



**UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO**

FACULTAD DE CIENCIAS

**MODELOS DE CAMINATAS ALEATORIAS
EN REDES COMPLEJAS**

T E S I S

QUE PARA OBTENER EL TÍTULO DE:

FÍSICO

P R E S E N T A:

RODRIGO LEAL CERVANTES



**DIRECTOR DE TESIS:
DR DAVID PHILIP SANDERS**
Ciudad de México 2016



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Hoja de Datos del Jurado

1. Datos del alumno

Apellido paterno
Apellido materno
Nombre(s)
Teléfono
Universidad Nacional Autónoma de México
Facultad de Ciencias
Carrera
Número de cuenta

1. Datos del alumno

Leal
Cervantes
Rodrigo
56 61 47 01
Universidad Nacional Autónoma de México
Facultad de Ciencias
Física
411002099

2. Datos del tutor

Grado
Nombre(s)
Apellido paterno

2. Datos del tutor

Dr
David Philip
Sanders

3. Datos del sinodal 1

Grado
Nombre(s)
Apellido paterno
Apellido materno

3. Datos del sinodal 1

Dr
Francisco Javier
Sevilla
Pérez

4. Datos del sinodal 2

Grado
Nombre(s)
Apellido paterno
Apellido materno

4. Datos del sinodal 2

Dra
Natalia Bárbara
Mantilla
Beniers

5. Datos del sinodal 3

Grado
Nombre(s)
Apellido paterno
Apellido materno

5. Datos del sinodal 3

Dr
Isaac
Pérez
Castillo

6. Datos del sinodal 4

Grado
Nombre(s)
Apellido paterno
Apellido materno

6. Datos del sinodal 4

Dr
Enrique
Hernández
Lemus

7. Datos del trabajo escrito

Título
Número de páginas
Año

7. Datos del trabajo escrito
Modelos de caminatas aletorias en redes complejas
100
2016

Índice general

1. Introducción	4
1.1. Descripción del problema	6
1.1.1. Movimiento browniano	6
1.1.2. Generalización de las caminatas aleatorias	7
1.1.3. Tiempos de primer paso y de primer encuentro	7
1.1.4. Cómo tratar el problema	9
1.2. Plan de la tesis	10
I Antecedentes	12
2. Caminatas y redes	13
2.1. Caminatas aleatorias en el espacio real	13
2.1.1. Espacio discreto	13
2.1.2. Espacio continuo	14
2.1.3. Tiempo	14
2.1.4. El teorema del límite central	15
2.2. Redes complejas	16
2.2.1. La red de mundo pequeño	18
2.2.2. Producto cartesiano de dos redes	21
2.3. Caminatas en redes	22
2.3.1. Formalismo de las caminatas aleatorias	22
2.3.2. La ecuación maestra	24
2.4. Tiempos de primer paso y de primer encuentro	24
2.4.1. Condiciones de terminación y promedios de las dis- tribuciones	24
2.4.2. Reformulación del problema de primer encuentro como un problema de primer paso	25

2.4.3.	Ecuación de recurrencia para el MFPT	28
2.4.4.	Ecuación de recurrencia para el MFET	31
2.5.	Promedios	31
2.5.1.	Tiempos promedio generales	32
2.5.2.	Promedios sobre configuraciones	32
3.	Anillos y métodos numéricos	34
3.1.	Resultados exactos para anillos	34
3.1.1.	El anillo simple	34
3.1.2.	El anillo general con $k > 1$	39
3.2.	Métodos numéricos	40
3.2.1.	La simulación Monte Carlo	40
3.2.2.	Enumeración exacta	43
3.2.3.	Método matricial	48
3.2.4.	Ventajas y desventajas de los métodos	52
II	Validación y resultados	54
4.	Validación de los métodos numéricos	55
4.1.	Distribución de los tiempos de primer paso y de primer en- cuentro	55
4.2.	Tiempos promedio	56
4.3.	El anillo simple	59
4.3.1.	Conclusión de la validación	62
5.	Resultados en <i>una</i> red de mundo pequeño	64
5.1.	Anillos	64
5.2.	Dos tipos de MFPT	67
5.3.	Arcos de anillo entre dos atajos	70
5.3.1.	Ajuste de los tiempos	70
5.3.2.	Resultados semi-analíticos para los arcos de anillo .	70
5.4.	Comparación entre el MFPT al origen y el MFET	73
6.	Tiempos promedio generales y promedios sobre configuraciones	76
6.1.	GFPT y GFET	76
6.2.	Promedios sobre configuraciones	81

7. Relación entre la distancia y los MFPT y MFET	83
7.1. Correlación del MFPT al origen y MFET con la distancia geodésica	83
7.2. La distancia dinámica	85
7.2.1. Caminata aleatoria en el espacio de distancia	86
7.3. Visualización del espacio de distancia	87
7.3.1. El espacio de distancia en una dimensión	87
7.3.2. La distancia en la red $G \times G$	89
7.4. Resumen	91
8. Conclusiones	93

Capítulo 1

Introducción

Un área de gran interés en las matemáticas discretas es el estudio de las redes complejas. Esto se debe en parte porque numerosos sistemas se presentan en dicha forma [15]: el internet y las redes sociales, la redes de colaboración científica y numerosos sistemas de la biología como las redes neuronales, metabólicas o de interacciones entre proteínas, son unos cuantos ejemplos.

El papel de las redes se ha estudiado desde la década de 1930 en las ciencias sociales, sobre todo desde el punto de vista de la centralidad y conectividad de los nodos de redes conformadas por decenas, o en el mejor de los casos, centenas de nodos [15]. Sin embargo, desde finales del siglo XX el interés se ha trasladado hacia la investigación de las propiedades estadísticas de redes mucho más grandes. Esto ha sido posible por el uso de la computadora, mismo que ha permitido recopilar información de redes que tienen millones, o incluso miles de millones de nodos, con igual número de conexiones.

De acuerdo con Newman [15], un factor importante que a menudo se menosprecia para explicar este cambio de enfoque es que, mientras que uno puede dibujar con puntos y líneas una red de decenas de nodos y responder preguntas sólo de la apariencia que tiene, lo mismo no se puede hacer para una red conformada por millones de nodos. El desarrollo reciente de herramientas estadísticas es un intento para encontrar algo que juegue el papel de la vista en los análisis que se hacían en el siglo pasado. Los métodos estadísticos son una respuesta a la pregunta: "¿Cómo puedo decir a qué se parece esta red, si no la puedo ver realmente?".

Newman divide los estudios sobre redes complejas en tres áreas de in-

investigación relativamente desconexas [15]. En primer lugar, es importante encontrar las propiedades estadísticas de las redes, distribuciones de longitudes de caminos y de grados, así como las herramientas apropiadas para medir estas cantidades. En segundo lugar, se desarrollan modelos de redes que pretenden capturar un aspecto u otro de las redes del mundo real, y que nos ayudan a entender el significado de las propiedades del primer campo de estudio. Finalmente, se pretende predecir el comportamiento de procesos dinámicos que se desarrollen sobre las redes, reales o que provienen de un modelo, a partir de las propiedades estructurales de éstas. Por ejemplo, los investigadores se preguntan cómo afecta la estructura de la red al tráfico en Internet, o cuáles son las propiedades dinámicas de sistemas sociales o biológicos estructurados como una red. Lahtinen et al. [10] explica que se espera que la topología de la red debería de tener un impacto mayor en prácticamente cualquier fenómeno que ocurra en ella.

Por otro lado, en un campo que parecería ajeno a las redes complejas, los modelos de caminatas aleatorias sirven para describir diferentes procesos en campos del conocimiento tan diversos como la probabilidad y las finanzas, la física, la química y la biología. En mecánica estadística, por ejemplo, se utilizan para modelar un movimiento browniano o para hacer modelos de polímeros [22].

De las múltiples propiedades de una caminata aleatoria que pueden ser de interés, se estudia frecuentemente al tiempo que requiere una caminata para llegar *por primera vez* a un sitio de interés. Varios problemas en biología se formulan en términos de dicho tiempo, por ejemplo: la tasa de transmisión de enfermedades entre animales territoriales [7], problemas de transporte a través de una membrana celular [20], o modelos del disparo de una neurona cuando el potencial, que realiza una caminata aleatoria sesgada, llega por primera vez a cierto umbral [20]. Por otro lado, uno puede calcular el tiempo de ruina de un jugador en diferentes juegos probabilísticos al encontrar dicho tiempo [13]. Esta clasificación importante de problemas se conoce como problemas de primer paso.

Las caminatas aleatorias sirven precisamente como herramientas que permiten estudiar la estructura de las redes complejas. En este trabajo se estudian caminatas aleatorias que se desarrollan sobre un tipo particular de red compleja, y se pretende hacer una aportación a la tercer área de investigación que se presentó antes. La idea no es nueva [8, 9, 10, 21, 17, 18, 5]; se han estudiado cantidades como el tiempo de cobertura, el tiempo de

primer paso o el tiempo de primer retorno, que se relacionan con las propiedades de recepción y transmisión de información [17] o con la relación de los tiempos de primer paso con la distribución de caminos geodésicos [18, 21]. En todos estos casos, se estudian caminantes que tienen que llegar *por primera vez* a un sitio meta.

El problema de los tiempos de encuentro de dos caminantes en una red es un problema poco estudiado, a pesar de que se formula naturalmente en la red que resulta del producto cartesiano de la red original consigo misma [1], y por lo tanto sirve para estudiar las propiedades de este producto. Además, es también un problema de interés comparar los tiempos de encuentro con los de primer paso: unos y otros muestran relaciones interesantes [14].

Motivados por esta situación, nos dedicamos a estudiar los tiempos de primer encuentro de dos caminantes en una red compleja. Para conseguirlo, también tenemos que estudiar los tiempos de primer paso, y cómo se comparan unos con otros. A continuación introducimos las ideas básicas del problema.

1.1. Descripción del problema

Una caminata aleatoria es un proceso donde una partícula se mueve en cierto espacio, brincando entre los sitios disponibles de forma aleatoria a lo largo del tiempo. La posición de la partícula cambia discontinuamente; no nos interesa cómo llega de una posición a otra, pero sí nos importa qué posiciones ocupa antes y después del brinco. Una caminata aleatoria es un modelo matemático, y como tal procura representar sólo las características más importantes del sistema real que pretende modelar.

1.1.1. Movimiento browniano

El ejemplo clásico donde se utilizan caminatas aleatorias, y que da una buena intuición del modelo, proviene de la física estadística. Las caminatas son muy útiles para modelar el movimiento browniano, la trayectoria de una partícula inmersa en un gas y que sufre colisiones con las partículas que la rodean.

La partícula que seguimos se mueve en línea recta con velocidad constante, hasta que se encuentre con otra partícula: la colisión le imprime

un momento y la hace cambiar de dirección y de velocidad. Además, la distancia que recorre entre dos colisiones es variable (aunque podemos hablar de un camino libre medio que tiene que ver con la densidad de partículas). La trayectoria de la partícula se ve por lo tanto como una sucesión de segmentos rectos, de longitud variable, con orientaciones relativas que cambian aleatoriamente. Esta trayectoria se puede reproducir más o menos fielmente según los detalles que se decidan incorporar con una caminata aleatoria en el espacio real. En la figura ?? se representa una caminata aleatoria en dos dimensiones.

Aquí iría como ejemplo una figura del movimiento de un gas de esferas rígidas, donde una de ellas tiene un marcador y deja impresa la trayectoria que sigue.

1.1.2. Generalización de las caminatas aleatorias

Ciertos problemas de probabilidad se pueden formular como una caminata aleatoria en un espacio discreto [13, 20]. Siguiendo este camino, una clase de espacios discretos que se pueden estudiar por medio de caminatas aleatorias son las redes complejas. En una red, los sitios disponibles para el caminante son los nodos. En el modelo más simple, una vez que el caminante ocupa cierto nodo puede brincar a los vecinos de éste siguiendo las conexiones de la red.. Dos ejemplos se encuentran en la figura 1.1.

El tipo de red que nos interesa en este trabajo es la red de mundo pequeño, figura (b) en la figura 1.1. En este tipo de red cada nodo está conectado con cierto número de nodos cercanos, y además hay un número de “atajos” que conectan nodos alejados. La red de mundo pequeño es un modelo que pretende incorporar la propiedad de mundo pequeño, mucho mejor conocida como el fenómeno de los “seis grados de separación” [22] (la creencia de que uno puede relacionarse con una persona que parecería muy alejada, utilizando como conexión a muy pocos individuos, por lo general no más de seis).

1.1.3. Tiempos de primer paso y de primer encuentro

La trayectoria de *una* caminata aleatoria es, valga la redundancia, aleatoria. Sin embargo, se puede decir cuál es la probabilidad de que siga cierta trayectoria y también se puede decir la estadística que tiene que seguir un conjunto grande de caminatas. En este trabajo nos interesan el tiempo

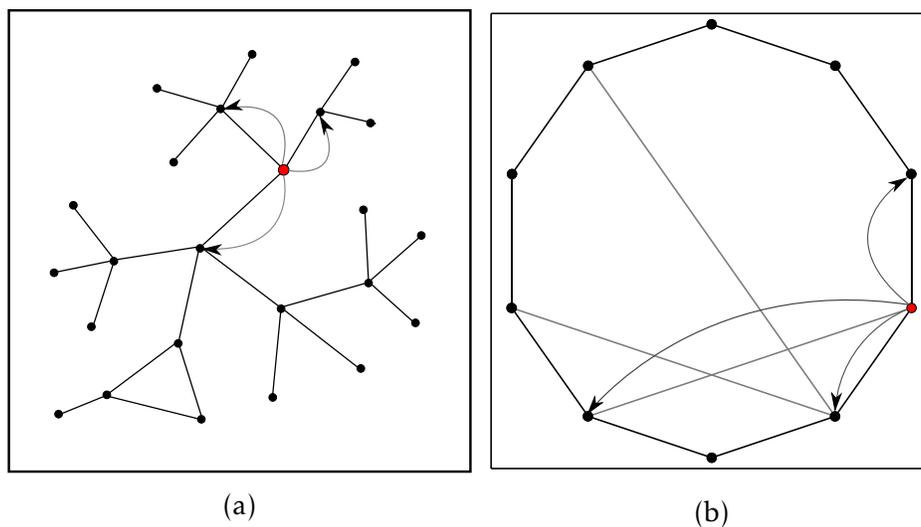


Figura 1.1: **Caminatas aleatorias en redes:** Las caminatas aleatorias no se restringen al espacio real, se pueden formular en espacios discreto como en la redes que mostramos aquí. (a) El caminante (punto rojo) puede brincar al cualquiera de sus tres vecinos en esta gráfica matemática. (b) En esta red de mundo pequeño, el caminante también tiene tres nodos vecinos a los que puede brincar.

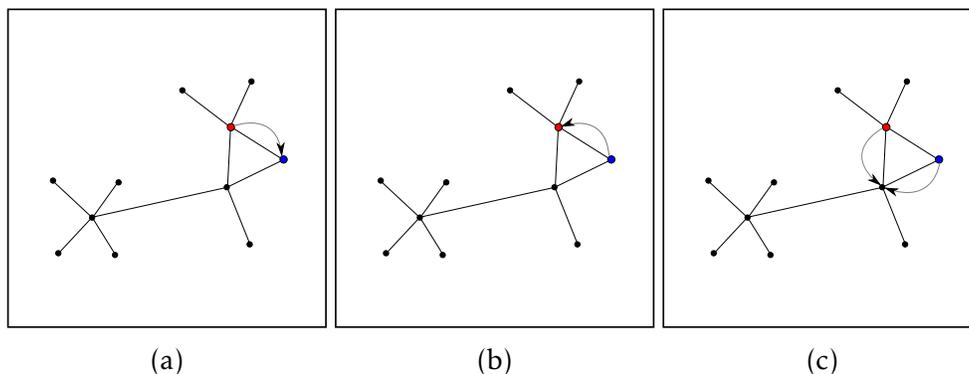


Figura 1.2: **Encuentros de dos caminantes:** Tres posibles maneras en las que dos caminantes (puntos azul y rojo) se pueden encontrar en una red.

que tarda un caminante en llegar a cierto sitio y el que le toma a dos caminantes encontrarse. (Las diferentes maneras en que dos caminantes se pueden encontrar se explican en la fig. 1.2.) Ambos casos son ejemplos de una clasificación importante de problemas que se llaman de primer paso. Para *un* caminante en el primer caso, o para *una* pareja de caminantes en el segundo, no podemos saber cuál es el tiempo que vamos a encontrar. No obstante, podemos calcular cuál es el tiempo que esperamos encontrar, el tiempo que en promedio deberíamos obtener.

1.1.4. Cómo tratar el problema

Nos acercamos a la cuestión de tres maneras distintas. En primer lugar, podemos utilizar la computadora para simular directamente una caminata aleatoria. Hacemos un gran número de caminatas (“gran número” significa cientos de miles o millones de repeticiones), y tomamos el promedio de los tiempos que encontramos. Este enfoque se conoce como simulación Monte Carlo. En segundo lugar, podemos decir exactamente cómo evoluciona a lo largo del tiempo la probabilidad de encontrar al caminante en cierto nodo de la red. Una vez conocida la probabilidad, la utilizamos para encontrar los promedios que buscamos. Esta idea se llama enumeración exacta. Finalmente, podemos plantear un sistema de ecuaciones, donde los tiempos promedio están expresados unos en función de otros. Es posible resolver el sistema utilizando álgebra lineal y encontrar con un solo cálculo los promedios para todos los nodos de la red. Este método se llama

el método matricial.

1.2. Plan de la tesis

La tesis se divide en dos partes que separan por un lado, el trabajo que ya existía sobre el tema, y por otro, nuestras aportaciones. Hay, sin embargo, algunos temas que se cuelan en la sección que no pertenecen porque antepone la coherencia lógica del trabajo.

La parte I es una síntesis del trabajo previo al nuestro. En el capítulo 2 presentamos las caminatas aleatorias y las redes complejas. Más adelante, describimos la red de mundo pequeño, nuestro sujeto de estudio, y damos las herramientas matemáticas para trabajar con las caminatas aleatorias en este tipo de red. También definimos los tiempos de primer paso y de primer encuentro, y los tiempos promedio generales y los promedios sobre configuraciones.

El capítulo 3 comienza con los resultados analíticos que se obtienen para la red de mundo pequeño más simple. Después presentamos los métodos numéricos que sirven para encontrar los tiempos de primer paso y primer encuentro, en redes que son más complicadas. Discutimos brevemente al final del capítulo, cuáles son las ventajas y aplicaciones de los diferentes métodos, y con esto terminamos la primer parte.

La parte II comienza con el capítulo 4 donde hacemos la validación de los métodos numéricos. El capítulo 5 contiene los resultados que obtenemos al estudiar ejemplares específicos de redes. Al principio discutimos los resultados que se obtienen para redes que no tienen atajos; más adelante, para redes que tienen atajos, estudiamos un caso que deja claro que las caminatas de primer paso no son simétricas respecto a la condición inicial y el sitio meta, y comparamos los tiempos de primer paso con los tiempos de primer encuentro, encontrando un parecido en un caso particular.

Los tiempos generales y los promedios sobre configuraciones se encuentran en el capítulo 6. Es aquí donde comparamos diferentes redes, e investigamos cómo afecta a los tiempos de la densidad de atajos. Encontramos una relación interesante entre los promedios generales y la longitud media de las redes. Al tomar el promedio sobre configuraciones, encontramos de nuevo un parecido entre los tiempos de primer paso y los de primer encuentro.

En el capítulo 7, investigamos la correlación de los tiempos prome-

dio con la distancia que separa los sitios físicos de la red, y también la distancia que separa en cada momento a los caminantes que tienen que encontrarse. Mostramos que esta última hace una caminata aleatoria de primer paso, y con ello encontramos una explicación para el parecido de los tiempos de uno y dos caminantes.

Por último, damos un resumen de los resultados más destacados del trabajo, y direcciones que puede ser interesante investigar más adelante.

Parte I
Antecedentes

Capítulo 2

Caminatas y redes

En este capítulo presentamos las caminatas aleatorias y las definiciones básicas de redes, la red de mundo pequeño y las herramientas para estudiar caminatas en redes. Definimos las dos cantidades que nos interesan a lo largo de la investigación: el tiempo promedio de primer paso, y el tiempo promedio de primer encuentro.

2.1. Caminatas aleatorias en el espacio real

Quizás las caminatas aleatorias más sencillas, o por lo menos las que podemos entender mejor porque para ellas tenemos la mayor intuición, son las caminatas en el espacio real. Es interesante que no hay un único tipo de caminata en el espacio real, hay múltiples definiciones que capturan mejor o peor un aspecto u otro de lo que sea que se quiera estudiar. Para comenzar, hay dos familias importantes, según se utilice una versión discreta o continua del espacio.

2.1.1. Espacio discreto

Para fijar ideas, consideremos dos dimensiones. En el plano, una caminata discreta tiene lugar en una cuadrícula. A partir del sitio que ocupa, que llamamos el centro, el caminante se puede mover a cualquiera de sus vecinos. Lo más común es tomar a los vecinos más cercanos del centro, esto es, en forma de cruz griega $+$. Esta vecindad se llama vecindad de *Von Neumann* [2]. Un ejemplo de caminata que se obtiene con esta definición

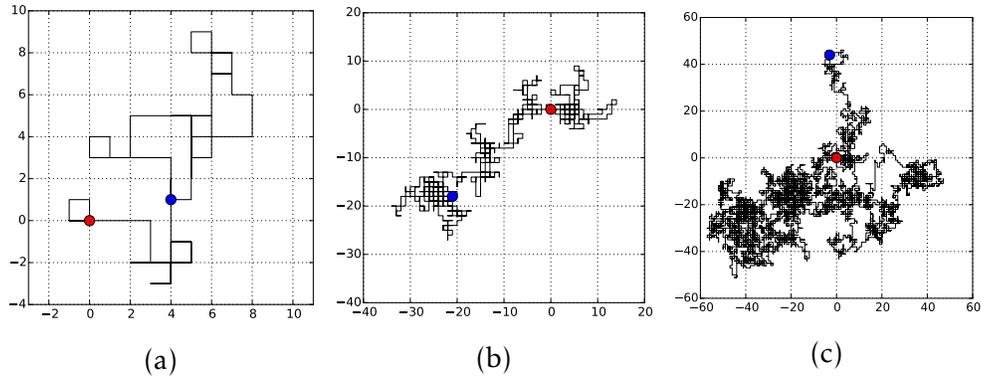


Figura 2.1: Trayectoria de una caminata aleatoria discreta después de (a) 100, (b) 1 000 y (c) 10 000 pasos. Se colorea en rojo el punto donde empezó la caminata (el origen) y en azul el punto donde termina.

se muestra en la fig. 2.1. También es común incluir en la vecindad a los sitios en las diagonales del centro, de manera a rodear completamente al centro. Esta vecindad se conoce como vecindad de *Moore* [2].

2.1.2. Espacio continuo

Cuando se utiliza espacio continuo, los brincos son vectores $\vec{s} \in \mathbb{R}^d$ (d es el número de dimensiones del espacio). Hay diferentes maneras de definir cómo se escoge el paso \vec{s} ; dos casos típicos son tomar los pasos Δx y Δy uniformemente en el intervalo $[-a, a]$, o expresar al paso $\vec{s} = (r, \theta)$ en coordenadas polares y escoger el ángulo uniformemente en $[0, 2\pi)$. A menudo se toma el radio constante, $r = 1$. Vale la pena decir que siempre hay transformaciones que nos llevan de un tipo de caminata a la otra. La segunda definición con $\vec{s} = (1, \theta)$ se usa para la caminata de la fig. 2.2.

2.1.3. Tiempo

A menudo se obvia que las caminatas aleatorias no ocurren solamente en el espacio, pero también en el tiempo. De la misma forma que para el espacio, las caminatas se pueden definir para tiempo discreto o continuo. En la primera versión (que es la que se utiliza en los modelos de este trabajo), el caminante da un brinco en cada paso de tiempo. En la de tiempo

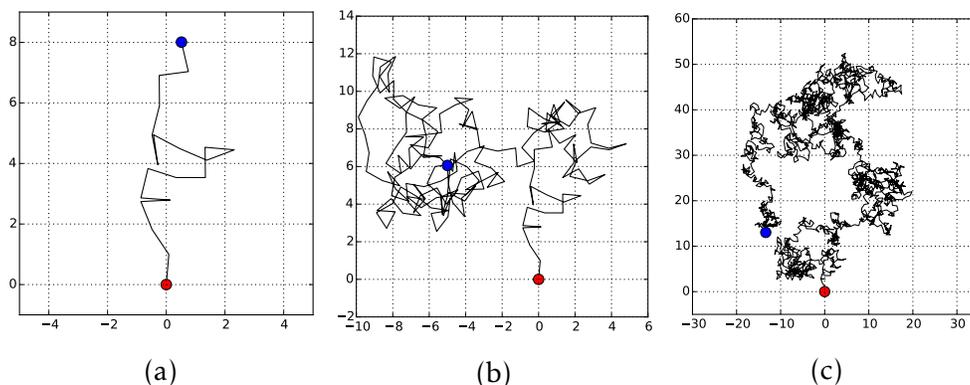


Figura 2.2: Trayectoria de una caminata aleatoria continua después de (a) 20, (b) 200 y (c) 2 000 pasos. Los pasos tienen norma unitaria y $\theta \in [0, 2\pi)$. Se colorea en rojo el punto donde empezó la caminata (el origen) y en azul el punto donde termina.

continuo, la partícula espera un tiempo Δt entre cada brinco. Δt se obtiene a partir de una distribución de probabilidad definida por una tasa (es común utilizar una distribución exponencial).

2.1.4. El teorema del límite central

Una de las propiedades más interesantes de las caminatas aleatorias es que no importa demasiado cómo se definen—si se utilizan pasos discretos o continuos, y en este caso de qué manera se toma el paso \vec{s} —ya que, para un número suficientemente grande de pasos todas las caminatas se parecen [22]. Esto se puede ver cualitativamente al comparar las figuras 2.2c y 2.1c. Sin embargo, el parecido es más que cualitativo; todas las caminatas que satisfacen cierto criterio técnico—que la distribución del tamaño de los pasos tiene una varianza finita—satisfacen el teorema del límite central. En la figura 2.3a, podemos apreciar que un número grande de caminantes del tipo $\vec{s} = (r, \theta)$ en coordenadas polares se distribuye con una simetría radial después de unos pocos pasos. Los caminantes se distribuyen siguiendo una distribución normal: el perfil de la coordenada x en la figura 2.3b se dibujó junto con el ajuste de dicha distribución y el acuerdo es muy bueno. Este comportamiento no es especial del paso polar $\vec{s} = (r, \theta)$ y para las otras definiciones de caminatas, incluida la caminata discreta, se puede obtener

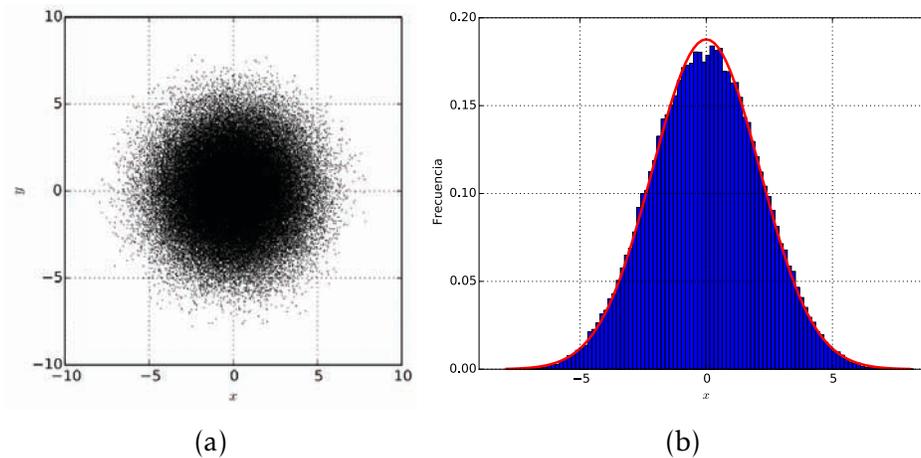


Figura 2.3: Teorema del límite central: (a) posición de diez mil caminantes distintos, después de 10 pasos, usando la definición polar de pasos continuos y (b) histograma de la coordenada x de los puntos, con un ajuste normal en rojo.

algo completamente análogo.

2.2. Redes complejas

Una **red** es un par ordenado $G = (V, E)$ donde V es un conjunto de **nodos** y $E \subseteq V \times V$ es un conjunto de **conexiones**¹ que unen a los nodos [3]. De vez en cuando utilizaremos **sitio** como sinónimo de nodo

Normalmente los nodos se asocian con un número entero, lo cuál es práctico para identificarlos con un índice. Uno puede representar una red acomodando los nodos en una circunferencia. En este trabajo los numeramos siguiendo el círculo en el sentido de las manecillas del reloj, utilizando la convención de que el primer nodo lleva el índice 0. Así, si la red G está formada por N nodos, el último lleva el índice $N - 1$:

$$V = \{0, 1, \dots, N - 1\}. \quad (2.1)$$

Al nodo 0 lo llamaremos el **origen de la red**.

¹Los símbolos V y E vienen de *vertex* y *edge*, vértice y arista en inglés, que se utilizan en teoría de gráficas. G mismo viene de *graph*.

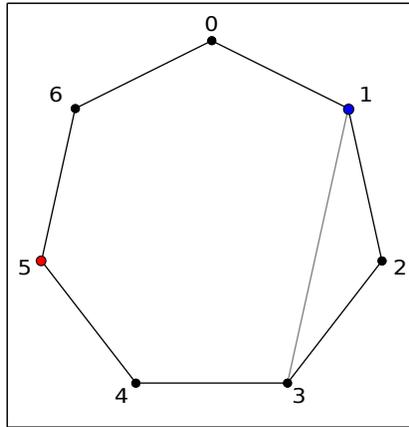


Figura 2.4: **Distancia geodésica:** Hay tres caminos que conectan los sitios 1 y 5: el primero pasa por 2, 3 y 4 y es de longitud 4, el segundo toma el atajo y pasa por 3 y 4 y el último pasa por 0 y 6. Los dos últimos son de longitud 3, por lo tanto, la distancia geodésica es $d(1, 5) = 3$.

Dado E , el conjunto de conexiones de la red G , definimos el **conjunto de vecinos** del sitio i como

$$N(i) = \{j \in \{0, \dots, N-1\} : (i, j) \in E\}.$$

El **grado** de un sitio es el número de conexiones que tiene:

$$\text{deg}(i) = |N(i)|.$$

donde $|\dots|$ denota la cardinalidad de un conjunto.

La **distancia geodésica** $d(i, m)$ que separa los nodos i y m es la longitud, en número de conexiones, del camino más corto que une dichos nodos. Puede que este camino no sea único, de hecho es frecuente que haya más de un camino de longitud mínima entre una pareja de nodos, por ejemplo, en la figura 2.4.

La **longitud media** de la red $L(G)$ es el promedio de las distancias geodésicas:

$$L(G) = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{m=0}^{N-1} d(i, m) \quad (2.2)$$

Un subconjunto no disconexo $T \subseteq V$ formado de tres nodos es un **triángulo** si cada nodo en T está conectado con los otros dos.

2.2.1. La red de mundo pequeño

Una red de mundo pequeño tiene la propiedad de que la mayoría de los pares de nodos están conectados por un camino corto, al mismo tiempo que la la fracción de nodos agrupados en triángulos se mantiene alta. La primera propiedad es típica de una gráfica aleatoria, la segunda lo es de una gráfica regular. Es por este motivo que una red de mundo pequeño tiene una estructura muy particular. Una familia de redes con diferente número de nodos N se define como una familia de redes de mundo pequeño si la longitud media L escala como

$$L \sim \log N$$

La figura 2.5 muestra este comportamiento.

Para modelar el efecto de mundo pequeño, Watts y Strogatz [24] inventaron un modelo simple que da lugar a una familia de redes, generadas por un parámetro p , que abarca desde una gráfica regular hasta algo parecido, aunque no del todo, a una gráfica aleatoria.

Se comienza con una red regular con N nodos y $2k$ vecinos cercanos, la red con parámetro $p = 0$ que es exactamente la de la figura 2.6a. Para obtener al resto de las redes de la familia, se recorre cada conexión y se decide si se modifica con probabilidad p . Modificar la conexión es dejar un extremo fijo y escoger el nuevo nodo al que se conecta el otro extremo de manera uniforme en toda la red con la restricción de que se prohíbe duplicar conexiones ya existentes.

En otra versión, se puede reconectar los dos extremos de cada unión, permitiendo repetir una conexión y también que un nodo sea su propio vecino. Esto tiene la ventaja de que para $p = 1$ se obtiene una familia de verdaderas gráficas aleatorias de Erdős–Rényi.

Sethna [22], como lo hicieron antes Monasson [12] y Newman y Watts Newman and Watts [16], propone una variante más del modelo. En su versión se mantienen todas las conexiones con los vecinos cercanos, y se agregan pNk “atajos” o nuevas conexiones de largo alcance para las cuales cada extremo se escoge uniformemente en toda la red. La única condición que se prohíbe es conectar a un nodo consigo mismo, pero se puede dupli-

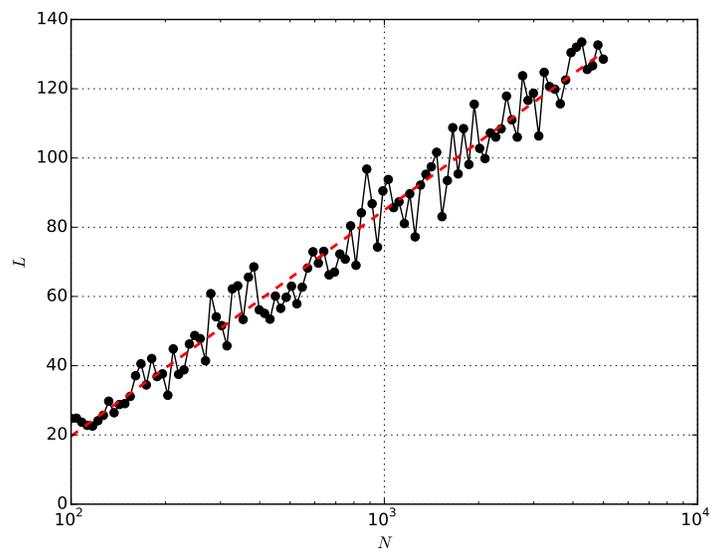


Figura 2.5: **La propiedad de mundo pequeño:** Una familia de redes con N nodos tienen una longitud media L que escala como $\log(N)$. Obtenemos las diferentes redes con el modelo de nuevas conexiones que se define un poco más adelante. Cada red tiene número de vecinos $k = 1$ y densidad de atajos $p = 0.01$.

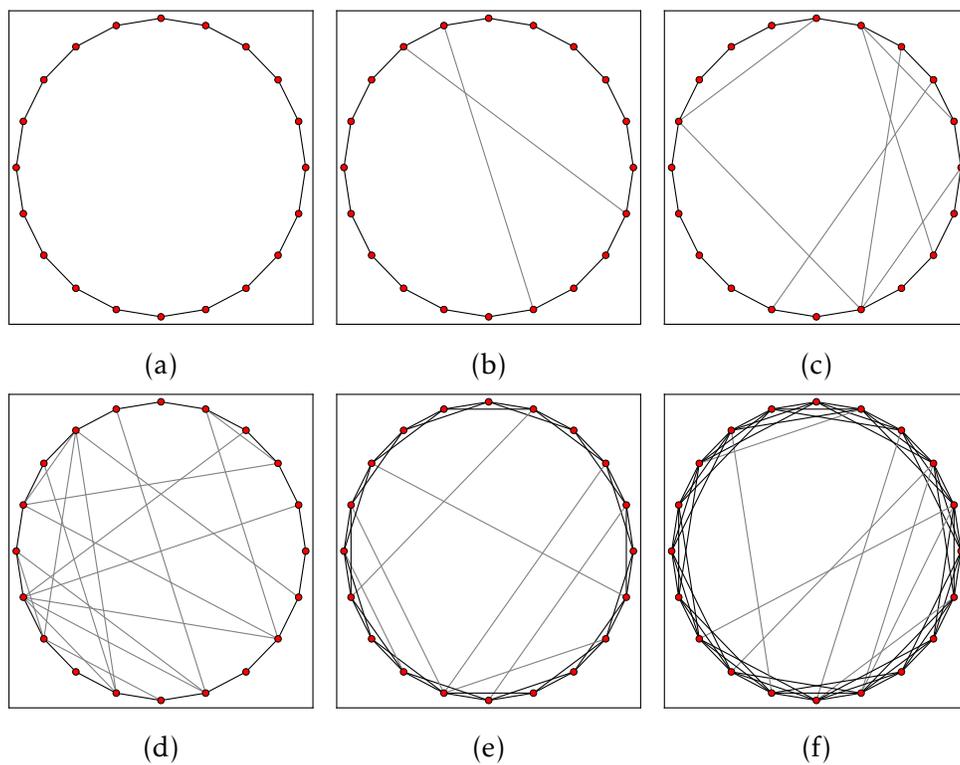


Figura 2.6: Redes de mundo pequeño con $N = 20$ nodos. Las primeras cuatro redes (de derecha a izquierda y de arriba a abajo) tienen $k = 1$ vecinos cercanos y densidad de atajos p (a) 0, (b) 0.1, (c) 0.4 y (d) 1. Las últimas tienen $p = 0.2$ y (e) $k = 2$ y (f) $k = 3$ vecinos cercanos.

car una conexión que ya existía. De esta manera, del total de conexiones $N \times 2k/2 = Nk$, p es la fracción de nuevos atajos que se van a agregar:

$$\frac{\# \text{ nuevas conexiones}}{\# \text{ conexiones entre vecinos cercanos}} = \frac{pNk}{Nk} = p.$$

En esta versión p no es una probabilidad—el número de atajos está fijo—pero el parámetro p codifica la densidad de atajos y su definición lo hace lo más parecido posible a la probabilidad de Watts y Strogatz.

Como los atajos pueden repetir una unión que ya existía, el número real de nuevas conexiones puede ser igual o menor a pNk .

Una ventaja de este modelo es que ningún nodo puede terminar desconectado del resto, pero una desventaja es que la familia de gráficas que tienen $p = 1$ ya no son gráficas aleatorias. Sin embargo, las diferentes versiones son lo suficientemente parecidas cuando el parámetro p es pequeño [15].

La versión de *Sethna* es la que se utiliza en este trabajo. En la figura 2.6, podemos encontrar diferentes versiones que se obtienen para distintas combinaciones de los parámetros p y k .

2.2.2. Producto cartesiano de dos redes

De Abdullah et al. [1] tomamos la siguiente definición: para un par de redes G y H finitas, se define el **producto cartesiano** $P = G \times H$ tal que

- $V_P = V_G \times V_H$
- $((a, x), (b, y)) \in E_P$
 - $(a, b) \in E_G$ y $x = y$ o
 - $a = b$ y $(x, y) \in E_H$.

Uno puede imaginar el producto P como en la figura 2.7: hacemos una copia nueva de una de las dos redes originales, por ejemplo G , una vez por cada nodo de la otra, H . Para la copia de G que corresponde al nodo $x \in V_H$, G_x , y para $a \in V_G$, agregamos la conexión que une $a \in G_x$ con $a \in G_y$ para todos los nodos $y \in V_H$ tales que $(x, y) \in E_H$.

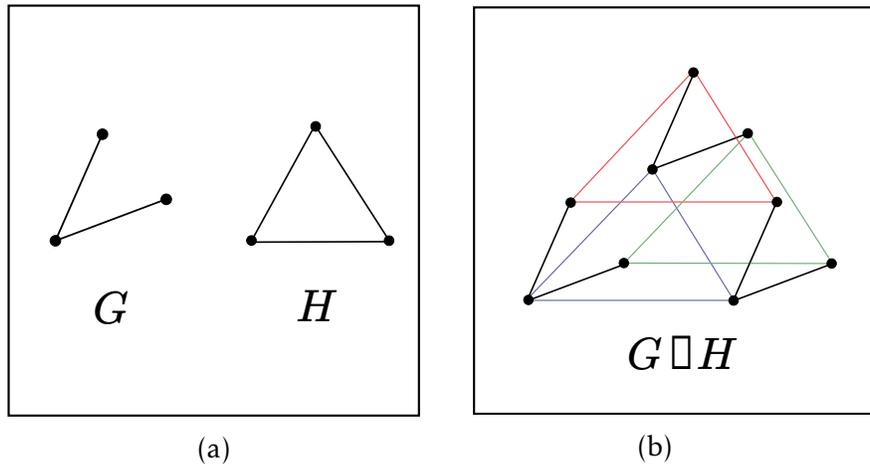


Figura 2.7: **Producto cartesiano de dos redes:** (a) Las redes originales G y H y (b) su producto cartesiano.

2.3. Caminatas en redes

Regresamos al tema de las caminatas aleatorias. Queremos ahora discutir la caminata que tiene lugar sobre una red de mundo pequeño. En esta sección, vamos a dar las herramientas matemáticas que nos sirven.

2.3.1. Formalismo de las caminatas aleatorias

Nuestro objeto de estudio es la secuencia de variables aleatorias discretas X_0, X_1, \dots , que denotamos $(X_n)_{n \geq 0}$. El valor que toma una la variable aleatoria X_n es la posición que ocupa un caminante en la red G , al tiempo n , y por lo tanto pedimos que $X_n \in V \ \forall n \geq 0$.

La secuencia $(X_n)_{n \geq 0}$ es una caminata aleatoria si se cumple:

(I) $\mathbb{P}(X_0) = i_0$

(II) $\mathbb{P}(X_{n+1} | X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \mathbb{P}(X_{n+1} | X_n = i_n)$.

La condición (ii) se llama propiedad de Markov, y formaliza el hecho de que el proceso no tiene memoria: el paso que toma el caminante en un tiempo dado no depende de la trayectoria que siguió para llegar al lugar que ocupa.

Puesto que una partícula en el sitio i escoge su destino en el conjunto $N(i)$, se tiene que cumplir además la restricción $x_{i+1} \in N(x_i)$. Esto se reduce a pedir que un caminante se mueva solamente a través de las conexiones de la red.

Nos interesa calcular la probabilidad de las diferentes trayectorias que puede seguir una caminata que empezó en el sitio x_0 . La probabilidad de que ocupe el sitio i al tiempo t es $\mathbb{P}(X_t = i | X_0 = x_0)$.

Normalmente queda claro por el contexto que estamos hablando de las trayectorias que se obtienen a partir de la condición inicial x_0 , y entonces omitiremos hacer explícito que ésta es una probabilidad condicional, escribiendo simplemente

$$\mathbb{P}(X_t = i).$$

Una trayectoria puede pasar múltiples veces por el sitio i . Como justificamos en la introducción, nos interesa estudiar la probabilidad de que el caminante pase *por primera vez* por el sitio meta i , que denotamos

$$\mathbb{F}(X_t = i).$$

Queda implícita en esta notación, como con la probabilidad \mathbb{P} , la condición inicial x_0 . Esta probabilidad en términos de \mathbb{P} , es la probabilidad de que al tiempo t el caminante ocupe el sitio meta sin haberlo ocupado antes al tiempo t , $\forall t < t$. Formalmente:

$$\mathbb{F}(X_t = i) = \mathbb{P}(X_t = i, X_{t-1} \neq i, \dots, X_0 \neq i). \quad (2.3)$$

Se le puede dar un peso distinto a cada una de las conexiones (esto se hace por ejemplo en modelos de circuitos eléctricos, ver el libro de Redner [20]) pero en este trabajo vamos a asignarles a todas el mismo peso. En la práctica esto significa que cada elemento de $N(i)$ tiene la misma probabilidad de ser seleccionado, y la probabilidad de transición entre dos sitios es

$$\mathbb{P}(i \rightarrow j) := W_{i,j} = \begin{cases} \frac{1}{\deg(i)} & j \in N(i) \\ 0 & j \notin N(i). \end{cases} \quad (2.4)$$

De esta definición queda claro que en general las probabilidades de transición no son simétricas:

$$W_{i,j} \neq W_{j,i}$$

a menos de que $\deg(i) = \deg(j)$.

En este trabajo, requerimos que los caminantes se muevan en cada paso de tiempo. Esto es equivalente a pedir $i \notin N(i)$.

2.3.2. La ecuación maestra

Trabajando con un caminante, partimos de la probabilidad de que éste se encuentre en el sitio i , al tiempo t , dado que empezó en el sitio x_0 , $\mathbb{P}(X_t = i | X_0 = x_0) = \mathbb{P}(X_t = i)$. Una caminata aleatoria es una cadena de Markov, así que $\mathbb{P}(X_t)$ depende sólo de las probabilidades al tiempo $t - 1$. El teorema de la probabilidad total nos da

$$\begin{aligned} \mathbb{P}(X_t = i) &= \sum_{j \in N(i)} \mathbb{P}(X_{t-1} = j) W_{i,j} \\ \mathbb{P}(X_t = i) &= \frac{1}{\deg(i)} \sum_{j \in N(i)} \mathbb{P}(X_{t-1} = j) \end{aligned} \quad (2.5)$$

Esta ecuación se llama **ecuación maestra**. La condición inicial, cuando el caminante empieza en el sitio x_0 , es

$$\mathbb{P}(X_0 = i) = \delta_{x_0,i}. \quad (2.6)$$

2.4. Tiempos de primer paso y de primer encuentro

En esta sección definimos las dos cantidades que nos interesa encontrar en las simulaciones de caminatas aleatorias, en redes de mundo pequeño.

2.4.1. Condiciones de terminación y promedios de las distribuciones

Para estudiar los procesos de primer paso, necesitamos agregar condiciones de terminación a las caminatas aleatorias. En primer lugar queremos estudiar una caminata que empieza en un sitio i y que termina cuando

llega por primera vez a un sitio meta m . El tiempo que le toma a un caminante cumplir esta condición se llama tiempo de primer paso, y lo denotamos por $\tau_P(i \rightarrow m)$. Un número infinito de simulaciones de la caminata define una distribución de la variable aleatoria $\tau_P(i \rightarrow m)$. El promedio de dicha distribución se denota por $\langle \tau_P(i \rightarrow m) \rangle$, y es la cantidad que queremos obtener. Vamos a referirnos a esta cantidad, como es costumbre, con las siglas en inglés de tiempo promedio de primer paso, MFPT (*Mean First Passage Time*).

En segundo lugar queremos estudiar el primer encuentro de dos caminatas que empiezan respectivamente en los sitios i y m . Los tiempos que les toma encontrarse se llaman tiempos de primer encuentro, $\tau_E(i, m)$. Como en el caso anterior, nos interesa el promedio de la distribución, $\langle \tau_E(i, m) \rangle$. En lugar de tiempo promedio de primer encuentro, lo llamamos MFET (*Mean First Encounter Time*).

Los caminantes que se mueven en la misma red hasta encontrarse son indistinguibles (un detalle al que tenemos que poner cuidado en las simulaciones), y se satisface

$$\tau_E(i, m) = \tau_E(m, i) . \quad (2.7)$$

Sin embargo, la probabilidad de transición entre sitios vecinos $W_{i,j}$ no es simétrica en general, y para primer paso tenemos:

$$\tau_P(i \rightarrow m) \neq \tau_P(m \rightarrow i) . \quad (2.8)$$

2.4.2. Reformulación del problema de primer encuentro como un problema de primer paso

Cuando queremos calcular explícitamente la probabilidad de que dos caminantes se encuentren, es mucho más fácil formular problema de primer encuentro como un problema de primer paso en un espacio de dimensión mayor. Queremos definir, por tanto, una caminata aleatoria en dos dimensiones.

Un caminante en una red G tiene en cada paso de tiempo $x \in V(G) = \{0, \dots, N - 1\}$. Si denotamos la posición del segundo caminante por $y \in V(G)$, definimos, en cada paso de tiempo, el vector posición

$$\vec{X} = (X, Y) \in \{(x, y) : x, y \in \{0, \dots, N - 1\}\} \subset \mathbb{Z}^2. \quad (2.9)$$

El vector \vec{X} es un nuevo caminante en \mathbb{Z}^2 , y sus coordenadas son los sitios que ocupan cada una de las partículas en la red original.

La dinámica del nuevo caminante es como sigue. En el caso más simple en el que la red tiene $2k = 2$ primeros vecinos y no tiene atajos, $p = 0$ (la red de la figura 2.6a), un caminante sólo puede brincar al sitio a su izquierda o a su derecha. Cuando lo hace el otro caminante se queda fijo, de modo que si el primer caminante es el que brinca, la coordenada x aumenta o disminuye en una unidad, y y no cambia; a la inversa, si brinca el segundo, y aumenta o disminuye una unidad y x no cambia. Por lo tanto, en \mathbb{Z}^2 , la vecindad del sitio \vec{X} es una vecindad de Von Neumann.

Si el número de primeros vecinos aumenta, $k > 1$, la vecindad se agranda e incorpora k sitios arriba, k sitios abajo, k a la izquierda y k a la derecha de \vec{X} : se mantiene la forma de cruz griega, pero el tamaño de cada brazo de la cruz crece. Si $p > 0$, además de los vecinos de la cruz, si x o y tienen atajos en G , hay vecinos alejados de \vec{X} .

Los caminantes X y Y tienen un encuentro en la red original G si y sólo si $X = Y$. Para el vector \vec{X} , decimos que hay un encuentro si \vec{X} pertenece al conjunto

$$E = \{(x, x) : x \in V(G)\}. \quad (2.10)$$

En \mathbb{Z}^2 , el conjunto E es el conjunto de sitios de la diagonal. Podemos encontrar esta representación en la figura 2.8.

El problema de encontrar los tiempos de primer encuentro de dos caminantes se convierte, por lo tanto, en el problema de encontrar los tiempos de primer paso del caminante en dos dimensiones, para llegar al conjunto E . Si los sitios i y m son las condiciones iniciales de los caminantes en G ,

$$\tau_E(i, m) = \tau_P(\vec{X}_0 \rightarrow E). \quad (2.11)$$

donde

$$\vec{X}_0 = (i, m).$$

Como en el caso de un caminante, definimos la probabilidad de primer encuentro al tiempo t ,

$$\mathbb{F}(\vec{X}_t \in E).$$

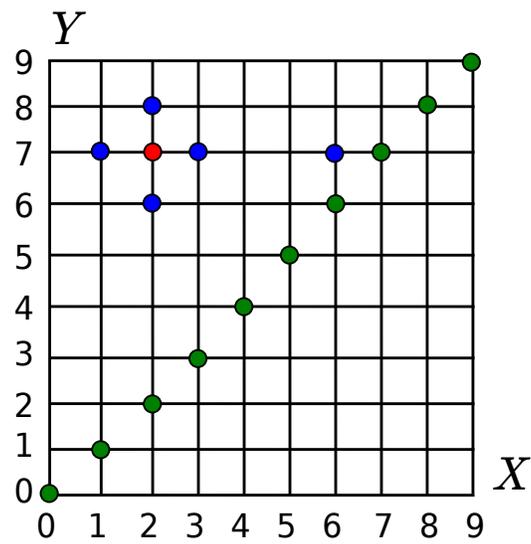


Figura 2.8: **Reformulación del problema:** Reformulamos el problema de encuentro de dos caminantes en una red como el primer paso del vector \vec{X} por los sitios del conjunto E . Representamos al sistema como una cuadrícula en \mathbb{Z}^2 : el caminante es el punto rojo, sus vecinos son los sitios azules y los sitios del conjunto E son verdes. Notemos que la vecindad del caminante es una cruz griega o vecindad de *Von Neumann*, pero además incluye un sitio alejado: en la red original hay un atajo que une los nodos 2 y 6.

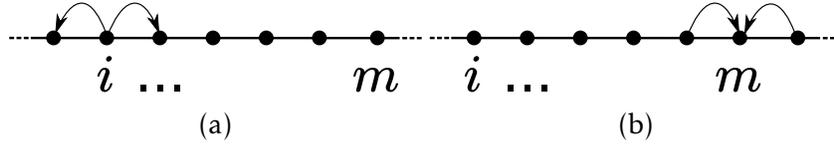


Figura 2.9: (a) La meta está fija y nos fijamos en el primer paso que hace el caminante desde la condición inicial i . (b) La condición inicial está fija y nos interesa el último paso que hace el caminante para llegar a la meta m .

Queda implícita en esta notación la condición inicial; sabemos que se refiere a la probabilidad condicional $\mathbb{P}(\vec{X}_t \in E | \vec{X}_0 = (i, m))$.

La probabilidad de primer encuentro se expresa en términos de la probabilidad de ocupación \mathbb{P} , como

$$\mathbb{P}(\vec{X}_t \in E) = \mathbb{P}(\vec{X}_t \in E, \vec{X}_{t-1} \notin E, \dots, \vec{X}_0 \notin E). \quad (2.12)$$

Si G era nuestra red original, la caminata aleatoria en dos dimensiones formalmente tiene lugar en el producto cartesiano $G \times G$. Hablamos de la caminata en dos dimensiones en \mathbb{Z}^2 de la misma manera que decimos que la caminata en G tiene lugar en una dimensión (en el conjunto $V(G) \cong \mathbb{Z}$). Es más fácil imaginar la caminata en dos dimensiones en el plano discreto, con todo y los brincos a sitios lejanos cuando uno de los caminantes toma un atajo, que en la red $G \times G$.

2.4.3. Ecuación de recurrencia para el MFPT

Siguiendo a Redner [20], podemos expresar los tiempos promedio unos en función de otros. Ésto resulta ser útil porque podemos encontrarlos todos al resolver un sistema lineal de ecuaciones.

Nos interesamos primero por el tiempo promedio de primer paso, $\tau_P(i, m)$. Podemos fijar el sitio meta m e investigar cómo se comportan los tiempos de primer paso desde diferentes condiciones iniciales i , o podemos hacer exactamente lo contrario, fijar la condición inicial y tomar distintos sitios meta. Analizamos los dos casos por separado. Ambos casos se encuentran representados en la figura 2.9.

Sitio meta fijo

El punto de partida es el valor esperado condicional. Dadas dos variables aleatorias X y Y , se tiene

$$E[X] = \sum_y E[X|Y = y] \mathbb{P}(Y = y). \quad (2.13)$$

Identificamos a la variable aleatoria X con el tiempo de primer paso $\tau_P(i \rightarrow m)$, y a la variable aleatoria Y como el sitio vecino al que brinca el caminante desde la condición inicial i , en el siguiente paso de tiempo. La ecuación anterior se convierte en

$$E[\tau_P(i \rightarrow m)] = \tau_P(i \rightarrow m) = \sum_{j \in N(i)} E[\tau_P(i \rightarrow m) | i \rightarrow j] W_{i,j}.$$

Hemos cometido un abuso de notación al escribir la condición $i \rightarrow j$. Queremos indicar simplemente que buscamos el valor esperado de la variable aleatoria $\tau_P(i \rightarrow m)$, dado que el caminante que empieza en i se mueve al sitio j en su primer movimiento. Recordamos que $N(i)$ denota el conjunto de los sitios vecinos a i .

Concentremos nuestra atención en el sumando $E[\tau_P(i \rightarrow m) | i \rightarrow j]$. Si el caminante brinca de i a j , lo que tendremos como nueva variable aleatoria es el tiempo que tarda de ir de j a la misma meta m . Sin embargo, hay que tomar en cuenta que el caminante “perdió” una unidad de tiempo en ir del sitio i a j :

$$\begin{aligned} E[\tau_P(i \rightarrow m) | i \rightarrow j] &= E[\tau_P(j \rightarrow m) + 1] \\ &= \tau_P(j \rightarrow m) + 1. \end{aligned}$$

Para evitar cargar con demasiada notación, definimos $\mu_{i,m} := \tau_P(i \rightarrow m)$. Usando el valor de la probabilidad de transición dada en (2.4), encontramos

$$\begin{aligned} \mu_{i,m} &= \sum_{j \in N(i)} \frac{1 + \mu_{j,m}}{\deg(i)} \\ \mu_{i,m} &= 1 + \frac{1}{\deg(i)} \sum_{j \in N(i)} \mu_{j,m}. \end{aligned} \quad (2.14)$$

Las condiciones de frontera de esta ecuación vienen implícitas en el conjunto de vecinos $N(i)$ de cada sitio i que tomemos como condición inicial de la caminata. De cualquier manera, ya hemos visto que en la red de mundo pequeño todos los sitios están conectados por lo menos con $2k$ vecinos cercanos y que el origen, 0 , es vecino del sitio $N - 1$. Tenemos, por lo tanto, condiciones de frontera periódicas.

Condición inicial fija

También en este caso partimos del valor esperado condicional, (2.13). Identificamos la variable aleatoria X con el tiempo de primer paso, pero ahora identificamos a la variable Y con el sitio desde el cual el caminante brinca para llegar a la meta m . Entonces:

$$E[\tau_P(i \rightarrow m)] = \tau_P(i \rightarrow m) = \sum_{j \in N(m)} E[\tau_P(i \rightarrow m) | j \rightarrow m] W_{j,m}.$$

El valor esperado condicional del lado derecho, nos da una nueva variable aleatoria: el tiempo que tarda el caminante que empezó en i , en llegar por primera vez a j . Una vez en j , el caminante toma una unidad de tiempo más en llegar a la meta m , por lo tanto

$$\begin{aligned} E[\tau_P(i \rightarrow m) | j \rightarrow m] &= E[\tau_P(j \rightarrow m) + 1] \\ &= \tau_P(i \rightarrow j) + 1. \end{aligned}$$

Usando la notación que definimos antes, $\mu_{i,m} := \tau_P(i \rightarrow m)$, encontramos la ecuación de recurrencia

$$\mu_{i,m} = \sum_{j \in N(m)} \frac{1 + \mu_{i,j}}{\deg(j)} \quad (2.15)$$

Comparada con (2.14), esta expresión es un poco más complicada porque esta vez el factor $1/\deg(j)$ no es común a todos los sumandos, y no lo podemos factorizar.

2.4.4. Ecuación de recurrencia para el MFET

Para encontrar el tiempo promedio de primer encuentro de dos caminantes, aprovechamos la reformulación del problema que hicimos antes, la ecuación (2.11).

Para simplificar la notación, definimos $\nu_{i,m} := \tau_E(i, m)$. La ecuación del valor esperado condicional aplica igualmente en este caso, y encontramos que el tiempo promedio de un sitio está relacionado con los tiempos promedio de sus vecinos exactamente de la misma manera para el MFPT donde el sitio meta está fijo:

$$\nu_{i,m} = 1 + \frac{1}{\deg(i, m)} \sum_{j \in N(i, m)} \nu_{j,m}. \quad (2.16)$$

La diferencia con los tiempos de primer paso, debemos insistir, es que estamos usando sitios en dos dimensiones y nos referimos a ellos con dos índices. $N(i, m)$ se refiere al conjunto de vecinos en dos dimensiones del sitio (i, m) , y $\deg(i, m)$ es el grado del mismo sitio en la red $G \times G$.

Como en el caso de primer paso, las condiciones de frontera vienen implícitas en el conjunto de vecinos $N(i, m)$. Cada uno de los caminantes se mueve en una red de mundo pequeño que tiene fronteras cíclicas, por lo tanto la caminata en \mathbb{Z}^2 tiene fronteras periódicas.

2.5. Promedios

Queremos tener herramientas que nos permitan comparar diferentes redes de mundo pequeño. Es natural pensar en tomar promedios de los tiempos de primer paso y de primer encuentro que podamos encontrar. En primer lugar, podemos tomar un promedio de los tiempos que se obtienen para las diferentes condiciones iniciales. Al hacerlo obtenemos un solo número para cada configuración de red, y podemos comparar dos redes con los valores obtenidos. Por otro lado, queremos conocer el efecto que tienen los parámetros de la red, en especial la densidad de atajos p , sobre un conjunto grande de redes. En este caso, utilizamos un promedio sobre distintas configuraciones.

2.5.1. Tiempos promedio generales

Nos interesa resumir toda la distribución de tiempos en un solo número para comparar dos ejemplares de redes que tienen los mismos parámetros. Para una red compuesta por N nodos, definimos los tiempos generales de primer encuentro y de primer paso,

$$\tau_{E \ G} = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{m=0}^{N-1} \tau_E(i, m) \quad (2.17)$$

y

$$\tau_{P \ G} = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{m=0}^{N-1} \tau_P(i \ m) \quad (2.18)$$

Los llamamos respectivamente GFET y GFPT, por la abreviación de los nombres en inglés. Hacemos notar que el GFPT está relacionado con el tiempo promedio de cobertura de la red,

$$\frac{1}{N} \sum_{m=0}^{N-1} \tau_P(0 \ m). \quad (2.19)$$

Sin embargo, el GFPT que toma en cuenta todas las condiciones iniciales, y le resta protagonismo al origen comparado con el tiempo de cobertura. Consideramos que esto es importante, porque es completamente arbitrario cuál es el nodo que funge como el origen.

2.5.2. Promedios sobre configuraciones

Los resultados que tenemos para una red en particular son para parejas de condiciones iniciales. Podemos fijar, al inicio de la simulación, a un caminante o al sitio meta en el origen; de esta manera obtenemos funciones de una sola variable i (el segundo caminante y la condición inicial del tiempo de primer paso): $\tau_E(0, i)$ y $\tau_P(i \ 0)$.

Calculamos los tiempos de primer paso y de primer encuentro para cada sitio i , y posteriormente tomamos los promedios sobre i en un conjunto grande de redes. De esta manera podemos estudiar el efecto que tienen los parámetros de la red, especialmente la densidad de atajos p .

Vamos a reducir la notación del promedio temporal al tomar $c = \tau$, para denotar por c al promedio sobre configuraciones:

$$c_E(0, i)$$

se refiere al promedio de los tiempos de primer encuentro, y

$$c_P(0, i)$$

al promedio de los tiempos de primer paso. Es importante destacar que al promediar sobre configuraciones obtenemos los mismos resultados para $c_P(0 \rightarrow i)$ y $c_P(i \rightarrow 0)$, pues estamos promediando el “desorden” de la red. Eran justamente los atajos acomodados de forma “desordenada” los que impedían que las transiciones fueran simétricas, $W_{i,j} \neq W_{j,i}$. No tenemos por tanto que distinguir entre la dirección de las simulaciones de primer paso.

Capítulo 3

Anillos y métodos numéricos

En este capítulo discutimos el caso más sencillo de la red de mundo pequeño, el anillo simple con número de vecinos cercanos $k = 1$ y densidad de atajos $p = 0$, para el que podemos encontrar de forma analítica los MFPT y MFET. Mostramos también que para anillos que tienen $k > 1$ hay una relación exacta entre unos y otros. Posteriormente presentamos los métodos numéricos que sirven para el resto de las redes que no son el anillo simple, y que no tienen solución analítica.

3.1. Resultados exactos para anillos

Cuando la densidad de atajos p es cero, podemos mostrar que el MFET es la mitad del MFPT cuando se comparan sitio por sitio. Además, si $k = 1$, podemos encontrar la expresión analítica de los tiempos promedio.

3.1.1. El anillo simple

La red de mundo pequeño con N nodos, $2k$ vecinos cercanos, y densidad de atajos p toma una forma muy sencilla cuando $k = 1$ y $p = 0$. En esta situación, el sistema es unidimensional y tienen condiciones de frontera periódicas: cualquier sitio tiene dos vecinos inmediatamente a su lado. Habíamos encontrado esta red en la figura 2.6a cuando introducimos el modelo de mundo pequeño. Para referencia, la incluimos de nuevo en 3.1, señalando cuál es el origen. Llamamos a esta red, un anillo simple.

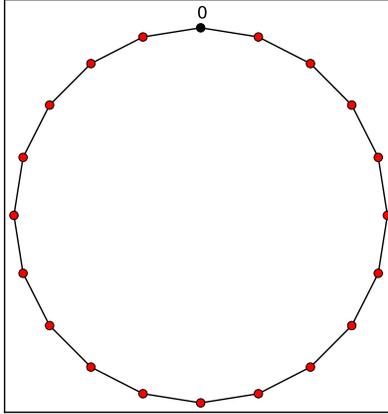


Figura 3.1: Anillo simple formado por $N = 20$ nodos.

Expresión exacta para el MFPT

En el capítulo 2 encontramos la ecuación de recurrencia (2.14) para el valor esperado de los tiempos de primer paso cuando el sitio meta m está fijo:

$$\mu_{i,m} = 1 + \frac{1}{\deg(i)} \sum_{j \in N(i)} \mu_{j,m}.$$

Para el anillo simple, $\deg(i) = 2k = 2$, y es fácil escribir quiénes son los vecinos del sitio i :

$$\mu_{i,m} = 1 + \frac{1}{2} (\mu_{i-1,m} + \mu_{i+1,m}).$$

Podemos hacer un cambio de variable: es claro que por la simetría del anillo $\mu_{i,m} = \mu_{0,m-i} = \mu_{m-i}$ (obviamos el cero en esta notación, y el tiempo de encuentro es función sólo de la variable $m - i$). Con esta notación, reacomodando términos tenemos

$$\mu_{m-(i+1)} - 2\mu_{m-i} + \mu_{m-(i-1)} = -2. \quad (3.1)$$

Del lado izquierdo de la ecuación, reconocemos una derivada discreta de segundo orden. La solución a esta ecuación es

$$\mu_{m-i} = \mu_{m,i} = -(m-i)^2 + A(m-i) + B.$$

Determinamos las constantes A y B con las condiciones de frontera: $\mu_{m,m} = 0$, $B = 0$ y $\mu_{m-N,m} = 0$, $A = N$. La solución completa está dada por

$$\mu_{i,m} = -(m-i)^2 + N(m-i), \quad m \geq i. \quad (3.2)$$

Esta expresión es la solución que encuentran Pandit y Amritkar [18].

Por definición el MFPT es positivo, así que la fórmula de arriba está definida sólo para m mayor que i . Esto resulta ser un ligero inconveniente, ya que en principio no está definido el tiempo de primer paso al origen, $\tau_P(i=0)$, cuando usamos la numeración usual para los índices, $i \in \{0, \dots, N-1\}$. Sin embargo, no es un inconveniente grave; si queremos la expresión cuando el origen es el sitio meta, podemos tomar i negativo. Es bastante claro cuál es el nodo a que nos referimos de esta forma: empezando en el origen recorreremos el anillo en el sentido contrario a las manecillas del reloj hasta encontrar el sitio $-i$.

Dos casos que nos interesan particularmente, son los casos en que la caminata parte del origen o llega al origen. Respectivamente encontramos:

$$\mu_{0,m} = -m^2 + Nm, \quad m \geq 0 \quad (3.3)$$

y

$$\mu_{i,0} = -i^2 - Ni, \quad i \leq 0. \quad (3.4)$$

El tiempo promedio de primer paso crece cuadráticamente con la distancia que separa la condición inicial y el sitio meta: ésta es la característica más importante de una caminata aleatoria en una dimensión. Pero además, como consecuencia de la frontera periódica, encontramos que la contribución lineal es proporcional al tamaño N de la red. En este trabajo, veremos que esta propiedad es también muy importante.

Expresión exacta para el MFET

Nos interesa ahora el tiempo promedio de primer encuentro, el MFET. Para encontrarlo, fijémonos en la distancia geodésica d que separa a los caminantes en el anillo. Para un par de caminantes en los sitios i y m , con $m \geq i$,

$$d(i, m) = \min\{m-i, N+i-m\}. \quad (3.5)$$

Es claro que $d(i, m) = 0$ si $i = m$, y también que la distancia está acotada:

$$d_{\max} = \begin{cases} N/2 & \text{si } N \text{ es par,} \\ \frac{N-1}{2} & \text{si } N \text{ es impar.} \end{cases} \quad (3.6)$$

Los caminantes nunca se mueven simultáneamente; en cada paso de tiempo se mueve uno después del otro, o se mueve dos veces el mismo caminante. En cualquier caso, la distancia d que los separa aumenta o disminuye en una unidad cada vez que camina uno de ellos. Por lo tanto, d hace también una caminata aleatoria en una dimensión en el conjunto $\{0, \dots, d_{\max}\}$. El tiempo de primer encuentro de los caminantes es igual al tiempo de primer paso de la distancia d por el origen:

$$\tau_E(i, m) = \tau_P(d = 0). \quad (3.7)$$

En principio, esta nueva caminata se hace en un sistema más pequeño que el original, que sólo tiene $d_{\max} + 1$ sitios, pero vamos a ver con detenimiento qué sucede en la frontera $d = d_{\max}$. De nuevo, tenemos que distinguir los casos N par o impar.

Denotemos por d_t la distancia que separa a los caminantes al tiempo t . Si N es par y $d_t = d_{\max}$, la distancia disminuye forzosamente en el siguiente paso: $d_{t+1} = d_{\max} - 1$. Esto se encuentra representado en la figura 3.2a. La frontera es absolutamente reflejante.

Por otro lado, si N es impar, en el siguiente paso la distancia puede quedarse como d_{\max} o puede disminuir. Ambos escenarios ocurren con probabilidad $1/2$. La figura 3.2b deja claro este comportamiento.

Tanto para N par como impar, podemos reflejar el sistema y tomamos el tiempo de primer paso en cada extremo.

Si N es par, reflejamos exactamente sobre $d = d_{\max}$ (fig. 3.3a); el nuevo sistema contiene $2d_{\max} + 1 = N + 1$ sitios. Si N es impar reflejamos a la derecha de $d = d_{\max}$ de manera a que este sitio tenga su reflejo (fig. 3.3b). El nuevo sistema también contiene $2d_{\max} + 2 = N + 1$ sitios.

Nos interesamos, entonces, en el primer paso por el primer y último sitio. Podemos identificarlos como el mismo sitio al tomar condiciones de frontera cíclicas, y esto nos deja con un anillo con N sitios. Así, encontramos que se puede considerar que la caminata aleatoria de la distancia también ocurre en un anillo simple de tamaño N . De esta manera conclu-

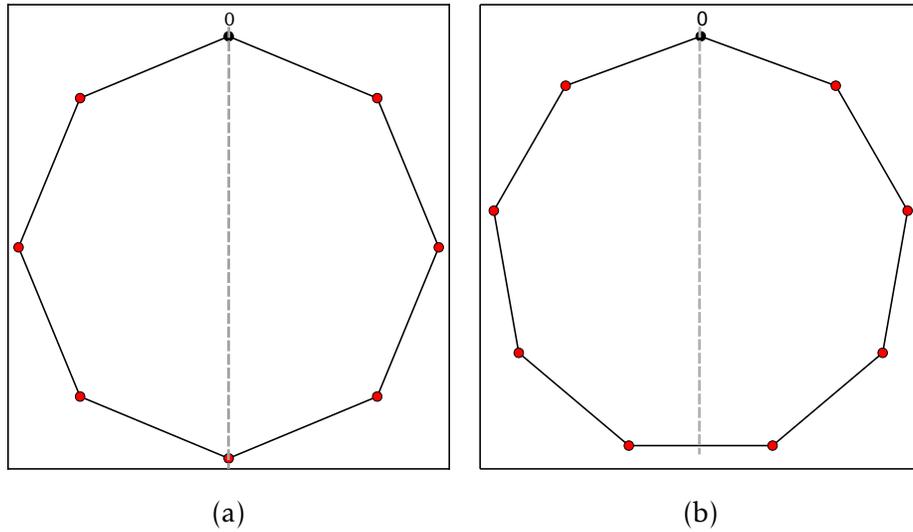


Figura 3.2: **Anillos simples par e impar:** (a) Anillo simple con $N = 8$ nodos y (b) con $N = 9$ nodos. En ambos casos $d_{\max} = 4$ pero a partir de la raya punteada vemos que la situación es diferente según la paridad; si N es par hay un sólo sitio separado d_{\max} del origen, si N es impar hay dos.

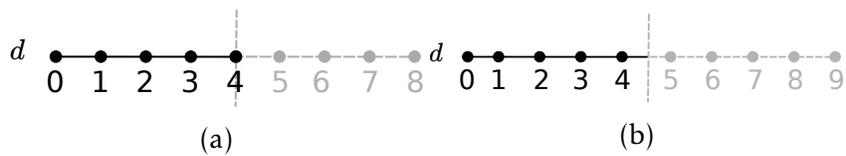


Figura 3.3: **Reflejamos el espacio d :** (a) Para un anillo con número par de nodos, reflejamos sobre d_{\max} ; (b) para un anillo impar reflejamos a la derecha de d_{\max} . En las dos figuras $d_{\max} = 4$. El intervalo tiene fronteras absorbentes en 0 y $N + 1$; podemos identificar ambos sitios tomando condiciones de frontera periódicas y recuperamos un anillo simple de tamaño N .

mos que el tiempo de primer encuentro de dos caminantes es equivalente al tiempo de primer paso por el origen, de un solo caminante.

En toda la discusión anterior, obviamos por simplicidad un punto importante. De la manera en que definimos las caminatas en las simulaciones de MFET, en cada paso de tiempo ocurren dos brincos. Ésto implica que en la caminata equivalente en el espacio de distancia que presentamos, el resultado es dos veces más grande del que debería de ser según nuestra definición. Puesto de otra manera, el resultado que se obtiene para el primer encuentro es la mitad de lo que se obtiene para una caminata de primer paso por el origen.

Encontramos, entonces, para el anillo simple:

$$\begin{aligned}\tau_E(i, m) &= \frac{1}{2}\mu_{i,m} \\ \tau_E(i, m) &= -\frac{1}{2}(m-i)^2 + \frac{N}{2}(m-i), \quad m \geq i.\end{aligned}\quad (3.8)$$

3.1.2. El anillo general con $k > 1$

Si el número de vecinos cercanos es $k > 1$, la estrategia que utilizamos para resolver (3.1)-tomar el lado izquierdo de la ecuación como la aproximación discreta de la segunda derivada—ya no funciona. De acuerdo con Pandit y Amritkar [18], el anillo general que tiene $k > 1$ no tiene solución analítica (aunque ellos proponen que escala de una cierta forma y encuentran un ajuste relativamente bueno con los datos numéricos). Sin embargo, el MFPT y el MFET están relacionados de la misma manera que en el caso del anillo simple, esto es, $\tau_E(i, m) = \frac{1}{2} \tau_P(d(i, m)) - 0$.

Para demostrar esta expresión, no vamos a utilizar el argumento de la distancia geodésica que funcionó para el anillo simple. Para k general

$$d_{\max} = \left\lceil \frac{N-1}{2k} \right\rceil, \quad (3.9)$$

y el espacio de distancia es cada vez más pequeño cuando k crece, por lo que no es tan fácil recuperar una caminata en un espacio de tamaño N .

Podemos recurrir a un argumento más sencillo y más poderoso (que por supuesto aplica para el anillo simple, pero era importante introducir el argumento del espacio de distancia que utilizaremos más tarde). Cuando simulamos dos caminantes, por la simetría del anillo, podemos escoger

a uno de ellos y mantenerlo fijo, pensando que los dos brinco en cada paso de tiempo los hace el otro caminante. Podemos pensar esta versión de la siguiente manera: en la caminata normal, siempre que el caminante que escogimos es el que brinca, después de que brinca rotamos al sistema para que recupere la posición que ocupaba antes. Así el caminante que escogimos no cambia de posición, y aparentemente es el otro el que se mueve. Si fijamos al primer caminante en el origen, es evidente que el segundo está haciendo una caminata de primer paso que termina cuando llega al origen, con la acotación de que en cada paso de tiempo se mueve dos veces y por lo tanto los valores que encontramos para MFET son la mitad de los de MFPT.

3.2. Métodos numéricos

Para los anillos que tiene $k > 1$ encontramos la relación entre el MFET y el MFPT, pero no tenemos una solución analítica para ninguno de ellos. Aquí, y en general para cualquier red que no sea el anillo simple, necesitamos métodos numéricos para calcular los tiempos promedio. En esta sección presentamos tres métodos alternativos.

3.2.1. La simulación Monte Carlo

La simulación Monte Carlo es la base en este trabajo de la simulación numérica. Ocupa en cierto sentido el lugar del experimento en el laboratorio, y a lo largo del trabajo los otros métodos numéricos se validan al compararlos con los "datos" que se obtienen con esta simulación.

Más que un solo tipo de simulación, Monte Carlo se refiere a una extensa familia de simulaciones. La idea detrás de todas ellas es utilizar números aleatorios o pseudo-aleatorios para modelar procesos que son complicados, o que no son deterministas. Las caminatas aleatorias, como un proceso estocástico, entran en la segunda categoría.

En la computadora utilizamos números aleatorios para decidir la trayectoria de un caminante en la red de mundo pequeño. A partir de ésta, podemos "medir" directamente cuánto tiempo tardó en llegar a un sitio meta, y encontramos el tiempo de primer paso de esa caminata en particular. De la misma forma, usando dos trayectorias para dos caminantes,

encontramos el tiempo de primer encuentro para dos ejemplares de caminatas.

Los tiempos promedio de primer paso MFPT, y de primer encuentro MFET, se encuentran al calcular el promedio de un número L de corridas como las que acabamos de mencionar. Hay que decir, sin embargo, que esta simulación es costosa computacionalmente. Los tiempos de cada realización son variables aleatorias independientes, tomadas de distribuciones bien comportadas para las que existe el primer y segundo momento. Podemos calcular la varianza del promedio, y encontramos que es proporcional al inverso del número de corridas L , por lo tanto, el error relativo que se comete con las simulaciones Monte Carlo, disminuye como

$$e_{\text{rel}} \propto \frac{1}{L}.$$

Se necesita un número de corridas L muy grande para que el resultado de la simulación sea preciso: obtener una cifra significativa más requiere cien veces más corridas.

Sincronización de las caminatas en el tiempo

¿Cómo asignamos una escala de tiempo a las simulaciones? La simulación de MFPT no presenta ningún problema, sólo depende de un caminante y es natural definir que un paso de tiempo corresponde exactamente a un brinco del caminante.

El MFET que depende de dos caminatas tiene que tratarse con más cuidado. La primera idea que uno tiene es definir el tiempo como en el caso de primer encuentro, haciendo simplemente que los dos caminantes brinquen una vez en cada tiempo, pero esta idea tiene problemas. Si las caminatas se hacen simultáneamente, dos partículas en sitios vecinos pueden intercambiar sus posiciones sin encontrarse. Para que esto no suceda, los caminantes tienen que moverse uno por uno, y no pueden moverse siempre en el mismo orden porque introduciríamos una distinción entre dos caminantes que eran indistinguibles en principio. Tenemos que escoger al azar cuál de los dos caminantes es el que se mueve.

Una segunda idea es por lo tanto que un paso de tiempo sea un brinco de un caminante. Sin embargo, con esta solución la escala de tiempo se dilata de manera poco realista. Imaginemos que hacemos no dos sino diez caminatas simultáneas. En cada paso de tiempo, escogemos entre los diez

caminantes quién es el que se va a mover, cuando lo natural sería que en promedio los diez caminantes se movieran en cada paso de tiempo. La solución es seleccionar a un caminante para que se mueva, en cada tiempo, tantas veces como se tienen caminantes. De esta manera permitimos que nuestra simulación de tiempo discreto sea lo más parecida a la de tiempo continuo, donde cada caminante brincaba con una cierta tasa. En nuestro caso de dos partículas, escogemos dos veces a un caminante. Puede ocurrir que sea dos veces el mismo caminante el que se mueva, como también puede suceder que se muevan ambos caminantes.

Algoritmos

A continuación presentamos los algoritmos para llevar a cabo una realización de tiempo de primer paso o de primer encuentro.

Tiempo de primer paso La posición del caminante se guarda en la variable `nodo_actual` y el sitio meta en la variable `nodo_meta`. El algoritmo es el siguiente:

1. Inicializa el tiempo como $t = 0$.
2. Si `nodo_actual` es distinto de `nodo_meta`, pasa a 3. De lo contrario pasa a 4.
3. Mueve al caminante: escoge dentro de los vecinos de `nodo_actual` un sitio y almacénalo en lugar del valor que contenía `nodo_actual`. Incrementa el tiempo t en una unidad, vuelve al paso 2.
4. Regresa t .

Tiempo de primer encuentro Almacenamos la posición de los caminantes en dos variables: `primer_nodo` y `segundo_nodo`. El algoritmo es como sigue:

1. Inicializa el tiempo como $t = 0$.
2. Si `primer_nodo` es distinto de `segundo_nodo`, pasa a 3. De lo contrario regresa t y termina el algoritmo. (Los caminantes empezaron en el mismo sitio.)

3. Repite dos veces las siguientes instrucciones. Si el algoritmo no termina, pasa a 4.
 - a) Escoge un caminante al azar. Si escogiste al primero, actualiza la variable `primer_nodo`: escoge de entre sus vecinos a un nuevo sitio, y almacénalo en lugar de `primer_nodo`. Si escogiste al segundo, haz lo propio para `segundo_nodo`.
 - b) Si `primer_nodo` es igual a `segundo_nodo`, regresa $t + 1$ y termina el algoritmo.
4. Incrementa el tiempo t en una unidad, y vuelve al paso 3.

3.2.2. Enumeración exacta

Hacer simulaciones de tipo Monte Carlo es costoso computacionalmente. Una alternativa es utilizar el método de enumeración exacta, que correspondería al límite $L \rightarrow \infty$ de Monte Carlo. (L es el número de corridas sobre las que se promedia.) De acuerdo con Majid et al. [11], se obtiene mucha más “información” para la misma cantidad de recursos computacionales con el método de enumeración exacta que con una simulación Monte Carlo.

Un caminante

El método de enumeración exacta consiste en resolver numéricamente la ecuación maestra (2.5)

$$\mathbb{P}(X_t = i) = \frac{1}{\text{deg}(i)} \sum_{j \in N(i)} \mathbb{P}(X_{t-1} = j),$$

dada la condición inicial (2.6)

$$\mathbb{P}(X_0 = i) = \delta_{x_0, i}.$$

La condición inicial sugiere que utilicemos un vector para almacenar la distribución de probabilidad al tiempo t , cuya componente i -ésima sea la probabilidad de que el caminante ocupe el sitio i . Si la red tiene N nodos, utilizamos un vector de N componentes:

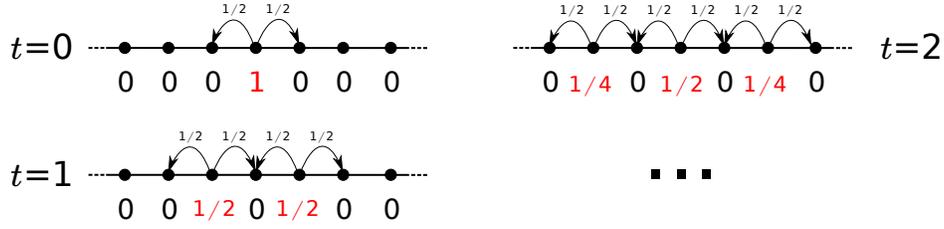


Figura 3.4: **Enumeración exacta para un caminante:** Para el anillo simple, podemos enviar en cada paso de tiempo la 1/2 de la probabilidad que hay en un sitio a cada uno de sus vecinos, a la izquierda y a la derecha.

$$\vec{p}_t \in \mathbb{R}^N, \quad (p_t)_i = \mathbb{P}(X_t = i). \quad (3.10)$$

Como ejemplo, tomemos como red el anillo simple con N nodos y $2k$ vecinos cercanos. El caminante en un sitio puede brincar a sus vecinos a la izquierda o a la derecha con probabilidad 1/2 así que obtenemos la sucesión de la figura 3.4.

En la práctica es tedioso llevar a cabo los cálculos repetitivos para encontrar \vec{p}_t a partir de \vec{p}_{t-1} , pero en la computadora esto no es un problema.

El algoritmo de enumeración exacta nos permite calcular, de manera numéricamente exacta, la distribución de probabilidades hasta cualquier tiempo t arbitrariamente grande (aunque finito).

MFPT

Si lo que nos interesa son los tiempos de primer paso por el sitio meta m , la receta que acabamos de explicar no nos sirve del todo, ya que las iteraciones del vector de probabilidad contribuyen varias veces en tiempos diferentes a la fracción de probabilidad que encontramos en el sitio que nos interesa. Necesitamos una modificación para descartar la fracción que correspondería a los caminantes que ya llegaron al sitio meta.

Siguiendo a Condamin et al. [4], después de actualizar las probabilidades con el método de enumeración exacta, hacemos cero la entrada m de \vec{p}_t , y guardamos esta cantidad por separado como $\mathbb{F}(X_t = m)$. Esta es la probabilidad de primer paso por el sitio m al tiempo t , siempre que desde $t = 1$ (desde la primera iteración) hayamos hecho cero la entrada m de \vec{p}_t . Al tratar de esta manera nuestro vector, obtenemos una distribución modificada que podemos denotar como $\vec{\mathbb{P}}$ y que está definida para cada sitio

de nuestra red,

$$\tilde{\mathbb{P}}(X_t = i) = (p_t)_i. \quad (3.11)$$

Por construcción, para el sitio meta

$$\tilde{\mathbb{P}}(X_t = m) = 0 \quad \forall t \geq 0.$$

Tenemos entonces que $\tilde{\mathbb{P}}$ por sí sola ya no es una probabilidad; al sumar sobre el conjunto de sitios de la red, no encontramos uno. Sin embargo, si tomamos en cuenta la probabilidad que hemos extraído en cada paso, la normalización funciona correctamente. En efecto, se cumple

$$\sum_{i \in V(G)} \tilde{\mathbb{P}}(X_t = i) + \mathbb{F}(X_t = m) = 1 \quad \forall t \geq 0. \quad (3.12)$$

Queremos ahora obtener el tiempo promedio de primer paso. En principio necesitamos la distribución infinita de \mathbb{F} :

$$\tau_P(x_0 \quad m) = \sum_{t=0}^{\infty} t \mathbb{F}(X_t = m). \quad (3.13)$$

El problema es que aunque el algoritmo sirva para alcanzar t arbitrariamente grande, tenemos que detenernos en algún tiempo finito. Sin embargo, podemos aprovechar que la distribución tiene una cola exponencial, $\mathbb{F}(X_t = m) \sim e^{-\alpha t}$ si es t suficientemente grande [4, 23]. Utilizamos el algoritmo de enumeración exacta para calcular \mathbb{F} hasta un tiempo t_0 , y aproximamos la contribución del resto de los términos:

$$\begin{aligned} \tau_P(x_0 \quad m) &= \sum_{t=0}^{t_0-m} t \mathbb{F}(X_t = m) + \sum_{t=t_0}^{\infty} t \cdot C \exp(-\alpha(t-t_0)) \\ &= \sum_{t=0}^{t_0-1} t \mathbb{F}(X_t = m) + \mathbb{F}(X_{t_0} = m) \sum_{t=t_0}^{\infty} t \cdot \exp(-\alpha(t-t_0)). \end{aligned}$$

La constante C se encontró con $\mathbb{F}(X_{t_0} = m) = C e^{-\alpha(t_0-t_0)}$. Vamos a limpiar un poco la notación definiendo $\mathbb{F}(X_{t_0} = m) = f$. Usando la serie geométrica, encontramos

$$\tau_P(x_0 = m) = \sum_{t=0}^{t_0-1} t \mathbb{F}(X_t = m) + \frac{f t_0}{1 - e^{-\alpha}} + \frac{f e^{-\alpha}}{(1 - e^{-\alpha})^2}.$$

Si α es pequeño, podemos conservar sólo los primeros dos términos de la serie de la exponencial, y tenemos:

$$\tau_P(x_0 = m) = \sum_{t=0}^{t_0-1} t \mathbb{F}(X_t = m) + \frac{f t_0}{\alpha} + \frac{f}{\alpha^2}. \quad (3.14)$$

Encontramos α a partir de la distribución que ya calculamos. Usando la aproximación de la cola exponencial para los tiempos t_0 y $t_0 - 10$,

$$\mathbb{F}(X_{t_0} = m) - \mathbb{F}(X_{t_0-10} = m) = \mathbb{F}(X_{t_0} = m)(1 - e^{10\alpha})$$

de donde

$$\alpha = \frac{1}{10} \ln \frac{\mathbb{F}(X_{t_0-10} = m)}{\mathbb{F}(X_{t_0} = m)}. \quad (3.15)$$

Para encontrar un t_0 que nos sirva, corremos algunas pruebas y tomamos un valor lo suficientemente grande para que el resultado de tomar $t_0 + 1000$ no difiera en más de algunas cifras decimales. El error se mantiene tan pequeño como se quiera, pues es igual a la diferencia entre los dos valores.

Dos caminantes

El algoritmo que servía antes para las probabilidades de primer paso no sirve para encontrar el tiempo de primer encuentro de dos caminantes. El primer problema es que no hay un sitio meta como tal, el encuentro puede suceder en cada sitio de la red. Además, los caminantes son independientes y si tenemos sus respectivas probabilidades de ocupar el sitio i al tiempo t , la probabilidad de que se encuentren es

$$\mathbb{P}(X_i = i, Y_t = i) = \mathbb{P}(X_t = i) \cdot \mathbb{P}(Y_t = i).$$

Esto es un segundo problema porque sólo una fracción de la probabilidad de ocupación de un caminante contribuye a un encuentro. Si hacemos una

entrada de su vector igual a cero quitaríamos más probabilidad de la que contribuyó al encuentro.

Para resolver el problema, utilizamos la caminata aleatoria en dos dimensiones que se definió en el capítulo anterior. El problema de encontrar los tiempos de primer encuentro de los dos caminantes se reformula como encontrar los tiempos de primer paso del caminante en 2D, al conjunto E :

$$\tau_E(i, m) = \tau_P(\vec{X}_0 \in E).$$

En \mathbb{Z}^2 los sitios del conjunto E son sitios meta de la caminata, y así tratamos el problema de la misma manera que las probabilidades de primer paso de la sección anterior. Como antes, definimos las probabilidades de primer paso por este conjunto,

$$\mathbb{F}(\vec{X}_t \in E),$$

y utilizamos el algoritmo que ya conocemos donde absorbemos probabilidades.

En lugar de los vectores \vec{p}_t , almacenamos las probabilidades en matrices

$$\mathbf{p}_t \in \mathbf{M}_{n \times n}, \quad (p_t)_{i,j} = \mathbb{P}(X_t = i)\mathbb{P}(Y_t = j) = \mathbb{P}(\vec{X}_t = (i, j)). \quad (3.16)$$

En cada paso de tiempo, después de actualizar a la matriz \mathbf{p}_t (esto es mandar la fracción de probabilidad que corresponde de un sitio a sus vecinos), almacenamos la probabilidad que llegó a la diagonal y que contribuyó al primer encuentro:

$$\mathbb{F}(\vec{X}_t \in E) = \sum_{i=0}^{N-1} (p_t)_{i,i}.$$

Después igualamos estas mismas entradas a cero. Obtenemos una distribución modificada

$$\tilde{\mathbb{P}}(\vec{X}_t = (i, j)) = (p_t)_{i,j}. \quad (3.17)$$

De manera equivalente a (3.12), la suma de $\tilde{\mathbb{P}}$ sobre todos los sitios en \mathbb{Z}^2 más la probabilidad de primer encuentro es igual a uno.

Avanzando paso a paso, obtenemos \mathbf{p}_t a partir de \mathbf{p}_{t-1} . Encontramos el tiempo promedio, de la misma forma que el MFPT, a partir de la distribución de probabilidades y una aproximación para la cola exponencial. La ecuación equivalente a (3.14) sólo necesita indicar que el primer paso es por el conjunto E de los sitios de la diagonal:

$$\tau_P(\vec{X}_0 \in E) = \sum_{t=0}^{t_0-1} t \mathbb{F}(\vec{X}_t \in E) + \frac{\mathbb{F}(\vec{X}_{t_0} \in E) t_0}{\alpha} + \frac{\mathbb{F}(\vec{X}_{t_0} \in E)}{\alpha^2}. \quad (3.18)$$

Ejemplo aquí.

Para que el método de enumeración exacta sea compatible con la simulación Monte Carlo, definimos los vecinos en 2D como

$$N(i, m) = \{(i, r) : r \in N(m)\} \cup \{(r, m) : r \in N(i)\}. \quad (3.19)$$

Recordemos que en la simulación Monte Carlo, en un paso de tiempo los caminantes se mueven dos veces. Sin embargo, hemos usado aquí que \mathbf{p}_t se encuentra a partir de \mathbf{p}_{t-1} utilizando la vecindad (3.19), donde está implícito que sólo se mueve un caminante en cada paso de tiempo. (Recordemos que si $p = 0$, ésta es la vecindad de Von Neuman, y en esta vecindad es fácil verlo). Así, los tiempos que se obtienen con el método de enumeración exacta que describimos son el doble de los tiempos de Monte Carlo. Al implementar el método en la computadora, es fácil resolver este problema llevando a cabo dos iteraciones del algoritmo en cada paso de tiempo. Cada uno de los dos valores de p que se absorben en la diagonal se guarda, y en un paso de tiempo se regresa la suma. De esta manera, las dos simulaciones se hacen equivalentes.

3.2.3. Método matricial

En este trabajo, nos interesa sobre todo calcular MFPT y MFET. Monte Carlo y el método de enumeración exacta nos permiten encontrar un promedio, después de haber obtenido una parte de la distribución de los tiempos de primer paso o de primer encuentro. En ese sentido estos dos métodos hacen más de lo que requerimos; podemos encontrar directamente los tiempos promedio. Escribimos un sistema lineal de ecuaciones para

todos los tiempos promedio (de primer paso o primer encuentro) y encontramos la solución de forma numéricamente exacta. Esta solución se llama método matricial [19].

MFPT al origen

Partimos de la ecuación de recurrencia para un sitio meta fijo (2.14), que se encontró en el capítulo 2. Nos interesan en particular los tiempos de primer paso por el origen, así que tomamos $m = 0$. Reacomodamos (2.14) para obtener

$$\mu_{i,0} - \frac{1}{\deg(i)} \sum_{j \in N(i)} \mu_{j,0} = 1. \quad (3.20)$$

Evidentemente,

$$\mu_{0,0} = 0. \quad (3.21)$$

Estas dos ecuaciones definen un sistema lineal de ecuaciones:

$$\mathbf{M} \cdot \vec{\mu}_{i,0} = \vec{c}, \quad (3.22)$$

donde \vec{c} es un vector constante y \mathbf{M} es la matriz de coeficiente de los tiempos promedio. \vec{c} tiene todas sus entrada iguales a uno, excepto la primera, c_0 , que se anula porque el origen es el blanco de la caminata. \mathbf{M} tiene unos en toda la diagonal y $M_{i,j} = -1/\deg(i) \quad \forall j \in N(i)$.

La solución es el vector $\vec{\mu}_{i,0} = (\mu_0, \mu_1, \dots, \mu_{N-1})$, dado por

$$\vec{\mu}_{i,0} = \mathbf{M}^{-1} \cdot \vec{c}. \quad (3.23)$$

La solución $\vec{\mu}_{i,0}$ se encuentra numéricamente.

Es conveniente definir a \mathbf{M} como una matriz dispersa (*sparse*), porque la mayoría de sus entradas son iguales a cero. El lenguaje Julia, en el que se programó toda la investigación de este trabajo, tiene implementada una función especializada para resolver ecuaciones del tipo de (3.22), y que utiliza diferentes algoritmos según la estructura de \mathbf{M} .

MFPT desde el origen

Nos interesa el tiempo de primer paso cuando la condición inicial está fija y hay múltiples sitios meta. En particular, queremos que los caminantes empiecen en el origen: fijamos $i = 0$ en la expresión de $\mu_{i,m}$. Podemos reordenar (2.15) para obtener:

$$\mu_{0,m} - \sum_{j \in N(m)} \frac{\mu_{0,j}}{\deg(j)} = \sum_{j \in N(m)} \frac{1}{\deg(j)}. \quad (3.24)$$

También en este caso vale (3.21), $\mu_{0,0} = 0$, y escribimos al sistema de ecuaciones en su forma vectorial,

$$\mathbf{M} \cdot \vec{\mu}_{0,m} = \vec{c}, \quad (3.25)$$

La primer entrada de \vec{c} es cero, y para el resto de las entradas, $c_m = \sum_{j \in N(m)} 1/\deg(j)$. Notemos que esta suma, en general, no es igual a uno. La matriz \mathbf{M} tiene unos en la diagonal, y para el resto de sus entradas, $M_{m,j} = -1/\deg(j) \quad \forall j \in N(m)$.

MFET

Partimos de la ecuación de recurrencia (2.16), que reacomodamos de la siguiente manera:

$$v_{i,m} - \frac{1}{\deg(i,m)} \sum_{j \in N(i,m)} v_{j,m} = 1.$$

Queremos escribir una ecuación de la forma de (3.22) para los tiempos promedio de primer encuentro, y para conseguirlo necesitamos ordenar las N^2 condiciones iniciales $\vec{X}_0 = (i, m) \in \mathbb{Z}^2$, de manera que tengamos un vector. Usamos el mapeo

$$s = (i + 1) + Nm, \quad (3.26)$$

que es especialmente útil para trabajar con el sistema de índices de Julia. Aplicándolo a los sitios de la ecuación anterior, encontramos

$$v_s - \frac{1}{\deg(s)} \sum_{j \in N(s)} v_j = 1. \quad (3.27)$$

Como hicimos antes para primer paso, reescribimos esta ecuación en forma vectorial: queremos encontrar \vec{v} , solución de

$$\mathbf{M} \cdot \vec{v} = \vec{c}. \quad (3.28)$$

Veamos qué forma tienen la matriz \mathbf{M} y el vector constante \vec{c} . Para el MFET, todos los sitios de la diagonal, es decir del conjunto E , son los blancos de la caminata. Estos sitios tienen $i = m$ y por tanto satisfacen $s_E = (N + 1)i + 1$. Para los blancos, $v_{s_E} = 0 \forall s_E$. Las entradas del vector constante para estos sitios son cero, $c_{s_E} = 0$.

Por (3.27), el resto de las entradas del vector \vec{c} son iguales a uno. La matriz \mathbf{M} tiene la misma forma que antes, unos en la diagonal y $-1/\text{deg}(s)$ en las entradas que corresponden a los vecinos del sitio s .

De particular interés son los tiempos de primer encuentro donde un caminante comienza siempre en el origen. (Esto es para comparar con el tiempo de primer paso por el origen). Recordando que los tiempos de primer encuentro son simétricos y se tiene $\tau_E(0, m) = \tau_E(m, 0)$, nos interesan especialmente las primeras N entradas de \vec{v} .

Falta decir unas palabras sobre el conjunto de vecinos, $N(s) = N(i, m)$. Para implementar el método matricial, utilizamos el mismo conjunto que en enumeración exacta (3.19). Por lo tanto, heredamos de este conjunto que sólo se mueve un caminante por paso de tiempo, y los tiempos de encuentro del método matricial que describimos no son los mismos que los de Monte Carlo. No podemos utilizar la solución que funcionó para enumeración exacta, porque no estamos moviendo a los caminantes, y tampoco es buena idea utilizar otro conjunto de vecinos (uno que incorpore los sitios posibles después de dos movimientos) porque es más probable encontrar al caminante en algunos sitios de la vecindad que en otros, y se dificulta escribir la ecuación de recurrencia (3.27).

Nos conformamos entonces con que los resultados del método matricial correspondan a una simulación en que sólo se hace un movimiento en cada paso de tiempo, y al final dividimos los resultados entre dos. Al tomar $\frac{1}{2}\vec{v}$ como el resultado final, obtenemos tres métodos que son equivalentes.

Para terminar, hay que hacer notar que para una red con N nodos, el problema se traduce a dos dimensiones con N^2 sitios. Después se construye la matriz \mathbf{M} de tamaño $N^2 \times N^2 = N^4$, así que \mathbf{M} crece mucho con el número de nodos de la red. De este comportamiento queda claro que, en especial para el tiempo de encuentro, es sumamente importante utilizar

matrices dispersas.

3.2.4. Ventajas y desventajas de los métodos

Al desarrollar nuestro trabajo, fuimos construyendo uno a uno los métodos numéricos. Es ciertamente muy útil tener diferentes soluciones al mismo problema, porque cada método tiene aplicaciones y limitaciones diferentes.

La primer herramienta de la que dispusimos fue Monte Carlo, y la mayoría del código que desarrollamos se basa en las primeras ideas que tuvimos con este método. Monte Carlo tiene la gran ventaja de ser completamente general y fácil de implementar, pero tiene también el inconveniente de ser costoso en términos de recursos computacionales.

El método de enumeración exacta es a su manera costoso computacionalmente, pero el resultado que brinda es preciso; ésta es su principal ventaja sobre Monte Carlo. Sin embargo, tiene un problema con la paridad por la aproximación que se utiliza para la cola exponencial. Expliquémonos: en un anillo simple con un número par de nodos, un caminante que se mueve en cada paso sólo puede acceder a un sitio meta en un número par o impar de pasos. Este comportamiento se describe en la figura 3.5. Esto no es un problema para calcular el tiempo promedio de primer paso, los tiempos pares (impares) contribuyen al promedio, los impares (pares) no, y la suma infinita del promedio converge. Pero, en enumeración exacta utilizamos una aproximación para la cola exponencial que supone que todos los valores se sitúan sobre la cola.

El método matricial, aunque tiene una utilidad diferente porque no sirve para encontrar la distribución de tiempos, es muy eficiente en términos de recursos, cuando sólo nos interesan los tiempos promedio. No obstante, su escalamiento como N^4 en el caso de primer encuentro lo hace inutilizable si el sistema es grande. Empíricamente, el límite para este método en el caso de MFET es poco más de $N = 200$ nodos (aunque esta cifra depende del número de vecinos de cada sitio, y por lo tanto de los parámetros k y p). Sistemas más grandes pueden congelar la computadora en la que se realizan los cálculos. Para calcular el MFPT, el método matricial se puede utilizar para sistemas más grandes, no tiene esta limitación.

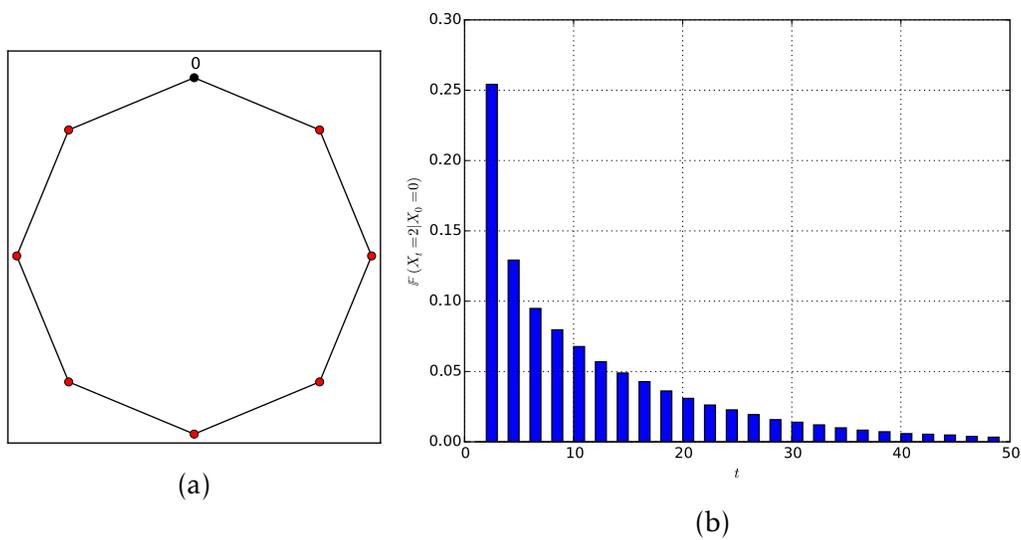


Figura 3.5: **Tiempos prohibidos en un anillo simple:** Por efectos de paridad de la caminata, los tiempos de primer paso a un sitio sólo pueden ser pares o impares: (a) anillo simple conformado por $N = 8$ nodos y (b) distribución de los tiempos de primer paso de una caminata que va del origen al sitio 2; es imposible que la caminata llegue al sitio meta en un tiempo impar.

Parte II

Validación y resultados

Capítulo 4

Validación de los métodos numéricos

En este capítulo mostramos la validación de los métodos numéricos que nos sirven para encontrar los tiempos promedio de primer paso y primer encuentro. La simulación Monte Carlo juega en este trabajo el papel de los “datos experimentales”; los resultados de los otros métodos se compararan con los resultados de Monte Carlo para validarlos. Además comparamos los resultados numéricos con la solución analítica del anillo simple.

4.1. Distribución de los tiempos de primer paso y de primer encuentro

Tanto Monte Carlo como el método de enumeración exacta sirven para encontrar la distribución (o al menos una parte de la distribución) de los tiempos de primer paso y primer encuentro. Monte Carlo lo hace indirectamente: para un número grande de repeticiones calculamos la frecuencia con que aparece cada tiempo. Enumeración exacta da por su parte un resultado que debería ser exacto. En la figura 4.1 comparamos los resultados de uno y otro método en el caso de **primer paso**: graficamos el histograma de los tiempos que se obtienen con Monte Carlo, y la distribución de probabilidades $\mathbb{F}(X_t = 0)$ de enumeración exacta, para la misma red y la misma condición inicial. El acuerdo entre los dos métodos es muy bueno: la curva limpia de enumeración exacta siempre se encuentra en medio de

las fluctuaciones de Monte Carlo.

Comparamos los resultados para **primer encuentro** en la figura 4.2. Ambos métodos vuelven a dar un buen acuerdo. Llamamos la atención del lector sobre el hecho de que tanto en el caso de FPT como de FET, en la figura (b), para la que se utiliza una escala logarítmica en el eje de las ordenadas, las curvas acaban en una línea recta. Ésta es la cola exponencial de la distribución, y es lo que nos permite calcular el tiempo promedio con el método de enumeración exacta, utilizando la aproximación de Condamin et al. [4].

4.2. Tiempos promedio

Los tres métodos numéricos se utilizan para calcular los valores esperados MFPT y MFET. Calculamos los tiempos promedio, tanto de primer paso como de primer encuentro, usando la red de la figura 4.3 como soporte de las caminatas.

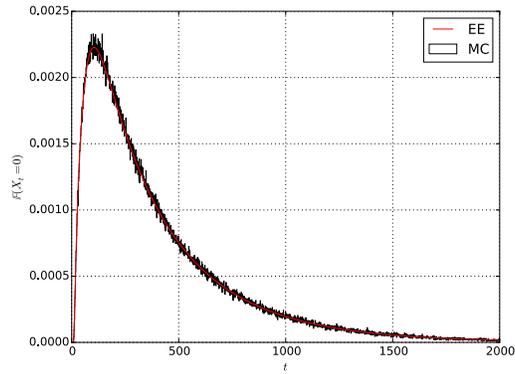
Las simulaciones de MFPT tienen una condición inicial y un sitio meta, esto es, tienen dos grados de libertad. Las de MFET tienen dos grados de libertad también, por la condición inicial de los dos caminantes. Para facilitar la comparación de los tiempos que encontremos, fijamos alguno de los grados de libertad.

Los resultados del método matricial, enumeración exacta y Monte Carlo se comparan en la figura 4.4 para el MFPT, cuando el sitio meta se fija como el origen. Calculamos, por lo tanto, $\tau_p(i, 0)$ como función de i . La comparación para el MFET se encuentra en la figura 4.5; un caminante comienza siempre en el origen y el otro empieza en el sitio i . De esta manera calculamos $\tau_E(0, i)$ como función de i .

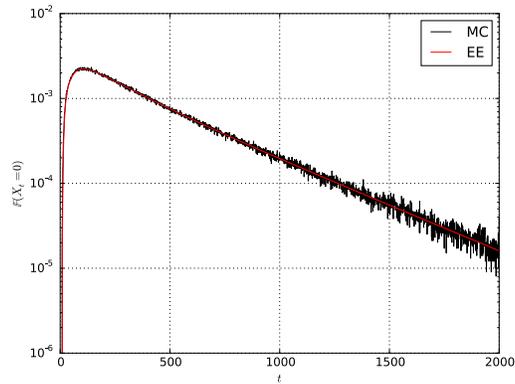
Los valores que se obtienen con el método matricial y con enumeración exacta se parecen tanto que es difícil distinguirlos en la misma gráfica. Por esta razón calculamos el error relativo de Monte Carlo y enumeración exacta respecto al método matricial, en las figuras 4.4b y 4.5b.

Los tres métodos tienen un muy buen acuerdo cuando se calcula primer paso: en la figura 4.4b constatamos que las fluctuaciones de Monte Carlo están centradas alrededor de los resultados del método matricial; los resultados de enumeración exacta, por su parte, presentan una oscilación interesante, que también está centrada.

Para las simulaciones de primer encuentro, los resultados de Monte

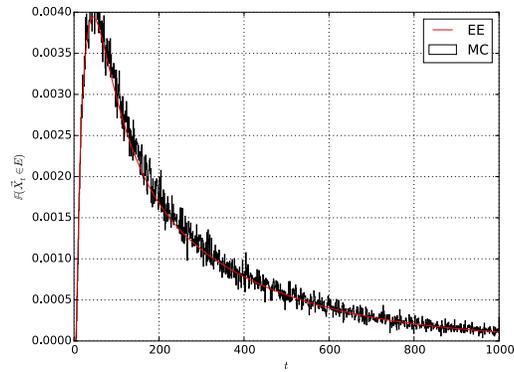


(a)

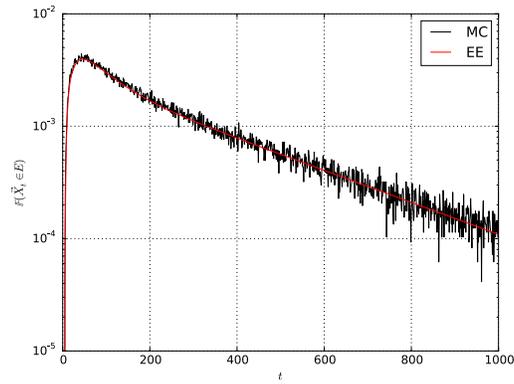


(b)

Figura 4.1: Comparación de la distribución de probabilidades de primer paso, usando Monte Carlo con $N = 10^6$ corridas (negro) y enumeración exacta (rojo): (a) escala natural y (b) escala logarítmica. Las simulaciones se hacen en la misma red, formada por $N = 100$ nodos, $2k = 4$ vecinos cercanos y densidad de atajos $p = 0.1$, y se utiliza la misma condición inicial: e caminante empieza en $x_0 = 33$ y la caminata termina cuando llega al origen.



(a)



(b)

Figura 4.2: Comparación de la distribución de probabilidades de primer encuentro, usando Monte Carlo con $N = 10^5$ corridas (negro) y enumeración exacta (rojo): (a) escala natural y (b) escala logarítmica. Las simulaciones se hacen sobre la red que se mencionó en la fig. 4.1; un caminantes empieza en el sitio $i = 33$ y el otro en el origen.

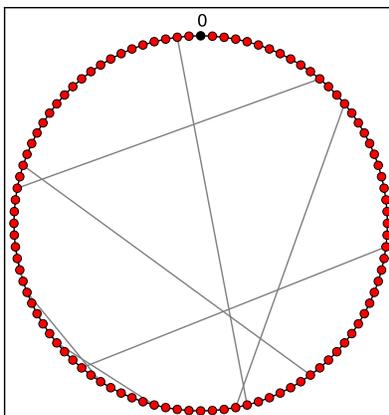


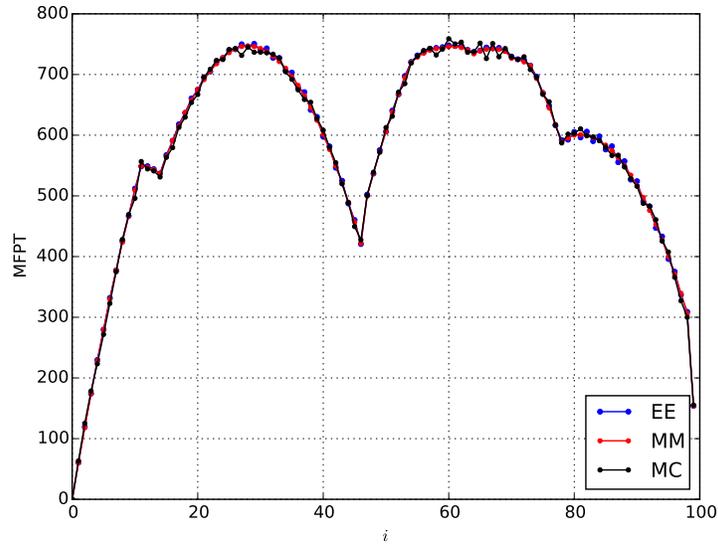
Figura 4.3: Los tiempos promedio se calculan para esta red, que tiene $N = 100$ nodos, $2k = 2$ vecinos cercanos y densidad de atajos $p = 0.1$.

Carlo están consistentemente abajo de los de los otros métodos. Concluimos que tenemos una fuente de errores sistemáticos al calcular el MFET, y sugerimos que puede encontrarse en la definición de la vecindad de un sitio, $N(i)$. Hemos visto que esta vecindad define camitas donde sólo se mueve un caminata en cada paso de tiempo, y que tenemos que arreglar la escala temporal de los métodos de enumeración exacta y matricial con artilugios.

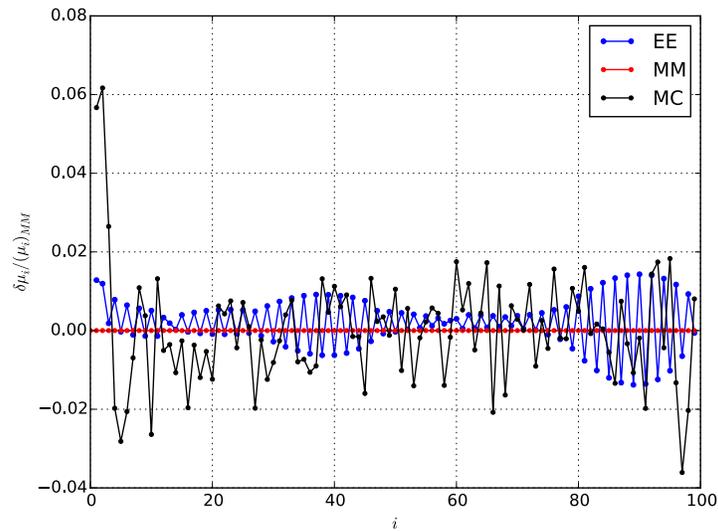
Por su parte, los resultados de enumeración exacta son muy cercanos a los del método matricial, mucho más que los de Monte Carlo, pero se encuentran consistentemente por arriba de los valores de referencia. En este caso tenemos otro tipo de error sistemático, que no se puede explicar por el conjunto $N(i)$ ya que ambos métodos lo utilizan. El error sistemático proviene de la aproximación de la cola exponencial en el método de enumeración exacta. Recordemos que la distribución que se obtiene con este método es numéricamente exacta, pero para encontrar el valor esperado de la distribución cometemos necesariamente un error, que resulta ser sistemático.

4.3. El anillo simple

Verificamos a continuación que las relaciones entre el MFET y el MFPT en los anillos del capítulo 3 son válidas.

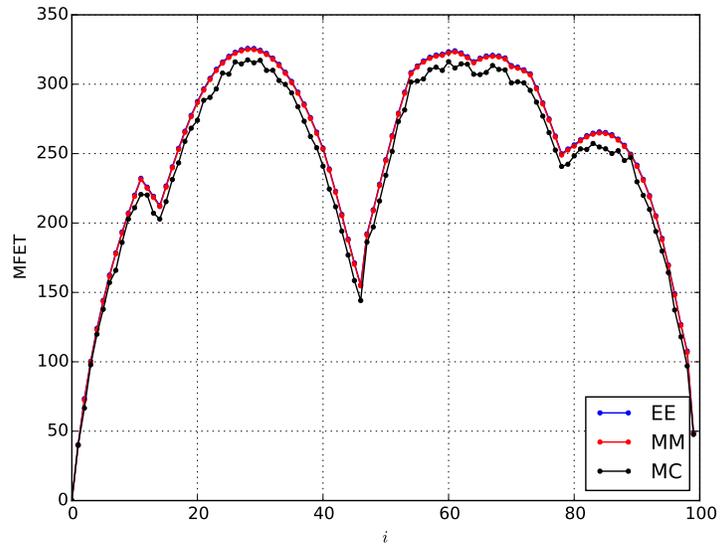


(a)

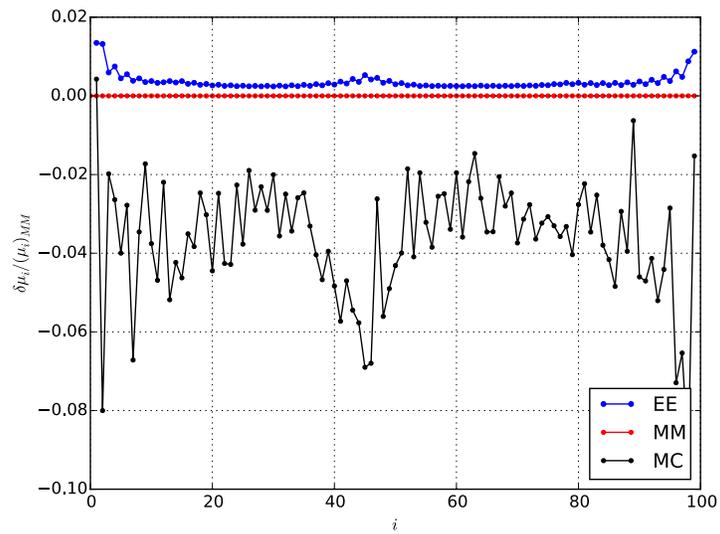


(b)

Figura 4.4: (a) Comparación de los tiempos promedio de **primer paso** (MFPT) usando los tres métodos y (b) error relativo respecto a los resultados del método matricial. En las leyendas “MM” se refiere al método matricial, “EE” es enumeración exacta y “MC” es Monte Carlo. Los resultados de los primeros dos son tan parecidos que casi no se puede ver la curva verde por debajo de la roja. Se utiliza la red de la fig. 4.3. El caminante empieza en el sitio i y la caminata termina al llegar al origen.



(a)



(b)

Figura 4.5: (a) Comparación de los tiempos promedio de **primer encuentro** (MFET) usando los tres métodos y (b) error relativo respecto a los resultados del método matricial. Se utiliza la red de la fig. 4.3. Un caminate comienza en el origen y el otro en el sitio i .

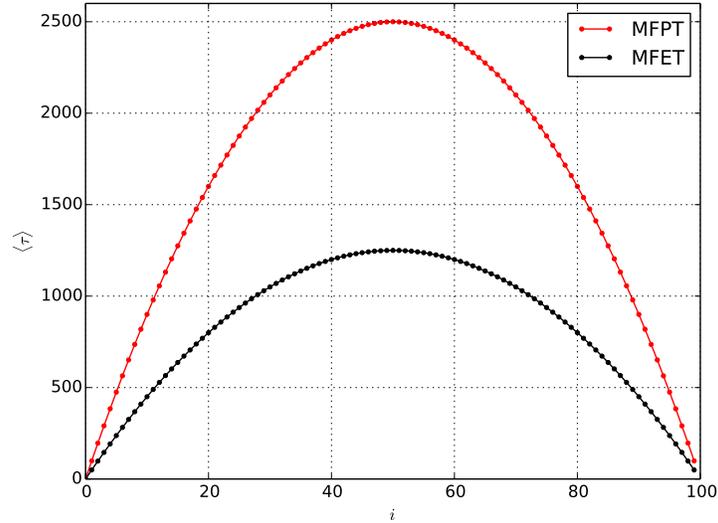


Figura 4.6: MFET y MFPT para un anillo simple con $N = 100$ nodos. Se utilizó el método matricial para obtener los resultados.

Numéricamente, el resultado (3.3) se verifica en la figura 4.6. Al tomar el cociente punto a punto, encontramos numéricamente que τ_E es exactamente la mitad de τ_P .

Además, ajustamos un polinomio de grado dos a los datos del método matricial, para el tiempo de primer paso. Encontramos:

$$\tau_p(0, i) = -1.0i^2 + 100.0i - 5.90519 \times 10^{-13}, \quad i \in V(G).$$

De esta expresión es claro que, dentro del error de redondeo numérico, encontramos exactamente los coeficientes -1 , $-N$ y 0 para los términos cuadrático, lineal, y constante respectivamente, predichos en la solución analítica (3.3).

4.3.1. Conclusión de la validación

La correspondencia entre las distribuciones de tiempos de Monte Carlo y enumeración exacta es muy buena (figs. 4.1 y 4.2) y podemos confiar que el método de enumeración exacta realmente produce una distribución numéricamene exacta.

Cuando comparamos el valor esperado, para primer paso los tres métodos dan resultados prácticamente idénticos (fig. 4.4). Sin embargo, para primer encuentro la validación nos indica que cometemos un par de errores sistemático. El error del método de enumeración exacta nos queda claro, y a pesar de que es sistemático, no nos preocupa porque se puede hacer tan chico como se quiera. Por su parte, el error sistemático de los dos métodos indirectos comparados con Monte Carlo es un poco más molesto; sólo tenemos una adivinanza de dónde se comete el error. No obstante, destacamos que no es un error alarmante: la discrepancia que se obtiene entre Monte Carlo y el método matricial ronda en promedio el 3 o 4%, y el valor máximo no pasa de 8%. Desde un punto de vista práctico, el error es suficientemente pequeño para decir que nuestros métodos funcionan correctamente.

Por todas estas razones, la validación de los resultados nos deja satisfechos.

Capítulo 5

Resultados en *una* red de mundo pequeño

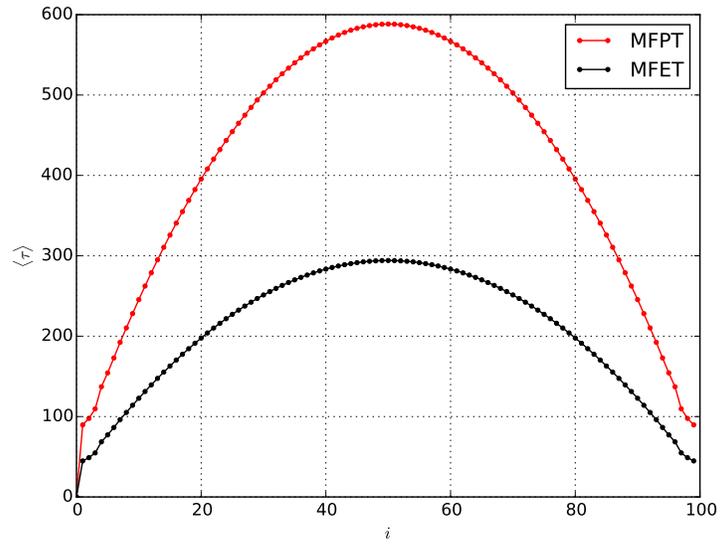
La terna de parámetros N , el número de nodos, k el número de vecinos cercanos y p , la densidad de atajos, define una familia de redes. Cada elemento de la familia tiene una configuración distinta de atajos. En este capítulo nos interesa tomar una red en particular, para comparar los tiempos de primer paso y de primer encuentro, sin preocuparnos demasiado de cuál es el efecto de escoger diferentes parámetros de red.

5.1. Anillos

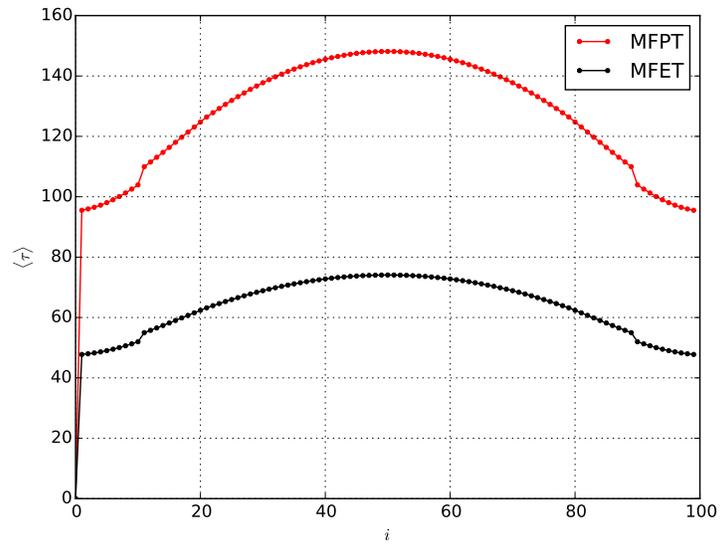
La figura 5.1 presenta una comparación del MFET y MFPT para anillos que no tienen atajos y tienen un número de vecinos cercanos de $k > 1$. Cabe mencionar que el cociente punto a punto τ_E / τ_P es exactamente 0.5, y esto valida la predicción que habíamos hecho en el capítulo 3.

Más allá de la validación, la forma de cada una de las curvas en la figura 5.1, es por sí misma interesante. Los primeros k puntos y los últimos k puntos están separados de la parte central de los datos por una discontinuidad. Esto era de esperarse, porque los primeros k vecinos de cada lado del origen están conectados directamente con éste, a diferencia del resto de los sitios de la red. Por lo tanto, debe de ser mucho más tardado llegar a los primeros sitios que ya no son vecinos del origen.

Fijemos nuestra atención sobre la curva de MFPT (curva roja) en la figura 5.1b, donde la red tiene $k = 10$ vecinos cercanos. La parte central,



(a)



(b)

Figura 5.1: MFET y MFPT para dos anillos que no son simples, y que tienen $N = 100$ nodos, $p = 0$ y (a) $k = 3$, (b) $k = 10$ vecinos cercanos.

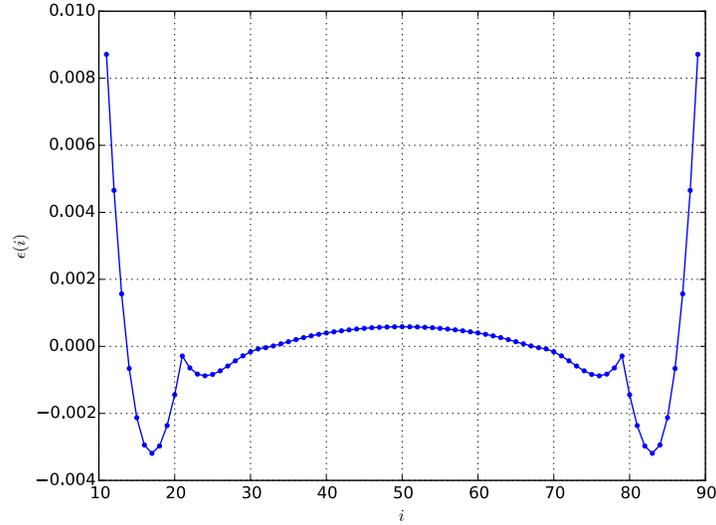


Figura 5.2: Error relativo entre el ajuste cuadrático y los datos de la parte central de la curva de MFPT de la fig. 5.1b. El ajuste es bueno cerca del centro, pero no tanto en las orillas, cerca de los primeros puntos que dejan de estar conectados con el origen.

entre las dos discontinuidades parece tener una forma parabólica. Para verificar este comportamiento, ajustamos los resultados con un polinomio de grado 2. La figura 5.2 muestra el error relativo $\epsilon(i)$ que se tiene entre el ajuste y los datos. En efecto, el comportamiento es aproximadamente cuadrático excepto para los puntos más cerca de la discontinuidad (los primeros puntos que dejan de estar directamente conectados con el origen). Sin embargo, los coeficientes que se obtienen para el ajuste ya no tienen ninguna relación con los que predice la solución del anillo simple—el coeficiente -1 para el término cuadrático y el factor N para el término lineal. En este caso en particular, se encontró

$$\tau_p(0 \quad i) \quad -0.026i^2 + 2.57i + 83.85, \quad i \in \{11, \dots, 89\}.$$

En esta ecuación se utilizó la notación del MFPT que comienza en el origen; sin embargo, debido a que no hay atajos en los anillos, la probabilidad de transición es simétrica y las dos versiones de MFPT, la que comienza en el origen y la que termina en el origen, dan el mismo resultado. Más aún, por la simetría del anillo es irrelevante qué sitio es el que nom-

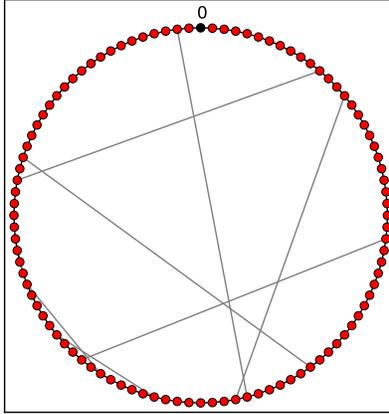


Figura 5.3: Red formada por $N = 100$ nodos, $2k = 2$ vecinos cercanos y densidad de atajos $p = 0.1$.

bramos el origen. Hubiéramos podido escribir igualmente $\tau_p(i = 0)$, o incluso mejor, $\tau_p(0, i)$ para insistir en que los dos tipos de MFPT son iguales.

Para terminar esta discusión, también vale la pena notar que el tiempo promedio para los puntos que son vecinos cercanos del origen crece con una curvatura cóncava, lo contrario a lo que sería el comportamiento en el anillo simple. También aquí era de esperarse un comportamiento muy diferente al resto de la red, porque además de estar directamente conectados con el origen, estos sitios tienen un alto agrupamiento (*clustering*) con respecto al origen, que es el centro del grupo.

La conclusión que obtenemos de estudiar los anillos generales es que a pesar de que no hay resultado analítico, el comportamiento del sistema es relativamente sencillo. Esto justifica que nos interese principalmente por las redes que tienen atajos, $p \neq 0$, y bastará concentrarnos en el caso caso $k = 1$.

5.2. Dos tipos de MFPT

Para todos los resultados de este capítulo, retomamos la red de la figura 4.3 que se utilizó para la validación de los métodos numéricos. Para futura referencia, la volvemos a incluir en este punto.

Cuando estudiamos el tiempo de primer encuentro, queremos indagar

cuál es el efecto de la estructura de la red. Queremos comparar los sitios de la red, y para hacerlo nos gustaría obtener funciones que sólo dependieran de un sitio. El problema es que tenemos dos caminantes, cada uno con su condición inicial. Lo natural para deshacernos de la libertad en una variable, es fijar una de las condiciones iniciales: uno de los caminantes comenzará siempre en el origen, y el segundo ocupará en distintas simulaciones todos los otros sitios i de la red. De esta manera estudiamos el comportamiento de $\tau_E(0, i)$ como función de i .

En el caso de primer paso, estudiamos a un único caminante, pero también aquí tenemos dos variables: la condición inicial y el sitio meta. Utilizamos la misma estrategia y fijamos una de las dos variables; hacemos que el sitio meta sea el origen o al contrario, que sea la condición inicial, de todas las simulaciones.

Una de las consecuencias de que la probabilidad de transición entre dos sitios no es en general simétrica, $W_{i,j} \neq W_{j,i}$, es que el tiempo promedio de primer paso de un sitio i a un sitio m no es igual al de m a i . Este comportamiento está plasmado en la ecuación (2.8) en el capítulo 2. Por dicho comportamiento nos interesa comparar las dos opciones que tenemos para fijar una de las variables, queremos ver qué tan diferentes son los valores que se obtienen cuando el origen es el sitio meta o es la condición inicial.

Esto se encuentra en la figura 5.4. Para futura referencia, los nodos de la red que están conectados con atajos están coloreados en rojo. Es interesante que $\tau_p(0 \rightarrow i)$ presenta más variaciones que $\tau_p(i \rightarrow 0)$. Podemos referirnos a las ecuaciones (2.14) y (2.15) para entender este fenómeno. Supongamos que i es un sitio conectado con un atajo. Para el primer paso que llega al origen $\tau_p(i \rightarrow 0)$, el atajo sólo afecta directamente a i , no a sus vecinos. En efecto, el factor $1/\text{deg}(i)$ sólo aparece en la ecuación del sitio i . Sin embargo, para el primer paso que comienza en el origen $\tau_p(0 \rightarrow i)$, el atajo también afecta directamente a los vecinos de i ya que el mismo factor $1/\text{deg}(i)$ aparece en la ecuación de todos los sitios vecinos. Por esta razón el segundo caso es más sensible a la presencia de atajos, y observamos variaciones más importantes cerca de los puntos en rojo.

Además de que el primer paso desde el origen es mucho más sensible a la presencia de atajos que el MFPT al origen, en la comparación de los dos tipos de MFPT (fig. 5.4) parecería que en ambos casos, entre dos puntos consecutivos conectados por atajos (puntos rojos), los tiempos se ordenan en parábolas.

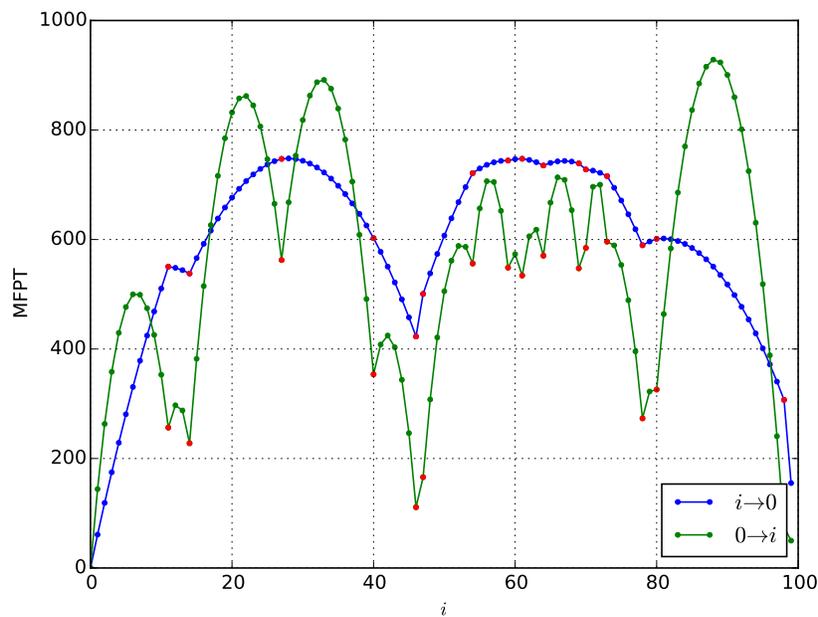


Figura 5.4: Comparación de $\tau_p(i \rightarrow 0)$ y $\tau_p(0 \rightarrow i)$, utilizando el método de enumeración exacta. Se utiliza la red de la fig. 4.3, que tiene $N = 100$ nodos, $2k = 2$ vecinos cercanos y densidad de atajos $p = 0.1$. Las curvas se obtuvieron con enumeración exacta.

5.3. Arcos de anillo entre dos atajos

Podemos verificar si es cierto que los tiempos se acomodan en parábolas, al tomar el caso más sencillo posible: un anillo simple que tiene solamente un atajo. Para esta red, el número de vecinos cercanos es $k = 1$, y entre los nodos conectados por el atajo, tenemos un arco de un anillo simple. Esperaríamos entonces que entre los sitios conectados por el atajo, el tiempo promedio de primer paso satisficiera la solución que encontramos para el anillo (3.2). Éste debe ser el caso sólo del MFPT al origen, no del MFPT desde el origen, $\tau_p(0 \rightarrow i)$. En efecto, Las ecuaciones que satisfacen este segundo no son las del anillo simple, porque los vecinos de un atajo j tienen un término $-1/\text{deg}(j)$.

5.3.1. Ajuste de los tiempos

Tenemos un ejemplo de una red con $k = 1$ vecinos cercanos y un único atajo en la figura 5.5a. A su derecha, la figura 5.5b es la curva de primer paso por el origen, $\tau_p(i \rightarrow 0)$. El atajo conecta los nodos 55 y 88 (que coloreamos en negro). En líneas punteadas graficamos las parábolas que se obtienen del ajuste de los tiempos de cada sección.

Los términos lineales del ajuste se encuentran en la tabla 5.1. De acuerdo con (3.3), en un anillo simple este término es el número de nodos N sobre el que tiene lugar la caminata. Entre el origen y el primer atajo hay 56 sitios (contando los extremos). Entre los dos atajos hay 34 nodos, y entre el segundo atajo y el origen hay 14 sitios. Por lo tanto, el número de nodos N de cada ajuste no tiene nada que ver con el número de sitios que realmente conforman cada arco. Tenemos que interpretar al término lineal N del ajuste como el número *efectivo* de sitios en cada arco: se trata del número de nodos que tendría que tener un anillo simple para producir las secciones de parábola que encontramos entre los atajos de la red 5.5a.

Intuitivamente, N debe depender de la posición de los atajos y de los largos que sean estos, de cuántos sitios “brinquen”.

5.3.2. Resultados semi-analíticos para los arcos de anillo

Consideremos una red que tiene N nodos, $k = 1$ vecinos cercanos, y densidad de atajo p . Supongamos que dicha red tiene r sitios a los que llega uno o más atajos, y que se trata de los nodos a_1, a_2, \dots, a_r . Sin pérdida de

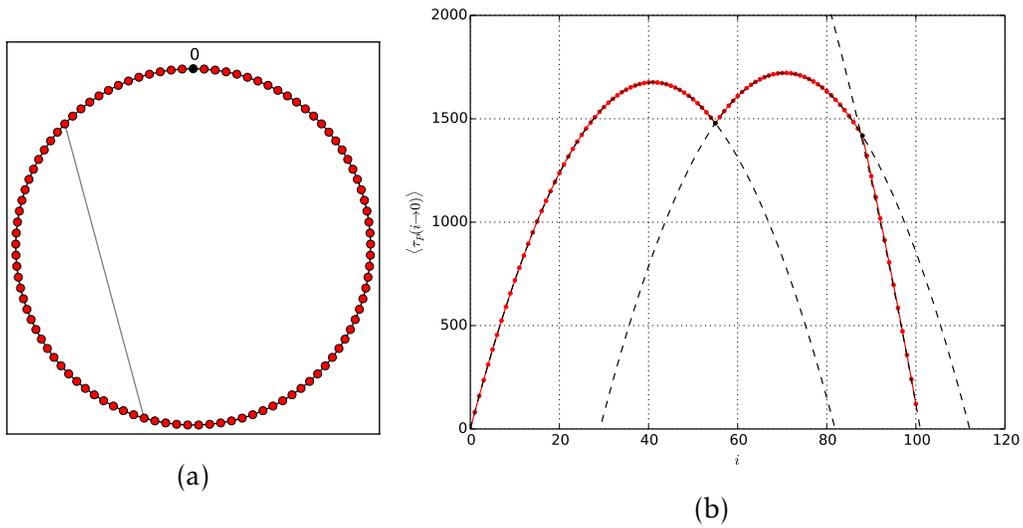


Figura 5.5: Red que tiene solamente un atajo. (a) La red conformada por $N = 101$ nodos, con $k = 1$ vecinos cercanos y densidad de atajos $p = 0.01$. (b) MFPT al origen de esta red, junto con la parábolas que se obtienen del ajuste de los puntos. En la tabla 5.1 podemos encontrar el coeficiente lineal del ajuste.

	N
$i \in \{0, \dots, 55\}$	81.9
$i \in \{55, \dots, 88\}$	31.2
$i \in \{88, \dots, 100\}$	-96.1

Tabla 5.1: Coeficiente lineal del ajuste de las parábolas de la fig. 5.5b. De acuerdo con (3.4), este término es el número de nodos N de un anillo.

generalidad, supongamos también que el origen no tiene atajos, ésto es, que $a_1 > 0$. El resultado de la sección anterior es que entre éstos sitios, los MFPT al origen se acomodan en parábolas que satisfacen la expresión (3.2), que repetimos aquí:

$$\mu_{i,m} = -(m-i)^2 + N(m-i)$$

Por lo tanto, si conocemos el valor del MFPT al origen en los sitios a_1, \dots, a_r , los tiempos de las parábolas deberían de quedar determinados de manera única. Denotemos a los tiempos en los sitios conectados por atajos por

$$\mu_{a_s,0} = \tau_s.$$

Para un nodo en el s -ésimo arco, $a_s \quad i \quad a_{s+1}$, se tiene que cumplir:

$$\mu_{i,0} = -(i-a_s)^2 + N_s(i-a_s) + \tau_s. \quad (5.1)$$

En esta ecuación queda por determinar el número de atajos *efectivo*, N_s . Lo obtenemos de la condición de frontera de la parábola en su extremo derecho: tenemos que recuperar el tiempo del siguiente atajo, $\mu_{a_{s+1},0} = \tau_{s+1}$. Por lo tanto

$$\begin{aligned} -(a_{s+1}-a_s)^2 + N_s(a_{s+1}-a_s) + \tau_s &= \tau_{s+1} \\ N_s &= \frac{(\tau_{s+1}-\tau_s) + (a_{s+1}-a_s)^2}{a_{s+1}-a_s} \end{aligned} \quad (5.2)$$

Esta expresión nos da la dependencia de N . Nuestra intuición era correcta y N_s depende de la distancia $(a_{s+1}-a_s)$ entre atajos consecutivos, y también de la diferencia de alturas $(\tau_{s+1}-\tau_s)$.

Ahora, el problema se reduce a encontrar los tiempos de los atajos, τ_1, \dots, τ_r , y todo el sistema queda determinado. Pero nos hace falta un dato más. Hasta este punto hemos tratado igual a todos los sitios a_1, \dots, a_r que tienen un atajo, pero en realidad estos sitios están apareados de dos en dos debido a que están conectados por $q = r/2$ atajos. Denotamos a la pareja (el vecino vía un atajo) de a_s como a_{p_s} , donde $p_s \in \{1, \dots, r\} \setminus \{s\}$.

Escribimos entonces la ecuación de recurrencia (2.14) para el sitio a_s :

$$\tau_s = 1 + \frac{1}{\deg(a_s)} (\tau_{a_{s-1}} + \tau_{a_{s+1}} + \tau_{p_s}).$$

Los vecinos a la izquierda y a la derecha de a_s pertenecen a dos parábolas donde los tiempos están descritos por (5.1),

$$\deg(a_s)\tau_s = \deg(a_s) + \left[\left(-(a_s - 1 - a_{s-1})^2 + N_{s-1}(a_s - 1 - a_{s-1}) + \tau_{s-1} \right) + \left(-1 + N_s + \tau_s \right) + \tau_{p_s} \right].$$

Utilizando (5.3.2), sustituimos el valor de N_s y N_{s-1} en la ecuación anterior. Después de un poco de álgebra encontramos:

$$\left[\deg(a_s) - K_s \right] \tau_s - \frac{1}{a_s - a_{s-1}} \tau_{s-1} - \frac{1}{a_{s+1} - a_s} \tau_{s+1} - \tau_{p_s} = \deg(a_s) + a_{s+1} - a_{s-1} - 2, \quad (5.3)$$

donde

$$K_s = 2 - \frac{a_{s+1} - a_{s-1}}{(a_{s+1} - a_s)(a_s - a_{s-1})} \quad (5.4)$$

Encontramos una ecuación que relaciona τ_s con los tiempos de los atajos anterior y posterior, τ_{s-1} y τ_{s+1} , y de su pareja τ_{p_s} . De esta manera conseguimos un sistema lineal de r ecuaciones que es mucho más rápido resolver (recordemos que el tamaño de la matriz de coeficientes crece como el cuadrado del número de ecuaciones que se tengan). Esta solución puede ser útil para sistemas muy grandes que tienen una proporción relativamente pequeña de atajos.

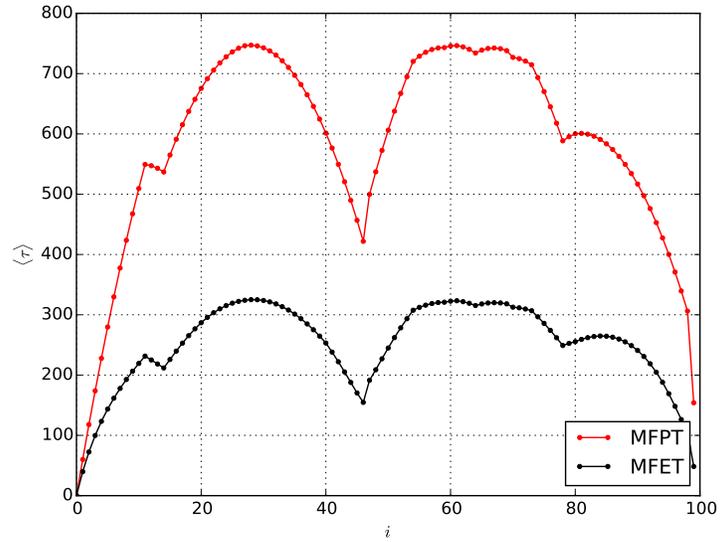
5.4. Comparación entre el MFPT al origen y el MFET

Las curvas de MFPT desde el origen o al origen resultan ser muy diferentes (fig. 5.4). Sin embargo, podemos referirnos a la validación de los métodos y darnos cuenta que las curvas de primer encuentro y de primer paso son similares cuando el origen es el sitio meta de las simulaciones de primer paso, i.e. cuando graficamos $\tau_p(i = 0)$. Este parecido es bastante

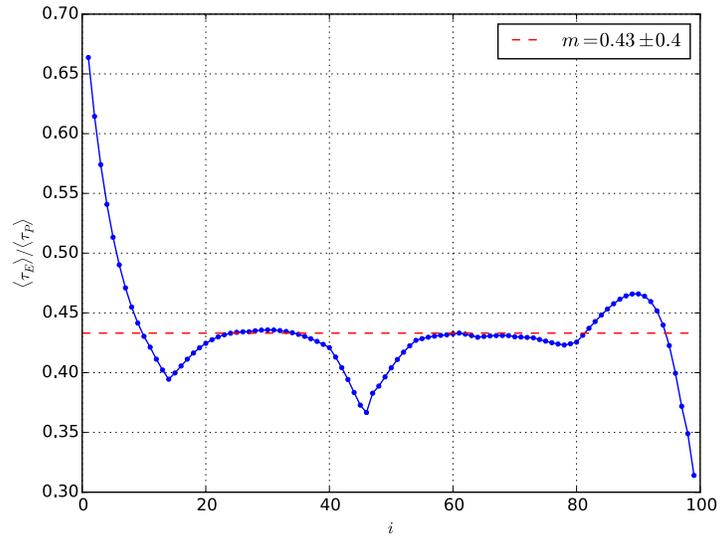
sorprendente: uno esperaría que el comportamiento de un solo caminante fuera muy diferente del de dos caminantes.

Comparamos ambas curvas en la figura 5.6a. En 5.6b graficamos el cociente $\tau_E(0, i) / \tau_P(i = 0)$ para cada sitio i . Además calculamos el promedio de este cociente sobre toda la red (excluyendo el origen para el cual el MFPT y el MFET son ambos cero). El promedio corresponde a la línea roja punteada en la figura 5.6b. Encontramos $m = 0.43$, con una desviación estándar $\sigma = 0.4$.

Hemos visto ya que en anillos, el MFET es en cada punto exactamente un medio del MFPT, y es por eso que es importante la figura 5.6b. Con excepción de los primeros 5 puntos ($i = 1, \dots, 5$), la curva del cociente se encuentra completamente por debajo de 0.5, y de hecho el promedio $m = 0.43$ es inferior a este valor. Si bien la presencia de atajos reduce todos los tiempos promedio, tiene un efecto más importante sobre el tiempo de primer encuentro que sobre el de primer paso.



(a)



(b)

Figura 5.6: (a) Comparación del MFET y MFPT cuando éste se toma desde i al origen, $\tau_P(i=0)$. (b) El cociente τ_E / τ_P . La línea roja punteada es el promedio. Para las simulaciones se utiliza la red de la fig. 4.3

Capítulo 6

Tiempos promedio generales y promedios sobre configuraciones

Nos interesamos en este capítulo en dos cantidades que sirven para comparar diferentes redes de mundo pequeño. En primer lugar, nos preguntamos si dos redes con los mismos parámetros dan lugar a tiempos de primer paso y de primer encuentro que sean similares. Para contestar esta pregunta, utilizamos los tiempos promedio generales. Por otro lado, queremos conocer cuál es el comportamiento promedio de los tiempos de primer paso y primer encuentro, para un conjunto de redes que tienen los mismos parámetros. En este caso, investigamos los promedios sobre configuraciones.

6.1. GFPT y GFET

En la figura 6.1 comparamos los tiempos generales de dos redes que tienen los mismos parámetros. Los promedios generales no son iguales. Tanto el GFPT como el GFET son menores para la red 6.1b y esto apunta a que la distribución de atajos de esta red hace que los sitios sean más accesibles unos a partir de otros.

Nos preguntamos cómo cambian los tiempos generales en función del parámetro p . Nos conviene comparar los resultados que calculemos para redes con atajos contra los tiempos generales de un anillo. En efecto, los anillos que tienen densidad de atajos $p = 0$, tienen asociado un único GFET y un único GFPT. Además, fijados los otros parámetros N y k , el tiempo

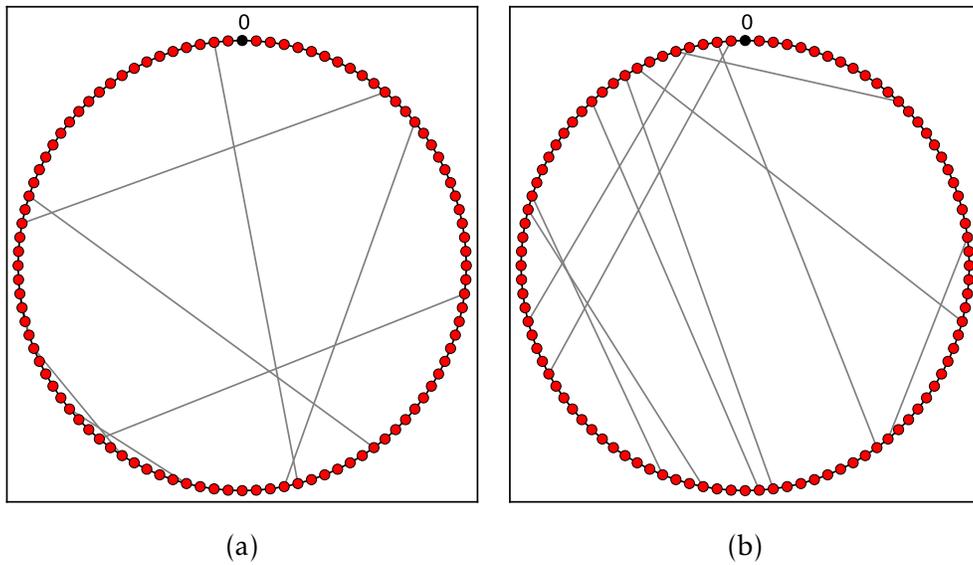


Figura 6.1: Dos redes que se utilizan para comparar los tiempos generales GFPT y GFET. Ambas tienen $N = 100$ nodos, $2k = 2$ vecinos cercanos y densidad de atajos $p = 0.1$. Para (a), redondeando a la unidad, encontramos $\tau_{P-G} = 567$ y $\tau_{E-G} = 250$; para (b) tenemos $\tau_{P-G} = 488$ y $\tau_{E-G} = 212$.

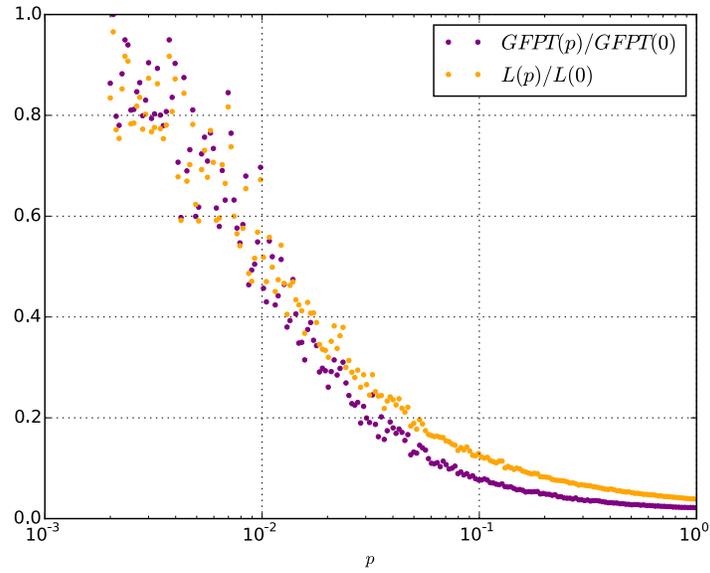
general máximo se obtiene para el anillo. Al trabajar con redes de mundo pequeño, los atajos reducen la distancia geodésica que separa cualesquiera dos sitios y es intuitivo que esto se tiene que reflejar en que bajen los tiempos de primer paso y de primer encuentro. La comparación con los tiempos del anillo da entonces un número en $[0, 1]$.

Para redes con $N = 500$ nodos, y $2k = 2$ vecinos cercanos, dibujamos los tiempos GFPT como función de p en la figura 6.2a. También incluimos la longitud media de la red $L(p)$, definida en (2.2) (comparada con la longitud media máxima que es la del anillo, $L(0)$). Los GFET de redes con $N = 100$ y $2k = 2$, como función de p , se encuentran en la figura 6.3a.

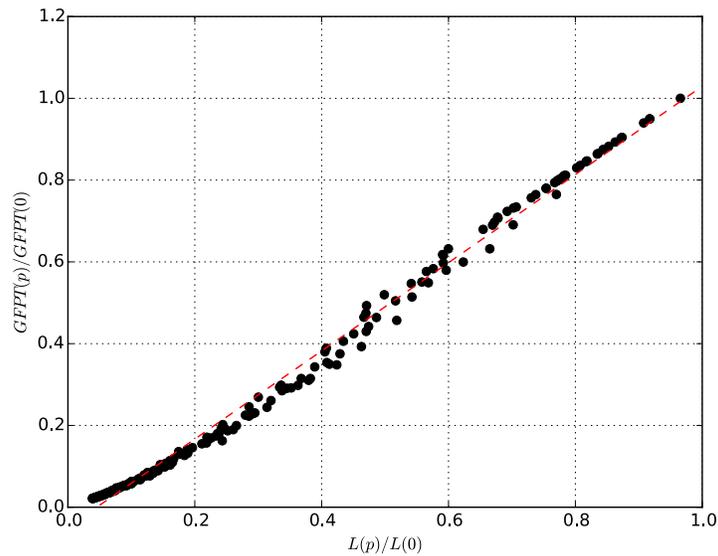
Recordamos que el número de atajos de una red de mundo pequeño es $n_a = pNk$. En nuestras simulaciones hemos utilizado $k = 1$, así que no tenemos ningún atajo si $p < 1/N$. En este punto es poco realista nuestro modelo porque no tenemos la libertad de tomar p arbitrariamente chica; al pasar el umbral $1/N$ siempre obtenemos el anillo. Esta es la razón por la que los puntos de la figura 6.2a terminan abruptamente del lado izquierdo; para p más pequeña todos los datos son exactamente $GFPT(0)$.

Tanto en la figura 6.2a como en 6.2a salta a la vista el parecido de los tiempos generales con la longitud media $L(p)$. Graficamos por lo tanto $GFPT(p)$ contra $L(p)$ en la figura 6.2b, y $GFET(p)$ contra $L(p)$ en la figura 6.3b. Obtenemos un par de curvas muy bien definidas, que muestran una clara correlación positiva entre los promedios generales y la longitud media. Los puntos no se acomodan todos sobre una recta: para valores pequeños de $L(p)$ aparece una ligera curvatura, pero de todas formas hacemos un ajuste lineal de los puntos (líneas rojas punteadas) y en ambos casos encontramos una pendiente de 1.08. Los tiempos generales y la longitud media varían prácticamente juntas.

Watts y Strogatz [24] dibujan precisamente la curva $L(p)/L(0)$ en el artículo en el que introducen el modelo de la red de mundo pequeño, y concluyen que la presencia de atajos disminuye fuertemente la longitud media. Por otro lado, Pandit y Amritkar [18] muestran que el tiempo de acceso al punto diametralmente al origen (que argumentan, está correlacionado con el tiempo promedio de cobertura (2.19)), tiene un comportamiento similar a $L(p)/L(0)$. Nosotros verificamos que este es también el comportamiento del tiempo general de **primer encuentro**. Así, nos queda más claro que nunca que la topología de una red tiene un gran impacto en los procesos dinámicos que ocurren en ella.

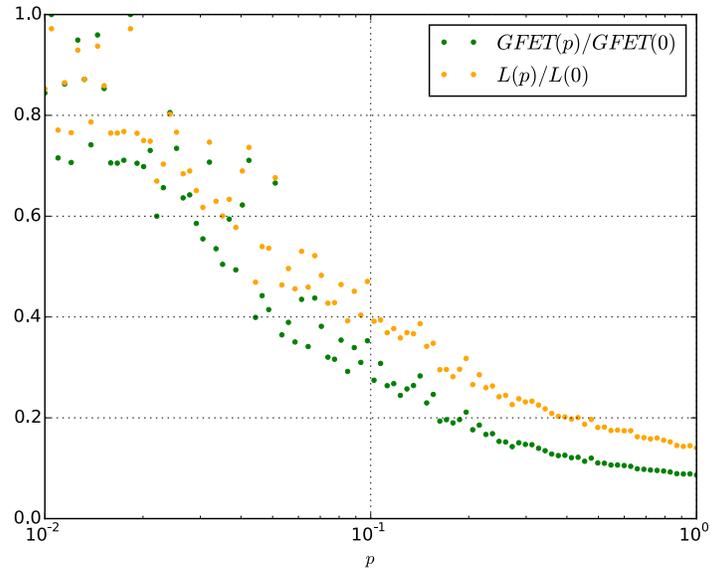


(a)

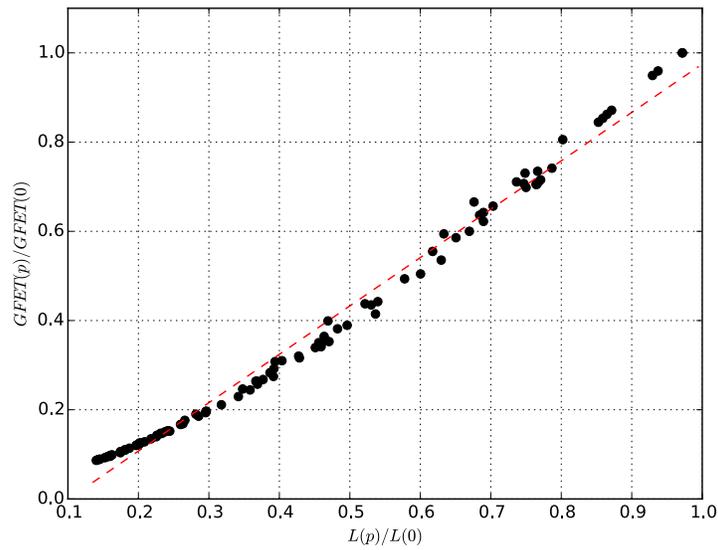


(b)

Figura 6.2: Tiempo general de primer paso: (a) $GFPT(p)$ y $L(p)$ como función de la densidad de atajos p , en redes que tienen $N = 500$ nodos y $k = 1$ vecinos cercanos. (b) $GFPT(p)$ contra $L(p)$. La línea roja punteada es un ajuste lineal de los puntos. Encontramos una pendiente de 1.08. En ambas figuras, los datos están normalizados respecto a los valores del anillo, $GFPT(0)$ y $L(0)$.



(a)



(b)

Figura 6.3: Tiempo general de primer encuentro: (a) $GFET(p)$ y $L(p)$ como función de p , en redes que tienen $N = 100$ y $k = 1$. (b) $GFET(p)$ contra $L(p)$. La línea roja punteada es un ajuste lineal de los puntos. También en este caso encontramos una pendiente de 1.08. En ambas figuras, los datos están normalizados respecto a los valores del anillo, $GFET(0)$ y $L(0)$.

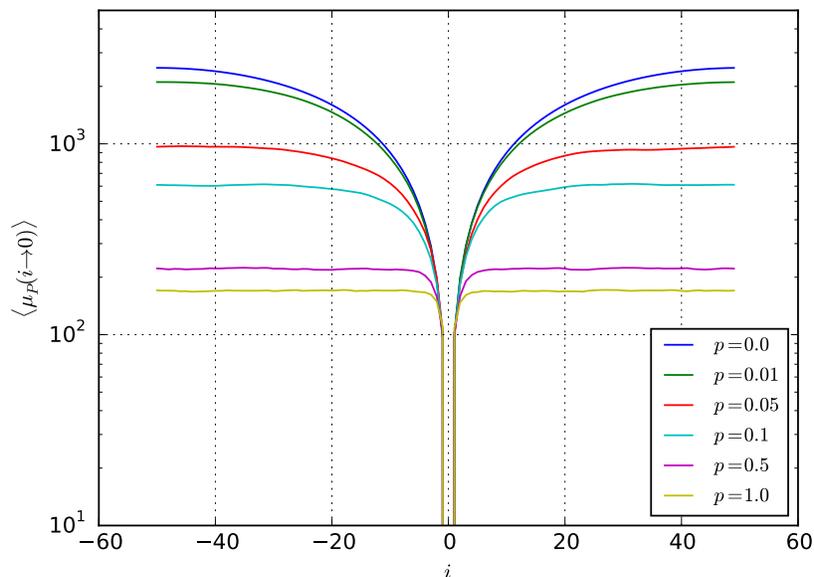


Figura 6.4: Promedio sobre 100 configuraciones de redes, de los tiempos de primer paso al origen. Se utilizan redes con $N = 100$ nodos y $k = 1$.

6.2. Promedios sobre configuraciones

Queremos estudiar el comportamiento de la familia de redes con parámetro p , con otra herramienta además de los tiempos generales.

En la figura 6.4 dibujamos los promedios de los tiempos de primer paso, colocando el origen en el centro de la curva. De esta manera queda claro que, al promediar el “desorden” de los atajos, tenemos una simetría respecto al origen: el tiempo que toma llegar a $-i$ es idéntico que el de i . Esta gráfica sirve también de validación de nuestros resultados, ya que tiene el mismo aspecto que la que obtuvieron Pandit y Amritkar [18]. En la figura 6.5 hacemos lo mismo para los tiempos de primer encuentro.

Los tiempos de primer paso son mayores que los de primer encuentro, pero es muy interesante que, salvo por esta diferencia, las curvas de las dos figuras tienen el mismo aspecto. Es sorprendente que el tiempo de encuentro de dos caminantes tenga un comportamiento similar al tiempo de primer paso; éste último concierne sólo a un caminante.

Identificamos dos comportamientos distintos. Cuando $|i|$ es pequeña, los tiempos se parecen al caso regular (el anillo) que es la curva azul $p = 0$.

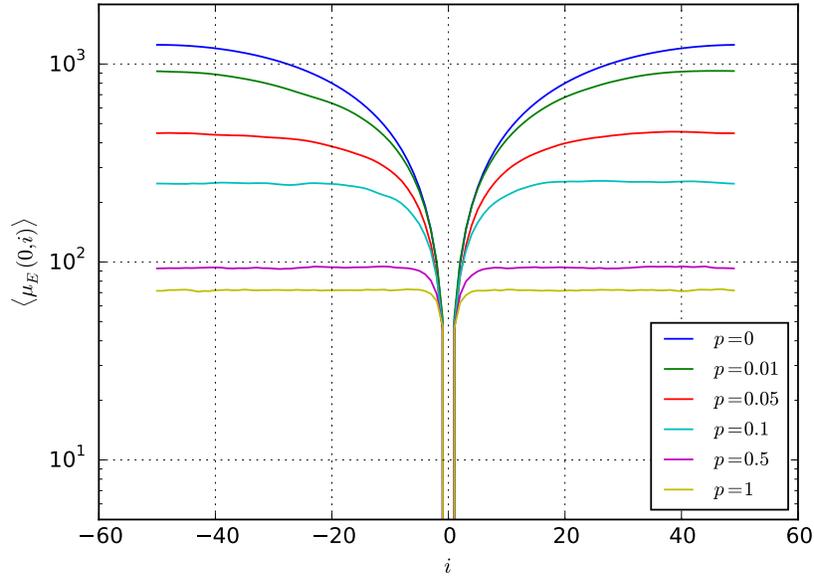


Figura 6.5: Promedio sobre 100 configuraciones de redes, de los tiempos de primer encuentro. Se utilizan redes con $N = 100$ nodos y $k = 1$.

Para $|i|$ más grande, los tiempos se saturan y la curva se vuelve constante. Pandit y Amritkar [18] encuentran que para $p = 1$, $c_P = 1/(N - 1)$. Éste es el comportamiento de una gráfica aleatoria.

Cuando p aumenta, la transición entre el comportamiento regular y la saturación ocurre antes, además de que la saturación se vuelve más prominente.

Capítulo 7

Relación entre la distancia y los MFPT y MFET

Uno de los resultados más interesantes que hemos encontrado es que, en un ejemplar de red, el MFPT al origen y el MFET son cualitativamente similares. También lo son cuando tomamos promedios sobre configuraciones (y aquí no tenemos que distinguir el tipo de MFPT, los dos son idénticos), y por lo tanto el encuentro de *dos* caminantes se comporta de forma similar al primer paso de *un* caminante. En este capítulo exploramos, a través de la distancia entre dos sitios, cómo la topología muy particular de la red de mundo pequeño resulta jugar un papel fundamental para explicar este comportamiento.

7.1. Correlación del MFPT al origen y MFET con la distancia geodésica

Es natural pensar que la distancia geodésica que separa dos sitios de la red afecta los tiempos promedio que se obtienen. En este punto retomamos las curvas de la figura 5.6 y las dibujamos con la distancia $d(0, i)$ que separa los diferentes sitios de la red y el origen, figura 7.1. La correlación con las curvas salta a la vista.

En la figura 7.2 dibujamos los mismos datos como función de la distancia $d(0, i)$. Constatamos que efectivamente existe una correlación positiva entre los tiempos promedio, se trate de MFET o MFPT al origen, y la distancia. Llama también nuestra atención el hecho de que los puntos

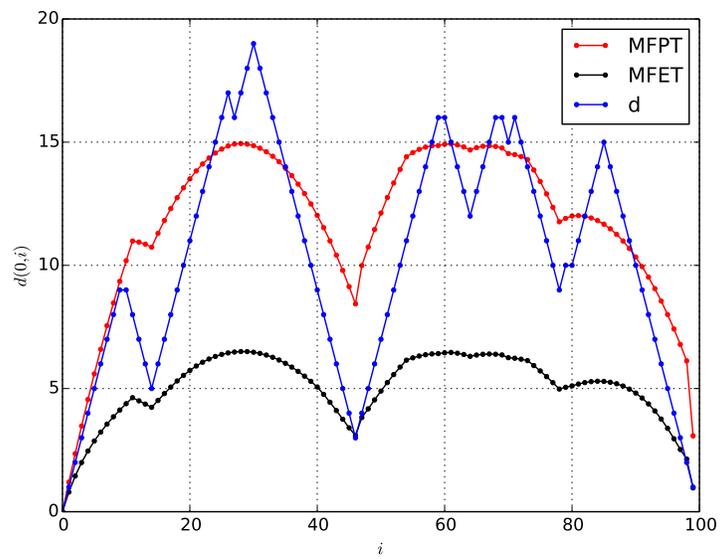


Figura 7.1: Curvas de MFET y MFPT de la fig. 5.6 junto con la distancia $d(0, i)$ de los sitios de la red al origen. Los tiempos promedio se encuentran dibujados con un factor de escala 1/50 para que la escala de la figura sea la verdadera escala de la distancia.

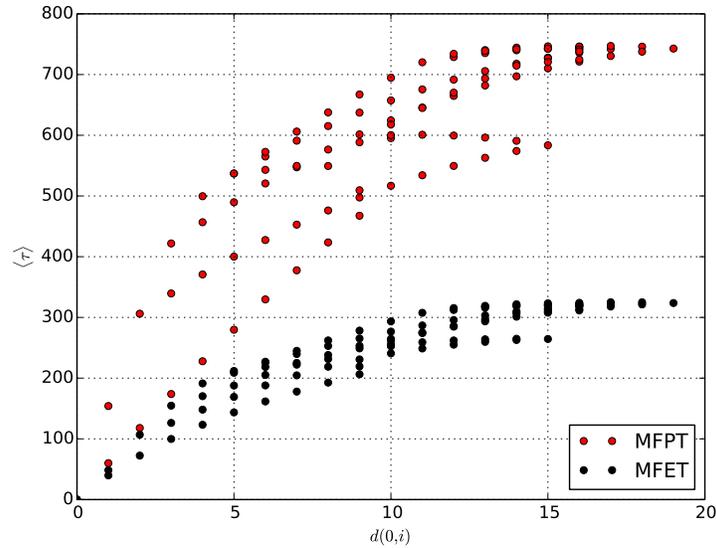


Figura 7.2: MFET y MFPT al origen como función de la distancia $d(0, i)$.

se acomodan estructuradamente sobre arcos. Sin embargo esto no nos sorprende porque recordamos las parábolas del capítulo 5. Esta estructura es consecuencia de dichas parábolas.

Este primer acercamiento con la distancia es instructivo, pero no es del todo satisfactorio. Si bien la distancia $d(i, 0)$ tiene la forma de las curvas de MFET y MFPT al origen, no es así para el MFPT que comienza en el origen. Además, sabemos que una caminata aleatoria es un proceso de Markov, y rápidamente pierde la memoria de su condición inicial. Por lo tanto, esperamos que las condiciones iniciales no jueguen un papel importante para las caminatas que no se encontraron rápidamente, esto es, en un tiempo comparable a la distancia inicial $\tau \sim d(0, i)$. Se sigue que tenemos que mirar en una dirección distinta. Resulta que es más interesante rastrear la distancia geodésica que separa a los dos caminantes en la simulación del MFET, **en cada paso de tiempo**.

7.2. La distancia dinámica

En el capítulo 3 demostramos que para los anillos se cumple siempre que el MFET es punto a punto un medio del MFPT. En el caso general,

cuando $p \neq 0$ y hay atajos presentes, la relación entre el MFET el MFPT al origen (si hay atajos hay dos tipos de MFPT) es más complicada. Podemos referirnos a la figura 5.6b, que muestra el cociente punto a punto del MFET y el MFPT al origen. No obstante, no deja de llamar la atención de que las dos caminatas dan resultados cualitativamente similares, pero no así el MFPT que parte del origen. Basándonos en las demostraciones que hicimos para anillos, queremos justificar que este comportamiento tiene que ver con la caminata de primer paso que se obtiene en el espacio de distancia, cuando estudiamos dos caminantes.

7.2.1. Caminata aleatoria en el espacio de distancia

En el anillo simple, la distancia que separa a los caminantes hace exactamente una caminata de primer paso, también sobre un anillo simple, y del mismo tamaño. La primer idea que podemos tener es que el MFPT al origen y el MFET son similares porque la distancia d que separa a los dos caminantes de MFET hace una caminata de primer paso en un espacio que es una red de mundo pequeño con algunos parámetros cambiados. Desafortunadamente, esta idea está equivocada: en el espacio de distancia, no podemos tener una red de mundo pequeño. Veamos por qué.

En los anillos generales con número de vecinos cercanos $k > 1$, la caminata de d no sigue de manera exacta lo que hacen los caminantes. Por ejemplo, supongamos que tenemos la situación de la figura 7.3, una red que tiene $k = 3$ vecinos cercanos, y que en un tiempo dado un caminante se encuentra en i y el otro en $i + 3$. La distancia que separa a los caminantes es 1, porque los sitios son vecinos. Sin embargo, si el segundo caminante brinca a $i + 1$ o $i + 2$, la distancia geodésica no cambia. Por lo tanto, en el espacio de distancia puede suceder que d se quede inmóvil cierto tiempo, a diferencia de los caminantes en una red de mundo pequeño, que se mueven en cada paso temporal.

Por otro lado, en una red que tiene atajos, la caminata aleatoria de d también es distinta de la de los caminantes. Supongamos que en un momento dado, un caminante se encuentra en el sitio i y el otro en m , y que m está conectado con un atajo. Si el segundo caminante es el que se mueve y toma el atajo, su posición cambia discontinuamente en la numeración de la red. Pero la distancia d nunca puede cambiar en más de una unidad, puesto que m es vecino del nuevo sitio que el caminante ocupa. Esto se encuentra en la figura 7.4. De hecho también es posible en este caso que d

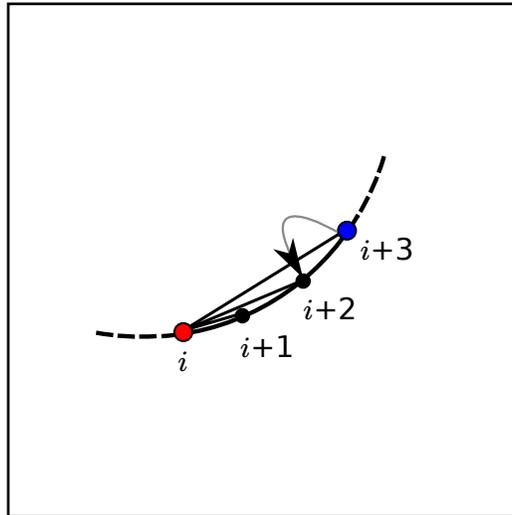


Figura 7.3: Sección de una red que tiene $k = 3$ vecinos cercanos, con un caminante en el sitio i y el otro en el $i + 3$. Para no saturar el diagrama sólo hemos dibujado las conexiones de vecinos cercanos del sitio i , pero el resto de los nodos que están conectados de la misma forma que i a sus vecinos.

no cambie. (Ésto ocurriría si hay otra ruta para llegar al nuevo sitio que no pasa por el atajo y que da exactamente la misma distancia que la que separaba a los caminantes antes del brinco.) La característica principal de las redes de mundo pequeño, los atajos, se pierde por completo en el espacio de distancia.

7.3. Visualización del espacio de distancia

En general, el espacio sobre el que ocurre la caminata aleatoria de la distancia que separa a los caminantes no es una red de mundo pequeño. No es trivial visualizar este espacio, pero puesto que es importante para explicar el parecido del MFPT al origen y el el MFET, proponemos dos opciones.

7.3.1. El espacio de distancia en una dimensión

Podemos imaginar al espacio de distancia como un intervalo discreto en una dimensión. A la izquierda tenemos al origen, a su derecha al sitio

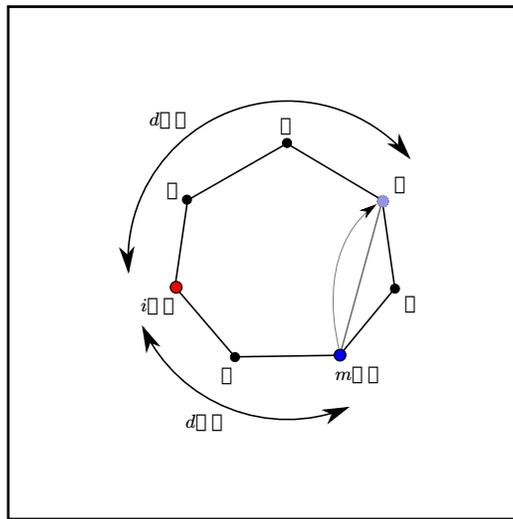


Figura 7.4: El caminante azul se mueve por el atajo de la red al sitio 2. A pesar de que su posición en la red cambia discontinuamente, la distancia que separa a los caminantes sólo aumenta en una unidad. Podemos observar que en su nueva posición la distancia que separa al caminante azul del rojo es 3, medido tanto por el arco de anillo que pasa por 6 y 0 como en la otra dirección, por los sitios 4 y 3 y el atajo.

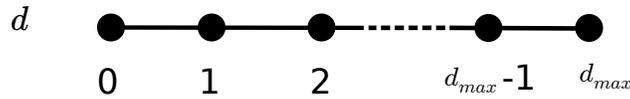


Figura 7.5: Intervalo discreto en el espacio de distancia.

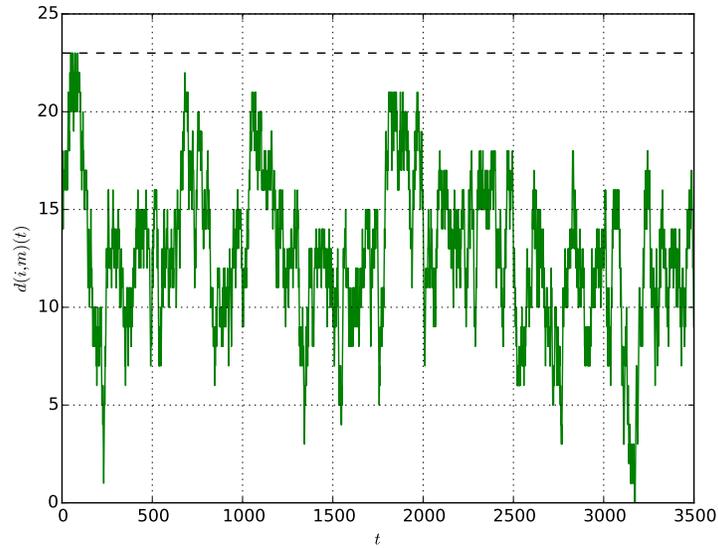
1 (los caminantes están separados en una unidad), a su derecha el sitio 2 (dos unidades), etc, y la cadena termina en el sitio d_{\max} . El origen es un sitio absorbente, y d_{\max} es reflejante con una cierta probabilidad. En general, las probabilidades de transición entre los sitios no son iguales en toda la cadena, ni tampoco lo son en las dos direcciones. Dependen de la cantidad de pares de sitios que estén separados $1, 2, \dots, d_{\max}$ unidades. Puesto que hay muchas parejas de sitios separados la misma distancia d , hay además una probabilidad de que la caminata de la distancia se quede en el sitio que ya ocupa.

Sobre la red de la figura 4.3 que tiene $N = 100$ nodos, $k = 1$ y $p = 0.1$, calculamos la distancia que separa a dos caminantes, a lo largo de un millón de pasos. En cada paso de tiempo, los caminantes se mueven dos veces, así que en el espacio de distancia tenemos dos millones de posiciones. Los primeros pasos se encuentran en la figura 7.6a, donde es evidente que la distancia hace una caminata aleatoria.

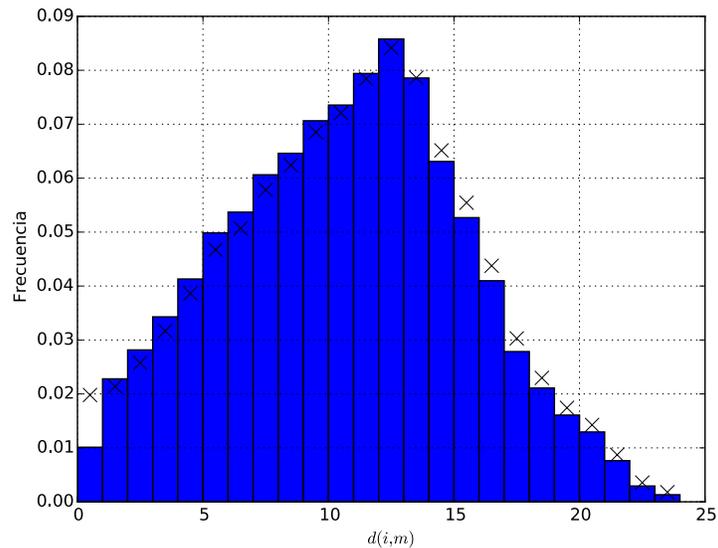
En la figura 7.6b dibujamos la frecuencia con la que la caminata visita cada posición en el espacio de distancia. Si se tratara de una red de mundo pequeño, el histograma se vería como la distribución de grados de la red, y éste no es el caso. También calculamos la fracción de las parejas de sitios (i, m) que están separados una distancia d . Los valores que encontramos para $d = 0, 1, \dots, d_{\max}$ definen la topología de la red; no son una propiedad dinámica, pero dan un buen acuerdo con las frecuencias de la caminata. Es este histograma de las posiciones de la caminata de d el que nos dice qué tan probable es ocupar los sitios de la cadena que presentábamos arriba como una herramienta para visualizar a la caminata en el espacio de distancias.

7.3.2. La distancia en la red $G \times G$

Una alternativa para imaginar el espacio dónde ocurre la caminata aleatoria de la distancia es dibujar el espacio \mathbb{Z}^2 , coloreando un sitio (i, m) según la distancia $d(i, m)$ en la red original G . Ésta es una manera indirecta



(a)



(b)

Figura 7.6: La distancia entre dos caminantes hace una caminata aleatoria. En la red de la fig. 4.3 calculamos la distancia entre dos caminantes a lo largo de un millón de pasos. (a) La caminata de la distancia para los primeros 3500 pasos. La línea negra punteada es la distancia $d_{\max} = 23$. Nótese que cerca del principio la distancia máxima se alcanza un par de veces, y que los caminantes se encuentran cerca del tiempo 3200, pues la curva toca el cero. (b) En azul, la frecuencia con la que la caminata visita cada distancia $d(i,m)$. Las cruces negras que aparecen junto con las barras son la fracción de parejas (i,m) que están separados una distancia d .

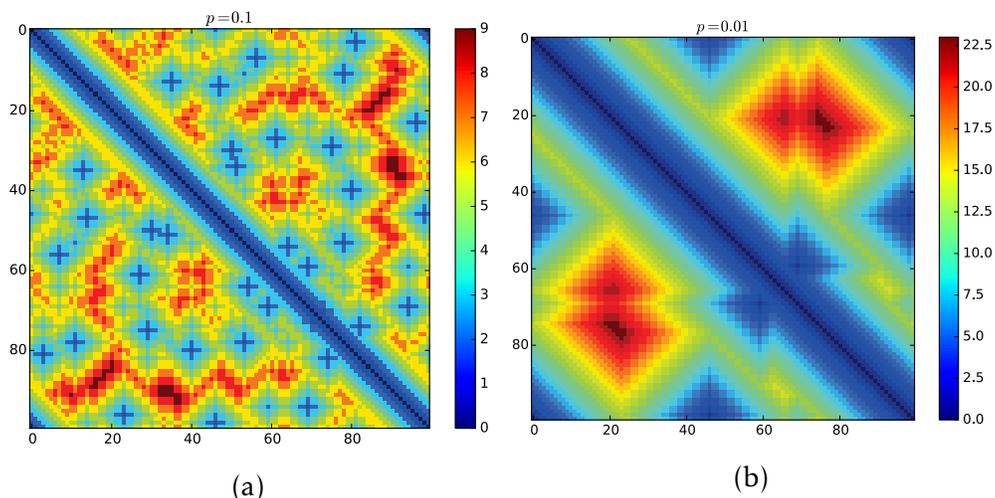


Figura 7.7: Distancia en la red \mathbb{Z}^2 para dos redes que tiene $N = 100$ nodos, $k = 2$ vecinos cercanos, y densidad de atajos (a) $p = 0.1$; (b) $p = 0.01$.

ta para imaginar el aspecto que tiene $G \times G$, donde ocurre la caminata de primer encuentro.

La figura 7.7 muestra esta idea para dos redes que tienen 100 nodos, $k = 2$ vecinos cercanos, y densidad de atajos $p = 0.1$ o $p = 0.01$. Los colores más fríos son valores de d bajos—el azul oscuro es $d = 0$ —y los más cálidos son valores más altos—el rojo oscuro corresponde a $d = d_{\max}$.

La desventaja de esta representación es que perdemos la noción de quiénes son vecinos de qué sitio, porque los vecinos conectados por atajos no están representados uno cerca del otro. Por lo tanto, esta representación es más bien cualitativa y no encontramos un uso práctico para ella.

7.4. Resumen

Hemos encontrado la razón por la que el MFPT al origen se parece al MFET: la simulación de primer encuentro se puede traducir a una caminata de primer paso por el origen en el espacio de distancia.

Para el anillo, la caminata en dicho espacio ocurre en un anillo idéntico, pero este no es el caso para una red de mundo pequeño general, que tiene densidad de atajos $p > 0$: el espacio de distancia no vuelve a ser una red de mundo pequeño. De hecho la característica principal de las redes de

mundo pequeño que es la presencia de atajos que unen sitios lejanos, se pierde por completo en el espacio de distancia.

Capítulo 8

Conclusiones

El objetivo principal del presente trabajo fue estudiar los tiempos promedio de primer encuentro (MFET) de dos caminantes aleatorios en la familia de redes de mundo pequeño. En el camino trabajamos con el tiempo promedio de primer paso (MFPT), que está cercanamente relacionado.

En la primera parte de la tesis resolvimos el caso más sencillo de la red, el anillo simple, y encontramos los tiempos de forma analítica. También demostramos que en la ausencia de atajos el MFPT es exactamente el doble del MFET, aún cuando el número de vecinos cercanos es mayor que $k = 1$.

El resto de los resultados de la investigación se encuentran en la segunda parte del trabajo. El capítulo 4 contiene la validación de los métodos numéricos. Encontramos que:

- Monte Carlo y enumeración exacta, que sirven para encontrar la distribución de tiempos, concuerdan satisfactoriamente. Sin embargo, desde la validación vemos la utilidad superior del método de enumeración exacta, puesto que permite conseguir un resultado numéricamente exacto.
- Los tres métodos dan resultados muy similares para el tiempo promedio de primer paso.
- Para el tiempo de primer encuentro cometemos un par de errores sistemáticos. El primero es una sobrevaloración por parte del método de enumeración exacta, que proviene de la aproximación que se utiliza para la cola exponencial. Este error se puede hacer tan pequeño

como se quiera, y no nos preocupa. El segundo error da una diferencia entre Monte Carlo y los otros dos métodos. Intuimos que proviene del conjunto de vecinos $N(i)$ que se utiliza tanto para enumeración exacta como para el método matricial. En promedio el error ronda 4% y tomando un enfoque práctico decidimos que no es demasiado preocupante.

En el capítulo 5 discutimos diferentes resultados que se obtienen en una versión de red:

- En primer lugar estudiamos las redes que no tienen atajos pero sí tienen más de un vecino cercano de cada lado. Para MFPT y MFET encontramos dos comportamientos para los nodos que están conectados con el origen, y el resto de nodos de la red que no lo están.
- En redes que tienen atajos, comparamos las dos direcciones de las simulaciones de primer paso. Se obtienen diferentes curvas si el origen es el sitio meta o la condición inicial de las caminatas. En el segundo caso, los resultados presentan variaciones más importantes. Esta diferencia se explica a partir de las ecuaciones de recurrencia que describen una y otra caminata. Además observamos que, en ambas caminatas, los tiempos de los sitios que no están conectados con atajos se organizan en parábolas, como en el caso de un anillo simple.
- Comprobamos este comportamiento utilizando una red que tiene solamente un atajo. Ajustamos tres secciones de parábolas del tiempo de primer paso que llega al origen, y encontramos la expresión predicha del anillo simple. El ajuste define el número efectivo de nodos N , de cada arco que corresponde a una parábola.
- Con un tratamiento semi-analítico encontramos una expresión para N en términos de los tiempos a los que llegan los atajos, y concluimos que el problema se puede por lo tanto reducir a un sistema con un número menor de incógnitas.
- Por último, comparamos el MFPT que llega al origen con el MFET. Encontramos un parecido bastante importante, por lo menos cualitativamente. Calculamos punto a punto el cociente del MFET entre el MFPT y encontramos valores que por lo general están debajo de

un medio. La presencia de atajos disminuye el tiempo de primer encuentro aún más de lo que disminuye el de primer paso.

En el capítulo siguiente, capítulo 6, trabajamos con los tiempos promedio generales y con los promedios sobre configuraciones. Los principales resultados que encontramos son:

- Dos redes con los mismos parámetros no tienen necesariamente el mismo tiempo general de primer paso (GFPT), ni tampoco el mismo tiempo general de primer encuentro (GFET).
- Los tiempos generales aumentan cuando la densidad de atajos p disminuye. El tiempo general máximo para N nodos y k vecinos cercanos se consigue en ausencia de atajos, esto es, cuando $p = 0$.
- Esta variación nos hace pensar en la longitud media de la red $L(p)$, que es una propiedad topológica de la red. Encontramos una correlación casi exacta entre los tiempos generales y $L(p)$.
- Cuando tomamos el promedio del MFPT o el MFET sobre distintas configuraciones, las curvas que se obtienen según la densidad de atajos p que se tome, son muy parecidas para una y otra simulación: el comportamiento de dos caminantes es similar al de uno solo.
- Las curvas muestran dos tipos de comportamiento: para sitios cercanos al origen, los tiempos se comportan como en una gráfica regular; para sitios lejanos, tienen el comportamiento de una gráfica aleatoria y se saturan en un valor constante. Cuando el parámetro p aumenta, la saturación ocurre antes y se da de forma más pronunciada.

Finalmente en el capítulo 7, tratamos de explicar por qué los tiempos de primer paso y primer encuentro se parecen, utilizando la distancia que separa a los dos caminantes en el segundo caso.

- La distancia de cada nodo al origen presenta variaciones similares a las de MFET y MFPT que llega al origen. Al dibujar los tiempos promedio contra esta distancia, encontramos una correlación positiva. Esta distancia estática, sin embargo, no explica por qué es diferente la curva de MFPT que parte del origen.

- Nos fijamos, entonces, en la distancia d que separa a los caminantes en la simulación de MFET, en cada paso de tiempo. d hace una caminata aleatoria de primer paso por el origen: los caminantes se encuentran cuando la distancia llega por primera vez al valor cero.
- La caminata de d no ocurre en una red de mundo pequeño. Llegamos a esta conclusión ya que la distancia nunca cambia en más de una unidad (y por lo tanto no puede haber atajos en el espacio en el que se mueve), además de que d puede quedarse en el sitio que ocupa.
- Para imaginar el aspecto del espacio de distancia en el que ocurre la caminata de d , proponemos un intervalo discreto en una dimensión. Podemos imaginar aproximadamente los valores de las probabilidades de transición en dicho intervalo a partir de un histograma de frecuencias que dibujamos.
- También podemos colorear la cuadrícula de $G \times G$, donde ocurre la caminata de primer encuentro, con colores que dependen de la distancia que separa cada pareja de sitios. Esta forma de representar el espacio no es mucho más que cualitativa, porque no captura ningún aspecto de los atajos de la red original.

Direcciones para continuar el trabajo

La conclusión más interesante de la investigación es que el MFET se parece al MFPT que llega al origen y, por lo tanto, que en las redes de mundo pequeño dos caminantes se comportan de forma similar a uno solo. La explicación de este parecido requiere que recurramos a la distancia que separa a los caminantes que se deben de encontrar, y la caminata aleatoria que hace esta distancia es relativamente bien comportada por la topología de la red.

Los vecinos cercanos tienen una influencia primordial en nuestra versión de la red de mundo pequeño. Sería instructivo verificar que se obtiene el mismo comportamiento en la versión original de Watts y Strogatz de la red de mundo pequeño, porque en ese modelo se modifican las conexiones con los vecinos cercanos.

Por otro lado, una pregunta obvia que nos podemos hacer es: ¿Qué sucede en otros tipos de redes? ¿Cómo se comparan los tiempos de primer paso y primer encuentro en una familia de redes completamente di-

ferente? Una elección interesante pueden ser las redes libres de escala de Barabasi-Albert.

Bibliografía

- [1] M. Abdullah, C. Cooper, and T. Radzik. The cover time of cartesian product graphs. In *Combinatorial Algorithms*, pages 377–389. Springer, 2011.
- [2] F. Berto and J. Tagliabue. Cellular automata. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2012 edition, 2012.
- [3] G. Chartrand and P. Zhang. *A first course in graph theory*. Dover, 2012.
- [4] S. Condamin, O. Bénichou, and M. Moreau. Random walks and brownian motion: A method of computation for first-passage times and related quantities in confined geometries. *Physical Review E*, 75(2):021111, 2007.
- [5] S. Condamin, O. Bénichou, V. Tejedor, R. Voituriez, and J. Klafter. First-passage times in complex scale-invariant media. *Nature*, 450(7166):77–80, 2007.
- [6] W. Feller. *An Introduction to Probability Theory and Its Applications*. John Wiley & Sons, 2008.
- [7] L. Giuggioli, S. R. Pérez-Becker, and D. P. Sanders. Encounter times in overlapping domains: application to epidemic spread in a population of territorial animals. *Physical Review Letters*, 110(5):058103, 2013.
- [8] S. Jespersen, I. M. Sokolov, and A. Blumen. Relaxation properties of small-world networks. *Physical Review E*, 62(3):4405, 2000.

- [9] J. Lahtinen, J. Kertész, and K. Kaski. Scaling of random spreading in small world networks. *Physical Review E*, 64(5):057105, 2001.
- [10] J. Lahtinen, J. Kertész, and K. Kaski. Random spreading phenomena in annealed small world networks. *Physica A: Statistical Mechanics and its Applications*, 311(3):571–580, 2002.
- [11] I. Majid, D. Ben-Avraham, S. Havlin, and H. E. Stanley. Exact-enumeration approach to random walks on percolation clusters in two dimensions. *Physical Review B*, 30(3):1626–1628, 1984.
- [12] R. Monasson. Diffusion, localization and dispersion relations on “small-world” lattices. *The European Physical Journal B - Condensed Matter and Complex Systems*, 12(4):555–567, 1999.
- [13] E. W. Montroll and M. F. Shlesinger. On the wonderful world of random walks. In *Nonequilibrium phenomena II: From stochasticity to hydrodynamics*, pages 1–121. North-Holland Physics Publishing, 1984.
- [14] M. Moreau, G. Oshanin, O. Bénichou, and M. Coppey. Pascal principle for diffusion-controlled trapping reactions. *Physical Review E*, 67(4):045104, 2003.
- [15] M. E. J. Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.
- [16] M. E. J. Newman and D. J. Watts. Renormalization group analysis of the small-world network model. *Physics Letters A*, 263(4):341–346, 1999.
- [17] J. D. Noh and H. Rieger. Random walks on complex networks. *Physical review letters*, 92(11):118701, 2004.
- [18] S. A. Pandit and R. E. Amritkar. Random spread on the family of small-world networks. *Physical Review E*, 63(4):041104_1–041104_7, 2001.
- [19] S. Pérez Becker. Tiempos de primer encuentro en caminantes aleatorios confinados. Master’s thesis, Universidad Nacional Autónoma de México (UNAM), 2011.

- [20] S. Redner. *A guide to first-passage processes*. Cambridge University Press, 2001.
- [21] B. Ribeiro, P. Basu, and D. Towsley. Multiple random walks to uncover short paths in power law networks. In *Computer Communications Workshops (INFOCOM WKSHPS), 2012 IEEE Conference on*, pages 250–255, 2012.
- [22] J. P. Sethna. *Statistical Mechanics: Entropy, Order Parameters, and Complexity*. Oxford University Press, 2006.
- [23] P. Solís Carmona. Enumeración exacta para tiempos de encuentro de caminantes aleatorios confinados. Master’s thesis, Universidad Nacional Autónoma de México (UNAM), 2015.
- [24] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442, 1998.