

Redes Complejas: Estructura, Dinámica y Evolución

Maximino Aldana
(Dated: Diciembre 2011)

“La materia no sólo interactúa, también se organiza. Conocemos básicamente todas las leyes de interacción de la materia, pero no sabemos casi nada sobre sus leyes de organización”.

Albert L. Lehninger (1917-1986)

I. INTRODUCCIÓN

En los últimos 11 años hemos sido testigos de una explosión en el estudio de las propiedades estructurales y dinámicas de las redes complejas. Durante este tiempo se han publicado cientos de artículos sobre este tema en revistas de investigación científica internacionales de diferentes disciplinas, que abarcan física, biología, sociología, neurología, economía, medicina, por mencionar algunos ejemplos. Este interés en las redes complejas radica en que nos hemos dado cuenta de que dichas redes abundan en la naturaleza, son parte de nuestra vida diaria y se presentan a diferentes niveles de organización. Por ejemplo, algunas redes biológicas que encontramos en el nivel microscópico son las *redes de regulación genética*, *redes de proteínas*, *redes neuronales*, *redes metabólicas*. Por otro lado, a un nivel de organización mucho mayor, encontramos *redes de comunicación e informáticas* (la red internet, la red www, redes telefónicas, etc.), *redes sociales* (amistades, contactos sexuales, colaboradores científicos, propagación de enfermedades, etc.), *redes ecológicas* (interacciones tróficas en un ecosistema). Las redes complejas son ubicuas, están por todos lados. Incluso se ha estudiado la red de super héroes en el Universo de Marvel, siendo el hombre ara na el super héroe más popular con la mayor conectividad [1]. Es un hecho sobresaliente el que todas estas redes, tan diferentes en naturaleza y en tama no, tengan muchas propiedades estructurales similares. Este hecho, tan simple como sorprendente, hace posible que podamos formular modelos matemáticos para entender y explicar las propiedades estructurales (y en algunos casos también las propiedades dinámicas) de las redes complejas.

En estas notas presento algunas de las herramientas matemáticas y conceptuales que se han desarrollado a lo largo de varios a nos para analizar la estructura y dinámica de las redes complejas. Los que estén interesados podrán encontrar un estudio mucho más completo del que presento aquí en las referencias [2–5], las cuales son trabajos de revisión excelentes que sirven como punto de partida para adentrarse en el fascinante mundo de las redes complejas.

II. ALGUNAS DEFINICIONES

A. Sistemas complejos

Comenzemos nuestro estudio describiendo qué son los sistemas complejos. Como ocurre con la gran mayoría de los conceptos científicos, no podemos definir los sistemas complejos en un simple enunciado. En lugar de ésto, vamos a enumerar las características más importantes que son comunes a todos los sistemas complejos:

1. Están compuestos de muchas partes que interactúan entre sí. De hecho, el adjetivo “Complejo” en este contexto no significa solamente que el sistema sea complicado, sino también que está compuesto de muchas partes, como un *complejo* industrial.
2. Cada parte tiene su propia estructura interna y está encargada de llevar a cabo una función específica.
3. Lo que ocurra a una parte del sistema afecta de manera **altamente no lineal** a todo el sistema.
4. Presentan **comportamientos emergentes**, de tal manera que el *el todo no es la simple suma de sus partes*.

Como un ejemplo típico de sistema complejo consideremos a la célula. Evidentemente la célula está compuesta de muchas partes (ribosomas, mitocondrias, núcleo, membrana, retículo endoplasmático, ADN, ARN, etc.), y cada una de estas partes se encarga de realizar alguna función específica dentro de la célula. Las partes de la célula responden de forma no lineal ante perturbaciones externas. Por ejemplo, algunas veces una mutación en el ADN no tiene ningún efecto en la célula, mientras que otras veces una sola mutación puede ser fatal[14]. Además, la célula presenta comportamientos emergentes que no pueden explicarse en términos de las propiedades de sus partes individuales. Así, podemos hablar de una célula enferma, pero no podemos decir que un ribosoma o una proteína estén enfermos. La enfermedad es una propiedad que emerge como resultado de la organización colectiva de todos los constituyentes de la célula.

B. Redes Complejas

Las redes complejas son conjuntos de muchos nodos conectados que interactúan de alguna forma. A los nodos de una red también se les llama *vértices* o *elementos* y los representaremos por los símbolos v_1, v_2, \dots, v_N ,

donde N es el número total de nodos en la red. Si un nodo v_i está conectado con otro nodo v_j , esta conexión se representa por una pareja ordenada (v_i, v_j) . La definición matemática de una red (también llamada *grafo* por los matemáticos) es la siguiente:

Definición 1 Una red \mathcal{R} consiste de un conjunto de nodos $\mathcal{V} = \{v_1, v_2, \dots, v_N\}$, y un conjunto de parejas ordenadas $\mathcal{E} = \{(v_i, v_j)\} \subset \mathcal{V} \times \mathcal{V}$. Cada pareja ordenada (v_i, v_j) se llama **conexión dirigida** del nodo v_i al nodo v_j . La red \mathcal{R} se llama **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**. Llamaremos a todos los nodos que estén conectados directamente a un nodo v_i , los **vecinos** de v_i . Finalmente, el número k_i de vecinos del nodo v_i (es decir, el número de conexiones de v_i) se llama la **conectividad** de v_i , y el promedio de estas conectividades, $\langle k \rangle = N^{-1} \sum_{i=1}^N k_i$, es la **conectividad de la red**.

Aunque la definición formal de una red es útil en el desarrollo matemático de la teoría, para nuestros propósitos basta con considerar que una red es un montón de nodos entre los que existen conexiones. En la naturaleza se pueden encontrar muchos tipos de redes, es decir, muchos tipos de nodos y conexiones. Por ejemplo, en una red social los nodos son las personas y las conexiones pueden ser los lazos de amistad que existan entre ellas: dos personas están conectadas si son amigos. En la misma sociedad podemos definir las conexiones de forma distinta, por ejemplo, dos personas están conectadas si han tenido relaciones sexuales. Claramente, la red definida a través de amistades es diferente a la red definida a través de contactos sexuales, ya que el hecho de que dos personas sean amigas no significa que hayan tenido relaciones sexuales, y viceversa.

Notemos entonces que incluso en un mismo conjunto de nodos podemos definir redes diferentes dependiendo de como hayamos definido las conexiones, lo cual, por supuesto, depende del fenómeno que nos interese estudiar. Por ejemplo, si estuviésemos interesados en analizar como se propaga una enfermedad como el SIDA en una sociedad, claramente nos convendría estudiar la red de interacciones sexuales, mientras que si estamos interesados en encontrar a un asesino, lo que nos conviene es estudiar la red de amistades, ya que son sus amigos los que pueden darnos información sobre su paradero. El Cuadro I muestran diferentes tipos de redes que se encuentran en la naturaleza.

En los ejemplos anteriores hay redes dirigidas y redes no dirigidas. Por ejemplo, la red de contactos sexuales es *no dirigida*, ya que si A tuvo relaciones con B, entonces evidentemente B tuvo relaciones con A. Pero la red de transmisión de la gripe es *dirigida*, ya que si A contagió de gripe a B, no necesariamente B contagió también a A. Otra red dirigida es la World Wide Web (WWW). En mi página web yo tengo una liga a la página del periódico *la jornada*, pero en la página de *la jornada* no hay ninguna liga a mi página web. En este sentido, hay una conexión

Redes sociales	
Sexuales	Dos personas están conectadas si han tenido por lo menos una relación sexual
Actores	Dos actores están conectados si han aparecido en la misma película
Amistades	Dos personas están conectadas si son amigas
Científicos	Dos científicos están conectados si han sido coautores en algún artículo
Familiares	Dos personas están conectadas si son familiares cercanos
Enfermedades	Dos personas están conectadas si una contagió de una enfermedad a la otra
Redes informáticas	
Internet	Dos computadoras están conectadas si hay un cable que las conecta
WWW	Dos páginas web están conectadas si hay un hipervínculo de una a la otra
Palabras	Dos palabras están conectadas si en el diccionario una aparece en la definición de la otra
Palabras	Dos palabras están conectadas si son sinónimos
Redes biológicas	
Protéicas	Dos proteínas están conectadas si participan en la misma reacción química
Genéticas	Dos genes están conectados si uno regula la expresión del otro
Ecológicas	Dos especies están conectadas si una se come a la otra
Neuronales	Dos neuronas están conectadas si existe una conexión sináptica entre ellas

Cuadro I. Diferentes tipos de redes.

de mi página hacia la de *la jornada*, pero no hay una conexión de regreso.

Intuitivamente, una red no dirigida puede pensarse como aquella en la que las conexiones entre los nodos siempre son simétricas (si A está conectado con B, entonces B está conectado con A), mientras que en una red dirigida no todas las conexiones son simétricas, es decir, siempre existen conexiones asimétricas (A está conectado con B pero B no está conectado con A).

Otro concepto importante es el de *islas* (o *sub redes*) de una red. Notemos que la definición de red que dimos arriba **no dice** que todos los nodos deben estar conectados unos con otros. Ni siquiera dice que todos los nodos deben tener conexiones. La definición matemática nos dice que una red es un conjunto de nodos entre los que existen *algunas* conexiones. Esto quiere decir que en la red pueden existir nodos que no tengan conexiones, es decir, nodos aislados. También pueden existir grupos de nodos que estén conectados entre sí pero que no estén conectados con el resto de la red. Como un ejemplo concreto pensemos en una red social en la que dos individuos están conectados sin ser familiares cercanos (específicamente, hermanos, medios hermanos, primos, padres, hijos, esposos, tíos, sobrinos, abuelos y nietos). Esta es evidentemente

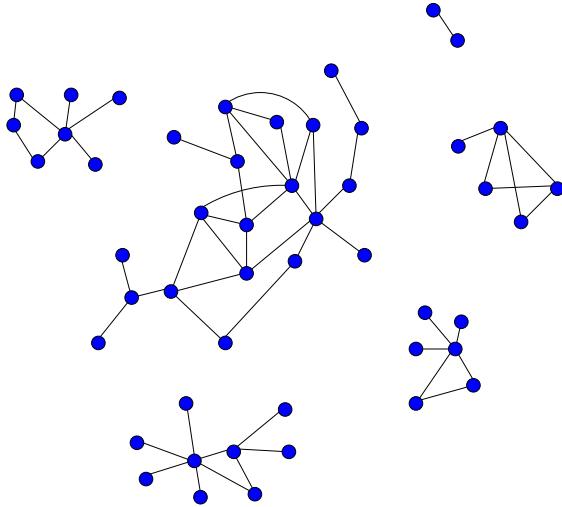


Figura 1. Una red puede estar compuesta de varias islas, como en el ejemplo mostrado en esta figura. La isla más grande se denomina la *isla gigante*.

mente una red no dirigida, ya que si A es familiar de B, entonces B también es familiar de A. Sin embargo, muy probablemente en una sociedad grande esta red estaría fracturada en islas o sub redes, debido a que claramente no todas las personas en una sociedad son familiares cercanos de todos los demás. En mi caso particular, mi apellido es “Aldana González”, por lo que es natural pensar que estoy conectado directamente con algunas de las familias Aldana y algunas las familias González en México. Pero muy probablemente ningún miembro de mi familia está conectado con alguien de la familia Azcárraga, o con la familia Zabludovsky. Por lo tanto, los miembros de mi familia conforman una isla o sub red dentro de la cual estamos conectados entre nosotros, pero esta isla está desconectada de otras familias de la sociedad. La Fig. 1 muestra un ejemplo de una red compuesta de varias islas.

Las islas en una red pueden tener diferentes tamaños, que van desde 1 (un sólo nodo que no está conectado a nadie) hasta el tamaño de toda la red (todos los nodos están conectados con todos), en cuyo caso la red consiste de una sola isla, que es ella misma. Es importante enfatizar que el hecho de que una isla no esté conectada al cuerpo principal de la red no significa que dicha isla no pertenezca a la red. La red no está determinada sólo por las conexiones, sino también por los nodos que conforman al sistema. Esto puede parecer poco intuitivo, pero desde el punto de vista matemático es conveniente considerar que todos los nodos del sistema pertenecen a la red, independientemente de que haya o no conexiones entre ellos. Como veremos más adelante, las islas juegan un papel importante en la teoría de redes.

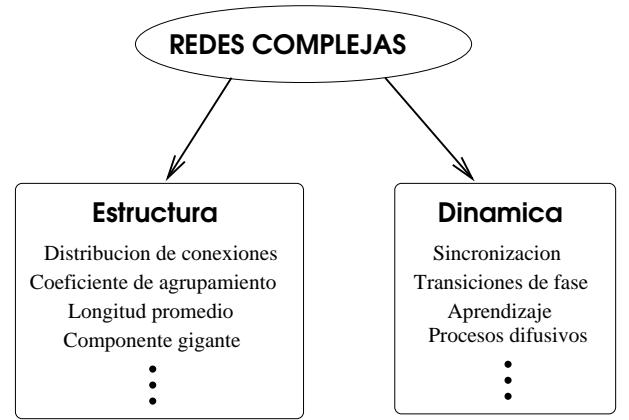


Figura 2. El estudio de las redes complejas puede dividirse en dos partes: (a) el estudio de sus propiedades estructurales y (b) el estudio de sus propiedades dinámicas.

III. ESTUDIO DE LAS REDES COMPLEJAS

Ejemplos de redes dirigidas y no dirigidas abundan en la naturaleza. Ahora ustedes pueden comenzar a pensar en la red que más les guste, ya sea una red biológica, social, informática o cualquier otra. Las redes se presentan en diferentes tamaños, colores y sabores. Pero, ¿qué hacemos con las redes? ¿Cómo las estudiamos?

El estudio general de las redes complejas puede dividirse en dos campos diferentes y complementarios: *Estructura* y *Dinámica* (ver la Fig. 2). En el primer campo de estudio uno está interesado en determinar las propiedades estructurales (o topológicas) de la red, es decir, en las propiedades que nos dicen cómo están conectados los nodos unos con otros. Algunas de las propiedades más importantes que determinan la estructura (o topología) de una red son las siguientes:

1. **La distribución de conexiones (o vecinos)**
 $P(k)$: Es la probabilidad de que un nodo escogido al azar tenga k conexiones (o vecinos). Por ejemplo, en una red de contactos sexuales $P(k)$ es la probabilidad de que una persona escogida al azar en una sociedad haya tenido k parejas sexuales distintas a lo largo de su vida.
2. **El coeficiente de agregación C** : Es la probabilidad de que dos nodos conectados directamente a un tercer nodo, estén conectados entre sí (ver la Fig. 3(a)). Por ejemplo, en una red de amistades, es la probabilidad de que dos de mis amigos sean ellos mismos amigos uno del otro.
3. **La longitud mínima L_{ij} entre dos nodos v_i y v_j** : Es el número mínimo de “brincos” que se tienen que dar para llegar de un nodo v_i de la red a otro nodo v_j de la red. Por ejemplo, en la red mostrada en la Fig. 3(b), aunque existen varios caminos para

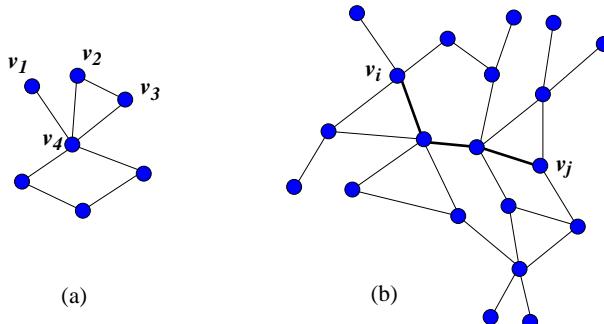


Figura 3. (a) Los nodos v_1 , v_2 y v_3 están conectados al nodo v_4 . Sin embargo, los nodos v_1 y v_2 no están conectados entre sí, mientras que los nodos v_2 y v_3 sí lo están. Esto significa que no todos los “amigos” de v_4 son amigos entre sí, lo cual disminuye el coeficiente de agregación. (b) Aún cuando existen varios caminos para llegar del nodo v_i al nodo v_j , el camino de longitud mínima consiste de tres pasos, indicados con líneas gruesas en la figura.

llegar de v_i a v_j , el camino mínimo consiste de tres pasos (indicado con líneas gruesas).

4. **La longitud promedio de la red L :** Es el promedio de las longitudes mínimas L_{ij} entre todas las posibles parejas de nodos (v_i, v_j) de la red.
5. **La distribución de tamaños de islas $P(s)$:** Es la probabilidad de que una isla esté compuesta por s nodos.
6. **El tamaño de la isla más grande,** al que denotaremos por S_∞ .

En una red, los nodos además de estar conectados también interactúan, y las interacciones pueden dar lugar a fenómenos dinámicos muy interesantes. Por lo tanto, además de estudiar las propiedades estructurales de una red también es importante estudiar sus propiedades dinámicas una vez que sabemos de qué manera interactúan los nodos. Por ejemplo, las enfermedades en una sociedad no son estáticas, sino que se propagan por toda la población dando lugar a epidemias. Las neuronas en el cerebro están conectadas físicamente unas con otras por medio de las uniones entre dendritas y axones. A través de dichas uniones las neuronas se transmiten señales eléctricas que se propagan por todo el cerebro y que dan lugar a una serie de fenómenos dinámicos interesantísimos, entre los cuales destacan el reconocimiento de imágenes y sonido, la motricidad de los músculos, el lenguaje, el pensamiento y finalmente la conciencia. Otros ejemplos son la propagación de virus informáticos en la red internet, o la comunicación entre los peces que da lugar a grupos enormes de peces moviéndose todos en la misma dirección. En fin, existen tantos fenómenos dinámicos en redes complejas como interacciones físicas, químicas, informáticas, o sociales se puedan imaginar, y

cada día aparecen más y más artículos en la literatura científica donde se estudian nuevos procesos dinámicos sobre redes complejas.

En estas notas introductorias nos enfocaremos más al estudio de las propiedades estructurales de las redes complejas. Sin embargo, en el último capítulo veremos un poco las propiedades dinámicas de redes neuronales.

IV. ESTRUCTURA DE LAS REDES COMPLEJAS

A. Distribución de vecinos

Tal vez la propiedad más importante que caracteriza la estructura de una red compleja es la distribución de vecinos $P(k)$, que nos dà la probabilidad de que un nodo escogido al azar tenga k conexiones (o vecinos). En los trabajos recientes que se han llevado a cabo para caracterizar a las redes complejas se ha encontrado que existen tres tipos de distribuciones $P(k)$ importantes, las que determinan tres estructuras o topologías[15] diferentes:

$$\text{Topología de Poisson } P(k) = e^{-z} \frac{z^k}{k!}, \quad (1)$$

$$\text{Topología Exponencial } P(k) = Ce^{-\alpha k}, \quad (2)$$

$$\text{Topología Libre de Escala } P(k) = Ck^{-\gamma}. \quad (3)$$

Las redes con topología de Poisson son importantes principalmente por razones históricas, ya que dichas redes fueron las primeras que se analizaron matemáticamente. Este análisis lo llevaron a cabo los matemáticos húngaros Paul Erdős (1913-1996) y Alfréd Rényi (1921-1970) en la década de los 50s. Ellos también reportaron la primera transición de fase topológica observada en redes con topología de Poisson. Por lo tanto, a estas redes también se les conoce como redes tipo Erdős-Rényi. 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. No fue sino hasta 1998 que se comenzó el estudio sistemático de las propiedades topológicas de las redes complejas reales. En este estudio participaron principalmente y de forma independiente Albert László-Barabási, Ricard Solé y Mark J. Newman. Ellos encontraron que la topología exponencial aparece algunas veces en las redes reales. Pero el resultado más sorprendente de sus estudios fue la ubicuidad de las redes con topología libre de escala, la cual aparece prácticamente en todos lados, desde las pequeñas redes metabólicas dentro de la célula, hasta las grandes redes informáticas como la red Internet. En el Cuadro II se listan algunas de las redes con topologías libres de escala que se han encontrado en los últimos 10 años.

Red	Núm. de nodos	Núm. de conexiones	γ_i
dominio www.nd.edu	325,729	1,469,680	2.1
Páginas de WWW encontradas por Altavista	$2,711 \times 10^9$	$2,130 \times 10^9$	2.1
Dominios en la WWW	$2,60 \times 10^5$	—	1.9
Nivel de inter-dominio de la Internet	4,389	8,256	2.2
Sistemas autónomos en la Internet	6,374	13,641	2.2
Nivel de ruteador en la Internet	150,000	200,000	2.3
Citas en la base de datos ISI	783,339	6,716,198	3.0
Citatas de la revista <i>Phys. Rev. D</i>	24,296	351,872	2.3
Red de colaboraciones de actores de Hollywood	212,250	61,085,555	2.3
Red de colaboradores en las revistas <i>Medline</i>	1,388,989	$1,028 \times 10^7$	2.5
Colaboradores en las revistas de matemáticas	70,975	$1,32 \times 10^5$	2.1
Colaboradores en las revistas de neurociencias	209,293	$1,21 \times 10^6$	2.4
Red de interacciones metabólicas en (<i>E. coli</i>)	778	$\sim 1500 - 3000$	2.2
Red de interacciones protéicas en levadura	1,870	2,240	2.5
Co-ocurrencia de palabras	470,000	17,000,000	2.7
Red de palabras sinónimas	22,311	—	2.8
Circuitos digitales	2×10^4	4×10^4	3.0
Llamadas telefónicas	47×10^6	8×10^7	2.1
Red de interacciones sexuales en humanos	2,810	—	3.4
Redes alimenticias (interacciones tróficas)	154	405	1.0

Cuadro II. Algunas de las redes libres de escala que se han encontrado en la naturaleza. Sólo se muestra el exponente de entrada para las redes dirigidas. Los cuadros con líneas “—” indican que yo no tenía el dato correspondiente al momento de escribir estas notas.

Seguramente se estarán preguntando qué es lo sorprendente respecto a la topología libre de escala. Bueno, por un lado sorprende que esta topología se encuentre en redes tan diferentes y de tan gran variedad como las listadas en el Cuadro II. El hecho de que la topología libre de escala aparezca por todos lados sugiere que podría existir un mecanismo simple que genera este tipo de redes a diferentes niveles de organización, desde las pequeñas redes intracelulares hasta las grandes redes sociales o informáticas. ¿Podría ser esto posible? ¿Será cierto que la formación de redes tan diferentes como la red de interacciones protéicas de *S. cerevisiae* (levadura), la red Internet y la red de colaboraciones científicas en las revistas de neurociencias, esté gobernada por la misma ley fundamental? No lo sabemos aún.

Por otro lado, la topología libre de escala es sorprendente porque no se esperaba que existiera. En la naturaleza existen muchos procesos aleatorios que generan distri-

buciones de Poisson o distribuciones exponenciales, pero existen muy pocos procesos conocidos que generan distribuciones libres de escala como la dada en la Eq. (3)[16]. De hecho, el trabajo de Erdős y Rényi demostró que las redes que se construyen a naciendo nuevos nodos y conexiones al azar presentan topologías de Poisson o exponenciales. Y ésta es precisamente la paradoja, que la red internet, la red de colaboraciones científicas y la red de contactos sexuales, por ejemplo, son redes que se formaron aleatoriamente a naciendo nuevos nodos y nuevas conexiones a lo largo del tiempo. Entonces, ¿cómo es posible que estas redes que se formaron al azar no presenten topologías de Poisson o topologías exponenciales como lo habían predicho Erdős y Rényi?

Las redes con topología de Poisson son muy diferentes estructuralmente a las redes con topología libre de escala. La Fig. 4 muestra una red de Poisson (arriba) y una red libre de escala (abajo). Como puede observarse,

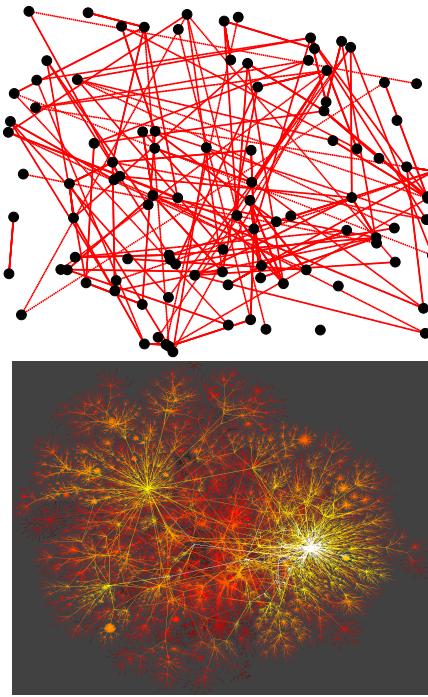


Figura 4. La figura superior muestra una red aleatoria con topología de Poisson. La red mostrada en el panel inferior es la red Internet al nivel de ruteadores, la cual es una red con topología libre de escala.

la red de Poisson se ve más aleatoria y más homogénea que la red libre de escala[17]. En las redes de Poisson todos los nodos tienen más o menos el mismo número de conexiones. Algunos nodos estarán más conectados que otros, pero en promedio todos tienen la misma conectividad, es decir, las conexiones en una red de Poisson están distribuidas homogéneamente entre sus nodos. Por el contrario, la característica más importante de las redes libres de escala es su alta heterogeneidad, ya que existen nodos con muy pocas conexiones, nodos medianamente conectados y nodos extremadamente conectados. Los nodos altamente conectados se denominan los *núcleos* o *centros* de la red[18]. Con una red libre de escala uno no puede decir que todos los nodos tienen “más o menos” la misma conectividad. Por el contrario, hay nodos con una sola conexión y también hay nodos con miles de conexiones.

Todavía no sabemos cuáles son los procesos que conducen a la formación de redes libres de escala. Sin embargo, en los últimos 8 años se han llevado a cabo avances sustanciales en la comprensión de estas redes. En las secciones que siguen veremos algunos de los formalismos matemáticos de crecimiento de redes que se han desarrollado para generar las tres topologías mencionadas con anterioridad (Poisson, exponencial y libre de escala).

V. REDES DE TIPO ERDÖS-RÉNYI

Imaginemos un conjunto de N botones de pantalón distribuidos aleatoriamente sobre una mesa e inicialmente desconectados. Al tiempo $t = 0$ escogemos aleatoriamente una pareja de botones y los hilvanamos con un hilo. Después de haber enlazado a esta pareja, la dejamos sobre la mesa y escogemos aleatoriamente otra pareja para hilvanar. Podemos escoger botones que están conectados con otros botones, pero si la pareja que escogemos ya está conectada entre sí, la descartamos y escogemos otra pareja. Lo que no se vale es hilvanar más de una vez a la misma pareja de botones. Repetimos este proceso sucesivamente M veces, escogiendo aleatoriamente una pareja de botones cada vez. Al final del proceso habremos establecido M enlaces entre M parejas diferentes de botones, generando así una red de botones. Intuitivamente es claro que si M (el número total de enlaces) es pequeño comparado con N (el número total de botones), entonces la red resultante estará desmembrada en varias islas pequeñas. Dentro de cada isla los botones estarán hilvanados entre sí, pero estarán desconectados de las otras islas. Sin embargo, si M es muy grande comparado con N , terminaremos con casi todos los botones hilvanados unos con otros. Probablemente haya islas muy pequeñas desconectadas de la red principal, pero seguramente la gran mayoría de botones formarán parte de una isla principal: la isla gigante.

Después de haber hilvanado M parejas en un conjunto total de N botones, ¿cuál es la distribución de conexiones $P(k)$ en la red resultante? Como veremos en un momento, la red que resulta de este proceso tiene una distribución $P(k)$ de Poisson. Pero antes de dar la prueba de este resultado, vale la pena mencionar que durante muchos años se pensó que este mecanismo de formación de redes en el cual parejas de nodos se enlanzan 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 o los contactos sexuales se dan por el encuentro casual y aleatorio de las personas que viven en una sociedad. Por lo tanto, era natural pensar que el mecanismo de “hilvanar parejas de botones escogidas al azar” reproducía lo que realmente ocurre en las redes sociales. Sí, era “natural” pensar lo, pero era incorrecto.

Calculemos ahora la probabilidad $P(k)$ para nuestra red de botones hilvanados. Para comenzar notemos que el número total N_p de parejas que se pueden formar en un conjunto de N botones es

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

Como enlazamos M parejas de botones, la probabilidad p_e de que una pareja arbitraria seleccionada al azar esté enlazada es

$$p_e = \frac{M}{N_p} = \frac{2M}{N(N - 1)} \quad (4)$$

Ahora enfoquemos nuestra atención sobre un nodo particular v_j de la red, escogido al azar. El número total de parejas que podrían contener a v_j es $N - 1$, ya que v_j se podría haber hilvanado con los $N - 1$ nodos restantes de la red. Sin embargo, en los M enlaces que se llevaron a cabo, no necesariamente escogimos al nodo v_j todas las veces posibles que se podría haber escogido. Supongamos entonces que de las M parejas que se escogieron, el nodo v_j estaba solamente en k de ellas. La probabilidad de que v_j esté contenido en k parejas de las $N - 1$ posibles es

$$P(k) = \binom{N-1}{k} (p_e)^k (1-p_e)^{N-1-k} \quad (5)$$

Esta es una distribución binomial para N y M finitas. Pero si consideramos ahora que la red es muy grande y tomamos el límite $N \rightarrow \infty$ y $M \rightarrow \infty$ de tal forma que la cantidad

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

permanezca finita, entonces la distribución (5) se transforma en

$$P(k) = e^{-z} \frac{z^k}{k!} \quad (6)$$

lo cual es la distribución de Poisson con promedio z . En el apéndice A muestro los pasos algebraicos que conducen de la ecuación (5) a la ecuación (6).

VI. CRECIMIENTO DE REDES

En la sección anterior suponímos que teníamos una población fija de N nodos y un número fijo M conexiones a nadidas aleatoriamente. Sin embargo, en la realidad esto no ocurre, las redes no están fijas. Por el contrario, las redes complejas evolucionan y crecen en el tiempo a través de la adición simultánea tanto de conexiones como de nodos. Pensemos por ejemplo en la red internet. En octubre de 1969 la “redinternet” consistía de sólo dos computadoras, una en la Universidad de California en Los Angeles (UCLA) y la otra en el Instituto de Investigaciones de Stanford (SRI). El primer mensaje que se transmitieron estas computadoras fue “LOGWIN”[19]. A lo largo de los años más y más computadoras se sumaron a la red internet, y para el año 2000, las primeras dos computadoras de la UCLA y el SRI que se dijeron “LOGWIN”, se habían convertido ya en sistemas que conectaban a 170 países y a más de 300 millones de personas.

Así como la red internet nació siendo pequeña y después creció con el paso del tiempo, también las redes metabólicas y las redes genéticas dentro de la célula, y muchas otras redes en la naturaleza, han crecido y evolucionado a lo largo del tiempo. Por lo tanto, es importante que nuestros modelos de formación de redes incorporen

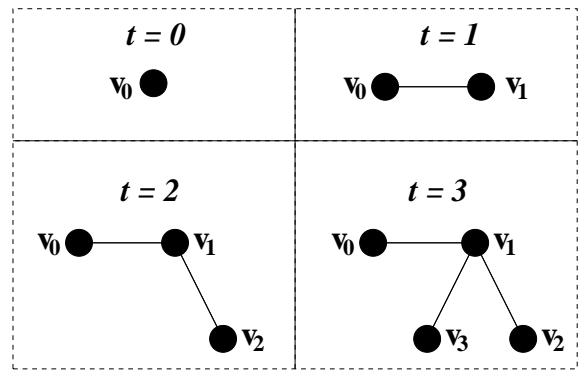


Figura 5. Crecimiento de redes. Al tiempo $t = 0$ hay un sólo nodo v_0 . En cada paso de tiempo subsecuente a nadimos un nuevo nodo que se conectarán a alguno de los nodos ya existentes con una probabilidad $\Pi(k, t)$ que depende de la conectividad k al tiempo t del nodo con el que se pretende establecer la conexión.

el hecho de que nuevos nodos y nuevas conexiones se pueden aadir a la red. También debe tomarse en cuenta el que nodos y conexiones ya existentes pueden eliminarse.

En el modelo más simple de crecimiento de redes a nadimos un nuevo nodo en cada paso de tiempo. Este nuevo nodo puede conectarse con alguno de los nodos ya existentes. Cada uno de los nodos ya existentes pueden ser seleccionados para la conexión con una probabilidad $\Pi(k_i, t)$, siendo k_i la conectividad al tiempo t del i -ésimo nodo ya existente.

El proceso comienza con un único nodo inicial v_0 al tiempo $t = 0$ (ver la Fig. 5). Al tiempo $t = 1$ a nadimos un nuevo nodo v_1 , que se conectarán al único nodo ya existente v_0 con probabilidad 1. Después, al tiempo $t = 2$ a nadimos el nodo v_2 que se conectarán con cualquiera de los nodos ya existentes v_1 y v_2 con la misma probabilidad 1/2. En este momento los nodos v_0 , v_1 y v_2 ya no tienen todos la misma conectividad: alguno de ellos tendrá dos conexiones mientras que los otros dos nodos tendrán sólo una conexión. Al tiempo $t = 3$ a nadimos al nodo v_3 , que se conectarán con alguno de los nodos ya existentes v_0 , v_1 o v_2 con una probabilidad que es función de sus conectividades k_0 , k_1 y k_2 . Continuando con este proceso, al tiempo $t+1$ a nadimos al nodo v_{t+1} que se conectarán a cualquiera de los nodos ya existentes v_0, v_1, \dots, v_t , el cual será seleccionado con una probabilidad $\Pi(k_i, t)$, siendo k_i la conectividad al tiempo t del nodo seleccionado.

Denotemos por $P(k, t)$ a la probabilidad de que al tiempo t un nodo arbitrario de la red tenga k conexiones. Es claro que esta probabilidad depende del tiempo. Sin embargo, si continuamos a nadiendo nodos por un tiempo suficientemente largo esperamos que la función $P(k, t)$ alcance un estado estacionario independiente del tiempo. Esto no significa que la red alcance un estado estacionario. La red sigue creciendo mientras continuemos a nadiendo nodos. Es únicamente la distribución de conectividades $P(k, t)$ la que llega a un estado estacionario

en el cual $P(k, t+1) = P(k, t) = P(k)$ [20].

Existen varios métodos para calcular la distribución de conectividades $P(k, t)$. En estas notas veremos el método de la *ecuación maestra*, pero existen por lo menos otros dos métodos distintos: el *método continuo*, inventado por Barabási, y el *método cinético*, introducido por Krapivsky, Redner, y Leyvraz [2]. Los tres métodos dan resultados equivalentes.

Para aquellos que no conocen el formalismo de la ecuación maestra, en el apéndice B doy una breve introducción al planteamiento de dicha ecuación, la cual es muy fácil de escribir pero, la mayoría de las veces, muy difícil de resolver.

Denotemos por $P(n, k, t)$ la probabilidad de que el nodo v_n tenga k conexiones al tiempo t . Notemos que esta probabilidad está asociada al nodo específico v_n . Sin embargo, podemos obtener la probabilidad $P(k, t)$ de que un nodo *arbitrario* tenga k conexiones al tiempo t a través de la siguiente ecuación:

$$P(k, t) = \frac{1}{N(t)} \sum_{n=0}^t P(n, k, t), \quad (7)$$

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

La ecuación que determina la evolución temporal de $P(n, k, t)$ se obtiene notando que en cada paso de tiempo hay dos contribuciones a dicha probabilidad:

1. Al tiempo t el nodo v_n tenía $k-1$ conexiones y fue seleccionado (con probabilidad $\Pi(k-1, t)$) para conectarse con el nuevo nodo a nadido 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 a nadido (lo cual ocurre con probabilidad $1 - \Pi(k, t)$). Por lo tanto, al tiempo $t+1$ el nodo v_n seguirá teniendo k conexiones.

Tomando en cuenta estas dos contribuciones, la ecuación maestra que determina la evolución temporal de $P(n, k, t)$ es

$$\begin{aligned} P(n, k, t+1) &= \underbrace{P(n, k-1, t)}_{\substack{k-1 \text{ conexiones} \\ \text{al tiempo } t}} \overbrace{\Pi(k-1, t)}^{v_n \text{ fue} \\ \text{seleccionado}} \\ &\quad + \underbrace{P(n, k, t)}_{\substack{k \text{ conexiones} \\ \text{al tiempo } t}} \overbrace{[1 - \Pi(k, t)]}^{v_n \text{ no fue} \\ \text{seleccionado}}. \end{aligned} \quad (8)$$

Como el nodo v_n “nació” al tiempo $t = n$ con una sola conexión, la condición inicial para resolver la ecuación anterior es

$$P(n, k, t)|_{t=n} = \delta_{k,1}. \quad (9)$$

Sumando sobre n en la Eq. (8) desde $n = 0$ hasta $n = t$ y tomando en consideración la Eq. (7), obtenemos

$$\begin{aligned} \sum_{n=0}^t P(n, k, t+1) &= N(t) \left\{ P(k-1, t) \Pi(k-1, t) \right. \\ &\quad \left. + P(k, t) [1 - \Pi(k, t)] \right\} \end{aligned} \quad (10)$$

Ahora bien, como en cada paso de tiempo a nadimos un nuevo nodo (comenzando con el nodo v_0 al tiempo $t = 0$), entonces $N(t) = t + 1$. Por lo tanto, la ecuación anterior puede escribirse como

$$\begin{aligned} \sum_{n=0}^t P(n, k, t+1) &= (t+1) \left\{ P(k-1, t) \Pi(k-1, t) \right. \\ &\quad \left. + P(k, t) [1 - \Pi(k, t)] \right\} \end{aligned} \quad (11)$$

En el lado izquierdo de la ecuación anterior podemos completar la suma hasta $t+1$ de la siguiente manera:

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

donde hemos utilizado el hecho de que

$$P(t+1, k, t+1) = \delta_{k,1}$$

tal y como lo establece la Eq. (9). También utilizamos la Eq. (7) (evaluada en $t+1$) y el hecho de que $N(t+1) = t+2$. Sustituyendo estos resultados en la Eq. (11) obtenemos

$$\begin{aligned} (t+2)P(k, t+1) - \delta_{k,1} &= (t+1) \left\{ P(k-1, t) \Pi(k-1, t) \right. \\ &\quad \left. + P(k, t) [1 - \Pi(k, t)] \right\} \end{aligned} \quad (12)$$

Notemos que para poder resolver la Eq. (12) necesitamos 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 a nade al tiempo t . Como veremos en las siguientes secciones, diferentes formas de la probabilidad $\Pi(k, t)$ conducen a topologías diferentes.

VII. TOPOLOGÍA EXPONENCIAL

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

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

donde $N(t) = t+1$ es el número total de nodos al tiempo t . Sustituyendo esta forma de $\Pi(k)$ en la Eq. (12) obtenemos

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

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 dicho estado estacionario la última ecuación se transforma en

$$P(k) = \frac{1}{2}(P(k-1) + \delta_{k,1}). \quad (13)$$

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

$$P(k) = 2^{-k} \quad (14)$$

que no es más que la distribución exponencial. Es importante señalar que esta distribución aparece en el contexto del crecimiento de redes como el resultado de lo que se llama *enlace igualitario*, es decir, en una situación en la cual cada nodo nuevo que se añade a la red se puede enlazar a cualquiera de los nodos ya existentes con la misma probabilidad. En este sentido el nuevo enlace que se forma es igualitario, porque no discrimina entre los nodos ya existentes.

VIII. TOPOLOGÍA LIBRE DE ESCALA

En la vida real las conexiones entre diferentes nodos no se dan de manera igualitaria. Por ejemplo, si tenemos una computadora nueva y queremos conectarla a internet, no vamos a contratar el servicio de internet de alguna compañía elegida al azar, sino que buscaremos la compañía que ofrezca el mejor servicio y al mejor precio, y probablemente será esta compañía la que tenga más clientes. En una escuela los varones no buscan a su pareja al azar, sino que buscarán salir con la chica más bonita, o tal vez con la más inteligente, y será esta muchacha la que tenga más pretendientes. Por esta razón, Barabási inventó el concepto de *enlace preferencial* en el cual los nuevos nodos que se añaden a la red se conectarán preferentemente con los nodos ya existentes que tengan el mayor número de conexiones. Intuitivamente podemos pensar que el enlace preferencial consiste en que uno siempre trata de estar conectado con los nodos más “populares”, es decir, con los nodos de mayor conectividad.

Para incorporar este comportamiento, Barabási sugirió que la probabilidad de enlace $\Pi(k, t)$ debe tomar la forma

$$\Pi(k, t) = \left(\sum_{n=0}^{N(t)} k_n \right)^{-1} k \quad (15)$$

donde k_n es la conectividad del n -ésimo nodo ya existente al tiempo t . El factor $\left(\sum_{n=0}^{N(t)} k_n \right)^{-1}$ es simplemente para garantizar que la probabilidad $\Pi(k, t)$ esté normalizada. Al hacer que $\Pi(k, t)$ sea proporcional a k , como lo propuso Barabási, tenemos enlace preferencial, ya que de esta forma, entre más grande sea la conectividad k de un nodo, mayor será la probabilidad de conectarse con él. Como en cada paso de tiempo a nadamos un nuevo nodo con una conexión, comenzando con cero conexiones al tiempo $t = 0$, entonces para cualquier tiempo $t > 0$ se tiene

$$\sum_{n=0}^{N(t)} k_n = 2t,$$

y por lo tanto

$$\Pi(k, t) = \frac{k}{2t} \quad (16)$$

Substituyendo este resultado en la Eq. (12) obtenemos

$$(t+2)P(k, t+1) - \delta_{k,q} = \frac{t+1}{2t} \left\{ (k-1)P(k-1, t) + [2t-k]P(k, t) \right\}$$

En el límite $t \rightarrow \infty$ el sistema alcanza el estado estacionario. Por lo tanto, la distribución de conexiones estacionaria $P(k)$ se obtiene de la ecuación anterior tomando el límite $t \rightarrow \infty$, lo que conduce a

$$P(k) + \frac{1}{2} [kP(k) - (k-1)P(k-1)] = \delta_{k,1} \quad (17)$$

Como todos los nodos tienen por lo menos una conexión, para resolver la ecuación (17) utilizamos la condición inicial $P(0) = 0$, lo cual nos da la solución

$$P(k) = \frac{4}{k(k+1)(k+2)} \quad (18)$$

Aun cuando esta no es una ley de potencias “perfecta” (ver la Fig. 6), para valores grandes de k la distribución $P(k)$ se comporta como $P(k) \sim k^{-3}$. Por un lado, este es un resultado muy interesante: el enlace preferencial genera distribuciones de conectividades con colas libres de escala. Son estas colas largas las responsables de que existan elementos altamente conectados. Por otro lado, este resultado también es desalentador ya que el proceso de enlace preferencial *siempre* da el exponente $\gamma = 3$, el cual no se encuentra frecuentemente en la naturaleza

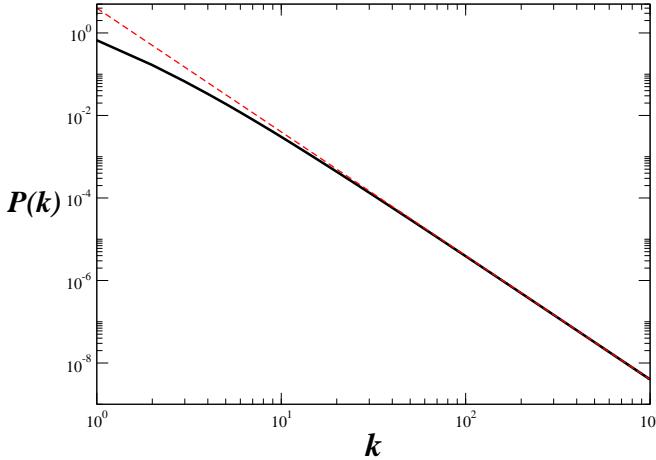


Figura 6. Gráfica log-log de la distribución de conectividades $P(k)$ dada en la ecuación (18). La línea punteada roja es la gráfica de la ley de potencias “perfecta” $f(k) = 4k^{-3}$. Nota que $P(k) \approx 4k^{-3}$ para valores grandes de k .

(ver los exponentes en el Cuadro II). En otras palabras, aun cuando el método de enlace preferencial nos da *una* ley de potencias, es incapaz de reproducir la gran variedad de exponentes encontrados en la naturaleza. Por esta razón, se han propuesto diferentes formas de la función $\Pi(k, t)$ que corresponden a diferentes tipos de enlace preferencial. Por ejemplo, Krapivsky, Redner, y Leyvraz propusieron una función de enlace $\Pi(k, t)$ *no lineal* (ver la referencia [6]), de la siguiente forma

$$\Pi(k, t) = \left(\sum_{n=0}^{N(t)} k_n^\alpha \right)^{-1} k^\alpha \quad (19)$$

donde α es un exponente arbitrario. Desafortunadamente, sólo para $\alpha = 1$ se tiene que la distribución $P(k)$ tiene una cola libre de escala. El trabajo de Krapivsky, Redner, y Leyvraz mostró que la naturaleza libre de escala de la red queda destruida cuando el enlace preferencial obedece una regla no lineal como la dada en la Eq. (19) cuando $\alpha \neq 1$. Es importante mencionar que todavía no tenemos modelos de crecimiento de redes que generen todos los exponentes listados en el Cuadro II.

IX. ¿NODOS QUE SE HACEN VIEJOS?

Otro problema con el modelo de enlace preferencial propuesto por Barabási es que predice que los nodos más viejos, es decir lo que se a nadieron primero a la red, son los que eventualmente adquirirán el mayor número de conexiones. En otras palabras, en una red libre de escala podríamos identificar fácilmente a los nodos más viejos: son aquellos altamente conectados. Y entre más viejo sea un nodo, mayor será su conectividad.

Para ver que esto efectivamente es un resultado del modelo de enlace preferencial, consideremos un nodo par-

ticular v_n de la red. Su conectividad k_n va a cambiar a una tasa que es proporcional a la probabilidad $\Pi(k_n, t)$ de que este nodo adquiera más conexiones. Es decir,

$$\frac{dk_n}{dt} = \Pi(k_n, t)$$

Para el caso particular en el que $\Pi(k, t)$ está dada como en la Eq. (16), tenemos

$$\frac{dk_n}{dt} = \frac{k_n}{2t}$$

Como el nodo v_n nació al tiempo $t = n$ con una conexión, la condición inicial para la ecuación anterior es $k_n(n) = 1$, lo cual conduce a la solución

$$k_n(t) = \left(\frac{t}{n} \right)^{1/2} \quad (20)$$

De la ecuación anterior, es claro que a cualquier tiempo t se cumple la siguiente desigualdad

$$k_1(t) > k_2(t) > k_3(t) > \dots > k_n(t) \quad (21)$$

lo cual significa que los nodos que nacieron primero tendrán, en promedio, conectividades mayores que los nodos que llegaron después.

Sin embargo, este comportamiento predicho por el modelo del enlace preferencial, en el que los nodos más viejos son los más conectados, no siempre se observa en la naturaleza. Por ejemplo, Lada A. Adamic y Bernardo Huberman estudiaron una red que consistía en 260,000 páginas www [7], considerando que dos páginas estaban conectadas si una contenía un hipervínculo a la otra. Adamic y Huberman encontraron que, aún cuando esta red tiene topología libre de escala con un exponente $\gamma \sim 2$ (ver la Fig. 7(a)), no existe correlación entre la conectividad de los nodos y su edad. La Fig. 7(b) muestra la gráfica de la conectividad de los nodos como función de su edad. Como puede observarse, estas dos cantidades no parecen estar relacionadas.

Probablemente el contra ejemplo más contundente que contradice los resultados de las Eqs. (20) y (21) es Google, el robot buscador de la red WWW. Recordemos que Altavista y Yahoo ya existían y estaban bien establecidos en 1998, a no en que apareció Google. Rápidamente Google tomó la delantera, convirtiéndose en el robot buscador más popular en el mundo informático. ¿A qué se debió este éxito repentino de Google? Probablemente a que estaba mejor diseñado, era más rápido y más eficiente que sus competidores Altavista y Yahoo. En otras palabras, Google nació con características que lo hacían un robot “mejor adaptado” que sus competidores.

Esta observación hizo que Barabási propusiera el concepto de *adaptabilidad* en el contexto de las redes complejas. Así, cada uno de los nodos v_n , además de tener una conectividad k_n , también tenía un parámetro de adaptabilidad asociado w_n . Este parámetro es una medida de qué tan bien estaba adaptado el nodo v_n a su entorno: entre más grande es el valor de w_n , mayor es el

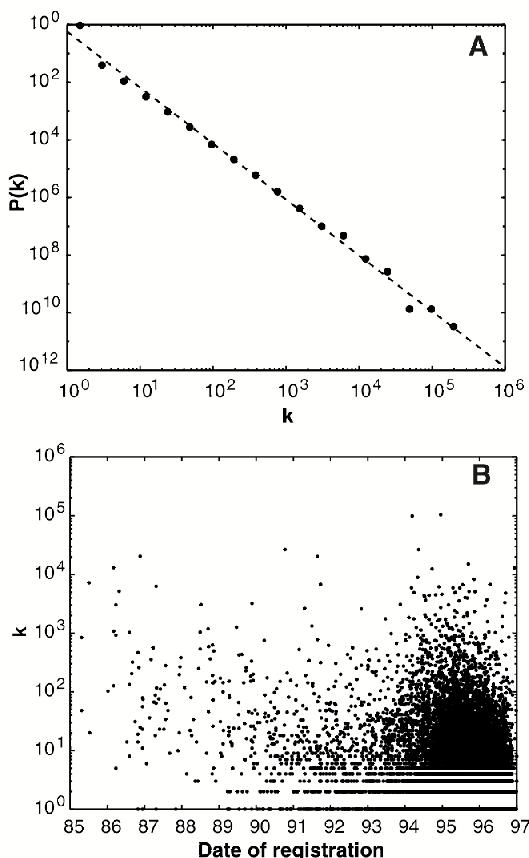


Figura 7. (A) Gráfica log-log de la distribución de conectividades $P(k)$ para una red WWW consistente en 260,000 páginas. (B) Conectividad k de los nodos de la red como función de su edad. Las figuras fueron tomadas (sin permiso) de la referencia [7].

grado de adaptabilidad de v_n . Específicamente, lo que Barabási hizo fue proponer una función de enlace preferencial $\Pi(k, w, t)$ que, además de ser proporcional a la conectividad k de los nodos ya existentes, también es proporcional a su adaptabilidad w (ver la referencia [8]). De esta forma, la probabilidad de que un nuevo nodo a nadado a la red se conecte con el nodo v_n ya existente, cuya conectividad es k_n y cuya adaptabilidad es w_n , queda dada por

$$\Pi(k_n, w_n, t) = Cw_n k_n$$

donde C es una constante de normalización. Las adaptabilidades w_n asociadas a los nodos de la red son variables aleatorias que se escogen de una distribución $A(w)$. Cada nodo nace con su propia adaptabilidad, la cual no cambia en el tiempo.

Un resultado sorprendente del modelo de enlace preferencial con adaptabilidad es que puede ser mapeado exactamente a un gas de Bose-Einstein. En este mapeo cada nodo de la red representa un nivel energético del gas, mientras que las conectividades de los nodos representan los números de ocupación de los respectivos nive-

les de energía. El mapeo entre el modelo de red y el gas de Bose-Einstein es más que una simple metáfora. Es un mapeo matemático preciso y, bajo ciertas condiciones, la red en crecimiento puede presentar “condensación de Bose-Einstein”, lo cual significa que repentinamente un nodo puede adquirir la mayoría de las conexiones de la red, independientemente de su edad (como sucedió con Google). Los que estén interesados podrán encontrar los detalles de este trabajo en la referencia [8]. Hasta donde yo sé, no existe acuerdo cuantitativo entre los resultados del modelo de enlace preferencial con adaptabilidad, y los datos experimentales que se tienen de redes reales. Sin embargo, la condensación de Bose-Einstein predicha en este modelo podría ser el primer paso para entender mejor el “fenómeno Google”.

X. LA ISLA GIGANTE

Regresemos al modelo de los botones descrito en la sección V. Inicialmente todos los botones están sobre la mesa desconectados, pero conforme vamos hilvanando parejas de botones, se comienzan a formar islas de botones conectados. Al principio las islas son pequeñas, pero al ir a nadando más y más conexiones entre los botones, las islas crecen y se conectan entre sí. Eventualmente se formará una isla gigante, es decir, una isla que es mucho más grande que todas las demás.

El tamaño de la isla gigante es importante en el estudio de la propagación de epidemias en una sociedad, en donde en lugar de tener botones hilvanados, tenemos personas que se contagian unas a otras. Las islas consisten en los grupos de personas que están infectadas, y la enfermedad se convierte en una epidemia cuando se forma una isla gigante que abarca a la mayoría de la sociedad.

La pregunta importante es: dado un conjunto de elementos, ¿cuántos enlaces (contagios) tienen que establecerse para que se forme la isla gigante? Esta pregunta fue contestada por Erdős y Rényi, quienes fueron los primeros en mostrar la existencia de una transición de fase en la teoría de redes. Dicha transición de fase consiste precisamente en la formación de la isla gigante. Aunque su trabajo original lo llevaron a cabo para redes con topología de Poisson, es fácil generalizar el análisis que hicieron para extenderlo a redes con topologías arbitrarias.

Para calcular el tamaño de la isla gigante utilizaremos un razonamiento de consistencia. Sea q la probabilidad de que un nodo v_n escogido aleatoriamente *no pertenezca* a la isla gigante. Supongamos que v_n tiene k vecinos. Claramente, v_n no pertenece a la isla gigante si y sólo si ninguno de sus k vecinos tampoco pertenece a la isla gigante. Por lo tanto, la probabilidad q debe satisfacer la ecuación de consistencia

$$q = \sum_k P(k)q^k \quad (22)$$

El lado izquierdo de esta ecuación es simplemente la probabilidad q de que v_n no pertenezca a la isla gigante. El

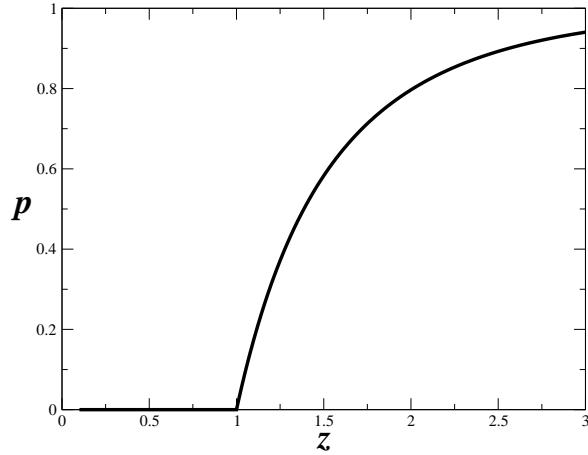


Figura 8. Probabilidad p de pertenecer a la isla gigante como función de la conectividad promedio z de la red, pare el caso de la topología de Poisson.

lado derecho es la probabilidad $P(k)$ de que v_n tenga k vecinos, multiplicada por la probabilidad q^k de que ninguno de estos k vecinos pertenezca a la isla gigante. La suma sobre k toma en cuenta todos los posibles vecinos que v_n pudiera llegar a tener.

En el caso en que $P(k)$ es una distribución de Poisson, la Eq. (22) se transforma en

$$q = \sum_{k=0}^{\infty} e^{-z} \frac{(zq)^k}{k!} = e^{z(q-1)}$$

Denotando como $p = 1 - q$ la probabilidad de que un nodo arbitrario *si* pertenezca a la isla gigante, obtenemos

$$p = 1 - e^{-zp} \quad (23)$$

La ecuación anterior es trascendental y no puede resolverse de forma analítica. Sin embargo, es fácil resolverla de forma numérica y encontrar p para cada valor de z . La Fig. 8 muestra p como función de z . Notemos que para $z < 1$ la probabilidad de pertenecer a la isla gigante es cero. En otras palabras, no hay isla gigante para $z < 1$. Conforme aumentamos el valor de z (lo cual es equivalente a a nadir más conexiones a la red), la isla gigante comienza a formarse y justo en $z = 1$ aparece. Este es un resultado interesante: se necesita, en promedio, sólo una conexión por nodo para que aparezca la isla gigante. Si seguimos aumentando el valor de z , el tamaño de la isla gigante aumenta más y más, como se puede observar en la Fig. 8, la cual muestra el comportamiento característico de una transición de fase de segundo orden que ocurre en $z = 1$.

Consideraremos ahora el caso de las redes con topología libre de escala, para las que la distribución $P(k)$ toma la forma

$$P(k) = \frac{1}{\zeta(\gamma)} k^{-\gamma}$$

donde $\zeta(\gamma) = \sum_{k=1}^{\infty} k^{-\gamma}$ es la función zeta de Riemann la cual, en este caso simplemente juega el papel de la constante de normalización. Para esta distribución la ecuación de consistencia Eq. (22) se convierte en

$$\begin{aligned} q &= \frac{1}{\zeta(\gamma)} \sum_{k=1}^{\infty} \frac{q^k}{k^{\gamma}} \\ &= \frac{1}{\zeta(\gamma)} \text{Li}_{\gamma}(q) \end{aligned}$$

donde $\text{Li}_{\gamma}(q) = \sum_{k=1}^{\infty} k^{-\gamma} q^k$ es la función poli-logarítmica de orden γ . En términos de la probabilidad $p = 1 - q$, la ecuación anterior queda como

$$p = 1 - \frac{1}{\zeta(\gamma)} \text{Li}_{\gamma}(1 - p)$$

Como $\text{Li}_{\gamma}(0) = 0$, es claro $p = 1$ siempre es solución de la ecuación anterior. Por lo tanto, en una red libre de escala la componente gigante siempre existe y la probabilidad de pertenecer a ella es $p = 1$. Este interesante resultado es una consecuencia de la existencia de los elementos altamente conectados en redes libres de escala. Dichos elementos mantienen a toda la red conectada, impidiendo que se fracture en pequeñas islas.

El resultado anterior puede tener implicaciones importantes en la propagación de epidemias en una sociedad, en la que podemos considerar que la isla gigante es el conjunto de personas que se han contagiado de alguna enfermedad. Los resultados presentados en esta sección muestran que en redes con topología de Poisson siempre podemos vacunar a un número suficiente de personas para garantizar que la conectividad promedio z de la red de individuos infectados se mantenga inferior a 1. En tal caso podemos detener la epidemia, ya que no habrá isla gigante, sólo pequeñas islas de personas infectadas.

Sin embargo, las redes sociales no tienen topología de Poisson, sino que exhiben topologías libres de escala. Para tales topologías siempre existe una isla gigante, y la probabilidad de pertenecer a ella es $p = 1$. Por lo tanto, en redes libres de escala las enfermedades tarde o temprano se propagan a todas las personas que no hayan sido vacunadas. Es decir, la única forma de detener la propagación de una epidemia en redes libres de escala es vacunando a *todas* las personas de la sociedad. Los efectos aterradores del resultado anterior pueden observarse claramente en África, donde comunidades enteras han desaparecido a causa del SIDA.

Más información sobre la propagación de epidemias en redes sociales puede encontrarse en las referencias [9, 10].

XI. CRITICALIDAD, ROBUSTEZ Y EVOLUCIÓN EN REDES GENÉTICAS

“Uno se pregunta [...] cómo es posible que los organismos complejos siquiera evolucionen. Parecen tener tantos genes, tantos efectos múltiples o pleitrópicos de cada gen, tantas posibilidades de que hayan mutaciones letales en el desarrollo embrionario, y todo tipo de problemas debido a su largo su desarrollo”.

J.T. Bonner. *The evolution of complexity by means of natural selection.* (1988).

Hasta ahora hemos visto las propiedades estructurales o topológicas de las redes complejas. Estas propiedades nos dicen cómo están conectados los elementos de la red y su estudio nos muestra de qué forma procesos aleatorios y aparentemente independientes pueden dar lugar a estructuras complejas e inesperadas (como la topología libre de escala). No obstante, sabemos que una vez que los elementos de una red están conectados, éstos pueden comenzar a interactuar generando procesos dinámicos que se propagan a través de toda la red. Consideremos, por ejemplo, las neuronas en el cerebro, las cuales están conectadas unas con otras por medio de axones y dendritas. A través de estas conexiones las neuronas interaccionan enviándose impulsos eléctricos que se propagan por todo el cerebro y que generan patrones de actividad eléctrica en las diferentes capas cerebrales responsables de la visión, el olfato, el tacto, etc. Son precisamente estos patrones de actividad eléctrica neuronal, asociados con nuestros sentidos y con nuestra conciencia, lo que llamamos “procesos dinámicos” en el cerebro.

Otro ejemplo muy interesante de procesos dinámicos en redes complejas lo encontramos en las redes de regulación genética que determinan las características fenotípicas de los organismos vivos. Este es precisamente el tema de este capítulo, donde veremos uno de los modelos dinámicos de red genética más interesantes que se han inventado, y lo utilizaremos para tratar de entender cómo se lleva a cabo la evolución de los seres vivos. Más específicamente, queremos entender qué es lo que está evolucionando. Déjenme ser más explícito. La teoría de la evolución consta de dos partes. Por un lado está la *evidencia irrefutable*, acumulada en el registro fósil por arqueólogos y antropólogos de todo el planeta, de que las especies de organismos vivos han ido cambiando a lo largo de millones de años. Por lo tanto, el hecho de que la evolución ha ocurrido es incuestionable con toda la evidencia que tenemos ahora. Por otro lado, la teoría de la evolución también trata de explicar los *mecanismos* por medio de los cuales las especies han ido evolucionando. Tales mecanismos aún no están completamente comprendidos y todavía existe mucha controversia respecto a cómo se ha dado la evolución de los seres vivos. Un mecanismo muy importante de evolución lo propuso Darwin, y es la famosa *selección natural*, la cual se ha popularizado de manera trivial diciendo que la selección natural consiste en “*la sobrevivencia del más apto*”. Esta formulación de lo que es la selección natural está mal por varias razones, pero la más

importante es que genera un razonamiento circular que no conduce a ninguna parte. Porque, ¿cómo se define al “más apto”? Pues precisamente como el que sobrevive. Pero entonces “*la sobrevivencia del más apto*” significaría “*la sobrevivencia del que sobrevive*”, lo cual es una definición circular sin ningún valor científico. Como veremos en este capítulo, entender la evolución de los organismos vivos es mucho más complicado que lo que trae consigo la supuesta sobrevivencia del más apto. Lo que tenemos que entender es cuáles son las características que se seleccionan y cómo se van transformando. Enfocaremos nuestra atención en la evolución de las redes genéticas que determinan los fenotípos de los organismos. A primera vista, podría parecer que esto es un trabajo inacabable, en vista de la gran diversidad de formas de vida que existen en nuestro planeta: plantas, insectos, reptiles, mamíferos; y dentro de cada reino hay cientos de miles o millones de especies distintas. Sin embargo, hay dos propiedades que todos estos organismos tienen en común: han evolucionado en entornos cambiantes, lo cual los ha hecho al mismo tiempo *robustos* y *adaptables*. Comenzaremos entonces definiendo estas dos propiedades con más detenimiento.

La “robustez” y la “evolucionabilidad” (la capacidad de evolucionar) son dos propiedades centrales de los organismos vivos. Sin embargo, siempre es problemático utilizar estas palabras juntas en el mismo título, o incluso en el mismo párrafo, a menos que sea para decir que son propiedades (aparentemente) antagónicas. Por un lado, los organismos vivos son robustos porque pueden mantener su forma y funcionamiento ante la presencia de una gran variedad de perturbaciones, que van desde cambios físicos o químicos transitorios en el entorno, hasta mutaciones permanentes en su material genético. Por lo tanto, la idea de robustez se asocia siempre con la invarianza del fenotípico ante perturbaciones internas (mutaciones genéticas) o externas (cambios ambientales). Por otro lado, los organismos vivos también evolucionan, es decir, cambian su fenotipo adquiriendo nuevas formas y/o funciones, como resultado de las mutaciones genéticas y de las presiones selectivas impuestas por los cambios ambientales. Tanto la robustez como la evolucionabilidad se han observado en diferentes niveles de organización biológica, desde la regulación genética hasta el comportamiento dinámico de ecosistemas enteros. Sin embargo, a pesar de la importancia central de estas dos propiedades para el entendimiento del funcionamiento y evolución de los sistemas biológicos, no es claro aún cuáles son los mecanismos estructurales y dinámicos que generan organismos complejos que son robustos, es decir que no cambian ante perturbaciones, pero que al mismo tiempo son evolucionables, es decir, que tienen la capacidad de eventualmente cambiar y adaptarse como resultado de las mismas perturbaciones.

Estamos tan acostumbrados a que los organismos vivos sean al mismo tiempo robustos y evolucionables, que rara vez consideramos que la existencia simultánea de estas dos propiedades en un mismo sistema dinámico sea un problema. Sin embargo, si lo contextualizamos en el campo de la informática, o de la ingeniería mecánica, por

ejemplo, rápidamente veremos que no es sencillo diseñar ni programas computacionales ni máquinas que sean robustas y evolucionables simultáneamente, es decir que mantengan su funcionamiento ante una gran cantidad de “errores” de los usuarios, y que al mismo tiempo tengan la capacidad de adquirir nuevas funciones (además de las ya existentes) que les permitan adaptarse a dichos errores.

Es importante, por lo tanto, entender cómo la robustez y la evolucionabilidad, dos propiedades aparentemente antagónicas, pueden coexistir en los organismos vivos. Notemos que el problema no consiste en describir mecanismos para la existencia de robustez, o de evolucionabilidad, por separado. Es relativamente sencillo hacer máquinas robustas ante fallas, por ejemplo, por medio de la redundancia, que consiste en duplicar o triplicar las partes que se espera que fallen con mayor frecuencia o que son fundamentales para el funcionamiento de la máquina. Así, si queremos que el sistema de frenos ABS en un coche nunca falle, ponemos dos o tres circuitos controladores idénticos, de tal forma que si uno falla, hay otros para reemplazarlo. Análogamente, no es complicado diseñar sistemas que cambien y se adapten a nuevos retos ambientales. Una red neuronal informática entrenada para reconocer imágenes de botellas se puede volver a entrenar para reconocer imágenes de edificios, por ejemplo.

Así, robustez y evolucionabilidad se pueden tener por separado. Lo que es difícil es entender cómo es que ambas propiedades se manifiestan en un mismo sistema simultáneamente. Un sistema diseñado para ser robusto no cambia, y uno que cambia no es robusto. Sin embargo, los organismos vivos son robustos y sin embargo evolucionan. Como veremos, es la criticalidad dinámica de las redes genéticas la que hace posible esta coexistencia. Estudios teóricos llevados a cabo durante años sobre las propiedades dinámicas de modelos de redes genéticas han mostrado que dichas redes pueden operar en tres fases dinámicas distintas, que se han denominado fase ordenada, fase crítica y fase caótica. En la fase ordenada el sistema es muy robusto a las perturbaciones, pero es incapaz de cambiar. Por otro lado, en la fase caótica incluso perturbaciones muy pequeñas alteran el comportamiento de todo el sistema, es cual es por tanto muy inestable ante perturbaciones. La fase crítica es la que separa la fase ordenada de la fase caótica. Es precisamente en la fase crítica donde se encuentra el balance entre robustez y evolucionabilidad necesario para describir el funcionamiento de los organismos vivos. En la fase crítica el sistema es robusto ante perturbaciones, pero no es tan robusto como para estar “congelado” evolutivamente. Estas observaciones llevaron a Stuart Kauffman a proponer la hipótesis de “la vida al borde del caos” hace casi 40 años. Desde entonces, no había existido evidencia experimental que soportara esta hipótesis. En este trabajo describiremos la importancia de la criticalidad dinámica a nivel genético para la evolución de los organismos y mostraremos evidencia experimental que confirma la

existencia de dicha criticalidad.

A. Robustez y evolucionabilidad

No existen definiciones precisas para robustez y evolucionabilidad universalmente aceptadas. Estos términos se definen dependiendo del contexto y del nivel de organización bajo consideración. Por lo tanto, antes de comenzar nuestro estudio, es necesario definir qué entendemos por robustez y evolucionabilidad en el contexto de las redes genéticas. A continuación presentamos lo que parece ser el consenso dentro de la comunidad de Biología de Sistemas:

Definición 2 *Robustez es la invarianza de los fenotipos ante la presencia de perturbaciones internas o externas al organismo.*

Definición 3 *Un sistema biológico se dice que es evolucionable si puede adquirir fenotipos nuevos a través del cambio genético (mutaciones), fenotipos que pueden ayudar al organismo a sobrevivir y reproducirse.*

Para que estas definiciones sean útiles desde un punto de vista cuantitativo, es necesario también definir operacionalmente fenotipo y perturbación. En cuanto a la definición de perturbación no hay problema. Es fácil definirla como la tasa de mutación en el ADN, o como la intensidad del ruido en el nivel de expresión genética. Nosotros en este trabajo utilizaremos una perturbación muy particular, que es la duplicación y divergencia genética. Este proceso consiste en que durante la replicación celular frecuentemente se cometan “errores” en la replicación del ADN, uno de los cuales consiste en que algunos genes se copian por duplicado. El genoma final del organismo tiene entonces dos copias idénticas del mismo gen (o del mismo grupo de genes), lo cual provee cierta redundancia, y por lo tanto cierta robustez, al genoma del organismo, ya que si alguna de las dos copias del gen sufre mutaciones, aún está la otra copia para llevar a cabo la función celular correspondiente. De hecho, rápidamente después de la duplicación genética, una de las dos copias comienza a mutar, lo cual hace que el gen original y el gen duplicado diverjan en funcionalidad.

Susumu Ohno fue uno de los primeros en percibirse de la importancia de la duplicación y divergencia genéticas en la evolución, ya que este mecanismo constituye una fuente importante de material para la innovación genética. Se estima que al menos el 50 % de los genes en células procariotas, y más del 90 % en células eucariotas, son producto de la duplicación y divergencia genéticas. Por lo tanto, la “perturbación” que estudiaremos en este trabajo consistirá precisamente en duplicar y mutar genes en una determinada red genética, y ver qué consecuencias tiene esto en el fenotipo. Esto nos lleva al segundo punto a definir. Es decir, para saber si el fenotipo es robusto o no ante duplicaciones genéticas, es necesario definirlo de tal forma que se puedan cuantificar sus cambios.

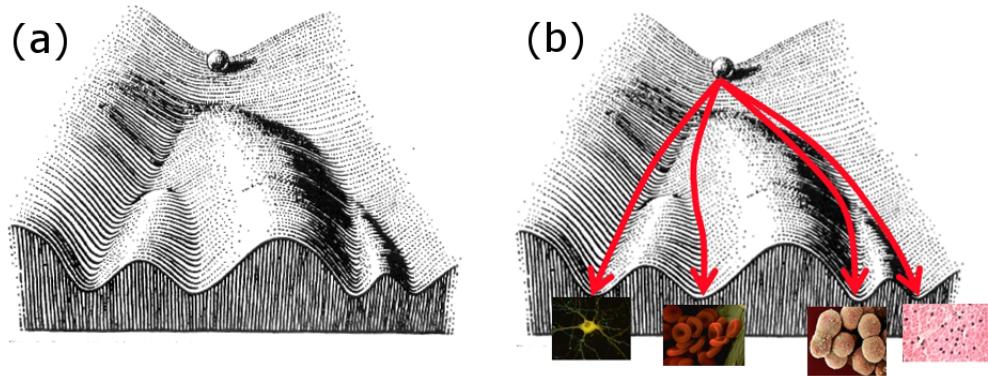


Figura 9. (a) El paisaje epigenético imaginado por Waddington. La célula madre embrionaria corresponde a una canica que está a punto de rodar cuesta abajo por un terreno no uniforme con diversos valles. Dependiendo de la condición inicial (temperatura, presión, acidez, etc.), la canica terminará en alguno de los mínimos estables del terreno. (b) Estos mínimos corresponden a diferentes tipos o destinos celulares, y los valles para llegar a ellos corresponden a las rutas de diferenciación.

Desde hace varias décadas se formuló la hipótesis de que las características fenotípicas de los organismos están determinadas por los atractores dinámicos de la red genética subyacente. El primero en formular esta hipótesis fue Conrad Waddington, al introducir la metáfora de los paisajes epigenéticos. En esta metáfora, la célula madre embrionaria puede considerarse como una pequeña canica que está a punto de rodar cuesta abajo por un terreno con varias colinas y valles (ver la Fig. 9a). Dependiendo de la condición inicial, la canica puede rodar por un valle o por otro, terminando en un sumidero del que no podrá salir. En el contexto de la célula, cada valle representa una ruta de diferenciación, mientras que cada sumidero representa un destino final o estado funcional final de la célula (ver la Fig. 9b). Así, Waddington se imagina al desarrollo embrionario como un proceso dinámico que tiene diversos atractores y diferentes rutas para llegar a ellos.

La metáfora del paisaje epigenético fue formalizada posteriormente por Stuart Kauffman, quien demostró que modelos simples de redes genéticas efectivamente exhiben atractores dinámicos y diferentes rutas para llegar a ellos. Siguiendo a Waddington, Kauffman formuló la hipótesis de que cada atractor dinámico corresponde a un tipo celular o a un destino celular. Por lo tanto, la información del fenotipo celular no está contenida sólo en la secuencia de bases del ADN, sino también en los atractores dinámicos que se generan a partir de la interacción colectiva de todos los genes en el genoma. Por lo tanto, una definición operativa del fenotipo del organismo puede darse en términos del paisaje epigenético, o lo que es equivalente en el formalismo de Kauffman, en términos del paisaje de atractores dinámicos de la red genética. Volveremos a este punto más adelante.

B. El modelo de Kauffman

Probablemente, el modelo de red genética más sencillo que describe adecuadamente los procesos de regulación genética y diferenciación celular es el que propuso Kauffman en 1969. En este modelo uno está interesado en la evolución temporal del estado de expresión de los genes y no de las concentraciones de sus productos. Por lo tanto, cada gen se representa con una variable binaria σ que puede adquirir los valores $\sigma = 1$ si el gen que representa está expresado, y $\sigma = 0$ si dicho gen no está expresado. El genoma de un organismo se representa entonces por un conjunto de N de estas variables $\{\sigma_1, \sigma_2, \dots, \sigma_N\}$, cada una correspondiendo a un gen.

Los genes interactúan unos con otros a través de las proteínas reguladoras (factores de transcripción) que generan. Los factores de transcripción se unen de forma específica a las regiones reguladoras de los genes regulados, activando o reprimiendo su expresión. Esta serie de interacciones, mediadas por los factores de transcripción, da lugar a una red dirigida en la que dos genes están “conectados” si la expresión de uno regula, positiva o negativamente, la expresión del otro. La Fig. 10 muestra la estructura topológica de la red de regulación transcripcional de la bacteria *Escherichia coli*, de acuerdo con la información más reciente en la base de datos del regulonDB.

En el modelo de Kauffman, a cada gen σ_n se le asocia un conjunto de k_n reguladores, $\{\sigma_{n_1}, \sigma_{n_2}, \dots, \sigma_{n_{k_n}}\}$, que son los que determinan el estado de expresión de σ_n . Una vez que se han establecido los reguladores de todos y cada uno de los genes, la dinámica de la red genética queda dada por la ecuación

$$\sigma_n(t + \tau) = F_n(\sigma_{n_1}(t), \sigma_{n_2}(t), \dots, \sigma_{n_{k_n}}(t)) \quad (24)$$

Esta ecuación nos dice que el estado de expresión de un gen σ_n al tiempo $t + \tau$ está determinado úni-

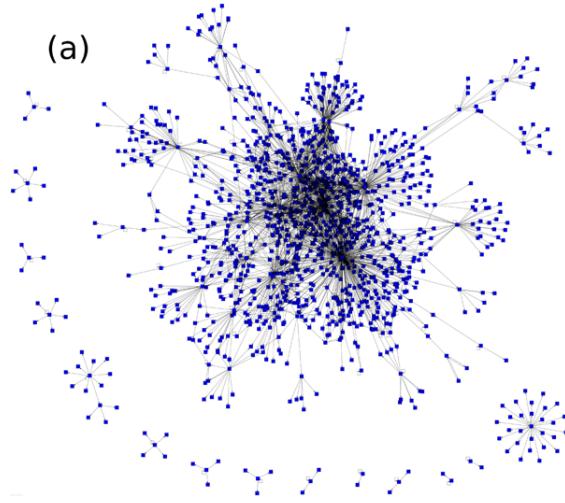


Figura 10. Estructura topológica de la red de regulación transcripcional de *E. coli*. Cada nodo representa un gen, y cada enlace representa una interacción de regulación. Nótese la famosa estructura libre de escala, en la que la gran mayoría de los nodos están poco conectados, mientras que algunos cuantos nodos tienen un gran número de conexiones. Las conexiones son dirigidas porque indican cuáles son los genes reguladores y cuáles son los genes regulados. Sin embargo, la dirección de las flechas no se alcanza a ver en la imagen.

camente por el estado de expresión de sus reguladores $\{\sigma_{n_1}, \sigma_{n_2}, \dots, \sigma_{n_{k_n}}\}$ evaluado en un tiempo anterior t . El parámetro τ es una medida del tiempo típico de respuesta de la expresión genética de los genes. La función booleana F_n determina la interacción entre σ_n y sus reguladores, y se construye tomando en cuenta la naturaleza activadora o represora de los reguladores de σ_n . Al iterar la Eq. (24) en el tiempo cada gen, siguiendo en todo momento su propia función booleana, pasa por una serie de estados prendido/apagado hasta que toda la red entra en un estado de expresión periódico (ver la Fig. 11). Algunos genes alcanzan una expresión constante que ya no cambia en el tiempo, mientras que otros genes se siguen prendiendo y apagando pero de forma periódica. Este estado de expresión periódico es el atractor dinámico de la red al que nos referimos antes. Generalmente, una misma red genética tiene varios atractores dinámicos. Cada atractor dinámico queda identificado de forma única por el conjunto de genes que están prendidos en él. En otras palabras, en atractores distintos se expresan conjuntos de genes diferentes. Pero ésta es precisamente la característica que identifica a los diferentes tipos celulares en un organismo, es decir, *conjuntos distintos de genes se expresan en tipos celulares distintos*. Por tal motivo, Kauffman formuló la hipótesis de que *los atractores dinámicos de la red genética corresponden a los diferentes tipos o destinos celulares observados en el organismo*.

Esta hipótesis ha sido confirmada recientemente con estudios numéricos y experimentales llevados a cabo por diversos grupos de investigación, entre los que destacan el de la Dra. Elena lvarez-Buylla del Instituto de Ecología

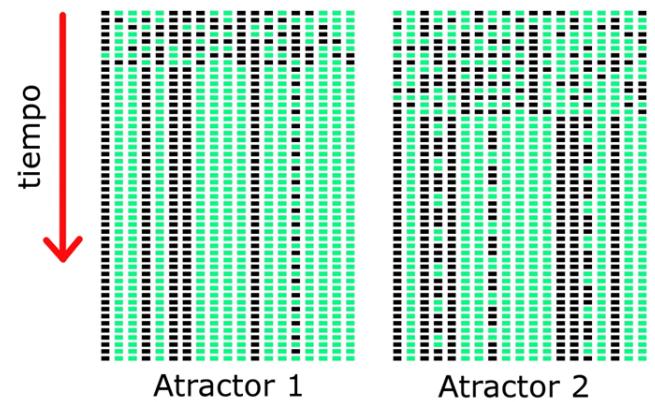


Figura 11. Representación visual de la dinámica en el modelo de Kauffman. Cada cuadrado representa un gen, en color claro si el gen está expresado y en negro si no está expresado. Cada fila horizontal representa el estado de todo el genoma en un instante de tiempo. La primera fila es la condición inicial con la que se comienza la dinámica, la segunda fila es el estado del genoma después de una iteración de la Eq. (24), la tercera fila después de dos iteraciones, y así sucesivamente. Puede observarse que después de un tiempo transitorio el genoma entra en un patrón de expresión periódico, que es el atractor dinámico. Los dos paneles mostrados corresponden a diferentes condiciones iniciales que terminan en diferentes atractores.

de la UNAM, quienes encontraron que los atractores de la red de desarrollo floral de la planta *Arabidopsis thaliana* corresponden a los diferentes patrones de expresión genética observados experimentalmente en los distintos órganos de la planta. Estudios similares llevados a cabo por el grupo de la Dra. Reka Albert de la Universidad de Pensilvania, el grupo del Dr. Sui Huang del Harvard Children's Hospital, y el grupo del Dr. Chao Tang de la Universidad de California en San Francisco, muestran también que los atractores de la red genética de diferentes organismos (*Drosophila melanogaster*, células de tejido endotelial, y *Saccharomyces cerevisiae*) corresponden a los patrones de expresión observados en diferentes tipos o destinos celulares. Por lo tanto, estos trabajos muestran de forma contundente que características fenotípicas importantísimas del organismo están codificadas en los atractores dinámicos de la red genética subyacente.

C. El paisaje de atractores

El paisaje de atractores en el modelo de Kauffman es equivalente al espacio fase de un sistema dinámico. Es una manera de representar gráficamente las diferentes trayectorias que sigue el sistema al moverse de acuerdo con la ecuación que dicta su dinámica, en nuestro caso la Eq. (24). Para ilustrar la construcción del paisaje de atractores en el modelo de Kauffman, supongamos que tenemos una red genética de sólo 4 genes. Existen $\Omega = 2^4 = 16$ estados dinámicos posibles en los que la red se puede encontrar, desde el estado 0000 en el que to-

dos los genes están apagados, hasta el estado 1111 en el que todos los genes están prendidos, pasando por todos los estados intermedios. Todos estos estados posibles de expresión en los que se puede encontrar la red en algún momento se muestran en la Fig. 12a. A lo largo del tiempo, el estado de expresión de cada gen cambia de acuerdo a la Eq. (24), pero por el momento no consideraremos los cambios que sufre cada gen, sino los cambios que sufre la red completa, es decir, consideraremos el efecto neto de todas las funciones booleanas sobre el estado de toda la red. Supongamos como indica la Fig. 12b que, al iniciar la dinámica con la condición inicial 0000, después de un paso de tiempo la red cambia al estado 0101, el cual a su vez cambia al estado 1100 en el siguiente paso de tiempo, el cual se convierte en 0110 que después vuelve a cambiar a 0101. A partir de este momento se vuelve a repetir el ciclo ya que el estado 0101 ya se había encontrado antes. En este ejemplo, la red sigue la trayectoria en el espacio de estados

$$0000 \rightarrow 0101 \rightarrow 0110 \rightarrow 1100 \rightarrow 0101 \rightarrow 0110 \rightarrow 1100 \rightarrow \dots$$

El atractor consiste de los tres estados 0101, 0110 y 1100. Si ahora comenzáramos la dinámica con la condición inicial 1110, se tendría la trayectoria

$$1110 \rightarrow 1011 \rightarrow 0010 \rightarrow 0011 \rightarrow 0011 \rightarrow 0011 \rightarrow \dots$$

En este otro caso el atractor consiste de un sólo estado, 0011. La Fig. 12c muestra el diagrama al que se llegaría si conectamos con flechas todos los posibles estados dinámicos de la red conforme se van sucediendo uno al otro en el tiempo hasta llegar al atractor. Todos los estados que eventualmente convergen al mismo atractor conforman la cuenca de atracción de ese atractor. Como se ve en la Fig. 12c, la misma red puede tener más de un atractor. La Fig. 12d muestra un ejemplo similar pero para una red con 10 genes, la cual tiene entonces $\Omega = 2^{10} = 1024$ estados (una red con N genes tendrá $\Omega = 2^N$ estados posibles).

Nótese como la dinámica generada por la Eq. (24) divide al espacio de estados en los diferentes atractores y sus respectivas cuencas de atracción. Las estructuras que resultan de esta división, como las mostradas en la Fig. 12c y 12d, es lo que se conoce como el paisaje de atractores. Desde el punto de vista biológico, los estados que conforman a los atractores son patrones de expresión genéticas que contienen la información del tipo celular (riñón, hígado, pulmón, neurona, etc.), mientras que las “ramas” que convergen a esos atractores son las diferentes rutas de diferenciación. Es hasta cierto punto sorprendente que un modelo tan sencillo como el propuesto por Kauffman pueda reproducir acertadamente dos de los mecanismos fundamentales de los sistemas biológicos, que es la regulación genética y la diferenciación celular.

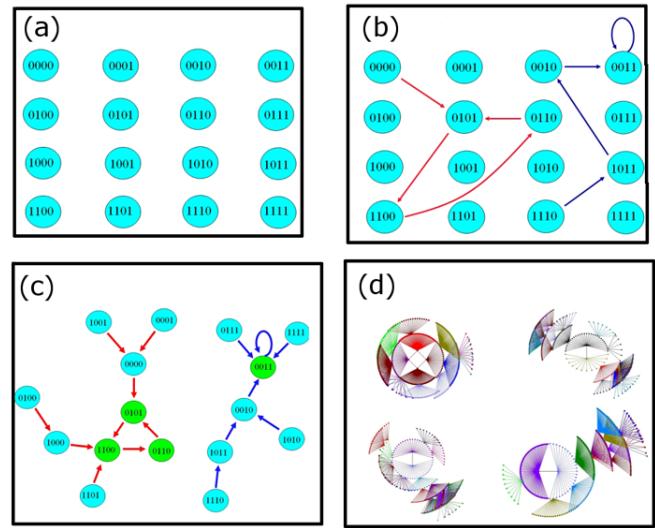


Figura 12. Representación gráfica del paisaje de atractores. (a) Una red con sólo cuatro genes puede estar en cualquiera de los 16 estados posibles listados aquí. (b) La dinámica generada por la Eq. (24) conecta los estados en el tiempo, de tal manera que dos estados están conectados si uno es sucesor del otro. (c) Al conectar todos los posibles estados con sus sucesores, observamos que hay sólo dos atractores: uno que consiste de los tres estados $\{0101, 1100, 0110\}$, y otro que consiste de un solo estado $\{0011\}$. Los atractores se pintan en color verde y las cuencas de atracción en color azul. (d) Un ejemplo similar pero para una red con 10 genes, la cual tiene entonces $\Omega = 2^{10} = 1024$ estados. Es esta caso hubieron cuatro atractores. Las estructuras en forma de abanico reflejan el hecho de que varios estados pueden tener el mismo sucesor.

D. Criticalidad en el modelo de Kauffman

Las redes de Kauffman pueden operar en tres regímenes dinámicos distintos: la fase ordenada, la fase crítica y la fase caótica. La fase crítica es donde se da la transición entre el orden y el caos. Antes de describir en qué consisten estas tres fases, es importante mencionar que existe evidencia experimental y teórica de que muchos sistemas complejos dinámicos, además de las redes genéticas, operan en la fase crítica. Ejemplos típicos son avalanchas de nieve o de arena, precipitación atmosférica, mercados financieros, terremotos, medios granulares, y la actividad neuronal en el cerebro, por mencionar sólo algunos. El interés en los sistemas críticos, es decir los que operan al borde del orden y del caos, radica en que tales sistemas presentan propiedades notables que serían muy difíciles de explicar en ausencia de la criticalidad. Por ejemplo, tales sistemas pueden integrar, procesar y transferir información más rápidamente y de forma más eficiente que sistemas no críticos. O pueden detectar y responder a estímulos cuyas intensidades varían en varios órdenes de magnitud[21]. Estas propiedades de los sistemas críticos son principalmente una consecuencia de las correlaciones de largo alcance que emergen en la fase crítica, generando comportamientos colectivos del sistema como un todo.

Así, la criticalidad le confiere al sistema la propiedad de responder y adaptarse como un todo a las fluctuaciones o perturbaciones que ocurren en entornos rápidamente cambiantes.

En el modelo de Kauffman, la fase dinámica en la que opera la red se caracteriza por el comportamiento temporal de la distancia Hamming $x(t)$ entre dos trayectorias diferentes en el espacio de estados. Supongamos que inicializamos la red en alguna condición en la que algunos genes están prendidos y otros apagados. Llamaremos a esta condición la condición silvestre. Bajo de dinámica dada por la Eq. (24), la red seguirá una trayectoria como la indicada en el panel izquierdo de la Fig. 13, a la que llamaremos trayectoria silvestre (el primer renglón de dicha trayectoria es la condición inicial). Ahora supongamos que inicializamos la red en una condición ligeramente diferente a la anterior, en la que algunos genes tienen un estado de expresión diferente respecto a la condición silvestre. Esto se puede lograr aumentando o disminuyendo artificialmente el nivel de expresión de algunos genes. De cualquier forma, los genes cuya expresión ha sido alterada respecto a la condición silvestre se han marcado con rojo en el panel derecho de la Fig. 13 (los dos genes rojos en el primer renglón). Esta nueva condición “perturbada” generará una trayectoria en la que la perturbación inicial puede cambiar la expresión de otros genes a lo largo del tiempo, propagándose a otras partes de la red. Es decir, la perturbación inicial generará una avalancha de perturbaciones que puede o no propagarse a otros elementos de la red. En el ejemplo mostrado en la Fig. 13, los genes que han sido perturbados respecto a la trayectoria silvestre se han marcado en rojo.

Definición 4 La distancia Hamming $x(t)$ es la fracción de genes perturbados al tiempo t en toda la red, es decir, es la fracción de genes en el genoma que han cambiado su nivel de expresión (respecto a la trayectoria silvestre) debido a la perturbación inicial.

De ahora en adelante nos referiremos a $x(t)$ como el tamaño (fraccional) de la avalancha de perturbaciones. La evolución temporal de $x(t)$ está dada por un mapeo dinámico $x(t + \tau) = M(x(t))$ el cual relaciona el tamaño de la avalancha de perturbaciones en dos pasos de tiempo consecutivos. Dada una perturbación inicial $x(0)$ al tiempo $t = 0$, iteraciones sucesivas de este mapeo eventualmente convergerán a un valor final $x_\infty = \lim_{t \rightarrow \infty} x(t)$, que es el valor final del tamaño de la avalancha. Es precisamente este valor final x_∞ el que caracteriza la fase dinámica en la que opera la red. Si $x_\infty = 0$ entonces la avalancha de perturbaciones eventualmente desaparece, y por lo tanto la expresión del genoma regresa a su estado silvestre no perturbado. Esta es la fase ordenada en la cual las perturbaciones transitorias no tienen ningún efecto en el estado de expresión estable de la red. Por el contrario, si $x_\infty > 0$ entonces la perturbación inicial de unos cuantos genes se propaga a toda la red, alterando eventualmente la expresión de una fracción considerable de genes. Esta es la fase caótica en la que la expresión

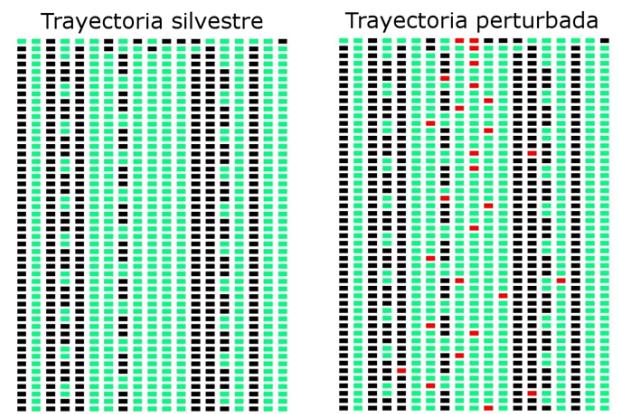


Figura 13. La distancia Hamming es el tamaño fraccional de la avalancha de perturbaciones producida por una perturbación inicial. El panel izquierdo muestra la “trayectoria silvestre” respecto de la cual se medirán las perturbaciones. En el panel derecho se muestra la trayectoria perturbada. Los genes en rojo son aquellos que han cambiado su nivel de expresión, respecto de la trayectoria silvestre, debido a la perturbación inicial (que consiste en los dos cuadritos rojos en el primer renglón).

de todo el genoma cambia erráticamente ante perturbaciones. Finalmente, en la fase crítica, el comportamiento típico de la avalancha de perturbaciones es que no crece ni decrece, sino que se queda más o menos del mismo tamaño. Así, una pequeña perturbación inicial permanecerá pequeña a lo largo del tiempo (o desaparecerá), pero no se propagará a todo el sistema. En esta fase la dinámica de la red no es extremadamente sensible a perturbaciones, pero tampoco está congelada. El comportamiento de la avalancha de perturbaciones se ilustra gráficamente en la Fig. 14.

En 1986 Y. Pomeau y B. Derrida formularon una teoría de campo medio que demuestra la existencia de la transición de fase entre orden y caos en el modelo de Kauffman. En esta teoría se asume que cada gene tiene en promedio K reguladores, seleccionados aleatoriamente con probabilidad uniforme de entre todos los genes que conforman la red (esta elección da lugar a la topología homogénea aleatoria). Además, se asume que todos los genes son estadísticamente independientes y estadísticamente equivalentes. Con estas suposiciones, se puede demostrar que el mapeo $M(x)$ que determina la dinámica temporal del tamaño de la avalancha de perturbaciones tiene las siguientes propiedades (la demostración matemática de estas propiedades puede encontrarse en la referencia [[11]]):

- $M(x)$ es una función continua monótonamente creciente en el intervalo $x \in [0, 1]$.
- $M(0) = 0$, es decir, si no hay perturbación al tiempo $t = 0$ entonces no habrá perturbación a ningún tiempo posterior.
- $0 < M(1) < 1$, es decir, el mapeo siempre se mantiene acotado en el intervalo $x \in [0, 1]$.

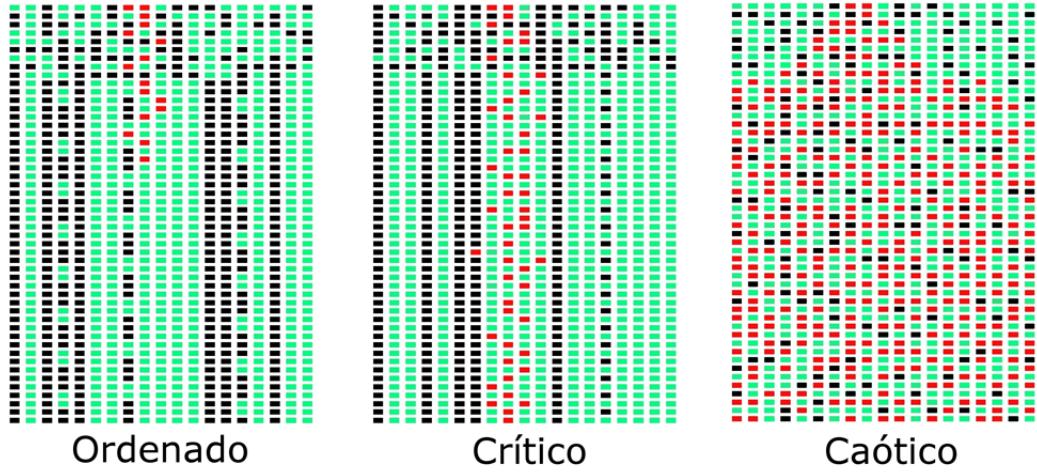


Figura 14. Las tres fases se caracterizan por la propagación de las perturbaciones a través del sistema. En la fase ordenada, una pequeña perturbación inicial (cuadritos rojos en el primer renglón) rápidamente desaparece en el tiempo. En la fase crítica la perturbación inicial no desaparece, pero tampoco crece, se mantiene confinada. En la fase caótica la perturbación se propaga a todo el sistema.

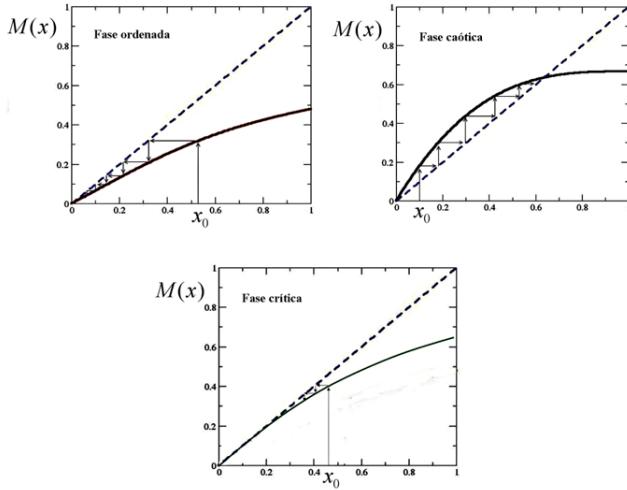


Figura 15. Formas típicas del mapeo $M(x)$ en las tres fases dinámicas. En cada caso las flechas indican la evolución temporal de la perturbación al iterar el mapeo. En la fase ordenada una perturbación inicial x_0 desaparece rápidamente. En la fase caótica la pequeña perturbación inicial x_0 se amplifica. En la fase crítica la perturbación inicial x_0 comienza a disminuir, pero rápidamente queda “atrapada” entre la recta identidad y el mapeo, pareciendo como si no disminuyera ni aumentara.

La forma general de este mapeo se muestra en la Fig. 15 para tres casos ilustrativos. En el primer caso (primer panel de la Fig. 15) el mapeo $M(x)$ va completamente por debajo de la identidad. Por lo tanto, dada una perturbación inicial x_0 , iteraciones sucesivas de este mapeo convergerán rápidamente hacia $x_\infty = 0$ y la perturbación desaparecerá. Esta es la fase ordenada. En el segundo caso (segundo panel de la Fig. 15), el mapeo $M(x)$ co-

mienza en el origen por arriba de la diagonal y la cruza en algún punto $x_\infty < 1$. Por lo tanto, iniciando con una pequeña perturbación inicial x_0 , iteraciones sucesivas de este mapeo amplifican la perturbación hasta que alcanza un tamaño final x_∞ que es más grande que x_0 . Esta es la fase caótica. El tercer caso (tercer panel de la Fig. 15) es cuando el mapeo $M(x)$ es tangente a la diagonal en el origen. En este último caso una pequeña perturbación inicial x_0 comienza disminuyendo, pero debido a que el mapeo es tangente a la diagonal en el origen, tardará muchísimo tiempo en desaparecer. Para todos los tiempos finitos parece que la perturbación inicial x_0 no crece ni decrece en el tiempo y se tendrá entonces que $x_\infty \approx x_0 \approx 0$. Esta es precisamente la característica de la fase crítica.

Las observaciones anteriores muestran que lo que determina la fase dinámica en la cual opera la red es la pendiente del mapeo $M(x)$ en el origen, es decir, en $x = 0$. Debido a que este es un parámetro importante, se le ha dado un nombre: *la susceptibilidad de la red*, y se le denota con el símbolo S . Si $S < 1$ la red está en la fase ordenada (el mapeo comienza por debajo de la identidad); si $S > 1$ la red está en la fase caótica (el mapeo comienza por arriba de la identidad). La fase crítica se obtiene cuando $S = 1$ (el mapeo comienza tangente a la identidad). La teoría de campo medio predice que la susceptibilidad de la red está determinada por la ecuación

$$S = 2p(1 - p)K \quad (25)$$

donde K es el número promedio de reguladores por gen en toda la red, y p es la fracción promedio de frases de regulación activadoras. El parámetro K está relacionado con la topología de la red, mientras que el parámetro p tiene que ver con las funciones booleanas que regulan la actividad de los genes. Así, tanto la topología de la red

genética como la naturaleza de las interacciones entre los genes determinan si la red opera en la fase ordenada, en la crítica o en la caótica.

E. Criticalidad en redes genéticas de organismos reales

La teoría de campo medio desarrollada por Pomeau y Derrida hace suposiciones que evidentemente no son ciertas para los organismos reales. En particular, las suposiciones de que los genes son estadísticamente independientes y estadísticamente equivalentes no se cumplen. Como muestra la Fig. 10, hay genes que están altamente conectados y genes que están muy poco conectados. Por lo tanto, los genes no son equivalentes. Además, los genes altamente conectados, también llamados reguladores globales, regulan la expresión de cientos de otros genes, lo cual introduce correlaciones en la expresión genética de gran parte de la red. La existencia de los reguladores globales hace que la suposición de campo medio sobre la independencia estadística de los genes tampoco sea cierta. Por lo tanto, es importante determinar, en primer lugar, si la transición de fase entre orden y caos que describimos anteriormente, y que predice la teoría de campo medio, sigue siendo válida para redes cuya topología es como la observada en las redes genéticas de organismos reales. La Fig. 16 muestra el valor final del tamaño de la avalancha de perturbaciones, x_∞ , como función de la susceptibilidad S de la red. La línea negra sólida corresponde a lo predicho por la teoría de campo medio, mientras que la línea roja quebrada es el resultado que se obtiene al simular en la computadora la dinámica de Kauffman en redes con topologías “realistas”, es decir, construidas con las mismas propiedades topológicas que las observadas en redes de organismos reales (como la mostrada en la Fig. 10). Puede observarse en la Fig. 16 que la transición de fase predicha por la teoría de campo medio es idéntica a la obtenida para redes con topologías realistas, aún cuando para tales redes las hipótesis del campo medio no se cumplen. Este resultado es bastante inesperado y muestra un ejemplo en el cual la teoría de campo medio conduce a resultados correctos incluso con suposiciones francamente equivocadas.

Pare determinar si las redes genéticas de organismos reales operan en la fase crítica, lo que hicimos fue calcular el mapeo $M(x)$ utilizando los datos experimentales que se tienen hasta el momento de las redes genéticas de 5 organismos: la planta *Arabidopsis thaliana*, la mosca *Drosophila melanogaster*, la bacteria *Escherichia coli*, la bacteria *Bacillus subtilis* y el hongo *Saccharomyces cerevisiae*. En el caso de la planta utilizamos la red de desarrollo floral que determina el surgimiento de los diferentes órganos de la planta (pétales, estambres, carpelos, etc.). Para la mosca utilizamos la red que determina la segmentación polar del embrión en las diferentes partes que eventualmente darán lugar a distintos órganos. Estas dos redes están completamente documentadas y basadas

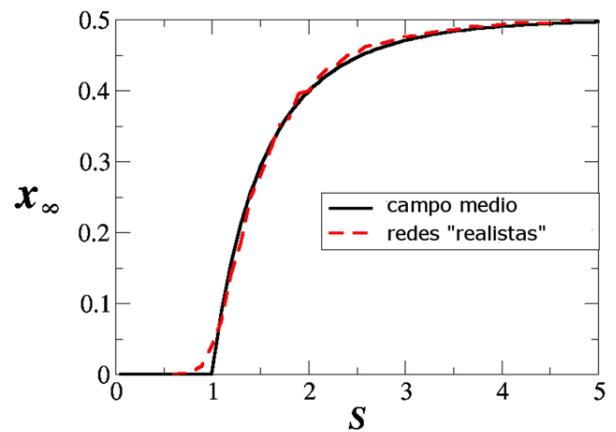


Figura 16. Transición de fase predicha por la teoría de campo medio (curva negra continua) y la obtenida numéricamente para redes con topologías “realistas” (curva roja quebrada). Nótese como la transición de fase es idéntica en ambos casos.

en datos experimentales, tanto la topología como las funciones booleanas reguladoras. No son redes grandes[22], pero son módulos genéticos representativos dentro del genoma encargados de llevar a cabo funciones bien caracterizadas.

Por otro lado, las redes que utilizamos para los organismos unicelulares *E. coli*, *B. subtilis* y *S. cerevisiae* son redes mucho más grandes[23] que intentan describir todas las interacciones transcripcionales en los genomas de esos organismos. Aunque dichas redes todavía no están completas, las partes que ya se tienen en las bases de datos son lo suficientemente grandes como para ser estadísticamente representativas de la estructura de toda la red. El problema es que para estas redes sólo se conoce la topología, pero no se conocen las funciones booleanas. Por lo tanto, en ausencia de este conocimiento lo que hicimos para calcular el mapeo $M(x)$ fue implementar funciones booleanas aleatorias. Sin embargo, las funciones aleatorias que utilizamos fueron construidas utilizando como modelo la expresión genética a gran escala observada en experimentos de microarreglo. Dichos experimentos muestran imágenes “instantáneas” del estado de expresión genética de todos los genes en un organismo (ver la Fig. 17). Recabamos todos los experimentos de microarreglo que pudimos encontrar en las bases de datos públicas (fueron más de 400 en total), y de ellos inferimos el valor del parámetro p que aparece en la Eq. (25), es decir, el valor de la fracción de frases de regulación activadoras en todo el genoma. Con este valor de p construimos aleatoriamente las funciones booleanas que después implementamos en las redes de *E. coli*, *B. subtilis* y *S. cerevisiae*. Por lo tanto, aún cuando las funciones booleanas que utilizamos eran aleatorias con un sesgo dado por p , las construimos de la forma más compatible posible con los datos experimentales existentes.

El mapeo $M(x)$ obtenido para las redes genéticas de los

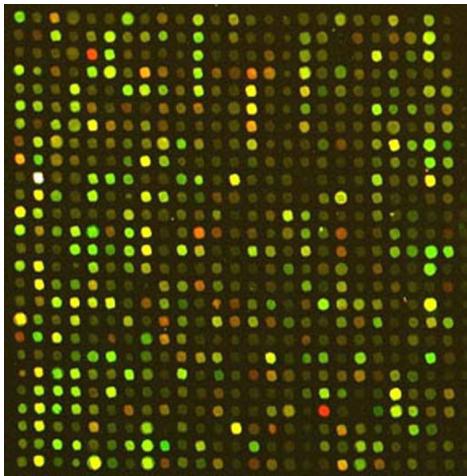


Figura 17. Con la tecnología existente se puede observar, en tiempo real, el estado de expresión de todos los genes en el genoma. En esta imagen se muestra un experimento de microarreglo en el que cada punto brillante indica el estado de expresión de un gen distinto en *E. coli*. Entre más brillante sea el punto, más activo está el gen correspondiente. Utilizando estadística bayesiana y la topología conocida de la red genética, se puede inferir, a partir de estos experimentos, la fracción de frases booleanas activadoras en todo el genoma.

5 organismos mencionados anteriormente se muestra en la Fig. 18. Como puede observarse, en todos los casos el mapeo es consistente con dinámicas críticas. Aún cuando la sensitividad S de la red no es exactamente igual al valor crítico $S = 1$ en ningún caso, está bastante cerca de dicho valor crítico. Esto es particularmente cierto para las redes de *E. coli* y *S. cerevisiae*, que son las más completas que tenemos hasta el momento. Los datos reportados en la Fig. 18 constituyen la evidencia más sólida de criticalidad en la dinámica genética de organismos reales encontrada hasta la fecha.

F. Criticalidad, robustez y evolución

La evidencia que hemos presentado hasta el momento sobre la existencia de criticalidad dinámica en redes genéticas está basada principalmente en el estudio de la respuesta de la red ante perturbaciones transitorias en el estado de expresión de los genes. Es decir, la fase dinámica en la cual opera la red determina si ésta será capaz o no de recuperarse de la sub-expresión o de la sobre-expresión de unos cuantos genes inducida por fluctuaciones en el medio ambiente. Sin embargo, existe una manifestación mucho más fundamental de la criticalidad, y tiene que ver con la transformación del paisaje de atractores ante mutaciones permanentes de la red.

Como mencionamos en la introducción, el principal mecanismo de evolución y crecimiento genómico es la duplicación genética seguida por divergencia. Sin embargo, el paisaje de atractores depende de forma no trivial del

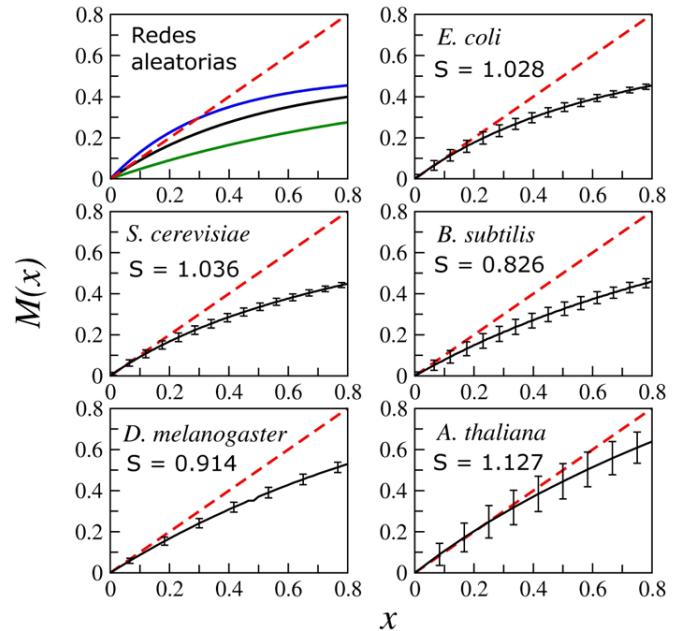


Figura 18. El mapeo $M(x)$ muestra criticalidad dinámica en las redes genéticas de 5 organismos modelo ampliamente estudiados en biología, los cuales representan 4 reinos distintos de la vida. El valor de la susceptibilidad S (la pendiente del mapeo en el origen) se indica. Para facilitar la comparación, el panel superior izquierdo muestra el mapeo $M(x)$ para distintas redes aleatorias operando en las tres fases dinámicas.

número de genes que hay en la red y de todas las interacciones entre ellos. Por lo tanto, duplicar un gen y modificarlo puede alterar fuertemente el paisaje de atractores. No queremos que esto ocurra ya que el paisaje de atractores contiene información de las características fenotípicas más relevantes del organismo. Por lo tanto, es importante determinar hasta qué punto el proceso de duplicación y divergencia genética modifica el paisaje de atractores. Esto nos lleva a la siguiente definición operacional de la robustez y evolucionabilidad genéticas:

Definición 5 Una red de Kauffman con N_a atractores es robusta si todos sus M_a atractores se conservan después de duplicar y mutar un sólo gen. Adicionalmente, diremos que la red es evolucionable si, como consecuencia de la duplicación y divergencia de un gen, nuevos atractores emergen.

Vale la pena enfatizar que el paisaje de atractores depende de las interacciones específicas que han entre todos los genes de la red, es decir, que el paisaje de atractores no es una propiedad particular de un solo gen, o de un grupo reducido de genes, sino que emerge del comportamiento colectivo de toda la red. Por lo tanto, la robustez y evolucionabilidad definidas anteriormente son también propiedades emergentes colectivas asociadas a toda la red, o como lo diría A. Wagner, son propiedades *distribuidas*.

La duplicación genética seguida de divergencia se puede implementar fácilmente en el modelo de Kauffman.

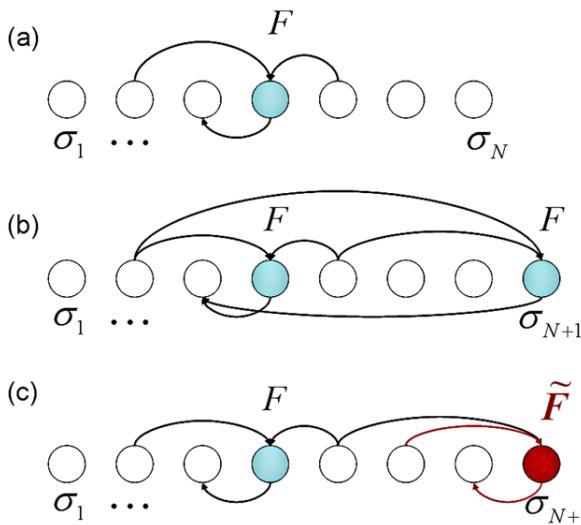


Figura 19. Duplicación y divergencia genética en el modelo de Kauffman. (a) La red original con N genes. Se selecciona un gen particular, el cual se muestra con sus inputs, outputs y función booleana F . (b) El gen seleccionado se duplica, lo cual da lugar a una red con $N + 1$ genes. El gen duplicado es idéntico al gen padre en el sentido de que tiene los mismos inputs, outputs y la misma función booleana F . (c) Algunas características del gen duplicado se cambian, como sus inputs, sus outputs o su función booleanas, lo cual hace que el gen duplicado ya no sea igual al gen padre. La red mutada no sólo tiene un gen adicional, sino que dicho gen puede ser funcionalmente diferente a todos los genes de la red original.

Comenzamos con una red que tiene N genes, cada uno de los cuales tiene asociado un conjunto de reguladores y una función booleana. Esta red, a la cual llamaremos *la red original*, tiene un determinado paisaje de atractores que determinan las características fenotípicas del organismo. De la red original seleccionamos un gen al azar y lo duplicamos, por lo tanto, el gen duplicado tendrá los mismos reguladores (inputs), los mismos genes regulados (outputs) y la misma función booleanas que el gen del que proviene, tal y como se ilustra en la Fig. 19. Esto producirá una red con $N + 1$ genes, siendo el último gen una copia idéntica de alguno de los genes originales. Después simulamos el proceso de divergencia genética mutando al gen duplicado, es decir, cambiando permanentemente alguno de sus reguladores, alguno de sus regulados, o su función booleanas (o todo al mismo tiempo, lo cual equivaldría a añadir un gen completamente nuevo a la red). Nos referiremos a la red que resulta después de este proceso de duplicación y divergencia como *la red mutada*. No hay razón *a priori* para suponer que los atractores de la red original serán iguales a los de la red mutada. Se sabe, por ejemplo, que el número promedio de atractores aumenta con el tamaño de la red, así que esperamos que la red mutada tenga, en promedio, más atractores que la red original[24].

Lo que nos interesa es comparar el paisaje de atrac-

tores de la red original con el paisaje de atractores de la red mutada, y cuantificar los cambios. En particular, queremos saber cuántos atractores de la red original se conservan, y cuántos cambian. Para esto, generamos en la computadora poblaciones de decenas de miles de redes distintas, cada red con su propio paisaje de atractores. En cada red implementamos el proceso descrito en la Fig. 19 y comparamos los atractores de la red original con los atractores de la red mutada. La Fig. 20 muestra la fracción $P(q)$ de redes en la población en las cuales un porcentaje q de los atractores originales se conservaron en la correspondiente red mutada. Por ejemplo, si $P(30) = 0,4$, esto quiere decir que en dos quintas partes de la población de redes, el 30 % de los atractores originales se conservaron después de la duplicación y divergencia, mientras que el 70 % de los atractores originales cambiaron. La Fig. 20 muestra resultados para poblaciones de redes operando en las tres fases dinámicas. Puede observarse en esta figura que para las redes ordenadas y para las críticas sobresalen dos picos en la gráfica de $P(q)$, uno en $q = 0\%$ y el otro en $q = 100\%$, y casi no hay nada entre estos dos extremos. Esto quiere decir que para tales redes, después de la duplicación y divergencia de un solo gen, ningún atractor se conservó, lo cual corresponde al pico en $q = 0\%$, o bien todos los atractores originales se conservaron, lo cual corresponde al pico en $q = 100\%$. Casi no hay casos intermedios. O se conserva todo, o cambia todo. Sin embargo, notemos de la Fig. 20 que la probabilidad de que se conserven todos los atractores es más grande que la probabilidad de que ninguno se conserve. Por lo tanto, para redes operando en las fases ordenada o crítica, lo más probable es que, después de un evento de duplicación y divergencia se mantengan las características fenotípicas del organismo codificadas en los atractores dinámicos de la red. Esta es la robustez del paisaje epigenético que mencionamos anteriormente. Dicha robustez no está presente en las redes operando en la fase caótica, como puede observarse en la Fig. 20, ya que en dicha fase el pico en $q = 100\%$ desaparece, mientras que el pico en $q = 0\%$ no sólo no desaparece, sino que es el más prominente. Por lo tanto, en la fase caótica no hay robustez ya que el paisaje de atractores nunca se conserva.

¿Y qué pasa con la evolucionabilidad? Como habíamos mencionado anteriormente, el número promedio de atractores crece con el tamaño de la red. Por lo tanto, es probable que la red mutada tenga más atractores que la red original. Así, además de los atractores originales, el paisaje epigenético de la red mutada puede contener nuevos atractores, que corresponderían a nuevas fenotipos en el organismo. En la Fig. 21 mostramos la probabilidad de que la red mutada, además de tener exactamente los mismos atractores que la red original, tenga más. En otras palabras, esta es la probabilidad de desarrollar nuevos fenotipos sin cambiar los ya existentes. La Fig. 21 se muestra dicha probabilidad para redes operando en las tres fases dinámicas, y como vemos, la probabilidad es máxima para redes críticas.

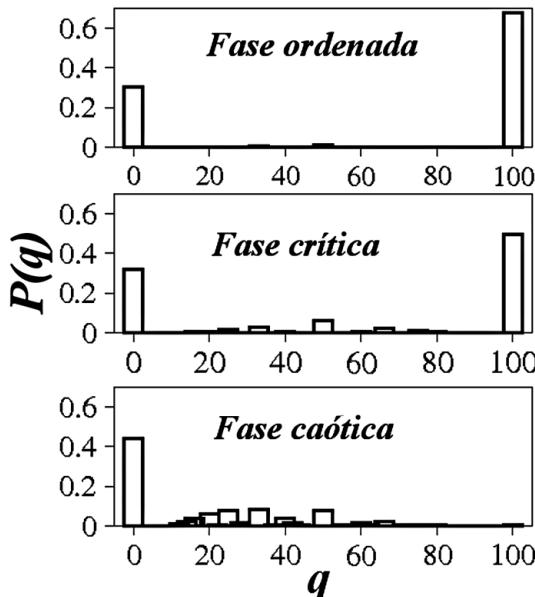


Figura 20. Probabilidad $P(q)$ de que se conserve un porcentaje q de los atractores originales después de la duplicación y divergencia de un solo gen en la red de Kauffman. En las fases ordenada y crítica ocurre con la máxima probabilidad que todos los atractores originales se conservan, mientras que en la fase caótica lo más probable es que ninguno de los atractores originales se conserva.

El resultado mostrado en la Fig. 21 es de la mayor importancia. No sólo estamos pidiendo que la red sea robusta al conservar todos sus atractores, sino que además estamos pidiendo que sea evolucionable ya que nuevos atractores emergen. Esta coexistencia entre robustez y evolucionabilidad hace posible que los organismos reutilicen las soluciones que se han desarrollado para adaptarse a diversos retos ambientales, y que al mismo tiempo sean capaces de generar nuevas soluciones para adaptarse a nuevos retos. Los organismos no se reinventan en cada paso evolutivo, sino que utilizan las soluciones ya existentes y generan nuevas. Y esta coexistencia entre robustez (mantener las soluciones ya existentes) y evolucionabilidad (generar nuevas soluciones) ocurre con la mayor probabilidad en redes con dinámicas críticas. Una representación visual de dicha coexistencia aparece en la Fig. 22.

G. Importancia de la criticalidad genética

Utilizando los datos experimentales más completos que se han producido hasta el momento en cuanto a la estructura y expresión genómicos, hemos demostrado que las redes genéticas de cinco organismos de reinos diferentes exhiben dinámicas críticas. El hecho de que estas redes muestren todas criticalidad dinámica sugiere fuertemente que la criticalidad es una característica genética que ha surgido en la evolución porque de alguna manera confie-

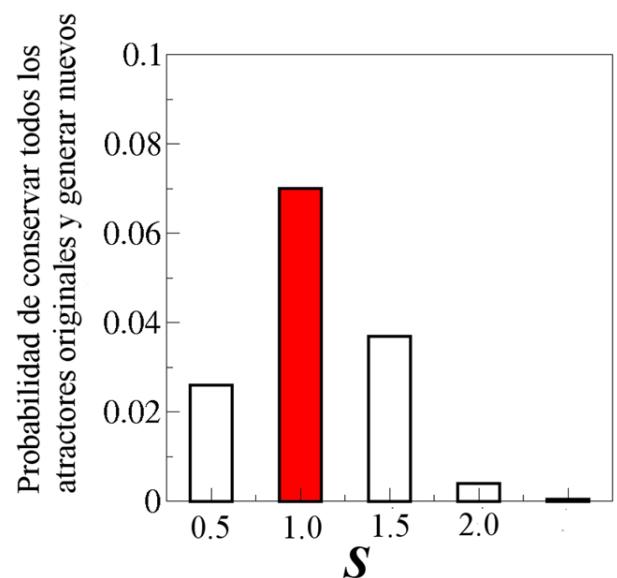


Figura 21. Probabilidad de que la red mutada conserve todos los atractores de la red original y que al mismo tiempo tenga nuevos atractores, como función de la susceptibilidad de la red. Nótese que es para la fase crítica ($S = 1$) para la que se tiene la máxima probabilidad de coexistencia entre robustez y evolucionabilidad.

re ventajas adaptativas a los organismos. En particular, estudiando el comportamiento del paisaje de atractores en poblaciones de redes Booleanas sujetas a mutaciones, hemos encontrado que es justo en la fase crítica donde coexisten con la máxima probabilidad la robustez y la evolucionabilidad observadas en los organismos vivos. Esta coexistencia hace posible, por un lado, que los organismos conserven las características fenotípicas ya adquiridas para adaptarse a sus entornos, y por otro lado les da la plasticidad necesaria para desarrollar nuevas características. Las nuevas características adquiridas con la emergencia de nuevos atractores en el paisaje epigenético pueden ser favorables (o no) para que el organismo se enfrente a nuevos desafíos evolutivos. En cualquier caso, son precisamente las características del fenotipo que emergen con los nuevos atractores dinámicos de la red, las que funguen como el material sobre el cual actúa la selección natural para generalizar organismos adaptados a sus entornos. Es importante notar la diferencia entre la nueva función que adquiere el gen duplicado después de la divergencia, y los nuevos fenotipos que adquiere toda la red. En la literatura científica generalmente se dice que la divergencia genética hace que el gen duplicado adquiera una nueva función (puede generar una nueva proteína, o regular a un conjunto diferente de genes, etc.). Sin embargo, hemos visto que este tipo de “selección relajada”, en el cual el gen duplicado se muta y se retiene, puede cambiar la función de la red genética completa, es decir, puede generar atractores nuevos en los que se codifican nuevas características fenotípicas de todo el organismo. Esto puede ser el mecanismo por el cual ocurre el “equili-

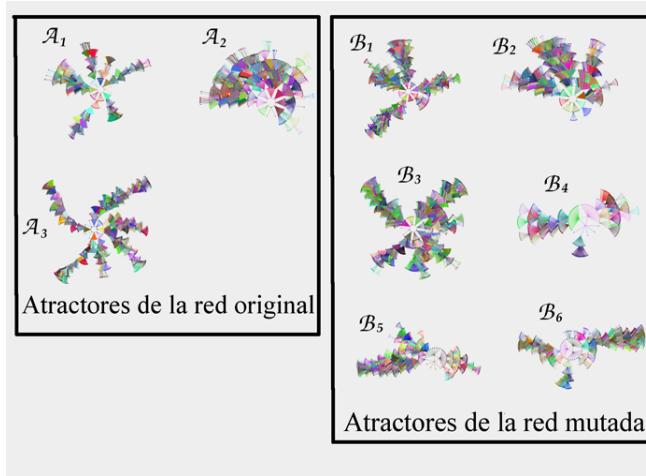


Figura 22. Coexistencia de robustez y evolucionabilidad. El cuadro de la izquierda muestra el paisaje de atractores de la red original, mientras que el cuadro de la derecha muestra el correspondiente paisaje de la red mutada. Los atractores B_1, B_2 y B_3 son idénticos a los atractores A_1, A_2 y A_3 , por lo que el fenotípico codificado en dichos atractores se conserva después de la duplicación y divergencia. Sin embargo, nuevos atractores emergen (B_4, B_5 y B_6) en el paisaje de la red mutada, lo cual le confiere al organismo la posibilidad de evolucionar.

brio puntuado” en la evolución, proceso propuesto por N. Eldredge y S.J. Gould que establece que muy frecuentemente los cambios evolutivos no se dan de forma gradual, poco a poco, sino que ocurren en eventos rápidos que generan ramificación en las especies.

Así, la duplicación y divergencia genética es un mecanismo que constantemente está reestructurando el paisaje epigenético, expandiéndolo y explorando nuevas soluciones ante retos ambientales cambiantes. La criticalidad dinámica hace posible que esta exploración y reestructuración del paisaje epigenético se lleve a cabo de forma robusta, sin que el organismo pierda las características ya adquiridas, generándose así la gran variedad de funcionamientos estables que observamos en la naturaleza que nos rodea.

XII. EPÍLOGO

La teoría de redes complejas es un campo de gran actividad científica en la actualidad. Basta con hacer una búsqueda en Google con la frase “Complex Networks” para darnos cuenta de la gran cantidad de investigadores que se encuentran trabajando en este campo. En estas notas introductorias hemos visto sólo *algunas* de las propiedades importantes de las redes complejas, pero la teoría no se termina aquí. Actualmente uno de los principales desafíos es caracterizar las propiedades dinámicas de las redes complejas dado que conocemos su topología. Por ejemplo, la estructura de la red de regulación genética de *Escherichia coli* es conocida. Sin embargo, aún no

podemos predecir los diferentes fenotípicos que resultan de la dinámica de dicha red. Probablemente la persona que resuelva este problema se ganará el premio Nobel. Pero lo que nos interesa ultimadamente es encontrar las leyes de organización de la materia, si es que tales leyes existen. El hecho de que redes tan diferentes presenten el mismo tipo de topología libre de escala sugiere que puede existir una ley fundamental que determine este tipo de estructuras. Igualmente, el hecho de que especies de organismos tan distintas como las bacterias, los insectos y las plantas hayan evolucionado hacia dinámicas críticas en la red genética sugiere que existen leyes dinámicas que determinan las características fenotípicas de los organismos. Dichas leyes no se han encontrado aún y tal vez nunca se encuentren, pero el viaje hacia su descubrimiento ha resultado más que divertido. Espero con estas notas haberlos motivado para que se unan al estudio del fascinante mundo de las redes complejas y a la búsqueda de las leyes de organización de la materia. ¡En hora buena!

APÉNDICES

Apéndice A: Distribución de Poisson como límite de una distribución binomial

En este apéndice mostramos cómo obtener la Eq. (6) a partir de la Eq. (5), 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 e^a :

$$n! \approx e^{-n} n^n \text{ para } n \text{ grande,}$$

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

Utilizando la fórmula de Stirling, el coeficiente binomial $\binom{N-1}{k}$ que aparece en la ecuación (5) puede escribirse, para N grande, como

$$\begin{aligned} \binom{N-1}{k} &= \frac{(N-1)!}{k!(N-1-k)!} \\ &\approx \frac{e^{-(N-1)}(N-1)^{N-1}}{k! e^{-(N-1-k)}(N-1-k)^{N-1-k}} \\ &= \frac{e^{-k}(N-1-k)^k}{k!} \left(\frac{N-1}{N-1-k}\right)^{N-1} \end{aligned}$$

Por lo tanto, la distribución binomial (5) queda, para N grande, como

$$\begin{aligned} P(k) &= \frac{e^{-k}(N-1-k)^k}{k!} \\ &\times \left(\frac{N-1}{N-1-k}\right)^{N-1} (p_e)^k (1-p_e)^{N-1} \quad (\text{A1}) \end{aligned}$$

Para continuar, notemos que la cantidad $z = 2M/N$ es la conectividad promedio de cada nodo en la red. En términos de esta cantidad, la probabilidad p_e puede escribirse

como $p_e = z/(N - 1)$, con lo cual la ecuación (A1) queda como

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

lo cual, después de arreglar algunos términos, puede escribirse como

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

tomando el límite $N \rightarrow \infty$ y $M \rightarrow \infty$ de tal forma que la conectividad promedio $z = 2M/N$ permanezca constante, y utilizando el hecho de que $e^a = \lim_{n \rightarrow \infty} (1 + a/n)^n$, obtenemos finalmente

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

Por lo tanto, para una red muy grande, el proceso de “hilvanar” parejas de botones escogidas al azar genera una distribución de conexiones de Poisson.

Apéndice B: Ecuación Maestra

Supongamos que tenemos un sistema que puede estar en cualquiera de los estados E_1, E_2, \dots, E_N . Sea $W_{m \rightarrow n}(t)$ la probabilidad condicional de que el sistema, dado que estaba en el estado E_m al tiempo t , “brinque” al estado E_n . Las probabilidades $W_{m \rightarrow n}(t)$ se denominan *probabilidades de transición*.

Sea $P(n, t)$ la probabilidad de que el sistema se encuentre en el estado E_n al tiempo t . Esta probabilidad puede cambiar en el tiempo por dos factores:

1. El sistema estaba en el estado E_n al tiempo t y brincó a otro estado E_m , lo cual claramente disminuye la probabilidad $P(n, t+1)$ de encontrar al sistema en el estado E_n al tiempo $t+1$.
2. El sistema estaba en algún estado E_m al tiempo t y brincó al estado E_n , lo cual aumenta la probabilidad $P(n, t+1)$ de encontrar al sistema en el estado E_n al siguiente instante de tiempo.

Tomando en cuenta estos dos factores, la probabilidad de que el sistema se encuentre en el estado E_n al tiempo $t+1$ está dada por

$$P(n, t+1) = \sum_{m=1}^N P(m, t) W_{m \rightarrow n}(t) - \sum_{m=1}^N P(n, t) W_{n \rightarrow m}(t) \quad (B1)$$

El primer término del lado derecho nos dice que al tiempo t el sistema pudo estar en el estado E_m y brincar al estado E_n , aumentando así la probabilidad de estar en E_n al tiempo $t+1$. El segundo término toma en cuenta el hecho de que al tiempo t el sistema pudo estar en el estado E_n y salirse de allí, brincando al estado E_m y disminuyendo la probabilidad de estar en E_n al tiempo $t+1$. Las sumas sobre m toman en cuenta todos los posibles estados hacia o desde los que puede brincar el sistema.

La ecuación (B1) es la ecuación maestra del sistema. El objetivo es resolverla para encontrar $P(n, t)$. Si las probabilidades de transición $W_{m \rightarrow n}$ son independientes del tiempo, la ecuación se puede resolver más o menos fácilmente. Pero la cosa se complica enormemente si las probabilidades de transición dependen del tiempo. Afortunadamente, las ecuaciones maestras que aparecen en la teoría de crecimiento de redes, al menos en los casos sencillos presentados en estas notas, son relativamente fáciles de resolver.

- [1] En la red del universo de Marvel, dos super héroes están conectados si han aparecido por lo menos una vez en el mismo comic. De este estudio, que puede encontrarse en <http://dmi.uib.es/~joe/marvel.html>, resulta que es el hombre ara na el super héroe con el máximo número de conexiones.
- [2] Statistical Mechanics of Complex Networks. Réka Albert y Albert-László Barabási. *Reviews of Modern Physics* **74**(1):47-97 (2002).
- [3] The Structure and Function of Complex Networks. Mark E.J. Newman. *SIAM Review* **45**(2):167-256 (2003).
- [4] Evolution of Networks. S.N. Dorogovtsev and J.F.F. Mendes. *Advances in Physics* **51**(4):1079-1187 (2002).
- [5] Evolution of Networks: From Biological Nets to the Internet and WWW. S.N. Dorogovtsev and J.F.F. Mendes. Oxford University Press, Oxford. ISBN: 0198515901. (2003). Este libro puede encontrarse **GRATIS** en <http://www.fyslab.hut.fi/~sdo/>
- [6] Connectivity of Growing Random Networks. P. L. Krapivsky, S. Redner, y F. Leyvraz. *Phys. Rev. Lett.* **85**:4629-4632 (2000).
- [7] Power-Law Distribution of the Word Wide Web. L. A. Adamic y B. Huberman. *Science* **287**(24):2115a (2000).
- [8] Bose-Einstein condensation in complex networks. G. Bianconi y A.-L. Barabási. *Phys. Rev. Lett.* **86**:5632-5635 (2001).
- [9] Epidemic Spreading in Scale-Free Networks. Romualdo Pastor-Satorras y Alessandro Vespignani. *Phys. Rev. Lett.* **86**:3201-3203 (2001).
- [10] Spread of Epidemic Disease on Networks. Mark E. J. Newman. *Phys. Rev. E* **66**:16128 (2002).
- [11] Boolean dynamics of networks with scale-free topology. Aldana M. *Physica D* **185** (2003) 4566
- [12] Robustness and evolvability in genetic regulatory networks. Aldana, M., Balleza E., Kauffman S y Resendiz O. *Journal of Theoretical Biology* **245**: 433-448 (2007).
- [13] Critical dynamics in genetic regulatory networks: Examples from four kingdoms. Balleza E., Alvarez-Buylla ER, Chaos A., Kauffman S, Shmulevich I y Aldana M. *PLoS ONE* **3**(6):

e2456 (2998).

- [14] La *anemia falciforme*, por ejemplo, es una enfermedad de los glóbulos rojos de la sangre que se origina por una sola mutación en uno de los más de 600 aminoácidos que conforman a la proteína β -globina encargada de capturar oxígeno.
- [15] En el contexto de las redes complejas, la palabra “topología”.
- [16] A las distribuciones libres de escala también se les llama distribuciones de potencia.
- [17] Las redes con topología exponencial se parecen mucho a las de Poisson.
- [18] En inglés se les llama *hubs*.
- [19] LOGWIN es la concatenación de “log”.
- [20] Análogamente, la distribución de estaturas en los habitantes del D.F. se ha mantenido estacionaria por muchos años aún

cuando la población de chilangos no ha dejado de aumentar.

- [21] Por ejemplo, el cerebro puede detectar y responder a sonidos cuyas intensidades varían desde unos 5 decibeles (ruído de una hoja movida por el viento) hasta 130 decibeles (bocinas en una tocada callejera de rock). Esto representa una diferencia en intensidades de 12 órdenes de magnitud. Lo mismo pasa con la vista.
- [22] La red del desarrollo floral de *A. thaliana* tiene 15 genes, y la red de segmentación polar de *D. melanogaster* tiene 60 genes.
- [23] Actualmente, las bases de datos cuentan con 1400, 900 y 3500 genes, respectivamente.
- [24] Esto es consistente con el hecho de que organismos más complejos (con más características fenotípicas) tienen en general genomas más grandes.