Aprendizaje Automático para Datos en Grafos

Modelos de Grafos Aleatorios - Parte II

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

1. Modelos de Variables Latentes

2. Random dot product graphs

Modelos de Variables Latentes

- Variables Latentes ampliamente usado para modelar datos con observaciones parciales

  Ex: Hidden Markov Models, análisis factorial

- Esta idea se aplica en análisis estadístico de redes en (básicamente) dos variantes:
  - Modelos de clases latentes: la pertenencia a cierta clase (no observada) es la que marca la tendencia a conectarse
  - Modelos de vectores latentes: las conexiones son más probables en función de que tan "cerca" estén los nodos en cierto espacio latente

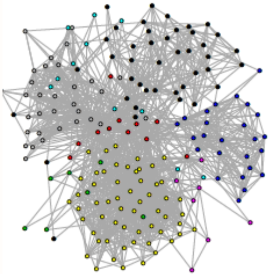
- Si bien en redes existen muchas variantes, nos enfocaremos en

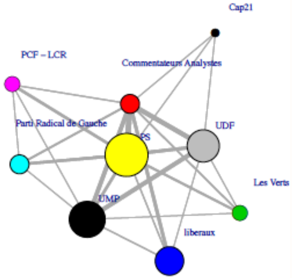
  $\Rightarrow$ Stochastic block models (SBMs)

  $\Rightarrow$ Una variante general y no-paramétrica denominada grafones

Ejemplo 1: Blogs políticos franceses

- Red de blogs políticos franceses de Octubre 2006 [Kolaczyk'17]
  - ⇒ Es un grafo con $N_v = 192$ blogs conectados por $N_e = 1431$ aristas
  - ⇒ Cada color indica la afiliación a cierto partido político

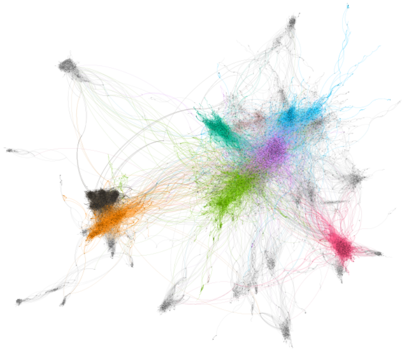




- Visualmente está claro que hay una mezcla de sub-grafos densamente conectados
  - ⇒ Diferentes tasas de conexión entre blogs dependiendo del partido
  - ⇒ Erdős-Rényi con p fijo no puede capturar esta estructura

Ejemplo 2: Actores que comparten películas

- Red de colaboración de actores según IMDb entre 2017 y 2021 (ver EVA)
  - ⇒ Es un grafo con $N_v = 21617$ actores conectados por $N_e = 73702$ aristas
  - ⇒ Cada color indica comunidad estimada con la modularidad



- Diferentes tasas de conexión entre actores dependiendo de dónde trabajan: Hollywood, independientes, Bollywood, Nollywood, etc.
- Un configuration model tampoco puede capturar esta estructura

Stochastic block models

- Los stochastic block models son explícitos en esta noción
  - ⇒ Grupos, módulos o comunidades $\mathcal{C}_1, \ldots, \mathcal{C}_Q$
  - ⇒ Tasas de conexión π_{qr} de vértices inter/intra grupos

Stochastic block models

- Los stochastic block models son explícitos en esta noción
  - $\Rightarrow$ Grupos, módulos o comunidades $\mathcal{C}_1, \ldots, \mathcal{C}_Q$
  - $\Rightarrow$ Tasas de conexión π_{qr} de vértices inter/intra grupos

Modelo generativo para un grafo no-dirigido $G(\mathcal{V}, \mathcal{E})$

- Cada vértice $i \in \mathcal{V}$ pertenece de manera independiente a $\mathcal{C}_q$ con probabilidad α_q

$$\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_Q]^\top, \quad \mathbf{1}^\top \boldsymbol{\alpha} = 1$$

- Para cada par de vértices $i, j \in \mathcal{V}$, con $i \in \mathcal{C}_q$ y $j \in \mathcal{C}_r$ $\Rightarrow (i, j) \in \mathcal{E}$ con proba π_{qr}

P. W. Holland et al., "Stochastic block-models: First steps," *Social Networks*, vol. 5, pp. 109-137, 1983

Especificación del modelo y flexibilidad

- En otras palabras, con $Z_{iq} = \mathbb{I}\{i \in \mathcal{C}_q\}$ y $\mathbf{Z}_i = [Z_{i1}, \ldots, Z_{iQ}]^\top$

$$\mathbf{Z}_i \overset{\text{i.i.d.}}{\sim} \text{Multinomial}(1, \boldsymbol{\alpha}),$$
$$A_{ij} \,\big|\, \mathbf{Z}_i = \mathbf{z}_i, \mathbf{Z}_j = \mathbf{z}_j \sim \text{Bernoulli}(\pi_{\mathbf{z}_i,\mathbf{z}_j})$$

  para $1 \leq i, j \leq N_v$, donde $A_{ij} = A_{ji}$ y $A_{ii} \equiv 0$

- Parámetros: Q proporciones de grupos (α_q) y $Q(Q+1)/2$ probas de conexión (π_{qr})

Especificación del modelo y flexibilidad

- En otras palabras, con $Z_{iq} = \mathbb{I}\{i \in \mathcal{C}_q\}$ y $\mathbf{Z}_i = [Z_{i1}, \ldots, Z_{iQ}]^\top$

$$\mathbf{Z}_i \overset{\text{i.i.d.}}{\sim} \text{Multinomial}(1, \boldsymbol{\alpha}),$$
$$A_{ij} \,\big|\, \mathbf{Z}_i = \mathbf{z}_i, \mathbf{Z}_j = \mathbf{z}_j \sim \text{Bernoulli}(\pi_{\mathbf{z}_i,\mathbf{z}_j})$$

  para $1 \leq i, j \leq N_v$, donde $A_{ij} = A_{ji}$ y $A_{ii} \equiv 0$

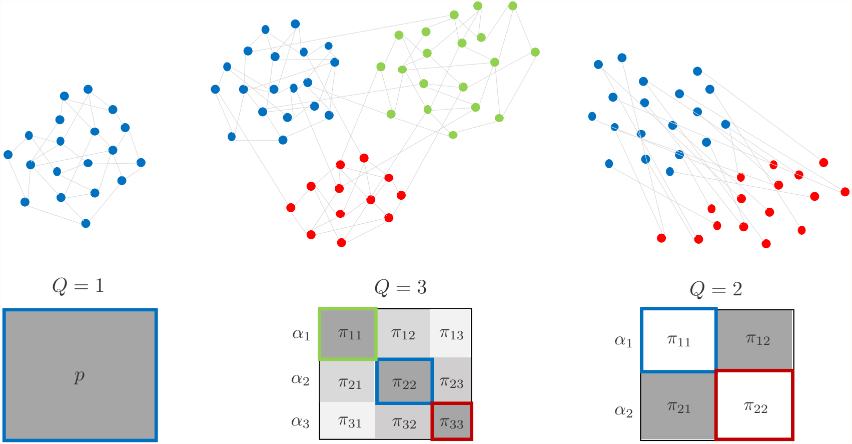
- Parámetros: Q proporciones de grupos (α_q) y $Q(Q+1)/2$ probas de conexión (π_{qr})
- Es una mezcla de grafos aleatorios clásicos

$$\mathrm{P}\left(A_{ij} = 1\right) = \mathrm{P}\left(\bigcup_{1 \leq q,r \leq Q} (i \in \mathcal{C}_q) \cap (j \in \mathcal{C}_r) \cap ((i,j) \in \mathcal{E})\right) = \sum_{1 \leq q,r \leq Q} \alpha_q \alpha_r \pi_{qr}$$

- Es más flexible para capturar la estructura de grafos observados
  - $\Rightarrow$ Hay temas de identificabilidad [Allman et al'11]
  - $\Rightarrow$ Transición de fase en la aparición de una componente gigante [Söderberg'02'03abc]

FACULTAD DE INGENIERÍA UDELAR

Especificación del modelo y flexibilidad (cont.)



- Mezcla de modelos Erdős-Rényi puede ser sorprendentemente flexible

Grafones y $f-$random graphs

- Una variante no-parámetrica de SBM se puede definir así:

$$U_1, \ldots, U_{N_v} \overset{\text{i.i.d.}}{\sim} \text{Uniform}[0,1],$$
$$A_{ij} \,\big|\, U_i = u_i, U_j = u_j \sim \text{Bernoulli}(f(u_i, u_j))$$

para $1 \leq i, j \leq N_v$, donde $A_{ij} = A_{ji}$ y $A_{ii} \equiv 0$

- Grafón: función simétrica y medible $f : [0,1]^2 \mapsto [0,1]$

  $\Rightarrow$ El grafo resultante G se denomina $f-$random graph

Grafones y $f-$random graphs

- Una variante no-parámetrica de SBM se puede definir así:

$$U_1, \ldots, U_{N_v} \overset{\text{i.i.d.}}{\sim} \text{Uniform}[0,1],$$
$$A_{ij} \,\big|\, U_i = u_i, U_j = u_j \sim \text{Bernoulli}(f(u_i, u_j))$$

  para $1 \leq i, j \leq N_v$, donde $A_{ij} = A_{ji}$ y $A_{ii} \equiv 0$

- Grafón: función simétrica y medible $f : [0,1]^2 \mapsto [0,1]$
  - $\Rightarrow$ El grafo resultante G se denomina $f-$random graph
- Variables aleatorias latentes U_i uniformes en $[0,1]$ dan la posición de cada nodo
  - $\Rightarrow$ El grafón $f(u_i, u_j)$ especifica la probabilidad de conexión entre i, j
- SBM: variables latentes $\mathbf{Z}_i$ dan la membresía de los vértices a uno de los Q grupos
  - $\Rightarrow$ La probabilidad π_{qr} define la proabilidad de conexión entre $i \in \mathcal{C}_q, j \in \mathcal{C}_r$

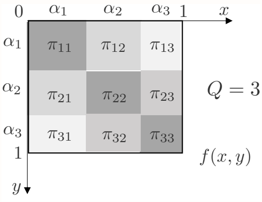
L. Lovász, "Large Networks and Graph Limits," *AMS Colloquium Publications*, vol. 60, 2012

Ejemplo: grafones SBM

- El modelo $f-$random incluye a SBM (paramétrico). ¿Cómo?

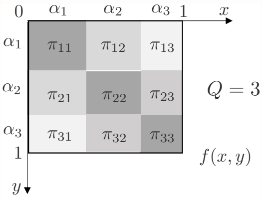
Ejemplo: grafones SBM

- El modelo $f-$random incluye a SBM (paramétrico). ¿Cómo?
  (i) Partimos $[0,1]$ en Q sub-intervalos de largo $\alpha_1, \ldots, \alpha_Q$
  (ii) Tomamos el producto cartesiano para partir $[0,1]^2$ en Q^2 bloques
  (iii) Definimos f para ser constante a tramos en los bloques: el bloque qr tiene "altura" π_{qr}



Ejemplo: grafones SBM

- El modelo $f-$random incluye a SBM (paramétrico). ¿Cómo?
  - (i) Partimos $[0,1]$ en Q sub-intervalos de largo $\alpha_1, \ldots, \alpha_Q$
  - (ii) Tomamos el producto cartesiano para partir $[0,1]^2$ en Q^2 bloques
  - (iii) Definimos f para ser constante a tramos en los bloques: el bloque qr tiene "altura" π_{qr}



- Se puede aproximar cualquier función medible por una constante a tramos

  $\Rightarrow$ Podemos aproximar cualquier $f-$random graph (en distribución) con un SBM

  $\Rightarrow$ Peeeero el número de bloques Q puede ser enorme...

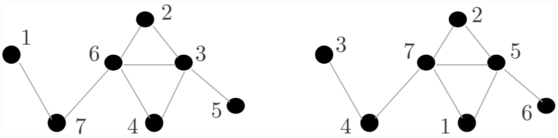
Ejemplo: Generación de grafo

- Consideremos un grafo $f-$random con $f(x,y) = \min(x,y)$ [Lovász'12]
  - $\Rightarrow$ La gráfica de la izquierda muestra el grafón en escala de grises f
- **Q:** ¿Qué pinta tienen los grafos generados? Ejemplo con $N_v = 40$

- Gráfica del centro muestra una realización de la matriz de adyacencia $\mathbf{A}$
  - $\Rightarrow$ Dado sólo $\mathbf{A}$, es imposible reconocer el origen del grafo
- Pero ordenando los vértices según la conectividad $d_{(1)}, \ldots, d_{(40)}$ aparece un patrón
  - $\Rightarrow$ Parece posible una estimación no-paramétrica del grafón

Intercambiabilidad de vértices

- **Def:** una matriz aleatoria $\mathbf{A} = [A_{ij}]_{i,j\in\mathcal{V}}$ es intercambiable en vértices si $\mathbf{A}_\sigma := [A_{\sigma(i)\sigma(j)}]_{i,j\in\mathcal{V}} \overset{D}{=} \mathbf{A}$ para cada permutación $\sigma : \mathcal{V} \mapsto \mathcal{V}$

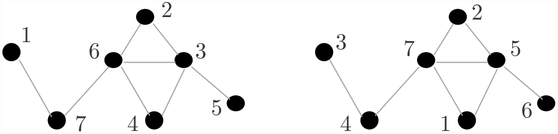


- Modelos intercambiables asignan la misma probabilidad a grafos isomorfos

  $\Rightarrow$ Estos modelos son más naturales para grafos no etiquetados

Intercambiabilidad de vértices

- **Def:** una matriz aleatoria $\mathbf{A} = [A_{ij}]_{i,j\in\mathcal{V}}$ es intercambiable en vértices si $\mathbf{A}_\sigma := [A_{\sigma(i)\sigma(j)}]_{i,j\in\mathcal{V}} \overset{D}{=} \mathbf{A}$ para cada permutación $\sigma : \mathcal{V} \mapsto \mathcal{V}$



- Modelos intercambiables asignan la misma probabilidad a grafos isomorfos
  - $\Rightarrow$ Estos modelos son más naturales para grafos no etiquetados
- Como en SBMs, uno puede demostrar que los grafos f-random son intercambiables
- Notablemente, cada modelo intercamiable es una mezcla de grafos f-random
  - $\Rightarrow$ El teorema de Aldous-Hoover extiende el resultado sobre secuencias de de Finetti

D. J. Aldous, "Representations for partially exchangeable arrays of random variables," *Journal of Mulivariate Analysis*, vol. 11, 1981

Todo grafo f-random es intercambiable

- La distribución de un grafo $f-$random con N_v vértices es

$$\mathrm{P}\left(\mathbf{A} = \mathbf{a}\right) =$$

Todo grafo f-random es intercambiable

- La distribución de un grafo f−random con N_v vértices es

$$\mathrm{P}\left(\mathbf{A}=\mathbf{a}\right)=\int_{[0,1]^{N_v}}\prod_{1\leq i\neq j\leq N_v} f(u_i,u_j)^{a_{ij}}(1-f(u_i,u_j))^{1-a_{ij}}du_1\ldots du_{N_v}$$

Todo grafo f-random es intercambiable

- La distribución de un grafo f−random con N_v vértices es

$$\mathrm{P}(\mathbf{A} = \mathbf{a}) = \int_{[0,1]^{N_v}} \prod_{1 \leq i \neq j \leq N_v} f(u_i, u_j)^{a_{ij}} (1 - f(u_i, u_j))^{1-a_{ij}} du_1 \ldots du_{N_v}$$

- Para permutaciones arbitrarias $\sigma : \mathcal{V} \mapsto \mathcal{V}$ resulta

$$\begin{aligned}
\mathrm{P}(\mathbf{A}_\sigma = \mathbf{a}_\sigma) &= \int_{[0,1]^{N_v}} \prod_{1 \leq i \neq j \leq N_v} f(u_i, u_j)^{a_{\sigma(i)\sigma(j)}} (1 - f(u_i, u_j))^{1-a_{\sigma(i)\sigma(j)}} du_1 \ldots du_{N_v} \\
&= \int_{[0,1]^{N_v}} \prod_{1 \leq i \neq j \leq N_v} f(u_{\sigma^{-1}(i)}, u_{\sigma^{-1}(j)})^{a_{ij}} (1 - f(u_{\sigma^{-1}(i)}, u_{\sigma^{-1}(j)}))^{1-a_{ij}} \\
&\quad \times du_{\sigma^{-1}(1)} \ldots du_{\sigma^{-1}(N_v)} \\
&= \int_{[0,1]^{N_v}} \prod_{1 \leq i \neq j \leq N_v} f(u_i, u_j)^{a_{ij}} (1 - f(u_i, u_j))^{1-a_{ij}} du_1 \ldots du_{N_v} \\
&= \mathrm{P}(\mathbf{A} = \mathbf{a})
\end{aligned}$$

Problemas de identidad

- Dos f distintas pueden producir grafos $f-$random con la misma distribución $\Rightarrow$ No identificable

  Ej: grafones $f(x,y)$ y $f(1-x,1-y)$ resultan en el mismo modelo

$$U \overset{D}{=} 1-U \text{ para } U \sim \text{Uniform}[0,1]$$

  Ej: dos grafones $f(x,y)$ y $f(\phi(x),\phi(y))$ para un ϕ que preserve la medida, i.e.,

$$\phi : [0,1] \mapsto [0,1] \text{ para el cual } \phi(U) \sim \text{Uniform}[0,1]$$

Problemas de identidad

- Dos f distintas pueden producir grafos $f-$random con la misma distribución $\Rightarrow$ No identificable

  Ej: grafones $f(x, y)$ y $f(1-x, 1-y)$ resultan en el mismo modelo

$$U \stackrel{D}{=} 1 - U \text{ para } U \sim \text{Uniform}[0,1]$$

  Ej: dos grafones $f(x, y)$ y $f(\phi(x), \phi(y))$ para un ϕ que preserve la medida, i.e.,

$$\phi : [0,1] \mapsto [0,1] \text{ para el cual } \phi(U) \sim \text{Uniform}[0,1]$$

- Son necesarias ciertas restricciones en el grafón f para su modelado estadístico
- **Def:** f es estrictamente monótono si $\exists\, \phi$ tal que $\tilde{f}(x, y) = f(\phi(x), \phi(y))$ tiene una función de grados $\tilde{g}(x) = \int_{[0,1]} \tilde{f}(x, y) dy$ estrictamente creciente

  $\Rightarrow$ Esta restricción a $\tilde{f}$ resulta en un modelo identificable [Bickel-Chen'09]

FACULTAD DE INGENIERÍA UDELAR

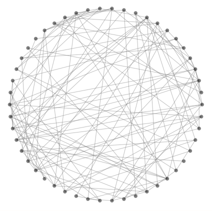
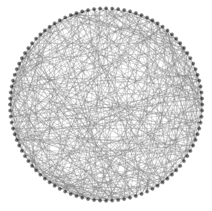
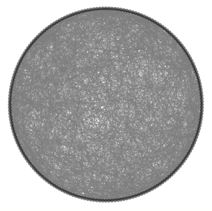
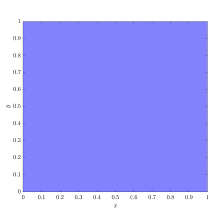
Límite de un grafo

- Una secuencia de grafos $G_n(\mathcal{V}_n, \mathcal{E}_n)$ con una cantidad de vértices creciente $N_v = n$

  Q: ¿Cuándo podemos decir que $\{G_n\}_{n=1}^{\infty}$ converge a un límite?
  Q: ¿Qué sentido tiene esta convergencia?
  Q: ¿Qué tipo de objeto es este límite?

- Spoiler: Si la secuencia $\{G_n\}_{n=1}^{\infty}$ converge, el límite es un grafón f

  Ej: La secuencia $\mathrm{ER}(n, p)$ a medida que $n \to \infty$

$n = 50$ $\rightarrow$ $n = 100$ $\rightarrow$ $n = 200$ $\rightarrow$ Grafón $f(x, y) \equiv p$

Densidad de homomorfismo

Def: Homomorfismos h son aquellos mapeos del motif $F(\mathcal{V}', \mathcal{E}')$ (i.e. un grafo pequeño) a un grafo $G(\mathcal{V}, \mathcal{E})$ que preservan la adyacencia

$$h : \mathcal{V}' \mapsto \mathcal{V} \text{ tal que } (i,j) \in \mathcal{E}' \text{ implica } (h(i), h(j)) \in \mathcal{E}$$

- Existen $\binom{|\mathcal{V}|}{|\mathcal{V}'|}$ mapeos posibles de F a G, pero sólo $\hom(F, G)$ son homomorfismos

  $\Rightarrow$ **Def:** Densidad de homomorfismos del motif F en el grafo G es

$$t(F, G) = \frac{\hom(F, G)}{\binom{|\mathcal{V}|}{|\mathcal{V}'|}}$$

Densidad de homomorfismo

Def: Homomorfismos h son aquellos mapeos del motif $F(\mathcal{V}', \mathcal{E}')$ (i.e. un grafo pequeño) a un grafo $G(\mathcal{V}, \mathcal{E})$ que preservan la adyacencia

$$h : \mathcal{V}' \mapsto \mathcal{V} \text{ tal que } (i, j) \in \mathcal{E}' \text{ implica } (h(i), h(j)) \in \mathcal{E}$$

- Existen $\binom{|\mathcal{V}|}{|\mathcal{V}'|}$ mapeos posibles de F a G, pero sólo $\hom(F, G)$ son homomorfismos
  $\Rightarrow$ **Def:** Densidad de homomorfismos del motif F en el grafo G es

$$t(F, G) = \frac{\hom(F, G)}{\binom{|\mathcal{V}|}{|\mathcal{V}'|}}$$

- Medida relativa del número de formas en que F puede mapearse a G
- Interpretación: si se sortea uniforme aleatorio un mapeo, cuál es la probabilidad de que sea un homomorfismo?

Densidad de homomorfismo

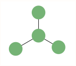
Def: Homomorfismos h son aquellos mapeos del motif $F(\mathcal{V}', \mathcal{E}')$ (i.e. un grafo pequeño) a un grafo $G(\mathcal{V}, \mathcal{E})$ que preservan la adyacencia

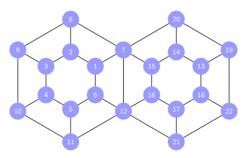
$$h : \mathcal{V}' \mapsto \mathcal{V} \text{ tal que } (i, j) \in \mathcal{E}' \text{ implica } (h(i), h(j)) \in \mathcal{E}$$

- Existen $\binom{|\mathcal{V}|}{|\mathcal{V}'|}$ mapeos posibles de F a G, pero sólo $\hom(F, G)$ son homomorfismos
  $\Rightarrow$ **Def:** Densidad de homomorfismos del motif F en el grafo G es

$$t(F, G) = \frac{\hom(F, G)}{\binom{|\mathcal{V}|}{|\mathcal{V}'|}}$$

- Medida relativa del número de formas en que F puede mapearse a G
- Interpretación: si se sortea uniforme aleatorio un mapeo, cuál es la probabilidad de que sea un homomorfismo?





Densidad de homomorfismo

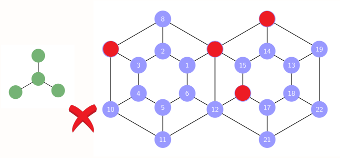
Def: Homomorfismos h son aquellos mapeos del motif $F(\mathcal{V}', \mathcal{E}')$ (i.e. un grafo pequeño) a un grafo $G(\mathcal{V}, \mathcal{E})$ que preservan la adyacencia

$$h : \mathcal{V}' \mapsto \mathcal{V} \text{ tal que } (i, j) \in \mathcal{E}' \text{ implica } (h(i), h(j)) \in \mathcal{E}$$

- Existen $\binom{|\mathcal{V}|}{|\mathcal{V}'|}$ mapeos posibles de F a G, pero sólo $\hom(F, G)$ son homomorfismos
  - $\Rightarrow$ **Def:** Densidad de homomorfismos del motif F en el grafo G es

$$t(F, G) = \frac{\hom(F, G)}{\binom{|\mathcal{V}|}{|\mathcal{V}'|}}$$

- Medida relativa del número de formas en que F puede mapearse a G
- Interpretación: si se sortea uniforme aleatorio un mapeo, cuál es la probabilidad de que sea un homomorfismo?



Densidad de homomorfismo

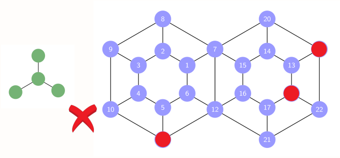
Def: Homomorfismos h son aquellos mapeos del motif $F(\mathcal{V}', \mathcal{E}')$ (i.e. un grafo pequeño) a un grafo $G(\mathcal{V}, \mathcal{E})$ que preservan la adyacencia

$$h : \mathcal{V}' \mapsto \mathcal{V} \text{ tal que } (i, j) \in \mathcal{E}' \text{ implica } (h(i), h(j)) \in \mathcal{E}$$

- Existen $\binom{|\mathcal{V}|}{|\mathcal{V}'|}$ mapeos posibles de F a G, pero sólo $\hom(F, G)$ son homomorfismos
  $\Rightarrow$ **Def:** Densidad de homomorfismos del motif F en el grafo G es

$$t(F, G) = \frac{\hom(F, G)}{\binom{|\mathcal{V}|}{|\mathcal{V}'|}}$$

- Medida relativa del número de formas en que F puede mapearse a G
- Interpretación: si se sortea uniforme aleatorio un mapeo, cuál es la probabilidad de que sea un homomorfismo?



Densidad de homomorfismo

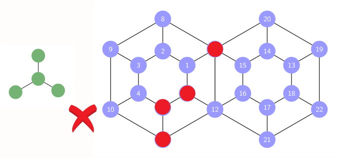
Def: Homomorfismos h son aquellos mapeos del motif $F(\mathcal{V}', \mathcal{E}')$ (i.e. un grafo pequeño) a un grafo $G(\mathcal{V}, \mathcal{E})$ que preservan la adyacencia

$$h : \mathcal{V}' \mapsto \mathcal{V} \text{ tal que } (i, j) \in \mathcal{E}' \text{ implica } (h(i), h(j)) \in \mathcal{E}$$

- Existen $\binom{|\mathcal{V}|}{|\mathcal{V}'|}$ mapeos posibles de F a G, pero sólo $\hom(F, G)$ son homomorfismos
  $\Rightarrow$ **Def:** Densidad de homomorfismos del motif F en el grafo G es

$$t(F, G) = \frac{\hom(F, G)}{\binom{|\mathcal{V}|}{|\mathcal{V}'|}}$$

- Medida relativa del número de formas en que F puede mapearse a G
- Interpretación: si se sortea uniforme aleatorio un mapeo, cuál es la probabilidad de que sea un homomorfismo?



Densidad de homomorfismo

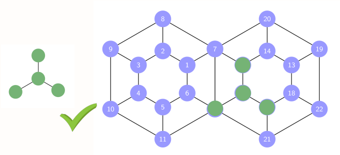
Def: Homomorfismos h son aquellos mapeos del motif $F(\mathcal{V}', \mathcal{E}')$ (i.e. un grafo pequeño) a un grafo $G(\mathcal{V}, \mathcal{E})$ que preservan la adyacencia

$$h : \mathcal{V}' \mapsto \mathcal{V} \text{ tal que } (i,j) \in \mathcal{E}' \text{ implica } (h(i), h(j)) \in \mathcal{E}$$

- Existen $\binom{|\mathcal{V}|}{|\mathcal{V}'|}$ mapeos posibles de F a G, pero sólo $\hom(F, G)$ son homomorfismos
  - $\Rightarrow$ **Def:** Densidad de homomorfismos del motif F en el grafo G es

$$t(F, G) = \frac{\hom(F, G)}{\binom{|\mathcal{V}|}{|\mathcal{V}'|}}$$

- Medida relativa del número de formas en que F puede mapearse a G
- Interpretación: si se sortea uniforme aleatorio un mapeo, cuál es la probabilidad de que sea un homomorfismo?



Convergencia de secuencias de grafos

- **Def:** Una secuencia de grafos $\{G_n\}_{n=1}^{\infty}$ converge cuando para todo motif F, la secuencia de densidad de homomorfismos $\{t(F, G_n)\}_{n=1}^{\infty}$ converge
- Algunas aspectos importantes de la definición
  - Si la secuencia se vuelve constante, entonces converge
  - Secuencias de grafos isomórficos converge
  - La convergencia es en cantidades normalizadas, no en número de aristas, triángulos, ...
  - Los resultados son para secuencias de grafos densos, i.e., $|\mathcal{E}_n| = \Omega(n^2)$

Convergencia de secuencias de grafos

- **Def:** Una secuencia de grafos $\{G_n\}_{n=1}^{\infty}$ converge cuando para todo motif F, la secuencia de densidad de homomorfismos $\{t(F, G_n)\}_{n=1}^{\infty}$ converge
- Algunas aspectos importantes de la definición
  - Si la secuencia se vuelve constante, entonces converge
  - Secuencias de grafos isomórficos converge
  - La convergencia es en cantidades normalizadas, no en número de aristas, triángulos, ...
  - Los resultados son para secuencias de grafos densos, i.e., $|\mathcal{E}_n| = \Omega(n^2)$
- Respondimos las dos primeras preguntas. Sólo nos queda la tercera y última

  $\Rightarrow$ El límite de una secuencia de grafos no es necesariamente un grafo

  $\Rightarrow$ **Q:** ¿Qué tipo de objeto es este límite?

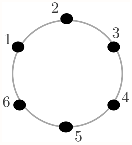
L. Lovász and B. Szegedy, "Limits of dense graph sequences," *Journal of Combinatorial Theory, Series B*, vol. 96, 2006

Grafón inducido

- Todo grafo admite una representación denominada grafón inducido
  - Consideremos un grafo $G(\mathcal{V}, \mathcal{E})$ con matriz de adyacencia $\mathbf{A}$
  - Hagamos una partición uniforme de $[0, 1]$ en N_v sub-intervalos $\Rightarrow I_i = [\frac{i-1}{N_v}, \frac{i}{N_v})$

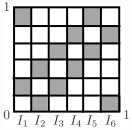
Def: El grafón inducido f_G de G es

$$f_G(x, y) = \sum_{1 \leq i,j \leq N_v} A_{ij} \mathbb{I}\{x \in I_i\} \mathbb{I}\{y \in I_j\}$$



Grafo cíclico G con $N_v = 6$ nodos

$\rightarrow$



Grafón f_G inducido por el grafo G

El límite es un grafón

- **Propiedad:** La densidad de homomorfismo del motif F en el grafo G está dado por

$$t(F,G) = \int_{[0,1]^{|\mathcal{V}'|}} \prod_{(i,j)\in\mathcal{E}'} f_G(u_i,u_j)du_1 \dots du_{|\mathcal{V}'|}$$

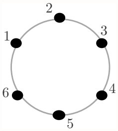
El límite es un grafón

- **Propiedad:** La densidad de homomorfismo del motif F en el grafo G está dado por

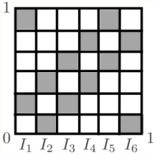
$$t(F,G) = \int_{[0,1]^{|\mathcal{V}'|}} \prod_{(i,j)\in\mathcal{E}'} f_G(u_i,u_j)\,du_1\ldots du_{|\mathcal{V}'|}$$

Ej: Sea F el motif más simple: una arista

- $t(F,G)$ responde la pregunta "Cuántas aristas tiene G (dividido entre las $N_v \times N_v$ posibilidades)?"



$\rightarrow$



- Podemos contar las aristas o simplemente calcular $\int_{[0,1]^2} f_G(u_1,u_2)du_1du_2$ (el área gris en el grafón inducido)

El límite es un grafón

- Esto sigue siendo cierto en el límite. Si la secuencia $\{G_n\}_{n=1}^{\infty}$ converge, entonces

$$\lim_{n\to\infty} t(F, G_n) = \int_{[0,1]^{|\mathcal{V}'|}} \prod_{(i,j)\in\mathcal{E}'} f(u_i, u_j) du_1 \dots du_{|\mathcal{V}'|}$$

para alguna función $f : [0,1]^2 \mapsto [0,1]$ simétrica y medible

FACULTAD DE INGENIERÍA UDELAR

El límite es un grafón

- Esto sigue siendo cierto en el límite. Si la secuencia $\{G_n\}_{n=1}^{\infty}$ converge, entonces

$$\lim_{n\to\infty} t(F, G_n) = \int_{[0,1]^{|\mathcal{V}'|}} \prod_{(i,j)\in\mathcal{E}'} f(u_i, u_j) du_1 \dots du_{|\mathcal{V}'|}$$

  para alguna función $f : [0,1]^2 \mapsto [0,1]$ simétrica y medible
- La intuición es la misma de antes: probabilidad de mapear el motif

Ej Sea F el motif tipo estrella de antes

$$t(F, G) = \int_{[0,1]^4} f(u_1, u_2) f(u_1, u_3) f(u_1, u_4) du_1 du_2 du_3 du_4$$

El límite es un grafón

- Esto sigue siendo cierto en el límite. Si la secuencia $\{G_n\}_{n=1}^{\infty}$ converge, entonces

$$\lim_{n\to\infty} t(F, G_n) = \int_{[0,1]^{|\mathcal{V}'|}} \prod_{(i,j)\in\mathcal{E}'} f(u_i, u_j) du_1 \dots du_{|\mathcal{V}'|}$$

  para alguna función $f : [0,1]^2 \mapsto [0,1]$ simétrica y medible
- La intuición es la misma de antes: probabilidad de mapear el motif

Ej Sea F el motif tipo estrella de antes

$$t(F, G) = \int_{[0,1]^4} \underbrace{f(u_1, u_2) f(u_1, u_3) f(u_1, u_4)}_{\substack{\text{probabilidad de la estrella} \\ \textit{dados } (u_1, u_2, u_3, u_4)}} du_1 du_2 du_3 du_4$$

El límite es un grafón

- Esto sigue siendo cierto en el límite. Si la secuencia $\{G_n\}_{n=1}^{\infty}$ converge, entonces
$$\lim_{n\to\infty} t(F,G_n) = \int_{[0,1]^{|\mathcal{V}'|}} \prod_{(i,j)\in\mathcal{E}'} f(u_i,u_j)du_1\dots du_{|\mathcal{V}'|}$$
para alguna función $f:[0,1]^2 \mapsto [0,1]$ simétrica y medible
- La intuición es la misma de antes: probabilidad de mapear el motif

Ej Sea F el motif tipo estrella de antes

$$t(F,G) = \underbrace{\int_{[0,1]^4} \underbrace{f(u_1,u_2)f(u_1,u_3)f(u_1,u_4)}_{\substack{\text{probabilidad de la estrella} \\ \textit{dados } (u_1,u_2,u_3,u_4)}} du_1du_2du_3du_4}_{\text{promediada entre los } (u_1,u_2,u_3,u_4) \text{ sorteados uniformes}}$$

El límite es un grafón

- Esto sigue siendo cierto en el límite. Si la secuencia $\{G_n\}_{n=1}^{\infty}$ converge, entonces

$$\lim_{n\to\infty} t(F, G_n) = \int_{[0,1]^{|\mathcal{V}'|}} \prod_{(i,j)\in\mathcal{E}'} f(u_i, u_j) du_1 \dots du_{|\mathcal{V}'|}$$

  para alguna función $f : [0,1]^2 \mapsto [0,1]$ simétrica y medible
- La intuición es la misma de antes: probabilidad de mapear el motif

Ej Sea F el motif tipo estrella de antes

$$\underbrace{t(F,G) = \int_{[0,1]^4} \underbrace{f(u_1,u_2)f(u_1,u_3)f(u_1,u_4)}_{\substack{\text{probabilidad de la estrella} \\ \textit{dados } (u_1,u_2,u_3,u_4)}} du_1 du_2 du_3 du_4}_{\text{promediada entre los } (u_1,u_2,u_3,u_4) \text{ sorteados uniformes}}$$

- Identificamos el objeto límite – el grafón – con f

¿Y esto para qué sirve?

Impacto matemático

- Trae herramientas de análisis a lo que a priori es un contexto puramente combinatorio

Impacto en inferencia estadística

- Realizaciones grandes se vuelven representativas del proceso generativo

  $\Rightarrow$ Inferir el mecanismo de generación de los datos examinando una única realización

Impacto en machine learning

- Estudiar filtros de grafos y GNNs en el límite de grandes grafos

  $\Rightarrow$ Transferabilidad e.g., entrenando en un grafo más pequeño que donde se aplica

L. Ruiz et al, “Graphon neural networks and the transferability of graph neural networks,” *NeurIPS*, 2020

Plausibilidad

- 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

Plausibilidad

- 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

Q: ¿Qué tan apropiados son modelos de variables latentes? ¿Son plausibles?

Q: ¿Podemos aproximar un grafo observado G^{obs} usando un SBM?

$\Rightarrow$ Una variante del lema de regularidad de Szemerédi resulta útil aquí

C. Borgs et al, "Graph limits and parameter testing," *Syposium on Theory of Computing*, 2006

Cut distance

- Discutir nociones de aproximación requiere una distancia entre grafos

Def: La cut distance entre los grafos $G(\mathcal{V},\mathcal{E})$ y $G'(\mathcal{V}',\mathcal{E}')$ (con $|\mathcal{V}| = |\mathcal{V}'| = N_v$) es

$$d_\square(G,G') = \frac{1}{N_v^2} \max_{\mathcal{S},\mathcal{T}\in\{1,\ldots,N_v\}} \left| \sum_{i\in\mathcal{S}} \sum_{j\in\mathcal{T}} (A_{ij} - A'_{ij}) \right|$$

$\Rightarrow$ Se puede demostrar que $d_\square(\cdot,\cdot)$ efectivamente es una métrica

- La definición y estudio de propiedades de distancias en grafos es un tema súper actual

B. Bollobás and O. Riordan, "Sparse graphs: Metrics and random models," *Random Structures & Algorithms*, vol. 39, 2011

Un resultado sobre aproximaciones

- Sea $\mathcal{P} = \{\mathcal{V}_1, \ldots, \mathcal{V}_Q\}$ una partición de los vértices $\mathcal{V}$ de G en Q clases
- Definamos un grafo completo G_P con vértices $\mathcal{V}$ y pesos

$$w_{ij}(G_P) = \frac{1}{|\mathcal{V}_q||\mathcal{V}_r|} \sum_{u \in \mathcal{V}_q} \sum_{v \in \mathcal{V}_r} A_{uv}, \quad i \in \mathcal{V}_q, j \in \mathcal{V}_r$$

$\Rightarrow$ Básicamente la aproximación del grafo G por un SBM de Q clases

$\Rightarrow$ La probabilidad de que una arista una i, j sería $w_{ij}(G_P)$

Un resultado sobre aproximaciones

- Sea $\mathcal{P} = \{\mathcal{V}_1, \ldots, \mathcal{V}_Q\}$ una partición de los vértices $\mathcal{V}$ de G en Q clases
- Definamos un grafo completo G_P con vértices $\mathcal{V}$ y pesos

$$w_{ij}(G_P) = \frac{1}{|\mathcal{V}_q||\mathcal{V}_r|} \sum_{u \in \mathcal{V}_q} \sum_{v \in \mathcal{V}_r} A_{uv}, \quad i \in \mathcal{V}_q, j \in \mathcal{V}_r$$

$\Rightarrow$ Básicamente la aproximación del grafo G por un SBM de Q clases

$\Rightarrow$ La probabilidad de que una arista una i, j sería $w_{ij}(G_P)$

Teorema: Para todo $\epsilon > 0$ y todo grafo $G(\mathcal{V}, \mathcal{E})$, existe una partición $\mathcal{P}$ de $\mathcal{V}$ en $Q \leq 2^{\frac{2}{\epsilon^2}}$ clases tal que $d_{\square}(G, G_P) \leq \epsilon$.

- Justificación de que un SBM puede aproximar cualquier grafo

$\Rightarrow$ La cota superior en Q puede ser gigantesca

¿Y los grafos f-random?

- Los grafos $f-$random son apropiados sólo para redes densas

> **Teorema:** Si un grafo G es la restricción a los vértices $\{1,\dots,N_v\}$ de un grafo aleatorio infinito intercambiable, entonces es denso o está vacío.

Intuitivamente: La proporción de aristas en $G(\mathcal{V},\mathcal{E})$ es

$$\varphi = \int_{[0,1]^2} f(u_1,u_2)du_1du_2$$

$\Rightarrow$ Si $\varphi = 0$ entonces $f = 0$ c.s. y G está vacío. Esparso, pero inútil

$\Rightarrow$ Si $\varphi > 0$ entonces (en media) $|\mathcal{E}| = \varphi \times \binom{N_v}{2} = \Omega(N_v^2)$

- El principal problema es la intercambiabilidad de vértices
  - Apropiado para grafos sin etiquetas...
  - Si tengo etiquetas, se podrían incorporar al modelo [Sweet'15]

Estimación de los parámetros de un SBM

- Los parámetros de un SBM son $\{\alpha_q\}_{q=1}^{Q}$ y $\{\pi_{qr}\}_{1\leq q,r\leq Q}$
- Conceptualmente puede simplificar pensar en dos conjuntos de 'observaciones'
  - $\Rightarrow$ Clases: $\mathbf{Z} = \{\{Z_{iq}\}_{q=1}^{Q}\}_{i\in\mathcal{V}}$, donde $Z_{iq} = \mathbb{I}\{i \in \mathcal{C}_q\}$
  - $\Rightarrow$ Enlaces: $\mathbf{A} = [A_{ij}]$, donde $A_{ij} = \mathbb{I}\{(i,j) \in \mathcal{E}\}$
- **Pero sólo observamos $\mathbf{A}$** ($\mathbf{Z}$ típicamente latente). Asumamos Q conocida
  - $\Rightarrow$ Interesa la estimación de los parámetros pero especialmente agrupar los vértices

Detección de comunidades basada en modelo

Supongamos que G fue generado por un SBM con Q clases. Predecir las pertenencia a clases $\mathbf{Z} = \{\{Z_{iq}\}_{q=1}^{Q}\}_{i\in\mathcal{V}}$, dada la observación $\mathbf{A} = \mathbf{a}$.

Estimación de Máxima Verosimilitud

- Si observáramos $\mathbf{A} = \mathbf{a}$ y $\mathbf{Z} = \mathbf{z}$, la log-likelihood sería (con $\boldsymbol{\theta} = \{\{\alpha_q\}, \{\pi_{qr}\}\}$)

$$\ell_{\boldsymbol{\theta}}(\mathbf{a}, \mathbf{z}) = \ell_{\boldsymbol{\theta}}(\mathbf{z}) + \ell_{\boldsymbol{\theta}}(\mathbf{a}|\mathbf{z})$$

$$\text{con} \quad \ell_{\boldsymbol{\theta}}(\mathbf{z}) = \log P\left(\mathbf{Z} = \mathbf{z}\right) = \log \left(\prod_i \left(\prod_q \alpha_q^{z_{iq}} \right) \right) = \sum_i \sum_q z_{iq} \log \alpha_q$$

$$\ell_{\boldsymbol{\theta}}(\mathbf{a}|\mathbf{z}) = \frac{1}{2} \sum_{i \neq j} \sum_{q \neq r} z_{iq} z_{jr} \log \left(\pi_{qr}^{a_{ij}} \left(1 - \pi_{qr}\right)^{1 - a_{ij}} \right)$$

Estimación de Máxima Verosimilitud

- Pero no observamos $\mathbf{Z}$. Hay que trabajar con la verosimilitud de los datos observados

$$\ell_{\boldsymbol{\theta}}(\mathbf{a}) = \log\left(\sum_{\mathbf{z}} \exp\left\{\ell_{\boldsymbol{\theta}}(\mathbf{a}, \mathbf{z})\right\}\right)$$

  - ¿Cuántos términos puede haber en esa sumatoria? $\Rightarrow$ calcular $\ell_{\boldsymbol{\theta}}(\mathbf{a})$ es inviable

Estimación de Máxima Verosimilitud

- Pero no observamos $\mathbf{Z}$. Hay que trabajar con la verosimilitud de los datos observados

$$\ell_{\boldsymbol{\theta}}(\mathbf{a}) = \log\left(\sum_{\mathbf{z}} \exp\left\{\ell_{\boldsymbol{\theta}}(\mathbf{a}, \mathbf{z})\right\}\right)$$

  - ¿Cuántos términos puede haber en esa sumatoria? $\Rightarrow$ calcular $\ell_{\boldsymbol{\theta}}(\mathbf{a})$ es inviable

- Ver el modelo como una mezcla sugiere usar Expectation Maximization [Snijders'97]

  $\Rightarrow$ Alternar entre estimar $\mathbb{E}\left[Z_{iq} \,\middle|\, \mathbf{A} = \mathbf{a}\right]$ y $\boldsymbol{\theta}$

  $\Rightarrow$ No escala más allá de $Q = 2$, $\mathrm{P}\left(\mathbf{Z} \,\middle|\, \mathbf{A} = \mathbf{a}\right)$ es computacionalmente caro

Máxima verosimilitud variacional

- Método variacional para optimizar una cota inferior de $\ell_{\boldsymbol{\theta}}(\mathbf{a})$:

$$J(R_{\mathbf{a}};\boldsymbol{\theta}) = \ell_{\boldsymbol{\theta}}(\mathbf{a}) - \mathrm{KL}(R_{\mathbf{a}}(\mathbf{Z}), \mathrm{P}\left(\mathbf{Z} \,\middle|\, \mathbf{A} = \mathbf{a}\right))$$

  - KL es la divergencia de Kullback–Leibler
  - $R_{\mathbf{a}}(\mathbf{Z})$ es una aproximación "manipulable" de $\mathrm{P}\left(\mathbf{Z} \,\middle|\, \mathbf{A} = \mathbf{a}\right)$

- Aproximación de campo medio de la distribución condicional

$$R_{\mathbf{a}}(\mathbf{Z}) = \prod_{i=1}^{N_v} h(\mathbf{Z}_i; \boldsymbol{\tau}_i)$$

  - $h(\cdot;\boldsymbol{\tau}_i)$: distribución multinomial con parámetro $\boldsymbol{\tau}_i = [\tau_{i1}, \ldots, \tau_{iQ}]^\top$

J. J. Daudin et al, "A mixture model for random graphs," *Stat. Comput.*, vol. 18, 2008

Algoritmo de maximización alternada

Proposición: Dado $\boldsymbol{\theta}$, el parámetro variacional óptimo $\{\hat{\boldsymbol{\tau}}_i\} = \operatorname{argmax}_{\{\boldsymbol{\tau}_i\}} J(R_{\mathbf{a}}; \{\boldsymbol{\tau}_i\}, \boldsymbol{\theta})$ satisface la siguiente ecuación de punto fijo

$$\hat{\tau}_{iq} \propto \alpha_q \prod_{j \neq i} \prod_r b(A_{ij}; \pi_{qr})^{\hat{\tau}_{jr}}$$

Dados $\{\boldsymbol{\tau}_i\}$, los valores de $\boldsymbol{\theta}$ que maximizan $J(R_{\mathbf{a}}; \{\boldsymbol{\tau}_i\}, \boldsymbol{\theta})$ son

$$\hat{\alpha}_q = \frac{1}{N_v} \sum_i \hat{\tau}_{iq}, \quad \hat{\pi}_{qr} = \sum_{i \neq j} \hat{\tau}_{iq} \hat{\tau}_{jr} A_{ij} \Big/ \sum_{i \neq j} \hat{\tau}_{iq} \hat{\tau}_{jr}$$

Algoritmo de maximización alternada

Proposición: Dado $\boldsymbol{\theta}$, el parámetro variacional óptimo $\{\hat{\boldsymbol{\tau}}_i\} = \text{argmax}_{\{\boldsymbol{\tau}_i\}} J(R_{\mathbf{a}}; \{\boldsymbol{\tau}_i\}, \boldsymbol{\theta})$ satisface la siguiente ecuación de punto fijo

$$\hat{\tau}_{iq} \propto \alpha_q \prod_{j \neq i} \prod_r b(A_{ij}; \pi_{qr})^{\hat{\tau}_{jr}}$$

Dados $\{\boldsymbol{\tau}_i\}$, los valores de $\boldsymbol{\theta}$ que maximizan $J(R_{\mathbf{a}}; \{\boldsymbol{\tau}_i\}, \boldsymbol{\theta})$ son

$$\hat{\alpha}_q = \frac{1}{N_v} \sum_i \hat{\tau}_{iq}, \quad \hat{\pi}_{qr} = \sum_{i \neq j} \hat{\tau}_{iq} \hat{\tau}_{jr} A_{ij} \Big/ \sum_{i \neq j} \hat{\tau}_{iq} \hat{\tau}_{jr}$$

- El algoritmo alterna entre actualizaciones de $\boldsymbol{\theta}$ y $\{\boldsymbol{\tau}_i\}$:

$$\boldsymbol{\theta}[k+1] = \underset{\boldsymbol{\theta}}{\text{argmax}}\, J(R_{\mathbf{a}}; \{\boldsymbol{\tau}_i[k]\}, \boldsymbol{\theta})$$

$$\{\boldsymbol{\tau}_i[k+1]\} = \underset{\{\boldsymbol{\tau}_i\}}{\text{argmax}}\, J(R_{\mathbf{a}}; \{\boldsymbol{\tau}_i\}, \boldsymbol{\theta}[k+1])$$

  - La secuencia de los valores de J es no-decreciente [Daudin et al'08]
  - Resultados de consistencia para $N_v \to \infty$, Q fijo [Celisse et al'12]

¿Cuántas clases usar?

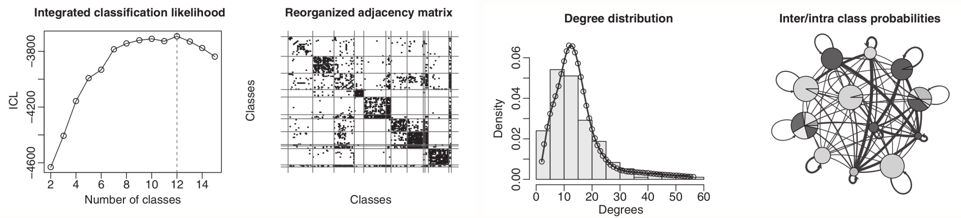
- El número de clases Q es generalmente desconocido y debe ser estimado
  - $\Rightarrow$ Se pueden usar técnicas bayesianas para selección de modelos
  - $\Rightarrow$ Distribución a priori $g(\boldsymbol{\theta} \,|\, m_Q)$ de $\boldsymbol{\theta}$ dado que el SBM m_Q tiene Q clases
- El criterio Integrated Classification Likelihood (ICL) resulta

$$\begin{aligned}\mathrm{ICL}(m_Q) = & \max_{\boldsymbol{\theta}} \log \ell_{\boldsymbol{\theta}, m_Q}(\mathbf{a}, \hat{\mathbf{z}}(\boldsymbol{\theta})) \\ & - \frac{Q(Q+1)}{4} \log \frac{N_v(N_v - 1)}{2} - \frac{Q-1}{2} \log N_v\end{aligned}$$

donde $\ell_{\boldsymbol{\theta}, m_Q}(\mathbf{a}, \hat{\mathbf{z}}(\boldsymbol{\theta}))$ es como antes pero con m_Q explícito

Evaluando la bondad del ajuste

- Goodness-of-fit $\Rightarrow$ mayormente computacionales y basados en visualización
- Ej: Red de blogs políticos franceses de octubre 2006 [Kolaczyk'17]
  - $\Rightarrow$ Ajuste a un SBM usando el método variacional (**mixer** en R)



- Valor óptimo $\hat{Q} = 12$, pero $Q \in [8, 12]$ parece razonable (9 partidos políticos)
  - $\Rightarrow$ Permutar la matriz de adyacencia muestra estructura (margen para agrupar)
- La distribución de grados ajusta razonablemente bien

Estimación del grafón

- **Objetivo:** estimar el grafón f de la realización observada G^{obs}
- Regresión no-paramétrica: estimar f dados $\{A_{ij}, U_i, U_j\}_{i,j\in\mathcal{V}}$
  $\Rightarrow$ Desafío es que los puntos $U_1, \ldots, U_{N_v}$ son latentes

Aproximación por SBM

C. Gao et al, "Rate-optimal graphon estimation," *Annals of Statistics*, vol. 43, 2015

Estimación por histograma (orden y suavizado)

S. H. Chan and E. M. Airoldi, "A consistent histogram estimator for exchangeable graph models," *ICML*, 2014

Modelo como proceso Gaussiano

P. Orbanz and D. M. Roy, "Bayesian models of graphs, arrays and other exchangeable random structures," *IEEE Trans. PAMI*, vol. 37, 2015

Extensiones de SBMs

Degree-corrected SBMs

- Comunidades con una distribución de grados "amplia"

B. Karrer B and M. E. Newman, "Stochastic blockmodels and community structure in networks," *Physical Review E.*, vol. 83, 2011

Mixed-membership SBMs

- Los nodos pueden pertenecer a más de una clase

E. M. Airoldi, "Mixed membership stochastic blockmodels," *J. Machine Learning Research*, vol. 9, 2008

SBMs jerárquicos

- Clustering jerárquico combinado con SBMs

A. Clauset et al, "Hierarchical structure and the prediction of missing links in networks," *Nature*, vol. 453, 2008

Modelos de Grafos Aleatorios

1. Modelos de Variables Latentes

2. Random dot product graphs

Random dot product graphs

- Consideremos un espacio latente $\mathcal{X}_d \subset \mathbb{R}^d$ tal que para todo

$$\mathbf{x}, \mathbf{y} \in \mathcal{X}_d \quad \Rightarrow \quad \mathbf{x}^\top \mathbf{y} \in [0,1]$$

  $\Rightarrow$ Distribución de producto interno $F : \mathcal{X}_d \mapsto [0,1]$

- Random dot product graphs (RDPGs):

$$\mathbf{x}_1, \ldots, \mathbf{x}_{N_v} \overset{\text{i.i.d.}}{\sim} F,$$
$$A_{ij} \,\big|\, \mathbf{x}_i, \mathbf{x}_j \sim \text{Bernoulli}(\mathbf{x}_i^\top \mathbf{x}_j)$$

  para $1 \leq i, j \leq N_v$, donde $A_{ij} = A_{ji}$ y $A_{ii} \equiv 0$

- Tipo particularmente intuitivo y "tratable" de grafos aleatorios con posiciones latentes

  $\Rightarrow$ Posición de los vértices $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_v}]^T \in \mathbb{R}^{N_v \times d}$

S. J. Young and E. R. Scheinerman, "Random dot product graph models for social networks," *WAW*, 2007

Conexión a otros modelos

- RDPG incluye varios modelos clásicos de grafos aleatorios

  Ej: Erdös-Renyi $\mathrm{ER}(N_v, p)$ es un RDPG con $d = 1$ y $\mathcal{X}_d = \{\sqrt{p}\}$

  Ej: SBM es un RDGP tal que F es

$$\mathrm{P}\left(\mathbf{X} = \mathbf{x}_q\right) = \alpha_q, \quad q = 1, \ldots, Q$$

  con d y $\mathbf{x}_1, \ldots, \mathbf{x}_Q$ tal que $\pi_{qr} = \mathbf{x}_q^\top \mathbf{x}_r$

- Esto último implica que los RDPGs son al menos tan expresivos como los SBMs

Conexión a otros modelos

- RDPG incluye varios modelos clásicos de grafos aleatorios

  Ej: Erdös-Renyi ER(N_v, p) es un RDPG con $d = 1$ y $\mathcal{X}_d = \{\sqrt{p}\}$

  Ej: SBM es un RDGP tal que F es

$$\mathrm{P}\left(\mathbf{X} = \mathbf{x}_q\right) = \alpha_q, \quad q = 1, \ldots, Q$$

  con d y $\mathbf{x}_1, \ldots, \mathbf{x}_Q$ tal que $\pi_{qr} = \mathbf{x}_q^\top \mathbf{x}_r$

- Esto último implica que los RDPGs son al menos tan expresivos como los SBMs
- RDPGs son un caso especial de modelos de posiciones latentes [Hoff et al'02]

$$A_{ij} \,\big|\, \mathbf{x}_i, \mathbf{x}_j \sim \text{Bernoulli}(\kappa(\mathbf{x}_i, \mathbf{x}_j))$$

$\Rightarrow$ RDPG aproxima cualquier $\kappa(\cdot)$ con un d suficientemente grande [Tang et al'13]

Estimación de las posiciones latentes

Q: Dado G de un RDPG ¿cuál es el "mejor" $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_v}]^\top$?

- MLE parece una buena idea, pero es impensado para N_v grandes

$$\hat{\mathbf{X}}_{ML} = \underset{\mathbf{X}}{\operatorname{argmax}} \prod_{i<j} (\mathbf{x}_i^\top \mathbf{x}_j)^{A_{ij}} (1 - \mathbf{x}_i^\top \mathbf{x}_j)^{1-A_{ij}}$$

Estimación de las posiciones latentes

Q: Dado G de un RDPG ¿cuál es el "mejor" $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_{N_v}]^\top$?

- MLE parece una buena idea, pero es impensado para N_v grandes

$$\hat{\mathbf{X}}_{ML} = \underset{\mathbf{X}}{\operatorname{argmax}} \prod_{i<j} (\mathbf{x}_i^\top \mathbf{x}_j)^{A_{ij}} (1 - \mathbf{x}_i^\top \mathbf{x}_j)^{1-A_{ij}}$$

- Sea $P_{ij} = \mathrm{P}\left((i,j) \in \mathcal{E}\right)$ y definamos la matriz $\mathbf{P} = [P_{ij}] \in [0,1]^{N_v \times N_v}$
  - $\Rightarrow$ El modelo RDPG especifica que $\mathbf{P} = \mathbf{X}\mathbf{X}^\top$
  - $\Rightarrow$ **Clave:** La $\mathbf{A}$ observada es una realización ruidosa de $\mathbf{P}$ $(\mathbb{E}[A] = \mathbf{P})$
- Sugiere aplicar mínimos cuadrados para hallar $\mathbf{X}$

$$\hat{\mathbf{X}}_{LS} = \underset{\mathbf{X}}{\operatorname{argmin}} \|\mathbf{X}\mathbf{X}^\top - \mathbf{A}\|_F^2$$

Adjacency spectral embedding

- Como $\mathbf{A}$ es real y simétrica, podemos descomponerla como $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$
  - $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_{N_v}]$ es la matriz ortogonal de vectores propios
  - $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{N_v})$, con valore propios $\lambda_1 \geq \ldots \geq \lambda_{N_v}$

Adjacency spectral embedding

- Como $\mathbf{A}$ es real y simétrica, podemos descomponerla como $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$
  - $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_{N_v}]$ es la matriz ortogonal de vectores propios
  - $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_{N_v})$, con valore propios $\lambda_1 \geq \ldots \geq \lambda_{N_v}$
- Definamos $\hat{\mathbf{\Lambda}} = \mathrm{diag}(\lambda_1^+, \ldots, \lambda_d^+)$ y $\hat{\mathbf{U}} = [\mathbf{u}_1, \ldots, \mathbf{u}_d]$ $(\lambda^+ := \max(0, \lambda))$
- La mejor aproximación de rango d semi-definida positiva de $\mathbf{A}$ es $\hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{U}}^\top$

$\Rightarrow$ Ajacency spectral embedding (ASE) es $\hat{\mathbf{X}}_{LS} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}^{1/2}$ dado que

$$\mathbf{A} \approx \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{U}}^\top = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}^{1/2}\hat{\mathbf{\Lambda}}^{1/2}\hat{\mathbf{U}}^\top = \hat{\mathbf{X}}_{LS}\hat{\mathbf{X}}_{LS}^\top$$

Adjacency spectral embedding

- Como $\mathbf{A}$ es real y simétrica, podemos descomponerla como $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$
  - $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_{N_v}]$ es la matriz ortogonal de vectores propios
  - $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1, \ldots, \lambda_{N_v})$, con valore propios $\lambda_1 \geq \ldots \geq \lambda_{N_v}$
- Definamos $\hat{\boldsymbol{\Lambda}} = \operatorname{diag}(\lambda_1^+, \ldots, \lambda_d^+)$ y $\hat{\mathbf{U}} = [\mathbf{u}_1, \ldots, \mathbf{u}_d]$ $(\lambda^+ := \text{máx}(0, \lambda))$
- La mejor aproximación de rango d semi-definida positiva de $\mathbf{A}$ es $\hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{U}}^\top$

$\Rightarrow$ Ajacency spectral embedding (ASE) es $\hat{\mathbf{X}}_{LS} = \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}^{1/2}$ dado que

$$\mathbf{A} \approx \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}\hat{\mathbf{U}}^\top = \hat{\mathbf{U}}\hat{\boldsymbol{\Lambda}}^{1/2}\hat{\boldsymbol{\Lambda}}^{1/2}\hat{\mathbf{U}}^\top = \hat{\mathbf{X}}_{LS}\hat{\mathbf{X}}_{LS}^\top$$

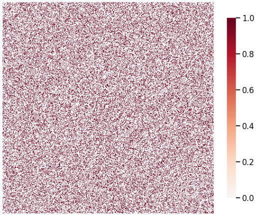
Q: ¿La solución es única? Nop, dado que el producto interno es invariante a rotaciones

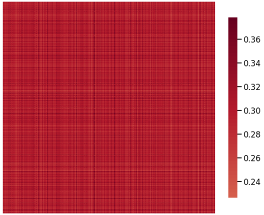
$$\mathbf{P} = \mathbf{X}\mathbf{W}(\mathbf{X}\mathbf{W})^\top = \mathbf{X}\mathbf{X}^\top, \quad \mathbf{W}\mathbf{W}^\top = \mathbf{I}_d$$

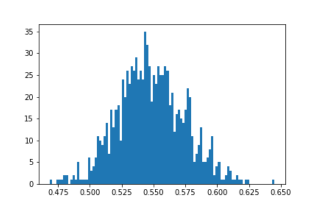
$\Rightarrow$ El embedding de un RDPG es identificable modulo rotaciones

Embedding de un grafo Erdös-Renyi

- Ej: Grafo Erdös-Renyi $\mathrm{ER}(1000, 0{,}3)$ (realización de $\mathbf{A}$ a la izquierda)







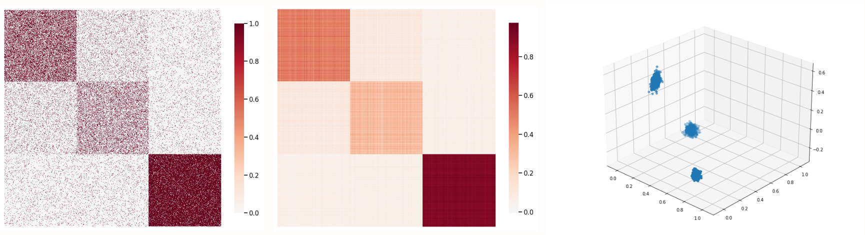
- Para $d = 1$ calculamos el ASE $\hat{\mathbf{x}}_{LS}$ y graficamos $\hat{\mathbf{x}}_{LS}\hat{\mathbf{x}}_{LS}^\top$ (centro)
  - $\Rightarrow$ Buena aproximación de la matriz constante $\mathbf{P} = 0{,}3 \times \mathbf{1}\mathbf{1}^\top$
  - $\Rightarrow$ El histograma de las entradas de $\hat{\mathbf{x}}_{LS}$ da más evidencia (derecha, $\sqrt{p} = 0{,}547$)
- Resultados sobre consistencia y distribuciones límites disponibles

A. Athreya et al., “Statistical inference on random dot product graphs: A survey,” *J. Mach. Learn. Res.*, vol. 18, pp. 1-92, 2018

Embedding de un grafo SBM

- Ej: SBM con $N_v = 1500$, $Q = 3$ y parámetros

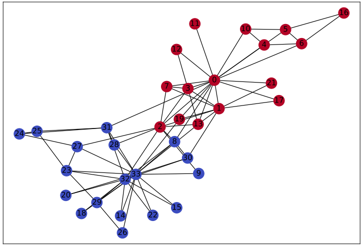
$$\boldsymbol{\alpha} = \begin{bmatrix} 1/3 \\ 1/3 \\ 1/3 \end{bmatrix}, \quad \boldsymbol{\Pi} = \begin{bmatrix} 0{,}5 & 0{,}1 & 0{,}05 \\ 0{,}1 & 0{,}3 & 0{,}05 \\ 0{,}05 & 0{,}05 & 0{,}9 \end{bmatrix}$$

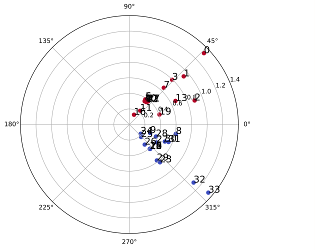


- Realización de la adyacencia (izquierda), $\hat{\mathbf{X}}_{LS}\hat{\mathbf{X}}_{LS}^{\top}$ (centro), filas de $\hat{\mathbf{X}}_{LS}$ (derecha)
- Justifica el uso de métodos geométricos (e.g. KNN o GMM del espectro)

Interpretabilidad de los embeddings

- Ex: Zachary's karate club con $N_v = 34$, $N_e = 78$ (izquierda)





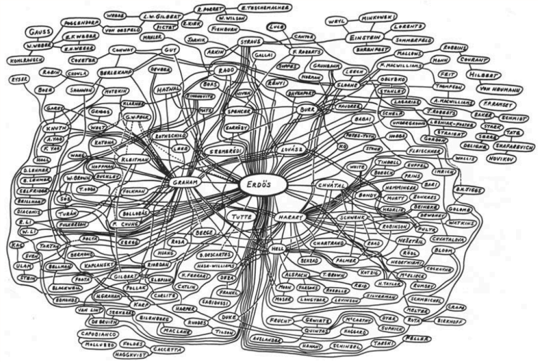
- Embeddings (filas de $\hat{\mathbf{X}}_{LS}$) para $d = 2$ (derecha)
  - El administrador del club ($i = 0$) y el instructor ($j = 33$) son ortogonales
- La interpretabilidad es una característica muy valiosa de los RDPGs

  $\Rightarrow$ Alineación del vector indica afinidad entre nodos (comunidad)

  $\Rightarrow$ Magnitud del vector indica conectividad del nodo

Grafo de colaboración entre matemáticos

- Ej: Red de colaboración entre matemáticos centrado en Paul Erdős



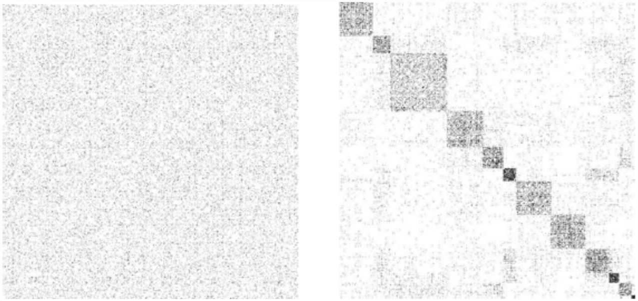
- La mayoría de los matemáticos tienen un número de Erdős de no más de 4 o 5

  ⇒ Dibujo creado por R. Graham en 1979

Grafo de colaboración entre matemáticos

- Grafo de co-autoría G, $N_v = 4301$ nodos con número de Erdős ≤ 2

  $\Rightarrow$ No hay una estructura clara de la matriz de adyacnecia "cruda" $\mathbf{A}$ (izquierda)



- La estructura de comunidades se revela al permutar la matriz (derecha)

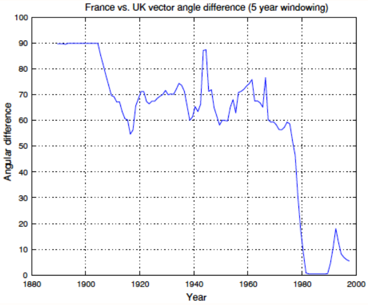
  (i) Se obtiene el ASE $\hat{\mathbf{X}}_{LS}$ de los matemáticos

  (ii) K-means angular en las filas de $\hat{\mathbf{X}}_{LS}$ [Scheinerman-Tucker'10]

Relaciones internacionales

- Ej: Red dinámica G_t de relaciones entre naciones

  $\Rightarrow (i,j) \in \mathcal{E}_t$ si las naciones tuvieron un tratado de alianza en el año t



- Ángulo entre el embedding del Reino Unido y Francia entre 1890 y 1995
  - Ortogonal durante fines del siglo XIX
  - Se acercan durante las guerras, y se vuelve alejar durante la ocupación nazi en la segunda guerra
  - Fuertemente alineados a partir de los 70 hasta la creación de la UE

Extensiones

- Ignoramos la diagonal $[\hat{\mathbf{X}}_{LS}\hat{\mathbf{X}}_{LS}^{\top}]_{ii} = 0$. Se puede resolver mediante algoritmos iterativos o descenso por gradiente

  E. R. Scheinerman and K. Tucker, "Modeling graphs using dot product representations," *Comput. Stat.*, vol. 25, pp. 1-16, 2010

  M. Fiori et al., "Algorithmic Advances for the Adjacency Spectral Embedding," *EUSIPCO*, 2022

- ¿$\mathbf{A}$ no es semi-definida positiva? Extensión conocida como generalized RDPG

  P. Rubin-Delanchy et al, "A statistical interpretation of spectral embedding: The generalised random dot product graph," *arXiv:1709.05506 [stat.ML]*, 2017

- Considerar el caso con pesos y/o dirigido es posible

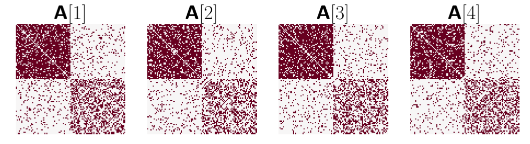
  B. Marenco et al., "Online Change Point Detection for Weighted and Directed Random Dot Product Graphs," *IEEE TSIPN*, 2022

- Permite varias aplicaciones en testing, clustering, **change-point detection**, ...

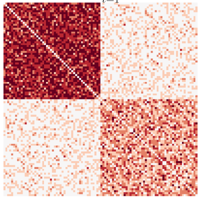
Online change point detection: entrenamiento

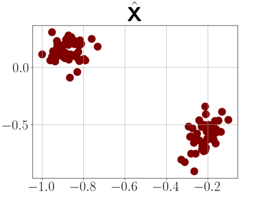
- **Idea:** Usar "estimating function" [Kirsch-Tadjuidje'15]

  $\Rightarrow$ Entrenar con m grafos "limpios" (sin cambios)



$$\bar{\mathbf{A}} = \frac{1}{m}\sum_{t=1}^{m}\mathbf{A}[t]$$

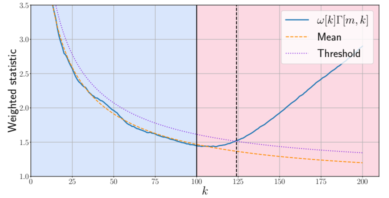




Online change point detection: Monitoreo

- Observar las matrices secuencialmente $\mathbf{A}[m+1], \mathbf{A}[m+2], \ldots$
  - Monitorear la suma acumulada $\mathbf{S}[m,k] = \sum_{t=m+1}^{m+k} \left(\hat{\mathbf{X}}\hat{\mathbf{X}}^\top - \mathbf{A}[t]\right)$

Proposición: Para k grande y bajo la hipótesis nula de no-cambio, $\Gamma[m,k] := \|\mathbf{S}[m,k]\|^2$ tiene una distribución χ^2 generalizada.

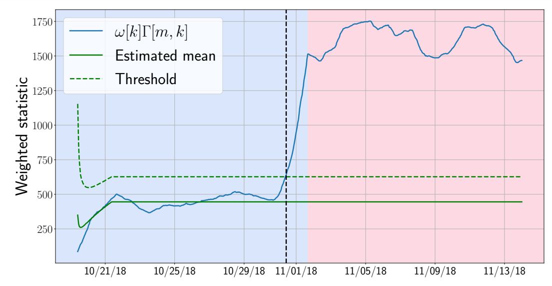


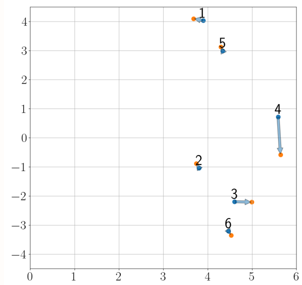
- Liviano: requisito de memoria y computacional de $O(N^2)$

Monitoreo de una red inalámbrica

- Extensión del modelo para incluir grafos con pesos y dirigidos
- Red inalámbrica del Plan Ceibal. Medidas@hora del RSSI para $N = 6$ APs

  $\Rightarrow$ Según el administrador de red: *AP 4 fue movido el 30 de octubre*

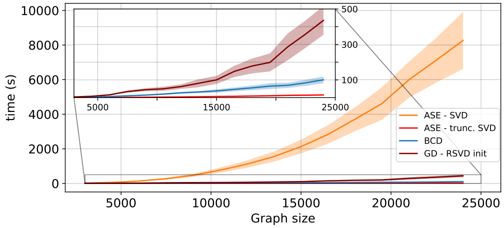


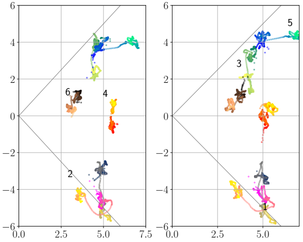


- Explicar el cambio a través de la interpretabilidad del ASE
- Reproducibilidad $\Rightarrow$ Código disponible en @ `https://github.com/git-artes/cpd_rdpg`

Futuro

- Descenso por gradiente para hallar el ASE: escalabilidad y tracking
  - Manejo de datos faltantes, alineación mediante inicialización (M. Fiori et al. "Gradient-Based Spectral Embeddings of Random Dot Product Graphs", arXiv preprint arXiv:2307.13818, submitted)





- Propiedades estadísticas del modelo no-paramétrico del weighted RDPG

$$\mathbb{E}\left[e^{tA_{ij}}|\mathbf{X}\right] = \sum_{m=0}^{\infty} \frac{t^m \mathbb{E}\left[A_{ij}^m\right]}{m!} = 1 + \sum_{m=1}^{\infty} \frac{t^m \mathbf{x}_i^\top[m]\mathbf{x}_j[m]}{m!}$$