

JH-FC  
AG-F  
959  
1



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS  
DEPARTAMENTO DE FÍSICA

# Optimización de redes complejas bidimensionales

TESIS ENTREGADA A LA UNIVERSIDAD DE CHILE EN CUMPLIMIENTO  
PARCIAL DE LOS REQUISITOS PARA OPTAR AL GRADO DE:

**MAGÍSTER EN CIENCIAS CON MENCIÓN EN FÍSICA**

**SEBASTIÁN GUILLIER TORRES**

NOVIEMBRE, 2010



**DIRECTORES DE TESIS:**

JUAN ALEJANDRO VALDIVIA  
VICTOR MUÑOZ  
JOSE ROGAN

# INFORME DE APROBACIÓN TESIS DE MAGÍSTER

Se informa a la Escuela de Postgrado de la Facultad de Ciencias que la Tesis de Magíster presentada por el candidato.

## SEBASTIÁN GUILLIER TORRES

Ha sido aprobada por la comisión de Evaluación de la tesis como requisito para optar al grado de Magíster en Ciencias con mención en Física, en el examen de Defensa de Tesis rendido el día 5 de noviembre de 2010.

### DIRECTORES DE TESIS:

Dr. Juan Alejandro Valdivia

Dr. Victor Muñoz

Dr. Jose Rogan

*JAV*  
.....  
*Victor Muñoz*  
.....  
*J. Rogan*  
.....

### COMISIÓN DE EVALUACION DE LA TESIS:

Dr. Mario Molina

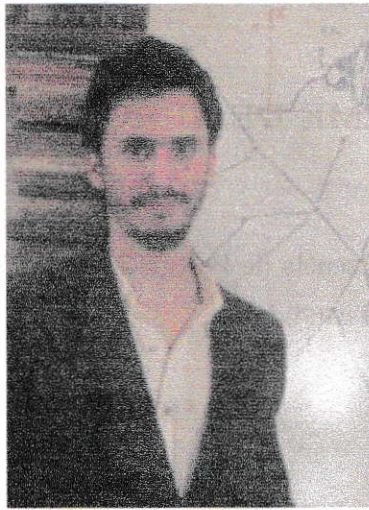
Dr. Miguel Kiwi

Dr. Malcel Clerc

*C. Molina*  
.....  
*M. Kiwi*  
.....  
*M. Clerc*  
.....



Nacido en la ciudad de Santiago de Chile el 31 de julio de 1982, hijo de Isabel y Cristián y segundo de cuatro hermanos. Desde niño tuvo una notoria inclinación por las matemáticas la que perduro hasta su adolescencia, en donde sus habilidades fueron destacadas por cada uno de los profesores del área que tuvo. Sin embargo, este hecho estuvo opacado por una constante indisciplina escolar, la



que lo llevo a pasar por seis colegios diferentes, finalizando incluso la enseñanza media por medio de exámenes libres. Vuelve a mostrar su habilidad en una buena P.A.A. destacando 770 y 784 puntos en las pruebas de aptitud y específica matemáticas respectivamente, posterior a ello decide en el año 2001 entrar a la Universidad de Chile a estudiar Licenciatura en Física. Egresó el año 2005 y a mediados del 2006 ingresa a estudiar Ingeniería Civil Industrial en la Universidad de Valparaíso de donde egresa en forma destacada a comienzos del 2009, desarrollándose durante ese mismo año como académico en esta universidad. Paralelamente a la ingeniería entra a estudiar magíster en física en la Universidad de Chile la que concluye con la tesis titulada *Optimización de Redes Complejas Bidimensionales*. Actualmente se encuentra escribiendo su tesis de Ingeniería y desde febrero de 2010 hasta ahora se encuentra trabajando de tiempo completo en la gerencia de riesgo y análisis cuantitativo en Celfin Capital.



## AGRADECIMIENTOS



Agradezco al departamento de física por entregarme la beca que me permitió cursar el programa de magíster.

Agradezco a los profesores J. A. Valdivia, V. Muñoz y J. Rogan a quienes responsabilizo de esta tesis.

Quiero aprovechar esta página como última instancia para agradecer a cada uno de los profesores que tuve durante todo mi paso por la facultad de ciencias, tanto en licenciatura como en magíster. Durante estos años cada enseñanza y cada dialogo estimularon mi poco experimentado intelecto, asombrándome por la naturaleza y convirtiendo a la universidad en una de las mejores etapas de mi vida.

Agradezco también a cada funcionario de la universidad por la particular buena relación con los alumnos, en especial a Marisol Maira, quien siempre me estuvo dispuesta a colaborar.

Por último, agradezco a mi familia quienes, en todo lo transversal de la vida, son mi pilar fundamental.





## RESUMEN

El estudio de las redes complejas ha sido de gran interés en este último tiempo puesto que son capaces de describir una variedad importante de fenómenos. En particular, en este trabajo estudiaremos aquellos fenómenos que pueden ser representados por redes complejas espaciales, es decir, en aquellas en que los nodos y enlaces tiene una ubicación bien definida en el espacio, como por ejemplo una red de carreteras. Nuestro interés, es proponer modelos que reproduzcan en alguna medida las características fundamentales de estos sistemas. Para ello, en primer lugar, en el Cap. 2 se hace una revisión de los principales tópicos de las redes complejas, así como también de los principales modelos existentes.

El eje fundamental en el cual se sustentan los modelos propuestos en ésta tesis es la optimización. Esta optimización esta asociada, por un lado, a la minimización de los costos de fabricación de las redes, y por otro, a la minimización de los costos de ineficiencia de estas. La manera por la cual es medida esta ineficiencia es la que da paso a diferentes modelos. En el Cap. 3 se hace una revisión a algunos modelos de optimización encontrados en la literatura y se profundiza el estudio de sus características. Se muestra, que bajo ciertas condiciones, las redes resultantes de estos modelos presentan características *small-world*, hecho que no ha sido evidenciado de forma explicita por otros autores. Sin embargo, estas redes no son capaces de reproducir las distribuciones de conectividad libre de escala vista en muchas redes reales.

Por último, en el Cap. 4 se estudia una evolución a los modelos anteriores, la que consiste en asumir la construcción de la red como un proceso paulatino de crecimiento, aún que sin dejar de lado la optimización. Las redes resultado de estos modelos pierden en parte sus características *small-world*, pero a cambio, sus distribuciones cumplen leyes de potencia (con constantes  $\gamma = 2.2$  y  $\gamma = 2.5$ ), lo cual se asemeja mucho más a muchas redes encontradas en la naturaleza.



# Índice general

<b>1. Introducción</b>	<b>1</b>
<b>2. Redes complejas</b>	<b>4</b>
2.1. Definiciones generales . . . . .	4
2.2. Distribución de la conectividad . . . . .	7
2.3. Redes aleatorias . . . . .	9
2.4. Crecimiento de redes . . . . .	11
2.4.1. Topología exponencial . . . . .	14
2.4.2. Topología libre de escala . . . . .	16
2.5. Redes tipo <i>small-world</i> . . . . .	18
2.5.1. Redes <i>small-world</i> reales . . . . .	18
2.5.2. Modelo probabilístico de Watts y Strogatz . . . . .	18
2.6. Redes espaciales bidimensionales . . . . .	20
2.6.1. Condiciones de borde periódicas . . . . .	21
2.7. Algoritmo de Floyd-Warshall . . . . .	22
<b>3. Optimización: Mínimo global</b>	<b>26</b>
3.1. Modelo de minimización . . . . .	26
3.1.1. Algoritmo de minimización $\mathcal{M}$ . . . . .	27
3.1.2. Algoritmo genético . . . . .	29
3.2. Modelo global I: Geodésica geométrica . . . . .	31
3.3. Modelo global II: Geodésica . . . . .	36
<b>4. Optimización: Modelos de crecimiento</b>	<b>41</b>
4.1. Crecimiento con ahorro de camino (CAC) . . . . .	42
4.2. Crecimiento Central I . . . . .	43
4.3. Crecimiento Central II . . . . .	46
4.4. Crecimiento Total I . . . . .	50
4.5. Crecimiento Total II . . . . .	55
<b>5. Conclusiones</b>	<b>60</b>
<b>A. Distribución de Poisson como límite de una distribución Binomial</b>	<b>62</b>
<b>Bibliografía</b>	<b>64</b>

# Capítulo 1

## Introducción

Las redes complejas proporcionan un marco teórico útil para la representación y entendimiento de muchos sistemas tecnológicos, biológicos y sociales, y es por ello que han recibido una sustancial atención en los últimos años [1–4].

Una parte importante de estos modelos de redes tiene una representación geométrica, en la cual los nodos tienen una posición bien definida en el espacio, este tipo de redes será el foco de esta tesis. Ejemplo de redes espaciales o geométricas son las redes de transporte eléctrico, redes de proteínas, redes de trenes, etc [5–8]. Se ha encontrado que muchos de estos sistemas a menudo presentan características comunes, tales como distribuciones libres de escala en su conectividad, alto *factor de clustering* y baja distancia promedio entre nodos [9–11] <sup>1</sup>.

En vista de lo anterior, muchos investigadores han desarrollado trabajos que intentan explicar estas características y es en esa dirección en la cual avanza esta tesis. Siendo más precisos, estamos interesados de proponer modelos que repliquen de mejor forma las principales características halladas en redes reales. Para ello, en primer lugar trabajaremos con modelos de redes espaciales que surgen como resultado de algún criterio de optimización. Esta optimización está basada en la minimización de una energía efectiva, la cual pondera entre dos factores, uno asociado al costo de construcción de la red y otro asociado a su eficiencia. La medida de esta eficiencia radica en la distancia promedio entre los nodos. Por otro lado, el costo asociado a la construcción de la red está determinado por la cantidad de enlaces y sus longitudes. Más detalles sobre esta energía efectiva se entregan en las secciones 3.1, 3.2 y 3.3. Sin embargo, a pesar de lo intuitivo que resultan estos modelos, no son capaces de explicar

---

<sup>1</sup>Estas y otras propiedades de las redes serán revisadas con mayor atención en el capítulo 2



características claves de las redes como la distribución de conectividad, es por ello que finalmente se opta por introducir modelos de crecimiento. La dinámica de éstos es la continua expansión de la red producto de la constante incorporación de nuevos nodos al sistema. Es particular, sin embargo, que a pesar de ser redes dinámicas sus principales características tienden a estados estacionarios cuando el número de nodos es suficientemente grande.

Varios autores han propuesto modelos de crecimiento de redes que en alguna medida se ajustan a redes reales, sin embargo, distan de una completa descripción, además a veces carecen de realismo e intuición. Para solucionar estos puntos, los modelos de crecimiento aquí propuestos están nuevamente basados en la optimización de una energía específica. Este hecho no solo busca darle mayor realismo e intuición a los modelos sino que también busca la obtención de mejores resultados.

Los modelos aquí tratados se desarrollan teniendo en mente redes de transporte “humanas”. Con esto nos referimos a redes construidas por personas, ya sean de flujo físico o de información, como por ejemplo, redes de carreteras, de potencia eléctrica, de Internet, de rutas de vuelo, de cableado telefónico, etc. Sin embargo, los mismos análisis pueden ser útiles para describir otros tipos de redes.

Esta tesis está dividida en tres partes principales:

**Redes Complejas (Cap. 2):** Esta sección es un resumen de los conceptos básicos relacionados a las redes complejas, acompañado de algunos de los modelos más recurrentes en la literatura. Aquí también se define la terminología y nomenclatura que se utilizará en el resto de la tesis. Finalmente se expone parte del algoritmo que se utilizará en las simulaciones posteriores.

**Mínimo Global (Cap. 3):** Aquí se trabajan dos modelos de optimización que ya han sido tratados por otros autores. Se detallan y se entregan resultados que podrían ser de interés, como por ejemplo, características *small-world* [10]. Los dos modelos aquí tratados (que llamaremos Global I y Global II) buscan el diseño óptimo de una red para una configuración espacial de nodos fija. En particular, lo que se busca es el mínimo global de costos para una energía determinada, la cual dependerá de la configuración de los enlaces. Los resultados aquí obtenidos servirán como un punto de comparación con otros modelos publicados, así como



también servirán como comparación con los modelos de crecimiento que se tratarán en el capítulo siguiente.

**Crecimiento (Cap. 4):** La introducción del concepto de crecimiento de redes surge como una necesidad de crear modelos que describan mejor a las redes reales. Los modelos tratados son cuatro. Los primeros dos tienen que ver con una red que crece en torno a un único nodo central, despreciando la importancia de los otros nodos de la red. A estos modelos los llamaremos crecimiento Central I y II. Los otros dos modelos, por otra parte, ponderan por igual la importancia de todos los nodos en la red, A éstos últimos los denominaremos Crecimiento Total I y Crecimiento Total II. Los cuatro modelos aquí tratados optimizan costos paso a paso según la configuración que haya al momento de agregar cada nodo, sin tomar en cuenta la configuración futura, es decir, se está realizando una optimización local, sin considerar el costo global de todo el proceso. Esto concuerda con la lógica de crecimiento de algunas redes reales, en donde el sentido de lo urgente (corto plazo) puede imponerse por sobre el sentido de lo importante (largo plazo).

# Capítulo 2

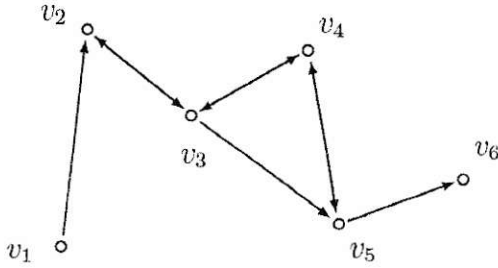
## Redes complejas

En este capítulo introduciremos algunos conceptos útiles para caracterizar redes complejas. También discutiremos ejemplos específicos de redes que han sido definidas y estudiadas por otros autores, poniendo énfasis en las diferentes topologías que estos obtuvieron. Posteriormente se introducirá el concepto de redes espaciales bidimensionales, y por último, se explicará el algoritmo de Floyd Marshall el cual será utilizado en los capítulos siguientes para determinar distancias mínimas, las que tendrán un papel fundamental en esta tesis.

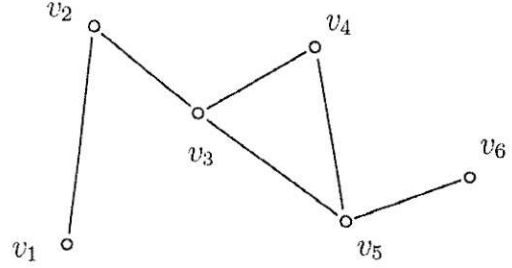
### 2.1. Definiciones generales

Para comenzar el análisis de redes, es preciso hacer algunas definiciones [1–3].

- Def. 1** **Nodo (vertex):** *Es la unidad fundamental de una red. A cada nodo individual lo denotaremos  $v_i$ , y al conjunto total de nodos de la red lo denominaremos  $\mathcal{V} = \{v_i\}$ .*
- Def. 2** **Enlace (edge):** *Par ordenado de nodos  $(v_i, v_j)$ . Gráficamente puede visualizarse como una línea recta que une el nodo  $v_i$  con el nodo  $v_j$ . Ocasionalmente será denominado como  $e_{ij}$ .*
- Def. 3** **Red:** *Conjunto de  $N$  nodos  $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$ , y un conjunto de  $M$  parejas ordenadas (enlaces)  $\mathcal{E} = \{(v_i, v_j)\} \subset \mathcal{V} \times \mathcal{V}$ . La red se denomina **no dirigida** si para cada pareja  $(v_i, v_j) \in \mathcal{E}$  también existe la pareja  $(v_j, v_i) \in \mathcal{E}$ . De lo contrario, la red se denomina **dirigida**. Al número de enlaces  $M$  se le denomina el orden de la red.*



(a) Red dirigida



(b) Red no dirigida

Figura 2.1: Gráfica de los dos tipos de redes mencionadas. Una representación típica de red dirigida es por medio de flechas tal como se muestra en la parte (a). En cambio para las redes no dirigidas basta con mostrar la existencia o no del enlace [parte (b)]. Observemos que para la red dirigida podemos movernos desde el nodo  $v_1$  al  $v_6$ , pero no podemos llegar del nodo  $v_6$  al  $v_1$ . En contraste, en la red no dirigida todos los caminos son reversibles.

**Def. 4** **Conectividad:** También conocida como el grado de coordinación, es el número de enlaces conectados a un nodo. La conectividad del nodo  $v_i$  es denotada por  $k_i$  y al promedio  $\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i$  se le denomina la **conectividad de la red**.

**Def. 5** **Matriz de adyacencia:** La matriz de adyacencia  $\mathbf{A}$  es una representación matricial del conjunto de enlaces  $\mathcal{E}$ . Sus términos son tales que:

$$\mathbf{A}_{ij} = \begin{cases} 1 & \text{si } (v_i, v_j) \in \mathcal{E} \\ 0 & \text{si } (v_i, v_j) \notin \mathcal{E} \end{cases}$$

Como ejemplo, a continuación se muestran las matrices de adyacencia correspondientes a la figura 2.1.

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Red dirigida [Fig. 2.1.a]

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad (2.1)$$

Red no dirigida [Fig. 2.1.b]

Notemos que la matriz de la derecha es simétrica, característica de todas las matrices de adyacencia de redes no dirigidas. En esta tesis sólo consideraremos redes **no** dirigidas, de

modo que cuando nos referimos a “red”, siempre queremos decir “red no dirigida”.

**Def. 6 Geodésica:** Conjunto de enlaces que conforman el camino más corto (menor número de enlaces) entre un nodo y otro a través de la red. Es el camino que tiene menos “estaciones” entre un nodo y otro.

**Def. 7 Matriz  $\mathbf{H}$ :** Las componentes  $\mathbf{H}_{ij}$  de la matriz corresponden al número de enlaces que conforman la geodésica entre los nodos  $v_i$  y  $v_j$ . Si no hay forma de conectar  $v_i$  con  $v_j$  entonces  $\mathbf{H}_{ij} = \infty$ .

**Def. 8 Factor de Clustering:** Número entre cero y uno que indica qué tan “agrupados” están los nodos. El factor de clustering de un nodo  $v_n$  se define como el cociente entre el número de enlaces que hay entre los “nodos vecinos” de  $v_n$ , y el máximo número de estos que puede haber ( $k_n(k_n - 1)/2$ ). Esto es:

$$C_n = \frac{\sum_{i,j} \mathbf{A}_{ij}}{k_n(k_n - 1)/2} \quad \forall i, j \text{ tal que } \mathbf{A}_{in} = 1 \text{ y que } \mathbf{A}_{jn} = 1, \quad (2.2)$$

donde  $k_n$  es la conectividad del nodo  $v_n$ . El factor de clustering de la red se define como:

$$C = \frac{1}{N} \sum_{n=1}^N C_n. \quad (2.3)$$

Es importante señalar que todas las propiedades mencionadas hasta ahora quedan determinadas a partir de la matriz de adyacencia. Por ejemplo, para determinar la conectividad  $k_i$  de un nodo podemos escribir

$$k_n = \sum_{i=1}^N \mathbf{A}_{in} = \sum_{i=1}^N \mathbf{A}_{ni}. \quad (2.4)$$

Para el caso de la matriz de adyacencia mostrada en la ecuación (2.1) (red no dirigida) se obtienen los valores mostrados en la Tabla 2.1. También la matriz  $\mathbf{H}$  se puede obtener a partir de la matriz de adyacencia, como se mostrará en la Sec. 2.7.



Nodo	Conectividad
$k_1$	1
$k_2$	2
$k_3$	3
$k_4$	2
$k_5$	3
$k_6$	1
$\langle k \rangle$	2

Tabla 2.1: Conectividad de la red no dirigida mostrada en la Figura 2.1. Notemos que para este ejemplo  $N = 6$  y  $M = 6$ , y en general se tendrá que  $\langle k \rangle = 2M/N$ .

## 2.2. Distribución de la conectividad

Una propiedad muy importante que caracteriza la estructura de una red compleja es la distribución de conectividad  $P(k)$ , que es la probabilidad de que un nodo escogido al azar tenga  $k$  conexiones (o vecinos). Diversos trabajos [1-3,12-25] nos permiten establecer principalmente tres tipos de distribuciones  $P(k)$ , las que determinan tres estructuras o topologías diferentes. Estas son:

$$P(k) = e^{-z} \cdot \frac{z^k}{k!} \quad \text{Distribución de Poisson.} \quad (2.5)$$

$$P(k) = Ce^{-\alpha k} \quad \text{Distribución exponencial.} \quad (2.6)$$

$$P(k) = Ck^{-\gamma} \quad \text{Distribución con ley de potencia.} \quad (2.7)$$

En la figura 2.2, se grafican las distribuciones de Poisson, Exponencial y Ley de potencia. Las gráficas relevantes que se utilizarán más adelante son la distribución de Poisson en escala LIN-LIN, Exponencial en LOG-LIN y Libre de Escala en LOG-LOG (ver figura 2.2).

Al final de los años 50, y principios de los 60, Erdős y Rényi fueron los primeros en estudiar la distribución de una red aleatoria [26-28] (ver Sec. 2.3), obteniendo distribuciones de Poisson. Sin embargo, a pesar de su importancia histórica las redes con topología de Poisson están lejos de ser una representación realista de las redes reales observadas en la naturaleza.

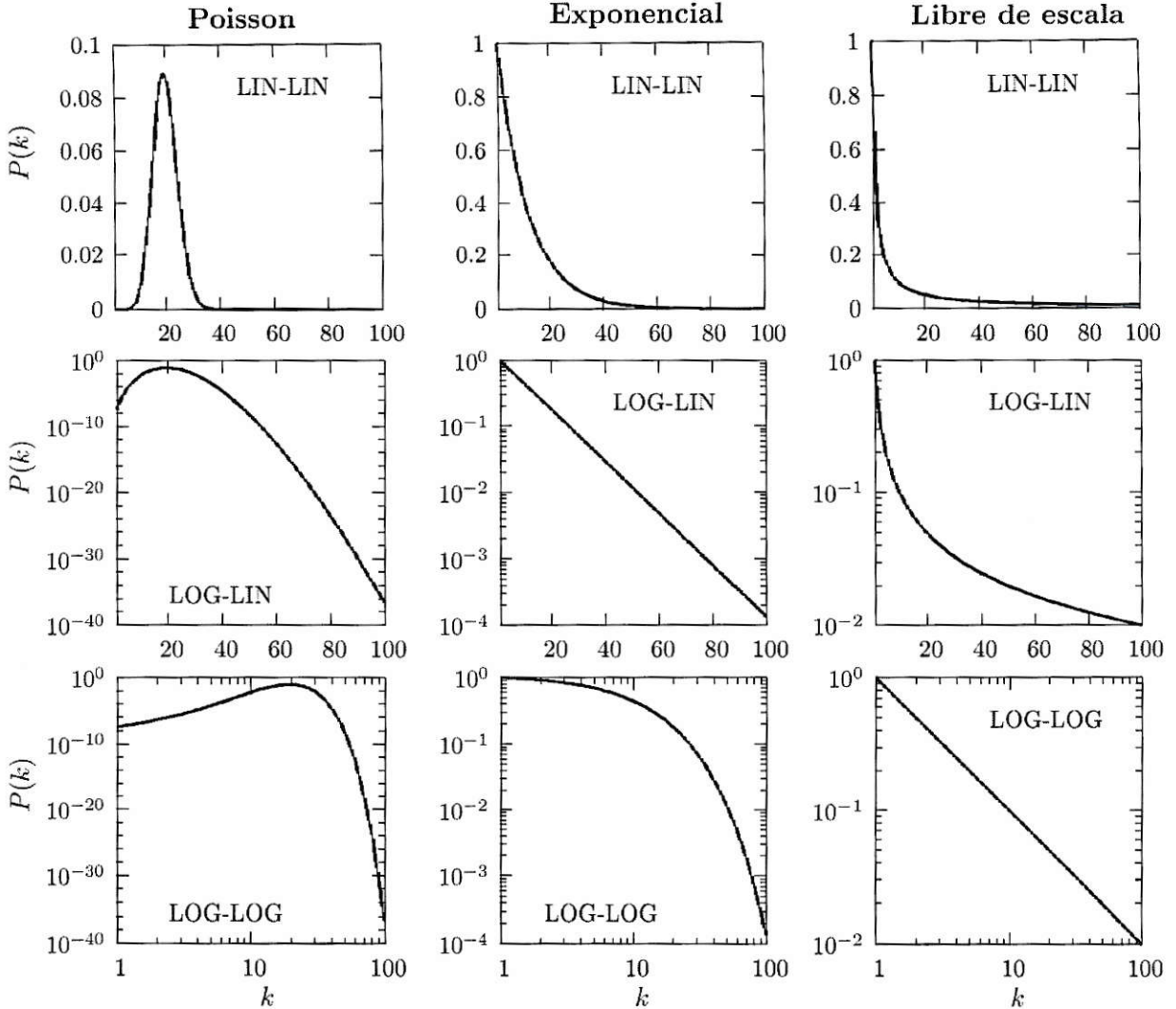


Figura 2.2: Distribuciones Libre de Escala, Poisson y Exponencial graficadas en diferentes escalas. Las representaciones más útiles para nosotros serán LIN-LIN para la dist. Poisson, LOG-LIN para la dist. exponencial y LOG-LOG para la dist. con ley de potencia.

Recién en 1998 se comenzó el estudio sistemático de las propiedades topológicas de las redes reales. Estos estudios mostraron que un número considerable de redes reales existentes en la naturaleza presentan distribuciones de conectividad que siguen leyes de potencia [12–25]. Esto se puede apreciar en la tabla 2.2 que muestra algunos ejemplos de redes reales y el valor del exponente  $\gamma$  [Ec. (2.7)] para dichas redes.

A menudo, las distribuciones con leyes de potencia son llamadas distribuciones libre de escala, esto es debido a la distribución es la misma en todas las escalas, por ejemplo, para un rango de  $k$  tenemos que  $P(k) = Ck^\gamma$  ahora bien si cambiamos la escala en un factor 10 (sin pérdida de generalidad) tenemos  $P(k\gamma 10) = C10^\gamma \cdot k^\gamma \propto k^\gamma$ .

red	$N$	$M$	$\gamma$	Referencia
Red de llamadas telefónicas	47 000 000	80 000 000	2.1	[13, 21]
Redes P2P	880	1 296	2.1	[14, 25]
Redes metabólicas	307	2 359	2.2	[19]
Red de actores de cine	449 913	25 516 482	2.3	[20]
Interacción de proteínas	2 115	2 240	2.4	[17]
Internet	10 697	31 992	2.5	[12, 18]
Red de palabras co-ocurridas	460 902	17 000 000	2.7	[23, 24]
Circuitos electrónicos	24 097	53 248	3.0	[15]
Red de contactos sexuales	2 810	-	3.2	[16, 22]

Tabla 2.2: Ejemplos de redes reales con distribución de conectividad libre de escala. Se indica el valor del  $N^{\circ}$  de nodos  $N$ ,  $N^{\circ}$  de enlaces  $M$  y el exponente  $\gamma$  de la distribución [Ec. (2.7)].

## 2.3. Redes aleatorias

Como primer ejemplo de una red compleja, estudiaremos las redes aleatorias, también llamadas redes tipo Erdős-Rényi. Estas se construyen de la siguiente forma:

- Se define un set de  $N$  nodos.
- Se escoge al azar una pareja de nodos  $v_i$  y  $v_j$ , y si no están conectados se conectan.
- Se repite el proceso anterior hasta obtener  $M$  enlaces, en donde  $M$  será el orden de la red.

Se puede demostrar que toda red aleatoria tiene una conectividad que sigue una distribución de Poisson. Para ello, notemos primero que el número total  $N_p$  de enlaces que se pueden formar en un conjunto de  $N$  nodos es

$$N_p = \frac{1}{2}N(N-1).$$

Como se enlazaron  $M$  parejas de nodos, la probabilidad  $p_e$  de que una pareja arbitraria de nodos esté enlazada es

$$p_e = \frac{M}{N_p} = \frac{2M}{N(N-1)}. \quad (2.8)$$



Consideremos un nodo particular  $v_j$  de la red, escogido al azar. El número total de enlaces que podrían conectar a  $v_j$  es  $N - 1$ , ya que  $v_j$  tiene  $N - 1$  posibles nodos con los cuales conectarse. Como lo indica la ecuación (2.8) la probabilidad de que exista cada enlace en la red es  $p_e$ . Por lo tanto la probabilidad  $P(k)$  de que el nodo  $v_j$  tenga  $k$  de los  $N - 1$  enlaces posibles, suponiendo que la probabilidad de existencia de cada uno de ellos es  $p_e$ , viene dada por la siguiente expresión:

$$P(k) = \binom{N-1}{k} (p_e)^k (1-p_e)^{N-1-k}. \quad (2.9)$$

Ésta es la distribución de probabilidad general para cualquier valor de  $N$  y  $M$ . Pero si consideramos que la red es muy grande,  $N \rightarrow \infty$   $M \rightarrow \infty$ , de tal forma que la cantidad

$$z = \frac{2M}{N}$$

permanezca finita, entonces (ver apéndice A) la distribución (2.9) se transforma en

$$P(k) = e^{-z} \frac{z^k}{k!}. \quad (2.10)$$

Ésta corresponde a una distribución de Poisson con promedio  $z$  [Eq. (2.5)]. Con esto concluimos que, para una red muy grande, el proceso de enlazar pares de nodos en forma aleatoria conduce a una distribución de Poisson. Las figuras 2.3.a, 2.3.b y 2.3.c muestran una representación gráfica de redes aleatorias con  $N = 100$  y con  $M = 50$ ,  $M = 200$  y  $M = 400$  respectivamente, mientras que las figuras 2.3.d, 2.3.e y 2.3.f muestran sus distribuciones de Poisson respectivas

Cabe señalar que durante muchos años se pensó que este mecanismo de formación de redes, en el cual cada pareja de nodos se enlaza aleatoriamente, era adecuado para describir el origen de ciertas redes sociales como las redes de amistades o las redes de contactos sexuales. Después de todo, las amistades se dan por el encuentro casual y aleatorio de las personas que viven en una sociedad. Por tanto, era natural esperar que el modelo aleatorio reprodujera lo que realmente ocurre en las redes sociales, sin embargo, esto resultó ser incorrecto, como lo muestra la tabla 2.2



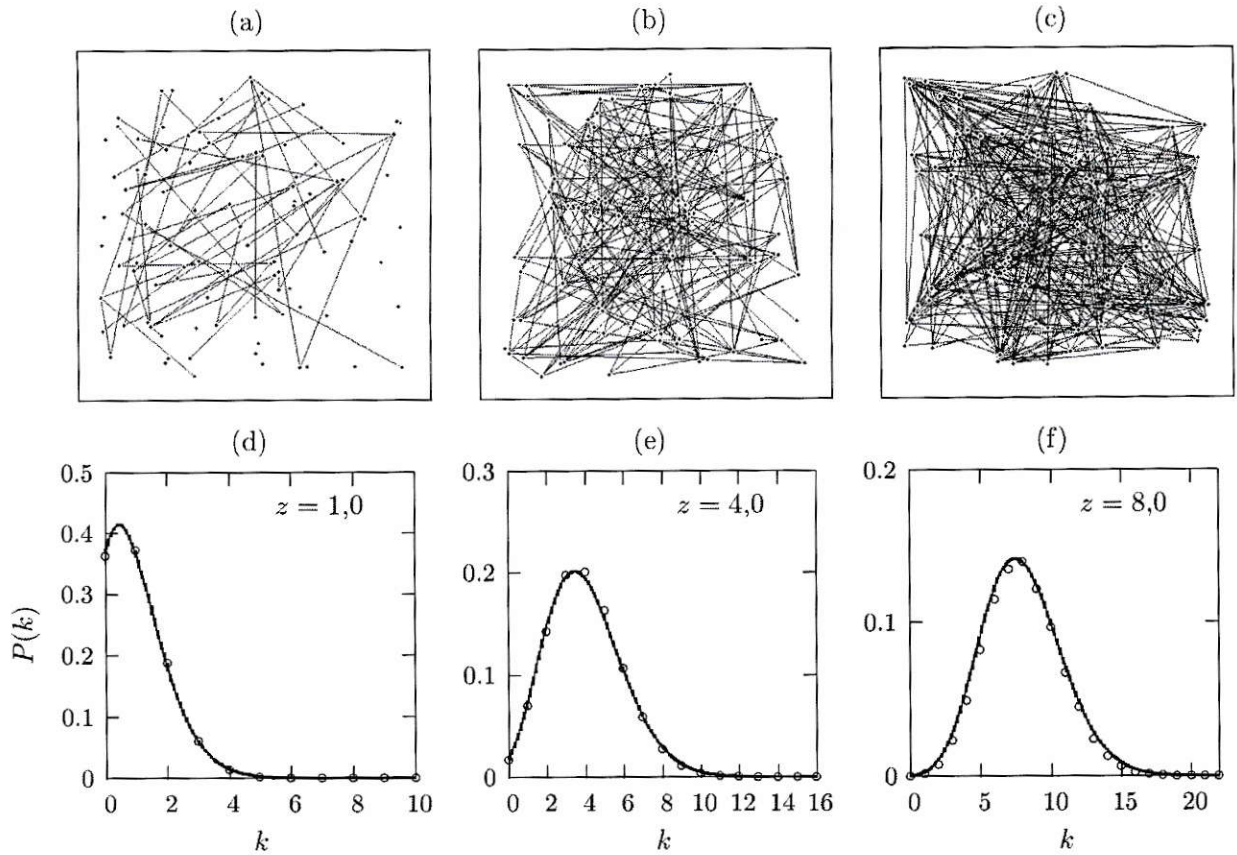


Figura 2.3: Ejemplos de redes aleatorias con  $N = 100$  y con  $M = 50$ ,  $M = 200$  y  $M = 400$  [figuras (a), (b) y (c)]. Junto a ellas se muestran las distribuciones de conectividad ajustadas a su distribución de Poisson respectiva [figuras (d), (e) y (f)].

## 2.4. Crecimiento de redes

El concepto de crecimiento de redes surge en la literatura como una búsqueda de realizar modelos que se ajusten más a las características de las redes reales. En este caso, la red tiene inicialmente un cierto  $N^0$  de nodos y enlaces, y luego se van agregando nuevos nodos y enlaces, siguiendo algún criterio adecuado. Consideremos en particular el siguiente modelo de crecimiento, utilizando un parámetro temporal  $t$  para rotular la evolución de la red:

- Se comienza el proceso de construcción en  $t = 0$  solamente con un nodo  $v_0$ .
- En el instante  $t = 1$  se añade el nodo  $v_1$ , que es conectado al nodo  $v_0$ .
- En  $t = 2$  se añade un nuevo nodo  $v_2$  y se enlaza a  $v_0$  con una probabilidad  $\Pi_0$  o a  $v_1$  con una probabilidad  $\Pi_1$ , con  $\Pi_0 + \Pi_1 = 1$ .

- En  $t = 3$  se añade un nodo  $v_3$  y se enlaza a uno de los nodos  $v_0, v_1$  o  $v_2$ , con probabilidades  $\Pi_0, \Pi_1$  y  $\Pi_2$  respectivamente. Aquí, los valores de  $\Pi_i$  en general son diferentes a los de  $t = 2$  y se cumple que  $\Pi_0 + \Pi_1 + \Pi_2 = 1$ .
- En general, se tiene que en el instante  $t + 1$  se añade el nodo  $v_{t+1}$  el cual se conectará solo con uno de los nodos  $v_0, v_1, v_2, \dots, v_t$  ya existente, en donde la probabilidad de que el nodo  $v_i$  sea seleccionado es  $\Pi_i$ . Debido a que  $\Pi_i$  varía en cada iteración se tiene que  $\Pi_i = \Pi_i(t)$ , debiendo cumplirse que  $\sum_{i=0}^t \Pi_i(t) = 1$ .
- Se repite este proceso muchas veces y se espera a que la distribución de conectividad del sistema tienda a un estado estacionario ( $N$  suficientemente grande).

Notemos que el caso estacionario aludido no implica que la red deje de crecer, sino simplemente que su distribución de conectividad sea una función independiente del tiempo.

Un aspecto que resulta crucial para determinar la topología de la red resultante es el criterio con el cual se determina a qué nodo preexistente se conecta cada nuevo nodo de la red. Una elección posible es que la probabilidad de conexión a un nodo dependa de la conectividad de dicho nodo, es decir:

$$\Pi_i(t) = \Pi(k_i, t) . \quad (2.11)$$

Bajo esta condición, es posible encontrar una ecuación para la evolución de la distribución de conectividad  $P(k)$ , lo cual haremos a continuación.

Denotemos por  $P(n, k, t)$  la probabilidad de que el nodo  $v_n$  tenga  $k$  conexiones en el tiempo  $t$ . Entonces la probabilidad  $P(k, t)$  de que un nodo arbitrario tenga  $k$  conexiones en el tiempo  $t$  es:

$$P(k, t) = \frac{1}{N(t)} \sum_{n=0}^t P(n, k, t) , \quad (2.12)$$

donde  $N(t)$  es el número de nodos de la red en el tiempo  $t$ .

La ecuación que determina la evolución temporal de  $P(n, k, t)$  se obtiene notando que, para que el nodo  $v_n$  cualquiera de la red tenga  $k$  conexiones en el tiempo  $t + 1$ , hay dos procesos que dan origen a ello, estos son:

1. Al tiempo  $t$  el nodo  $v_n$  tenía  $k - 1$  conexiones y fue seleccionado [con una probabilidad  $\Pi(k - 1, t)$ ] para conectarse con el nuevo nodo añadido a la red. Por lo tanto, al tiempo  $t + 1$  el nodo  $v_n$  tendrá  $k$  conexiones.
2. El nodo  $v_n$  ya tenía  $k$  conexiones al tiempo  $t$  y no fue seleccionado para conectarse con el nuevo nodo añadido [lo cual ocurre con un probabilidad  $1 - \Pi(k, t)$ ]. Por lo tanto, en el tiempo  $t + 1$  el nodo  $v_n$  seguirá teniendo  $k$  conexiones.

Tomando ambas contribuciones, la ecuación que determina la evolución de  $P(n, k, t)$  es:

$$\begin{aligned}
 & \underbrace{\hspace{10em}}_{\substack{\text{Probabilidad} \\ \text{de que } v_n \text{ sea} \\ \text{seleccionado}}} & & \underbrace{\hspace{10em}}_{\substack{\text{Probabilidad} \\ \text{de que } v_n \text{ no sea} \\ \text{seleccionado}}} \\
 P(n, k, t + 1) = & \underbrace{P(n, k - 1, t)}_{\substack{\text{Probabilidad} \\ \text{de que } v_n \text{ tenga} \\ k - 1 \text{ conexiones} \\ \text{al tiempo } t}} \cdot \Pi(k - 1, t) + \underbrace{P(n, k, t)}_{\substack{\text{Probabilidad} \\ \text{de que } v_n \text{ tenga} \\ k \text{ conexiones} \\ \text{al tiempo } t}} \cdot (1 - \Pi(k, t)) . \quad (2.13)
 \end{aligned}$$

Por su parte el nodo  $v_n$  apareció en el tiempo  $t = n$  con una sola conexión, la cual da la siguiente condición:

$$P(n, k, t)|_{t=n} = \delta_{k,1} . \quad (2.14)$$

Si sumamos sobre  $n$  a ambos lados de la ecuación (2.13), y utilizando la ecuación (2.12), nos queda

$$\begin{aligned}
 \sum_{n=0}^t P(n, k, t + 1) &= \sum_{n=0}^t \{P(n, k - 1, t)\Pi(k - 1, t) + P(n, k, t)(1 - \Pi(k, t))\} \\
 &= N(t)\{P(n, k - 1, t)\Pi(k - 1, t) + P(n, k, t)(1 - \Pi(k, t))\} . \quad (2.15)
 \end{aligned}$$

Notemos en que el modelo implica que en cada paso se añade un nuevo nodo (comenzando con el nodo  $v_0$  en  $t = 0$ ). Esto implica que  $N(t) = t + 1$ , con lo cual la ecuación (2.15) queda:

$$\sum_{n=0}^t P(n, k, t + 1) = (t + 1)\{P(k - 1, t)\Pi(k - 1, t) + P(k, t)(1 - \Pi(k, t))\} . \quad (2.16)$$

Por otro lado, el lado izquierdo de la ecuación (2.16) se puede reescribir, usando (2.14) y (2.16) queda:

$$\begin{aligned} \sum_{n=0}^t P(n, k, t + 1) &= \sum_{n=0}^{t+1} P(n, k, t + 1) - P(t + 1, k, t + 1) \\ &= \sum_{n=0}^{t+1} P(n, k, t + 1) - \delta_{k,1} \\ &= N(t + 1)P(k, t + 1) - \delta_{k,1} \\ &= (t + 2)P(k, t + 1) - \delta_{k,1} . \end{aligned} \quad (2.17)$$

De (2.16) y (2.17), nos queda finalmente que:

$$(t + 2)P(k, t + 1) - \delta_{k,1} = (t + 1)\{P(k - 1, t)\Pi(k - 1, t) + P(k, t)(1 - \Pi(k, t))\} . \quad (2.18)$$

Ésta es una ecuación de recurrencia que determina  $P(k, t + 1)$  a partir de  $P(k, t)$  y  $P(k, t - 1)$ . Para resolverla es necesario conocer explícitamente la función  $\Pi(k, t)$ , que es la probabilidad de que un nodo ya existente con conectividad  $k$  sea seleccionado para conectarse con el nuevo nodo que se añade en el tiempo  $t$ . Como se verá en las siguientes secciones, diferentes elecciones de la probabilidad  $\Pi(k, t)$  conducen a topologías diferentes de la red resultante.

### 2.4.1. Topología exponencial

En el caso de que cualquiera de los nodos existentes pueda escogerse con la misma probabilidad para conectarse con el nuevo nodo añadido, la probabilidad  $\Pi(k, t)$  es independiente de  $k$  y queda dada por



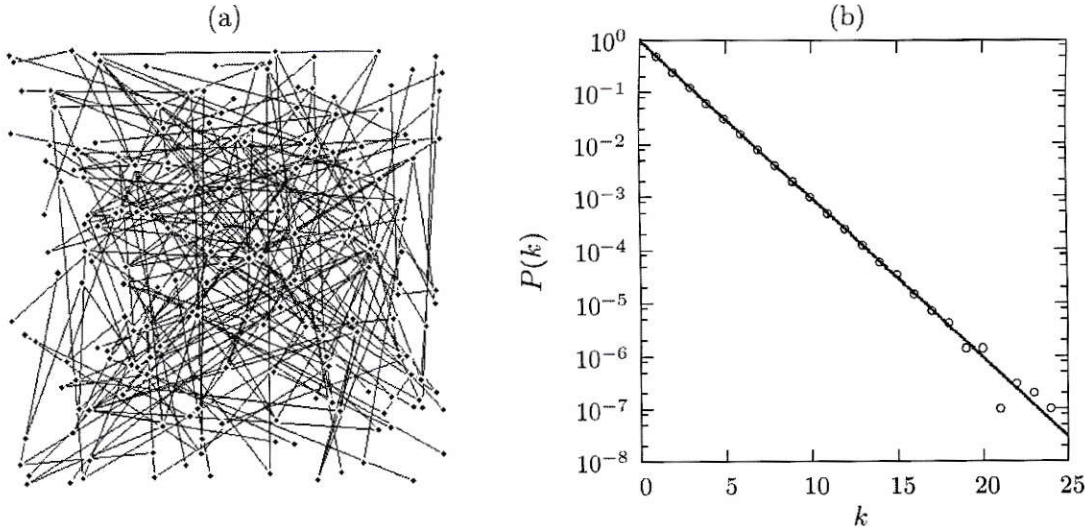


Figura 2.4: (a) Representación gráfica de una red resultante con un modelo de topología exponencial (Sec. 2.4.1) con  $N = 200$  nodos. (b) Distribución de conectividad para el mismo modelo con  $N = 10^6$ . Puntos: resultados de la simulación. Línea sólida: Ec. (2.6).

$$\Pi(k, t) = \frac{1}{N(t)} = \frac{1}{t+1}. \quad (2.19)$$

Sustituyendo (2.19) en (2.18), queda

$$(t+2)P(k, t+1) - \delta_{k,1} = P(k-1, t) + tP(k, t). \quad (2.20)$$

Después de un tiempo muy largo, la función  $P(k, t)$  alcanza un estado estacionario en el que  $P(k, t+1) = P(k, t) = P(k)$ . En este caso la última ecuación queda como

$$P(k) = \frac{1}{2}[P(k-1) + \delta_{k,1}]. \quad (2.21)$$

Como todos los nodos de la red tienen por lo menos una conexión, es claro que  $P(0) = 0$ . Con esta condición inicial, la ecuación de recurrencia anterior tiene la solución

$$P(k) = 2^{-k} = e^{-\ln(2)k}. \quad (2.22)$$

En la figura 2.4 se tienen resultados de simulaciones de este modelo, mostrando claramente cómo la distribución de conectividad tiende a la ecuación (2.22).

## 2.4.2. Topología libre de escala

La topología exponencial no es muy buena cuando se quieren describir algunos tipos de redes reales, por ejemplo, sociales. Esto es debido a que un nuevo individuo no se vincula con alguno de los ya existentes equialeatoriamente, sino que existen ciertas preferencias, Por ejemplo, si un nuevo actor debe decidir entre dos obras en las cuales trabajar, es muy posible que decida participar en el elenco más “reconocido”, ya que tendrá más visibilidad, o por que probablemente signifique que la obra sea de mejor calidad, etc. Entonces, el actor creará con mayor probabilidad, un vinculo con actores más “populares”. Para incorporar este comportamiento, el modelo Barbási-Albert [1,29,30] sugiere que la “popularidad” de un nodo está determinada segun su conectividad, por lo que sugieren que la probabilidad de enlace  $\Pi(k, t)$  debe tener la forma:

$$\Pi(k, t) = \left( \sum_{i=0}^{N(t)} k_i \right)^{-1} k, \quad (2.23)$$

donde  $k_i$  representa la conectividad del  $i$ -ésimo nodo ya existente en el tiempo  $t$ . El factor  $\left( \sum_{n=0}^{N(t)} k_n \right)^{-1}$  simplemente garantiza que la probabilidad  $\Pi(k, t)$  esté normalizada. Al hacer que  $\Pi(k, t)$  sea proporcional a  $k$ , tenemos enlace preferencial, ya que de esta forma, mientras más grande sea la conectividad  $k$  de un nodo, mayor será la probabilidad de conectarse con él. Como en cada paso de tiempo se añade un nuevo nodo a la red con un nuevo enlace, y como se comienza con cero enlaces en  $t = 0$ , se tendrán  $t$  enlaces en el tiempo  $t$ . Debido a que cada enlace aporta dos conexiones, se tiene que para cualquier tiempo  $t > 0$  se cumple que

$$\sum_{i=0}^{N(t)} k_i = 2t, \quad (2.24)$$

y por lo tanto

$$\Pi(k, t) = \frac{k}{2t}. \quad (2.25)$$

Sustituyendo (2.25) en (2.18), nos queda

$$(t + 2)P(k, t + 1) - \delta_{k,1} = (t + 1)\{(k - 1)P(k - 1, t) + (2t - k)P(k, t)\}. \quad (2.26)$$

La distribución de conectividad estacionaria  $P(k)$  se obtiene del límite  $t \rightarrow \infty$  de la última

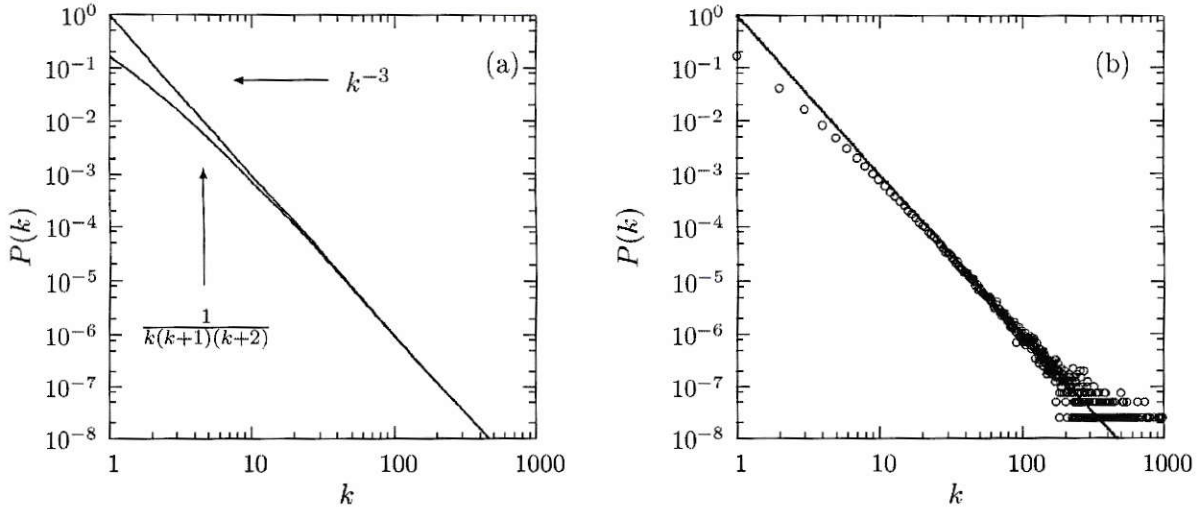


Figura 2.5: (a) Función mostrada en la ecuación (2.28) junto con la función  $f(k) = k^{-3}$ . Se puede ver que para valores altos de  $k$  ambas tienen el mismo comportamiento. (b) Distribución de conectividad de una simulación del modelo aquí expuesto. La simulación fue hecha para  $N = 10^6$  nodos. Nuevamente podemos ver que para valores altos de  $k$  la distribución tiende a la función  $f(k) = k^{-3}$ .

ecuación, lo que conduce a

$$P(k) + \frac{1}{2}\{(kP(k) - (k-1)P(k-1))\} = \delta_{k,1}. \quad (2.27)$$

Como todos los nodos tienen al menos una conexión, podemos imponer que  $P(0) = 0$ , lo cual nos da la solución

$$P(k) = \frac{4}{k(k+1)(k+2)}. \quad (2.28)$$

Aunque esto no es exactamente una ley de potencias, para valores grandes de  $k$  la distribución se comporta como  $P(k) \sim k^{-3}$ . La figura 2.5 muestra una gráfica de la ecuación (2.28), y también se muestran resultados de una simulación del modelo.

Si bien, el modelo de Barbási-Albert fue inspirado en redes sociales, es de gran utilidad para representar otros procesos de crecimiento de redes, en los que los nuevos nodos se enlazan a los nodos con alta “popularidad”. Por ejemplo pensemos en una red de servidores de internet; cada nuevo miembro no se enlaza al servidor más cercano, sino que pondera entre la cercanía y accesibilidad del servidor al resto de la red, accesibilidad que es una medida de la “popularidad” del servidor.



## 2.5. Redes tipo *small-world*

Duncan J. Watts y S. Strogatz publicaron en la revista *Nature* en el año 1998 un artículo en donde se describían por primera vez las llamadas redes de mundo pequeño (*small-world*) [10], poniendo en evidencia la existencia de éstas en la naturaleza y en la sociedad. Ellos definen una red *small-world* como aquella cuya *distancia media*  $\langle \mathbf{H}_{ij} \rangle$  es similar a la de una red aleatoria, pero cuyo *factor de clustering* (Sec. 8) es mayor. Además, en el artículo, proponen un modelo probabilístico mediante el cual es posible obtener este tipo de redes.

### 2.5.1. Redes *small-world* reales

La tabla 3.1 [10], muestra tres tipos de redes. La primera, es el sistema nervioso de la especie *Caenorhabditis elegans*, en donde los nodos representan neuronas y los enlaces, conexiones sinápticas. El número de nodos de esta red es de  $N=282$  y su conectividad promedio es de  $k = 14$ . En segundo lugar se analizó la red de colaboraciones cinematográficas de los EE.UU. Y finalmente la red de distribución de una compañía eléctrica de la zona oeste de los Estados Unidos, en la que cada transformador o subestación representa un nodo y los tendidos de alto voltaje representan los enlaces. Podemos ver a partir de estos resultados que la *distancia media* en los tres casos es similar a la de su análogo en redes aleatorias, sin embargo, en los tres casos el *factor de clustering* es significativamente mayor.

	$N$	$k$	$\langle \mathbf{H} \rangle$	$\langle \mathbf{H} \rangle_a$	$C$	$C_a$
<i>C. elegans</i>	282	14	2.65	2.25	0.28	0.05
Red de Actores	225226	61	3.65	2.99	0.79	0.00027
Red eléctrica	4941	2.67	18.7	12.4	0.08	0.005

Tabla 2.3: Valores de la *distancia media*  $\langle \mathbf{H} \rangle$  y del *factor de clustering*  $C$  para tres redes reales. También se muestran estos mismos términos ( $\langle \mathbf{H} \rangle$  y  $C$ ) obtenidos para redes aleatorias de las mismas características (igual número de nodos  $N$  y de enlaces  $M$ ), los cuales han sido rotulados como  $\langle \mathbf{H} \rangle_a$  y  $C_a$ .

### 2.5.2. Modelo probabilístico de Watts y Strogatz

Watts y Strogatz [10] muestran un modelo probabilístico capaz de generar redes regulares, redes *small-world* o bien redes aleatorias. El modelo funciona de la siguiente manera: se comienza con una red regular como la mostrada en la figura 2.6.a, la cual tiene  $N = 20$



nodos y conectividad igual para todos los nodos  $k_i = 4$ . Luego, con una probabilidad  $p$ , cada enlace es movido de su lugar. De esta forma si  $p = 0$ , cada enlace permanecerá donde mismo, y si  $p = 1$  entonces cada enlace migrará de posición a alguna otra posición escogida al azar, conformando finalmente una red completamente aleatoria. Para valores intermedios de  $p$  surgen estructuras con características *small-world* (baja distancia promedio  $\langle H \rangle$ , alto *factor de clustering*). La gráfica de la figura 2.7 muestra la distancia promedio  $H$  y el factor de clustering  $C$  en función de  $p$ , ambos normalizados a sus respectivos valores para  $p = 0$ .

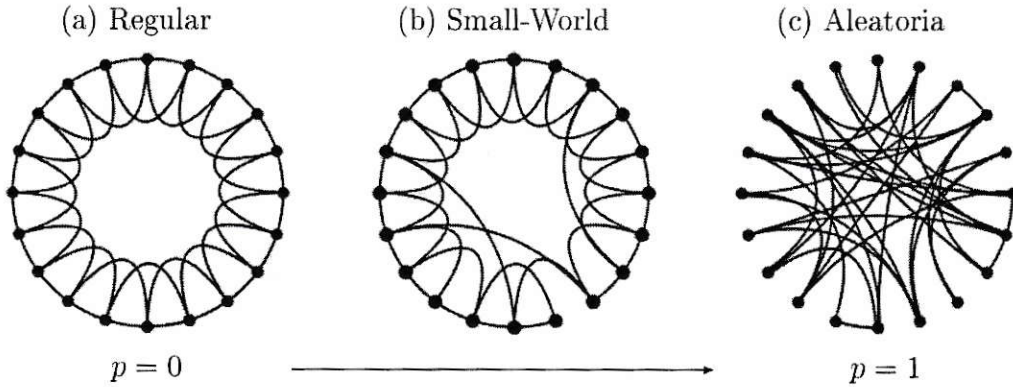


Figura 2.6: Descripción gráfica del modelo planteado por Watts y Strogatz. Se puede ver que para  $p = 0$  la red es completamente ordenada, mientras que para  $p = 1$  la red es completamente aleatoria. Para valores intermedios tenemos redes tipo *small-world*.

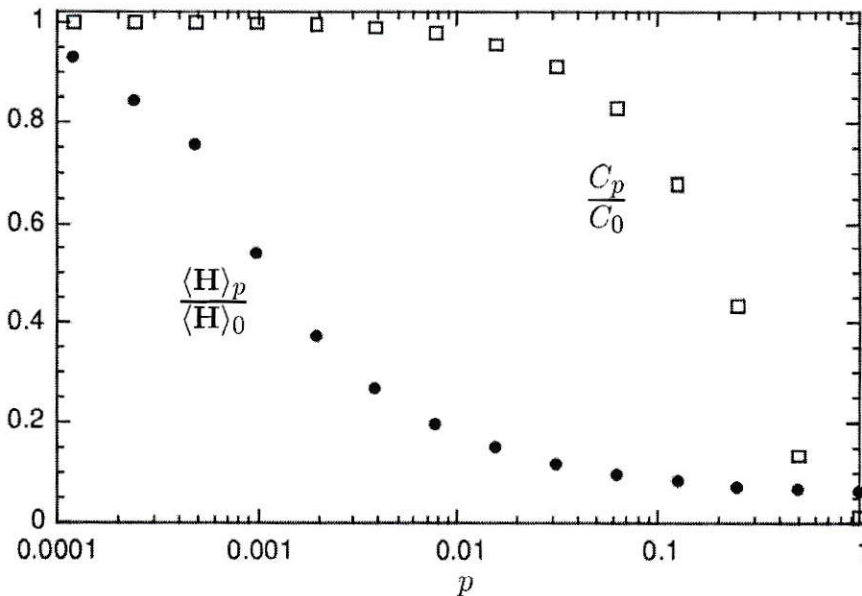


Figura 2.7: Gráfico que indica valores de la *distancia media*  $\langle H \rangle$  y el *factor de clustering*  $C$  en función de  $p$  para el modelo de Watts y Strogatz (Sec. 2.5.2). Ambos valores están normalizados a sus respectivos valores para  $p = 0$ .

## 2.6. Redes espaciales bidimensionales

Una red espacial es una red en la cual cada nodo tiene asociado una coordenada en el espacio. En particular en esta tesis se considerarán redes espaciales bidimensionales, es decir, cada nodo tendrá asociado un punto en el plano  $x-y$ . Debido a esto, a cada par de nodos se les puede asociar una longitud que represente la distancia euclidiana entre ambos, con lo cual se puede introducir la siguiente definición:

**Def. 9** *Matriz D: Matriz de distancias. Sus términos  $D_{ij}$  indican la distancia euclidiana entre los nodos  $v_i$  y  $v_j$ .*

Con respecto a los enlaces de la red, por simplicidad se asumirá que éstos sólo pueden ser líneas rectas que van desde un nodo a otro. Ahora observemos que la Def. 6 (Sec. 2.1) entrega una definición de geodésica que tiene que ver con el mínimo numero de enlaces entre dos nodos. Sin embargo, cuando estamos frente a redes espaciales surge de forma natural una definición alternativa al camino más corto entre nodos, el cual corresponde a la menor distancia geométrica que hay que recorrer para llegar de un nodo a otro siguiendo enlaces de la red. A este camino lo llamaremos *geodésica geométrica*.

**Def. 10 Geodésica geométrica:** *Conjunto de enlaces que conforman el camino más corto (menor longitud geométrica) entre un nodo y otro a través de la red.*

Conviene entonces definir una matriz análoga a  $H_{ij}$ , pero basada en el concepto de *geodésica geométrica*:

**Def. 11 Matriz L:** *Las componentes  $L_{ij}$  de la matriz corresponden a la longitud de la geodésica geométrica entre los nodos  $v_i$  y  $v_j$ . Si no hay forma de conectar  $v_i$  con  $v_j$ , entonces  $L_{ij} = \infty$ .*

La figura 2.8 muestra un diagrama en el cual se ilustran las tres definiciones de distancia que hemos introducido,  $H_{ij}$ ,  $L_{ij}$  y  $D_{ij}$ .

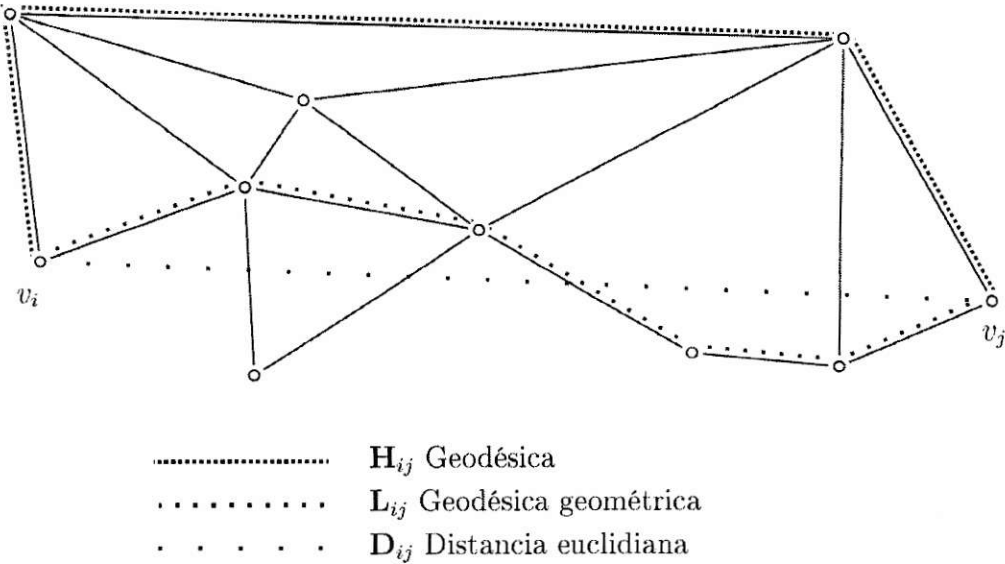


Figura 2.8: Comparación entre  $\mathbf{H}_{ij}$ ,  $\mathbf{L}_{ij}$  y  $\mathbf{D}_{ij}$ , para los nodos  $v_i$  y  $v_j$ .

### 2.6.1. Condiciones de borde periódicas

En este trabajo consideraremos condiciones de borde periódicas sobre el plano que contiene a la red, con lo cual eliminamos la existencia de nodos en el contorno, los cuales pueden presentar características particulares. Algo similar conseguiríamos si utilizáramos redes con condiciones de borde **no** periódicas, pero con  $N \gg 1$ , ya que en ese caso el número de nodos en el contorno de la red se haría muy pequeño en comparación al número de nodos al interior y su efecto sobre las propiedades estadísticas de la red sería despreciable. Sin embargo, para nuestras simulaciones, del orden de  $N \approx 100$  nodos, no podemos suponer esta condición.

Para planos con condiciones de borde **no** periódicas se tiene una sola opción de establecer un enlace entre dos nodos (recordar que los enlaces sólo pueden ser líneas rectas). En cambio, si las fronteras del plano son periódicas, se tiene más de una forma de conectar dos nodos. La figura 2.9 muestra un ejemplo a escala real en el que se indican dos posibles caminos para unir dos nodos. Podemos notar que se genera ambigüedad con respecto a cual es la distancia entre  $v_i$  y  $v_j$  (Sec. 2.6, Def. 9), ya que en el ejemplo esta podría ser de 3.61 cm o bien de 2.24 cm. Para evitar imprecisiones futuras con respecto a la matriz  $\mathbf{D}$  la redefiniremos de la siguiente forma:



**Def. 12 Matriz D:** Se le denominará como la matriz de distancias y sus términos  $D_{ij}$  indican la longitud del camino más corto que une  $v_i$  con  $v_j$  de entre todos los caminos posibles.

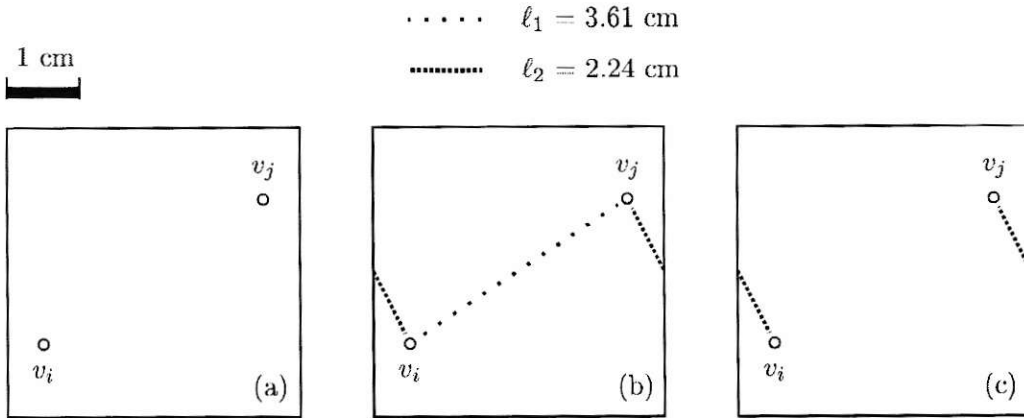


Figura 2.9: (a) Dos nodos de una red cualquiera; (b) dos posibles caminos rectos que conectan los nodos (dos posibles valores de  $D_{ij}$ ); (c) distancia mínima entre los dos nodos, con ello decimos que  $D_{ij} = 2.24 \text{ cm}$ .

## 2.7. Algoritmo de Floyd-Warshall

En esta sección estudiaremos el algoritmo de Floyd-Warshall, el cual será utilizado para determinar las distancias mínimas dentro de la red. Cuando hablamos de distancias mínimas nos referimos a  $H$  y  $L$  que corresponden a la matriz de distancias geodésicas y a la matriz de distancias geodésicas geométricas definidas en las secciones 2.1 y 2.6 (definiciones 7 y 11 respectivamente). Estas matrices serán clave en los cálculos de optimización que se llevarán a cabo en los capítulos 3 y 4.

El algoritmo de Floyd-Warshall [31, 32] (conocido a veces como Algoritmo de Roy-Floyd, puesto que Bernard Roy describió este algoritmo en 1959) es un algoritmo que en general trabaja sobre redes abstractas y fijas <sup>1</sup>, las cuales se caracterizan por ser ponderadas, esto es, en las que cada enlace tiene un costo asociado, y el costo total de moverse de un nodo a otro por medio de la red es la suma de los costos de los enlaces que conforman el camino que los une. Bajo este criterio, el algoritmo de Floyd es utilizado para encontrar cuál es el camino

<sup>1</sup>El algoritmo se limita a hacer mediciones y cálculos sobre una red fija (no dinámica) en la que no hay entrada ni salida de nodos ni enlaces.



más económico que una cada par de nodos de la red. En redes geométricas es común asumir que el “costo” de cada enlace es su longitud, bajo este criterio el algoritmo entrega cuál es el camino más corto entre cada par de nodos.

El algoritmo es totalmente eficiente ya que compara todas las trayectorias posibles a través de la red entre cada par de nodos. Puede hacer esto con solamente  $N^3$  comparaciones (esto es notable considerando que puede haber hasta  $N^2$  enlaces en la red, y cada combinación de enlaces debe ser evaluada para cada par de nodos). Es importante señalar, que debido a que el algoritmo compara todas las trayectorias posibles, el resultado entregado corresponde al mínimo total y no a mínimos locales.

El algoritmo opera de la siguiente forma:

- Primero se define una matriz  $W_{N \times N}$  en la que cada componente representa el costo de cada enlace. Entre los pares de nodos donde no hay enlace el costo se considera infinito. Además,  $W_{ii} = 0 \forall i$ .

- Posteriormente, se realiza la siguiente elección para cada uno de los componentes de la matriz  $W$ :

$$W_{ij} = \text{mín}(W_{ij}, W_{i1} + W_{1j}) .$$

En palabras; se escoge entre “caminar” por el enlace  $e_{ij}$  el cual tiene un costo  $W_{ij}$ , o bien por el camino con “escala” en  $v_1$  ( $e_{i1} + e_{1j}$ ) el cual tiene un costo de  $W_{i1} + W_{1j}$ . Entre ambas posibilidades nos quedamos con la más económica. Esta elección se realiza para todos los pares de nodos de la red.

- Repetimos el proceso anterior, pero esta vez consideramos la conveniencia de hacer “escala” en  $v_2$  y no en  $v_1$ , es decir, se realiza la siguiente elección para cada uno de los componentes de la matriz  $W$ :

$$W_{ij} = \text{mín}(W_{ij}, W_{i2} + W_{2j}) .$$

La entrada de la derecha de esta función es la matriz  $W$  resultante del paso anterior (elección en torno a  $v_1$ ).

- Luego, lo mismo para  $v_3$ :

$$W_{ij} = \text{mín}(W_{ij}, W_{i3} + W_{3j}) .$$

Aquí nuevamente el input de la derecha de la función es la matriz  $W$  resultante del paso anterior.

- La secuencia anterior se continúa para  $v_4, v_5, v_6, \dots$ , hasta  $v_N$ .

- El resultado de este proceso entrega una matriz  $W$  final, en la cual cada componente  $W_{ij}$  representa el menor costo posible para moverse desde  $v_i$  a  $v_j$  por medio de la red. Si no se puede llegar desde  $v_i$  a  $v_j$  por medio de la red, el algoritmo naturalmente nos entrega  $W_{ij} = \infty$ .

El algoritmo de Floyd tiene una infinidad de aplicaciones en diversas áreas [33], como lo son la investigación de operaciones, en la economía, en las telecomunicaciones, en transporte, etc. En general es útil cada vez que podamos modelar la problemática por medio de redes y podamos asociarle un costo a cada enlace. Como ya dijimos, en esta tesis utilizaremos el algoritmo para determinar distancias mínimas (matrices  $\mathbf{H}$  y  $\mathbf{L}$ ).

Para determinar la matriz  $\mathbf{H}$  el algoritmo asume que, si existe enlace entre  $v_i$  y  $v_j$  el costo de transitar por él es *uno*, mientras que, si no existe enlace el costo de transitar por él es *infinito*. Por ello utiliza como semilla una matriz  $W^{inicial}$ , en la cual su componente  $W_{ij}$  es  $W_{ij} = 1$  si es que  $A_{ij} = 1$  o bien  $W_{ij} = \infty$  si es que  $A_{ij} = 0$ . En la diagonal de la matriz todos sus términos son  $W_{ii} = 0$  (no hay costo de moverse de  $v_i$  a  $v_i$ ). De esta forma el algoritmo entregará como resultado final una matriz  $W^{final}$  en la que cada una de sus componentes  $W_{ij}$  indica el mínimo número de enlaces posibles que hay para llegar de  $v_i$  a  $v_j$  (distancia geodésica).

Para el caso de la matriz  $\mathbf{L}$  el algoritmo utiliza como semilla una matriz  $W^{inicial}$  en la cual sus componentes  $W_{ij}$  serán la distancia euclidiana entre  $v_i$  y  $v_j$  si es que existe el enlace ( $A_{ij} = 1 \rightarrow W_{ij} = \mathbf{D}_{ij}$ ), de lo contrario, serán *infinito* ( $A_{ij} = 0 \rightarrow W_{ij} = \infty$ ) y nuevamente los términos de la diagonal serán ceros ( $W_{ii} = 0$ ). De esta forma el algoritmo entregará como resultado final una matriz  $W^{final}$  en la que cada una de sus componentes  $W_{ij}$  indica el mínimo valor geodésico geométrico posible que hay entre  $v_i$  a  $v_j$  (distancia geodésica).

La figura 2.10 muestra un ejemplo a escala real de como opera el algoritmo. Las matrices  $\mathbf{A}$  y  $\mathbf{D}$  están hechas a partir de la red ahí mostrada con su respectiva escala. Luego la figura muestra las matrices semilla que utiliza el algoritmo, y por último, se muestran las matrices  $\mathbf{L}$  y  $\mathbf{H}$  resultantes.

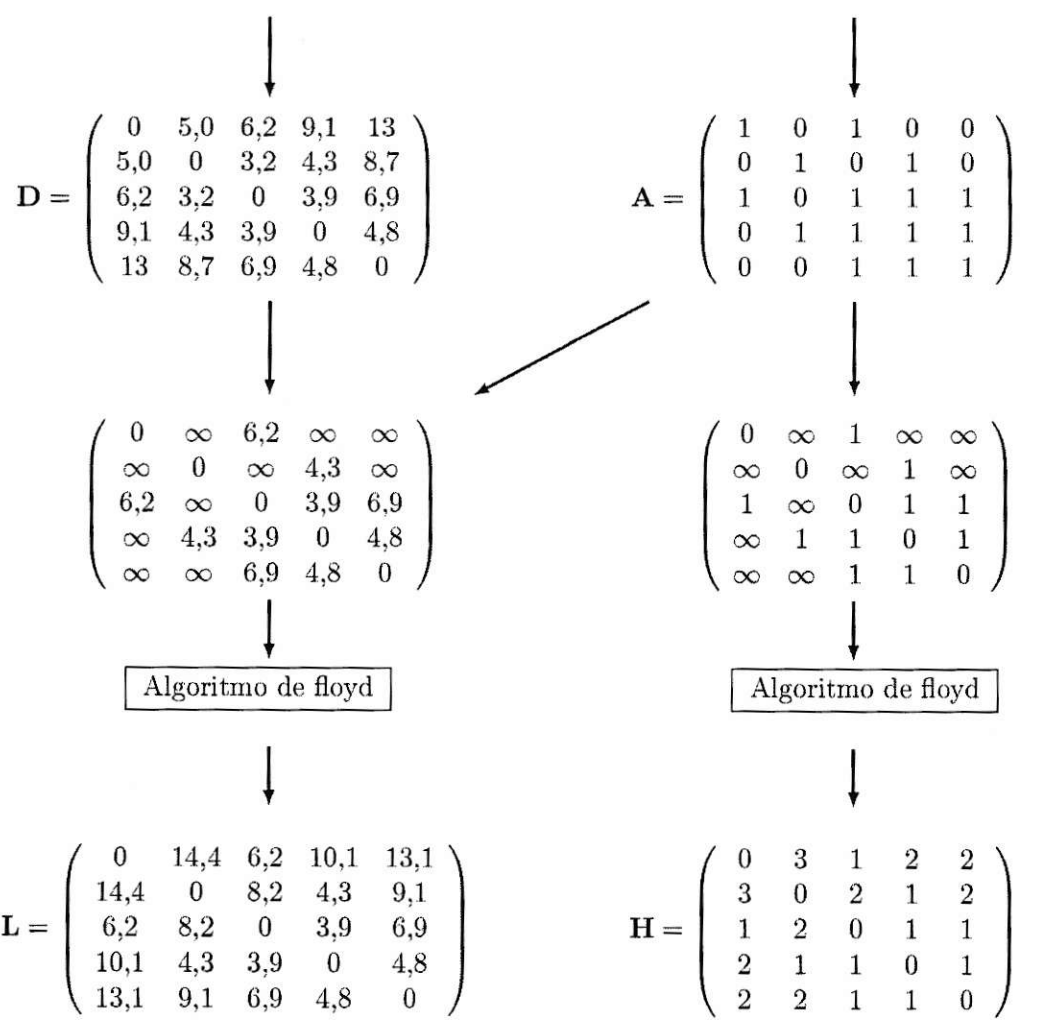
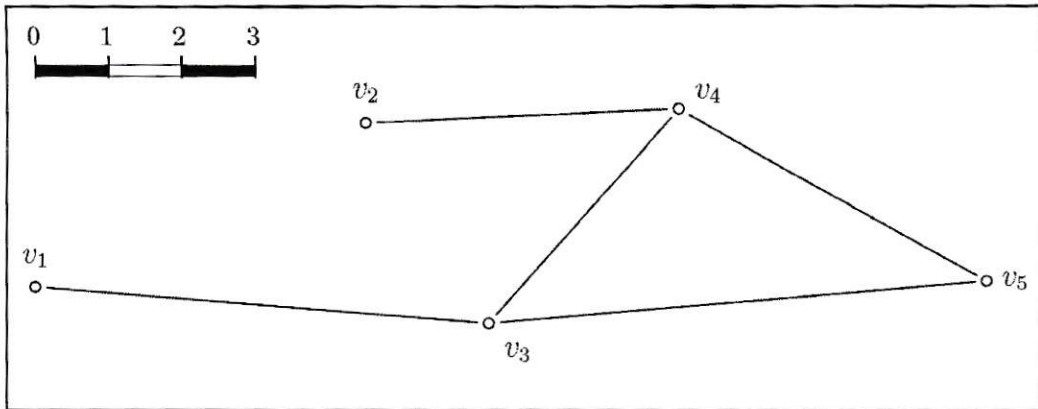


Figura 2.10: Esquema que grafica el uso del algoritmo de floyd.

## Capítulo 3

# Optimización: Mínimo global

Como una manera de proponer modelos que puedan explicar de mejor forma las principales características de las redes reales, en este capítulo, se hace una revisión de algunos modelos basados en optimización. Estos modelos se caracterizan por buscar una configuración de enlaces óptima para un conjunto de nodos fijos en un plano. La manera en que se mide que tan óptima es cada configuración es por medio de la definición de una energía específica la que pondera entre costos de construcción y costos por ineficiencia de la red. Debido a lo intuitivo de los modelos tratados, es de esperar, que sus resultados logren en alguna medida replicar parte de las características de las redes reales.

El capítulo está dividido en tres secciones, en la primera se explican las metodologías usadas para buscar los mínimos de la energía específica señalada, la segunda sección muestra los resultados obtenidos para el primer modelo (Mínimo global I: geodésica geométrica). Por último, la tercera sección muestra los resultados obtenidos para el segundo modelo (Mínimo global II: geodésica).

### 3.1. Modelo de minimización

El problema que se quiere resolver es el establecer los enlaces de un conjunto de  $N$  nodos fijos en un plano periódico. El criterio para establecer los enlaces estará dado por la minimización de una energía específica del tipo:

$$\text{Energía} = E(\lambda, \mathbf{A}, \mathbf{D}) = E_{\lambda}(\mathbf{A}), \quad (3.1)$$



$N$	Combinaciones	$N$	Combinaciones	$N$	Combinaciones	$N$	Combinaciones
1	0	10	$3,5 \times 10^{13}$	40	$6,9 \times 10^{234}$	100	$1,3 \times 10^{1490}$
2	1	15	$4,1 \times 10^{31}$	50	$5,8 \times 10^{368}$	125	$9,9 \times 10^{2518}$
3	8	20	$1,6 \times 10^{57}$	60	$6,7 \times 10^{532}$	150	$1,0 \times 10^{3364}$
4	64	25	$2,0 \times 10^{90}$	70	$9,8 \times 10^{726}$	175	$1,6 \times 10^{4583}$
5	1024	30	$8,9 \times 10^{130}$	80	$1,8 \times 10^{951}$	200	$3,4 \times 10^{5988}$

Tabla 3.1: Tabla que muestra el número de configuraciones posibles de la matriz  $\mathbf{A}$  en función de su tamaño. Dicho de otra forma, la tabla indica la cantidad de redes posibles de construir a partir de  $N$  nodos.

en donde  $\lambda$  es un parámetro, el cual permanecerá fijo en cada simulación y que permitirá ponderar la importancia de cada tipo de costo y  $\mathbf{D}$  es la matriz de distancias, la cual también permanece constante. Con esto, la energía de la ecuación (3.1) sólo depende de la matriz de adyacencia  $\mathbf{A}$ .

En principio la resolución de este problema implica la evaluación de la energía de todas las configuraciones posibles de la matriz  $\mathbf{A}$ , un problema inmanejable considerando la tabla 3.1 que muestra la cantidad de combinaciones de  $\mathbf{A}$  en función de su tamaño  $N \times N$  ( $N =$  número de nodos). Es por ello que se utilizará un método de carácter estadístico, disminuyendo significativamente los tiempos de cálculo, pero que sólo nos brindará una aproximación a la solución exacta del problema.

### 3.1.1. Algoritmo de minimización $\mathcal{M}$

Como primera aproximación en la resolución del problema se utilizará un método iterativo, el cual designaremos por la letra  $\mathcal{M}$ . Este algoritmo parte de una condición inicial  $\mathbf{A}$  y entrega como resultado una matriz  $\bar{\mathbf{A}}$  cumpliéndose siempre que:

$$\mathcal{M}(\mathbf{A}) = \bar{\mathbf{A}} \quad \implies \quad E_\lambda(\bar{\mathbf{A}}) \leq E_\lambda(\mathbf{A}) \ .$$

El algoritmo mapea el espacio de matrices y se va moviendo en la dirección en la cual la derivada de la función  $E(\mathbf{A})$  es negativa, deteniéndose en algún mínimo de la función. Como es de esperar, es altamente probable que este mínimo represente uno local y no uno global, siendo esta probabilidad dramáticamente creciente con el número de nodos  $N$ . A continuación se explica brevemente el proceso iterativo:

1. Se comienza con la matriz de adyacencia elemental  $\mathbf{A}_0$ , la cual asume que todos los nodos están conectados entre sí.

$$\mathbf{A}_0 = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 1 & 1 & & 1 \\ 1 & 1 & 1 & & 1 \\ \vdots & & & \ddots & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}$$

2. Se elimina un enlace “al azar”.
3. Se calcula la energía de la nueva configuración.
4. Si la energía aumenta se repone el enlace eliminado; si no, se mantiene la configuración.
5. Se escoge nuevamente un enlace al azar, que no haya sido escogido anteriormente <sup>1</sup> se elimina y se vuelve al paso 3.
6. Después de recorrer en forma aleatoria todos los enlaces de la red, se termina el proceso.

La figura 3.1 muestra un esquema del proceso. Debido a que la secuencia de selección de los enlaces es aleatoria, el proceso no siempre converge a la misma solución. Es más, para los valores de  $N$  utilizados en esta tesis, la probabilidad de llegar a la misma solución en dos procesos se hace prácticamente nula; diferentes secuencias conducen a diferentes mínimos. Este tipo de algoritmo es comúnmente utilizado por otros autores para encontrar mínimos en espacios multidimensionales, sin embargo, varias veces es introducido un factor aleatorio a la energía con el fin de evitar quedar “atrapado” en mínimos locales (Montecarlo con temperatura) [34]. Nosotros, para evitar esto, haremos uso de un algoritmo genético, el cual es explicado en la próxima subsección.

---

<sup>1</sup>Debido a las características que tendrá la energía específica, no hay razón para que un enlace que antes no se haya eliminado ahora sí se justifique hacerlo.

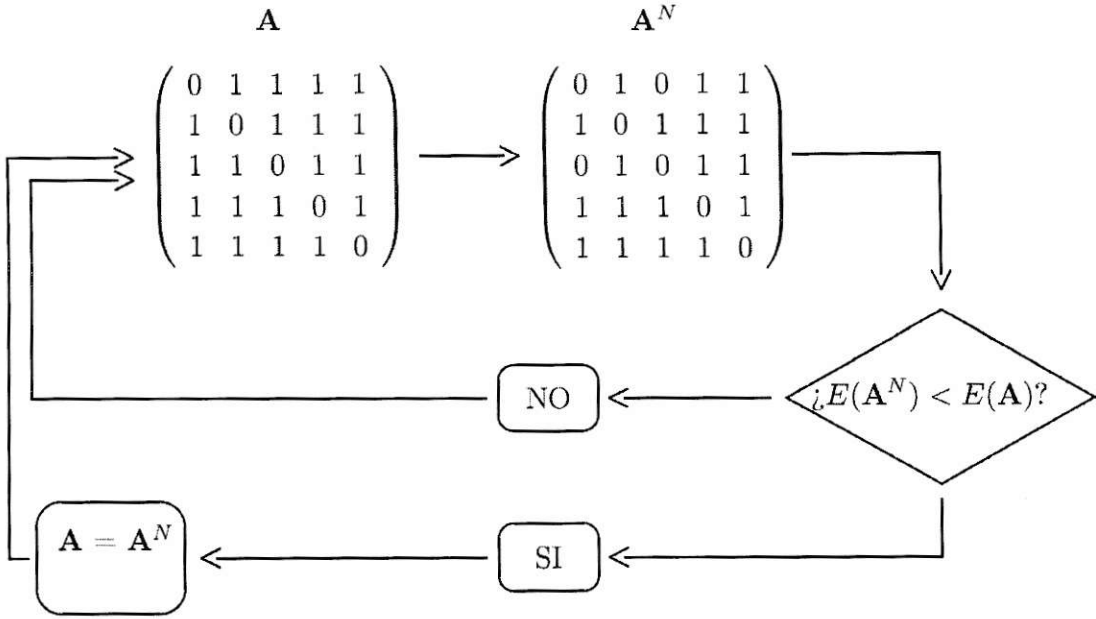


Figura 3.1: Esquema del algoritmo de minimización (Sec. 3.1.1)

### 3.1.2. Algoritmo genético

Como ya se indicó, las soluciones entregadas por  $\mathcal{M}$  no representan la solución óptima al problema, tratándose sólo de un mínimo local. Para mejorar, en parte, este problema, se utilizó un algoritmo genético, el cual busca una mejor solución al problema a partir de un conjunto de soluciones locales. El uso de este tipo de algoritmos para la búsqueda de soluciones en espacios multidimensionales es bastante común debido a su eficacia [35–37]. A continuación se explican brevemente los pasos seguidos por el algoritmo genético aquí utilizado:

1. Se busca una población de  $n$  soluciones locales por medio de la función  $\mathcal{M}(\mathbf{A}_0)$ . A este conjunto se le denominará  $\{\mathbf{A}_i\}$  con  $i = 1, \dots, n$ . Recordemos que la función  $\mathcal{M}$  es de carácter aleatorio, por lo que se tiene que  $\mathbf{A}_i \neq \mathbf{A}_j$  (con  $i \neq j$ ). Al conjunto  $\{\mathbf{A}_i\}$  se le denominará *población semilla*.
2. Se cruzan entre sí todos los pares de matrices semilla. El cruce o la unión entre dos matrices aquí es definida de la siguiente forma:

$$\mathbf{A} \equiv \mathbf{B} \cup \mathbf{C} \quad \iff \quad \mathbf{A}_{ij} = \max(\mathbf{B}_{ij}, \mathbf{C}_{ij})$$

3. Cada matriz resultante del paso anterior es llevada a algún mínimo cercano por medio del algoritmo  $\mathcal{M}$ .
4. En definitiva, el paso anterior entrega una población de  $n(n-1)/2$  matrices, de las cuales se seleccionan las  $n$  matrices más óptimas, que pasan a conformar el nuevo conjunto semilla para el paso 2.
5. Después de algunas generaciones, cuando la dispersión de  $E_\lambda$  no varía mucho en la población, se detiene el proceso y se selecciona la matriz óptima de la población.

La figura 3.2 muestra un esquema de cómo funciona el algoritmo genético utilizado. Es importante mencionar que, si bien este algoritmo genético nos entrega mejores mínimos, nada nos asegura encontrar la solución óptima al problema.

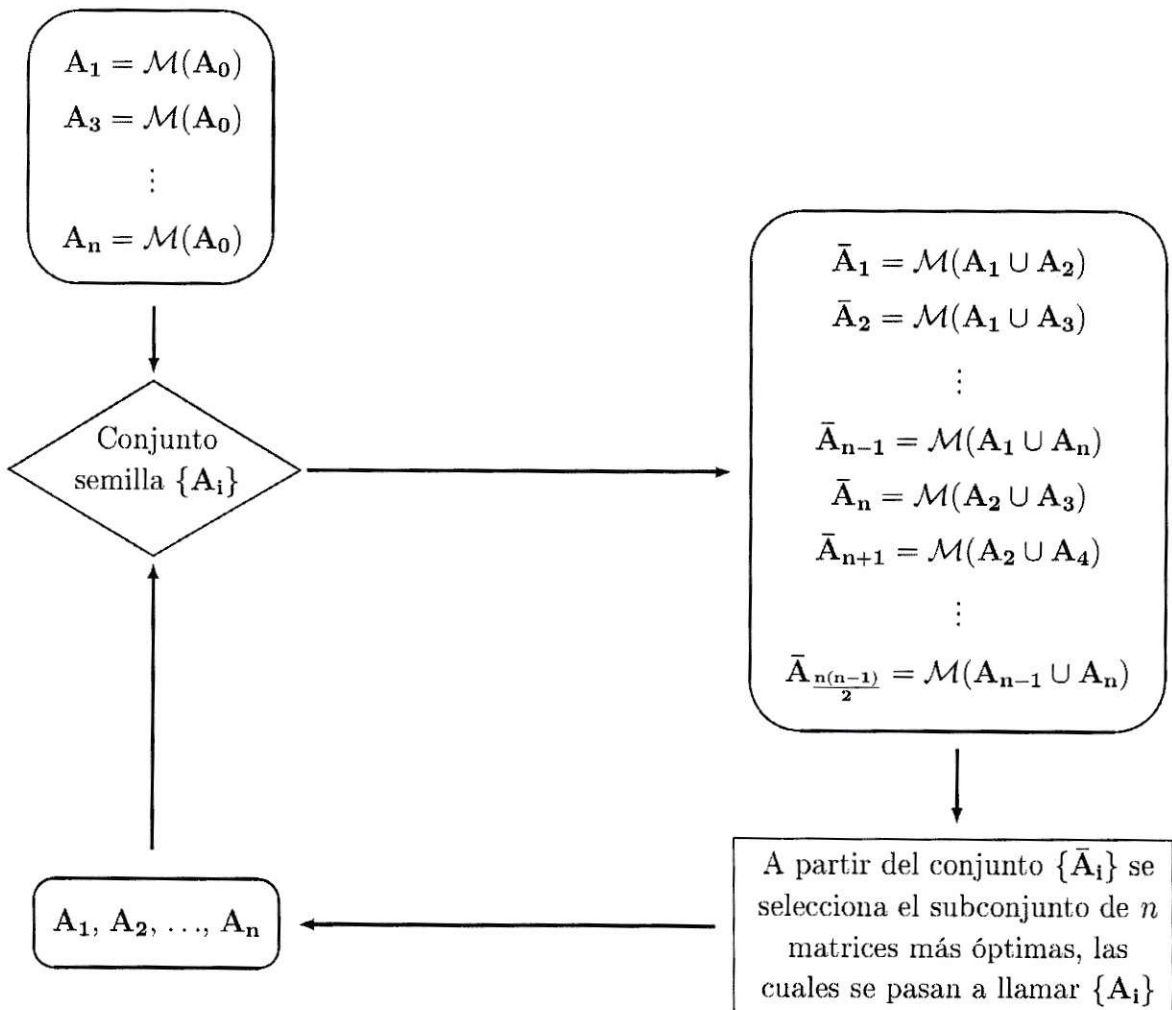


Figura 3.2: Esquema del funcionamiento del algoritmo genético 3.1.2.



## 3.2. Modelo global I: Geodésica geométrica

La primera serie de simulaciones está asociada a encontrar la matriz de adyacencia que minimiza la energía dada por la siguiente ecuación [5, 38, 39]

$$E = (1 - \lambda)D + \lambda L, \quad (3.2)$$

donde

$$D = \sum_{i < j} \mathbf{A}_{ij} \mathbf{D}_{ij}, \quad (3.3)$$

$$L = \sum_{i < j} \mathbf{L}_{ij}. \quad (3.4)$$

Donde  $\mathbf{A}$  es la matriz de adyacencia (Def. 5, Sec. 2.1),  $\mathbf{D}$  es la matriz de distancias (Def. 12, Sec. ??) y  $\mathbf{L}$  es la matriz de *geodésicas geométricas* (Def. 11, Sec. 2.6. Buscar una configuración de  $\mathbf{A}$  que minimice ésta energía, es buscar una red que optimice ponderando entre costos de construcción y costos por ineficiencia, donde los costos por construcción son medidos por el volumen total de enlaces y la eficiencia por la suma de las distancias que hay que recorrer entre un nodo y otro. Es por ello que el modelo es útil cuando se tiene una red de transporte en la cual la eficiencia depende de que los trayectos que deba recorrer la información sean lo más cortos posibles, no importando el número de “escalas” que la información deba hacer. Un ejemplo de esto es una red de carreteras. En ésta, lo relevante desde el punto de vista de la eficiencia son los tiempos de viaje, los cuales, dependerán del largo de las distancias totales recorridas, siendo despreciable el efecto del número de intersecciones.

La ecuación (3.2) depende del parámetro  $D$ , el cual depende de la matriz  $\mathbf{A}$  y  $\mathbf{D}$ , esto es  $D = D(\mathbf{A}, \mathbf{D})$ . Por otro lado, el término  $L$  depende de la matriz  $\mathbf{L}$ , la cual depende de  $\mathbf{A}$  y de  $\mathbf{D}$ <sup>2</sup>, por lo que también tenemos que  $L = L(\mathbf{A}, \mathbf{D})$ . Por último, como ya se mencionó en el capítulo anterior, la matriz  $\mathbf{D}$  es fija, de modo que

$$E = (1 - \lambda)D + \lambda L = E(\lambda, \mathbf{A}, \mathbf{D}) = E_\lambda(\mathbf{A}), \quad (3.5)$$

que es la forma (3.1). La figura 3.3 muestra una serie de redes que resultan de las simulaciones para diferentes valores de  $\lambda$ . Se puede apreciar que el número de enlaces aumenta con  $\lambda$ ,

<sup>2</sup>Recordar la sección 2.7 en donde por medio del algoritmo de Floyd se obtiene  $\mathbf{L}$  a partir de  $\mathbf{A}$  y de  $\mathbf{D}$

resultado que es de esperar, debido a que la ecuación (3.2) indica que el costo de los enlaces disminuye a medida que aumenta  $\lambda$ .

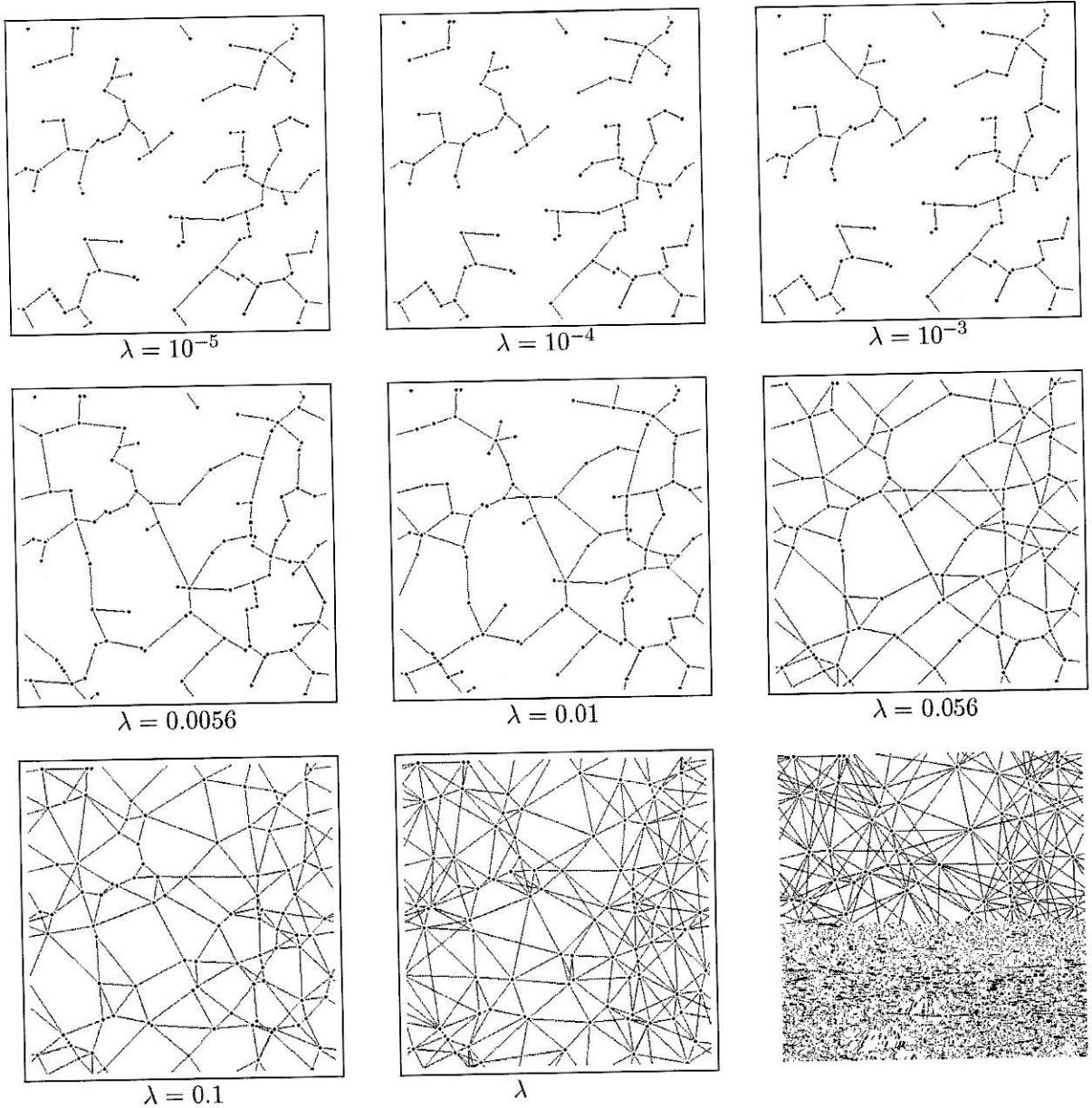


Figura 3.3: Árboles resultado de la minimización de los costos para distintos valores de  $\lambda$  ( $N=80$ ). Podemos ver que el número de "caminos" utilizados es bajo (bajo valor de  $L$ ) cuando los caminos son más largas entre un nodo y otro (alto valor de  $D$ ), lo que es "barato" de construir pero "ineficiente". En cambio, cuando los caminos son más cortos (bajo valor de  $D$ ), el número de caminos es mayor (alto valor de  $L$ ), por lo que la red es más costosa pero "eficiente". Entre medio, tenemos toda una gama de configuraciones intermedias de los dos factores de costo. Cada panel muestra la misma configuración de nodos.

En todos los casos mostrados en la figura 3.3, la conectividad de la red sigue una distribución poissoniana. El máximo de la distribución se desplaza hacia la derecha a medida que aumenta el valor de  $\lambda$ , esto es debido a que a mayor valor de  $\lambda$  es mayor el número de enlaces, lo que a su vez implica un mayor valor de la conectividad media  $\langle k \rangle$  de los nodos.

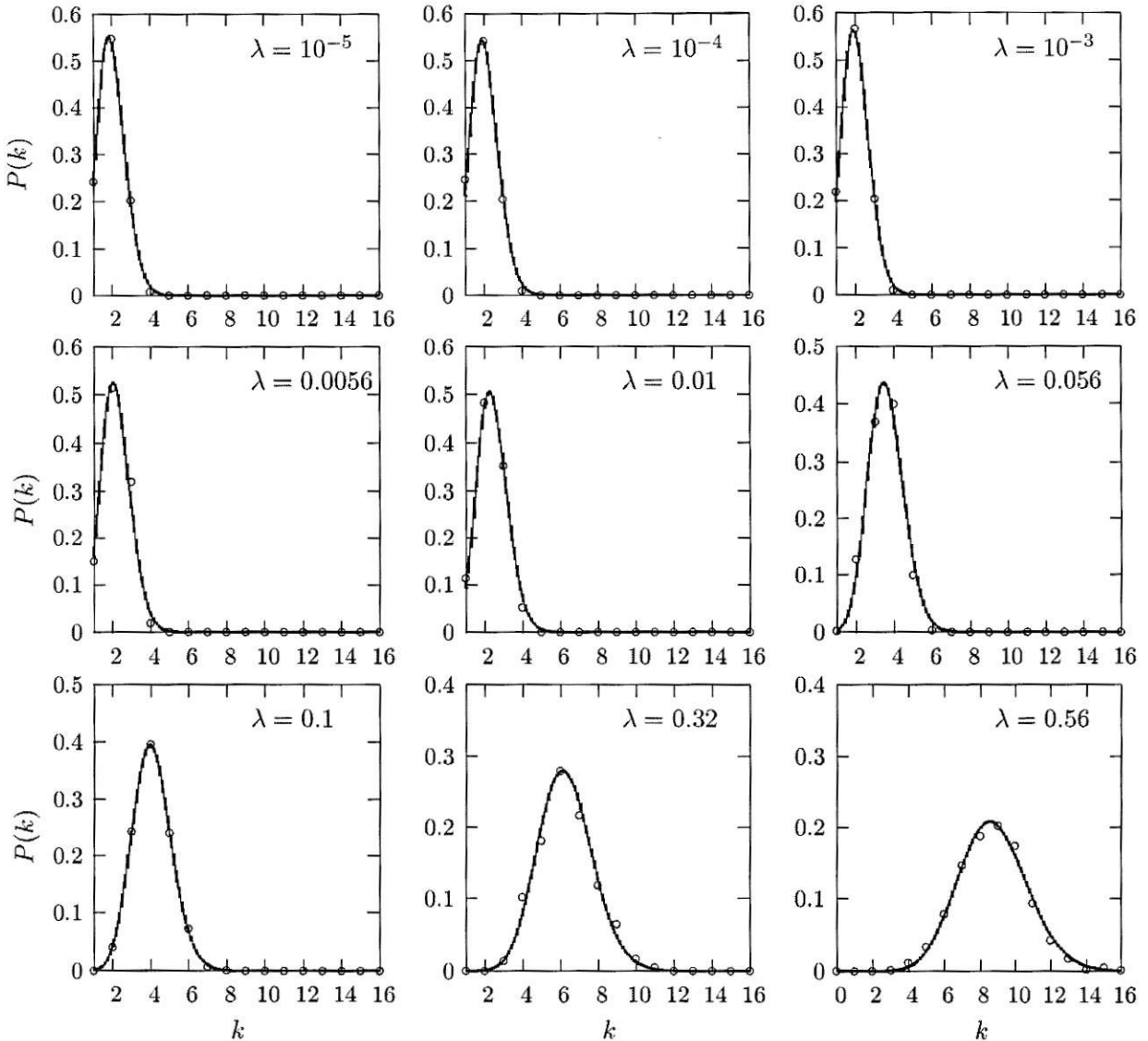


Figura 3.4: Distribución de la conectividad para distintos valores de  $\lambda$  (mismos que figura 3.3). Cada distribución es el promedio de 20 simulaciones, y para cada una de ellas se utilizaron  $N = 80$  nodos. En cada una de las gráficas se muestra un ajuste a una curva poissoniana (ver ecuación 2.5). El máximo se desplaza hacia la derecha a medida que aumenta  $\lambda$ .



La figura 3.5 muestra los valores de  $D$  y  $L$  en función de  $\lambda$ , normalizados a su máximo valor. Como es de esperar  $D$  aumenta y  $L$  disminuye con  $\lambda$ , debido a que la ecuación (3.2) pondera de forma inversa entre ellos. En tanto, en la figura 3.6 se grafica el coeficiente de clustering  $C$  y la distancia media  $\langle H \rangle$  en función de  $\lambda$ .

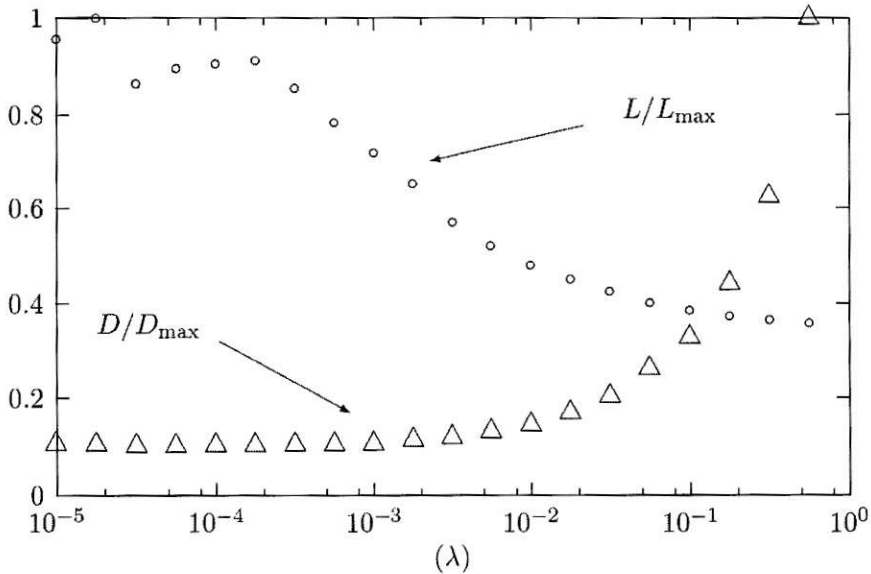


Figura 3.5: Gráfica de  $D$  y  $L$  en función de  $\lambda$ . Cada punto es el promedio de 20 simulaciones y para cada una de ellas se utilizaron  $N = 80$  nodos.

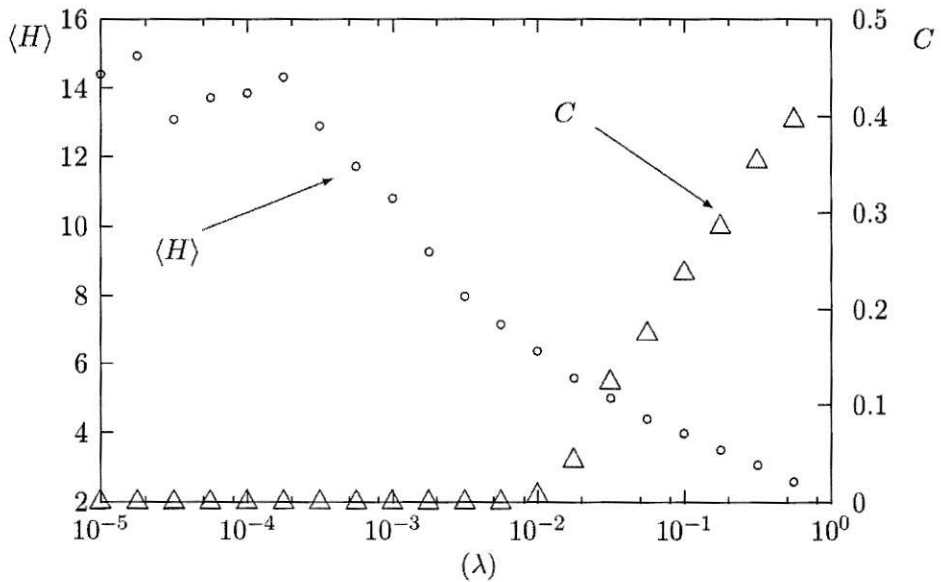


Figura 3.6: Gráfica de  $\langle H \rangle$  y  $C$  (coeficiente de clustering) en función de  $\lambda$ . Cada punto es el promedio de 20 simulaciones y para cada una de ellas se utilizaron  $N = 80$  nodos.



De la figura 3.6 se sigue que, a partir de  $\lambda \approx 10^{-2}$ , la red adquiere características de mundo pequeño. Para poner esto en evidencia, la tabla 3.2 muestra los valores de  $\langle \mathbf{H} \rangle$  y  $C$  para  $\lambda > 10^{-2}$ . También se muestran los valores de  $\langle \mathbf{H} \rangle$  y  $C$  para redes aleatorias (sección 2.3) y de redes del modelo Watts-Strogatz para  $p = 0.1$  (sección 2.5.2) con igual número de nodos  $N$  y enlaces  $M$ . Se puede observar que el factor de clustering es mucho mayor que el de las redes aleatorias, y sin embargo la distancia promedio es bastante similar. Ahora, comparando estos mismos términos con los del modelo Watts-Strogatz, se puede observar que los resultados son similares, mostrando así que para estos valores de  $\lambda$  la red es de mundo pequeño.

Modelo Global I			Red Aleatoria		Red Modelo W-S		$M$
$\lambda$	$\langle \mathbf{H} \rangle$	$C$	$\langle \mathbf{H} \rangle$	$C$	$\langle \mathbf{H} \rangle$	$C$	
0.562	2.619	0.396	2.238	0.109	2.523	0.382	348
0.316	3.103	0.354	2.553	0.079	2.964	0.343	253
0.177	3.511	0.286	2.877	0.062	3.451	0.317	199
0.100	3.976	0.238	3.229	0.049	3.927	0.269	164
0.056	4.391	0.174	3.633	0.042	4.732	0.249	139

Tabla 3.2: Valores de  $\langle \mathbf{H} \rangle$  y  $C$  en función de  $\lambda$ . Se puede ver que se tiene un alto factor de *clustering* y una baja distancia promedio, propiedades típicas de redes *small-world*. Conjuntamente se muestran los valores de  $\langle \mathbf{H} \rangle$  y  $C$  de redes aleatorias (sección 2.3) y de redes del modelo Watts-Strogatz con  $p = 0.1$  (sección 2.5.2) del mismo tamaño. Por último se indica el orden  $M$  de las redes en cada caso ( $N=80$  para todas las redes y todos los casos).

### 3.3. Modelo global II: Geodésica

En esta serie de simulaciones se está interesado en encontrar la matriz de adyacencia que minimiza la energía dada por la siguiente ecuación [5, 40]:

$$E(\lambda) = (1 - \lambda)D + \lambda H , \quad (3.6)$$

donde,

$$D = \sum_{i < j} \mathbf{A}_{ij} \mathbf{D}_{ij} , \quad (3.7)$$

$$H = \sum_{i < j} \mathbf{H}_{ij} . \quad (3.8)$$

Donde  $\mathbf{A}$  es la matriz de adyacencia (Def. 5, Sec. 2.1),  $\mathbf{D}$  es la matriz de distancias (Def. 12, Sec. ??) y  $\mathbf{H}$  es la matriz de *geodésicas* (Def. 7, Sec. 2.1. Este modelo, a diferencia del de la sección anterior, es relevante cuando se tienen redes de transporte en las cuales la eficiencia depende de que la información haga el menor número de “escalas” en su viaje, no importando el largo final del trayecto que ésta deba recorrer. Un ejemplo de esto puede ser una red de rutas de vuelo pequeña (por ejemplo al interior de un país). En ésta, los tiempos de vuelo son pequeños en comparación a los tiempos de espera cada vez que el avión hace una escala. Además, los pasajeros siempre tienden a escoger vuelos en el que el número de escalas sea menor y el sólo hecho de aterrizar implica costos de operación.

Nuevamente, se tiene que  $D = D(\mathbf{A}, \mathbf{D})$ . Y también se tiene que  $H = H(\mathbf{A}, \mathbf{D})$ <sup>3</sup>. Por lo que nuevamente podemos decir que:

$$E = (1 - \lambda)D + \lambda H = E(\lambda, \mathbf{A}, \mathbf{D}) = E_\lambda(\mathbf{A}) . \quad (3.9)$$

Lo que la convierte en una función del tipo (3.1). La figura 3.7 muestra una serie de redes resultado de la minimización hecha en función de la energía mostrada en la ecuación (3.6) para diferentes valores de  $\lambda$ . A diferencia de las redes obtenidas por el modelo anterior, éstas tienden a formar estrellas. Éstas configuraciones tipo estrella, son las que minimizan  $H$ .

---

<sup>3</sup>Recordar la sección 2.7 en donde por medio del algoritmo de Floyd se obtiene  $\mathbf{H}$  a partir de  $\mathbf{A}$

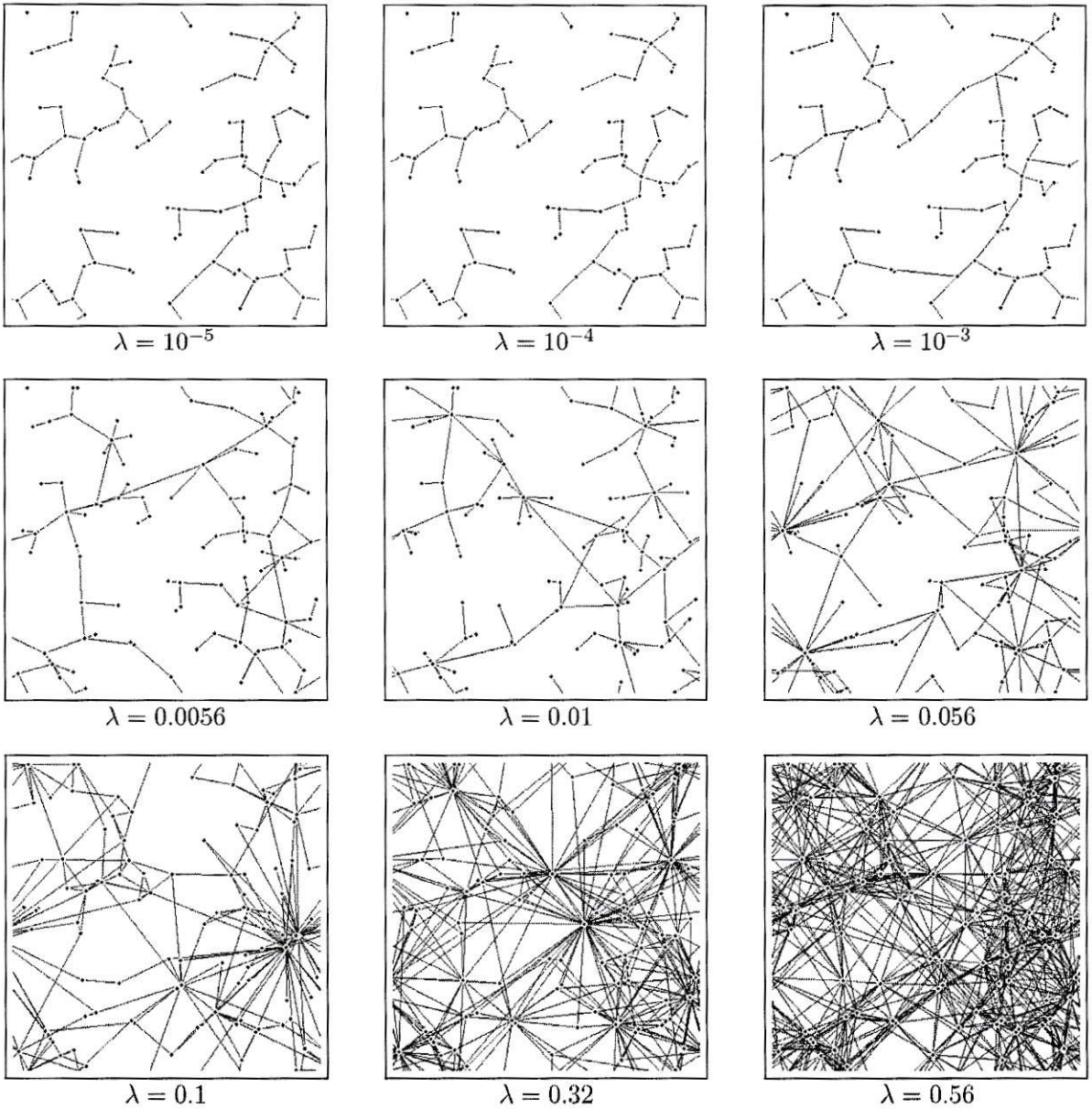


Figura 3.7: Árboles resultado de la minimización de la ecuación 3.6 para distintos valores de  $\lambda$  ( $N=80$ ). La conducta exhibida es similar a la de la figura 3.3, en el sentido de que para valores de  $\lambda \approx 0$  el volúmen de “camino” utilizado es bajo (bajo valor de  $D$ ), en contraste con  $\lambda \approx 1$  en donde el volúmen de “camino” es mayor (alto valor de  $D$ ). Por otro lado, tenemos que la suma de las geodésicas ( $H$ ) crece con  $\lambda$ . Ambas observaciones eran de esperar debido a que  $\lambda$  pondera inversamente entre  $D$  y  $H$  (crece una y disminuye la otra, y viceversa).

En la figura 3.8 se muestran las distribuciones de conectividad para los valores de  $\lambda$  de la figura 3.7. La distribución es similar, aunque no igual, a una distribución poissoniana, cuyo máximo se desplaza a la derecha al aumentar lambda. Esto se debe a que la conectividad promedio aumenta con  $\lambda$ .



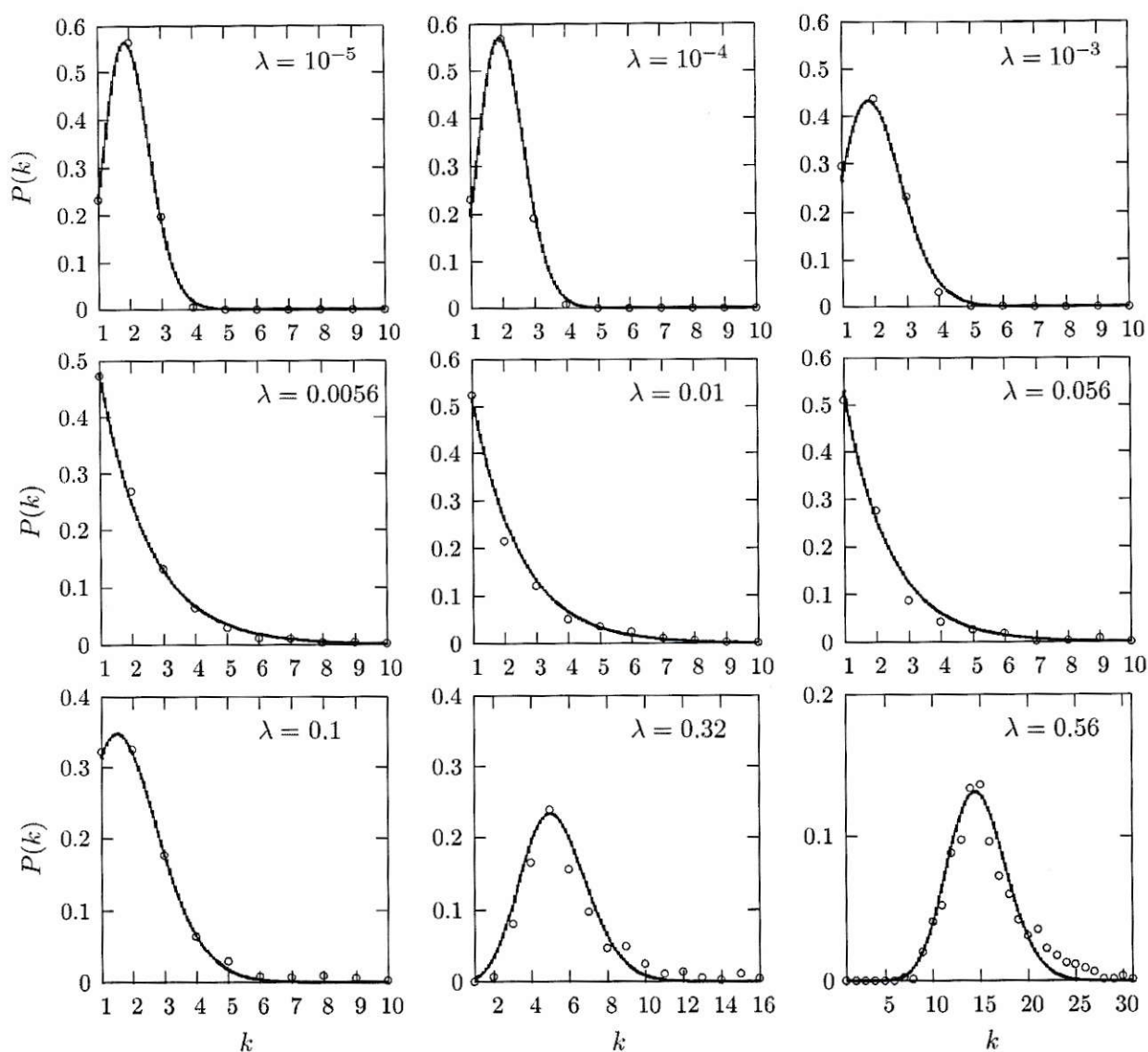


Figura 3.8: Distribución de la conectividad para distintos valores de  $\lambda$ . Cada distribución es el promedio de 20 simulaciones y para cada una de ellas se utilizaron  $N=80$  nodos. En cada una de las gráficas se muestra un ajuste a una curva poissoniana [Ec. (2.5)].

La figura 3.9 muestra los valores de  $D$  y  $H$  en función de  $\lambda$ , normalizados a su máximo valor. Como es de esperar,  $D$  aumenta y  $H$  disminuye con  $\lambda$ , debido a que la ecuación (3.2) pondera de forma inversa entre ellos. En tanto, en la figura 3.10 se grafica el factor de clustering  $C$  y la distancia media  $\langle H \rangle$  en función de  $\lambda$ .



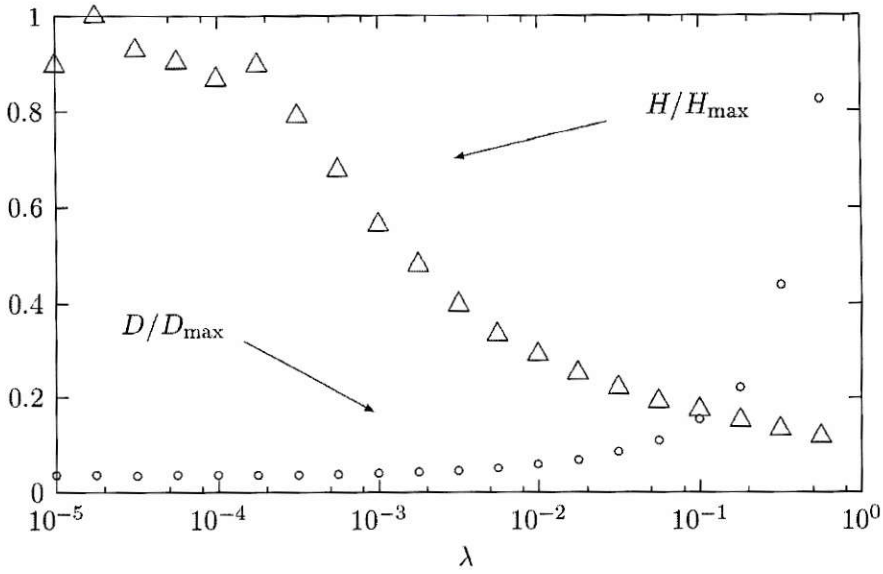


Figura 3.9: Gráfica de  $D$  y  $H$  en función de  $\lambda$ . Cada punto es el promedio de 20 simulaciones y para cada una de ellas se utilizaron  $N=80$  nodos.

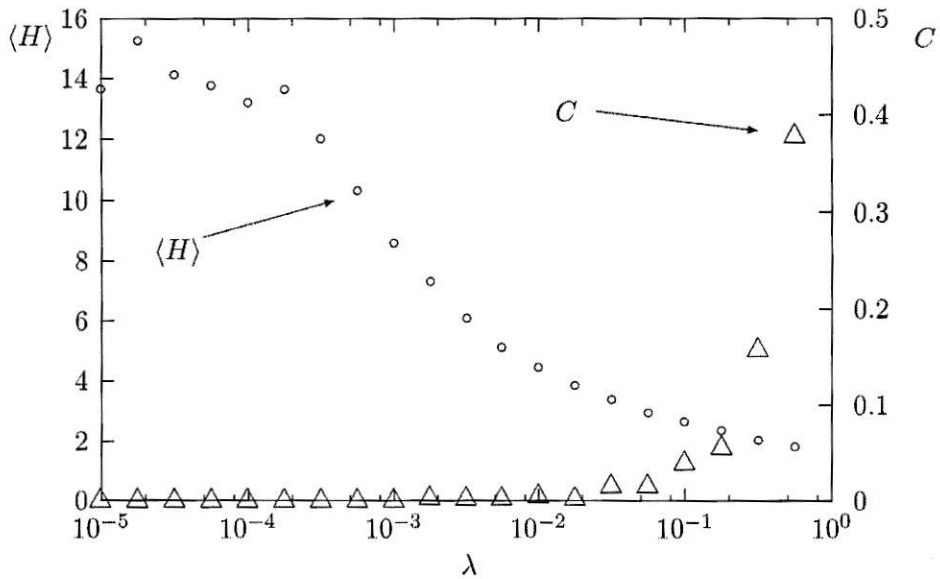


Figura 3.10: Gráfica de  $\langle H \rangle$  y  $C$  (coeficiente de clustering) en función de  $\lambda$ . Cada punto es el promedio de 20 simulaciones y para cada una de ellas se utilizaron  $N=80$  nodos.

En la figura 3.10 se aprecia que  $H$  desciende más rápido con  $\lambda$  que en el modelo anterior (Fig. 3.6). Esto es debido a que la energía aquí utilizada (Ec. 3.6) incluye de forma explícita el término  $H$  el cual, obviamente, es proporcional a  $\langle H \rangle$ . Por otra parte  $C$  pareciera aumentar más lentamente. De todas formas, también podemos decir en este modelo, que cuando  $\lambda$  se aproxima a  $\lambda = 1$  el modelo presenta características *small world*. Para poner esto en evidencia la tabla 3.10 muestra resultados de  $\langle H \rangle$  y  $C$  para algunos valores de  $\lambda$ . Como punto de comparación, en la tabla se incluyen los resultados de redes aleatorias y redes del modelo Watts-Strogatz. Comparando los resultados de la tabla con el modelo anterior (Sec. 3.2), aquí se pueden ver valores de  $\langle H \rangle$  notoriamente menores incluso que los proporcionados por redes aleatorias. Sin embargo, las propiedades *small-world* no son evidentes en el factor de clustering, el cual, tiende a ser menor que para el modelo de la Sec. 3.2.

Modelo Global II			Red Aleatoria		Red Modelo W-S		$M$
$\lambda$	$\langle H \rangle$	$C$	$\langle H \rangle$	$C$	$\langle H \rangle$	$C$	
0.562	1.806	0.378	1.837	0.197	1.988	0.441	625
0.316	2.031	0.158	2.327	0.099	2.635	0.366	315
0.177	2.337	0.055	3.310	0.047	4.066	0.269	158
0.100	2.657	0.039	3.979	0.034	5.491	0.238	125
0.056	2.931	0.015	5.286	0.012	8.235	0.166	98

Tabla 3.3: Valores de  $\langle H \rangle$  y  $C$  en función de  $\lambda$ . Se puede ver que se tiene un alto factor de *clustering* y una baja distancia promedio, propiedades típicas de redes *small-world*. Conjuntamente se muestran los valores de  $\langle H \rangle$  y  $C$  de redes aleatorias (sección 2.3) y de redes del modelo Watts-Strogatz con  $p = 0.1$  (sección 2.5.2) del mismo tamaño. Por último se indica el orden  $M$  de las redes en cada caso ( $N=80$  para todas las redes y todos los casos).

Otros autores ya habían estudiado los modelos Global I y II, sin embargo, ninguno de ellos había hecho notar la capacidad que estos tienen para reproducir las características *small-world* definidas por Watts y Strogatz. La única aproximación al respecto fue hecha por Mathias y Gopal [40], quienes proponen un modelo basado en la minimización de la energía (3.6) (Modelo Global II) capaz de generar redes *small-world*, pero a diferencia de lo realizado aquí, el modelo propuesto por ellos asume nodos distribuidos uniformemente sobre un anillo, lo cual lo hace un caso muy particular. Además, ellos no mencionan nada con respecto a la energía (3.2) (Modelo Global I), la cual es representativa en un número importante de los casos.

# Capítulo 4

## Optimización: Modelos de crecimiento

Los modelos tratados en la sección anterior, por su naturaleza óptima, parecen ser una manera razonable de modelar redes reales. Sin embargo, como ya se vio en la sección 2.2, muchas de las redes reales responden a distribuciones con leyes de potencia y no a distribuciones poissonianas, que fueron las obtenidas en todos los casos estudiados en el capítulo 3. Como una vía de solucionar este problema, en este capítulo se trabajará con modelos basados en el crecimiento de redes, y en donde también se considerarán las reglas de optimización vistas en el capítulo anterior.

Los modelos de crecimiento que se expondrán, al igual que los mostrados en la sección 2.4, comenzarán con un solo nodo  $v_0$ , y cada nuevo nodo  $v_n$  que surge se enlaza a algún nodo  $v_m$  existente en la red. Al igual que el capítulo 3, los nodos están ubicados en un plano bidimensional, con condiciones de borde periódicas, y la posición en cada nuevo nodo  $v_n$  se determina aleatoriamente. La decisión clave que determina la configuración final de la red es:

*¿a cual nodo  $v_m$  existente de la red se debe enlazar el nuevo nodo  $v_n$ ?*

A continuación, se expondrán cinco modelos que pretenden responder de diferentes formas esta pregunta. Notar siempre que  $v_m \in (v_0, v_1, \dots, v_{n-1})$ , y se impondrá que el proceso de crecimiento se detiene cuando  $n = N$ . Es importante señalar que cada vez que un nodo surge en una posición queda en ella para siempre y cada enlace que se forma nunca más es removido. De esta forma, en cada paso, tenemos un nuevo nodo y un nuevo enlace, los que serán definitivos, y al final del proceso tendremos  $N$  enlaces y  $N + 1$  nodos (incluyendo el nodo  $v_0$  inicial). Por otro lado, debido a que siempre se tendrá  $N \gg 1$ , podemos pensar que el

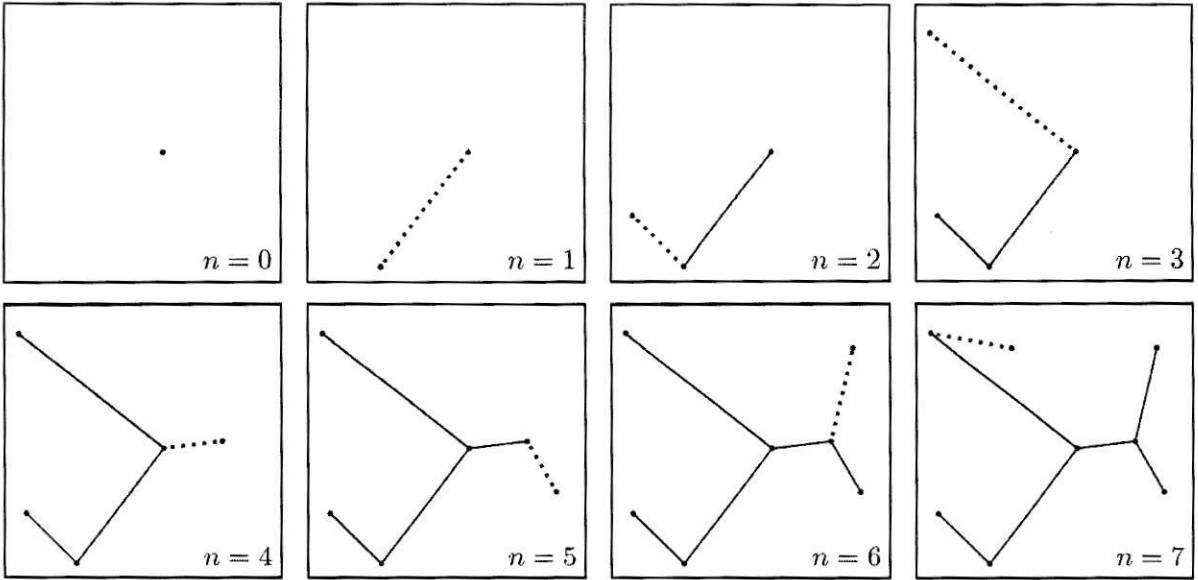


Figura 4.1: Diagrama de crecimiento para el modelo CAC (Sec. 4.1).

caso  $n = N$  representa un estado estacionario. Nos referimos con estacionario, a que tanto las distribuciones de conectividad, como otras características de la red, tienden a un equilibrio, no importando si escogemos  $N = 10^3$  ó  $N = 10^4$ .

## 4.1. Crecimiento con ahorro de camino (CAC)

Uno de los modelos más simples de crecimiento basado en optimización que podemos pensar, es aquél que propone ahorrar camino en cada paso, esto es, cada nuevo nodo que surge en el plano se enlaza simplemente al nodo más cercano a él (ver figura 4.1). En este caso, estamos asumiendo que la única energía en el sistema que se pretende minimizar, es la referente al costo de los enlaces:

$$E = \sum_{i < j} D_{ij} A_{ij} . \quad (4.1)$$



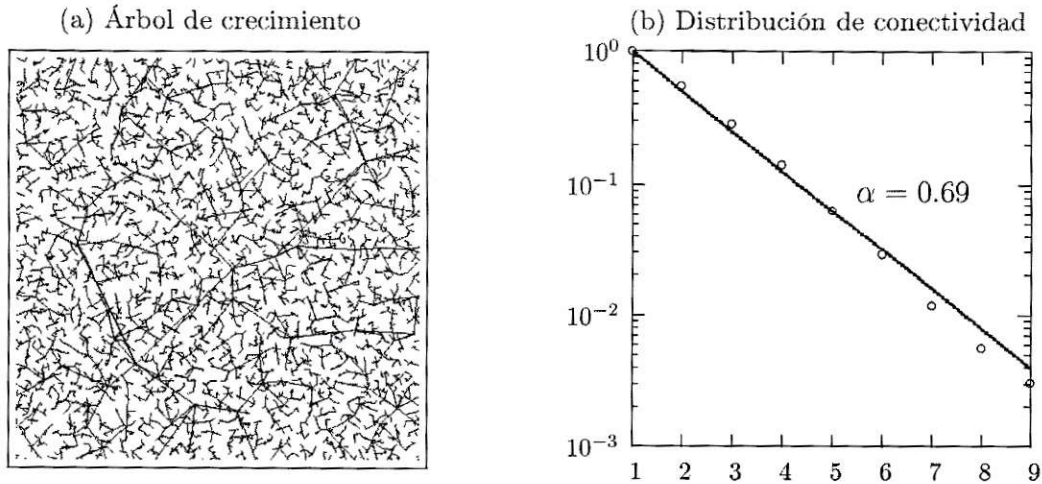


Figura 4.2: Resultados del modelo de crecimiento CAC, con  $N = 5000$ .

Con esta definición el incremento de energía del  $n$ -ésimo paso es:

$$E_n = D_{nm} , \quad (4.2)$$

donde  $v_n$  es el nuevo nodo en el  $n$ -ésimo paso y  $v_m$  el nodo al cual se enlaza. Esta además decir que  $\sum E_n = E$ . La figura 4.2 muestra resultados de una simulación del modelo. Podemos entender la distribución de la Fig. 4.2.b notando lo siguiente: cada nuevo nodo  $v_n$  se enlaza simplemente al nodo  $v_m$  más cercano a él, pero debido a que la aparición del nodo  $v_n$  es aleatoria en el plano, el nodo  $v_m$  también será aleatorio, es decir, cada nuevo nodo escoge al azar un nodo de la red para enlazarse a él. Esta dinámica consigue reproducir en alguna medida la dinámica del modelo de crecimiento expuesto en la subsección 2.4.1, el cual converge a una distribución exponencial con  $\alpha = 0.69$  [Ec. (2.6)].

## 4.2. Crecimiento Central I

Gastner y Newman [5, 41] sugieren un modelo de crecimiento basado en optimización similar a los que aquí se exponen. Ellos proponen que la energía de la red es del tipo:

$$E = (1 - \lambda)D + \lambda L_0 , \quad (4.3)$$

donde

$$D = \sum_{i < j} \mathbf{A}_{ij} \mathbf{D}_{ij}, \quad (4.4)$$

$$L_0 = \sum_i \mathbf{L}_{i0}, \quad (4.5)$$

donde  $\lambda$  es un parámetro que pondera entre  $D$  y  $L_0$ . Notar que ésta definición de energía es igual a la mostrada en la sección 3.2 [ec. (3.2)], pero a diferencia de ella, aquí sólo se consideran las geodésicas que llegan al nodo  $v_0$ . Es decir, cada nuevo nodo  $v_n$  escoge al nodo  $v_m$  tal que minimiza la siguiente expresión:

$$E_n = (1 - \lambda) \cdot \mathbf{D}_{nm} + \lambda \cdot \mathbf{L}_{n0}, \quad (4.6)$$

siendo el término  $E_n$  el incremento de la energía en el  $n$ -ésimo paso y  $\sum_n E_n = E$ . Sin embargo, a diferencia de lo que proponemos en este trabajo, Gastner y Newman no seleccionan al azar la posición del nuevo nodo  $v_n$ , sino que su modelo propone que la ubicación de todos los nodos debe estar preestablecida. Así, el nuevo nodo  $v_n$  será escogido de este grupo como el nodo que genere el menor costo de adhesión.

Nosotros consideramos que el surgimiento aleatorio de nuevos nodos es más representativo de la realidad en un porcentaje importante de los casos, ya que, por lo general, cada nuevo individuo no espera ser el más económico de entre todos los desconectados para poder adherirse, sino que más bien, son factores aleatorios externos los que influyen en su decisión, como por ejemplo, el descubrimiento de una mina de oro en un pequeño poblado lo puede obligar a conectarse con la red de carreteras que llegan a la ciudad. Es por ello que hacemos énfasis en que tanto el modelo “central I” como todos los modelos vistos en este capítulo asumen que el surgimiento de nuevos nodos en el plano es aleatorio.

La definición de energía mostrada en la ecuación (4.3) asume que cada nodo  $v_i$  solo mantiene flujo de información con el nodo central  $v_0$ , despreciando el intercambio de información con los otros nodos de la red. Este modelo puede considerarse útil cuando se tienen organizaciones completamente centralizadas, como por ejemplo, una capital de provincia con mucha



importancia, rodeada de muchos pueblos menores en donde se desprecia la interacción que pueda haber entre ellos, o una configuración de computadores en donde todos están obligados a acceder a una base de datos central.

En la figura 4.3 se muestran la gráfica de algunos árboles de crecimiento resultado del modelo con  $N = 1000$ , para diferentes valores de  $\lambda$ . Es interesante notar que para  $\lambda \approx 1$  todos los nodos llegan de forma casi directa al nodo central. Por otro lado, la figura 4.4 muestra las distribuciones de conectividad para los mismo valores de  $\lambda$ . Podemos notar que las distribuciones tienen un comportamiento exponencial.

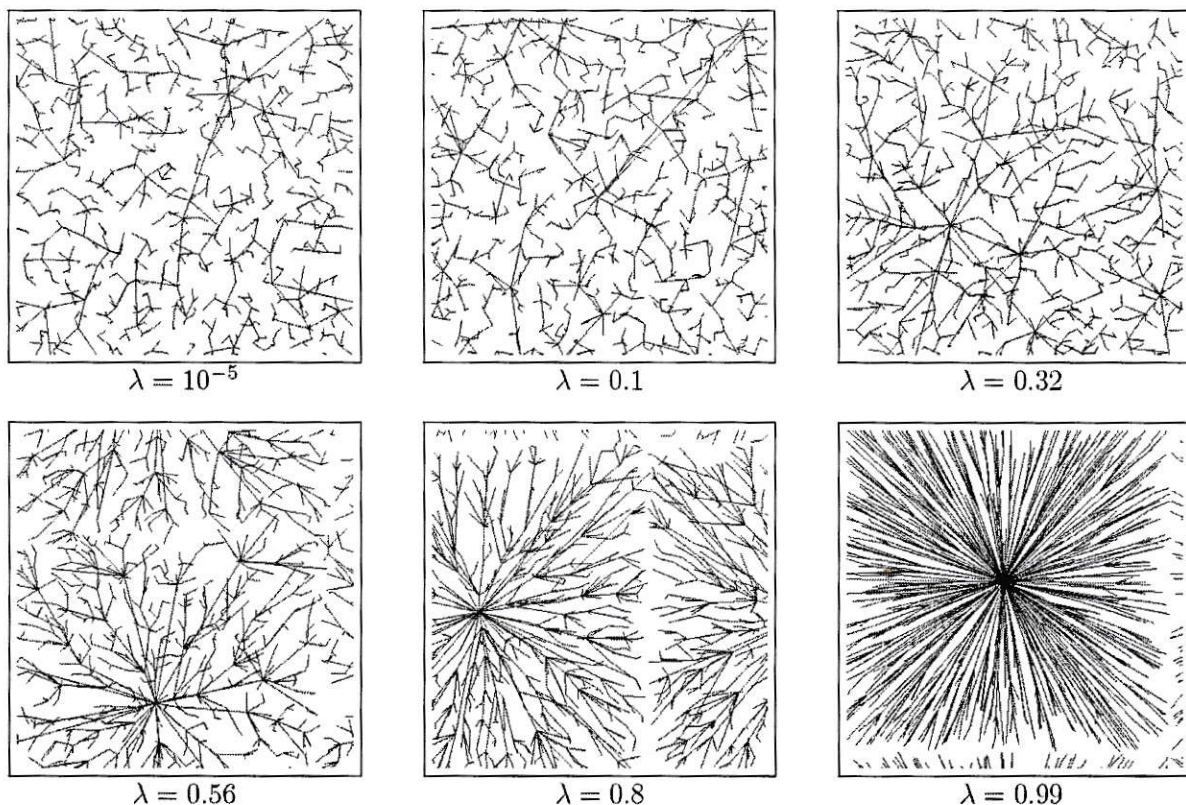


Figura 4.3: Árboles resultado del crecimiento de redes minimizando en cada paso la ecuación (4.6). Se muestran simulaciones para diferentes valores de  $\lambda$  y en cada una de ellos el número total de nodos utilizados fue de  $N = 1000$ . Se puede ver que nuevamente se cumple el patrón de la figura 3.3, el cual muestra que, para  $\lambda \approx 0$  es menor el volúmen de caminos y mayor la longitud de las rutas (geodésicas geométricas) entre nodo y nodo, mientras que para  $\lambda \approx 1$  se produce lo contrario. Sin embargo, esta vez podemos ver que los nodos “buscan” al nodo central  $v_0$ . Esto se debe a que la ecuación (4.3) sólo considera las geodésicas que llegan a  $v_0$ .

Es interesante observar, de las distribuciones de la figura 4.4, que cuando  $\lambda \rightarrow 0$  obtenemos

distribuciones exponenciales con  $\lambda \approx 0.69$ , al igual que para el modelo CAC planteado en la sección anterior (Sec. 4.1). Esto es de esperar ya que el límite cuando  $\lambda \rightarrow 0$  de la energía de la ecuación (4.3) es precisamente la energía del modelo CAC mostrada en la ecuación (4.1).

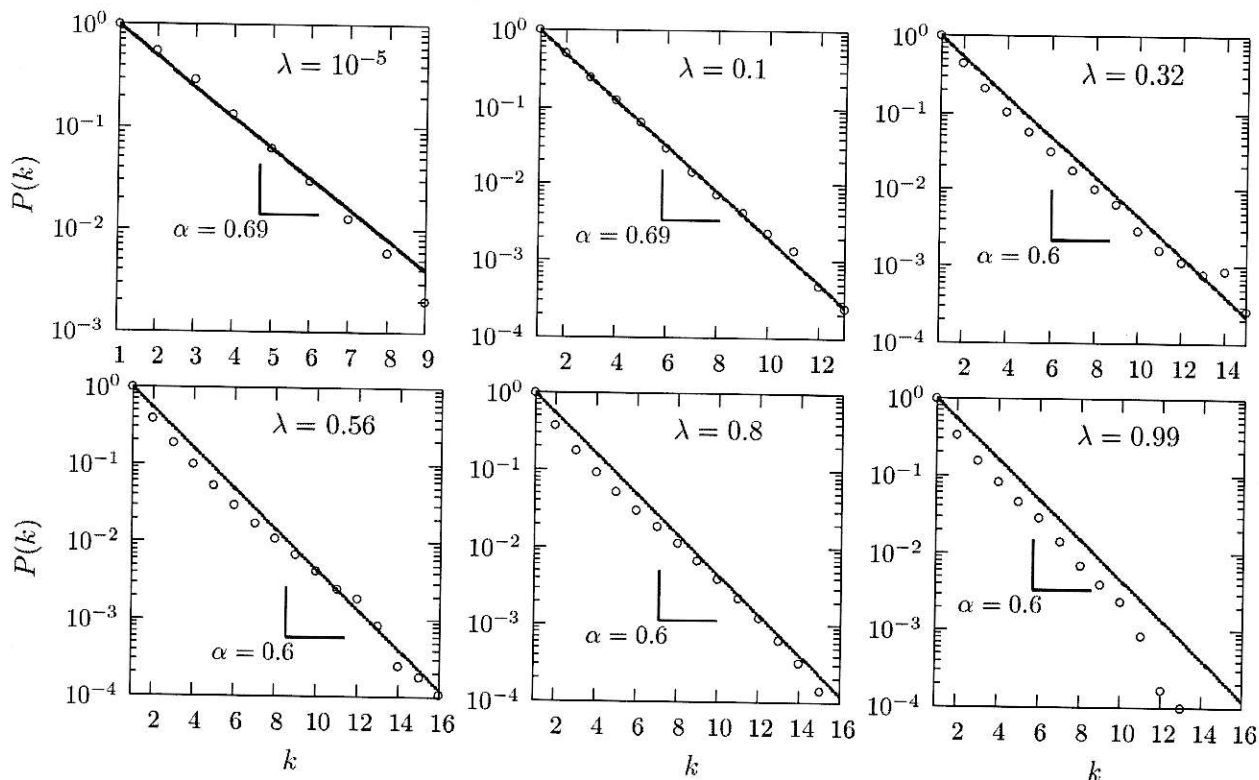


Figura 4.4: Distribuciones de conectividad de redes de crecimiento obtenidas mediante la optimización de la ecuación (4.6) para diferentes valores de  $\lambda$ . Cada gráfico muestra el promedio de 50 simulaciones diferentes, cada una con  $N = 1000$ .

### 4.3. Crecimiento Central II

Una variación a lo que propone el modelo anterior es suponer que la energía del sistema está dada por:

$$E = (1 - \lambda)D + \lambda H_0, \quad (4.7)$$

donde



$$D = \sum_{i < j} \mathbf{A}_{ij} \mathbf{D}_{ij}, \quad (4.8)$$

$$H_0 = \sum_i \mathbf{H}_{i0}, \quad (4.9)$$

donde  $\lambda$  es un parámetro que pondera entre los costos de los caminos ( $D$ ) y la suma de las geodésicas que llegan a  $v_0$  ( $H_0$ ). Esta energía es muy similar a la mostrada en la sección anterior pero, ésta vez, utilizamos  $\mathbf{H}$  y no  $\mathbf{L}$  (al igual que para el modelo Global II visto en la sección 3.3). Con esta definición, y siguiendo la dinámica de crecimiento, en la que en cada nuevo paso se busca optimizar costo, tenemos que cada nuevo nodo  $v_n$  escogerá enlazarse al nodo  $v_m$  tal que minimice se la siguiente función:

$$E_n = (1 - \lambda) \cdot \mathbf{D}_{nm} + \lambda \cdot \mathbf{H}_{n0}, \quad (4.10)$$

siendo nuevamente el término  $E_n$  el incremento de la energía en el  $n$ -ésimo paso ( $\sum_n E_n = E$ ). Este modelo fue planteado originalmente por Fabrikant *et al* [42], y su revisión aquí se justifica en el sentido en que representa un marco útil a los modelos que se verán en secciones posteriores.

Al igual que para el caso anterior (modelo central I, Sec. 4.2) la definición de energía de la ecuación (4.7) asume que cada nodo  $v_i$  sólo mantiene flujo de información con el nodo central  $v_0$ , despreciando el intercambio de información con los otros nodos de la red.

En la figura 4.5 se muestran la gráfica de algunos resultados del modelo con  $N = 1000$  para diferentes valores de  $\lambda$ . Debido a la similitud de la ecuación de energía aquí planteada [eq. (4.7)] con la planteada en la sección 3.3 (modelo global II), nuevamente se produce la formación de patrones estrella, aunque ésta vez, son más explícitos. Esto es debido a que los patrones estrellas minimizan el término  $H_0$ .

Por otro lado, la figura 4.6 muestra las distribuciones de conectividad para algunos valores de  $\lambda$ . Notemos que nuevamente se cumple que cuando  $\lambda \rightarrow 0$  se tiende al modelo CAC, por ello

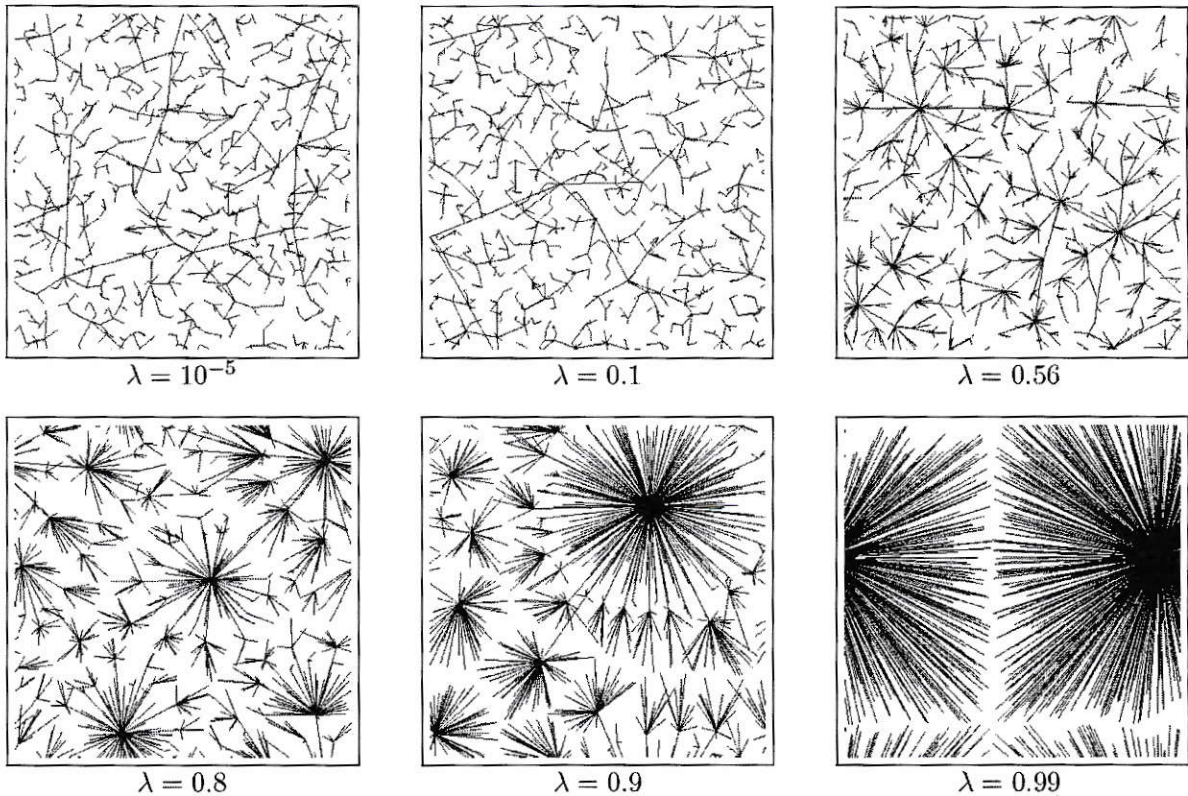


Figura 4.5: Árboles resultado del crecimiento de redes utilizando la ecuación (4.10) para diferentes valores de  $\lambda$ . El número total de nodos utilizado en cada árbol fue de  $N = 1000$ . Como es de esperar, se cumple el patrón de mayor “volumen” de caminos para  $\lambda \approx 1$  y menor para  $\lambda \approx 0$ . Por otro lado, podemos notar la formación de estrellas, situación similar exhibida en el modelo “Global II”.

que las distribuciones de  $\lambda \approx 0$  siguen curvas exponenciales con  $\alpha \approx 0.69$ . Pero a diferencia de las distribuciones del modelo Central I, aquí, podemos ver leyes de potencia. Este resultado nos acerca más a las redes reales como las resumidas en la tabla 2.2. Sin embargo, el exponente de ley de potencia ( $\gamma \approx 1.6$ ) está un poco alejado de los valores mostrados en la tabla 2.2, los cuales oscilan entre  $\gamma = 2.1$  y  $\gamma = 3.2$ . Además, el hecho de sólo considerar el flujo de información con el nodo central, claramente limita la generalidad de los modelos “centrales” vistos en las secciones 4.2 y 4.3.

Como una forma de buscar modelos aplicables a una mayor cantidad de casos reales, en las próximas secciones se eliminará la restricción de “flujo central” impuesta de forma implícita, tanto por Gastner-Newman [5,41] como por Fabrikant [42]. Por lo que consideraremos modelos en que la interacción relevante no solo será la de cada nodo con el nodo  $v_0$ , sino que será la interacción entre todos los pares de nodos de la red.

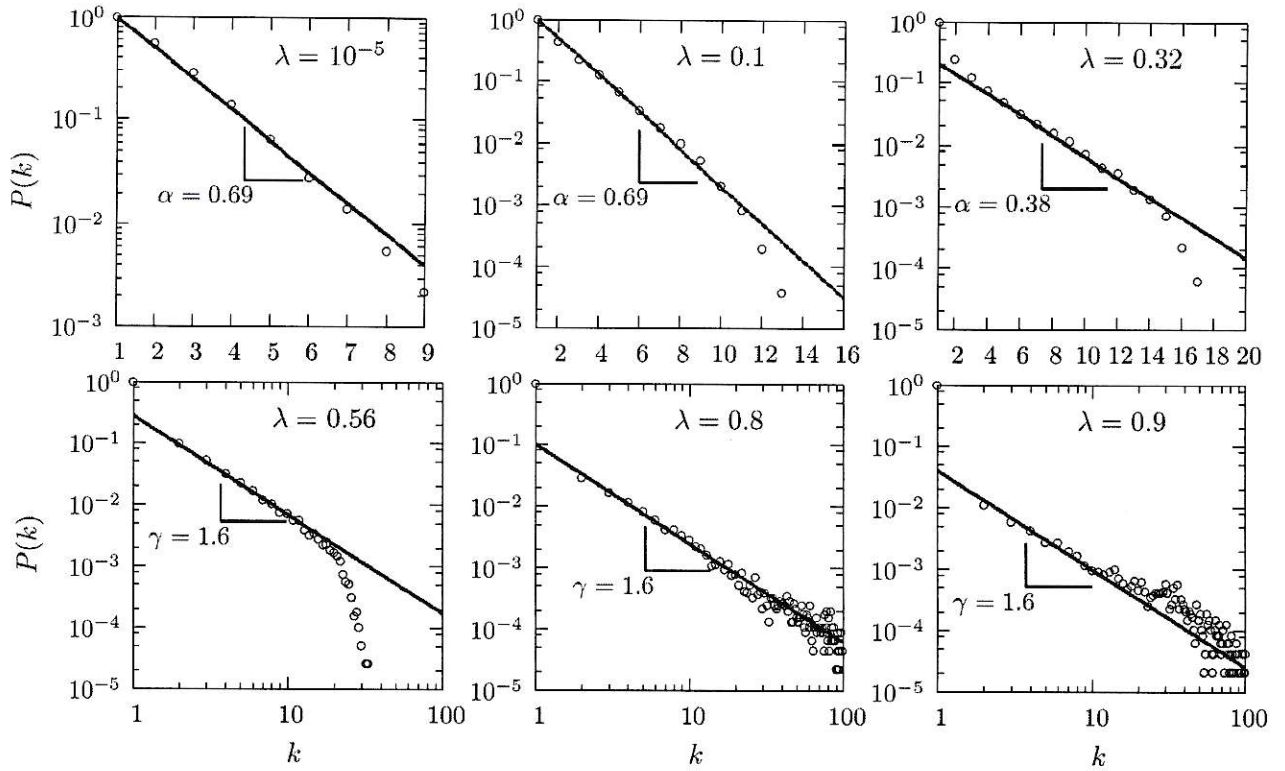


Figura 4.6: Distribución de conectividad de redes de crecimiento obtenidas mediante la optimización de la ecuación 4.10 para diferentes valores de  $\lambda$ . Cada gráfico muestra el promedio de 50 simulaciones diferentes cada una hecha para  $N = 1000$ .



## 4.4. Crecimiento Total I

Los modelos denominados “crecimiento total” que aquí se expondrán, corresponden, en el contexto local, a la continuación y generalización de los modelos de “crecimiento central”, sin embargo, en el contexto general, responden a nuestra propuesta original de modelos capaces de generar redes libres de escala similares a las obtenidas por Barabasi (Sec. 2.4.2), pero sustentadas sobre la intuición de optimización que hemos desarrollado durante este trabajo.

Consideremos la misma situación propuesta en los modelos anteriores, en la que se comienza con un solo nodo  $v_0$  y en forma aleatoria van apareciendo nuevos nodos en el cuadro unitario periódico, los cuales se van enlazando a algún nodo ya perteneciente a la red, pero esta vez, la energía que queremos minimizar en cada paso esta dada por:

$$E = (1 - \lambda)D + \lambda L , \quad (4.11)$$

donde

$$D = \sum_{i < j} \mathbf{A}_{ij} \mathbf{D}_{ij} , \quad (4.12)$$

$$L = \sum_{i < j} \mathbf{L}_{ij} . \quad (4.13)$$

Esta energía es exactamente la misma que la energía enunciada en la ecuación (3.2) (Modelo Global I). Con esta definición decimos que cada nuevo nodo  $v_n$  escoje al nodo  $v_m$  tal que minimiza la siguiente ecuación:

$$E_n = (1 - \lambda) \cdot \mathbf{D}_{nm} + \lambda \cdot \sum_{i=0}^{n-1} \mathbf{L}_{ni} , \quad (4.14)$$

o equivalentemente:

$$\begin{aligned} E_n &= (1 - \lambda) \cdot \mathbf{D}_{nm} + \lambda \cdot \left( n \cdot \mathbf{D}_{nm} + \sum_{i=0}^{n-1} \mathbf{L}_{mi} \right) , \\ E_n &= [1 + \lambda(n - 1)] \cdot \mathbf{D}_{nm} + \lambda \cdot \sum_{i=0}^{n-1} \mathbf{L}_{mi} . \end{aligned} \quad (4.15)$$

El término  $\mathbf{D}_{nm}$  es la distancia euclidiana entre los nodos  $v_n$  y  $v_m$ , y el término  $\mathbf{L}_{mi}$  es

la geodésica geométrica entre los nodos  $v_m$  y  $v_i$ . La optimización de la ecuación (4.15) nos indica que en cada paso, cada nuevo nodo  $v_n$  se enlaza a un nodo  $v_m$  de la red, que por un lado, se encuentre cerca a él (bajo valor  $D_{nm}$ ), pero por otro, que esté en una “posición privilegiada” en la red. Cuando decimos “posición privilegiada”, nos referimos a un nodo que tenga alta accesibilidad a todos los otros nodos de la red (bajo valor  $\sum_i L_{mi}$ ). Esta noción de “posición privilegiada” es en alguna medida un símil a la noción de “popularidad” expuesta por Barabasi (Sec. 2.4.2). Aun que cabe señalar que, para nuestro modelo, la existencia de la dualidad “ahorro-eficiencia” o dicho de otra forma “costo-popularidad” lo hace más representativa en un gran número de casos.

Algunas situaciones reales útiles para representar con este modelo pueden ser una red de servidores de internet, o de estaciones de potencia, o de carreteras, etc. Pero que necesariamente sean resultado de un “proceso desordenado de crecimiento” y no de un diseño planeado y dirigido. Con “proceso desordenado de crecimiento” nos referimos a que, en cada paso no se tiene certeza de la configuración futura. Esto en la práctica suele ocurrir, puesto que el crecimiento va siendo determinado por las demandas futuras, las cuales muchas veces son totalmente impredecibles.

En la figura 4.7 se muestran algunos resultados de este modelo con  $N = 1000$  y para diferentes valores de  $\lambda$ , mientras que en la figura 4.8 se muestran las distribuciones de conectividad de diferentes simulaciones para los mismos valores de  $\lambda$ . Podemos notar que cuando  $\lambda \approx 0$  las distribuciones tienden a ajustarse a una exponencial negativa con  $\alpha = 0.69$  [Ec. (2.6)] lo que concuerda con el modelo CAC <sup>1</sup>, y para valores ligeramente mayores de  $\lambda$ , éstas tienden a ajustarse a leyes de potencia, siendo la constante característica  $\gamma = 2.5$  [Ec. (2.7)]. Este último resultado es de sumo interés, ya que se encuentra en el rango de parámetros exhibidos por redes reales (Tabla 2.2).

La figura 4.9 muestra valores de  $D$  y  $L$  para diferentes valores de  $\lambda$ . Estos valores están

---

<sup>1</sup>Para este caso también se tiene que cuando  $\lambda \rightarrow 0$  el modelo converge al modelo CAC.



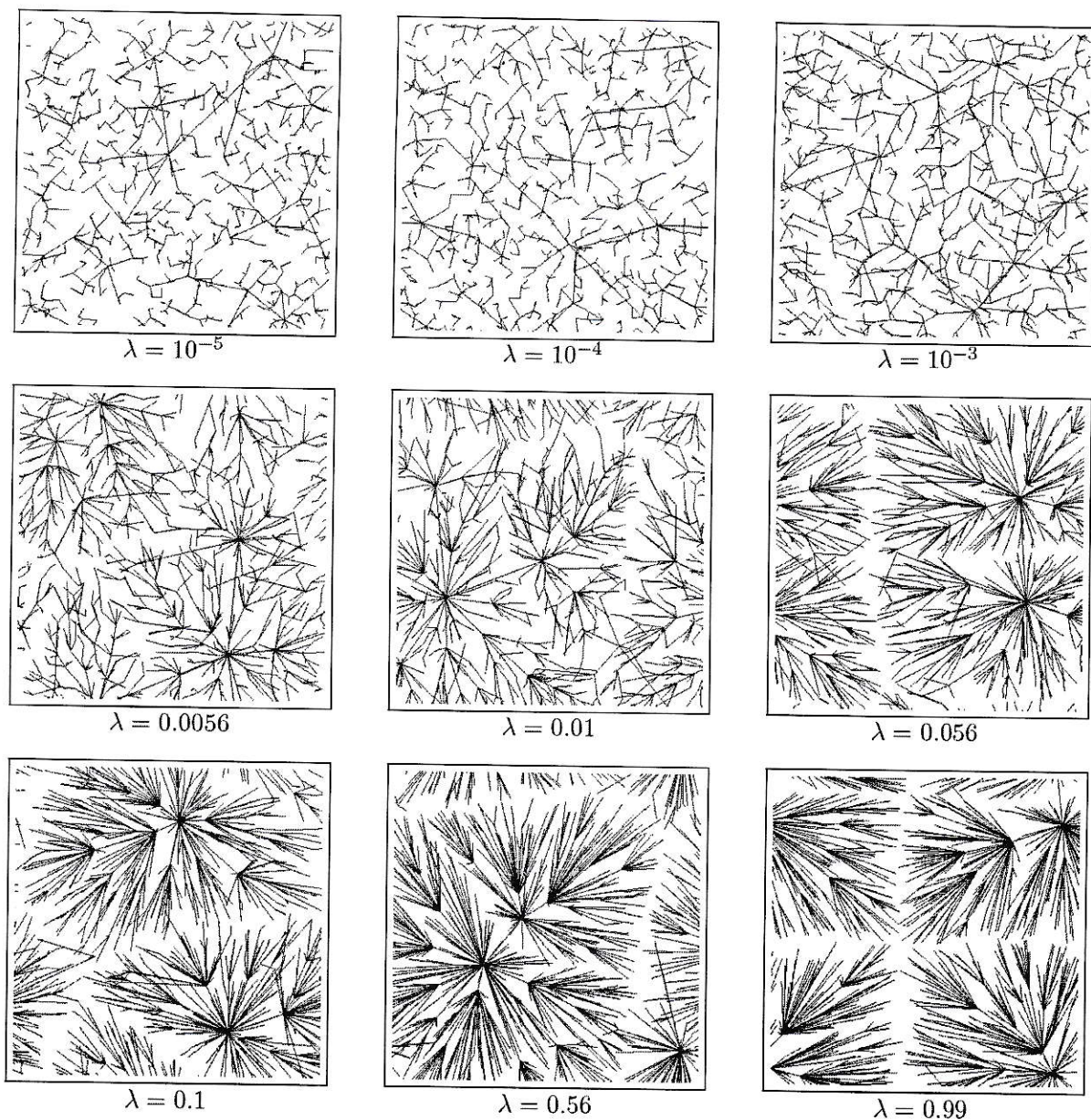


Figura 4.7: Árboles resultado del crecimiento de redes utilizando la ecuación (4.16) para diferentes valores de  $\lambda$ . El número total de nodos utilizado en cada árbol fue de  $N = 1000$ .

normalizados al mayor valor de  $D$  y  $L$ , respectivamente. Es claro notar que mientras que  $D$  es creciente con  $\lambda$ ,  $L$  es decreciente, lo cual era de esperar debido a que el rol de  $\lambda$ , es precisamente, ponderar el costo entre el uno y el otro.



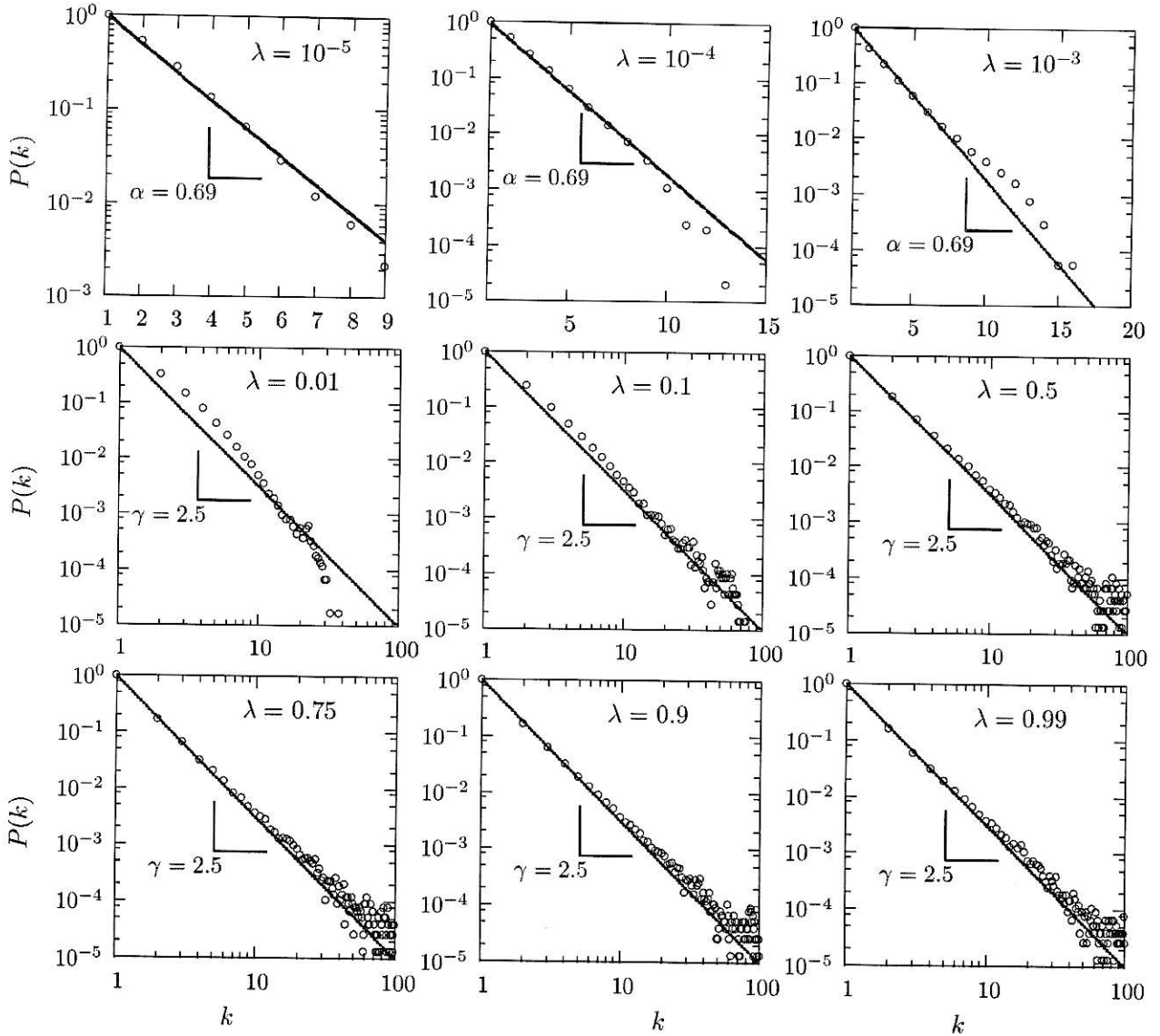


Figura 4.8: Distribución de la conectividad de redes, resultado del modelo de crecimiento total I para diferentes valores de  $\lambda$ . Cada gráfico muestra el promedio de 100 simulaciones diferentes, cada una hecha para  $N = 1000$ .

En la figura 4.10 se muestra el valor  $H$  en función de  $\lambda$ . Notemos que el gráfico muestra una curva decreciente, la cual comienza con valores de  $H \approx 11.2$  y termina con valores de  $H \approx 5.5$ . Estos valores son bastante pequeños, similares a las redes tipo *small-world* vistas en la sección 2.5, pero a diferencia de ellas, aquí por las características del modelo y la definición del *factor de clustering* se tiene que  $C = 0$ . Esto lo podemos explicar de la siguiente forma: el término  $C$  es directamente proporcional al número de triángulos de la red, pero debido a que en la red si hay  $n$  enlaces hay  $n + 1$  nodos, se hace imposible la formación de triángulos,

esto, siempre y cuando no se permita la formación de islas (caso de todos los modelos vistos en esta tesis) <sup>2</sup>.

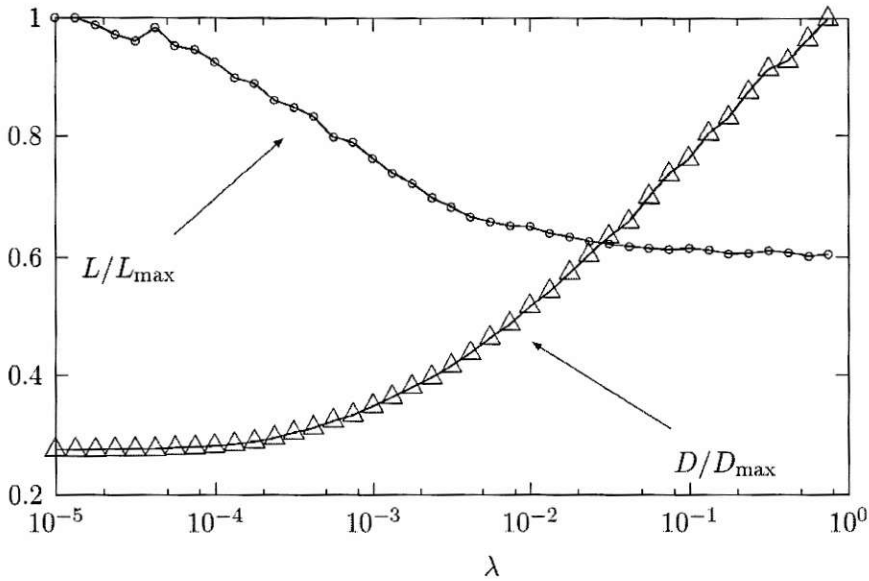


Figura 4.9: Valores de  $D$  y  $L$  para diferentes valores de  $\lambda$ . Estos valores están normalizados al mayor valor de  $D$  y  $L$  respectivamente. Cada punto en la figura es el promedio de 50 simulaciones diferentes con  $N = 1000$ .

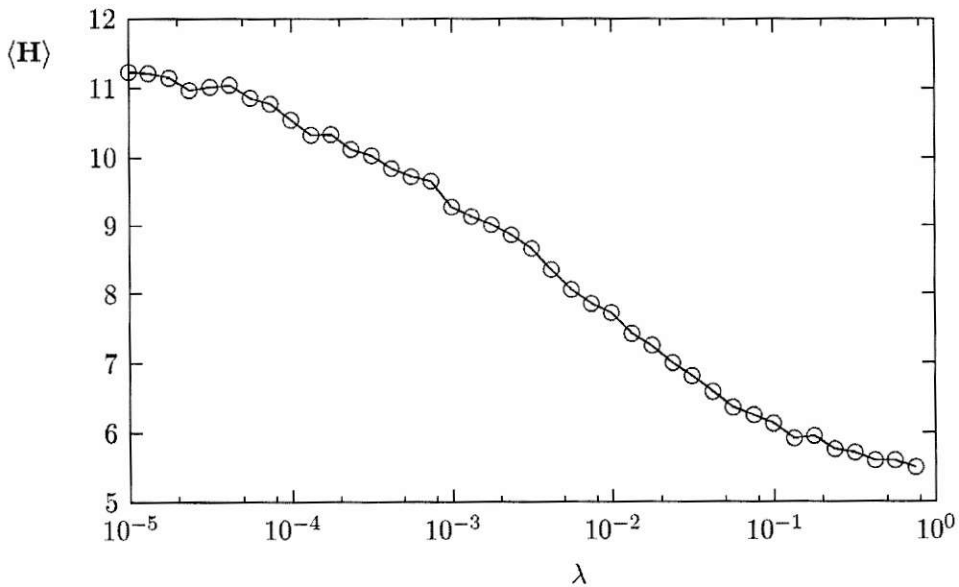


Figura 4.10: Valor de  $\langle H \rangle$  en función de  $\lambda$ . Cada punto en el gráfico es el promedio de 50 simulaciones diferentes con  $N = 1000$ .

<sup>2</sup>Insitamos al lector a que intente realizar una red con  $n$  enlaces y  $n + 1$  nodos en la que exista al menos un triángulo y descubrirá que es imposible, si no se permite la formación de islas en su red.

## 4.5. Crecimiento Total II

Aquí estudiamos una variación del caso anterior, en donde reemplazamos el “camino geométrico más corto”  $L_{ij}$  por el “camino más corto”  $H_{ij}$ . Con esto, la energía queda dada por:

$$E = (1 - \lambda)D + \lambda H , \quad (4.16)$$

donde

$$D = \sum_{i < j} A_{ij} D_{ij} , \quad (4.17)$$

$$H = \sum_{i < j} H_{ij} . \quad (4.18)$$

Esta es la misma definición que la planteada en el modelo global II (Ec. 3.6). Entonces, tenemos que cada nuevo nodo  $v_n$  escoge al nodo  $v_m$ , tal que, minimiza la siguiente ecuación:

$$E_n = (1 - \lambda) \cdot D_{nm} + \lambda \cdot \sum_{i=0}^{n-1} H_{ni} , \quad (4.19)$$

o equivalentemente

$$E_n = (1 - \lambda) \cdot D_{nm} + \lambda \cdot \left( n \cdot D_{nm} + \sum_{i=0}^{n-1} H_{mi} \right) ,$$

$$E_n = (1 + \lambda(n - 1)) \cdot D_{nm} + \lambda \cdot \sum_{i=0}^{n-1} H_{mi} . \quad (4.20)$$

El término  $D_{nm}$  es la distancia euclidiana entre los nodos  $v_n$  y  $v_m$  y el término  $H_{mi}$  es la geodésica entre los nodos  $v_m$  y  $v_i$ . Esta vez tenemos una situación similar a la del modelo de la Sec 4.4, en que, en cada paso, cada nuevo nodo  $v_n$  se enlazará a un nodo  $v_m$  de la red que cumpla las dos siguientes condiciones, por un lado, que se encuentre cerca de él (bajo valor  $D_{nm}$ ), y por otro, que esté en una “posición privilegiada en la red”. Pero esta vez, con



“posición privilegiada en la red”, nos referimos a un nodo que tenga baja distancia geodésica con todos los nodos de la red (bajo valor  $\sum_i \mathbf{H}_{mi}$ ).

En la figura 4.7 se muestran algunos resultados de este modelo con  $N = 1000$  y para diferentes valores de  $\lambda$ . Nuevamente al igual que para los modelos Global II y Central II podemos ver el surgimiento de estructuras tipo estrella.

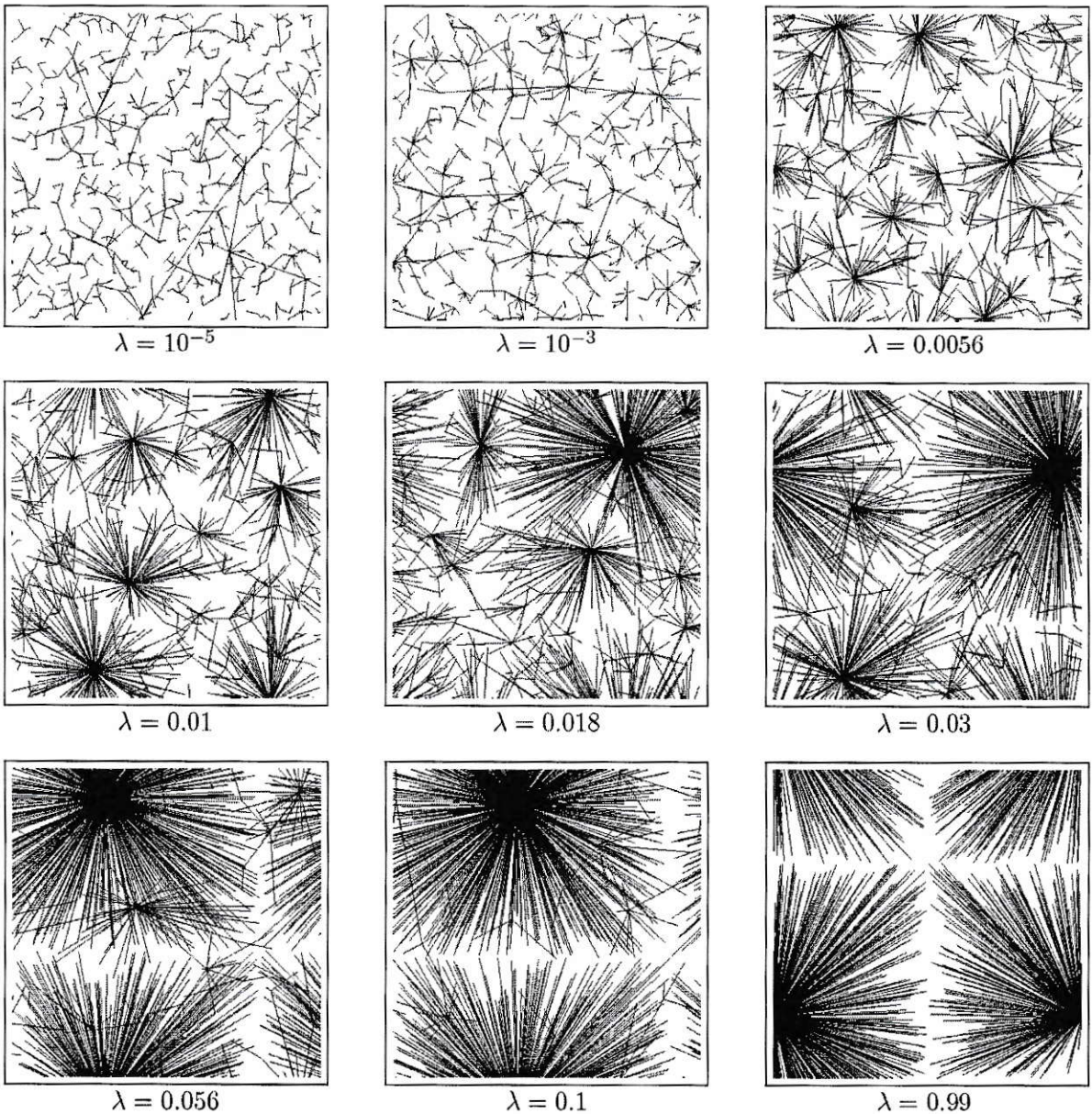


Figura 4.11: Árboles resultado del crecimiento de redes utilizando la ecuación 4.19 para diferentes valores de  $\lambda$ . El número total de nodos utilizado en cada árbol fue de  $N = 1000$ .

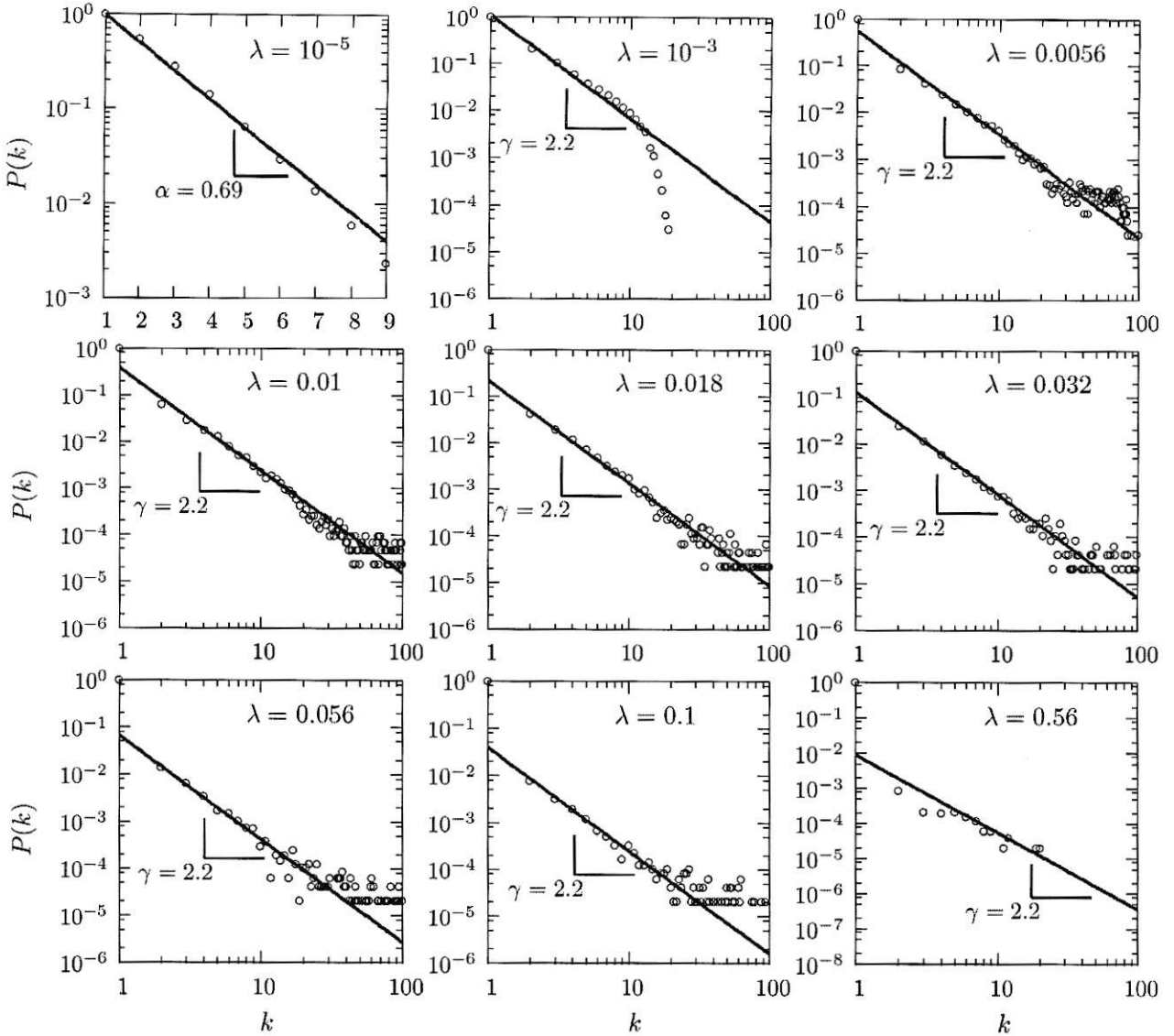


Figura 4.12: Distribucion de la conectividad de redes resultado del modelo de crecimiento total II para diferentes valores de  $\lambda$ . Cada gráfico muestra el promedio de 100 simulaciones diferentes, cada una hecha para  $N = 1000$ .

La figura 4.12 muestra las distribuciones de conectividad de diferentes simulaciones para los mismo valores de  $\lambda$ . Al igual que para el modelo Total I, podemos notar que para valores de  $\lambda$  cercanos a 0 las distribuciones de conectividad tienden a ajustarse a una exponencial negativa, mientras que para valores ligeramente mayores, éstas tienden a ajustarse a leyes de potencia, siendo la constante característica  $\gamma = 2,2$  [Ec. (2.7)], lo que es consistente con el rango de resultados exhibidos en la tabla 2.2.



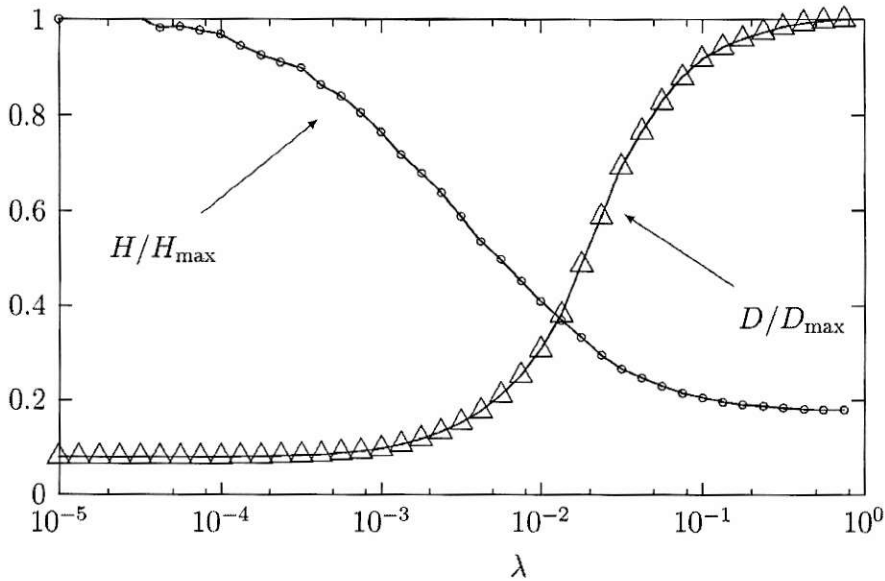


Figura 4.13: Valores de  $D$  y  $L$  para diferentes valores de  $\lambda$ . Estos valores están normalizados al mayor valor de  $D$  y  $L$  respectivamente. Cada punto en la figura es el promedio de 50 simulaciones diferentes con  $N = 1000$ .

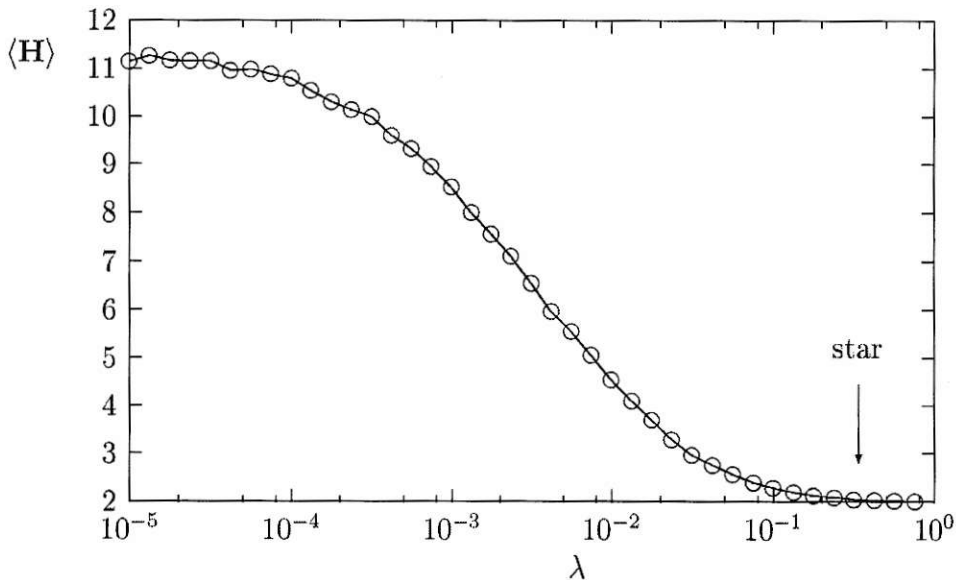


Figura 4.14: Valor de  $\langle H \rangle$  en función de  $\lambda$ . Para valores de  $\lambda \approx 1$  se hace clara la formación de estrellas. Cada punto en el gráfico es el promedio de 50 simulaciones diferentes con  $N = 1000$ .

La figura 4.13 muestran valores de  $D$  y  $L$  para diferentes valores de  $\lambda$ . Estos valores están normalizados al mayor valor de  $D$  y  $L$ , respectivamente. Por otro lado la figura 4.14 se muestra el valor de  $H$  en función de  $\lambda$ . Notemos que el gráfico muestra una curva decreciente,



la cual comienza con valores de  $H \approx 11.0$  y termina con valores de  $H \approx 2.0$ . Estos valores son bastante pequeños, similares a las redes *small-world*. Notemos de la figura 4.11 que para  $\lambda \approx 1$  tenemos una ertrella casi perfecta, lo que concuerda con  $H \approx 2.0$  de la figura 4.14, ya que, en una configuración estrella, todos los nodos están unidos a un único nodo central, lo que implica que, cada nodo dista sólo de dos enlaces de cualquier otro nodo de la red.

# Capítulo 5

## Conclusiones

En este trabajo hemos estudiado el problema de la construcción de una red bidimensional capaz de replicar las principales características encontradas en redes reales. Para ello, en primer lugar, consideramos redes en las que su construcción comienza con un conjunto de nodos fijos distribuidos de forma aleatoria en un plano, y cuya configuración de enlaces se define bajo un criterio de optimización consistente en un compromiso entre un bajo número de enlaces y bajas distancias de geodésicas. Esto dio origen a los modelos Global I y II (Sec. 3.2 y Sec. 3.3), los que responden a dos definiciones de geodésicas diferentes (Def. 6 y Def. 10).

A partir de las simulaciones pudimos notar que la principal riqueza de estos modelos radica en su capacidad de reproducir las características *Small-World* definidas por Watts y Strogatz [10], y si bien, estos modelos han sido estudiados por diferentes autores, ninguno de ellos había puesto en evidencia este hecho. Solo existe una aproximación por parte de Mathias y Gopal [40], pero su propuesta de modelo es bastante restrictiva al limitar los nodos a distribuirse sobre un anillo.

Debido a que en los modelos Global I y II la configuración de nodos es fija y sabida desde el comienzo, pueden ser útiles para representar diseños de construcción de redes planificadas, algo que a priori parece adecuado cuando queremos modelar redes de transporte hechas por el hombre. Sin embargo, las distribuciones de conectividad obtenidas son curvas poissonianas, lo cual no concuerda con la naturaleza libre de escala encontrada en diversas redes reales.

Con el fin de obtener las citadas distribuciones libre de escala es que se decidió asumir la construcción de la red como resultado de un proceso de crecimiento, similar a lo hecho por Barabási y Albert [1, 29, 30], este hecho llevó en una primera instancia a la creación de los modelos “Crecimiento Central I y II”, los cuales ya fueron tratados en alguna medida por

Gastner-Newman [5, 41] y Fabrikant [42] respectivamente. Y, su revisión en este trabajo se justifica en el sentido que brindan un marco conceptual útil y una evolución hacia los modelos “Crecimiento Total I y II”.

Debido a la falta de generalidad de los modelos de crecimiento central y su incapacidad de explicar las topologías de redes reales es que surgen los modelos “Crecimiento Total I y II”. Cabe señalar que estos modelos son el aporte más significativo de este trabajo, ya que, no han sido tratados anteriormente en la literatura. En ellos, se busca modelar las características estructurales de redes de transporte en las que la optimización es realizada en cada paso, y no en el diseño general de la red. Este es el caso de muchas redes reales, en las que la construcción de un nuevo enlace está determinada por la demanda y por los costos de construcción instantáneos, no evaluando costos de enlaces, ni demandas futuras. Un resultado interesante, y quizá el más importante de este trabajo, es que tanto el modelo “Crecimiento Total I” como el modelo “Crecimiento Total II” logran reproducir en forma satisfactoria distribuciones de conectividad libres de escala, en donde los exponentes característicos son  $\gamma = 2.5$  y  $\gamma = 2.2$ , los cuales están dentro del rango de las redes reales (Tabla 2.2). Otros modelos, como el propuesto por Barábasi y Albert, logran resultados similares, pero a diferencia de ellos, los aquí propuestos son de una naturaleza determinista ya que la red se contruye en función de una minimización y no en función de una elección aleatoria como el propuesto por ellos, lo cual lo hace más representatvo en un número importante de casos.

Otra característica interesante de los modelos “Crecimiento Total I y II”, es que el valor de la media de las geodésicas es bajo (en comparación con redes aleatorias) lo cual, es característico de redes *small-world*. Sin embargo, son incapaces de reproducir el alto *factor de clustering* propio de este tipo de redes. Este último punto representa un desafío y un punto de partida de futuros trabajos sobre modelos de redes complejas bidimensionales, basados en optimización. De todos modos, este trabajo, permite mostrar que las propiedades estadísticas de redes reales, pueden ser descritas satisfactoriamente con modelos sencillos de crecimiento basados en optimización, y que dichas propiedades, dependen críticamente del criterio de optimización utilizado.



# Apéndice A

## Distribución de Poisson como límite de una distribución Binomial

Aquí se mostrará un metodo para llegar de la ecuación (2.9) a la ecuación (2.10), tomando el límite  $N \rightarrow \infty$  y  $M \rightarrow \infty$ . Para esto utilizaremos dos resultados bien conocidos: la fórmula de Stirling y la definición de la exponencial:

$$k! \approx e^{-k} k^k \quad \text{para } k \text{ grandes,}$$

$$e^a = \lim_{k \rightarrow \infty} \left(1 + \frac{a}{k}\right)^k.$$

La distribución binomial mostrada en la ecuación (2.9) es:

$$P(k) = \binom{n}{k} (p_e)^k (1 - p_e)^{n-k} \quad (\text{A.1})$$

donde se ha definido  $n = N - 1$ . Utilizando la formula de Stirling, el coeficiente binomial de la ecuación (A.1) para  $n$  grande puede escribirse como

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (\text{A.2})$$

$$\approx \frac{e^{-n} n^n}{k! e^{-(n-k)} (n-k)^{(n-k)}} = \frac{e^{-k} (n-k)^k}{k!} \left(\frac{n}{n-k}\right)^n.$$

Con esto la distribución binomial (A.1) queda, para  $n$  grandes, como

$$P(k) = \frac{e^{-k}(n-k)^k}{k!} \left(\frac{n}{n-k}\right)^n (p_e)^k (1-p_e)^{n-k} . \quad (\text{A.3})$$

Notemos que el término  $z = 2M/N$  es la conectividad promedio de cada nodo de la red. En términos de este número, la probabilidad  $p_e$  puede escribirse como  $p_e = z/n$ , con lo cual la ecuación (A.3) queda como

$$P(k) = \frac{e^{-k}(n-k)^k}{k!} \left(\frac{n}{n-k}\right)^n \left(\frac{z}{n}\right)^k \left(1 - \frac{z}{n}\right)^{n-k} , \quad (\text{A.4})$$

$$P(k) = \frac{e^{-k}}{k!} z^k \left(1 - \frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{-n} \left(1 - \frac{z}{n}\right)^{n-k} ,$$

Así, cuando  $n \rightarrow \infty$ ,

$$P(k) = \frac{z^k}{k!} \cdot e^{-z} . \quad (\text{A.5})$$

Por lo tanto, para una red muy grande, el proceso de enlazar pares de nodos en forma aleatoria conduce a una distribución de Poisson.

# Bibliografía

- [1] R. ALBERT and A.-L. BARABÁSI, *Reviews of Modern Physics* **74**(1), 47 (2002).
- [2] M. E. J. NEWMAN, *SIAM Review* **45**(2), 167 (2003).
- [3] S.Ñ. DOROGOVTSSEV and J. F. F. MENDES, *Advances in Physics* **51**, 1079 (2002).
- [4] S.Ñ. DOROGOVTSSEV and J. F. F. MENDES, *Evolution of Networks: From Biological Nets to the Internet and WWW*, Oxford University Press, Oxford, ISBN:0198515901, 2003.
- [5] M. T. GASTNER, *Spatial distributions: density-equalizing map projections, facility location, and two-dimensional network*, PhD thesis, University of Michigan, 2005.
- [6] M. T. GASTNER and M. E. J. NEWMAN, *Physical Review E* **74** (2006).
- [7] S. LI YANG, N.-S. HSU, P. W. F. LOUIE, and W. W.-G. YEH, *Journal of Infrastructure Systems* **2**, 54 (1996).
- [8] M. BARTHELEMY and A. FLAMMINI, *Journal of Statistical Mechanics* (2006).
- [9] P. SEN, S. DASGUPTA, A. CHATTERJEE, P. A. SREERAM, G. MUKHERJEE, and S. S. MANNA, *Physical Review E* **67** (2003).
- [10] D. J. WATTS and S. H. STROGATZ, *Nature* **393**, 440 (1998).
- [11] R. GUIMERA, S. MOSSA, A. TURTSHI, and L. A.Ñ. AMARAL, *Preprint cond-mat/0312535* (2003).
- [12] M. FALOUTSOS, P. FALOUTSOS, and C. FALOUTSOS, *Proceedings of the conference on Applications, technologies, architectures, and protocols for computer communication* , 251 (1999).
- [13] W. AIELLO, F. CHUNG, and L. LU, *Handbook of massive data sets* , 97 (2002).
- [14] L. ADAMIC, R. M. LUKOSE, A. R. PUNIYANI, and B. A. HUBERMAN, *Physical review E* **64** (2001).
- [15] R. F. I CANCHO, C. JANSSEN, and R. V. SOLÉ, *Physical review E* **64** (2001).
- [16] F. LIJEROS, C. R. EDLING, and L. A.Ñ. AMARAL, *Microbes and Infection* **5**, 189 (2003).
- [17] H. JEONG, H. M. A.-L. BARABÁSI, and Z.Ñ. OLFVAI, *Nature* **411**, 41 (2001).
- [18] Q. CHENG, H. CHANG, R. GOVINDAN, S. JAMIN, S. J. SHENKER, and W. WILLINGER, *Proceedings of the 21st annual join conference of the IEEE computer and communication Societies* (2002).
- [19] H. JEONG, B. TOMBOR, R. ALBERT, Z.Ñ. OLFVAI, and A.-L. BARABÁSI, *Nature* **407**, 651 (2000).



- [20] L. A.Ñ. AMARAL, A. SCALA, M. BARTHÉLÉMY, and H. E. STANLEY, *Proceedings of the National Academy of Sciences of the USA* **97**, 11149 (1999).
- [21] W. AIELLO, F. CHUNG, and L. LU, *Proceedings of the thirty-second annual ACM symposium on Theory of computing* **97**, 171 (2000).
- [22] F. LIJEROS, C. R. EDLING, L. A.Ñ. AMARAL, H. E. STANLEY, and Y. ABERG, *Nature* **411**, 907 (2001).
- [23] R. F. I CANCHO and R. V. SOLÉ, *Proc. Royal Soc. London B* **268**, 2261 (2001).
- [24] S.Ñ. DOROGOVITSEV and J. F. F. MENDES, *Proc. Royal Soc. London B* **268**, 2603 (2001).
- [25] M. RIPEANU, I. FOSTER, and A. IAMNITCHI, *IEEE Internet Computing* **6**, 50 (2002).
- [26] E. P. and R. A., *Publ. Math. Debrecen* **6**, 290 (1959).
- [27] E. P. and R. A., *Publ. Math. Inst. Hung. Acad. Sci* **5**, 17 (1960).
- [28] E. P. and R. A., *Bull. Inst. Internat. Tokio* **38**, 343 (1961).
- [29] A.-L. BARABÁSI and R. ALBERT, *Science* **286**, 509 (1999).
- [30] A.-L. BARABÁSI, R. ALBERT, and H. JEONG, *Physica A* **272**, 173 (1999).
- [31] R. W. FLOYD, *Communication of the ACM* **5**, 345 (1962).
- [32] T. H. CORMEN, C. E. LEISERSON, and R. L. RIVEST, *Introduction to algorithms*, MIT Press y McGraw-Hill. ISBN 0-262-03141-8, priemra edición edition, 1990.
- [33] H. A. TAHA, *Investigación de Operaciones*, Pearson Educación, 7 español edition, 2004.
- [34] A. J. SCOTT, *Statistical Mechanics of Complex Networks, Lecture Notes in Physics, Springer (Berlin)* **625**, 114 (2003).
- [35] M. MITCHELL, *An introduction to genetic algorithms*, The MIT press, first edition edition, 1998.
- [36] J.ÑOCEDAL and S. J. WRIGHT, *Numerical optimization*, Springer, second edition edition, 2006.
- [37] D. GOLDBERG, *Genetic Algorithms*, Addison-Wesley Longman Publishing Co., 1989.
- [38] A. J. SCOTT, *Transportation Research* **3**, 201 (1969).
- [39] J. W. BILLHEIMER and P. GRAY, *Transportation science* **7**, 49 (1973).
- [40] N. MATHIAS and V. GOPAL, *Physical reveiw E* **63** (2001).
- [41] M. T. GASTNER and M. E. J. NEWMAN, *Journal of Statistical Mechanics* (2006).
- [42] A. FABRIKANT, E. KOUTSOUPIAS, and C. H. PAPADIMITRIOU, *In the Proceedings of the International Conference on Automata, Languages, and Programming (ICALP) 2002*, , 110 (2002).