FACULTAD DE INGENIERÍA

UNIVERSIDAD DE LA REPÚBLICA URUGUAY

Aprendizaje Automático para Datos en Grafos

Modelos de Grafos Aleatorios - Parte I

Federico 'Larroca' La Rocca
Muy basado en transparencias de **Gonzalo Mateos**

flarroca@fing.edu.uy
http://iie.fing.edu.uy/personal/flarroca

Modelos de Grafos Aleatorios

¿Para qué modelar estadísticamente grafos?

- Los modelos de grafos aleatorios tienen varios usos:
  1) ¿Qué mecanismos pueden explicar propiedades que observamos en grafos reales?
     Ejemplo: los efectos tipo small-world, o distribuciones de grado con ley de potencia

¿Para qué modelar estadísticamente grafos?

- Los modelos de grafos aleatorios tienen varios usos:
  1) ¿Qué mecanismos pueden explicar propiedades que observamos en grafos reales?
     Ejemplo: los efectos tipo small-world, o distribuciones de grado con ley de potencia
  2) Crear un modelo nulo para testear propiedades de un grafo
     Ejemplo: ¿El valor del coeficiente de clustering de mi grafo es raro?

¿Para qué modelar estadísticamente grafos?

- Los modelos de grafos aleatorios tienen varios usos:
  1) ¿Qué mecanismos pueden explicar propiedades que observamos en grafos reales?
     Ejemplo: los efectos tipo small-world, o distribuciones de grado con ley de potencia
  2) Crear un modelo nulo para testear propiedades de un grafo
     Ejemplo: ¿El valor del coeficiente de clustering de mi grafo es raro?
  3) Evaluación de factores que puedan predecir relaciones entre los nodos
     Ejemplo: ¿En este grafo hay efectos transitivos o de reciprocidad?

¿Para qué modelar estadísticamente grafos?

- Los modelos de grafos aleatorios tienen varios usos:
  1) ¿Qué mecanismos pueden explicar propiedades que observamos en grafos reales?
     Ejemplo: los efectos tipo small-world, o distribuciones de grado con ley de potencia
  2) Crear un modelo nulo para testear propiedades de un grafo
     Ejemplo: ¿El valor del coeficiente de clustering de mi grafo es raro?
  3) Evaluación de factores que puedan predecir relaciones entre los nodos
     Ejemplo: ¿En este grafo hay efectos transitivos o de reciprocidad?
  4) Generación de grafos
     Ejemplo: ¿Grafos para evaluar mi algoritmo de aprendizaje en condiciones "controladas"?

¿Para qué modelar estadísticamente grafos?

- Los modelos de grafos aleatorios tienen varios usos:
  1) ¿Qué mecanismos pueden explicar propiedades que observamos en grafos reales?
     Ejemplo: los efectos tipo small-world, o distribuciones de grado con ley de potencia
  2) Crear un modelo nulo para testear propiedades de un grafo
     Ejemplo: ¿El valor del coeficiente de clustering de mi grafo es raro?
  3) Evaluación de factores que puedan predecir relaciones entre los nodos
     Ejemplo: ¿En este grafo hay efectos transitivos o de reciprocidad?
  4) Generación de grafos
     Ejemplo: ¿Grafos para evaluar mi algoritmo de aprendizaje en condiciones "controladas"?
- Nos vamos a enfocar en la construcción de varios de los modelos más populares, algunas propiedades y su uso. Quedará mucho por el camino (ver el EVA para más material).

Modelos de Grafos Aleatorios

- **Definición:** Un modelo de grafos aleatorios es una colección

$$\{\mathrm{P}_\theta(G), G \in \mathcal{G} : \theta \in \Theta\}$$

  - $\mathcal{G}$ es el conjunto de grafos posibles bajo este modelo
  - $\mathrm{P}_\theta(\cdot)$ es la distribución de probabilidad sobre $\mathcal{G}$ (casi siempre va a figurar como $\mathrm{P}(\cdot)$ directamente)
  - θ es el vector de parámetros del modelo, perteneciente a cierto espacio Θ

Modelos de Grafos Aleatorios

- **Definición:** Un modelo de grafos aleatorios es una colección

$$\{\mathrm{P}_\theta(G), G \in \mathcal{G} : \theta \in \Theta\}$$

  - $\mathcal{G}$ es el conjunto de grafos posibles bajo este modelo
  - $\mathrm{P}_\theta(\cdot)$ es la distribución de probabilidad sobre $\mathcal{G}$ (casi siempre va a figurar como $\mathrm{P}(\cdot)$ directamente)
  - θ es el vector de parámetros del modelo, perteneciente a cierto espacio Θ

- La riqueza y la utilidad del modelo dependen de cómo especifiquemos $\mathrm{P}_\theta(\cdot)$

  $\Rightarrow$ Como veremos en estas clases, los métodos van de lo simple a lo complejo

Especificación del modelo

1) $\mathrm{P}(\cdot)$ uniforme en $\mathcal{G}$, tomando en cuenta restricciones estructurales sobre los grafos en $\mathcal{G}$

   Ejemplo: Grafos tipo Erdös-Renyi o el denominado modelo Configurational

Especificación del modelo

1) $\mathrm{P}(\cdot)$ uniforme en $\mathcal{G}$, tomando en cuenta restricciones estructurales sobre los grafos en $\mathcal{G}$

   Ejemplo: Grafos tipo Erdös-Renyi o el denominado modelo Configurational

2) Inducir $\mathrm{P}(\cdot)$ a través de la aplicación de un modelo generativo

   Ejemplo: small world, preferential attachment, copying models, deep generative models

Especificación del modelo

1) $\mathrm{P}(\cdot)$ uniforme en $\mathcal{G}$, tomando en cuenta restricciones estructurales sobre los grafos en $\mathcal{G}$
   Ejemplo: Grafos tipo Erdös-Renyi o el denominado modelo Configurational
2) Inducir $\mathrm{P}(\cdot)$ a través de la aplicación de un modelo generativo
   Ejemplo: small world, preferential attachment, copying models, deep generative models
3) Modelar propiedades estructurales y sus efectos sobre la topología de G
   Ejemplo: exponential random graph models

Especificación del modelo

1) P(·) uniforme en $\mathcal{G}$, tomando en cuenta restricciones estructurales sobre los grafos en $\mathcal{G}$

   Ejemplo: Grafos tipo Erdös-Renyi o el denominado modelo Configurational

2) Inducir P(·) a través de la aplicación de un modelo generativo

   Ejemplo: small world, preferential attachment, copying models, deep generative models

3) Modelar propiedades estructurales y sus efectos sobre la topología de G

   Ejemplo: exponential random graph models

4) Modelar la tendencia de los nodos a conectarse mediante variables latentes

   Ejemplo: stochastic block models, graphons, random dot product graphs

Especificación del modelo

1) $\mathrm{P}(\cdot)$ uniforme en $\mathcal{G}$, tomando en cuenta restricciones estructurales sobre los grafos en $\mathcal{G}$
   Ejemplo: Grafos tipo Erdös-Renyi o el denominado modelo Configurational
2) Inducir $\mathrm{P}(\cdot)$ a través de la aplicación de un modelo generativo
   Ejemplo: small world, preferential attachment, copying models, deep generative models
3) Modelar propiedades estructurales y sus efectos sobre la topología de G
   Ejemplo: exponential random graph models
4) Modelar la tendencia de los nodos a conectarse mediante variables latentes
   Ejemplo: stochastic block models, graphons, random dot product graphs

- Costo computacional de algoritmos de inferencia y generación son aspectos importantes

Modelos de Grafos Aleatorios

Grafos Aleatorios Clásicos

- Se le asigna la misma probabilidad a todos los grafos (no-dirigidos) de cierto orden y tamaño

Grafos Aleatorios Clásicos

- Se le asigna la misma probabilidad a todos los grafos (no-dirigidos) de cierto orden y tamaño
  - Se especifica la colección $\mathcal{G}_{N_v,N_e}$ como aquellos grafos $G(V,E)$ con $|V| = N_v$, $|E| = N_e$
  - $\mathrm{P}(G) = \binom{N}{N_e}^{-1}$ para cada $G \in \mathcal{G}_{N_v,N_e}$, donde $N = |V^{(2)}| = \binom{N_v}{2}$

Grafos Aleatorios Clásicos

- Se le asigna la misma probabilidad a todos los grafos (no-dirigidos) de cierto orden y tamaño
  - Se especifica la colección $\mathcal{G}_{N_v,N_e}$ como aquellos grafos $G(V,E)$ con $|V| = N_v$, $|E| = N_e$
  - $\mathrm{P}(G) = \binom{N}{N_e}^{-1}$ para cada $G \in \mathcal{G}_{N_v,N_e}$, donde $N = |V^{(2)}| = \binom{N_v}{2}$
- La variante más común es conocida como modelo de Erdös-Renyi(-Gilbert) y la notaremos como $\mathrm{ER}(n,p)$ (también se usa $G_{n,p}$)

  $\Rightarrow$ Grafo no dirigido con $N_v = n$ nodos (igual que antes)

  $\Rightarrow$ La arista (u,v) existe con probabilidad p independiente del resto

Grafos Aleatorios Clásicos

- Se le asigna la misma probabilidad a todos los grafos (no-dirigidos) de cierto orden y tamaño
  - Se especifica la colección $\mathcal{G}_{N_v,N_e}$ como aquellos grafos $G(V,E)$ con $|V| = N_v$, $|E| = N_e$
  - $\mathrm{P}(G) = \binom{N}{N_e}^{-1}$ para cada $G \in \mathcal{G}_{N_v,N_e}$, donde $N = |V^{(2)}| = \binom{N_v}{2}$
- La variante más común es conocida como modelo de Erdös-Renyi(-Gilbert) y la notaremos como $\mathrm{ER}(n,p)$ (también se usa $G_{n,p}$)

  $\Rightarrow$ Grafo no dirigido con $N_v = n$ nodos (igual que antes)

  $\Rightarrow$ La arista (u,v) existe con probabilidad p independiente del resto
- Simulación: sortear $N = \binom{N_v}{2} \approx N_v^2/2$ i.i.d. $\mathrm{Ber}(p)$ VAs
  - Claramente ineficiente cuando $p \sim N_v^{-1}$ $\Rightarrow$ grafo esparso, la mayoría de las veces vamos a sacar un 0
  - Mejor sortear $\mathrm{Geo}(p)$ i.i.d. para ver qué aristas generar (corre en tiempo $O(N_v + N_e)$)

Propiedades de $\mathrm{ER}(n,p)$

- $\mathrm{ER}(n,p)$ es un modelo ampliamente estudiado, y se conocen expresiones para varios indicadores. Algunas propiedades interesantes:

Propiedades de ER(n, p)

- ER(n, p) es un modelo ampliamente estudiado, y se conocen expresiones para varios indicadores. Algunas propiedades interesantes:

P1) Distribución de grados ¿Cuánto vale $P(D_i = d) = P(d)$, la probabilidad de que el nodo i tenga d vecinos?

Propiedades de $\mathrm{ER}(n,p)$

- $\mathrm{ER}(n,p)$ es un modelo ampliamente estudiado, y se conocen expresiones para varios indicadores. Algunas propiedades interesantes:

P1) Distribución de grados ¿Cuánto vale $P(D_i = d) = P(d)$, la probabilidad de que el nodo i tenga d vecinos?

$$\left.\begin{array}{r} D_i = \displaystyle\sum_{\substack{j=1,\ldots,n \\ j\neq i}} A_{i,j} \\ (A_{i,j})_{j\neq i} \sim \mathrm{Ber}(p)\,\text{i.i.d.} \end{array}\right\} \Rightarrow D_i \sim \mathrm{Bin}(n-1,p)$$

Propiedades de ER(n, p)

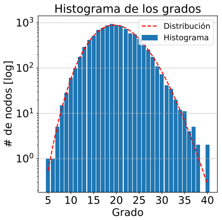
- ER(n, p) es un modelo ampliamente estudiado, y se conocen expresiones para varios indicadores. Algunas propiedades interesantes:

P1) Distribución de grados ¿Cuánto vale $P(D_i = d) = P(d)$, la probabilidad de que el nodo i tenga d vecinos?

$$\left.\begin{array}{r} D_i = \displaystyle\sum_{\substack{j=1,\dots,n \\ j\neq i}} A_{i,j} \\ (A_{i,j})_{j\neq i} \sim \text{Ber}(p)\,\text{i.i.d.} \end{array}\right\} \Rightarrow D_i \sim \text{Bin}(n-1, p) \underset{\text{Hoeffding}}{\Rightarrow} P(D_i - p(n-1) \geq t) \leq e^{-\frac{2t^2}{n-1}}$$

$\Rightarrow$ P (d) está concentrado alrededor de $p(n-1)$ con colas exponenciales

Ejemplo Histograma de grados de un ER(n, p) con $n = 10{,}000$ y $p = 20/10{,}000$



Propiedades de $\mathrm{ER}(n, p)$

P2) Transición de fase en la aparición de una componente gigante

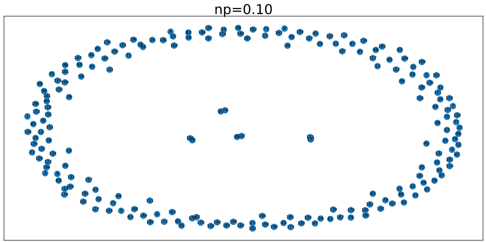
- Si $np > 1$, $\mathrm{ER}(n, p)$ tiene una componente gigante de tamaño $O(n)$ w.h.p.
- Si $np < 1$, $\mathrm{ER}(n, p)$ tienen componentes de tamaño $O(\log n)$ w.h.p.

Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Si $np > 1$, ER(n, p) tiene una componente gigante de tamaño $O(n)$ w.h.p.
- Si $np < 1$, ER(n, p) tienen componentes de tamaño $O(\log n)$ w.h.p.

Ejemplo ER(n, p) para $n = 200$ a medida que cambia p

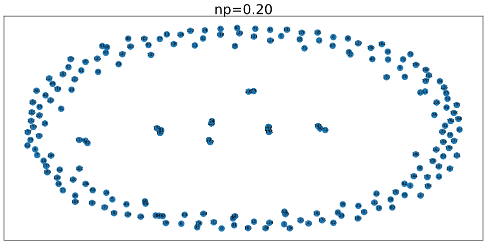


Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Si $np > 1$, ER(n, p) tiene una componente gigante de tamaño $O(n)$ w.h.p.
- Si $np < 1$, ER(n, p) tienen componentes de tamaño $O(\log n)$ w.h.p.

Ejemplo ER(n, p) para $n = 200$ a medida que cambia p

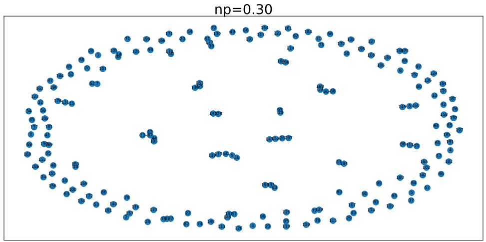


Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Si $np > 1$, ER(n, p) tiene una componente gigante de tamaño $O(n)$ w.h.p.
- Si $np < 1$, ER(n, p) tienen componentes de tamaño $O(\log n)$ w.h.p.

Ejemplo ER(n, p) para $n = 200$ a medida que cambia p

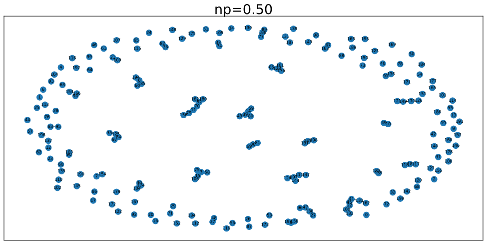


Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Si $np > 1$, ER(n, p) tiene una componente gigante de tamaño $O(n)$ w.h.p.
- Si $np < 1$, ER(n, p) tienen componentes de tamaño $O(\log n)$ w.h.p.

Ejemplo ER(n, p) para $n = 200$ a medida que cambia p

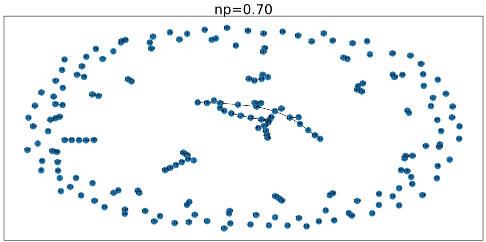


Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Si $np > 1$, ER(n, p) tiene una componente gigante de tamaño $O(n)$ w.h.p.
- Si $np < 1$, ER(n, p) tienen componentes de tamaño $O(\log n)$ w.h.p.

Ejemplo ER(n, p) para $n = 200$ a medida que cambia p

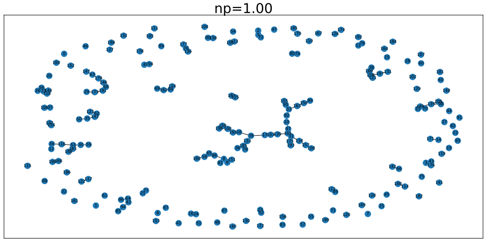


Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Si $np > 1$, ER(n, p) tiene una componente gigante de tamaño $O(n)$ w.h.p.
- Si $np < 1$, ER(n, p) tienen componentes de tamaño $O(\log n)$ w.h.p.

Ejemplo ER(n, p) para $n = 200$ a medida que cambia p

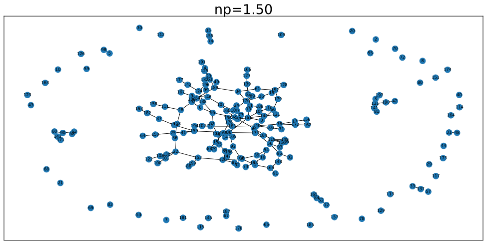


Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Si $np > 1$, ER(n, p) tiene una componente gigante de tamaño $O(n)$ w.h.p.
- Si $np < 1$, ER(n, p) tienen componentes de tamaño $O(\log n)$ w.h.p.

Ejemplo ER(n, p) para $n = 200$ a medida que cambia p

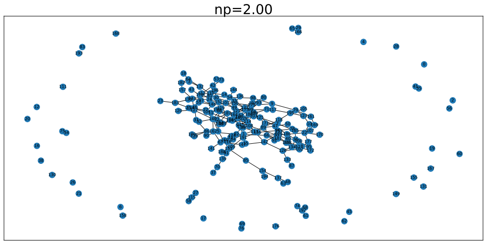


Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Si $np > 1$, ER(n, p) tiene una componente gigante de tamaño $O(n)$ w.h.p.
- Si $np < 1$, ER(n, p) tienen componentes de tamaño $O(\log n)$ w.h.p.

Ejemplo ER(n, p) para $n = 200$ a medida que cambia p

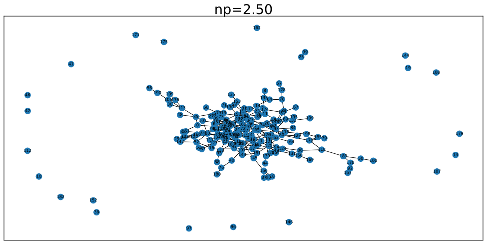


Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Si $np > 1$, ER(n, p) tiene una componente gigante de tamaño $O(n)$ w.h.p.
- Si $np < 1$, ER(n, p) tienen componentes de tamaño $O(\log n)$ w.h.p.

Ejemplo ER(n, p) para $n = 200$ a medida que cambia p

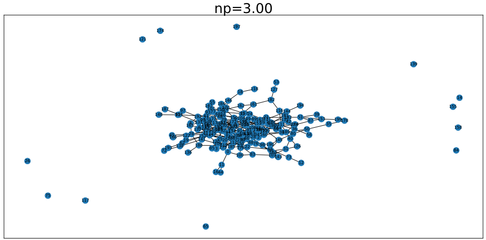


Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Si $np > 1$, ER(n, p) tiene una componente gigante de tamaño $O(n)$ w.h.p.
- Si $np < 1$, ER(n, p) tienen componentes de tamaño $O(\log n)$ w.h.p.

Ejemplo ER(n, p) para $n = 200$ a medida que cambia p

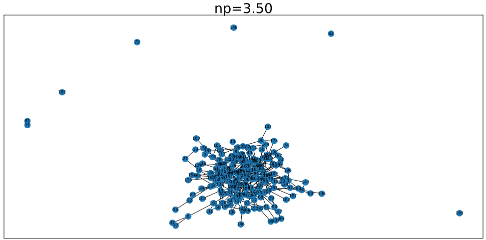


Propiedades de $\mathrm{ER}(n, p)$

P2) Transición de fase en la aparición de una componente gigante

- Si $np > 1$, $\mathrm{ER}(n, p)$ tiene una componente gigante de tamaño $O(n)$ w.h.p.
- Si $np < 1$, $\mathrm{ER}(n, p)$ tienen componentes de tamaño $O(\log n)$ w.h.p.

Ejemplo $\mathrm{ER}(n, p)$ para $n = 200$ a medida que cambia p

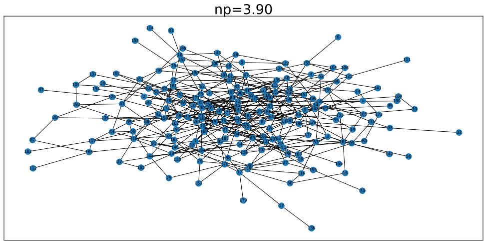


Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Si $np > 1$, ER(n, p) tiene una componente gigante de tamaño $O(n)$ w.h.p.
- Si $np < 1$, ER(n, p) tienen componentes de tamaño $O(\log n)$ w.h.p.

Ejemplo ER(n, p) para $n = 200$ a medida que cambia p



Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- ¿Cómo podemos demostrarlo?
- Definamos un proceso de exploración en el grafo que recorra la componente conexa arrancando de algún nodo al azar y veamos con qué tamaño termina

Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- ¿Cómo podemos demostrarlo?
- Definamos un proceso de exploración en el grafo que recorra la componente conexa arrancando de algún nodo al azar y veamos con qué tamaño termina
- Nodos en uno de tres estados: activo ($\mathcal{A}_t$), inactivo ($\mathcal{I}_t$) o explorado ($\mathcal{E}_t$)

Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- ¿Cómo podemos demostrarlo?
- Definamos un proceso de exploración en el grafo que recorra la componente conexa arrancando de algún nodo al azar y veamos con qué tamaño termina
- Nodos en uno de tres estados: activo ($\mathcal{A}_t$), inactivo ($\mathcal{I}_t$) o explorado ($\mathcal{E}_t$)

$G = (V, E) \sim \text{ER}(n, p)$, $V = 1, \ldots, n$ % Genero un grafo ER
$\mathcal{A}_0 = \{1\}, \mathcal{I}_0 = \{2, \ldots, n\}, \mathcal{E}_0 = \emptyset$ % Inicialización: un activo cualquiera
while $\mathcal{A}_t \neq \emptyset$ **do**
  Esperar tiempo exponencial de media 1
  Elegir $v \in \mathcal{A}_t$ % Explorar algún nodo activo
  foreach $w \in \mathcal{I}_t \cap \mathcal{N}(v)$ **do**
    $\mathcal{A}_t \leftarrow \mathcal{A}_t \cup \{w\}$ % Agrego vecinos inactivos de v
  end
  $\mathcal{E}_t \leftarrow \mathcal{E}_t \cup \{v\}$ % Paso v a los explorados
  $\mathcal{A}_t \leftarrow \mathcal{A}_t \setminus \{v\}$
end
return $|\mathcal{E}_t|$

Propiedades de $\mathrm{ER}(n, p)$

P2) Transición de fase en la aparición de una componente gigante

- Nos interesa $|\mathcal{E}_t| = E_t$. Junto con $|\mathcal{A}_t| = A_t$ forman una cadena de Markov en tiempo continuo:

Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Nos interesa $|\mathcal{E}_t| = E_t$. Junto con $|\mathcal{A}_t| = A_t$ forman una cadena de Markov en tiempo continuo:
  - El vector (E_t, A_t) permanece en cada estado durante un tiempo exponencial
  - En los momentos que hay cambio:

$$(E_{t+}, A_{t+}) = \Big(\underbrace{E_t + 1}_{\substack{\text{Un explorado} \\ \text{más } (v)}} , \underbrace{A_t + X_t - 1}_{\substack{\text{Los vecinos inactivos} \\ \text{de } v \text{ ahora son activos} \\ \text{(y } v \text{ ahora es explorado)}}} \Big)$$

Propiedades de $\mathrm{ER}(n,p)$

P2) Transición de fase en la aparición de una componente gigante

- Nos interesa $|\mathcal{E}_t| = E_t$. Junto con $|\mathcal{A}_t| = A_t$ forman una cadena de Markov en tiempo continuo:
  - El vector (E_t, A_t) permanece en cada estado durante un tiempo exponencial
  - En los momentos que hay cambio:

$$(E_{t+}, A_{t+}) = \Big(\underbrace{E_t + 1}_{\substack{\text{Un explorado} \\ \text{más } (v)}} , \underbrace{A_t + X_t - 1}_{\substack{\text{Los vecinos inactivos} \\ \text{de } v \text{ ahora son activos} \\ (\text{y } v \text{ ahora es explorado})}} \Big)$$

$$X_t \sim \mathrm{Bin}(|\mathcal{I}_t|, p) = \mathrm{Bin}(n - A_t - E_t, p)$$

Propiedades de ER(n,p)

P2) Transición de fase en la aparición de una componente gigante

- Nos interesa $|\mathcal{E}_t| = E_t$. Junto con $|\mathcal{A}_t| = A_t$ forman una cadena de Markov en tiempo continuo:
  - El vector (E_t, A_t) permanece en cada estado durante un tiempo exponencial
  - En los momentos que hay cambio:

$$(E_{t^+}, A_{t^+}) = \Big(\underbrace{E_t + 1}_{\substack{\text{Un explorado} \\ \text{más } (v)}} \;,\; \underbrace{A_t + X_t - 1}_{\substack{\text{Los vecinos inactivos} \\ \text{de } v \text{ ahora son activos} \\ (\text{y } v \text{ ahora es explorado})}} \Big)$$

$$X_t \sim \text{Bin}(|\mathcal{I}_t|, p) = \text{Bin}(n - A_t - E_t, p)$$

- ¿Cuánto vale E_t cuando A_t se hace cero (a medida que $n \to \infty$)?

Propiedades de ER(n, p)

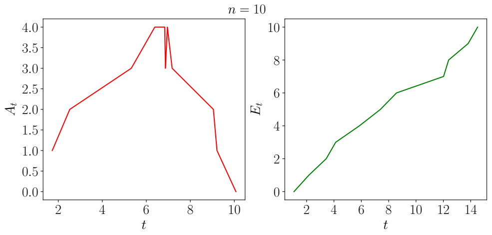
P2) Transición de fase en la aparición de una componente gigante

- Nos interesa $|\mathcal{E}_t| = E_t$. Junto con $|\mathcal{A}_t| = A_t$ forman una cadena de Markov en tiempo continuo:
  - El vector (E_t, A_t) permanece en cada estado durante un tiempo exponencial
  - En los momentos que hay cambio:

$$(E_{t+}, A_{t+}) = \Big(\underbrace{E_t + 1}_{\text{Un explorado más } (v)} \ , \ \underbrace{A_t + X_t - 1}_{\text{Los vecinos inactivos de } v \text{ ahora son activos (y } v \text{ ahora es explorado)}} \Big)$$

$$X_t \sim \text{Bin}(|\mathcal{I}_t|, p) = \text{Bin}(n - A_t - E_t, p)$$

- ¿Cuánto vale E_t cuando A_t se hace cero (a medida que $n \to \infty$)? Ej: $p = 2/n$



Propiedades de $\mathrm{ER}(n,p)$

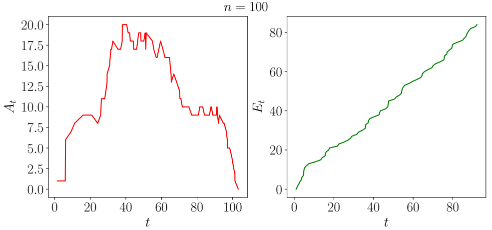
P2) Transición de fase en la aparición de una componente gigante

- Nos interesa $|\mathcal{E}_t| = E_t$. Junto con $|\mathcal{A}_t| = A_t$ forman una cadena de Markov en tiempo continuo:
  - El vector (E_t, A_t) permanece en cada estado durante un tiempo exponencial
  - En los momentos que hay cambio:

$$(E_{t+}, A_{t+}) = \Big(\underbrace{E_t + 1}_{\text{Un explorado más } (v)},\ \underbrace{A_t + X_t - 1}_{\substack{\text{Los vecinos inactivos} \\ \text{de } v \text{ ahora son activos} \\ \text{(y } v \text{ ahora es explorado)}}}\Big)$$

$$X_t \sim \mathrm{Bin}(|\mathcal{I}_t|, p) = \mathrm{Bin}(n - A_t - E_t, p)$$

- ¿Cuánto vale E_t cuando A_t se hace cero (a medida que $n \to \infty$)? Ej: $p = 2/n$



Propiedades de $\mathrm{ER}(n,p)$

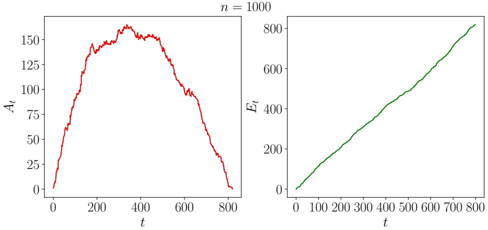
P2) Transición de fase en la aparición de una componente gigante

- Nos interesa $|\mathcal{E}_t| = E_t$. Junto con $|\mathcal{A}_t| = A_t$ forman una cadena de Markov en tiempo continuo:
  - El vector (E_t, A_t) permanece en cada estado durante un tiempo exponencial
  - En los momentos que hay cambio:

$$(E_{t+}, A_{t+}) = \Big(\underbrace{E_t + 1}_{\substack{\text{Un explorado}\\ \text{más } (v)}},\ \underbrace{A_t + X_t - 1}_{\substack{\text{Los vecinos inactivos}\\ \text{de } v \text{ ahora son activos}\\ (\text{y } v \text{ ahora es explorado})}}\Big)$$

$$X_t \sim \mathrm{Bin}(|\mathcal{I}_t|, p) = \mathrm{Bin}(n - A_t - E_t, p)$$

- ¿Cuánto vale E_t cuando A_t se hace cero (a medida que $n \to \infty$)? Ej: $p = 2/n$



Propiedades de $\mathrm{ER}(n,p)$

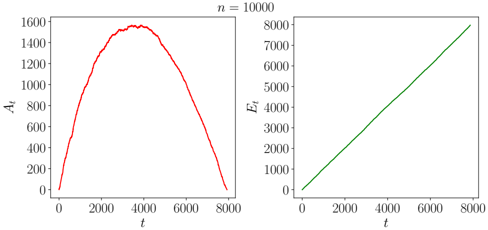
P2) Transición de fase en la aparición de una componente gigante

- Nos interesa $|\mathcal{E}_t| = E_t$. Junto con $|\mathcal{A}_t| = A_t$ forman una cadena de Markov en tiempo continuo:
  - El vector (E_t, A_t) permanece en cada estado durante un tiempo exponencial
  - En los momentos que hay cambio:

$$(E_{t+}, A_{t+}) = \Big(\underbrace{E_t + 1}_{\text{Un explorado más } (v)} \;,\; \underbrace{A_t + X_t - 1}_{\text{Los vecinos inactivos de } v \text{ ahora son activos (y } v \text{ ahora es explorado)}} \Big)$$

$$X_t \sim \mathrm{Bin}(|\mathcal{I}_t|, p) = \mathrm{Bin}(n - A_t - E_t, p)$$

- ¿Cuánto vale E_t cuando A_t se hace cero (a medida que $n \to \infty$)? Ej: $p = 2/n$



Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Los procesos escalado $\frac{E_{\tau n}}{n} = e_\tau$ y $\frac{A_{\tau n}}{n} = a_\tau$ parecen tener un comportamiento determinístico con $n \to \infty$: límite fluido

Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Los procesos escalado $\frac{E_{\tau n}}{n} = e_\tau$ y $\frac{A_{\tau n}}{n} = a_\tau$ parecen tener un comportamiento determinístico con $n \to \infty$: límite fluido
  - Una cadena de Markov admite la siguiente descomposición:

$$Y(t) = Y(0) + \int_0^t Q(Y(s))ds + M(t)$$

con $Q(l) = \sum_m (l-m)q(l,m) = \sum_m (l-m)p_{l,m}\lambda_l$ el drift del proceso (el cambio promedio del proceso en l) y $M(t)$ una martingala (ruido)

Propiedades de $\mathrm{ER}(n,p)$

P2) Transición de fase en la aparición de una componente gigante

- Los procesos escalado $\frac{E_{\tau n}}{n} = e_\tau$ y $\frac{A_{\tau n}}{n} = a_\tau$ parecen tener un comportamiento determinístico con $n \to \infty$: límite fluido
  - Una cadena de Markov admite la siguiente descomposición:

$$Y(t) = Y(0) + \int_0^t Q(Y(s))ds + M(t)$$

con $Q(l) = \sum_m (l-m)q(l,m) = \sum_m (l-m)p_{l,m}\lambda_l$ el drift del proceso (el cambio promedio del proceso en l) y $M(t)$ una martingala (ruido)

- Para este caso en particular resulta

$$E_t = 0 + \int_0^t 1ds + M_E(t)$$

$$A_t = 1 + \int_0^t \left(p(n - A_s - E_s) - 1\right) ds + M_A(t)$$

FACULTAD DE INGENIERÍA UDELAR

Propiedades de $\mathrm{ER}(n,p)$

P2) Transición de fase en la aparición de una componente gigante

- Re-escalando, usando $p = c/n$ (cantidad de vecinos promedio $\approx c$)...

$$e_\tau = \frac{E_{\tau n}}{n} = 0 + \frac{1}{n}\int_0^{\tau n} 1 ds + \frac{M_E(\tau n)}{n}$$

$$a_\tau = \frac{A_{\tau n}}{n} = \frac{1}{n} + \frac{1}{n}\int_0^{\tau n} \left(\frac{c}{n}(n - A_s - E_s) - 1\right) ds + \frac{M_A(\tau n)}{n}$$

Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Re-escalando, usando $p = c/n$ (cantidad de vecinos promedio $\approx c$)...

$$e_\tau = \frac{E_{\tau n}}{n} = 0 + \frac{1}{n}\int_0^{\tau n} 1 ds + \frac{M_E(\tau n)}{n}$$

$$a_\tau = \frac{A_{\tau n}}{n} = \frac{1}{n} + \frac{1}{n}\int_0^{\tau n} \left(\frac{c}{n}(n - A_s - E_s) - 1\right) ds + \frac{M_A(\tau n)}{n}$$

... y suponiendo que las martinagalas re-escaladas van a cero en n

$$e_\tau = \tau$$

$$\frac{da_\tau}{d\tau} = \frac{1}{n}\left(\frac{c}{n}(n - A_{\tau n} - E_{\tau n}) - 1\right) n$$

$$a_0 = 0$$

Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Re-escalando, usando $p = c/n$ (cantidad de vecinos promedio $\approx c$)...

$$e_\tau = \frac{E_{\tau n}}{n} = 0 + \frac{1}{n}\int_0^{\tau n} 1 ds + \frac{M_E(\tau n)}{n}$$

$$a_\tau = \frac{A_{\tau n}}{n} = \frac{1}{n} + \frac{1}{n}\int_0^{\tau n} \left(\frac{c}{n}(n - A_s - E_s) - 1\right) ds + \frac{M_A(\tau n)}{n}$$

... y suponiendo que las martinagalas re-escaladas van a cero en n

$$e_\tau = \tau$$
$$\frac{da_\tau}{d\tau} = \frac{1}{n}\left(\frac{c}{n}(n - A_{\tau n} - E_{\tau n}) - 1\right) n = c(1 - a_\tau - e_\tau) - 1$$
$$a_0 = 0$$

... que tiene solución analítica!

Propiedades de ER(n, p)

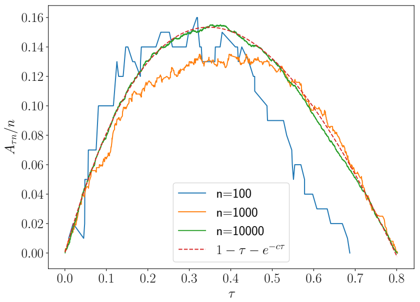
P2) Transición de fase en la aparición de una componente gigante

- Finalmente, en el límite $n \to \infty$

$$e_\tau = \tau$$

$$a_\tau = 1 - \tau - e^{-c\tau}$$

Ejemplo: $p = 2/n$



Propiedades de $\mathrm{ER}(n, p)$

P2) Transición de fase en la aparición de una componente gigante

- Finalmente, en el límite $n \to \infty$

$$
\begin{aligned}
e_\tau &= \tau \\
a_\tau &= 1 - \tau - e^{-c\tau}
\end{aligned}
$$

$\Rightarrow$ Habrá una componente gigante $\Leftrightarrow c > 1$ y su tamaño (proporcional a n) será la solución positiva de $1 - \tau = e^{-c\tau}$

Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Finalmente, en el límite $n \to \infty$

$$
\begin{aligned}
e_\tau &= \tau \\
a_\tau &= 1 - \tau - e^{-c\tau}
\end{aligned}
$$

$\Rightarrow$ Habrá una componente gigante $\Leftrightarrow c > 1$ y su tamaño (proporcional a n) será la solución positiva de $1 - \tau = e^{-c\tau}$

- Ver Remco van der Hofstad, "Random Graphs and Complex Networks. Volume One." Cambridge Series in Statistical and Probabilistic Mathematics (2017) para más formalidad

Propiedades de ER(n, p)

P2) Transición de fase en la aparición de una componente gigante

- Finalmente, en el límite $n \to \infty$

$$
\begin{aligned}
e_\tau &= \tau \\
a_\tau &= 1 - \tau - e^{-c\tau}
\end{aligned}
$$

$\Rightarrow$ Habrá una componente gigante $\Leftrightarrow c > 1$ y su tamaño (proporcional a n) será la solución positiva de $1 - \tau = e^{-c\tau}$

- Ver Remco van der Hofstad, "Random Graphs and Complex Networks. Volume One." Cambridge Series in Statistical and Probabilistic Mathematics (2017) para más formalidad

P3) Clustering coefficient pequeño de orden $O(n^{-1})$ y diámetro corto de orden $O(\log n)$ w.h.p.

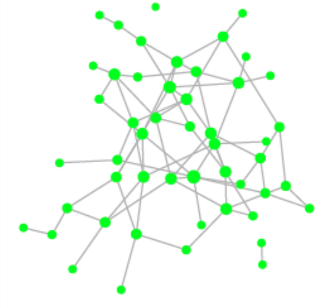
Modelos de Grafos Aleatorios

FACULTAD DE INGENIERÍA UDELAR

Modelos para redes reales

- Quizá la inovación más importante en los modelos modernos de grafos

Grafos aleatorios "tradicionales" → Modelos que imitan las propiedades observadas



Generalizando los grafos aleatorios

- Receta para generalizar el modelo de Erdös-Renyi

  $\Rightarrow$ Especificar $\mathcal{G}$ conteniendo los grafos de orden N_v, pero que además posean cierta característica

  $\Rightarrow$ Asignamos probabilidad uniforme para todos los grafos $G \in \mathcal{G}$

Generalizando los grafos aleatorios

- Receta para generalizar el modelo de Erdös-Renyi

  $\Rightarrow$ Especificar $\mathcal{G}$ conteniendo los grafos de orden N_v, pero que además posean cierta característica

  $\Rightarrow$ Asignamos probabilidad uniforme para todos los grafos $G \in \mathcal{G}$
- ¿Con qué dato contamos sobre nuestra red?
  - Lo único que sabemos es la cantidad media de vecinos c

    $\Rightarrow$ $\mathrm{ER}(n, c/n)$ parece una buena elección

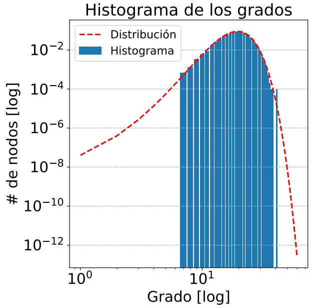
Generalizando los grafos aleatorios

- Receta para generalizar el modelo de Erdös-Renyi

  $\Rightarrow$ Especificar $\mathcal{G}$ conteniendo los grafos de orden N_v, pero que además posean cierta característica

  $\Rightarrow$ Asignamos probabilidad uniforme para todos los grafos $G \in \mathcal{G}$
- ¿Con qué dato contamos sobre nuestra red?
  - Lo único que sabemos es la cantidad media de vecinos c

    $\Rightarrow$ $\mathrm{ER}(n, c/n)$ parece una buena elección
  - ¿Y si sabemos la secuencia o la distribución de grados? ¿Sigue pareciendo?



Generalizando los grafos aleatorios

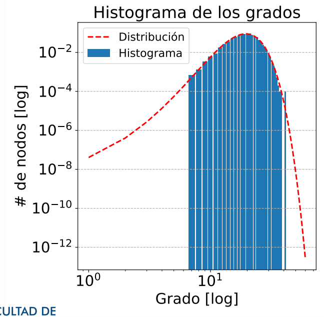
- Receta para generalizar el modelo de Erdös-Renyi

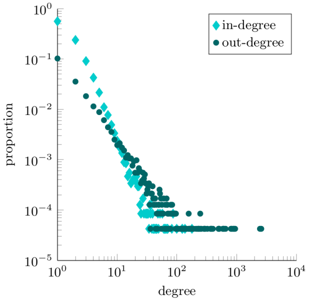
  $\Rightarrow$ Especificar $\mathcal{G}$ conteniendo los grafos de orden N_v, pero que además posean cierta característica

  $\Rightarrow$ Asignamos probabilidad uniforme para todos los grafos $G \in \mathcal{G}$

- ¿Con qué dato contamos sobre nuestra red?
  - Lo único que sabemos es la cantidad media de vecinos c

    $\Rightarrow$ ER$(n, c/n)$ parece una buena elección
  - ¿Y si sabemos la secuencia o la distribución de grados? ¿Sigue pareciendo?





Conectividad entre Autonomous Systems (ASs)

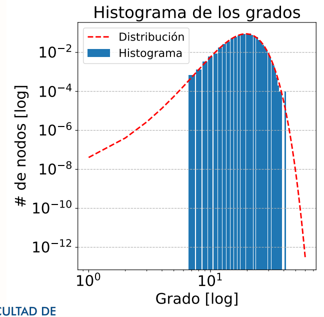
Generalizando los grafos aleatorios

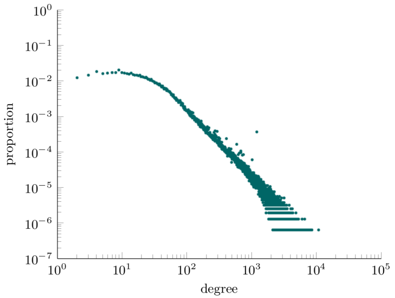
- Receta para generalizar el modelo de Erdös-Renyi

  $\Rightarrow$ Especificar $\mathcal{G}$ conteniendo los grafos de orden N_v, pero que además posean cierta característica

  $\Rightarrow$ Asignamos probabilidad uniforme para todos los grafos $G \in \mathcal{G}$
- ¿Con qué dato contamos sobre nuestra red?
  - Lo único que sabemos es la cantidad media de vecinos c

    $\Rightarrow$ ER$(n, c/n)$ parece una buena elección
  - ¿Y si sabemos la secuencia o la distribución de grados? ¿Sigue pareciendo?





Actores que comparten películas en IMdB

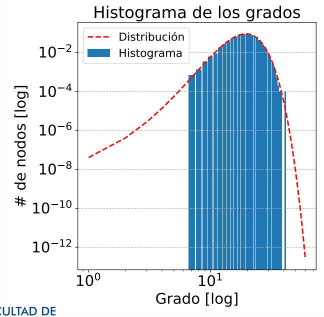
Generalizando los grafos aleatorios

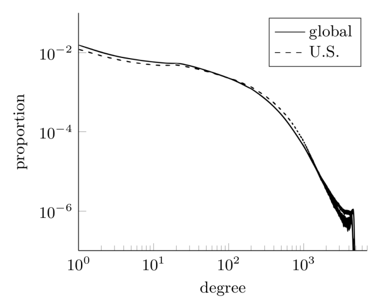
- Receta para generalizar el modelo de Erdös-Renyi

  $\Rightarrow$ Especificar $\mathcal{G}$ conteniendo los grafos de orden N_v, pero que además posean cierta característica

  $\Rightarrow$ Asignamos probabilidad uniforme para todos los grafos $G \in \mathcal{G}$
- ¿Con qué dato contamos sobre nuestra red?
  - Lo único que sabemos es la cantidad media de vecinos c

    $\Rightarrow$ ER$(n, c/n)$ parece una buena elección
  - ¿Y si sabemos la secuencia o la distribución de grados? ¿Sigue pareciendo?





Conectividad entre amigos en Facebook

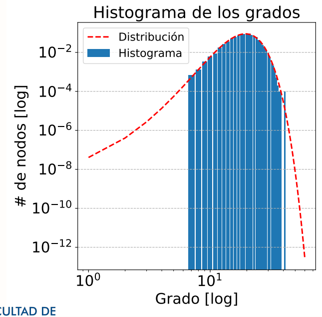
Generalizando los grafos aleatorios

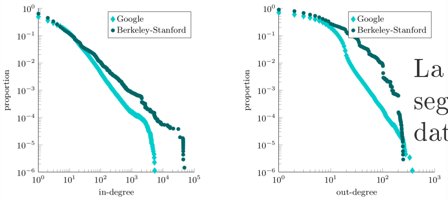
- Receta para generalizar el modelo de Erdös-Renyi

  $\Rightarrow$ Especificar $\mathcal{G}$ conteniendo los grafos de orden N_v, pero que además posean cierta característica

  $\Rightarrow$ Asignamos probabilidad uniforme para todos los grafos $G \in \mathcal{G}$
- ¿Con qué dato contamos sobre nuestra red?
  - Lo único que sabemos es la cantidad media de vecinos c

    $\Rightarrow$ ER$(n, c/n)$ parece una buena elección
  - ¿Y si sabemos la secuencia o la distribución de grados? ¿Sigue pareciendo?





La WWW según dos datasets

Generalizando los grafos aleatorios

- **Configuration model:** secuencia de grados dada $\mathbf{d} = \{d_1, \ldots, d_{N_v}\}$
  - El tamaño también queda dado bajo este modelo: $N_e = \sum_i d_i/2 = \bar{d}N_v/2 \Rightarrow \mathcal{G} \subset \mathcal{G}_{N_v,N_e}$
  - O sea, que es equivalente a especificar el modelo mediante una distribución condicional en $\mathcal{G}_{N_v,N_e}$

Generalizando los grafos aleatorios

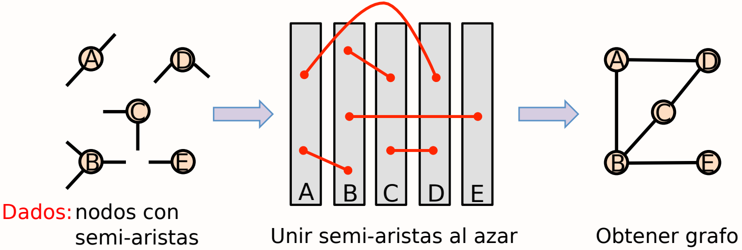
- Configuration model: secuencia de grados dada $\mathbf{d} = \{d_1, \ldots, d_{N_v}\}$
  - El tamaño también queda dado bajo este modelo: $N_e = \sum_i d_i/2 = \bar{d}N_v/2 \Rightarrow \mathcal{G} \subset \mathcal{G}_{N_v,N_e}$
  - O sea, que es equivalente a especificar el modelo mediante una distribución condicional en $\mathcal{G}_{N_v,N_e}$
- Modelos tipo configuration son útiles como 'nulos'

  Ej: comparar grafo observado G con $G' \in \mathcal{G}$ con power-law $\mathrm{P}(d)$
- También son útiles para imponer restricciones de conectividad

  Ej: La mayoría de los grafos no tienen nodos aislados

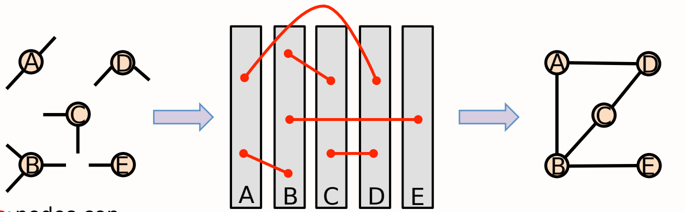
Simulando Configuration Models $\mathrm{CM}_n(\mathbf{d})$

- Matching algorithm. Ejemplo: $\mathbf{d} = \{1, 2, 2, 2, 3\}$



Simulando Configuration Models $\mathrm{CM}_n(\mathbf{d})$

- Matching algorithm. Ejemplo: $\mathbf{d} = \{1, 2, 2, 2, 3\}$

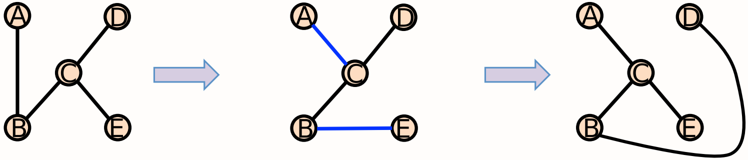


Dados: nodos con semi-aristas

Unir semi-aristas al azar

Obtener grafo

- Switching algorithm.



Inicializar

Aleatoriamente cambiar aristas
Repetir ~100Ne veces

Obtener grafo

Simulando Configuration Models

- Muchas veces me interesa generar grafos con cierta distribución de grados $P(d)$
- Puedo sortear la secuencia $\mathbf{d} = \{d_1, d_2, \ldots, N_v\}$ iid de $P(d)$

Simulando Configuration Models

- Muchas veces me interesa generar grafos con cierta distribución de grados $P(d)$
- Puedo sortear la secuencia $\mathbf{d} = \{d_1, d_2, \ldots, N_v\}$ iid de $P(d)$
  - ¿Cualquier secuencia es válida?
    $\Rightarrow$ Al menos necesito que $\sum_i d_i$ sea par. ¿Qué más?

Simulando Configuration Models

- Muchas veces me interesa generar grafos con cierta distribución de grados $P(d)$
- Puedo sortear la secuencia $\mathbf{d} = \{d_1, d_2, \ldots, N_v\}$ iid de $P(d)$
  - ¿Cualquier secuencia es válida?
    - $\Rightarrow$ Al menos necesito que $\sum_i d_i$ sea par. ¿Qué más?
    - $\Rightarrow$ Ejemplo: ¿$\mathbf{d} = \{5, 3, 1, 1, 1, 1\}$ puede generar un grafo simple?

Simulando Configuration Models

- Muchas veces me interesa generar grafos con cierta distribución de grados $P(d)$
- Puedo sortear la secuencia $\mathbf{d} = \{d_1, d_2, \ldots, N_v\}$ iid de $P(d)$
  - ¿Cualquier secuencia es válida?
    - $\Rightarrow$ Al menos necesito que $\sum_i d_i$ sea par. ¿Qué más?
    - $\Rightarrow$ Ejemplo: ¿$\mathbf{d} = \{5, 3, 1, 1, 1, 1\}$ puede generar un grafo simple?
  - Teorema de Erdös-Gallai Condición necesaria y suficiente para que la secuencia sea gráfica: ser par y además

$$\underbrace{\sum_{i=1}^{k} d_i}_{\substack{\text{Semi-aristas de los} \\ \text{primero } k \text{ nodos}}} \leq \underbrace{k(k-1)}_{\substack{2\times\text{máximo número de} \\ \text{aristas entre } k \text{ nodos} \\ \text{(los } k \text{ interconectados)}}} + \underbrace{\sum_{i=k+1}^{N_v} \min(d_i, k)}_{\substack{\text{Máximo número de aristas} \\ \text{desde los nodos } i > k \\ \text{(los } k \text{ conectados hacia afuera)}}} \qquad \forall 1 \leq k \leq N_v$$

con $d_1 \leq d_2 \leq \ldots \leq d_{N_v}$

Simulando Configuration Models

- Muchas veces me interesa generar grafos con cierta distribución de grados $P(d)$
- Puedo sortear la secuencia $\mathbf{d} = \{d_1, d_2, \ldots, N_v\}$ iid de $P(d)$
  - ¿Cualquier secuencia es válida?
    - $\Rightarrow$ Al menos necesito que $\sum_i d_i$ sea par. ¿Qué más?
    - $\Rightarrow$ Ejemplo: ¿$\mathbf{d} = \{5, 3, 1, 1, 1, 1\}$ puede generar un grafo simple?
  - Teorema de Erdös-Gallai Condición necesaria y suficiente para que la secuencia sea gráfica: ser par y además

$$\underbrace{\sum_{i=1}^{k} d_i}_{\substack{\text{Semi-aristas de los} \\ \text{primero } k \text{ nodos}}} \leq \underbrace{k(k-1)}_{\substack{2\times\text{máximo número de} \\ \text{aristas entre } k \text{ nodos} \\ \text{(los } k \text{ interconectados)}}} + \underbrace{\sum_{i=k+1}^{N_v} \min(d_i, k)}_{\substack{\text{Máximo número de aristas} \\ \text{desde los nodos } i > k \\ \text{(los } k \text{ conectados hacia afuera)}}} \qquad \forall 1 \leq k \leq N_v$$

  con $d_1 \leq d_2 \leq \ldots \leq d_{N_v}$

- Por suerte, no hace falta chequear esta condición para el algoritmo de matching

Simulando Configuration Models

- Problema Incluso si la secuencia $\mathbf{d} = \{d_1, d_2, \ldots, d_{N_v}\}$ es gráfica, el algoritmo de matching puede generar multi-aristas y loops

  $\Rightarrow$ Borrémoslos del grafo resultante y llamemos a esta nueva secuencia $\mathbf{d}^{(er)} \Rightarrow$ típicamente $\mathbf{d} \neq \mathbf{d}^{(er)}$

Simulando Configuration Models

- Problema Incluso si la secuencia $\mathbf{d} = \{d_1, d_2, \ldots, d_{N_v}\}$ es gráfica, el algoritmo de matching puede generar multi-aristas y loops

  $\Rightarrow$ Borrémoslos del grafo resultante y llamemos a esta nueva secuencia $\mathbf{d}^{(er)}$ $\Rightarrow$ típicamente $\mathbf{d} \neq \mathbf{d}^{(er)}$

  - Por suerte, si $P(d)$ tiene media finita y $P(D \geq 1) = 1$ entonces

$$P\left(\sum_{d=1}^{N_v} |P^{(er)}(d) - P(d)| > \epsilon\right) \to 0$$

Simulando Configuration Models

- Problema Incluso si la secuencia $\mathbf{d} = \{d_1, d_2, \ldots, d_{N_v}\}$ es gráfica, el algoritmo de matching puede generar multi-aristas y loops

  $\Rightarrow$ Borrémoslos del grafo resultante y llamemos a esta nueva secuencia $\mathbf{d}^{(er)}$ $\Rightarrow$ típicamente $\mathbf{d} \neq \mathbf{d}^{(er)}$

  - Por suerte, si $P(d)$ tiene media finita y $P(D \geq 1) = 1$ entonces

$$P\left(\sum_{d=1}^{N_v} |P^{(er)}(d) - P(d)| > \epsilon\right) \to 0$$

- Otro problema ¿Pero $\mathrm{CM}_n(\mathbf{d})$ genera al azar uniforme entre todos los que tiene secuencia de grados $\mathbf{d}$?

Simulando Configuration Models

- Problema Incluso si la secuencia $\mathbf{d} = \{d_1, d_2, \ldots, d_{N_v}\}$ es gráfica, el algoritmo de matching puede generar multi-aristas y loops

  $\Rightarrow$ Borrémoslos del grafo resultante y llamemos a esta nueva secuencia $\mathbf{d}^{(er)} \Rightarrow$ típicamente $\mathbf{d} \neq \mathbf{d}^{(er)}$

  - Por suerte, si $P(d)$ tiene media finita y $P(D \geq 1) = 1$ entonces

$$P\left(\sum_{d=1}^{N_v} |P^{(er)}(d) - P(d)| > \epsilon\right) \to 0$$

- Otro problema ¿Pero $\mathrm{CM}_n(\mathbf{d})$ genera al azar uniforme entre todos los que tiene secuencia de grados $\mathbf{d}$?

  ✓ Condicionado al evento $\{\mathrm{CM}_n(\mathbf{d})$ es un grafo simple$\}$

  $\Rightarrow$ si además además $\mathrm{Var}(D) < \infty$ entonces la probabilidad de obtener un grafo simple bajo $\mathrm{CM}_n(\mathbf{d})$ converge en N_v a

$$e^{-\nu/2-\nu^2/4} \text{ con } \nu = \mathbb{E}\{D(D-1)\}/\mathbb{E}\{D\}$$

Simulando Configuration Models

- ¿Para qué sirve lo anterior?
  - Quiero estudiar una propiedad de los grafos uniformes con secuencia de grados $\mathbf{d}$ ($\mathrm{UG}_n(\mathbf{d})$) del estilo

$$\lim_{n\to\infty} P\left(\mathrm{UG}_n(\mathbf{d}) \in \mathcal{E}_n\right) \to 1$$

  - Puedo hacer el siguiente razonamiento, y estudiar si $P\left(\mathrm{CM}_n(\mathbf{d}) \in \mathcal{E}_n^c\right) \to 0$ (que es mucho más fácil)

$$\begin{aligned} P\left(\mathrm{UG}_n(\mathbf{d}) \in \mathcal{E}_n^c\right) &= P\left(\mathrm{CM}_n(\mathbf{d}) \in \mathcal{E}_n^c | \mathrm{CM}_n(\mathbf{d}) \text{ simple}\right) \\ &= \frac{P\left(\mathrm{CM}_n(\mathbf{d}) \in \mathcal{E}_n^c, \mathrm{CM}_n(\mathbf{d}) \text{ simple}\right)}{P\left(\mathrm{CM}_n(\mathbf{d}) \text{ simple}\right)} \\ &\leq \frac{P\left(\mathrm{CM}_n(\mathbf{d}) \in \mathcal{E}_n^c\right)}{P\left(\mathrm{CM}_n(\mathbf{d}) \text{ simple}\right)} \to 0 \end{aligned}$$

Algunos Resultados de $\mathrm{CM}_n(\mathbf{d})$

P1) Transición de fase en la aparición de una componente gigante

- Se puede hacer un análisis con ecuaciones diferenciales como hicimos para $\mathrm{ER}(n, c/n)$
  - Hay que considerar el proceso dado por $A_t(j)$ (nodos activos de grado j), E_t (nodos explorados, lo que me interesa contar) y U_t (semi-aristas sin juntar)
- Componente gigante de tamaño $O(n)$ como en $\mathrm{ER}(n, c/n)$
- Condición: $\nu = \mathbb{E}\{D(D-1)\}/\mathbb{E}\{D\} > 1$ (caso $\mathrm{ER}(n, c/n)$ equivale a $c(1+c) > 2c$)

Algunos Resultados de $\mathrm{CM}_n(\mathbf{d})$

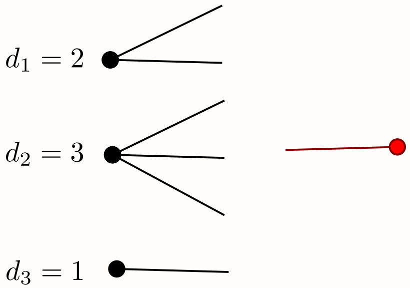
P1) Transición de fase en la aparición de una componente gigante

- Se puede hacer un análisis con ecuaciones diferenciales como hicimos para $\mathrm{ER}(n, c/n)$
  - Hay que considerar el proceso dado por $A_t(j)$ (nodos activos de grado j), E_t (nodos explorados, lo que me interesa contar) y U_t (semi-aristas sin juntar)
- Componente gigante de tamaño $O(n)$ como en $\mathrm{ER}(n, c/n)$
- Condición: $\nu = \mathbb{E}\{D(D-1)\}/\mathbb{E}\{D\} > 1$ (caso $\mathrm{ER}(n, c/n)$ equivale a $c(1+c) > 2c$)
  - Intuitivamente: necesito que el grado (menos uno) de un vecino del nodo sea mayor a 1

Algunos Resultados de $\mathrm{CM}_n(\mathbf{d})$

P1) Transición de fase en la aparición de una componente gigante

- Se puede hacer un análisis con ecuaciones diferenciales como hicimos para $\mathrm{ER}(n, c/n)$
  - Hay que considerar el proceso dado por $A_t(j)$ (nodos activos de grado j), E_t (nodos explorados, lo que me interesa contar) y U_t (semi-aristas sin juntar)
- Componente gigante de tamaño $O(n)$ como en $\mathrm{ER}(n, c/n)$
- Condición: $\nu = \mathbb{E}\{D(D-1)\}/\mathbb{E}\{D\} > 1$ (caso $\mathrm{ER}(n, c/n)$ equivale a $c(1+c) > 2c$)
  - Intuitivamente: necesito que el grado (menos uno) de un vecino del nodo sea mayor a 1

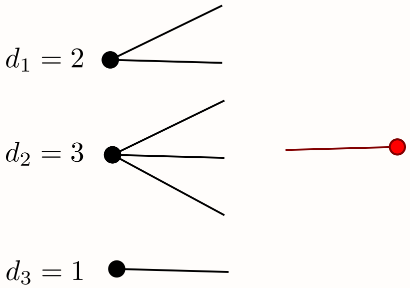


$$P\left(\text{grado del vecino} = j\right) = \frac{\sum_i d_i \mathbb{I}_{d_i = j}}{\sum_i d_i}$$

Algunos Resultados de $\mathrm{CM}_n(\mathbf{d})$

P1) Transición de fase en la aparición de una componente gigante

- Se puede hacer un análisis con ecuaciones diferenciales como hicimos para $\mathrm{ER}(n, c/n)$
  - Hay que considerar el proceso dado por $A_t(j)$ (nodos activos de grado j), E_t (nodos explorados, lo que me interesa contar) y U_t (semi-aristas sin juntar)
- Componente gigante de tamaño $O(n)$ como en $\mathrm{ER}(n, c/n)$
- Condición: $\nu = \mathbb{E}\{D(D-1)\}/\mathbb{E}\{D\} > 1$ (caso $\mathrm{ER}(n, c/n)$ equivale a $c(1+c) > 2c$)
  - Intuitivamente: necesito que el grado (menos uno) de un vecino del nodo sea mayor a 1

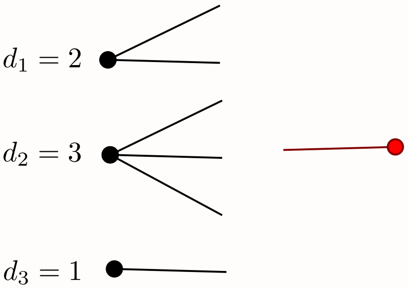


$$P\,(\text{grado del vecino} = j) = \frac{\sum_i d_i \mathbb{I}_{d_i=j}}{\sum_i d_i} = \frac{j\sum_i \mathbb{I}_{d_i=j}}{\sum_i d_i} = \frac{j n_j}{\sum_i i n_i} = \frac{j p_j}{\sum_i i p_i}$$

Algunos Resultados de $\mathrm{CM}_n(\mathbf{d})$

P1) Transición de fase en la aparición de una componente gigante

- Se puede hacer un análisis con ecuaciones diferenciales como hicimos para $\mathrm{ER}(n, c/n)$
  - Hay que considerar el proceso dado por $A_t(j)$ (nodos activos de grado j), E_t (nodos explorados, lo que me interesa contar) y U_t (semi-aristas sin juntar)
- Componente gigante de tamaño $O(n)$ como en $\mathrm{ER}(n, c/n)$
- Condición: $\nu = \mathbb{E}\{D(D-1)\}/\mathbb{E}\{D\} > 1$ (caso $\mathrm{ER}(n, c/n)$ equivale a $c(1+c) > 2c$)
  - Intuitivamente: necesito que el grado (menos uno) de un vecino del nodo sea mayor a 1



$$P\,(\text{grado del vecino} = j) = \frac{\sum_i d_i \mathbb{I}_{d_i=j}}{\sum_i d_i} =$$

$$= \frac{j \sum_i \mathbb{I}_{d_i=j}}{\sum_i d_i} = \frac{j n_j}{\sum_i i n_i} = \frac{j p_j}{\sum_i i p_i}$$

$$\mathbb{E}\{\text{grado del vecino} - 1\} = \sum_j (j-1) \frac{j p_j}{\sum_i i p_i} =$$

$$= \frac{\mathbb{E}\{D(D-1)\}}{\mathbb{E}\{D\}}$$

Algunos Resultados de $\mathrm{CM}_n(\mathbf{d})$

P2) El clustering coefficient también se va a 0 como en $\mathrm{ER}(n, c/n)$

Algunos Resultados de $\mathrm{CM}_n(\mathbf{d})$

P2) El clustering coefficient también se va a 0 como en $\mathrm{ER}(n, c/n)$

P3) Distancias típicas dependen de $P(d)$

- Si $Var(D) < \infty \Rightarrow$ distancias típicas de orden $O(\log_\nu n)$
- Caso particular de una power-law $\mathrm{P}\,(d) \sim Cd^{-\alpha}$ con $\alpha \in (2,3)$ (i.e. media finita y varianza infinita)$\Rightarrow$ distancias típicas de orden $O(\log\log(n))$

Algunos Resultados de $\mathrm{CM}_n(\mathbf{d})$

P2) El clustering coefficient también se va a 0 como en $\mathrm{ER}(n, c/n)$

P3) Distancias típicas dependen de $P(d)$

- Si $Var(D) < \infty \Rightarrow$ distancias típicas de orden $O(\log_\nu n)$
- Caso particular de una power-law $\mathrm{P}(d) \sim Cd^{-\alpha}$ con $\alpha \in (2,3)$ (i.e. media finita y varianza infinita)$\Rightarrow$ distancias típicas de orden $O(\log\log(n))$

P4) Se puede calcular cotas sobre el independence number de estos grafos analizando procesos de exploración mediante límites fluidos.

- Brightwell, Graham, Svante Janson, and Malwina Luczak. “The greedy independent set in a random graph with given degrees.” *Random Structures & Algorithms* 51, no. 4 (2017): 565-586.
- Bermolen, Paola, Matthieu Jonckheere, and Pascal Moyal. “The jamming constant of uniform random graphs.” *Stochastic Processes and their Applications* 127, no. 7 (2017): 2138-2178.

Aplicación 1: Indicadores Estadísticamente Significativos

- Consideremos un grafo G^{obs} que se obtuvo de una observación
- Q: Es el indicador estructural $\eta(G^{obs})$ estadísticamente significativo, i.e., inusual?
  - $\Rightarrow$ Para evaluar si es significativo se necesita un marco de referencia: un modelo nulo

Aplicación 1: Indicadores Estadísticamente Significativos

- Consideremos un grafo G^{obs} que se obtuvo de una observación
- Q: Es el indicador estructural $\eta(G^{obs})$ estadísticamente significativo, i.e., inusual?
  - $\Rightarrow$ Para evaluar si es significativo se necesita un marco de referencia: un modelo nulo
  - $\Rightarrow$ Es habitual usar grafos aleatorios como marco de referencia

Aplicación 1: Indicadores Estadísticamente Significativos

- Consideremos un grafo G^{obs} que se obtuvo de una observación
- Q: Es el indicador estructural $\eta(G^{obs})$ estadísticamente significativo, i.e., inusual?
  - $\Rightarrow$ Para evaluar si es significativo se necesita un marco de referencia: un modelo nulo
  - $\Rightarrow$ Es habitual usar grafos aleatorios como marco de referencia
- Definamos una colección $\mathcal{G}$, y comparemos $\eta(G^{obs})$ con los valores $\{\eta(G) : G \in \mathcal{G}\}$
  - $\Rightarrow$ Más formalmente, construyamos una distribución de referencia

$$\mathrm{P}_{\eta,\mathcal{G}}(t) = \frac{|\{G \in \mathcal{G} : \eta(G) \leq t\}|}{|\mathcal{G}|}$$

- Si $\eta(G^{obs})$ es suficientemente improbable bajo $\mathrm{P}_{\eta,\mathcal{G}}(t)$
  - $\Rightarrow$ Evidencia contra la hipótesis nula H_0: G^{obs} es una muestra uniforme de $\mathcal{G}$

Ejemplo: club de karate de Zachary

- El club de karate de Zachary tiene coeficiente de clustering $\mathrm{cl}(G^{obs}) = 0{,}2257$
  - $\Rightarrow$ Usemos grafos aleatorios para verificar si este es un valor inusual

Ejemplo: club de karate de Zachary

- El club de karate de Zachary tiene coeficiente de clustering $\mathrm{cl}(G^{obs}) = 0{,}2257$

  $\Rightarrow$ Usemos grafos aleatorios para verificar si este es un valor inusual

- Construimos dos referencias

  1) Conjunto $\mathcal{G}_1$ de grafos aleatorios con el mismo $N_v = 34$ y $N_e = 78$ (i.e. $\mathrm{ER}(N_v, N_e)$)
  2) Agregar la restricción que los grafos en $\mathcal{G}_2$ tengan la misma secuencia de grados que G^{obs} (i.e. $\mathrm{CM}_{N_v}(\mathbf{d})$)

Ejemplo: club de karate de Zachary

- El club de karate de Zachary tiene coeficiente de clustering $\mathrm{cl}(G^{obs}) = 0{,}2257$

  $\Rightarrow$ Usemos grafos aleatorios para verificar si este es un valor inusual

- Construimos dos referencias
  1) Conjunto $\mathcal{G}_1$ de grafos aleatorios con el mismo $N_v = 34$ y $N_e = 78$ (i.e. $\mathrm{ER}(N_v, N_e)$)
  2) Agregar la restricción que los grafos en $\mathcal{G}_2$ tengan la misma secuencia de grados que G^{obs} (i.e. $\mathrm{CM}_{N_v}(\mathbf{d})$)

- $|\mathcal{G}_1| \approx 8{,}4 \times 10^{96}$ y $|\mathcal{G}_2|$ mucho más pequeño, pero igual enorme

  $\Rightarrow$ Hallar todos los grafos $\mathcal{G}_1$ o $\mathcal{G}_2$ para obtener las distribuciones $\mathrm{P}_{\eta,\mathcal{G}_1}(t)$ o $\mathrm{P}_{\eta,\mathcal{G}_2}(t)$ es imposible

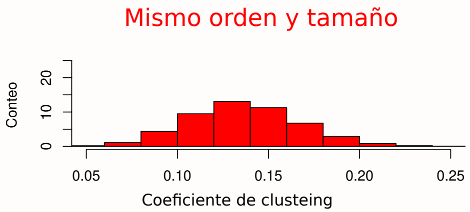
- Alternativa: usa simulaciones para aproximar las distribuciones

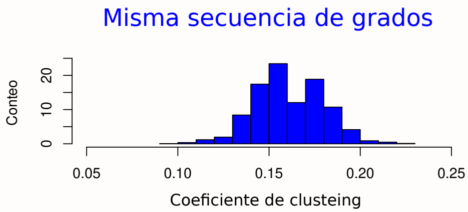
  $\Rightarrow$ Generamos 10.000 muestras uniformes G de $\mathcal{G}_1$ y $\mathcal{G}_2$

  $\Rightarrow$ Calculamos $\eta(G) = \mathrm{cl}(G)$ para cada muestra y graficamos los histogramas

Ejemplo: club de karate de Zachary (cont.)

- Gráfica de histogramas para aproximar la distribución





- El valor $\mathrm{cl}(G^{obs}) = 0{,}2257$ es muy poco probable para ambos modelos

  Ej: solo 3 de las 10.000 muestras de $\mathcal{G}_1$ tenía $\mathrm{cl}(G) > 0{,}2257$

- La evidencia nos hace concluir que G^{obs} no fue obtenida de una muestra uniforme en $\mathcal{G}_1$ o $\mathcal{G}_2$

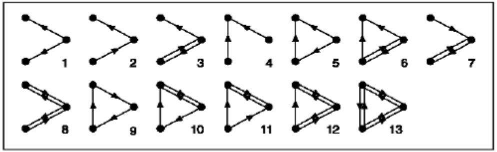
Aplicación 2: Detectar motifs en grafos

- Otra aplicación de grafos aleatorios es la detección de network motifs

  ⇒ Hallar las estructuras de las que se construye el grafo

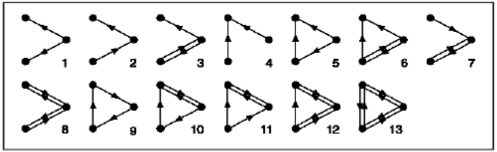
Aplicación 2: Detectar motifs en grafos

- Otra aplicación de grafos aleatorios es la detección de network motifs
  - ⇒ Hallar las estructuras de las que se construye el grafo
- **Def:** Network motifs son sub-grafos pequeños que ocurren mucho más frecuentemente en el grafo observado que en un grafo aleatorio comparable
- Ej: Hay $L_3 = 13$ grafos dirigidos (y conexos) distintos con $k = 3$ vértices



Aplicación 2: Detectar motifs en grafos

- Otra aplicación de grafos aleatorios es la detección de network motifs
  - $\Rightarrow$ Hallar las estructuras de las que se construye el grafo
- **Def:** Network motifs son sub-grafos pequeños que ocurren mucho más frecuentemente en el grafo observado que en un grafo aleatorio comparable
- Ej: Hay $L_3 = 13$ grafos dirigidos (y conexos) distintos con $k = 3$ vértices

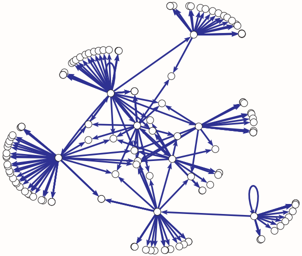


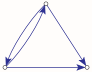
- Sea N_i^k el número de sub-grafos de k vértices del tipo i en G, $i = 1, \ldots, L_k$
  - $\Rightarrow$ Cada valor N_i^k puede compararse con referencia $\mathrm{P}_{N_i,\mathcal{G}}$
  - $\Rightarrow$ Si N_i^k es extremo $\Rightarrow$ sub-grafo es network motif

Ejemplo: red de blogs sobre SIDA

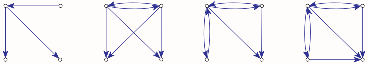
- Red de blogs sobre SIDA G^{obs} con $N_v = 146$ bloggers y $N_e = 183$ enlaces

  $\Rightarrow$ Se buscan motifs con $k = 3$ and 4 vertices





motif de 3 vértices



motifs de 4 vértices

- Se simulan 10.000 grafos dirigidos usando el algoritmo de switching

  $\Rightarrow$ Se respetó la secuencia de grados de entrada y salida de G^{obs}, así como las aristas mutuas

  $\Rightarrow$ Se estimaron distribuciones de referencia $\mathrm{P}_{N_i,\mathcal{G}}(t)$

- Ej: dos bloggers con una arista mutua y una 'autoridad' en común

Desafíos para detectar motifs

- Los motifs generalmente se intersectan con copias de sí mismos

  $\Rightarrow$ Un requisito extra puede ser que sean subgrafos frecuentes pero también disjuntos

Desafíos para detectar motifs

- Los motifs generalmente se intersectan con copias de sí mismos

  $\Rightarrow$ Un requisito extra puede ser que sean subgrafos frecuentes pero también disjuntos

- A medida que el grafo crece, los desafíos computacionales son cada vez mayores

  $\Rightarrow$ Por ejemplo, el número de potenciales motifs L_k crece muy rápidamente con k

  Ej: Subgrafos dirigidos $L_3 = 13$, $L_4 = 199$, $L_5 = 9364$

Desafíos para detectar motifs

- Los motifs generalmente se intersectan con copias de sí mismos

  $\Rightarrow$ Un requisito extra puede ser que sean subgrafos frecuentes pero también disjuntos

- A medida que el grafo crece, los desafíos computacionales son cada vez mayores

  $\Rightarrow$ Por ejemplo, el número de potenciales motifs L_k crece muy rápidamente con k

  Ej: Subgrafos dirigidos $L_3 = 13$, $L_4 = 199$, $L_5 = 9364$

- Una estrategia posible es muestrear el grafo

Modelos de Grafos Aleatorios

Redes Dinámicas

- Un buen modelo estadístico debería ser [Robbins-Morris'07]
  - ✓ Estimable a partir de y razonablemente representativo de los datos observados
  - ✓ Plausible teóricamente sobre los efectos que pueden haber producido la red
  - ✓ Capaz de discriminar entre los distintos efectos que mejor explican los datos

Redes Dinámicas

- Un buen modelo estadístico debería ser [Robbins-Morris'07]
  - ✓ Estimable a partir de y razonablemente representativo de los datos observados
  - ✓ Plausible teóricamente sobre los efectos que pueden haber producido la red
  - ✓ Capaz de discriminar entre los distintos efectos que mejor explican los datos
- El modelo $\mathrm{CM}_n(\mathbf{d})$ puede generar grafos tomando en cuenta la distribución de grados, incluso varios casos de power-law
  - ¿Pero qué mecanismo lleva a esta configuración?
  - Modelos como el Preferential Attachement dan una explicación sencilla pero que brinda intuición

Redes Dinámicas

- Muchas redes crecen o al menos evolucionan en el tiempo

  Ej: Web, citas científicas, Twitter, genoma ...
- Posible encare para modelos que imitan el crecimiento de un grafo
  - Especificar mecanismos simples para la dinámica del grafo
  - Estudiar las características estructurales que emergen a medida que el tiempo $t \to \infty$
- Q: Estas propiedades son las que se observan en grafos del mundo real?

Redes Dinámicas

- Muchas redes crecen o al menos evolucionan en el tiempo

  Ej: Web, citas científicas, Twitter, genoma ...

- Posible encare para modelos que imitan el crecimiento de un grafo
  - Especificar mecanismos simples para la dinámica del grafo
  - Estudiar las características estructurales que emergen a medida que el tiempo $t \to \infty$

- Q: Estas propiedades son las que se observan en grafos del mundo real?

- Los dos métodos de este tipo más populares por lejos son

  $\Rightarrow$ Preferential attachment models

  $\Rightarrow$ Copying models

- Mecanismos que pueden explicar popularidad y duplicación de genes respectivamente

Preferential attachment model

- Modelo simple para la creación de, por ejemplo, enlaces entre páginas web
  - Los vértices se crean de a uno, y los notamos $1, \ldots, N_v$

FACULTAD DE INGENIERÍA UDELAR

Preferential attachment model

- Modelo simple para la creación de, por ejemplo, enlaces entre páginas web
  - Los vértices se crean de a uno, y los notamos $1, \ldots, N_v$
  - Cuando el nodo j es creado tiene un único enlace hacia i, $1 \leq i < j$

Preferential attachment model

- Modelo simple para la creación de, por ejemplo, enlaces entre páginas web
  - Los vértices se crean de a uno, y los notamos $1, \ldots, N_v$
  - Cuando el nodo j es creado tiene un único enlace hacia i, $1 \leq i < j$
  - La creación del enlace (j, i) es aleatorio:
    - Con probabilidad p, j se conecta con i elegido uniforme al azar
    - Con probabilidad $1 - p$, j se conecta con i con probabilidad $\propto d_i^{in}$

Preferential attachment model

- Modelo simple para la creación de, por ejemplo, enlaces entre páginas web
  - Los vértices se crean de a uno, y los notamos $1, \ldots, N_v$
  - Cuando el nodo j es creado tiene un único enlace hacia i, $1 \leq i < j$
  - La creación del enlace (j, i) es aleatorio:
    - Con probabilidad p, j se conecta con i elegido uniforme al azar
    - Con probabilidad $1 - p$, j se conecta con i con probabilidad $\propto d_i^{in}$
- El grafo resultante es dirigido, donde cada vértice tiene $d_v^{out} = 1$

Preferential attachment model

- Modelo simple para la creación de, por ejemplo, enlaces entre páginas web
  - Los vértices se crean de a uno, y los notamos $1, \ldots, N_v$
  - Cuando el nodo j es creado tiene un único enlace hacia i, $1 \leq i < j$
  - La creación del enlace (j, i) es aleatorio:
    - Con probabilidad p, j se conecta con i elegido uniforme al azar
    - Con probabilidad $1 - p$, j se conecta con i con probabilidad $\propto d_i^{in}$
- El grafo resultante es dirigido, donde cada vértice tiene $d_v^{out} = 1$
- El modelo preferential attachment lleva a una dinámica de "rico se vuelve más rico"

  $\Rightarrow$ Los enlaces se forman preferentemente hacia los nodos que (actualmente) son los más populares

  $\Rightarrow$ La probabilidad de que el nodo i aumente su popularidad $\propto$ a la popularidad actual del nodo i

Preferential attachment resulta en power laws

Teorema

El preferential attachment model resulta en una distribución de grados del tipo power-law con exponente $\alpha = 1 + \frac{1}{1-p}$*, i.e.,*

$$\mathrm{P}\left(d^{in} = d\right) \propto d^{-\left(1+\frac{1}{1-p}\right)}$$

Preferential attachment resulta en power laws

Teorema

El preferential attachment model resulta en una distribución de grados del tipo power-law con exponente $\alpha = 1 + \frac{1}{1-p}$*, i.e.,*

$$\mathrm{P}\left(d^{in} = d\right) \propto d^{-\left(1+\frac{1}{1-p}\right)}$$

- La clave: "j se conecta con i con probabilidad $\propto d_i^{in}$" es básicamente copiar, i.e., "j elige una arista uniforme al azar y copia su destino", i.e., "copiar la decisión de otro nodo elegido uniforme al azar"
- Importante: Copiar las decisiones de otros vs. decisiones totalmente independientes en $\mathrm{ER}(n,p)$ o $\mathrm{CM}_n(\mathbf{d})$

Preferential attachment resulta en power laws

Teorema

El preferential attachment model resulta en una distribución de grados del tipo power-law con exponente $\alpha = 1 + \frac{1}{1-p}$, *i.e.,*

$$\mathrm{P}\left(d^{in} = d\right) \propto d^{-\left(1+\frac{1}{1-p}\right)}$$

- La clave: "j se conecta con i con probabilidad $\propto d_i^{in}$" es básicamente copiar, i.e., "j elige una arista uniforme al azar y copia su destino", i.e., "copiar la decisión de otro nodo elegido uniforme al azar"
- Importante: Copiar las decisiones de otros vs. decisiones totalmente independientes en $\mathrm{ER}(n,p)$ o $\mathrm{CM}_n(\mathbf{d})$
- A medida que $p \to 0$ $\Rightarrow$ La copia se vuelve más frecuente $\Rightarrow$ Menor $\alpha \to 2$
  - Intuitivamente: mucho más probable encontrarse con páginas web extremadamente populares (colas más pesadas en la distribución de grados)

Aproximación Continua usando Límites Fluidos

- Grado entrante $d_i^{in}(t)$ del nodo i en tiempo $t \geq i$ es una cadena de Markov:
  1) Condición inicial: $d_i^{in}(i) = 0$ dado que el nodo i se crea en tiempo $t = i$
  2) Dinámica: $d_i^{in}(t)$ aumenta en uno con probabilidad (cf. hay t aristas creadas en tiempo t)

$$\mathrm{P}\left((t+1, i) \in E\right) = p \times \frac{1}{t} + (1-p) \times \frac{d_i^{in}(t)}{t}$$

Aproximación Continua usando Límites Fluidos

- Grado entrante $d_i^{in}(t)$ del nodo i en tiempo $t \geq i$ es una cadena de Markov:
  1) Condición inicial: $d_i^{in}(i) = 0$ dado que el nodo i se crea en tiempo $t = i$
  2) Dinámica: $d_i^{in}(t)$ aumenta en uno con probabilidad (cf. hay t aristas creadas en tiempo t)

$$\mathrm{P}\left((t+1, i) \in E\right) = p \times \frac{1}{t} + (1-p) \times \frac{d_i^{in}(t)}{t}$$

- Hagamos una aproximación por límite fluido como ya hicimos con $\mathrm{ER}(n, c/n)$
  - Tiempo continuo $t \in [0, N_v]$
  - Los grados ahora son continuos e indexados por i: $x_i^{in}(t) : [i, N_v] \mapsto \mathbb{R}_+$

Aproximación Continua usando Límites Fluidos

- Grado entrante $d_i^{in}(t)$ del nodo i en tiempo $t \geq i$ es una cadena de Markov:
  1) Condición inicial: $d_i^{in}(i) = 0$ dado que el nodo i se crea en tiempo $t = i$
  2) Dinámica: $d_i^{in}(t)$ aumenta en uno con probabilidad (cf. hay t aristas creadas en tiempo t)

$$\mathrm{P}\left((t+1, i) \in E\right) = p \times \frac{1}{t} + (1-p) \times \frac{d_i^{in}(t)}{t}$$

- Hagamos una aproximación por límite fluido como ya hicimos con $\mathrm{ER}(n, c/n)$
  - Tiempo continuo $t \in [0, N_v]$
  - Los grados ahora son continuos e indexados por i: $x_i^{in}(t) : [i, N_v] \mapsto \mathbb{R}_+$
- Como siempre crece en uno, el drift resulta en la siguiente ecuación diferencial:

$$\frac{dx_i^{in}(t)}{dt} = \frac{p}{t} + \frac{(1-p)x_i^{in}(t)}{t}, \quad x_i^{in}(i) = 0$$

Resolviendo la ecuación diferencial

- La ecuación diferencial es (con $q = 1 - p$)

$$\frac{dx_i^{in}(t)}{dt} = \frac{p + qx_i^{in}(t)}{t}$$

Resolviendo la ecuación diferencial

- La ecuación diferencial es (con $q = 1 - p$)

$$\frac{dx_i^{in}(t)}{dt} = \frac{p + qx_i^{in}(t)}{t}$$

- Divido ambos lados por $p + qx_i^{in}(t)$ e integro en t

$$\int \frac{1}{p + qx_i^{in}} \frac{dx_i^{in}}{dt} dt = \int \frac{1}{t} dt$$

Resolviendo la ecuación diferencial

- La ecuación diferencial es (con $q = 1 - p$)

$$\frac{dx_i^{in}(t)}{dt} = \frac{p + qx_i^{in}(t)}{t}$$

- Divido ambos lados por $p + qx_i^{in}(t)$ e integro en t

$$\int \frac{1}{p + qx_i^{in}} \frac{dx_i^{in}}{dt} dt = \int \frac{1}{t} dt$$

- Resolviendo las integrales resulta (c es una constante)

$$\ln\left(p + qx_i^{in}\right) = q \ln(t) + c$$

Resolviendo la ecuación diferencial (cont.)

- Tomando exponentes y usando $K = e^c$ se obtiene

$$\ln\left(p + q x_i^{in}(t)\right) = q \ln(t) + c \;\Rightarrow\; x_i^{in}(t) = \frac{1}{q}\left(K t^q - p\right)$$

Resolviendo la ecuación diferencial (cont.)

- Tomando exponentes y usando $K = e^c$ se obtiene

$$\ln\left(p + q x_i^{in}(t)\right) = q \ln(t) + c \;\Rightarrow\; x_i^{in}(t) = \frac{1}{q}\left(K t^q - p\right)$$

- Para determinar la constante K usamos la condición inicial

$$0 = x_i^{in}(i) = \frac{1}{q}\left(K i^q - p\right) \;\Rightarrow\; K = \frac{p}{i^q}$$

Resolviendo la ecuación diferencial (cont.)

- Tomando exponentes y usando $K = e^c$ se obtiene

$$\ln\left(p + q x_i^{in}(t)\right) = q \ln(t) + c \;\Rightarrow\; x_i^{in}(t) = \frac{1}{q}\left(K t^q - p\right)$$

- Para determinar la constante K usamos la condición inicial

$$0 = x_i^{in}(i) = \frac{1}{q}\left(K i^q - p\right) \;\Rightarrow\; K = \frac{p}{i^q}$$

- Y finalmente el límite fluido de $d_i^{in}(t)$ resulta

$$x_i^{in}(t) = \frac{1}{q}\left(\frac{p}{i^q} \times t^q - p\right) = \frac{p}{q}\left[\left(\frac{t}{i}\right)^q - 1\right]$$

¿Y la distribución de grados?

- Q: En tiempo t qué fracción $\bar{F}(d)$ de los nodos tiene grado $\geq d$?

  En términos del límite fluido: ¿Qué fracción de las funciones cumple $x_i^{in}(t) \geq d$ en tiempo t?

$$x_i^{in}(t) = \frac{p}{q}\left[\left(\frac{t}{i}\right)^q - 1\right] \geq d$$

¿Y la distribución de grados?

- Q: En tiempo t qué fracción $\bar{F}(d)$ de los nodos tiene grado $\geq d$?

  En términos del límite fluido: ¿Qué fracción de las funciones cumple $x_i^{in}(t) \geq d$ en tiempo t?

$$x_i^{in}(t) = \frac{p}{q}\left[\left(\frac{t}{i}\right)^q - 1\right] \geq d$$

- Se puede re-escribir en términos de i como

$$i \leq t\left[\left(\frac{q}{p}\right)d + 1\right]^{-1/q}$$

¿Y la distribución de grados?

- Q: En tiempo t qué fracción $\bar{F}(d)$ de los nodos tiene grado $\geq d$?

  En términos del límite fluido: ¿Qué fracción de las funciones cumple $x_i^{in}(t) \geq d$ en tiempo t?

$$x_i^{in}(t) = \frac{p}{q}\left[\left(\frac{t}{i}\right)^q - 1\right] \geq d$$

- Se puede re-escribir en términos de i como

$$i \leq t\left[\left(\frac{q}{p}\right)d + 1\right]^{-1/q}$$

- En tiempo t hay exactamente t nodos en el grafo, así que la fracción es

$$\bar{F}(d) = \left[\left(\frac{q}{p}\right)d + 1\right]^{-1/q} = 1 - F(d)$$

¿Y la distribución de grados?

- La distribución de grados está dada por la PDF $P(d)$
- Recordemos que la PDF, la CDF y la CCDF tienen la siguiente relación

$$p(x) = \frac{dF(x)}{dx} = -\frac{d\bar{F}(x)}{dx}$$

¿Y la distribución de grados?

- La distribución de grados está dada por la PDF $P(d)$
- Recordemos que la PDF, la CDF y la CCDF tienen la siguiente relación

$$p(x) = \frac{dF(x)}{dx} = -\frac{d\bar{F}(x)}{dx}$$

- Tomando la derivada de $\bar{F}(d) = \left[\left(\frac{q}{p}\right) d + 1\right]^{-1/q}$ resulta

$$P(d) = \frac{1}{p}\left[\left(\frac{q}{p}\right) d + 1\right]^{-(1+\frac{1}{q})}$$

$\Rightarrow$ "Mostramos" que $P(d) \propto d^{-(1+1/q)}$, una distribución tipo power law con exponente $\alpha = 1 + \frac{1}{1-p}$

Modelo de Barabási-Albert

- El modelo de Barabási-Albert (BA) es para grafos no-dirigidos
- El grafo inicial $G_{BA}(0)$ tiene $N_v(0)$ vértices y $N_e(0)$ aristas $(t = 0)$

FACULTAD DE INGENIERÍA UDELAR

Modelo de Barabási-Albert

- El modelo de Barabási-Albert (BA) es para grafos no-dirigidos
- El grafo inicial $G_{BA}(0)$ tiene $N_v(0)$ vértices y $N_e(0)$ aristas $(t = 0)$
- Para $t = 1, 2, \ldots$ el grafo actual $G_{BA}(t-1)$ "crece" a $G_{BA}(t)$:
  - Se agrega un vertice u de grado $d_u(t) = m \geq 1$
  - Las m aristas nuevas son incidentes a m vértices diferentes en $G_{BA}(t-1)$
  - El nuevo vertice u se conecta a $v \in V(t-1)$ con probabilidad

$$\mathrm{P}\left((u,v) \in E(t)\right) = \frac{d_v(t-1)}{\sum_{v'} d_{v'}(t-1)}$$

- El nuevo vértice prefiere nodos con grado más alto
- A. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509-512, 1999

FACULTAD DE INGENIERÍA UDELAR

Linearized Chord Diagram

- El modelo de BA tiene varios problemas formales (y no tanto):
  - No se especifica el grafo inicial
  - El número de aristas medio es 1 y no m
  - Se sortea un grupo de m vértices, pero no se dan las marginales de cada nodo
- El Linearized chord diagram (LCD) model (y otros) elimina estas ambigüedades

Linearized Chord Diagram

- El modelo de BA tiene varios problemas formales (y no tanto):
  - No se especifica el grafo inicial
  - El número de aristas medio es 1 y no m
  - Se sortea un grupo de m vértices, pero no se dan las marginales de cada nodo
- El Linearized chord diagram (LCD) model (y otros) elimina estas ambigüedades
- Para $m = 1$, se comienza con $G_{LCD}(0)$: un único nodo con un self-loop

Linearized Chord Diagram

- El modelo de BA tiene varios problemas formales (y no tanto):
  - No se especifica el grafo inicial
  - El número de aristas medio es 1 y no m
  - Se sortea un grupo de m vértices, pero no se dan las marginales de cada nodo
- El Linearized chord diagram (LCD) model (y otros) elimina estas ambigüedades
- Para $m = 1$, se comienza con $G_{LCD}(0)$: un único nodo con un self-loop
- Para $t = 1, 2, \ldots$ el grafo $G_{LCD}(t-1)$ "crece" a $G_{LCD}(t)$:
  - Se agrega un vértice v_t con una arista hacia $v_s \in V(t)$
  - El vértice destino v_s, $1 \leq s \leq t$ se elige con probabilidad

$$\mathrm{P}\left(s = j\right) = \begin{cases} \frac{d_{v_j}(t-1)}{2t-1}, & \text{si } 1 \leq j \leq t-1, \\ \frac{1}{2t-1}, & \text{si } j = t \end{cases}$$

Linearized Chord Diagram

- El modelo de BA tiene varios problemas formales (y no tanto):
  - No se especifica el grafo inicial
  - El número de aristas medio es 1 y no m
  - Se sortea un grupo de m vértices, pero no se dan las marginales de cada nodo
- El Linearized chord diagram (LCD) model (y otros) elimina estas ambigüedades
- Para $m = 1$, se comienza con $G_{LCD}(0)$: un único nodo con un self-loop
- Para $t = 1, 2, \ldots$ el grafo $G_{LCD}(t-1)$ "crece" a $G_{LCD}(t)$:
  - Se agrega un vértice v_t con una arista hacia $v_s \in V(t)$
  - El vértice destino v_s, $1 \leq s \leq t$ se elige con probabilidad

$$\mathrm{P}\,(s = j) = \begin{cases} \frac{d_{v_j}(t-1)}{2t-1}, & \text{si } 1 \leq j \leq t-1, \\ \frac{1}{2t-1}, & \text{si } j = t \end{cases}$$

- Para $m > 1$ simplemente se repite lo anterior m veces en cada t
  - Los m vértices creados se "colapsan" en un único vértice, manteniendo las aristas
- A. Bollobás et al, "The degree sequence of a scale-free random graph process," *Random Struct. and Alg.*, vol. 18, pp. 279-290, 2001

Algunas propiedades del modelo PA

P1) Cuando $t \to \infty$, $G_{LCD}(t)$ tiene distribución de grados tipo power-law con $\alpha = 3$

$\Rightarrow$ Se genera $\alpha = 3 + \delta/m > 2$ agregando un bias $\delta > -m$ a las probabilidades

$\Rightarrow$ Ver Remco van der Hofstad, "Random Graphs and Complex Networks. Volume One." Cambridge Series in Statistical and Probabilistic Mathematics (2017)

Algunas propiedades del modelo PA

P1) Cuando $t \to \infty$, $G_{LCD}(t)$ tiene distribución de grados tipo power-law con $\alpha = 3$

$\Rightarrow$ Se genera $\alpha = 3 + \delta/m > 2$ agregando un bias $\delta > -m$ a las probabilidades

$\Rightarrow$ Ver Remco van der Hofstad, "Random Graphs and Complex Networks. Volume One." Cambridge Series in Statistical and Probabilistic Mathematics (2017)

P2) El grafo $G_{LCD}(t)$ es conexo w.h.p.

Algunas propiedades del modelo PA

P1) Cuando $t \to \infty$, $G_{LCD}(t)$ tiene distribución de grados tipo power-law con $\alpha = 3$

$\Rightarrow$ Se genera $\alpha = 3 + \delta/m > 2$ agregando un bias $\delta > -m$ a las probabilidades

$\Rightarrow$ Ver Remco van der Hofstad, "Random Graphs and Complex Networks. Volume One." Cambridge Series in Statistical and Probabilistic Mathematics (2017)

P2) El grafo $G_{LCD}(t)$ es conexo w.h.p.

P3) Diámetro pequeño

$$\mathrm{diam}(G_{LCD}(t)) = \begin{cases} O(\log N_v(t)), & m = 1 \\ O\left(\frac{\log N_v(t)}{\log\log N_v(t)}\right), & m > 1 \end{cases}$$

Algunas propiedades del modelo PA

P1) Cuando $t \to \infty$, $G_{LCD}(t)$ tiene distribución de grados tipo power-law con $\alpha = 3$

$\Rightarrow$ Se genera $\alpha = 3 + \delta/m > 2$ agregando un bias $\delta > -m$ a las probabilidades

$\Rightarrow$ Ver Remco van der Hofstad, "Random Graphs and Complex Networks. Volume One." Cambridge Series in Statistical and Probabilistic Mathematics (2017)

P2) El grafo $G_{LCD}(t)$ es conexo w.h.p.

P3) Diámetro pequeño

$$\mathrm{diam}(G_{LCD}(t)) = \begin{cases} O(\log N_v(t)), & m = 1 \\ O(\frac{\log N_v(t)}{\log \log N_v(t)}), & m > 1 \end{cases}$$

P4) Clustering pequeño, pues para $m > 1$

$$\mathbb{E}\left[\mathrm{cl}(G_{LCD}(t))\right] \approx \frac{m-1}{8} \frac{(\log N_v(t))^2}{N_v(t)}$$

$\Rightarrow$ Apenas mejor que el $O(N_v^{-1})$ de los grafos aleatorios clásicos

Modelos de copia

- La copia es otro mecanismo de interés

  Ej: la duplicación de genes para re-usar información en la evolución de organismos

- Ya vimos que copiar lleva distribuciones tipo power law
  - Preferential Attachment genera $\alpha > 2$, pero muchas redes biológicas tienen $1 < \alpha < 2$
  - Copiar varias decisiones de otro nodo puede generar esos exponentes

Modelos de copia

- La copia es otro mecanismo de interés

  Ej: la duplicación de genes para re-usar información en la evolución de organismos

- Ya vimos que copiar lleva distribuciones tipo power law
  - Preferential Attachment genera $\alpha > 2$, pero muchas redes biológicas tienen $1 < \alpha < 2$
  - Copiar varias decisiones de otro nodo puede generar esos exponentes

- Inicializo con el grafo $G_C(0)$ $(t = 0)$

Modelos de copia

- La copia es otro mecanismo de interés

  Ej: la duplicación de genes para re-usar información en la evolución de organismos

- Ya vimos que copiar lleva distribuciones tipo power law
  - Preferential Attachment genera $\alpha > 2$, pero muchas redes biológicas tienen $1 < \alpha < 2$
  - Copiar varias decisiones de otro nodo puede generar esos exponentes

- Inicializo con el grafo $G_C(0)$ $(t = 0)$
- Para $t = 1, 2, \ldots$ el grafo actual $G_C(t-1)$ "crece" a $G_C(t)$:
  - Agregar un vértice nuevo u
  - Elegir otro vértice $v \in V(t-1)$ aleatorio uniforme (i.e. con proba $\frac{1}{N_v(t-1)}$)
  - Unir el nuevo vértice u con los vecinos de v independiente con probabilidad p

Modelos de copia

- La copia es otro mecanismo de interés

  Ej: la duplicación de genes para re-usar información en la evolución de organismos

- Ya vimos que copiar lleva distribuciones tipo power law
  - Preferential Attachment genera $\alpha > 2$, pero muchas redes biológicas tienen $1 < \alpha < 2$
  - Copiar varias decisiones de otro nodo puede generar esos exponentes

- Inicializo con el grafo $G_C(0)$ $(t = 0)$
- Para $t = 1, 2, \dots$ el grafo actual $G_C(t-1)$ "crece" a $G_C(t)$:
  - Agregar un vértice nuevo u
  - Elegir otro vértice $v \in V(t-1)$ aleatorio uniforme (i.e. con proba $\frac{1}{N_v(t-1)}$)
  - Unir el nuevo vértice u con los vecinos de v independiente con probabilidad p

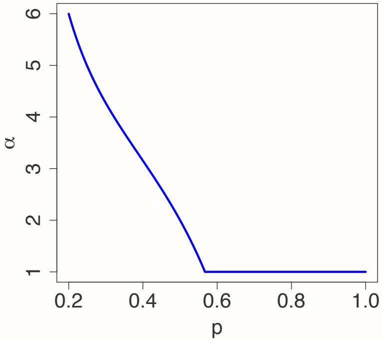
- El caso $p = 1$ lleva a copiar totalmente las aristas de un nodo pre-existente

- F. Chung et al, "Duplication models for biological networks," *Journal of Computational Biology*, vol. 10, pp. 677-687, 2003

Distribución de grados asintótica

- La distribución de grados tiende a una power law w.h.p. [Chung et al'03]
  $\Rightarrow$ El exponente α en la curva es la solución de la ecuación

$$p(\alpha - 1) = 1 - p^{\alpha - 1}$$



- Duplicación total ($p = 1$) no genera power-law; aunque sí lo hace si
  $\Rightarrow$ se copia parcialmente una fracción $q \in (0, 1)$ de veces

Modelos de Grafos Aleatorios

¿Un mundo pequeño?

- Seis grados de separación se volvió popular a partir de una obra de teatro [Guare'90]
  - ⇒ Caminos cortos entre nosotros y cualquiera en el mundo
  - ⇒ El término es relativamente nuevo, no así el concepto

¿Un mundo pequeño?

- Seis grados de separación se volvió popular a partir de una obra de teatro [Guare'90]
  - ⇒ Caminos cortos entre nosotros y cualquiera en el mundo
  - ⇒ El término es relativamente nuevo, no así el concepto
- El escritor húngaro Frigyes Karinthy lo describe en uno de sus cuentos de 1929
  - ⇒ El mundo moderno se está 'achicando' debido al aumento en la conectividad humana
  - ⇒ Apuesta que puede encontrar a cualquier persona usando no más de 5 individuos, siendo uno de ellos un conocido suyo, y todos usando su red contactos personales

¿Un mundo pequeño?

- Seis grados de separación se volvió popular a partir de una obra de teatro [Guare'90]
  - ⇒ Caminos cortos entre nosotros y cualquiera en el mundo
  - ⇒ El término es relativamente nuevo, no así el concepto
- El escritor húngaro Frigyes Karinthy lo describe en uno de sus cuentos de 1929
  - ⇒ El mundo moderno se está 'achicando' debido al aumento en la conectividad humana
  - ⇒ Apuesta que puede encontrar a cualquier persona usando no más de 5 individuos, siendo uno de ellos un conocido suyo, y todos usando su red contactos personales
- Primer tratamiento matemático en [Kochen-Pool'50]
  - ⇒ Modela formalmente los mecanismos de las redes sociales
  - ⇒ Pero la cuestión de los 'grados de separación' quedó sin responder

¿Un mundo pequeño?

- Seis grados de separación se volvió popular a partir de una obra de teatro [Guare'90]

  ⇒ Caminos cortos entre nosotros y cualquiera en el mundo

  ⇒ El término es relativamente nuevo, no así el concepto

- El escritor húngaro Frigyes Karinthy lo describe en uno de sus cuentos de 1929

  ⇒ El mundo moderno se está 'achicando' debido al aumento en la conectividad humana

  ⇒ Apuesta que puede encontrar a cualquier persona usando no más de 5 individuos, siendo uno de ellos un conocido suyo, y todos usando su red contactos personales

- Primer tratamiento matemático en [Kochen-Pool'50]

  ⇒ Modela formalmente los mecanismos de las redes sociales

  ⇒ Pero la cuestión de los 'grados de separación' quedó sin responder

  ⇒ Este trabajo a su vez inspira el famoso experimento en [Milgram'67]

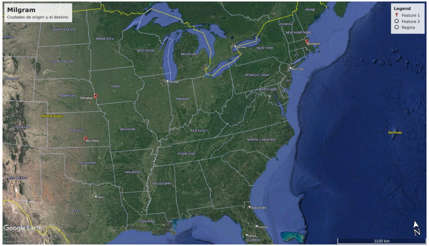
- Mucha más info e historia en Schnettler, Sebastian. "A structured overview of 50 years of small-world research." Social networks 31, no. 3 (2009).

Experimento de Milgram

- Q1: ¿Cuál es la distancia típica entre dos personas?
  - ⇒ La idea es medirlo en la red social (global) de amistades
  - ⇒ Dado que no puedo obtener la red completa, es necesario estimar las distancias

Experimento de Milgram

- Q1: ¿Cuál es la distancia típica entre dos personas?
  - $\Rightarrow$ La idea es medirlo en la red social (global) de amistades
  - $\Rightarrow$ Dado que no puedo obtener la red completa, es necesario estimar las distancias
- La (muy ingeniosa) solución de S. Milgram en su famoso experimento de 1967
  - Se enviaron 296 cartas a personas en Wichita, Kansas y Omaha, Nebraska
  - Las cartas indicaban una (única) persona de contacto en Boston, Massachusetts
  - Las instrucciones eran hacer llegar la carta al contacto siguiendo las siguientes reglas



Experimento de Milgram

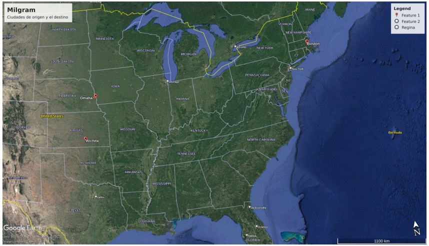
- Q1: ¿Cuál es la distancia típica entre dos personas?
  - ⇒ La idea es medirlo en la red social (global) de amistades
  - ⇒ Dado que no puedo obtener la red completa, es necesario estimar las distancias
- La (muy ingeniosa) solución de S. Milgram en su famoso experimento de 1967
  - Se enviaron 296 cartas a personas en Wichita, Kansas y Omaha, Nebraska
  - Las cartas indicaban una (única) persona de contacto en Boston, Massachusetts
  - Las instrucciones eran hacer llegar la carta al contacto siguiendo las siguientes reglas

- Amigo: alguien que tratamos por nombre de pila

  Regla 1: Si el contacto es un amigo, enviarlo al amigo; sino

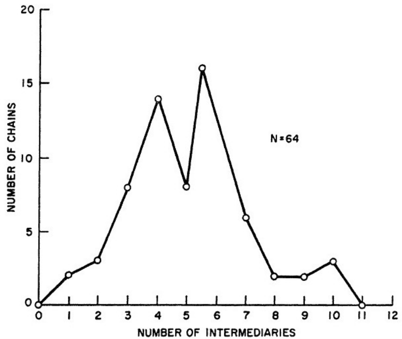
  Regla 2: Re-enviarlo al amigo que más probablemente sea amigo del contacto

- Q2: ¿Cuántas llegaron? ¿Cuánto demoraron?



Resultados del experimento de Milgram

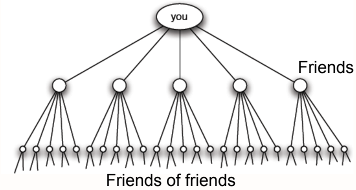
- 64 de las 296 cartas llegaron al destino, con un largo promedio de $\bar{\ell} = 6{,}2$
  - $\Rightarrow$ Esto a su vez inspiró la obra de Guare
- Conclusión: caminos cortos conectan dos personas cualesquiera

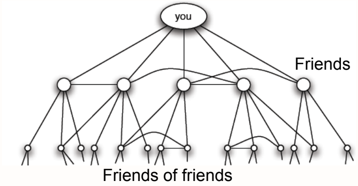


- S. Milgram, "The small-world problem," *Psychology Today*, vol. 2, pp. 60-67, 1967

¿Qué pasa acá?

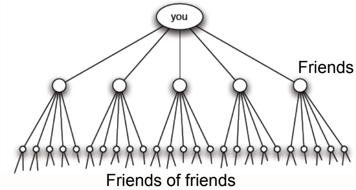
- Milgram muestra que los caminos cortos existen y en abundancia
- Q: ¿Es la teoría del mundo pequeño razonable? Por supuesto si asumimos:
  - Tenemos 100 amigos, cada cual con otros 100 amigo, ...
  - Después de 5 grados tenemos 10^{10} personas $>$ doble de la población de la Tierra

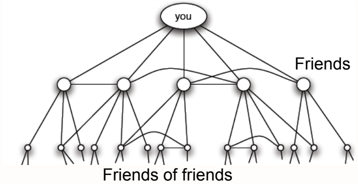




¿Qué pasa acá?

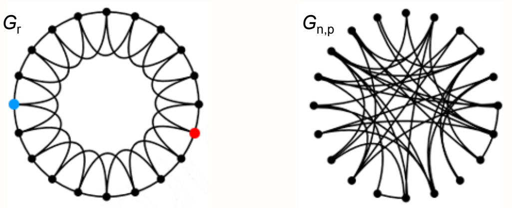
- Milgram muestra que los caminos cortos existen y en abundancia
- Q: ¿Es la teoría del mundo pequeño razonable? Por supuesto si asumimos:
  - Tenemos 100 amigos, cada cual con otros 100 amigo, ...
  - Después de 5 grados tenemos 10^{10} personas > doble de la población de la Tierra





- No parece un modelo razonable para una red social, que típicamente es:
  - $\Rightarrow$ Homofilia [Lazarzfeld'54]
  - $\Rightarrow$ Abundantes triángulos cerrados [Rapoport'53]
- Q: ¿Una red muy estructurada localmente **y** globalmente pequeña? ¿Cómo?

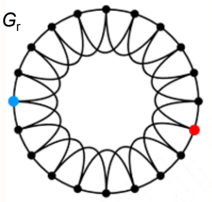
Estructura y aleatoriedad como extremos



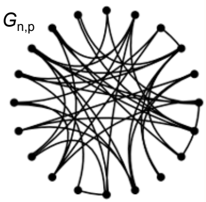
High clustering and diameter   Low clustering and diameter

- Lattice regular uni-dimensional G_r de N_v vértices
  - Cada vértice está conectado a sus $2r$ vecinos más cercanos (r a cada lado)

Estructura y aleatoriedad como extremos



High clustering and diameter



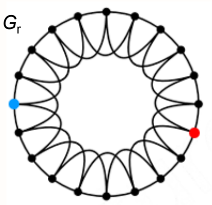
Low clustering and diameter

- Lattice regular uni-dimensional G_r de N_v vértices
  - Cada vértice está conectado a sus $2r$ vecinos más cercanos (r a cada lado)

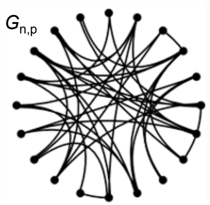
Tanta estructura resulta en clustering alto, pero también en un diámetro alto

$$\mathrm{cl}(G_r) = \frac{3r-3}{4r-2} \text{ y } \mathrm{diam}(G_r) = \frac{N_v}{2r}$$

Estructura y aleatoriedad como extremos



High clustering and diameter



Low clustering and diameter

- Lattice regular uni-dimensional G_r de N_v vértices
  - Cada vértice está conectado a sus $2r$ vecinos más cercanos (r a cada lado)

  Tanta estructura resulta en clustering alto, pero también en un diámetro alto

$$\mathrm{cl}(G_r) = \frac{3r-3}{4r-2} \text{ y } \mathrm{diam}(G_r) = \frac{N_v}{2r}$$

- El otro extremo es un grafo aleatorio $\mathrm{ER}(N_v, p)$ con $p = O(N_v^{-1})$

  Ya vimos que esta aleatoriedad generar un diámetro pequeño, pero un bajo clustering

$$\mathrm{cl}(G_{N_v,p}) = O(N_v^{-1}) \text{ y } \mathrm{diam}(G_{N_v,p}) = O(\log N_v)$$

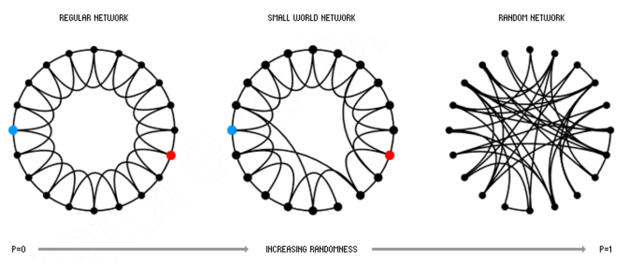
El modelo Watts-Strogatz

- Small-world model: mezcla estructura con una pizca de aleatoriedad

  1: Inicializamos con un lattice regular que tenga el clustering buscado

  2: La aleatoriedad se genera introduciendo atajos en el grafo

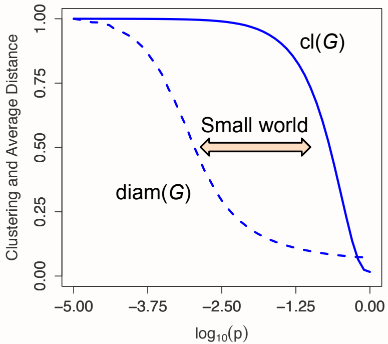
  $\Rightarrow$ Cada arista se re-conecta aleatoriamente con probabilidad (pequeña) p



- La reconexión interpola entre los extremos regular y puramente aleatorio

Resultados numéricos

- Simulación del modelo Watts-Strogatz con $N_v = 1,000$ y $r = 6$
  - Probabilidad de reconexión p variando de 0 (lattice G_r) a 1 ($\mathrm{ER}(N_v, 2r/N_v)$)
  - $\mathrm{cl}(G)$ y $\mathrm{diam}(G)$ normalizados respecto al valor máximo ($p = 0$)



- Intervalo importante de $p \in [10^{-3}, 10^{-1}]$ resulta en $\mathrm{diam}(G)$ pequeño y $\mathrm{cl}(G)$ grande

El modelo Watts-Strogatz

- Propiedades estructurales del modelo de Watts-Strogatz

  P1: Para N_v grande el clusetring coefficient resulta

$$\mathrm{cl}(G) \approx \frac{3r-3}{4r-2}(1-p^3) = \mathrm{cl}(G_r)(1-p^3)$$

  P2: La distribución de grados se concentra alrededor de su media $2r$

El modelo Watts-Strogatz

- Propiedades estructurales del modelo de Watts-Strogatz

  P1: Para N_v grande el clusetring coefficient resulta

$$\mathrm{cl}(G) \approx \frac{3r-3}{4r-2}(1-p^3) = \mathrm{cl}(G_r)(1-p^3)$$

  P2: La distribución de grados se concentra alrededor de su media $2r$

- Los modelos y propiedades tipo small-world son importantes en varias disciplinas
- Particularmente en “comunicación” en el sentido amplio

  $\Rightarrow$ Dispersión de rumores, chismes, noticias (falsas)

  $\Rightarrow$ Dispersión de enfermedades y epidemias

  $\Rightarrow$ Búsqueda de información en redes

Modelos de Grafos Aleatorios

Exponential random graph models

- Un buen modelo estadístico debería ser [Robbins-Morris'07]
  - ✓ Estimable a partir de y razonablemente representativo de los datos observados
  - ✓ Plausible teóricamente sobre los efectos que pueden haber producido la red
  - ✓ Capaz de discriminar entre los distintos efectos que mejor explican los datos

Exponential random graph models

- Un buen modelo estadístico debería ser [Robbins-Morris'07]
  - ✓ Estimable a partir de y razonablemente representativo de los datos observados
  - ✓ Plausible teóricamente sobre los efectos que pueden haber producido la red
  - ✓ Capaz de discriminar entre los distintos efectos que mejor explican los datos
- Los modelos del tipo Exponential random graph models (ERGMs) (también conocidos como p^* models) se diseñaron específicamente con estos criterios en mente
- G. Robbins et al., "An introduction to exponential random graph (p^*) models for social networks," *Social Networks*, vol. 29, pp. 173-191, 2007

Distribuciones exponenciales

- **Def:** Un vector aleatorio discreto $\mathbf{Z} \in \mathcal{Z}$ pertenece a la familia de las exponenciales si

$$\mathrm{P}_\theta(\mathbf{Z} = \mathbf{z}) = \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{z}) - \psi(\theta)\right\}$$

  - $\boldsymbol{\theta} \in \mathbb{R}^p$ es un vector de parámetros y $\mathbf{g} : \mathcal{Z} \mapsto \mathbb{R}^p$ es una función
  - $\psi(\theta)$ es simplemente para normalizar, mateniendo $\sum_{\mathbf{z}\in\mathcal{Z}} \mathrm{P}_\theta(\mathbf{z}) = 1$
  - Ej: Bernoulli, binomial, Poisson, geométrica

Distribuciones exponenciales

- **Def:** Un vector aleatorio discreto $\mathbf{Z} \in \mathcal{Z}$ pertenece a la familia de las exponenciales si

$$\mathrm{P}_\theta(\mathbf{Z} = \mathbf{z}) = \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{z}) - \psi(\theta)\right\}$$

  - $\boldsymbol{\theta} \in \mathbb{R}^p$ es un vector de parámetros y $\mathbf{g} : \mathcal{Z} \mapsto \mathbb{R}^p$ es una función
  - $\psi(\theta)$ es simplemente para normalizar, mateniendo $\sum_{\mathbf{z}\in\mathcal{Z}} \mathrm{P}_\theta(\mathbf{z}) = 1$
  - Ej: Bernoulli, binomial, Poisson, geométrica

- Las distribuciones continuas tienen una forma similar

  Ej: gaussiana, Pareto, chi-cuadrado

- Estas distribuciones comparten propiedades algebraicas y geométricas útiles

  $\Rightarrow$ Las hace matemáticamente conveniente para inferencia y simulación

Exponential random graph model

- Sea $G(V, E)$ un grafo aleatorio no-dirigido, con $Y_{ij} := \mathbb{I}\{(i, j) \in E\}$
  - La matriz $\mathbf{Y} = [Y_{ij}]$ es la de adyacencia (aleatoria), $\mathbf{y} = [y_{ij}]$ una realización

Exponential random graph model

- Sea $G(V,E)$ un grafo aleatorio no-dirigido, con $Y_{ij} := \mathbb{I}\{(i,j) \in E\}$
  - La matriz $\mathbf{Y} = [Y_{ij}]$ es la de adyacencia (aleatoria), $\mathbf{y} = [y_{ij}]$ una realización
- Un ERGM especifica una forma exponencial para la distribución de $\mathbf{Y}$, i.e.,

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\left\{\sum_H \theta_H g_H(\mathbf{y})\right\}, \quad \text{donde}$$

(i) cada H es una configuración: un conjunto posible de aristas en G;
(ii) $g_H(\mathbf{y})$ es el estadístico de red correspondiente a la configuración H

$$g_H(\mathbf{y}) = \prod_{y_{ij} \in H} y_{ij} = \mathbb{I}\{H \text{ está en } \mathbf{y}\}$$

(iii) $\theta_H \neq 0$ solo si las aristas de H son condicionalmente dependendientes; y
(iv) $\kappa(\boldsymbol{\theta})$ es una constante normalizadora que mantiene $\sum_{\mathbf{y}} \mathrm{P}_\theta(\mathbf{y}) = 1$

Discusión

- Los vértices del grafo N_v y su orden están fijos y dados, sólo las aristas son aleatorias
  - $\Rightarrow$ Asumimos aristas no-dirigidas y sin pesos. Extensiones posibles

Discusión

- Los vértices del grafo N_v y su orden están fijos y dados, sólo las aristas son aleatorias
  - $\Rightarrow$ Asumimos aristas no-dirigidas y sin pesos. Extensiones posibles
- ERGMs describen el grafo aleatorio a partir de patrones locales
  - Estas configuraciones son las características estructurales de interés
  - Ej: ¿Hay efectos de reciprocidad? Agregar arcos mutuos en las configuraciones
  - Ej: ¿Hay efectos de transitividad? Agreguemos triángulos

Discusión

- Los vértices del grafo N_v y su orden están fijos y dados, sólo las aristas son aleatorias

  $\Rightarrow$ Asumimos aristas no-dirigidas y sin pesos. Extensiones posibles

- ERGMs describen el grafo aleatorio a partir de patrones locales
  - Estas configuraciones son las características estructurales de interés
  - Ej: ¿Hay efectos de reciprocidad? Agregar arcos mutuos en las configuraciones
  - Ej: ¿Hay efectos de transitividad? Agreguemos triángulos

- La (in)dependencia es condicional en todas las otras variables (aristas) en G

  $\Rightarrow$ Se controla las configuraciones relevantes (i.e., $\theta_H \neq 0$) para el modelo

- Ciertas dependencias implican modelos particulares

¿Cómo construir un modelo?

- Al usar un ERGM para una red los pasos son básicamente los siguientes
  - ⇒ Cuidado: hay varias decisiones explícitas que tienen un fuerte impacto

Paso 1: Cada arista (relación) es una variable aleatoria

Paso 2: Se propone una hipótesis de dependencia

Paso 3: La hipótesis de dependencia implica una forma particular para el modelo

Paso 4: El modelo se simplifica, e.g., a través de la homogeneización

Paso 5: Estimar e interpretar los resultados

Ejemplo: grafos aleatorios de Bernoulli

- Hipótesis: cada arista está presente independientemente de todas las demás (e.g., ER(n, p))
  - $\Rightarrow$ La hipótesis más simple (e irrealista) de dependencia

Ejemplo: grafos aleatorios de Bernoulli

- Hipótesis: cada arista está presente independientemente de todas las demás (e.g., ER(n, p))

  $\Rightarrow$ La hipótesis más simple (e irrealista) de dependencia

- Para cada (i, j), suponemos Y_{ij} independiente de Y_{uv}, para todo $(u, v) \neq (i, j)$

  $\Rightarrow$ $\theta_H = 0$ para todo H que involucre dos o más aristas

- Sólo la configuración de arista, i.e., $g_H(\mathbf{y}) = y_{ij}$, es relevante, y el ERGM resulta

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\left\{\sum_{i,j} \theta_{ij} y_{ij}\right\}$$

Ejemplo: grafos aleatorios de Bernoulli

- Hipótesis: cada arista está presente independientemente de todas las demás (e.g., ER(n, p))

  $\Rightarrow$ La hipótesis más simple (e irrealista) de dependencia

- Para cada (i, j), suponemos Y_{ij} independiente de Y_{uv}, para todo $(u, v) \neq (i, j)$

  $\Rightarrow$ $\theta_H = 0$ para todo H que involucre dos o más aristas

- Sólo la configuración de arista, i.e., $g_H(\mathbf{y}) = y_{ij}$, es relevante, y el ERGM resulta

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\left\{\sum_{i,j} \theta_{ij} y_{ij}\right\}$$

$\Rightarrow$ Cada arista existe de forma independiente con probabilidad $\propto\ e^{\theta_{ij}}$ y no existe con probabilidad $\propto\ 1$

$$\Rightarrow p_{ij} = \frac{\exp(\theta_{ij})}{1 + \exp(\theta_{ij})}$$

Restricción de los parámetros: homogeneidad

- Demasiado parámetros hacen la estimación a partir de un único $\mathbf{y}$ imposible
  - $\Rightarrow$ En este caso, tenemos N_v^2 parámetros $\{\theta_{ij}\}$. ¿Reducción?

Restricción de los parámetros: homogeneidad

- Demasiado parámetros hacen la estimación a partir de un único $\mathbf{y}$ imposible
  - $\Rightarrow$ En este caso, tenemos N_v^2 parámetros $\{\theta_{ij}\}$. ¿Reducción?
- Homogeneidad en G, i.e., $\theta_{ij} = \theta$ para todo (i, j) resulta en

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\{\theta L(\mathbf{y})\}$$

  - El estadístico suficiente en este caso es el número de aristas observadas $L(\mathbf{y}) = \sum_{i,j} y_{ij}$
  - Este ejemplo de ERGM resulta en un $\mathrm{ER}(n, p)$, con $p = \dfrac{\exp\theta}{1 + \exp\theta}$

Restricción de los parámetros: homogeneidad

- Demasiado parámetros hacen la estimación a partir de un único $\mathbf{y}$ imposible
  - $\Rightarrow$ En este caso, tenemos N_v^2 parámetros $\{\theta_{ij}\}$. ¿Reducción?
- Homogeneidad en G, i.e., $\theta_{ij} = \theta$ para todo (i, j) resulta en

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\{\theta L(\mathbf{y})\}$$

  - El estadístico suficiente en este caso es el número de aristas observadas $L(\mathbf{y}) = \sum_{i,j} y_{ij}$
  - Este ejemplo de ERGM resulta en un ER(n, p), con $p = \dfrac{\exp\theta}{1 + \exp\theta}$
- Otro ejemplo: sabemos a priori que hay dos tipos de nodos y cuáles son
  - $\Rightarrow$ Podemos imponer homogeneidad en las aristas intra e inter-conjuntos, i.e.,

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\{\theta_1 L_1(\mathbf{y}) + \theta_{12} L_{12}(\mathbf{y}) + \theta_2 L_2(\mathbf{y})\}$$

Más ejemplos: Markov random graphs

- Dependencia tipo Markov para grafos [Frank-Strauss'86]
  - Asumimos que dos aristas son dependientes si comparten un vértice
  - La existencia de la arista (Y_{ij}) depende de todas las otras aristas que involucren i o j
  - Intuición: Los amigos de dos personas nos dicen mucho sobre si son amigos o no

Más ejemplos: Markov random graphs

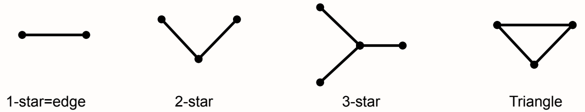
- Dependencia tipo Markov para grafos [Frank-Strauss'86]
  - Asumimos que dos aristas son dependientes si comparten un vértice
  - La existencia de la arista (Y_{ij}) depende de todas las otras aristas que involucren i o j
  - Intuición: Los amigos de dos personas nos dicen mucho sobre si son amigos o no

Teorema

Suponiendo homogeneidad, G es un Markov random graph sii

$$P_\theta(\mathbf{Y}=\mathbf{y}) = \left(\frac{1}{\kappa(\boldsymbol{\theta})}\right) \exp\left\{\sum_{k=1}^{N_v-1} \theta_k S_k(\mathbf{y}) + \theta_\tau T(\mathbf{y})\right\}, \quad \textit{donde}$$

$S_k(\mathbf{y})$ es el número de k-estrellas, y $T(\mathbf{y})$ el número de triángulos



Estadísticos Alternativos

- Incluir términos para k alto es claramente un desafío
  - $\Rightarrow$ Solución típica: efectos de las estrellas grandes se omiten, e.g., $\theta_k = 0$, $k \geq 4$
  - $\Rightarrow$ Pero esto resulta en modelos que ajustan pobremente los datos. ¿Qué hacemos?

Estadísticos Alternativos

- Incluir términos para k alto es claramente un desafío
  - $\Rightarrow$ Solución típica: efectos de las estrellas grandes se omiten, e.g., $\theta_k = 0$, $k \geq 4$
  - $\Rightarrow$ Pero esto resulta en modelos que ajustan pobremente los datos. ¿Qué hacemos?
- Idea: Tomemos una forma paramétrica para $\theta_k \propto (-1)^k \lambda^{2-k}$ [Snijders et al'06]
  - Combinamos $S_k(\mathbf{y})$, $k \geq 2$ en un único estadístico denominado alternating k-star, i.e.,

$$\mathrm{AKS}_\lambda(\mathbf{y}) = \sum_{k=2}^{N_v-1} (-1)^k \frac{S_k(\mathbf{y})}{\lambda^{k-2}}, \quad \lambda > 1$$

Estadísticos Alternativos

- Incluir términos para k alto es claramente un desafío
  - $\Rightarrow$ Solución típica: efectos de las estrellas grandes se omiten, e.g., $\theta_k = 0$, $k \geq 4$
  - $\Rightarrow$ Pero esto resulta en modelos que ajustan pobremente los datos. ¿Qué hacemos?
- Idea: Tomemos una forma paramétrica para $\theta_k \propto (-1)^k \lambda^{2-k}$ [Snijders et al'06]
  - Combinamos $S_k(\mathbf{y})$, $k \geq 2$ en un único estadístico denominado alternating k-star, i.e.,

$$\mathrm{AKS}_\lambda(\mathbf{y}) = \sum_{k=2}^{N_v-1} (-1)^k \frac{S_k(\mathbf{y})}{\lambda^{k-2}}, \quad \lambda > 1$$

  - Se puede probar que $\mathrm{AKS}_\lambda(\mathbf{y}) \propto$ la media exponencial de grados

$$\mathrm{GWD}_\gamma(\mathbf{y}) = \sum_{d=0}^{N_v-1} e^{-\gamma d} N_d(\mathbf{y}), \quad \gamma > 0$$

$\Rightarrow$ $N_d(\mathbf{y})$ es el número de vértices de grado d

Incorporando atributos de los vértices

- Es relativamente sencillo incorporar atributos de los vértices a los ERGMs
  Ej: género, jerarquía en la organización, función dentro del sistema

Incorporando atributos de los vértices

- Es relativamente sencillo incorporar atributos de los vértices a los ERGMs

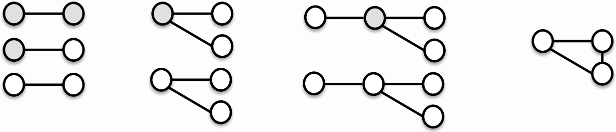
  Ej: género, jerarquía en la organización, función dentro del sistema

  - Tomamos una realización $\mathbf{x}$ de un vector aleatorio $\mathbf{X} \in \mathbb{R}^{N_v}$ definido en V
  - Especificamos una familia exponencial para la distribución condicional

$$\mathrm{P}_\theta(\mathbf{Y} = \mathbf{y} \,|\, \mathbf{X} = \mathbf{x})$$

$\Rightarrow$ Se incluyen estadísticas adicionales $g(\cdot)$ sobre $\mathbf{y}$ y $\mathbf{x}$

  - Ej: configuraciones para Markov, con atributos binarios



Estimación de los parámetros del ERGM

- MLE para el vector $\boldsymbol{\theta}$ en un ERGM es

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ \boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\boldsymbol{\theta}) \right\}, \quad \text{donde } \psi(\boldsymbol{\theta}) := \log \kappa(\boldsymbol{\theta})$$

Estimación de los parámetros del ERGM

- MLE para el vector $\boldsymbol{\theta}$ en un ERGM es

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ \boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\boldsymbol{\theta}) \right\}, \qquad \text{donde } \psi(\boldsymbol{\theta}) := \log \kappa(\boldsymbol{\theta})$$

- Pero $\psi(\boldsymbol{\theta})$ implica la suma de $2^{\binom{N_v}{2}}$ posibles valores de $\mathbf{y}$ para cada posible $\boldsymbol{\theta}$

  $\Rightarrow$ Se necesitan métodos numéricos para aproximarlo

Estimación de los parámetros del ERGM

- MLE para el vector $\boldsymbol{\theta}$ en un ERGM es

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ \boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\boldsymbol{\theta}) \right\}, \qquad \text{donde } \psi(\boldsymbol{\theta}) := \log \kappa(\boldsymbol{\theta})$$

- Pero $\psi(\boldsymbol{\theta})$ implica la suma de $2^{\binom{N_v}{2}}$ posibles valores de $\mathbf{y}$ para cada posible $\boldsymbol{\theta}$

  $\Rightarrow$ Se necesitan métodos numéricos para aproximarlo

- Idea: para valores fijos $\boldsymbol{\theta}_0$, maximizar log-likelihood ratio

$$r(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = \ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta}_0) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{y}) - [\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)]$$

- Identidad clave: (ver transparencias al final)

$$\exp\{\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)\} = \mathbb{E}_{\theta_0}\left[\exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{Y})\right\}\right]$$

Markov chain Monte Carlo MLE

- En limpio: para valores fijos $\boldsymbol{\theta}_0$, maximizar log-likelihood ratio

$$r(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{y}) - \log\left(\mathbb{E}_{\theta_0}\left[\exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{Y})\right\}\right]\right)$$

Markov chain Monte Carlo MLE

- En limpio: para valores fijos $\boldsymbol{\theta}_0$, maximizar log-likelihood ratio

$$r(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = (\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{y}) - \log\left(\mathbb{E}_{\theta_0}\left[\exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{Y})\right\}\right]\right)$$

- Markov chain Monte Carlo MLE para buscar $\boldsymbol{\theta}$

  Paso 1: obtener muestras $\mathbf{y}_1, \ldots, \mathbf{y}_n$ del ERGM usando $\boldsymbol{\theta}_0$

  Paso 2: aproximar $\mathbb{E}_{\theta_0}[\cdot]$ a través de la media de los $\exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{y}_k)\right\}$

  Paso 3: aproximar $\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)$ mediante el logaritmo de lo anterior

  Paso 4: evaluar el log-likelihood ratio $r(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ aproximado

Estimación de los parámetros del ERGM revisited

- MLE para el vector $\boldsymbol{\theta}$ en un ERGM es

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ \boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\boldsymbol{\theta}) \right\}, \quad \text{donde } \psi(\boldsymbol{\theta}) := \log \kappa(\boldsymbol{\theta})$$

Estimación de los parámetros del ERGM revisited

- MLE para el vector $\boldsymbol{\theta}$ en un ERGM es

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ \boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\boldsymbol{\theta}) \right\}, \qquad \text{donde } \psi(\boldsymbol{\theta}) := \log \kappa(\boldsymbol{\theta})$$

- Derivando e igualando a cero

$$\mathbf{g}(\mathbf{y}) = \nabla \psi(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

  - ¿Cambiamos una expresión horrible por su derivada?

Estimación de los parámetros del ERGM revisited

- MLE para el vector $\boldsymbol{\theta}$ en un ERGM es

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \left\{ \boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\boldsymbol{\theta}) \right\}, \quad \text{donde } \psi(\boldsymbol{\theta}) := \log \kappa(\boldsymbol{\theta})$$

- Derivando e igualando a cero

$$\mathbf{g}(\mathbf{y}) = \nabla \psi(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

  - ¿Cambiamos una expresión horrible por su derivada?
- Usando que $\mathbb{E}_{\theta}[\mathbf{g}(\mathbf{Y})] = \nabla \psi(\boldsymbol{\theta})$ (ver slides al final), el MLE es la solución de

$$\mathbb{E}_{\hat{\theta}}[\mathbf{g}(\mathbf{Y})] = \mathbf{g}(\mathbf{y})$$

que puede resolverse usando el método clásico de Robbins y Monro

Evaluación de un modelo

- Mejor ajuste elegido de una cierta clase de modelos ...

Evaluación de un modelo

- Mejor ajuste elegido de una cierta clase de modelos . . . pero puede ser un mal ajuste a los datos si la clase de modelos no es suficientemente rica

Evaluación de un modelo

- Mejor ajuste elegido de una cierta clase de modelos . . . pero puede ser un mal ajuste a los datos si la clase de modelos no es suficientemente rica
- ¿Cómo evaluamos un modelo (ERGM)?

  Paso 1: simulamos muchas realizaciones de grafos aleatorios bajo el modelo ajustado

  Paso 2: comparamos características de alto nivel con el G^{obs}

  Ej: distribución de grados, centralidad, diámetro
- Si hay diferencias significativas con G^{obs}, concluimos

  $\Rightarrow$ El modelo es incapaz de expresar los datos

  $\Rightarrow$ Lamentablemente no se han desarrollado indicadores de bondad de ajuste (goodness-of-fit)

Evaluación de un modelo

- Mejor ajuste elegido de una cierta clase de modelos ...pero puede ser un mal ajuste a los datos si la clase de modelos no es suficientemente rica
- ¿Cómo evaluamos un modelo (ERGM)?

  Paso 1: simulamos muchas realizaciones de grafos aleatorios bajo el modelo ajustado

  Paso 2: comparamos características de alto nivel con el G^{obs}

  Ej: distribución de grados, centralidad, diámetro
- Si hay diferencias significativas con G^{obs}, concluimos

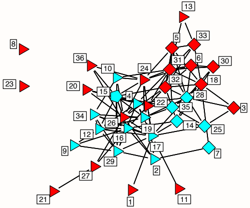
  $\Rightarrow$ El modelo es incapaz de expresar los datos

  $\Rightarrow$ Lamentablemente no se han desarrollado indicadores de bondad de ajuste (goodness-of-fit)
- **Importante:** especificar el modelo para los ERGMs es difícil (y muy importante)

  $\Rightarrow$ Si bien ERGM tiene limitaciones matemáticas importantes, ilustra varios aspectos importantes de modelado estadístico

Ejemplo: red de colaboración de abogados

- Grafo G^{obs} de relaciones laborales entre abogados [Lazega'01]
  - Los nodos son $N_v = 36$ "partners", y las aristas indican si trabajaron juntos



- Además, se incluye varios atributos de cada nodo:
  - Antigüedad (la etiqueta indica el orden en el ranking)
  - Ubicación de la oficina (triángulo, cuadrado o pentágono)
  - Tipo de práctica, i.e., litigación (rojo) o corporativa (cían)
  - Género (sólo tres partners son femeninas, etiquetadas con 27, 29 y 34)
- Objetivo: estudiar la cooperación entre actores sociales en una organización

Modelo para la red de colaboración de abogados

- Evaluación del efecto de la red: usemos $S_1(\mathbf{y}) = N_e$ y el alternating k-triangles

$$\mathrm{AKT}_\lambda(\mathbf{y}) = 3T_1(\mathbf{y}) + \sum_{k=2}^{N_v-2} (-1)^{k+1} \frac{T_k(\mathbf{y})}{\lambda^{k-1}}$$

$\Rightarrow$ similar al TKS, pero $T_k(\mathbf{y})$ cuenta cuántas veces k triángulos comparten una base

Modelo para la red de colaboración de abogados

- Evaluación del efecto de la red: usemos $S_1(\mathbf{y}) = N_e$ y el alternating k-triangles

$$\mathrm{AKT}_\lambda(\mathbf{y}) = 3T_1(\mathbf{y}) + \sum_{k=2}^{N_v-2} (-1)^{k+1} \frac{T_k(\mathbf{y})}{\lambda^{k-1}}$$

$\Rightarrow$ similar al TKS, pero $T_k(\mathbf{y})$ cuenta cuántas veces k triángulos comparten una base

- Testear los siguientes efectos exógenos:

$$h^{(1)}(\mathbf{x}_i, \mathbf{x}_j) = \text{antigüedad}_i + \text{antigüedad}_j, \quad h^{(2)}(\mathbf{x}_i, \mathbf{x}_j) = \text{práctica}_i + \text{práctica}_j$$
$$h^{(3)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{práctica}_i = \text{práctica}_j\right\}, \quad h^{(4)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{género}_i = \text{género}_j\right\}$$
$$h^{(5)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{oficina}_i = \text{oficina}_j\right\}, \quad \mathbf{h}(\mathbf{x}_i, \mathbf{x}_j) := [h^{(1)}(\mathbf{x}_i, \mathbf{x}_j), \ldots, h^{(5)}(\mathbf{x}_i, \mathbf{x}_j)]^T$$

Modelo para la red de colaboración de abogados

- Evaluación del efecto de la red: usemos $S_1(\mathbf{y}) = N_e$ y el alternating k-triangles

$$\mathrm{AKT}_\lambda(\mathbf{y}) = 3T_1(\mathbf{y}) + \sum_{k=2}^{N_v-2} (-1)^{k+1} \frac{T_k(\mathbf{y})}{\lambda^{k-1}}$$

$\Rightarrow$ similar al TKS, pero $T_k(\mathbf{y})$ cuenta cuántas veces k triángulos comparten una base

- Testear los siguientes efectos exógenos:

$$h^{(1)}(\mathbf{x}_i, \mathbf{x}_j) = \text{antigüedad}_i + \text{antigüedad}_j, \quad h^{(2)}(\mathbf{x}_i, \mathbf{x}_j) = \text{práctica}_i + \text{práctica}_j$$
$$h^{(3)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{práctica}_i = \text{práctica}_j\right\}, \quad h^{(4)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{género}_i = \text{género}_j\right\}$$
$$h^{(5)}(\mathbf{x}_i, \mathbf{x}_j) = \mathbb{I}\left\{\text{oficina}_i = \text{oficina}_j\right\}, \quad \mathbf{h}(\mathbf{x}_i, \mathbf{x}_j) := [h^{(1)}(\mathbf{x}_i, \mathbf{x}_j), \ldots, h^{(5)}(\mathbf{x}_i, \mathbf{x}_j)]^T$$

- ERGM resultante

$$\mathbb{P}_{\boldsymbol{\theta},\boldsymbol{\beta}}(\mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}) = \frac{1}{\kappa(\boldsymbol{\theta}, \boldsymbol{\beta})} \exp\left\{\theta_1 S_1(\mathbf{y}) + \theta_2 \mathrm{AKT}_\lambda(\mathbf{y}) + \boldsymbol{\beta}^T \mathbf{g}(\mathbf{y}, \mathbf{x})\right\}$$
$$\mathbf{g}(\mathbf{y}, \mathbf{x}) = \sum_{i,j} y_{ij} \mathbf{h}(\mathbf{x}_i, \mathbf{x}_j)$$

Resultado del ajuste

- Resultados del ajuste usando el método de MCMC MLE

Parameter	Estimate	'Standard Error'
Density (θ_1)	-6.2073	0.5697
Alternating k-triangles (θ_2)	0.5909	0.0882
Seniority Main Effect (β_1)	0.0245	0.0064
Practice Main Effect (β_2)	0.3945	0.1103
Same Practice (β_3)	0.7721	0.1973
Same Gender (β_4)	0.7302	0.2495
Same Office (β_5)	1.1614	0.1952

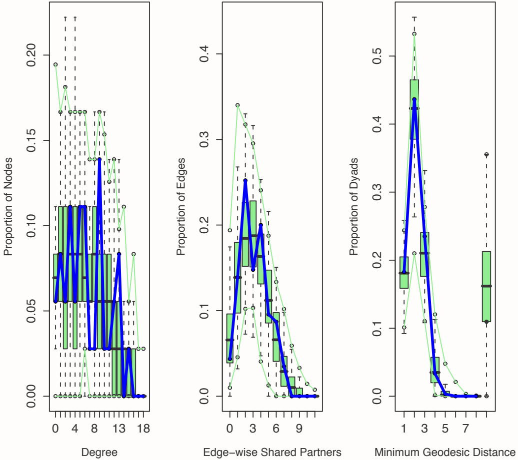
$\Rightarrow$ Errores estándar obtenidos de manera heurística

- Se pueden identificar varios factores que incrementan la probabilidad de cooperación
  Ej: misma práctica, género y lugar de la oficina duplican las chances
- Gran evidencia de transitividad dado que $\hat{\theta}_2 \gg \mathrm{se}(\hat{\theta}_2)$

  $\Rightarrow$ Homofilia ya está tomada en cuenta, por lo que hay algo que no tuvimos en cuenta que explica esto

“Goodness-of-fit”

- Evaluemos el modelo respecto a G^{obs}
  - Tomemos muestras del ERGM ajustado
- Comparamos la distribución de
  - Grados
  - Número de vecinos en común
  - Distancia
- Las gráficas muestran un buen ajuste



FACULTAD DE INGENIERÍA UDELAR

Anexo 1: Propiedad clave

- Recordemos que $\exp \psi(\boldsymbol{\theta}) = \sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}$ para escribir

$$\exp\{\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)\} = \frac{\sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}}{\exp \psi(\boldsymbol{\theta}_0)}$$

Anexo 1: Propiedad clave

- Recordemos que $\exp\psi(\boldsymbol{\theta}) = \sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}$ para escribir

$$\exp\left\{\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)\right\} = \frac{\sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}}{\exp\psi(\boldsymbol{\theta}_0)}$$

- Multiplicando y diviendo entre $\exp\left\{\boldsymbol{\theta}_0^\top \mathbf{g}(\mathbf{y})\right\} > 0$ resulta en

$$\begin{aligned}
\exp\left\{\psi(\boldsymbol{\theta}) - \psi(\boldsymbol{\theta}_0)\right\} &= \sum_{\mathbf{y}} \exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{y})\right\} \times \frac{\exp\left\{\boldsymbol{\theta}_0^\top \mathbf{g}(\mathbf{y})\right\}}{\exp\psi(\boldsymbol{\theta}_0)} \\
&= \sum_{\mathbf{y}} \exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{y})\right\} \mathrm{P}_{\theta_0}(\mathbf{Y} = \mathbf{y}) \\
&= \mathbb{E}_{\theta_0}\left[\exp\left\{(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^\top \mathbf{g}(\mathbf{Y})\right\}\right]
\end{aligned}$$

- Se usó que $\exp\left\{\boldsymbol{\theta}_0^\top \mathbf{g}(\mathbf{y}) - \psi(\boldsymbol{\theta}_0)\right\}$ es la densidad de las exponenciales $\mathrm{P}_{\theta_0}(\mathbf{Y} = \mathbf{y})$

Anexo 2: prueba de $\mathbb{E}\left[g(\mathbf{Y})\right] = \nabla\psi(\theta)$

- La densidad de $\mathbf{Y}$ es $\mathrm{P}_\theta(\mathbf{Y}=\mathbf{y}) = \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}$, por lo que

$$\begin{aligned}\mathbb{E}_\theta[g(\mathbf{Y})] &= \sum_{\mathbf{y}} g(\mathbf{y}) \mathrm{P}_\theta(\mathbf{Y}=\mathbf{y}) \\ &= \sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}\end{aligned}$$

Anexo 2: prueba de $\mathbb{E}\left[g(\mathbf{Y})\right] = \nabla\psi(\theta)$

- La densidad de $\mathbf{Y}$ es $\mathrm{P}_\theta(\mathbf{Y}=\mathbf{y}) = \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}$, por lo que

$$
\begin{aligned}
\mathbb{E}_\theta[g(\mathbf{Y})] &= \sum_{\mathbf{y}} g(\mathbf{y}) \mathrm{P}_\theta(\mathbf{Y}=\mathbf{y}) \\
&= \sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}
\end{aligned}
$$

- Recordemos que $\psi(\boldsymbol{\theta}) = \log \sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}$ y usando la regla de la cadena

$$
\begin{aligned}
\nabla\psi(\boldsymbol{\theta}) &= \frac{\sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}}{\sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}} = \frac{\sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}}{\exp\psi(\theta)} \\
&= \sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}
\end{aligned}
$$

Anexo 2: prueba de $\mathbb{E}\left[g(\mathbf{Y})\right] = \nabla\psi(\theta)$

- La densidad de $\mathbf{Y}$ es $\mathrm{P}_\theta(\mathbf{Y}=\mathbf{y}) = \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}$, por lo que

$$\mathbb{E}_\theta[g(\mathbf{Y})] = \sum_{\mathbf{y}} g(\mathbf{y})\mathrm{P}_\theta(\mathbf{Y}=\mathbf{y})$$
$$= \sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}$$

- Recordemos que $\psi(\boldsymbol{\theta}) = \log \sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}$ y usando la regla de la cadena

$$\nabla\psi(\boldsymbol{\theta}) = \frac{\sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}}{\sum_{\mathbf{y}} \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}} = \frac{\sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y})\right\}}{\exp \psi(\theta)}$$
$$= \sum_{\mathbf{y}} g(\mathbf{y}) \exp\left\{\boldsymbol{\theta}^\top \mathbf{g}(\mathbf{y}) - \psi(\theta)\right\}$$

- La sumas en rojo y azul son iguales $\Rightarrow \mathbb{E}_\theta[g(\mathbf{Y})] = \nabla\psi(\boldsymbol{\theta})$