo, en nuestra opinión, parten de una interpretación errónea del trabajo de Gkantsidis. Es cierto que sus resultados parecen apoyar su modelo, pero creemos que esto se debe a los pequeños valores de T usados para sus experimentos (1/100 del tamaño de la red).

3. Escenario y supuestos

Sea G un grafo con N nodos. La distribución del grado es dada por p_k, donde p_k es la probabilidad de que un nodo cualquiera de la red tenga grado k. n_k denota el número de nodos que tienen grado k ($\sum_k n_k = N$, $\forall k\ p_k = n_k/N$). El grado medio es dado por $\overline{k} = \sum_k kp_k$. Definimos *conexión* como cada uno de los extremos de un enlace. S es la suma de todas las conexiones de la red $S = \sum_k kn_k$.

Todos los nodos de la red tienen la misma probabilidad de ser el origen o destino de una búsqueda. Cada nodo puede responder por sus vecinos, por lo que una búsqueda habrá terminado con éxito cuando se haya encontrado el nodo destino o cualquiera de los vecinos de este.

La red no tiene bucles (no hay enlaces conectando a un nodo consigo mismo) ni multienlaces (si dos nodos están conectados, es por un único enlace).

Finalmente, se asume que en la red se cumplen estas dos condiciones:

- La *probabilidad de llegar a un nodo de grado k en un salto, $P_A(k)$ es proporcional* a k, y se calcula mediante la siguiente expresión [1]:

$$P_A(k) = \frac{kn_k}{S} = \frac{kp_kN}{\sum_j jp_jN} = \frac{kp_k}{\sum_j jp_j} = \frac{kp_k}{\overline{k}} \quad (2)$$

- Además, *todos los nodos de igual grado tienen la misma probabilidad de ser visitados por el camino aleatorio.* Es decir, si el camino aleatorio visita un nodo de grado k, entonces la probabilidad de que sea cada uno en particular de los n_k nodos con grado k es $1/n_k$. Por lo tanto, la posibilidad de visitar un nodo específico en un salto, si ese nodo tiene grado k, es $P_A(k)/n_k$.

En principio, estas suposiciones dependen de cómo se haya construído la red. Sin embargo, creemos que son lo suficientemente 'naturales' como para que se cumplan en la mayoría de los casos.

Como dato de partida para los cálculos, nuestro modelo asume que conocemos únicamente la distribución de grado $n_k \forall k$ de la red.

4. Modelo

Nuestro modelo se basa en una serie de métricas que estiman cómo evoluciona el conocimiento acerca de la red que se va acumulando a lo largo de un camino aleatorio. Cada métrica mide, respectivamente, el número de *nodos visitados*, el número de *conexiones comprobadas* y el número de *nodos conocidos* en cada salto del camino. Explicaremos cada métrica con más detalle en las siguientes secciones.

Además, estos valores serán usados para intentar estimar la longitud media de las búsquedas medida en número de saltos, $\overline{l}$. Una búsqueda consiste en localizar, partiendo desde un nodo origen y siguiendo un camino aleatorio, a otro nodo destino. Tanto el nodo origen como el nodo destino son elegidos al azar de manera uniforme entre los miembros de la red (dicho de otro modo, todos los nodos pueden ser origen o destino de una búsqueda con igual probabilidad). La búsqueda habrá tenido éxito cuando el camino aleatorio visite el nodo buscado o alguno de sus vecinos (en realidad, salvo en el caso de que el nodo origen y el destino sean el mismo, las búsquedas siempre terminarán con la visita a un vecino del destino).

4.1. Nodos visitados

Esta métrica mide el número medio de nodos *diferentes* de grado k visitados por el camino aleatorio tras efectuar l saltos. La denotamos V_k^l.

Recordemos que todos los nodos de la red tienen la misma probabilidad de ser el origen de la búsqueda, por lo tanto la probabilidad de comenzar la búsqueda en un nodo de grado k es p_k. Usamos $l = 0$ para expresar el momento inicial de la búsqueda. Así, estimamos V_k^0 como:

$$V_k^0 = 1 \cdot p_k + 0 \cdot (1 - p_k) = p_k \quad (3)$$

Para el primer salto, $l = 1$, tenemos que:

$$V_k^1 = V_k^0 + (1 \cdot P_A(k) + 0 \cdot (1 - P_A(k))) = V_k^0 + P_A(k) \quad (4)$$

A partir del primer salto, cuando $l > 1$, debemos tener en cuenta la probabilidad de que, al llegar a un nodo de grado k, dicho nodo ya haya sido visitado antes por el camino aleatorio. Para calcular esa probabilidad, necesitamos definir primero dos valores:

- $P_{\text{NoVisitado}}(k, l)$. Representa la probabilidad de que, si el camino aleatorio llega a un nodo de grado k en el salto l, dicho nodo no haya sido ya visitado. Recordemos que el número de nodos de grado k es n_k, y el número de nodos visitados en los saltos anteriores viene dado por V_k^{l-1}. Así, una primera aproximación a la probabilidad de estar revisitando un nodo, si el nodo al que se llega tiene grado k, sería V_k^{l-1}/n_k. Sin embargo, V_k^{l-2}/n_k se ajusta más a la realidad, ya que el nodo al que se llegó en el salto $l-1$ no puede ser revisitado en el salto l (no hay bucles en la red). De esta forma, finalmente, la probabilidad de que el nodo no haya sido visitado antes se calcula como:

$$P_{\text{NoVisitado}}(k, l) = \left(1 - \frac{V_k^{l-2}}{n_k}\right) \quad (5)$$

- P_{Retorno}. Es la probabilidad de que en el salto l el nodo se esté moviendo 'hacia atrás', de vuelta al nodo del que vino en el salto $l-1$. El nodo visitado en el salto $l-1$ tiene grado j con probabilidad $P_A(j)$, por lo tanto el camino aleatorio volverá por el mismo enlace por el que vino con probabilidad $1/j$. Así, definimos P_{Retorno} de esta forma:

$$P_{\text{Retorno}} = \sum_j P_A(j)\frac{1}{j} = \frac{1}{\overline{k}} \quad (6)$$

Usando esas probabilidades, definimos V_k^l para $l > 2$ como:

$$V_k^l = V_k^{l-1} + P_A(k) P_{\text{NoVisitado}} (1 - P_{\text{Retorno}}) \quad (7)$$

Desarrollando esa expresión nos queda que:

$$V_k^l \;=\; V_k^{l-1} \;+\; \frac{kp_k}{\overline{k}}\left(1-\frac{1}{\overline{k}}\right)\left(1-\frac{V_k^{l-2}}{n_k}\right) \qquad (8)$$

4.2. Conexiones comprobadas

Esta métrica representa el número medio de conexiones *diferentes* que han sido comprobadas hasta el salto l, inclusive. Denominamos conexión comprobada al extremo que está al otro lado de cada uno de los enlaces de un nodo visitado. Denotamos a esta métrica L^l.

En general, si el camino aleatorio llega hasta un nodo de grado k, y el nodo no ha sido visitado antes, entonces se están comprobando k conexiones. Así, L^l es sencillo de calcular a partir de V_k^l:

$$L^l \;=\; \sum_k kV_k^l \qquad (9)$$

4.3. Nodos conocidos

Esta métrica estima el número de nodos de grado k diferentes que han sido conocidos durante el camino aleatorio hasta el salto l, inclusive. La denotamos C_k^l.

Para el momento inicial de la búsqueda, $l = 0$, tenemos que:

$$C_k^0 \;=\; p_k \;+\; \sum_j p_j j P_A(k) \qquad (10)$$

En esta fórmula, el primer término p_k representa el nodo visitado al inicio, y el segundo sumando representa el número de conexiones del nodo inicial que apuntan a nodos de grado k, promediado sobre cada posible grado que puede tener dicho nodo inicial. Si tiene grado j, cosa que sucederá con probabilidad p_j, entonces, en promedio, $jP_A(k)$ enlaces suyos apuntarán a nodos de grado k, que pasarán a ser conocidos.

Para cada salto l, $V_j^l - V_j^{l-1}$ representa el número medio de nodos de grado j que son visitados por primera vez en ese salto. Para calcular C_k^l sólo nos interesan esos nodos, ya que revisitar un nodo no incrementa el número de pares conocidos.

Ahora, nos fijamos en el número de conexiones que serán comprobadas por primera vez en el salto l. Puede considerarse que representa el número de *intentos* en el salto l. Llamamos a esta métrica In^l. Una primera aproximación sería calcular In^l como:

$$\mathrm{In}^l \;=\; L^l - L^{l-1} \;=\; \sum_k (V_k^l - V_k^{l-1})\, k \qquad (11)$$

Sin embargo, para una mayor precisión, debemos darnos cuenta de que uno de los k enlaces del nodo al que se llega apuntarán al nodo del que viene el camino aleatorio. Por lo tanto, ese enlace no puede llevar al nodo buscado ni incrementar el número de nodos conocidos. Así, es más exacto expresar In^l como sigue:

$$\mathrm{In}^l \;=\; \sum_k (V_k^l - V_k^{l-1})\,(k-1) \qquad (12)$$

Finalmente, necesitamos calcular la probabilidad de que cada enlace del nodo visitado en l apunte a un nodo de grado k que no ha sido conocido antes (asumiendo que el nodo visitado lo es por primera vez, si no esa probabilidad sería 0). Estimamos esa probabilidad de la siguiente forma:

$$\frac{k(n_k - C_k^{l-1})}{S - L^{l-1}} \qquad (13)$$

La justificación de esta expresión es la siguiente. La suma de conexiones de nodos de grado k no conocidos viene dada por $k(n_k - C_k^{l-1})$. Por otro lado, el número de posibles conexiones a las que cada enlace del nodo visitado puede apuntar es $S - L^{l-1}$. S es el número total de conexiones, de la que restamos aquellas conexiones ya comprobadas L^{l-1}. La razón de esta substracción es que el nodo visitado en el salto l no ha sido visitado antes, y por lo tanto ninguno de sus enlaces pueden apuntar a una conexión comprobada previamente. Dividiendo ambas cantidades, tenemos una aproximación a la probabilidad que buscábamos: la probabilidad de que una de las conexiónes del nodo visitado en el salto l apunte a un nodo desconocido de grado k.

Con esta probabilidad, podemos calcular C_K^l como sigue:

$$C_k^l \;=\; C_k^{l-1} \;+\; \left(\frac{k(n_k - C_k^{l-1})}{S - L^{l-1}}\right)\mathrm{In}^l \qquad (14)$$

4.4. Longitud media de las búsquedas

Finalmente, usando las magnitudes anteriores, vamos a dar una expresión que estime la longitud media de una búsqueda en la red, $\overline{l}$. Recordemos que una búsqueda consiste en localizar, partiendo desde un nodo origen y siguiendo un camino aleatorio, a otro nodo destino, y que la búsqueda habrá tenido éxito cuando se haya visitado el nodo buscado o alguno de sus vecinos.

4.4.1. Probabilidad de éxito

La probabilidad de encontrar el nodo buscado en cada salto viene dada por los valores C_k^l. Sea C^l el número total de nodos conocidos en el salto l:

$$C^l \;=\; \sum_k C_k^l \qquad (15)$$

Entonces, la probabilidad de éxito en el salto l, P_{Exito}^l está definida por:

$$P_{\mathrm{Exito}}^l \;=\; \frac{C^l - C^{l-1}}{N - C^{l-1}} \qquad (16)$$

Donde $N - C^{l-1}$ es el número de nodos que son aún desconocidos en el salto l, que a su vez son los únicos posibles candidatos de ser el nodo destino, y $C^l - C^{l-1}$ representa el número de nodos que serán conocidos en el salto l, que puede interpretarse como el número de *intentos* en ese salto. Así, la probabilidad de éxito viene dada por la cantidad de intentos dividida por el número de candidatos.

4.4.2. Longitud media

Finalmente, $\bar{l}$ es dado por la siguiente expresión:

$$\bar{l} = \sum_{l}^{\infty} l P^l_{\text{Fin}} \tag{17}$$

donde P^l_{Fin} es la probabilidad de que la búsqueda sea finalizada en el salto l. Esto es, la probabilidad de que la búsqueda tenga éxito en el salto l y no haya tenido éxito en los $l-1$ saltos anteriores.

$$P^l_{\text{Fin}} = P^l_{\text{Exito}} \prod_{i=0}^{l-1} (1 - P^i_{\text{Exito}}) \tag{18}$$

Usando Eq. 16 podemos reescribir P^l_{Fin} como:

$$P^l_{\text{Fin}} = \frac{C^l - C^{l-1}}{N} \tag{19}$$

y por lo tanto $\bar{l}$ pasaría a escribirse de la siguiente forma:

$$\bar{l} = \frac{1}{N} \sum_{l}^{\infty} l(C^l - C^{l-1}) \tag{20}$$

5. Resultados experimentales

Para probar la bondad de nuestro modelo en redes con ley de potencias, hemos ejecutado una serie de experimentos, cada uno centrado en una métrica en particular.

Para cada uno de estos experimentos se construyó una red diferente. Todas las redes tenían un tamaño de 10^5 nodos. Todas, a su vez, fueron construídas de la misma forma. Primero, se generó una red siguiendo el mecanismo propuesto por Barabasi [2]. Este algoritmo parte de un pequeño número inicial de nodos $n_0, n_0 << N$. Después, se va insertando en cada paso un nuevo nodo que establecerá conexiones con los nodos ya presentes de forma *preferencial*: la probabilidad de formar el enlace con un nodo i, pe_i, es proporcional a su grado k_i ($\text{pe}_i = k_i / \sum_j k_j$). Este mecanismo asegura que los grados de los nodos de la red obtenida siguen una distribución de ley de potencias. Para nuestras redes, se partía de un número inicial de 5 nodos y se creaban 5 nuevos enlaces por cada nodo añadido, de tal forma que la red resultante tenía un grado medio de $\bar{k} = 10$ vecinos por nodo.

Después, de la red obtenida se extraía la distribución de grado, y se construía una nueva red mediante el algoritmo descrito por Newman [12]. Este algoritmo es sencillo. Inicialmente, ninguna conexión de ningún nodo está conectada. Después, de forma iterativa, se van eligiendo al azar y de manera uniforme pares de conexiones aún sin conectar y se enlazan entre sí. En nuestro caso, sólo hemos introducido la variante de que los bucles y multienlaces son evitados. La razón de rehacer la red usando el mecanismo de Newman es que este puede generar, con idéntica probabilidad, todas las redes posibles con la distribución de grado dada, algo que Barabasi no asegura.

5.1. Experimento sobre el número de nodos visitados

Aquí mostramos los resultados del experimento que trata de medir la precisión de la métrica V^l_k, es decir, del número de nodos visitados en el salto l. Para ello, nos fijamos en dos valores en particular: el número *total* de nodos visitados, $\sum_k V^l_k$, y el número de nodos visitados de grado 10, V^l_{10}, en cada salto l. Para obtener los resultados reales, se ejecutaron 10^4 caminos aleatorios sobre la red, con una longitud de 10^5 saltos cada uno. Después se calculó para cada salto l, el valor medio de los distintos V^l_k obtenidos en cada camino aleatorio.

Los resultados, hasta el salto 10^5, se muestran en la Fig. 1 (por claridad, se muestran los resultados cada 2000 saltos). La gráfica muestra que hay un buen ajuste entre los valores predichos por el modelo y los obtenidos por la simulación.

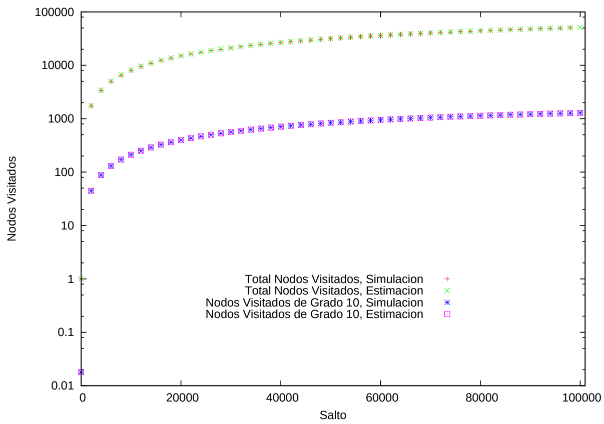


Figura 1: Nodos visitados, escala log.

5.2. Experimento sobre el número de conexiones comprobadas

En este apartado nos centramos en la métrica L^l, número de conexiones distintas comprobadas en el salto l. Igual que antes, se ejecutaron 10^4 caminos aleatorios de 10^5 saltos de longitud para obtener los resultados experimentales.

Los resultados pueden verse en la Fig. 2. Como en el caso anterior, la estimación teórica da valores similares a los obtenidos por simulación.

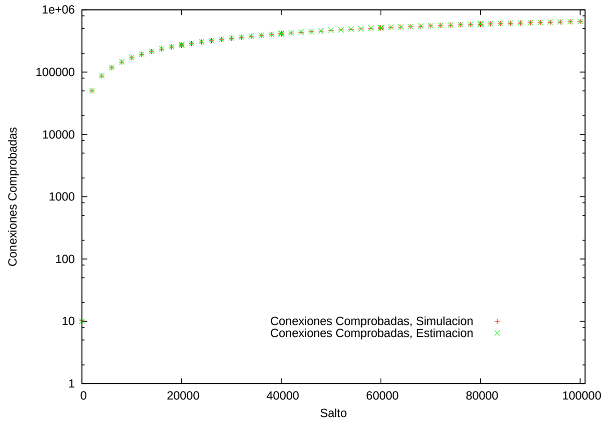


Figura 2: Con. comprobadas, escala log.

5.3. Experimento sobre el número de nodos conocidos

Aquí, nuestros experimentos intentan medir la precisión de la estimación del número de nodos conocidos por el camino aleatorio, el grado de cobertura de la red. Para ello, estudiamos el número total de nodos desconocidos, $N - \sum_k C_k^l$, y el número de nodos desconocidos de grado 10, $n_{10} - C_{10}^l$, para cada salto l. De nuevo, los resultados experimentales se calculan como la media de los datos obtenidos de 10^4 caminos aleatorios de 10^5 saltos de longitud.

La gráfica en la Fig. 3 muestra nuestros resultados. Una vez más, hay un buen ajuste entre los resultados predichos por el modelo y los obtenidos experimentalmente.

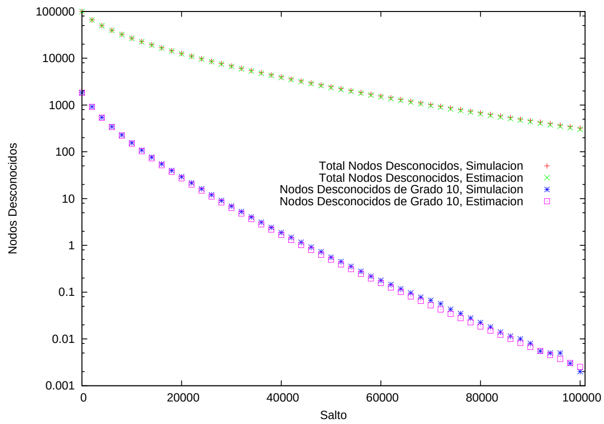


Figura 3: Nodos conocidos, escala log.

A partir de los resultados se observa que el conocimiento acerca de los nodos en la red aumenta rápidamente al inicio del camino aleatorio. Por ejemplo, alrededor del salto 10^4 ya se conoce el 70 % de la red. Esto se debe a que los caminos aleatorios visitan los nodos de grado alto con mayor probabilidad. Estos nodos, por su alto número de vecinos, incrementan mucho el conocimiento acerca de la red de los caminos que los visitan por primera vez. Sin embargo, también se ve en la gráfica como según se realizan más saltos siguiendo el camino aleatorio, cada vez se conocen menos nodos nuevos (cada salto aporta menos conocimiento). La razón son las revisitas: la tendencia a visitar una y otra vez los mismos nodos de grado alto (llegar a un nodo ya visitado, claro está, no aumenta el conocimiento acerca de la red). Estos resultados están de acuerdo con lo observado por Adamic [1].

5.4. Longitud del camino medio

Finalmente, mostraremos los resultados de la estimación de la longitud del camino medio $\bar{l}$ de las búsquedas. En este caso, se muestran los resultados para redes de distintos tamaños: 10^4, $2{,}5 \cdot 10^4$, $5 \cdot 10^4$, 10^5, $2 \cdot 10^5$, $5 \cdot 10^5$ y 10^6 nodos. Todas las redes fueron construídas siguiendo el mecanismo descrito al inicio de esta sección. Sobre cada red se ejecutaron 10^4 búsquedas. Cada búsqueda comenzaba en un nodo elegido al azar de manera uniforme entre todos los miembros de la red, y tenía como destino otro nodo elegido de la misma forma. Cuando el nodo destino era conocido por el camino aleatorio (un vecino era visitado), la búsqueda se daba por finalizada y se anotaba el número total de saltos realizados.

En la Fig. 4 mostramos los resultados de los experimentos junto con las estimaciones obtenidas usando Eq. 20.

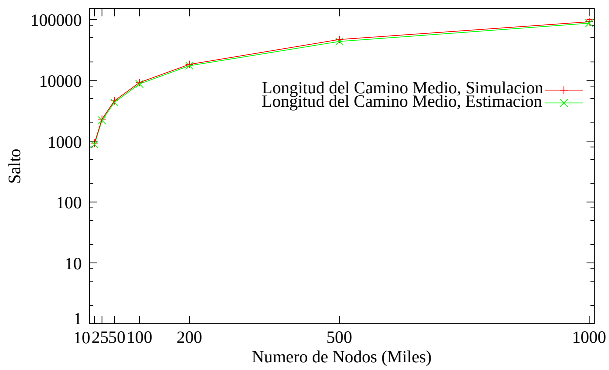


Figura 4: Longitud de las búsquedas, escala log.

La Fig. 4 nos indica que el error se mantiene en una proporción constante al tamaño de la red. En concreto, el error es en todos los casos de aproximadamente el 5 % del tamaño de la red, lo que creemos es una buena estimación.

6. Conclusiones y trabajo futuro

En este artículo hemos introducido nuestra propuesta de modelo para caracterizar caminos aleatorios en redes con ley de potencias. Este modelo consiste de varias métricas que nos indican cómo evolucionará el conocimiento acerca de la red que va acumulando un camino aleatorio. A partir de esas métricas, hallamos una expresión que estima la longitud del camino medio de las búsquedas por caminos aleatorios en este tipo de redes. El único dato que necesita el modelo es la distribución del grado de los nodos de la red.

Asimismo, para validar el modelo, hemos presentado también los resultados de una serie de experimentos. Estos resultados parecen confirmar que los valores estimados a partir del modelo son muy cercanos a la realidad.

Como trabajo futuro, las líneas de investigación que pueden seguirse a partir del modelo son varias. Por ejemplo, podría estudiarse su aplicabilidad a otros tipos de redes distintas a las redes con ley de potencias. También puede servir de base para estudiar la eficiencia de diversas topologías de red para resolver búsquedas.

Referencias

[1] Lada A. Adamic, Bernardo A. Huberman, Rajan M. Lukose, and Amit R. Puniyani. Search in power law networks. *Physical Review E*, 64:46135–46143, October 2001.

[2] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286:509,512, 1999.

[3] Nabhendra Bisnik and Alhussein Abouzeid. Modeling and analysis of random walk search algorithms in p2p networks. In *Proceedings of the Second International Workshop on Hot Topics in Peer-to-Peer Systems (Hot-P2P 2005)*, pages 95–103. IEEE Computer Society, July 2005.

[4] Yatin Chawathe, Sylvia Ratnasamy, Nick Lanham, and Scott Shenker. Making Gnutella-like P2P systems scalable. In *Proceedings of the 2003 conference on applications, technologies, architectures, and protocols for computer communications (SIGCOMM 2003)*, pages 407–418, Karlsruhe, Germany, August 2003.

[5] Stephen Dill, S. Ravi Kumar, Kevin S. McCurley, Sridhar Rajagopalan, D. Sivakumar, and Andrew Tomkins. Self-similarity in the web. In *Proceedings of the 27th International Conference on Very Large Data Bases*, pages 69–78. ACM Press, 2001.

[6] Michalis Faloutsos, Petros Faloutsos, and Christos Faloutsos. On power-law relationships of the internet topology. In *Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication*, pages 251–262. ACM Press, 1999.

[7] Christos Gkantsidis, Milena Mihail, and Amin Saberi. Random walks in peer-to-peer networks. In *Proceedings of the Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies, INFOCOM 2004*, volume 1, pages 120–130, Hong Kong, March 2004.

[8] Mihajlo A. Jovanovic, Fred S. Annexstein, and Kenneth A. Berman. Modeling peer-to-peer network topologies through 'small-world' models and power laws. In *Proceedings of the IX. Telecommunications Forum (TELFOR 2001)*, 2001.

[9] Mihajlo A. Jovanovic, Fred S. Annexstein, and Kenneth A. Berman. Scalability issues in large peer-to-peer networks - a case study of gnutella, 2001.

[10] Qin Lv, Pei Cao, Edith Cohen, Kai Li, and Scott Shenker. Search and replication in unstructured peer-to-peer networks. In *Proceedings of the 16th international conference on Supercomputing*, pages 84–95, New York, New York, United States, June 2005.

[11] Qin Lv, Sylvia Ratnasamy, and Scott Shenker. Can heterogeneity make Gnutella scalable? In *Revised Papers from the First International Workshop on Peer-to-Peer Systems*, pages 94–103, Cambridge, United States, March 2002.

[12] Mark E. J. Newman, Steven H. Strogatz, and Duncan J. Watts. Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64(2):026118–1,026118–17, Jul 2001.

[13] Sylvia Ratnasamy, Paul Francis, Mark Handley, Richard Karp, and Scott Schenker. A scalable content-addressable network. In *Proceedings of the 2001 conference on applications, technologies, architectures, and protocols for computer communications (SIGCOMM 2001)*, pages 161–1672, San Diego, California, United States, 2001.

[14] Herbert S. Wilf. *Generatingfunctionology*. Academic Press, 1994.