Conectividad en un grafo aleatorio

Juan Piccini

June 3, 2013

En este material trabajaremos en el ejercicio 5) del práctico 8: Tenemos una red con n nodos conectados entre sí (cada nodo tiene un enlace directo con los otros n-1 nodos). Esta red se conoce como el grafo completo K_n. La cantidad de enlaces (o aristas) es $C_2^n = \frac{n(n-1)}{2}$.

Cada arista puede estar funcionando o no, supondremos que el que una arista funcione o no es independiente del estado de las restantes aristas, y además que todas ellas tienen la misma probabilidad p de funcionamiento (y $1-p$ de no funcionamiento).

Se desea estimar $P(n)$, la probabilidad que la red permanezca conectada (que exista al menos un camino entre cada par de nodos).

Lo haremos sorteando una cantidad N sufcientemente grande de tales redes, utilizando el método de Montecarlo para hallar un estimador de $P(n)$. Nos basaremos en **la matriz de adyacencia** del grafo: es una matriz $A_{n\times n}$ tal que $A_{i,j}$ vale 1 si hay una arista entre los nodos i, j y vale 0 en caso contrario. En la diagonal principal ponemos ceros (grafo sin lazos).

Para el grafo completo K_n esta matriz tendría unos en todas sus casillas excepto la diagonal principal. Pero si cada arista puede fallar con probabilidad 1-p, entonces lo que hacemos es inicializar una matriz $n \times n$ de ceros, y luego para cada lugar i,j sorteamos una variable $X \sim U[0,1]$. Si $X <= p$ significa que hay una arista entre los nodos i,j y pondremos un 1 en el lugar i,j de dicha matriz. Si $X > p$ dejaremos el 0 en el lugar i,j.

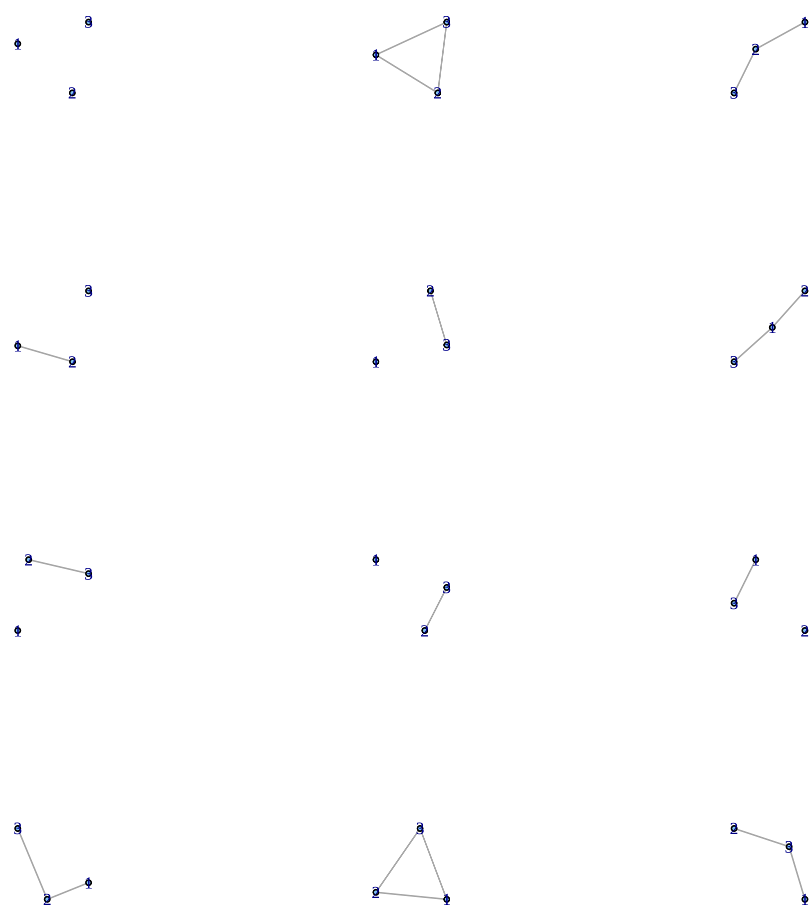
Como la matriz es simétrica (el grafo es no dirigido), el ciclo que implementa esto es un loop doble, el primero recorre las filas desde la primera a la penúltima, el segundo recorre las columnas desde la segunda hasta la última, y cada vez que ponemos un 1 en el lugar i,j, hacemos lo mismo en el lugar j,i.

Una vez que tenemos esta matriz de adyacencia que representa un grafo de n nodos donde la existencia de aristas entre sus nodos es aleatoria e independiente con probabilidad p, testeamos si dicho grafo es conexo o no. Recordemos que para que el grafo sea conexo no tiene porqué existir un enlace directo entre cada par de nodos, sino que alcanza con que exista un camino para cada par de nodos.

Si repetimos esto N veces (n y p fijos), tendremos N realizaciones de nuestro grafo, donde para cada pareja de nodos el que exista una arista es algo aleatorio. En algunas de estas N realizaciones, dos nodos i,j estarán unidos por un enlace, en otras no.

A continuación mostramos algunas realizaciones de estos grafos, en ellas se aprecia que al aumentar p aumenta la frecuencia con la que dichas realizaciones son grafos conexos. Para n=3 y p=0.5 tenemos que $P(n) = 3p^2 - 2p^3$. Este polinomio se obtuvo mediante la fórmula de inclusión-exclusión.

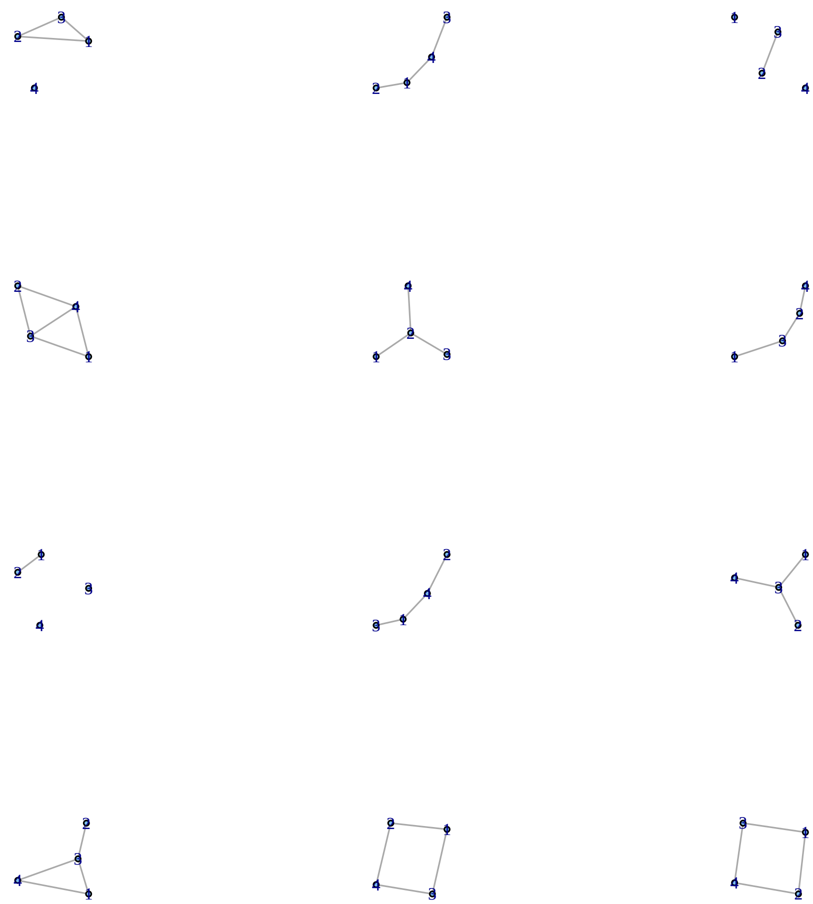
Figure 1: Realizaciones con p=0.5 y n=3



Si ahora mantenemos p=0.5 pero n=4, tenemos que la proporción de realizaciones que son conexas aumenta.

Contamos luego cuantos de esos N grafos resultaron conexos (para ello a

Figure 2: Realizaciones con p=0.5 y n=4



cada grafo se le corre un algoritmo que testea si es conexo o no) y dividimos dicha cantidad entre N. Este será $\overline{P}(N)$, nuestro estimador para P(n).

Si repetimos esto para varios valores de p y varios valores de n, obtendremos una tabla con información interesante.

A continuación mostramos los resultados obtenidos cuando $p = 0.5, 0.6, 0.7, 0.8, 0.9$, y $n = 2, 3, \ldots, 14$. La siguiente tabla muestra como evoluciona $\overline{P}(N)$ en función de p y n. El gráfico 3 se construye ploteando cada fila con un color distinto.

Tabla del estimador según p y n

p/n	2	3	4	5	6	7	8	9	10	11	12	13	14
0.5	0.485	0.515	0.577	0.723	0.818	0.882	0.942	0.959	0.980	0.989	0.994	0.994	0.998
0.6	0.629	0.638	0.779	0.895	0.937	0.965	0.986	0.994	0.998	0.999	0.997	1	1
0.7	0.695	0.761	0.886	0.961	0.985	0.998	0.998	1	1	1	1	1	1
0.8	0.781	0.895	0.960	0.990	0.998	1	1	1	1	1	1	1	1
0.9	0.892	0.977	0.998	1	1	1	1	1	1	1	1	1	1

Cuando el grafo solamente tiene 2 nodos, la probabilidad de que sea conexo es p, de ahi que P(N) esté muy cerca de p. Al aumentar el número de nodos n, la probabilidad de obtener un grafo desconectado baja, o lo que es lo mismo, aumenta la probabilidad de tener un grafo conexo.

Fijados n y p, en cada una de las N simulaciones lo que tenemos es una variable $X_i \sim Ber(P(n), i = 1, \ldots, N)$ (el éxito es que la realización i-sima sea un grafo conexo, lo que ocurre con probabilidad $P(n)$, o más precisamente $P(p, n)$).

Por la Ley Fuerte de los Grandes Números, nuestro estimador $\overline{P}(N)$ tiende a $P(n)$. Como cada X_i tiene esperanza $\mu = P(n)$ y varianza $\sigma^2 = P(n)(1 - P(n))$, podemos aplicar el Teorema Central del Lmite para hallar un intervalo de confianza centrado en $\overline{P}(N)$ para $P(n)$.

Si elegimos un intervalo al 90%, sea $z_{0.95} \simeq 1.645$ el punto bajo la campana de Gauss que deja el 5% del área a su derecha.

Entonces el intervalo es

$$[\overline{P}(N) - \frac{z_{0.95}\sqrt{\overline{P}(N)}(1 - \sqrt{\overline{P}(N)})}{\sqrt{N}}, \overline{P}(N) + \frac{z_{0.95}\sqrt{\overline{P}(N)}(1 - \sqrt{\overline{P}(N)})}{\sqrt{N}}]$$

Por ejemplo, para N=500 (el N usado para simular las realizaciones) y $\overline{P}(N) = 0.990$ correspondiente a p=0.8 y n=5, tenemos que el IC al 90% es $[0.982, 0.997]$. Este intervalo tiene un 90% de probabilidad de capturar al verdadero valor $P(0.8, 5)$.

Figure 3: Probabilidad de conexión en función de n y p