

Curso Redes Complejas, La Plata, 2023, Clase 2^o

Juan I. Perotti*

Instituto de Física Enrique Gaviola (IFEG-CONICET),

Ciudad Universitaria, 5000 Córdoba, Argentina and

Facultad de Matemática, Astronomía, Física y Computación,

Universidad Nacional de Córdoba, Ciudad Universitaria, 5000 Córdoba, Argentina

(Dated: September 1, 2023)

Modelado de la estructura de redes. Redes aleatorias. Modelo configuracional. Modelos de máxima entropía. Detección de comunidades y modelos bayesianos. Generación dinámica de redes.

I. REDES ALEATORIAS

En general, con redes aleatorias nos referimos a cualquier ensamble de redes, típicamente de un dado número de nodos n . Podemos especificar el ensamble de grafos \mathcal{G} utilizando una distribución de probabilidades $P(\mathcal{G} = G)$. Alternativamente, podemos utilizar una distribución de matrices de adyacencia $P(A = a)$. Es importante resaltar, sin embargo, que un grafo de n nodos admite $n!$ matrices de adyacencia equivalentes. A saber, una por cada enumeración de sus nodos.

Las propiedades estadísticas de los grafos de un ensamble se pueden determinar calculando el valor esperado de observables $f \in \mathcal{G} \rightarrow \mathbb{R}$. Por ejemplo,

$$\langle\langle k \rangle\rangle = \sum_{G \in \mathcal{G}} \langle k(G) \rangle P(G)$$

determina el promedio de los grados medios de un ensamble.

A. Grafos de Erdős-Rényi

Posiblemente, el modelo de grafos aleatorios que podamos imaginar viene definido por la distribución de probabilidades uniforme sobre todo el conjunto de grafos con n nodos y m links. En este ensamble, denotado por $\mathcal{G}(n, m)$, la probabilidad de obtener un grafo particular G es

$$P(G) = \frac{1}{\binom{\binom{n}{2}}{m}}$$

Ejercicio 1: Demuestre la anterior aseveración.

En el ensamble $G(n, m)$ todos los grafos poseen m links. Por ende, $\langle\langle k \rangle\rangle = 2m/n$ trivialmente. Sin embargo, una gran variedad de cálculos resultan difíciles y hasta imposibles con este ensamble. Por ende, P. Erdős y A. Rényi introdujeron a fines de los 50's y principios de los 60's un ensamble esencialmente equivalente [1], denotado por $G(n, p)$ y definido por

$$P(a) = \prod_{i < j} P(a_{ij})$$

donde $P(a_{ij} = 1) = p$, $P(a_{ij} = 0) = 1 - p$ y se asumió una enumeración arbitraria de los nodos de entre todas las $n!$ posibles. Es decir, es un ensamble en donde las probabilidades de que los links existan son independientes e iguales a p . En este ensamble, se puede ver que

$$P(a) = p^m (1 - p)^{n - m}$$

Ejercicio 2: Demuestre la anterior aseveración.

En $G(n, p)$, el número promedio de links viene dado por

$$\langle m \rangle = \binom{n}{2} p$$

puesto que $\binom{n}{p} = n(n-1)/2$ es el número máximo de links que un grafo del ensamble puede tener. De manera similar, el grado promedio viene dado por

$$\langle\langle k \rangle\rangle = \left\langle \frac{2m}{n} \right\rangle = \frac{2}{n} \langle m \rangle = \frac{2}{n} \frac{n(n-1)}{2} p = (n-1)p$$

La distribución de grados de las redes de un ensamble se define

$$P(k) = \sum_a \frac{1}{n} \sum_r \sum_{k_r} \delta(k - k_r) P(k_r|a) P(a)$$

y para el caso del ensamble de grafos de Erdős-Rèni resulta en

$$\begin{aligned} P(k) &= \sum_a \frac{1}{n} \sum_r \sum_{k_r} \delta(k - k_r) \delta\left(k_r - \sum_s a_{rs}\right) \prod_i \delta(a_{ii}) \prod_{i < j} p^{a_{ij}} (1-p)^{1-a_{ij}} \delta(a_{ji} - a_{ij}) \\ &= \frac{1}{n} \sum_r \sum_{a_{r1}, \dots, a_{rn}} \delta\left(k - \sum_s a_{rs}\right) \prod_i \delta(a_{ii}) \prod_{i < j} p^{a_{ij}} (1-p)^{1-a_{ij}} \delta(a_{ji} - a_{ij}) \\ &= \frac{1}{n} \sum_r \binom{n-1}{k} p^k (1-p)^{n-1-k} \\ &= \binom{n-1}{k} p^k (1-p)^{n-1-k} \end{aligned}$$

Ejercicio 3: Demuestre que para el caso de redes sparse, la anterior expresión se reduce a la distribución de Poisson

$$P(k) \approx e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$$

cuando $n \gg \langle k \rangle$.

Ejercicio 4: Muestre que el coeficiente de clustering promedio de los grafos del ensamble $G(n, p)$ es

$$\langle\langle c \rangle\rangle = \frac{\langle\langle k \rangle\rangle}{n-1}$$

Cuando $p = 0$, los grafos en $G(n, p)$ no tienen links, y por ende están constituidos por n componentes de un nodo cada una. Cuando $p = 1$, los grafos en $G(n, p)$ resultan totalmente conectados [2], y por ende están constituidos por una única componente. Que pasa para valores intermedios de p ? En general, un grafo del ensamble tendrá varias componentes de tamaños $s_1 \geq s_2 \geq \dots \geq s_t$. Llamamos componente gigante a la primera componente. Nos interesa calcular $\langle s_1 \rangle$ en función de p en el límite $n \rightarrow \infty$.

La probabilidad de que un nodo i elegido al azar no pertenezca a la componente gigante es $u = 1 - s_1/n$. Si i no pertenece a la componente gigante, luego ninguno de sus vecinos puede hacerlo. Entonces, la probabilidad de que i no esté conectado a la componente gigante a través de algún otro nodo arbitrario j tiene dos contribuciones: *a*) o bien i no está conectado a j , *b*) o bien i está conectado a j pero j no pertenece a la componente gigante. La probabilidad de que ocurra *a*) es $1-p$ y la probabilidad de que ocurra *b*) es pu . Entonces, la probabilidad de que i no esté conectado a la componente gigante vía j es $1-p+pu$. Esta probabilidad hay que tenerla en cuenta para todos los nodos diferentes de i . Luego, la probabilidad de que i no esté conectado a la componente gigante a través de ningún otro nodo es $(1-p+up)^{n-1}$. De esta manera, igualando la probabilidad de que i no pertenezca la componente gigante con la probabilidad de que i no esté conectado a la componente gigante a través de ningún otro nodo, obtenemos

$$u = (1-p+up)^{n-1}$$

Para $n \rightarrow \infty$ mientras $\langle k \rangle$ se mantiene constante, la expresión se reduce a

$$u = e^{-\langle k \rangle(1-u)}$$

o, alternativamente, a

$$s_1 = 1 - e^{-\langle k \rangle s_1}$$

Esta es una ecuación trascendental cuyas soluciones podemos analizar gráficamente. Para ello graficamos el lado izquierdo y el lado derecho en función de s_1 , tal como se muestra en la Fig. 1. Vemos que las curvas siempre se intersecan en $s_1 = 0$. Para $\langle k \rangle \leq 1$, es la única intersección posible. Para $\langle k \rangle > 1$ aparece una nueva intersección para $s_1 > 0$. Esta solución es la que resulta probabilísticamente dominante para $\langle k \rangle > 1$. Para demostrar esta última aseveración, necesitamos apelar a la teoría de funciones generatrices.

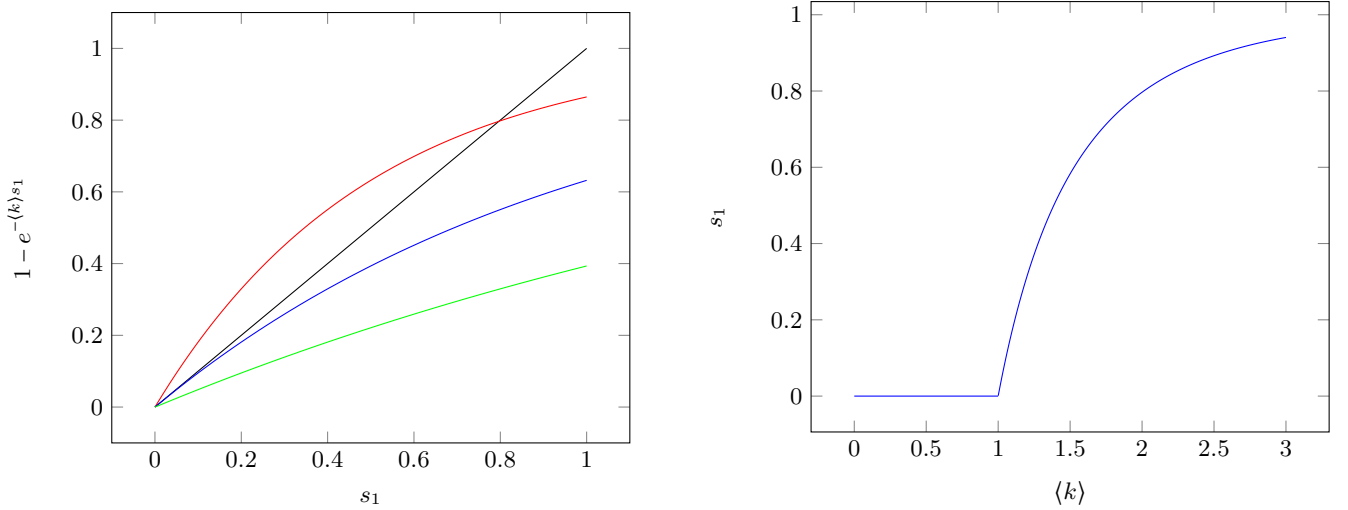


FIG. 1. Solución gráfica de la ecuación $s_1 = 1 - e^{-(k)s_1}$. **Izquierda:** En negro se grafica s_1 vs s_1 . En rojo, azul y verde se grafica $1 - e^{-(k)s_1}$ vs s_1 para $\langle k \rangle$ igual a 2, 1 y $1/2$, respectivamente. La curva roja presenta dos puntos de cruce o soluciones. Una en $s_1 = 0$ y para $s_1 > 0$. Las curvas azul y verde presentan sólo una solución en $s_1 = 0$. La solución para $s_1 > 0$ sólo existe cuando $\langle k \rangle > 1$. **Derecha:** Se grafica s_1 vs $\langle k \rangle$ donde cada valor de s_1 corresponde a la solución probabilísticamente dominante de la ecuación $s_1 = 1 - e^{-(k)s_1}$. Se aprecia una transición de fase continua en $\langle k \rangle = 1$.

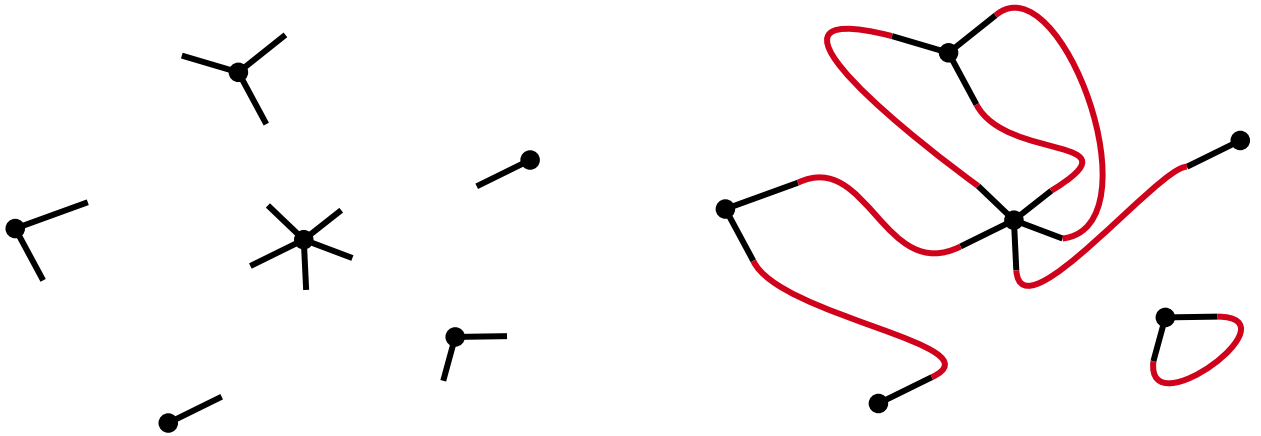


FIG. 2. Construcción de una red utilizando el modelo configuracional. **Izquierda:** Asignamos stubs a los nodos de acuerdo a una secuencia de grados arbitraria. **Derecha:** Conectamos los stubs apareandolos de manera aleatoria.

II. EL MODELO CONFIGURACIONAL

El modelo configuracional permite generar multigrafos no dirigidos de una manera que, en el límite $n \rightarrow \infty$ mientras $\langle k \rangle \sim 1$, básicamente se tornan grafos no dirigidos con una distribución de grados arbitraria, puesto que en dicho límite la probabilidad de que el modelo configuracional genere grafos con multilinks se torna despreciable.

Para generar un grafo de acuerdo al modelo configuracional, comenzamos con n nodos aislados. Luego asignamos un grado k_i a cada nodo $i \in \{1, \dots, n\}$ de manera arbitraria, tal como en el ejemplo del panel izquierdo de la Fig. 2. Es importante que la suma $\sum_i k_i$ sea para para que tenga sentido la expresión $\sum_i k_i = 2m$. Luego, anexamos k_i stubs (o cabitos) al nodo i para cada $i \in \{1, \dots, n\}$ tal como en el ejemplo del panel derecho de la Fig. 2. Notar que puede aparecer selflinks y multilinks.

Ejercicio 3: Desarrolle un algoritmo eficiente para la generación de grafos utilizando el modelo configuracional a partir de una secuencia de grados arbitraria y válida.

A. El modelo de Chung-Lu

El modelo configuracional es análogo al modelo $G(n, m)$ de grafos aleatorios. Existe una versión del modelo configuracional análogo a $G(n, p)$? Si, existen varias. Una posibilidad es la utilizada por Chung-Lu [3], y consiste en asumir links independientes, elegir $P(a_{ij} = 1) = \frac{k_i k_j}{2m}$ para $i \neq j$, $P(a_{ii} = 1) = \frac{k_i^2}{4m}$ y $P(a_{ij} \neq 1) = 1 - P(a_{ij} = 1)$ para todo i, j . Esta formulación funciona relativamente bien pero tiene problemas para redes con hubs grandes, porque la suposición de independencia entre links no resulta adecuada en dicho caso.

Ejercicio 4: Estime la probabilidad de obtener dos links entre i y j . Para ello, use que la probabilidad de que el nodo i se conecte al nodo j por medio de conectar uno de sus stubs con uno de los stubs de j es $P(a_{ij} = 1) \approx k_i \frac{k_j}{2m-1}$, y que la probabilidades correspondientes a diferentes links son aproximadamente independientes. Luego, para $m \gg 1$, estime el número de links dobles que pueden aparecer en la red. Finalmente, estime el número de autolinks simples que pueden aparecer en la red. ¿Tienen sentido estas estimaciones cuando $\langle k^2 \rangle \rightarrow \infty$?

B. Distribución de grados remanentes

Consideremos el modelo configuracional correspondiente a una secuencia de grados $\{k_1, \dots, k_n\}$ para $n \gg 1$, $m \gg 1$ y $\langle k \rangle \sim 1$. La probabilidad de que un nodo elegido al azar de una red generada por el modelo tenga grado k es p_k donde $p_k = \frac{1}{n} \sum_i \delta_{kk_i}$.

Consideremos el siguiente evento. Primero, generamos una red con el modelo, luego elegimos un nodo al azar, digamos el nodo i , y finalmente elegimos uno de sus k_i links. ¿Con que probabilidad el nodo al otro extremo del link tendrá grado k ? Dada la naturaleza del modelo configuracional, el link elegido se conecta a uno de los $2m - 1$ stubs disponibles, de los cuales $k_i - 1$ de ellos pertenecen al nodo i y k_j de ellos pertenecen a cualquier otro nodo j . Luego, para $m \gg 1$ se puede despreciar la probabilidad de que se formen autolinks, y por ende, el nodo al otro extremo del link es j con probabilidad $\approx k_j / 2m$. Luego, el nodo al otro extremo del link tiene grado k con probabilidad

$$\approx \frac{1}{n} \sum_j \delta_{kk_j} \frac{k_j}{2m} = \frac{kp_k}{\langle k \rangle}$$

La probabilidad de que se formen autolinks es despreciable para $m \gg 1$ puesto que la mayoría de los $2m$ stubs no pertenecen al nodo i .

Ejercicio 5: La *friendship paradox* (en promedio, los coautores de tus papers tienen mas citas que vos). Demuestre que, según el modelo configuracional, el grado medio de los vecinos de un nodo elegido al azar es $\sigma_k^2 / \langle k \rangle$. ¿Que ocurre para redes con distribuciones de grado de cola larga?

Ejercicio 6: Verifique numéricamente el resultado del ejercicio 5, implementando el modelo configuracional en su lenguaje de programación preferido, es decir <https://julia.org/>.

El grado remanente es el número de links que tiene un nodo vecino además del link por el cual se lo contactó (ver Fig. 3). Por ende, la probabilidad de que un vecino elegido al azar de entre todos los vecinos de un nodo elegido al azar tenga grado remanente k es

$$q_k = \frac{(k+1)p_{k+1}}{\langle k \rangle}$$

Ejercicio 7: Use la distribución de grados remanente para calcular el coeficiente de clustering promediado sobre el ensamble de grafos generados por el modelo configuracional para una dada secuencia de grados.

Ejercicio 8: Genere un grafo a partir del modelo configuracional para $n = 1000$ y a partir de una secuencia de grados generada con una distribución de Poisson de promedio $\langle k \rangle = 1.2$.

Ejercicio 9: Calcule las componentes del grafo generado en 8. ¿Cuántas componentes ve? ¿Cuál es la distribución de los tamaños las componentes?

C. Distribución de segundos vecinos

La anterior observación nos permite relacionar el modelo configuracional con la teoría de funciones generatrices y los procesos de ramificación (ver. apéndice A). Para ello, consideremos las Probability Generating Functions (PGFs) de los coeficientes p_k y q_k . Es decir, consideramos

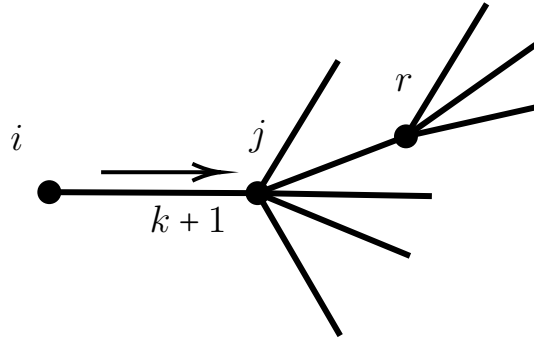


FIG. 3. Ilustración de la noción de grado remanente. El grado remanente de llegar a j desde el nodo i es k , ya que el grado de j es $k + 1$. La noción puede aplicarse nuevamente al ir de j a r , y así sucesivamente.

$$f_0(x) := \sum_{k=0}^{\infty} p_k x^k$$

y

$$f_1(x) := \sum_{k=0}^{\infty} q_k x^k$$

Es facil ver que $f_1(x) = f'_0(x)/\langle k \rangle$ y $f'_0(1) = \langle k \rangle$, y por ende $f_1(x) = f'_0(x)/f'_0(1)$.

El número de segundos vecinos de un nodo es el número de nodos que existen a una distancia topológica 2 del mismo. La probabilidad $p_z^{(2)}$ de que un nodo posea r segundos vecinos satisface

$$p_z^{(2)} = \sum_{k=0}^{\infty} p_k P^{(2)}(z|k)$$

donde $P^{(2)}(z|k)$ denota la probabilidad condicional de que el nodo en cuestión i posea z segundos vecinos dado que tiene k (primeros) vecinos. A su vez

$$P^{(2)}(z|k) = \sum_{y_1=0}^{\infty} \dots \sum_{y_k=0}^{\infty} \delta_{z, \sum_{r=1}^k y_r} \prod_{r=1}^k q_{y_r}$$

puesto que el número de segundos vecinos es igual al grado remanente de cada uno de los primeros vecinos. A decir verdad, esta última relación es una aproximación que desprecia los casos en los que se forman loops, de manera análoga a cómo se desprecian autolinks al calcular los grados remanentes q_k . La aproximación es buena para redes grandes y ralas, i.e. redes en que $n \gg 1$ y $\langle k \rangle \sim 1$, puesto que, bajo estas condiciones, el número de stubs en la vecindad de i es del orden de z , mientras que el número de stubs fuera de la vecindad es del orden de $m \gg z$. Luego, la probabilidad de que un link emergente de las vecindad de i se conecte con otro stub de la vecindad es del orden de $z/m \gg 1$. En otras palabras, la vecindad de i es un grafo que, estadísticamente, es un proceso de ramificaciones.

Consideremos ahora la PGF asociada a la distribución del número de segundos vecinos

$$\begin{aligned}
f^{(2)}(x) &:= \sum_{z=0}^{\infty} p_z^{(2)} x^z \\
&= \sum_{z=0}^{\infty} x^z \sum_{k=0}^{\infty} p_k \sum_{y_1=0}^{\infty} \dots \sum_{y_k=0}^{\infty} \delta_{z, \sum_{r=1}^k y_r} \prod_{r=1}^k q_{y_r} \\
&= \sum_{k=0}^{\infty} p_k \sum_{y_1=0}^{\infty} \dots \sum_{y_k=0}^{\infty} x^{\sum_{r=1}^k y_r} \prod_{r=1}^k q_{y_r} \\
&= \sum_{k=0}^{\infty} p_k \sum_{y_1=0}^{\infty} \dots \sum_{y_k=0}^{\infty} \prod_{r=1}^k q_{y_r} x^{y_r} \\
&= \sum_{k=0}^{\infty} p_k \left(\sum_{y=0}^{\infty} q_y x^y \right)^k \\
&= \sum_{k=0}^{\infty} p_k (g_1(x))^k \\
&= f_0(f_1(x))
\end{aligned}$$

lo cual constituye un caso particular de la relación de composición de PGFs.

Ejercicio 10: Muestre que, cuando la probabilidad de que surjan loops es despreciable, la PGF $f^{(\ell)}$ de la distribución del número de ℓ -ésimos vecinos (i.e. vecinos a distancia topológica ℓ) satisface

$$f^{(\ell)}(x) = f^{(\ell-1)}(f_1(x)) = f_0(f_1(\dots f_1(x) \dots))$$

donde, en el último término, f_1 aparece $\ell - 1$ veces.

Ejercicio 11: Muestre que el número promedio $\langle k^{(\ell)} \rangle$ de ℓ -ésimos vecinos satisface

$$\langle k^{(\ell)} \rangle = \left(\frac{\langle k^{(2)} \rangle}{\langle k \rangle} \right)^{\ell-1} \langle k \rangle$$

D. PGF de la distribución de tamaños de componentes pequeñas

Un tamaño s es macroscópico relativo a un tamaño n cuando $s/n \sim 1$ para $n \rightarrow \infty$, y es microscópico cuando $s/n \rightarrow 0$.

Un grafo de $n \gg 1$ nodos y $\langle k \rangle \sim 1$ links generado por un modelo configuracional, está típicamente constituido por una componente gigante de tamaño macroscópico y una multitud de pequeñas componentes de tamaños microscópicos. Por argumentos similares a los utilizados en la derivación de $p_z^{(\ell)}$, estas componentes pequeñas son predominantemente árboles, i.e. no tienen loops. Por esta observación, intuimos que conviene utilizar las propiedades de las PGFs para ganar información sobre la distribución de tamaños de las pequeñas componentes.

Consideremos un nodo i perteneciente a una componente pequeña de tamaño s . Si el nodo tiene k vecinos j_1, \dots, j_k , su remoción fragmenta la componente en k componentes menores de tamaños w_1, \dots, w_k , donde $s = 1 + w_1 + \dots + w_k$ (ver Fig. 4).

Sea p_s la probabilidad de que un nodo i elegido al azar pertenezca a una componente pequeña de tamaño s . Luego, $p_s = \sum_k p_k P(s|k)$ donde $P(s|k)$ es la probabilidad condicional de que la componente de i tenga tamaño s , dado que el mismo tiene grado k . Si

$$h_0(x) := \sum_{s=1}^{\infty} p_s x^s$$

es la PGF correspondiente a la distribución s , luego

$$\begin{aligned}
h_0(x) &= \sum_{k=0}^{\infty} p_k \sum_{s=1}^{\infty} P(s|k) x^s \\
&= x \sum_{k=0}^{\infty} p_k \sum_{s=1}^{\infty} P(s|k) x^{s-1} \\
&= x \sum_{k=0}^{\infty} p_k \sum_{s=0}^{\infty} P(s+1|k) x^s
\end{aligned}$$

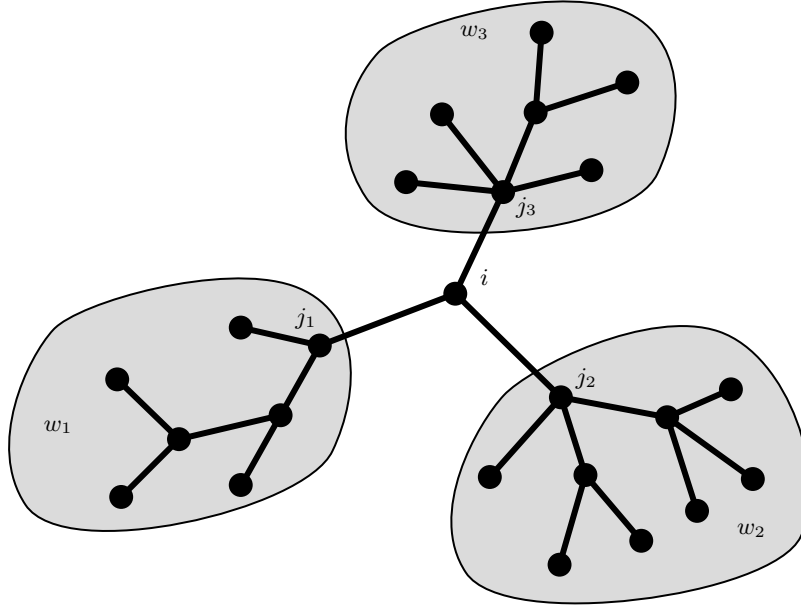


FIG. 4. El nodo i pertenece a una componente pequeña de tamaño $s = 1 + w_1 + w_2 + w_3$ conectándose a ellas a través de k links a correspondientes nodos j_1 , j_2 y j_3 . Si removemos el nodo i , la componente se fragmenta en 3 subcomponentes pequeñas de tamaños w_1 , w_2 y w_3 .

donde

$$v(x) := \sum_{s=0}^{\infty} P(s+1|k)x^s$$

es la PGF de la distribución de probabilidades de que las subcomponentes emergentes de remover i , tengan tamaños que sumen s . Dado el grado k del i , los tamaños w_1, \dots, w_k de las subcomponentes son estadísticamente independientes, tal como ocurre con la estadística de las diferentes ramas de un proceso de ramificación. Luego, teniendo en cuenta la relación de composición de PGFs, podemos escribir

$$v(x) = (h_1(x))^k$$

para alguna PGF $h_1(x) = \sum_{w=1}^{\infty} p_w x^w$ de la distribución de probabilidades p_w del tamaño w de una subcomponente. Luego, juntando resultados, obtenemos que

$$h_0(x) = x \sum_{k=0}^{\infty} (h_1(x))^k = x f_0(h_1(x))$$

lo cual constituye una ecuación que determina h_0 en términos de f_0 y h_1 .

Si el vecino j de i tiene grado remanente k_j , entonces pertenecerá a una subcomponente de tamaño w con probabilidad $P(s = w|k_j)$. Resaltamos aquí que $P(s = w|k_j)$ es la misma distribución que se usó en el cálculo de h_0 , lo cual tiene sentido ya que el tamaño de la (sub)componente depende del tamaño de las (subsub)componentes de sus vecinos. Entonces, procediendo de manera análoga al cálculo de h_0 , obtenemos

$$h_1(x) = x \sum_{k_j=0}^{\infty} q_{k_j} \sum_{w=0}^{\infty} P(w+1|k_j)x^w = x f_1(h_1(x))$$

con la diferencia que se utilizó q_{k_j} , en vez de p_k , porque los k_j son grados remanentes en vez de grados. Esta es una ecuación para la PGF h_1 en términos de la PFG f_1 . No siempre se puede resolver, y cuando se resuelve, no siempre se puede obtener la distribución p_w completa a partir de la solución. De todas formas, estas ecuaciones se pueden utilizar para obtener información útil aún cuando no se las pueda resolver.

E. Componente gigante

La probabilidad de que un nodo elegido al azar pertenezca a una componente pequeña es

$$\sum_{s=1}^{\infty} p_s = h_0(1)$$

Esto pone de manifiesto que, estrictamente hablando, h_0 no es una PGF, ya que $h_0(1) < 1$. De todas formas, las propiedades de h_0 y h_1 utilizadas para obtener las ecuaciones que las determinan no hacen uso de la propiedad de normalización de las PGFs.

La fracción de nodos en la componente gigante es

$$S = 1 - h_0(1) = 1 - f_0(h_1(1))$$

De los resultados de la sección anterior

$$h_1(1) = f_1(h_1(1))$$

Luego, definiendo $u := h_1(1)$, obtenemos

$$S = 1 - f_0(u)$$

y

$$u = f_1(u)$$

Es decir, podemos calcular el tamaño de la componente gigante en términos de f_0 y un punto fijo u de f_1 . El punto fijo de f_1 puede obtenerse de manera gráfica, de manera análoga a como ya lo hicimos para el modelo de Erdős-Rényi en la Fig. 1.

III. MODELOS DE MÁXIMA ENTROPÍA

Consideremos una distribución de probabilidades $P(X)$ sobre una variable estocástica X , y una larga secuencia de observaciones x_1, \dots, x_r de X . Se puede probar que el mínimo número de bits necesario para comunicar las r observaciones (p. ej., usando código Morse) es

$$\geq r \frac{S(X)}{\ln 2}$$

donde

$$S(X) = - \sum_x P(X = x) \ln P(X = x)$$

es la entropía de Shannon de la variable X . En otras palabras, la entropía acota por debajo la cantidad mínima de bits requeridos por observación.

La entropía cuantifica ignorancia. Es decir, cuantifica cuanta información en bits nos falta para determinar cuales fueron las observaciones de X . Luego, la maximización de la entropía respecto a $P(X)$ restringiendo a X a que satisfaga ciertas condiciones o *constraints*, nos permite obtener la distribución $P(X)$ menos informativa posible que satisface las condiciones solicitadas. Este es el enfoque detrás del principio de Máxima Entropía o, simplemente, Max-Ent.

A. Grafos exponenciales

Sea \mathcal{G} el conjunto de todos los grafos $G \in \mathcal{G}$ de tamaño n . La distribución $P(\mathcal{G})$ menos informativa sobre grafos en \mathcal{G} que resulta consistente con ciertos valores esperados

$$x_r = \langle f_r(G) \rangle = \sum_{G \in \mathcal{G}} P(G) f_r(G)$$

para correspondientes observables $f_r \in (\mathcal{G} \rightarrow \mathbb{R})$, es la que maximiza la funcional

$$\Lambda(P, \alpha, \beta) = - \sum_{G \in \mathcal{G}} P(G) \ln P(G) + \alpha \left(1 - \sum_{G \in \mathcal{G}} P(G) \right) + \sum_r \beta_r \left(\langle f_r(G) \rangle - \sum_{G \in \mathcal{G}} P(G) f_r(G) \right)$$

con respecto a la distribución $P(\mathcal{G})$, el escalar α y el vector β de componentes β_r . Es fácil ver que la $P(\mathcal{G})$ que minimiza Λ es la distribución de Boltzmann

$$P(G|\beta) = \frac{e^{E_\beta(G)}}{Z(\beta)}$$

donde Z es una función partición dada por

$$Z(\beta) = \sum_{G \in \mathcal{G}} e^{E_\beta(G)}$$

E_β una función potencial dada por

$$E_\beta(G) = \sum_r \beta_r f_r(G)$$

y los valores de los parámetros β quedan determinados por las ecuaciones

$$\langle f_r(G) \rangle = \frac{1}{Z} \frac{\partial Z}{\partial \beta_r} = \frac{\partial \ln Z}{\partial \beta_r} = \frac{\partial F}{\partial \beta_r}$$

donde se introdujo la *energía libre* F dada por

$$F(\beta) = \ln Z(\beta)$$

1. Versión microcanónica

Es similar a la anterior sólo que formulada utilizando

$$Z = \sum_{G \in \mathcal{G}} e^{\sum_r \beta_r f_r(G)} \prod_r \delta(y_r - g_r(G))$$

donde

$$y_r = g_r(G)$$

son condiciones *duras* sobre el ensamble, y puede repetir o no las *estadísticas* $x_r = \langle f_r(G) \rangle$.

2. Observables estadísticas

Se puede eliminar una condición estadística eligiendo $\beta_r = 0$. Esto puede hacerse a posteriori de calcular la derivada de Z que nos devuelve $\langle f_r(G) \rangle$, y así considerar observables que no afectan la estadística.

B. Ejemplos

El ejemplo más usado es la versión Max-Ent del modelo de Chung-Lu [4]. Para ello, conviene utilizar ensambles A de matrices de adjacencia simétricas a , en vez de ensambles \mathcal{G} de grafos G . En otras palabras, se considera la función partición dada por

$$Z = \sum_{a \in A} e^{E_\beta(a)}$$

Luego, para este modelo, se pide como condiciones fijar una secuencia de grados esperados

$$x_i = \langle k_i \rangle$$

Para ello, se define

$$f_i(a) := k_i = \sum_j a_{ij}$$

para todo $i \in \{1, \dots, n\}$. Esto corresponde a elegir la función potencial

$$E_\beta(a) = \sum_i \beta_i \sum_j a_{ij} = \sum_{ij:i < j} (\beta_i + \beta_j) a_{ij}$$

Luego, se tiene que

$$Z(\beta) = \sum_{a \in A} e^{\sum_{ij:i < j} (\beta_i + \beta_j) a_{ij}} = \prod_{ij:i < j} \sum_{a_{ij} \in \{0,1\}} e^{(\beta_i + \beta_j) a_{ij}} = \prod_{ij:i < j} \left(1 + e^{\beta_i + \beta_j} \right)$$

De esta manera, sumando un término auxiliar a la función potencial,

$$E_\beta(a) = \sum_{ij:i < j} (\beta_i + \beta_j) a_{ij} + \mu_{ij} a_{ij}$$

se puede calcular el valor esperado de a_{rs} . Es decir

$$\langle a_{rs} \rangle = \frac{1}{Z} \frac{\partial Z}{\partial \mu_{rs}} \Big|_{\mu=0} = \frac{1}{Z} \sum_{a \in A} a_{rs} e^{\sum_{ij:i < j} (\beta_i + \beta_j) a_{ij}} = \frac{1}{1 + e^{-(\beta_r + \beta_s)}}$$

Este valor esperado puede interpretarse como la probabilidad p_{rs} de que exista un link entre los nodos r y s en los grafos del ensamble. Notar que para $e^{-(\beta_r + \beta_s)} \gg 1$, se tiene que

$$p_{rs} \approx e^{\beta_r + \beta_s} = e^{\beta_r} e^{\beta_s} = \frac{\langle k_r \rangle \langle k_s \rangle}{2 \langle m \rangle}$$

de manera análoga a lo que se pide en el modelo de Chung-Lu.

Ejercicio 12: Muestre que el ensamble de grafos de Erdős-Rényi puede obtenerse del formalismo de Max-Ent eligiendo $E(G) = m$. ¿Que papel juega β en este caso?

IV. DETECCIÓN DE COMUNIDADES Y MÉTODOS BAYESIANOS

A. Detección de comunidades

El problema de detección de las comunidades de un grafo, consiste en identificar grupos de nodos altamente conectados entre si. La idea consiste en encontrar una partición de nodos para la cuál la conectividad entre los grupos o comunidades resulte significativamente menor que la conectividad interna de cada comunidad.

Este problema fué introducido al área de las redes complejas por M. Girvan y M.E.J. Newman. Como ejemplo, ellos estudian la estructura de comunidades de la red Karate Club (ver Fig. 5). En la red, los links representan relaciones de amistad entre los miembros de un club universitario de Karate en el que el estudiante Wayne W. Zachary participaba. Como consecuencia de una disputa entre el instructor y el presidente del club, el club eventualmente se divide en dos comunidades. Girvan y Newman desarrollan un algoritmo que les permite identificar estas comunidades utilizando la estructura previa a la ruptura del club, con la excepción de un nodo. El algoritmo consiste en ir eliminando, secuencialmente, los links de mayor betweeness hasta que la red se fragmente en comunidades. El algoritmo resulta poco conveniente por el gran costo computacional requerido para recalculer el betweeness de los links, luego de eliminar cada link. Además, el procedimiento de remoción de links puede aplicarse hasta que no queden links y cada nodo se convierta en su propia comunidad, por lo que no queda claro en que punto hay que detenerlo e identificar los fragmentos obtenidos como las *auténticas* comunidades de la red.

B. Modularidad de Girvan-Newman

Posteriormente, Newman proponen detectar la estructura de comunidades de una red, maximizando la llamada *modularidad*

$$Q_c(a) = \sum_{ij} \left(a_{ij} - \frac{k_i k_j}{2m} \right)$$

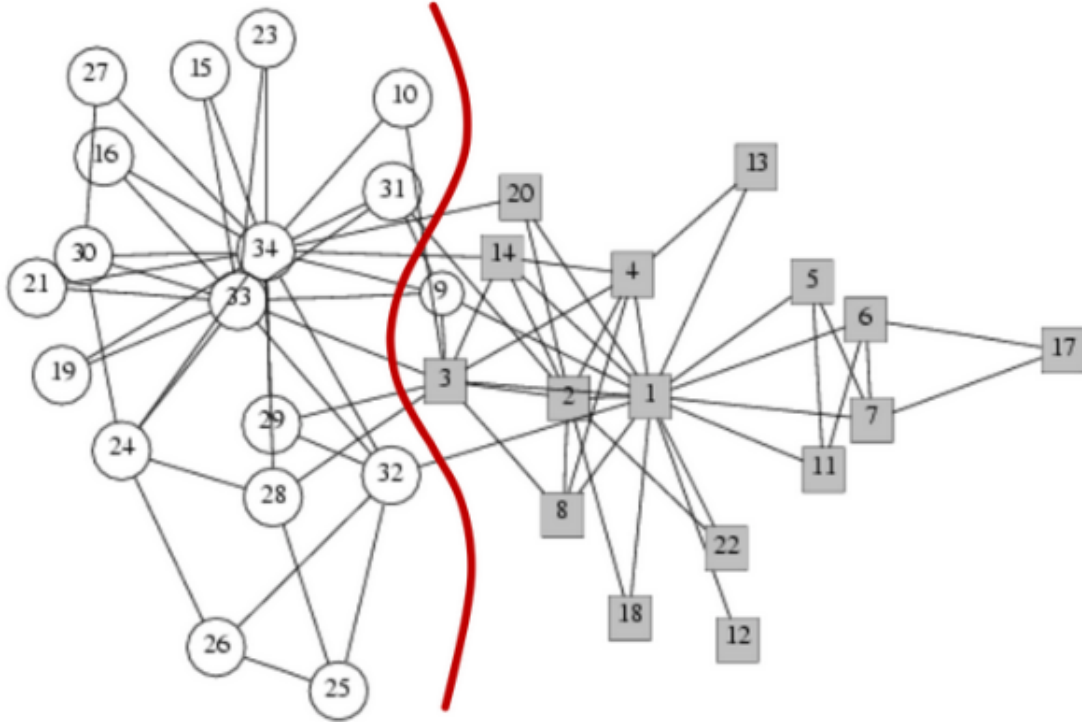


FIG. 5. La estructura de comunidades de la red Karate-Club. El nodo 1 es el instructor y el 34 el presidente del club. La curva roja indica por donde se fragmentó la red. El nodo 9 queda mal identificado por el algoritmo de detección de Girvan y Newman [7].

como función de c [5]. En principio, este enfoque resuelve uno de los problemas del método anterior. A saber, el problema de determinar en que punto el algoritmo debe detenerse para que los fragmentos obtenidos correspondan a las *verdaderas* comunidades de la red. Como se podrán imaginar, sobre todo, quienes tengan experiencia minimizando la energía del modelo de Potts, la maximización de Q es computacionalmente un problema NP-hard. De todas formas, distintas heurísticas pueden utilizarse para encontrar aproximaciones razonables a la estructura de comunidades que maximiza Q .

En 2007, Fortunato y Barthélemy publican un paper en donde muestran que utilizar Q para detectar comunidades presenta problemas [6]. Esencialmente, ellos muestran que la maximización de Q introduce un bias que tiende a favorecer la detección de comunidades de un dado tamaño. Como consecuencia, el método basado en la maximización de Q frecuentemente fusiona módulos pequeños o fragmenta módulos grandes, aún cuando estos están compuestos por subgrafos totalmente conectados.

El trabajo de Fortunato y Barthélemy ha dado lugar a interesantes debates científicos, donde se ha discutido y aún se discute acerca de qué métodos son apropiados para identificar estructuras de comunidades y otras características mesoscópicas en redes complejas. Con el tiempo, ha quedado claro que diversas nociones de la estadística tradicional y la teoría de la información, juegan un rol fundamental en el problema. A continuación, discutimos algunas de estas ideas.

C. Métodos Bayesianos

Consideremos una categorización de los nodos de un grafo de n nodos. La categorización queda determinada por un vector c de componentes $c_i \in \{1, \dots, n\}$. Una categorización define una partición o estructura de comunidades de los nodos. Consideremos ahora ensambles de grafos determinados por una distribución condicional

$$P(A|c)$$

Pensamos a este ensamble como un modelo generativo, en donde $P(a|c)$ es la probabilidad de observar una red a cuya estructura de comunidades sea c .

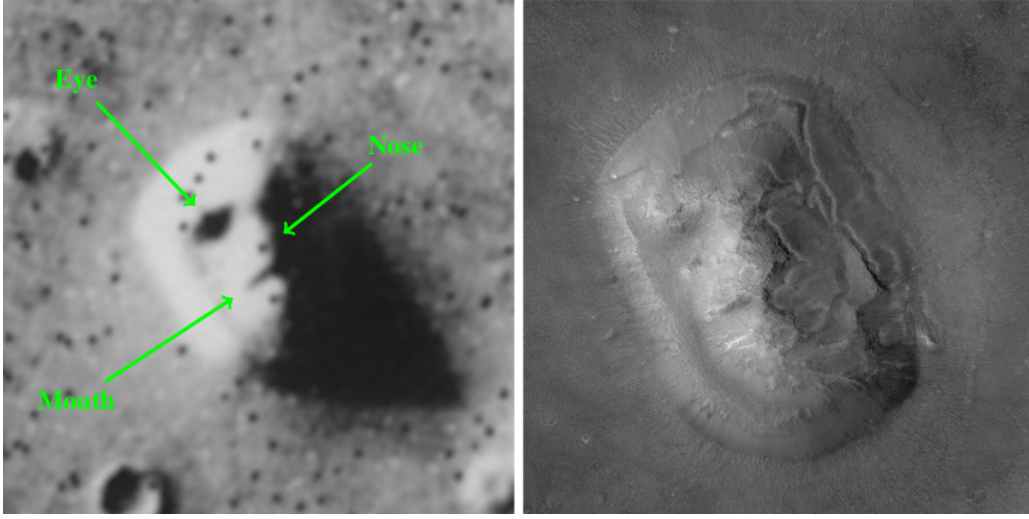


FIG. 6. Sobre el concepto de inferencia estadística y la importancia de considerar distribuciones prior adecuadas. **Izquierda:** Foto tomada en 1976 por la nave espacial Viking 1 de la famosa *cara de Marte*. **Derecha:** Foto más reciente tomada por la nave espacial Mars Global Surveyor en 2007. Fuente [https://en.wikipedia.org/wiki/Cydonia_\(Mars\)](https://en.wikipedia.org/wiki/Cydonia_(Mars)).

El teorema de Bayes nos permite asociar a la distribución $(A|c)$, una distribución posterior $P(C|a)$ dada por

$$P(c|a) = \frac{P(a|c)P(c)}{P(a)}$$

donde

$$P(C)$$

es una distribución de probabilidades llamada prior y

$$P(A) = \sum_{c \in C} P(A|c)P(c)$$

es una distribución llamada evidence. La distribución prior $P(c)$ juega un rol crucial. Se la puede interpretar como una creencia a priori sobre cuales son las estructuras de comunidades viables. Para adquirir una idea intuitiva del rol de una prior, considere la fig. 6. En el panel de la izquierda se observa una foto de Cydonia Mensae, una región de montañosa del planeta marte tomada por la nave espacial Viking 1 en 1976. En la foto se aprecia una figura con la forma de una cara. Algunos conspiranoicos u oportunistas, hablaron de la existencia de civilizaciones perdidas en marte. En el panel de la derecha se observa una foto más reciente, en donde queda mucho más claro que la estructura humanoide no es otra cosa que una montaña. ¿O no? En este ejemplo, se asocian imagenes a a palabras c . Cuando la distribución prior $P(C)$ prioriza palabras antropocentristas, luego tiene sentido que la distribución posterior $P(C|a)$ tienda asociar la palabra cara con la imagen. Si por en cambio $P(C)$ prioriza palabras geográficas, luego tiene sentido que $P(C|a)$ tienda a asociar la palabra montaña con la imagen. El rol de la priors no puede evitarse. Las priors representan el contexto en el que se realizan inferencias. Una versión bayesiana del teorema No Free Lunch nos indica que no existe un modelo generativo universalmente óptimo, en el sentido que siempre es posible mejorar el desempeño de un modelo seleccionando adecuadamente una prior.

Todo método de detección de comunidades puede reformularse en términos de inferencia bayesiana. Para ver esto, supongamos que alguien nos provee de una quality function $Q(a, c)$ que, según nos informan, nos permite detectar la estructura de comunidades de una red arbitraria a , si la maximizamos con respecto a c para el dado a . Luego, interpretamos Q como un potencial en función de c y parametrizado por a , y consideramos la distribución de Boltzmann

$$P(c|a) = \frac{e^{\beta Q_a(c)}}{Z(\beta, a)}$$

donde

$$Z(\beta, a) := \sum_c e^{\beta Q_a(c)}$$

es una función partición y β es un parámetro arbitrario. En el límite $\beta \rightarrow \infty$ recuperamos la maximización de $Q_a(c)$ con respecto a c . En el límite $\beta \rightarrow 0$, la distribución $P(C|a)$ resulta uniforme. Si aplicamos el teorema de Bayes, obtenemos

$$P(a|c) = \frac{e^{\beta Q_a(c)} P(a)}{Z(\beta, a) P(c)} \quad (1)$$

donde $P(A) = \sum_c P(A|c)P(c)$ es la distribución marginal sobre redes y $P(C)$ es una distribución prior sobre particiones. Las distribuciones $P(A|C)$, $P(A)$ y $P(C)$ no son independientes, porque la normalización de $P(A|c)$ impone la siguiente condición

$$\sum_a \frac{e^{\beta Q_a(c)}}{Z(\beta, a)} P(a) = P(c) \quad (2)$$

Aún así, existe cierta libertad de elección. Por ejemplo, podemos elegir la opción de máxima entropía para la distribución conjunta

$$P(a, b) = \frac{e^{\beta Q(c, a)}}{\Omega(\beta)}$$

donde

$$\Omega(\beta) := \sum_{c, a} e^{\beta Q(c, a)}$$

Esta elección resulta en

$$P(a) = \frac{Z(\beta, a)}{\Omega(\beta)}$$

y

$$P(c) = \frac{W(\beta, c)}{\Omega(\beta)}$$

donde

$$W(\beta, c) := \sum_a e^{\beta Q_c(a)}$$

En resumen, detectar comunidades maximizando $Q_a(c)$ con respecto a c se torna bayesianamente análogo a considerar el modelo generativo de la Ec. 1 con la prior de la Ec. 2.

En términos de la teoría de la información, detectar comunidades corresponde a minimizar la Minimum Description Length necesaria para describir conjuntamente la partición elegida y la red

$$L(c, a) = -\ln P(c, a) = -\beta Q(c, a) + \ln \sum_{c', a'} e^{\beta Q(c', a')}$$

V. GENERACIÓN DINÁMICA DE REDES

Hasta ahora hemos discutido como modelar ensambles de redes con características relativamente realistas. Pero no hemos discutido cómo o porqué las redes adquieren la estructura que exhiben. En esta sección presentaremos algunos de los modelos mecanicistas más simples pero importantes que abordan esta cuestión.

A. El modelo de Barabási-Albert

¿Porqué muchas redes adquieren una estructura consistente con una distribución de grados de cola larga? La emergencia de una distribución de grados tipo ley de potencias en redes complejas puede explicarse utilizando el *preferential attachment*. Como se menciona en la clase 1, mecanismos similares fueron previamente estudiadas por

Yule en 1925, Herbert Simon en 1955 y Derek J. de Solla Price en 1976. Pero fué recién en 1999 que la idea fué esencialmente redescubierta e introducida al mundo de las redes complejas por Barabási y Albert [8].

El mecanismo puede describirse con pocas palabras. El mismo consiste en crecer una red semilla de unos pocos nodos y links, agregándole de a un nodo a la vez. Cada nuevo nodo se conecta con los ya preexistentes utilizando m nuevos links. Así, a un tiempo ordinal $t = 0$ se parte de una red semilla de n_0 nodos y m_0 links. Luego, a tiempo $t = 1$ se agrega un primer nodo y q nuevos links. A tiempo $t = 2$ un segundo nodo y m nuevos links, y así sucesivamente. A tiempo t la red tendrá $n_0 + t$ nodos y $m_0 + qt$ links. Crucialmente, y de aquí el nombre, se asume que la probabilidad $p_i(t)$ de que el nodo $n_0 + t + 1$ se conecte a un nodo preexistente $i \in \{1, \dots, n_0 + t\}$ es proporcional al grado $k_i(t)$ que posee dicho nodo i al tiempo t . En otras palabras, que

$$p_i(t) = \frac{k_i(t)}{\sum_{j=1}^{n_0+t} k_j(t)}$$

Para $t \gg n_0, m_0, q$, el proceso esencialmente olvida la forma particular de la red semilla, y exhibe una distribución de grados

$$P_t(k) \sim k^{-\gamma}$$

de exponente $\gamma = 3$ para $k \gg 1$. Este resultado se puede probar analíticamente utilizando diversas aproximaciones.

Ejercicio 13: Implemente el modelo de Barabási-Albert. Genere redes de $n = 1000, 10000$ y 100000 nodos para $m = 3$ y 10 . Use como red semilla, en cada caso, una red totalmente conectada de $n_0 = 2q + 1$ nodos, de manera que $m_0 = n_0(n_0 - 1)/2 = n_0 q$. Calcule, para cada caso, la distribución de grados. Grafique con puntos, simultaneamente y usando diferentes colores, todas las distribuciones de grados en escala log-log. Agregue, como guía al ojo, una curva $\sim k^{-3}$.

Existen una gran cantidad de variaciones del modelo de Barabási-Albert. Algunas de estas variaciones demuestran la generalidad y robustez de la capacidad del mecanismo de generar redes con distribuciones de grados tipo ley de potencias. Por ejemplo, el modelo de Bianconi y Barabási [9], en donde se asigna a cada nodo un fitness intrínseco η_i obtenido de alguna distribución $\rho(\eta)$, de manera que $p_i(t) \sim \eta_i k_i(t)$. Con esta modificación, se pueden obtener distribuciones de grados de exponentes γ distintos a 3. Este modelo exhibe una transición de fase análoga a la observada en la condensación de Bosé-Einstein.

Appendix A: Funciones generatrices

Considerar la siguiente función

$$\begin{aligned} f(x) &= (x^0 + x^1 + x^2)(x^1 + x^2 + x^3 + x^4)(x^0 + x^3) \\ &= x^0 x^1 x^0 + \\ &\quad x^1 x^1 x^0 + x^0 x^2 x^0 + \\ &\quad x^0 x^3 x^0 + x^1 x^2 x^0 + x^2 x^1 x^0 + \dots \end{aligned}$$

En la primera línea de la última igualdad, recopilamos los términos de orden 1, en el segundo los de orden 2, en el tercero los de orden 3, y así. Los diferentes términos de cada renglón ilustran como podemos combinar factores, rojos, azules y verdes para generar un término de una dada potencia. Vemos, entonces, que podemos multiplicar polinomios para algebraizar cálculos combinatorios. La cantidad de términos de orden k nos permite calcular de cuantas maneras podemos combinar 0, 1 o 2 elementos rojos, 1, 2, 3 o 4, elementos azules y, 0 o 3 elementos verdes en un conjunto de k elementos. Son combinaciones porque el orden en que se seleccionan los elementos de un mismo color no se tiene en cuenta.

Consideremos, ahora, la siguiente función

$$f(x) = \sum_{k=0}^{\infty} a_k x^k$$

Es fácil ver que

$$\frac{1}{m!} \frac{d^m f}{dx^m}(x=0) = a_m$$

Entonces, podemos usar esta expresión para calcular los números de combinaciones. Por esta razón, se denomina a f la Generating Function (GF) de los coeficientes a_k [10]. En español, f es una función generatriz.

1. Relaciones útiles

Una relación muy útil es la siguiente

$$\frac{1 - x^{n+1}}{1 - x} = 1 + x + x^2 + \dots + x^n$$

Por ende, estas expresiones corresponden a la función generatriz de coeficientes $a_0 = a_1 = a_2 = \dots = a_n = 1$, también denotada por $(1, 1, 1, \dots, 1, 0, 0, 0, \dots)$ donde el último 1 aparece en la posición $(n + 1)$ -ésima.

Ejercicio A.1: Derive la anterior relación.

Asumiendo $|x| < 1$ y $n \rightarrow \infty$, de la relación anterior se deduce

$$\frac{1}{1 - x} = 1 + x + x^2 + \dots + x^n + \dots$$

que corresponde a la secuencia $(1, 1, 1, \dots)$. También se deduce

$$\frac{1}{1 + x} = 1 - x + x^2 - x^3 + \dots + (-1)^n x^n + \dots$$

que corresponde a la secuencia $(1, -1, 1, -1, \dots)$. De estas relaciones podemos obtener

$$x^k + x^{k+1} + x^{k+2} + \dots + x^{k+n} = x^k (1 + x + x^2 + \dots + x^n) = x^k \frac{1 - x^{n+1}}{1 - x}$$

que corresponde a la secuencia $(0, 0, \dots, 0, 1, 1, \dots, 1, 0, 0, \dots)$ o

$$\frac{1}{1 - ax} = 1 + ax + a^2 x^2 + \dots + a^n x^n + \dots$$

que corresponde a la secuencia $(1, a, a^2, \dots, a^n, \dots)$, o

$$\frac{1}{(1 - x)^2} = \frac{d}{dx} \frac{1}{1 - x} = 0 + 1 + 2x + 3x^2 + \dots$$

que corresponde a la secuencia $(0, 1, 2, 3, \dots)$, y cosas del estilo.

2. Funciones generatrices de probabilidades

Las Probability Generating Functions (PGF) son funciones generatrices cuyos coeficientes representan una distribución de probabilidades de una variable aleatoria K con realizaciones $k \in \{0, 1, 2, \dots\}$. Más precisamente, consideramos

$$f_K(x) = p_0 + p_1 x + p_2 x^2 + \dots + p_k x^k + \dots = \sum_{k=0}^{\infty} p_k x^k = E(x^K)$$

Trivialmente, $f(x = 1) = 1$.

Ejercicio A.2: Muestre que $f'_K(1) = \langle k \rangle$ y que $f''_K(1) = \langle k \rangle^2 - \langle k \rangle$. En general, muestre que

$$\left(\left(x \frac{d}{dx} \right)^r f_K \right) (1) = \left(\left(\frac{d}{d \ln x} \right)^r f_K \right) (1) = \langle k^r \rangle$$

donde

$$x \frac{d}{dx} = \frac{d}{d \ln x}$$

es un operador diferencial.

Una propiedad importante de las PGFs es la siguiente. Sean A y B dos variables aleatorias sobre $\{0, 1, 2, \dots\}$ estadísticamente independientes. Sea $C := A + B$. Luego

$$\begin{aligned}
 f_C(x) &= E(x^C) \\
 &= \sum_{c=0}^{\infty} P(C=c)x^c \\
 &= \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} P(A=a, B=b)x^{a+b} = E(x^A \cdot x^B) \\
 &= \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} P(A=a)P(B=b)x^a x^b \\
 &= \sum_{a=0}^{\infty} P(A=a)x^a \sum_{b=0}^{\infty} P(B=b)x^b \\
 &= \left(\sum_{a=0}^{\infty} P(A=a)x^a \right) \left(\sum_{b=0}^{\infty} P(B=b)x^b \right) \\
 &= E(x^A)E(x^B) \\
 &= f_A(x)f_B(x)
 \end{aligned}$$

Más generalmente, si N es una variable aleatoria en $\{0, 1, 2, \dots\}$ y A_1, A_2, \dots son variables aleatorias independientes en $\{0, 1, 2, \dots\}$, luego $C = A_1 + A_2 + \dots + A_N$ es una variable aleatoria en $\{0, 1, 2, \dots\}$ cuya PGF satisface

$$\begin{aligned}
 f_C(x) &= E(x^C) \\
 &= \sum_{c=0}^{\infty} P(C=c)x^c \\
 &= \sum_{n=0}^{\infty} P(N=n) \sum_{c=0}^{\infty} P(C=c|N=n)x^c \\
 &= \sum_{n=0}^{\infty} P(N=n) \sum_{a_1=0}^{\infty} \dots \sum_{a_n=0}^{\infty} P(A_1=a_1, \dots, A_n=a_n|N=n)x^{a_1+\dots+a_n} \\
 &= \sum_{n=0}^{\infty} P(N=n) \sum_{a_1=0}^{\infty} \dots \sum_{a_n=0}^{\infty} P(A_1=a_1|N=n) \dots P(A_n=a_n|N=n)x^{a_1+\dots+a_n} \\
 &= \sum_{n=0}^{\infty} P(N=n) \sum_{a_1=0}^{\infty} P(A_1=a_1|N=n)x^{a_1} \dots \sum_{a_n=0}^{\infty} P(A_n=a_n|N=n)x^{a_n} \\
 &= \sum_{n=0}^{\infty} P(N=n) \left(\sum_{a_1=0}^{\infty} P(A_1=a_1|N=n)x^{a_1} \right) \dots \left(\sum_{a_n=0}^{\infty} P(A_n=a_n|N=n)x^{a_n} \right) \\
 &= \sum_{n=0}^{\infty} P(N=n) E(x^{A_1}) \dots E(x^{A_n}) \\
 &= \sum_{n=0}^{\infty} P(N=n) f_{A_1|n}(x) \dots f_{A_n|n}(x)
 \end{aligned}$$

En el caso particular que A_1, \dots, A_n siguen una misma distribución $P(A_i=a) = P(A=a)$, luego

$$\begin{aligned}
 f_C(x) &= \sum_{n=0}^{\infty} P(N=n) f_{A_1|n}(x) \dots f_{A_n|n}(x) \\
 &= \sum_{n=0}^{\infty} P(N=n) (f_A(x))^n \\
 &= f_N(f_A(x))
 \end{aligned}$$

Este resultado constituye la *relación de composición de PGFs*.

3. PGF de la distribución de Bernoulli

La distribución de probabilidades de Bernoulli $\text{Bern}(p)$ concierne una variable $B \in \{0, 1\}$ donde $P(B = 0) = 1 - p = P(B = 1) = p$. La PGF de Bernoulli viene dada por

$$f_B(x) = P(B = 0)x^0 + P(B = 1)x^1 = 1 - p + px$$

4. PGF de la distribución binomial

La distribución binomial de probabilidades $\text{B}(n, p)$ concierne una variable $K \in \{0, 1, 2, \dots, n\}$ donde

$$P(K = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

La PGF de la distribución binomial es

$$\begin{aligned} f_K(x) &= \sum_{k=0}^n P(K = k) x^k \\ &= \sum_{k=0}^n \binom{n}{k} p^k (1 - p)^{n-k} x^k \\ &= (1 - p + px)^n \\ &= (f_B(x))^n \end{aligned}$$

5. Procesos de ramificación

Repasamos la teoría de branching processes o procesos de ramificación. Consideramos un tiempo ordinal $t \in \{0, 1, 2, \dots\}$ que denota los niveles de un árbol o “las generaciones de un árbol genealógico”. Sea N_t una variable estocástica que denota el número de individuos en la generación t . Sea $K_{i,t}$ una variable aleatoria que denota el número de descendientes que genera el i -ésimo individuo de la generación t . Con esto en mente, tenemos

$$N_{t+1} = K_{1,t} + K_{2,t} + \dots + K_{N_t,t}$$

Luego, podemos escribir

$$f_{N_{t+1}}(x) = f_{N_t}(f_K(x))$$

donde asumimos que $K_{i,t} = K$ para todo t y $i \in \{1, \dots, N_t\}$ y f_K denota la PGF una variable estocástica K . Es decir, cualquier individuo del árbol genealógico genera k descendientes con probabilidad $P(K = k)$ de manera idéntica y estadísticamente independiente. A la $P(K)$ se la denomina *offspring distribution*.

Podemos pensar que en el proceso que genera el árbol de descendientes, los individuos no mueren y sólo tienen descendientes una sola vez. Alternativamente, podemos pensar que los individuos mueren en cada generación, en cuyo caso N_t representa el número de individuos vivos en la generación t -ésima. O alternativamente, podemos pensar que $K = 0$ representa la muerte de un individuo y que $K > 1$ representa que el individuo continúa y tiene $K - 1$ descendientes o muere y tiene K descendientes. Estas interpretaciones son útiles para relacionar los correspondientes procesos a través de una única formulación.

Para $t = 0$, tenemos $N_1 = K_{1,0} = K$, y por ende

$$f_{N_1} = f_K(x)$$

Para $t = 1$, la anterior ecuación adopta la forma

$$f_{N_2}(x) = f_{N_1}(f_K(x)) = f_K(f_K(x))$$

En general,

$$f_{N_t}(x) = f_K(f_K(\dots f_K(x) \dots))$$

donde f_K aparece t veces. Alternativamente, podemos probar que

$$f_{N_{t+1}}(x) = f_K(f_{N_t}(x))$$

Ejemplo 1: Supongamos que f_K está asociado a la distribución de probabilidades $p_0 = 1$ y $p_k = 0$ para todo $k > 0$. Luego, el branching process resulta trivial. Un individuo inicial no tiene descendencia.

Ejemplo 2: Suponemos $p_0 = 1/100$, $p_1 = 99/100$ y $p_k = 0$ para $k > 1$. En este caso, un individuo tiene grandes chances de tener un descendiente y pocas chances de no hacerlo. El árbol genealógico generado por un individuo consta entonces de una sola línea que termina alrededor de la 100-ésima generación.

Ejemplo 3: Suponemos $p_0 = p_1 = 0$ y $p_2 = 1$. Luego, en cada generación el número de descendientes duplica a la anterior.

Ejercicio A.3: Teniendo en cuenta que los individuos no mueren, cuantos individuos existen en la generación t -ésima del ejemplo 3, si partimos de un sólo individuo?

Ejercicio A.4: Demuestre que $p_e = 1$ si $\langle k \rangle \leq 1$.

Ejercicio A.5: Muestre que $\langle N_t \rangle = \langle K \rangle^t$.

a. Probabilidad de extinción

La representación en serie de la PGF para N_t es

$$f_{N_t}(x) = P(N_t = 0) + P(N_t = 1)x + P(N_t = 2)x^2 + \dots$$

Luego,

$$f_{N_t}(x = 0) = P(N_t = 0)$$

es la probabilidad de que en la t -ésima generación el número de descendientes sea 0. En otras palabras, representa la probabilidad de que la población haya dejado de tener descendientes en alguna generación anterior o igual a t .

Sea

$$f_{N_\infty}(x) := \lim_{t \rightarrow \infty} f_{N_t}(x)$$

Luego,

$$p_e := f_{N_\infty}(x = 0)$$

es la probabilidad de que el proceso genere un árbol de descendientes de tamaño finito. En el caso en que el árbol se interpreta como individuos que pueden morir en cada generación, p_e representa la probabilidad de extinción de la población.

Utilizando que

$$\lim_{t \rightarrow \infty} f_{N_{t+1}}(x) = \lim_{t \rightarrow \infty} f_{N_t}(x) = f_{N_\infty}(x = 0)$$

y un resultado previo, se obtiene que

$$f_{N_\infty}(x) = f_K(f_{N_\infty}(x))$$

lo cual constituye una ecuación para $f_{N_\infty}(x)$ en términos de $f_K(x)$. Esta ecuación establece relaciones de recurrencia para los coeficientes de la representación en serie

$$f_{N_\infty}(x) = P(N_\infty = 0) + P(N_\infty = 1)x + P(N_\infty = 2)x^2 + \dots$$

en términos de los coeficientes de la representación en serie

$$f_K(x) = P(K = 0) + P(K = 1)x + P(K = 2)x^2 + \dots$$

Además, replazando x por 0, obtenemos que

$$p_e = f_K(p_e)$$

i.e., la probabilidad de extinción p_e es un punto fijo de la PGF f_K .

Ejercicio A.3: Considere $p_0 = 1/4$, $p_1 = 1/2$, $p_2 = 1/4$ y $p_k = 0$ para $k > 2$. Cuanto vale p_e ?

Ejercicio A.4: Muestre que $f_K(x)$ es creciente cuando $p_k > 0$ para algún $k > 0$ y convexa cuando $p_k > 0$ para algún $k > 1$.

Ejercicio A.5: Muestre que $p_e = 1$ si $\langle k \rangle \leq 1$.

Ejercicio A.6: Sea $M_t := N_0 + N_1 + \dots + N_t$ donde $N_0 = 1$, i.e. asumimos que el branching process inicia con un único individuo. Calcule $\langle M_t \rangle$. Muestre que $\langle M_\infty \rangle := \lim_{t \rightarrow \infty} \langle M_t \rangle = \frac{1}{1 - \langle k \rangle}$ cuando $\langle k \rangle < 1$. De manera similar, muestre que $\langle M_\infty \rangle = \infty$ cuando $\langle k \rangle \geq 1$.

Ejercicio A.7: Cuando $\langle k \rangle = 1$, es interesante observar que $\langle M_\infty \rangle = \infty$ a pesar de que $p_e = 1$. ¿Que cree Ud. que ocurre?

* juan.perotti@unc.edu.ar

- [1] A decir verdad, aparentemente fueron originalmente introducidos por Solomonoff y Rapoport, y posteriormente popularizados por Erdős y Rényi.
- [2] Es decir, todos los links existen.
- [3] F. Chung and L. Lu, The average distances in random graphs with given expected degrees, Proceedings of the National Academy of Sciences **99**, 15879 (2002).
- [4] G. Bianconi, Entropy of network ensembles, Phys. Rev. E **79**, 036114 (2009).
- [5] M. E. Newman, Modularity and community structure in networks, Proceedings of the national academy of sciences **103**, 8577 (2006).
- [6] S. Fortunato and M. Barthelemy, Resolution limit in community detection, Proceedings of the national academy of sciences **104**, 36 (2007).
- [7] M. Girvan and M. E. Newman, Community structure in social and biological networks, Proceedings of the national academy of sciences **99**, 7821 (2002).
- [8] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, science **286**, 509 (1999).
- [9] G. Bianconi and A.-L. Barabási, Competition and multiscaling in evolving networks, Europhysics letters **54**, 436 (2001).
- [10] H. S. Wilf, *generatingfunctionology* (A. K. Peters, Ltd., USA, 2006).