Procesos de Generación Aleatoria de Grafos

Miguel Raggi

ENES Morelia

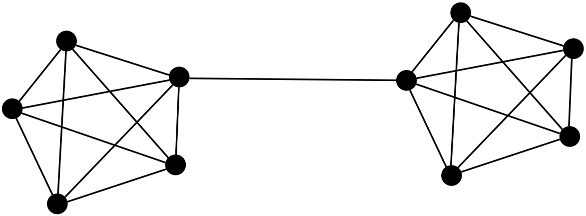
21 de febrero de 2018

Índice:

Índice:

Modelos

- Ahora vamos a intentar atacar la siguiente pregunta: Tenemos una red que obtuvimos del mundo real y nos preguntamos cómo se “formó”.
- Es decir, queremos tratar de entender qué procesos influyeron en su formación.
- Es un arte, porque no podemos estar seguros.
- Pero por ejemplo, si vemos una gráfica así, no pensaríamos que se formó completamente aleatoriamente:



- Para empezar a entender esto, debemos ver cómo tienden a ser las gráficas completamente aleatorias.

Índice:

Formación Erdős-Renyi

Definición

El modelo de Erdős-Renyi de formación de una red tiene dos parámetros: n, el número de vértices, y p, la probabilidad de cada arista de "estar".

- *Cada arista está con probabilidad de p, independientemente de las demás aristas.*
- *Como si lanzáramos una moneda (que tiene probabilidad p de salir sol) por cada arista y si sale sol, ponemos la arista.*

Propiedades de ER

- En promedio, ¿cuántas aristas hay?

Propiedades de ER

- En promedio, ¿cuántas aristas hay? Pues $p\binom{n}{2}$.

Propiedades de ER

- En promedio, ¿cuántas aristas hay? Pues $p\binom{n}{2}$.
- El número de aristas sigue una distribución binomial con probabilidad p y número $M =$

Propiedades de ER

- En promedio, ¿cuántas aristas hay? Pues $p\binom{n}{2}$.
- El número de aristas sigue una distribución binomial con probabilidad p y número $M = \binom{n}{2}$.
- Así que la probabilidad de tener exactamente k aristas es de

$$P(|E(G)| = k) =$$

Propiedades de ER

- En promedio, ¿cuántas aristas hay? Pues $p\binom{n}{2}$.
- El número de aristas sigue una distribución binomial con probabilidad p y número $M = \binom{n}{2}$.
- Así que la probabilidad de tener exactamente k aristas es de

$$P(|E(G)| = k) = \binom{M}{k} p^k (1-p)^{M-k}$$

Propiedades de ER

- En promedio, ¿cuántas aristas hay? Pues $p\binom{n}{2}$.
- El número de aristas sigue una distribución binomial con probabilidad p y número $M = \binom{n}{2}$.
- Así que la probabilidad de tener exactamente k aristas es de

$$P(|E(G)| = k) = \binom{M}{k} p^k (1-p)^{M-k}$$

- ¿Cuál es la varianza del número de aristas?

Propiedades de ER

- En promedio, ¿cuántas aristas hay? Pues $p\binom{n}{2}$.
- El número de aristas sigue una distribución binomial con probabilidad p y número $M = \binom{n}{2}$.
- Así que la probabilidad de tener exactamente k aristas es de

$$P(|E(G)| = k) = \binom{M}{k} p^k (1-p)^{M-k}$$

- ¿Cuál es la varianza del número de aristas? $Mp(1-p)$

Propiedades de ER

- En promedio, ¿cuántas aristas hay? Pues $p\binom{n}{2}$.
- El número de aristas sigue una distribución binomial con probabilidad p y número $M = \binom{n}{2}$.
- Así que la probabilidad de tener exactamente k aristas es de

$$P(|E(G)| = k) = \binom{M}{k} p^k (1-p)^{M-k}$$

- ¿Cuál es la varianza del número de aristas? $Mp(1-p)$
- Entonces podemos comparar una red dada con la red que generaría ese mismo número de aristas (despejando p) y después comparar propiedades de la red aleatoria contra propiedades de la red que tienes.

Propiedades de ER

- En promedio, ¿cuántas aristas hay? Pues $p\binom{n}{2}$.
- El número de aristas sigue una distribución binomial con probabilidad p y número $M = \binom{n}{2}$.
- Así que la probabilidad de tener exactamente k aristas es de

$$P(|E(G)| = k) = \binom{M}{k} p^k (1-p)^{M-k}$$

- ¿Cuál es la varianza del número de aristas? $Mp(1-p)$
- Entonces podemos comparar una red dada con la red que generaría ese mismo número de aristas (despejando p) y después comparar propiedades de la red aleatoria contra propiedades de la red que tienes.
- Por ejemplo, ¿cuántos triángulos tiene (en promedio) el modelo ER?

Propiedades de ER

- En promedio, ¿cuántas aristas hay? Pues $p\binom{n}{2}$.
- El número de aristas sigue una distribución binomial con probabilidad p y número $M = \binom{n}{2}$.
- Así que la probabilidad de tener exactamente k aristas es de

$$P(|E(G)| = k) = \binom{M}{k} p^k (1-p)^{M-k}$$

- ¿Cuál es la varianza del número de aristas? $Mp(1-p)$
- Entonces podemos comparar una red dada con la red que generaría ese mismo número de aristas (despejando p) y después comparar propiedades de la red aleatoria contra propiedades de la red que tienes.
- Por ejemplo, ¿cuántos triángulos tiene (en promedio) el modelo ER? Pues $p^3\binom{n}{3}$.

Propiedades de ER

- En promedio, ¿cuántas aristas hay? Pues $p\binom{n}{2}$.
- El número de aristas sigue una distribución binomial con probabilidad p y número $M = \binom{n}{2}$.
- Así que la probabilidad de tener exactamente k aristas es de

$$P(|E(G)| = k) = \binom{M}{k} p^k (1-p)^{M-k}$$

- ¿Cuál es la varianza del número de aristas? $Mp(1-p)$
- Entonces podemos comparar una red dada con la red que generaría ese mismo número de aristas (despejando p) y después comparar propiedades de la red aleatoria contra propiedades de la red que tienes.
- Por ejemplo, ¿cuántos triángulos tiene (en promedio) el modelo ER? Pues $p^3\binom{n}{3}$.
- Si tu red tiene bastantes más triángulos que eso (o bastante menos), probablemente hay algo "raro".

Umbrales

Definición

Dada una propiedad (de grafos) Q *tal que* $Q(G)$ *es verdadero implica que* $Q(G + \text{una arista})$ *también es verdadero,*

Umbrales

Definición

Dada una propiedad (de grafos) Q tal que $Q(G)$ es verdadero implica que $Q(G + \text{una arista})$ también es verdadero, un umbral para Q es un número $u = u(n)$ tal que para todo $U > u$

$$\lim_{n\to\infty} P(Q(ER(n, U))) \to 1$$

pero no es cierto para $U < u$.

Umbrales

Definición

Dada una propiedad (de grafos) Q tal que $Q(G)$ es verdadero implica que $Q(G+\text{una arista})$ también es verdadero, un umbral para Q es un número $u = u(n)$ tal que para todo $U > u$

$$\lim_{n\to\infty} P(Q(ER(n,U))) \to 1$$

pero no es cierto para $U < u$.

- En palabras, un umbral U es la mínima probabilidad que necesitas para que la probabilidad de que ocurra cierta cosa sea alta.

Umbrales

Umbrales

Veamos en sage ejemplos de esto con $n = 50$:

- Si $p \approx \frac{1}{n^2}$, el grafo comienza a tener aristas.

Umbrales

Veamos en sage ejemplos de esto con $n = 50$:

- Si $p \approx \frac{1}{n^2}$, el grafo comienza a tener aristas.
- Si $p \approx \frac{1}{n^{3/2}}$, el grafo comienza a tener una componente con al menos 3 aristas.

Umbrales

Veamos en sage ejemplos de esto con $n = 50$:

- Si $p \approx \frac{1}{n^2}$, el grafo comienza a tener aristas.
- Si $p \approx \frac{1}{n^{3/2}}$, el grafo comienza a tener una componente con al menos 3 aristas.
- Si $p \approx \frac{1}{n}$, ya tiene ciclos, y tendrá una componente gigante.

Umbrales

Veamos en sage ejemplos de esto con $n = 50$:

- Si $p \approx \frac{1}{n^2}$, el grafo comienza a tener aristas.
- Si $p \approx \frac{1}{n^{3/2}}$, el grafo comienza a tener una componente con al menos 3 aristas.
- Si $p \approx \frac{1}{n}$, ya tiene ciclos, y tendrá una componente gigante.
- Si $p \approx \frac{ln(n)}{n}$, ya será conexa.

Histogramas grados

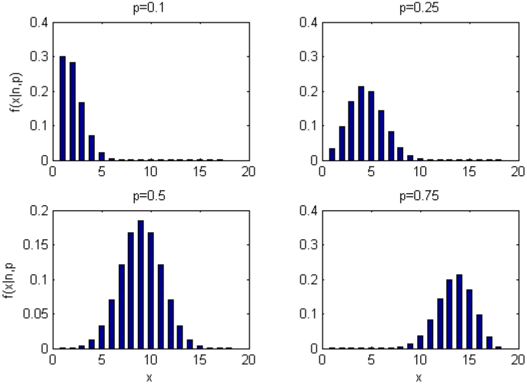
- ¿Cómo es la distribución de grados en este modelo?

Histogramas grados

- ¿Cómo es la distribución de grados en este modelo?
- Pues es binomial. Es decir, la probabilidad de que un vértice tenga grado k es $\binom{n}{k}p^k(1-p)^{n-k}$

Histogramas grados

- ¿Cómo es la distribución de grados en este modelo?
- Pues es binomial. Es decir, la probabilidad de que un vértice tenga grado k es $\binom{n}{k}p^k(1-p)^{n-k}$
- Por ejemplo:



Otros modelos de ER

Claro, hay otros modelos de ER, que no estudiaremos a fondo:

Otros modelos de ER

Claro, hay otros modelos de ER, que no estudiaremos a fondo:

- Un grafo aleatorio bipartito

Otros modelos de ER

Claro, hay otros modelos de ER, que no estudiaremos a fondo:

- Un grafo aleatorio bipartito
- En general, dado un grafo H, puedes escoger un subgrafo aleatorio tomando cada arista con probabilidad p.

Otros modelos de ER

Claro, hay otros modelos de ER, que no estudiaremos a fondo:

- Un grafo aleatorio bipartito
- En general, dado un grafo H, puedes escoger un subgrafo aleatorio tomando cada arista con probabilidad p.
- Un grafo aleatorio de n vértices y un número fijo de aristas.

Para ahorita

- P: ¿Cómo generamos un grafo aleatorio con exactamente m aristas?

Para ahorita

- P: ¿Cómo generamos un grafo aleatorio con exactamente m aristas?
- R: Depende: si m es chico, ponemos aristas aleatorias hasta juntar m.

Para ahorita

- P: ¿Cómo generamos un grafo aleatorio con exactamente m aristas?
- R: Depende: si m es chico, ponemos aristas aleatorias hasta juntar m. Si m es grande, tomamos la lista de todas las aristas y las vamos escogiendo y quitando de una por una.
- Programar: Erdős-Renyi y Erdős-Renyi con exactamente m aristas.
- P: ¿Cómo generamos un árbol aleatorio?

Para ahorita

- P: ¿Cómo generamos un grafo aleatorio con exactamente m aristas?
- R: Depende: si m es chico, ponemos aristas aleatorias hasta juntar m. Si m es grande, tomamos la lista de todas las aristas y las vamos escogiendo y quitando de una por una.
- Programar: Erdős-Renyi y Erdős-Renyi con exactamente m aristas.
- P: ¿Cómo generamos un árbol aleatorio?
- R: Empieza con un vértice y ve agregando vértices de uno por uno, conectando exactamente con otro vértice.

Para ahorita

- P: ¿Cómo generamos un grafo aleatorio con exactamente m aristas?
- R: Depende: si m es chico, ponemos aristas aleatorias hasta juntar m. Si m es grande, tomamos la lista de todas las aristas y las vamos escogiendo y quitando de una por una.
- Programar: Erdős-Renyi y Erdős-Renyi con exactamente m aristas.
- P: ¿Cómo generamos un árbol aleatorio?
- R: Empieza con un vértice y ve agregando vértices de uno por uno, conectando exactamente con otro vértice.
- ¡Hay que programar esto!

Para ahorita

- P: ¿Cómo generamos un grafo aleatorio con exactamente m aristas?
- R: Depende: si m es chico, ponemos aristas aleatorias hasta juntar m. Si m es grande, tomamos la lista de todas las aristas y las vamos escogiendo y quitando de una por una.
- Programar: Erdős-Renyi y Erdős-Renyi con exactamente m aristas.
- P: ¿Cómo generamos un árbol aleatorio?
- R: Empieza con un vértice y ve agregando vértices de uno por uno, conectando exactamente con otro vértice.
- ¡Hay que programar esto!
- P*: ¿Cómo produzco árboles binarios?

Para ahorita

- P: ¿Cómo generamos un grafo aleatorio con exactamente m aristas?
- R: Depende: si m es chico, ponemos aristas aleatorias hasta juntar m. Si m es grande, tomamos la lista de todas las aristas y las vamos escogiendo y quitando de una por una.
- Programar: Erdős-Renyi y Erdős-Renyi con exactamente m aristas.
- P: ¿Cómo generamos un árbol aleatorio?
- R: Empieza con un vértice y ve agregando vértices de uno por uno, conectando exactamente con otro vértice.
- ¡Hay que programar esto!
- P*: ¿Cómo produzco árboles binarios?
- R: Tarea!

Clustering Global

Definición

Dado un grafo G, su clustering global $cl(G)$ es su proporción de triángulos formados entre triángulos posibles.

Clustering Global

Definición

Dado un grafo G, su clustering global $cl(G)$ es su proporción de triángulos formados entre triángulos posibles. Es decir,

$$cl(G) = \frac{3\#\ \text{de}\ \triangle}{\#\ \text{de}\ \wedge}$$

Clustering Global

Definición

Dado un grafo G, su clustering global $cl(G)$ es su proporción de triángulos formados entre triángulos posibles. Es decir,

$$cl(G) = \frac{3\#\ \textit{de}\ \triangle}{\#\ \textit{de}\ \wedge}$$

- El 3 es porque cada $\triangle$ tiene 3 $\wedge$'s.

Clustering Global

Definición

Dado un grafo G, su clustering global $cl(G)$ es su proporción de triángulos formados entre triángulos posibles. Es decir,

$$cl(G) = \frac{3\# \text{ de } \triangle}{\# \text{ de } \wedge}$$

- El 3 es porque cada $\triangle$ tiene 3 $\wedge$'s.
- Si los grados de los vértices son $d_1, d_2, ..., d_n$, entonces

$$\# \text{ de } \wedge\text{'s} = \binom{d_1}{2} + \binom{d_2}{2} + ... + \binom{d_n}{2}$$

Clustering: Ejemplos

¿Cuánto vale el clustering en los siguientes grafos?

- K_n:

Clustering: Ejemplos

¿Cuánto vale el clustering en los siguientes grafos?

- K_n: 1

Clustering: Ejemplos

¿Cuánto vale el clustering en los siguientes grafos?

- K_n: 1
- C_n:

Clustering: Ejemplos

¿Cuánto vale el clustering en los siguientes grafos?

- K_n: 1
- C_n: 0 si $n \neq 3$, 1 si $n = 3$

Clustering: Ejemplos

¿Cuánto vale el clustering en los siguientes grafos?

- K_n: 1
- C_n: 0 si $n \neq 3$, 1 si $n = 3$
- $K_{n,m}$:

Clustering: Ejemplos

¿Cuánto vale el clustering en los siguientes grafos?

- K_n: 1
- C_n: 0 si $n \neq 3$, 1 si $n = 3$
- $K_{n,m}$: 0

Clustering: Ejemplos

¿Cuánto vale el clustering en los siguientes grafos?

- K_n: 1
- C_n: 0 si $n \neq 3$, 1 si $n = 3$
- $K_{n,m}$: 0
- K_4-arista:

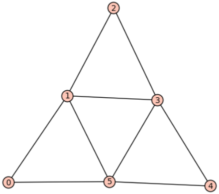
Clustering: Ejemplos

¿Cuánto vale el clustering en los siguientes grafos?

- K_n: 1
- C_n: 0 si $n \neq 3$, 1 si $n = 3$
- $K_{n,m}$: 0
- K_4-arista: $6/8$

Clustering: Ejemplos

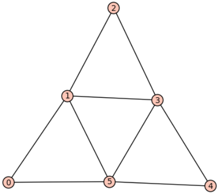
¿Cuánto vale el clustering en los siguientes grafos?

- K_n: 1
- C_n: 0 si $n \neq 3$, 1 si $n = 3$
- $K_{n,m}$: 0
- K_4-arista: $6/8$
- 

:

Clustering: Ejemplos

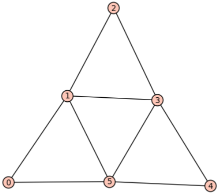
¿Cuánto vale el clustering en los siguientes grafos?

- K_n: 1
- C_n: 0 si $n \neq 3$, 1 si $n = 3$
- $K_{n,m}$: 0
- K_4-arista: $6/8$
- 

: $12/15$

Clustering: Ejemplos

¿Cuánto vale el clustering en los siguientes grafos?

- K_n: 1
- C_n: 0 si $n \neq 3$, 1 si $n = 3$
- $K_{n,m}$: 0
- K_4-arista: $6/8$
- 

: $12/15$
- En promedio, en $ER(n,p)$: p

Clustering local

Definición

Dado un grafo G y un vértice v tal que $\delta(v) \geq 2$. Su Clustering local $cl(G, v)$ es la proporción de sus parejas de vecinos que forman triángulo.

Clustering local

Definición

Dado un grafo G y un vértice v tal que $\delta(v) \geq 2$. Su Clustering local $cl(G, v)$ es la proporción de sus parejas de vecinos que forman triángulo. Es decir,

$$cl(G, v) = \frac{\#\text{ aristas } x - y \text{ con } x, y \text{ vecinos de } v}{\binom{\delta(v)}{2}}$$

Clustering: Ejemplos

¿Cuánto vale el clustering local en los siguientes grafos?

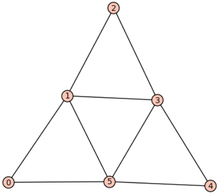
Clustering: Ejemplos

¿Cuánto vale el clustering local en los siguientes grafos?

- K_4-arista:

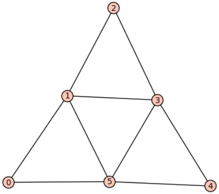
Clustering: Ejemplos

¿Cuánto vale el clustering local en los siguientes grafos?

- K_4-arista: 1 para 2 de ellos y $2/3$ para los otros dos.
- 
:

Clustering: Ejemplos

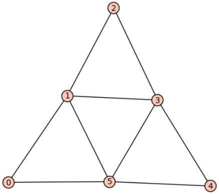
¿Cuánto vale el clustering local en los siguientes grafos?

- K_4-arista: 1 para 2 de ellos y $2/3$ para los otros dos.
- 

: 1 y $1/2$ para pares, impares respectivamente.
- En promedio, en $ER(n,p)$:

Clustering: Ejemplos

¿Cuánto vale el clustering local en los siguientes grafos?

- K_4-arista: 1 para 2 de ellos y $2/3$ para los otros dos.
- 

: 1 y $1/2$ para pares, impares respectivamente.
- En promedio, en $ER(n, p)$: p

¿Cómo se relacionan el clustering global y el local?

- Pues el clustering global es un promedio pesado del local de los vértices, donde el peso de un vértice v es $\binom{\delta(v)}{2}$
- Lo cual quiere decir que

$$\min_v(cl(G,v)) \leq cl(G) \leq \max_v(cl(G,v))$$

- También se habla del clustering average, que es simplemente el promedio normal del clustering de los vértices.

Índice:

Introducción

- Vimos que muchas veces tendía a haber distribuciones de vértices más “extremas” que en ER.

Introducción

- Vimos que muchas veces tendía a haber distribuciones de vértices más "extremas" que en ER.
- A continuación veremos varios modelos que producen distribuciones así, o que producen más triángulos, etc.

Albert Barabanasi

Otro modelo:

Albert Barabanasi

Otro modelo:

- Comenzamos poniendo los vértices de uno por uno. Comenzamos con k vértices y sin aristas.

Albert Barabanasi

Otro modelo:

- Comenzamos poniendo los vértices de uno por uno. Comenzamos con k vértices y sin aristas.
- Después, cada vértice nuevo lo unimos con otros k, escogidos aleatoriamente.

Albert Barabanasi

Otro modelo:

- Comenzamos poniendo los vértices de uno por uno. Comenzamos con k vértices y sin aristas.
- Después, cada vértice nuevo lo unimos con otros k, escogidos aleatoriamente.
- Eso lo que nos da es que los vértices más "nuevos" tendrán menos aristas, y los más "viejos" tendrán más aristas.

Albert Barabanasi

Otro modelo:

- Comenzamos poniendo los vértices de uno por uno. Comenzamos con k vértices y sin aristas.
- Después, cada vértice nuevo lo unimos con otros k, escogidos aleatoriamente.
- Eso lo que nos da es que los vértices más "nuevos" tendrán menos aristas, y los más "viejos" tendrán más aristas.
- ¿Cuántas aristas tendrá un vértice que nació en tiempo $m > k$, en promedio, después de n pasos?

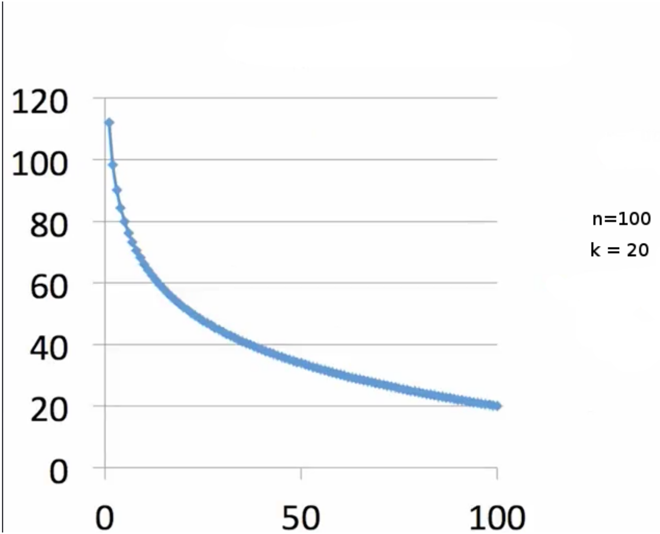
Albert Barabanasi

Otro modelo:

- Comenzamos poniendo los vértices de uno por uno. Comenzamos con k vértices y sin aristas.
- Después, cada vértice nuevo lo unimos con otros k, escogidos aleatoriamente.
- Eso lo que nos da es que los vértices más "nuevos" tendrán menos aristas, y los más "viejos" tendrán más aristas.
- ¿Cuántas aristas tendrá un vértice que nació en tiempo $m > k$, en promedio, después de n pasos?

$$k + \frac{k}{m} + \frac{k}{m+1} + \frac{k}{m+2} + ... + \frac{k}{m} \approx k(1 + \ln(n/m))$$

Histogramas



Preferential Attachment

- Ahora haremos lo siguiente: Al agregar un vértice, en vez de escoger aleatoriamente los puntos, le daremos preferencia a los puntos con grado muy alto:

Preferential Attachment

- Ahora haremos lo siguiente: Al agregar un vértice, en vez de escoger aleatoriamente los puntos, le daremos preferencia a los puntos con grado muy alto:
- Es decir, la probabilidad de formar una liga con un vértice v será proporcional al grado de v.

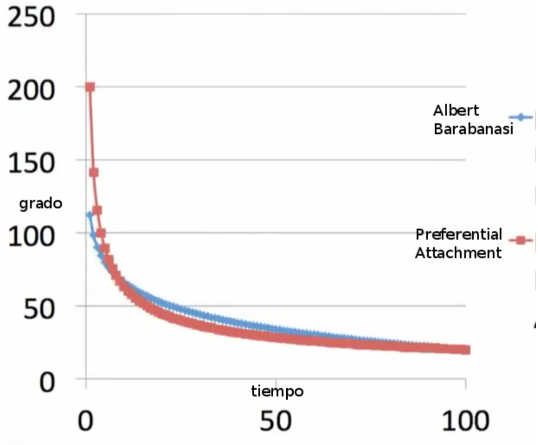
Preferential Attachment

- Ahora haremos lo siguiente: Al agregar un vértice, en vez de escoger aleatoriamente los puntos, le daremos preferencia a los puntos con grado muy alto:
- Es decir, la probabilidad de formar una liga con un vértice v será proporcional al grado de v.
- En total, en un tiempo n hay $\approx kn$ aristas (pues cada vértice nuevo agrega k aristas)

Preferential Attachment

- Ahora haremos lo siguiente: Al agregar un vértice, en vez de escoger aleatoriamente los puntos, le daremos preferencia a los puntos con grado muy alto:
- Es decir, la probabilidad de formar una liga con un vértice v será proporcional al grado de v.
- En total, en un tiempo n hay $\approx kn$ aristas (pues cada vértice nuevo agrega k aristas)
- Así que la probabilidad de formar una arista con un vértice de grado d es $\frac{d}{2kn}$.

Preferential Attachment



Amigos de Amigos

- Ahora haremos lo siguiente: Cada nuevo nodo se une a k otros nodos, y después se une a un amigo aleatorio de cada uno de los nuevos nodos que se unió.

Amigos de Amigos

- Ahora haremos lo siguiente: Cada nuevo nodo se une a k otros nodos, y después se une a un amigo aleatorio de cada uno de los nuevos nodos que se unió.
- Es decir, es probable que te hagas amigo de tus amigos.

Amigos de Amigos

- Ahora haremos lo siguiente: Cada nuevo nodo se une a k otros nodos, y después se une a un amigo aleatorio de cada uno de los nuevos nodos que se unió.
- Es decir, es probable que te hagas amigo de tus amigos.
- Podemos ponerle varios parámetros, como la probabilidad de unirte a cada amigo, o incluso unirte a amigos de amigos de amigos.

Amigos de Amigos

- Ahora haremos lo siguiente: Cada nuevo nodo se une a k otros nodos, y después se une a un amigo aleatorio de cada uno de los nuevos nodos que se unió.
- Es decir, es probable que te hagas amigo de tus amigos.
- Podemos ponerle varios parámetros, como la probabilidad de unirte a cada amigo, o incluso unirte a amigos de amigos de amigos.
- Se vuelve muy complicado el análisis teórico, pero se pueden hacer simulaciones.

Amigos de Amigos

- Ahora haremos lo siguiente: Cada nuevo nodo se une a k otros nodos, y después se une a un amigo aleatorio de cada uno de los nuevos nodos que se unió.
- Es decir, es probable que te hagas amigo de tus amigos.
- Podemos ponerle varios parámetros, como la probabilidad de unirte a cada amigo, o incluso unirte a amigos de amigos de amigos.
- Se vuelve muy complicado el análisis teórico, pero se pueden hacer simulaciones.
- Hay muchos modelos más que se forman nodo por nodo en base a lo que ya se formó: Walter-Strogatz, Jackson-Rogers, etc.