



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Tesis de Licenciatura

Algoritmos para detección de comunidades

Nicolás Agote

Directora: Inés Armendáriz

Fecha de Presentación: 10 de abril de 2024

Resumen

El modelo estocástico de bloques (SBM por su sigla en inglés) es un grafo aleatorio donde los vértices se encuentran divididos en clases (o comunidades) que comparten una etiqueta, y la probabilidad de conexión entre cada par de vértices está determinada de sus etiquetas. Es uno de los modelos más populares utilizados en diversas aplicaciones en redes biológicas, sociales y tecnológicas para estudiar el problema de identificación de comunidades, en el que se trata de identificar la etiqueta de los vértices a partir de la observación de una realización del grafo.

En esta tesis repasamos resultados fundacionales en la literatura, como la existencia de una transición de fase que determina cuándo es posible identificar las comunidades. Además, exploramos y demostramos la consistencia de algunos métodos espectrales, métodos de minimización de cortes y métodos por extensiones armónicas como posibles enfoques para resolver el problema, y realizamos aportes originales en las dos últimas familias de métodos. Por último, analizamos el desempeño de estos estimadores a través de simulaciones.

Agradecimientos

"Y Frodo no habría llegado muy lejos sin Sam, ¿verdad, papá?"

El Señor de los Anillos: Las dos torres.

Fue un largo camino el que desembocó en esta tesis, trayecto que fue un gran regalo. Es muchísimo lo que recibí. No hablo solo en cuanto a la matemática: iremos citando grandes trabajos e ideas sobre las cuales esta tesis espera aportar su granito de arena. Hablo también de cariño, dedicación y acompañamiento, de paciencia, generosidad y conocimiento, de confianza, validación y sustento... de muchas personas por las cuales podemos estar disfrutando de leer este trabajo en este momento. Ojalá haya sabido reflejar esas tantas cosas buenas que me dieron, que aprendí y confío que seguiré aprendiendo de todos ustedes.

¿Como se dio ese camino? Empezó alguna vez en una reunión en la oficina de Inés Armendáriz, en la cual junto con Pablo Ferrari y Florencia Leonardi nos animamos a explorar algunas ideas para detectar comunidades, y entre ellos los estimadores por extensiones armónicas. No tardé mucho tiempo en perderme en el problema para seguir profundizando en entre futuras reuniones. Fue el disparador de un año de trabajo que trajo como fruto esta tesis que estás leyendo. En ese tiempo, tuve la suerte además de compartir estas ideas y nutrirme de comentarios de mucha gente, que completó la investigación que hizo a esta tesis. De Julio Rossi, con quien exploramos las nociones de cortes y posibles aplicaciones. De charlas y pósters en seminarios y congresos: en el seminario de Modelos Estocásticos organizado por Pablo Groisman, en un ciclo de charlas organizado por Matthieu Jonckheere; en el CLAPEM, en la UMA, en "Probabilistic and Game Theoretic interpretations of PDEs", en la escuela de probabilidades de Chile y en el PISA. En charlas con amigos, docentes, alumnos... ¡e incluso familiares! Y también de valiosos aportes de los jurados, Dani Galicer y Nico Saintier, para llegar a esta versión final. Entre otros, gracias a ellos tenemos una sección más amplia acerca de la correctitud y complejidad de los algoritmos.

Pero... el camino empezó aún antes. Cuando llegué a la facultad de Exactas, que supo convertirse rápidamente en un lugar muy querido. Donde recibí ampliamente, durante toda la carrera, pero también de infiltrado en materias de computación y del profesorado, en congresos y grupos de seminario, en la banda Lambda, y durante el CBC. Y como ayudante, con docentes, pares y alumnos. Con grupos con los cuales pude atreverme a soñar, noches de juegos dentro y fuera de la facultad, Learning Experiment, CIMA, TAP, tenis, cine, básquet, fútbol con el CALI... Con el aventón que recibí del colegio como alumno y docente, y de mi paso por las olimpiadas de matemática. Y, desde ya, con el constante y fundamental apoyo de mucha gente, de amigos y de familia: es un regalo precioso todo lo que me han querido y acompañado. Ojalá me hayan encontrado y me sepan agradecido.

Gracias, muchas gracias.

Y gracias a Vos, que estas leyendo esta tesis: cuando veas algo bueno, ya sabés de donde vino, y ojalá encuentres Vos también algo lindo para llevar.

Índice general

1. Introducción	6
2. Grafos, grafos aleatorios y el modelo estocástico de bloques	9
2.1. Grafos	9
2.2. Algunas preliminares	11
2.3. El grafo de Erdős-Renyi	14
2.4. El modelo estocástico de bloques	15
2.5. Notas	23
3. Detección de comunidades en el modelo estocástico de bloques	24
3.1. El problema de detección de comunidades	24
3.2. Estimadores maximales y límites de la estimación	28
3.3. Métodos espectrales	34
3.4. Consistencia débil para métodos espectrales	38
3.5. Consistencia fuerte para métodos espectrales	39
3.6. Notas	50
4. Estimadores de minimización de corte	53
4.1. Cortes balanceados y estimadores de minimización de corte	53
4.2. Consistencia para el minimizante del corte de Cheeger	56
4.3. Consistencia para el minimizante del corte ratio	71
4.4. Notas	73
5. Estimadores por extensiones armónicas	75
5.1. Motivación, funciones armónicas y paseos aleatorios	75
5.2. El estimador por extensiones armónicas para dos comunidades balanceadas	84
5.3. Consistencia débil del estimador por extensiones armónicas	89
5.4. El estimador por extensiones armónicas en el caso general	93
5.5. Consistencia fuerte del estimador por extensiones armónicas	100
5.6. Notas	112
6. Simulaciones	114
6.1. Algoritmos para los calcular los estimadores	114

0. Índice general	5
6.2. Consistencia y complejidad de los algoritmos para aproximar \hat{C}_{HEV}	118
6.3. Implementación de los estimadores y simulaciones	135
7. Conclusiones y trabajo futuro	139
Bibliografía	140
Lista de teoremas	140
Índice de figuras	145
A. Anexo	147
A.1. Demostraciones del capítulo 2	147
A.2. Demostraciones del capítulo 4	158
A.3. Demostraciones del capítulo 5	170
A.3.1. Consistencia fuerte para los estimadores por extensiones armónicas en dos comunidades balanceadas	170
A.3.2. Consistencia débil para los estimadores por extensiones armónicas en el caso general	171

1. Introducción

¿En qué consiste este trabajo? En resumidas palabras, vamos a explorar distintos métodos para atacar el problema de detección de comunidades en modelos estocásticos de bloques, realizando contribuciones originales tanto en el estudio de los métodos de minimización de cortes como en el entendimiento de algunos métodos semi-supervisados que hacen uso de paseos al azar y pueden dar una mejora respecto de métodos en la literatura al estimar la comunidad de subconjuntos chicos de nodos. Pero antes de que nos adentremos en estas cuestiones, como este problema tiene su historia, veamos de qué se tratan estos conceptos y cómo llegamos aquí a estudiarlos.

Comenzaremos entonces nombrando los **grafos 2.1.1**, también conocidos como redes. Estos son importantes herramientas matemáticas y permiten modelar fenómenos muy variados: desde rutas entre diversos puntos en una ciudad, como los encontró Euler en 1736 al resolver el problema de los puentes de Königsberg, hasta posibles estados y acciones en un juego, o también redes sociales. No es de extrañar entonces que sean objetos esenciales de estudio en áreas como aprendizaje automático y ciencias de la computación pero también álgebra abstracta.

Partiendo con Erdős y Renyi en 1959 se comenzaron a estudiar **grafos aleatorios**. Por ejemplo, en el modelo que lleva sus nombres **2.3.1** se fija un número de nodos n y se agregan aristas entre cada par de ellos con una probabilidad p fija de forma independiente. Como cuenta Remco van der Hofstad en la brillante introducción a su libro [vdH17], rápidamente estos grafos aleatorios cobraron relevancia en el estudio y comprensión de redes complejas. Parte de la relevancia del estudio de estos objetos radica en describir el comportamiento global de estructuras muy complejas, que no admiten una representación especialmente esclarecedora al considerarlos en su totalidad, partiendo de los entornos de cada nodo, que muchas veces sí admiten una descripción más sencilla. Por ejemplo, en los grafos de k vecinos más cercanos o el ε -grafo los nodos se colocan en un espacio de acuerdo a alguna distribución y luego se agregan aristas mediante algún criterio geométrico (fijo un nodo, de estar entre los k nodos más cercanos a él o a distancia más cercana que ε); es mediante esta descripción local que se pueden estudiar nociones globales como conexión, más generalmente la homología, pero también percolación y otras variantes.

Entre los diversos problemas que se estudian en grafos (tanto determinísticos como aleatorios) está el problema de **clustering**, que consiste en particionar de alguna manera a los nodos de un grafo en subconjuntos de nodos similares. Por ejemplo, nubes de nodos bien

separadas de otras en un grafo geométrico, o perfiles similares en una red social.

No todos los grafos presentan una estructura con subconjuntos de nodos similares, pero realmente disímiles entre grupos: de hecho es una noción elusiva de determinar en general. Por ejemplo, en el modelo de Erdős y Renyi podemos observar que todos los nodos son esencialmente iguales, y en general no deberíamos encontrar estos cúmulos. Es así que generalizando llegamos finalmente al **modelo estocástico de bloques 2.4.1** (SBM por sus siglas en inglés), propuesto en 1983 por Holland, Larsky, Blackmond y Leinhardt y objeto de gran relevancia en el estudio actual como caso fundamental de grafos cuyos nodos muestran un comportamiento similar de a grupos. Más detalladamente, el modelo consiste de un grafo cuyos n nodos pertenecen a alguna de k comunidades, y donde se agregan aristas aleatoriamente entre cada par de nodos solo dependiendo de las comunidades a las que pertenecen. Ya en la literatura se han estudiado las propiedades del modelo estocástico de bloques: algunas relativas a nociones globales como su conexión o el tamaño de las componentes conexas más grandes partiendo de las nociones conocidas para los grafos de Erdős y Renyi; y otras relativas al estudio del problema de clustering, o de **detección de comunidades 3.1.1** como se conoce en este contexto. El objetivo es dado un grafo que proviene de este modelo, y solo con la información que nos dan las conexiones que tiene cada nodo, predecir cuáles nodos forman parte de la misma comunidad.

Hay muchos enfoques para atacar este problema, como son por ejemplo métodos de maximización de modularidad [CL23], métodos de programación semidefinida [Abb18], métodos espectrales [Abb18, AFWZ20, vL07], métodos de corte mínimo [HS11], métodos que usan la técnica de “partir el grafo” [Abb18, MNS16], métodos que utilizan paseos aleatorios y métodos de propagación de etiquetas. Crucialmente y bajo hipótesis adecuadas se ha demostrado que es posible recuperar las comunidades.

¿Qué significa recuperar las comunidades? Hay muchos criterios, pero en este trabajo nos concentraremos fundamentalmente en dos: **consistencia fuerte** y **consistencia débil 3.1.2**; en castellano son descubrir exactamente los subconjuntos de nodos que conforman cada una de las comunidades con alta probabilidad para grafos grandes, o recuperar los subconjuntos salvo una cantidad no representativa de errores con alta probabilidad para grafos grandes. Remarcamos que los criterios son asintóticos: necesitamos una alta cantidad de nodos para asegurar que las estimaciones funcionen bien. Sin embargo, como observamos en simulaciones y aplicaciones, más allá de garantías teóricas vemos en la práctica que con un número moderado de nodos los estimadores ya tienen una buena precisión.

¿En qué consiste luego este trabajo?

Primero, en una introducción más detallada de los grafos, grafos aleatorios y del modelo estocástico de bloques, cosa que haremos en el Capítulo 2. En particular, estudiaremos la propiedad de conexión para estos grafos aleatorios, que se dará en una transición de fase, como veremos en el Teorema 2.4.8. Durante este Capítulo nos basaremos principalmente en los resultados en [Abb18, MNS16].

A continuación, usaremos estas herramientas e intuiciones para definir y explorar el pro-

blema de detección de comunidades 3.1.1, que introduciremos formalmente junto con algunos resultados destacados de la literatura en el Capítulo 3. Más detalladamente, exploraremos la transición de fase que indica cuándo es posible detectar las comunidades 3.1.4, basándonos nuevamente en la exposición en [Abb18]. Introduciremos algunos estimadores maximales en cuanto a la precisión y también los métodos espectrales, que resultarán ser fuertemente consistentes bajo hipótesis adecuadas y motivarán intuiciones en el Capítulo 5.

A partir de aquí continuaremos exponiendo trabajo original, bajo la dirección de Inés Armendáriz, y en colaboración con Pablo Ferrari, Florencia Leonardi y Julio Rossi.

En el Capítulo 4 exploraremos los métodos de minimización de corte 3.2.3. Estos buscan explotar la intuición de que, bajo hipótesis adecuadas, las comunidades arman cúmulos bien interconectados en su interior pero con escasas conexiones entre sí para tener buenas estimaciones. Mostraremos que algunas variantes, el corte de Cheeger y el corte ratio 4.1.3, son fuertemente consistentes apelando a argumentos probabilísticos.

A continuación, en el Capítulo 5 introduciremos estimadores basados en recuperar información de las comunidades realizando paseos aleatorios en el grafo, los estimadores por extensiones armónicas 5.2.1. A diferencia de los estimadores anteriores, para estos necesitaremos conocer con anterioridad las comunidades de algunos nodos en el grafo, a partir de los cuales buscaremos predecir las comunidades de los nodos restantes. Mostraremos que tiene motivaciones muy naturales, y como contribución original veremos que son débilmente consistentes y también daremos una variante 5.5.3 que es además fuertemente consistente.

Por último, en el Capítulo 6 estudiaremos distintas maneras de implementar los métodos, mostraremos su correctitud y haremos unas simulaciones para evaluar su desempeño. Mencionamos en este punto que aquí se encuentra una de las grandes virtudes de los estimadores por extensiones armónicas, que preservan la información ya conocida y pueden implementarse mediante métodos de Montecarlo que, como es habitual en la práctica, conjeturamos que resultarán más rápidos si solamente se requiere estimar la comunidad de pocos nodos.

Finalmente, como conclusión y antes de la bibliografía, el índice de teoremas y el anexo, en el Capítulo 7 damos algunas direcciones en las que podemos continuar nuestro estudio. En estas líneas mencionamos las notas al final de cada capítulo que introducirán posibles líneas de investigación o mejora y conectarán con artículos en la literatura.

Sin más preámbulos comencemos entonces nuestra exploración de los modelos estocásticos de bloques y algoritmos para detección de comunidades.

2. Grafos, grafos aleatorios y el modelo estocástico de bloques

2.1 Grafos

Como adelantamos en la introducción a la tesis, estaremos tratando con un problema modelado sobre grafos (aleatorios), así que en este capítulo repasaremos algunas de las definiciones fundamentales y fijaremos la notación que usaremos a lo largo del trabajo.

[2.1.1] Definición: Grafos

Un grafo no dirigido y en adelante simplemente **grafo** $G = (V, E)$ está dado por un conjunto de **nodos** V y un conjunto de **aristas** $E \subseteq \{\{v, w\} \mid v, w \in V\}$ que conectan pares de nodos.

Decimos que la arista $e = \{v, w\}$ **incide** en v y en w y que éstos son **extremos** de la arista e . Llamaremos en este caso a v y w nodos **vecinos** y notaremos $v \sim w$; llamaremos además $N(v)$ al conjunto de vecinos de v . Diremos además que dos aristas son **adyacentes** cuando tienen algún extremo en común.

El caso degenerado de arista $\{v\}$ se conoce como un **bucle**, y diremos que un grafo es **simple** cuando no cuenta con bucles.

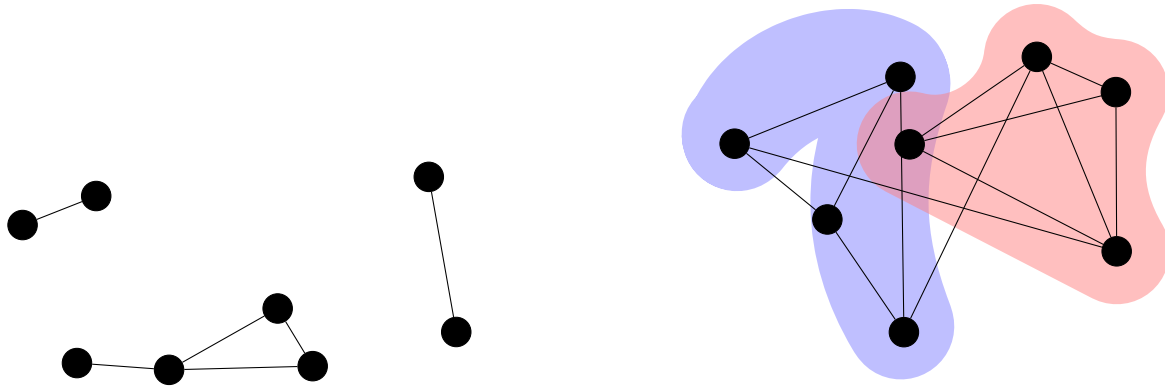
Llamamos **grado** de un nodo v a $\deg(v)$ la cantidad de veces que v es extremo de una arista en E , donde para cada bucle el nodo será extremo dos veces. Cuando el grafo es simple, el grado de v coincide con la cantidad de aristas que inciden en él. Diremos que un nodo v es **aislado** cuando $\deg(v) = 0$.

En adelante, trabajaremos siempre con grafos simples, a los cuales representaremos, como en la Figura 2.1, mediante diagramas con puntos para los nodos y segmentos para las aristas.

Precisaremos también algunas definiciones complementarias:

[2.1.2] Definición: Caminos, ciclos, conexión y distancia en grafos

Dado un grafo $G = (V, E)$, llamaremos **camino** entre v y w nodos en V a una secuencia alternada $v_0 e_1 v_1 \dots v_{k-1} e_k v_k$ de nodos $v_0, \dots, v_k \in V$ y aristas $e_1, \dots, e_k \in E$ que cumplen que $e_i = \{v_{i-1}, v_i\}$ para cada $1 \leq i \leq k$ y donde $v_0 = v$ y $v_k = w$, para



(a) Este primer grafo es simple y desconexo. Podemos partirlo en tres componentes conexas, donde solamente la más grande tiene un ciclo. En las componentes conexas de dos nodos, cada nodo tiene grado 1.

(b) Este segundo grafo también es simple, y esta vez es conexo. La distancia más grande entre nodos de este grafo es de 3, y entre sus muchos ciclos podemos encontrar uno que pasa por todos los nodos exactamente una vez.

Figura 2.1: Algunos ejemplos de grafos.

algún $k \in \mathbb{N}_0$.

$$v_0 \xrightarrow{e_1} v_1 \xrightarrow{e_2} \dots \xrightarrow{e_{k-1}} v_{k-1} \xrightarrow{e_k} v_k \quad (2.1)$$

Más aún, en ese caso diremos que el camino tiene **longitud** k . Además, diremos que es un **camino simple** cuando no repite vértices, salvo tal vez el primero v_0 y el último v_k .

Llamaremos **ciclo** a un camino $v_0 e_1 v_1 \dots v_k e_{k-1} v_k$ donde $v_i \neq v_j$ si $0 \leq i < k$ y además $v_0 = v_k$.

Diremos que un grafo es **conexo** cuando para cada par de nodos distintos v y w hay un camino entre v y w .

En grafos conexos tenemos una noción de **distancia** entre dos nodos v, w , dada por

$$\text{dist}(v, w) = \inf \{ \text{longitud}(\gamma) \mid \gamma \text{ es un camino entre } v \text{ y } w \text{ en } G \}. \quad (2.2)$$

Cuando un grafo no es conexo puede tomarse la distancia como infinito si no hay caminos entre un par de nodos dado, con lo que equivalentemente un grafo es conexo cuando todas las distancias entre nodos son finitas.

En general nuestros conjuntos de nodos no tendrán especial relevancia, y por eso adoptaremos la siguiente notación:

[2.1.3] Notación: Conjunto de nodos

Dado un grafo $G = (V, E)$, en adelante llamaremos $n = \#V$ a la cantidad de nodos en él, y los pensaremos numerados del 1 al n .

Será útil para descubrir ciertas características de nuestros grafos que estos pueden representarse además mediante matrices:

[2.1.4] Definición: Matriz de adyacencia de un grafo

Dado un grafo $G = (V, E)$ definimos su **matriz de adyacencia** asociada $A \in \{0, 1\}^{n \times n}$ como

$$A_{v,w} = \begin{cases} 1 & \text{si } v \sim w, \text{ y} \\ 0 & \text{en caso contrario.} \end{cases} \quad (2.3)$$

Remarcamos que la matriz de adyacencia de un grafo lo determina unívocamente. Además, notamos que estas matrices serán simétricas pues que dos nodos sean vecinos no depende del orden en que los nombremos, y tendrán diagonal cero pues no tienen bucles.

[2.1.5] Observación: Grafos pesados

Mencionaremos una generalización de los grafos que nos permitirá visualizar algunos “grafos” que aparecerán en futuros cálculos durante el trabajo y motivar ciertas notaciones, que son los grafos pesados. Ellos son grafos $G = (V, E)$ donde cada arista tiene asociada un número no negativo, su peso, y donde pensamos a 0 como la ausencia de la arista; equivalentemente podemos pensar también en una matriz de adyacencia generalizada $W \in \mathbb{R}_{\geq 0}^{n \times n}$ en donde cada coeficiente es el peso asociado a la arista. Además, se puede generalizar la noción de grado de un nodo como la suma de los pesos de las aristas adyacentes, y la distancia entre cada par de nodos como el ínfimo sobre la suma de los pesos de cada camino entre ellos.

Pasaremos en la siguiente sección a introducir algunas preliminares de teoría de probabilidades, que nos permitirán, junto con estas nociones de grafos, trabajar con grafos aleatorios.

2.2 Algunas preliminares

Antes que nada, introduciremos notación con la que nos manejaremos en el resto del trabajo:

[2.2.1] Notación: Repaso y notación de probabilidades

Llamaremos $(\Omega, \mathcal{F}, \mathbb{P})$ a los **espacios de probabilidad**, donde \mathbb{P} es una **medida de probabilidad**. Notaremos además \mathbb{E} para la **esperanza**, Var para la **varianza** y Cov para la **covarianza** asociadas a variables aleatorias. Dada X una variable aleatoria, notaremos como \mathbb{P}_X a la **distribución** asociada, F_X a la **función de distribución acumulada**, p_X a la función de **probabilidad puntual** en caso de ser discreta y f_X a la **densidad** en caso de ser absolutamente continua. Notaremos además $\sigma(X)$ a la **sigma-álgebra generada** por X , y usaremos \sim para notar que dos variables aleatorias o medidas tienen la misma distribución, y $\stackrel{\text{i.i.d.}}{\sim}$ cuando además son independientes. Recordamos que un **proceso estocástico** es una colección de variables aleatorias $(X_t)_{t \in I}$ definidas en un mismo espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$; nuestros procesos

estocásticos serán a tiempos discretos, e $I = \mathbb{N}_0$.

Notaremos $\xrightarrow{\text{a.s.}}$ para **convergencia casi segura**, $\xrightarrow{\mathbb{P}}$ para **convergencia en probabilidad**, $\xrightarrow{\mathcal{D}}$ para **convergencia en distribución**, y fijo algún espacio L^p notaremos $\xrightarrow{L^p}$ para **convergencia en L^p** .

Diremos además que una familia de eventos $(\mathcal{A}_n)_{n \in \mathbb{N}}$ ocurre **asintóticamente casi seguramente** o **con alta probabilidad** y notaremos a.a.s. cuando

$$\mathbb{P}(\mathcal{A}_n) \xrightarrow{n \rightarrow \infty} 1, \quad (2.4)$$

aunque dependiendo del contexto con alta probabilidad podrá referirse a que la velocidad de convergencia a 1 también sea suficientemente alta y estará claro en el contexto.

Por último, notaremos por $\mathbb{1}_{\mathcal{A}}$ a la función indicadora del evento \mathcal{A} .

A continuación fijaremos la notación que usaremos para medir el comportamiento asintótico de sucesiones de números reales:

[2.2.2] Notación: Notación de comportamiento asintótico

Dadas funciones $f, g : \mathbb{N} \rightarrow \mathbb{R}$, notaremos $g = o(f)$ cuando $\frac{g(n)}{f(n)} \xrightarrow{n \rightarrow \infty} 0$ y también diremos en ese caso que $f = \omega(g)$. Por otro lado, diremos que $g = \mathcal{O}(f)$ cuando $\limsup_{n \rightarrow \infty} \left| \frac{g(n)}{f(n)} \right| < \infty$, y una vez más diremos en este caso que $f = \Omega(g)$. Diremos que $g = \Theta(f)$ cuando ocurren ambas que $g = \mathcal{O}(f)$ y que $g = \Omega(f)$.

A lo largo de los cálculos, excepto que se indique lo contrario C será una constante que podría variar de línea en línea. Además, usaremos los símbolos \asymp , \lesssim y \gtrsim para referirnos a igualdades o desigualdades que valen salvo una constante, y a veces también para n suficientemente grandes.

Enunciaremos también algunas desigualdades clásicas de probabilidad que usaremos a lo largo del trabajo, junto con las desigualdades de Markov y Tchebyshev:

[2.2.3] Proposición: Desigualdad de Paley-Zygmund

Sea $Z \neq 0$ una variable aleatoria no negativa tal que $\text{Var}[Z] < \infty$ y sea $\theta \in [0, 1]$.

Luego, vale que

$$\mathbb{P}(\{Z > \theta \mathbb{E}[Z]\}) \geq \frac{(1 - \theta)^2 \mathbb{E}[Z]^2}{\mathbb{E}[Z^2]}. \quad (2.5)$$

Demostración. Para cada $\theta \in [0, 1]$, podemos demostrar la desigualdad de Paley-Zygmund comenzando por notar que por Cauchy-Schwarz vale que

$$\begin{aligned} \mathbb{E}[Z] &= \mathbb{E}[Z \mathbb{1}_{\{Z \leq \theta \mathbb{E}[Z]\}}] + \mathbb{E}[Z \mathbb{1}_{\{Z > \theta \mathbb{E}[Z]\}}] \\ &\leq \theta \mathbb{E}[Z] + \sqrt{\mathbb{E}[Z^2]} \sqrt{\mathbb{P}(\{Z > \theta \mathbb{E}[Z]\})}, \end{aligned} \quad (2.6)$$

porque entonces despejando $\frac{(1-\theta)\mathbb{E}[Z]}{\sqrt{\mathbb{E}[Z^2]}} \leq \sqrt{\mathbb{P}(\{Z > \theta \mathbb{E}[Z]\})}$ y elevándolo al cuadrado para completamos la demostración. Aclaremos que en este último paso usamos que Z es no

negativa para preservar la desigualdad al elevar al cuadrado. ■

[2.2.4] Proposición: Desigualdad de Bernstein - [Ver18, Teorema 2.8.4.]

Sea $(X_n)_{n \in \mathbb{N}}$ una familia de variables aleatorias independientes tales que $X_n \leq C$ casi seguramente para cada $n \in \mathbb{N}$. Luego, para cada $t > 0$ y para cada $n \in \mathbb{N}$ vale que

$$\mathbb{P} \left(\left\{ \left| \sum_{i=1}^n X_i - \mathbb{E} [X_i] \right| > t \right\} \right) \leq 2 \exp \left(-\frac{1}{2} \frac{t^2}{\text{Var} [\sum_{i=1}^n X_i] + \frac{C}{3}t} \right). \quad (2.7)$$

Mencionamos que para algunos resultados la desigualdad de Tchebyshev o el teorema central del límite serían suficientes, pero la desigualdad de Bernstein da cotas más finas que nos serán de utilidad cuando queremos asegurar que ocurren muchos eventos simultáneamente. Como esto ocurrirá frecuentemente a lo largo del trabajo, mencionamos además la siguiente observación:

[2.2.5] Observación: Contables eventos que ocurren con alta probabilidad, ocurren simultáneamente con alta probabilidad

Si I es un conjunto de índices contable y $(\mathcal{A}_i)_{i \in I}$ son eventos tales que $\mathbb{P}(\mathcal{A}_i) \geq 1 - o(f_i)$ para ciertas funciones $f_i : \mathbb{N} \rightarrow [0, 1]$ para cada $i \in I$, tendremos que

$$\mathbb{P}(\cap_{i \in I} \mathcal{A}_i) \geq 1 - o \left(\sum_{i \in I} f_i \right), \quad (2.8)$$

pues usando una cota por la unión de eventos para los complementos tendremos que

$$\mathbb{P}(\cap_{i \in I} \mathcal{A}_i) = 1 - \mathbb{P}(\cup_{i \in I} \mathcal{A}_i^C) \geq 1 - \sum_{i \in I} o(f_i) \geq 1 - o \left(\sum_{i \in I} f_i \right). \quad (2.9)$$

Análogamente, si $\mathbb{P}(\mathcal{A}_i) \geq 1 - \mathcal{O}(f_i)$ para cada $i \in I$, tendremos que

$$\mathbb{P}(\cap_{i \in I} \mathcal{A}_i) \geq 1 - \mathcal{O} \left(\sum_{i \in I} f_i \right). \quad (2.10)$$

Fijaremos también alguna notación que usaremos de ahora en más cuando trabajemos con vectores y matrices, tanto determinísticos como aleatorios:

[2.2.6] Notación: Notación para vectores y matrices

Para $n, m \in \mathbb{N}$, notaremos $\mathbf{0}_n$ al vector 0 en \mathbb{R}^n , $\mathbf{1}_n$ al vector con toda coordenada 1 en \mathbb{R}^n , \mathbf{I}_n a la **matriz identidad** en $\mathbb{R}^{n \times n}$, $\mathbf{J}_{n \times m}$ a la matriz con todos los coeficientes 1 en $\mathbb{R}^{n \times m}$, y e_1, \dots, e_n a los vectores de la **base canónica** de \mathbb{R}^n . También anotaremos $\mathbf{0}$, $\mathbf{1}$, \mathbf{I} y \mathbf{J} cuando las dimensiones estén claras en el contexto.

Notaremos Ker e Im para el **núcleo** y la **imágen** de una matriz respectivamente, y ρ para su **radio espectral**, el mayor módulo entre sus autovalores.

Notaremos $\|\bullet\|$ para la **norma euclídea** o norma 2 para operadores, y $\|\bullet\|_p$ para las

normas \mathbf{p} tanto para vectores como para operadores.

Además, fija M una matriz, escribiremos $M_{j,\bullet}$ para referirnos a su **j -ésima fila** y $M_{\bullet,j}$ para referirnos a su **j -ésima columna**, y $M^{(i)}$ para referirnos a la matriz a la que le dejamos como $\mathbf{0}$ la i -ésima fila y la i -ésima columna.

Notaremos también $\mathbb{E}[X]$ para la esperanza de una matriz aleatoria o un vector aleatorio, recordando que no es más que la matriz o el vector que se obtiene tomando esperanza coordinada a coordinada.

Ahora sí, podremos pasar a introducir nuestros primeros modelos de grafos aleatorios, que nos ocuparán el resto del capítulo.

2.3 El grafo de Erdős-Renyi

Como mencionamos en la introducción, un primer modelo de grafos aleatorios es el estudiado por Erdős y Renyi ya para 1959. En este grafo tenemos una cantidad de nodos fija, y las aristas se agregan de manera aleatoria e independiente entre cada par de nodos. En este trabajo consideraremos un modelo que se comporta de manera similar al modelo de Erdős-Renyi, y es así que incluimos esta sección para motivar las características que veremos a continuación.

Más formalmente:

[2.3.1] Definición: Grafo de Erdős-Renyi

Dado $n \in \mathbb{N}$, $p \in [0, 1]$ decimos que $G = (V, E) \sim \text{Erdős-Renyi}(n, p)$ cuando:

- G tiene n nodos, y
- si i, j son nodos distintos luego $\mathbb{1}_{\{\{i,j\} \in E\}} \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$.

Cuando trabajemos con familias de estos grafos, notaremos simplemente $G \sim \text{Erdős-Renyi}(n, p_n)$ y quedará implícito que cada modelo dependerá de la cantidad de nodos n considerada.

Cuando hay una cantidad de nodos suficientemente grande los grafos de Erdős-Renyi comienzan a mostrar ciertas características globales con probabilidad alta. Por ejemplo, cuando hay suficiente cantidad de aristas quedará determinada una única componente conexa grande, como enunciamos a continuación:

[2.3.2] Teorema: Existencia de una única componente conexa lineal - [Dur07, Teorema 2.3.2.]

Sea $G \sim \text{Erdős-Renyi}\left(n, \frac{\lambda}{n}\right)$ una familia de grafos aleatorios para cierto $\lambda \in \mathbb{R}$ no negativo y n suficientemente grandes tales que $\frac{\lambda}{n} \leq 1$. Luego, existe una constante β tal que G tiene una única componente conexa con más que $\beta \log(n)$ nodos a.a.s..

Más aún, esa será su componente conexa con máxima cantidad de elementos y tendrá

$(1 - \rho(\lambda))n$ nodos a.a.s., donde $\rho(\lambda)$ es la probabilidad de extinción de un proceso de ramificación con progenie Poisson (λ).

Solamente como comentario, los procesos de ramificación aparecen al estudiar el comportamiento local del grafo, que en vecindades de cada nodo fijo se parecerá a dichos procesos asintóticamente. También para explorar cada componente conexa se estudia qué ocurre al descubrir el grafo partiendo de algún nodo como un paseo al azar o como un proceso de contacto, y de ahí que aparezcan constantes asociadas a estos procesos en el enunciado.

Resulta que a medida que la densidad de aristas crece la componente conexa más grande va aumentando en tamaño y eventualmente se llega a que el grafo es conexo con alta probabilidad, en una transición de fase como enunciamos a continuación:

[2.3.3] Teorema: Transición de fase para la conexión en grafos de Erdős–Renyi - [Dur07, Teoremas 2.8.1., 2.8.2. y 2.8.3.]

Sea $G \sim \text{Erdős-Renyi} \left(n, a \frac{\log(n)}{n} \right)$. Luego,

- si $a > 1$, G es conexo a.a.s., y
- si $a < 1$, G tiene $\Theta(n^{1-a})$ nodos aislados a.a.s..

En particular, G es conexo a.a.s. si $a > 1$ y no es conexo a.a.s. si $a < 1$.

Más aún, si $a > \frac{1}{2}$, luego G consiste de una componente conexa de cardinal máximo y nodos aislados a.a.s.. Si para cierto $b \in \mathbb{R}$ no negativo vale que $G \sim \text{Erdős-Renyi} \left(n, \frac{\log(n)+b+o(1)}{n} \right)$ luego G tiene Poisson (b) nodos aislados a.a.s..

Podremos ahora entonces considerar una generalización de estos grafos que nos acompañará en adelante a lo largo del trabajo.

2.4 El modelo estocástico de bloques

En un grafo de Erdős-Renyi cada nodo es esencialmente idéntico en cuanto a su comportamiento. Para modelar un poco más de complejidad, se pueden agrupar los nodos en una cantidad k de clases que vamos a llamar comunidades, donde los nodos en cada clase tienen un comportamiento idéntico entre sí. Este será el modelo estocástico de bloques, que introduciremos a continuación:

[2.4.1] Definición: Modelo estocástico de bloques SBM

Decimos que $G_{\mathcal{C}}$ se distribuye como el **modelo estocástico de bloques** y notamos $G_{\mathcal{C}} \sim \text{SBM}(n, k, m_{\mathcal{C}}, p, q)$ cuando $G = (V, E)$ es un grafo aleatorio simple de n nodos y \mathcal{C} es un vector aleatorio que en cada coordenada toma valores en $\{1, \dots, k\}$ y cumplen que:

- cada nodo i pertenece a la **comunidad** $\mathcal{C}(i) \stackrel{\text{i.i.d.}}{\sim} m_{\mathcal{C}}$ donde $m_{\mathcal{C}}$ es una medida

de probabilidad soportada en el conjunto de comunidades $\{1, \dots, k\}$; y

- condicionando por el vector de comunidades \mathcal{C} , para cada par de nodos distintos i, j vale que

$$\mathbb{1}_{\{\{i,j\} \in E\}} \sim \begin{cases} \text{Bernoulli}(p) & \text{si } \mathcal{C}(i) = \mathcal{C}(j), \text{ y} \\ \text{Bernoulli}(q) & \text{si } \mathcal{C}(i) \neq \mathcal{C}(j), \end{cases} \quad (2.11)$$

donde hay independencia para aristas distintas.

Decimos que el modelo es **asortativo** cuando $p > q$.

Una vez más, cuando trabajemos con familias de estos grafos, notaremos simplemente $G_{\mathcal{C}} \sim \text{SBM}(n, k, m_{\mathcal{C}}, p_n, q_n)$ y quedará implícito que cada modelo dependerá de la cantidad de nodos n considerada.

Podemos entender la construcción de un grafo del modelo estocástico de bloques de la siguiente manera: dada una cantidad de nodos n , primero le asignamos una comunidad a cada nodo de acuerdo a la distribución $m_{\mathcal{C}}$ de manera independiente para cada nodo, y después para cada par de nodos distintos agregamos una arista de acuerdo a una variable Bernoulli cuyo parámetro depende de las comunidades sorteadas previamente, y de manera independiente para cada par de aristas distintas, como muestra la Figura 2.2.

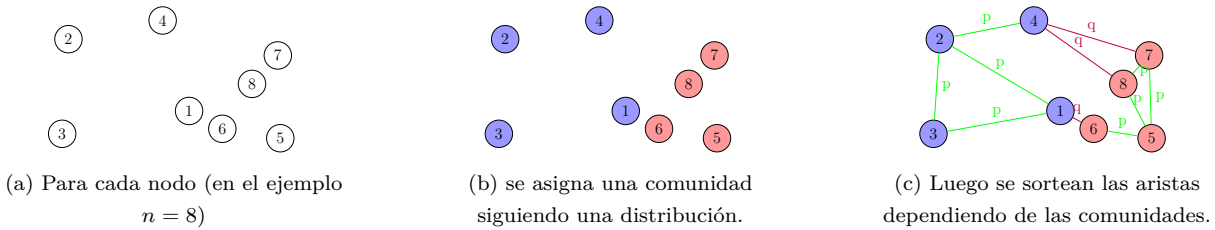


Figura 2.2: Construcción de un modelo SBM.

Agregamos algunas notaciones para simplificar la escritura de lo que sigue:

[2.4.2] **Notación:** Algunas notaciones para las comunidades

Dada una comunidad l , notamos

- $\mathcal{C}_l = \mathcal{C}^{-1}(\{l\})$ al **conjunto de nodos** que pertenecen a ella,
- $n_l = \#\mathcal{C}_l = |\mathcal{C}_l|$ a la **cantidad de nodos** que forman parte de ella, y
- $\alpha_l = \frac{n_l}{n}$ a la **proporción de nodos** que forman parte de ella.

Además notaremos $P_{\mathcal{C}} \in [0, 1]^{k \times k}$ a la **matriz de probabilidades de conexión** para la cual $(P_{\mathcal{C}})_{l_1, l_2}$ es la probabilidad de que haya una arista entre un nodo de comunidad l_1 y otro de comunidad l_2 . Observamos que luego vale que para cada par de nodos distintos i, j

$$A_{i,j} = \mathbb{1}_{\{\{i,j\} \in E\}} \sim \text{Bernoulli}\left((P_{\mathcal{C}})_{\mathcal{C}(i), \mathcal{C}(j)}\right), \quad (2.12)$$

donde A es la matriz de adyacencia del grafo, y para el modelo SBM la matriz está dada por $P_{\mathcal{C}} = (p - q)\mathbf{I}_k + q\mathbf{J}_{k \times k}$.

Reescribiendo la construcción gráfica del modelo SBM en lenguaje formal obtenemos las probabilidades puntuales del modelo de manera específica:

[2.4.3] Observación: Probabilidad puntual para modelos SBM

Si $G_{\mathcal{C}} \sim \text{SBM}(n, k, m_{\mathcal{C}}, p, q)$ luego para cada grafo $\tilde{G} = (\tilde{V}, \tilde{E})$, al cual pensamos partido en comunidades $\tilde{\mathcal{C}}$, y con \tilde{n}_l nodos de comunidad l para cada $1 \leq l \leq k$, y con \tilde{n}_{l_1, l_2} aristas entre nodos de las comunidades l_1 y l_2 , vale que:

$$p_{G_{\mathcal{C}}}(\tilde{G}_{\tilde{\mathcal{C}}}) = \left(\prod_{l=1}^k (p_{m_{\mathcal{C}}}(l))^{\tilde{n}_l} \right) \left(\prod_{l_1=1}^k \prod_{l_2=1}^{l_1-1} q^{\tilde{n}_{l_1, l_2}} (1 - q)^{\tilde{n}_{l_1} \tilde{n}_{l_2} - \tilde{n}_{l_1, l_2}} \right) \quad (2.13)$$

$$\left(\prod_{l=1}^k p^{\tilde{n}_{l, l}} (1 - p)^{\frac{\tilde{n}_l^2 - \tilde{n}_l}{2} - \tilde{n}_{l, l}} \right). \quad (2.14)$$

En particular, la probabilidad puntual para modelos SBM solo depende de las cantidades de nodos en cada comunidad y de la cantidad de aristas dentro de la misma comunidad y entre comunidades.

En efecto, podemos llegar a la expresión primero condicionando a que las comunidades estén dadas por una configuración $\hat{\mathcal{C}}$ en específico

$$\mathbb{P}(\{G_{\mathcal{C}} = \tilde{G}_{\tilde{\mathcal{C}}}\}) = \mathbb{P}(\{G = \tilde{G}\} | \{\mathcal{C} = \tilde{\mathcal{C}}\}) \mathbb{P}(\{\mathcal{C} = \tilde{\mathcal{C}}\}); \quad (2.15)$$

a continuación ordenando las aristas por las comunidades de los nodos que conectan y usando independencia para obtener

$$\mathbb{P}(\{G = \tilde{G}\} | \{\mathcal{C} = \tilde{\mathcal{C}}\}) = \left(\prod_{l_1=1}^k \prod_{l_2=1}^{l_1-1} q^{\tilde{n}_{l_1, l_2}} (1 - q)^{\tilde{n}_{l_1} \tilde{n}_{l_2} - \tilde{n}_{l_1, l_2}} \right) \quad (2.16)$$

$$\left(\prod_{l=1}^k p^{\tilde{n}_{l, l}} (1 - p)^{\frac{\tilde{n}_l^2 - \tilde{n}_l}{2} - \tilde{n}_{l, l}} \right);$$

y por último usando que la independencia en la elección de comunidad para cada nodo nos da que $\mathbb{P}(\{\mathcal{C} = \tilde{\mathcal{C}}\}) = \prod_{i=1}^k (p_{m_{\mathcal{C}}}(i))^{\tilde{n}_i}$.

Una primera observación que podemos hacer del modelo es que los tamaños de las comunidades podrían ir variando, pero (casi) nunca van a ser demasiado chicos. Esto será de importancia más adelante en el trabajo, así que lo formalizaremos en el siguiente resultado:

[2.4.4] Observación: Las comunidades son de tamaño lineal con alta probabilidad

Si $G_{\mathcal{C}} \sim \text{SBM}(n, k, m_{\mathcal{C}}, p_n, q_n)$ es una familia de grafos donde k y $m_{\mathcal{C}}$ no varían con

n , entonces para cada $\varepsilon > 0$ fijo valdrá que:

$$\mathbb{P} \left(\left\{ n_l = p_{m_c}(l)n + \mathcal{O} \left(n^{\frac{1+\varepsilon}{2}} \right) \text{ para cada } 1 \leq l \leq k \right\} \right) \geq 1 - o(1). \quad (2.17)$$

En particular, $n_l \asymp n$ para cada $1 \leq l \leq k$.

En efecto, podremos demostrarlo usando la desigualdad de Bernstein. En primer lugar, podemos reducirnos a mostrar que la cota vale para cada comunidad l individualmente por una cota por la unión de eventos, porque al ser finitas comunidades podremos recuperar el decaimiento que queremos:

$$\begin{aligned} & \mathbb{P} \left(\left\{ |n_l - np_{m_c}(l)| \leq n^{\frac{1+\varepsilon}{2}} \text{ para cada } 1 \leq l \leq k \right\} \right) \\ & \geq 1 - \sum_{l=1}^k \mathbb{P} \left(\left\{ |n_l - np_{m_c}(l)| \leq n^{\frac{1+\varepsilon}{2}} \right\} \right) = 1 - \mathcal{O}(k \exp(-Cn^\varepsilon)). \end{aligned} \quad (2.18)$$

Restará entonces mostrar que vale la cota para cada comunidad $1 \leq l \leq k$, lo cual es una aplicación inmediata de la desigualdad de Bernstein 2.2.4:

$$\begin{aligned} \mathbb{P} \left(\left\{ |n_l - np_{m_c}(l)| > n^{\frac{1+\varepsilon}{2}} \right\} \right) & \leq 2 \exp \left(-\frac{1}{2} \frac{\left(n^{\frac{1+\varepsilon}{2}} \right)^2}{np_{m_c}(l)(1-p_{m_c}(l)) + \frac{1}{3}n^{\frac{1+\varepsilon}{2}}} \right) \\ & = o(1). \end{aligned} \quad (2.19)$$

Algunos casos particulares del modelo de bloques estocástico son el modelo simétrico y el modelo balanceado, que nos servirán como primeros casos de estudio para los algoritmos a ser presentados.

[2.4.5] **Definición: Modelo estocástico de bloques simétrico SSBM**

Decimos que $G \sim \text{SBM}(n, k, m_c, p, q)$ es un **modelo estocástico de bloques simétrico** cuando m_c es la distribución uniforme en las comunidades, y notaremos $G \sim \text{SSBM}(n, k, p, q)$.

[2.4.6] **Definición: Modelo estocástico de bloques con tamaño de comunidades fijas y el modelo balanceado**

Llamaremos **modelo estocástico de bloques simétrico balanceado** a $G \sim \text{SBM}(n, k, m_c, p, q)$ condicionado al evento en el cual todas las comunidades son del mismo tamaño $\left\{ n_l = \frac{n}{k} \text{ para cada } 1 \leq l \leq k \right\}$, donde $k|n$. Notaremos $G \sim \text{BSBM}(n, k, p, q)$ en este caso.

Más generalmente, notaremos $G \sim \text{SBM}(n, k, (\tilde{n}_l)_{1 \leq l \leq k}, p, q)$ al modelo condicionado para que cada comunidad tenga un tamaño fijo $\left\{ n_l = \tilde{n}_l \text{ para cada } 1 \leq l \leq k \right\}$.

Condicionar al tamaño de las comunidades tiene la ventaja de que podemos simplificar la notación aún más, como veremos a continuación.

[2.4.7] Observación: Condicionar para fijar las comunidades de cada nodo

Como en el modelo balanceado, va a ser útil trabajar condicionando a que los nodos pertenezcan a comunidades fijas \mathcal{C} . Como renombrar los nodos (y las aristas afectadas) no afecta las probabilidades puntuales del modelo (podemos verificarlo a partir de las probabilidades puntuales halladas en la Observación 2.4.3) notamos que cuando condicionamos por

$$\{n_{\bullet} = \tilde{n}_{\bullet}\} = \{ \text{En } \mathcal{C} \text{ hay } \tilde{n}_l \text{ nodos con comunidad } i \text{ para cada } 1 \leq l \leq k \}, \quad (2.20)$$

a cada asignación de las comunidades válida le corresponde la misma probabilidad. En ese caso supondremos sin pérdida de generalidad que

$$\mathcal{C}(i) = l \text{ si y sólo si } \sum_{l'=0}^{l-1} n_{l'} < i \leq \sum_{l'=0}^l n_{l'}, \quad (2.21)$$

es decir los primeros n_1 nodos pertenecerán a la comunidad 1, los siguientes n_2 a la comunidad 2, y así sucesivamente hasta que los últimos n_k nodos corresponden a la comunidad k .

En particular, estudiar esta configuración privilegiada para el vector de comunidades nos determinará el comportamiento del modelo estocástico de bloques balanceado.

Como habíamos adelantado previamente, el modelo estocástico de bloques puede pensarse como una generalización de los grafos de Erdős-Renyi que nos permite diferenciar el comportamiento para clases de nodos. Es así que el comportamiento de este grafo va a seguir siendo similar al del grafo de Erdős-Renyi, al menos en cuanto a conexión se refiere, reemplazando el coeficiente a de densidad de aristas en el Teorema 2.3.2 por un coeficiente promedio de los nodos entre distintas comunidades, y obtendremos de manera análoga una transición de fase. Daremos distintas versiones, correspondientes a cada variante de los modelos estocásticos de bloques: presentaremos primero una para el modelo balanceado 2.4.6, después pasaremos a otra para tamaños de comunidades fijos, y terminaremos por el modelo original presentado en 2.4.1.

[2.4.8] Teorema: Transición de fase para la conexión en modelos BSBM

Sea $G_{\mathcal{C}} \sim \text{BSBM} \left(n, k, a \frac{\log(n)}{n}, b \frac{\log(n)}{n} \right)$. Luego,

- si $\frac{a+(k-1)b}{k} > 1$, G es conexo a.a.s., y
- si $\frac{a+(k-1)b}{k} < 1$, G tiene $\Theta \left(n^{1-\frac{a+(k-1)b}{k}} \right)$ nodos aislados a.a.s. para cada comunidad.

En particular, G es conexo a.a.s. si $\frac{a+(k-1)b}{k} > 1$ y no es conexo a.a.s. si $\frac{a+(k-1)b}{k} < 1$.

[2.4.9] Teorema: Transición de fase para la conexión en modelos SBM con tamaños de comunidades fijos

Sea $G_{\mathcal{C}} \sim \text{SBM}\left(n, k, (n_l)_{1 \leq l \leq k}, a \frac{\log(n)}{n}, b \frac{\log(n)}{n}\right)$ una familia de modelos estocásticos de bloques. Luego, notemos $\lambda_n = \min_{1 \leq l \leq k} \left\{ \frac{n-n_l}{n} b + \frac{n_l}{n} a \right\}$, y tendremos que

- si $\liminf_{n \rightarrow \infty} \lambda_n > 1$, G es conexo a.a.s., y
- si $\liminf_{n \rightarrow \infty} \lambda_n < 1$, G tiene $\Theta\left(n^{1-\liminf_{n \rightarrow \infty} \lambda_n}\right)$ nodos aislados a.a.s. para cada comunidad l que verifique que $\liminf_{n \rightarrow \infty} \lambda_{n,l} < 1$, donde $\lambda_{n,l} = \frac{n-n_l}{n} b + \frac{n_l}{n} a$.

En particular, G es conexo a.a.s. si $\liminf_{n \rightarrow \infty} \lambda_n > 1$ y no es conexo a.a.s. si $\liminf_{n \rightarrow \infty} \lambda_n < 1$.

[2.4.10] Teorema: Transición de fase para la conexión en modelos SBM

Sea $G_{\mathcal{C}} \sim \text{SBM}\left(n, k, m_{\mathcal{C}}, a \frac{\log(n)}{n}, b \frac{\log(n)}{n}\right)$ una familia de modelos estocásticos de bloques. Luego, notemos $\lambda = \min_{1 \leq l \leq k} \left\{ \left(1 - \frac{p_{m_{\mathcal{C}}}(l)}{\log(n)}\right) b + \frac{p_{m_{\mathcal{C}}}(l)}{\log(n)} a \right\}$, y tendremos que

- si $\lambda > 1$, G es conexo a.a.s., y
- si $\lambda < 1$, G tiene $\Theta\left(n^{1-\lambda}\right)$ nodos aislados a.a.s. para cada comunidad l que verifique que

$$\lambda_l = \liminf_{n \rightarrow \infty} \left(1 - \frac{p_{m_{\mathcal{C}}}(l)}{\log(n)}\right) b + \frac{p_{m_{\mathcal{C}}}(l)}{\log(n)} a < 1$$

En particular, G es conexo a.a.s. si $\lambda > 1$ y no es conexo a.a.s. si $\lambda < 1$.

Estos resultados aparecen mencionados en distintos lugares de la literatura, como por ejemplo en [Abb18], pero no encontramos una demostración escrita como para referir a ella. Es así que en el anexo A.1 realizamos las demostraciones pertinentes, que obtuvimos adaptando las ideas que se usan para demostrar estas mismas propiedades para los grafos de Erdős-Renyi en el libro [Dur07] de Durrett. Solamente a modo de comentario, y en resumidas palabras, la demostración de conexión consiste en determinar que cada nodo está conectado a una cantidad grande de nodos con probabilidad alta, en cuyo caso que dos de esos cúmulos no se conecten ocurre con probabilidad muy baja; y la otra implicación podemos atacarla usando desigualdades de concentración para entender la cantidad de nodos aislados para cada comunidad.

Como herramienta teórica clave en estos argumentos está el entender el grado de cada nodo en estos modelos. Como además nos será de utilidad en futuros capítulos, presentaremos el siguiente resultado, que nos indica que al ser las probabilidades de conexión todas del orden $\frac{\log(n)}{n}$, éstos se conectarán a $\log(n)$ nodos.

[2.4.11] Proposición: Los grados se concentran en el grado esperado

Sea $G_{\mathcal{C}} \sim \text{SBM}(n, k, (n_l)_{1 \leq l \leq k}, p_n, q_n)$ una familia de modelos estocásticos de bloques donde $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para ciertos reales positivos $a > b$.

Llamemos $\lambda_n = \min_{1 \leq l \leq k} \left\{ \frac{n-n_l}{n} b + \frac{n_l}{n} a \right\}$, y supongamos que $\liminf_{n \rightarrow \infty} \lambda_n > 1$.

Luego, vale que existen $0 < x < 1 < y$ números reales tales que para cada nodo i se verifica que

$$\mathbb{P} \left(\left\{ x \deg_n(i) \leq \widehat{\deg}_n(i) \leq y \deg_n(i) \right\} \right) \geq 1 - o(n^{-1}), \quad (2.22)$$

donde notamos por $\widehat{\deg}_n(i)$ al grado de i en G y a $\deg_n(i) = \mathbb{E}[\widehat{\deg}_n(i)]$ su grado esperado.

En particular, vale que

$$\mathbb{P} \left(\left\{ x \deg_n(i) \leq \widehat{\deg}_n(i) \leq y \deg_n(i) \text{ para cada nodo } i \in V \right\} \right) \geq 1 - o(1) \quad (2.23)$$

y $\widehat{\deg}_n(i) = \Theta(\log(n))$ para cada $i \in V$ con alta probabilidad.

Observamos que en el modelo balanceado la condición sobre λ_n equivale a pedir que $\frac{a+(k-1)b}{k} > 1$.

Desde ya, esto es lo mismo que podríamos esperar para nodos en un grafo de Erdős-Renyi con la misma densidad de aristas: aquí la única dificultad adicional es técnica, y consiste en que las probabilidades varían dependiendo de las comunidades de los nodos. Para ayudarnos a salvar este obstáculo, emplearemos la siguiente desigualdad de concentración:

[2.4.12] Lema: Desigualdad de concentración para la conectividad de modelos SBM

Sea $Z = X_1 + \dots + X_t$, donde X_s es suma de binomiales independientes $X_s \sim \sum_{l=1}^k \text{Binomial}(\lceil (1-\delta)n_l \rceil, p_{s,l})$ para probabilidades $p_{s,l}$, $1 \leq s \leq t$, $1 \leq l \leq k$.

Luego, vale que si $\gamma(x) = x \log(x) - x + 1$, tenemos que:

1. para cada $0 < x < 1$ vale que

$$\mathbb{P}(\{Z \leq x\mu\}) \leq e^{-\gamma(x)\mu}, \quad (2.24)$$

2. y para cada $1 < y$ vale que

$$\mathbb{P}(\{Z \geq y\mu\}) \leq e^{-\gamma(y)\mu}. \quad (2.25)$$

Vamos a remitir al anexo A.1 para la demostración de este lema, y en su lugar lo usaremos para demostrar que los grados suelen tener el orden esperado como enunciamos en 2.4.11.

Demostración. Basta que demostremos la desigualdad (2.22), porque luego en el espíritu de la Observación 2.2.5 podemos usar una cota por la unión de eventos para obtener la segunda desigualdad (2.23).

Es así que usaremos dos veces el Lema 2.4.12 para acotar el grado por arriba y por abajo en orden por el grado esperado.

Primero mostraremos la cota inferior para los grados empíricos. Como $\liminf_{n \rightarrow \infty} \lambda_n > 1$ y $\lim_{x \rightarrow 0^+} \gamma(x) = 1$, tomando $\eta = \frac{1}{2} (\liminf_{n \rightarrow \infty} \lambda_n - 1)$ tendremos que existe $0 < x < 1$ tal que $\gamma(x) (\liminf_{n \rightarrow \infty} \lambda_n - \eta) > 1$. En particular, usando el lema con $\delta = 0$ y para este x obtenemos que para cada nodo i fijo vale que para n suficientemente grandes se verifica que

$$\mathbb{P} \left(\left\{ \widehat{\deg}_n(i) \leq x \deg_n(i) \right\} \right) \leq e^{-\gamma(x) \deg_n(i)} \leq e^{-\gamma(x) (\liminf_{n \rightarrow \infty} \lambda_n - \eta + \mathcal{O}(\frac{1}{n})) \log(n)} = o(n^{-1}). \quad (2.26)$$

que muestra lo que queríamos ver, donde usamos que para n suficientemente grandes valdrá que $\deg_n(i) = (n - n_{\mathcal{C}(i)}) \left(b \frac{\log(n)}{n}\right) + (n_{\mathcal{C}(i)} - 1) \left(a \frac{\log(n)}{n}\right) \geq (\liminf_{n \rightarrow \infty} \lambda_n) \log(n) + \mathcal{O}\left(\frac{1}{n}\right) \log(n)$ para la segunda desigualdad.

Ahora podemos proceder análogamente para mostrar y acotar el grado empírico por arriba, porque como $\lim_{y \rightarrow \infty} \gamma(y) = \infty$ tenemos que podremos hallar $1 < y$ que verifique que $\gamma(y) (\liminf_{n \rightarrow \infty} \lambda_n - \eta) > 1$, con lo que para cada nodo i fijo y n suficientemente grandes tendremos que

$$\mathbb{P} \left(\left\{ \widehat{\deg}_n(i) \geq y \deg_n(i) \right\} \right) \leq e^{-\gamma(y) \deg_n(i)} \leq e^{-\gamma(y) (\liminf_{n \rightarrow \infty} \lambda_n - \eta + \mathcal{O}(\frac{1}{n})) \log(n)} = o(n^{-1}). \quad (2.27)$$

Solo restará combinar ambos resultados como en la Observación 2.2.5 para ver que además ambas cotas ocurren simultáneamente con alta probabilidad, y obtendremos (2.22) como queríamos.

Por último, como $\deg_n(i) = (n - n_{\mathcal{C}(i)}) \left(b \frac{\log(n)}{n}\right) + (n_{\mathcal{C}(i)} - 1) \left(a \frac{\log(n)}{n}\right)$ tal cual calculamos anteriormente, obtenemos que $\deg_n = \Theta(\log(n))$. Completamos así la demostración, observando además que se logra en realidad un orden de decaimiento $1 - o(n^{-\xi})$ para $\xi > 0$ suficientemente pequeño al tomar el mínimo coeficiente para $-\log(n)$ en las cotas (2.26) y (2.27). ■

Más allá de que entendamos lo que ocurre con los grados de los nodos, nos interesa entender cualitativamente si logramos el objetivo de que los nodos de la misma comunidad se conectan entre sí más frecuentemente que a nodos de otras comunidades en general. El siguiente resultado nos indica que cuando la densidad de aristas es suficientemente alta esto efectivamente ocurre para todos los nodos del grafo en simultáneo:

[2.4.13] Teorema: Cada nodo tiene una mayoría de vecinos de su misma comunidad - [MNS16, Proposiciones 2.7 y 2.8]

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos tales que $p_n > q_n$ para cada $n \in \mathbb{N}$. Luego, tenemos que

- con alta probabilidad cada nodo salvo $o(n)$ tiene entre sus vecinos una mayoría estricta de su misma comunidad si y sólo si $P(n, p_n, q_n) = 1 - o(1)$, y
- con alta probabilidad cada nodo tiene entre sus vecinos una mayoría estricta de

su misma comunidad si y sólo si $P(n, p_n, q_n) = 1 - o(n^{-1})$

para $P(n, p_n, q_n)$ la probabilidad en $G_{\mathcal{C}}$ de que un nodo fijo tenga entre sus vecinos una mayoría estricta de nodos de su comunidad

$$P(n, p_n, q_n) = \mathbb{P}(\{\#(N(i) \cap \mathcal{C}_l) \text{ se maximice estrictamente en } l = \mathcal{C}(i)\}). \quad (2.28)$$

Más aún, cuando $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para a y b reales no negativos con $a > b$ valen además que $P(n, p_n, q_n) = 1 - o(1)$ y que

$$P(n, p_n, q_n) = 1 - o(n^{-1}) \text{ si y sólo si } \frac{a+b}{2} > 1 + \sqrt{ab}. \quad (2.29)$$

Este resultado indica que un estudio detallado de las aristas en el grafo nos dará información acerca de las comunidades de cada nodo, pues en general esperaríamos que nodos muy interconectados entre sí pertenezcan a las mismas comunidades. En el siguiente capítulo exploraremos más precisamente como podemos explotar estas características del grafo para encontrarlas.

2.5 Notas

[2.5.1] **Nota:** Modelos estocásticos de bloques más generales en la literatura.

El modelo estocástico de bloques puede ser definido de forma más general para que las comunidades no tengan todas comportamientos idénticos entre sí, permitiendo matrices $P_{\mathcal{C}}$ que toman más que dos valores. Este enfoque es usual en la literatura, y por ejemplo puede encontrarse en [Abb18]. En particular, es posible estudiar los fenómenos que estaremos trabajando en grafos no asortativos ($q > p$). Por ejemplo, la versión original del Teorema 2.4.13 vale en el caso no asortativo, y se puede encontrar en [MNS16].

[2.5.2] **Nota:** Simulaciones para modelos estocásticos de bloques

En el Capítulo 6 realizaremos algunas simulaciones con estos grafos aleatorios y proveeremos acceso al código desarrollado, pero vale la pena mencionar como recurso [este sitio](#), recomendado en [vdH17], en el cual se puede acceder a simulaciones interactivas del grafo de Erdős-Renyi, para ganar intuición.

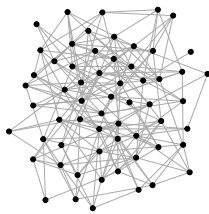
3. Detección de comunidades en el modelo estocástico de bloques

3.1 El problema de detección de comunidades

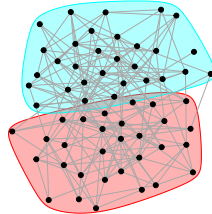
Introducido el modelo estocástico de bloques, el problema que nos ocupará en el resto del trabajo es el siguiente:

[3.1.1] Problema: El problema de detección de comunidades

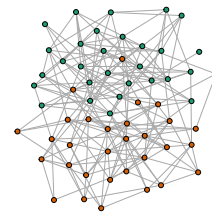
Dado $G_C \sim \text{SSBM}(n, k, p, q)$ un grafo, queremos hallar una predicción \hat{C} de las comunidades de cada nodo que sea precisa.



(a) A partir de un grafo y sus conexiones



(b) realizamos una estimación de las comunidades



(c) que queremos que sea precisa al comparar contra las comunidades reales.

Figura 3.1: El problema de detección de comunidades en el modelo SBM.

Desde ya, en general será imposible predecir las comunidades de cada nodo acertando en cada uno para cada grafo fijo.

Una primera razón es que con mayor o menor probabilidad es posible obtener un mismo grafo para distintas asignaciones de comunidades, con lo cual cualquier estimación fija va a cometer errores en este modelo. Es así que necesitaremos definir criterios de precisión asintóticos, que estudian lo que ocurre cuando n tiende a infinito para una familia de grafos: al contar con muchos nodos podremos empezar a notar las propiedades del modelo que exploramos en el Capítulo 2 y explotarnos para predecir las comunidades de los nodos.

Por otro lado, necesitaremos definir criterios de precisión que contemplen a las comunidades como particiones del conjunto de nodos más que como vectores como vinimos haciendo hasta el momento, porque lo interesante en el modelo no es el nombre que tiene cada comunidad si no que es hallar los nodos que forman parte de cada una. Esto conlleva a la

definición de coincidencia, que nos independiza de los nombres que le hayamos puesto a cada comunidad en nuestra estimación, y que usaremos para definir los criterios de consistencia débil y consistencia fuerte que estudiaremos en lo que sigue en el trabajo:

[3.1.2] Definición: Consistencia fuerte y consistencia débil

Definimos la **coincidencia** entre dos asignaciones a k comunidades para n nodos \mathcal{C} y $\tilde{\mathcal{C}}$ como

$$\text{coincidencia}(\mathcal{C}, \tilde{\mathcal{C}}) = \max_{\sigma \in S_k} \left\{ \frac{1}{n} \cdot \# \{1 \leq i \leq n \mid \mathcal{C}(i) = \sigma(\tilde{\mathcal{C}}(i))\} \right\} \quad (3.1)$$

donde S_k son las permutaciones de k elementos.

Dada $G_{\mathcal{C}} \sim \text{SBM}(n, k, m_{\mathcal{C}}, p_n, q_n)$ una familia de modelos diremos que una familia de estimaciones $\hat{\mathcal{C}}$ es **fuertemente consistente** si

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\{ \text{coincidencia}(\mathcal{C}, \hat{\mathcal{C}}) = n \} \right) = 1, \quad (3.2)$$

y diremos que es **débilmente consistente** si

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\{ \text{coincidencia}(\mathcal{C}, \hat{\mathcal{C}}) = n - o(n) \} \right) = 1. \quad (3.3)$$

Desde ya, el ideal será poder realizar estimaciones que sean fuertemente consistentes. Sin embargo, no siempre podremos hacerlo: como anticipan los resultados 2.4.8 y 2.4.13, los grafos van a presentar características distintas dependiendo de la densidad de aristas dentro y entre comunidades, y éstas van a tener un impacto grande en las estimaciones que podremos lograr. Por ejemplo, cuando $p_n = q_n$ y hay la misma probabilidad de agregar aristas entre comunidades o dentro de cada comunidad no será posible recuperarlas; de hecho, en este caso el modelo se reduce a un grafo de Erdős-Renyi y todos los nodos se comportarán de manera idéntica. Por otro lado, un primer resultado en esta dirección es que necesitamos grafos conexos como para captar suficiente información relativa a las comunidades, como precisamos a continuación:

[3.1.3] Proposición: Consistencia es imposible sin conexión.

Sea $((G_{\mathcal{C}})_n)_{n \in \mathbb{N}}$ una familia de grafos tales que $(G_{\mathcal{C}})_n \sim \text{BSBM} \left(n, k, p_n = a \frac{\log(n)}{n}, q_n = b \frac{\log(n)}{n} \right)$ con a, b reales positivos tales que $\frac{a+b}{2} < 1$.

Entonces no existen estimadores fuertemente consistentes para las comunidades.

Demostración. Supongamos que $\hat{\mathcal{C}}$ es alguna estimación de \mathcal{C} , y veamos que con probabilidad alta no podrá recuperar exactamente las comunidades. Nuestra estrategia consistirá en usar que además de no ser conexo los grafos en este caso tienen una cantidad alta de nodos aislados con alta probabilidad, como muestra el Teorema 2.4.8, donde para estos nodos no es evidente a qué comunidad podrían haber pertenecido.

Para empezar, vamos a suponer sin pérdida de generalidad que, en caso de que G tenga dos nodos aislados o más, $\hat{\mathcal{C}}$ asigna a dos comunidades distintas algunos nodos aislados casi

seguramente. En efecto, si \mathcal{C} es el conjunto de grafos que tienen nodos aislados para los cuales $\hat{\mathcal{C}}$ les asigna a todos la misma comunidad,

$$\mathbb{P} \left(\left\{ \text{coincidencia} (\mathcal{C}, \hat{\mathcal{C}}) = n, G \in \mathcal{C} \right\} \right) \quad (3.4)$$

$$\leq \mathbb{P} \left(\bigcup_{G \in \mathcal{C}} \bigcup_{l=1}^k \left\{ \mathcal{C}_l \text{ no tiene nodos aislados en } G \right\} \right) \quad (3.5)$$

$$\leq \mathbb{P} \left(\bigcup_{l=1}^k \left\{ \mathcal{C}_l \text{ no tiene nodos aislados} \right\} \right) \xrightarrow{n \rightarrow \infty} 0$$

por el Teorema 2.4.8, con lo cual si $\mathbb{P}(\{G \in \mathcal{C}\}) > 0$ es imposible que $\hat{\mathcal{C}}$ sea fuertemente consistente.

Luego, llamemos

$$\mathcal{A} = \{G \text{ grafo de } n \text{ nodos} \mid G \text{ tiene 2 o más nodos aislados}\}, \quad (3.6)$$

y si $\tilde{G} \in \mathcal{A}$ entonces valdrá que no será posible recuperar las comunidades para \tilde{G} con probabilidad alta. Más concretamente, podremos ver que para cada estimación $\hat{\mathcal{C}}$,

$$\mathbb{P} \left(\left\{ \mathcal{C} = \hat{\mathcal{C}} \right\} \mid \left\{ G = \tilde{G} \right\} \right) \leq \frac{1}{2}, \quad (3.7)$$

y así utilizando probabilidad total sobre los grafos en \mathcal{A} y además empleando el Teorema 2.4.8 para los que no tendremos que

$$\begin{aligned} & \mathbb{P} \left(\left\{ \text{coincidencia} (\mathcal{C}, \hat{\mathcal{C}}) = n \right\} \right) \quad (3.8) \\ &= \sum_{\tilde{G} \in \mathcal{A}} \mathbb{P} \left(\left\{ \text{coincidencia} (\mathcal{C}, \hat{\mathcal{C}}) = n \right\} \mid \left\{ G = \tilde{G} \right\} \right) \mathbb{P} \left(\left\{ G = \tilde{G} \right\} \right) \\ & \quad + \mathbb{P} \left(\left\{ \text{coincidencia} (\mathcal{C}, \hat{\mathcal{C}}) = n \right\} \mid \left\{ G \notin \mathcal{A} \right\} \right) \mathbb{P} (G \notin \mathcal{A}) \\ &\leq \sum_{\tilde{G} \in \mathcal{A}} \frac{1}{2} \mathbb{P} \left(\left\{ G = \tilde{G} \right\} \right) + \mathbb{P} \left(\left\{ G \notin \mathcal{A} \right\} \right) \\ &= \frac{1}{2} \mathbb{P} \left(\left\{ G \in \mathcal{A} \right\} \right) + \mathbb{P} \left(\left\{ G \notin \mathcal{A} \right\} \right) \\ &\leq \frac{1}{2} + \mathbb{P} \left(\left\{ G \notin \mathcal{A} \right\} \right) \xrightarrow{n \rightarrow \infty} \frac{1}{2} < 1, \end{aligned}$$

lo cual muestra que será imposible que las estimaciones $\hat{\mathcal{C}}$ sean fuertemente consistentes.

Mostremos luego la afirmación (3.7). Para eso, en primer lugar notemos para cada asignación de comunidades $\tilde{\mathcal{C}}$ fija como $\mathcal{B}_{\tilde{\mathcal{C}}}$ al conjunto de asignaciones $\tilde{\mathcal{C}}'$ sobre \tilde{G} que tienen la misma cantidad de nodos para cada comunidad y la misma cantidad de aristas entre cada combinación de comunidades. Luego, notemos que $\mathbb{P}(\{\mathcal{C} = \bullet\} \mid G = \tilde{G})$ es constante sobre $\mathcal{B}_{\tilde{\mathcal{C}}}$; podemos deducir esto de que $\mathbb{P}(\{G = \tilde{G}, \mathcal{C} = \tilde{\mathcal{C}}\}) = \mathbb{P}(\{G = \tilde{G}, \mathcal{C} = \tilde{\mathcal{C}}'\})$ usando la fórmula de probabilidades puntuales hallada en la Observación 2.4.3 y que $\tilde{\mathcal{C}}$ y $\tilde{\mathcal{C}}'$ viven en $\mathcal{B}_{\tilde{\mathcal{C}}}$. Por último, notemos que cuando \tilde{G} tiene por lo menos dos nodos aislados, las asignaciones de comunidades que mandan algún nodo aislado a cada comunidad tienen probabilidad menor o igual que $\frac{1}{2}$, pues si $\tilde{\mathcal{C}}$ es una de ellas vale que la asignación $\tilde{\mathcal{C}}'$ que intercambia la comunidad para ese par de nodos aislados asignados a comunidades distintas cumple que

$\tilde{\mathcal{C}}' \in \mathcal{B}_{\tilde{\mathcal{C}}}$ y entonces

$$\begin{aligned} \mathbb{P}(\{\mathcal{C} = \tilde{\mathcal{C}}\} | \{G = \tilde{G}\}) &= \frac{1}{2} [\mathbb{P}(\{\mathcal{C} = \tilde{\mathcal{C}}\} | \{G = \tilde{G}\}) \\ &\quad + \mathbb{P}(\{\mathcal{C} = \tilde{\mathcal{C}}'\} | \{G = \tilde{G}\})] \\ &\leq \frac{1}{2}. \end{aligned} \tag{3.9}$$

■

Como nuestro objetivo va a ser encontrar estimadores fuertemente consistentes, de ahora en más siempre trabajaremos por encima de esta densidad de aristas que nos garantiza la conexión en el modelo con probabilidad alta. Afortunadamente, resulta que podremos encontrar algoritmos eficientes que construyan predicciones fuertemente consistentes de las comunidades cuando la densidad de aristas es suficientemente alta. Este fenómeno se dará en una transición de fase, como ocurrió con la conexión de los grafos en los Teoremas 2.3.3 y 2.4.8, y podemos caracterizarlo de la siguiente manera:

[3.1.4] Teorema: Factibilidad de consistencia fuerte en modelos BSBM - [Abb18, Teorema 5.] o [MNS16, Teorema 2.5.]

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos estocásticos de bloques, donde $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para $a > b > 0$ números reales positivos. Luego vale que

- si $\frac{a+b}{2} > 1 + \sqrt{ab}$ entonces existe algún algoritmo fuertemente consistente eficiente para resolver el problema de detección de comunidades, y
- si $\frac{a+b}{2} < 1 + \sqrt{ab}$ entonces no existe ningún algoritmo fuertemente consistente que resuelva el problema de detección de comunidades.

Notamos que las hipótesis que requiere este teorema son un poco más fuertes que las hipótesis requeridas para tener conexión en el modelo. Una primera manera en que se vuelven más fuertes es en que exigen que haya cierta diferencia entre p_n y q_n , con lo cual distinguen al modelo estocástico de bloques del grafo de Erdős-Renyi: en efecto, como $\frac{a+b}{2} > 1 + \sqrt{ab}$ si y sólo si $(\sqrt{a} - \sqrt{b})^2 > 2$ en ese caso a y b deben estar suficientemente separados, con lo que p_n y q_n también. Más allá de la heurística, estas hipótesis nos permiten tener garantías más fuertes sobre las propiedades del grafo, y no casualmente la transición entre posibilidad e imposibilidad de recuperar las comunidades con consistencia fuerte coincide con la transición entre que cada nodo tenga la mayoría de sus vecinos de su propia comunidad o no, que estudiamos en el Teorema 2.4.13. Este estudio más minucioso de cómo las propiedades del grafo afectan a los estimadores permitirá demostrar esta transición de fase y motivará algunos enfoques a lo largo del trabajo, por lo que le dedicaremos las próximas secciones.

3.2 Estimadores maximales y límites de la estimación

En primer lugar, exploraremos qué pasa debajo del umbral en el cual no habrán estimadores fuertemente consistentes y daremos una idea de la demostración en ese caso. A grandes rasgos, el problema estará en que la asignación real de las comunidades no será la más probable a partir de la información que conocemos, lo cual causará que incluso los mejores estimadores deban fallar en la predicción de la comunidad de algún nodo con probabilidad alta.

Como herramienta teórica, será útil encontrar una familia de estimadores que sea maximal en cuanto a su habilidad de predecir las comunidades exactamente. Esto nos va a permitir trabajar con una familia de estimadores en concreto al demostrar que no es fuertemente consistente en vez que tener que considerar cada familia arbitraria en la demostración. Afortunadamente hay una tal familia, la de los estimadores que en estadística bayesiana se conocen como máximos a posteriori o MAP, que nos permitirán deducir propiedades concretas acerca de las mejores estimaciones posibles y nos será de gran utilidad en la demostración:

[3.2.1] Definición: Estimadores máximos a posteriori

Dado $G_{\mathcal{C}} \sim \text{SBM}(n, k, m_{\mathcal{C}}, p, q)$, un estimador **máximo a posteriori** de \mathcal{C} es un estimador $\hat{\mathcal{C}}$ que maximiza

$$\mathbb{P}(\{\text{coincidencia}(\mathcal{C}, \tilde{\mathcal{C}}) = n\} | \{G = \tilde{G}\}) \mathbb{P}(\{G = \tilde{G}\}) \quad (3.10)$$

para cada grafo \tilde{G} de n nodos sobre las asignaciones de comunidades $\tilde{\mathcal{C}}$. Notaremos al estimador en este caso $\hat{\mathcal{C}}_{\text{MAP}}$.

Dada una familia de modelos donde $G_{\mathcal{C}} \sim \text{SBM}(n, k, m_{\mathcal{C}}, p_n, q_n)$ diremos que una familia de estimadores $\hat{\mathcal{C}}$ es **familia de estimadores máximos a posteriori** cuando cada estimador lo es para cada tamaño de grafo n fijo.

[3.2.2] Observación: El estimador MAP determina la existencia de estimadores fuertemente consistentes

Si $\hat{\mathcal{C}}$ es otro estimador, usando la ley de probabilidad total sobre los grafos \tilde{G} de n nodos obtenemos que

$$\begin{aligned} & \mathbb{P}(\{\text{coincidencia}(\mathcal{C}, \hat{\mathcal{C}}) = n\}) \\ &= \sum_{\tilde{G}} \mathbb{P}(\{\text{coincidencia}(\mathcal{C}, \hat{\mathcal{C}}) = n\} | \{G = \tilde{G}\}) \mathbb{P}(\{G = \tilde{G}\}) \\ &\leq \sum_{\tilde{G}} \mathbb{P}(\{\text{coincidencia}(\mathcal{C}, \hat{\mathcal{C}}_{\text{MAP}}) = n\} | \{G = \tilde{G}\}) \mathbb{P}(\{G = \tilde{G}\}) \\ &= \mathbb{P}(\{\text{coincidencia}(\mathcal{C}, \hat{\mathcal{C}}_{\text{MAP}}) = n\}), \end{aligned} \quad (3.11)$$

lo cual nos indica que el estimador de máximo a posteriori maximiza la probabilidad de recuperar las comunidades con consistencia fuerte.

Así, si existe algún estimador fuertemente consistente para una familia de modelos SBM, el estimador máximo a posteriori será fuertemente consistente, y si una familia de estimadores de máximo a posteriori no es fuertemente consistente no habrá ningún estimador que lo sea.

Además, en el caso del modelo BSBM y bajo hipótesis suficientes los estimadores máximos a posteriori también coincidirán con otros estimadores clásicos como el de máxima verosimilitud y el de corte mínimo, como detallaremos a continuación.

[3.2.3] Definición: Estimadores de máxima verosimilitud y de corte mínimo

Dado $G_{\mathcal{C}} \sim \text{SBM}(n, k, m_{\mathcal{C}}, p, q)$, un estimador de **máxima verosimilitud** de \mathcal{C} es un estimador $\hat{\mathcal{C}}$ que maximiza

$$\mathbb{P} \left(\{ G = \tilde{G} \} \mid \{ \text{coincidencia}(\mathcal{C}, \tilde{\mathcal{C}}) = n \} \right) \quad (3.12)$$

para cada grafo \tilde{G} de n nodos sobre las asignaciones de comunidades $\tilde{\mathcal{C}}$. Notaremos al estimador en este caso $\hat{\mathcal{C}}_{\text{MLE}}$.

Además, fijo un grafo \tilde{G} , definiremos para una partición en comunidades $\tilde{\mathcal{C}}$ su **corte** como

$$\text{Corte}(\tilde{\mathcal{C}}) = \# \{ e = \{i, j\} \in \tilde{E} \mid \tilde{\mathcal{C}}(i) \neq \tilde{\mathcal{C}}(j) \} \quad (3.13)$$

la cantidad de aristas que conectan nodos de distintas comunidades. Luego, dado $G_{\mathcal{C}} \sim \text{SBM}(n, k, m_{\mathcal{C}}, p, q)$, llamaremos estimador de **corte mínimo** de \mathcal{C} a cada estimador $\hat{\mathcal{C}}$ que cumpla que condicionando a que $G = \tilde{G}$

$$\text{Corte}(\hat{\mathcal{C}}) = \text{mín} \{ \text{Corte}(\tilde{\mathcal{C}}) \} \quad (3.14)$$

para cada grafo \tilde{G} de n nodos, sobre las asignaciones de comunidades $\tilde{\mathcal{C}}$: de restringirnos a una clase de asignaciones lo aclararemos. En el caso particular del modelo BSBM, consideraremos solamente las asignaciones de comunidades $\tilde{\mathcal{C}}$ balanceadas, que asignan la misma cantidad de nodos a cada comunidad. Notaremos al estimador en estos casos $\hat{\mathcal{C}}_{\text{CUT}}$.

Nuevamente, diremos que una familia de estimadores son de máxima verosimilitud o de corte mínimo cuando para cada n fijo lo son.

Antes de continuar, en la Figura 3.2 podremos observar el corte entre dos comunidades estimadas, donde las aristas que cruzan entre comunidades distintas están marcadas en rojo. Que el estimador de mínima bisección sea óptimo (en el caso asortativo) nos confirma la intuición de que, como las comunidades más densamente conectadas en el interior que en el exterior, para minimizar el corte conviene no partir las comunidades.

Pasaremos entonces a mostrar la equivalencia, mencionando que volveremos a estudiar los cortes y algunas variantes en el Capítulo 4:

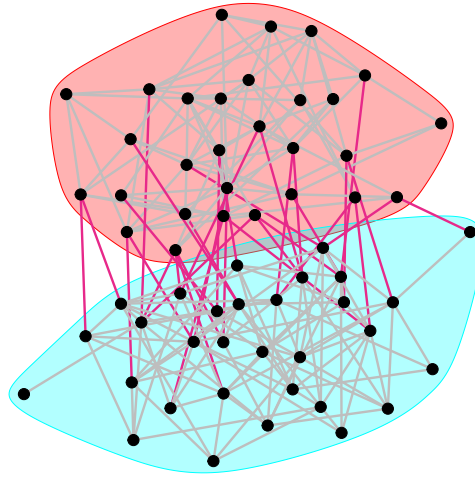


Figura 3.2: Visualización de un corte entre comunidades en el modelo BSBM.

[3.2.4] **Observación:** En el modelo BSBM el estimador MAP coincide con el estimador de máxima verosimilitud, y en el caso asortativo con el estimador de mínima bisección.

Para el modelo BSBM (n, k, p_n, q_n) , todas las asignaciones de comunidades balanceadas ocurren con la misma probabilidad, como notamos en la Observación 2.4.7, con lo que al multiplicar por la constante $\mathbb{P}(\{\text{coincidencia}(\mathcal{C}, \tilde{\mathcal{C}}) = n\})$ a la función de verosimilitud $\mathbb{P}(\{G = \tilde{G}\} | \{\text{coincidencia}(\mathcal{C}, \tilde{\mathcal{C}}) = n\})$ obtenemos que equivalentemente el estimador de máxima verosimilitud maximiza sobre $\tilde{\mathcal{C}}$ la probabilidad

$$\begin{aligned} & \mathbb{P}(\{G = \tilde{G}, \text{coincidencia}(\mathcal{C}, \tilde{\mathcal{C}}) = n\}) \\ &= \mathbb{P}(\{\text{coincidencia}(\mathcal{C}, \tilde{\mathcal{C}}) = n\} | \{G = \tilde{G}\}) \mathbb{P}(\{G = \tilde{G}\}) \end{aligned} \quad (3.15)$$

y así es también un estimador de máximo a posteriori.

Más aún, tendremos que a partir de la Observación 2.4.3 vale que si $\tilde{\mathcal{C}}$ es una asignación de comunidades que asigna $\tilde{n}_l = \frac{n}{k}$ nodos en la comunidad k para cada $1 \leq l \leq k$ y $\tilde{G} = (\tilde{V}, \tilde{E})$ tiene \tilde{n}_{l_1, l_2} aristas entre nodos de las comunidades l_1 y l_2 , para cada $1 \leq l_1, l_2 \leq k$, entonces podremos escribir la función de verosimilitud asociando de la siguiente manera:

$$\begin{aligned} & \mathbb{P}(\{G = \tilde{G}\} | \{\text{coincidencia}(\mathcal{C}, \tilde{\mathcal{C}}) = n\}) \\ &= \left(\prod_{l_1=1}^k \prod_{l_2=1}^{l_1-1} q^{\tilde{n}_{l_1, l_2}} (1-q)^{\tilde{n}_{l_1} \tilde{n}_{l_2} - \tilde{n}_{l_1, l_2}} \right) \left(\prod_{l=1}^k p^{\tilde{n}_{l, l}} (1-p)^{\frac{\tilde{n}_l^2 - \tilde{n}_l}{2} - \tilde{n}_{l, l}} \right) \\ &= \left(\prod_{l_1=1}^k \prod_{l_2=1}^{l_1-1} (1-q)^{\tilde{n}_{l_1} \tilde{n}_{l_2}} \right) \left(\prod_{l=1}^k (1-p)^{\frac{\tilde{n}_l^2 - \tilde{n}_l}{2}} \right) \left(\frac{p}{1-p} \right)^{\#\tilde{E} - \text{Corte}(\tilde{\mathcal{C}})} \left(\frac{q}{1-q} \right)^{\text{Corte}(\tilde{\mathcal{C}})} \\ &= \left[(1-q)^{\frac{k(k-1)}{2} \left(\frac{n}{k}\right)^2} (1-p)^{k \left(\frac{n}{k}\right)^2 - \frac{n}{k}} \left(\frac{p}{1-p} \right)^{\#\tilde{E}} \right] \left(\frac{q(1-p)}{p(1-q)} \right)^{\text{Corte}(\tilde{\mathcal{C}})} ; \end{aligned} \quad (3.16)$$

en la deducción anterior observamos que $\text{Corte}(\tilde{\mathcal{C}}) = \sum_{l_1=1}^k \sum_{l_2=1}^{l_1-1} \tilde{n}_{l_1, l_2}$ y $\#\tilde{E} - \text{Corte}(\tilde{\mathcal{C}}) = \sum_{l=1}^k \tilde{n}_l$ cuentan la cantidad de aristas entre las comunidades y dentro de las comunidades respectivamente.

Notamos ahora que el factor entre corchetes en (3.16) está fijo por lo que conocemos del modelo, con lo que la probabilidad solo depende del corte asociado a $\tilde{\mathcal{C}}$. Luego como en el caso asortativo $p > q$ obtenemos que $\frac{q(1-p)}{p(1-q)} < 1$ y entonces maximizar la verosimilitud del grafo equivale a encontrar una partición en comunidades que minimice la cantidad de conexiones entre comunidades distintas (y en el caso no asortativo cuando $q < p$ que la maximice).

Así, en el modelo BSBM en el caso asortativo los estimadores de máxima verosimilitud y corte mínimo son equivalentes, y también entonces los estimadores de máximo a posteriori, como queríamos ver.

Ahora entendiendo mejor nuestro estimador maximal, para que este coincida con las comunidades reales estas deben maximizar la probabilidad de generar el grafo... asintóticamente para casi todo grafo. Por eso nos concentraremos en mostrar que en este caso hay muchos ejemplos de grafos que es más probable que hayan sido generados por otras asignaciones de comunidades, y conducirá a la demostración de una implicación del Teorema 3.1.4:

Demostración de la imposibilidad en el Teorema 3.1.4. Tomemos la familia de modelos $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n = a \frac{\log(n)}{n}, q = b \frac{\log(n)}{n})$ con $\frac{a+b}{2} < 1 + \sqrt{ab}$, y mostremos que no existen familias de estimadores fuertemente consistentes.

En primer lugar, por la Observación 3.2.2 basta ver que el estimador de máximo a posteriori no es fuertemente consistente en este caso. Como el estimador MAP coincide con los estimadores MLE y CUT en el modelo BSBM, lo cual vimos en la Observación 3.2.4, basta ver que para cada grafo fijo \tilde{G} podemos encontrar una partición en comunidades $\tilde{\mathcal{C}}$ que sea distinta a la partición real \mathcal{C} y que obtenga una mayor verosimilitud o un menor corte - diremos en ese caso que $\tilde{\mathcal{C}}$ mejora a \mathcal{C} - con probabilidad superior a una constante. Para eso, buscaremos la partición entre aquellas que se obtienen intercambiando de comunidades a dos nodos en la partición de comunidades real.

Supondremos en esta demostración sin pérdida de generalidad que las comunidades reales están dadas de manera ordenada como en la Observación 2.4.7: $\mathcal{C}_1 = \{1, \dots, \frac{n}{2}\}$ y $\mathcal{C}_2 = \{\frac{n}{2} + 1, \dots, n\}$.

Llamemos $\mathcal{C}_{[i \leftrightarrow j]}$ al vector de comunidades en el que las comunidades de i y j están intercambiadas y el resto son como en \mathcal{C} , y además notemos \mathcal{A}_i al evento que i tenga estrictamente más vecinos fuera de su comunidad. Luego, tenemos que

$$\mathbb{P}\left(\cup_{i \in \mathcal{C}_1, j \in \mathcal{C}_2} \left\{ \mathcal{C}_{[i \leftrightarrow j]} \text{ mejora a } \mathcal{C} \right\}\right) \geq 2\mathbb{P}\left(\cup_{i \in \mathcal{C}_1} \mathcal{A}_i\right) - 1. \quad (3.17)$$

En efecto, cuando tenemos nodos $i \in \mathcal{C}_1, j \in \mathcal{C}_2$ tales que tanto i como j tienen una mayoría estricta de vecinos fuera de su comunidad, vale que intercambiando i y j tenemos

que la cantidad de conexiones entre las comunidades no crece, con lo que $\mathcal{C}_{[i \leftrightarrow j]}$ inducirá más verosimilitud que \mathcal{C} al resultar en un corte más chico. Más detalladamente, si X_i y X_j son las cantidades de vecinos de i y j en su respectiva comunidad, Y_i e Y_j son los vecinos de i y j fuera de su respectiva comunidad y distintos de i y de j y $A_{i,j} = \mathbb{1}_{\{\{i,j\} \in E\}}$ indica si i y j son vecinos, luego en $\mathcal{C}_{[i \leftrightarrow j]}$ valdrá que i y j tendrán $X_1 + X_2$ vecinos fuera de sus respectivas (nuevas) comunidades. Como vale que

$$X_i + X_j \leq (Y_i - 1) + (Y_j - 1) \leq (Y_i - A_{i,j}) + (Y_j - A_{i,j}), \quad (3.18)$$

obtenemos que el corte solo puede reducirse, como queríamos ver. Luego tenemos que

$$\begin{aligned} \mathbb{P} \left(\bigcup_{i \in \mathcal{C}_1, j \in \mathcal{C}_2} \left\{ \mathcal{C}_{[i \leftrightarrow j]} \text{ mejora a } \mathcal{C} \right\} \right) &\geq \mathbb{P} \left(\bigcup_{i \in \mathcal{C}_1, j \in \mathcal{C}_2} (\mathcal{A}_i \cap \mathcal{A}_j) \right) \\ &= \mathbb{P} \left(\bigcup_{i \in \mathcal{C}_1} \mathcal{A}_i \right) + \mathbb{P} \left(\bigcup_{j \in \mathcal{C}_2} \mathcal{A}_j \right) - \mathbb{P} \left(\bigcup_{i \in V} \mathcal{A}_i \right) \\ &\geq 2\mathbb{P} \left(\bigcup_{i \in \mathcal{C}_1} \mathcal{A}_i \right) - 1 \end{aligned} \quad (3.19)$$

usando además inclusión-exclusión y acotando por 1. Bastará ver entonces que vale que $\liminf_{n \rightarrow \infty} \mathbb{P} \left(\bigcup_{i \in \mathcal{C}_1} \mathcal{A}_i \right) > \frac{1}{2}$ para concluir que $\mathbb{P} \left(\bigcup_{i \in \mathcal{C}_1, j \in \mathcal{C}_2} \left\{ \mathcal{C}_{[i \leftrightarrow j]} \text{ mejora a } \mathcal{C} \right\} \right)$ es asintóticamente más grande que una constante y concluir que consistencia fuerte es imposible.

Nos concentraremos entonces en mostrar que más aún $\mathbb{P} \left(\bigcup_{i \in \mathcal{C}_1} \mathcal{A}_i \right) \xrightarrow{n \rightarrow \infty} 1$, es decir algún nodo de \mathcal{C}_1 tiene la mayoría de vecinos fuera de su propia comunidad con probabilidad alta. En primer lugar, por la desigualdad de Paley-Zygmund 2.2.3 tenemos que

$$\mathbb{P} \left(\bigcup_{i \in \mathcal{C}_1} \mathcal{A}_i \right) \geq \frac{\left(\frac{n}{2} \mathbb{P}(\mathcal{A}_1) \right)^2}{\frac{n}{2} \mathbb{P}(\mathcal{A}_1) + \frac{n}{2} \left(\frac{n}{2} - 1 \right) \mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2)} \quad (3.20)$$

expresando $\mathbb{P} \left(\bigcup_{i \in \mathcal{C}_1} \mathcal{A}_i \right) = \mathbb{P} \left(\sum_{i=1}^n \mathbb{1}_{\mathcal{A}_i} > 0 \right)$ y usando que las variables $\mathbb{1}_{\mathcal{A}_i} > 0$ se distribuyen idénticamente. Sin embargo, notamos que no son independientes: como intuición y fijando los nodos 1 y 2 en \mathcal{C}_1 tenemos que para verificar si estos nodos tienen la mayoría de vecinos fuera de \mathcal{C}_1 precisamos conocer si los propios nodos 1 y 2 son vecinos. Sin embargo, cuando n es suficientemente grande esa posible información acerca de 1 vecino deja de hacerse relevante y podremos mostrar que

$$\mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2) = (1 + o(1)) \mathbb{P}(\mathcal{A}_1) \mathbb{P}(\mathcal{A}_2) \quad (3.21)$$

una independencia asintótica entre \mathcal{A}_1 y \mathcal{A}_2 . Reemplazando en (3.20) obtenemos que

$$\begin{aligned} \mathbb{P} \left(\bigcup_{i \in \mathcal{C}_1} \mathcal{A}_i \right) &\geq \frac{\left(\frac{n}{2} \mathbb{P}(\mathcal{A}_1) \right)^2}{\frac{n}{2} \mathbb{P}(\mathcal{A}_1) + (1 + o(1)) \frac{n}{2} \left(\frac{n}{2} - 1 \right) \mathbb{P}(\mathcal{A}_1) \mathbb{P}(\mathcal{A}_2)} \\ &= \frac{\left(\frac{n}{2} \mathbb{P}(\mathcal{A}_1) \right)^2}{\left(\frac{n}{2} \mathbb{P}(\mathcal{A}_1) \right) + (1 + o(1)) \left(\frac{n}{2} \mathbb{P}(\mathcal{A}_1) \right)^2} \end{aligned} \quad (3.22)$$

Aquí podemos notar que el límite de la cota depende exclusivamente de la convergencia de $n\mathbb{P}(\mathcal{A}_1)$, como muestra la última igualdad. Afortunadamente en nuestro caso la probabilidad

$\mathbb{P}(\mathcal{A}_1)$ no decrece suficientemente rápido y podremos ver que $n\mathbb{P}(\mathcal{A}_1) \xrightarrow{n \rightarrow \infty} \infty$ con lo cual $\mathbb{P}(\cup_{i \in \mathcal{C}_1} \mathcal{A}_i) \xrightarrow{n \rightarrow \infty} 1$ como queríamos demostrar. Para demostrar que $n\mathbb{P}(\mathcal{A}_1) \rightarrow \infty$ haremos uso de una desigualdad técnica que podemos encontrar en [Abb18, Lema 4.], que establece que si $X \sim \text{Binomial}\left(\frac{n}{2}, a \frac{\log(n)}{n}\right)$ e $Y \sim \text{Binomial}\left(\frac{n}{2}, b \frac{\log(n)}{n}\right)$ son independientes luego

$$\mathbb{P}(\{Y \geq X + \mathcal{O}(1)\}) = n^{-(\frac{a-b}{2} - \sqrt{ab}) + o(1)}. \quad (3.23)$$

Notando entonces nuevamente para el nodo 1 en \mathcal{C}_1 por $X_1 \sim \text{Binomial}\left(\frac{n}{2} - 1, p_n\right)$ a su cantidad de vecinos en \mathcal{C}_1 e $Y_1 \sim \text{Binomial}\left(\frac{n}{2}, q_n\right)$ a su cantidad de vecinos fuera de \mathcal{C}_1 , tendremos que $\mathcal{A}_i = \{Y_1 \geq X_1 + 1\}$ con X_1 e Y_1 independientes. Luego, considerando $X'_1 \sim \text{Binomial}\left(\frac{n}{2}, p_n\right)$ obtenida sumando a X_1 una variable Bernoulli (p_n) independiente a X_1 e Y_1 , como $X'_1 \geq X_1$ obtenemos que $\{Y_1 \geq X_1 + 1\} \supseteq \{Y_1 \geq X'_1 + 1\}$ y así usando (3.23) podemos mostrar lo que queríamos:

$$\begin{aligned} n\mathbb{P}(\mathcal{A}_i) &\geq n\mathbb{P}(\{Y_1 \geq X'_1 + 1\}) \\ &\geq n^{1 - (\frac{a-b}{2} - \sqrt{ab}) + o(1)} \xrightarrow{n \rightarrow \infty} \infty. \end{aligned} \quad (3.24)$$

Observemos que usamos que por hipótesis $\frac{a+b}{2} < 1 + \sqrt{ab}$ para concluir el límite.

Luego solo queda mostrar la independencia asintótica enunciada en (3.21), para lo cual como adelantamos la idea clave es que los nodos 1 y 2 cuando n es grande tienen probabilidad muy alta de no ser vecinos y en ese caso hay independencia real. En efecto, podemos ver que para $A_{1,2} = \mathbb{1}_{\{\{1,2\} \in E\}}$ la variable que indica si 1 y 2 son vecinos en G vale que

$$(1 - p_n)\mathbb{P}(\mathcal{A}_1 | \{A_{1,2} = 0\}) \leq \mathbb{P}(\mathcal{A}_1) \leq \mathbb{P}(\mathcal{A}_1 | \{A_{1,2} = 0\}), \quad (3.25)$$

con un resultado análogo para $\mathbb{P}(\mathcal{A}_2)$, que resulta en

$$(1 - p_n)^3 \mathbb{P}(\mathcal{A}_1) \mathbb{P}(\mathcal{A}_2) \leq \mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2) \leq \mathbb{P}(\mathcal{A}_1) \mathbb{P}(\mathcal{A}_2) \quad (3.26)$$

y muestra lo que queríamos ver.

Más en detalle, primero notaremos que si X_1 y X_2 son las cantidades de vecinos de los nodos 1 y 2 en $\mathcal{C}_1 \setminus \{1, 2\}$ respectivamente, e Y_1 e Y_2 son las cantidades de vecinos de 1 y 2 fuera de \mathcal{C}_1 , entonces X_1 y $X_2 \sim \text{Binomial}\left(\frac{n}{2} - 2, p_n\right)$, e Y_1 e $Y_2 \sim \text{Binomial}\left(\frac{n}{2}, q_n\right)$, y además X_1 , X_2 , Y_1 , Y_2 y $A_{1,2}$ serán independientes. Luego, reescribiremos en términos de estas variables $\mathcal{A}_1 = \{Y_1 \geq X_1 + A_{1,2} + 1\}$ y $\mathcal{A}_2 = \{Y_2 \geq X_2 + A_{1,2} + 1\}$. Así, tenemos que vale que

$$\begin{aligned} \mathbb{P}(\mathcal{A}_1) &= \mathbb{P}(\{Y_1 \geq X_1 + 1, A_{1,2} = 0\}) \\ &\quad + \mathbb{P}(\{Y_1 \geq X_1 + 2, A_{1,2} = 1\}) \\ &= \mathbb{P}(\{Y_1 \geq X_1 + 1\}) \mathbb{P}(\{A_{1,2} = 0\}) \\ &\quad + \mathbb{P}(\{Y_1 \geq X_1 + 2\}) \mathbb{P}(\{A_{1,2} = 1\}) \\ &= \mathbb{P}(\{Y_1 \geq X_1 + 1\}) (1 - p_n) + \mathbb{P}(\{Y_1 \geq X_1 + 2\}) p_n, \end{aligned} \quad (3.27)$$

con lo cual acotando por abajo por el primer término y usando para acotar por arriba que $\mathbb{P}(\{Y_1 \geq X_1 + 2\}) \leq \mathbb{P}(\{Y_1 \geq X_1 + 1\})$ llegamos a la desigualdad en (3.25), donde observamos que $\mathbb{P}(\{Y_1 \geq X_1 + 1\}) = \mathbb{P}(\mathcal{A}_1 | \{A_{1,2} = 0\})$. Además, usando probabilidad total e independencia al condicionar por $\{A_{1,2} = 0\}$ obtenemos que

$$\begin{aligned} \mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2) &= \mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2 | \{A_{1,2} = 0\}) \mathbb{P}(\{A_{1,2} = 0\}) \\ &\quad + \mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2 | \{A_{1,2} = 1\}) \mathbb{P}(\{A_{1,2} = 1\}) \\ &= \mathbb{P}(\mathcal{A}_1 | \{A_{1,2} = 0\}) \mathbb{P}(\mathcal{A}_2 | \{A_{1,2} = 0\}) (1 - p_n) \\ &\quad + \mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2 | \{A_{1,2} = 1\}) p_n \end{aligned} \tag{3.28}$$

y entonces podremos llegar a la desigualdad en (3.26) usando primero (3.25) para acotar por debajo, y usando que $\mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2 | \{A_{1,2} = 1\}) \leq \mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2 | \{A_{1,2} = 0\})$ y nuevamente la desigualdad en (3.25) para acotar por arriba. ■

Quedará entonces demostrar la otra implicación del teorema, lo cual haremos en la siguiente sección mostrando unos ejemplos de algoritmos eficientes que recuperan las comunidades: los métodos espectrales.

3.3 Métodos espectrales

Miraremos ahora qué ocurre ahora cuando estamos por arriba del umbral donde vimos que es imposible conseguir estimadores fuertemente consistentes. Hay diversas estrategias que se pueden seguir para predecir las comunidades con esta precisión, como discutimos en la Nota 3.6.3, pero en este trabajo elegimos mostrar como ejemplo los estimadores espectrales, porque motivarán algunas de las ideas de los capítulos que siguen y son métodos extendidos en la literatura. La idea principal es que la información con la que contamos acerca del grafo es su matriz de adyacencia, y entender cómo son sus características nos permitirá descubrir la información necesaria para revelar las comunidades en el grafo.

[3.3.1] Observación: Descomposición espectral de la matriz de adyacencia teórica en un modelo BSBM y motivación de los métodos espectrales

Consideremos una familia de modelos balanceados $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$, donde a partir de ahora en esta sección trabajaremos con modelos en los que además agregamos bucles independientemente en cada nodo con probabilidad p_n para una menor complejidad en las cuentas. Como en la Observación 2.4.7, llamemos $\tilde{\mathcal{C}}$ a la partición donde las comunidades son $\tilde{\mathcal{C}}_1 = \{1, \dots, \frac{n}{2}\}$ y $\tilde{\mathcal{C}}_2 = \{\frac{n}{2} + 1, \dots, n\}$.

Llamemos \hat{A} a la matriz de adyacencia de G y tendremos que al condicionar por $\{\text{coincidencia}(\mathcal{C}, \tilde{\mathcal{C}}) = n\}$ podremos escribir a \hat{A} en bloques de tamaño $\frac{n}{2} \times \frac{n}{2}$ de la siguiente manera

$$A = \begin{pmatrix} A^{1,1} & A^{1,2} \\ A^{2,1} & A^{2,2} \end{pmatrix}, \tag{3.29}$$

donde los coeficientes en $A^{1,1}$ y $A^{2,2}$ serán variables Bernoulli (p_n) y los coeficientes en $A^{1,2}$ y $A^{2,1}$ serán variables Bernoulli (q_n), y además tendremos independencia para los coeficientes en la diagonal y por sobre ella - muchos estarán repetidos pues la matriz es simétrica.

Desde ya, esta matriz podría no tener una descomposición espectral sencilla, pero valdrá que al tomar esperanza tendremos que podremos escribirla en bloques como

$$A = \mathbb{E} [\hat{A}] = \begin{pmatrix} p_n \mathbf{J} & q_n \mathbf{J} \\ q_n \mathbf{J} & p_n \mathbf{J} \end{pmatrix}, \quad (3.30)$$

donde la esperanza se toma coeficiente a coeficiente; ahora sí tenemos una matriz no muy intimidante. En adelante llamaremos a \hat{A} la matriz de adyacencia empírica y a A la matriz de adyacencia teórica para distinguirlas.

Volviendo a la descomposición espectral, inmediatamente vemos que

- $u_1 = \frac{1}{\sqrt{n}} \begin{pmatrix} \mathbf{1} \\ \mathbf{1} \end{pmatrix}$ es autovector de A con autovalor $n \frac{p_n + q_n}{2}$, y
- $u_2 = \frac{1}{\sqrt{n}} \begin{pmatrix} \mathbf{1} \\ -\mathbf{1} \end{pmatrix}$ es autovector de A con autovalor $n \frac{p_n - q_n}{2}$,

y más aún estos son los únicos con autovalor no nulo. Esto es así pues el rango de A es 2, siendo $\text{Im}(A) \subseteq \left\langle \begin{pmatrix} p_n \mathbf{1} \\ q_n \mathbf{1} \end{pmatrix}, \begin{pmatrix} q_n \mathbf{1} \\ p_n \mathbf{1} \end{pmatrix} \right\rangle$ y nuestros autovectores (o esa) una base.

Retomando nuestro objetivo de distinguir las comunidades, el primer autovector no es muy interesante, pues no distingue información alguna de los nodos; sin embargo, el segundo autovector sí: nos parte a los nodos en dos grupos según el signo de su respectiva coordenada. Así, tendremos que podemos estimar las comunidades mediante

$$\hat{C}(i) = \phi(\text{signo}((u_2)_i)) \quad (3.31)$$

y de hecho recuperarlas de manera exacta, donde ϕ es la función que manda $\phi(x) = 1$ si $x \geq 0$ y $\phi(x) = 2$ si no.

Parece entonces razonable dada la matriz de adyacencia del modelo \hat{A} encontrar el segundo autovector \hat{u}_2 y clasificar a los nodos según el signo. Pero... ¡cuidado! Esto es válido para A la esperanza de la matriz de adyacencia, y no para \hat{A} la matriz que vamos a tener a disposición.

Afortunadamente, podremos mostrar que no va a haber mucha diferencia entre el autovector de \hat{A} y el autovector de A cuando n es grande.

En primer lugar, tenemos un resultado de álgebra lineal que podremos usar para cuantificar qué tan lejos van a estar los autovectores de A y de \hat{A} , el teorema de Davis-Kahan. Originalmente el teorema muestra que el ángulo entre los dos autovectores es chico (y de

ahí el nombre teorema del seno de Davis-Kahan), pero inmediatamente se puede recuperar una cota para la distancia a partir de la cota para el ángulo teniendo tal vez que re-orientar alguno de los autovectores; esta es la versión que presentaremos acá y se ajusta mejor a lo que vamos a necesitar. Sin más preámbulos:

[3.3.2] Teorema: Teorema de Davis-Kahan - [Abb18, Teorema 9.]

Sea M una matriz simétrica en $\mathbb{R}^{n \times n}$ con autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ y autovectores ortonormales asociados u_1, \dots, u_n .

Sea E una matriz simétrica en $\mathbb{R}^{n \times n}$ y consideremos la perturbación de M por E , $M + E$, para la cual llamaremos $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_n$ a sus autovalores y $\hat{u}_1, \dots, \hat{u}_n$ a sus respectivos autovectores ortonormales asociados.

Fijemos j un índice con $1 \leq j \leq n$, y llamemos $\Delta(M, j) = \min \{ \lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1} \}$ la mínima separación entre el j -ésimo autovalor de M y los contiguos, donde para los casos en los extremos definiremos $\lambda_0 = \infty$ y $\lambda_{n+1} = -\infty$.

Luego, existe alguna constante $C > 0$ tal que

$$\min_{s=\pm 1} \{ \|s\hat{u}_j - u_j\|_2 \} \leq C \frac{\|E\|_2}{\Delta(M, j)} \quad (3.32)$$

Más aún, si vale que $\|E\|_2 \leq \Delta(M, j)$, entonces además podremos acotar para alguna constante $C > 0$ como a continuación:

$$\min_{s=\pm 1} \{ \|s\hat{u}_j - u_j\|_2 \} \leq C \frac{\|Eu_j\|_2}{\Delta(M, j)}. \quad (3.33)$$

Vemos entonces que mostrar que los segundos autovectores van a ser próximos se reduce a mostrar que A y \hat{A} son parecidas, en cuyo caso la perturbación entre ambas será chica, y que los autovalores de A se distancian razonablemente del segundo autovalor, entonces tendremos que la diferencia entre los autovectores será acotada. Por la Observación 3.3.1 tenemos que la máxima distancia entre el segundo autovalor y los contiguos es

$$\max \left\{ n \frac{p_n + q_n}{2} - n \frac{p_n - q_n}{2}, n \frac{p_n - q_n}{2} - 0 \right\} = \max \left\{ nq_n, n \frac{p_n - q_n}{2} \right\}; \quad (3.34)$$

resulta que esta separación será suficiente en el regimen que es de nuestro interés, donde tendremos que $p_n, q_n \asymp \frac{\log(n)}{n}$. Más aún, tenemos un resultado que nos permitirá ver que los autovalores de la matriz empírica son parecidos a los de la matriz teórica y entonces garantizar que pasa algo parecido para los autovalores de la matriz de adyacencia empírica, el teorema de Weyl.

[3.3.3] Teorema: Teorema de Weyl

Sea M una matriz simétrica en $\mathbb{R}^{n \times n}$ con autovalores $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$.

Dada además E una matriz simétrica en $\mathbb{R}^{n \times n}$, consideremos la perturbación de M por E , $M + E$, que será simétrica con autovalores $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_n$.

Luego, vale que

$$|\hat{\lambda}_i - \lambda_i| \leq \|E\|_2 \quad (3.35)$$

para cada $1 \leq i \leq n$.

Bastará entonces ver que A y \hat{A} son próximas para convencernos de que tiene sentido usar esta estrategia para estimar las comunidades, y es aquí donde el siguiente resultado nos garantizará que esto ocurre:

[3.3.4] Teorema: La matriz de adyacencia empírica está cerca de la matriz de adyacencia teórica - [LR15, Teorema 5.2.]

Sea \hat{A} la matriz de adyacencia de un grafo aleatorio de n nodos para el cual las aristas se agregan de manera independiente. Luego, llamemos $A = \mathbb{E}[\hat{A}]$ y supongamos que cumple que $nA_{i,j} \leq d$ para cada $1 \leq i, j \leq n$ y para algún número $d \geq c_0 \log(n)$ con $c_0 > 0$.

Entonces, para cada $r > 0$ existe una constante $C = C(r, c_0)$ tal que

$$\mathbb{P} \left(\left\{ \|\hat{A} - A\|_2 \leq C\sqrt{d} \right\} \right) \geq 1 - n^{-r} \quad (3.36)$$

Remarcamos que la única forma en que A y \hat{A} van a ser cercanas es en cuanto a normas de operadores, pues coeficiente a coeficiente esto no ocurrirá: en efecto, \hat{A} tiene coeficientes 0 o 1 y A tiene coeficientes p_n o q_n y así las diferencias $1 - p_n$ o $1 - q_n$ siempre serán significativas. Sin embargo, siendo la norma de operadores la que nos interesaba resulta que nuevamente en el regimen que estamos estudiando esta cota bastará. Desde ya, vamos a hacer esto más riguroso en la siguiente sección, pero por ahora queda entonces motivada la siguiente definición:

[3.3.5] Definición: Estimadores espectrales $\hat{\mathcal{C}}_{\text{SP}}$

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos balanceados. Entonces, definimos el **estimador espectral** para \mathcal{C} que notaremos $\hat{\mathcal{C}}_{\text{SP}}$ como

$$\hat{\mathcal{C}}_{\text{SP}}(i) = \phi(\text{signo}((\hat{u}_2)_i)) \quad (3.37)$$

para cada $1 \leq i \leq n$, donde \hat{u}_2 es el autovector asociado al segundo autovalor de \hat{A} la matriz de adyacencia del modelo ordenados de mayor a menor y ϕ es la función que manda $\phi(x) = 1$ si $x \geq 0$ y $\phi(x) = 2$ si no.

Como adelantamos, para demostrar la implicación restante del Teorema 3.1.4 mostraremos que los métodos espectrales son fuertemente consistentes. Lo haremos en dos partes: primero demostraremos consistencia débil para los estimadores haciendo riguroso el argumento que esbozamos previamente y luego en un segundo paso más técnico veremos que con argumentos más sutiles se puede mostrar que en realidad también este estimador es

fuertemente consistente.

3.4 Consistencia débil para métodos espectrales

Comenzaremos primero demostrando una desigualdad técnica, que nos permitirá controlar la distancia entre los correspondientes autovectores:

[3.4.1] Lema: Los segundos autovectores están cerca

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos balanceados donde $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ con $a > b > 0$.

Luego, vale que para cada $r \in \mathbb{N}_0$ se tiene que

$$\mathbb{P} \left(\left\{ \min_{s=\pm 1} \{ \|s\hat{u}_2 - u_2\|_2 \} = \mathcal{O} \left(\frac{1}{\sqrt{\log(n)}} \|u_2\|_2 \right) \right\} \right) \geq 1 - o(n^{-r}). \quad (3.38)$$

Demostración. Hace falta únicamente dejar por escrito lo ya anticipado. Por el Teorema 3.3.2 de Davis-Kahan, tendremos que

$$\min_{s=\pm 1} \{ \|s\hat{u}_2 - u_2\|_2 \} \leq \frac{\|\hat{A} - A\|_2}{\max \left\{ b, \frac{a-b}{2} \right\} \log(n)} \quad (3.39)$$

Ahora usando el Teorema 3.3.4, $\|\hat{A} - A\|_2 \lesssim \sqrt{\log(n)}$ con probabilidad mayor que $1 - o(n^{-r})$, y usando además que ya calculamos $\Delta(A, 2) = \max \left\{ b, \frac{a-b}{2} \right\} \log(n)$ en (3.34) tendremos que con alta probabilidad vale que

$$\min_{s=\pm 1} \{ \|s\hat{u}_2 - u_2\|_2 \} \lesssim \frac{1}{\max \left\{ b, \frac{a-b}{2} \right\} \sqrt{\log(n)}}. \quad (3.40)$$

Por último, como $\|u_2\|_2 = 1$, con probabilidad mayor que $1 - o(n^{-r})$ tenemos que

$$\min_{s=\pm 1} \{ \|s\hat{u}_2 - u_2\|_2 \} \lesssim \frac{1}{\sqrt{\log(n)}} \|u_2\|_2, \quad (3.41)$$

y completamos la demostración. ■

De aquí a la demostración de consistencia débil, basta notar que los nodos mal clasificados cumplirán que tendrán coordenadas muy distintas en el autovalor empírico a las del autovalor teórico, y como esto no es usual deben ser pocos:

[3.4.2] Teorema: Consistencia débil para métodos espectrales

Sea $G_{\mathcal{C}} \sim \text{BSBM} \left(n, 2, p_n = a \frac{\log(n)}{n}, q_n = b \frac{\log(n)}{n} \right)$ una familia de modelos con $a > b > 0$ números positivos donde $\frac{a+b}{2} > 1 + \sqrt{ab}$.

Entonces, la familia de estimadores $\hat{\mathcal{C}}_{\text{SP}}$ estima de manera débilmente consistente a \mathcal{C} .

Más aún, el conjunto de nodos mal clasificados $\mathcal{B}(\hat{\mathcal{C}}_{\text{SP}})$ cumple que para cada $r \in \mathbb{N}_0$

vale que

$$\mathbb{P} \left(\left\{ \#(\mathcal{B}(\widehat{\mathcal{C}}_{\text{SP}})) \leq \mathcal{O} \left(\frac{1}{\log(n)} n \right) \right\} \right) \geq 1 - o(n^{-r}). \quad (3.42)$$

Demostración. Basta ver que $\mathcal{B}(\widehat{\mathcal{C}}_{\text{SP}})$ tiene menos que $\frac{1}{\log(n)}n$ miembros, pues esa cantidad es $o(n)$ y entonces el estimador será débilmente consistente.

Sin pérdida de generalidad supongamos que el mínimo se realiza para $s = 1$ al considerar la distancia entre los segundos autovectores $\min_{s=\pm 1} \{ \|s\hat{u}_2 - u_2\|_2 \}$.

Notemos ahora que si $i \in \mathcal{B}(\widehat{\mathcal{C}}_{\text{SP}})$, luego su signo debe diferir del que tiene u_2 , con lo que en esa coordenada i debe valer que $|(\hat{u}_2)_i - (u_2)_i| > |(u_2)_i| = \frac{1}{\sqrt{n}}$ usando la expresión para u_2 que obtuvimos en la Observación 3.3.1. Luego podremos acotar usando el Lema 3.4.1 para conseguir que con probabilidad mayor que $1 - o(n^{-r})$ vale que

$$\#(\mathcal{B}(\widehat{\mathcal{C}}_{\text{SP}})) \left(\frac{1}{\sqrt{n}} \right)^2 \leq \sum_{i=1}^n |(\hat{u}_2)_i - (u_2)_i|^2 = \|\hat{u}_2 - u_2\|_2^2 \lesssim \frac{1}{\log(n)} \|u_2\|_2^2. \quad (3.43)$$

Entonces, como $\|u_2\|_2 = 1$, vale la cota que queríamos con suficientemente alta probabilidad. ■

Habiendo demostrado que los estimadores espectrales son débilmente consistentes, como adelantamos estudiaremos más cuidadosamente en la siguiente sección por qué también son estimadores fuertemente consistentes.

3.5 Consistencia fuerte para métodos espectrales

Para demostrar la consistencia fuerte del estimador espectral, bastaría con ver que coordenada a coordenada el segundo autovector empírico \hat{u}_2 está cerca del segundo autovector teórico u_2 . Sin embargo, en [Abb18, pág. 40.] se explica que no siempre valdrá que estos dos están suficientemente cerca, y será necesario adaptar nuestra estrategia.

Afortunadamente podemos encontrar otro estimador fuertemente consistente que se va a parecer lo suficiente al estimador espectral distinto al estimador teórico. Retomando lo aprendido en el Teorema 2.4.13, cada nodo tiene una mayoría estricta entre sus vecinos de su comunidad, lo que indica que si podemos clasificar suficientemente bien la vecindad de un nodo entonces tendremos información suficiente para clasificarlo. Suponiendo en primer lugar que contamos con un oráculo que nos revela las comunidades de todos los nodos salvo dos, uno de cada comunidad, ¿podríamos predecir la comunidad de los nodos restantes? Como sabemos que la mayoría de los vecinos son clasificados correctamente, podríamos guiarnos por ese criterio y definir un estimador que le asigne a cada uno la comunidad que tiene la mayoría de sus vecinos. Esa es la intuición detrás del estimador espectral ayudado por un oráculo, que definiremos a continuación:

[3.5.1] Definición: Estimador espectral ayudado por un oráculo $\hat{\mathcal{C}}_{\text{SPA}}$

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos balanceados. Entonces, definimos el **estimador espectral ayudado por un oráculo** para \mathcal{C} que notaremos $\hat{\mathcal{C}}_{\text{SPA}}$ como

$$\hat{\mathcal{C}}_{\text{SPA}}(i) = \phi \left(\text{signo} \left(\left(\frac{1}{\lambda_2} \hat{A} u_2 \right)_i \right) \right) \quad (3.44)$$

para cada $1 \leq i \leq n$, donde \hat{u}_2 es el autovector asociado al segundo autovalor de \hat{A} la matriz de adyacencia del modelo ordenados de mayor a menor, y u_2 es el segundo autovector de A la matriz de adyacencia teórica con autovalor asociado λ_2 , donde nuevamente los autovectores están ordenados de mayor a menor, y ϕ es la función que manda $\phi(x) = 1$ si $x \geq 0$ y $\phi(x) = 2$ si no.

Resulta que el estimador espectral ayudado por un oráculo va a ser fuertemente consistente, como detallaremos en la Proposición 3.5.6. Sin embargo, este estimador es puramente teórico, pues no vamos a conocer el oráculo u_2 .

Afortunadamente, podremos conseguir un estimador más que nos servirá de puente entre el estimador ayudado por el oráculo y el estimador espectral. Motivados por la estrategia, podemos estimar las comunidades con el estimador espectral para todos los nodos salvo uno, confiando que la predicción va a ser bastante precisa, y después nuevamente asignarle al nodo separado la comunidad de más peso entre sus vecinos. Esta estrategia será suficientemente semejante a la que emplea el oráculo y nos permitirá obtener otro estimador fuertemente consistente, que definimos a continuación:

[3.5.2] Definición: Estimador espectral por vecindades $\hat{\mathcal{C}}_{\text{SPV}}$

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos balanceados.

Llamemos $\hat{A}^{(i)}$ a la matriz de adyacencia que se obtiene al dejar al i -ésimo nodo aislado poniendo en la i -ésima columna y en la i -ésima fila de \hat{A} ceros, y análogamente llamemos $A^{(i)} = \mathbb{E} [\hat{A}^{(i)}]$ a la matriz teórica de este nuevo grafo aleatorio, para cada $1 \leq i \leq n$.

Entonces, definimos el **estimador espectral por vecindades** para \mathcal{C} que notaremos $\hat{\mathcal{C}}_{\text{SPV}}$ como

$$\hat{\mathcal{C}}_{\text{SPV}}(i) = \phi \left(\text{signo} \left(\left(\frac{1}{\lambda_2} \hat{A} \hat{u}_2^{(i)} \right)_i \right) \right) \quad (3.45)$$

para cada $1 \leq i \leq n$, donde $\hat{u}_2^{(i)}$ es el autovector asociado al segundo autovalor de $\hat{A}^{(i)}$, donde los autovalores se ordenan de mayor a menor, y λ_2 es el segundo autovalor de A la matriz de adyacencia teórica, y ϕ es la función que manda $\phi(x) = 1$ si $x \geq 0$ y $\phi(x) = 2$ si no.

Desde ya, la utilidad de este estimador también será teórica, pues para predecir las comunidades de todos los nodos habrá que hacer la estimación espectral n veces lo que

será computacionalmente costoso. Sin embargo, como ahora sí este estimador estará suficientemente cerca del estimador espectral definido previamente, ambos serán fuertemente consistentes, y no hará falta usar los estimadores espectrales por vecindades.

Por último, antes de demostrar la consistencia fuerte de estos estimadores presentamos unos lemas técnicos que nos van a ser de utilidad para explotar las mayorías de vecinos de la propia comunidad que le corresponden a cada nodo. El primero justifica que en las condiciones en las cuales cada nodo tiene una mayoría entre sus vecinos de su propia comunidad, la estimación ayudada por un oráculo va a funcionar:

[3.5.3] Lema: Las mayorías no son chicas con probabilidad alta - [AFWZ20, Lema 8.]

Supongamos que $a > b$ son números reales, y que $(W_i)_{1 \leq i \leq \frac{n}{2}}$ y $(Z_i)_{1 \leq i \leq \frac{n}{2}}$ son variables independientes con $W_i \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p_n)$ y $Z_i \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(q_n)$ para $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$. Luego, para cada $\eta \in \mathbb{R}$ vale que

$$\mathbb{P} \left(\left\{ \sum_{i=1}^{\frac{n}{2}} W_i - Z_i \leq \eta \log(n) \right\} \right) \leq n^{-\frac{(\sqrt{a}-\sqrt{b})^2}{2} + \frac{\eta}{2} \log\left(\frac{a}{b}\right)}. \quad (3.46)$$

El segundo lema nos permitirá mostrar que no perdemos mucho usando la estimación espectral sobre el resto de los nodos en lugar que la estimación con el oráculo; esto quedará más claro esto en las demostraciones:

[3.5.4] Lema: Los estimadores por vecindades se portan parecido a los estimadores ayudados por un oráculo - [AFWZ20, versión simplificada del Lema 7.]

Si $w \in \mathbb{R}^n$ es un vector fijo, y $(X_i)_{1 \leq i \leq n}$ son variables aleatorias Bernoulli (p_i) independientes, y $p \geq \max_{1 \leq i \leq n} \{p_i\}$ y $\eta \geq 0$, entonces para cada $\gamma \in (0, 1)$ vale que

$$\mathbb{P} \left(\left\{ \left| \sum_{i=1}^n w_i (X_i - \mathbb{E}[X_i]) \right| \geq \frac{(2 + \eta)pn \max \left\{ \|w\|_\infty, \frac{1}{\gamma\sqrt{n}} \|w\|_2 \right\}}{\max \left\{ 1, \log\left(\frac{1}{\gamma}\right) \right\}} \right\} \right) \leq 2e^{-\eta np}. \quad (3.47)$$

El tercer y último lema lo agregamos para afianzar las intuiciones acerca de los ordenes de los objetos con los que vamos a estar tratando y aumentar la legibilidad de las demostraciones. Algunas ya las habíamos mencionado anteriormente, pero las dejamos escritas acá nuevamente para mayor comodidad luego:

[3.5.5] Lema: Preliminares para la Proposición (3.5.7)

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos balanceados donde $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para ciertos $a > b > 0$ números positivos.

Llamamos $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ a los autovalores de A y $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_n$ a los autovalores de \hat{A} . Luego,

1. $\|A_{j,\bullet}\|_2 = \mathcal{O}\left(\frac{\log(n)}{\sqrt{n}}\right)$ para cada $j \in V$,
2. $\lambda_1 = \frac{a+b}{2} \log(n)$, $\lambda_2 = \frac{a-b}{2} \log(n)$ y $\lambda_j = 0$ para cada $j > 2$, y
3. $\Delta(A, 2) = \max\left\{b, \frac{a-b}{2}\right\} \log(n) \asymp \log(n)$.

Además, fijo $r \in \mathbb{N}_0$, con probabilidad mayor que $1 - o(n^{-r})$ y donde i es cualquier nodo fijo valen

1. $\|\widehat{A} - A\|_2$, $\|\widehat{A}^{(i)} - A\|_2$ y $\|\widehat{A} - \widehat{A}^{(i)}\|_2$ todas son $\mathcal{O}\left(\sqrt{\log(n)}\right)$,
2. $\|A_{j,\bullet}\|_2$ y $\|A_{j,\bullet}^{(i)}\|_2$ todas son $\mathcal{O}(\log(n))$ para cada $1 \leq j \leq n$,
3. $|\widehat{\lambda}_2 - \lambda_2| = \mathcal{O}\left(\sqrt{\log(n)}\right)$ y $\widehat{\lambda}_2 = \Theta(\lambda_2) = \Theta(\log(n))$,
4. $\Delta(\widehat{A}, 2) = \Theta(\log(n))$ donde Δ es como en el Teorema 3.3.2 de Davis-Kahan.

Demostración del Lema 3.5.5. Mostraremos primero las cotas determinísticas.

La cota $\|A_{j,\bullet}\|_2 = \mathcal{O}\left(\frac{\log(n)}{\sqrt{n}}\right)$ vale por un cálculo directo

$$\begin{aligned}
\|A_{j,\bullet}\|_2 &= \sqrt{\sum_{i=1}^{\frac{n}{2}} p_n^2 + \sum_{i=\frac{n}{2}+1}^n q_n^2} \\
&= \sqrt{\frac{n}{2} p_n^2 + \frac{n}{2} q_n^2} \\
&= \sqrt{\frac{a^2 + b^2 \log(n)^2}{2} \frac{1}{n}} \\
&\asymp \frac{\log(n)}{\sqrt{n}},
\end{aligned} \tag{3.48}$$

usando la expresión para A obtenida en la Observación 3.3.1.

Las expresiones en 2. se deducen inmediatamente de las que habíamos obtenido en la Observación 3.3.1, y la derivación de 3. la podemos encontrar en (3.34).

Pasemos ahora a la parte que vale con alta probabilidad. En primer lugar, nuestra herramienta fundamental será el Teorema 3.3.4, que nos da que $\|\widehat{A} - A\|_2 \lesssim \sqrt{\log(n)}$ con probabilidad mayor que $1 - o(n^{-3})$.

A partir de esta desigualdad, conseguimos que $\|\widehat{A}_j\|_2 \lesssim \sqrt{\log(n)}$ porque

$$\begin{aligned}
\|\widehat{A}_i\|_2 &\leq \|\widehat{A}_i - A_i\|_2 + \|A_i\|_2 \\
&\leq \|\widehat{A} - A\|_2 + \|A_i\|_2 \\
&\lesssim \sqrt{\log(n)} + \frac{\log(n)}{\sqrt{n}} \\
&\asymp \sqrt{\log(n)}
\end{aligned} \tag{3.49}$$

Luego, podemos acotar

$$\|\widehat{A}^{(i)} - \widehat{A}\|_2 \leq \|\widehat{A}^{(i)} - \widehat{A}\|_F \leq \sqrt{2} \|\widehat{A}_i\|_2 \lesssim \sqrt{\log(n)}, \quad (3.50)$$

donde $\|\bullet\|_F$ denota la norma Frobenius que consta en tratar a la matriz como un vector y tomarle la norma vectorial 2 usual. Como la resta $\widehat{A}^{(i)} - \widehat{A}$ tiene coordenadas no nulas solo para la fila i -ésima y la columna i -ésima y es simétrica, tenemos que al tomar norma Frobenius cada coeficiente en \widehat{A}_i aparece sumado al cuadrado a lo sumo dos veces y de ahí se desprende la desigualdad.

Usando que $\|\widehat{A}^{(i)} - \widehat{A}\|_2 \lesssim \sqrt{\log(n)}$ y además $\|\widehat{A} - A\|_2 \lesssim \sqrt{\log(n)}$ por desigualdad triangular obtenemos que $\|\widehat{A}^{(i)} - A\|_2 \lesssim \sqrt{\log(n)}$, lo que muestra 1..

Además, análogamente que como en (3.49) deducimos que $\|\widehat{A}^{(j)}\|_2 \lesssim \sqrt{\log(n)}$, y mostramos 2..

Veamos ahora que $|\hat{\lambda}_2 - \lambda_2| \lesssim \sqrt{\log(n)}$ y $\hat{\lambda}_2 \asymp \lambda_2 \asymp \log(n)$. Lo primero vale porque por el Teorema 3.3.3 de Weyl tenemos que

$$|\hat{\lambda}_2 - \lambda_2| \leq \|\widehat{A} - A\|_2 \lesssim \sqrt{\log(n)} \quad (3.51)$$

usando además 1. para concluir. Luego existe alguna constante $C > 0$ tal que

$$\lambda_2 - C\sqrt{\log(n)} \lesssim \hat{\lambda}_2 \leq \lambda_2 + C\sqrt{\log(n)} \quad (3.52)$$

y así como $\lambda_2 \asymp \log(n)$ como vimos anteriormente tendremos que $\hat{\lambda}_2 \asymp \log(n)$ también. Esto muestra 3..

Por último, para ver que $\Delta(\widehat{A}, 2) \asymp \log(n)$ usaremos nuevamente el Teorema 3.3.3 de Weyl para acotar la distancia entre los autovalores empíricos y teóricos por la norma. En efecto, como

$$\hat{\lambda}_1 - \hat{\lambda}_2 \geq (\lambda_1 - \|\widehat{A} - A\|_2) - (\lambda_2 + \|\widehat{A} - A\|_2) \gtrsim b \log(n) - 2\sqrt{\log(n)} \asymp \log(n) \text{ y} \quad (3.53)$$

$$\hat{\lambda}_1 - \hat{\lambda}_2 \leq (\lambda_1 + \|\widehat{A} - A\|_2) - (\lambda_2 - \|\widehat{A} - A\|_2) \lesssim b \log(n) + 2\sqrt{\log(n)} \asymp \log(n)$$

obtenemos que $\hat{\lambda}_1 - \hat{\lambda}_2 \asymp \log(n)$, y análogamente

$$\hat{\lambda}_2 - \hat{\lambda}_3 \geq (\lambda_2 - \|\widehat{A} - A\|_2) - (0 + \|\widehat{A} - A\|_2) \gtrsim \frac{a-b}{2} \log(n) - 2\sqrt{\log(n)} \asymp \log(n) \text{ y}$$

$$\hat{\lambda}_2 - \hat{\lambda}_3 \leq (\lambda_2 + \|\widehat{A} - A\|_2) - (0 - \|\widehat{A} - A\|_2) \gtrsim \frac{a-b}{2} \log(n) + 2\sqrt{\log(n)} \asymp \log(n), \quad (3.54)$$

de lo que se sigue que el máximo entre ambas diferencias tiene orden exactamente $\log(n)$ y muestra 4.. ■

Ahora sí, ya sin más preámbulos hagamos rigurosas las afirmaciones presentadas previamente, para lo cual haremos uso sin dar más explicaciones del Lema 3.5.5. En primer lugar,

[3.5.6] Proposición: El estimador espectral ayudado por un oráculo $\widehat{\mathcal{C}}_{\text{SPA}}$ es fuertemente consistente.

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos balanceados donde $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para $a > b > 0$ números positivos que cumplen además que $\frac{a+b}{2} > 1 + \sqrt{ab}$.

Luego, $\widehat{\mathcal{C}}_{\text{SPA}}$ es fuertemente consistente.

Además, vale que

$$\mathbb{P} \left(\left\{ \left| \left(\frac{1}{\lambda_2} \widehat{A}u_2 \right)_i \right| \geq \Omega(1) \frac{1}{\sqrt{n}} \text{ para cada } 1 \leq i \leq n \right\} \right) \geq 1 - o(1). \quad (3.55)$$

Demostración. Por la Observación 2.2.5 basta que demos demos que vale la cota para cada nodo individualmente con probabilidad mayor que $1 - o(n^{-1})$, pues son n eventos que queremos que valgan simultáneamente. La demostración para nodos de cada comunidad será análoga, así que supondremos sin pérdida de generalidad en adelante que $i \in \mathcal{C}_1$.

Ahora bien, tenemos que como $u_2 = \frac{1}{\sqrt{n}} \begin{pmatrix} \mathbf{1} \\ -\mathbf{1} \end{pmatrix}$ y además $\lambda_2 = \frac{a-b}{2} \log(n)$ un cálculo directo nos muestra que

$$\left(\frac{1}{\lambda_2} \widehat{A}u_2 \right)_i = \frac{1}{\frac{a-b}{2} \log(n)} \frac{1}{\sqrt{n}} \left[\sum_{j=1}^{\frac{n}{2}} \widehat{A}_{i,j} - \sum_{j=\frac{n}{2}+1}^n \widehat{A}_{i,j} \right], \quad (3.56)$$

donde el signo de esta expresión depende solamente de la suma, que es una resta de binomiales.

Eso es precisamente lo que el Lema 3.5.3 controla, y es así que usándolo obtenemos que para

$$\mathbb{P} \left(\left\{ \sum_{j=1}^{\frac{n}{2}} \widehat{A}_{i,j} - \sum_{j=\frac{n}{2}+1}^n \widehat{A}_{i,j} \leq \eta \log(n) \right\} \right) \leq n^{-\frac{(\sqrt{a}-\sqrt{b})^2}{2} + \frac{\eta}{2} \log\left(\frac{a}{b}\right)} \quad (3.57)$$

que para η suficientemente chico es $o(n^{-1})$ bajo nuestra hipótesis de que $\frac{a+b}{2} > 1 + \sqrt{ab}$, que equivalentemente nos dice que $\frac{(\sqrt{a}-\sqrt{b})^2}{2} > 1$.

Esto muestra que $\widehat{\mathcal{C}}_{\text{SPA}}$ es fuertemente consistente.

La cota en (3.55) se obtiene al combinar la expresión en (3.56) y la desigualdad en (3.57), mediante lo cual podremos ver que con probabilidad mayor que $1 - o(n^{-1})$ vale que

$$\left| \left(\frac{1}{\lambda_2} \widehat{A}u_2 \right)_i \right| \geq \frac{1}{\frac{a-b}{2} \log(n)} \frac{1}{\sqrt{n}} [\eta \log(n)] = \frac{2\eta}{a-b} \frac{1}{\sqrt{n}} \quad (3.58)$$

y terminamos la demostración. ■

Podremos ahora mostrar que los estimadores de interés práctico estiman de forma similar al estimador espectral ayudado por un oráculo.

[3.5.7] Proposición: Los vectores \hat{u}_2 , $\left(\left(\frac{1}{\lambda_2}\hat{A}\hat{u}_2^{(i)}\right)_i\right)$ y $\frac{1}{\lambda_2}\hat{A}u_2$ están cerca coeficiente a coeficiente.

Sea $G_C \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de estimadores con $p_n = a\frac{\log(n)}{n}$ y $q_n = b\frac{\log(n)}{n}$ para $a > b > 0$ que cumplen además que $\frac{a+b}{2} > 1 + \sqrt{ab}$.

Luego,

1. dado i un nodo cualquiera, podemos suponer sin pérdida de generalidad que simultáneamente $s = 1$ minimiza

$$\min_{s=\pm 1} \left\{ \left\| s\hat{u}_2 - \hat{u}_2^{(i)} \right\|_2 \right\}, \min_{s=\pm 1} \left\{ \left\| su_2 - \hat{u}_2^{(i)} \right\|_2 \right\}, \text{ y } \min_{s=\pm 1} \left\{ \|su_2 - \hat{u}_2\|_2 \right\}; \text{ y} \quad (3.59)$$

2. en ese caso, vale que para cada $r \in \mathbb{N}_0$ valen ambas

$$\mathbb{P} \left(\left\{ \left\| \hat{u}_2 - \frac{1}{\lambda_2}\hat{A}u_2 \right\|_\infty \leq \mathcal{O} \left(\frac{1}{\log(\log(n))} \right) \frac{1}{\sqrt{n}} \right\} \right) \geq 1 - o(n^{-r}) \quad (3.60)$$

y

$$\mathbb{P} \left(\left\{ \left| \left(\frac{1}{\lambda_2}\hat{A}\hat{u}_2^{(i)} \right)_i - \left(\frac{1}{\lambda_2}\hat{A}u_2 \right)_i \right| \leq \mathcal{O} \left(\frac{1}{\log(\log(n))} \right) \frac{1}{\sqrt{n}} \right\} \right) \geq 1 - o(n^{-r}). \quad (3.61)$$

Demostración. Mostraremos primero que podemos suponer sin pérdida de generalidad que fijo i los tres autovectores se orientan parecido: que $s = 1$ minimiza $\min_{s=\pm 1} \left\{ \left\| s\hat{u}_2 - \hat{u}_2^{(i)} \right\|_2 \right\}$, $\min_{s=\pm 1} \left\{ \left\| su_2 - \hat{u}_2^{(i)} \right\|_2 \right\}$ y $\min_{s=\pm 1} \left\{ \|su_2 - \hat{u}_2\|_2 \right\}$.

Supongamos primero sin pérdida de generalidad que es $s = 1$ la orientación que minimiza $\min_{s=\pm 1} \left\{ \|s\hat{u}_2 - u_2\|_\infty \right\}$ y también la que minimiza $\min_{s=\pm 1} \left\{ \left\| su_2 - \hat{u}_2^{(i)} \right\|_2 \right\}$, y veamos que luego también minimiza la restante.

Usando el Teorema 3.3.2 de Davis-Kahan tenemos que

$$\|\hat{u}_2 - u_2\|_2 \leq \frac{\|\hat{A} - A\|_2}{\Delta(A, 2)} \lesssim \frac{\sqrt{\log(n)}}{\log(n)} = \frac{1}{\sqrt{\log(n)}} \quad (3.62)$$

y también

$$\left\| \hat{u}_2^{(i)} - u_2 \right\|_2 \leq \frac{\|\hat{A}^{(i)} - A\|_2}{\Delta(A, 2)} \lesssim \frac{\sqrt{\log(n)}}{\log(n)} = \frac{1}{\sqrt{\log(n)}}. \quad (3.63)$$

Combinando ambas, tenemos que por la desigualdad triangular podemos acotar

$$\|\hat{u}_2^{(i)} - \hat{u}_2\|_2 \leq \|\hat{u}_2^{(i)} - u_2\|_2 + \|u_2 - \hat{u}_2\|_2 \lesssim \frac{1}{\sqrt{\log(n)}}. \quad (3.64)$$

Además, sabemos que para $s = \pm 1$ vale que

$$\|s\hat{u}_2^{(i)} - \hat{u}_2\|_2^2 = \|\hat{u}_2^{(i)}\|_2^2 + \|\hat{u}_2\|_2^2 - 2s(\hat{u}_2^{(i)})^t \hat{u}_2 = 2 - 2s(\hat{u}_2^{(i)})^t \hat{u}_2, \quad (3.65)$$

y como esta expresión puede ser menor que 1 solo para un valor de s y ya lo es para $s = 1$ (pues asintóticamente $\|\hat{u}_2^{(i)} - \hat{u}_2\|_2 \xrightarrow{n \rightarrow \infty} 0$) debe ser que este minimiza también la norma restante, como queríamos ver.

Mostraremos ahora que vale (3.60), suponiendo que todos los autovectores se orientan parecido en el sentido que mostramos recién.

En ese caso,

$$\begin{aligned} \left\| \hat{u}_2 - \frac{1}{\lambda_2} \hat{A} u_2 \right\|_\infty &\leq \left\| \hat{u}_2 - \frac{1}{\lambda_2} \hat{A} \hat{u}_2 \right\|_\infty + \left\| \frac{1}{\lambda_2} \hat{A} \hat{u}_2 - \frac{1}{\lambda_2} \hat{A} u_2 \right\|_\infty \\ &\leq \left\| \hat{u}_2 - \frac{\hat{\lambda}_2}{\lambda_2} \hat{u}_2 \right\|_\infty + \left\| \frac{1}{\lambda_2} \hat{A} (\hat{u}_2 - u_2) \right\|_\infty \\ &\leq \left| 1 - \frac{\hat{\lambda}_2}{\lambda_2} \right| \|\hat{u}_2\|_\infty + \left\| \frac{1}{\lambda_2} \hat{A} (\hat{u}_2 - u_2) \right\|_\infty. \end{aligned} \quad (3.66)$$

Ahora bien, recordando que los autovalores correspondientes van a ser similares podemos ver que con probabilidad mayor que $1 - o(n^{-r})$:

$$\left| 1 - \frac{\hat{\lambda}_2}{\lambda_2} \right| = \frac{|\lambda_2 - \hat{\lambda}_2|}{\lambda_2} \lesssim \frac{\sqrt{\log(n)}}{\log(n)} \asymp \frac{1}{\sqrt{\log(n)}}. \quad (3.67)$$

Adelantando que vamos a poder acotar el segundo término también en términos de $\|\hat{u}_2\|_\infty$ con probabilidad mayor que $1 - o(n^{-r})$ como a continuación

$$\left\| \frac{1}{\lambda_2} \hat{A} (\hat{u}_2 - u_2) \right\|_\infty \lesssim \frac{1}{\log(\log(n))} \|\hat{u}_2\|_\infty, \quad (3.68)$$

podremos llegar a la cota en (3.60). Efectivamente, vamos a poder reemplazar (3.67) y (3.68) en (3.66) para obtener

$$\left\| \hat{u}_2 - \frac{1}{\lambda_2} \hat{A} u_2 \right\|_\infty \lesssim \frac{1}{\sqrt{\log(n)}} \|\hat{u}_2\|_\infty + \frac{1}{\log(\log(n))} \|\hat{u}_2\|_\infty \lesssim \frac{1}{\log(\log(n))} \|\hat{u}_2\|_\infty, \quad (3.69)$$

que se parece al resultado que queremos pero todavía tenemos que reemplazar la norma $\|\hat{u}_2\|_\infty$ por $\frac{1}{\sqrt{n}}$. Afortunadamente,

$$\begin{aligned} \|\hat{u}_2\|_\infty &\leq \left\| \hat{u}_2 - \frac{1}{\lambda_2} \hat{A} u_2 \right\|_\infty + \left\| \frac{1}{\lambda_2} \hat{A} u_2 \right\|_\infty \\ &\lesssim \frac{1}{\log(\log(n))} \|\hat{u}_2\|_\infty + \frac{1}{\sqrt{n}} \end{aligned} \quad (3.70)$$

aprovechando que controlamos $\left\| \hat{u}_2 - \frac{1}{\lambda_2} \hat{A} u_2 \right\|_\infty$ por (3.69) y $\left\| \frac{1}{\lambda_2} \hat{A} u_2 \right\|_\infty$ por la Proposición 3.5.6. De este modo, despejando podremos llegar a que para alguna constante $C > 0$ vale que

$$\left(1 - C \frac{1}{\log(\log(n))} \right) \|\hat{u}_2\|_\infty \leq C \frac{1}{\sqrt{n}}, \quad (3.71)$$

y al combinarlo con (3.69) mostramos (3.60):

$$\left\| \hat{u}_2 - \frac{1}{\lambda_2} \hat{A} u_2 \right\|_\infty \lesssim \frac{1}{\log(\log(n))} \frac{1}{\sqrt{n}}. \quad (3.72)$$

Nos falta entonces solamente ver que vale la cota en (3.68) para mostrar la primera afirmación del lema, que va a ser más trabajoso. A partir de ahora acotaremos primero la distancia coordinada a coordinada con probabilidad mayor que $1 - o(n^{-(r+1)})$, y la Observación 2.2.5 nos dirá que luego con probabilidad mayor que $1 - o(n^{-r})$ todas las cotas valen simultáneamente y acotaremos la norma infinito. Fijemos entonces $i \in \mathcal{C}_1$ - la demostración es análoga para nodos en \mathcal{C}_2 - y usando la desigualdad triangular introducimos a $\hat{u}_2^{(i)}$, que sin pérdida de generalidad viene orientado correctamente:

$$\begin{aligned} \left| \left(\frac{1}{\lambda_2} \hat{A} (\hat{u}_2 - u_2) \right)_i \right| &= \left| \frac{1}{\lambda_2} \hat{A}_{i,\bullet} (\hat{u}_2 - u_2) \right| \\ &\leq \left| \frac{1}{\lambda_2} \hat{A}_{i,\bullet} (\hat{u}_2 - \hat{u}_2^{(i)}) \right| + \left| \frac{1}{\lambda_2} \hat{A}_{i,\bullet} (\hat{u}_2^{(i)} - u_2) \right| \\ &\leq \frac{1}{\lambda_2} \|\hat{A}_{i,\bullet}\|_2 \|\hat{u}_2 - \hat{u}_2^{(i)}\|_2 + \left| \frac{1}{\lambda_2} \hat{A}_{i,\bullet} (\hat{u}_2^{(i)} - u_2) \right|. \end{aligned} \quad (3.73)$$

Ahora acotaremos cada término en (3.73).

Primero, para el primer término acotaremos cada factor. En primer lugar, recordemos que $\frac{1}{\lambda_2} \asymp \frac{1}{\log(n)}$ y que $\|\hat{A}_{i,\bullet}\|_2 \lesssim \sqrt{\log(n)}$ con probabilidad mayor que $1 - o(n^{-(r+1)})$. A continuación, podemos usar el Teorema 3.3.2 de Davis-Kahan para acotar la distancia entre los autovectores:

$$\|\hat{u}_2^{(i)} - \hat{u}_2\|_2 \leq \frac{\|(\hat{A}^{(i)} - \hat{A})\hat{u}_2\|_2}{\Delta(\hat{A}, 2)}. \quad (3.74)$$

En este punto tendremos que aprovechar tener a \hat{u}_2 multiplicando dentro de la norma en vez que únicamente la norma de la resta de las matrices para conseguir una cota más fina. Para eso, notamos que un cálculo directo nos dice que por un lado $((\hat{A}^{(i)} - \hat{A})\hat{u}_2)_i = -\hat{\lambda}_2 (\hat{u}_2)_i$, usando que \hat{u}_2 es autovector y que la i -ésima fila de $\hat{A}^{(i)}$ es $\mathbf{0}$; y por otro lado también nos dice que $((\hat{A}^{(i)} - \hat{A})\hat{u}_2)_j = -\hat{A}_{j,i} (\hat{u}_2)_i$ para cada nodo $j \neq i$, porque \hat{A} coincide con $\hat{A}^{(i)}$ sobre la j -ésima fila salvo por el coeficiente en la i -ésima columna que para $\hat{A}^{(i)}$ será 0. Con las coordenadas de este vector en mente, tendremos que

$$\begin{aligned} \|(\hat{A} - \hat{A}^{(i)})\hat{u}_2\|_2 &\leq \sqrt{\hat{\lambda}_2^2 |(\hat{u}_2)_i|^2 + \sum_{j=1, j \neq i}^n \hat{A}_{j,i}^2 |(\hat{u}_2)_i|^2} \\ &\leq \sqrt{\hat{\lambda}_2^2 + \|\hat{A}_{\bullet,i}\|_2} \|\hat{u}_2\|_\infty \\ &\lesssim \log(n) \|\hat{u}_2\|_\infty, \end{aligned} \quad (3.75)$$

donde para la última desigualdad recordamos que $\hat{\lambda}_2 \asymp \log(n)$ y que al ser \hat{A} simétrica vale que $\|\hat{A}_{\bullet,i}\|_2 = \|\hat{A}_{i,\bullet}\|_2 \lesssim \sqrt{\log(n)}$ con probabilidad mayor que $1 - o(n^{-(r+1)})$. Por otro lado tenemos que $\Delta(\hat{A}, 2) \asymp \log(n)$ y así reemplazando en la cota en (3.74) obtenemos que

$$\|\hat{u}_2^{(i)} - \hat{u}_2\|_2 \lesssim \frac{\log(n) \|\hat{u}_2\|_\infty}{\log(n)} = \|\hat{u}_2\|_\infty \quad (3.76)$$

y el primer término en (3.73) cumple que con probabilidad mayor que $1 - o(n^{-(r+1)})$ vale que

$$\frac{1}{\lambda_2} \|\widehat{A}_{i,\bullet}\|_2 \|\hat{u}_2 - \hat{u}_2^{(i)}\|_2 \lesssim \frac{1}{\log(n)} \sqrt{\log(n)} \|\hat{u}_2\|_\infty = \frac{1}{\sqrt{\log(n)}} \|\hat{u}_2\|_\infty \quad (3.77)$$

que nos alcanzará.

Ahora nos concentraremos en el segundo término en (3.73). En primer lugar, usando la desigualdad triangular obtendremos que

$$\left| \frac{1}{\lambda_2} \widehat{A}_{i,\bullet}(\hat{u}_2^{(i)} - u_2) \right| \leq \left| \frac{1}{\lambda_2} A_{i,\bullet}(\hat{u}_2^{(i)} - u_2) \right| + \left| \frac{1}{\lambda_2} (\widehat{A} - A)_{i,\bullet}(\hat{u}_2^{(i)} - u_2) \right|. \quad (3.78)$$

A continuación, acotando directamente tendremos que con probabilidad mayor que $1 - o(n^{-(r+1)})$ vale que

$$\begin{aligned} \left| \frac{1}{\lambda_2} A_{i,\bullet}(\hat{u}_2^{(i)} - u_2) \right| &\leq \frac{1}{\lambda_2} \|A_{i,\bullet}\|_2 \|\hat{u}_2^{(i)} - u_2\|_2 \\ &\lesssim \frac{1}{\log(n)} \frac{\log(n)}{\sqrt{n}} \frac{1}{\sqrt{\log(n)}} \\ &\asymp \frac{1}{\sqrt{n}} \frac{1}{\sqrt{\log(n)}}, \end{aligned} \quad (3.79)$$

donde para acotar $\|\hat{u}_2^{(i)} - u_2\|_2$ usamos el Teorema 3.3.2 de Davis-Kahan:

$$\|\hat{u}_2^{(i)} - u_2\|_2 \leq \frac{\|\widehat{A}^{(i)} - A\|_2}{\Delta(A, 2)} \lesssim \frac{\sqrt{\log(n)}}{\log(n)} = \frac{1}{\sqrt{\log(n)}}. \quad (3.80)$$

Para acotar el término restante en (3.78) necesitaremos argumentos más sutiles. Primero notemos que como $\hat{u}_2^{(i)}$ es autovector para la matriz $\widehat{A}^{(i)}$ que es independiente de la fila \widehat{A}_i , también $\hat{u}_2^{(i)}$ lo será. Luego, mostraremos que para cada vector fijo \hat{u}_2^i podemos acotar con probabilidad más grande que $1 - o(n^{-(r+1)})$ de la siguiente manera

$$\left| \frac{1}{\lambda_2} (\widehat{A} - A)_{i,\bullet}(\hat{u}_2^i - u_2) \right| \lesssim \frac{1}{\log(\log(n))} \|\hat{u}_2\|_\infty, \quad (3.81)$$

donde podremos suponer sin pérdida de generalidad que $\hat{u}_2^{(i)}$ es constante al acotar usando la independencia con respecto a $\widehat{A}_{i,\bullet}$. En efecto, tenemos que si llamamos $\mathbb{1}_{\mathcal{A}}$ a la indicadora del evento en el que ocurre la desigualdad en (3.81) para $\widehat{A}_{i,\bullet}$ y \hat{u}_2^i , tendremos que

$$\begin{aligned} \mathbb{P}(\mathcal{A}) &= \mathbb{E}_{(\widehat{A}_{i,\bullet}, \hat{u}_2^{(i)})} \left[\mathbb{1}_{\mathcal{A}}(\widehat{A}_i, \hat{u}_2^{(i)}) \right] \\ &= \mathbb{E}_{\hat{u}_2^{(i)}} \left[\mathbb{E}_{\widehat{A}_{i,\bullet}} \left[\mathbb{1}_{\mathcal{A}}(\widehat{A}_i, \hat{u}_2^{(i)}) \right] \right] \\ &\geq \mathbb{E}_{\hat{u}_2^{(i)}} \left[1 - o(n^{-(r+1)}) \right] \\ &= 1 - o(n^{-(r+1)}), \end{aligned} \quad (3.82)$$

donde usamos los subíndices en \mathbb{E} para denotar con respecto a qué elemento aleatorio tomamos esperanza. Ahora bien, usando el Lema 3.5.4 tenemos que para $\eta = \frac{r+2}{a}$, $w = \hat{u}_2^{(i)} - u_2$, $p = a \frac{\log(n)}{n}$ y $\gamma = \frac{1}{\sqrt{\log(n)}}$, con probabilidad mayor que $1 - o(n^{-(r+1)})$ - pues $-\eta np = -(r+2) \log(n)$ y así $1 - 2e^{-\eta np} = 1 - 2e^{-(r+2) \log(n)}$ - vale que:

$$\begin{aligned} \left| \sum_{j=1}^n \left((\hat{u}_2^{(i)})_j - (u_2)_j \right) (\hat{A}_{i,j} - A_{i,j}) \right| &= \left| \sum_{j=1}^n w_j (\hat{A}_{i,j} - A_{i,j}) \right| & (3.83) \\ &\lesssim (2 + \eta) np \frac{\max \left\{ \|w\|_\infty, \frac{1}{\gamma \sqrt{n}} \|w\|_2 \right\}}{\max \left\{ 1, \log \left(\frac{1}{\gamma} \right) \right\}} \\ &\lesssim \log(n) \frac{\max \left\{ \left\| \hat{u}_2^{(i)} - u_2 \right\|_\infty, \frac{1}{\sqrt{n}} \right\}}{\log(\log(n))}, \end{aligned}$$

donde recordamos que como vimos en (3.80) vale que $\left\| \hat{u}_2^{(i)} - u_2 \right\|_2 \lesssim \frac{1}{\sqrt{\log(n)}} = \gamma$. A continuación, queremos deshacernos de la norma infinito en la cota, para lo cual procederemos del siguiente modo:

$$\begin{aligned} \left\| \hat{u}_2^{(i)} - u_2 \right\|_\infty &\leq \left\| \hat{u}_2^{(i)} - \hat{u}_2 \right\|_\infty + \left\| \hat{u}_2 - u_2 \right\|_\infty & (3.84) \\ &\leq \left\| \hat{u}_2^{(i)} - \hat{u}_2 \right\|_2 + \left\| \hat{u}_2 \right\|_\infty + \left\| u_2 \right\|_\infty \\ &\lesssim \left\| \hat{u}_2 \right\|_\infty. \end{aligned}$$

Aclaremos que la última cota se obtiene usando que $\left\| \hat{u}_2^{(i)} - \hat{u}_2 \right\|_2 \lesssim \left\| \hat{u}_2 \right\|_\infty$ como vimos en (3.76), y que como $\left\| u_2 \right\|_\infty = \frac{1}{\sqrt{n}}$ esta es mínima sobre los vectores que además tienen norma euclídea 1 y entonces tendremos que inmediatamente $\left\| u_2 \right\|_\infty = \frac{1}{\sqrt{n}} \leq \left\| \hat{u}_2 \right\|_\infty$. Luego, combinando (3.84) con nuestra cota en (3.83), obtenemos que

$$\left| \sum_{j=1}^n \left((\hat{u}_2^{(i)})_j - (u_2)_j \right) (\hat{A}_{i,j} - A_{i,j}) \right| \lesssim \log(n) \frac{\left\| \hat{u}_2 \right\|_\infty}{\log(\log(n))}, \quad (3.85)$$

y entonces recordando que $\lambda_2 \asymp \log(n)$ deducimos que

$$\left| \frac{1}{\lambda_2} (\hat{A} - A)_{i,\bullet} (\hat{u}_2^i - u_2) \right| \lesssim \frac{1}{\log(n)} \left(\log(n) \frac{\left\| \hat{u}_2 \right\|_\infty}{\log(\log(n))} \right) = \frac{1}{\log(\log(n))} \left\| \hat{u}_2 \right\|_\infty \quad (3.86)$$

como queríamos ver.

Volviendo ahora a (3.78) y usando las cotas que obtuvimos en (3.79) y (3.86), tenemos que entonces

$$\left| \frac{1}{\lambda_2} \hat{A}_i (\hat{u}_2^{(i)} - \hat{u}_2) \right| \lesssim \frac{1}{\sqrt{n}} \frac{1}{\sqrt{\log(n)}} + \frac{1}{\log(\log(n))} \left\| \hat{u}_2 \right\|_\infty \lesssim \frac{1}{\log(\log(n))} \left\| \hat{u}_2 \right\|_\infty \quad (3.87)$$

donde recordamos que por minimalidad entre las normas infinito de vectores con norma 2 igual a 1 vale que $\frac{1}{\sqrt{n}} \leq \left\| \hat{u}_2 \right\|_\infty$.

Como adelantamos, esto muestra la cota que queríamos en (3.68) pues acotamos para cada nodo individualmente con probabilidad mayor que $1 - o(n^{-(r+1)})$, y así completamos la demostración de la cota (3.60).

Solo resta ahora demostrar la cota (3.61). Podremos hacerlo rápidamente con todo el trabajo que ya realizamos: combinando (3.87) y (3.71) obtenemos

$$\left| \frac{1}{\lambda_2} \widehat{A}_i(\widehat{u}_2^{(i)} - \widehat{u}_2) \right| \lesssim \frac{1}{\log(\log(n))} \|\widehat{u}_2\|_\infty \lesssim \frac{1}{\log(\log(n))} \frac{1}{\sqrt{n}}, \quad (3.88)$$

que es precisamente lo que queríamos ver. ■

Ahora sí, podremos concluir finalmente que el estimador espectral, y también el estimador espectral por vecindades, son fuertemente consistentes.

[3.5.8] Teorema: El estimador espectral \widehat{C}_{SP} y el estimador espectral por vecindades son fuertemente consistentes.

Sea $G_C \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos donde $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para ciertos $a > b > 0$ números positivos que cumplen además que $\frac{a+b}{2} > 1 + \sqrt{ab}$. Luego, los estimadores \widehat{C}_{SP} y \widehat{C}_{SPV} son fuertemente consistentes.

Demostración. Sabemos por la Proposición 3.5.6 que el estimador \widehat{C}_{SPA} es fuertemente consistente, es decir que $\phi(\text{signo}((\frac{1}{\lambda_2} \widehat{A}u_2)_i)) = \mathcal{C}(i)$ para cada nodo con probabilidad alta.

Bastará ver entonces que $(\frac{1}{\lambda_2} \widehat{A}u_2^{(i)})_i$ y $(\widehat{u}_2)_i$ comparten el signo con $(\frac{1}{\lambda_2} \widehat{A}u_2)_i$ con alta probabilidad para todo i , suponiendo que los vectores están orientados similarmente como en la Proposición 3.5.7... pero es precisamente lo que este resultado asegura. En efecto, como $(\frac{1}{\lambda_2} \widehat{A}u_2)_i \gtrsim \frac{1}{\sqrt{n}}$ con alta probabilidad por la Proposición 3.5.6 y además tenemos que tanto $\left| (\frac{1}{\lambda_2} \widehat{A}u_2^{(i)})_i - (\frac{1}{\lambda_2} \widehat{A}u_2)_i \right| \lesssim \frac{1}{\log(\log(n))} \frac{1}{\sqrt{n}}$ como $|\widehat{u}_2)_i - (\frac{1}{\lambda_2} \widehat{A}u_2)_i| \lesssim \frac{1}{\log(\log(n))} \frac{1}{\sqrt{n}}$ con alta probabilidad, será imposible que para n suficientemente grandes valga que los signos difieran, lo que completa la demostración. ■

Queda entonces demostrada la consistencia fuerte para los estimadores espectrales, y también la implicación restante en el Teorema 3.1.4. Con esto terminaremos este capítulo, para dar paso a explorar en los siguientes algunos otros estimadores y sus propiedades.

3.6 Notas

[3.6.1] Nota: Otras nociones de consistencia y límites para el caso general

Es posible estudiar en qué circunstancias se puede lograr algún criterio más relajado de precisión, como **detección débil** o **distinguibilidad**. Referimos a [Abb18] para más detalles.

Por otro lado, también se puede estudiar para qué valores de los parámetros se tiene consistencia fuerte en el caso general. Esto es nuevamente tratado en [Abb18] y también en [CL23]. Interesantemente en [Abb18] se discuten otras alternativas para demostrar el Teorema 3.1.4 y llegar a los límites teóricos de la estimación, como unas cotas directas y la técnica de “partir el grafo” que motiva los estimadores por vecindades y nos será de utilidad en el Capítulo 5.

[3.6.2] Nota: Extensión a k comunidades para algoritmos espectrales

Los métodos espectrales pueden generalizarse para hacer detección de comunidades en un contexto más general, cuando hay una cantidad arbitraria k de comunidades. En este caso una forma de generalizarlos es utilizar no solo el segundo autovector si no que todos los autovectores entre el segundo y el k -ésimo, y al realizar k -medias sobre el conjunto de vectores asociados a cada nodo podremos separarlos en k subconjuntos que aproximarán las comunidades. Podemos encontrar un tratamiento riguroso de esto en [AFWZ20, Sección C.4.], donde se muestra que de tener convergencia en k -medias a centros razonables el estimador será fuertemente consistente para 3 comunidades.

En la Figura 3.3 podemos ver las coordenadas en el autoespacio generado por el segundo y el tercer autovector de cada fila de la matriz de adyacencia empírica \hat{A} , donde están pintados de distinto color según son clasificados por el algoritmo de 3-medias; los puntos de color celeste son justamente las medias halladas.

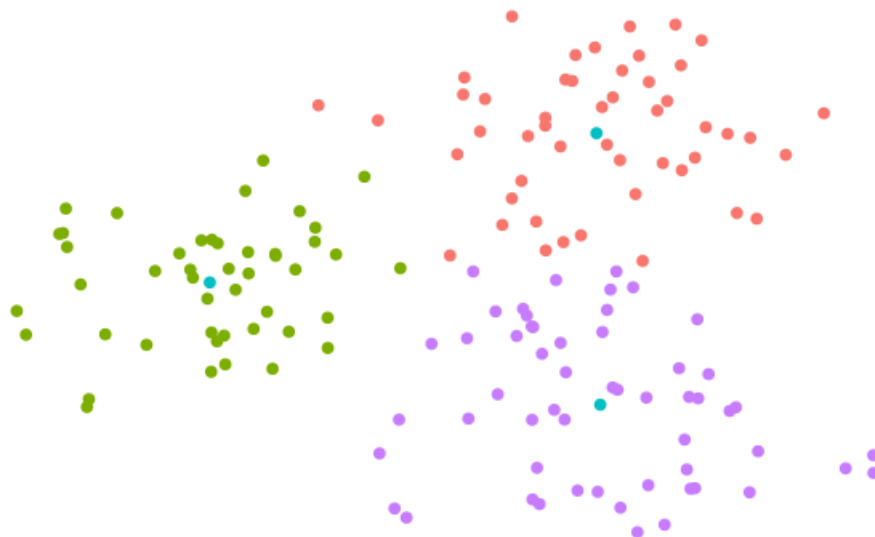


Figura 3.3: El método espectral para 3 comunidades balanceadas.

[3.6.3] Nota: Estimadores fuertemente consistentes en la literatura

Desde ya, los estimadores presentados en este capítulo no son los únicos que logran estimar de manera fuertemente consistente las comunidades. Recordando lo mencionado en la introducción, algunas otras avenidas para atacar el problema son: relajaciones a problemas SDP, como en [Abb18]; maximización de modularidad, como en [CL23]; y estimadores que usan la técnica de “partir el grafo” (graph splitting), como en [Abb18, MNS16].

[3.6.4] Nota: Detección de comunidades en modelos más generales

Aún cuando mejora por sobre los grafos de Erdős-Renyi, el modelo estocástico de bloques es a veces demasiado restrictivo para aplicaciones en la vida real, en que incluso dentro de las diversas comunidades cada individuo presenta algunas características propias. Hay modelos más generales que pueden usarse para capturar el comportamiento individual de cada nodo, entre los cuales podemos mencionar por ejemplo a los grafos aleatorios via productos internos

usados en [BFMM23], que al elegir vectores idénticos para nodos de la misma comunidad recuperan el modelo SBM.

4. Estimadores de minimización de corte

4.1 Cortes balanceados y estimadores de minimización de corte

En el Capítulo 3 mencionamos ya el concepto de corte para una partición de comunidades $\tilde{\mathcal{C}}$ en un grafo $\tilde{G} = (\tilde{V}, \tilde{E})$

$$\text{Corte}(\tilde{\mathcal{C}}) = \# \{e = \{i, j\} \in \tilde{E} \mid \tilde{\mathcal{C}}(i) \neq \tilde{\mathcal{C}}(j)\} \quad (4.1)$$

cuando en la Definición 3.2.3 introdujimos el estimador de mínima bisección para modelos BSBM con dos comunidades:

$$\hat{\mathcal{C}}_{\text{CUT}} \text{ minimiza } \{ \text{Corte}(\tilde{\mathcal{C}}) \} \text{ sobre particiones balanceadas.} \quad (4.2)$$

Recordaremos luego que minimizar cortes es una estrategia intuitiva para detectar las comunidades, pues en el caso asortativo estas tienen una gran densidad de aristas entre sus miembros y entonces separar nodos de la misma comunidad real en distintas comunidades estimadas aportará una gran cantidad de aristas al corte; podremos convencerlos también gráficamente por medio de la Figura 3.2.

Más aún, en el modelo balanceado la estimación además de ser intuitiva es certera. En efecto, al coincidir con el estimador de máximo a posteriori, el estimador de mínima bisección será fuertemente consistente cuando sea posible conseguir estimadores fuertemente consistentes:

[4.1.1] Observación: El estimador de corte mínimo $\hat{\mathcal{C}}_{\text{CUT}}$ es fuertemente consistente en el modelo BSBM asortativo

En efecto, vimos en el Teorema 3.5.8 que el estimador $\hat{\mathcal{C}}_{\text{SP}}$ es fuertemente consistente para modelos BSBM $(n, 2, p_n, q_n)$ donde $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ con $a > b$ y $\frac{a+b}{2} > 1 + \sqrt{ab}$. Luego, como notamos en la Observación 3.2.2, valdrá automáticamente que el estimador máximo a posteriori $\hat{\mathcal{C}}_{\text{MAP}}$ será también fuertemente consistente. Más aún, como el estimador de corte mínimo y el estimador de máximo a posteriori coinciden en estos modelos en el caso asortativo, como vimos en la Observación 3.2.4, también valdrá que el estimador de mínima bisección será fuertemente consistente.

Sin embargo, esta garantía tiene algunas limitaciones.

En primer lugar, una limitación técnica: pudimos demostrar consistencia fuerte solo para un modelo donde la asignación real de comunidades es estrictamente balanceada, con mitad de los nodos en cada comunidad. Afortunadamente, estudiando con cuidado la Observación 3.2.4 podremos salvarla siempre que conozcamos la cantidad de nodos en cada comunidad y nos restrinjamos a particiones en comunidades que preserven dichos tamaños, pues en este caso seguiremos siendo capaces de demostrar que coincide con el estimador máximo a posteriori.

En segundo lugar, una limitación esencial y que ya no podremos salvar: tendremos problemas al no conocer los tamaños de las comunidades, lo que no nos permitiría restringir las clases de particiones sobre las que vamos a minimizar. Haremos esta noción precisa en la siguiente observación:

[4.1.2] **Observación: Limitación del corte de mínima bisección**

Consideremos $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ donde $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$, con $a > b$ reales positivos. Luego, la asignación de comunidades dada por

$$\tilde{\mathcal{C}}(i) = \begin{cases} 1 & \text{si } i = 1, \text{ y} \\ 2 & \text{si no.} \end{cases} \quad (4.3)$$

cumple que

$$\mathbb{E} [\text{Corte}(\tilde{\mathcal{C}})] = \mathbb{E} \left[\frac{\text{deg}(1)}{1} \right] = \left(\frac{n}{2} - 1 \right) p_n + \frac{n}{2} q_n \asymp n \frac{\log(n)}{n} = \log(n), \quad (4.4)$$

pues las aristas que cuenta el corte son las que conectan al nodo 1 con el resto.

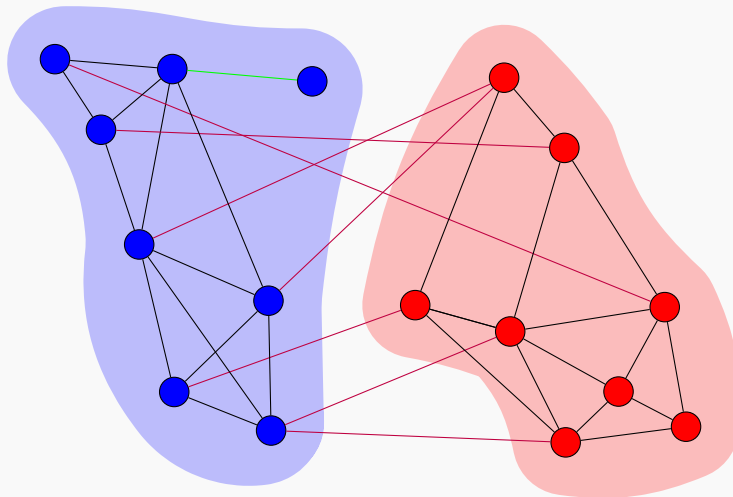


Figura 4.1: El estimador por mínima bisección debe ser restringido para no caer en malos casos.

Sin embargo, el corte de las comunidades reales cuenta a todas las aristas entre nodos de distintas comunidades, que serán más. En efecto,

$$\mathbb{E} [\text{Corte}(\mathcal{C})] = \frac{1}{2} \frac{n}{2} \left(\frac{n}{2} - 1 \right) q_n \asymp n^2 \frac{\log(n)}{n} = n \log(n), \quad (4.5)$$

con lo cual $\mathbb{E} [\text{Corte}(\hat{\mathcal{C}})] < \mathbb{E} [\text{Corte}(\mathcal{C})]$ para valores suficientemente grandes de n y no tendremos esperanzas de recuperar la partición real a partir de un minimizante. En la Figura 4.1 vemos un ejemplo concreto de esta situación, donde esta vez la partición que deja al nodo 5 solo tiene corte 1, que es menor que el corte asociado a la partición por las comunidades reales, en rojo y azul, de 7.

Es así que, como habíamos adelantado en el Capítulo 3, exploraremos ahora en más detalle cómo estimar las comunidades minimizando algunas variantes de cortes para particiones en el grafo.

Como moraleja de la Observación 4.1.2 sabremos que la noción de corte falla en que no tiene en cuenta el tamaño de las comunidades, que ya notamos en la Observación 2.4.4 que casi siempre será lineal. Podremos evitar particiones con comunidades de tamaño sublineal mejorando la noción de corte: por ejemplo agregando factores que penalicen tener comunidades muy chicas; o si no minimizando el corte sobre particiones en donde ninguna comunidad sea demasiado chica.

Aquí nos enfocaremos en la primera alternativa, que tendrá la consecuencia deseable que no tendremos que restringir las particiones sobre las cuales minimizamos, excepto que deseemos hacerlo. Más precisamente, estudiaremos el corte de Cheeger y el corte “ratio”:

[4.1.3] Definición: Corte de Cheeger y corte ratio - [HS11, Sección 3.]

Sea \tilde{G} un grafo con n nodos y consideremos $\tilde{\mathcal{C}}$ una partición en dos comunidades, digamos 1 y 2, para \tilde{G} . Luego, diremos que

- el **corte de Cheeger** de \tilde{G} está dado por:

$$\text{RCC}(\tilde{\mathcal{C}}) = \frac{\text{Corte}(\tilde{\mathcal{C}})}{\min\{\#\tilde{\mathcal{C}}_1, \#\tilde{\mathcal{C}}_2\}} \quad (4.6)$$

- y el **corte ratio** de \tilde{G} está dado por:

$$\text{RC}(\tilde{\mathcal{C}}) = n \frac{\text{Corte}(\tilde{\mathcal{C}})}{\#\tilde{\mathcal{C}}_1 \cdot \#\tilde{\mathcal{C}}_2} = \text{Corte}(\tilde{\mathcal{C}}) \left(\frac{1}{\#\tilde{\mathcal{C}}_1} + \frac{1}{\#\tilde{\mathcal{C}}_2} \right). \quad (4.7)$$

Dada $G_{\mathcal{C}}$ una familia de modelos estocásticos de bloques con dos comunidades denotaremos respectivamente $\hat{\mathcal{C}}_{\text{RCC}}$ y $\hat{\mathcal{C}}_{\text{RC}}$ a los estimadores que minimicen RCC y RC sobre todas las asignaciones $\tilde{\mathcal{C}}$ en estas dos comunidades 1 y 2 para \tilde{G} .

Remarcamos que, al dividir por el mínimo cardinal entre los conjuntos de la partición o por el producto de ambos tamaños, penalizamos comunidades chicas, que nos ayudará a evitar los problemas que discutimos en la observación 4.1.

Más aún, resulta que esta penalización es suficiente, y en la literatura se ha logrado dar resultados de consistencia para estos estimadores, pero hasta donde encontramos apelando

a métodos analíticos más que a recursos netamente probabilísticos, como elaboraremos en la Nota 4.4.1. Nuestro objetivo será, entonces, explicar con herramientas probabilísticas por qué estos estimadores son certeros.

Para comenzar, y aunque el caso interesante será cuando no nos restrinjamos a minimizar solamente sobre comunidades con tamaños fijos, podremos verificar que en este caso ya sabemos que serán fuertemente consistentes:

[4.1.4] **Observación:** Los estimadores por minimización de corte de Cheeger y corte ratio son equivalentes al de mínima bisección en el modelo BSBM sobre comunidades balanceadas

En efecto, si $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$, y $\tilde{\mathcal{C}}$ es una asignación de comunidades balanceada, tendremos que

$$\text{Corte}(\tilde{\mathcal{C}}) = \frac{n}{2} \text{RCC}(\tilde{\mathcal{C}}) = \frac{n}{2} \text{RC}(\tilde{\mathcal{C}}), \quad (4.8)$$

y así en particular los tres cortes se minimizan para las mismas particiones balanceadas. Como consecuencia, cuando $\hat{\mathcal{C}}_{\text{CUT}}$ es fuertemente consistente, por ejemplo en los casos que contempla la Observación 4.1.1, $\hat{\mathcal{C}}_{\text{RCC}}$ y $\hat{\mathcal{C}}_{\text{RC}}$ también lo son.

Antes de pasar al caso general, fijaremos un poco más de notación que usaremos durante el resto del capítulo:

[4.1.5] **Notación:** Tamaño y solapamientos para las comunidades estimadas

En cada grafo con n nodos y dos comunidades \mathcal{C} , digamos 1 y 2, para cada partición $\tilde{\mathcal{C}}$ en dos comunidades, notaremos:

1. $m_l = \#\tilde{\mathcal{C}}_l$ y $\beta_l = \frac{m_l}{n}$ para cada $1 \leq l \leq 2$ al cardinal y a la proporción que ocupa cada comunidad en la partición, y
2. $X_{i,j} = \#(\mathcal{C}_i \cap \tilde{\mathcal{C}}_j)$ y $x_{i,j} = \frac{X_{i,j}}{n}$ para cada $1 \leq i, j \leq 2$ al cardinal y a las proporciones que corresponden a cada solapamiento entre las particiones.

Recordemos que ya tenemos notación similar para \mathcal{C} (como n_i y α_i), que introducimos en 2.4.2.

Ahora sí, en la siguiente sección ya sin más preámbulos pasaremos a mostrar la consistencia fuerte para los estimadores. Remarcamos que el trabajo realizado aquí es original, aunque como mencionamos en la Nota 4.4.1 algunos resultados de consistencia para los cortes pueden conseguirse mediante otros enfoques.

4.2 Consistencia para el minimizante del corte de Cheeger

Como adelantamos, buscaremos en esta sección llegar al siguiente resultado de consistencia fuerte para el corte de Cheeger:

[4.2.1] Teorema: Consistencia fuerte de una realización de corte de Cheeger mínima

Sea $G \sim \text{BSBM} \left(n, 2, p_n = a \frac{\log(n)}{n}, q_n = b \frac{\log(n)}{n} \right)$ tal que $a > 2b$. Luego, si $a - 2b$ es suficientemente grande, vale que $\hat{\mathcal{C}}_{\text{RCC}}$ es un estimador fuertemente consistente de \mathcal{C} .

Remarcamos que, aunque es en el modelo balanceado, la mejora que tendremos comparando contra el estimador de mínima bisección es que no necesitamos restringir las particiones de comunidades sobre las cuales minimizamos para obtener nuestra partición.

La idea central para la demostración será que individualmente cualquier estimación fija que proponamos de las comunidades va a tener un corte de Cheeger más grande que el de las comunidades reales. Una primera manifestación concreta de este hecho es el siguiente lema, en el que mostramos que el valor esperado del corte de Cheeger de las comunidades reales es inferior que el de cualquier otra estimación fija para las comunidades:

[4.2.2] Lema: Cálculos de esperanza y varianza

Sea $G \sim \text{SBM} \left(n, 2, (n_i)_{1 \leq i \leq 2}, p_n = a \frac{\log(n)}{n}, q_n = b \frac{\log(n)}{n} \right)$, y sea $\tilde{\mathcal{C}}$ una partición de comunidades para G .

Luego, vale que

$$\begin{aligned} \mathbb{E} \left[\text{RCC}(\tilde{\mathcal{C}}) - \text{RCC}(\mathcal{C}) \right] &= (\text{máx}\{m_1, m_2\} - \text{máx}\{n_1, n_2\}) q_n & (4.9) \\ &+ \left(\frac{X_{1,1}X_{1,2} + X_{2,1}X_{2,2}}{\text{mín}\{m_1, m_2\}} \right) (p_n - q_n) \\ &= (\text{máx}\{\beta_1, \beta_2\} - \text{máx}\{\alpha_1, \alpha_2\}) n q_n \\ &+ \left(\frac{x_{1,1}x_{1,2} + x_{2,1}x_{2,2}}{\text{mín}\{\beta_1, \beta_2\}} \right) n (p_n - q_n) \end{aligned}$$

y

$$\begin{aligned} \text{Var} \left[\text{RCC}(\tilde{\mathcal{C}}) - \text{RCC}(\mathcal{C}) \right] & & (4.10) \\ &= \left(\frac{1}{\text{mín}\{m_1, m_2\}} - \frac{1}{\text{mín}\{n_1, n_2\}} \right)^2 (X_{1,1}X_{2,2} + X_{1,2}X_{2,1}) (q_n(1 - q_n)) \\ &+ \left(\frac{1}{\text{mín}\{n_1, n_2\}} \right)^2 (X_{1,1}X_{2,1} + X_{1,2}X_{2,2}) (q_n(1 - q_n)) \\ &+ \left(\frac{1}{\text{mín}\{m_1, m_2\}} \right)^2 (X_{1,1}X_{1,2} + X_{2,1}X_{2,2}) (p_n(1 - p_n)) \\ &= \left(\frac{1}{\text{mín}\{\beta_1, \beta_2\}} - \frac{1}{\text{mín}\{\alpha_1, \alpha_2\}} \right)^2 (x_{1,1}x_{2,2} + x_{1,2}x_{2,1}) (q_n(1 - q_n)) \\ &+ \left(\frac{1}{\text{mín}\{\alpha_1, \alpha_2\}} \right)^2 (x_{1,1}x_{2,1} + x_{1,2}x_{2,2}) (q_n(1 - q_n)) \\ &+ \left(\frac{1}{\text{mín}\{\beta_1, \beta_2\}} \right)^2 (x_{1,1}x_{1,2} + x_{2,1}x_{2,2}) (p_n(1 - p_n)). \end{aligned}$$

Más aún, vale que tanto para el caso balanceado $n_1 = n_2$ como cuando $a > 2b$,

$$\mathbb{E} [\text{RCC}(\tilde{\mathcal{C}}) - \text{RCC}(\mathcal{C})] > 0 \quad (4.11)$$

para cualquier partición $\tilde{\mathcal{C}}$ que no induzca la misma partición que \mathcal{C} .

Demostración. En primer lugar, calculemos la esperanza de la diferencia. Esta es combinación lineal de variables Bernoulli que nos indican la presencia de aristas, escaladas por los coeficientes que aparecen en el corte, como podemos apreciar en la Figura 4.2. Allí, el cuadrado es una especie de diagrama de Venn en el que podemos colocar cada nodo: las aristas azules formarán parte de ambos cortes, las rojas solo del corte asociado a las comunidades reales \mathcal{C} y las verdes solo al corte asociado a las comunidades estimadas $\tilde{\mathcal{C}}$.

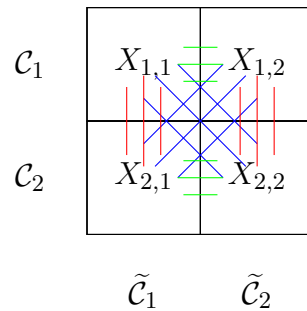


Figura 4.2: Un diagrama para visualizar la diferencia entre los cortes asociados a \mathcal{C} y $\tilde{\mathcal{C}}$.

En consecuencia, podemos usar la linealidad de la esperanza para obtener:

$$\begin{aligned} & \mathbb{E} [\text{RCC}(\tilde{\mathcal{C}}) - \text{RCC}(\mathcal{C})] \quad (4.12) \\ &= \left(\frac{1}{\min\{m_1, m_2\}} - \frac{1}{\min\{n_1, n_2\}} \right) (X_{1,1}X_{2,2} + X_{1,2}X_{2,1}) q_n \\ & - \left(\frac{1}{\min\{n_1, n_2\}} \right) (X_{1,1}X_{2,1} + X_{1,2}X_{2,2}) q_n \\ & + \left(\frac{1}{\min\{m_1, m_2\}} \right) (X_{1,1}X_{1,2} + X_{2,1}X_{2,2}) p_n \\ &= \left(\frac{1}{\min\{\beta_1, \beta_2\}} - \frac{1}{\min\{\alpha_1, \alpha_2\}} \right) (x_{1,1}x_{2,2} + x_{1,2}x_{2,1}) nq_n \\ & - \left(\frac{1}{\min\{\alpha_1, \alpha_2\}} \right) (x_{1,1}x_{2,1} + x_{1,2}x_{2,2}) nq_n \\ & + \left(\frac{1}{\min\{\beta_1, \beta_2\}} \right) (x_{1,1}x_{1,2} + x_{2,1}x_{2,2}) np_n. \end{aligned}$$

Una forma alternativa de expresar esta esperanza es la siguiente:

$$\begin{aligned} \mathbb{E} [\text{RCC}(\tilde{\mathcal{C}}) - \text{RCC}(\mathcal{C})] &= \left(\frac{m_1 m_2}{\min\{m_1, m_2\}} - \frac{n_1 n_2}{\min\{n_1, n_2\}} \right) q_n \\ &\quad + \left(\frac{X_{1,1} X_{1,2} + X_{2,1} X_{2,2}}{\min\{m_1, m_2\}} \right) (p_n - q_n) \\ &= \left(\frac{\beta_1 \beta_2}{\min\{\beta_1, \beta_2\}} - \frac{\alpha_1 \alpha_2}{\min\{\alpha_1, \alpha_2\}} \right) n q_n \\ &\quad + \left(\frac{x_{1,1} x_{1,2} + x_{2,1} x_{2,2}}{\min\{\beta_1, \beta_2\}} \right) n (p_n - q_n). \end{aligned} \quad (4.13)$$

En efecto, la expresión se obtiene a partir de la anterior sumando y restando $\left(\frac{X_{1,1} X_{1,2} + X_{2,1} X_{2,2}}{\min\{m_1, m_2\}} \right) q_n$ (o el análogo con proporciones), que puede interpretarse como sumar y restar aristas entre las parejas de nodos en $\tilde{\mathcal{C}}_1$ y $\tilde{\mathcal{C}}_2$ que viven en la misma comunidad, pero ahora con probabilidades q_n independientes. Obtenemos entonces la expresión en (4.9) que queríamos al notar que $\frac{\beta_1 \beta_2}{\min\{\beta_1, \beta_2\}} = \max\{\beta_1, \beta_2\}$ y análogamente $\frac{\alpha_1 \alpha_2}{\min\{\alpha_1, \alpha_2\}} = \max\{\alpha_1, \alpha_2\}$.

Así, solamente resta demostrar que la esperanza es no negativa y es además positiva bajo las hipótesis adecuadas. Primero supongamos que $a > 2b$, y usando que valdrá que

$$\frac{x_{1,1} x_{1,2} + x_{2,1} x_{2,2}}{\min\{\beta_1, \beta_2\}} \geq |\max\{\beta_1, \beta_2\} - \max\{\alpha_1, \alpha_2\}|, \quad (4.14)$$

tendremos que como $p_n > 2q_n$ obtendremos a partir de (4.9) que la esperanza es no negativa

$$\begin{aligned} \mathbb{E} [\text{RCC}(\tilde{\mathcal{C}}) - \text{RCC}(\mathcal{C})] &\geq (\max\{\beta_1, \beta_2\} - \max\{\alpha_1, \alpha_2\}) n q_n \\ &\quad + |\max\{\beta_1, \beta_2\} - \max\{\alpha_1, \alpha_2\}| n (p_n - q_n) \\ &\geq (\max\{\beta_1, \beta_2\} - \max\{\alpha_1, \alpha_2\}) n q_n \\ &\quad + |\max\{\beta_1, \beta_2\} - \max\{\alpha_1, \alpha_2\}| n q_n \\ &\geq 0. \end{aligned} \quad (4.15)$$

Más aún, con un análisis más minucioso podremos mostrar que bajo las hipótesis adecuadas la esperanza es positiva. En efecto, suponiendo además que \mathcal{C} y $\tilde{\mathcal{C}}$ inducen particiones distintas: si $\max\{\beta_1, \beta_2\} - \max\{\alpha_1, \alpha_2\} \geq 0$ luego como $p_n > q_n$ y además $\frac{x_{1,1} x_{1,2} + x_{2,1} x_{2,2}}{\min\{\beta_1, \beta_2\}} \neq 0$ (porque debe haber algún nodo mal clasificado) obtenemos la desigualdad estricta a partir de (4.9); y si $\max\{\beta_1, \beta_2\} - \max\{\alpha_1, \alpha_2\} < 0$ entonces la segunda desigualdad en (4.15) es estricta usando que $p_n > 2q_n$ por $a > 2b$, que muestra lo afirmado.

Por último, en el caso balanceado vale necesariamente que $\max\{\beta_1, \beta_2\} \geq \frac{1}{2} = \max\{\alpha_1, \alpha_2\}$ según lo cual también análogamente la esperanza es no negativa a partir de (4.9) y es además positiva cuando se verifica que coincidencia $(\tilde{\mathcal{C}}, \mathcal{C}) \neq n$ en cuyo caso $\frac{x_{1,1} x_{1,2} + x_{2,1} x_{2,2}}{\min\{\beta_1, \beta_2\}} \neq 0$.

Queda entonces solamente mostrar (4.14) para completar las demostraciones relativas a la esperanza, en donde para simplificar la notación vamos a suponer sin pérdida de generalidad que $\alpha_1 = \min\{\alpha_1, \alpha_2\}$ y que $\beta_1 = \min\{\beta_1, \beta_2\}$.

Si $\beta_2 \geq \alpha_2$ entonces mostrar la desigualdad (4.14) equivale a demostrar que

$$x_{1,1} x_{1,2} + x_{2,1} x_{2,2} \geq \beta_1 (\beta_2 - \alpha_2), \quad (4.16)$$

donde multiplicamos por β_1 para no tener más cocientes. Pero equivalentemente podríamos demostrar que

$$x_{1,1}(\alpha_1 - x_{1,1}) + x_{2,1}(1 - \alpha_1 - x_{2,1}) \geq (x_{1,1} + x_{2,1})(\alpha_1 - x_{1,1} - x_{2,1}), \quad (4.17)$$

que distribuyendo y reacomodando los términos podemos ver que equivale a mostrar que

$$x_{2,1}(1 - 2\alpha_1 + 2x_{1,1}) \geq 0, \quad (4.18)$$

que ocurre pues $x_{2,1} \geq 0$ y $1 - 2\alpha_1 + 2x_{1,1} \geq 1 - 2\alpha_1 \geq 0$ al ser $\alpha_1 + \alpha_2 = 1$ y ser α_1 el mínimo.

Si en cambio $\alpha_2 \geq \beta_2$, entonces mostrar la desigualdad equivale a ver que

$$x_{1,1}x_{1,2} + x_{2,1}x_{2,2} \geq \beta_1(\alpha_2 - \beta_2), \quad (4.19)$$

donde nuevamente multiplicamos por β_1 para no tener más cocientes. Una vez más equivalentemente podríamos demostrar que

$$x_{1,1}(\alpha_1 - x_{1,1}) + x_{2,1}(1 - \alpha_1 - x_{2,1}) \geq (x_{1,1} + x_{2,1})(x_{1,1} + x_{2,1} - \alpha_1), \quad (4.20)$$

que distribuyendo y reacomodando los términos podemos ver que equivale a mostrar que

$$2x_{1,1}(\alpha_1 - x_{1,1}) + x_{2,1}(1 - 2x_{1,1} - 2x_{2,1}) \geq 0, \quad (4.21)$$

que ocurre pues $x_{1,1}, x_{2,1} \geq 0$, $\alpha_1 - x_{1,1} = x_{1,2} \geq 0$ y $1 - 2x_{1,1} - 2x_{2,1} = 1 - 2\beta_1 \geq 0$ al ser $\beta_1 + \beta_2 = 1$ y ser β_1 el mínimo.

Finalmente, podremos hallar la fórmula para la varianza en (4.10) mediante un cálculo directo, similarmente a como hallamos la esperanza, expresando a la diferencia de cortes de Cheeger como una combinación lineal de variables aleatorias con distribución Bernoulli independientes, donde una vez más la Figura 4.2 puede ayudar a hacer sentido de los solapamientos que aparecen en la expresión. Esto completa la demostración. ■

De aquí a la demostración del teorema vamos a necesitar entender en más detalle cuánto se aleja el corte de Cheeger de las comunidades reales del de cada estimación, porque más que resultados del valor esperado queremos resultados en cuanto a la probabilidad. Más aún, al precisar que *simultáneamente* el corte de Cheeger de las comunidades sea mínimo necesitaremos que las probabilidades decaigan rápidamente para poder contrarrestar el hecho de que hay muchas estimaciones distintas fijas que podríamos considerar. Hay veces que el corte es evidentemente mucho más grande, para particiones muy alejadas de las comunidades reales, pero esto es menos claro para particiones casi precisas y para particiones donde alguna comunidad estimada cuenta con pocos nodos: en estos casos que haya suficiente varianza en la distribución de aristas podría ser detrimental. Controlar lo que ocurre en los casos borde ocupará gran parte de la demostración y es de carácter más bien técnico, aunque esencial.

A continuación en primer lugar enunciamos algunos lemas técnicos que usaremos. Para comenzar, este lema nos ayudará a lidiar con números combinatorios que debremos acotar:

[4.2.3] Lema: Cota para los coeficientes binomiales

Sean $n \geq k \in \mathbb{N}_0$. Luego

$$\binom{n}{k} \leq \left(\frac{en}{k}\right)^k = \exp(k \log(n) + k - k \log(k)) \leq \exp(k(\log(n) + 1)), \quad (4.22)$$

donde interpretamos $0 \log(0) = 0$. En particular

$$\binom{n}{k} \leq \exp(\min\{k, n-k\}(\log(n) + 1)). \quad (4.23)$$

Demostración. Para la primer cota, podemos proceder de la siguiente manera:

$$\binom{n}{k} = \frac{n!}{(n-k)!k!} \leq n^k \left(\frac{e}{k}\right)^k = \left(\frac{en}{k}\right)^k; \quad (4.24)$$

donde para la desigualdad usamos que $\frac{1}{k!} \leq \left(\frac{e}{k}\right)^k$ porque $e^k = \sum_{j \in \mathbb{N}_0} \frac{k^j}{j!} > \frac{k^k}{k!}$.

La segunda desigualdad es inmediata notando que $\exp(-k \log(k)) \leq 1$.

La tercera desigualdad se sigue de que $\binom{n}{k} = \binom{n}{n-k}$ según lo que tenemos que $\binom{n}{k}$ es menor que el mínimo que inducen ambas cotas, que es precisamente la expresión indicada. ■

Proseguiremos con este segundo lema, que es de carácter más bien técnico, y será de gran ayuda para tratar con algunos polinomios.

[4.2.4] Lema: Cota para cociente de polinomios

Recordemos que para cada multiíndice $\gamma \in \mathbb{N}_0^n$ definimos $x^\gamma = \prod_{i=1}^n x_i^{\gamma_i}$ para cada $x \in \mathbb{R}^n$.

Sea luego $\Gamma \subseteq \mathbb{N}_0^n$ un subconjunto de multiíndices, y sea $P \in \mathbb{R}[X_1, \dots, X_n]$ un polinomio de coeficientes positivos $P(X) = \sum_{\gamma \in \Gamma} a_\gamma X^\gamma$, con $P(0) = 0$ y $P(x) \neq 0$ para cada $0 < \|x\| \leq \varepsilon$ para cierto $\varepsilon > 0$.

Consideremos además $Q \in \mathbb{R}[X_1, \dots, X_n]$ otro polinomio de la forma $Q(X) = \sum_{\gamma \in \Gamma} Q_\gamma(X) X^\gamma$ para ciertos polinomios $Q_\gamma \in \mathbb{R}[X_1, \dots, X_n]$ y con $Q(0) = 0$.

Luego, vale que

$$\sup_{x \in C, \|x\| \leq \varepsilon} \left\{ \left| \frac{Q(x)}{P(x)} \right| \right\} < \infty, \quad (4.25)$$

y más aún si para cada $\gamma \in \Gamma$ vale que $Q_\gamma(0) = 0$ entonces

$$\lim_{x \rightarrow 0, x \in C} \left| \frac{Q(x)}{P(x)} \right| = 0, \quad (4.26)$$

donde $C = \{x \in \mathbb{R}^n \mid x_i \geq 0 \text{ para todo } 1 \leq i \leq n, x \neq 0\}$.

Del lema de cociente de polinomios. En primer lugar, notemos que por la desigualdad triangular

$$\left| \frac{Q(x)}{P(x)} \right| \leq \sum_{\gamma \in \Gamma} \left| \frac{Q_\gamma(x)x^\gamma}{P(x)} \right| \leq \sum_{\gamma \in \Gamma} \frac{|Q_\gamma(x)|}{a_\gamma}, \quad (4.27)$$

donde en la última desigualdad usamos que o bien $x^\gamma = 0$ y el término es 0 o bien $P(x) \geq a_\gamma x^\gamma$ con lo cual podemos acotar $\frac{x^\gamma}{P(x)} \leq \frac{1}{a_\gamma}$ para cada $x \in C$. Así podremos deducir la desigualdad (4.25), notando que $\overline{C} \cap \{ \|x\| \leq \varepsilon \}$ es un compacto en \mathbb{R}^n en el cual nuestro polinomio $\sum_{\gamma \in \Gamma} \frac{|Q_\gamma(x)|}{a_\gamma}$ será acotado.

Para el límite (4.26) basta notar solamente que cuando $Q_\gamma(0) = 0$ para cada $\gamma \in \Gamma$ tendremos que por continuidad $\lim_{x \rightarrow 0} \sum_{\gamma \in \Gamma} \frac{|Q_\gamma(x)|}{a_\gamma} = 0$ y habremos mostrado lo afirmado. ■

Continuaremos por este último lema, que nos ayudará a acomodar algunas expresiones al acotar:

[4.2.5] Lema: Cota para simplificar cocientes de polinomios

Dados $0 < A \leq B$ y $0 < C \leq D$, y también $x, y \geq 0$ donde no ambos son cero, entonces

$$\frac{(Ax + By)^2}{Cx + Dy} \geq \frac{A^2}{D}(x + y) \quad (4.28)$$

Demostración. Podemos acotar por abajo $Ax + By$ de la siguiente manera,

$$Ax + By = A(x + y) + (B - A)y \geq A(x + y), \quad (4.29)$$

y al combinarlo con esta cota superior de $Cx + Dy$

$$Cx + Dy = D(x + y) - (D - C)x \leq D(x + y) \quad (4.30)$$

obtenemos la cota deseada. ■

Por último, ya sin más preámbulos, demostraremos el Teorema 4.2.1.

Demostración. Vamos a mostrar que con alta probabilidad el corte de Cheeger de las comunidades reales es el único mínimo entre los cortes de Cheeger asociados a particiones del grafo. Esto alcanzará para demostrar consistencia fuerte, pues luego $\widehat{\mathcal{C}}_{\text{RCC}} = \mathcal{C}$ a.a.s., que es exactamente consistencia fuerte.

En primer lugar, usando una cota de unión de eventos logramos reducir el problema a mostrar que la probabilidad de que individualmente una partición tenga corte de Cheeger más chico que las comunidades reales decrezca suficientemente rápido. En efecto, como:

$$\begin{aligned} & \mathbb{P} \left(\left\{ \text{RCC}(\tilde{\mathcal{C}}_{\text{RCC}}) < \text{RCC}(\mathcal{C}) \right\} \right) \\ &= \mathbb{P} \left(\left\{ \text{RCC}(\tilde{\mathcal{C}}) < \text{RCC}(\mathcal{C}) \text{ para alguna partición } \tilde{\mathcal{C}} \right\} \right) \\ &\leq \sum_{\tilde{\mathcal{C}}, \text{coincidencia}(\tilde{\mathcal{C}}, \mathcal{C}) \neq n} \mathbb{P} \left(\left\{ \text{RCC}(\tilde{\mathcal{C}}) < \text{RCC}(\mathcal{C}) \right\} \right), \end{aligned} \quad (4.31)$$

basta ver que la suma va a 0 cuando n va a ∞ para demostrar que el complemento ocurre a.a.s..

Para demostrar el decaimiento rápido, usaremos la desigualdad de Bernstein 2.2.4. A través de esta obtendremos que

$$\begin{aligned} \mathbb{P} \left(\left\{ \text{RCC}(\mathcal{C}) > \text{RCC}(\tilde{\mathcal{C}}) \right\} \right) &\leq \mathbb{P} \left(\left\{ \left| \text{RCC}(\tilde{\mathcal{C}}) - \text{RCC}(\mathcal{C}) - E \right| > E \right\} \right) \\ &\leq 2 \exp \left(-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E} \right), \end{aligned} \quad (4.32)$$

donde notamos para esta demostración $E = \mathbb{E} [\text{RCC}(\tilde{\mathcal{C}}) - \text{RCC}(\mathcal{C})]$, $V = \text{Var} [\text{RCC}(\tilde{\mathcal{C}}) - \text{RCC}(\mathcal{C})]$, y M a una constante que acotará a cada uno de los términos en la combinación lineal de variables Bernoulli independientes que es la diferencia en cortes de Cheeger. Mencionamos que todas estas tres cantidades, E , V y M , dependerán de n , \mathcal{C} y $\tilde{\mathcal{C}}$, pero no hacemos explícita la dependencia para no cargar en demasía la notación, y como vamos a trabajar en acotar las probabilidades término a término podremos pensar que están fijas en cada cuenta en cuestión.

Para poder usar la desigualdad de Bernstein precisamos que $E > 0$, pero como demostramos en el Lema 4.2.2 esto ocurrirá cuando $\tilde{\mathcal{C}}$ no induce la misma partición que \mathcal{C} y $a > 2b$ como en los casos que nos interesan. Además, podemos hacer explícita la constante M : la diferencia entre los cortes es una combinación lineal de variables Bernoulli independientes, con coeficientes $\frac{1}{\min\{m_1, m_2\}} - \frac{2}{n}$ en las aristas que aparecen en ambos cortes, $\frac{1}{\min\{m_1, m_2\}}$ en las aristas que aparecen solamente en el corte de Cheeger de $\tilde{\mathcal{C}}$ y $-\frac{2}{n}$ en las aristas que aparecen solamente en el corte de Cheeger asociado a \mathcal{C} ; de esta manera la cota

$$M_{\mathcal{C}, \tilde{\mathcal{C}}} = \frac{1}{\min\{m_1, m_2\}} \quad (4.33)$$

acota uniformemente en valor absoluto a los coeficientes al ser $\min\{m_1, m_2\} \leq \frac{n}{2}$ pues $m_1 + m_2 = n$.

Ahora bien, las cantidades E , V y M dependerán únicamente de los solapamientos $X_{i,j}$. Por eso podremos indexar la suma en términos de estos solapamientos de cada estimación con las particiones reales, lo que nos permitirá acotar de la siguiente manera, al completar en (4.31):

$$\begin{aligned} \mathbb{P} \left(\left\{ \text{RCC}(\tilde{\mathcal{C}}_{\text{RCC}}) < \text{RCC}(\mathcal{C}) \right\} \right) &\quad (4.34) \\ &\leq \sum_{X_{1,1}=1}^{\frac{n}{2}} \sum_{X_{2,1}=1, X \text{ admisibles}}^{\frac{n}{2}} \binom{\frac{n}{2}}{X_{1,1}} \binom{\frac{n}{2}}{X_{2,1}} 2 \exp \left(-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E} \right), \end{aligned}$$

donde por X admisibles nos referimos a particiones en donde $m_1, m_2 > 0$. Notemos que tenemos una cantidad cuadrática de términos, con lo que si demostramos que cada uno es $o(n^2)$ obtendremos que efectivamente la suma convergerá a 0.

El resto de la demostración entonces se concentrará en mostrar que el decaimiento es así de rápido.

En primer lugar, por el Lema 4.2.3 tendremos que para cualquier elección de j, j' vale que:

$$\begin{aligned} \binom{\frac{n}{2}}{X_{1,1}} \binom{\frac{n}{2}}{X_{2,1}} &\leq \exp\left((X_{1,j} + X_{2,j'}) \left(\log\left(\frac{n}{2}\right) + 1\right)\right) \\ &\leq \exp\left((X_{1,j} + X_{2,j'}) (\log(n) + 1)\right) \end{aligned} \quad (4.35)$$

pues $X_{1,2} = \frac{n}{2} - X_{1,1}$ y $X_{2,2} = \frac{n}{2} - X_{2,1}$. Así, basta con garantizar un orden de decaimiento mayor que este para las probabilidades para completar la demostración.

Con ese fin, primero entenderemos el exponente de nuestra cota para cada estimación $\tilde{\mathcal{C}}$ fija en términos de las proporciones de sus solapamientos con las comunidades reales \mathcal{C} , $x_{i,j}$, y después separaremos en casos para entender su comportamiento.

Para comenzar, notamos que $n_1 = n_2$ (o $\alpha_1 = \alpha_2$), y supondremos sin pérdida de generalidad que $m_1 \leq m_2$ (o equivalentemente que $\beta_1 \leq \beta_2$) lo que nos permitirá calcular los mínimos que aparecen en la expresión del corte de Cheeger. En efecto, debido a la simetría del problema, si fuese $m_1 > m_2$ podemos renombrar las comunidades estimadas en $\tilde{\mathcal{C}}$ para obtener $m_1 \leq m_2$ y estar en la situación mencionada. En particular, en este caso obtenemos que $M = \frac{1}{n\beta_1}$ a partir de (4.33).

Luego, extraeremos los factores que dependen de n y los cocientes de la expresión para que nos queden únicamente sumas y productos de proporciones $x_{i,j}$ de la siguiente manera:

$$-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E} = -\frac{1}{2} n \log(n) \frac{\left(\frac{\beta_1}{\log(n)} E\right)^2}{\beta_1^2 \frac{n}{\log(n)} \left(V + \frac{M}{3}E\right)}. \quad (4.36)$$

A continuación, expresaremos cada factor en el cociente en términos de las proporciones. Obtenemos expandiendo las expresiones obtenidas para E y V en el Lema 4.2.2 y usando que $M = \frac{1}{n\beta_1}$ que podemos expresar los factores escalados como polinomios en las proporciones de solapamiento $x_{i,j}$ y $\frac{\log(n)}{n}$:

$$\begin{cases} \frac{\beta_1}{\log(n)} E &= P(x_{1,1}, x_{1,2}, x_{2,1}, x_{2,2}) \\ \beta_1^2 \frac{n}{\log(n)} \left(V + \frac{M}{3}E\right) &= Q(x_{1,1}, x_{1,2}, x_{2,1}, x_{2,2}) + R(x_{1,1}, x_{1,2}, x_{2,1}, x_{2,2}) \frac{\log(n)}{n}. \end{cases} \quad (4.37)$$

En adelante, anotaremos simplemente P , Q y R y los pensaremos evaluados en las proporciones.

Más precisamente, podemos desarrollar las expresiones para obtener que P , Q , y R están dados por:

$$P = ax_{1,1}x_{1,2} + bx_{1,2}x_{2,1} + bx_{1,1}x_{2,2} + ax_{2,1}x_{2,2} - \frac{1}{2}bx_{1,1} - \frac{1}{2}bx_{2,1}, \quad (4.38)$$

$$\begin{aligned} Q &= 4bx_{1,1}^3x_{2,1} + 4bx_{1,1}^2x_{1,2}x_{2,1} + 8bx_{1,1}^2x_{2,1}^2 + 8bx_{1,1}x_{1,2}x_{2,1}^2 \\ &\quad + 4bx_{1,1}x_{2,1}^3 + 4bx_{1,2}x_{2,1}^3 + 4bx_{1,1}^3x_{2,2} + 4bx_{1,1}^2x_{1,2}x_{2,2} \\ &\quad + 8bx_{1,1}^2x_{2,1}x_{2,2} + 8bx_{1,1}x_{1,2}x_{2,1}x_{2,2} + 4bx_{1,1}x_{2,1}^2x_{2,2} + 4bx_{1,2}x_{2,1}^2x_{2,2} \\ &\quad - 4bx_{1,1}x_{1,2}x_{2,1} - 4bx_{1,2}x_{2,1}^2 - 4bx_{1,1}^2x_{2,2} - 4bx_{1,1}x_{2,1}x_{2,2} \end{aligned} \quad (4.39)$$

$$\begin{aligned}
& + \frac{4}{3}ax_{1,1}x_{1,2} + \frac{4}{3}bx_{1,2}x_{2,1} + \frac{4}{3}bx_{1,1}x_{2,2} + \frac{4}{3}ax_{2,1}x_{2,2} \\
& - \frac{1}{6}bx_{1,1} - \frac{1}{6}bx_{2,1}, \\
R = & -4b^2x_{1,1}^3x_{2,1} - 4b^2x_{1,1}^2x_{1,2}x_{2,1} - 8b^2x_{1,1}^2x_{2,1}^2 - 8b^2x_{1,1}x_{1,2}x_{2,1}^2 \\
& - 4b^2x_{1,1}x_{2,1}^3 - 4b^2x_{1,2}x_{2,1}^3 - 4b^2x_{1,1}^3x_{2,2} - 4b^2x_{1,1}^2x_{1,2}x_{2,2} \\
& - 8b^2x_{1,1}^2x_{2,1}x_{2,2} - 8b^2x_{1,1}x_{1,2}x_{2,1}x_{2,2} - 4b^2x_{1,1}x_{2,1}^2x_{2,2} - 4b^2x_{1,2}x_{2,1}^2x_{2,2} \\
& + 4b^2x_{1,1}x_{1,2}x_{2,1} + 4b^2x_{1,2}x_{2,1}^2 + 4b^2x_{1,1}^2x_{2,2} + 4b^2x_{1,1}x_{2,1}x_{2,2} \\
& - a^2x_{1,1}x_{1,2} - b^2x_{1,2}x_{2,1} - b^2x_{1,1}x_{2,2} - a^2x_{2,1}x_{2,2}.
\end{aligned} \tag{4.40}$$

Antes de ponernos a analizarlos en detalle, consideremos $\varepsilon > 0$ que será un número suficientemente pequeño a fijar eventualmente y separaremos en tres casos:

1. cuando $\varepsilon \leq x_{1,1}, x_{1,2} \leq \frac{1}{2} - \varepsilon$ o $\varepsilon \leq x_{2,1}, x_{2,2} \leq \frac{1}{2} - \varepsilon$, es decir que la predicción global no es muy precisa y ninguna de las dos comunidades estimadas es muy chica;
2. cuando $x_{1,1} < \varepsilon$, $x_{1,2} > \frac{1}{2} - \varepsilon$, $x_{2,1} < \varepsilon$ y $x_{2,2} > \frac{1}{2} - \varepsilon$, es decir que una de las dos comunidades estimadas es muy chica y la otra es muy grande y en particular la estimación es imprecisa;
3. y cuando $x_{1,1} > \frac{1}{2} - \varepsilon$, $x_{1,2} < \varepsilon$, $x_{2,1} < \varepsilon$ y $x_{2,2} > \frac{1}{2} - \varepsilon$, es decir que ambas comunidades estimadas son suficientemente grandes y la predicción es casi totalmente certera.

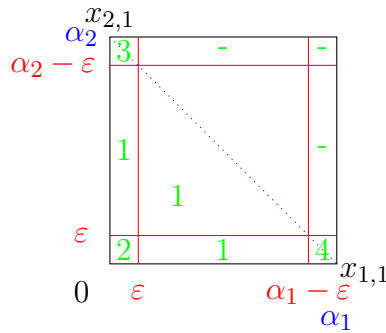


Figura 4.3: Un diagrama para entender los casos relevantes a considerar para la consistencia del corte de Cheeger.

Afirmamos que es suficiente considerar únicamente estos tres casos. En efecto, primero notemos que, como $x_{1,1} + x_{1,2} = \frac{1}{2}$ y $x_{2,1} + x_{2,2} = \frac{1}{2}$, podemos determinar cada caso unívocamente por los valores de $x_{1,1}$ y $x_{2,1}$, que serán números cada uno entre 0 y $\frac{1}{2}$. Convencidos de esto, bastará convencernos de que podremos lidiar con cada coordenada en el cuadrado de la Figura 4.2. Así, primero distinguimos las regiones que delimitan los casos 1., 2. y 3. en la figura, donde además el caso en el que $x_{1,1} < \varepsilon$ y $x_{2,1} > \frac{1}{2} - \varepsilon$ fue marcado con un 3. porque puede reducirse a él por simetría renombrando las comunidades reales. Esto cubre todo lo que está por debajo de la línea azul punteada, y entonces notando que no puede pasar que $x_{1,1}$ y $x_{2,1}$ sean simultáneamente grandes, pues supusimos previamente que $x_{1,1} + x_{2,1} = \beta_1 \leq \frac{1}{2}$

al ser esta la comunidad estimada más chica, vemos que no necesitamos considerar los casos sobre la línea azul punteada. Alternativamente podemos convencernos que los casos arriba de esta línea salen por simetría renombrando las comunidades estimadas. En definitiva vemos que estamos considerando todos los casos, como habíamos afirmado.

En el **caso 1.**, tendremos suficiente margen para acotar bruscamente $\left| \frac{P}{Q + R \frac{\log(n)}{n}} \right|$ para cada ε fijo, y concluir que las probabilidades decaen suficientemente rápido. En efecto, conseguiremos la siguiente cota para n suficientemente grandes, para cada $C \in (0, 1)$ constante

$$-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E} \leq -\log(n)(X_{1,1} + X_{2,1}) \left[C \frac{32\varepsilon^5(a-b)^2}{\frac{8}{3}a + 83b} \right], \quad (4.41)$$

y así combinando con la cota para los números combinatorios en (4.35) obtenemos que

$$\begin{aligned} & \binom{\frac{n}{2}}{X_{1,1}} \binom{\frac{n}{2}}{X_{2,1}} \exp \left(-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E} \right) \\ & \leq \exp \left(\log(n)(X_{1,1} + X_{2,1}) \left(\frac{\log(n) + 1}{\log(n)} - C \frac{32\varepsilon^5(a-b)^2}{\frac{8}{3}a + 83b} \right) \right), \end{aligned} \quad (4.42)$$

y entonces siempre que

$$(X_{1,1} + X_{2,1}) \left(\frac{\log(n) + 1}{\log(n)} - C \frac{32\varepsilon^5(a-b)^2}{\frac{8}{3}a + 83b} \right) < -2 \quad (4.43)$$

obtendremos el orden de decaimiento que queríamos de $o(n^2)$. Notemos que podremos asegurar que esto ocurrirá para valores de a suficientemente grandes, pues como $X_{1,1} + X_{2,1} = m_1 \geq 1$ bastará ver que

$$\left(\frac{\log(n) + 1}{\log(n)} - C \frac{32\varepsilon^5(a-b)^2}{\frac{8}{3}a + 83b} \right) < -2, \quad (4.44)$$

lo cual ocurre porque este segundo factor se comportará para n suficientemente grandes como $1 - C' \frac{a^2}{a} = 1 - C'a$ que se irá a $-\infty$ cuando $a \rightarrow \infty$, donde usamos que $a > 2b$ para ignorar la dependencia de b y además C' es una constante no necesariamente igual a la de la expresión anterior.

Resta entonces mostrar que vale (4.41) como fue anticipado. En efecto, podremos usar que valen ambas

$$P \geq \frac{\beta_1}{\log(n)} \left(2\varepsilon^2 \log(n)(a-b) \right) = 2\beta_1 \varepsilon^2 (a-b) \quad (4.45)$$

y también

$$Q + R \frac{\log(n)}{n} \leq \left[\frac{8}{3}a + 83b \right] \left(\frac{1}{2} \right)^4 + \left[2a^2 + 82b^2 \right] \left(\frac{1}{2} \right)^4 \frac{\log(n)}{n} \quad (4.46)$$

para acotar combinando las dos cotas del siguiente modo, para alguna constante $C \in (0, 1)$ y suponiendo que n es suficientemente grande:

$$\frac{P}{Q + R \frac{\log(n)}{n}} \geq \frac{\left(2\beta_1 \varepsilon^2 (a-b) \right)^2}{\left[\frac{8}{3}a + 83b \right] \left(\frac{1}{2} \right)^4 + \left[2a^2 + 82b^2 \right] \left(\frac{1}{2} \right)^4 \frac{\log(n)}{n}} \quad (4.47)$$

$$\begin{aligned}
&= \beta_1 \left(\frac{4\beta_1 \varepsilon^4 (a-b)^2}{\left[\frac{8}{3}a + 83b\right] \left(\frac{1}{2}\right)^4} \right) \left(\frac{1}{1 + \frac{[2a^2 + 82b^2] \left(\frac{1}{2}\right)^4 \frac{\log(n)}{n}}{\left[\frac{8}{3}a + 83b\right] \left(\frac{1}{2}\right)^4}} \right) \\
&> \beta_1 C \frac{4\varepsilon^5 (a-b)^2}{\left[\frac{8}{3}a + 83b\right] \left(\frac{1}{2}\right)^4} \\
&= \beta_1 C \frac{64\varepsilon^5 (a-b)^2}{\frac{8}{3}a + 83b}.
\end{aligned}$$

Aclaremos que para la última desigualdad usamos que $\beta_1 = x_{1,1} + x_{2,1} \geq \varepsilon$, y tomamos C una constante que acote inferiormente a

$$\frac{1}{1 + \frac{[2a^2 + 82b^2] \left(\frac{1}{2}\right)^4 \frac{\log(n)}{n}}{\left[\frac{8}{3}a + 83b\right] \left(\frac{1}{2}\right)^4}} = \frac{1}{1 + \frac{2a^2 + 82b^2 \log(n)}{\frac{8}{3}a + 83b} \frac{1}{n}} \xrightarrow{n \rightarrow \infty} 1, \quad (4.48)$$

que para n suficientemente grandes estará arbitrariamente cerca de 1. Notemos luego que al multiplicar por $-\frac{1}{2}n \log(n)$ para obtener el exponente de nuestra cota por Bernstein como nos indica (4.36) podemos deducir la desigualdad que buscábamos (4.41), notando que $n\beta_1 = (X_{1,1} + X_{2,1})$.

Resta solamente luego verificar que valen las cotas (4.45) y (4.46) para terminar de acotar cada término en (4.34) correspondiente al caso 1.

En primer lugar, notemos que P es no negativo pues hemos demostrado en el Lema 4.2.2 que $E > 0$ cuando la comunidad estimada no coincide con la comunidad real, con lo cual $|P| = P$ sobre los solapamientos considerados. Así, usando la expresión de la esperanza obtenida en dicho lema obtenemos que

$$\begin{aligned}
E &= \left(\beta_2 - \frac{1}{2}\right) nq_n + \left(\frac{x_{1,1}x_{1,2} + x_{2,1}x_{2,2}}{\beta_1}\right) n(p_n - q_n) \\
&\geq \left(\frac{x_{1,1}x_{1,2} + x_{2,1}x_{2,2}}{\beta_1}\right) n(p_n - q_n) \\
&\geq 2\varepsilon^2 n(p_n - q_n) \\
&= 2\varepsilon^2 \log(n)(a - b).
\end{aligned} \quad (4.49)$$

Aclaremos que en la primera cota usamos que el primer término es no negativo, pues $\beta_2 \geq \frac{1}{2}$ al estar asociada a la comunidad más grande entre las dos comunidades estimadas. Más aún, añadimos que en la segunda cota usamos ambas que, como $x_{1,1}, x_{1,2} \geq \varepsilon$ o $x_{2,1}, x_{2,2} \geq \varepsilon$, vale que $x_{1,1}x_{1,2} + x_{2,1}x_{2,2} \geq \varepsilon^2$; además notamos que $\beta_1 \leq \frac{1}{2}$ y así $\frac{1}{\beta_1} \geq 2$, al estar β_1 asociada a la comunidad más chica entre las comunidades estimadas. Finalmente, escalando E para obtener P como en (4.37) vemos que podemos acotar P como afirmábamos en (4.45).

Por otro lado, acotando mediante la desigualdad triangular y usando que $x_{i,j} \leq \frac{1}{2}$ para cada par de índices i, j obtenemos que

$$\left| Q(x_{1,1}, x_{1,2}, x_{2,1}, x_{2,2}) + R(x_{1,1}, x_{1,2}, x_{2,1}, x_{2,2}) \frac{\log(n)}{n} \right| \quad (4.50)$$

$$\begin{aligned}
&\leq [4b + 4b + 8b + 8b + 4b + 4b + 4b + 4b + 8b + 8b + 4b + 4b \\
&\quad + 4b + 4b + 4b + 4b + \frac{4}{3}a + \frac{4}{3}b + \frac{4}{3}b + \frac{4}{3}a + \frac{1}{6}b + \frac{1}{6}b] \left(\frac{1}{2}\right)^4 \\
&\quad + [4b^2 + 4b^2 + 8b^2 + 8b^2 + 4b^2 + 4b^2 + 4b^2 + 4b^2 + 8b^2 + 8b^2 + 4b^2 + 4b^2 \\
&\quad + 4b^2 + 4b^2 + 4b^2 + 4b^2 + a^2 + b^2 + b^2 + a^2] \left(\frac{1}{2}\right)^4 \frac{\log(n)}{n} \\
&= \left[\frac{8}{3}a + 83b\right] \left(\frac{1}{2}\right)^4 + [2a^2 + 82b^2] \left(\frac{1}{2}\right)^4 \frac{\log(n)}{n}
\end{aligned}$$

Por último notemos que nuevamente por el Lema 4.2.2 vale que $E > 0$, y que además $M \geq 0$ y $V \geq 0$, con lo cual tenemos que $Q + R \frac{\log(n)}{n} \geq 0$ y es igual a su valor absoluto y de esta manera mostramos que vale (4.46) como queríamos, y obtenemos el decaimiento precisado en este caso.

En el **caso 2**. necesitaremos un análisis más minucioso, pues pasará que P podrá ser cercano a 0 para algunas proporciones $x_{i,j}$, y deberemos ver entonces que el orden de P coincide con el orden de $Q + R \frac{\log(n)}{n}$ como polinomios en las proporciones para que el orden de nuestra cota para el exponente supere al orden de la cota para los números combinatorios.

Con ese objetivo en mente, expresamos a P , Q y R en términos de las variables que estarán próximas a 0, reemplazando $x_{1,2} = \frac{1}{2} - x_{1,1}$ y $x_{2,2} = \frac{1}{2} - x_{1,1}$ para obtener

$$P = -ax_{1,1}^2 - 2bx_{1,1}x_{2,1} - ax_{2,1}^2 + \frac{1}{2}ax_{1,1} + \frac{1}{2}ax_{2,1}, \quad (4.51)$$

$$\begin{aligned}
Q &= 8bx_{1,1}^2x_{2,1} + 8bx_{1,1}x_{2,1}^2 - \frac{4}{3}ax_{1,1}^2 - bx_{1,1}^2 \\
&\quad - \frac{14}{3}bx_{1,1}x_{2,1} - \frac{4}{3}ax_{2,1}^2 - bx_{2,1}^2 + \frac{2}{3}ax_{1,1} \\
&\quad + \frac{1}{2}bx_{1,1} + \frac{2}{3}ax_{2,1} + \frac{1}{2}bx_{2,1},
\end{aligned} \quad (4.52)$$

$$\begin{aligned}
R &= -8b^2x_{1,1}^2x_{2,1} - 8b^2x_{1,1}x_{2,1}^2 + a^2x_{1,1}^2 + b^2x_{1,1}^2 \\
&\quad + 4b^2x_{1,1}x_{2,1} + a^2x_{2,1}^2 + b^2x_{2,1}^2 - \frac{1}{2}a^2x_{1,1} \\
&\quad - \frac{1}{2}b^2x_{1,1} - \frac{1}{2}a^2x_{2,1} - \frac{1}{2}b^2x_{2,1}.
\end{aligned} \quad (4.53)$$

A continuación definimos P_1, Q_1 como la suma de los monomios de grado 1 en P y Q respectivamente,

$$P_1 = \frac{1}{2}ax_{1,1} + \frac{1}{2}ax_{2,1} \quad (4.54)$$

$$Q_1 = \frac{2}{3}ax_{1,1} + \frac{1}{2}bx_{1,1} + \frac{2}{3}ax_{2,1} + \frac{1}{2}bx_{2,1}. \quad (4.55)$$

La intuición al realizar esto es que para proporciones $x_{1,1}$ y $x_{1,2}$ cercanas a 0, al no tener término independiente será que P será cercano a P_1 y Q a Q_1 ; más aún además $Q + R \frac{\log(n)}{n}$ también será cercano a Q_1 y así mediante P_1 y Q_1 entenderemos el exponente en la cota de Bernstein.

En efecto, podremos formalizar la intuición mediante el Lema 4.2.4 como detallaremos a

continuación. Primero, expresamos a P y $Q + R\frac{\log(n)}{n}$ en términos de P_1 y Q_1 :

$$P = P_1 \left(1 + \frac{P - P_1}{P_1} \right) \quad (4.56)$$

$$Q + R\frac{\log(n)}{n} = Q_1 \left(1 + \frac{Q - Q_1}{Q_1} + \frac{R \log(n)}{Q_1 n} \right). \quad (4.57)$$

A continuación, notamos que los coeficientes de P_1 y Q_1 son todos positivos, $P(0) = Q(0) = R(0) = P_1(0) = Q_1(0) = 0$, $P_1(x) = 0$ solamente cuando $x = 0$ para proporciones en el primer cuadrante y análogamente $Q_1(x) = 0$ solamente cuando $x = 0$ para proporciones en el primer cuadrante, y por último como P_1 y Q_1 tienen como monomios a múltiplos de x e y vale que P , Q y R se expresan como combinación lineal de polinomios por estos monomios: así estamos en las condiciones del Lema 4.2.4. Haciendo uso de dicho lema podremos mostrar que para valores de ε suficientemente chicos será que $\left(1 + \frac{P - P_1}{P_1}\right)$ y $\left(1 + \frac{Q - Q_1}{Q_1} + \frac{R \log(n)}{Q_1 n}\right)$ son cercanos a 1, pues valen que

$$\sup_{x_{1,1}, x_{2,1} \leq \varepsilon} \left\{ \left| \frac{P(x_{1,1}, x_{2,1}) - P_1(x_{1,1}, x_{2,1})}{P_1(x_{1,1}, x_{2,1})} \right| \right\} \xrightarrow{\varepsilon \rightarrow 0} 0, \quad (4.58)$$

que

$$\sup_{x_{1,1}, x_{2,1} \leq \varepsilon} \left\{ \left| \frac{Q(x_{1,1}, x_{2,1}) - Q_1(x_{1,1}, x_{2,1})}{Q_1(x_{1,1}, x_{2,1})} \right| \right\} \xrightarrow{\varepsilon \rightarrow 0} 0, \quad (4.59)$$

y que

$$\sup_{x_{1,1}, x_{2,1} \leq \varepsilon} \left\{ \left| \frac{R(x_{1,1}, x_{2,1})}{Q_1(x_{1,1}, x_{2,1})} \right| \right\} < \infty; \quad (4.60)$$

agregamos que como $\frac{R}{Q_1}$ es acotado obtendremos que para valores suficientemente grandes de n , $\frac{R \log(n)}{Q_1 n}$ será uniformemente chico en las proporciones. Combinando todo lo anterior, obtenemos que para constantes $C_1 < 1 < C_2$ tan próximas a 1 como se quiera, podremos hallar $\varepsilon > 0$ suficientemente chicos y n suficientemente grandes tales que

$$C_1 \frac{(P_1)^2}{Q_1} \leq \frac{P^2}{Q + R\frac{\log(n)}{n}} \leq C_2 \frac{(P_1)^2}{Q_1} \quad (4.61)$$

para todo $x_{1,1}, x_{1,2} \leq \varepsilon$.

Pero luego como

$$\begin{aligned} \frac{(P_1)^2}{Q_1} &= \left(\frac{1}{2}ax_{1,1} + \frac{1}{2}ax_{2,1} \right) \frac{\frac{1}{2}ax_{1,1} + \frac{1}{2}ax_{2,1}}{\frac{2}{3}ax_{1,1} + \frac{1}{2}bx_{1,1} + \frac{2}{3}ax_{2,1} + \frac{1}{2}bx_{2,1}} \\ &\geq \frac{1}{2}a\beta_1 \frac{\frac{1}{2}a}{\frac{2}{3}a + \frac{1}{2}b} \\ &= \beta_1 \frac{\left(\frac{1}{2}a\right)^2}{\frac{2}{3}a + \frac{1}{2}b} \end{aligned} \quad (4.62)$$

obtenemos al multiplicar por $-\frac{1}{2}n \log(n)$, como nos indica (4.37), y al combinar después con (4.61) y además con la cota para los números combinatorios en (4.35) que cada término en la

suma (4.34) correspondiente al caso 2. puede controlarse por la siguiente expresión, si $\varepsilon > 0$ es suficientemente chico, $C < 1$ es una constante a elección y n es suficientemente grande:

$$\begin{aligned} & \binom{\frac{n}{2}}{X_{1,1}} \binom{\frac{n}{2}}{X_{2,1}} 2 \exp \left(-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E} \right) \\ & \leq 2 \exp \left(\log(n) (X_{1,1} + X_{2,1}) \left(\frac{\log(n) + 1}{\log(n)} - \frac{1}{2} C \frac{\left(\frac{1}{2}a\right)^2}{\frac{2}{3}a + \frac{1}{2}b} \right) \right). \end{aligned} \quad (4.63)$$

Entonces análogamente al caso 1. tendremos que la cota es $o(n^{-2})$ cuando

$$(X_{1,1} + X_{2,1}) \left(\frac{\log(n) + 1}{\log(n)} - \frac{1}{2} C \frac{\left(\frac{1}{2}a\right)^2}{\frac{2}{3}a + \frac{1}{2}b} \right) < -2 \quad (4.64)$$

que una vez más podremos ver de la misma manera que ocurre para a suficientemente grande.

En el **caso 3.**, nuevamente tendremos que hacer un análisis más exhaustivo y procederemos como en el caso 2., adoptando notación idéntica. Tendremos que, y esta vez expresados en términos de $x_{1,2}$ y $x_{2,1}$ que son las proporciones que serán chicas ahora, P , Q y R serán de la forma:

$$P = -ax_{1,2}^2 + 2bx_{1,2}x_{2,1} - ax_{2,1}^2 + \frac{1}{2}ax_{1,2} + \frac{1}{2}ax_{2,1} - bx_{2,1}, \quad (4.65)$$

$$\begin{aligned} Q &= 8bx_{1,2}^2x_{2,1} - 8bx_{1,2}x_{2,1}^2 - \frac{4}{3}ax_{1,2}^2 - bx_{1,2}^2 \\ &\quad - \frac{10}{3}bx_{1,2}x_{2,1} - \frac{4}{3}ax_{2,1}^2 + 3bx_{2,1}^2 + \frac{2}{3}ax_{1,2} \\ &\quad + \frac{1}{2}bx_{1,2} + \frac{2}{3}ax_{2,1} + \frac{1}{6}bx_{2,1}, \end{aligned} \quad (4.66)$$

$$\begin{aligned} R &= -8b^2x_{1,2}^2x_{2,1} + 8b^2x_{1,2}x_{2,1}^2 + a^2x_{1,2}^2 + b^2x_{1,2}^2 \\ &\quad + 4b^2x_{1,2}x_{2,1} + a^2x_{2,1}^2 - 3b^2x_{2,1}^2 - \frac{1}{2}a^2x_{1,2} \\ &\quad - \frac{1}{2}b^2x_{1,2} - \frac{1}{2}a^2x_{2,1} - \frac{1}{2}b^2x_{2,1}, \end{aligned} \quad (4.67)$$

y así P_1 y Q_1 serán de la forma

$$P_1 = \frac{1}{2}ax_{1,2} + \left(\frac{1}{2}a - b\right)x_{2,1}, \quad (4.68)$$

$$Q_1 = \left(\frac{2}{3}a + \frac{1}{2}b\right)x_{1,2} + \left(\frac{2}{3}a + \frac{1}{6}b\right)x_{2,1}. \quad (4.69)$$

Notando que P_1 y Q_1 tienen coeficientes no negativos pues $2a > b$ y que una vez más se verifican las hipótesis del Lema 4.2.4, podremos usarlo esta vez junto con el Lema 4.2.5 para ver que para cualquier constante $C < 1$ podremos hallar ε suficientemente chico y n

suficientemente grande tales que

$$\begin{aligned} \frac{P^2}{Q + R \frac{\log(n)}{n}} &\geq C \frac{P_1^2}{Q_1} \\ &= C \frac{\left(\frac{1}{2}ax_{1,2} + \left(\frac{1}{2}a - b\right)x_{2,1}\right)^2}{\left(\frac{2}{3}a + \frac{1}{2}b\right)x_{1,2} + \left(\frac{2}{3}a + \frac{1}{6}b\right)x_{2,1}} \\ &\geq C(x_{1,2} + x_{2,1}) \frac{\left(\frac{1}{2}a - b\right)^2}{\frac{2}{3}a + \frac{1}{2}b}. \end{aligned} \quad (4.70)$$

Así en definitiva obtenemos que al escalar multiplicando por $-\frac{1}{2}n \log(n)$ como indica (4.37) podremos controlar el exponente de la cota de Bernstein como a continuación

$$-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E} \leq -\frac{1}{2} \log(n)(X_{1,2} + X_{2,1}) \left[C \frac{\left(\frac{1}{2}a - b\right)^2}{\frac{2}{3}a + \frac{1}{2}b} \right], \quad (4.71)$$

y entonces cada término en (4.34) correspondiente al caso 3. puede acotarse por la siguiente expresión

$$\begin{aligned} &\binom{\frac{n}{2}}{X_{1,1}} \binom{\frac{n}{2}}{X_{2,1}} 2 \exp\left(-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E}\right) \\ &\leq 2 \exp\left(\log(n)(X_{1,2} + X_{2,1}) \left(\frac{\log(n) + 1}{\log(n)} - \frac{1}{2} C \frac{\left(\frac{1}{2}a - b\right)^2}{\frac{2}{3}a + \frac{1}{2}b}\right)\right), \end{aligned} \quad (4.72)$$

y para $a - 2b$ suficientemente grande obtenemos que cada término será $o(n^{-2})$ de manera análoga a los casos previos. ■

Habiendo mostrado la consistencia para el corte de Cheeger, continuaremos en la siguiente sección demostrándola para el corte ratio.

4.3 Consistencia para el minimizante del corte ratio

Como adelantamos, tendremos un resultado análogo para el corte ratio:

[4.3.1] Teorema: Consistencia fuerte de una realización de corte ratio mínima

Sea $G \sim \text{SBM}(n, 2, (n_l)_{1 \leq l \leq 2}, p_n, q_n)$ donde $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para $a > b > 0$ números positivos. Supongamos además que $n_l = \Omega(n)$ para cada $1 \leq l \leq 2$.

Luego, si $a - b$ es suficientemente grande, vale que $\hat{\mathcal{C}}_{\text{RC}}$ es un estimador fuertemente consistente de \mathcal{C} .

La demostración usa las mismas técnicas que ya usamos para mostrar el Teorema 4.2.1, pudiendo determinar que no necesitamos restringirnos a particiones balanceadas. Sin embargo, además cuenta con la particularidad de que, debido a la simetría que presenta el

corte ratio, podremos demostrar la consistencia fuerte para el modelo general con tamaños de comunidades fijas en vez que solamente para el modelo balanceado. Aún así, como las ideas ya fueron presentadas anteriormente remitiremos para las demostraciones al anexo A.2. Mencionamos además que la condición de que los tamaños de las comunidades sean asintóticamente lineales es necesaria en la demostración pero además refleja que si las comunidades fuesen demasiado chicas entonces los factores que ponemos para balancearlas terminarían logrando el efecto contrario y no tendríamos una estimación exacta.

Por otro lado, una segunda y última particularidad es que al haber mostrado que el resultado vale para una gama amplia de tamaños de comunidades fijos, podremos extender el teorema anterior al modelo SBM general en el cual sorteamos la comunidad de cada nodo individualmente, como detallaremos en el siguiente teorema:

[4.3.2] Teorema: Consistencia fuerte del estimador de corte ratio en el caso general

Sea $G \sim \text{SBM}(n, 2, m_{\mathcal{C}}, p_n, q_n)$ donde $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para $a > b > 0$ números positivos.

Luego, si $a - b$ es suficientemente grande vale que $\hat{\mathcal{C}}_{\text{RC}}$ es un estimador fuertemente consistente de \mathcal{C} .

Demostración. Primero, notamos que por la Observación 2.4.4 sabemos que $n_l = np_{m_{\mathcal{C}}}(l) + \mathcal{O}\left(n^{\frac{1+\varepsilon}{2}}\right)$ para cada comunidad l con probabilidad más alta que $1 - \mathcal{O}(n^{-\varepsilon})$, en particular $1 - o(1)$, fijo $\varepsilon > 0$ suficientemente chico.

Llamemos luego en esta demostración para cada vector $(\tilde{n}_l)_{1 \leq l \leq k} \in \mathbb{N}_0^k$ con tamaños para las comunidades, que cumpla además que $\sum_{l=1}^k \tilde{n}_l = n$, $\mathcal{A}_{\tilde{n}}$ al evento dado por

$$\mathcal{A}_{\tilde{n}} = \{ n_l = \tilde{n}_l \text{ para cada } 1 \leq l \leq k \}, \quad (4.73)$$

y digamos que $(\tilde{n}_l)_{1 \leq l \leq k}$ es admisible si vale que $|\tilde{n}_l - np_{m_{\mathcal{C}}}(l)| \leq n^{\frac{3}{4}}$ para cada $n \in \mathbb{N}$ suficientemente grande y para cada $1 \leq l \leq k$. Notemos que si $(\tilde{n}_l)_{1 \leq l \leq k}$ no es admisible luego valdrá que en particular $\tilde{n}_l \neq np_{m_{\mathcal{C}}}(l) + \mathcal{O}\left(n^{\frac{1+\varepsilon}{2}}\right)$.

Entonces mediante la ley de probabilidad total obtenemos que, además separando los casos que ya notamos que son poco probables, podemos mostrar que la probabilidad de que haya algún error en la predicción es asintóticamente arbitrariamente chica y tendremos

consistencia fuerte, como queríamos mostrar. En efecto:

$$\begin{aligned}
& \mathbb{P} \left(\left\{ \text{coincidencia} \left(\widehat{\mathcal{C}}_{\text{RC}}, \mathcal{C} \right) \neq n \right\} \right) \\
& \leq \sum_{(\tilde{n}_l)_{1 \leq l \leq k} \text{ admisible}} \mathbb{P} \left(\left\{ \text{coincidencia} \left(\widehat{\mathcal{C}}_{\text{RC}}, \mathcal{C} \right) \neq n \right\} \mid \mathcal{A}_{\tilde{n}} \right) \cdot \mathbb{P} \left(\mathcal{A}_{\tilde{n}} \right) \\
& \quad + \mathbb{P} \left(\left\{ n_l \neq np_{m_c}(l) + \mathcal{O} \left(n^{\frac{1+\varepsilon}{2}} \right) \right\} \right) \\
& \leq \max_{(\tilde{n}_l)_{1 \leq l \leq k} \text{ admisible}} \left\{ \mathbb{P} \left(\left\{ \text{coincidencia} \left(\widehat{\mathcal{C}}_{\text{RC}}, \mathcal{C} \right) \neq n \right\} \mid \mathcal{A}_{\tilde{n}} \right) \right\} \\
& \quad + \mathbb{P} \left(\left\{ n_l \neq np_{m_c}(l) + \mathcal{O} \left(n^{\frac{1+\varepsilon}{2}} \right) \right\} \right) \\
& \lesssim o(1),
\end{aligned} \tag{4.74}$$

donde la cota para el orden del primer término se deduce de que vale que

$$\mathbb{P} \left(\left\{ \text{coincidencia} \left(\widehat{\mathcal{C}}_{\text{RC}}, \mathcal{C} \right) \neq n \right\} \mid \mathcal{A}_{\tilde{n}} \right) = o(1) \tag{4.75}$$

para cada \tilde{n} admisible usando el Teorema 4.3.1.

Aclaremos en este sentido que podemos asegurar las cotas fijo $a - b$ suficientemente grande para cada \tilde{n} admisible, porque podremos acotar uniformemente sobre los \tilde{n} admisibles los tamaños \tilde{n}_l . Más precisamente, al ser en ese caso $|\tilde{n}_l - np_{m_c}(l)| \leq n^{\frac{3}{4}}$ para cada $1 \leq l \leq k$, tendremos que

$$\left(p_{m_c}(l) - n^{-\frac{1}{4}} \right) n \leq \tilde{n}_l \leq \left(p_{m_c}(l) + n^{-\frac{1}{4}} \right) n \tag{4.76}$$

lo cual muestra lo afirmado. ■

Finalizaremos entonces nuestro estudio de estimadores por cortes mínimos dando paso a las notas en las cuales elaboramos sobre posibles generalizaciones, relaciones con los estimadores espectrales y las limitaciones técnicas que resta sortear para que podamos generalizar el resultado de cortes de Cheeger al modelo SBM como pudimos hacer con el corte ratio.

A continuación, en el próximo capítulo introduciremos una variante para el modelo estocástico de bloques en contextos semi-supervisados, y una nueva familia de estimadores para encarar el problema de detección de comunidades en este caso.

4.4 Notas

[4.4.1] Nota: Los estimadores de corte ratio y corte de Cheeger en la literatura
 En [vL07] se deduce que el corte ratio es equivalente es equivalente a una versión de los métodos espectrales que estudiamos en la sección 3.3. Allí en particular se discuten estrategias analíticas para demostrar la consistencia de estas minimizaciones de corte pero motivadas principalmente por grafos geométricos.

[4.4.2] Nota: Otras variantes, y generalizaciones a más que 2 comunidades
 El corte de Cheeger y el corte ratio 4.1.3 no son las únicas variantes de mínima bisección para

penalizar particiones desbalanceadas. Por ejemplo, existen también los cortes normalizados, donde se penaliza además que el subconjunto elegido no tenga una densidad de aristas alta al dividir en lugar de por el cardinal, por el volumen de cada conjunto en la partición - este consiste, para cada comunidad estimada, en la cantidad de aristas dentro de ella.

Otro aspecto en el que se podría generalizar es en cuanto a la cantidad de comunidades. Mientras que la noción de corte cuenta con una definición más aceptada, no es así para los cortes de Cheeger. Sin embargo, han sido estudiadas en la literatura en ciertas ocasiones, especialmente en el contexto de grafos geométricos.

[4.4.3] Nota: Dificultades al extender los resultados para el corte de Cheeger al caso general

Mencionamos aquí que intentamos de extender los resultados que encontramos para el corte de Cheeger al caso general. Lamentablemente, al contar con comunidades no exactamente balanceadas ocurre que, utilizando notación de la demostración del Teorema 4.2.1, tanto P como $Q + R \frac{\log(n)}{n}$ pueden contar con un término independiente que depende de la diferencia entre las proporciones de nodos que tiene cada comunidad $\alpha_2 - \alpha_1$ pero no depende de las proporciones $x_{i,j}$. Esto lleva a que no pueda utilizarse la estrategia empleada en la demostración y queda pendiente un estudio más detallado de cómo generalizar los resultados a este contexto.

5. Estimadores por extensiones armónicas

5.1 Motivación, funciones armónicas y paseos aleatorios

En los Capítulos 3 y 4 introdujimos algunos métodos para realizar detección de comunidades en modelos SBM, incluso algunos óptimos desde el punto de vista del uso de la información disponible para realizar dicha detección. Todas tenían en común que antes de realizar la predicción la única información que se tenía del modelo era el grafo (y los tamaños de las comunidades en caso de condicionar a que tenían tamaño fijo como en el modelo balanceado). ¿Qué pasa si, antes de realizar la estimación, contamos con un poco de información acerca de las comunidades?

Podemos inspirarnos de algunos razonamientos en cierto sentido naturales, que nos limitaremos a presentar en lenguaje natural por ahora y formalizaremos luego:

[5.1.1] Observación: Juegos de “tira y afloja”

Supongamos que tenemos un grupo grande de gente, partido en dos comunidades de tamaños considerables, 1 y 2. Debido al tamaño de las comunidades, no se conocen todos sus miembros entre sí dentro de cada comunidad, pero en general las personas conocen más gente dentro de sus respectivas comunidades que fuera de ellas. Si ahora conociésemos algunas pocas personas destacadas dentro de cada comunidad, ¿cómo podríamos predecir a qué comunidad pertenece el resto?

Esta situación naturalmente puede ser modelada mediante un grafo, donde los nodos son las personas y hay una arista entre cada par de personas que se conocen. Como intuición, para cada persona podríamos ver qué ocurre con sus conocidos: como conoce más gente dentro de su comunidad, podríamos decir que la comunidad de cada uno es la de la mayoría de sus vecinos, que es una idea que ya apareció previamente en este trabajo. Sin embargo, no conocemos la comunidad de mucha gente, así que no vamos a poder decidir tan simplemente por mayoría.

En este contexto, una idea superadora es ir eligiendo sucesivamente un conocido, y un conocido del conocido, y así continuar hasta llegar a una de las personas destacadas, confiando en que cada persona debería ser más cercana a las personas destacadas de su propia comunidad que de las de la comunidad ajena. Para evitar las malas

rachas, podemos realizar esta secuencia una buena cantidad de veces, y la proporción de intentos en los que llegamos desde cada persona hacia alguno de los referentes de su comunidad debería ser mayor que la proporción de veces que llegamos a los de la comunidad ajena. En este sentido, la ley de los grandes números nos diría que lo que queremos es comparar las probabilidades de que se llegue primero a una persona de la propia comunidad contra las probabilidades de llegar primero a una persona de alguna comunidad diferente. Es precisamente mediante aquellas probabilidades que podremos estimar las comunidades.

En la figura 5.1 se realiza este procedimiento para dos de las personas en el grafo, donde sabemos que 1 pertenece a la comunidad roja y 2 a la azul. Al llegar en general antes a 1 desde i , lo clasificamos a la comunidad roja, y análogamente al llegar antes a 2 desde j , lo clasificamos a la comunidad azul. Repitiendo este proceso, en color quedan marcadas las comunidades estimadas que podrían resultar de hacer una predicción de esta manera.

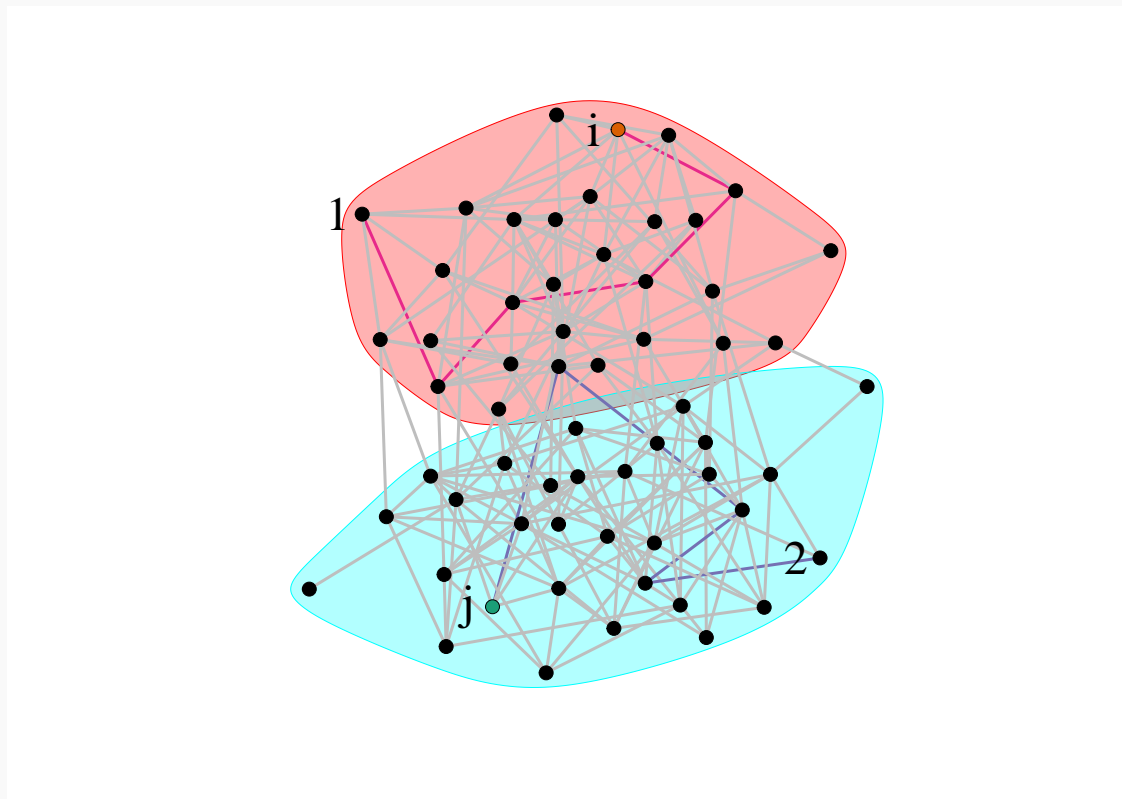


Figura 5.1: El estimador por extensiones armónicas como un juego de “tira y afloja”.

[5.1.2] Observación: Modelo del votante

Consideremos ahora unas elecciones entre dos partidos, que llamaremos 1 y 2, en las cuales participará una gran cantidad de votantes. Podemos armar con ellos un grafo, donde cada votante es un nodo y agregamos aristas entre dos votantes si se tratan entre sí. Inicialmente cada votante tiene una preferencia por alguno de los dos partidos, pero a medida que pasa el tiempo, al comentar sus preferencias con sus conocidos, pueden

cambiar de opinión al ser convencidos por alguno de ellos. Eso sí: están los miembros de cada partido que, desde ya, nunca cambiarán su parecer y siempre preferirán su partido. En estas condiciones: ¿cómo podríamos estimar la preferencia de cada persona?

En primer lugar, para los miembros de los partidos políticos está claro que debemos predecir el partido al cual están afiliados.

Pensando ahora en el resto de los votantes, sabemos que a lo largo del tiempo su opinión va a cambiar de acuerdo a la opinión de su entorno, y en nuestra predicción deberíamos tener esto en cuenta. Una vez más, el entorno a su vez dependerá de otros entornos, y como estas inter-relaciones se vuelven complejas se vuelve necesaria alguna manera elegante de capturarlas.

Podemos luego pensar en una noción que en cierto sentido implícitamente tenga en cuenta esta red de dependencias: es aquí que podemos proponer mirar la proporción de tiempo que pasa cada votante prefiriendo cada partido. En efecto, las preferencias del entorno de cada votante están guiando qué tan seguido cambia su favor hacia cada partido y luego determinarán las proporciones. Mirando intervalos de tiempo suficientemente grandes como para evitar malas rachas, obtenemos entonces una manera para estimar la preferencia de cada votante.

Resulta que estas ideas desembocan en una misma estrategia para realizar las predicciones, porque en el fondo hay una noción esencial que es que cada predicción o preferencia depende de las del entorno. ¿Cómo definir nuestro estimador, y ver que estas estrategias coinciden? Vamos a hacerlo como en el juego de “tira y afloje” discutido en la Observación 5.1.1, y daremos en la Nota 5.6.5 algunas ideas que permitirán derivar la equivalencia entre esta estrategia y la empleada en el modelo del votante que presentamos en la Observación 5.1.2.

Antes de proceder con la definición precisaremos entonces formalizar las intuiciones del “tira y afloje”, y es así que comencemos presentando los paseos al azar. Mientras que estos deben su nombre a la interpretación de realizar caminatas en un grafo de manera azarosa, en nuestro caso nos permitirán entender el proceso de sucesivas elecciones de conocidos en el juego:

[5.1.3] Definición: Paseos al azar

Dado $G = (V, E)$ un grafo, decimos que un proceso estocástico $(X_t)_{t \in \mathbb{N}_0}$ es un **paseo al azar** en G si

1. para cada tiempo t , $X_t \in V$ es un nodo en el grafo, y
2. para cada tiempo t , $\{X_t, X_{t+1}\}$ es una arista en E .

Más aún, diremos que es un **paseo al azar simétrico** en G si vale que

$$\mathbb{P}(X_{t+1} = i_{t+1} | X_t = i_t) = \begin{cases} \frac{1}{\deg(i_t)} & \text{si } i_t \sim i_{t+1}, \text{ y} \\ 0 & \text{si no.} \end{cases} \quad (5.1)$$

Si miramos con atención como evoluciona el paseo al azar simétrico, notaremos que cumple con una cierta propiedad de pérdida de memoria: para continuar una trayectoria por el grafo a partir de cierto lugar no importa por donde ya se paseó en el pasado, si no que solamente importa el lugar desde donde estamos parados actualmente. Esta propiedad caracteriza a los paseos al azar simétricos como cadenas de Markov, que nos permitirá un gran arsenal de herramientas teóricas para entender sus propiedades: de hecho, en este trabajo nos concentraremos únicamente en paseos al azar que cumplan con esta propiedad, más allá de no ser quizás simétricos. Antes de elaborar más, introduciremos primero esta importante clase de procesos estocásticos:

[5.1.4] Definición: Cadenas de Markov

Diremos que un proceso estocástico $(X_n)_{n \in \mathbb{N}_0}$ es una **cadena de Markov** en el **espacio de estados** I , con **distribución inicial** μ y **probabilidades de transición** $P \in [0, 1]^{I \times I}$ si

- $\mathbb{P}(\{X_0 = i_0\}) = \mu(\{i_0\})$ para cada $i_0 \in I$, y
- para cada $n \in \mathbb{N}$ y para cada $i_0, \dots, i_{n+1} \in I$ tales que $\mathbb{P}(\{X_j = i_j, 0 \leq j \leq n\}) > 0$ vale que

$$\begin{aligned} \mathbb{P}(\{X_{n+1} = i_{n+1}\} | \{X_j = i_j, 0 \leq j \leq n\}) & \quad (5.2) \\ &= \mathbb{P}(\{X_{n+1} = i_{n+1}\} | \{X_n = i_n\}) \\ &= P_{i_n, i_{n+1}} \end{aligned}$$

En este caso notaremos $(X_n)_n \sim \text{Markov}(\mu, P)$.

Además, escribiremos \mathbb{P}_i y \mathbb{E}_i para las probabilidades y las esperanzas asociadas a la cadena cuando la distribución inicial μ indique que $\{X_0 = i\}$ ocurre casi seguramente.

No es el foco de este trabajo presentar la teoría de estos procesos exhaustivamente, así que nos limitaremos a nombrar las propiedades que nos son relevantes; en particular nos restringiremos al caso de espacios de estados finitos y a tiempos discretos.

Ahora sí, volviendo a la Observación 5.1.1 y los juegos de “tira y afloja”, introduciremos los tiempos de parada, que nos permitirán formalizar qué significa parar el proceso de selección de conocidos al llegar a una persona a la cual le conocemos la comunidad:

[5.1.5] Definición: Tiempos de parada y tiempos de primera visita

Dada $(X_t)_{t \geq 0}$ una cadena de Markov en el espacio de probabilidad $(\Omega, \mathcal{F}, \mathbb{P})$, decimos que una variable aleatoria $T \in \mathbb{N}_0 \cup \{\infty\}$ en ese espacio es un **tiempo de parada** para $(X_t)_{t \geq 0}$ cuando vale que $\{T = n\} \in \sigma(X_1, \dots, X_n)$, es decir que $\{T = n\}$ solo depende de X_1, \dots, X_n , para cada $n \in \mathbb{N}_0$.

Cuando T es un tiempo de parada para $(X_t)_{t \in \mathbb{N}_0}$ que cumple que $\{T < \infty\}$ ocurre casi seguramente, definiremos a la cadena a tiempo T como

$$X_T(\omega) = X_{T(\omega)}(\omega) \quad (5.3)$$

para cada $\omega \in \Omega$.

Además, definiremos τ_J el **tiempo de primera visita** a $\emptyset \subsetneq J \subseteq I$ como

$$\tau_J = \begin{cases} \inf_{t \geq 0} \{X_t \in J\} & \text{si existe } t \geq 0 \text{ tal que } X_t \in J, \text{ y} \\ \infty & \text{en caso contrario.} \end{cases} \quad (5.4)$$

Una primera y necesaria observación es que esta vez los nombres no engañan, y los tiempos de visita efectivamente son tiempos de parada:

[5.1.6] Observación: Los tiempos de primera visita son tiempos de parada

Sea $(X_t)_{t \in \mathbb{N}_0} \sim \text{Markov}(\mu, P)$ con espacio de estados I , y sea τ_J el tiempo de primera visita a $\emptyset \subsetneq J \subseteq I$. Luego, podemos ver que τ_J depende únicamente de los primeros estados de la cadena, pues para cada $n \in \mathbb{N}_0$ vale que

$$\{\tau_J = n\} = \{X_0 \notin J, X_1 \notin J, \dots, X_{n-1} \notin J, X_n \in J\}. \quad (5.5)$$

En consecuencia valdrá que τ_J efectivamente es un tiempo de parada.

Para poder realizar nuestras predicciones, necesitaremos luego estudiar qué ocurre con X_τ para τ el tiempo de primera visita a las personas conocidas. Será esencial entender entonces qué ocurre con las cadenas de Markov a tiempos aleatorios, y aquí la propiedad fuerte de Markov es esencial, extendiendo la noción de pérdida de memoria también a tiempos de parada:

[5.1.7] Teorema: Propiedad fuerte de Markov - [Nor98, Teorema 1.4.2.]

Sea $(X_t)_{t \in \mathbb{N}_0} \sim \text{Markov}(\mu, P)$ y consideremos T un tiempo de parada para la cadena. Luego, condicionando a que $T \leq \infty$ y que $X_T = i$ tenemos que valen ambas que $(X_{T+t})_{t \in \mathbb{N}_0}$ es una cadena de Markov con distribución Markov (δ_i, P) , es decir comenzando en i casi seguramente, y que es independiente de X_0, \dots, X_T .

Ahora bien, para poder contar con X_τ bien definido y poder usar la propiedad fuerte de Markov precisaremos encontrar en tiempo finito a las personas destacadas. Para convencernos de que esto efectivamente ocurrirá podemos estudiar qué tan frecuentemente son visitados los

estados durante un paseo al azar, para lo cual precisaremos introducir algunas definiciones complementarias:

[5.1.8] Definición: Estados recurrentes, transitorios y cadenas de Markov irreducibles

Dada $(X_t)_{t \geq 0} \sim \text{Markov}(\mu, P)$ en el espacio de estados I , cada estado $i \in I$ se dice **recurrente** si

$$\mathbb{P}_i(\{X_t = i \text{ para infinitos } t \in \mathbb{N}_0\}) = 1 \quad (5.6)$$

y **transitorio** si

$$\mathbb{P}_i(\{X_t = i \text{ para infinitos } t \in \mathbb{N}_0\}) = 0. \quad (5.7)$$

Más aún, diremos que la cadena de Markov es recurrente, o transitoria, si todo estado en ella lo es.

Diremos que una cadena de Markov con matriz de transiciones P es **irreducible** cuando para cada par de estados $i, j \in I$ existe un camino $i = i_0, i_1, \dots, i_N = j$ de i a j tal que $P_{i_{t-1}, i_t} > 0$ para cada $1 \leq t \leq N$, o equivalentemente cuando existe $t \in \mathbb{N}$ tal que $(P^t)_{i,j} > 0$.

Ahora sí, ya podemos introducir un primer resultado para entender qué tan frecuentes son las visitas a cada estado es que en cadenas de Markov, que nos dirá que cada uno o bien es recurrente o bien es transitorio:

[5.1.9] Teorema: Los estados son o bien recurrentes o bien transitorios - [Nor98, Teorema 1.5.3.]

Sea $(X_t)_{t \geq 0} \sim \text{Markov}(\mu, P)$ una cadena de Markov en un espacio de estados I . Luego, si $i \in I$ vale que o bien i es recurrente o bien i es transitorio.

Más aún, con esta caracterización a disposición podremos mostrar que en una cadena irreducible y finita los tiempos de primera visita a cualquier estado fijo son casi seguramente finitos:

[5.1.10] Teorema: Propiedades de cadenas de Markov irreducibles en espacios de estados finitos - [Nor98, Teoremas 1.5.6. y 1.5.7.]

Sea $(X_t)_{t \geq 0} \sim \text{Markov}(\mu, P)$ una cadena de Markov irreducible en un espacio de estados I finito.

Luego, vale que todo estado $i \in I$ es recurrente, y para cada $i, j \in I$ vale que

$$\mathbb{P}_i(\{\tau_{\{j\}} < \infty\}) = 1. \quad (5.8)$$

Por último, solo basta notar que, como adelantamos, los paseos al azar que estudiaremos cumplen con los requisitos teóricos y entonces tienen tiempos de primera visita casi

seguramente finitos:

[5.1.11] Proposición: Los tiempos de primera visita son casi seguramente finitos para paseos al azar irreducibles en grafos finitos

Sea $G = (V, E)$ un grafo finito. Luego, si $(X_t)_{t \in \mathbb{N}_0} \sim \text{Markov}(\mu, P)$ es una cadena de Markov irreducible que además es un paseo al azar en G , vale que será recurrente y todo tiempo de primera visita será finito casi seguramente.

En particular, si G es conexo, el paseo al azar simétrico en G es recurrente y tiene tiempos de primera visita finitos casi seguramente.

Demostración. Lo primero es consecuencia directa del Teorema 5.1.10, notando que si $\emptyset \subsetneq J \subseteq I$ luego vale que τ_J es finito al acotar, fijando $j \in J$, de la siguiente manera:

$$\begin{aligned} \mathbb{P}(\{\tau_J \leq \infty\}) &= \sum_{i \in I} \mathbb{P}_i(\{\tau_J \leq \infty\}) \mathbb{P}(\{X_0 = i\}) \\ &\geq \sum_{i \in I} \mathbb{P}_i(\{\tau_{\{j\}} \leq \infty\}) \mathbb{P}(\{X_0 = i\}) \\ &= \sum_{i \in I} 1 \cdot \mathbb{P}(\{X_0 = i\}) \\ &= 1. \end{aligned} \tag{5.9}$$

Luego solo resta estudiar el caso del paseo al azar simétrico en G un grafo conexo. Bastará ver que estamos nuevamente en las condiciones del teorema, habiendo aclarado ya en (5.9) que todo tiempo de primera visita será finito casi seguramente en esas circunstancias. Pero como tenemos que ser irreducible equivale a que haya un camino entre cada par de nodos $i, j \in I$, lo cual ocurre pues G es conexo, mostramos lo afirmado. En efecto, esta equivalencia se debe a que como para cada $\{v, w\} \in E$ arista vale que $P_{v,w} = \frac{1}{\deg(v)} > 0$, obtendremos que cada camino entre i y j tendrá probabilidad positiva de ocurrir y luego basta hallar para cada par de nodos uno solo; por otro lado ser irreducible implica que entre cada par de nodos i y j existe un camino con probabilidad positiva de ocurrir, y en particular nos da la existencia de un camino entre ellos. ■

Volviendo a las Observaciones 5.1.1 y 5.1.2, nos queda solamente un ingrediente más, que es la noción de del entorno con la que ambas estrategias contaban y que más aún causaba que coincidiesen. Esta se formalizará mediante las que llamaremos funciones armónicas, y más generalmente Q -armónicas, en un grafo:

[5.1.12] Definición: Funciones armónicas en un grafo

Sea $G = (V, E)$ un grafo y fijemos $W \subseteq V$ un subconjunto de nodos. Luego, diremos que una función $f : V \rightarrow \mathbb{R}$ es **armónica** con **condición de borde** $g : W \rightarrow \mathbb{R}$ si

$$f(v) = \begin{cases} \frac{1}{\deg(v)} \sum_{w \sim v} f(w) & \text{si } v \notin W, \text{ y} \\ g(v) & \text{si } v \in W. \end{cases} \tag{5.10}$$

Más generalmente, diremos que f es **Q-armónica** con **condición de borde** g en W

si para Q una matriz de transición para una cadena de Markov en V vale que

$$f(v) = \begin{cases} \sum_{w \in V} Q_{v,w} f(w) & \text{si } v \notin W, \text{ y} \\ g(v) & \text{si } v \in W. \end{cases} \quad (5.11)$$

En nuestro caso, nos interesará para realizar nuestras predicciones conseguir funciones armónicas que preserven la información que conocemos, de los nodos a los cuales les conocemos la comunidad. Eso se traducirá en buscar extensiones armónicas, que afortunadamente podemos mostrar que siempre vamos a poder encontrar. De hecho, y retomando los juegos de “tira y afloja” en la Observación 5.1.1, esa misma estrategia nos hallará las extensiones, como desarrollaremos a continuación:

[5.1.13] Proposición: Extensiones armónicas

Dado $G = (V, E)$ un grafo con un subconjunto de nodos distinguidos $W \subseteq V$ y una función $g : W \rightarrow \mathbb{R}$, definimos a f en $V \setminus W$ como:

$$f(i) = \mathbb{E}_i [g(X_{\tau_W})], \quad (5.12)$$

donde $(X_t)_{t \geq 0}$ es un paseo aleatorio simétrico en G empezando en i y para τ el tiempo de primera visita a W ,

$$\tau_W = \inf_{t \geq 0} \{ X_t \in W \}. \quad (5.13)$$

Luego, $f : V \rightarrow \mathbb{R}$ es una función armónica, con condición de borde g en W . Más aún, f es la única función que verifica esta propiedad.

Más generalmente, si Q es una matriz de probabilidades de transición para una cadena de Markov en V entonces también existe una única extensión Q -armónica para g que va a estar dada por la misma fórmula f pero donde las transiciones de $(X_t)_{t \in \mathbb{N}_0}$ están dadas por Q .

Demostración. En primer lugar, notemos que f está bien definida porque τ_W es finito casi seguramente, por la Proposición 5.1.11.

Veamos ahora que f es armónica. En efecto, si $i \notin W$ luego

$$\begin{aligned} f(i) &= \mathbb{E}_i [g(X_{\tau_W})] & (5.14) \\ &= \sum_{j \sim i} \mathbb{E}_i [g(X_{\tau_W}) | \{ X_1 = j \}] \mathbb{P}(\{ X_1 = j \}) \\ &= \frac{1}{\deg(i)} \sum_{j \sim i} \mathbb{E}_i [g(X_{\tau_W}) | \{ X_1 = j \}] \\ &= \frac{1}{\deg(i)} \sum_{j \sim i} \mathbb{E}_j [g(X_{\tau_W})] \\ &= \frac{1}{\deg(i)} \sum_{j \sim i} f(j). \end{aligned}$$

Aclaremos que para la anteúltima igualdad usamos que, por la propiedad fuerte de Markov 5.1.7, condicionar a que a tiempo $T = 1$ vale $X_T = j$ obtenemos una nueva cadena de Markov con las mismas probabilidades de transición P pero empezando en j ; aquí $T = 1$ es inmediatamente además un tiempo de parada. Además, como el nodo i no vivía en W , condicionando a que a tiempo 1 estamos en j tendremos que el nodo X_{τ_W} en el cual el paseo se detiene será el mismo empezando por i a tiempo 0 o por j a tiempo 1. Por otro lado, inmediatamente $f = g$ en W , con lo cual obtenemos que f es armónica en V con condición de borde g en W .

Además, f es la única función que satisface la condición. Para mostrar esto, primero notaremos que f puede pensarse también como la solución de un sistema de ecuaciones lineales que podemos expresar como $f = \Pi Q f + v$, donde

$$Q_{i,j} = \frac{1}{\deg(i)} A_{i,j} \quad (5.15)$$

para A la matriz de adyacencia de G , donde Π es la proyección a S el subespacio dado por

$$S = \left\{ x \in \mathbb{R}^{\#V} \mid x_i = 0 \text{ para cada } i \in W \right\}, \quad (5.16)$$

y donde v es al vector dado por

$$v_i = \mathbb{1}_W(i) G(i) = \begin{cases} G(i) & \text{si } i \in W, \text{ y} \\ 0 & \text{si no.} \end{cases} \quad (5.17)$$

Luego, bastará ver que $\Delta = \mathbf{I} - \Pi Q$ es inversible para ver que f es única.

Con este objetivo en mente veremos que Δ es definido positivo, con lo cual será en particular monomorfismo y entonces como $\mathbb{R}^{\#V}$ es de dimensión finita también será isomorfismo e inversible y habremos mostrado la unicidad que queríamos. Así, notemos primero que, usando que $(\Pi Q)_{i,\bullet} = \mathbf{0}$ si $i \in W$, vale que:

$$\begin{aligned} \langle \Delta x, x \rangle &= \sum_{i \in V} (\Delta x)_i x_i & (5.18) \\ &= \sum_{i \in W} x_i^2 + \sum_{i \notin W} \sum_{j \in V} (\mathbf{I} - \Pi Q)_{i,j} x_i x_j. \end{aligned}$$

A continuación, podemos trabajar con la segunda suma para obtener que, como $\sum_{j \in V} (\Pi Q)_{i,j} = 1$ si $i \notin W$,

$$\sum_{j \in V} (\mathbf{I} - \Pi Q)_{i,j} x_i x_j = x_i^2 - \sum_{j \in V} (\Pi Q)_{i,j} x_i x_j = \sum_{j \in V} (\Pi Q)_{i,j} (x_i^2 - x_i x_j) \quad (5.19)$$

y así

$$\begin{aligned}
\sum_{i \notin W} \sum_{j \in V} (\mathbf{I} - \Pi Q)_{i,j} x_i x_j &= \sum_{i \in V, j \in V} \mathbb{1}_{V \setminus W}(i) (\Pi Q)_{i,j} (x_i^2 - x_i x_j) \\
&= \frac{1}{2} \sum_{i \in V, j \in V} \mathbb{1}_{V \setminus W}(i) (\Pi Q)_{i,j} (x_i^2 - x_i x_j) \\
&\quad + \frac{1}{2} \sum_{j \in V, i \in V} \mathbb{1}_{V \setminus W}(j) (\Pi Q)_{j,i} (x_j^2 - x_i x_j) \\
&= \sum_{i \in V, j \in V} \frac{1}{2} \left[\mathbb{1}_{V \setminus W}(i) (\Pi Q)_{i,j} + \mathbb{1}_{V \setminus W}(j) (\Pi Q)_{j,i} \right] (x_i^2 - 2x_i x_j + x_j^2) \\
&= \sum_{i \in V, j \in V} \frac{1}{2} \left[\mathbb{1}_{V \setminus W}(i) (\Pi Q)_{i,j} + \mathbb{1}_{V \setminus W}(j) (\Pi Q)_{j,i} \right] (x_i - x_j)^2.
\end{aligned} \tag{5.20}$$

(5.21)

Combinando ambas cosas, tenemos que

$$\langle \Delta x, x \rangle = \sum_{i \in W} x_i^2 + \sum_{i \in V, j \in V} \frac{1}{2} \left[\mathbb{1}_{V \setminus W}(i) (\Pi Q)_{i,j} + \mathbb{1}_{V \setminus W}(j) (\Pi Q)_{j,i} \right] (x_i - x_j)^2,$$

que es no negativa.

Por otro lado, además podemos ver que $\langle \Delta x, x \rangle = 0$ solamente cuando $x = \mathbf{0}$. En primer lugar, notamos que si $\langle \Delta x, x \rangle = 0$, entonces $x_i = 0$ para cada $i \in W$ pues al ser una suma de términos no negativos, cada término en la primera suma debe ser 0. A continuación, notamos que por conexión existe para cada $i \notin W$ algún camino que termine en un nodo $j \in W$, digamos $i = i_N, i_{N-1}, \dots, i_0 = j$, donde podremos suponer sin pérdida de generalidad que $i_t \notin W$ si $t > 0$ renombrando como $i_0 = j$ al primer nodo en el camino partiendo de i que viva en W . Notamos luego que $\mathbb{1}_{V \setminus W}(i_t) (\Pi Q)_{i_t, i_{t-1}} > 0$ para cada $1 \leq t \leq N$, y así tenemos que necesariamente $(x_{i_t} - x_{i_{t-1}})^2 = 0$ al mirar el término correspondiente en la segunda suma, que debe ser 0, y tendremos inductivamente que $x_{i_t} = 0$ para cada $1 \leq t \leq N$. Esto muestra que $x = \mathbf{0}$, y entonces la extensión armónica es única.

Al generalizar, las demostraciones son análogas, y debemos reemplazar $\frac{1}{\deg(i)} A_{i,j}$ por $Q_{i,j}$ cuando no estuviese hecho ya explícitamente. En el argumento de unicidad, lo que permite hallar el camino para el argumento inductivo es la irreducibilidad en vez que la conexión. ■

Con esto ya tenemos todas las preliminares necesarias, así que sin más preámbulos pasaremos a estudiar los estimadores por extensiones armónicas, desarrollo original de la tesis, que nos ocuparán el resto del capítulo.

5.2 El estimador por extensiones armónicas para dos comunidades balanceadas

Finalmente, definamos estimadores por extensiones armónicas. Comenzaremos por estudiarlos para dos comunidades balanceadas, en el espíritu de la Observación 5.1.1, pero esto es solamente para facilitar la intuición; pasaremos al caso general en la Sección 5.4:

[5.2.1] Definición: Estimador por extensiones armónicas para dos comunidades balanceadas

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos balanceados, y supongamos que conocemos la comunidad de los nodos en $\mathcal{K} \subseteq V$ y que $G_{\mathcal{C}}$ es conexo con alta probabilidad. Entonces, definimos el **estimador por extensiones armónicas para dos comunidades balanceadas** para \mathcal{C} que notaremos $\hat{\mathcal{C}}_{\text{HE2}}$ para cada $i \in V$ como

$$\hat{\mathcal{C}}_{\text{HE2}}(i) = \phi\left(\text{signo}(\hat{h}_i)\right), \quad (5.22)$$

donde \hat{h} es una extensión armónica de $\mathbb{1}_{\mathcal{K}_1} - \mathbb{1}_{\mathcal{K}_2}$, para $\mathcal{K}_l = \mathcal{K} \cap \mathcal{C}_l$ para cada comunidad l , y donde ϕ es la función que manda $\phi(x) = 1$ si $x \geq 0$ y $\phi(x) = 2$ si no.

Antes de seguir, un comentario técnico: el requisito de que los grafos sean conexos con alta probabilidad nos va a garantizar que la extensión armónica que pedimos en la definición sea única. Si no, podría ser que hubiesen muchas posibles, si en alguna componente conexa del grafo no contásemos con una condición de borde: en ese caso, podemos tomar \hat{h} como 0 cuando hayan muchas posibilidades. En todo caso, vamos a poder trabajar suponiendo que nuestros grafos son conexos al estudiar la precisión, porque nos interesan resultados asintóticos, y estaremos en las hipótesis del Teorema 2.4.8.

Volviendo a los estimadores por extensiones armónicas, podemos ver que codifican a cuál de las comunidades tiene más cerca cada nodo. Para cada paseo al azar se le asignará un 1 si se llega antes a nodos de la comunidad 1 y un -1 si llega antes a nodos de la comunidad 2; el valor esperado nos dirá qué referentes tiene más cerca el nodo dependiendo del signo. Podemos imaginarlo como una competencia entre los nodos distinguidos de cada comunidad en ser visitados primero, como evidencia la siguiente observación:

[5.2.2] Observación: El estimador por extensiones armónicas para dos comunidades balanceadas como un “tira y afloja”

Tenemos que por la Proposición 5.1.13, la extensión armónica \hat{h} de $\mathbb{1}_{\mathcal{K}_1} - \mathbb{1}_{\mathcal{K}_2} : \mathcal{K} \rightarrow \mathbb{R}$ cumple que

$$\hat{h}_i = \mathbb{E}_i[\mathbb{1}_{\mathcal{K}_1}(X_{\tau_{\mathcal{K}}}) - \mathbb{1}_{\mathcal{K}_2}(X_{\tau_{\mathcal{K}}})] = \mathbb{E}_i[\mathbb{1}_{\mathcal{K}_1}(X_{\tau_{\mathcal{K}}})] - \mathbb{E}_i[\mathbb{1}_{\mathcal{K}_2}(X_{\tau_{\mathcal{K}}})] \quad (5.23)$$

para cada $i \notin \mathcal{K}$.

Luego, si fijamos una comunidad l y condicionamos por si $\tau_{\mathcal{K}_1} = \tau_{\mathcal{K}}$, o si $\tau_{\mathcal{K}_2} = \tau_{\mathcal{K}}$, obtenemos que para cada $i \notin \mathcal{K}$ vale que

$$\begin{aligned} \mathbb{E}_i[\mathbb{1}_{\mathcal{K}_l}(X_{\tau_{\mathcal{K}}})] &= \mathbb{E}_i[\mathbb{1}_{\mathcal{K}_l}(X_{\tau_{\mathcal{K}}}) \mid \{\tau_{\mathcal{K}_1} = \tau_{\mathcal{K}}\}] \cdot \mathbb{P}_i(\{\tau_{\mathcal{K}_1} = \tau_{\mathcal{K}}\}) \\ &\quad + \mathbb{E}_i[\mathbb{1}_{\mathcal{K}_l}(X_{\tau_{\mathcal{K}}}) \mid \{\tau_{\mathcal{K}_2} = \tau_{\mathcal{K}}\}] \cdot \mathbb{P}_i(\{\tau_{\mathcal{K}_2} = \tau_{\mathcal{K}}\}) \\ &= \mathbb{P}_i(\{\tau_{\mathcal{K}_l} = \tau_{\mathcal{K}}\}), \end{aligned} \quad (5.24)$$

y $\mathbb{E}_i[\mathbb{1}_{\mathcal{K}_l}(X_{\tau_{\mathcal{K}}})]$ es una medida de qué tan cerca está el nodo i de los nodos distinguidos

en \mathcal{K}_i comparado al resto de los nodos en \mathcal{K} .

En particular, deducimos que

$$\hat{h}_i = \mathbb{P}_i(\{\tau_{\mathcal{K}_1} = \tau_{\mathcal{K}}\}) - \mathbb{P}_i(\{\tau_{\mathcal{K}_2} = \tau_{\mathcal{K}}\})$$

y es así que \hat{h}_i nos da el resultado de la pulseada entre las personas conocidas de la comunidad 1 y las personas conocidas de la comunidad 2 de mostrar que están más cerca de i , determinando el signo el resultado del “tira y afloje”.

Nos gustaría ahora convencernos rigurosamente de que, más allá de que el método haya surgido de manera muy natural, sus estimaciones son consistentes. Como hicimos en las secciones 3.3, 3.4 y 3.5 al estudiar los métodos espectrales, una vez más será de utilidad teórica estudiar qué ocurre cuando trabajamos en vez que con el grafo empírico con el cual contamos, con el grafo pesado que resulta de tomarle esperanza a la matriz de adyacencia, comenzando como en la Observación 3.3.1. Si además recuperásemos alguna manera de asegurar que las soluciones empíricas se parecerán asintóticamente a las soluciones teóricas, y estas fuesen suficientemente precisas, vamos a obtener la consistencia del estimador. Notamos que como las extensiones armónicas no son autovectores no vamos a tener a disposición las herramientas teóricas que sí teníamos anteriormente y hará falta desarrollar las propias. Sin embargo, el Teorema 3.3.4 que nos habla de cercanía de matrices de adyacencia, y el Lema 3.5.4 que nos habla acerca de la vecindad de cada nodo volverán a jugar un papel importante en este capítulo.

Antes de estudiar las soluciones teóricas, primero fijaremos un poco de notación que nos será de utilidad en adelante:

[5.2.3] Notación: Extensiones teóricas y empíricas, y nodos conocidos

Dado $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$, llamaremos nuevamente \hat{A} a la matriz de adyacencia del modelo, la matriz de adyacencia empírica, y $A = \mathbb{E}[\hat{A}]$ a la esperanza de la matriz de adyacencia, la matriz teórica.

Llamaremos \hat{D} a la matriz de grados, una matriz diagonal con $\hat{D}_{i,i} = \widehat{\text{deg}}_n(i)$ para cada nodo i , y llamaremos $\hat{Q} = \hat{D}^{-1}\hat{A}$ a la matriz de probabilidades de transición de un paseo aleatorio simétrico por G .

Como contrapartida, llamaremos D a la matriz diagonal de grados esperados, con $D_{i,i} = \text{deg}_n(i)$ para cada nodo i . Además, notaremos $Q = D^{-1}A$ a la matriz de probabilidades de transición teórica, para un paseo aleatorio en el grafo completo.

Recordemos además que \mathcal{K} es el subconjunto de los nodos a los cuales les conocemos la comunidad, y que además $\mathcal{K}_l = \mathcal{C}_l \cap \mathcal{K}$ son los nodos conocidos en la comunidad l . Luego, como en los resultados que siguen $\#\mathcal{K}_l$ será constante en l , notaremos también $\#\mathcal{K}_l = g(n)$ y $\varepsilon_n = \frac{g(n)}{n}$ a la cantidad y la fracción de nodos conocidos por comunidad respectivamente.

Por último, llamaremos

$$S = \{x \in \mathbb{R}^n \mid x_i = 0 \text{ para cada } i \in \mathcal{K}\} \quad (5.25)$$

y Π a la proyección a dicho subespacio.

Será conveniente a lo largo del capítulo aprovechar que contaremos con expresiones matriciales para lidiar con las extensiones armónicas. Condicionando convenientemente, podremos lograr expresiones más agradables, como observamos a continuación:

[5.2.4] Observación: Extensiones armónicas en notación matricial

Recordando la Proposición 5.1.13, encontrar la extensión armónica no es más que resolver un sistema de ecuaciones lineales. Para obtener expresiones agradables, en el espíritu de la Observación 2.4.7, podremos condicionar para fijar los nodos conocidos además de los nodos de cada comunidad.

En efecto, basta entender el modelo fijando que los primeros $\frac{n}{2}$ nodos forman \mathcal{C}_1 y los últimos $\frac{n}{2}$ forman \mathcal{C}_2 , y que los primeros $g(n)$ nodos sean \mathcal{K}_1 y los últimos $\frac{n}{2}$ sean \mathcal{K}_2 , para obtener resultados para el modelo BSBM.

Podremos luego recordar que al ser una extensión armónica de $\mathbb{1}_{\mathcal{K}_1} - \mathbb{1}_{\mathcal{K}_2} : \mathcal{K} \rightarrow \mathbb{R}$ y como mencionamos en la demostración 5.1.13, \hat{h} será solución de un sistema de ecuaciones. En este caso, verificamos que el sistema es $\hat{h} = \Pi \hat{Q} \hat{h} + v$ para v que codifica el dato de borde y estará dado según como ordenamos los nodos por:

$$v = \begin{pmatrix} \mathbf{1}_{g(n)} \\ \mathbf{0}_{\frac{n}{2}-g(n)} \\ \mathbf{0}_{\frac{n}{2}-g(n)} \\ -\mathbf{1}_{g(n)} \end{pmatrix}. \quad (5.26)$$

Además, notamos que las expresiones para A y Π con este orden para los nodos serán

$$A = \begin{pmatrix} p_n \left(\mathbf{J}_{\frac{n}{2} \times \frac{n}{2}} - \mathbf{I}_{\frac{n}{2}} \right) & q_n \mathbf{J}_{\frac{n}{2} \times \frac{n}{2}} \\ q_n \mathbf{J}_{\frac{n}{2} \times \frac{n}{2}} & p_n \left(\mathbf{J}_{\frac{n}{2} \times \frac{n}{2}} - \mathbf{I}_{\frac{n}{2}} \right) \end{pmatrix}, \quad (5.27)$$

y

$$\Pi = \begin{pmatrix} \mathbf{0}_{g(n) \times g(n)} & \mathbf{0}_{g(n) \times (\frac{n}{2}-g(n))} & \mathbf{0}_{g(n) \times (\frac{n}{2}-g(n))} & \mathbf{0}_{g(n) \times g(n)} \\ \mathbf{0}_{(\frac{n}{2}-g(n)) \times g(n)} & \mathbf{I}_{\frac{n}{2}-g(n)} & \mathbf{0}_{(\frac{n}{2}-g(n)) \times (\frac{n}{2}-g(n))} & \mathbf{0}_{(\frac{n}{2}-g(n)) \times g(n)} \\ \mathbf{0}_{(\frac{n}{2}-g(n)) \times g(n)} & \mathbf{0}_{(\frac{n}{2}-g(n)) \times (\frac{n}{2}-g(n))} & \mathbf{I}_{\frac{n}{2}-g(n)} & \mathbf{0}_{(\frac{n}{2}-g(n)) \times g(n)} \\ \mathbf{0}_{g(n) \times g(n)} & \mathbf{0}_{g(n) \times (\frac{n}{2}-g(n))} & \mathbf{0}_{g(n) \times (\frac{n}{2}-g(n))} & \mathbf{0}_{g(n) \times g(n)} \end{pmatrix}. \quad (5.28)$$

Ahora sí, podemos hallar una fórmula para la extensión armónica teórica, que una vez más distinguirá correctamente los nodos de cada comunidad:

[5.2.5] Lema: Extensión armónica teórica para dos comunidades balanceadas

En la notación de 5.2.3, y suponiendo que $g(n) > 0$, vale que la única solución a $h = \Pi Qh + v$ está dada por

$$h = \begin{pmatrix} \mathbf{1}_{g(n)} \\ \frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)} \mathbf{1}_{\frac{n}{2} - g(n)} \\ -\frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)} \mathbf{1}_{\frac{n}{2} - g(n)} \\ -\mathbf{1}_{g(n)} \end{pmatrix} = \begin{pmatrix} \mathbf{1}_{g(n)} \\ \frac{g(n)(a-b)}{nb + g(n)(a-b)} \mathbf{1}_{\frac{n}{2} - g(n)} \\ -\frac{g(n)(a-b)}{nb + g(n)(a-b)} \mathbf{1}_{\frac{n}{2} - g(n)} \\ -\mathbf{1}_{g(n)} \end{pmatrix} \quad (5.29)$$

en bloques.

En particular, como $0 < \frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)}$, el estimador dado por

$$\widehat{C}(i) = \phi(\text{signo}(h_i)) \quad (5.30)$$

detecta exactamente las comunidades.

Demostración. Sabemos que h existe y es única porque podemos aplicar la Proposición 5.1.13 para un paseo al azar en el grafo completo con matriz de probabilidades de transición Q . Luego, basta mostrar que h es solución del sistema de ecuaciones para hallar la solución teórica.

Para ello, propondremos h en bloques de la forma:

$$h = \begin{pmatrix} \mathbf{1}_{g(n)} \\ \lambda \mathbf{1}_{\frac{n}{2} - g(n)} \\ -\lambda \mathbf{1}_{\frac{n}{2} - g(n)} \\ -\mathbf{1}_{g(n)} \end{pmatrix} \quad (5.31)$$

para $\lambda \in \mathbb{R}$ y veremos que, despejando λ , obtendremos la expresión deseada.

En efecto, tomando por ejemplo la ecuación correspondiente a la coordenada $g(n) + 1$ -ésima de $h = \Pi Qh + v$ obtenemos que

$$\begin{aligned} \lambda &= \sum_{i=1}^{g(n)} \frac{1}{\deg_n(i)} p_n + \sum_{i=g(n)+2}^{\frac{n}{2}} \lambda \frac{1}{\deg_n(i)} p_n \\ &\quad - \sum_{i=\frac{n}{2}+1}^{n-g(n)} \lambda \frac{1}{\deg_n(i)} q_n - \sum_{i=n-g(n)+1}^n \frac{1}{\deg_n(i)} q_n, \end{aligned} \quad (5.32)$$

y así despejando λ tendremos que

$$\lambda = \frac{\sum_{i=1}^{g(n)} \frac{1}{\deg_n(i)} p_n - \sum_{i=n-g(n)+1}^n \frac{1}{\deg_n(i)} q_n}{1 - \sum_{i=g(n)+2}^{\frac{n}{2}} \frac{1}{\deg_n(i)} p_n - \sum_{i=\frac{n}{2}+1}^{n-g(n)} \frac{1}{\deg_n(i)} q_n}, \quad (5.33)$$

que a su vez multiplicando arriba y abajo por $\deg_n(i) = \left(\frac{n}{2} - 1\right) p_n + \left(\frac{n}{2}\right) q_n$ nos permitirá obtener como queríamos que:

$$\begin{aligned} \lambda &= \frac{g(n)p_n - g(n)q_n}{\left[\left(\frac{n}{2} - 1\right) p_n + \left(\frac{n}{2}\right) q_n\right] - \left(\frac{n}{2} - g(n) - 1\right) p_n + \left(\frac{n}{2} - g(n)\right) q_n} \\ &= \frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)}. \end{aligned} \quad (5.34)$$

■

Continuando con la estrategia propuesta, nos queda entonces mostrar que la extensión armónica empírica estará suficientemente cerca de la extensión armónica teórica asintóticamente en n , cosa que haremos en la siguiente sección.

5.3 Consistencia débil del estimador por extensiones armónicas

Como ya comentamos anteriormente, no tendremos a disposición las herramientas teóricas relacionadas con autovalores del Capítulo 3, así que para mostrar la consistencia débil del estimador tendremos que mostrar nuestra propia cota para la distancia entre las extensiones empíricas y la teórica:

[5.3.1] Proposición: La solución empírica se parece a la solución teórica

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ donde $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ son tales que $a > b$ son reales no negativos y que $\frac{a+b}{2} > 1$. Supongamos además que $\varepsilon_n = \omega\left(\frac{1}{\sqrt{\log(n)}}\right)$. Luego, vale que

$$\mathbb{P}\left(\left\{\|\hat{h} - h\|_2 \leq \mathcal{O}\left(\frac{1}{\varepsilon_n \sqrt{\log(n)}}\right) \|h\|_2\right\}\right) \geq 1 - o(1). \quad (5.35)$$

De aquí a la consistencia débil, un argumento análogo a 3.4.2 pero ahora usando la cota del lema anterior completa la demostración:

[5.3.2] Teorema: Consistencia débil del estimador por extensiones armónicas para dos comunidades balanceadas $\hat{\mathcal{C}}_{\text{HE2}}$.

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos balanceados tales que $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para ciertos números positivos $a > b$. Supongamos además que $\frac{a+b}{2} > 1$, y también $\varepsilon_n = \omega\left(\frac{1}{\log(n)^{\frac{1}{3}}}\right)$.

Luego, tendremos que el estimador $\hat{\mathcal{C}}_{\text{HE2}}$ es débilmente consistente.

Más aún, el conjunto de nodos mal clasificados $\mathcal{B}(\hat{h})$ cumple que

$$\mathbb{P}\left(\left\{\#\left(\mathcal{B}(\hat{h})\right) \leq \mathcal{O}\left(\frac{1}{\varepsilon_n^3 \log(n)}\right) n\right\}\right) \geq 1 - o(1). \quad (5.36)$$

Demostración. Llamemos $\mathcal{B}(\hat{h})$ al conjunto de nodos mal clasificados por nuestra estimación, y veamos que tiene $o(n)$ miembros con probabilidad alta.

Si $i \in \mathcal{B}(\hat{h})$, entonces como $\mathcal{C}(i) \neq \hat{\mathcal{C}}_{\text{HE2}}(i)$ tendremos que \hat{h}_i y h_i tienen signos distintos, y así necesariamente $|\hat{h}_i - h_i| \geq |h_i|$. Luego, usando la cota que nos provee el Lema 5.3.1, que $|h_i| = \Theta(\varepsilon_n)$ por el Lema 5.2.5 y además que

$$\|h\|_2^2 = 2g(n) + (n - 2g(n)) \left(\frac{g(n)(a-b)}{nb + g(n)(a-b)} \right)^2 \lesssim n\varepsilon_n + n\varepsilon_n^2 \lesssim n\varepsilon_n, \quad (5.37)$$

tenemos que

$$\varepsilon_n^2 \#(\mathcal{B}(\hat{h})) \lesssim \sum_{i=1}^n |\hat{h}_i - h_i|^2 = \|\hat{h} - h\|_2^2 \lesssim \frac{\|h\|_2^2}{\varepsilon_n^2 \log(n)} \lesssim \frac{n\varepsilon_n}{\varepsilon_n^2 \log(n)} = \frac{1}{\varepsilon_n \log(n)} n. \quad (5.38)$$

Luego podemos despejar que

$$\#(\mathcal{B}(\hat{h})) \lesssim \frac{1}{\varepsilon_n^3 \log(n)} n = o(n), \quad (5.39)$$

como por hipótesis $\varepsilon_n^3 \log(n) \xrightarrow{n \rightarrow \infty} \infty$, lo cual muestra lo que queríamos. ■

Ahora sí, sin más preámbulos demostremos la Proposición 5.3.1.

Demostración. La demostración es técnica, y consistirá en acotar repetidas veces hasta obtener la cota deseada. Aclaremos que, como $\frac{a+b}{2} > 1$, el grafo será conexo con alta probabilidad por el Teorema 2.4.8, y entonces supondremos que estamos trabajando con grafos conexos durante la demostración, o lo que es lo mismo, mostraremos que

$$\mathbb{P} \left(\left\{ G \text{ es conexo} \right\} \cap \left\{ \|\hat{h} - h\|_2 \leq \mathcal{O} \left(\frac{1}{\varepsilon_n \sqrt{\log(n)}} \right) \|h\|_2 \right\} \right) \geq 1 - o(1). \quad (5.40)$$

Además, la dividiremos en pasos para facilitar su lectura.

Paso 1: en primer lugar, buscaremos llegar a la desigualdad

$$\varepsilon_n \|\hat{h} - h\|_2^2 \leq |(\hat{h} - h)^t (\mathbf{I} - \Pi Q) (\hat{h} - h)|. \quad (5.41)$$

Para aquello, primero elevamos al cuadrado para tener un mejor manejo de la norma de la diferencia y entonces podemos sumar y restar ΠQ para obtener

$$\begin{aligned} \|\hat{h} - h\|_2^2 &= (\hat{h} - h)^t (\hat{h} - h) \\ &= (\hat{h} - h)^t (\mathbf{I} - \Pi Q) (\hat{h} - h) + (\hat{h} - h)^t \Pi Q (\hat{h} - h). \end{aligned} \quad (5.42)$$

Luego, como $\hat{h} - h \in S$ tenemos que $\Pi(\hat{h} - h) = \hat{h} - h$ y entonces podemos acotar el segundo término en (5.42) del siguiente modo:

$$|(\hat{h} - h)^t \Pi Q (\hat{h} - h)| = |(\hat{h} - h)^t \Pi Q \Pi (\hat{h} - h)| \leq \|\Pi Q \Pi\|_2 \|\hat{h} - h\|_2^2. \quad (5.43)$$

Ahora bien, $\Pi Q \Pi$ puede expresarse en bloques como

$$\Pi Q \Pi = D^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & B^{1,1} & B^{1,2} & \mathbf{0} \\ \mathbf{0} & B^{1,2} & B^{2,2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \quad (5.44)$$

para B una matriz en $\mathbb{R}^{(n-2g(n)) \times (n-2g(n))}$ dada en bloques $B^{t,l} \in \mathbb{R}^{(\frac{n}{2}-g(n)) \times (\frac{n}{2}-g(n))}$ por:

$$B^{t,l} = \begin{cases} tp_n (\mathbf{J} - \mathbf{I}) & \text{si } t = l, \\ q_n \mathbf{J} & \text{si } t \neq l; \end{cases} \quad (5.45)$$

es decir

$$B = \left(\begin{array}{cccc|cccc} 0 & p_n & \dots & p_n & q_n & q_n & \dots & q_n \\ p_n & 0 & \dots & p_n & q_n & q_n & \dots & q_n \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ p_n & \dots & p_n & 0 & q_n & q_n & \dots & q_n \\ \hline q_n & q_n & \dots & q_n & 0 & p_n & \dots & p_n \\ q_n & q_n & \dots & q_n & p_n & 0 & \dots & p_n \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ q_n & q_n & \dots & q_n & p_n & \dots & p_n & 0 \end{array} \right). \quad (5.46)$$

Llamamos $\beta_n = \left(\frac{n}{2} - g(n) - 1\right) p_n + \left(\frac{n}{2} - g(n)\right) q_n$ a la suma de los coeficientes en cada fila de B . Luego notando que $D^{-1} = \text{deg}_n(1)^{-1} \mathbf{I}$, porque el grado teórico es el mismo para cada nodo, y que $\frac{1}{\beta_n} B$ tiene norma más chica que 1 al ser una matriz estocástica, pues sus filas suman 1, valdrá que

$$\|\Pi Q \Pi\|_2 \leq \|D^{-1}\|_2 \beta_n \left\| \frac{1}{\beta_n} B \right\|_2 \leq \frac{\beta_n}{\text{deg}_n(1)} \lesssim 1 - \varepsilon_n. \quad (5.47)$$

Entonces acotando por arriba el segundo término en (5.42) por las cotas conseguidas en (5.43) y (5.47), obtenemos que

$$\|\hat{h} - h\|_2^2 \leq |(\hat{h} - h)^t (\mathbf{I} - \Pi Q) (\hat{h} - h)| + (1 - \varepsilon_n) \|\hat{h} - h\|_2^2, \quad (5.48)$$

con lo que restando el segundo término en ambos lados obtenemos la primera expresión (5.41) que buscábamos.

Paso 2: ahora buscaremos simplificar la cota en (5.41) aún más a

$$\varepsilon_n \|\hat{h} - h\|_2^2 \leq |(\hat{h} - h)^t (\Pi \hat{Q} - \Pi Q) \hat{h}|. \quad (5.49)$$

En primer lugar, notamos que como $\hat{h} - h \in S$, luego $(\hat{h} - h)^t v = 0$. Así, como además $h = \Pi Q h + v$, $(\mathbf{I} - \Pi Q) h = v$ y entonces

$$\begin{aligned} |(\hat{h} - h)^t (\mathbf{I} - \Pi Q) (\hat{h} - h)| &\leq |(\hat{h} - h)^t (\mathbf{I} - \Pi Q) \hat{h}| + |(\hat{h} - h)^t (\mathbf{I} - \Pi Q) h| \\ &= |(\hat{h} - h)^t (\mathbf{I} - \Pi Q) \hat{h}|. \end{aligned} \quad (5.50)$$

Análogamente, también $(\mathbf{I} - \Pi\widehat{Q})\hat{h} = v$ y podemos sumar y restar $\Pi\widehat{Q}$ para obtener

$$\begin{aligned} |(\hat{h} - h)^t(\mathbf{I} - \Pi Q)\hat{h}| &\leq |(\hat{h} - h)^t(\mathbf{I} - \Pi\widehat{Q})\hat{h}| + |(\hat{h} - h)^t(\Pi\widehat{Q} - \Pi Q)\hat{h}| \\ &= |(\hat{h} - h)^t(\Pi\widehat{Q} - \Pi Q)\hat{h}|. \end{aligned} \quad (5.51)$$

Juntando ambas cotas, llegamos a la expresión (5.49) como queríamos.

Paso 3: Simplificamos una última vez partiendo de (5.49) para conseguir

$$\varepsilon_n \|\hat{h} - h\|_2 \leq \|\widehat{D}^{-1}\|_2 \|\widehat{A} - A\|_2 \|\hat{h}\|_2 + \|\Pi(\widehat{D}^{-1} - D^{-1})A\hat{h}\|_2, \quad (5.52)$$

notando que como

$$\Pi\widehat{Q} - \Pi Q = \Pi\widehat{D}^{-1}(\widehat{A} - A) + \Pi(\widehat{D}^{-1} - D^{-1})A \quad (5.53)$$

podemos reemplazar en (5.49) y acotar

$$\begin{aligned} \varepsilon_n \|\hat{h} - h\|_2^2 &\leq |(\hat{h} - h)^t [\Pi\widehat{D}^{-1}(\widehat{A} - A) + \Pi(\widehat{D}^{-1} - D^{-1})A] \hat{h}| \\ &\leq \|\hat{h} - h\|_2 \left(\|\Pi\|_2 \|\widehat{D}^{-1}\|_2 \|\widehat{A} - A\|_2 \|\hat{h}\|_2 + \|\Pi(\widehat{D}^{-1} - D^{-1})A\hat{h}\|_2 \right) \end{aligned} \quad (5.54)$$

para obtener (5.52), donde usamos que Π al ser proyección tiene norma acotada por 1.

Paso 4: A continuación a partir de (5.52) pasaremos a acotar cada término por cotas del orden $\frac{1}{\sqrt{\log(n)}} \|\hat{h}\|_2$ que valgan con suficientemente alta probabilidad para concluir el resultado. En efecto, en ese caso tendremos que

$$\varepsilon_n \|\hat{h} - h\|_2 \lesssim \frac{1}{\sqrt{\log(n)}} \|\hat{h}\|_2 \leq \frac{1}{\sqrt{\log(n)}} (\|\hat{h} - h\|_2) + \frac{1}{\sqrt{\log(n)}} \|h\|_2 \quad (5.55)$$

con lo cual agrupando las normas $\|\hat{h} - h\|_2$ vemos que existe alguna constante $C > 0$ tal que

$$\left(\varepsilon_n - C \frac{1}{\sqrt{\log(n)}} \right) \|\hat{h} - h\|_2 \leq C \frac{1}{\sqrt{\log(n)}} \|h\|_2 \quad (5.56)$$

y usando nuestra hipótesis sabremos que $\varepsilon_n \sqrt{\log(n)} \xrightarrow[n \rightarrow \infty]{} \infty$ y en consecuencia tenemos que

$$(1 - o(1)) \|\hat{h} - h\|_2 = \left(1 - C \frac{1}{\varepsilon_n \sqrt{\log(n)}} \right) \|\hat{h} - h\|_2 \leq C \frac{1}{\varepsilon_n \sqrt{\log(n)}} \|h\|_2 \quad (5.57)$$

de lo cual se sigue la cota que queríamos demostrar.

Volviendo a la expresión en (5.52), acotaremos cada factor en el primer término. La matriz \widehat{D}^{-1} es diagonal con coeficientes $\widehat{D}_{i,i}^{-1} = \widehat{\deg}_n(i)^{-1}$ para cada $1 \leq i \leq n$, con lo que su norma estará dada por el mayor coeficiente en módulo entre ellos, correspondiente al mínimo grado de un nodo en el grafo. Por la Proposición 2.4.11 sabemos que cada grado es de orden $\log(n)$ y luego $\|\widehat{D}^{-1}\|_2 \lesssim \frac{1}{\log(n)}$, con probabilidad mayor que $1 - o(1)$. Además, el Teorema 3.3.4 nos garantiza que $\|\widehat{A} - A\|_2 \lesssim \sqrt{\log(n)}$ con probabilidad $1 - o(1)$. Concluimos que con probabilidad mayor que $1 - o(1)$ vale como queríamos ver que

$$\|\widehat{D}^{-1}\|_2 \|\widehat{A} - A\|_2 \|h\|_2 \lesssim \frac{1}{\log(n)} \sqrt{\log(n)} \|h\|_2 = \frac{1}{\sqrt{\log(n)}} \|h\|_2. \quad (5.58)$$

Resta solamente acotar el segundo término en (5.52) con alta probabilidad. Un cálculo directo nos muestra que

$$\left(\Pi \left(\widehat{D}^{-1} - D^{-1}\right) A\hat{h}\right)_i = \begin{cases} \frac{\widehat{\deg}_n(i) - \deg_n(i)}{\widehat{\deg}_n(i) \deg_n(i)} \rho_i & \text{si } g(n) < i \leq n - g(n), \\ 0 & \text{si no.} \end{cases} \quad (5.59)$$

para $\rho_i = \sum_{j \in \mathcal{C}_{\mathcal{C}(i)}, j \neq i} p_n \hat{h}_j + \sum_{j \in \mathcal{C}_{\mathcal{C}(i)}} q_n \hat{h}_j$. Ahora usando la desigualdad de Cauchy-Schwarz y que $a > b$, obtenemos que

$$|\rho_i| \leq a \frac{\log n}{\sqrt{n}} \|\hat{h}\|_2. \quad (5.60)$$

Luego, usando el cálculo explícito en (5.59) junto con estas cotas podemos ver que

$$\left\| \Pi \left(\widehat{D}^{-1} - D^{-1}\right) A\hat{h} \right\|_2 \leq \max_{1 \leq i \leq n} \left\{ \frac{1}{\deg_n(i) \widehat{\deg}_n(i)} \right\} \left\| (\widehat{D} - D) \mathbf{1}_n \right\|_2 \left(a \frac{\log n}{\sqrt{n}} \|\hat{h}\|_2 \right), \quad (5.61)$$

donde $(\widehat{D} - D) \mathbf{1}_n$ no es más que la matriz diagonal que tiene la resta de los grados. Ahora bien, por el Lema 2.4.11, tanto $\deg_n(i)$ como $\widehat{\deg}_n(i)$ son de orden $\log(n)$ con probabilidad mayor que $1 - o(1)$, por lo que el máximo se acota por un factor de orden $\frac{1}{\log(n)^2}$. Más aún, usando que $(\widehat{D} - D) \mathbf{1}_n = (\widehat{A} - A) \mathbf{1}_n$ y el Teorema 3.3.4 podremos acotar con probabilidad mayor que $1 - o(n^{-r})$ para conseguir que

$$\begin{aligned} \left\| (\widehat{D} - D) \mathbf{1}_n \right\|_2 &\leq \left\| \widehat{A} - A \right\|_2 \|\mathbf{1}_n\|_2 \\ &\lesssim \sqrt{\log(n)} \sqrt{n}. \end{aligned} \quad (5.62)$$

Finalmente juntando todas estas cotas auxiliares y acotando en (5.61) obtendremos que con probabilidad mayor que $1 - o(n^{-r})$ vale que

$$\begin{aligned} \left\| \Pi \left(\widehat{D} - D\right) A\hat{h} \right\|_2 &\lesssim \frac{1}{\log(n)^2} \left(\sqrt{\log(n)} \sqrt{n} \right) \left(a \frac{\log(n)}{\sqrt{n}} \|\hat{h}\|_2 \right) \\ &\lesssim \frac{1}{\sqrt{\log(n)}} \|\hat{h}\|_2, \end{aligned} \quad (5.63)$$

lo que completa la demostración. ■

Habiendo mostrado que el estimador por extensiones armónicas para dos comunidades balanceadas es débilmente consistente, nos resta como habíamos prometido pasar al caso general con k comunidades no necesariamente balanceadas, que haremos en la siguiente sección. Esta generalización resultará interesante no solo por sí misma si no que porque nos permitirá también mostrar resultados de consistencia débil para el modelo SBM y, más adelante, de consistencia fuerte.

5.4 El estimador por extensiones armónicas en el caso general

Como adelantamos, buscaremos ahora estimadores por extensiones armónicas pero para una cantidad arbitraria k de comunidades, y no necesariamente balanceadas. Para poder

generalizar correctamente, precisaremos una noción de cercanía a cada comunidad, lo cual nos llevará a precisar extensiones armónicas en \mathbb{R}^k más que en la recta real, como discutiremos a continuación:

[5.4.1] Observación: Generalización a k comunidades

Recordemos que el estimador por extensiones armónicas formaliza una competencia entre los nodos conocidos de cada comunidad a ver cuáles están más cerca de cada nodo fijo. Un ejemplo claro de esto se halla en la Observación 5.2.2, pues (5.25) nos dice que

$$\hat{h}_i = \mathbb{P}_i(\{\tau_{\mathcal{K}_1} < \tau_{\mathcal{K}_2}\}) - \mathbb{P}_i(\{\tau_{\mathcal{K}_2} < \tau_{\mathcal{K}_1}\})$$

para cada $i \notin \mathcal{K}$. Sin embargo, mirando con cuidado las cuentas en dicha observación verificamos que la información fundamental que precisamos para poder conocer esta diferencia es la probabilidad de que el paseo al azar visite primero los nodos en \mathcal{K}_l antes que el resto de los nodos en \mathcal{K} , para cada comunidad l , que deducimos en (5.24) que está dada por la fórmula

$$\mathbb{E}_i[\mathbb{1}_{\mathcal{K}_l}(X_{\tau_{\mathcal{K}}})] = \mathbb{P}_i[\{\tau_{\mathcal{K}_l} = \tau_{\mathcal{K}}\}]. \quad (5.64)$$

Contando con estas cantidades, podremos en el espíritu del estimador para dos comunidades balanceadas estimar que la comunidad más cercana es aquella para la cual esta probabilidad sea más alta.

Naturalmente, esta información no se representa adecuadamente en una recta pues consta de k cantidades distintas; tampoco podremos utilizar solamente una diferencia porque esta nos ayudará a distinguir solamente entre dos comunidades, pero precisamos distinguir entre cada par. Sin embargo, sí podremos representarla en el espacio \mathbb{R}^k , tomando

$$\hat{h}_{i,l} = \mathbb{E}_i[\mathbb{1}_{\mathcal{K}_l}(X_{\tau_{\mathcal{K}}})] = \mathbb{P}_i[\{\tau_{\mathcal{K}_l} = \tau_{\mathcal{K}}\}]. \quad (5.65)$$

Más aún, recordando que también podemos pensar en extensiones armónicas para funciones que llegan allí, tendremos que $\hat{h}_{i,\bullet}$ será la extensión armónica a V de $\sum_{l=1}^k \mathbb{1}_{\mathcal{K}_l} e_l : \mathcal{K} \rightarrow \mathbb{R}^k$, donde e_1, \dots, e_k es la base canónica de \mathbb{R}^k . En efecto, como para cada $i \in V$ vale que

$$\hat{h}_{i,\bullet} = \sum_{l=1}^k \hat{h}_{i,l} e_l = \sum_{l=1}^k \mathbb{E}_i[\mathbb{1}_{\mathcal{K}_l}(X_{\tau_{\mathcal{K}}})] e_l = \mathbb{E}_i\left[\sum_{l=1}^k \mathbb{1}_{\mathcal{K}_l}(X_{\tau_{\mathcal{K}}}) e_l\right], \quad (5.66)$$

la Proposición 5.1.13 nos garantiza que esta es la única extensión armónica con esta condición de borde al estar dada por la fórmula correcta.

Queda entonces motivado que la siguiente es una forma natural de generalizar el estimador:

[5.4.2] Definición: Estimador por extensiones armónicas en el caso general

Sea $G_{\mathcal{C}}$ una familia de modelos estocásticos de bloques con k comunidades, y supongamos que conocemos la comunidad de los nodos en $\mathcal{K} \subseteq V$ y que $G_{\mathcal{C}}$ son conexos con alta probabilidad. Entonces, definimos el **estimador por extensiones armónicas** para \mathcal{C} que notaremos $\hat{\mathcal{C}}_{\text{HE}}$ para cada $i \in V$ como

$$\hat{\mathcal{C}}_{\text{HE}}(i) = \underset{1 \leq l \leq k}{\operatorname{argmín}} \left\{ \|e_l - \hat{h}_{i,\bullet}\|_2 \right\} = \underset{1 \leq l \leq k}{\operatorname{argmáx}} \{ \hat{h}_{i,l} \}, \quad (5.67)$$

donde \hat{h} es una extensión armónica a V de $\sum_{l=1}^k \mathbf{1}_{\mathcal{K}_l} e_l : \mathcal{K} \rightarrow \mathbb{R}^k$, para e_1, \dots, e_k la base canónica de \mathbb{R}^k .

Una vez más, el requisito de que los modelos sean conexos con alta probabilidad es para garantizar la unicidad para la extensión armónica, que podemos tomar como $\hat{h} = 0$ en otro caso, y que nuevamente resultará meramente en algunas aclaraciones en las demostraciones. Antes de seguir, podemos convencernos de que ambas maneras de definir el estimador son, en efecto, equivalentes:

[5.4.3] Observación: Equivalencia entre las dos nociones de cercanía para el estimador por extensiones armónicas en el caso general

Notemos que $\hat{h}_{i,l} \in [0, 1]$ para cada $i \in V$ y $1 \leq l \leq k$ porque $\hat{h}_{i,l}$ es la esperanza de una función que vive entre 0 y 1.

Luego, mostremos la igualdad en la definición 5.4.2 viendo que $\hat{h}_{i,l} \leq \hat{h}_{i,l'}$ si y solo si $\|e_l - \hat{h}_{i,\bullet}\|_2 \geq \|e_{l'} - \hat{h}_{i,\bullet}\|_2$. En efecto, desarrollando las normas al cuadrado y después usando diferencia de cuadrados para simplificar la expresión, obtenemos que

$$\begin{aligned} & \|e_l - \hat{h}_{i,\bullet}\|_2^2 - \|e_{l'} - \hat{h}_{i,\bullet}\|_2^2 & (5.68) \\ &= \left(\sum_{t=1, t \neq l}^k \hat{h}_{i,t}^2 + (1 - \hat{h}_{i,l})^2 \right) - \left(\sum_{t=1, t \neq l'}^k \hat{h}_{i,t}^2 + (1 - \hat{h}_{i,l'})^2 \right) \\ &= [\hat{h}_{i,l'}^2 - \hat{h}_{i,l}^2] + [(1 - \hat{h}_{i,l})^2 - (1 - \hat{h}_{i,l'})^2] \\ &= (\hat{h}_{i,l'} - \hat{h}_{i,l})(\hat{h}_{i,l'} + \hat{h}_{i,l}) + [(1 - \hat{h}_{i,l}) - (1 - \hat{h}_{i,l'})][(1 - \hat{h}_{i,l}) + (1 - \hat{h}_{i,l'})] \\ &= (\hat{h}_{i,l'} - \hat{h}_{i,l})[(\hat{h}_{i,l'} + \hat{h}_{i,l}) + (2 - \hat{h}_{i,l} - \hat{h}_{i,l'})] \\ &= 2(\hat{h}_{i,l'} - \hat{h}_{i,l}) \end{aligned}$$

a partir de lo cual se deduce inmediatamente lo que queríamos ver.

Retomando la intuición que de esta manera conservamos pulseadas individuales entre comunidades, en la siguiente observación formalizaremos aquello y mostraremos la relación entre esta generalización y el estimador que estudiamos en las secciones anteriores:

[5.4.4] **Observación:** Relación entre el estimador por extensiones armónicas para dos comunidades balanceadas y el estimador en el caso general

Tomemos la transformación lineal $T : \mathbb{R}^2 \rightarrow \mathbb{R}$ dada por $T(x_1, x_2) = x_1 - x_2$.

Llamemos esta vez \hat{f} a la extensión armónica en 5.2.1 y \hat{h} a la extensión armónica en la definición general 5.4.2 y valdrá que $\hat{f}_i = T(\hat{h}_{i,\bullet})$ para cada $i \in V$.

En efecto, tenemos que basta mostrar que $i \mapsto T(\hat{h}_{i,\bullet})$ es una extensión armónica de $\mathbb{1}_{\mathcal{K}_1} - \mathbb{1}_{\mathcal{K}_2} : \mathcal{K} \rightarrow \mathbb{R}$ usando que mostramos su unicidad en la Proposición 5.1.13.

Por un lado, tenemos que si $i \in \mathcal{K}_l$ vale que $\hat{h}_{i,\bullet} = e_l$ y así $T(\hat{h}_{i,\bullet}) = (-1)^{l+1}$ para cada $1 \leq l \leq 2$, como queríamos ver.

Por otro lado, tenemos que si $i \notin \mathcal{K}$ vale que

$$\begin{aligned} T(\hat{h}_{i,\bullet}) &= \hat{h}_{i,1} - \hat{h}_{i,2} & (5.69) \\ &= \left(\frac{1}{\widehat{\deg}_n(i)} \sum_{j \sim i} \hat{h}_{j,\bullet} \right)_1 - \left(\frac{1}{\widehat{\deg}_n(i)} \sum_{j \sim i} \hat{h}_{j,\bullet} \right)_2 \\ &= \frac{1}{\widehat{\deg}_n(i)} \sum_{j \sim i} \hat{h}_{j,1} - \hat{h}_{j,2} \\ &= \frac{1}{\widehat{\deg}_n(i)} \sum_{j \sim i} T(\hat{h}_{j,\bullet}), \end{aligned}$$

lo cual muestra lo que queríamos ver.

En particular, como $\hat{f}_i = T(\hat{h}_{i,\bullet}) = \hat{h}_{i,1} - \hat{h}_{i,2}$ para cada $i \in V$, tendremos que $\phi(\text{signo}(\hat{f}_i)) = \text{argmáx}_{1 \leq l \leq 2} \{ \hat{h}_{i,l} \}$ para cada $i \in V$, lo cual muestra que los dos estimadores coinciden en el caso que hayan 2 comunidades.

Nuevamente, nos gustaría deducir que la generalización es débilmente consistente, donde esta vez además realizaremos la demostración para el modelo SBM general. Afortunadamente podremos hacerlo, y como los argumentos son similares y solamente se agregan dificultades técnicas referiremos para las demostraciones al anexo A.3.2. Sin embargo, sí será interesante considerar las generalizaciones, porque introducir los posibles desbalances traerá beneficios concretos y nos servirá más adelante en el capítulo, en particular garantizando la consistencia débil en el modelo SBM general pero sin condicionar al cardinal de cada comunidad.

Introduciremos luego las soluciones teóricas del método general, para lo cual en primer lugar mencionaremos como adaptamos la notación matricial para dos comunidades balanceadas que fijamos en la Observación 5.2.4 al caso general:

[5.4.5] **Observación:** Notación matricial en k comunidades

También ahora podremos entender a \hat{h} la extensión armónica empírica y a h la extensión armónica teórica como un sistema de ecuaciones lineales que notaremos también $\hat{h} = \Pi \hat{Q} \hat{h} + v$ en el caso empírico y $h = \Pi Q h + v$ en el caso teórico, donde ahora

vamos a tomar a Π como la proyección al subespacio S dado por

$$S = \left\{ x \in \mathbb{R}^{n \times k} \mid x_{i,\bullet} = \mathbf{0} \text{ para cada } i \in \mathcal{K} \right\}, \quad (5.70)$$

y v codifica la información que conocemos y está dado por

$$v_{i,\bullet} = \begin{cases} e_{\mathcal{C}(i)} & \text{si } i \in \mathcal{K}, \text{ y} \\ \mathbf{0} & \text{en caso contrario.} \end{cases} \quad (5.71)$$

Nuevamente, condicionando podemos suponer que los nodos están ordenados para obtener expresiones más agradables. En este caso, supondremos que los primeros n_1 nodos hacen a la comunidad 1, los siguientes n_2 nodos hacen a la comunidad 2, y así sucesivamente. Dentro de cada comunidad l , los primeros $g(n)$ nodos forman los nodos conocidos \mathcal{K}_l . Así, por ejemplo, podemos expresar a v en bloques como

$$v = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{1} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{pmatrix}. \quad (5.72)$$

Con la notación ya fijada, podemos presentar las soluciones en el caso general:

[5.4.6] Lema: Extensión armónica teórica en el caso general

En la notación de 5.4.5, suponiendo que $g(n) > 0$, la solución h para la ecuación $h = \Pi Q h + v$ está dada por

$$h = \begin{pmatrix} \mathbf{1} & \mathbf{0} & \dots & \mathbf{0} \\ H_{1,1}\mathbf{1} & H_{1,2}\mathbf{1} & \dots & H_{1,k}\mathbf{1} \\ \mathbf{0} & \mathbf{1} & \dots & \mathbf{0} \\ H_{2,1}\mathbf{1} & H_{2,2}\mathbf{1} & \dots & H_{2,k}\mathbf{1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{1} \\ H_{n,1}\mathbf{1} & H_{n,2}\mathbf{1} & \dots & H_{n,k}\mathbf{1} \end{pmatrix} \quad (5.73)$$

en bloques, o más detalladamente

$$h_{i,l} = \begin{cases} 1 & \text{si } i \in \mathcal{K}_l, \\ 0 & \text{si } i \in \mathcal{K} \setminus \mathcal{K}_l, \\ H_{l,l} & \text{si } i \in \mathcal{C}_l \setminus \mathcal{K}_l, \text{ y} \\ H_{s,l} & \text{si } i \in \mathcal{C}_s \setminus \mathcal{K}_s; \end{cases} \quad (5.74)$$

donde

$$H_{l,l} = \frac{(nq_n + n_l(p_n - q_n))q_n}{(p_n + (k-1)q_n)(nq_n + g(n)(p_n - q_n))} + \frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)} \quad (5.75)$$

y para cada $s \neq l$

$$H_{l,s} = \frac{(nq_n + n_s(p_n - q_n))q_n}{(p_n + (k-1)q_n)(nq_n + g(n)(p_n - q_n))}. \quad (5.76)$$

En particular, obtenemos que el estimador dado por

$$\hat{\mathcal{C}}(i) = \underset{1 \leq l \leq k}{\operatorname{argmín}} \left\{ \|e_l - h_{i,\bullet}\|_2 \right\} = \underset{1 \leq l \leq k}{\operatorname{argmáx}} \left\{ h_{i,l} \right\} \quad (5.77)$$

coincide con las comunidades \mathcal{C} si y solo si

$$\varepsilon_n > \frac{\max_{l \neq s} \{n_l - n_s\}}{n} \frac{q_n}{p_n + (k-1)q_n}, \quad (5.78)$$

y en ese caso si $\varepsilon_n = \frac{\max_{l \neq s} \{n_l - n_s\}}{n} \frac{q_n}{p_n + (k-1)q_n} + \delta_n$ para una separación $\delta_n > 0$ valdrá que $H_{l,l} - H_{l,s} \geq \Omega(\delta_n)$.

Notamos que cuando las comunidades no son balanceadas ahora se requiere que ε_n sea suficientemente grande para que el estimador teórico asociado coincida con las comunidades reales; esto se debe a que el desbalance introduce un término de orden constante y solo tenemos un término de orden ε_n para compensar, como podemos apreciar en (5.78).

Nuevamente, podremos mostrar que las soluciones empíricas están cerca de las soluciones teóricas, como consecuencia de la cercanía de las matrices de adyacencia que tendremos por el Teorema 3.3.4:

[5.4.7] Proposición: La solución empírica se parece a la solución teórica: versión para k comunidades.

Sea $G_{\mathcal{C}} \sim \text{SBM}(n, k, (n_l)_{1 \leq l \leq k}, p_n, q_n)$ donde $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ son tales que $a > b$ son reales no negativos y que $\liminf_{n \rightarrow \infty} \frac{\min_{i \in V} \{\deg_n(i)\}}{\log(n)} > 1$.

Supongamos además que $\varepsilon_n = \omega\left(\frac{1}{\sqrt{\log(n)}}\right)$.

Luego, vale que

$$\mathbb{P} \left(\left\{ \|\hat{h} - h\|_2 \leq \mathcal{O} \left(\frac{1}{\varepsilon_n \sqrt{\log(n)}} \right) \|h\|_2 \right\} \right) \geq 1 - o(1). \quad (5.79)$$

A partir de esto se deduce análogamente la consistencia débil del estimador:

[5.4.8] Teorema: Consistencia débil del estimador por extensiones armónicas $\hat{\mathcal{C}}_{\text{HE}}$.

Sea $G_{\mathcal{C}} \sim \text{SBM}(n, k, (n_l)_{1 \leq l \leq k}, p_n, q_n)$ una familia de modelos estocásticos de bloques

donde $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para números positivos $a > b > 0$ y donde $n_l = \Omega(n)$ para cada $1 \leq l \leq k$. Supongamos además que $\liminf_{n \rightarrow \infty} \frac{\min_{i \in V} \{\deg_n(i)\}}{\log(n)} > 1$. Por último, supongamos que para una separación $\delta_n > 0$ valen que

$$\varepsilon_n > \frac{\max_{t \neq l} \{n_t - n_l\}}{n} \frac{q_n}{p_n + (k-1)q_n} + \delta_n \quad (5.80)$$

y que $\delta_n^2 \varepsilon_n \log(n) \xrightarrow{n \rightarrow \infty} \infty$.

Luego, tendremos que el estimador $\hat{\mathcal{C}}_{\text{HE}}$ es débilmente consistente.

Más aún, el conjunto de nodos mal clasificados $\mathcal{B}(\hat{h})$ cumple que

$$\mathbb{P} \left(\left\{ \#(\mathcal{B}(\hat{h})) \leq \mathcal{O} \left(\frac{1}{\delta_n^2 \varepsilon_n \log(n)} \right) n \right\} \right) \geq 1 - o(1). \quad (5.81)$$

Por último, podemos generalizar aún más como habíamos adelantado al modelo SBM general:

[5.4.9] Teorema: Consistencia débil del estimador por extensiones armónicas para el modelo SBM general

Sea $G_{\mathcal{C}} \sim \text{SBM}(n, k, m_{\mathcal{C}}, p_n, q_n)$ una familia de modelos estocásticos de bloques donde $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para $a > b > 0$ números positivos con $\frac{\min_{i \in V} \{\deg_n(i)\}}{\log(n)} > 1$. Supongamos que para una separación $\delta_n > 0$ vale que

$$\varepsilon_n > \max_{t \neq l} \{p_{m_{\mathcal{C}}}(t) - p_{m_{\mathcal{C}}}(l)\} \frac{q_n}{p_n + (k-1)q_n} + \delta_n \quad (5.82)$$

donde además $\delta_n^2 \varepsilon_n \log(n) \xrightarrow{n \rightarrow \infty} \infty$.

Luego, tendremos que el estimador $\hat{\mathcal{C}}_{\text{HE}}$ es débilmente consistente.

Más aún, el conjunto de nodos mal clasificados $\mathcal{B}(\hat{h})$ cumple que

$$\mathbb{P} \left(\left\{ \#(\mathcal{B}(\hat{h})) \leq \mathcal{O} \left(\frac{1}{\delta_n^2 \varepsilon_n \log(n)} \right) n \right\} \right) \geq 1 - o(1).$$

Demostración. Procederemos aquí análogamente a la demostración del Teorema 4.3.2, aprovechando que conocemos las probabilidades de tener consistencia débil al condicionar para concluir mediante la ley de probabilidad total que también hay consistencia débil en el caso general.

Primero, notamos una vez más que por la Observación 2.4.4 sabemos que $n_l = np_{m_{\mathcal{C}}}(l) + \mathcal{O}(n^{\frac{1+\varepsilon}{2}})$ para cada comunidad l con probabilidad más alta que $1 - o(1)$, fijo $\varepsilon > 0$ suficientemente chico.

Otra vez, para cada vector de tamaños $(\tilde{n}_l)_{1 \leq l \leq k} \in \mathbb{N}_0^k$ que cumpla además que $\sum_{l=1}^k \tilde{n}_l = n$ llamemos $\mathcal{A}_{\tilde{n}}$ al evento dado por

$$\mathcal{A}_{\tilde{n}} = \{n_l = \tilde{n}_l \text{ para cada } 1 \leq l \leq k\}, \quad (5.83)$$

y digamos que $(\tilde{n}_l)_{1 \leq l \leq k}$ es admisible si vale que $|\tilde{n}_l - np_{m_C}(l)| \leq n^{\frac{3}{4}}$ para cada $n \in \mathbb{N}$ suficientemente grande y para cada $1 \leq l \leq k$. Una vez más, tendremos que si $(\tilde{n}_l)_{1 \leq l \leq k}$ no es admisible luego valdrá que en particular $\tilde{n}_l \neq np_{m_C}(l) + \mathcal{O}\left(n^{\frac{1+\varepsilon}{2}}\right)$.

Por otro lado, notemos que para configuraciones admisibles valdrá que

$$\frac{|\tilde{n}_t - \tilde{n}_l|}{n} \frac{q_n}{p_n + (k-1)q_n} \leq (|p_{m_C}(t) - p_{m_C}(l)| + o(1)) \frac{q_n}{p_n + (k-1)q_n} \quad (5.84)$$

según lo cual para n suficientemente grandes recuperamos que

$$\frac{\max_{t \neq l} \{ \tilde{n}_t - \tilde{n}_l \}}{n} \frac{q_n}{p_n + (k-1)q_n} + \delta_n < \varepsilon_n \quad (5.85)$$

Entonces procediendo como adelantamos, mediante la ley de probabilidad total y separando los casos poco probables, podemos verificar que el estimador por extensiones armónicas para dos comunidades balanceadas es débilmente consistente:

$$\begin{aligned} & \mathbb{P} \left(\left\{ \#(\mathcal{B}(\hat{h})) \not\leq \mathcal{O} \left(\frac{1}{\delta_n^2 \varepsilon_n \log(n)} \right) n \right\} \right) \quad (5.86) \\ & \leq \sum_{(\tilde{n}_l)_{1 \leq l \leq k} \text{ admisible}} \mathbb{P} \left(\left\{ \#(\mathcal{B}(\hat{h})) \not\leq \mathcal{O} \left(\frac{1}{\delta_n^2 \varepsilon_n \log(n)} \right) n \right\} \middle| \mathcal{A}_{\tilde{n}} \right) \cdot \mathbb{P}(\mathcal{A}_{\tilde{n}}) \\ & \quad + \mathbb{P} \left(\left\{ n_l \neq np_{m_C}(l) + \mathcal{O} \left(n^{\frac{1+\varepsilon}{2}} \right) \right\} \right) \\ & \leq \max_{(\tilde{n}_l)_{1 \leq l \leq k} \text{ admisible}} \left\{ \mathbb{P} \left(\left\{ \#(\mathcal{B}(\hat{h})) \not\leq \mathcal{O} \left(\frac{1}{\delta_n^2 \varepsilon_n \log(n)} \right) n \right\} \middle| \mathcal{A}_{\tilde{n}} \right) \right\} \\ & \quad + \mathbb{P} \left(\left\{ n_l \neq np_{m_C}(l) + \mathcal{O} \left(n^{\frac{1+\varepsilon}{2}} \right) \right\} \right) \\ & \lesssim o(1). \end{aligned}$$

Aclaremos que para tener que $\mathbb{P} \left(\left\{ \#(\mathcal{B}(\hat{h})) \not\leq \mathcal{O} \left(\frac{1}{\delta_n^2 \varepsilon_n \log(n)} \right) n \right\} \middle| \mathcal{A}_{\tilde{n}} \right) = o(1)$ para cada \tilde{n} admisible usamos que vimos que estamos en las hipótesis del Teorema 5.4.8, y que como tendremos que podremos acotar \tilde{n}_l uniformemente en \tilde{n} para cada $1 \leq l \leq k$ como en (4.76) tendremos que las constantes en los teoremas que dependían de los tamaños de las comunidades estarán controladas por estas cotas uniformes y podremos usar los resultados sin más problemas. ■

Demostrada la consistencia débil de nuestros estimadores por extensiones armónicas, continuaremos entonces en la siguiente sección buscando mejorar nuestra garantía de consistencia a consistencia fuerte.

5.5 Consistencia fuerte del estimador por extensiones armónicas

Antes de empezar el estudio de la consistencia fuerte de nuestros operadores, mencionamos que volveremos en esta sección a tratar el caso de dos comunidades balanceadas, y discutiremos posibles generalizaciones al final del capítulo.

Notaremos también que, nuevamente, para demostrar la consistencia fuerte del estimador una primera idea puede ser ver que coordenada a coordenada, más que en norma euclídea, vale que el estimador empírico está cerca del estimador teórico. Sin embargo, análogamente a lo que ocurría con el estimador espectral, nos va a resultar conveniente pasar por otro estimador intermedio que también será fuertemente consistente y además será suficientemente cercano a nuestro estimador empírico: una vez más recurriremos al estimador ayudado por un oráculo, pero ahora lo construiremos partiendo de nuestras extensiones armónicas. En este contexto, podremos definirlo de la siguiente manera:

[5.5.1] Definición: Estimador por extensiones armónicas ayudado por un oráculo para dos comunidades balanceadas

Sea $G_{\mathcal{C}} \sim \text{SSBM}(n, 2, p_n, q_n)$ una familia de modelos estocásticos de bloques. Luego, definimos el **estimador por extensiones armónicas ayudado por un oráculo para dos comunidades balanceadas** para \mathcal{C} que notaremos como $\hat{\mathcal{C}}_{\text{HEA2}}$ como

$$\hat{\mathcal{C}}_{\text{HEA2}}(i) = \phi(\text{signo}((\Pi\hat{Q}\Pi h + v)_i))$$

para cada $i \in V$, donde h es la extensión armónica de $\mathbb{1}_{\mathcal{K}_1} - \mathbb{1}_{\mathcal{K}_2} : \mathcal{K} \rightarrow \mathbb{R}$ para Q y donde ϕ es la función que manda $\phi(x) = 1$ si $x \geq 0$ y $\phi(x) = 2$ si no.

Recordemos que el estimador ayudado por un oráculo predecía la comunidad de cada nodo a partir de las comunidades de sus vecinos. Mencionamos que en este caso añadimos la proyección Π por una cuestión meramente técnica, para evitar que los nodos en \mathcal{K} tengan demasiada influencia al predecir las comunidades, porque como serán pocos podría haber demasiada varianza y que alguno de los nodos se conecte a demasiados referentes de la otra comunidad y, aunque es posible estimar las comunidades sin la segunda proyección, requeriríamos hipótesis extra. Por otro lado, al necesitar la extensión armónica teórica no va a ser un estimador que vamos a poder usar en la práctica. Análogamente a como procedimos con los estimadores espectrales, tendremos una variante que sí podremos calcular y que volverá a parecerse al estimador ayudado por un oráculo, que es la de los estimadores por vecindades. Antes de introducirlos, un poco de notación que nos será de ayuda:

[5.5.2] Notación: Notación para los estimadores por vecindades

Dado $U \subseteq V$, notamos $n^{(U)} = n - \#U$, y $n_i^{(U)} = \#(\mathcal{C}_i \cap U)$. Llamaremos además $\nu_n = \frac{\#U}{n}$ a la proporción de nodos en U .

Además, notamos $\hat{A}^{(U)}$, $A^{(U)}$, $\hat{Q}^{(U)}$, $Q^{(U)}$ a las matrices correspondientes pero donde la i -ésima fila y columna las fijamos en $\mathbf{0}$ para cada $i \in U$. Notaremos también $\widehat{\text{deg}}_n^{(U)}(i) = \sum_{j \in V} \hat{A}_{i,j}^{(U)}$ y $\text{deg}_n^{(U)}(i) = \sum_{j \in V} A_{i,j}^{(U)}$ a los grados de cada nodo en este contexto.

Por último, notaremos $\hat{h}^{(U)}$, $h^{(U)}$ a las extensiones armónicas empírica y teórica respectivamente.

Ya sin más preámbulos, en este contexto podremos definir los estimadores por vecindades

de la siguiente manera:

[5.5.3] Definición: Estimador por extensiones armónicas por vecindades para dos comunidades balanceadas

Sea $G_{\mathcal{C}} \sim \text{SSBM}(n, 2, p_n, q_n)$ una familia de modelos balanceados conexos con alta probabilidad. Luego, definimos el **estimador por extensiones armónicas por vecindades para dos comunidades balanceadas** para \mathcal{C} que notaremos como $\hat{\mathcal{C}}_{\text{HEV2}}$ como

$$\hat{\mathcal{C}}_{\text{HEV2}}(i) = \begin{cases} \phi \left(\text{signo} \left(\left(\Pi \hat{Q} \Pi \hat{h}^{(i)} \right)_i \right) \right) & \text{si } i \notin \mathcal{K} \\ \phi(\text{signo}(v_i)) & \text{si } i \in \mathcal{K} \end{cases} \quad (5.87)$$

para cada $i \in V$, donde $\hat{h}^{(i)}$ es una extensión $\hat{Q}^{(i)}$ -armónica de $\mathbb{1}_{\mathcal{K}_1} - \mathbb{1}_{\mathcal{K}_2} : \mathcal{K} \rightarrow \mathbb{R}$ para $\hat{Q}^{(i)}$ la matriz que resulta de fijar la i -ésima fila y columna de \hat{Q} como $\mathbf{0}$ y donde ϕ es la función que manda $\phi(x) = 1$ si $x \geq 0$ y $\phi(x) = 2$ si no.

Más aún, dada \mathcal{U} una partición de los nodos en $V \setminus \mathcal{K}$ definimos más generalmente el estimador por vecindades como

$$\hat{\mathcal{C}}_{\text{HEV2}}(i) = \begin{cases} \phi \left(\text{signo} \left(\left(\Pi \hat{Q} \Pi \hat{h}^{(U)} \right)_i \right) \right) & \text{si } i \in U \text{ para un cierto } U \in \mathcal{U}, \text{ y} \\ \phi(\text{signo}(v_i)) & \text{si } i \in \mathcal{K}, \end{cases} \quad (5.88)$$

donde $\hat{h}^{(U)}$ es una extensión $\hat{Q}^{(U)}$ -armónica de $\mathbb{1}_{\mathcal{K}_1} - \mathbb{1}_{\mathcal{K}_2} : \mathcal{K} \rightarrow \mathbb{R}$ y donde ϕ es la función que manda $\phi(x) = 1$ si $x \geq 0$ y $\phi(x) = 2$ si no.

Desde ya, tenemos nuevamente la cuestión técnica de la conexión, porque los estimadores por vecindades están bien definidos cuando el grafo es conexo al quitar cada $U \in \mathcal{U}$ y tenemos una única extensión armónica. Sin embargo, esto ocurrirá para U suficientemente chicos, y al mismo tiempo \mathcal{U} suficientemente chico, al elegir una partición aleatoriamente, que será nuestro caso, y podremos nuevamente proceder sin inconvenientes.

Volviendo a los estimadores por vecindades, recordemos que la intuición al proceder de este modo es aprovechar que las estimaciones de los estimadores por extensiones armónicas ya son considerablemente buenas, entonces podemos intentar de predecir la comunidad de cada nodo a partir de las estimaciones de las comunidades de los vecinos.

Sin embargo, actualmente todavía no sabemos si el estimador por vecindades coincide con el estimador por extensiones armónicas usual. Una pieza clave en la demostración de que el estimador por vecindades y el estimador por extensiones armónicas usual coincidían con alta probabilidad era la desigualdad de Davis Kahan, que como solo es válida para autovectores no tenemos a disposición y todavía no logramos suplir con ese grado de precisión. Sin embargo, el estimador por vecindades ya nos otorga una manera de predecir las comunidades que no precisa acceso a elementos teóricos, y puede ser usado en la práctica, así que es digno de estudiarse.

Añadimos que la decisión de considerar subconjuntos $U \subseteq V \setminus \mathcal{K}$ más generales que los

que tienen un solo nodo $\{i\}$ con $i \in V \setminus \mathcal{K}$ nos dejará obtener un método que es computacionalmente más rápido, al hacer menos estimaciones previas para predecir las comunidades de todos los nodos, como analizaremos más adelante con las simulaciones en el Capítulo 6.

Antes de demostrar la consistencia fuerte de estos nuevos estimadores, comenzaremos observando que ya conocemos las extensiones armónicas para el estimador por vecindades, porque estudiamos las extensiones armónicas también en el caso general:

[5.5.4] Observación: Solución teórica del estimador por vecindades en el modelo BSBM

Como en la definición 5.5.3, supongamos que $U \subseteq V \setminus \mathcal{K}$, y hallemos fórmulas para las extensiones $Q^{(U)}$ -armónicas teóricas asociadas $h^{(U)}$.

En primer lugar, notemos que $h_i^{(U)} = 0$ para cada $i \in U$. Podemos deducir esto de que en $h^{(U)} = \Pi Q^{(U)} h^{(U)} + v$ la i -ésima fila nos da la ecuación $h_i^{(U)} = 0$ para cada $i \in U$.

Luego, podemos pensar equivalentemente que para el estimador por vecindades primero hacemos la extensión armónica en un grafo en el cual eliminamos los nodos en U , y después usamos esa extensión para predecir. Usando entonces los resultados para el caso general, porque no podemos garantizar que U tenga la misma cantidad de nodos de cada comunidad, podemos primero hallar la extensión $Q^{(U)}$ -armónica teórica de $\mathbb{1}_{\mathcal{K}_1} e_1 + \mathbb{1}_{\mathcal{K}_2} e_2 : \mathcal{K} \rightarrow \mathbb{R}^2$ mediante el Lema 5.4.6 y luego procediendo análogamente a la Observación 5.4.4 podemos transformarla a la extensión armónica asociada a $\mathbb{1}_{\mathcal{K}_1} - \mathbb{1}_{\mathcal{K}_2} : \mathcal{K} \rightarrow \mathbb{R}$. De esta manera, obtenemos que $h^{(U)}$ está dada por

$$\frac{(n_1^{(U)} - n_2^{(U)})(p_n - q_n)q_n}{(p_n + (k-1)q_n)(n^{(U)}q_n + g(n)(p_n - q_n))} + \frac{g(n)(p_n - q_n)}{n^{(U)}q_n + g(n)(p_n - q_n)} \quad (5.89)$$

para nodos en $\mathcal{C}_1 \setminus \mathcal{K}$, y el opuesto cuando el nodo está en $\mathcal{C}_2 \setminus \mathcal{K}$.

Además, pondremos para rápido acceso algunas estimaciones que aparecerán recurrentemente en las demostraciones.

[5.5.5] Lema: Cotas útiles para los estimadores por extensiones armónicas.

Sea $G_{\mathcal{C}} \sim \text{SBM}(n, 2, p_n, q_n,)$ una familia de modelos con $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para $a > b > 0$, donde además $\frac{a+b}{2} > 1$.

Supongamos además que $\varepsilon_n = \omega\left(\frac{1}{\sqrt{\log(n)}}\right)$.

Si $U \subseteq V \setminus \mathcal{K}$ es tal que $\nu_n = o(1)$, vale que:

1. $\|A_{i,\bullet}\|_2 \leq \mathcal{O}\left(\frac{\log(n)}{\sqrt{n}}\right)$ para cada $i \in V$,
2. $\|h^{(U)}\|_2 \leq \mathcal{O}\left(\max\{\nu_n, \sqrt{\varepsilon_n}\}\right) \sqrt{n}$,
3. $\|h^{(U)} - h\|_2 \leq \mathcal{O}\left(\max\{\nu_n, \varepsilon_n \sqrt{\nu_n}\}\right) \sqrt{n}$,
4. $\|\hat{h}^{(U)} - h^{(U)}\|_2 \leq \mathcal{O}\left(\max\left\{\nu_n \frac{1}{\varepsilon_n \sqrt{\log(n)}}, \frac{1}{\sqrt{\varepsilon_n} \sqrt{\log(n)}}\right\}\right) \sqrt{n}$ con probabilidad mayor que $1 - o(1)$, y

$$5. \left\| \hat{h}^{(U)} - h \right\|_2 \leq \mathcal{O} \left(\max \left\{ \frac{1}{\sqrt{\varepsilon_n} \sqrt{\log(n)}}, \nu_n, \varepsilon_n \sqrt{\nu_n} \right\} \right) \sqrt{n} \text{ con probabilidad mayor que } 1 - o(1).$$

Demostración. Para mostrar estas cotas resta solamente acotar cuidadosamente.

En primer lugar, vemos que un cálculo directo nos muestra que vale 1.:

$$\|A_{i,\bullet}\|_2^2 = \left(\frac{n}{2} - 1\right) p_n^2 + \frac{n}{2} q_n^2 \lesssim \frac{\log(n)^2}{n}. \quad (5.90)$$

A continuación, para ver que vale 2., primero podemos encontrar el orden de cada coordenada para después acotar la norma. Para comenzar, si $i \in \mathcal{K}$ luego $|h_i^{(U)}| = 1$, y además si $i \in U$ luego $h_i^{(U)} = 0$. Por otro lado, si $i \notin U$, $i \notin \mathcal{K}$, supondremos sin pérdida de generalidad que $\mathcal{C}(i) = 1$. Luego, como mencionamos en el Lema 5.5.4,

$$h_i^{(U)} = \frac{(n_1^{(U)} - n_2^{(U)})(p_n - q_n)q_n}{(p_n + (k-1)q_n)(n^{(U)}q_n + g(n)(p_n - q_n))} + \frac{g(n)(p_n - q_n)}{n^{(U)}q_n + g(n)(p_n - q_n)}, \quad (5.91)$$

y acotaremos cada término. Para el primer término, usamos que $n_1^{(U)}$ y $n_2^{(U)}$ son aproximadamente la mitad de los nodos, salvo por eventualmente haber quitado más nodos en alguna comunidad que la otra, lo cual puede resultar en un desbalance de a lo sumo $\#U$, y así podemos acotar de la siguiente manera:

$$\begin{aligned} & \left| \frac{(n_1^{(U)} - n_2^{(U)})(p_n - q_n)q_n}{(p_n + (k-1)q_n)(n^{(U)}q_n + g(n)(p_n - q_n))} \right| \\ &= \frac{|n_1^{(U)} - n_2^{(U)}|}{n} \frac{(p_n - q_n)q_n}{(p_n + (k-1)q_n)\left(\frac{n^{(U)}}{n}q_n + \frac{g(n)}{n}(p_n - q_n)\right)} \\ &\lesssim \frac{\#U}{n} \cdot 1 \\ &= \nu_n. \end{aligned} \quad (5.92)$$

Ahora bien, como además $\frac{g(n)(p_n - q_n)}{n^{(U)}q_n + g(n)(p_n - q_n)} \asymp \varepsilon_n$, tendremos que

$$\left| h_i^{(U)} \right| \lesssim \max \{ \nu_n, \varepsilon_n \}, \quad (5.93)$$

y así combinando todo habremos hallado que, usando que $\varepsilon_n n \geq \varepsilon_n^2 n$ porque $\varepsilon_n \leq 1$,

$$\begin{aligned} \left\| h^{(U)} \right\|_2^2 &\lesssim 2g(n) \cdot 1 + (n - \#U - 2g(n)) \max \{ \nu_n, \varepsilon_n \}^2 \\ &\lesssim \max \{ \varepsilon_n n, \nu_n^2 n, \varepsilon_n^2 n \} \\ &= \max \{ \sqrt{\varepsilon_n}, \nu_n \}^2 n \end{aligned} \quad (5.94)$$

a partir de lo cual tomando raíz cuadrada se deduce lo afirmado.

Para mostrar 3. primero acotaremos el orden de $\left| h_i - h_i^{(U)} \right|$ coordenada a coordenada. Si $i \in \mathcal{K}$, ambas coinciden y entonces el valor absoluto de la diferencia es 0. Si en cambio $i \in U$ luego tenemos que $h_i^{(U)} = 0$ y entonces valdrá que

$$h_i - h_i^{(U)} = \frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)} \asymp \varepsilon_n. \quad (5.95)$$

Por último, si $i \notin U$, $i \notin \mathcal{K}$, supongamos primero sin pérdida de generalidad que $\mathcal{C}(i) = 1$, y en este caso tenemos que

$$\begin{aligned} h_i^{(U)} - h_i &= \left[\frac{(n_1^{(U)} - n_2^{(U)})(p_n - q_n)q_n}{(p_n + (k-1)q_n)(n^{(U)}q_n + g(n)(p_n - q_n))} + \frac{g(n)(p_n - q_n)}{n^{(U)}q_n + g(n)(p_n - q_n)} \right] \\ &\quad - \left[\frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)} \right] \\ &= \left[\frac{(n_1^{(U)} - n_2^{(U)})(p_n - q_n)q_n}{(p_n + (k-1)q_n)(n^{(U)}q_n + g(n)(p_n - q_n))} \right] \\ &\quad + \left[\frac{g(n)(p_n - q_n)}{n^{(U)}q_n + g(n)(p_n - q_n)} - \frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)} \right], \end{aligned} \quad (5.96)$$

donde podremos acotar cada término por separado. En primer lugar, análogamente a (5.92), podemos acotar

$$\left| \frac{(n_1^{(U)} - n_2^{(U)})(p_n - q_n)q_n}{(p_n + (k-1)q_n)(n^{(U)}q_n + g(n)(p_n - q_n))} \right| \lesssim \nu_n, \quad (5.97)$$

y además el segundo término podemos controlarlo de la siguiente manera:

$$\begin{aligned} &\frac{g(n)(p_n - q_n)}{n^{(U)}q_n + g(n)(p_n - q_n)} - \frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)} \\ &= \frac{g(n)(p_n - q_n) \left[(nq_n + g(n)(p_n - q_n)) - (n^{(U)}q_n + g(n)(p_n - q_n)) \right]}{\left(n^{(U)}q_n + g(n)(p_n - q_n) \right) (nq_n + g(n)(p_n - q_n))} \\ &= \frac{g(n)(p_n - q_n) \#U q_n}{\left(n^{(U)}q_n + g(n)(p_n - q_n) \right) (nq_n + g(n)(p_n - q_n))} \\ &\asymp \varepsilon_n \nu_n. \end{aligned} \quad (5.98)$$

Combinando ambas cotas, podemos ver que, como $\varepsilon_n \leq 1$,

$$\begin{aligned} \left| h_i^{(U)} - h_i \right| &\lesssim \nu_n + \varepsilon_n \nu_n \\ &\lesssim \nu_n, \end{aligned} \quad (5.99)$$

y así controlar la norma al cuadrado

$$\left\| h^{(U)} - h \right\|_2^2 \lesssim \#U \cdot \varepsilon_n^2 + (n - \#U - 2g(n)) (\nu_n)^2 \lesssim \max \{ \nu_n, \sqrt{\nu_n} \varepsilon_n \}^2 n; \quad (5.100)$$

resta solamente que tomemos raíz cuadrada para que obtengamos la cota que queríamos.

Continuaremos mostrando 4., que no es más que una consecuencia directa de la Proposición 5.4.7, pues al sacar los nodos de U estaremos posiblemente en el caso general antes que en el caso balanceado:

$$\|\hat{h}^{(U)} - h^{(U)}\|_2 \lesssim \frac{\|h^{(U)}\|_2}{\frac{g(n)}{n^{(U)}} \sqrt{\log(n^{(U)})}} \quad (5.101)$$

$$\begin{aligned} &\lesssim \frac{\max\{\nu_n, \sqrt{\varepsilon_n}\} \sqrt{n}}{\varepsilon_n \sqrt{\log(n)}} \\ &\lesssim \max\left\{ \nu_n \frac{1}{\varepsilon_n \sqrt{\log(n)}}, \frac{1}{\sqrt{\varepsilon_n} \sqrt{\log(n)}} \right\} \sqrt{n}. \end{aligned} \quad (5.102)$$

Notamos que aún al sacar los nodos de U se verifica que $\frac{\min_{i \in V} \{\deg_n^{(U)}(i)\}}{\log(n)} > 1$ para n suficientemente grandes, pues vale que para cada i fijo tendremos que

$$|\deg_n(i) - \deg_n^{(U)}(i)| \leq \#U p_n \asymp \nu_n \log(n) = o(\log(n)) \quad (5.103)$$

y entonces nuestra hipótesis $\frac{a+b}{2} > 1$ nos da lo que afirmamos, porque para n suficientemente grandes valdrá que

$$\frac{\min_{i \in V} \{\deg_n^{(U)}(i)\}}{\log(n)} (1 + o(1)) = \frac{\min_{i \in V} \{\deg_n(i)\}}{\log(n)} = \frac{a+b}{2} (1 + o(1)). \quad (5.104)$$

Finalmente, para ver 5. podemos combinar todas las cotas obtenidas sucesivamente, y tendremos que

$$\begin{aligned} \|\hat{h}^{(U)} - h\|_2 &\leq \|\hat{h}^{(U)} - h^{(U)}\|_2 + \|h^{(U)} - h\|_2 \quad (5.105) \\ &\lesssim \max\left\{ \nu_n \frac{1}{\varepsilon_n \sqrt{\log(n)}}, \frac{1}{\sqrt{\varepsilon_n} \sqrt{\log(n)}} \right\} \sqrt{n} + \max\{\nu_n, \varepsilon_n \sqrt{\nu_n}\} \sqrt{n} \\ &\lesssim \max\left\{ \frac{1}{\sqrt{\varepsilon_n} \sqrt{\log(n)}}, \nu_n \frac{1}{\varepsilon_n \sqrt{\log(n)}}, \nu_n, \varepsilon_n \sqrt{\nu_n} \right\} \sqrt{n} \\ &= \max\left\{ \frac{1}{\sqrt{\varepsilon_n} \sqrt{\log(n)}}, \nu_n, \varepsilon_n \sqrt{\nu_n} \right\} \sqrt{n}. \end{aligned}$$

En la última igualdad, usamos que como $\varepsilon_n = \omega\left(\frac{1}{\sqrt{\log(n)}}\right)$ vale que $\frac{1}{\varepsilon_n \sqrt{\log(n)}} \xrightarrow{n \rightarrow \infty} 0$ así que podemos eliminar $\nu_n \frac{1}{\varepsilon_n \sqrt{\log(n)}}$ en favor de ν_n como candidato a máximo, y completamos la demostración. ■

Ahora sí, a continuación demostraremos la consistencia fuerte de estos nuevos estimadores por extensiones armónicas. Nuevamente usaremos el Lema 3.5.4, porque como adelantamos previamente las estimaciones de lo que ocurre con la vecindad de cada nodo juegan un papel principal también en estas demostraciones. Precisaremos sin embargo generalizar el Lema 3.5.3 para obtener una descripción más precisa que la que hizo falta en el Capítulo 3. A continuación enunciamos el resultado que usaremos, y lo demostraremos en el anexo A.3.1:

[5.5.6] Lema: Las mayorías no son chicas con probabilidad alta

Supongamos que $a > b$ son números reales, y que $(W_i)_{1 \leq i \leq m}$ y $(Z_i)_{1 \leq i \leq m}$ son variables independientes con $W_i \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p_n)$ y $Z_i \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(q_n)$ para $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$, donde $a > b$ son números positivos. Luego, para cada $\eta \in \mathbb{R}$ vale que

$$\mathbb{P} \left(\left\{ \sum_{i=1}^m W_i - Z_i \leq \eta \log(n) \right\} \right) \leq n^{-\frac{m}{n}(\sqrt{a}-\sqrt{b})^2 + \frac{\eta}{2} \log(\frac{a}{b})} \quad (5.106)$$

Comenzaremos luego por el estimador ayudado por un oráculo:

[5.5.7] Proposición: El estimador por extensiones armónicas con ayuda de un oráculo $\widehat{\mathcal{C}}_{\text{HEA2}}$ es fuertemente consistente.

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos balanceados con $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para $a > b > 0$ números positivos que cumplen además que $\frac{a+b}{2} > 1 + \sqrt{ab}$.

Supongamos además que $0 < \varepsilon_n = o(1)$.

Luego, $\widehat{\mathcal{C}}_{\text{HEA2}}$ es fuertemente consistente.

Además, vale que

$$\mathbb{P} \left(\left\{ \min_{i \notin \mathcal{K}} \left\{ |(\Pi \widehat{Q} \Pi h + v)_i| \right\} \geq \Omega(\varepsilon_n) \right\} \right) = 1 - o(1). \quad (5.107)$$

Demostración. Para ver que el estimador $\widehat{\mathcal{C}}_{\text{HEA2}}$ es fuertemente consistente, veremos que las coordenadas asociadas a nodos en la comunidad \mathcal{C}_1 son positivas y mayores que $C\varepsilon_n$ y las coordenadas asociadas a la comunidad \mathcal{C}_2 son negativas y menores que $-C\varepsilon_n$ simultáneamente para alguna constante $C > 0$ con probabilidad alta. Excluimos a los nodos en \mathcal{K} , pues para aquellos $\widehat{\mathcal{C}}_{\text{HEA2}}(i) = 1$ si $i \in \mathcal{K}_1$ y $\widehat{\mathcal{C}}_{\text{HEA2}}(i) = -1$ si $i \in \mathcal{K}_2$, con lo que tenemos la estimación correcta asegurada. Como resulta que demostrar esto es la esencia de la demostración de la desigualdad (5.107) que queremos ver que vale con alta probabilidad, nos concentraremos en ella y obtendremos la consistencia fuerte en el camino.

Veamos entonces que vale (5.107). Más precisamente, veremos que si $i \notin \mathcal{K}$ valdrá que para una constante $C > 0$ que no depende de i ni de n :

$$\mathbb{P} \left(\left\{ |(\Pi \widehat{Q} \Pi h + v)_i| \geq C\varepsilon_n \right\} \right) = 1 - o(n^{-1}). \quad (5.108)$$

Supondremos sin pérdida de generalidad que $i \in \mathcal{C}_1 \setminus \mathcal{K}_1$ por comodidad, porque notando que $(\Pi \widehat{Q} \Pi h + v)_i$ se distribuye como $-(\Pi \widehat{Q} \Pi h + v)_j$ para cada $i \in \mathcal{C}_1 \setminus \mathcal{K}_1$ y para cada $j \in \mathcal{C}_2 \setminus \mathcal{K}_2$ podremos realizar una demostración análoga para cada nodo. Más aún, al ser todas idénticamente distribuidas la constante C podremos tomarla uniformemente igual a la que consigamos para el nodo i fijo. Por último, usando una cota de unión de eventos para los complementos como en la Observación 2.2.5 obtendremos que valen todos simultáneamente con probabilidad mayor que $1 - o(1)$, como queríamos.

Mostraremos entonces la desigualdad (5.108). Primero notamos que como $\widehat{Q} = \widehat{D}^{-1}\widehat{A}$ e $i \notin \mathcal{K}$ vale que

$$|(\Pi\widehat{Q}\Pi h + v)_i| = \left| \frac{(\widehat{A}\Pi h)_i}{\widehat{\deg}_n(i)} \right| \gtrsim \frac{|(\widehat{A}\Pi h)_i|}{\log(n)}, \quad (5.109)$$

donde por el Lema 2.4.11 obtenemos que con probabilidad mayor que $1 - o(1)$ vale que $\widehat{\deg}_n(i) \asymp \log(n)$ para este i fijo, más precisamente en (2.22). Luego, un cálculo directo de $(\widehat{A}\Pi h)_i$ nos indica que

$$\begin{aligned} (\widehat{A}\Pi h)_i &= \sum_{j=g(n)+1}^{n-g(n)} \widehat{A}_{i,j} h_j \\ &= \frac{g(n)(a-b)}{nb + g(n)(a-b)} \left(\sum_{j=g(n)+1, j \neq i}^{\frac{n}{2}} \widehat{A}_{i,j} - \sum_{j=\frac{n}{2}+1}^{n-g(n)} \widehat{A}_{i,j} \right) \end{aligned} \quad (5.110)$$

donde usamos para h la expresión obtenida en el Lema 5.2.5. Notemos que el primer factor es de orden ε_n , y concentrémosnos entonces en acotar el segundo. Para hacer esto, usaremos el Lema 5.5.6. Más precisamente, obtendremos que en términos de alguna constante $\eta > 0$ suficientemente chica que podremos acotar la expresión en (5.110) con probabilidad mayor que $1 - o(n^{-1})$ de la siguiente manera

$$\begin{aligned} (\widehat{A}\Pi h)_i &\geq \frac{g(n)(a-b)}{nb + g(n)(a-b)} \cdot \left(\frac{\eta}{2} \log(n) \right) \\ &= \left(\frac{\eta}{2} \frac{g(n)(a-b)}{nb + g(n)(a-b)} \right) \log(n) \\ &\asymp \varepsilon_n \log(n). \end{aligned} \quad (5.111)$$

Así, en particular $(\widehat{A}\Pi h)_i$ será no negativo y podremos junto con (5.109) concluir la demostración, pues

$$|(\Pi\widehat{Q}\Pi h + v)_i| \gtrsim \frac{\varepsilon_n \log(n)}{\log(n)} = \varepsilon_n \quad (5.112)$$

Veamos luego que el segundo factor en (5.110) es no negativo y mayor que $\frac{\eta}{2} \log(n)$ para alguna constante η suficientemente chica. Queremos aplicar el Lema 5.5.6, pero tenemos el detalle técnico no tenemos la misma cantidad de variables aleatorias sumando y restando. Para remediar lo esto, notamos que si tomamos $X \sim \text{Bernoulli}(p_n)$ una variable aleatoria independiente de $\widehat{A}_{i,j}$ para cada j tendremos que si n es suficientemente grande luego

$$\begin{aligned} &\left\{ \sum_{j=g(n)+1, j \neq i}^{\frac{n}{2}} \widehat{A}_{i,j} - \sum_{j=\frac{n}{2}+1}^{n-g(n)} \widehat{A}_{i,j} \leq \frac{\eta}{2} \cdot \log(n) \right\} \\ &\subseteq \left\{ X + \sum_{j=g(n)+1, j \neq i}^{\frac{n}{2}} \widehat{A}_{i,j} - \sum_{j=\frac{n}{2}+1}^{n-g(n)} \widehat{A}_{i,j} \leq \eta \cdot \log(n) \right\}, \end{aligned} \quad (5.113)$$

usando que $X \leq 1$ y $\frac{\eta}{2} \log(n) + 1 \leq \eta \log(n)$ para n suficientemente grandes. Ahora sí, aplicando el Lema 5.5.6 tendremos que

$$\begin{aligned}
& \mathbb{P} \left(\left\{ \sum_{j=g(n)+1, j \neq i}^{\frac{n}{2}} \hat{A}_{i,j} - \sum_{j=\frac{n}{2}+1}^{n-g(n)} \hat{A}_{i,j} \leq \frac{\eta}{2} \cdot \log(n) \right\} \right) \\
& \leq \mathbb{P} \left(\left\{ X + \sum_{j=g(n)+1, j \neq i}^{\frac{n}{2}} \hat{A}_{i,j} - \sum_{j=\frac{n}{2}+1}^{n-g(n)} \hat{A}_{i,j} \leq \eta \log(n) \right\} \right) \\
& \leq n^{-\frac{\frac{\eta}{2}-g(n)}{n} (\sqrt{a}-\sqrt{b})^2 + \frac{\eta}{2} \log\left(\frac{a}{b}\right)} \\
& = n^{-(1-2\varepsilon_n) \frac{(\sqrt{a}-\sqrt{b})^2}{2} + \frac{\eta}{2} \log\left(\frac{a}{b}\right)} \\
& = o\left(n^{-1}\right),
\end{aligned} \tag{5.114}$$

donde en la última igualdad usamos que como $\frac{a+b}{2} > 1 + \sqrt{ab}$ y también vale que $1 - 2\varepsilon_n = 1 - o(1)$ tendremos que para $\eta > 0$ suficientemente chicos y n suficientemente grandes luego $-(1 - 2\varepsilon_n) \frac{(\sqrt{a}-\sqrt{b})^2}{2} + \frac{\eta}{2} \log\left(\frac{a}{b}\right) < -1$ y entonces la probabilidad en (5.114) es $o(n^{-1})$, lo cual muestra lo que queríamos ver. ■

Continuaremos entonces mostrando que, como adelantamos, los estimadores ayudado por un oráculo y por vecindades utilizan vectores parecidos para realizar sus estimaciones:

[5.5.8] Proposición: El estimador ayudado por un oráculo y el estimador por vecindades están cerca coeficiente a coeficiente.

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos con $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para $a > b > 0$, donde además $\frac{a+b}{2} > 1$, y supongamos que $\varepsilon_n = \omega\left(\frac{1}{(\log(\log(n)))^{\frac{2}{3}}}\right)$.

Sea $U \subseteq V \setminus \mathcal{K}$ tal que $\nu_n \leq \mathcal{O}\left(\frac{1}{\log(n)}\right)$.

Luego, tendremos que vale que

$$\mathbb{P} \left(\left\{ |(\Pi \hat{Q} \Pi \hat{h}^{(U)} + v)_i - (\Pi \hat{Q} h \Pi + v)_i| \leq o(1) \varepsilon_n \text{ para cada } i \in U \right\} \right) \geq 1 - o(1). \tag{5.115}$$

En particular, para cada $i \in V$ fijo, vale que

$$\mathbb{P} \left(\left\{ |(\Pi \hat{Q} \Pi \hat{h}^{(i)} + v)_i - (\Pi \hat{Q} \Pi h + v)_i| \leq o(1) \varepsilon_n \right\} \right) \geq 1 - o(1). \tag{5.116}$$

Demostración. En primer lugar, fijo $i \in U$ podemos acotar de la siguiente manera

$$\begin{aligned}
|(\Pi \hat{Q} \Pi \hat{h}^{(U)} + v)_i - (\Pi \hat{Q} \Pi h + v)_i| &= |\hat{Q}_{i, \bullet} \cdot (\hat{h}^{(U)} - h)| \\
&\lesssim \frac{1}{\log(n)} |\hat{A}_{i, \bullet} \cdot (\hat{h}^{(U)} - h)| \\
&\leq \frac{1}{\log(n)} |A_{i, \bullet} \cdot (\hat{h}^{(U)} - h)| + \frac{1}{\log(n)} |(\hat{A} - A)_{i, \bullet} \cdot (\hat{h}^{(U)} - h)|,
\end{aligned} \tag{5.117}$$

donde usamos que $\Pi(\hat{h}^{(U)} - h) = \hat{h}^{(U)} - h$ pues en los nodos en \mathcal{K} vale que $\hat{h}^{(U)}$ y h coinciden. Luego, basta ver que cada término es de orden $o(1) \varepsilon_n$.

Para comenzar, trabajaremos con el primer término, que podremos controlar usando la desigualdad de Cauchy-Schwarz junto con que $\frac{1}{(\log(\log(n)))^{\frac{2}{3}}} \lesssim \varepsilon_n \leq 1$ y $\nu_n \lesssim \frac{1}{\log(n)}$:

$$\begin{aligned}
\frac{1}{\log(n)} \left| \sum_{j=1}^n A_{i,j} (\hat{h}_j^{(U)} - h_j) \right| &\leq \frac{1}{\log(n)} \|A_{i,\bullet}\|_2 \|\hat{h}^{(U)} - h\|_2 \\
&\lesssim \frac{1}{\log(n)} \left(\frac{\log(n)}{\sqrt{n}} \right) \left(\max \left\{ \frac{1}{\sqrt{\varepsilon_n} \sqrt{\log(n)}}, \nu_n, \varepsilon_n \sqrt{\nu_n} \right\} \sqrt{n} \right) \\
&= \max \left\{ \frac{1}{\sqrt{\varepsilon_n} \sqrt{\log(n)}}, \nu_n, \varepsilon_n \sqrt{\nu_n} \right\} \\
&\lesssim \frac{1}{\sqrt{\varepsilon_n} \sqrt{\log(n)}} \\
&\lesssim o(1) \varepsilon_n;
\end{aligned} \tag{5.118}$$

notamos que la última desigualdad vale pues

$$\frac{1}{\sqrt{\varepsilon_n} \sqrt{\log(n)}} = \frac{1}{\varepsilon_n^{\frac{3}{2}} \sqrt{\log(n)}} \varepsilon_n \lesssim \frac{\log(\log(n))}{\sqrt{\log(n)}} \varepsilon_n = o(1) \varepsilon_n. \tag{5.119}$$

Luego, buscaremos acotar el segundo término en (5.117), que será más laborioso.

Análogamente a (3.83), podremos usar la independencia entre $\hat{h}^{(U)} - h$ y $\hat{A}_{i,\bullet}$ para aplicar el Lema 3.5.4. Tomando en el enunciado del lema $\eta = \frac{2}{a}$, $p = a \frac{\log(n)}{n}$, $\gamma = \frac{1}{\sqrt{\log(n)}}$ y $x = \frac{\|\hat{h}^{(U)} - h\|_2}{\sqrt{n} \|\hat{h}^{(U)} - h\|_\infty}$ obtendremos que con probabilidad mayor o igual que $1 - o(n^{-1})$ valdrá que

$$\begin{aligned}
\left| \sum_{j=1}^n (\hat{A}_{i,j} - A_{i,j}) (\hat{h}_j^{(U)} - h_j) \right| &\lesssim \log(n) \frac{\max \left\{ \|\hat{h}^{(U)} - h\|_\infty, \frac{\|\hat{h}^{(U)} - h\|_2}{\gamma \sqrt{n}} \right\}}{\log\left(\frac{1}{\gamma}\right)} \\
&\asymp \log(n) \frac{\max \left\{ \|\hat{h}^{(U)} - h\|_\infty, \|\hat{h}^{(U)} - h\|_2 \frac{\sqrt{\log(n)}}{\sqrt{n}} \right\}}{\log(\log(n))}.
\end{aligned} \tag{5.120}$$

A continuación, podemos usar que por ser esperanza de funciones que toman valores entre -1 y 1 se verifica que

$$\|\hat{h}^{(U)} - h\|_\infty \leq \|\hat{h}^{(U)}\|_\infty + \|h\|_\infty \leq 2, \tag{5.121}$$

y además como $\varepsilon_n = \omega\left(\frac{1}{(\log(\log(n)))^{\frac{2}{3}}}\right)$ y $\nu_n \lesssim \frac{1}{\log(n)}$ podemos deducir que

$$\begin{aligned} \|\hat{h}^{(U)} - h\|_2 \frac{\sqrt{\log(n)}}{\sqrt{n}} &\lesssim \left[\max \left\{ \frac{1}{\sqrt{\varepsilon_n} \sqrt{\log(n)}}, \nu_n, \varepsilon_n \sqrt{\nu_n} \right\} \sqrt{n} \right] \left[\frac{\sqrt{\log(n)}}{\sqrt{n}} \right] \\ &\lesssim \frac{1}{\sqrt{\varepsilon_n}} \\ &\lesssim o(\log(\log(n))) \varepsilon_n \end{aligned} \quad (5.122)$$

$$\lesssim o(\log(\log(n))) \varepsilon_n \quad (5.123)$$

y así concluir que como queríamos ver

$$\frac{1}{\log(n)} \left| (\hat{A} - A)_{i,\bullet} \cdot (\hat{h}^{(U)} - h) \right| \lesssim \frac{1}{\log(n)} \left(\log(n) \frac{o(\log(\log(n))) \varepsilon_n}{\log(\log(n))} \right) \asymp o(1) \varepsilon_n, \quad (5.124)$$

lo que completa la demostración. Aclaremos que en (5.122) para acotar $\frac{1}{\sqrt{\varepsilon_n}}$ podemos proceder análogamente a (5.119) para obtener que

$$\frac{1}{\sqrt{\varepsilon_n}} = \frac{1}{\varepsilon_n^{\frac{3}{2}}} \varepsilon_n = o(\log(\log(n))) \varepsilon_n \quad (5.125)$$

como escribimos. ■

Finalmente, como corolario obtendremos la consistencia fuerte del estimador por vecindades:

[5.5.9] Teorema: Consistencia fuerte del estimador por vecindades $\hat{\mathcal{C}}_{\text{HEV}}$.

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos balanceados con $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para $a > b > 0$, donde además $\frac{a+b}{2} > 1 + \sqrt{ab}$, y supongamos que

$$\varepsilon_n = \omega\left(\frac{1}{(\log(\log(n)))^{\frac{2}{3}}}\right) \text{ y } \varepsilon_n = o(1).$$

Sea \mathcal{U} una partición de $V \setminus \mathcal{K}$ elegida independientemente y uniformemente al azar tal que para cada $U \in \mathcal{U}$, $\nu_n = \Theta\left(\frac{1}{\log(n)}\right)$.

Luego, tenemos que el estimador por vecindades es fuertemente consistente.

Demostración. Por la Proposición 5.5.7 sabemos que el estimador $\hat{\mathcal{C}}_{\text{HEA2}}$ es fuertemente consistente, y más aún existe $C > 0$ tal que $\left| (\Pi \hat{Q} \Pi h + v)_i \right| \geq C \varepsilon_n > 0$ para cada $1 \leq i \leq n$ con probabilidad mayor que $1 - o(1)$.

Basta luego ver que para cada conjunto $U \in \mathcal{U}$ y para cada $i \in U$, $(\Pi Q \Pi \hat{h}^{(U)} + v)_i$ está más cerca que $C \varepsilon_n$ de $(\Pi \hat{Q} \Pi h + v)_i$ para n suficientemente grandes con alta probabilidad para ver que los signos coinciden, y en consecuencia la estimación también. Pero eso es exactamente lo que afirma la Proposición 5.5.8, por la cual sabemos que con probabilidad mayor que $1 - o(1)$ vale que si $i \in U$ entonces tendremos que $\left| (\Pi Q \Pi \hat{h}^{(U)} + v)_i - (\Pi \hat{Q} \Pi h + v)_i \right| \leq o(1) \varepsilon_n$, lo cual muestra lo afirmado. Más aún, también es cierto con probabilidad mayor que $1 - o(n^{-\xi})$ para ξ suficientemente pequeño, porque el orden de decaimiento en la Proposición 5.5.8 depende esencialmente de los ordenes de decaimiento del Teorema 3.3.4, de la Proposición 2.4.11 y del Lema 5.5.6, que permiten ese orden ligeramente mayor: en consecuencia

por la Observación 2.2.5 conseguimos que simultáneamente valen las cotas para cada $U \in \mathcal{U}$ con probabilidad mayor que $1 - o\left(\frac{\log(n)}{n^\xi}\right) = 1 - o(1)$ y completamos la demostración. ■

Habiendo mostrado la consistencia fuerte de este estimador, terminaremos mencionando que todos los argumentos desarrollados aquí se extienden sin más problemas al caso de k comunidades balanceadas. Además, en las notas discutimos sobre los límites en la estimación en este nuevo contexto semi-supervisado, y algunas generalizaciones o variantes de los estimadores por extensiones armónicas.

En el siguiente capítulo exploraremos aspectos de índole más práctica y realizaremos algunas simulaciones para los estimadores presentados a lo largo del trabajo.

5.6 Notas

[5.6.1] **Nota: Otras maneras de lograr consistencia fuerte a partir de extensiones armónicas**

En [MNS16] los autores introducen un estimador que mejora las predicciones de un algoritmo débilmente consistente para hacerlo fuertemente consistente en el modelo BSBM pero cuando no se conoce previamente la comunidad de ningún nodo. Este estimador utiliza la técnica de “partir el grafo”: primero se particionan los nodos en una partición \mathcal{U} adecuada y después para cada $U \in \mathcal{U}$ se estima la comunidad de los nodos allí a partir de las estimaciones débilmente consistentes sobre $V \setminus U$; esto recordará la estrategia empleada en este capítulo con el estimador por vecindades.

Resulta que su estrategia puede adaptarse al caso semi-supervisado, donde sí conocemos las comunidades de algunos nodos. En efecto, lo que haremos será apartar los nodos para los cuales conocemos la comunidad \mathcal{K} y después particionar $V \setminus \mathcal{K}$ adecuadamente en \mathcal{U} , pero ahora para predecir las comunidades de los nodos en $U \in \mathcal{U}$ utilizaremos las estimaciones débilmente consistentes de nuestro algoritmo pero sobre $V \setminus U$ donde incorporamos nuevamente los nodos a los cuales les conocemos la comunidad. Análogamente a como se muestra en el trabajo, las estimaciones obtenidas de esta manera serán fuertemente consistentes, y vamos a tener entonces una manera alternativa de “partir el grafo” para lograr consistencia fuerte a partir de extensiones armónicas.

Mencionamos que al igual que en el trabajo original se puede usar cualquier otro estimador débilmente consistente y llegar al mismo resultado.

[5.6.2] **Nota: Otros laplacianos**

En la demostración de la Proposición 5.1.13 definimos un Laplaciano Δ . Esto no fue casualidad: hay teoría para funciones armónicas discretas que recupera construcciones típicas del estudio de ecuaciones diferenciales en derivadas parciales como funciones de Green, y teoremas como la desigualdad de Harmack o el teorema de Liouville. En particular, también se generalizan los p -Laplacianos al contexto discreto, y pueden obtenerse resultados similares. Así, podemos pensar en análogos a los estimadores por extensiones armónicas...

pero para extensiones p -armónicas. Queda pendiente explorar qué resultaría de utilizar estas generalizaciones.

[5.6.3] Nota: Límites de la predicción en contextos semi-supervisados

Mencionamos que aún cuando parece que contamos con más información, seguiremos teniendo los mismos límites teóricos para nuestras estimaciones que los que mencionamos en el Capítulo 3, obteniendo el siguiente resultado:

[5.6.4] Teorema: Factibilidad de consistencia fuerte para estimadores semi-supervisados en el modelo BSBM

Sea $G_{\mathcal{C}} \sim \text{BSBM}\left(n, 2, p = a \frac{\log(n)}{n}, q = b \frac{\log(n)}{n}\right)$ una familia de modelos balanceados para a, b números reales positivos. Supongamos además que para los nodos en $\mathcal{K} \subseteq V$ conocemos su comunidad, y que $\#\mathcal{C}_i \cap \mathcal{K}$ es constante en i , y con $\#(V \setminus \mathcal{K}) = \Omega(n)$. Luego vale que

- si $\frac{a+b}{2} > 1 + \sqrt{ab}$ entonces existe algún algoritmo fuertemente consistente eficiente para resolver el problema de detección de comunidades, y
- si $\frac{a+b}{2} < 1 + \sqrt{ab}$ entonces no existe ningún algoritmo fuertemente consistente que resuelva el problema de detección de comunidades.

Demostrar la imposibilidad se puede lograr de manera similar a la demostración en el Capítulo 3, y la dificultad es nomás técnica de llevar cuenta de que hay menos nodos que podemos clasificar erróneamente.

En cuanto a demostrar la factibilidad, podemos usar por ejemplo los estimadores por extensiones armónicas por vecindades, o también olvidarnos que conocemos las comunidades de algunos nodos y usar los métodos espectrales que ya son fuertemente consistentes bajo estas condiciones.

Aunque pareciese entonces que el estudio de estimadores en el caso semi-supervisado no aporta estrategias relevantes para atacar el problema de detección de comunidades, nos gustaría mencionar algunas grandes ventajas con las que cuentan estos métodos: en primer lugar, preservan la información que ya se tenía, lo cual no necesariamente ocurre con los estimadores anteriores; y en segundo lugar, como exploraremos en el Capítulo 6, en particular con los métodos por extensiones armónicas puede realizarse una implementación mediante métodos de Montecarlo que suelen en la práctica resultar rápidos para calcular la comunidad de un subconjunto pequeño de nodos.

[5.6.5] Nota: Modelo del votante

Podemos formalizar que la proporción de tiempo de permanencia en cada opinión para cada votante es la extensión armónica que queremos mediante la Proposición 5.1.13 simplemente verificando que es una función armónica que sobre \mathcal{K} coincide con $\sum_{l=1}^k \mathbb{1}_{\mathcal{K}_l} e_l$.

6. Simulaciones

6.1 Algoritmos para los calcular los estimadores

En primer lugar, discutamos algunas maneras de calcular los estimadores que estudiamos a lo largo del trabajo: los estimadores espectrales, los estimadores de minimización de corte y los estimadores por extensiones armónicas. Mientras que un análisis totalmente exhaustivo de los algoritmos escapa el alcance de esta tesis, nos concentraremos en demostrar su corrección y hacer un estudio de su complejidad teórica. Con este objetivo en mente, hagamos primero un breve repaso acerca de los que tendrán las operaciones usuales con vectores y matrices que queremos usar:

[6.1.1] Observación: Costos de operar con vectores y matrices

Las operaciones coordenada a coordenada pueden realizarse con complejidad temporal en el peor caso $\mathcal{O}(\text{costo}(\text{operación}) \cdot m)$ para matrices o vectores de m coeficientes en total. En particular, la suma entre dos matrices o vectores es lineal en el tamaño de estas estructuras. Notaremos simplemente $f(A)$ cuando aplicamos la función $f : \mathbb{R} \rightarrow \mathbb{R}$ a la matriz A en los algoritmos que siguen.

El producto entre matrices puede realizarse con complejidad temporal en el peor caso menor que $\mathcal{O}(nmr)$; en particular cuando $n = m = r$ obtenemos una complejidad de $\mathcal{O}(n^3)$.

Combinando ambas, obtenemos que la complejidad temporal de calcular normas o distancias es lineal en el tamaño de los vectores.

[6.1.2] Observación: Costos de operar con matrices y vectores en regímenes esparsos

Mencionamos que en nuestro caso al trabajar en regímenes donde hay pocas aristas muchas veces nuestras matrices serán algo esparsas y podremos obtener mejores cotas para la complejidad de los métodos. Para matrices con m coeficientes no nulos la multiplicación de esta matriz por un vector con n coordenadas puede ser implementada con complejidad temporal de $\mathcal{O}(\max\{m, n\})$.

En nuestro caso esto nos permitirá acotar la complejidad temporal con alta probabilidad por cotas de ordenes menores que en el caso general, pues siempre que usemos la matriz de adyacencia de un modelo estocástico de bloques que cumpla las conexiones

de conectividad por el Teorema 2.4.11 tendremos que los grados de cada nodo son del orden de $\Theta(\log(n))$ y así la matriz de adyacencia del grafo tendrá $\Theta(n \log(n))$ coeficientes no nulos con alta probabilidad.

Comencemos luego por los estimadores espectrales para dos comunidades. Recordando la definición del estimador en 3.3.5, tendremos que como en el algoritmo 1 habrá que encontrar el segundo autovector de la matriz de adyacencia del modelo y después estimar las comunidades de acuerdo al signo de sus coordenadas.

Algoritmo 1: Estimadores espectrales para dos comunidades

Datos: A la matriz de adyacencia del modelo y k la cantidad de comunidades

Resultado: C las comunidades estimadas

```

1  $u \leftarrow \text{autovectores}(A, 2)$ 
2  $C \leftarrow \phi(\text{signo}(u))$ 

```

Como comentario, para pasar a una cantidad arbitraria de comunidades k basta como en el algoritmo 2 utilizar los autovectores entre el segundo y el k -ésimo y ahora predecir las comunidades mediante k -medias, como discutimos en la Nota 3.6.2.

Algoritmo 2: Estimadores espectrales

Datos: A la matriz de adyacencia del modelo y k la cantidad de comunidades

Resultado: C las comunidades estimadas

```

1  $u \leftarrow \text{autovectores}(A, 2 \text{ a } k)$ 
2  $C \leftarrow \text{kMedias}(u, k)$ 

```

Desde ya, para analizar en detalle estos algoritmos precisamos especificar cómo vamos a hacer para encontrar los autovectores de la matriz de adyacencia. Nos limitaremos a notar que el problema puede resolverse con una complejidad temporal de $\mathcal{O}(n^3)$ en el peor caso, esencialmente la complejidad temporal del producto entre dos matrices. En el caso de la generalización para k comunidades, el algoritmo de **k-medias** tiene una complejidad temporal de $\mathcal{O}(\text{iteraciones} \cdot k \cdot n)$ pues calcular las distancias se realiza con complejidad lineal y debremos hacerlo una vez por nodo, que habrán n , y por centroide, que habrán k , para cada iteración. En total, obtenemos que los algoritmos espectrales podrían ser implementados con una complejidad de $\mathcal{O}(n^3)$ en el peor caso, aunque queda lugar para optimizar y también aproximar mediante algoritmos iterativos.

Continuemos luego por los algoritmos por extensiones armónicas. Siguiendo la Definición 5.2.1, el estimador por extensiones armónicas para dos comunidades puede ser implementado como en el algoritmo 3, mientras que análogamente la generalización a k comunidades puede ser implementada como en el algoritmo 4.

Vale la pena una vez más en este punto detenernos para comentar distintas estrategias para calcular las extensiones armónicas. En primer lugar, como discutimos en la demostración

Algoritmo 3: Estimadores por extensiones armónicas

Datos: A la matriz de adyacencia del modelo, k la cantidad de comunidades, $K \subseteq V$ los nodos con comunidad conocida y C_K las comunidades sobre ellos

Resultado: C las comunidades estimadas para todos los nodos de V

- 1 $D \leftarrow \text{diagonal}(A \cdot 1)$
- 2 $Q \leftarrow \text{inversa}(D) \cdot A$
- 3 $h \leftarrow \text{extensión armónica}(Q, 2, C_K)$
- 4 $C \leftarrow \phi(\text{signo}(h))$

Algoritmo 4: Estimadores por extensiones armónicas

Datos: A la matriz de adyacencia del modelo, k la cantidad de comunidades, $K \subseteq V$ los nodos con comunidad conocida y C_K las comunidades sobre ellos

Resultado: C las comunidades estimadas para todos los nodos de V

- 1 $D \leftarrow A \cdot 1$
- 2 $Q \leftarrow \text{inversa}(D) \cdot A$
- 3 $h \leftarrow \text{extensión armónica}(Q, k, C_K)$
- 4 $C \leftarrow \text{argmax}(h)$

de 5.1.13, podremos encontrar la extensión resolviendo un sistema de ecuaciones lineales como en el algoritmo 5. Como no tendremos en general buenas propiedades para $I - \Pi Q$ como simetría perdemos el acceso a algunos métodos especializados para resolverlo. En particular, podremos resolver el sistema de ecuaciones lineales con una complejidad temporal de $\mathcal{O}(n^3)$ en el peor caso.

Algoritmo 5: Extensiones armónicas mediante la inversión

Datos: Q una matriz de probabilidades de transición en V , $W \subseteq V$ el borde y g la condición de borde

Resultado: f la extensión armónica a V de g

- 1 $\Pi \leftarrow \text{proyección}(S)$
- 2 $f \leftarrow \text{resolver}(f = \Pi Q f + g)$

Es así que es de nuestro interés encontrar métodos más rápidos. En este caso contaremos con dos alternativas que nos permitirán aproximar la extensión en vez que hallarla precisamente: podremos hacerlo mediante algoritmos iterativos o a través de métodos de Montecarlo. Presentemos primero los algoritmos, y en la siguiente sección daremos algunos resultados de complejidad y consistencia.

En primer lugar, presentemos el algoritmo iterativo 6. Como consecuencia del teorema de punto fijo de Banach tendremos que la sucesión convergerá a un punto fijo, y con cuidado obtendremos cotas teóricas para la velocidad de convergencia.

En segundo lugar, en el algoritmo 7 tenemos un procedimiento en el espíritu de los

Algoritmo 6: Extensiones armónicas mediante algoritmos iterativos

Datos: Q una matriz de probabilidades de transición en V , $W \subseteq V$ el borde, g la condición de borde y f_0 un vector inicial

Resultado: f una aproximación de la extensión armónica a V de g

```

1  $f \leftarrow f_0$ 
2  $\Pi \leftarrow \text{proyección}(S)$ 
3 para  $i \leftarrow 1$  a iteraciones hacer
4   |  $f \leftarrow \Pi Q f + g$ 
5 fin

```

métodos de Montecarlo, que nos podrá recordar a la Observación 5.1.1. Por la Ley de los grandes números, la sucesión de promedios converge casi seguramente a la esperanza que queremos hallar, y luego para suficientes iteraciones esperaremos hallar una aproximación suficientemente precisa. Podemos demostrar que la cantidad de iteraciones necesarias será del orden de $\mathcal{O}\left(\frac{1}{\varepsilon_n}\right)$ con alta probabilidad, en cuyo caso obtenemos que la complejidad teórica de estas implementaciones será baja.

Algoritmo 7: Extensiones armónicas mediante el método de Montecarlo

Datos: Q una matriz de probabilidades de transición en V , $W \subseteq V$ el borde, g la condición de borde

Resultado: f una aproximación de la extensión armónica a V de g

```

1 para  $v \in V \setminus W$  hacer
2   | para  $i \leftarrow 1$  a iteraciones hacer
3     |  $\text{caminata}_i \leftarrow v$ 
4     | mientras  $\text{caminata}_i \notin W$  hacer
5       |  $\text{caminata}_i \leftarrow \text{avanzar}(Q, \text{caminata}_i)$ 
6     | fin
7   | fin
8   |  $f_v \leftarrow \text{promedio}(g_{\text{caminata}_i})$ 
9 fin

```

Los estimadores por extensiones armónicas por vecindades podemos construirlos a partir del estimador por extensiones armónicas, siguiendo la Definición 5.5.3, a partir de aproximar el valor de la extensión armónica empírica de los vecinos de cada nodo que querramos catalogar en el subgrafo adecuado, y eligiendo una partición \mathcal{U} uniformemente al azar entre aquellas que tienen $\lceil \log(n) \rceil$ subconjuntos de (casi) igual tamaño.

Por último, mencionamos que el problema de minimizar el corte asociado a particiones de un grafo es NP-completo cuando tenemos restricciones sobre el tamaño de los conjuntos en la partición, con lo cual es necesario recurrir a alguna aproximación. En este caso utilizaremos para nuestros estimadores los ingeniosos algoritmos desarrollados por Buehler y Hein que se

encuentran disponibles en [la página de su grupo de investigación](#).

Antes de pasar a un análisis más detallado de los algoritmos para estimar las comunidades mediante extensiones armónicas, mencionamos que la complejidad en el peor caso de los algoritmos que conocemos en la literatura para estimar las comunidades acotada con alta probabilidad por $\mathcal{O}(\mathbf{n} \log(\mathbf{n})^C)$ para alguna potencia $C \geq 1$ en este régimen de densidad de aristas, $p_n, q_n = \Theta\left(\frac{\log(n)}{n}\right)$, es decir, salvo factores logarítmicos en la cantidad de nodos, la cantidad de aristas que tiene el modelo estocástico de bloques con alta probabilidad.

6.2 Consistencia y complejidad de los algoritmos para aproximar $\hat{\mathcal{C}}_{\text{HEV}}$

Antes de mostrar algunos resultados de simulaciones, veamos algunos resultados de consistencia y eficiencia para los algoritmos para aproximar los estimadores por extensiones armónicas. Nos concentraremos en el caso de dos comunidades porque es el que tenemos descrito con mayor precisión, pero hay resultados que generalizan sin más cambios a una cantidad arbitraria de comunidades, especialmente en el caso balanceado.

En primer lugar, para los métodos iterativos conseguimos que:

[6.2.1] **Proposición: Consistencia y complejidad de la aproximación de $\hat{\mathcal{C}}_{\text{HEV}}$ mediante métodos iterativos.**

Sea $G_{\mathcal{C}} \sim \text{SBM}(n, 2, (n_l)_{1 \leq l \leq 2}, p_n, q_n)$ una familia de modelos estocásticos de bloques en las hipótesis de 5.3.2 o 5.4.8 que garantizan la consistencia débil de los estimadores por extensiones armónicas.

Definamos inductivamente para cada $\hat{h}^{(0)} \in \mathbb{R}^n$ la sucesión dada por

$$\hat{h}^{(t+1)} = \Pi \hat{Q} \hat{h}^{(t)} + v, \quad (6.1)$$

donde Π , \hat{Q} y v están dados como en 5.2.3 o 5.4.5 respectivamente.

Luego vale que si $t = \omega\left(\log\left(\frac{1}{\varepsilon_n}\right) \frac{1}{\varepsilon_n}\right)$ entonces

$$\|\hat{h}^{(t)} - \hat{h}\|_{\infty} \leq o(\varepsilon_n). \quad (6.2)$$

En particular, el estimador por extensiones armónicas calculado mediante el algoritmo iterativo 6 es débilmente consistente y puede calcularse con complejidad temporal de orden $\mathcal{O}(\mathbf{n} \log(\mathbf{n})^3)$ con alta probabilidad para el caso de dos comunidades balanceadas.

Más aún, si valen las hipótesis del Teorema 5.5.9 entonces el estimador por vecindades inducido es fuertemente consistente y puede calcularse con alta probabilidad con complejidad temporal de orden $\mathcal{O}(\mathbf{n} \log(\mathbf{n})^4)$.

Para los métodos de Montecarlo resulta que durante la demostración precisaremos un control más fino de los coeficientes de \hat{h} de la extensión armónica empírica que no tenemos

aún. Por ejemplo, sería suficiente conocer que $|\hat{h}_i|$ sea de orden ε_n con alta probabilidad en el caso de los estimadores de 2 comunidades balanceadas, o que $|\hat{h}_{i,l}|$ es de orden constante con alta probabilidad en el caso de los estimadores generales, como ocurre con sus contrapartes teóricas. Como vimos en el capítulo anterior, las Proposiciones 5.3.1 y 5.4.7 no son suficientemente finas como para asegurar que para todos los coeficientes de la extensión empírica están cerca de los coeficientes de la solución teórica y solo resultan en que todos salvo una fracción negligible de nodos lo verifican. El paso restante para obtener estas estimaciones sería entonces un mejor control coeficiente a coeficiente, más en el estilo de las cuentas que hicimos en la Sección 5.5 al probar la consistencia fuerte del estimador por vecindades. En efecto, creemos que con un estudio más minucioso de las trayectorias realizadas por los paseos al azar, podremos mostrar que $\hat{h}_j^{(i)}$ y \hat{h}_j serán muy similares sobre vecinos $j \sim i$, de lo cual obtendremos que también \hat{h}_i y $(\Pi\hat{Q}\hat{h}^{(i)} + v)_i$ serán cercanos y entonces podremos concluir usando la Proposición 5.5.8 que más aún \hat{h}_i y $(\Pi\hat{Q}h + v)_i$ tendrán el mismo orden. ¿Por qué esperamos que esto sea cierto? Recordamos que \hat{h} y $\hat{h}^{(i)}$ son esperanzas sobre probabilidades relacionadas a las trayectorias de paseos al azar simétricas en grafos que son casi idénticos, difiriendo solo en un nodo: es así que trayectorias que eviten la región del grafo cercana a i deberían obtener probabilidades similares de ocurrir en ambos grafos y un valor idéntico para la función que integramos, con lo que acotando qué tan probable es que las trayectorias pasen por una región chica en el grafo y teniendo en cuenta las pequeñas diferencias en los grafos obtendremos una cota para la diferencia entre los coeficientes. Lo anotamos entonces a continuación, por ahora como conjetura:

[6.2.2] Conjetura: Orden de los coeficientes de la extensión armónica empírica

Sea $G_{\mathcal{C}} \sim \text{SBM}(n, 2, (n_l)_{1 \leq l \leq 2}, p_n, q_n)$ una familia de modelos estocásticos de bloques con $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para $a > b > 0$ que cumplen que $\frac{a+b}{2} > 1 + \sqrt{ab}$.

Luego, en el caso del estimador para 2 comunidades balanceadas, existen constantes $C_1, C_2 > 0$ tales que para cada $i \notin \mathcal{K}$ vale que

$$C_1 \varepsilon_n \leq |\hat{h}_i| \leq C_2 \varepsilon_n. \quad (6.3)$$

Además, en el caso del estimador general, existen constantes $C_1, C_2 > 0$ tales que para cada $i \notin \mathcal{K}$ y cada $1 \leq l \leq 2$ vale que

$$C_1 \leq \hat{h}_{i,l} \leq C_2. \quad (6.4)$$

Además, precisaremos un control sobre la cantidad de pasos que se toma cada paseo para hallar los nodos conocidos, y así acotar la complejidad del algoritmo. Queda para trabajo futuro determinar y escribir la demostración para la cota adecuada, pero por ahora como conjetura vamos a proponer que:

[6.2.3] Conjetura: Cota para la cantidad de pasos que toma un paseo simétrico en llegar a \mathcal{K}

Sea $G_{\mathcal{C}} \sim \text{SBM}(n, 2, (n_l)_{1 \leq l \leq 2}, p_n, q_n)$ una familia de modelos estocásticos de bloques con $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$ para $a > b > 0$ suficientemente grandes y diferentes.

Supongamos que $\varepsilon_n = \omega\left(\frac{1}{\log(\log(n))^{\frac{2}{3}}}\right)$. Luego, vale que existe una función $f(n)$ tal que para todo $r > 0$ y para cada $i \in V$ existe $C > 0$ que verifica que para cada paseo al azar simétrico $(X_t)_{t \in \mathbb{N}_0}$ vale que

$$\mathbb{P}(\{\tau_{\mathcal{K}} \leq Cf(n)\} | \{X_0 = i\}) = 1 - o\left(\frac{1}{n^r}\right) \quad (6.5)$$

para $\tau_{\mathcal{K}}$ el tiempo de primera visita a \mathcal{K} .

Provisionalmente conjeturamos $f(n) \simeq \frac{\log(n)}{\varepsilon_n} \lesssim \log(n)^2$.

Dejamos este resultado como conjetura principalmente para no extender más aún la tesis, porque creemos que puede demostrarse a partir del Lema 6.2.7. En efecto, de allí puede deducirse que $\tau_{\mathcal{K}}$ tiene colas acotadas por las de una geométrica de parámetro $C\varepsilon_n$ para alguna constante $C > 0$, con alta probabilidad. Es así que procediendo cuidadosamente, por ejemplo siguiendo el enfoque de [Ver18] para variables sub-exponenciales, también vamos a poder dar cotas para $\tau_{\mathcal{K}}$ en el orden de $\frac{1}{\varepsilon_n}$ (la esperanza de la geométrica), con alta probabilidad. Este enfoque también puede generalizarse para acotar la cantidad de pasos que realizan múltiples paseos mediante la desigualdad de Bernstein, y así podemos obtener cotas para la complejidad teórica de las simulaciones.

Confiando entonces en que estas conjeturas se verifican como ciertas, obtendremos en ese caso que:

[6.2.4] Proposición: Correctitud y complejidad de la aproximación de \widehat{C}_{HEV} via métodos de Montecarlo.

Sea $G_{\mathcal{C}} \sim \text{BSBM}(n, 2, p_n, q_n)$ una familia de modelos estocásticos de bloques en las hipótesis de 5.5.9 que garantizan la consistencia fuerte de los estimadores por extensiones armónicas.

Fijemos $W \subseteq V \setminus \mathcal{K}$ un subconjunto de nodos.

Luego, el estimador por vecindades basado en el algoritmo 7 es fuertemente consistente, y halla las comunidades de los nodos en W con complejidad temporal $\mathcal{O}\left(\#W \cdot \log(\#W \log(n)) \cdot \log(n)^3 \cdot \frac{1}{\varepsilon_n^3}\right)$ en el peor caso, con alta probabilidad.

Observamos en este punto que cuando W es suficientemente chico, por ejemplo cuando $\#W = o\left(\frac{n}{\log(n)^4}\right)$, la complejidad teórica obtenida es $o(n)$ y en particular es de un orden menor que la complejidad teórica que tienen los demás algoritmos que encontramos en la literatura.

Notamos aquí que se pueden obtener resultados de convergencia débil, también en el

caso general de k comunidades, de manera análoga, pero con complejidades más altas que la implementación mediante algoritmos iterativos. Es por eso que nos limitamos a dar ideas para completar una demostración en el caso de subconjuntos chicos de nodos, para el cual es necesaria la consistencia fuerte para asegurar que ninguno de los nodos en el subconjunto elegido es alguno de los que cuenta con una estimación errónea.

Antes de demostrarlas, daremos algunos lemas auxiliares que nos permitirán entender cómo se ve un modelo estocástico de bloques mirando más allá de los vecinos de un nodo. Para facilitar la escritura de las demostraciones que siguen, fijemos la siguiente notación para los entornos de los nodos en el grafo.

[6.2.5] **Notación: Entornos en un grafo.**

Dado un grafo $G = (V, E)$ y una cantidad de pasos r , notaremos

$$B_r(i) = \{j \in V \mid \text{dist}(i, j) < r\} \quad (6.6)$$

a la **bola abierta de radio r centrada en i ,**

$$\bar{B}_r(i) = \{j \in V \mid \text{dist}(i, j) \leq r\} \quad (6.7)$$

a la **bola cerrada de radio r centrada en i , y**

$$\partial B_r(i) = \partial \bar{B}_r(i) = \bar{B}_r(i) \setminus B_r(i) \quad (6.8)$$

a la **esfera de radio r centrada en i .**

Este primero nos da un panorama más completo de lo que ocurre en las vecindades de los nodos, hasta 2 pasos de distancia. Esto será útil en el futuro porque, como hay más nodos a dos pasos de distancia, eventos que entre los vecinos de un nodo fijo son más infrecuentes aquí pasan con más alta probabilidad, lo que permitirá salvar algunos problemas técnicos.

[6.2.6] **Lema: Los nodos a distancia dos en el modelo SBM**

Sea $G_{\mathcal{C}} \sim \text{SBM}(n, k, (n_l)_{1 \leq l \leq k}, p_n, q_n)$ para $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$, con $0 < b < a$, $n_l = \Omega(n)$ para cada $1 \leq l \leq k$ y $\liminf_{n \rightarrow \infty} \min_{1 \leq l \leq k} \left\{ a \frac{n_l}{n} + b \left(1 - \frac{n_l}{n}\right) \right\} > 1$.

Luego, vale que existen constantes $C_1, C_2 > 0$ tales que para cada nodo $i \in V$

$$\mathbb{P} \left(\left\{ C_1 \log(n)^2 \leq \#(\partial \bar{B}_2(i)) \leq C_2 \log(n)^2 \right\} \right) \geq 1 - o(n^{-1}). \quad (6.9)$$

Más aún, vale también que todas las comunidades tienen el mismo orden de nodos a distancia dos. Es decir, para cada $1 \leq l \leq k$ existen constantes $C_1, C_2 > 0$ tales que para cada nodo $i \in V$

$$\mathbb{P} \left(\left\{ C_1 \log(n)^2 \leq \#(\partial \bar{B}_2(i) \cap \mathcal{C}_l) \leq C_2 \log(n)^2 \right\} \right) \geq 1 - o(n^{-1}). \quad (6.10)$$

Por último, si uniformemente al azar para cada comunidad se eligen $g(n) = \varepsilon_n n$ nodos de comunidad conocida, donde $\varepsilon_n = \omega\left(\frac{1}{\log(n)}\right)$, entonces vale que existen constantes

$C_1, C_2 > 0$ tales que

$$\mathbb{P} \left(\left\{ C_1 \varepsilon_n \log(n)^2 \leq \#(\partial \overline{B}_2(i) \cap \mathcal{K}_l) \leq C_2 \varepsilon_n \log(n)^2 \right\} \right) \geq 1 - o(n^{-1}). \quad (6.11)$$

Demostración. Demostraremos aquí la primera cota (6.9), y después mostraremos que la segunda puede concluirse análogamente antes de deducir la tercera.

En primer lugar, recordemos que por la Proposición 2.4.11 sabemos que existen constantes $0 < C_1 < 1 < C_2$ tales que para cada $i \in V$

$$\mathcal{A}_i = \left\{ C_1 \log(n) \leq \widehat{\text{deg}}_n(i) \leq C_2 \log(n) \right\} \quad (6.12)$$

ocurre con probabilidad más alta que $1 - o(n^{-1-\xi})$ para un $\xi > 0$ suficientemente chico.

Podemos así demostrar la cota superior: primero podemos usar que i va a tener con alta probabilidad una cantidad logarítmica de vecinos, y después condicionando podemos mostrar que a su vez cada uno de ellos va a tener una cantidad logarítmica de vecinos nuevamente. Así, combinando todo como en en la Observación 2.2.5, podemos concluir que con probabilidad mayor que $1 - o(n^{-1-\xi})$ hay menos que $C_2^2 \log(n)^2$ nodos a distancia dos de i dado que los nodos a distancia exactamente dos son vecinos de los vecinos de i .

Ahora mostremos que vale con alta probabilidad la cota inferior. Una vez más buscaremos explotar que las vecindades de cada nodo tienen tamaño logarítmico, y tenemos $\Theta(\log(n))$ de ellas, para concluir. No podemos contar como para la cota superior porque contaríamos posiblemente muchas veces a los mismos nodos, para aquellos nodos que son vecinos de múltiples vecinos de i a la vez. Por eso, cambiaremos el enfoque, y nos concentraremos en estudiar la probabilidad de que un nodo que no está en la vecindad de i se conecte a algún nodo en la vecindad, y luego hallaremos una estimación para la cantidad de nodos a distancia exactamente dos mediante desigualdades de concentración.

En primer lugar, mostraremos que condicionando podemos reducir el problema a estudiar la probabilidad de que la cantidad de vecinos sea suficientemente alta para vecindades fijas. Con este propósito, dada una familia de subconjuntos disjuntos de nodos en el grafo, $J = (J_l)_{1 \leq l \leq k}$, notaremos

$$\mathcal{B}_J = \left\{ \partial \overline{B}_1(i) \cap \mathcal{C}_l = J_l \text{ para cada } 1 \leq l \leq k \right\}, \quad (6.13)$$

y diremos que una familia J es admisible cuando vale que $C_1 \log(n) \leq \#J \leq C_2 \log(n)$. Como recordamos anteriormente, tendremos que las familias J que describen bien los entornos son típicamente admisibles

$$\mathbb{P}(\cup_{J \text{ admisibles}} \mathcal{B}_J) = \mathbb{P}(\mathcal{A}_i) = 1 - o(n^{-1-\xi}), \quad (6.14)$$

así que nos restringiremos a estudiar lo que pasa sobre ellas. En efecto, como para estas familias vamos a poder demostrar de manera uniforme en J que para $C > 0$ y $\xi > 0$ suficientemente chicos vale que

$$\mathbb{P} \left(\left\{ \#(\partial \overline{B}_2(i)) \geq C \log(n)^2 \right\} \mid \mathcal{B}_J \right) \geq 1 - o(n^{-1-\xi}), \quad (6.15)$$

entonces condicionando vamos a poder deducir la cota inferior como queríamos:

$$\begin{aligned}
& \mathbb{P} \left(\left\{ \#(\partial \bar{B}_2(i)) \geq C \log(n)^2 \right\} \right) & (6.16) \\
& \geq \sum_{J \text{ admisible}, \mathbb{P}(\mathcal{B}_J) > 0} \mathbb{P} \left(\left\{ \#(\partial \bar{B}_2(i)) \geq C \log(n)^2 \right\} \middle| \mathcal{B}_J \right) \mathbb{P}(\mathcal{B}_J) \\
& \geq \left(1 - o\left(n^{-1-\xi}\right)\right) \sum_{J \text{ admisibles}, \mathbb{P}(\mathcal{B}_J) > 0} \mathbb{P}(\mathcal{B}_J) \\
& = \left(1 - o\left(n^{-1-\xi}\right)\right) \mathbb{P}(\mathcal{A}_i) \\
& = 1 - o\left(n^{-1-\xi}\right).
\end{aligned}$$

Nos queda entonces demostrar (6.15), para lo cual resta solamente aprovechar la información que nos da J de la vecindad. Por un lado, tenemos que si un nodo j no es vecino de i y está en la comunidad l , tendremos que hay alguna arista que conecta a j con algún vecino de i con probabilidad

$$\begin{aligned}
& 1 - \left(1 - a \frac{\log(n)}{n}\right)^{\#J_l} \prod_{t=1, t \neq l}^k \left(1 - b \frac{\log(n)}{n}\right)^{\#J_t} & (6.17) \\
& = 1 - \left(1 - a \frac{\log(n)}{n}\right)^{\#J_l} \left(1 - b \frac{\log(n)}{n}\right)^{\#J_{\neq l}},
\end{aligned}$$

donde notamos por $\#J_{\neq l}$ a $\#J_{\neq l} = \sum_{t=1, t \neq l}^k \#J_t$ para simplificar la notación. Por otro lado, como J es admisible vale que hay por lo menos $n_l - \#J_l - 1 \geq n_l - C_2 \log(n) - 1 = n_l(1 - o(1))$ tales nodos de comunidad l , con lo que tendremos por lo menos $\lceil (1 - \eta)n_l \rceil$ nodos de cada comunidad $1 \leq l \leq k$ como candidatos a estar a distancia dos, para una constante $\eta > 0$ suficientemente chica. Es así que en estas condiciones si para cada $1 \leq l \leq k$ llamamos Z_l a la cantidad de nodos entre los $\lceil (1 - \eta)n_l \rceil$ nodos de índice más chico de comunidad l en $\partial \bar{B}_2(i)$ y además llamamos Z a su suma $Z = \sum_{l=1}^k Z_l$, tendremos que $\#(\partial \bar{B}_2(i)) \geq Z$ y además las variables Z_l son independientes y con distribución dada por

$$Z_l \sim \text{Binomial} \left(\lceil (1 - \eta)n_l \rceil, 1 - \left(1 - a \frac{\log(n)}{n}\right)^{\#J_l} \left(1 - b \frac{\log(n)}{n}\right)^{\#J_{\neq l}} \right) \quad (6.18)$$

para cada $1 \leq l \leq k$. En particular, tenemos que

$$\mathbb{P} \left(\left\{ \#(\partial \bar{B}_2(i)) \geq C \log(n)^2 \right\} \middle| \mathcal{B}_J \right) \geq \mathbb{P} \left(\left\{ Z \geq C \log(n)^2 \right\} \middle| \mathcal{B}_J \right), \quad (6.19)$$

y resta solamente analizar el comportamiento de Z .

Como adelantamos, concluiremos usando desigualdades de concentración, para lo cual basta ver que la esperanza tiene el orden adecuado y mostrar que es improbable que la

variable sea mucho más chica. En cuanto a lo primero, un cálculo directo nos muestra que

$$\begin{aligned} \mathbb{E}[Z] &= \sum_{l=1}^k [(1-\eta)n_l] \left(1 - a \frac{\log(n)}{n}\right)^{\#J_l} \left(1 - b \frac{\log(n)}{n}\right)^{\#J_{\neq l}} \\ &= \sum_{l=1}^k [(1-\eta)n_l] \left[(a\#J_l + b\#J_{\neq l}) \frac{\log(n)}{n} (1 + o(1)) \right] \\ &\geq (1-\eta) \left(\sum_{l=1}^k \frac{n_l}{n} \right) (bC_1 \log(n)^2 (1 + o(1))) \\ &= (1-\eta)bC_1 \log(n)^2 (1 + o(1)), \end{aligned} \quad (6.20)$$

donde usamos que $a\#J_l + b\#J_{\neq l} \geq b\#J \geq bC_1 \log(n)$ porque J es admisible, y una aproximación por Taylor para la primera igualdad. En efecto, usando que suficientemente cerca de $x = 0$ valen que

$$\log(1-x) = -x + o(x) = -x(1 + o(1)) \quad (6.21)$$

y

$$1 - e^{-x} = x + o(x) = x(1 + o(1)), \quad (6.22)$$

tenemos que

$$\begin{aligned} &1 - \left(1 - a \frac{\log(n)}{n}\right)^{\#J_l} \left(1 - b \frac{\log(n)}{n}\right)^{\#J_{\neq l}} \\ &= 1 - \exp\left(\#J_l \log\left(1 - a \frac{\log(n)}{n}\right) + \#J_{\neq l} \log\left(1 - b \frac{\log(n)}{n}\right)\right) \\ &= 1 - \exp\left(-(\#J_l a + \#J_{\neq l} b) \frac{\log(n)}{n} (1 + o(1))\right) \\ &= (\#J_l a + \#J_{\neq l} b) \frac{\log(n)}{n} (1 + o(1)), \end{aligned} \quad (6.23)$$

donde notamos que todas $a \frac{\log(n)}{n}$, $b \frac{\log(n)}{n}$, $(\#J_l a + \#J_{\neq l} b) \frac{\log(n)}{n} = \mathcal{O}\left(\frac{\log(n)^2}{n}\right)$ cumplen que tienden a 0 cuando $n \rightarrow \infty$ para J admisibles. En cuanto a lo segundo, podemos usar la desigualdad de concentración del Lema 2.4.12 para ver que para cada $0 < x < 1$ constante y cada $0 < C < x(1-\eta)C_1 b$ vale que

$$\mathbb{P}\left(\{Z \leq C \log(n)^2\}\right) \leq \mathbb{P}\left(\{Z \leq x \mathbb{E}[Z]\}\right) \leq e^{-\gamma(x) \mathbb{E}[Z]} = o\left(n^{-1-\xi}\right) \quad (6.24)$$

para n suficientemente grandes y mostrar lo que queríamos. Aclaremos aquí que para la primera desigualdad usamos que en este caso $C \log(n)^2 \leq x \mathbb{E}[Z]$ lo que nos da la contención de conjuntos, y que para acotar el orden de decaimiento usamos que

$$\gamma(x) \mathbb{E}[Z] \geq (\gamma(x)(1-\eta)bC_1 + \gamma(x)o(1)) \log(n)^2 \geq \log(n) \quad (6.25)$$

para n suficientemente grandes.

Así queda demostrada (6.9). Como adelantamos, la desigualdad (6.10) que refiere a cada comunidad en concreto puede deducirse de manera análoga, y la única diferencia al estudiarla es que en ese caso vamos a poder reducir el problema a estudiar Z_l en vez que Z . Sin embargo, como análogamente a (6.20) vamos a poder obtener la misma cota

$$\mathbb{E} [Z_l] \geq (1 - \eta) b C_1 \log(n)^2 (1 + o(1)), \quad (6.26)$$

a partir de ahí la demostración es igual.

En cuanto a la última desigualdad (6.11), podemos deducirla de (6.10) para cada comunidad fija $1 \leq l \leq k$. Una vez más nos será útil condicionar, y para eso podemos notar para cada subconjunto $H \subseteq V \setminus \{i\}$ como Γ_H al conjunto

$$\Gamma_H = \{ \partial \bar{B}_2(i) \cap \mathcal{C}_l = H \} \quad (6.27)$$

y llamar admisibles a los conjuntos H que verifiquen que $C_1 \log(n)^2 \leq \#H \leq C_2 \log(n)^2$, donde las constantes hacen que para algún $\xi > 0$ suficientemente pequeño vale que

$$\mathbb{P} (\cup_{H \text{ admisible}} \Gamma_H) = 1 - o(n^{-1-\xi}), \quad (6.28)$$

como demostramos. En este caso, como para los conjuntos H admisibles vamos a poder demostrar que existen ciertas constantes $\tilde{C}_1, \tilde{C}_2 > 0$ tales que:

$$\mathbb{P} (\{ \tilde{C}_1 \varepsilon_n \log(n)^2 \leq \#(\partial \bar{B}_2(i) \cap \mathcal{K}_l) \leq \tilde{C}_2 \varepsilon_n \log(n)^2 \} \mid \Gamma_H) \geq 1 - o(n^{-1-\xi}), \quad (6.29)$$

vamos a tener que

$$\begin{aligned} & \mathbb{P} (\{ \tilde{C}_1 \varepsilon_n \log(n)^2 \leq \#(\partial \bar{B}_2(i) \cap \mathcal{K}_l) \leq \tilde{C}_2 \varepsilon_n \log(n)^2 \}) \quad (6.30) \\ & \geq \sum_{H \text{ admisible}, \mathbb{P}(\Gamma_H) > 0} \mathbb{P} (\{ \tilde{C}_1 \varepsilon_n \log(n)^2 \leq \#(\partial \bar{B}_2(i) \cap \mathcal{K}_l) \leq \tilde{C}_2 \varepsilon_n \log(n)^2 \} \mid \Gamma_H) \cdot \mathbb{P} (\Gamma_H) \\ & \geq \sum_{H \text{ admisible}, \mathbb{P}(\Gamma_H) > 0} (1 - o(n^{-1-\xi})) \mathbb{P} (\Gamma_H) \\ & \geq (1 - o(n^{-1-\xi})) \mathbb{P} (\cup_{H \text{ admisibles}} \Gamma_H) \\ & = (1 - o(n^{-1-\xi})) \end{aligned}$$

como queríamos.

Es así que solo resta ver que vale (6.29). Tendremos que salvar un ligero inconveniente técnico: como tenemos que en el grafo hay $g(n)$ nodos conocidos de la comunidad l elegidos uniformemente al azar entre los nodos de esa comunidad, la cantidad de nodos conocidos en H sigue una distribución hipergeométrica con parámetros n_l , $\#H$ y $g(n)$, y no binomial o suma de variables Bernoulli independientes como solimos acotar anteriormente. Ahora bien, por un resultado demostrado por Hoeffding en su artículo [Hoe63, Teorema 4] tendremos que para cada función $f : \mathbb{R} \rightarrow \mathbb{R}$ continua y convexa valdrá que

$$\mathbb{E} [f(\#(H \cap \mathcal{K}_l))] \leq \mathbb{E} [f(Y)], \quad (6.31)$$

donde $Y \sim \text{Binomial}\left(\#H, \frac{g(n)}{n_l}\right)$. En particular, como $e^{\theta x}$ es convexa para cada $\theta \in \mathbb{R}$ tenemos que podemos realizar cotas análogas a las de la desigualdad de concentración 2.4.12 también en este caso, donde inmediatamente después de que usamos la desigualdad de Markov en (A.2) y (A.4) continuamos acotando con este resultado para pasar al caso de suma de variables Bernoulli y a partir de ahí terminamos la demostración análogamente. En efecto, como $\varepsilon_n = \omega\left(\frac{1}{\log(n)}\right)$ vale que

$$\mathbb{E}[\#(H \cap \mathcal{K}_l)] = \#H \frac{g(n)}{n_l} \simeq \varepsilon_n \log(n)^2 = \omega(\log(n)), \quad (6.32)$$

y así para \tilde{C}_2 suficientemente grande tendremos que $\tilde{C}_2 \varepsilon_n \log(n)^2 = y \mathbb{E}[\#(H \cap \mathcal{K}_l)]$ para $y > 1$ y entonces

$$\begin{aligned} \mathbb{P}\left(\left\{\#(H \cap \mathcal{K}_l) > \tilde{C}_2 \varepsilon_n \log(n)^2\right\}\right) &\leq \mathbb{P}\left(\left\{\#(H \cap \mathcal{K}_l) > y \mathbb{E}[\#(H \cap \mathcal{K}_l)]\right\}\right) \quad (6.33) \\ &\leq e^{-\gamma(y) \mathbb{E}[\#(H \cap \mathcal{K}_l)]} \\ &= o\left(n^{-1-\xi}\right), \end{aligned}$$

y para $\tilde{C}_1 > 0$ suficientemente chico tenemos que $C_1 \varepsilon_n \log(n)^2 = x \mathbb{E}[Y]$ para $0 < x < 1$ y así

$$\begin{aligned} \mathbb{P}\left(\left\{\#(H \cap \mathcal{K}_l) < \tilde{C}_1 \varepsilon_n \log(n)^2\right\}\right) &\leq \mathbb{P}\left(\left\{\#(H \cap \mathcal{K}_l) < x \mathbb{E}[\#(H \cap \mathcal{K}_l)]\right\}\right) \quad (6.34) \\ &\leq e^{-\gamma(x) \mathbb{E}[\#(H \cap \mathcal{K}_l)]} \\ &= o\left(n^{-1-\xi}\right). \end{aligned}$$

Terminamos entonces la demostración aclarando que para deducir el orden de decaimiento usamos que $\gamma(x), \gamma(y)$ son constantes y entonces el orden de los exponentes es $\omega(\log(n)) \geq (1 + \xi) \log(n)$ para cada $\xi > 0$. ■

Demostraremos ahora que mirando los paseos al azar de a dos pasos por vez podremos entender la proporción de caminos que terminan en nodos de \mathcal{K} con respecto a la cantidad total de caminos. Esto nos va a dar una noción de que es raro que evitemos pasar por nodos en \mathcal{K} tras muchos pasos de un paseo al azar simétrico, que va a poder emplearse para mostrar la convergencia de los algoritmos iterativos, y también para acotar la cantidad de operaciones necesarias para calcular la aproximación via métodos de Montecarlo en el peor caso.

[6.2.7] Lema: Los paseos al azar generalmente llegan rápido a nodos conocidos.

Sea $G_{\mathcal{C}} \sim \text{SBM}(n, k, (n_l)_{1 \leq l \leq k}, p_n, q_n)$ para $p_n = a \frac{\log(n)}{n}$ y $q_n = b \frac{\log(n)}{n}$, con $0 < b < a$, $n_l = \Omega(n)$ para cada $1 \leq l \leq n$ y $\liminf_{n \rightarrow \infty} \min_{1 \leq l \leq k} \left\{ a \frac{n_l}{n} + b \left(1 - \frac{n_l}{n}\right) \right\} > 1$.

Supongamos que elegimos uniformemente al azar $g(n) = \varepsilon_n n$ nodos en cada comunidad, a los cuales denotaremos como \mathcal{K} , con $\varepsilon_n = \omega\left(\frac{1}{\log(n)}\right)$.

Entonces, existe una familia de grafos G que tiene probabilidad mayor que $1 - o(n^{-\xi})$ para un $\xi > 0$ suficientemente pequeño, y para los cuales si $(X_t)_{t \in \mathbb{N}_0}$ es un paseo al

azar simétrico en G existe $C > 0$ una constante suficientemente chica tal que para cada $M > 0$ y para cada nodo $i \in V$ y n suficientemente grandes tenemos que

$$\mathbb{P}(\{\tilde{\tau}_{\mathcal{K}} > M + 1\} | \{X_0 = i, \tilde{\tau}_{\mathcal{K}} > M, G_{\mathcal{C}}\}) \leq 1 - C\varepsilon_n \quad (6.35)$$

para $\tilde{\tau}_{\mathcal{K}} = \inf\{t \in \mathbb{N} | X_{2t} \in \mathcal{K}\}$ el tiempo de primera visita en algún paso par a \mathcal{K} .

Demostración. En primer lugar, nos restringiremos únicamente a una familia de modelos estocásticos de bloques que verifican buenas condiciones para los entornos a dos pasos. Diremos que un grafo G es admisible si cumple que para ciertas constantes positivas $C_1, C_2, C_3, C_{4,l}, C_{5,l}, C_{6,l}, C_{7,l}$ donde $1 \leq l \leq k$ valen que:

1. Las vecindades son de tamaños cercanos al esperado. Es decir, para cada $i \in V$

$$C_1 \log(n) \leq |\partial \bar{B}_1(i)| \leq C_2 \log(n). \quad (6.36)$$

2. Para cada par de nodos distintos, no hay muchos vecinos comunes. Es decir, para cada $i, j \in V$ distintos,

$$|\partial \bar{B}_1(i) \cap \partial \bar{B}_1(j)| \leq C_3. \quad (6.37)$$

3. A dos pasos, hay una cantidad de nodos y nodos de comunidad conocida cercana a la esperada. Es decir, para cada nodo $i \in V$,

$$C_{4,l} \log(n)^2 \leq |\partial \bar{B}_2(i) \cap \mathcal{C}_l| \leq C_{5,l} \log(n)^2 \quad (6.38)$$

y además

$$C_{6,l} \log(n)^2 \varepsilon_n \leq |\partial \bar{B}_2(i) \cap \mathcal{K}_l| \leq C_{7,l} \log(n)^2 \varepsilon_n, \quad (6.39)$$

para cada $1 \leq l \leq k$.

Notaremos simplemente C_4, C_6 a los mínimos sobre l , y C_5, C_7 a los máximos sobre l de las constantes asociadas.

Como los valores estarán cerca de las cantidades esperadas, con alta probabilidad los modelos estocásticos de bloques serán admisibles. En efecto, la Proposición 2.4.11 y el Lema 6.2.6 nos indican que las condiciones 1. y 3. se verifican con probabilidad mayor que $1 - o(n^{-\xi})$ para algún $\xi > 0$ suficientemente chico fijando constantes adecuadas. Como podemos concluir que esto también ocurre para la condición 2. usando desigualdades de concentración, entonces según la Observación 2.2.5 valen todas simultáneamente también con probabilidad mayor que $1 - o(n^{-\xi})$.

En efecto, la cantidad de vecinos en común de dos nodos es una suma de variables Bernoulli, y haciendo uso de esta escritura podemos concluir lo afirmado usando desigualdades

de concentración. Suponiendo conocidas las comunidades de los nodos, tendremos que para nodos distintos i y j vale que

$$X = \#(\partial\overline{B}_1(i) \cap \overline{B}_1(j)) = \sum_{l=1}^k \text{Binomial}(n_l - \mathbb{1}_{\mathcal{C}_l}(i) - \mathbb{1}_{\mathcal{C}_l}(j), (P_{\mathcal{C}})_{\mathcal{C}(i),l}(P_{\mathcal{C}})_{\mathcal{C}(j),l}) \quad (6.40)$$

para binomiales independientes. De esta manera, la cantidad esperada de vecinos en común es

$$\mathbb{E}[X] \simeq \sum_{l=1}^k n_l \frac{\log(n)}{n} \frac{\log(n)}{n} \simeq \frac{\log(n)^2}{n}. \quad (6.41)$$

Así, la desigualdad de concentración 2.4.12 nos indica que para cada $C_3 > 0$ y n suficientemente grandes vale que

$$\mathbb{P}(\{X \geq C_3\} | \mathcal{C}) = \mathbb{P}\left(\left\{X \geq \frac{C_3}{\mathbb{E}[X]} \mathbb{E}[X]\right\} \middle| \mathcal{C}\right) \leq e^{-\gamma\left(\frac{C_3}{\mathbb{E}[X]}\right)\mathbb{E}[X]} \leq o(n^{-2-\xi}), \quad (6.42)$$

para lo cual usamos que por (6.41) podemos deducir que

$$\gamma\left(\frac{C_3}{\mathbb{E}[X]}\right)\mathbb{E}[X] = C_3\left(\log\left(\frac{C_3}{\mathbb{E}[X]}\right) - 1\right) + \mathbb{E}[X] \geq \frac{C_3}{2}\log(n) \geq (2 + \xi)\log(n) \quad (6.43)$$

y de acá se deduce lo afirmado. Para terminar, mediante probabilidad total condicionando sobre las asignaciones de comunidades obtenemos que

$$\mathbb{P}(\{X \geq C_3\}) \leq o(n^{-2-\xi}) \quad (6.44)$$

y procediendo como en la Observación 2.2.5 y teniendo en cuenta que hay no más que $\binom{n}{2} \leq n^2$ parejas de nodos distintos obtenemos que simultáneamente para cada pareja la cantidad de vecinos mutuos es baja con alta probabilidad como queríamos.

Concluimos entonces que con probabilidad más alta que $1 - o(n^{-\xi})$ vale que los grafos son admisibles, para $\xi > 0$ suficientemente chico.

Resta ver entonces que en estos grafos se llega suficientemente frecuentemente a nodos conocidos en dos pasos de un paseo al azar.

En primer lugar, si conociésemos el el nodo en el que el paseo al azar se encuentra a tiempo $2M$, estudiando los caminos de dos pasos en el grafo podremos demostrar que suponiendo que es admisible para cada $j \in G$ vale que

$$\mathbb{P}(\{\tilde{\tau}_{\mathcal{K}} > M + 1\} | \{X_0 = i, X_{2M} = j, \tilde{\tau}_{\mathcal{K}} > M, G_{\mathcal{C}}\}) \leq 1 - C\varepsilon_n \quad (6.45)$$

y así mediante probabilidad total condicionando a la posición del paseo a tiempo $2M$ demostraremos lo que queríamos

$$\mathbb{P}(\{\tilde{\tau}_{\mathcal{K}} > M + 1\} | \{X_0 = i, \tilde{\tau}_{\mathcal{K}} > M, G_{\mathcal{C}}\}) \leq 1 - C\varepsilon_n. \quad (6.46)$$

Demostremos entonces (6.45). En primer lugar, notemos que como el paseo al azar $(X_t)_t$ es simétrico, las probabilidades de llegar a $V \setminus \mathcal{K}$ partiendo de algún nodo fijo j en dos pasos

están dadas, salvo por un factor constante, por un cociente de cantidades de caminos en G . En efecto, tenemos que si $\mathcal{W} = \{(j_1, j_2) \mid j \sim j_1, j_1 \sim j_2\}$ es un subconjunto de posibles trayectorias en dos pasos desde j , entonces

$$\begin{aligned} \mathbb{P}(\{(X_1, X_2) \in \mathcal{W}\} \mid \{X_0 = j\}) &= \sum_{(j_1, j_2) \in \mathcal{W}} \frac{1}{\widehat{\deg}_n(j_1)} \frac{1}{\widehat{\deg}_n(j_2)} \\ &\leq \sum_{(j_1, j_2) \in \mathcal{W}} \frac{C_2^2}{C_1^2} \frac{1}{\#\{\text{Caminatas en dos pasos desde } j\}} \\ &\leq \frac{C_2^2}{C_1^2} \frac{\#\mathcal{W}}{\#\{\text{Caminatas en dos pasos desde } j\}}. \end{aligned}$$

Desde ya, también podemos acotar por debajo usando el factor $\frac{C_2^2}{C_1^2}$.

Estudiemos entonces las caminatas de dos pasos saliendo desde j . Dentro de ellas, las que no terminan en \mathcal{K} podemos entenderlas a través de separar en estos tres casos:

1. En primer lugar, las caminatas que comienzan y vuelven a j . De estas hay como mucho $C_2 \log(n)$ porque quedan determinadas al elegir el nodo en $\partial \bar{B}_1(j)$ por el que pasa la caminata antes de volver a j .
2. En segundo lugar, las que terminan en $\partial \bar{B}_1(j)$. De estas hay como mucho $C_2 \log(n) \cdot C_3$, pues debemos primero elegir un nodo en $\partial \bar{B}_1(j)$ para el primer paso, y después algún vecino en común de este con j para el segundo paso.
3. Por último lugar, el resto de las caminatas terminan en $\partial \bar{B}_2(j)$. Para acotarlas, acotaremos en su lugar inferiormente las caminatas que terminan además en \mathcal{K} , el complemento. Notemos que de estas hay no más que $C_6 \log(n)^2 \varepsilon_n$.

Además, notamos que hay como poco $C_4 \log(n)^2$ caminatas en dos pasos comenzando en j . Combinando todo lo anterior terminamos la demostración, porque podemos acotar la probabilidad por

$$\begin{aligned} &\mathbb{P}(\{\tilde{\tau}_{\mathcal{K}} > M + 1\} \mid \{X_0 = i, X_{2M} = j, \tilde{\tau}_{\mathcal{K}} > M, G_C\}) \tag{6.47} \\ &\leq \left(\frac{C_2}{C_1}\right)^2 \frac{C_2 \log(n)}{C_6 \log(n)^2} + \left(\frac{C_2}{C_1}\right)^2 \frac{C_2 C_3 \log(n)}{C_4 \log(n)^2} + \left(1 - \left(\frac{C_1}{C_2}\right)^2 \frac{C_6 \log(n)^2 \varepsilon_n}{C_4 \log(n)^2}\right) \\ &= 1 + \frac{C_2^3(1 + C_3)}{C_1^2 C_4} \frac{1}{\log(n)} - \frac{C_1^2 C_6}{C_2^2 C_4} \varepsilon_n \\ &\leq 1 - C \varepsilon_n \end{aligned}$$

para n suficiente grandes, usando que $\varepsilon_n = \omega\left(\frac{1}{\log(n)}\right)$. ■

Desde ya, los estimadores con los que estamos tratando están basados en aproximaciones de los estimadores empíricos discutidos en el Capítulo 5. El siguiente lema indica que, si las aproximaciones son suficientemente buenas, entonces no hay diferencia alguna entre los estimadores empíricos que discutíamos previamente y los nuevos estimadores inducidos, aunque sea en lo que a consistencia se refiere:

[6.2.8] Lema: Aproximaciones suficientemente precisas inducen estimadores fuertemente consistentes

Tomemos \hat{h} y $\hat{h}^{(U)}$ las extensiones armónicas empíricas definidas en el Capítulo 5, y supongamos que \tilde{h} y $\tilde{h}^{(U)}$ son vectores que las aproximan, donde $U \in \mathcal{U}$ para \mathcal{U} una partición elegida uniformemente al azar con $\#U = \Theta\left(\frac{n}{\log(n)}\right)$ para cada $U \in \mathcal{U}$.

Luego, si valen las hipótesis de 5.3.2 o 5.4.8 y además vale con alta probabilidad que

$$\|\tilde{h} - \hat{h}\|_{\infty} = o(\varepsilon_n), \quad (6.48)$$

entonces el estimador inducido por \tilde{h} es débilmente consistente.

Además, en el caso del modelo balanceado con dos comunidades, si valen las hipótesis de 5.5.9 y también

$$\|\tilde{h}^{(U)} - \hat{h}^{(U)}\|_{\infty} = o(\varepsilon_n) \quad (6.49)$$

para cada $U \in \mathcal{U}$ con probabilidad mayor que $1 - o\left(\frac{1}{|\mathcal{U}|}\right)$, entonces el estimador por vecindades inducido por \tilde{h} es fuertemente consistente.

Demostración. Mostremos primero que la aproximación también será débilmente consistente.

Vale que en ese caso tendremos que con alta probabilidad vale que

$$\begin{aligned} \|\tilde{h} - h\|_2 &\leq \|\tilde{h} - \hat{h}\|_2 + \|\hat{h} - h\|_2 \\ &\lesssim o(\varepsilon_n) \sqrt{n} + \frac{1}{\varepsilon_n \sqrt{\log(n)}} \|h\|_2 \\ &\leq \frac{1}{\varepsilon_n \sqrt{\log(n)}} \|h\|_2 \end{aligned} \quad (6.50)$$

y entonces procediendo como en 5.3.2 tendremos que el estimador inducido por \tilde{h} es débilmente consistente.

Veamos ahora que el estimador inducido por \tilde{h} es fuertemente consistente en el segundo caso. En primer lugar, como $\|\Pi\|_{\infty}, \|\hat{Q}\|_{\infty} = 1$, simultáneamente para cada $U \in \mathcal{U}$ se tiene que

$$\begin{aligned} \|(\Pi \hat{Q} \hat{h}^{(U)} + v) - (\Pi \hat{Q} \tilde{h}^{(U)} + v)\|_{\infty} &\leq \|\Pi \hat{Q} (\hat{h}^{(U)} - \tilde{h}^{(U)})\|_{\infty} \\ &\leq \|\Pi\|_{\infty} \|\hat{Q}\|_{\infty} \|\hat{h}^{(U)} - \tilde{h}^{(U)}\|_{\infty} \\ &\leq o(\varepsilon_n) \end{aligned} \quad (6.51)$$

con alta probabilidad. Entonces, también vale que para cada $i \in \mathcal{U}$

$$\begin{aligned} &|(\Pi \hat{Q} h + v)_i - (\Pi \hat{Q} \tilde{h}^{(U)} + v)_i| \\ &\leq |(\Pi \hat{Q} h + v)_i - (\Pi \hat{Q} \hat{h}^{(U)} + v)_i| + |(\Pi \hat{Q} \hat{h}^{(U)} + v)_i - (\Pi \hat{Q} \tilde{h}^{(U)} + v)_i| \\ &\leq o(\varepsilon_n) \end{aligned} \quad (6.52)$$

donde usamos la Proposición 5.5.8, y la cota vale simultáneamente para cada $i \in U$ y para cada $U \in \mathcal{U}$. Es así que una vez más como en 5.5.9 podemos usar que el estimador ayudado por un oráculo es fuertemente consistente para concluir que \tilde{h} es fuertemente consistente y completar la demostración. ■

Ahora sí, sin más preliminares, demostremos la Proposición 6.2.1. La demostración consistirá en acotar hasta que podamos darle una interpretación probabilística a nuestra cota, para concluir que las aproximaciones son suficientemente buenas y así conseguir el resultado.

Demostración. En primer lugar, notemos que podemos acotar el error cometido con la aproximación tras N iteraciones de la siguiente forma:

$$\begin{aligned} \|\hat{h}^{(N)} - \hat{h}\|_\infty &= \|(\Pi\hat{Q})^N(\hat{h}^{(0)} - \hat{h})\|_\infty \\ &\leq \|(\Pi\hat{Q})^N\|_\infty \|\hat{h}^{(0)} - \hat{h}\|_\infty \\ &\leq \|(\Pi\hat{Q})^N\|_\infty (\|\hat{h}^{(0)}\|_\infty + 1). \end{aligned} \quad (6.53)$$

Ahora bien, vale que para cantidades de iteraciones de la forma $2t + 1$ tendremos que

$$(\Pi\hat{Q})^{2t+1} = [(\Pi\hat{Q}\Pi)^2]^t (\Pi\hat{Q}), \quad (6.54)$$

según lo cual, usando que $\|\Pi\hat{Q}\|_\infty \leq \|\Pi\|_\infty \|\hat{Q}\|_\infty \leq 1 \cdot 1 = 1$, obtenemos que:

$$\|(\Pi\hat{Q})^{2t+1}\|_\infty \leq \|(\Pi\hat{Q}\Pi)^2\|_\infty^t. \quad (6.55)$$

Es así que inmediatamente obtenemos que para cantidades de iteraciones de la forma $2t + 2$ vale una desigualdad análoga:

$$\|(\Pi\hat{Q})^{2t+2}\|_\infty \leq \|\Pi\hat{Q}\|_\infty \|(\Pi\hat{Q})^{2t+1}\|_\infty \leq \|(\Pi\hat{Q}\Pi)^2\|_\infty^t. \quad (6.56)$$

En particular, obtenemos que al combinar con (6.53) vale que

$$\|\hat{h}^{(N)} - \hat{h}\|_\infty \leq (\|\hat{h}^{(0)}\|_\infty + 1) \|(\Pi\hat{Q}\Pi)^2\|_\infty^{\lfloor \frac{N-1}{2} \rfloor}. \quad (6.57)$$

Ahora bien, tenemos una interpretación para la norma $\|(\Pi\hat{Q}\Pi)^2\|_\infty$ mediante probabilidades: vale que esta norma nos dirá la probabilidad de que un camino de longitud 2 empezado desde algún nodo fuera de \mathcal{K} nunca pase por \mathcal{K} . En efecto, tenemos que

$$\|(\Pi\hat{Q}\Pi)^2\|_\infty = \max_{1 \leq i \leq n} \left\{ \sum_{j=1}^n (\Pi\hat{Q}\Pi)_{i,j}^2 \right\}. \quad (6.58)$$

Notamos que coordenadas (i, j) tales que $i \in \mathcal{K}$ o $j \in \mathcal{K}$ causan que el coeficiente asociado se anule, y por eso podemos reducirnos a estudiar filas asociadas a nodos $i \notin \mathcal{K}$, y más aún sumar sobre columnas $j \notin \mathcal{K}$. En ese caso, tenemos que

$$\begin{aligned} &\sum_{j=1, j \notin \mathcal{K}}^n (\Pi\hat{Q}\Pi)_{i,j}^2 \\ &= \sum_{l=1, l \notin \mathcal{K}}^n \frac{\hat{A}_{i,l}}{\widehat{\deg}_n(i)} \cdot \frac{\hat{A}_{l,j}}{\widehat{\deg}_n(l)} \\ &= \mathbb{P}(\{\tau_{\mathcal{K}} > 2\} | \{X_0 = i, \tau_{\mathcal{K}} > 0, G_C\}), \end{aligned} \quad (6.59)$$

si $(X_t)_{t \geq 0}$ es un paseo al azar simétrico en el grafo y $\tau_{\mathcal{K}}$ es el tiempo de primera visita a \mathcal{K} . En efecto, la suma no es más que el desarrollo mediante probabilidad total respecto de todos los posibles caminos de longitud dos que no pasan por \mathcal{K} , dada la matriz de adyacencia \widehat{A} del modelo $G_{\mathcal{C}}$.

Es así que, podemos usar el Lema 6.2.7 para acotar la norma, con probabilidad mayor que $1 - o(n^{-\xi})$ para algún $\xi > 0$ suficientemente pequeño. En concreto, el resultado nos indica que la norma será no menor que $1 - C\varepsilon_n$ para alguna constante suficientemente grande $C > 0$, y entonces combinando este resultado con (6.57) deducimos que

$$\|\widehat{h}^{(N)} - \widehat{h}\|_{\infty} \lesssim \sqrt{1 - \varepsilon_n}^{\lfloor \frac{N-1}{2} \rfloor} = e^{\log(1 - \varepsilon_n) \lfloor \frac{N-1}{2} \rfloor}. \quad (6.60)$$

Ahora bien, usando la aproximación mediante el polinomio de Taylor para $\log(1 + x) = x + o(|x|)$ tenemos que para cantidades de iteraciones $N = \omega\left(\frac{1}{\varepsilon_n} \log\left(\frac{1}{\varepsilon_n}\right)\right)$ vale que

$$\|\widehat{h}^{(N)} - \widehat{h}\|_{\infty} \lesssim e^{\log(1 - \varepsilon_n) \lfloor \frac{N-1}{2} \rfloor} \leq e^{(-\varepsilon_n + o(\varepsilon_n)) \frac{1}{\varepsilon_n} \log\left(\frac{1}{\varepsilon_n}\right) \omega(1)} = e^{\omega(1) \log(\varepsilon_n)} = o(\varepsilon_n) \quad (6.61)$$

y mostramos lo que queríamos.

Notemos luego que cuando $\|\widehat{h}^{(N)} - \widehat{h}\|_{\infty} = o(\varepsilon_n)$ vale que por el Lema 6.2.8 el estimador inducido es débilmente consistente. Una demostración análoga fijando cada $U \in \mathcal{U}$ nos indicará que $\|(\widehat{h}^{(U)})^{(N)} - \widehat{h}^{(U)}\|_{\infty} = o(\varepsilon_n)$ y el estimador por vecindades inducido es fuertemente consistente, y entonces notando que con alta probabilidad podemos asegurar las cotas simultáneamente queda demostrada la correctitud.

En cuanto a la complejidad, sabemos que con alta probabilidad ocurre que la matriz \widehat{Q} tiene únicamente $n \log(n)$ coeficientes no nulos, como indica la Proposición 2.4.11. Es así que multiplicar por \widehat{Q} un vector tiene solamente costo de $n \log(n)$ con alta probabilidad. Como realizamos en total N iteraciones, la complejidad del algoritmo será $n \log(n)N$, y usando que $\frac{1}{\varepsilon_n} \log\left(\frac{1}{\varepsilon_n}\right) = o(\log(n) \log(\log(n)))$ al valer $\varepsilon_n = \omega\left(\frac{1}{\log(n)}\right)$ obtenemos la cota afirmada para la cantidad de operaciones. En el caso del estimador por vecindades, precisamos usar el algoritmo iterativo $O(\log(n))$ veces, una por cada $U \in \mathcal{U}$. Es así que la complejidad en el peor caso empeora solamente por un factor logarítmico. ■

Terminemos entonces dando una idea de por qué creemos que las aproximaciones via métodos de Montecarlo son consistentes y eficientes como indica la Proposición 6.2.4, dependiendo de las Conjeturas 6.2.2 y 6.2.3.

Demostración. En primer lugar, veamos que con suficientes paseos al azar tenemos garantías razonables de convergencia hacia la extensión armónica empírica.

A ese efecto, fijemos $i \in V$ un nodo y consideremos una sucesión de paseos al azar independientes $(X_t^s)_{t \in \mathbb{N}_0, s \in \mathbb{N}}$ comenzando en i . Luego, notemos para cada $1 \leq l \leq k$ se tiene que

$$\tilde{h}_{i,l}^s = \mathbb{1}_l\left(\mathcal{C}\left(X_{\tau_{\mathcal{K}}^s}^s\right)\right) \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}\left(\widehat{h}_{i,l}\right) \quad (6.62)$$

donde $\hat{h}_{i,l}$ es como en el caso del estimador general para k comunidades. Vale entonces que la comunidad estimada por el algoritmo, cuando realizamos en total una cantidad $S(n)$ de paseos, será la comunidad l que maximice

$$\tilde{h}_{i,l} = \frac{1}{S(n)} \sum_{s=1}^{S(n)} \tilde{h}_{i,l}^s, \quad (6.63)$$

es decir la comunidad a la que se haya llegado más frecuentemente. Mostraremos que cuando $S(n)$ es suficientemente grande, donde $S(n) = \log(\#(W) \log(n)) \frac{1}{\varepsilon_n^2} \log(n)$ bastaría confiando en las conjeturas enunciadas, vale que los promedios están suficientemente cerca de la esperanza

$$\mathbb{P} \left(\left\{ \left| \tilde{h}_{i,l} - \hat{h}_{i,l} \right| \leq o(\varepsilon_n) \right\} \right) \geq 1 - o \left(\frac{1}{\#(W) \log(n)} \right) \quad (6.64)$$

para cada $i \in V$ y cada $1 \leq l \leq k$. Ahora bien, fijo algún nodo $i \in W$ tenemos que con alta probabilidad van a haber no más que $\mathcal{O}(\log(n))$ nodos que son vecinos de él. Entonces valdrá que para cada uno de sus vecinos j , vamos a tener que con probabilidad mayor que $1 - o\left(\frac{1}{\#(W)}\right)$ se tiene que

$$\left| \tilde{h}_{j,l}^{\{\{i\}\}} - \hat{h}_{j,l}^{\{\{i\}\}} \right| \leq o(\varepsilon_n) \quad (6.65)$$

aplicando el resultado anterior en el grafo $G \setminus \{i\}$, con la misma estructura de comunidades. En particular, la comunidad estimada por $\tilde{h}^{\{\{i\}\}}$ para i coincidirá con la estimada mediante $\hat{h}^{\{\{i\}\}}$, análogamente a como lo demostramos en el Lema 6.2.8. De esta manera, con alta probabilidad tenemos que el estimador por vecindades inducido es fuertemente consistente, y restará solamente acotar la complejidad teórica del algoritmo.

Veamos en un poco más de detalle como se llega a la cota (6.64). Fijemos $t_n = \frac{\varepsilon_n}{\sqrt{\log(n)}}$, y usando la desigualdad de Bernstein tendremos que

$$\begin{aligned} \mathbb{P} \left(\left\{ \left| \sum_{s=1}^{S(n)} (Y_s - \hat{h}_i) \right| > S(n)t_n \right\} \right) &\leq 2 \exp \left(-\frac{1}{2} \frac{t_n^2}{S(n)\hat{h}_i(1-\hat{h}_i) + \frac{1}{3}S(n)t_n} \right) \\ &\leq 2 \exp \left(-\frac{1}{4} S(n) \frac{t_n^2}{\hat{h}_i} \right) \\ &= o \left(\frac{1}{\#(W) \log(n)} \right) \end{aligned} \quad (6.66)$$

donde usamos que para nuestras elecciones de $S(n)$ y t_n vale que $S(n) \frac{t_n^2}{\hat{h}_i} = \omega(\log(\#(W) \log(n)))$, y que además

$$S(n)\hat{h}_i(1-\hat{h}_i) + \frac{1}{3}S(n)t_n \leq 2S(n)\hat{h}_i \quad (6.67)$$

para acotar. Remarcamos en que aquí es donde, para tener un control sobre el orden de \hat{h}_i , precisamos la Conjetura 6.2.2. En este caso, como cada \hat{h}_i tiene orden constante, basta tomar $S(n) = \omega\left(\log(\#W \log(n)) \frac{1}{t_n^2}\right)$ para cada nodo para obtener el decaimiento que queremos.

Es así que las elecciones de $S(n)$ y t_n vienen dadas para garantizar a la vez el orden anterior y que, por otro lado, t_n tiene que cumplir $t_n = o(\varepsilon_n)$ para tener la consistencia de las estimaciones.

Antes de pasar al análisis de la complejidad teórica, notamos que aunque el análisis anterior está hecho en términos del estimador por extensiones armónicas general, también sirve para el estimador por extensiones armónicas para dos comunidades balanceadas. En este caso las variables a estudiar ahora serán de la forma

$$\tilde{h}_{s,i} = \mathbb{1}_1(\mathcal{C}(X_{\tau_{\mathcal{K}}^s})) - \mathbb{1}_2(\mathcal{C}(X_{\tau_{\mathcal{K}}^s})) \stackrel{\text{i.i.d.}}{\sim} \begin{cases} 1 & \text{con probabilidad } \hat{h}_{i,1}, \text{ y} \\ -1 & \text{con probabilidad } \hat{h}_{i,2} \end{cases} \quad (6.68)$$

para cada nodo $i \in V$. Como se verifica que

$$\tilde{h}_{s,i} = 2\mathbb{1}_1(\mathcal{C}(X_{\tau_{\mathcal{K}}^s})) - 1 \quad (6.69)$$

tenemos que ver que los promedios asociados al caso general, es decir a las indicadoras, están cerca de su esperanza implica también que estos promedios están cerca de su esperanza, y análogamente se deduce la consistencia del estimador por vecindades inducido.

Ahora resta solamente acotar la complejidad del algoritmo en el peor caso, con alta probabilidad. Notamos que la cantidad de operaciones realizadas estará ligada a

1. la cantidad de pasos que realiza cada paseo al azar, donde cada paso se puede realizar en tiempo constante, y
2. a la estimación final que consiste en calcular $(\Pi\widehat{Q}\Pi\hat{h}^{(U)} + v)_i$ para cada $i \in W$ y quedarnos con la comunidad que maximice las coordenadas.

Analicemos entonces qué tan rápidamente puede realizarse cada etapa del algoritmo.

En primer lugar, resulta que la estimación final puede realizarse rápidamente, con complejidad temporal $\mathcal{O}(\#(W) \log(n))$. En efecto, fijando un nodo $i \in W$, como $\widehat{Q}_{i,j} = 0$ para los nodos $j \not\sim i$ tenemos que si $i \notin \mathcal{K}$ entonces vale que

$$(\Pi\widehat{Q}\Pi\hat{h}^{(U)} + v)_i = \sum_{j \sim i, j \notin \mathcal{K}} \widehat{Q}_{i,j} \hat{h}_j^{(U)}; \quad (6.70)$$

como i no tendrá más que $\mathcal{O}(\log(n))$ vecinos con alta probabilidad (debido a la Proposición 2.4.11), podemos realizar esta suma, y luego tomar máximo respecto a las coordenadas, en $\mathcal{O}(\log(n))$ operaciones. En consecuencia, la complejidad estará dada por la cantidad de pasos de paseos al azar que necesitamos realizar, que será en general mayor.

A continuación acotamos entonces la complejidad asociada a realizar los paseos al azar. Para comenzar, notamos que para la estimación final solamente necesitamos conocer $\hat{h}^{(U)}$ sobre los vecinos $j \sim i$, que una vez más serán $\mathcal{O}(\log(n))$ con alta probabilidad. En consecuencia, realizaremos $S(n)$ paseos al azar para aproximar extensiones armónicas empíricas para estos $\mathcal{O}(\#(W) \log(n))$ nodos, lo que nos dará un total de $\mathcal{O}(\#(W) \log(n) S(n))$ paseos realizados con alta probabilidad. Ahora bien, la Conjetura 6.2.3 nos dice que entre todos

estos paseos al azar no realizan más que $\mathcal{O}\left(\#(W) \log(n) S(n) \frac{\log(n)}{\varepsilon_n}\right)$ pasos en total con alta probabilidad, y en consecuencia la complejidad de esta parte del algoritmo está acotada con alta probabilidad por

$$\mathcal{O}\left(\#(W) \log(n) S(n) \frac{\log(n)}{\varepsilon_n}\right). \quad (6.71)$$

Usando entonces que $S(n) = \log(\#(W) \log(n)) \frac{1}{\varepsilon_n} \log(n)$, tendremos que la complejidad en peor caso del algoritmo va a ser del orden de

$$\mathcal{O}\left(\#(W) \log(n)^3 \log(\#(W) \log(n)) \frac{1}{\varepsilon_n^3}\right) \quad (6.72)$$

como queremos afirmar. ■

Pasemos luego a presentar algunos resultados de simulaciones que realizamos con estos algoritmos.

6.3 Implementación de los estimadores y simulaciones

Mencionaremos rápidamente algunos detalles relevantes de la implementación. El lenguaje elegido para la implementación de los algoritmos, la generación de datos y el análisis posterior fue R. Los grafos que utilizamos para evaluar el desempeño de los algoritmos fueron generados mediante los paquetes usuales de R para obtener números aleatorios y son en su totalidad datos sintéticos. En la implementación de los algoritmos pudimos acceder a los métodos de Buehler y Hein a través de una interfaz en R que nos permite acceder a los métodos que están programados en `Octave`.

Ahora sí, pasemos entonces al análisis de los algoritmos para realizar detección de comunidades. Presentaremos aquí solamente unas exploraciones preliminares, limitándonos a verificar que los métodos verifican tener una precisión alta.

En primer lugar, podemos observar que los estimadores por extensiones armónicas efectivamente tienen un buen desempeño en las simulaciones, aún sin recurrir a los estimadores por vecindades. En la Figura 6.1 mostramos el desempeño del método de extensiones armónicas para parámetros $a = 6$, $b = 2$ y n variando, donde los colores se usan para distintos valores de ε_n para cada n . Las cajas representan la precisión del estimador para los datos en los cuantiles más cercanos a la media, que es la raya en el medio de cada caja, y las líneas nos indican donde están los otros cuantiles - en total se realizaron 20 simulaciones para cada valor de ε_n , donde utilizamos la implementación mediante algoritmos iterativos con $250 \log(n)$ iteraciones máximas por modelo. Los puntos que aparecen aislados son simulaciones que obtuvieron resultados anómalos. Observamos que cuando n crece tenemos que la precisión aumenta, y dejamos de tener malos resultados excepcionales. La elección de parámetros $a = 6$ y $b = 2$ es una muy ajustada en cuanto a que $\frac{a+b}{2} > 1 + \sqrt{ab}$ sin demasiado margen, y así con estos datos estamos cerca del límite en el que puede realizarse detección de comunidades con consistencia fuerte. En cambio, en la Figura 6.2 presentamos

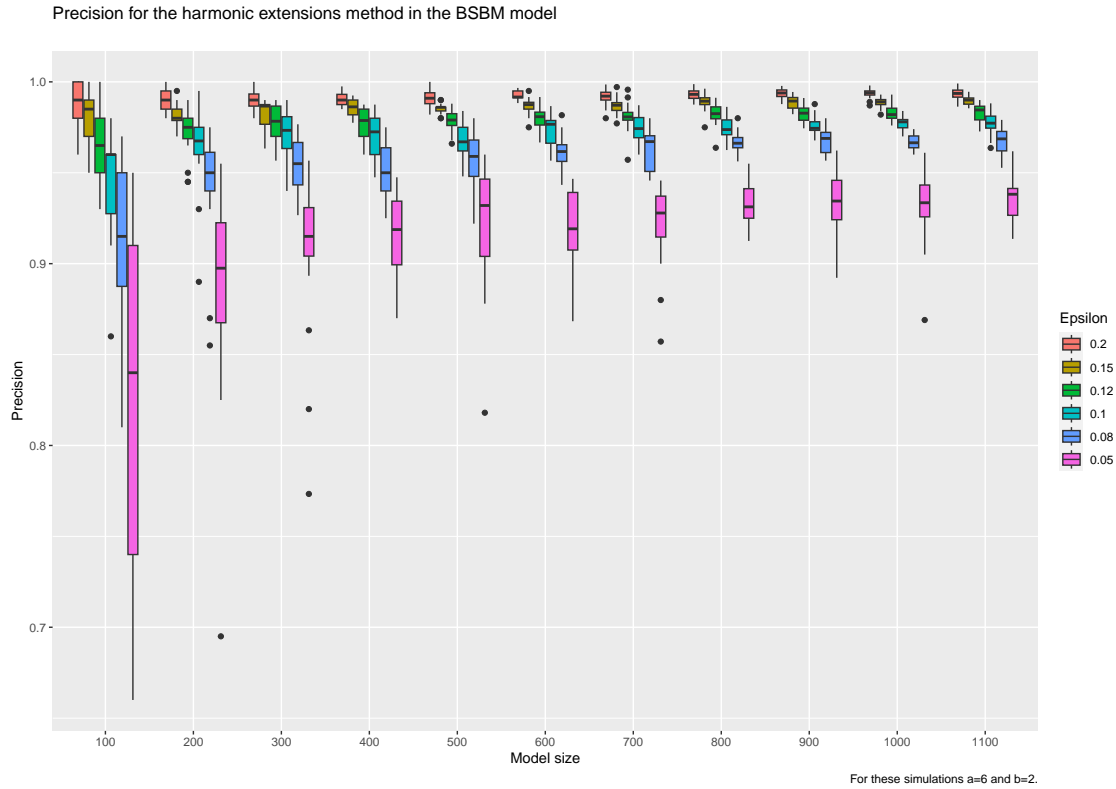


Figura 6.1: Simulación para el método de extensiones armónicas cerca del umbral teórico para tener consistencia fuerte.

los resultados del desempeño del método de extensiones armónicas ahora para parámetros $a = 6$, $b = 1$ y n variando, de la misma manera que las anteriores. Observamos que para estos datos se tiene un margen más holgado respecto del umbral mínimo para que pueda realizarse detección de comunidades fuertemente consistente, y podemos apreciar una mejoría en cuanto a la precisión.

Concluimos que el método parece efectivamente parece mejorar la precisión con n , aunque queda para trabajo futuro considerar tamaños de modelos más grandes porque la mejoría parece darse gradualmente.

En cuanto a los métodos de minimización de corte y métodos espectrales, rápidamente alcanzan una buena precisión. En las Figuras 6.3 y 6.4, mostramos nuevamente unos gráficos de cajas en los que se muestra para distintas combinaciones de a y b (en formato $a; b$) y modelos de tamaño $n = 50$ la precisión del estimador de corte mínimo de Cheeger en naranja y del estimador espectral en azul, realizando 20 simulaciones para cada uno. La única diferencia radica en que para la Figura 6.3 quitamos algunos datos anómalos (en total un resultado por combinación de a y b) de 6.4 para facilitar la legibilidad del gráfico. Confiando en que no son representativos, de lo que pasa en general, como vemos en términos de los cuantiles, vemos que incluso para modelos de tamaño chico la precisión de ambos métodos es excelente, y comparable. Por otro lado, observamos que al sortear las aristas con probabilidades $a \frac{\log(n)}{n}$ y $b \frac{\log(n)}{n}$ tenemos que contar con probabilidades altas ayuda, aún cuando a y b resulten solamente ligeramente por arriba del umbral en el cual se puede realizar

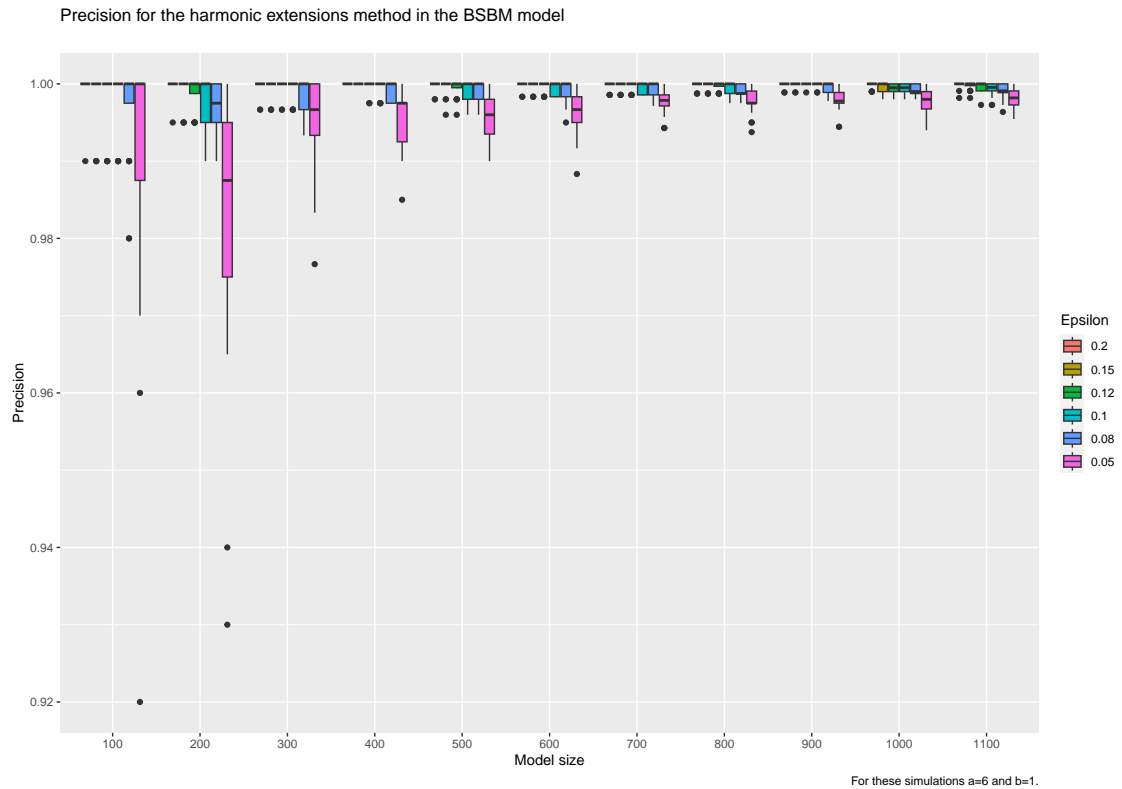


Figura 6.2: Simulación para el método de extensiones armónicas con margen con respecto al umbral teórico para tener consistencia fuerte.

detección de comunidades de manera fuertemente consistente.

Terminamos entonces mencionando algunas direcciones en que puede seguirse la experimentación. En cuanto a los estimadores por extensiones armónicas

- podemos verificar que las cantidades de iteraciones necesarias para aproximar la extensión armónica que obtenemos mediante nuestra garantía teórica sean ajustadas,
- podemos realizar la verificación análoga para la cantidad de pasos necesaria en los paseos al azar,
- y podemos verificar si efectivamente en la práctica las implementaciones mediante métodos de Montecarlo resultan más rápidas, más allá de que su complejidad teórica en el peor caso lo sugiere.

En cuanto a los estimadores de minimización de corte, resta evaluar el desempeño de los estimadores que buscan minimizar el corte ratio, y considerar modelos con cantidades más grandes de nodos. Además de comparaciones más detalladas entre las precisiones y velocidad de los distintos algoritmos, resta para considerar modelos no balanceados, y modelos con más que dos comunidades. Muchas de estas exploraciones ya están en condiciones de ser realizadas con el código provisto anteriormente.

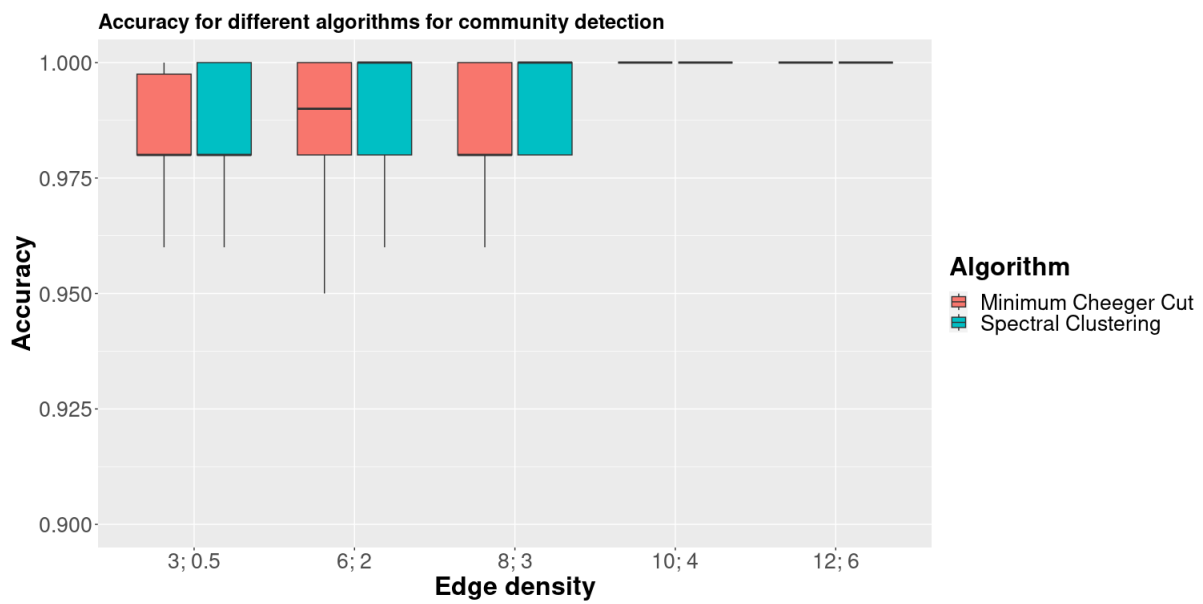


Figura 6.3: Comparación de la precisión alcanzada por el estimador minimizante del corte de Cheeger y los estimadores espectrales.

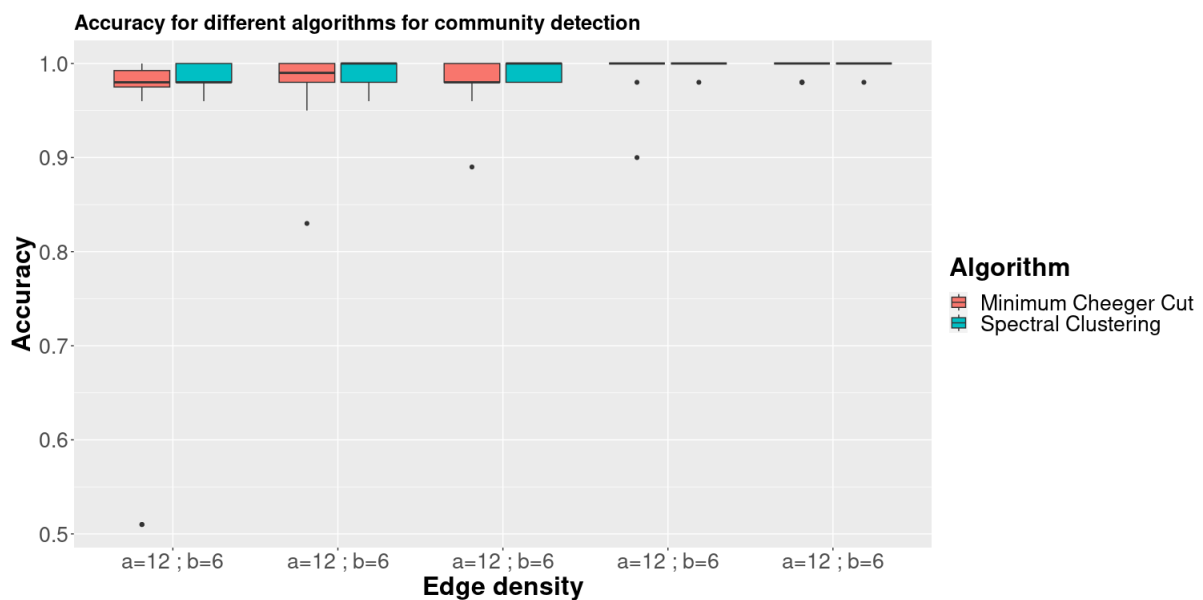


Figura 6.4: Comparación de la precisión alcanzada por el estimador minimizante del corte de Cheeger y los estimadores espectrales, mostrando los resultados anómalos.

7. Conclusiones y trabajo futuro

Llegamos entonces al final del trabajo, habiendo presentado los modelos estocásticos de bloques y algunos algoritmos para estimar las comunidades en ellos. Logramos mostrar mediante métodos probabilísticos que los estimadores de corte mínimo son consistentes, y lo verificamos mediante simulaciones. También hicimos lo propio para los estimadores por extensiones armónicas: mostramos que son consistentes y verificamos que estiman precisamente con algunas simulaciones.

Concluimos entonces el trabajo de tesis, pero nos quedan más cosas por explorar.

Dentro de los estimadores por minimización de corte, hay otras variantes de cortes balanceados que podríamos explorar, como mencionamos en la Nota 4.4.2. Además, queda pendiente extender el resultado de consistencia fuerte en cortes de Cheeger al modelo general, sorteando las dificultades técnicas que mencionamos en la Nota 4.4.3.

Dentro de los estimadores por extensiones armónicas, nos queda pendiente determinar si el estimador por extensiones armónicas base es fuertemente consistente, más allá de que el de vecindades lo sea. Además, queda pendiente escribir cómo se extienden los resultados a k comunidades desbalanceadas. Por último, queda verificar que en efecto hay una garantía teórica de que las implementaciones por Montecarlo son correctas y eficientes demostrando las conjeturas pendientes.

Finalmente, comentamos algunas nuevas direcciones en las cuales podríamos continuar.

Dentro de los métodos semi-supervisados, durante el trabajo de investigación resultó aparente que podrían generalizarse los algoritmos espectrales usando la descomposición en valores singulares de la matriz ΠQ para obtener estimadores en el caso semi-supervisado. Por otro lado, como en la Nota 5.6.2 queda pendiente explorar qué puede lograrse mediante extensiones p -armónicas.

Por último, mencionamos que al realizar extensiones armónicas para potencias de la matriz de adyacencia obtuvimos mejores resultados en nuestras simulaciones que para la matriz de adyacencia usual, incluso suponiendo conocidas proporciones de nodos mucho menores. Una intuición para entender este fenómeno está en que potencias de la matriz de adyacencia codifican la cantidad de caminos de cada longitud fija entre cada par de nodos, lo que en cierta medida le da más robustez. En este contexto resultó aparente que el tiempo que tarda un paseo al azar en llegar desde cada nodo a cierto nodo fijo i resultaba en general más chico dentro de la comunidad $\mathcal{C}(i)$, lo cual se tradujo en que estimadores basados en estas cantidades obtuviesen buenos resultados en simulaciones en el modelo BSBM.

Bibliografía

- [Abb18] Emmanuel Abbe, *Community detection and stochastic block models: Recent developments*, Journal of Machine Learning Research **18** (2018), no. 177, 1–86. ↑7, 8, 20, 23, 27, 33, 36, 39, 50, 51, 142
- [AFWZ20] Emmanuel Abbe, Jianqing Fan, Kaizheng Wang, and Yiqiao Zhong, *Entrywise eigenvector analysis of random matrices with low expected rank*, Ann. Statist. **48** (2020), no. 3, 1452–1474. También disponible en <https://arxiv.org/abs/1709.09565>. MR4124330 ↑7, 41, 51, 142
- [BFMM23] Paola Bermolen, Marcelo Fiori, Bernardo Marengo, and Gonzalo Mateos, *Gradient-based spectral embeddings of random dot product graphs*, 2023. Disponible en <https://arxiv.org/abs/2307.13818>. ↑52
- [CL23] Andressa Cerqueira and Florencia Leonardi, *Optimal recovery by maximum and integrated conditional likelihood in the general stochastic block model*, 2023. Disponible en <https://arxiv.org/abs/1709.09565>. ↑7, 50, 51
- [Dur07] Rick Durrett, *Random graph dynamics*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 20, Cambridge University Press, Cambridge, 2007. MR2271734 ↑14, 15, 20, 141
- [Hoe63] Wassily Hoeffding, *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), 13–30. MR144363 ↑125
- [HS11] Matthias Hein and Simon Setzer, *Beyond spectral clustering - tight relaxations of balanced graph cuts*, Advances in neural information processing systems, 2011. ↑7, 55, 143
- [LR15] Jing Lei and Alessandro Rinaldo, *Consistency of spectral clustering in stochastic block models*, Ann. Statist. **43** (2015), no. 1, 215–237. MR3285605 ↑37, 142
- [MNS16] Elchanan Mossel, Joe Neeman, and Allan Sly, *Consistency thresholds for the planted bisection model*, Electron. J. Probab. **21** (2016), Paper No. 21, 24. MR3485363 ↑7, 22, 23, 27, 51, 112, 142
- [Nor98] J. R. Norris, *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge, 1998. Reprint of 1997 original. MR1600720 ↑79, 80, 143
- [vdH17] Remco van der Hofstad, *Random graphs and complex networks. Vol. 1*, Cambridge Series in Statistical and Probabilistic Mathematics, [43], Cambridge University Press, Cambridge, 2017. MR3617364 ↑6, 23
- [Ver18] Roman Vershynin, *High-dimensional probability*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 47, Cambridge University Press, Cambridge, 2018. An introduction with applications in data science, With a foreword by Sara van de Geer. MR3837109 ↑13, 120, 141
- [vL07] Ulrike von Luxburg, *A tutorial on spectral clustering*, Stat. Comput. **17** (2007), no. 4, 395–416. También disponible en <https://arxiv.org/abs/0711.0189>. MR2409803 ↑7, 73

Índice de definiciones, observaciones y resultados

Capítulo 2

Definición: Grafos	9
Definición: Caminos, ciclos, conexión y distancia en grafos	9
Notación: Conjunto de nodos	10
Definición: Matriz de adyacencia de un grafo	11
Observación: Grafos pesados	11
Notación: Repaso y notación de probabilidades	11
Notación: Notación de comportamiento asintótico	12
Proposición: Desigualdad de Paley-Zygmund	12
Proposición: Desigualdad de Bernstein - [Ver18, Teorema 2.8.4.]	13
Observación: Contables eventos que ocurren con alta probabilidad, ocurren simultáneamente con alta probabilidad	13
Notación: Notación para vectores y matrices	13
Definición: Grafo de Erdős-Renyi	14
Teorema: Existencia de una única componente conexa lineal - [Dur07, Teorema 2.3.2.]	14
Teorema: Transición de fase para la conexión en grafos de Erdős-Renyi - [Dur07, Teoremas 2.8.1., 2.8.2. y 2.8.3.]	15
Definición: Modelo estocástico de bloques SBM	15
Notación: Algunas notaciones para las comunidades	16
Observación: Probabilidad puntual para modelos SBM	17
Observación: Las comunidades son de tamaño lineal con alta probabilidad	17
Definición: Modelo estocástico de bloques simétrico SSBM	18
Definición: Modelo estocástico de bloques con tamaño de comunidades fijas y el modelo balanceado	18
Observación: Condicionar para fijar las comunidades de cada nodo	19
Teorema: Transición de fase para la conexión en modelos BSBM	19
Teorema: Transición de fase para la conexión en modelos SBM con tamaños de comunidades fijos	20
Teorema: Transición de fase para la conexión en modelos SBM	20
Proposición: Los grados se concentran en el grado esperado	21

Lema: Desigualdad de concentración para la conectividad de modelos SBM	21
Teorema: Cada nodo tiene una mayoría de vecinos de su misma comunidad - [MNS16, Proposiciones 2.7 y 2.8]	22
Nota: Modelos estocásticos de bloques más generales en la literatura.	23
Nota: Simulaciones para modelos estocásticos de bloques	23

Capítulo 3

Problema: El problema de detección de comunidades	24
Definición: Consistencia fuerte y consistencia débil	25
Proposición: Consistencia es imposible sin conexión.	25
Teorema: Factibilidad de consistencia fuerte en modelos BSBM - [Abb18, Teorema 5.] o [MNS16, Teorema 2.5.]	27
Definición: Estimadores máximos a posteriori	28
Observación: El estimador MAP determina la existencia de estimadores fuertemente consistentes	28
Definición: Estimadores de máxima verosimilitud y de corte mínimo	29
Observación: En el modelo BSBM el estimador MAP coincide con el estimador de máxima verosimilitud, y en el caso asortativo con el estimador de mínima bisección.	30
Observación: Descomposición espectral de la matriz de adyacencia teórica en un modelo BSBM y motivación de los métodos espectrales	34
Teorema: Teorema de Davis-Kahan - [Abb18, Teorema 9.]	36
Teorema: Teorema de Weyl	36
Teorema: La matriz de adyacencia empírica está cerca de la matriz de adyacencia teórica - [LR15, Teorema 5.2.]	37
Definición: Estimadores espectrales $\hat{\mathcal{C}}_{\text{SP}}$	37
Lema: Los segundos autovectores están cerca	38
Teorema: Consistencia débil para métodos espectrales	38
Definición: Estimador espectral ayudado por un oráculo $\hat{\mathcal{C}}_{\text{SPA}}$	40
Definición: Estimador espectral por vecindades $\hat{\mathcal{C}}_{\text{SPV}}$	40
Lema: Las mayorías no son chicas con probabilidad alta - [AFWZ20, Lema 8.]	41
Lema: Los estimadores por vecindades se portan parecido a los estimadores ayuda- dos por un oráculo - [AFWZ20, versión simplificada del Lema 7.]	41
Lema: Preliminares para la Proposición (3.5.7)	41
Proposición: El estimador espectral ayudado por un oráculo $\hat{\mathcal{C}}_{\text{SPA}}$ es fuertemente consistente.	44
Proposición: Los vectores \hat{u}_2 , $\left(\left(\frac{1}{\lambda_2}\hat{A}\hat{u}_2^{(i)}\right)_i\right)$ y $\frac{1}{\lambda_2}\hat{A}u_2$ están cerca coeficiente a co- eficiente.	45
Teorema: El estimador espectral $\hat{\mathcal{C}}_{\text{SP}}$ y el estimador espectral por vecindades son fuertemente consistentes.	50
Nota: Otras nociones de consistencia y límites para el caso general	50

Nota: Extensión a k comunidades para algoritmos espectrales	51
Nota: Estimadores fuertemente consistentes en la literatura	51
Nota: Detección de comunidades en modelos más generales	51

Capítulo 4

Observación: El estimador de corte mínimo \widehat{C}_{CUT} es fuertemente consistente en el modelo BSBM asortativo	53
Observación: Limitación del corte de mínima bisección	54
Definición: Corte de Cheeger y corte ratio - [HS11, Sección 3.]	55
Observación: Los estimadores por minimización de corte de Cheeger y corte ratio son equivalentes al de mínima bisección en el modelo BSBM sobre comunidades balanceadas	56
Notación: Tamaño y solapamientos para las comunidades estimadas	56
Teorema: Consistencia fuerte de una realización de corte de Cheeger mínima	57
Lema: Cálculos de esperanza y varianza	57
Lema: Cota para los coeficientes binomiales	61
Lema: Cota para cociente de polinomios	61
Lema: Cota para simplificar cocientes de polinomios	62
Teorema: Consistencia fuerte de una realización de corte ratio mínima	71
Teorema: Consistencia fuerte del estimador de corte ratio en el caso general	72
Nota: Los estimadores de corte ratio y corte de Cheeger en la literatura	73
Nota: Otras variantes, y generalizaciones a más que 2 comunidades	73
Nota: Dificultades al extender los resultados para el corte de Cheeger al caso general	74

Capítulo 5

Observación: Juegos de “tira y afloja”	75
Observación: Modelo del votante	76
Definición: Paseos al azar	77
Definición: Cadenas de Markov	78
Definición: Tiempos de parada y tiempos de primera visita	79
Observación: Los tiempos de primera visita son tiempos de parada	79
Teorema: Propiedad fuerte de Markov - [Nor98, Teorema 1.4.2.]	79
Definición: Estados recurrentes, transitorios y cadenas de Markov irreducibles	80
Teorema: Los estados son o bien recurrentes o bien transitorios - [Nor98, Teorema 1.5.3.]	80
Teorema: Propiedades de cadenas de Markov irreducibles en espacios de estados finitos - [Nor98, Teoremas 1.5.6. y 1.5.7.]	80
Proposición: Los tiempos de primera visita son casi seguramente finitos para paseos al azar irreducibles en grafos finitos	81
Definición: Funciones armónicas en un grafo	81
Proposición: Extensiones armónicas	82

Definición: Estimador por extensiones armónicas para dos comunidades balanceadas	85
Observación: El estimador por extensiones armónicas para dos comunidades balanceadas como un “tira y afloja”	85
Notación: Extensiones teóricas y empíricas, y nodos conocidos	86
Observación: Extensiones armónicas en notación matricial	87
Lema: Extensión armónica teórica para dos comunidades balanceadas	88
Proposición: La solución empírica se parece a la solución teórica	89
Teorema: Consistencia débil del estimador por extensiones armónicas para dos comunidades balanceadas \hat{C}_{HE2} .	89
Observación: Generalización a k comunidades	94
Definición: Estimador por extensiones armónicas en el caso general	95
Observación: Equivalencia entre las dos nociones de cercanía para el estimador por extensiones armónicas en el caso general	95
Observación: Relación entre el estimador por extensiones armónicas para dos comunidades balanceadas y el estimador en el caso general	96
Observación: Notación matricial en k comunidades	96
Lema: Extensión armónica teórica en el caso general	97
Proposición: La solución empírica se parece a la solución teórica: versión para k comunidades.	98
Teorema: Consistencia débil del estimador por extensiones armónicas \hat{C}_{HE} .	98
Teorema: Consistencia débil del estimador por extensiones armónicas para el modelo SBM general	99
Definición: Estimador por extensiones armónicas ayudado por un oráculo para dos comunidades balanceadas	101
Notación: Notación para los estimadores por vecindades	101
Definición: Estimador por extensiones armónicas por vecindades para dos comunidades balanceadas	102
Observación: Solución teórica del estimador por vecindades en el modelo BSBM	103
Lema: Cotas útiles para los estimadores por extensiones armónicas.	103
Lema: Las mayorías no son chicas con probabilidad alta	107
Proposición: El estimador por extensiones armónicas con ayuda de un oráculo \hat{C}_{HEA2} es fuertemente consistente.	107
Proposición: El estimador ayudado por un oráculo y el estimador por vecindades están cerca coeficiente a coeficiente.	109
Teorema: Consistencia fuerte del estimador por vecindades \hat{C}_{HEV} .	111
Nota: Otras maneras de lograr consistencia fuerte a partir de extensiones armónicas	112
Nota: Otros laplacianos	112
Nota: Límites de la predicción en contextos semi-supervisados	113
Teorema: Factibilidad de consistencia fuerte para estimadores semi-supervisados en el modelo BSBM	113

Nota: Modelo del votante	113
------------------------------------	-----

Capítulo 6

Observación: Costos de operar con vectores y matrices	114
Observación: Costos de operar con matrices y vectores en regímenes esparsos . . .	114
Proposición: Consistencia y complejidad de la aproximación de $\hat{\mathcal{C}}_{\text{HEV}}$ mediante métodos iterativos.	118
Conjetura: Orden de los coeficientes de la extensión armónica empírica	119
Conjetura: Cota para la cantidad de pasos que toma un paseo simétrico en llegar a \mathcal{K}	120
Proposición: Correctitud y complejidad de la aproximación de $\hat{\mathcal{C}}_{\text{HEV}}$ via métodos de Montecarlo.	120
Notación: Entornos en un grafo.	121
Lema: Los nodos a distancia dos en el modelo SBM	121
Lema: Los paseos al azar generalmente llegan rápido a nodos conocidos.	126
Lema: Aproximaciones suficientemente precisas inducen estimadores fuertemente consistentes	130

Capítulo A

Definición: Martingalas, submartingalas y supermartingalas	148
Teorema: Teorema de parada opcional de Doob	148
Lema: Cálculo de esperanza y varianza para la diferencia de corte ratio.	158

Índice de figuras

2.1. Algunos ejemplos de grafos.	10
2.2. Construcción de un modelo SBM.	16
3.1. El problema de detección de comunidades en el modelo SBM.	24
3.2. Visualización de un corte entre comunidades en el modelo BSBM.	30
3.3. El método espectral para 3 comunidades balanceadas.	51
4.1. El estimador por mínima bisección debe ser restringido para no caer en malos casos.	54
4.2. Un diagrama para visualizar la diferencia entre los cortes asociados a \mathcal{C} y $\tilde{\mathcal{C}}$	58
4.3. Un diagrama para entender los casos relevantes a considerar para la consistencia del corte de Cheeger.	65
5.1. El estimador por extensiones armónicas como un juego de “tira y afloja”.	76
6.1. Simulación para el método de extensiones armónicas cerca del umbral teórico para tener consistencia fuerte.	136
6.2. Simulación para el método de extensiones armónicas con margen con respecto al umbral teórico para tener consistencia fuerte.	137
6.3. Comparación de la precisión alcanzada por el estimador minimizante del corte de Cheeger y los estimadores espectrales.	138
6.4. Comparación de la precisión alcanzada por el estimador minimizante del corte de Cheeger y los estimadores espectrales, mostrando los resultados anómalos.	138
A.1. Un diagrama para visualizar los casos en la demostración de consistencia fuerte para el corte ratio.	162

1. Anexo

A.1 Demostraciones del capítulo 2

En esta sección del anexo buscaremos demostrar los Teoremas 2.4.8, 2.4.9, y 2.4.10 que indican transiciones de fase para cuándo un modelo de bloques estocástico es conexo.

Para comenzar, demostraremos en primer lugar aquí la desigualdad en la Proposición 2.4.12, que será clave tanto para demostrar la transición de fase para la conexión como para controlar los grados de los nodos en la Proposición 2.4.11, que una vez más nos será de ayuda para lograr nuestro objetivo.

Demostración del Lema 2.4.12. En primer lugar, podemos mostrar que para cada $\theta \in \mathbb{R}$ fijo ocurre que

$$\mathbb{E} \left[e^{\theta Z} \right] \leq e^{\mathbb{E}[Z](e^\theta - 1)} \quad (\text{A.1})$$

mediante una cota directa, usando primero la independencia de las variables Bernoulli involucradas y después que $1 + x \leq e^x$ para cada $x \in \mathbb{R}$:

$$\begin{aligned} \mathbb{E} \left[e^{\theta Z} \right] &= \prod_{s=1}^t \prod_{l=1}^k \left[(1 - p_{s,l}) + p_{s,l} e^\theta \right]^{\lceil (1-\delta)n_l \rceil} \quad (\text{A.2}) \\ &\leq \prod_{s=1}^t \prod_{l=1}^k e^{\lceil (1-\delta)n_l \rceil p_{s,l} (e^\theta - 1)} \\ &= e^{\sum_{s=1}^t \left[\sum_{l=1}^k \lceil (1-\delta)n_l \rceil p_{s,l} \right] (e^\theta - 1)} \\ &= e^{\mathbb{E}[Z](e^\theta - 1)}. \quad (\text{A.3}) \end{aligned}$$

Luego, podemos concluir la desigualdad que queríamos usando la desigualdad de Markov y minimizando sobre $\theta \in \mathbb{R}$.

En efecto, cuando $\theta < 0$, la cota mostrada previamente nos da que para cada $x < 1$ vale que

$$\mathbb{P} \left(\{ Z \leq x \mathbb{E}[Z] \} \right) = \mathbb{P} \left(\{ e^{\theta Z} \geq e^{\theta x \mathbb{E}[Z]} \} \right) \leq \frac{\mathbb{E} \left[e^{\theta Z} \right]}{e^{\theta x \mathbb{E}[Z]}} \leq e^{(-\theta x + e^\theta - 1) \mathbb{E}[Z]} \quad (\text{A.4})$$

donde podemos mostrar que en $\theta = \log(x)$ la cota alcanza su mínimo derivando $-\theta x + e^\theta - 1$ para obtener la desigualdad enunciada.

Análogamente vale que para cada $1 < y$ tenemos que para cada $\theta > 0$ vale que

$$\mathbb{P}(\{Z \geq y\mathbb{E}[Z]\}) = \mathbb{P}(\{e^{\theta Z} \geq e^{\theta y\mathbb{E}[Z]}\}) \leq \frac{\mathbb{E}[e^{\theta Z}]}{e^{\theta y\mathbb{E}[Z]}} \leq e^{(-\theta y + e^\theta - 1)\mathbb{E}[Z]}, \quad (\text{A.5})$$

y nuevamente derivando mostramos que la cota alcanza su mínimo en $\theta = \log(y)$ como queríamos ver. ■

Notemos en primer lugar que el Teorema 2.4.8 no es más que un corolario del Teorema 2.4.9, según lo cual basta demostrar únicamente este último.

Antes de explicar la demostración en detalle, introduciremos unas familias de procesos estocásticos muy útiles en teoría de probabilidades, las martingalas, que ayudarán a entender cómo puede evolucionar una exploración del grafo de a un nodo por vez.

[A.1.1] Definición: Martingalas, submartingalas y supermartingalas

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, y consideremos $(M_t)_{t \in \mathbb{N}_0}$ y $(X_t)_{t \in \mathbb{N}_0}$ procesos estocásticos. Decimos que $(M_t)_t$ es una **supermartingala** con respecto a $(X_t)_t$ cuando valen que

1. $\mathbb{E}[|M_t|] < \infty$ para cada $t \in \mathbb{N}_0$,
2. $\mathbb{E}[M_{t+1} | X_1, \dots, X_t] \leq M_t$ para cada $t \in \mathbb{N}_0$, y
3. M_t es función de X_1, \dots, X_t para cada $t \in \mathbb{N}_0$.

Análogamente, diremos que (M_t) es **submartingala** con respecto a $(X_t)_t$ cuando valen que

1. $\mathbb{E}[|M_t|] < \infty$ para cada $t \in \mathbb{N}_0$,
2. $\mathbb{E}[M_{t+1} | X_1, \dots, X_t] \geq M_t$ para cada $t \in \mathbb{N}_0$, y
3. M_t es función de X_1, \dots, X_t para cada $t \in \mathbb{N}_0$.

Por último, diremos que $(M_t)_t$ es una **martingala** con respecto a $(X_t)_t$ cuando es a la vez una supermartingala y una submartingala.

Entre las muchas propiedades con las que cuentan las martingalas (y supermartingalas, y submartingalas) una muy útil es que para tiempos de parada suficientemente razonables podremos tener una noción de qué está ocurriendo con el proceso en esperanza.

[A.1.2] Teorema: Teorema de parada opcional de Doob

Sea $(\Omega, \mathcal{F}, \mathbb{P})$ un espacio de probabilidad, y consideremos $(M_t)_{t \in \mathbb{N}_0}$ una martingala, supermartingala o submartingala con respecto a $(X_t)_{t \in \mathbb{N}_0}$.

Luego, si τ es un tiempo de parada que verifica que $\tau \leq C < \infty$ casi seguramente para alguna constante $C > 0$ tenemos que

- $\mathbb{E}[M_\tau] \leq \mathbb{E}[M_0]$ si $(M_t)_t$ es una supermartingala con respecto a $(X_t)_t$,

■ $\mathbb{E} [M_\tau] \geq \mathbb{E} [M_0]$ si $(M_t)_t$ es una submartingala con respecto a $(X_t)_t$,
y en particular $\mathbb{E} [M_\tau] = \mathbb{E} [M_0]$ cuando $(M_t)_t$ es una martingala con respecto a $(X_t)_t$.

Ahora sí, sin más preámbulos demostremos el Teorema 2.4.9:

Demostración de conexión con alta probabilidad en el Teorema 2.4.9. La idea de la demostración será mostrar que las componentes conexas asociadas a cada nodo tienen un tamaño considerable con probabilidad alta. En este caso, para que dos nodos no se conecten también debe ocurrir que ningún nodo en su componente conexa se conecte a nodos de la otra componente conexa, y entonces para tener una desconexión va a hacer falta que en el grafo no hayan muchas aristas. Como consecuencia, la probabilidad de que dos nodos cualesquiera no se conecten va a ser baja y entonces como contrapartida la probabilidad de que el grafo sea conexo va a ser alta como queríamos.

Daremos luego una manera de explorar la componente conexa de cada nodo, de a un nodo por vez, y mostraremos que explorando de esta manera podremos ver que estas son grandes.

Así, llamaremos $t \in \mathbb{N}_0$ a los tiempos o pasos en la exploración, R_t a los nodos ya explorados, A_t a los nodos a explorar a continuación y U_t a los nodos no encontrados aún. Empezamos por explorar partiendo de $A_0 = A_{0,i}$ para cada $i \in V$, que en las n distintas exploraciones va a ser un subconjunto de la componente conexa asociada a ese nodo. Esto determina las condiciones iniciales, que estarán dadas por:

$$\begin{aligned} R_0 &= \emptyset, \\ A_0 &\subseteq V, \\ U_0 &= V \setminus A_0. \end{aligned} \tag{A.6}$$

A continuación, podemos definir inductivamente R_t , A_t y U_t o bien manteniéndolos constantes a tiempo $t + 1$ cuando $A_t = \emptyset$, en cuyo caso ya habremos agotado los nodos de la componente conexa, o si no cuando $A_t \neq \emptyset$ como:

$$\begin{aligned} R_{t+1} &= R_t \cup \{i_t\}, \\ A_{t+1} &= (A_t \setminus \{i_t\}) \cup \{y \in U_t \mid i_t \sim y\}, \\ U_{t+1} &= U_t \setminus \{y \in U_t \mid i_t \sim y\}, \end{aligned} \tag{A.7}$$

donde notamos i_t a algún nodo que eligiremos de manera medible a partir de A_t que nos indica cómo continuar la exploración. Remarcamos que como comenzamos con A_0 nodos en una misma componente conexa, este procedimiento solo explorará nodos dentro de ella, y que el proceso se estanca cuando agotamos todos los nodos de la componente. En este caso, $\#R_t + \#A_t = t + \#A_t$ nos da una cota inferior para la cantidad de nodos en la componente conexa que contiene a A_0 . En particular, el tiempo que el proceso sobreviva antes de que

$A_t = \emptyset$ nos dará una noción de qué tan grande es la componente conexa asociada a los nodos en A_0 . Es así que llamaremos

$$\tau = \inf_{t \geq 0} \{ \#A_t = 0 \} \quad (\text{A.8})$$

y tendremos que la cantidad de nodos en la componente conexa asociada a A_0 estará dada por $\tau + \#A_\tau = \tau$.

Ahora definimos también un proceso auxiliar que, al sobrevivir menos tiempo que la exploración que introducimos anteriormente, nos dará una cota inferior para τ y así una cota inferior para el tamaño de la componente conexa asociada a los nodos en A_0 . Nuevamente, \hat{R} , \hat{A} y \hat{U} cumplirán el rol de nodos ya explorados, a explorar a continuación, y todavía no explorados respectivamente. Como condiciones iniciales, fijaremos $\hat{A}_0 = A_0$, y esto determina que

$$\begin{aligned} \hat{R}_0 &= \emptyset, \\ \hat{A}_0 &\subseteq V, \\ \hat{U}_0 &= V \setminus \hat{A}_0. \end{aligned} \quad (\text{A.9})$$

Luego, continuaremos inductivamente manteniendo \hat{R}_t , \hat{A}_t y \hat{U}_t constantes en caso que \hat{A}_t sea vacío, pero en caso que no tomaremos $j_t = \min(\hat{A}_t)$ y para $\delta \in (0, 1)$ fijo exploraremos como a continuación:

$$\begin{aligned} \hat{R}_{t+1} &= \hat{R}_t \cup \{ j_t \}, \\ \hat{A}_{t+1} &= (\hat{A}_t \setminus \{ j_t \}) \cup \{ y \in \hat{U}_t^\delta \mid j_t \sim y \}, \\ \hat{U}_{t+1} &= \hat{U}_t \setminus \{ y \in \hat{U}_t^\delta \mid j_t \sim y \}. \end{aligned} \quad (\text{A.10})$$

Aquí \hat{U}_t^δ es el subconjunto que tiene los $\lceil (1 - \delta)n_l \rceil$ nodos con índice más chicos dentro de \hat{U}_t con comunidad l , o todos si hubiesen menos, para cada comunidad $1 \leq l \leq k$.

Remarcamos como intuición en que la diferencia crucial con el proceso anterior se encuentra en que nos limitamos a develar menos nodos en cada paso que en el la primera exploración. Así, tomando para cada $t \in \mathbb{N}_0$ a i_t como j_t en el primer proceso de exploración (A.7), para poder comparar ambas exploraciones, y suponiendo que $\hat{A}_0 = A_0$, entonces tendremos que vale que $\hat{A}_t \subseteq A_t$ por un razonamiento inductivo. En particular, en ese caso obtenemos que $\#A_t \geq \#\hat{A}_t$ para cada $t \geq 0$ y entonces definiendo

$$\hat{\tau} = \inf_{t \geq 0} \{ \#\hat{A}_t = 0 \} \quad (\text{A.11})$$

tendremos que el tamaño de la componente conexa asociada a A_0 estará acotado inferiormente por $t + \#\hat{A}_t$ y en también por $\hat{\tau}$.

El problema radicará entonces en ver que esta nueva exploración sobrevive el suficiente tiempo y que tiene ahí una buena cantidad de nodos a explorar aún. Para mostrar esto, definiremos un tiempo de parada auxiliar

$$T_W = \inf_{t \geq 0} \left\{ \#\hat{A}_t = 0 \text{ o } \#\hat{U}_t \cap \mathcal{C}_l < (1 - \delta)n_l \text{ para algún } 1 \leq l \leq k \right\} \quad (\text{A.12})$$

y un proceso estocástico $(W_t)_{t \in \mathbb{N}_0}$ dado por:

$$\begin{aligned} W_0 &= \#\hat{A}_0, \\ W_{t+1} &= \begin{cases} W_t - 1 + \sum_{y \in \hat{U}_t^\delta} \mathbb{1}_{N(j_t)}(y) & \text{si } 0 < t < T_W, \\ W_{t+1} = W_t - 1 + \xi^t & \text{si } T_W \geq t, \end{cases} \end{aligned} \quad (\text{A.13})$$

donde ξ^t son variables aleatorias independientes con distribución

$$\xi^t \sim \sum_{l=1}^k \text{Binomial}(\lceil (1-\delta)n_l \rceil, (P_C)_{\mathcal{C}(j_t), l}) \quad (\text{A.14})$$

donde ahora j_t son nodos elegidos de manera independiente y uniformemente al azar dentro de V . Así, tenemos que los incrementos $W_{t+1} - W_t$ se distribuyen de manera independiente de W_s para $s \neq t+1$ y con distribución dada por una suma de binomiales independientes $\sum_{l=1}^k \text{Binomial}(\lceil (1-\delta)n_l \rceil, (P_C)_{\mathcal{C}(j_t), l})$.

La intuición crucial es que W_t lleva cuenta de la cantidad de nodos a explorar $\#\hat{A}_t$ hasta tiempo T_W , con lo que entender este tiempo de parada y este nuevo proceso nos dará información del tamaño de la componente conexas como buscamos. Más precisamente, $\#\hat{A}_t = W_t$ para cada $t < T_W$, y además $T_W \leq \hat{\tau}$ según lo cual el tamaño de la componente conexas asociada a A_0 está acotado inferiormente por $T_W + \#\hat{A}_{T_W}$.

Querremos mostrar que T_W es grande, que implica ver que W_t es positivo por un tiempo largo y que sin embargo en la exploración no agotamos (casi todos) los nodos de ninguna comunidad.

Comenzaremos encontrando condiciones para las cuales $W_t > 0$ para cada $t \in \mathbb{N}_0$ con alta probabilidad, donde usaremos como herramienta auxiliar que para ciertos $\theta \in \mathbb{R}$ vale que $e^{-\theta W_t}$ es una supermartingala [A.1.1](#). En ese caso, definiendo $T_0 = \inf_{t \geq 0} \{W_t = 0\}$ y usando el teorema de parada opcional para supermartingalas [A.1.2](#) para $T_0 \wedge t = \min\{T_0, t\} \leq t < \infty$ obtenemos que

$$\mathbb{E} \left[e^{-\theta \#\hat{A}_0} \right] = \mathbb{E} \left[e^{-\theta W_0} \right] \geq \mathbb{E} \left[e^{-\theta W_{T_0 \wedge t}} \right] \geq \mathbb{E} \left[e^{-\theta W_{T_0}} \mathbb{1}_{\{T_0 \leq t\}} \right] \geq \mathbb{P}(\{T_0 \leq t\}), \quad (\text{A.15})$$

y tomando límite cuando t tiende a infinito acotamos la probabilidad de que $W_t = 0$ en algún momento

$$\mathbb{E} \left[e^{-\theta \#\hat{A}_0} \right] \geq \mathbb{P}(\{T_0 < \infty\}). \quad (\text{A.16})$$

En particular, cuando $\theta \#\hat{A}_0$ sea suficientemente grande con alta probabilidad vamos a haber mostrado lo que queríamos, y por eso nos dedicaremos a continuación entonces a estudiar condiciones sobre θ y $\#\hat{A}_0$ que garanticen una buena cota en [\(A.16\)](#).

En primer lugar, observamos que resulta que $e^{-\theta W_t}$ va a ser una supermartingala cuando $\mathbb{E} \left[e^{\theta(W_{t+1} - W_t)} \right] \leq 1$ para cada $t \in \mathbb{N}_0$. En efecto, $e^{-\theta W_t}$ son todas variables integrables, y además en ese caso vamos a tener que $\mathbb{E} \left[e^{-\theta W_{t+1}} \mid e^{-\theta W_t} \right] \leq e^{-\theta W_t}$ para cada $t \in \mathbb{N}_0$ como

vemos a continuación:

$$\begin{aligned}\mathbb{E} \left[e^{-\theta W_{t+1}} \mid e^{-\theta W_t} \right] &= \mathbb{E} \left[e^{-\theta(W_{t+1}-W_t)} \mid e^{-\theta W_t} \right] e^{-\theta W_t} \\ &= \mathbb{E} \left[e^{-\theta(W_{t+1}-W_t)} \right] e^{-\theta W_t} \\ &\leq e^{-\theta W_t}.\end{aligned}\tag{A.17}$$

Nos quedará entonces simplemente hallar valores de θ que hacen que $\mathbb{E} \left[e^{-\theta(W_{t+1}-W_t)} \right] \leq 1$ para cada $t \in \mathbb{N}_0$.

Con ese objetivo en mente, acotaremos la esperanza por una familia de funciones $f_{l,\delta} : \mathbb{R} \rightarrow \mathbb{R}$ que dependerán de $1 \leq l \leq k$ y $\delta \in (0, 1)$, que nos permitirán un análisis más detallado:

$$\begin{aligned}\mathbb{E} \left[e^{-\theta(W_{t+1}-W_t)} \right] &= e^\theta \prod_{l=1}^k \left[\left(1 - (P_C)_{C(j_t),l} \right) + (P_C)_{C(j_t),l} e^{-\theta} \right]^{\lceil (1-\delta)n_l \rceil} \\ &\leq e^\theta \prod_{l=1}^k e^{\lceil (1-\delta)n_l \rceil (P_C)_{C(j_t),l} (e^{-\theta} - 1)} \\ &= e^{\theta + \left[\sum_{l=1}^k \lceil (1-\delta)n_l \rceil (P_C)_{C(j_t),l} \right] (e^{-\theta} - 1)} = f_{C(j_t),\delta}(\theta).\end{aligned}\tag{A.18}$$

A continuación, podemos deducir algunas propiedades de estas funciones que nos permitirá acotarlas. En primer lugar, podemos ver que $f_{l,\delta}(0) = 1$ y $\lim_{\theta \rightarrow \infty} f_{l,\delta}(\theta) = \infty$ para cada $1 \leq l \leq k$ y para cada $\delta \in (0, 1)$. Además, vamos a poder ver que cada función $f_{l,\delta}$ es estrictamente convexa estudiando su segunda derivada

$$\begin{aligned}f''_{l,\delta}(\theta) &= \left(1 - \left[\sum_{s=1}^k \lceil (1-\delta)n_s \rceil (P_C)_{l,s} \right] e^{-\theta} \right)^2 e^{\theta + \left[\sum_{s=1}^k \lceil (1-\delta)n_s \rceil (P_C)_{l,s} \right] (e^{-\theta} - 1)} \\ &\quad + \left(\left[\sum_{s=1}^k \lceil (1-\delta)n_s \rceil (P_C)_{l,s} \right] e^\theta \right) e^{\theta + \left[\sum_{s=1}^k \lceil (1-\delta)n_s \rceil (P_C)_{l,s} \right] (e^{-\theta} - 1)},\end{aligned}\tag{A.19}$$

que es estrictamente positiva. Ahora bien, como $\sum_{s=1}^k n_s (P_C)_{l,s} \geq \lambda_n \log(n)$ para cada $1 \leq l \leq k$, y como también $\liminf_{n \rightarrow \infty} \lambda_n > 1$ para cualquier $\delta \in (0, 1)$ fijo, tendremos que

$$\sum_{s=1}^k \lceil (1-\delta)n_s \rceil (P_C)_{l,s} \geq (1-\delta) \sum_{s=1}^k n_s (P_C)_{l,s} \geq (1-\delta) \lambda_n \log(n) > 1\tag{A.20}$$

cuando n sea suficientemente grande. Luego, en este caso tendremos que la primera derivada de $f_{l,\delta}$

$$f'_{l,\delta}(\theta) = \left(1 - \left[\sum_{s=1}^k \lceil (1-\delta)n_s \rceil (P_C)_{l,s} \right] e^{-\theta} \right) e^{\theta + \left[\sum_{s=1}^k \lceil (1-\delta)n_s \rceil (P_C)_{l,s} \right] (e^{-\theta} - 1)}\tag{A.21}$$

cumple que $f'_{l,\delta}(0) < 0$. Así, por ser estrictamente convexas, para cada $1 \leq l \leq k$ y cada $\delta \in (0, 1)$ hay una única solución positiva de $f_{l,\delta}(\theta) = 1$ que llamaremos $\theta_{l,\delta}$, y nuevamente por esa razón tendremos que $f_{l,\delta}(\theta) < 1$ para cada $\theta \in (0, \theta_{l,\delta})$, cuando n es suficientemente grande. Más aún, fijo $\delta > 0$ podremos eliminar la dependencia en

l notando que para $\theta_\delta = \min_{1 \leq l \leq k} \{ \theta_{l,\delta} \}$ tendremos que $f_{l,\delta}(\theta) < 1$ para cada $\theta \in (0, \theta_\delta)$, para n suficientemente grandes. Por último, observamos que $\theta_\delta \geq 1$ de hecho porque $e^{1 + \left[\sum_{i=1}^k \lceil (1-\delta)n_i \rceil (P_C)_{C(j_i),l} \right] (e^{-1}-1)} < 1$ para n suficientemente grandes, porque el factor acompañando a $(e^{-1} - 1) < 0$ es de orden logarítmico, con lo que podemos imaginar que $\theta_\delta = 1$ en adelante o aunque sea contamos con la intuición de que $\frac{1}{\theta_\delta}$ no es un número grande.

Concluimos entonces que que $e^{-\frac{1}{2}\theta_\delta W_t}$ es una supermartingala para n suficientemente grandes. A modo de comentario que será útil en una futura demostración, observamos que la dependencia de n corresponde a que $\log(n)$ sea suficientemente grande como para que $(1-\delta)\lambda_n \log(n) > 1$ en A.20, y podríamos acotarlo uniformemente en λ_n si supiésemos qué tan rápido ocurre que λ_n está sobre 1 (cosa que eventualmente ocurrirá pues $\liminf_{n \rightarrow \infty} \lambda_n > 1$).

Recordando ahora la cota que obtuvimos en A.16, queremos que $\#A_0$ sea grande. Supongamos entonces por ahora que el evento

$$\left\{ \text{La componente conexa de } i \text{ tiene por lo menos } \left\lceil \frac{4}{\theta_\delta} \log(n) \right\rceil \right\} \quad (\text{A.22})$$

ocurre con probabilidad más alta que $1 - o(n^{-1})$ para cada $i \in V$. Luego, fijemos que $A_{0,i}$ sea el conjunto aleatorio que tendrá los $\left\lceil \frac{4}{\theta_\delta} \log(n) \right\rceil - 1$ nodos de índice más bajo en la componente conexa de i además que a i mismo, si hubiesen suficientes nodos, y si no todos los nodos de la componente conexa de i , para cada $i \in V$. Estos serán n conjuntos y como supusimos van a tener tamaño $\left\lceil \frac{4}{\theta_\delta} \log(n) \right\rceil$ con alta probabilidad. Comenzando la exploración a partir de ellos, obtendremos que para cada nodo $i \in V$ valdrá que

$$\begin{aligned} \mathbb{P}(\{T_0 < \infty\}) &\leq \mathbb{E} \left[e^{-\theta_\delta A_{0,i}} \right] & (\text{A.23}) \\ &\leq \mathbb{P} \left(\left\{ \#A_{0,i} < \left\lceil \frac{4}{\theta_\delta} \log(n) \right\rceil \right\} \right) \cdot 1 + \mathbb{P} \left(\left\{ \#A_{0,i} \geq \left\lceil \frac{4}{\theta_\delta} \log(n) \right\rceil \right\} \right) \cdot n^{-2} \\ &= o(n^{-1}) + 1 \cdot n^{-2} \\ &= o(n^{-1}). \end{aligned}$$

En particular, procediendo como en la Observación 2.2.5 vale entonces que las exploraciones comenzadas en alguno de estos n subconjuntos simultáneamente verifican que vale que $T_0 = \infty$ con probabilidad mayor que $1 - o(1)$.

Resta ver entonces que en efecto las componentes conexas de cada nodo tienen por lo menos esta cantidad $\left\lceil \frac{4}{\theta_\delta} \log(n) \right\rceil$ de nodos, en cuyo caso mejoraremos la cota a que tienen por lo menos $x(1-\delta)\sqrt{n} \log(n)$ nodos con alta probabilidad.

Fijemos entonces i un nodo. Llamemos \mathcal{A}_1 al evento

$$\mathcal{A}_1 = \left\{ C_1 \log(n) \leq \widehat{\deg}_n(j) \leq C_2 \log(n) \text{ para cada nodo } j \right\}, \quad (\text{A.24})$$

que valdrá con alta probabilidad por la Proposición 2.4.11 para ciertas constantes $0 < C_1 < 1 < C_2$. Es así que i tiene una cantidad logarítmica de vecinos con alta probabilidad, e iterando este razonamiento para sus vecinos obtenemos que cada vecino de i tiene también una

cantidad logarítmica de vecinos. Podremos entonces acotar la cantidad de nodos a distancia 2 de i a partir de la suma de los tamaños de las vecindades de vecinos de i , que tendrá un orden de $(\log(n))^2$... pero debemos compensar que es posible que tengan nodos repetidos entre sí, incluyendo a i . Sin embargo, podremos ver que no hay mucho solapamiento entre vecindades: más precisamente vamos a ver que \mathcal{A}_2 el evento dado por

$$\mathcal{A}_2 = \{ \# N(u) \cap N(v) \leq C \text{ para cada pareja de nodos } u, v \} \quad (\text{A.25})$$

ocurre con alta probabilidad, donde C es una constante suficientemente grande. Pero entonces, y restringiéndonos a considerar solamente los \tilde{C} vecinos de índice más chico de i , tendremos que

$$\# \{j \in V \mid \text{dist}(j, i) = 2\} \geq \tilde{C} C_1 \log(n) - \binom{\tilde{C}}{2} C^2 \geq \left\lceil \frac{4}{\theta_\delta} \log(n) \right\rceil \quad (\text{A.26})$$

para n suficientemente grandes, cuando \tilde{C} es suficientemente grande. En particular, podemos concluir como queríamos que

$$\mathbb{P} \left(\left\{ \# \{j \in V \mid \text{dist}(j, i) \leq 2\} \geq \left\lceil \frac{4}{\theta_\delta} \log(n) \right\rceil \right\} \right) \geq \mathbb{P}(\mathcal{A}_1 \cap \mathcal{A}_2) = 1 - o(1), \quad (\text{A.27})$$

al ser ambos eventos que ocurren con alta probabilidad.

Resta entonces convencernos de que es asintóticamente improbable que dos nodos fijos tengan muchos vecinos en común. Fijemos u, v dos nodos y tendremos que cada otro nodo j será vecino en común de ambos con probabilidad $(P_{\mathcal{C}})_{\mathcal{C}(u), \mathcal{C}(j)} \cdot (P_{\mathcal{C}})_{\mathcal{C}(v), \mathcal{C}(j)}$; luego por independencia tendremos que la cantidad de vecinos en común de ambos nodos $\# N(u) \cap N(v)$ será una suma de variables binomiales independientes

$$\sum_{l=1}^k \text{Binomial}(n_l - \mathbb{1}_{\{l\}}(\mathcal{C}(u)) - \mathbb{1}_{\{l\}}(\mathcal{C}(v)), (P_{\mathcal{C}})_{\mathcal{C}(u), l} \cdot (P_{\mathcal{C}})_{\mathcal{C}(v), l}). \quad (\text{A.28})$$

Notemos que entonces la esperanza y la varianza serán muy chicas, donde $\text{Var}[\# N(u) \cap N(v)] \leq \mathbb{E}[\# N(u) \cap N(v)]$ y son ambas de orden $\Theta\left(\frac{(\log(n))^2}{n}\right)$.

Luego usando el Lema 2.4.12 para $t = 1$, $\delta = 0$ e $y = \tilde{C} \frac{n}{(\log(n))^2}$ obtenemos que

$$\begin{aligned} & \mathbb{P}(\{ \# N(u) \cap N(v) \geq y \mathbb{E}[\# N(u) \cap N(v)] \}) \\ & \leq e^{-(y \log(y) - y + 1) \mathbb{E}[\# N(u) \cap N(v)]} \\ & = \mathcal{O}(n^{-3}) \end{aligned} \quad (\text{A.29})$$

donde

$$\begin{aligned} & (y \log(y) - y + 1) \mathbb{E}[\# N(u) \cap N(v)] \\ & \geq \tilde{C} \left[\frac{n}{(\log(n))^2} \mathbb{E}[\# N(u) \cap N(v)] \right] \left[\log \left(\tilde{C} \frac{n}{(\log(n))^2} \right) - 1 \right] \end{aligned} \quad (\text{A.30})$$

será asintóticamente más grande que $3 \log(n)$ para \tilde{C} suficientemente grandes. En particular, mediante una cota por la unión de eventos podremos mostrar que para aquella \tilde{C} constante

vale que simultáneamente cada pareja de nodos no tiene más que \tilde{C} vecinos en común con alta probabilidad, porque hay solamente $\binom{n}{2} = \mathcal{O}(n^2)$ tales parejas.

Ahora podremos ocuparnos de ver que en esta exploración tardamos tiempo en agotar los nodos de cada comunidad, que es la otra condición que precisamos para controlar T_W . Para eso, podemos usar el Lema 2.4.12 dos veces, para obtener que para tiempos $t = \sqrt{n}$ y valores $x < 1 < y$ tales que $\gamma(x), \gamma(y) > 0$ valen que

$$\mathbb{P} \left(\left\{ W_{\sqrt{n}} - W_0 \geq x \mathbb{E} [W_{\sqrt{n}} - W_0] \right\} \right) \leq e^{-\gamma(x) \mathbb{E} [W_{\sqrt{n}} - W_0]} \quad (\text{A.31})$$

y también

$$\mathbb{P} \left(\left\{ W_{\sqrt{n}} - W_0 \leq y \mathbb{E} [W_{\sqrt{n}} - W_0] \right\} \right) \leq e^{-\gamma(y) \mathbb{E} [W_{\sqrt{n}} - W_0]}. \quad (\text{A.32})$$

Como podemos calcular que

$$\begin{aligned} \mathbb{E} [W_{\sqrt{n}} - W_0] &= \mathbb{E} \left[\sum_{s=1}^{\sqrt{n}} W_s - W_{s-1} \right] \\ &= \sum_{s=1}^{\sqrt{n}} \mathbb{E} [W_s - W_{s-1}] \\ &\begin{cases} \geq -\sqrt{n} + \sqrt{n}(1 - \delta) \min_{1 \leq t \leq t} \left\{ \frac{n_t}{n} a + \frac{n-n_t}{n} b \right\} \log(n) \\ \leq -\sqrt{n} + \sqrt{n} \max_{1 \leq t \leq t} \left\{ \frac{n_t}{n} a + \frac{n-n_t}{n} b \right\} \log(n) \end{cases} \end{aligned} \quad (\text{A.33})$$

podemos obtener que para n suficientemente grandes vale que

$$\sqrt{n} [(1 - \delta) \lambda_n \log(n) - 1] \leq \mathbb{E} [W_{\sqrt{n}} - W_0] \leq \sqrt{n} [(a + b) \log(n) - 1]. \quad (\text{A.34})$$

En particular, a tiempo \sqrt{n} se habrá revelado una cantidad de nodos en el orden de $\sqrt{n} \log(n)$ con probabilidad mayor que $1 - n^{-C\sqrt{n}}$ donde C es alguna constante positiva, para n suficientemente grandes. Así, podemos concluir que con alta probabilidad $T_W > \sqrt{n}$, porque $\#A_0 = \left\lceil \frac{4}{\theta_\delta} \log(n) \right\rceil = o(n)$ y al haberse revelado una cantidad sublineal de nodos no podemos haber agotado los nodos de ninguna de las comunidades que son lineales.

Además, como $\#\hat{A}_{\sqrt{n}} = W_{\sqrt{n}}$ en este caso tendremos que la componente conexa asociada a A_0 tendrá más aún por lo menos $\sqrt{n} + \#\hat{A}_{\sqrt{n}} \geq x\sqrt{n} \log(n)$ nodos con probabilidad mayor que $1 - n^{-C\sqrt{n}}$ para alguna constante C y para n suficientemente grandes, donde usamos que eventualmente $\lambda_n > 1$ porque $\liminf_{n \rightarrow \infty} \lambda_n > 1$. Recordando nuestra cota A.23, tenemos que $\#A_0 = \left\lceil \frac{4}{\theta_\delta} \log(n) \right\rceil$, y es así que como en la Observación 2.2.5 al haber no más que $\binom{n}{\#A_0} = \mathcal{O} \left(n^{\frac{4}{\theta_\delta} \log(n) + 1} \right)$ subconjuntos posibles de ese tamaño $\left\lceil \frac{4}{\theta_\delta} \log(n) \right\rceil$ para empezar la exploración tendremos que el evento

$$\left\{ \#\hat{A}_t \geq x(1 - \delta)\sqrt{n} \log(n) \text{ para cada } A_0 \subseteq V \text{ de tamaño } \left\lceil \frac{4}{\theta_\delta} \log(n) \right\rceil \right\} \quad (\text{A.35})$$

ocurrirá con probabilidad más alta que $1 - o(1)$. En particular, combinando con las cotas anteriores obtenemos que el evento

$$\mathcal{A}_3 = \left\{ \text{Cada nodo tiene más que } x(1 - \delta)\sqrt{n} \log(n) \text{ nodos en su componente conexa} \right\} \quad (\text{A.36})$$

ocurre con probabilidad alta pues podemos iniciar la exploración con suficientes nodos y esto desembocará en que a tiempo \sqrt{n} hayamos explorado por lo menos esa cantidad de nodos en cada componente conexa ambas con alta probabilidad.

Finalmente, podemos mostrar que las componentes conexas asociadas a dos nodos distintos deben intersectarse. En efecto, como podemos acotar la probabilidad de que dos subconjuntos de tamaños mayores que $x(1-\delta)\sqrt{n}\log(n)$ fijos no se intersequen por

$$\left(1 - b \frac{\log(n)}{n}\right)^{(x(1-\delta)\sqrt{n}\log(n))^2} \leq e^{-bx^2(1-\delta)^2 \log(n)^3} = o(n^{-2}) \quad (\text{A.37})$$

obtenemos que por una cota por la unión de eventos vale que

$$\begin{aligned} & \mathbb{P}(\{\text{El grafo no es conexo}\}) \\ & \leq \mathbb{P}(\mathcal{A}_3^C) + \mathbb{P}(\mathcal{A}_3 \cap \{\exists i, j \in V \text{ cuyas componentes conexas no coinciden}\}) \\ & \leq o(1) + \binom{n}{2} o(n^{-2}) \\ & = o(1) \end{aligned} \quad (\text{A.38})$$

y completamos la demostración. ■

Daremos ahora la implicación en la otra dirección. La implicación aquí escrita sirve para el caso balanceado, pero puede adaptarse al caso general con tamaños de comunidades fijas:

Demostración de la cantidad de nodos aislados en 2.4.8. En este caso, la estrategia será mostrar que la cantidad esperada de nodos aislados para cada comunidad escala como $n^{1-\frac{a+(k-1)b}{k}}$ y luego usar cotas para la varianza para mostrar que con probabilidad alta el grado de cada nodo estará cerca del grado esperado, con lo que podremos concluir que efectivamente cada comunidad tiene una cantidad en ese orden de nodos aislados.

Para comenzar, supondremos sin pérdida de generalidad que los nodos están ordenados por comunidad, como mencionamos en la Observación 2.4.7 - si lo demostramos para esta asignación de comunidades fija entonces por probabilidad total valdrá para cualquier asignación balanceada. Más aún, sin pérdida de generalidad trabajaremos con la comunidad 1, dado que por un lado la cantidad de nodos aislados para el resto de las comunidades seguirá la misma distribución al ser las comunidades balanceadas; y además por otro lado como hay finitas comunidades tendremos por la Observación 2.2.5 que mostrando que con alta probabilidad vale la cota para cada una de ellas individualmente, también valdrá para todas simultáneamente.

Mostremos entonces que el grado esperado tiene el orden buscado. Primero notamos que fijo un nodo $i \in V$ cualquiera, su grado $\deg(i)$ tiene distribución $X + Y$ para $X \sim \text{Binomial}\left(\frac{1}{k}n - 1, p_n\right)$ e $Y \sim \text{Binomial}\left(\frac{k-1}{k}n, q_n\right)$ independientes, debido a que hay X aristas conectándolo con los otros nodos de su comunidad e Y aristas conectándolo con nodos de

las otras comunidades. Luego tenemos que para cada nodo i vale que

$$\begin{aligned} \mathbb{P}(\{\deg(i) = 0\}) &= \mathbb{P}(\{X = 0\}) \mathbb{P}(\{Y = 0\}) \\ &= (1 - p_n)^{\frac{1}{k}n-1} (1 - q_n)^{\frac{k-1}{k}n} \\ &\asymp n^{-\frac{a+(k-1)b}{k}}, \end{aligned} \quad (\text{A.39})$$

donde el orden se deduce de que por el límite clásico $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n}\right)^n = e$ tenemos que

$$\lim_{n \rightarrow \infty} (1 - p_n)^{\frac{1}{k}n-1} (n^{-\frac{a}{k}})^{-1} = 1 \quad (\text{A.40})$$

y también

$$\lim_{n \rightarrow \infty} (1 - q_n)^{\frac{k-1}{k}n} (n^{-\frac{(k-1)b}{k}})^{-1} = 1. \quad (\text{A.41})$$

Definiendo entonces $I_n = \#\{i \in \mathcal{C}_1 \mid \deg(i) = 0\}$ la cantidad de nodos aislados en \mathcal{C}_1 , como $I_n = \sum_{i=1}^{\frac{n}{k}} \mathbb{1}_{\{\deg(i)=0\}}$ podremos deducir que su esperanza es del orden adecuado:

$$\begin{aligned} \mathbb{E}[I_n] &= \frac{n}{k} \mathbb{P}(\{\deg(1) = 0\}) \\ &\asymp n^{1 - \frac{a+(k-1)b}{k}}. \end{aligned} \quad (\text{A.42})$$

Solamente resta ver entonces que I_n es cercano a su esperanza. En efecto, adelantando que $\text{Var}[I_n] = \Theta(\mathbb{E}[I_n])$, y además usando la desigualdad de Bernstein 2.2.4 y que $\mathbb{E}[I_n] = \Omega(\log(n))$ por (A.42) podremos acotar para ciertas constantes $C > 0$ variando de línea de la siguiente manera

$$\begin{aligned} \mathbb{P}\left(\left\{|I_n - \mathbb{E}[I_n]| > \sqrt{\log(n)\mathbb{E}[I_n]}\right\}\right) &\leq 2 \exp\left(-\frac{1}{2} \frac{\log(n)\mathbb{E}[I_n]}{\text{Var}[I_n] + \frac{1}{3}\sqrt{\log(n)\mathbb{E}[I_n]}}\right) \\ &\leq 2 \exp\left(-C \frac{\log(n)\mathbb{E}[I_n]}{\mathbb{E}[I_n] + \sqrt{\mathbb{E}[I_n]\mathbb{E}[I_n]}}\right) \\ &= 2 \exp(-C \log(n)) = 2n^{-C} = o(1), \end{aligned} \quad (\text{A.43})$$

con lo que obtenemos que $I_n = \mathbb{E}[I_n] + \mathcal{O}\left(\sqrt{\log(n)\mathbb{E}[I_n]}\right) = \Theta(\mathbb{E}[I_n])$ como queríamos.

Queda entonces solamente verificar que $\text{Var}[I_n] = \Theta(\mathbb{E}[I_n])$ para completar la demostración. Efectivamente, primero expresando la varianza como suma de covarianzas obtenemos que

$$\begin{aligned} \text{Var}[I_n] &= \frac{n}{k} \text{Var}\left[\mathbb{1}_{\{\deg(1)=0\}}\right] \\ &\quad + \frac{n}{k} \left(\frac{n}{k} - 1\right) \text{Cov}\left[\mathbb{1}_{\{\deg(1)=0\}}, \mathbb{1}_{\{\deg(2)=0\}}\right], \end{aligned} \quad (\text{A.44})$$

y a continuación usando (A.39) podremos ver que el primer término tiene orden

$$\begin{aligned} \frac{n}{k} \text{Var}\left[\mathbb{1}_{\{\deg(1)=0\}}\right] &= \frac{n}{k} \mathbb{P}(\{\deg(1) = 0\}) (1 - \mathbb{P}(\{\deg(1) = 0\})) \\ &\asymp n^{1 - \frac{a+(k-1)b}{k}} \end{aligned} \quad (\text{A.45})$$

y que el segundo término tiene orden

$$\begin{aligned} & \frac{n}{k} \left(\frac{n}{k} - 1 \right) \text{Cov} \left[\mathbb{1}_{\{\text{deg}(1)=0\}}, \mathbb{1}_{\{\text{deg}(2)=0\}} \right] \\ &= \frac{n}{k} \left(\frac{n}{k} - 1 \right) \left((1 - p_n)^{-1} - 1 \right) \mathbb{P}(\{\text{deg}(1) = 0\}) \mathbb{P}(\{\text{deg}(2) = 0\}), \\ &\asymp p_n n^{2 \left(1 - \frac{a+(k-1)b}{k} \right)}, \end{aligned} \quad (\text{A.46})$$

a partir de lo cual podremos deducir que la varianza tiene el mismo orden que la esperanza como queríamos ver

$$\begin{aligned} \text{Var} [I_n] &\asymp n^{1 - \frac{a+(k-1)b}{k}} + p_n n^{2 \left(1 - \frac{a+(k-1)b}{k} \right)} \\ &\asymp n^{1 - \frac{a+(k-1)b}{k}} = \Theta(\mathbb{E} [I_n]). \end{aligned} \quad (\text{A.47})$$

Para terminar, aclaramos para estas deducciones por un lado que en (A.47) usamos que, como $\frac{a+(k-1)b}{k} < 1$ y $p_n \xrightarrow{n \rightarrow \infty} 0$, el segundo orden es asintóticamente más chico que el primero; y por otro lado para el cálculo de la covarianza en (A.46) notamos que

$$\begin{aligned} & \mathbb{P}(\{\text{deg}(1) = 0, \text{deg}(2) = 0\}) \\ &= (1 - p_n)^{-1} \mathbb{P}(\{\text{deg}(1) = 0\}) \mathbb{P}(\{\text{deg}(2) = 0\}) \end{aligned} \quad (\text{A.48})$$

porque para que los nodos 1 y 2 tengan grado 0 se precisa que todas las aristas adyacentes no estén en el grafo, de las cuales hay exactamente una en común. ■

Por último, la demostración de la transición de fase en la conexión se da también en el modelo SBM general podemos obtenerla a partir de la ley de probabilidad total, y utilizando que ya sabemos que cuando los tamaños de las comunidades están fijos tenemos que vale el resultado, en el espíritu de las demostraciones de los Teoremas 4.3.2 y 5.4.9.

A.2 Demostraciones del capítulo 4

Con el objetivo de demostrar el Teorema 4.3.1, comenzaremos nuevamente notando que el corte ratio asociado a las comunidades reales es más chico que los cortes ratio asociados a cualquier otra partición, en valor esperado:

[A.2.1] Lema: Cálculo de esperanza y varianza para la diferencia de corte ratio.

Sea $G \sim \text{SBM} \left(n, 2, (n_i)_{1 \leq i \leq 2}, p_n = a \frac{\log(n)}{n}, q_n = b \frac{\log(n)}{n} \right)$, y sea $\tilde{\mathcal{C}}$ una partición de comunidades para G .

Luego, vale que

$$\begin{aligned} \mathbb{E} [\text{RC}(\tilde{\mathcal{C}}) - \text{RC}(\mathcal{C})] &= \left(\frac{n}{m_1 m_2} \right) (X_{1,1} X_{1,2} + X_{2,1} X_{2,2}) (p_n - q_n) \\ &= \frac{1}{\beta_1 \beta_2} (x_{1,1} x_{1,2} + x_{2,1} x_{2,2}) n (p_n - q_n), \end{aligned} \quad (\text{A.49})$$

y

$$\begin{aligned}
\text{Var} [\text{RC} (\tilde{\mathcal{C}}) - \text{RC} (\mathcal{C})] &= \left(\frac{n}{m_1 m_2} - \frac{n}{n_1 n_2} \right)^2 (X_{1,1} X_{2,2} + X_{1,2} X_{2,1}) (q_n (1 - q_n)) \\
&\quad + \left(\frac{n}{n_1 n_2} \right)^2 (X_{1,1} X_{2,1} + X_{1,2} X_{2,2}) (q_n (1 - q_n)) \\
&\quad + \left(\frac{n}{m_1 m_2} \right)^2 (X_{1,1} X_{1,2} + X_{2,1} X_{2,2}) (p_n (1 - p_n)) \\
&= \left(\frac{1}{\beta_1 \beta_2} - \frac{1}{\alpha_1 \alpha_2} \right)^2 (x_{1,1} x_{2,2} + x_{1,2} x_{2,1}) (q_n (1 - q_n)) \\
&\quad + \left(\frac{1}{\alpha_1 \alpha_2} \right)^2 (x_{1,1} x_{2,1} + x_{1,2} x_{2,2}) (q_n (1 - q_n)) \\
&\quad + \left(\frac{1}{\beta_1 \beta_2} \right)^2 (x_{1,1} x_{1,2} + x_{2,1} x_{2,2}) (p_n (1 - p_n)).
\end{aligned}$$

Más aún, vale que cuando $a > b$ tenemos que

$$\mathbb{E} [\text{RC} (\tilde{\mathcal{C}}) - \text{RC} (\mathcal{C})] > 0 \quad (\text{A.50})$$

para cualquier partición $\tilde{\mathcal{C}}$ que no induzca la misma partición que \mathcal{C} .

Demostración. Para el corte ratio se procede de forma análoga a como hicimos para el corte de Cheeger en el Lema 4.2.2, aprovechando que la diferencia entre los cortes ratio es una combinación lineal de variables Bernoulli:

$$\begin{aligned}
\mathbb{E} [\text{RC} (\tilde{\mathcal{C}}) - \text{RC} (\mathcal{C})] &= \left(\frac{n}{m_1 m_2} - \frac{n}{n_1 n_2} \right) (X_{1,1} X_{2,2} + X_{1,2} X_{2,1}) q_n \\
&\quad - \left(\frac{n}{n_1 n_2} \right) (X_{1,1} X_{2,1} + X_{1,2} X_{2,2}) q_n \\
&\quad + \left(\frac{n}{m_1 m_2} \right) (X_{1,1} X_{1,2} + X_{2,1} X_{2,2}) p_n \\
&= \left(\frac{1}{\beta_1 \beta_2} - \frac{1}{\alpha_1 \alpha_2} \right) (x_{1,1} x_{2,2} + x_{1,2} x_{2,1}) n q_n \\
&\quad - \left(\frac{1}{\alpha_1 \alpha_2} \right) (x_{1,1} x_{2,1} + x_{1,2} x_{2,2}) n q_n \\
&\quad + \left(\frac{1}{\beta_1 \beta_2} \right) (x_{1,1} x_{1,2} + x_{2,1} x_{2,2}) n p_n.
\end{aligned}$$

Por otro lado, podemos simplificar la expresión nuevamente sumando y restando

$$\frac{n}{m_1 m_2} (X_{1,1} X_{1,2} + X_{2,1} X_{2,2}) q_n, \quad (\text{A.51})$$

y obtenemos que gran parte de la expresión se cancela:

$$\begin{aligned}
\mathbb{E} [\text{RC} (\tilde{\mathcal{C}}) - \text{RC} (\mathcal{C})] &= \left(m_1 m_2 \left(\frac{n}{m_1 m_2} \right) - n_1 n_2 \left(\frac{n}{n_1 n_2} \right) \right) q_n \\
&\quad + \left(\frac{n}{m_1 m_2} \right) (X_{1,1} X_{1,2} + X_{2,1} X_{2,2}) (p_n - q_n) \\
&= \left(\frac{n}{m_1 m_2} \right) (X_{1,1} X_{1,2} + X_{2,1} X_{2,2}) (p_n - q_n).
\end{aligned} \quad (\text{A.52})$$

A continuación podemos expresar las esperanzas en términos de proporciones también, y con esto quedan halladas las expresiones del lema:

$$\mathbb{E} [\text{RC}(\tilde{\mathcal{C}}) - \text{RC}(\mathcal{C})] = \frac{1}{\beta_1 \beta_2} (x_{1,1} x_{1,2} + x_{2,1} x_{2,2}) n(p_n - q_n),$$

Por otro lado, notamos que en este caso solamente usando que $a > b$, y por lo tanto $p_n > q_n$, inmediatamente la esperanza es no negativa. Además, como $x_{1,1} x_{1,2} + x_{2,1} x_{2,2} \neq 0$ cuando hay algún nodo mal clasificado, la esperanza será en este caso positiva.

Finalmente, también aprovechando que la diferencia entre los cortes ratio es una combinación lineal de variables Bernoulli independientes podemos verificar las expresiones para la varianza. ■

Sin más preámbulos, demostremos ahora el Teorema 4.3.1, para lo cual imitaremos la estrategia que usamos para demostrar el Teorema 4.2.1:

Demostración. Nuevamente, para demostrar que el estimador es fuertemente consistente mostraremos que a.a.s. vale el corte inducido por las comunidades reales $\text{RC}(\mathcal{C})$ minimiza el corte ratio, en cuyo caso $\hat{\mathcal{C}}_{\text{RC}} = \mathcal{C}$ a.a.s..

También esta vez supondremos sin pérdida de generalidad que $m_1 \leq m_2$ (o $\beta_1 \leq \beta_2$) y $n_1 \leq n_2$ (o $\alpha_1 \leq \alpha_2$) durante la demostración, pero como las comunidades no son necesariamente balanceadas no necesariamente tendremos que $n_1 = n_2 = \frac{n}{2}$ (o $\alpha_1 = \alpha_2 = \frac{1}{2}$).

Una vez más, para demostrar la minimalidad usaremos primero una cota por la unión de eventos y a continuación la desigualdad de Bernstein y nos enfocaremos en acotar cada uno de los términos de la suma resultante, análogamente a (4.31), (4.32) y (4.34):

$$\begin{aligned} \mathbb{P} \left(\left\{ \text{RC}(\tilde{\mathcal{C}}_{\text{RC}}) < \text{RC}(\mathcal{C}) \right\} \right) &\leq \sum_{\tilde{\mathcal{C}}, \text{coincidencia}(\tilde{\mathcal{C}}, \mathcal{C}) \neq n} \mathbb{P} \left(\left\{ \text{RC}(\tilde{\mathcal{C}}) < \text{RC}(\mathcal{C}) \right\} \right) \quad (\text{A.53}) \\ &\leq \sum_{\tilde{\mathcal{C}}, \text{coincidencia}(\tilde{\mathcal{C}}, \mathcal{C}) \neq n} 2 \exp \left(-\frac{1}{2} \frac{E^2}{V + \frac{M}{3} E} \right) \\ &= \sum_{X_{1,1}=1}^{n_1} \sum_{X_{2,1}=1, X \text{ admisible}}^{n_2} \binom{n_1}{X_{1,1}} \binom{n_2}{X_{2,1}} 2 \exp \left(-\frac{1}{2} \frac{E^2}{V + \frac{M}{3} E} \right). \end{aligned}$$

Nuevamente, bastará ver que cada término es $o(n^2)$ pues hay una cantidad cuadrática de ellos, y también E , V y M dependen de n , $\tilde{\mathcal{C}}$ y \mathcal{C} aunque esto no está explícitamente manifiesto en la notación.

Notamos que podemos determinar M precisamente, como el máximo entre $\frac{n}{m_1 m_2}$, $\frac{n}{n_1 n_2}$ y $\left| \frac{n}{m_1 m_2} - \frac{n}{n_1 n_2} \right|$, que son los valores absolutos de los coeficientes en la combinación lineal de variables Bernoulli que es la diferencia de cortes ratio $\text{RC}(\tilde{\mathcal{C}}) - \text{RC}(\mathcal{C})$. Así, obtenemos que

$$M = \begin{cases} \frac{n}{m_1 m_2} & \text{si } m_1 \leq n_1, \text{ y} \\ \frac{n}{n_1 n_2} & \text{si } m_1 > n_1. \end{cases} \quad (\text{A.54})$$

Una vez más, comenzamos por acotar los números combinatorios. Si $j, j' \in \{1, 2\}$ mediante el Lema 4.2.3 podremos mostrar que

$$\begin{aligned} \binom{n_1}{X_{1,j}} \binom{n_2}{X_{2,j'}} &\leq \exp \left(X_{1,j}(\log(n_1) + 1) + X_{2,j'}(\log(n_2) + 1) \right) \\ &\leq \exp \left((X_{1,j} + X_{2,j'}) (\log(n) + 1) \right), \end{aligned} \quad (\text{A.55})$$

con lo cual nuevamente buscaremos acotar las probabilidades con una cota que contrarreste este orden.

A continuación, partiremos en 8 casos, que corresponden a dos grupos de cuatro casos idénticos: unos para cuando $m_1 \leq n_1$ que notaremos por 1., 2., 3. y 4.; y otros para cuando $m_1 > n_1$, que notaremos 5., 6., 7., y 8.. Listaremos los primeros, y los segundos se definirán análogamente en el mismo orden:

1. $x_{1,1}, x_{1,2} \geq \varepsilon$ o $x_{2,1}, x_{2,2} \geq \varepsilon$,
2. $x_{1,1}, x_{2,1} \leq \varepsilon$,
3. $x_{1,1} \leq \varepsilon$ y $x_{2,1} \geq \alpha_2 - \varepsilon$,
4. $x_{1,1} \geq \alpha_1 - \varepsilon$ y $x_{2,1} \leq \varepsilon$,

En contraste con la partición en casos que realizamos en 4.2.1 y que visualizamos mediante la Figura 4.2, en esta ocasión no tendremos tanta simetría al ser las comunidades desbalanceadas y tendremos que añadir un caso más; por otra parte tampoco podremos asegurar que $m_1 \leq n_1$ y por eso tendremos dos tandas de casos. Una vez más, expresando cada caso solamente en términos de $x_{1,1}$, que variará entre 0 y α_1 , y $x_{2,1}$, que variará entre 0 y α_2 , podremos representar en las coordenadas de un diagrama como en el de la Figura A.1 todos los casos a considerar. Nuevamente, no puede ocurrir que simultáneamente $x_{1,1}$ y $x_{2,1}$ sean grandes porque $x_{1,1} + x_{2,1} = \beta_1$ y $\beta_1 \leq \frac{1}{2}$ al ser la comunidad estimada más chica como supusimos previamente, con lo que basta que consideremos en nuestro análisis coordenadas en el cuadrado debajo de la línea punteada. Como estas efectivamente están cubiertas por los casos 1. a 4. (o 5. a 8.) bastará entonces mostrar que los términos en (A.53) asociados a cada uno son de orden $o(n^{-2})$ para demostrar la consistencia fuerte de $\widehat{\mathcal{C}}_{\text{RC}}$.

En primer lugar, tomaremos P , Q y R polinomios que pensaremos siempre evaluados en las proporciones $x_{i,j}$ tales que

$$\begin{cases} \frac{\beta_1 \beta_2}{\log(n)} E = P, \\ \frac{n}{\log(n)} (\alpha_1 \alpha_2 \beta_1 \beta_2)^2 \left(V + \frac{M}{3} E \right) = Q + R \frac{\log(n)}{n}, \end{cases} \quad (\text{A.56})$$

y de esta manera obtenemos una nueva forma de entender el exponente en la cota por Bernstein:

$$-\frac{1}{2} \frac{E^2}{V + \frac{M}{3} E} = -\frac{1}{2} n \log(n) \alpha_1^2 \alpha_2^2 \frac{P^2}{Q + R \frac{\log(n)}{n}}. \quad (\text{A.57})$$

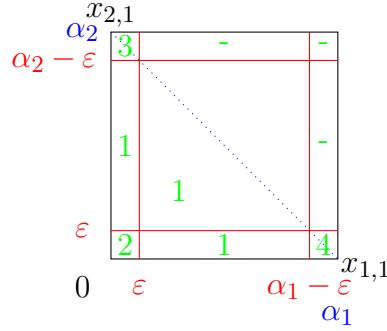


Figura A.1: Un diagrama para visualizar los casos en la demostración de consistencia fuerte para el corte ratio.

Supongamos primero entonces que $m_1 \leq n_1$, y demostremos que valen los casos 1. a 4..

Comencemos con el **caso 1.**, donde procediendo como en la demostración del Teorema 4.2.1 para el caso análogo usaremos las expresiones para E y V obtenidas en A.2.1 y también que $M = \frac{n}{m_1 m_2} = \frac{1}{n \beta_1 \beta_2}$ por (A.54) para obtener fórmulas concretas para P , Q y R :

$$P = ax_{1,1}x_{1,2} - bx_{1,1}x_{1,2} + ax_{2,1}x_{2,2} - bx_{2,1}x_{2,2}, \quad (\text{A.58})$$

$$Q = bx_{1,1}^3x_{1,2}^2x_{2,1} + bx_{1,1}^2x_{1,2}^3x_{2,1} + \frac{4}{3}ax_{1,1}^3x_{1,2}x_{2,1}^2 - \frac{1}{3}bx_{1,1}^3x_{1,2}x_{2,1}^2 \quad (\text{A.59})$$

$$\begin{aligned} & + \frac{8}{3}ax_{1,1}^2x_{1,2}^2x_{2,1}^2 - \frac{2}{3}bx_{1,1}^2x_{1,2}^2x_{2,1}^2 + \frac{4}{3}ax_{1,1}x_{1,2}^3x_{2,1}^2 - \frac{1}{3}bx_{1,1}x_{1,2}^3x_{2,1}^2 \\ & + bx_{1,1}^2x_{1,2}x_{2,1}^3 + bx_{1,1}x_{1,2}^2x_{2,1}^3 + bx_{1,1}^3x_{1,2}^2x_{2,2} + bx_{1,1}^2x_{1,2}^3x_{2,2} \\ & + \frac{8}{3}ax_{1,1}^3x_{1,2}x_{2,1}x_{2,2} - \frac{2}{3}bx_{1,1}^3x_{1,2}x_{2,1}x_{2,2} + \frac{16}{3}ax_{1,1}^2x_{1,2}^2x_{2,1}x_{2,2} - \frac{4}{3}bx_{1,1}^2x_{1,2}^2x_{2,1}x_{2,2} \\ & + \frac{8}{3}ax_{1,1}x_{1,2}^3x_{2,1}x_{2,2} - \frac{2}{3}bx_{1,1}x_{1,2}^3x_{2,1}x_{2,2} + bx_{1,1}^3x_{1,2}^2x_{2,2} + 4bx_{1,1}^2x_{1,2}x_{2,1}x_{2,2} \\ & + 4bx_{1,1}x_{1,2}^2x_{2,1}x_{2,2} + bx_{1,2}^3x_{2,1}x_{2,2} + \frac{4}{3}ax_{1,1}^2x_{1,2}^3x_{2,2} - \frac{1}{3}bx_{1,1}^2x_{1,2}^3x_{2,2} \\ & + \frac{8}{3}ax_{1,1}x_{1,2}x_{2,1}^3x_{2,2} - \frac{2}{3}bx_{1,1}x_{1,2}x_{2,1}^3x_{2,2} + \frac{4}{3}ax_{1,2}^2x_{2,1}^3x_{2,2} - \frac{1}{3}bx_{1,2}^2x_{2,1}^3x_{2,2} \\ & + \frac{4}{3}ax_{1,1}^3x_{1,2}x_{2,2}^2 - \frac{1}{3}bx_{1,1}^3x_{1,2}x_{2,2}^2 + \frac{8}{3}ax_{1,1}^2x_{1,2}^2x_{2,2}^2 - \frac{2}{3}bx_{1,1}^2x_{1,2}^2x_{2,2}^2 \\ & + \frac{4}{3}ax_{1,1}x_{1,2}^3x_{2,2}^2 - \frac{1}{3}bx_{1,1}x_{1,2}^3x_{2,2}^2 + bx_{1,1}^3x_{2,1}x_{2,2}^2 + 4bx_{1,1}^2x_{1,2}x_{2,1}x_{2,2}^2 \\ & + 4bx_{1,1}x_{1,2}^2x_{2,1}x_{2,2}^2 + bx_{1,2}^3x_{2,1}x_{2,2}^2 + \frac{8}{3}ax_{1,1}^2x_{1,2}^2x_{2,2}^2 - \frac{2}{3}bx_{1,1}^2x_{1,2}^2x_{2,2}^2 \\ & + \frac{16}{3}ax_{1,1}x_{1,2}x_{2,1}^2x_{2,2}^2 - \frac{4}{3}bx_{1,1}x_{1,2}x_{2,1}^2x_{2,2}^2 + \frac{8}{3}ax_{1,2}^2x_{2,1}^2x_{2,2}^2 - \frac{2}{3}bx_{1,2}^2x_{2,1}^2x_{2,2}^2 \\ & + bx_{1,1}x_{1,2}^3x_{2,2}^2 + bx_{1,2}^3x_{2,1}x_{2,2}^2 + bx_{1,1}^2x_{1,2}x_{2,2}^3 + bx_{1,1}x_{1,2}^2x_{2,2}^3 \\ & + \frac{4}{3}ax_{1,1}^2x_{2,1}x_{2,2}^3 - \frac{1}{3}bx_{1,1}^2x_{2,1}x_{2,2}^3 + \frac{8}{3}ax_{1,1}x_{1,2}x_{2,1}x_{2,2}^3 - \frac{2}{3}bx_{1,1}x_{1,2}x_{2,1}x_{2,2}^3 \\ & + \frac{4}{3}ax_{1,2}^2x_{2,1}x_{2,2}^3 - \frac{1}{3}bx_{1,2}^2x_{2,1}x_{2,2}^3 + bx_{1,1}x_{2,1}^2x_{2,2}^3 + bx_{1,2}x_{2,1}^2x_{2,2}^3, \end{aligned}$$

$$\begin{aligned} R = & -b^2x_{1,1}^3x_{1,2}^2x_{2,1} - b^2x_{1,1}^2x_{1,2}^3x_{2,1} - a^2x_{1,1}^3x_{1,2}x_{2,1}^2 - 2a^2x_{1,1}^2x_{1,2}^2x_{2,1}^2 \\ & - a^2x_{1,1}x_{1,2}^3x_{2,1}^2 - b^2x_{1,1}^2x_{1,2}x_{2,1}^3 - b^2x_{1,1}x_{1,2}^2x_{2,1}^3 - b^2x_{1,1}^3x_{1,2}^2x_{2,2} \end{aligned} \quad (\text{A.60})$$

$$\begin{aligned}
& -b^2 x_{1,1}^2 x_{1,2}^3 x_{2,2} - 2a^2 x_{1,1}^3 x_{1,2} x_{2,1} x_{2,2} - 4a^2 x_{1,1}^2 x_{1,2}^2 x_{2,1} x_{2,2} - 2a^2 x_{1,1} x_{1,2}^3 x_{2,1} x_{2,2} \\
& -b^2 x_{1,1}^3 x_{2,1}^2 x_{2,2} - 4b^2 x_{1,1}^2 x_{1,2} x_{2,1}^2 x_{2,2} - 4b^2 x_{1,1} x_{1,2}^2 x_{2,1}^2 x_{2,2} - b^2 x_{1,2}^3 x_{2,1}^2 x_{2,2} \\
& -a^2 x_{1,1}^2 x_{2,1}^3 x_{2,2} - 2a^2 x_{1,1} x_{1,2} x_{2,1}^3 x_{2,2} - a^2 x_{1,2}^2 x_{2,1}^3 x_{2,2} - a^2 x_{1,1}^3 x_{1,2} x_{2,2}^2 \\
& -2a^2 x_{1,1}^2 x_{1,2}^2 x_{2,2} - a^2 x_{1,1} x_{1,2}^3 x_{2,2}^2 - b^2 x_{1,1}^3 x_{2,1} x_{2,2}^2 - 4b^2 x_{1,1}^2 x_{1,2} x_{2,1} x_{2,2}^2 \\
& -4b^2 x_{1,1} x_{1,2}^2 x_{2,1} x_{2,2}^2 - b^2 x_{1,2}^3 x_{2,1} x_{2,2}^2 - 2a^2 x_{1,1}^2 x_{2,1}^2 x_{2,2}^2 - 4a^2 x_{1,1} x_{1,2} x_{2,1}^2 x_{2,2}^2 \\
& -2a^2 x_{1,2}^2 x_{2,1}^2 x_{2,2}^2 - b^2 x_{1,1} x_{2,1}^3 x_{2,2}^2 - b^2 x_{1,2} x_{2,1}^3 x_{2,2}^2 - b^2 x_{1,1}^2 x_{1,2} x_{2,2}^3 \\
& -b^2 x_{1,1} x_{1,2}^2 x_{2,2}^3 - a^2 x_{1,1}^2 x_{2,1} x_{2,2}^3 - 2a^2 x_{1,1} x_{1,2} x_{2,1} x_{2,2}^3 - a^2 x_{1,2}^2 x_{2,1} x_{2,2}^3 \\
& -b^2 x_{1,1} x_{2,1}^2 x_{2,2}^3 - b^2 x_{1,2} x_{2,1}^2 x_{2,2}^3.
\end{aligned}$$

A continuación, acotaremos a P por abajo y a $Q + R \frac{\log(n)}{n}$ por arriba para llegar a una cota para el exponente en la cota por Bernstein. En efecto, usando la expresión que obtuvimos para P tendremos que

$$P \geq (x_{1,1} x_{1,2} + x_{2,1} x_{2,2}) (a - b) \quad (\text{A.61})$$

y por otro lado acotando por desigualdad triangular y usando que cada proporción cumple que $x_{i,j} \leq 1$ tendremos que

$$\begin{aligned}
Q + R \frac{\log(n)}{n} & \leq \left[b + b + \frac{4}{3}a + \frac{1}{3}b + \frac{8}{3}a + \frac{2}{3}b + \frac{4}{3}a + \frac{1}{3}b + b + b + b + b \right. \\
& + \frac{8}{3}a + \frac{2}{3}b + \frac{16}{3}a + \frac{4}{3}b + \frac{8}{3}a + \frac{2}{3}b + b + 4b + 4b + b + \frac{4}{3}a + \frac{1}{3}b \\
& + \frac{8}{3}a + \frac{2}{3}b + \frac{4}{3}a + \frac{1}{3}b + \frac{4}{3} + \frac{1}{3}b + \frac{8}{3}a + \frac{2}{3}b + \frac{4}{3}a + \frac{1}{3}b + b + 4b \\
& + 4b + b + \frac{8}{3}a + \frac{2}{3}b + \frac{16}{3}a + \frac{4}{3}b + \frac{8}{3}a + \frac{2}{3}b + b + b + b + b \\
& \left. + \frac{4}{3}a + \frac{1}{3}b + \frac{8}{3}a + \frac{2}{3}b + \frac{4}{3}a + \frac{1}{3}b + b + b \right] \\
& + \left[b^2 + b^2 + a^2 + 2a^2 + a^2 + b^2 + b^2 + b^2 + b^2 + 2a^2 + 4a^2 + 2a^2 \right. \\
& + b^2 + 4b^2 + 4b^2 + b^2 + a^2 + 2a^2 + a^2 + a^2 + 2a^2 + a^2 + b^2 + 4b^2 \\
& + 4b^2 + b^2 + 2a^2 + 4a^2 + 2a^2 + b^2 + b^2 + b^2 + b^2 + a^2 + 2a^2 + a^2 \\
& \left. + b^2 + b^2 \right] \frac{\log(n)}{n} \\
& = \left[\frac{128}{3}a + \frac{128}{3}b \right] + \left[32a^2 + 32b^2 \right] \frac{\log(n)}{n},
\end{aligned} \quad (\text{A.62})$$

y es así que combinando ambas y por (A.57) llegaremos a que

$$\begin{aligned}
-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E} & \leq -\frac{1}{2} n \log(n) \alpha_1^2 \alpha_2^2 \frac{(x_{1,1} x_{1,2} + x_{2,1} x_{2,2})^2 (a - b)^2}{\left[\frac{128}{3}a + \frac{128}{3}b \right] + \left[32a^2 + 32b^2 \right] \frac{\log(n)}{n}} \\
& \leq -\log(n) (X_{1,1} + X_{2,1}) \frac{\frac{1}{2} \varepsilon^4 \alpha_1^2 \alpha_2^2 (a - b)^2}{\left[\frac{128}{3}a + \frac{128}{3}b \right] + \left[32a^2 + 32b^2 \right] \frac{\log(n)}{n}}.
\end{aligned} \quad (\text{A.63})$$

Aclaremos que en la última desigualdad usamos que $x_{1,1} x_{1,2} + x_{2,1} x_{2,2} \geq \varepsilon^2$, porque como estamos suponiendo en este caso ocurre alguno entre $x_{1,1}, x_{1,2} \geq \varepsilon$ o $x_{2,1}, x_{2,2} \geq \varepsilon$. Además, usamos que $1 \geq \beta_1$ para introducirlo en la cota y obtener $X_{1,1} + X_{2,1}$ a partir de $n\beta_1$.

Tendremos entonces que cada término en (A.53) asociado al caso 1. admite una cota de la forma

$$\begin{aligned} & \binom{n_1}{X_{1,1}} \binom{n_2}{X_{2,1}} 2 \exp\left(-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E}\right) \\ & \leq \exp\left(-\log(n)(X_{1,1} + X_{2,1}) \left(\frac{\log(n) + 1}{\log(n)} - \frac{\frac{1}{2}\varepsilon^4 \alpha_1^2 \alpha_2^2 (a-b)^2}{\left[\frac{128}{3}a + \frac{128}{3}b\right] + [32a^2 + 32b^2] \frac{\log(n)}{n}}\right)\right) \end{aligned} \quad (\text{A.64})$$

y para $a - b$ suficientemente grande demostramos lo que queríamos, notando que $\alpha_1, \alpha_2 = \Omega(1)$ están controlados por hipótesis.

Ahora veamos que valen los **casos 2., 3. y 4.** Nuevamente expresaremos P , Q y R en términos de las proporciones más cercanas a 0, y nos fijaremos en los monomios de orden 1 relevantes para entender el orden del cociente.

En el **caso 2.**,

$$P = a\alpha_1 x_{1,1} - \alpha_1 b x_{1,1} - a x_{1,1}^2 + b x_{1,1}^2 + a\alpha_2 x_{2,1} - \alpha_2 b x_{2,1} - a x_{2,1}^2 + b x_{2,1}^2, \quad (\text{A.65})$$

$$Q = \frac{4}{3} a \alpha_1^3 \alpha_2^2 x_{1,1} - \frac{1}{3} \alpha_1^3 \alpha_2^2 b x_{1,1} + \alpha_1^2 \alpha_2^3 b x_{1,1} - \frac{4}{3} a \alpha_1^2 \alpha_2^2 x_{1,1}^2 \quad (\text{A.66})$$

$$+ \alpha_1^3 \alpha_2 b x_{1,1}^2 + \frac{1}{3} \alpha_1^2 \alpha_2^2 b x_{1,1}^2 - \alpha_1 \alpha_2^3 b x_{1,1}^2 - 2 \alpha_1^2 \alpha_2 b x_{1,1}^3$$

$$+ \alpha_1 \alpha_2 b x_{1,1}^4 + \frac{4}{3} a \alpha_1^2 \alpha_2^3 x_{2,1} + \alpha_1^3 \alpha_2^2 b x_{2,1} - \frac{1}{3} \alpha_1^2 \alpha_2^3 b x_{2,1}$$

$$- 2 \alpha_1^2 \alpha_2^2 b x_{1,1} x_{2,1} + 2 \alpha_1 \alpha_2^2 b x_{1,1}^2 x_{2,1} - \frac{4}{3} a \alpha_1^2 \alpha_2^2 x_{2,1}^2 - \alpha_1^3 \alpha_2 b x_{2,1}^2$$

$$+ \frac{1}{3} \alpha_1^2 \alpha_2^2 b x_{2,1}^2 + \alpha_1 \alpha_2^3 b x_{2,1}^2 + 2 \alpha_1^2 \alpha_2 b x_{1,1} x_{2,1}^2 - 2 \alpha_1 \alpha_2 b x_{1,1}^2 x_{2,1}^2$$

$$- 2 \alpha_1 \alpha_2^2 b x_{2,1}^3 + \alpha_1 \alpha_2 b x_{2,1}^4,$$

$$R = -a^2 \alpha_1^3 \alpha_2^2 x_{1,1} - \alpha_1^2 \alpha_2^3 b^2 x_{1,1} + a^2 \alpha_1^2 \alpha_2^2 x_{1,1}^2 - \alpha_1^3 \alpha_2 b^2 x_{1,1}^2 \quad (\text{A.67})$$

$$+ \alpha_1 \alpha_2^3 b^2 x_{1,1}^2 + 2 \alpha_1^2 \alpha_2 b^2 x_{1,1}^3 - \alpha_1 \alpha_2 b^2 x_{1,1}^4 - a^2 \alpha_1^2 \alpha_2^3 x_{2,1}$$

$$- \alpha_1^3 \alpha_2^2 b^2 x_{2,1} + 2 \alpha_1^2 \alpha_2^2 b^2 x_{1,1} x_{2,1} - 2 \alpha_1 \alpha_2^2 b^2 x_{1,1}^2 x_{2,1} + a^2 \alpha_1^2 \alpha_2^2 x_{2,1}^2$$

$$+ \alpha_1^3 \alpha_2 b^2 x_{2,1}^2 - \alpha_1 \alpha_2^3 b^2 x_{2,1}^2 - 2 \alpha_1^2 \alpha_2 b^2 x_{1,1} x_{2,1}^2 + 2 \alpha_1 \alpha_2 b^2 x_{1,1}^2 x_{2,1}^2$$

$$+ 2 \alpha_1 \alpha_2^2 b^2 x_{2,1}^3 - \alpha_1 \alpha_2 b^2 x_{2,1}^4, \quad (\text{A.68})$$

y entonces

$$P_1 = a\alpha_1 x_{1,1} - \alpha_1 b x_{1,1} + a\alpha_2 x_{2,1} - \alpha_2 b x_{2,1} \quad (\text{A.69})$$

$$= [(a-b)\alpha_1] x_{1,1} + [(a-b)\alpha_2] x_{2,1}$$

$$Q_1 = \frac{4}{3} a \alpha_1^3 \alpha_2^2 x_{1,1} - \frac{1}{3} \alpha_1^3 \alpha_2^2 b x_{1,1} + \alpha_1^2 \alpha_2^3 b x_{1,1} + \frac{4}{3} a \alpha_1^2 \alpha_2^3 x_{2,1} + \alpha_1^3 \alpha_2^2 b x_{2,1} - \frac{1}{3} \alpha_1^2 \alpha_2^3 b x_{2,1}$$

$$= \left[\alpha_1^2 \alpha_2^2 \left(\frac{1}{3} \alpha_1 (a-b) + (a\alpha_1 + b\alpha_2) \right) \right] x_{1,1} + \left[\alpha_1^2 \alpha_2^2 \left(\frac{1}{3} \alpha_2 (a-b) + (a\alpha_2 + b\alpha_1) \right) \right] x_{2,1}.$$

Como P_1 y Q_1 tienen coeficientes no negativos, y se verificarán todas las hipótesis podemos usar el Lema (4.2.4) para acotar, fija una constante $0 < C < 1$ tan cercana a 1 como

querramos y para n suficientemente grandes, de la siguiente manera:

$$\begin{aligned} \frac{P^2}{Q + R \frac{\log(n)}{n}} &\geq C \frac{P_1^2}{Q_1} & (A.70) \\ &= C \frac{([(a-b)\alpha_1] x_{1,1} + [(a-b)\alpha_2] x_{2,1})^2}{\left[\alpha_1^2 \alpha_2^2 \left(\frac{1}{3} \alpha_1 (a-b) + (a\alpha_1 + b\alpha_2) \right) \right] x_{1,1} + \left[\alpha_1^2 \alpha_2^2 \left(\frac{1}{3} \alpha_2 (a-b) + (a\alpha_2 + b\alpha_1) \right) \right] x_{2,1}} \\ &\geq C \frac{[(a-b)\alpha_1]^2}{\alpha_1^2 \alpha_2^2 \left(\frac{1}{3} \alpha_2 (a-b) + (a\alpha_2 + b\alpha_1) \right)} (x_{1,1} + x_{2,1}). \end{aligned}$$

Aclaremos que para la segunda desigualdad usamos el Lema 4.2.5 que nos permite simplificar a una expresión más agradable, notando que como $\alpha_2 \geq \alpha_1 > 0$ y $a > b$ es que $(a-b)\alpha_2 \geq (a-b)\alpha_1$ y además

$$(a\alpha_2 + b\alpha_1) - (a\alpha_1 + b\alpha_2) = (a-b)(\alpha_2 - \alpha_1) \geq 0. \quad (A.71)$$

De esta manera, obtenemos una cota para el exponente en la cota por Bernstein

$$\begin{aligned} -\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E} &\leq -\frac{1}{2} n \log(n) \alpha_1^2 \alpha_2^2 C \frac{[(a-b)\alpha_1]^2}{\alpha_1^2 \alpha_2^2 \left(\frac{1}{3} \alpha_2 (a-b) + (a\alpha_2 + b\alpha_1) \right)} (x_{1,1} + x_{2,1}) & (A.72) \\ &\leq -\log(n) (X_{1,1} + X_{2,1}) \frac{\frac{1}{2} C [(a-b)\alpha_1]^2}{\frac{1}{3} \alpha_2 (a-b) + (a\alpha_2 + b\alpha_1)} \end{aligned}$$

que podremos emplear para acotar cada término asociado al caso 2. en la suma (A.53) de la siguiente manera:

$$\begin{aligned} &\binom{n_1}{X_{1,1}} \binom{n_2}{X_{2,1}} 2 \exp\left(-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E}\right) & (A.73) \\ &\leq \exp\left(-\log(n) (X_{1,1} + X_{2,1}) \left(\frac{\log(n) + 1}{\log(n)} - \frac{\frac{1}{2} C [(a-b)\alpha_1]^2}{\frac{1}{3} \alpha_2 (a-b) + (a\alpha_2 + b\alpha_1)}\right)\right). \end{aligned}$$

Luego, como $\alpha_1, \alpha_2 = \Theta(1)$, para $a-b$ suficientemente grande tendremos que efectivamente cada término es de orden $o(n^{-2})$.

Los **casos 3. y 4.** resultan en cuentas idénticas, salvo que hay que reemplazar $x_{1,1}$ por $x_{1,2}$ en el caso 3. y por $x_{1,2}$ por en el caso 4.

Ahora queda mostrar que pasa cuando $m_1 > n_1$, o $\beta_1 > \alpha_1$.

Comenzaremos mostrando el **caso 5.**, para el cual procederemos análogamente al caso 1.. En primer lugar, definimos P , Q y R polinomios en $x_{i,j}$ como en (A.56) y esta vez los podemos expresar de la siguiente manera:

$$P = ax_{1,1}x_{1,2} - bx_{1,1}x_{1,2} + ax_{2,1}x_{2,2} - bx_{2,1}x_{2,2}, \quad (A.74)$$

$$\begin{aligned} Q &= \frac{1}{3} ax_{1,1}^3 x_{1,2}^2 x_{2,1} + \frac{2}{3} bx_{1,1}^3 x_{1,2}^2 x_{2,1} + \frac{1}{3} ax_{1,1}^2 x_{1,2}^3 x_{2,1} + \frac{2}{3} bx_{1,1}^2 x_{1,2}^3 x_{2,1} & (A.75) \\ &\quad + ax_{1,1}^3 x_{1,2} x_{2,1}^2 + \frac{7}{3} ax_{1,1}^2 x_{1,2}^2 x_{2,1}^2 - \frac{1}{3} bx_{1,1}^2 x_{1,2}^2 x_{2,1}^2 + \frac{4}{3} ax_{1,1} x_{1,2}^3 x_{2,1}^2 \\ &\quad - \frac{1}{3} bx_{1,1} x_{1,2}^3 x_{2,1}^2 + bx_{1,1}^2 x_{1,2} x_{2,1}^3 + bx_{1,1} x_{1,2}^2 x_{2,1}^3 + \frac{1}{3} ax_{1,1}^3 x_{1,2}^2 x_{2,2} \end{aligned}$$

$$\begin{aligned}
& + \frac{2}{3}bx_{1,1}^3x_{1,2}^2x_{2,2} + \frac{1}{3}ax_{1,1}^2x_{1,2}^3x_{2,2} + \frac{2}{3}bx_{1,1}^2x_{1,2}^3x_{2,2} + \frac{7}{3}ax_{1,1}^3x_{1,2}x_{2,1}x_{2,2} \\
& - \frac{1}{3}bx_{1,1}^3x_{1,2}x_{2,1}x_{2,2} + \frac{14}{3}ax_{1,1}^2x_{1,2}^2x_{2,1}x_{2,2} - \frac{2}{3}bx_{1,1}^2x_{1,2}^2x_{2,1}x_{2,2} + \frac{7}{3}ax_{1,1}x_{1,2}^3x_{2,1}x_{2,2} \\
& - \frac{1}{3}bx_{1,1}x_{1,2}^3x_{2,1}x_{2,2} + bx_{1,1}^3x_{2,1}^2x_{2,2} + \frac{2}{3}ax_{1,1}^2x_{1,2}x_{2,1}^2x_{2,2} + \frac{10}{3}bx_{1,1}^2x_{1,2}x_{2,1}^2x_{2,2} \\
& + \frac{2}{3}ax_{1,1}x_{1,2}^2x_{2,1}^2x_{2,2} + \frac{10}{3}bx_{1,1}x_{1,2}^2x_{2,1}^2x_{2,2} + bx_{1,2}^3x_{2,1}^2x_{2,2} + ax_{1,1}^2x_{2,1}^3x_{2,2} \\
& + \frac{7}{3}ax_{1,1}x_{1,2}x_{2,1}^3x_{2,2} - \frac{1}{3}bx_{1,1}x_{1,2}x_{2,1}^3x_{2,2} + \frac{4}{3}ax_{1,2}^2x_{2,1}^3x_{2,2} - \frac{1}{3}bx_{1,2}^2x_{2,1}^3x_{2,2} \\
& + \frac{4}{3}ax_{1,1}^3x_{1,2}x_{2,2}^2 - \frac{1}{3}bx_{1,1}^3x_{1,2}x_{2,2}^2 + \frac{7}{3}ax_{1,1}^2x_{1,2}^2x_{2,2}^2 - \frac{1}{3}bx_{1,1}^2x_{1,2}^2x_{2,2}^2 \\
& + ax_{1,1}x_{1,2}^3x_{2,2}^2 + bx_{1,1}^3x_{2,1}x_{2,2}^2 + \frac{2}{3}ax_{1,1}^2x_{1,2}x_{2,1}x_{2,2}^2 + \frac{10}{3}bx_{1,1}^2x_{1,2}x_{2,1}x_{2,2}^2 \\
& + \frac{2}{3}ax_{1,1}x_{1,2}^2x_{2,1}x_{2,2}^2 + \frac{10}{3}bx_{1,1}x_{1,2}^2x_{2,1}x_{2,2}^2 + bx_{1,2}^3x_{2,1}x_{2,2}^2 + \frac{7}{3}ax_{1,1}^2x_{2,1}^2x_{2,2}^2 \\
& - \frac{1}{3}bx_{1,1}^2x_{2,1}^2x_{2,2}^2 + \frac{14}{3}ax_{1,1}x_{1,2}x_{2,1}^2x_{2,2}^2 - \frac{2}{3}bx_{1,1}x_{1,2}x_{2,1}^2x_{2,2}^2 + \frac{7}{3}ax_{1,2}^2x_{2,1}^2x_{2,2}^2 \\
& - \frac{1}{3}bx_{1,2}^2x_{2,1}^2x_{2,2}^2 + \frac{1}{3}ax_{1,1}x_{2,1}^3x_{2,2}^2 + \frac{2}{3}bx_{1,1}x_{2,1}^3x_{2,2}^2 + \frac{1}{3}ax_{1,2}x_{2,1}^3x_{2,2}^2 \\
& + \frac{2}{3}bx_{1,2}x_{2,1}^3x_{2,2}^2 + bx_{1,1}^2x_{1,2}x_{2,1}^3x_{2,2}^2 + bx_{1,1}x_{1,2}^2x_{2,1}^3x_{2,2}^2 + \frac{4}{3}ax_{1,1}^2x_{2,1}x_{2,2}^3 \\
& - \frac{1}{3}bx_{1,1}^2x_{2,1}x_{2,2}^3 + \frac{7}{3}ax_{1,1}x_{1,2}x_{2,1}x_{2,2}^3 - \frac{1}{3}bx_{1,1}x_{1,2}x_{2,1}x_{2,2}^3 + ax_{1,2}^2x_{2,1}x_{2,2}^3 \\
& + \frac{1}{3}ax_{1,1}x_{2,1}^2x_{2,2}^3 + \frac{2}{3}bx_{1,1}x_{2,1}^2x_{2,2}^3 + \frac{1}{3}ax_{1,2}x_{2,1}^2x_{2,2}^3 + \frac{2}{3}bx_{1,2}x_{2,1}^2x_{2,2}^3, \\
R = & -b^2x_{1,1}^3x_{1,2}^2x_{2,1} - b^2x_{1,1}^2x_{1,2}^3x_{2,1} - a^2x_{1,1}^3x_{1,2}x_{2,1}^2 - 2a^2x_{1,1}^2x_{1,2}^2x_{2,1}^2 \\
& - a^2x_{1,1}x_{1,2}^3x_{2,1}^2 - b^2x_{1,1}^2x_{1,2}x_{2,1}^3 - b^2x_{1,1}x_{1,2}^2x_{2,1}^3 - b^2x_{1,1}^3x_{1,2}^2x_{2,2} \\
& - b^2x_{1,1}^2x_{1,2}^3x_{2,2} - 2a^2x_{1,1}^3x_{1,2}x_{2,1}x_{2,2} - 4a^2x_{1,1}^2x_{1,2}^2x_{2,1}x_{2,2} - 2a^2x_{1,1}x_{1,2}^3x_{2,1}x_{2,2} \\
& - b^2x_{1,1}^3x_{2,1}^2x_{2,2} - 4b^2x_{1,1}^2x_{1,2}x_{2,1}^2x_{2,2} - 4b^2x_{1,1}x_{1,2}^2x_{2,1}^2x_{2,2} - b^2x_{1,2}^3x_{2,1}^2x_{2,2} \\
& - a^2x_{1,1}^2x_{2,1}^3x_{2,2} - 2a^2x_{1,1}x_{1,2}x_{2,1}^3x_{2,2} - a^2x_{1,2}^2x_{2,1}^3x_{2,2} - a^2x_{1,1}^3x_{1,2}x_{2,2}^2 \\
& - 2a^2x_{1,1}^2x_{1,2}^2x_{2,2}^2 - a^2x_{1,1}x_{1,2}^3x_{2,2}^2 - b^2x_{1,1}^3x_{2,1}x_{2,2}^2 - 4b^2x_{1,1}^2x_{1,2}x_{2,1}x_{2,2}^2 \\
& - 4b^2x_{1,1}x_{1,2}^2x_{2,1}x_{2,2}^2 - b^2x_{1,2}^3x_{2,1}x_{2,2}^2 - 2a^2x_{1,1}^2x_{2,1}^2x_{2,2}^2 - 4a^2x_{1,1}x_{1,2}x_{2,1}^2x_{2,2}^2 \\
& - 2a^2x_{1,2}^2x_{2,1}^2x_{2,2}^2 - b^2x_{1,1}x_{2,1}^3x_{2,2}^2 - b^2x_{1,2}x_{2,1}^3x_{2,2}^2 - b^2x_{1,1}^2x_{1,2}x_{2,2}^3 \\
& - b^2x_{1,1}x_{1,2}^2x_{2,2}^3 - a^2x_{1,1}^2x_{2,1}x_{2,2}^3 - 2a^2x_{1,1}x_{1,2}x_{2,1}x_{2,2}^3 - a^2x_{1,2}^2x_{2,1}x_{2,2}^3 \\
& - b^2x_{1,1}x_{2,1}^2x_{2,2}^3 - b^2x_{1,2}x_{2,1}^2x_{2,2}^3.
\end{aligned} \tag{A.76}$$

Ahora, acotamos a P por abajo y a $Q + R \frac{\log(n)}{n}$ por arriba para obtener una cota para el exponente. En primer lugar,

$$P \geq (x_{1,1}x_{1,2} + x_{2,1}x_{2,2})(a - b), \tag{A.77}$$

y por otro lado acotando por desigualdad triangular y usando que $x_{i,j} \leq 1$ obtenemos que

$$\begin{aligned}
Q + R \frac{\log(n)}{n} \leq & \left[\frac{1}{3}a + \frac{2}{3}b + \frac{1}{3}a + \frac{2}{3}b + a + \frac{7}{3}a + \frac{1}{3}b + \frac{4}{3}a + \frac{1}{3}b + b + b + \frac{1}{3}a \right. \\
& \left. + \frac{2}{3}b + \frac{1}{3}a + \frac{2}{3}b + \frac{7}{3}a + \frac{1}{3}b + \frac{14}{3}a + \frac{2}{3}b + \frac{7}{3}a + \frac{1}{3}b + b + \frac{2}{3}a + \frac{10}{3}b \right]
\end{aligned} \tag{A.78}$$

$$\begin{aligned}
& + \frac{2}{3}a + \frac{10}{3}b + b + a + \frac{7}{3}a + \frac{1}{3}b + \frac{4}{3}a + \frac{1}{3}b + \frac{4}{3}a + \frac{1}{3}b + \frac{7}{3}a + \frac{1}{3}b \\
& + a + b + \frac{2}{3}a + \frac{10}{3}b + \frac{2}{3}a + \frac{10}{3}b + b + \frac{7}{3}a + \frac{1}{3}b + \frac{14}{3}a + \frac{2}{3}b + \frac{7}{3}a \\
& + \frac{1}{3}b + \frac{1}{3}a + \frac{2}{3}b + \frac{1}{3}a + \frac{2}{3}b + b + b + \frac{4}{3}a + \frac{1}{3}b + \frac{7}{3}a + \frac{1}{3}b + a \\
& + \left. \frac{1}{3}a + \frac{2}{3}b + \frac{1}{3}a + \frac{2}{3}b \right] \\
& + \left[b^2 + b^2 + a^2 + 2a^2 + a^2 + b^2 + b^2 + b^2 + b^2 + 2a^2 + 4a^2 + 2a^2 \right. \\
& + b^2 + 4b^2 + 4b^2 + b^2 + a^2 + 2a^2 + a^2 + a^2 + 2a^2 + a^2 + b^2 + 4b^2 \\
& + 4b^2 + b^2 + 2a^2 + 4a^2 + 2a^2 + b^2 + b^2 + b^2 + b^2 + a^2 + 2a^2 + a^2 \\
& \left. + b^2 + b^2 \right] \frac{\log(n)}{n} \\
& = \left[\frac{128}{3}a + 32b \right] + \left[32a^2 + 24b^2 \right] \frac{\log(n)}{n}.
\end{aligned}$$

De esta manera, usando los mismos argumentos que en (A.63) podemos acotar el exponente en las cotas por Bernstein como a continuación:

$$\begin{aligned}
-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E} & \leq -\frac{1}{2} n \log(n) \alpha_1^2 \alpha_2^2 \frac{(x_{1,1}x_{1,2} + x_{2,1}x_{2,2})^2 (a-b)^2}{\left[\frac{128}{3}a + 32b \right] + \left[32a^2 + 24b^2 \right] \frac{\log(n)}{n}} \quad (\text{A.79}) \\
& \leq -\log(n)(X_{1,1} + X_{2,1}) \frac{\frac{1}{2}\varepsilon^4 \alpha_1^2 \alpha_2^2 (a-b)^2}{\left[\frac{128}{3}a + 32b \right] + \left[32a^2 + 24b^2 \right] \frac{\log(n)}{n}}.
\end{aligned}$$

Más aún, esta cota nos permitirá controlar cada término asociado al caso 5. en (A.53) pues vale que

$$\begin{aligned}
& \binom{n_1}{X_{1,1}} \binom{n_2}{X_{2,1}} 2 \exp\left(-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E}\right) \quad (\text{A.80}) \\
& \leq \exp\left(-\log(n)(X_{1,1} + X_{2,1}) \left(\frac{\log(n) + 1}{\log(n)} - \frac{\frac{1}{2}\varepsilon^4 \alpha_1^2 \alpha_2^2 (a-b)^2}{\left[\frac{128}{3}a + 32b \right] + \left[32a^2 + 24b^2 \right] \frac{\log(n)}{n}} \right)\right),
\end{aligned}$$

donde como $\alpha_1, \alpha_2 = \Theta(1)$ tendremos que para $a - b$ suficientemente grandes esta expresión será $o(n^{-2})$.

Ahora veamos que valen los **casos 6., 7. y 8.**, para los cuales procederemos como en el caso 2., expresando P , Q y R en términos de las proporciones más cercanas a 0 para después estudiar con cuidado los monomios de orden 1 relevantes para entender el orden de nuestras cotas.

En el **caso 6.**,

$$P = a\alpha_1 x_{1,1} - \alpha_1 b x_{1,1} - a x_{1,1}^2 + b x_{1,1}^2 + a\alpha_2 x_{2,1} - \alpha_2 b x_{2,1} - a x_{2,1}^2 + b x_{2,1}^2, \quad (\text{A.81})$$

$$\begin{aligned}
Q & = a\alpha_1^3 \alpha_2^2 x_{1,1} + \alpha_1^2 \alpha_2^3 b x_{1,1} + \frac{1}{3} a \alpha_1^3 \alpha_2 x_{1,1}^2 - \frac{2}{3} a \alpha_1^2 \alpha_2^2 x_{1,1}^2 \quad (\text{A.82}) \\
& + \frac{2}{3} \alpha_1^3 \alpha_2 b x_{1,1}^2 - \frac{1}{3} \alpha_1^2 \alpha_2^2 b x_{1,1}^2 - \alpha_1 \alpha_2^3 b x_{1,1}^2 - \frac{2}{3} a \alpha_1^2 \alpha_2 x_{1,1}^3 \\
& - \frac{1}{3} a \alpha_1 \alpha_2^2 x_{1,1}^3 - \frac{4}{3} \alpha_1^2 \alpha_2 b x_{1,1}^3 + \frac{1}{3} \alpha_1 \alpha_2^2 b x_{1,1}^3 + \frac{1}{3} a \alpha_1 \alpha_2 x_{1,1}^4
\end{aligned}$$

$$\begin{aligned}
& + \frac{2}{3}\alpha_1\alpha_2bx_{1,1}^4 + a\alpha_1^2\alpha_2^3x_{2,1} + \alpha_1^3\alpha_2^2bx_{2,1} + \frac{1}{3}a\alpha_1^3\alpha_2x_{1,1}x_{2,1} \\
& + \frac{2}{3}a\alpha_1^2\alpha_2^2x_{1,1}x_{2,1} + \frac{1}{3}a\alpha_1\alpha_2^3x_{1,1}x_{2,1} - \frac{1}{3}\alpha_1^3\alpha_2bx_{1,1}x_{2,1} - \frac{8}{3}\alpha_1^2\alpha_2^2bx_{1,1}x_{2,1} \\
& - \frac{1}{3}\alpha_1\alpha_2^3bx_{1,1}x_{2,1} - a\alpha_1^2\alpha_2x_{1,1}^2x_{2,1} - \frac{2}{3}a\alpha_1\alpha_2^2x_{1,1}^2x_{2,1} + \alpha_1^2\alpha_2bx_{1,1}^2x_{2,1} \\
& + \frac{8}{3}\alpha_1\alpha_2^2bx_{1,1}^2x_{2,1} + \frac{2}{3}a\alpha_1\alpha_2x_{1,1}^3x_{2,1} - \frac{2}{3}\alpha_1\alpha_2bx_{1,1}^3x_{2,1} - \frac{2}{3}a\alpha_1^2\alpha_2^2x_{2,1}^2 \\
& + \frac{1}{3}a\alpha_1\alpha_2^3x_{2,1}^2 - \alpha_1^3\alpha_2bx_{2,1}^2 - \frac{1}{3}\alpha_1^2\alpha_2^2bx_{2,1}^2 + \frac{2}{3}\alpha_1\alpha_2^3bx_{2,1}^2 \\
& - \frac{2}{3}a\alpha_1^2\alpha_2x_{1,1}x_{2,1}^2 - a\alpha_1\alpha_2^2x_{1,1}x_{2,1}^2 + \frac{8}{3}\alpha_1^2\alpha_2bx_{1,1}x_{2,1}^2 + \alpha_1\alpha_2^2bx_{1,1}x_{2,1}^2 \\
& + \frac{2}{3}a\alpha_1\alpha_2x_{1,1}^2x_{2,1}^2 - \frac{8}{3}\alpha_1\alpha_2bx_{1,1}^2x_{2,1}^2 - \frac{1}{3}a\alpha_1^2\alpha_2x_{2,1}^3 - \frac{2}{3}a\alpha_1\alpha_2^2x_{2,1}^3 \\
& + \frac{1}{3}\alpha_1^2\alpha_2bx_{2,1}^3 - \frac{4}{3}\alpha_1\alpha_2^2bx_{2,1}^3 + \frac{2}{3}a\alpha_1\alpha_2x_{1,1}x_{2,1}^3 \\
& - \frac{2}{3}\alpha_1\alpha_2bx_{1,1}x_{2,1}^3 + \frac{1}{3}a\alpha_1\alpha_2x_{2,1}^4 + \frac{2}{3}\alpha_1\alpha_2bx_{2,1}^4, \tag{A.83}
\end{aligned}$$

$$R = -a^2\alpha_1^3\alpha_2^2x_{1,1} - \alpha_1^2\alpha_2^3b^2x_{1,1} + a^2\alpha_1^2\alpha_2^2x_{1,1}^2 - \alpha_1^3\alpha_2b^2x_{1,1}^2 \tag{A.84}$$

$$\begin{aligned}
& + \alpha_1\alpha_2^3b^2x_{1,1}^2 + 2\alpha_1^2\alpha_2b^2x_{1,1}^3 - \alpha_1\alpha_2b^2x_{1,1}^4 - a^2\alpha_1^2\alpha_2^3x_{2,1} \\
& - \alpha_1^3\alpha_2^2b^2x_{2,1} + 2\alpha_1^2\alpha_2^2b^2x_{1,1}x_{2,1} - 2\alpha_1\alpha_2^2b^2x_{1,1}^2x_{2,1} + a^2\alpha_1^2\alpha_2^2x_{2,1}^2 \\
& + \alpha_1^3\alpha_2b^2x_{2,1}^2 - \alpha_1\alpha_2^3b^2x_{2,1}^2 - 2\alpha_1^2\alpha_2b^2x_{1,1}x_{2,1}^2 + 2\alpha_1\alpha_2b^2x_{1,1}^2x_{2,1}^2 \\
& + 2\alpha_1\alpha_2^2b^2x_{2,1}^3 - \alpha_1\alpha_2b^2x_{2,1}^4, \tag{A.85}
\end{aligned}$$

y entonces

$$P_1 = a\alpha_1x_{1,1} - \alpha_1bx_{1,1} + a\alpha_2x_{2,1} - \alpha_2bx_{2,1} \tag{A.86}$$

$$= [\alpha_1(a-b)]x_{1,1} + [\alpha_2(a-b)]x_{2,1},$$

$$\begin{aligned}
Q_1 & = a\alpha_1^3\alpha_2^2x_{1,1} + \alpha_1^2\alpha_2^3bx_{1,1} + a\alpha_1^2\alpha_2^3x_{2,1} + \alpha_1^3\alpha_2^2bx_{2,1} \\
& = [\alpha_1^2\alpha_2^2(a\alpha_1 + b\alpha_2)]x_{1,1} + [\alpha_1^2\alpha_2^2(a\alpha_2 + b\alpha_1)]x_{2,1}.
\end{aligned}$$

Usando los Lemas 4.2.4 y 4.2.5 podremos concluir que para cada constante $0 < C < 1$ si n es suficientemente grande valdrá que

$$\frac{P^2}{Q + R\frac{\log(n)}{n}} \geq C \frac{P_1^2}{Q_1} \tag{A.87}$$

$$\begin{aligned}
& = C \frac{([\alpha_1(a-b)]x_{1,1} + [\alpha_2(a-b)]x_{2,1})^2}{[\alpha_1^2\alpha_2^2(a\alpha_1 + b\alpha_2)]x_{1,1} + [\alpha_1^2\alpha_2^2(a\alpha_2 + b\alpha_1)]x_{2,1}} \\
& \geq C \frac{[\alpha_1(a-b)]^2}{\alpha_1^2\alpha_2^2(a\alpha_2 + b\alpha_1)}(x_{1,1} + x_{2,1}); \tag{A.88}
\end{aligned}$$

así podremos también acotar el exponente en cada cota por Bernstein

$$\begin{aligned}
-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E} & \leq -\frac{1}{2} n \log(n) \alpha_1^2 \alpha_2^2 C \frac{[\alpha_1(a-b)]^2}{\alpha_1^2 \alpha_2^2 (a\alpha_2 + b\alpha_1)} (x_{1,1} + x_{2,1}) \\
& \leq -\log(n) (X_{1,1} + X_{2,1}) \frac{\frac{1}{2} C [\alpha_1(a-b)]^2}{a\alpha_2 + b\alpha_1}, \tag{A.89}
\end{aligned}$$

y usar a su vez esta cota para acotar cada término asociado al caso 6. en [A.53](#):

$$\binom{n_1}{X_{1,1}} \binom{n_2}{X_{2,1}} 2 \exp\left(-\frac{1}{2} \frac{E^2}{V + \frac{M}{3}E}\right) \quad (\text{A.90})$$

$$\leq \exp\left(-\log(n)(X_{1,1} + X_{2,1}) \left(\frac{\log(n) + 1}{\log(n)} - \frac{\frac{1}{2}C[\alpha_1(a-b)]^2}{a\alpha_2 + b\alpha_1}\right)\right). \quad (\text{A.91})$$

Una vez más, como $\alpha_1, \alpha_2 = \Theta(1)$ tendremos que para $a - b$ suficientemente grande todos los términos serán $o(n^{-2})$ como queríamos.

En el **caso 7.**,

$$P = a\alpha_1 x_{1,1} - \alpha_1 b x_{1,1} - a x_{1,1}^2 + b x_{1,1}^2 + a\alpha_2 x_{2,2} - \alpha_2 b x_{2,2} - a x_{2,2}^2 + b x_{2,2}^2, \quad (\text{A.92})$$

$$Q = \frac{4}{3} a \alpha_1^3 \alpha_2^2 x_{1,1} - \frac{1}{3} \alpha_1^3 \alpha_2^2 b x_{1,1} + \alpha_1^2 \alpha_2^3 b x_{1,1} + \frac{1}{3} a \alpha_1^3 \alpha_2 x_{1,1}^2 \quad (\text{A.93})$$

$$- \frac{5}{3} a \alpha_1^2 \alpha_2^2 x_{1,1}^2 + \frac{2}{3} \alpha_1^3 \alpha_2 b x_{1,1}^2 + \frac{2}{3} \alpha_1^2 \alpha_2^2 b x_{1,1}^2 - \alpha_1 \alpha_2^3 b x_{1,1}^2$$

$$- \frac{2}{3} a \alpha_1^2 \alpha_2 x_{1,1}^3 + \frac{1}{3} a \alpha_1 \alpha_2^2 x_{1,1}^3 - \frac{4}{3} \alpha_1^2 \alpha_2 b x_{1,1}^3 - \frac{1}{3} \alpha_1 \alpha_2^2 b x_{1,1}^3$$

$$+ \frac{1}{3} a \alpha_1 \alpha_2 x_{1,1}^4 + \frac{2}{3} \alpha_1 \alpha_2 b x_{1,1}^4 + \frac{4}{3} a \alpha_1^2 \alpha_2^3 x_{2,2} + \alpha_1^3 \alpha_2^2 b x_{2,2}$$

$$- \frac{1}{3} \alpha_1^2 \alpha_2^3 b x_{2,2} - \frac{1}{3} a \alpha_1^3 \alpha_2 x_{1,1} x_{2,2} + \frac{2}{3} a \alpha_1^2 \alpha_2^2 x_{1,1} x_{2,2} - \frac{1}{3} a \alpha_1 \alpha_2^3 x_{1,1} x_{2,2}$$

$$+ \frac{1}{3} \alpha_1^3 \alpha_2 b x_{1,1} x_{2,2} - \frac{8}{3} \alpha_1^2 \alpha_2^2 b x_{1,1} x_{2,2} + \frac{1}{3} \alpha_1 \alpha_2^3 b x_{1,1} x_{2,2} + a \alpha_1^2 \alpha_2 x_{1,1}^2 x_{2,2}$$

$$- \frac{2}{3} a \alpha_1 \alpha_2^2 x_{1,1}^2 x_{2,2} - \alpha_1^2 \alpha_2 b x_{1,1}^2 x_{2,2} + \frac{8}{3} \alpha_1 \alpha_2^2 b x_{1,1}^2 x_{2,2} - \frac{2}{3} a \alpha_1 \alpha_2 x_{1,1}^3 x_{2,2}$$

$$+ \frac{2}{3} \alpha_1 \alpha_2 b x_{1,1}^3 x_{2,2} - \frac{5}{3} a \alpha_1^2 \alpha_2^2 x_{2,2}^2 + \frac{1}{3} a \alpha_1 \alpha_2^3 x_{2,2}^2 - \alpha_1^3 \alpha_2 b x_{2,2}^2$$

$$+ \frac{2}{3} \alpha_1^2 \alpha_2^2 b x_{2,2}^2 + \frac{2}{3} \alpha_1 \alpha_2^3 b x_{2,2}^2 - \frac{2}{3} a \alpha_1^2 \alpha_2 x_{1,1} x_{2,2}^2 + a \alpha_1 \alpha_2^2 x_{1,1} x_{2,2}^2$$

$$+ \frac{8}{3} \alpha_1^2 \alpha_2 b x_{1,1} x_{2,2}^2 - \alpha_1 \alpha_2^2 b x_{1,1} x_{2,2}^2 + \frac{2}{3} a \alpha_1 \alpha_2 x_{1,1}^2 x_{2,2}^2 - \frac{8}{3} \alpha_1 \alpha_2 b x_{1,1}^2 x_{2,2}^2$$

$$+ \frac{1}{3} a \alpha_1^2 \alpha_2 x_{2,2}^3 - \frac{2}{3} a \alpha_1 \alpha_2^2 x_{2,2}^3 - \frac{1}{3} \alpha_1^2 \alpha_2 b x_{2,2}^3 - \frac{4}{3} \alpha_1 \alpha_2^2 b x_{2,2}^3$$

$$- \frac{2}{3} a \alpha_1 \alpha_2 x_{1,1} x_{2,2}^3 + \frac{2}{3} \alpha_1 \alpha_2 b x_{1,1} x_{2,2}^3 + \frac{1}{3} a \alpha_1 \alpha_2 x_{2,2}^4 + \frac{2}{3} \alpha_1 \alpha_2 b x_{2,2}^4,$$

$$R = -a^2 \alpha_1^3 \alpha_2^2 x_{1,1} - \alpha_1^2 \alpha_2^3 b^2 x_{1,1} + a^2 \alpha_1^2 \alpha_2^2 x_{1,1}^2 - \alpha_1^3 \alpha_2 b^2 x_{1,1}^2 \quad (\text{A.94})$$

$$+ \alpha_1 \alpha_2^3 b^2 x_{1,1}^2 + 2 \alpha_1^2 \alpha_2 b^2 x_{1,1}^3 - \alpha_1 \alpha_2 b^2 x_{1,1}^4 - a^2 \alpha_1^2 \alpha_2^3 x_{2,2}$$

$$- \alpha_1^3 \alpha_2 b^2 x_{2,2} + 2 \alpha_1^2 \alpha_2^2 b^2 x_{1,1} x_{2,2} - 2 \alpha_1 \alpha_2^2 b^2 x_{1,1}^2 x_{2,2} + a^2 \alpha_1^2 \alpha_2^2 x_{2,2}^2$$

$$+ \alpha_1^3 \alpha_2 b^2 x_{2,2}^2 - \alpha_1 \alpha_2^3 b^2 x_{2,2}^2 - 2 \alpha_1^2 \alpha_2 b^2 x_{1,1} x_{2,2}^2 + 2 \alpha_1 \alpha_2 b^2 x_{1,1}^2 x_{2,2}^2$$

$$+ 2 \alpha_1 \alpha_2^2 b^2 x_{2,2}^3 - \alpha_1 \alpha_2 b^2 x_{2,2}^4,$$

y entonces

$$P_1 = a\alpha_1 x_{1,1} - \alpha_1 b x_{1,1} + a\alpha_2 x_{2,2} - \alpha_2 b x_{2,2} \quad (\text{A.95})$$

$$= [(a-b)\alpha_1] x_{1,1} + [(a-b)\alpha_2] x_{2,2},$$

$$Q_1 = \frac{4}{3} a\alpha_1^3 \alpha_2^2 x_{1,1} - \frac{1}{3} \alpha_1^3 \alpha_2^2 b x_{1,1} + \alpha_1^2 \alpha_2^3 b x_{1,1} + \frac{4}{3} a\alpha_1^2 \alpha_2^3 x_{2,2} + \alpha_1^3 \alpha_2^2 b x_{2,2} - \frac{1}{3} \alpha_1^2 \alpha_2^3 b x_{2,2}$$

$$= \left[\alpha_1^2 \alpha_2^2 \left(\frac{1}{3} \alpha_1 (a-b) + (\alpha_1 a + \alpha_2 b) \right) \right] x_{1,1} + \left[\alpha_1^2 \alpha_2^2 \left(\frac{1}{3} \alpha_2 (a-b) + (\alpha_2 a + \alpha_1 b) \right) \right] x_{2,2}.$$

Revisando los polinomios P_1, Q_1 con atención, vemos que estos son idénticos a lso del caso 2., donde va $x_{2,2}$ en lugar de $x_{2,1}$, y entonces podemos terminar la demostración análogamente.

Por último, el **caso 8.** resulta en cuentas idénticas al caso 7. pero reemplazando $x_{1,1}$ por $x_{1,2}$ y $x_{2,2}$ por $x_{2,1}$, lo que completa la demostración. ■

A.3 Demostraciones del capítulo 5

A.3.1. Consistencia fuerte para los estimadores por extensiones armónicas en dos comunidades balanceadas

Demostraremos aquí el Lema 5.5.6 que nos había quedado pendiente. La demostración se obtiene análogamente a la demostración de 3.5.3, y se puede escribir como a continuación:

Demostración. En primer lugar, notamos que usando primero la desigualdad de Markov y luego la independenciam entre las variables W_i y Z_i podemos acotar para obtener que vale que para cada $\lambda < 0$:

$$\mathbb{P} \left(\left\{ \sum_{i=1}^m W_i - Z_i \leq \eta \log(n) \right\} \right) = \mathbb{P} \left(\left\{ e^{\lambda \sum_{i=1}^m W_i - Z_i} \geq e^{\lambda \eta \log(n)} \right\} \right) \quad (\text{A.96})$$

$$\leq \frac{\mathbb{E} \left[e^{\sum_{i=1}^m W_i - Z_i} \right]}{n^{\lambda \eta}}$$

$$= n^{-\lambda \eta} \prod_{i=1}^m \mathbb{E} \left[e^{\lambda W_i} \right] \mathbb{E} \left[e^{-\lambda Z_i} \right].$$

Nos ocuparemos entonces de estudiar cada factor, y en adelante fijaremos $\lambda = \frac{1}{2} \log \left(\frac{b}{a} \right)$: la elección es técnica y permitirá llegar al exponente que precisamos para nuestra cota.

En primer lugar, podemos ver que vale que

$$\log \left(\mathbb{E} \left[e^{\lambda W_i} \right] \right) = \log \left(e^{\lambda a \frac{\log(n)}{n}} + \left(1 - a \frac{\log(n)}{n} \right) \right) \quad (\text{A.97})$$

$$= \log \left(1 + \left(e^{\lambda a \frac{\log(n)}{n}} - a \frac{\log(n)}{n} \right) \right)$$

$$\leq \left(e^{\lambda a \frac{\log(n)}{n}} - a \frac{\log(n)}{n} \right)$$

$$= (\sqrt{ab} - a) \frac{\log(n)}{n},$$

donde usamos que $\log(1+x) \leq x$ para cada $x \in \mathbb{R}$ en la desigualdad.

También podemos demostrar de la misma manera una desigualdad an

$$\begin{aligned} \log\left(\mathbb{E}\left[e^{-\lambda Z_i}\right]\right) &= \log\left(e^{-\lambda b \frac{\log(n)}{n}} + \left(1 - b \frac{\log(n)}{n}\right)\right) \\ &= \log\left(1 + \left(e^{-\lambda b \frac{\log(n)}{n}} - b \frac{\log(n)}{n}\right)\right) \\ &\leq \left(e^{-\lambda b \frac{\log(n)}{n}} - b \frac{\log(n)}{n}\right) \\ &= (\sqrt{ab} - b) \frac{\log(n)}{n}. \end{aligned} \tag{A.98}$$

Combinando ambas podemos ver que

$$\begin{aligned} \log\left(\mathbb{E}\left[e^{\lambda W_i}\right] \mathbb{E}\left[e^{-\lambda Z_i}\right]\right) &\leq (\sqrt{ab} - a) \frac{\log(n)}{n} + (\sqrt{ab} - b) \frac{\log(n)}{n} \\ &= -(\sqrt{a} - \sqrt{b})^2 \frac{\log(n)}{n}, \end{aligned} \tag{A.99}$$

y así juntando con la cota en (A.96) podemos llegar a la desigualdad que queríamos demostrar:

$$\begin{aligned} \mathbb{P}\left(\left\{\sum_{i=1}^m W_i - Z_i \leq \eta \log(n)\right\}\right) &\leq n^{-\lambda \eta} \prod_{i=1}^m \mathbb{E}\left[e^{\lambda W_i}\right] \mathbb{E}\left[e^{-\lambda Z_i}\right] \\ &\leq n^{-\lambda \eta} \prod_{i=1}^m e^{-(\sqrt{a} - \sqrt{b})^2 \frac{\log(n)}{n}} \\ &= n^{-\lambda \eta - \frac{m}{n} (\sqrt{a} - \sqrt{b})^2}. \end{aligned} \tag{A.100}$$

■

A.3.2. Consistencia débil para los estimadores por extensiones armónicas en el caso general

Comenzaremos demostrando que la extensión armónica teórica en el caso general cumple las propiedades que afirmamos en el Lema 5.4.6:

Demostración. Primero notaremos que la solución es única porque por la Proposición 5.1.13 tenemos que la extensión armónica es única. Alternativamente, podemos ver que $\mathbf{I} - \Pi Q$ es invertible, y de esta forma obtendremos la unicidad. En más detalle, basta mostrar que el radio espectral $\rho(\Pi Q)$ satisface $\rho(\Pi Q) < 1$. En efecto, como para cada autovector w de ΠQ vale que $w \in S$, tenemos además que $\Pi Q w = \Pi Q \Pi w$ y w es también autovector de $\Pi Q \Pi$, y así debe ser que $\rho(\Pi Q) \leq \rho(\Pi Q \Pi) \leq \|\Pi Q \Pi\|_\infty < 1$.

Ahora vamos a deducir las fórmula (5.74) proponiendo h en bloques como en el enunciado y despejando el valor de los valores $H_{i,l}$ usando el sistema de ecuaciones $h = \Pi Q h + v$. Más aún, resolveremos el sistema columna por columna, y sin pérdida de generalidad anotaremos a continuación como proceder con la primera columna solamente dado que el resto se resuelve

de forma análoga. Resolveremos luego $h_{\bullet,1} = \Pi Q h_{\bullet,1} + v_{\bullet,1}$, o más compactamente $(\mathbf{I} - \Pi Q)h_{\bullet,1} = v_{\bullet,1}$.

En primer lugar, notemos que si $i \in \mathcal{K}$ luego $(\Pi Q)_{i,\bullet} = \mathbf{0}$ con lo que necesariamente $h_{i,1} = v_{i,1}$ como queríamos.

En segundo lugar, para los nodos $i \notin \mathcal{K}$, despejaremos las fórmulas para $H_{l,1}$ enunciadas para cada comunidad $1 \leq l \leq k$. Como los nodos dentro de una misma comunidad inducen filas iguales para $\mathbf{I} - \Pi Q$ y para v , tendremos que el sistema anterior nos dará k ecuaciones para despejar los coeficientes. Cuando además $i \in \mathcal{C}_1$, obtenemos que la coordenada i -ésima del sistema nos da:

$$\begin{aligned} & \left[g(n) \left(-\frac{p_n}{\deg_n(1)} \right) + H_{1,1} + (n_1 - g(n) - 1) \left(-\frac{p_n}{\deg_n(1)} \right) H_{1,1} \right] \\ & + \sum_{t>1}^k (n_t - g(n)) \left(-\frac{q_n}{\deg_n(1)} \right) H_{t,1} = 0; \end{aligned} \quad (\text{A.101})$$

y cuando en cambio $i \in \mathcal{C}_l$ para algún $1 < l \leq k$:

$$\begin{aligned} & \left[g(n) \left(-\frac{q_n}{\deg_n(l)} \right) + (n_1 - g(n)) \left(-\frac{q_n}{\deg_n(l)} \right) H_{1,1} \right] \\ & + \left[H_{l,1} + (n_l - g(n) - 1) \left(-\frac{p_n}{\deg_n(l)} \right) H_{l,1} \right] \\ & + \sum_{t>1, t \neq l}^k (n_t - g(n)) \left(-\frac{q_n}{\deg_n(l)} \right) H_{t,1} = 0. \end{aligned} \quad (\text{A.102})$$

Ahora podemos restarle $\deg_n(l)$ veces (A.102) a $\deg_n(1)$ veces (A.101) para obtener:

$$\begin{aligned} 0 &= \deg_n(1) \left(\left[g(n) \left(-\frac{p_n}{\deg_n(1)} \right) + H_{1,1} + (n_1 - g(n) - 1) \left(-\frac{p_n}{\deg_n(1)} \right) H_{1,1} \right] \right. \\ & \quad \left. + (n_l - g(n)) \left(-\frac{q_n}{\deg_n(1)} \right) H_{l,1} \right) \\ & \quad - \deg_n(l) \left(\left[g(n) \left(-\frac{q_n}{\deg_n(l)} \right) + (n_1 - g(n)) \left(-\frac{q_n}{\deg_n(l)} \right) H_{1,1} \right] \right. \\ & \quad \left. + \left[H_{l,1} + (n_l - g(n) - 1) \left(-\frac{p_n}{\deg_n(l)} \right) H_{l,1} \right] \right) \\ &= -g(n)(p_n - q_n) + [\deg_n(1) - (n_1 - g(n) - 1)p_n + (n_1 - g(n))q_n] H_{1,1} \\ & \quad - [\deg_n(l) - (n_l - g(n) - 1)p_n + (n_l - g(n))q_n] H_{l,1}, \end{aligned} \quad (\text{A.103})$$

donde hay muchos términos que omitimos de entrada porque se cancelan. Ahora usando que para cada nodo $j \in \mathcal{C}_t$ vale que

$$\begin{aligned} & \deg_n(j) - (n_t - g(n) - 1)p_n + (n_t - g(n))q_n \\ &= [(n_t - 1)p_n + (n - n_t)q_n] - (n_t - g(n) - 1)p_n + (n_t - g(n))q_n \\ &= nq_n + g(n)(p_n - q_n) \end{aligned} \quad (\text{A.104})$$

es constante, podemos dividir por $nq_n + g(n)(p_n - q_n)$ y habremos obtenido que:

$$H_{1,1} - H_{l,1} = \frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)}. \quad (\text{A.105})$$

A continuación, podemos resolver para $H_{l,1}$ en términos de $H_{1,1}$ en (A.105) y entonces reemplazando en (A.101) podemos despejar una expresión para $H_{1,1}$:

$$H_{1,1} = \frac{p_n - \frac{(n-n_1-(k-1)g(n))(p_n-q_n)}{nq_n+g(n)(p_n-q_n)}q_n}{p_n + (k-1)q_n}, \quad (\text{A.106})$$

En efecto, reemplazando en (A.101) los coeficientes $H_{t,1}$, $1 < t \leq k$, y después multiplicando por $\deg_n(1)$ obtenemos:

$$\begin{aligned} & -p_n g(n) + \deg_n(1)H_{1,1} - (n_1 - g(n) - 1)p_n H_{1,1} \\ & - \sum_{t>1}^k (n_t - g(n))q_n \left(H_{1,1} - \frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)} \right) = 0; \end{aligned} \quad (\text{A.107})$$

y ahora despejando $H_{1,1}$ obtenemos que

$$\left[-g(n)p_n + g(n) \frac{(n - n_1 - (k-1)g(n))(p_n - q_n)}{nq_n + g(n)(p_n - q_n)} q_n \right] + [g(n)(p_n + (k-1)q_n)] H_{1,1} = 0, \quad (\text{A.108})$$

y entonces llegamos a (A.106). Aclaremos que en (A.108) usamos que

$$\deg_n(1) - (n_1 - g(n) - 1)p_n - \sum_{t \neq 1} (n_t - g(n))q_n = g(n)(p_n + (k-1)q_n) \quad (\text{A.109})$$

y además que

$$- \sum_{t>1}^k (n_t - g(n))q_n = -(n - n_1 - (k-1)g(n))q_n. \quad (\text{A.110})$$

Por último emprolijamos la expresión y llegamos a

$$H_{1,1} = \frac{(nq_n + n_1(p_n - q_n))q_n}{(p_n + (k-1)q_n)(nq_n + g(n)(p_n - q_n))} + \frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)}. \quad (\text{A.111})$$

En efecto, esta expresión se obtiene de desarrollar y despejar adecuadamente:

$$\begin{aligned} H_{1,1} &= \frac{p_n - \frac{(n-n_1-(k-1)g(n))(p_n-q_n)}{nq_n+g(n)(p_n-q_n)}q_n}{p_n + (k-1)q_n} \\ &= \frac{p_n(nq_n + g(n)(p_n - q_n)) - (n - n_1 - (k-1)g(n))(p_n - q_n)q_n}{(p_n + (k-1)q_n)(nq_n + g(n)(p_n - q_n))} \\ &= \frac{[(nq_n + n_1(p_n - q_n))q_n] + [g(n)(p_n + (k-1)q_n)(p_n - q_n)]}{(p_n + (k-1)q_n)(nq_n + g(n)(p_n - q_n))} \\ &= \frac{(nq_n + n_1(p_n - q_n))q_n}{(p_n + (k-1)q_n)(nq_n + g(n)(p_n - q_n))} + \frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)}. \end{aligned} \quad (\text{A.112})$$

Así, en particular, necesariamente por (A.105) vale que

$$H_{l,1} = \frac{(nq_n + n_1(p_n - q_n))q_n}{(p_n + (k-1)q_n)(nq_n + g(n)(p_n - q_n))}. \quad (\text{A.113})$$

Solo resta ver ahora cuándo $\widehat{\mathcal{C}}$ coincide con \mathcal{C} , pero esto es exactamente ver cuándo $H_{l,l} \geq H_{l,t}$ para cada par de comunidades l, t . Con las expresiones halladas en (A.111) y (A.113) deducimos que si $l \neq t$ tenemos que

$$H_{l,l} - H_{l,t} = \frac{(n_l - n_t)(p_n - q_n)q_n}{(p_n + (k-1)q_n)(nq_n + g(n)(p_n - q_n))} + \frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)}. \quad (\text{A.114})$$

Multiplicando a ambos lados por $\frac{nq_n + g(n)(p_n - q_n)}{p_n - q_n}$ tenemos que la expresión anterior es no negativa si y sólo si

$$g(n) > \frac{(n_t - n_l)q_n}{p_n + (k-1)q_n}, \quad (\text{A.115})$$

es decir

$$\varepsilon_n > \frac{n_t - n_l}{n} \frac{q_n}{p_n + (k-1)q_n}, \quad (\text{A.116})$$

con lo que, al precisar que valga para todo $l \neq t$, tomando máximo hallamos la condición que queríamos.

Finalmente, la observación acerca de la separación δ_n se deduce inmediatamente a partir de (A.114) y (A.116), pues si $\varepsilon_n = \frac{\max_{l \neq t} \{n_t - n_l\}}{n} \frac{q_n}{p_n + (k-1)q_n} + \delta_n$ con $\delta_n \geq 0$ entonces vale como queríamos ver que

$$\begin{aligned} H_{l,l} - H_{l,t} &= \frac{(n_l - n_t)(p_n - q_n)q_n}{(p_n + (k-1)q_n)(nq_n + g(n)(p_n - q_n))} + \frac{g(n)(p_n - q_n)}{nq_n + g(n)(p_n - q_n)} \\ &\gtrsim \delta_n \frac{n(p_n - q_n)}{nq_n + g(n)(p_n - q_n)} \\ &\asymp \delta_n, \end{aligned} \quad (\text{A.117})$$

lo cual completa la demostración. ■

Pasaremos ahora a demostrar la cota 5.4.7, para lo cual haremos uso de 2.4.11 y de 3.3.4:

Demostración. La demostración es análoga a la demostración de 5.3.1 con algunas diferencias técnicas, y la dividiremos en los pasos correspondientes para facilitar su lectura. Una vez más, supondremos sin pérdida de generalidad que los grafos son conexos, que podemos hacer por el Teorema 2.4.9.

Paso 0: primero, fijaremos sin pérdida de generalidad la columna 1 y a partir de ahora trabajaremos con las ecuaciones

$$\begin{aligned} \hat{h}_{\bullet,1} &= \Pi \widehat{Q} \hat{h}_{\bullet,1} + v_{\bullet,1}, \text{ y} \\ h_{\bullet,1} &= \Pi Q h_{\bullet,1} + v_{\bullet,1}. \end{aligned} \quad (\text{A.118})$$

Más aún, para no cargar excesivamente la notación anotaremos simplemente h , \hat{h} y v para las primeras columnas de las respectivas extensiones armónicas y condición de borde; nuevamente S pasará a ser el subespacio que definimos para dos comunidades en 5.2.4. Para el resto de las columnas, la cota será análoga.

Remarcamos que mostrar que la desigualdad ocurre con probabilidad mayor que $1 - o(n^{-1})$ para cada columna será suficiente, porque como hay una cantidad finita k de columnas por una cota por la unión de eventos también valdrá para todas simultáneamente con probabilidad mayor que $1 - o(1)$.

Paso 1: en primer lugar, buscaremos llegar a la desigualdad

$$\varepsilon_n \|\hat{h} - h\|_2^2 \lesssim |(\hat{h} - h)^t (\mathbf{I} - \Pi Q) (\hat{h} - h)|. \quad (\text{A.119})$$

Primero, sumamos y restamos ΠQ para obtener

$$\begin{aligned} \|\hat{h} - h\|_2^2 &= (\hat{h} - h)^t (\hat{h} - h) \\ &= (\hat{h} - h)^t (\mathbf{I} - \Pi Q) (\hat{h} - h) + (\hat{h} - h)^t \Pi Q (\hat{h} - h). \end{aligned} \quad (\text{A.120})$$

Luego, como $\hat{h} - h \in S$ tenemos que $\Pi(\hat{h} - h) = \hat{h} - h$ y entonces acotamos en (A.120) del siguiente modo:

$$|(\hat{h} - h)^t \Pi Q (\hat{h} - h)| = |(\hat{h} - h)^t \Pi Q \Pi (\hat{h} - h)| \leq \|\Pi Q \Pi\|_2 \|\hat{h} - h\|_2^2. \quad (\text{A.121})$$

Ahora bien, $\Pi Q \Pi$ salvo una permutación puede expresarse como

$$P \Pi Q \Pi P = D^{-1} \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & B \end{pmatrix} \quad (\text{A.122})$$

donde B es una matriz en $\mathbb{R}^{(n-kg(n)) \times (n-kg(n))}$ dada en bloques $B^{t,l} \in \mathbb{R}^{(n_t-g(n)) \times (n_l-g(n))}$ por:

$$B^{t,l} = \begin{cases} p_n (\mathbf{J} - \mathbf{I}) & \text{si } t = l, \\ q_n \mathbf{J} & \text{si } t \neq l; \end{cases} \quad (\text{A.123})$$

y más específicamente P es la matriz que pone primero los nodos conocidos y después los que no conocemos, ordenados por comunidad. Es decir, ordena los nodos en el siguiente orden: $\mathcal{K}_1, \mathcal{K}_2, \dots, \mathcal{K}_k, \mathcal{C}_1 \setminus \mathcal{K}_1, \mathcal{C}_2 \setminus \mathcal{K}_1, \dots, \mathcal{C}_k \setminus \mathcal{K}_k$.

Llamemos Λ a la matriz diagonal que tiene las sumas de las filas de B :

$$\begin{aligned} \Lambda_{i,i} &= (n_{\mathcal{C}(i)} - g(n) - 1) p_n + (n - n_{\mathcal{C}(i)} - (k-1)g(n)) q_n \\ &= n_{\mathcal{C}(i)} (p_n - q_n) - (g(n) + 1) p_n + (n - (k-1)g(n)) q_n. \end{aligned} \quad (\text{A.124})$$

Luego tenemos que como las filas de $\Lambda^{-1} B$ suman 1, esta es estocástica y tendrá norma menor que 1, y así valdrá que

$$\|\Pi Q \Pi\|_2 \leq \|D^{-1}\|_2 \|\Lambda\|_2 \|\Lambda^{-1} B\|_2 \leq \frac{1}{\min_{i \in V} \{\deg_n(i)\}} \cdot \max_{i \in V} \{\Lambda_{i,i}\} \cdot 1 \lesssim 1 - \varepsilon_n, \quad (\text{A.125})$$

notando que

$$\begin{aligned}
\max_{i \in V} \{ \Lambda_{i,i} \} &= \left(\max_{1 \leq l \leq k} \{ n_l \} - g(n) - 1 \right) p_n + \left(n - \max_{1 \leq l \leq k} \{ n_l \} - (k-1)g(n) \right) q_n \quad (\text{A.126}) \\
&= \left[\left(\max_{1 \leq l \leq k} \{ n_l \} - 1 \right) p_n + \left(n - \max_{1 \leq l \leq k} \{ n_l \} \right) q_n \right] - g(n)(p_n + (k-1)q_n) \\
&= \max_{i \in V} \{ \deg_n(i) \} - g(n)(p_n + (k-1)q_n) \\
&\lesssim \log(n) - \varepsilon_n \log(n) \quad (\text{A.127})
\end{aligned}$$

y por otro lado tendremos que

$$\begin{aligned}
\min_{i \in V} \{ \deg_n(i) \} &= \left(\min_{1 \leq l \leq k} \{ n_l \} - 1 \right) p_n + \left(n - \min_{1 \leq l \leq k} \{ n_l \} \right) q_n \quad (\text{A.128}) \\
&\asymp \log(n)
\end{aligned}$$

con lo que

$$\frac{\max_{i \in V} \{ \Lambda_{i,i} \}}{\min_{i \in V} \{ \deg_n(i) \}} \lesssim \frac{\log(n) - \varepsilon_n \log(n)}{\log(n)} = 1 - \varepsilon_n. \quad (\text{A.129})$$

Entonces acotando por arriba el segundo término en (A.120) por las cotas conseguidas en (A.121) y (A.125), obtenemos que

$$\|\hat{h} - h\|_2^2 \leq |(\hat{h} - h)^t (\mathbf{I} - \Pi Q)(\hat{h} - h)| + (1 - \varepsilon_n) \|\hat{h} - h\|_2^2, \quad (\text{A.130})$$

con lo que restando el segundo término en ambos lados obtenemos la primera expresión (A.119) que buscábamos.

Paso 2: ahora procedemos análogamente al paso correspondiente en la demostración de la Proposición 5.3.1 y de esta manera simplificamos la cota en (A.119) aún más a

$$\varepsilon_n \|\hat{h} - h\|_2^2 \lesssim |(\hat{h} - h)^t (\Pi \hat{Q} - \Pi Q) \hat{h}|. \quad (\text{A.131})$$

Más detalladamente, notamos que como $\hat{h} - h \in S$, luego $(\hat{h} - h)^t v = 0$. Así, como además $h = \Pi Q h + v$, $(\mathbf{I} - \Pi Q)h = v$ y entonces

$$\begin{aligned}
|(\hat{h} - h)^t (\mathbf{I} - \Pi Q)(\hat{h} - h)| &\leq |(\hat{h} - h)^t (\mathbf{I} - \Pi Q) \hat{h}| + |(\hat{h} - h)^t (\mathbf{I} - \Pi Q) h| \quad (\text{A.132}) \\
&= |(\hat{h} - h)^t (\mathbf{I} - \Pi Q) \hat{h}|.
\end{aligned}$$

Análogamente, también $(\mathbf{I} - \Pi \hat{Q}) \hat{h} = v$ y podemos sumar y restar $\Pi \hat{Q}$ para obtener

$$\begin{aligned}
|(\hat{h} - h)^t (\mathbf{I} - \Pi Q) \hat{h}| &\leq |(\hat{h} - h)^t (\mathbf{I} - \Pi \hat{Q}) \hat{h}| + |(\hat{h} - h)^t (\Pi \hat{Q} - \Pi Q) \hat{h}| \quad (\text{A.133}) \\
&= |(\hat{h} - h)^t (\Pi \hat{Q} - \Pi Q) \hat{h}|.
\end{aligned}$$

Juntando ambas cotas, llegamos a la expresión (A.131) como queríamos.

Paso 3: Podemos acotar una vez más partiendo de (A.131) para llegar a

$$\varepsilon_n \|\hat{h} - h\|_2 \lesssim \|\widehat{D}^{-1}\|_2 \|\widehat{A} - A\|_2 \|\hat{h}\|_2 + \|\Pi(\widehat{D}^{-1} - D^{-1})A\hat{h}\|_2, \quad (\text{A.134})$$

donde usamos que como

$$\Pi\widehat{Q} - \Pi Q = \Pi\widehat{D}^{-1}(\widehat{A} - A) + \Pi(\widehat{D}^{-1} - D^{-1})A \quad (\text{A.135})$$

podemos reemplazar en (A.131) y acotar

$$\begin{aligned} \varepsilon_n \|\widehat{h} - h\|_2^2 &\lesssim |(\widehat{h} - h)^t [\Pi\widehat{D}^{-1}(\widehat{A} - A) + \Pi(\widehat{D}^{-1} - D^{-1})A] \widehat{h}| \\ &\leq \|\widehat{h} - h\|_2 \left(\|\Pi\|_2 \|\widehat{D}^{-1}\|_2 \|\widehat{A} - A\|_2 \|\widehat{h}\|_2 + \|\Pi(\widehat{D}^{-1} - D^{-1})A\widehat{h}\|_2 \right) \end{aligned} \quad (\text{A.136})$$

para obtener (A.134) al valer que Π , al ser proyección, tiene norma acotada por 1.

Paso 4: A continuación a partir de (A.134) pasaremos a acotar cada término por cotas del orden $\frac{1}{\sqrt{\log(n)}} \|\widehat{h}\|_2$ que valgan con suficientemente alta probabilidad para concluir el resultado. En efecto, en ese caso tendremos que

$$\varepsilon_n \|\widehat{h} - h\|_2 \lesssim \frac{1}{\sqrt{\log(n)}} \|\widehat{h}\|_2 \leq \frac{1}{\sqrt{\log(n)}} \left(\|\widehat{h} - h\|_2 \right) + \frac{1}{\sqrt{\log(n)}} \|h\|_2 \quad (\text{A.137})$$

con lo cual agrupando las normas $\|\widehat{h} - h\|_2$ vemos que existe alguna constante $C > 0$ tal que

$$\left(\varepsilon_n - C \frac{1}{\sqrt{\log(n)}} \right) \|\widehat{h} - h\|_2 \lesssim \frac{1}{\sqrt{\log(n)}} \|h\|_2 \quad (\text{A.138})$$

y como por hipótesis tenemos que $\varepsilon_n \sqrt{\log(n)} \xrightarrow[n \rightarrow \infty]{} \infty$, podemos deducir que

$$(1 - o(1)) \|\widehat{h} - h\|_2 = \left(1 - C \frac{1}{\varepsilon_n \sqrt{\log(n)}} \right) \|\widehat{h} - h\|_2 \lesssim \frac{1}{\varepsilon_n \sqrt{\log(n)}} \|h\|_2 \quad (\text{A.139})$$

de lo cual se sigue la cota que queríamos demostrar.

Volviendo a la expresión en (A.134), acotaremos cada factor en el primer término. La matriz \widehat{D}^{-1} es diagonal con coeficientes $\widehat{D}_{i,i}^{-1} = \widehat{\deg}_n(i)^{-1}$ para cada $1 \leq i \leq n$, con lo que su norma estará dada por el mayor coeficiente en módulo entre ellos, correspondiente al mínimo grado de un nodo en el grafo. Por la Proposición 2.4.11 sabemos que cada grado es de orden $\log(n)$ y luego $\|\widehat{D}^{-1}\|_2 \lesssim \frac{1}{\log(n)}$, con probabilidad mayor que $1 - o(1)$. Además, el Teorema 3.3.4 nos garantiza que $\|\widehat{A} - A\|_2 \lesssim \sqrt{\log(n)}$ con probabilidad $1 - o(1)$. Concluimos que con probabilidad mayor que $1 - o(1)$ vale como queríamos ver que

$$\|\widehat{D}^{-1}\|_2 \|\widehat{A} - A\|_2 \|h\|_2 \lesssim \frac{1}{\log(n)} \log(n) \|h\|_2 = \frac{1}{\sqrt{\log(n)}} \|h\|_2. \quad (\text{A.140})$$

Resta solamente acotar el segundo término en (A.134) con alta probabilidad. Un cálculo directo nos muestra que

$$\left(\Pi(\widehat{D}^{-1} - D^{-1})A\widehat{h} \right)_i = \begin{cases} \frac{\widehat{\deg}_n(i) - \deg_n(i)}{\widehat{\deg}_n(i) \deg_n(i)} \rho_i & \text{si } i \notin \mathcal{K}, \\ 0 & \text{si } i \in \mathcal{K}. \end{cases} \quad (\text{A.141})$$

para ρ_i dada por $\rho_i = (A\hat{h})_i = \sum_{j \in \mathcal{C}_l, j \neq i} p_n \hat{h}_j + \sum_{j \notin \mathcal{C}_l} q_n \hat{h}_j$ cuando $i \in \mathcal{C}_l$ para $1 \leq l \leq k$. Ahora usando la desigualdad de Cauchy-Schwarz y que $a > b$, obtenemos que para $i \notin \mathcal{K}$ podemos acotar

$$|\rho_i| \leq a \frac{\log n}{\sqrt{n}} \|\hat{h}\|_2. \quad (\text{A.142})$$

Usando el cálculo explícito en (A.141) junto con estas cotas podemos ver que

$$\|\Pi(\widehat{D}^{-1} - D^{-1})A\hat{h}\|_2 \leq \max_{1 \leq i \leq n} \left\{ \frac{1}{\deg_n(i)\widehat{\deg}_n(i)} \right\} \|(\widehat{D} - D)\mathbf{1}_n\|_2 \left(a \frac{\log n}{\sqrt{n}} \|\hat{h}\|_2 \right), \quad (\text{A.143})$$

donde $(\widehat{D} - D)\mathbf{1}_n$ no es más que la matriz diagonal que tiene la resta de los grados. Ahora bien, por la Proposición 2.4.11, tanto $\deg_n(i)$ como $\widehat{\deg}_n(i)$ son de orden $\log(n)$ con probabilidad mayor que $1 - o(1)$ lo que el máximo se acota por un factor de orden $\frac{1}{\log(n)^2}$. Más aún, usando que $(\widehat{D} - D)\mathbf{1}_n = (\widehat{A} - A)\mathbf{1}_n$ y por el Teorema 3.3.4 podremos acotar con probabilidad mayor que $1 - o(1)$ para conseguir que

$$\begin{aligned} \|(\widehat{D}^{-1} - D^{-1})\mathbf{1}_n\|_2 &\leq \|\widehat{A} - A\|_2 \|\mathbf{1}_n\|_2 \\ &\lesssim \sqrt{\log(n)}\sqrt{n}. \end{aligned} \quad (\text{A.144})$$

Finalmente juntando todas estas cotas auxiliares y acotando en (A.143) obtendremos que con probabilidad mayor que $1 - o(1)$ vale que

$$\begin{aligned} \|\Pi(\widehat{D} - D)A\hat{h}\|_2 &\lesssim \frac{1}{\log(n)^2} \left(\sqrt{\log(n)}\sqrt{n} \right) \left(a \frac{\log(n)}{\sqrt{n}} \|\hat{h}\|_2 \right) \\ &\lesssim \frac{1}{\sqrt{\log(n)}} \|\hat{h}\|_2, \end{aligned} \quad (\text{A.145})$$

lo que completa la demostración. ■

Podremos ahora demostrar entonces la consistencia débil del estimador por extensiones armónicas en el caso general, como enunciamos en el Teorema 5.4.8:

Demostración. Nuevamente, llamemos $\mathcal{B}(\hat{h})$ al conjunto de nodos mal clasificados por nuestra estimación, y veamos que tiene $o(n)$ miembros con probabilidad alta, lo que demostrará la consistencia débil del estimador.

En primer lugar, notemos que $i \in \mathcal{B}(\hat{h})$ nos implica que

$$\|\hat{h}_{i,\bullet} - h_{i,\bullet}\|_2 \gtrsim \delta_n, \quad (\text{A.146})$$

donde la cota es la separación esperada para la solución teórica en 5.4.6. En efecto, tenemos que basta ver que

$$\|\hat{h}_{i,\bullet} - h_{i,\bullet}\|_2 \gtrsim \left| h_{i,\widehat{\mathcal{C}}(i)} - h_{i,\mathcal{C}(i)} \right| \asymp \delta_n \quad (\text{A.147})$$

para cada $i \in \mathcal{B}(\hat{h})$, es decir i tal que $\hat{\mathcal{C}}(i) \neq \mathcal{C}(i)$. En primer lugar, si $\hat{h}_{i,\hat{\mathcal{C}}(i)} \geq h_{i,\mathcal{C}(i)}$ tendremos que

$$\|\hat{h}_{i,\bullet} - h_{i,\bullet}\|_2 \geq \hat{h}_{i,\hat{\mathcal{C}}(i)} - h_{i,\hat{\mathcal{C}}(i)} \geq h_{i,\mathcal{C}(i)} - h_{i,\hat{\mathcal{C}}(i)} \quad (\text{A.148})$$

como queríamos. Si en cambio $\hat{h}_{i,\mathcal{C}(i)} \leq h_{i,\hat{\mathcal{C}}(i)}$, análogamente tendremos que

$$\|\hat{h}_{i,\bullet} - h_{i,\bullet}\|_2 \geq h_{i,\mathcal{C}(i)} - \hat{h}_{i,\mathcal{C}(i)} \geq h_{i,\mathcal{C}(i)} - h_{i,\hat{\mathcal{C}}(i)} \quad (\text{A.149})$$

y habremos mostrado lo afirmado. Cuando no ocurre ninguna de las dos anteriores, tendremos que $h_{i,\hat{\mathcal{C}}(i)} \leq \hat{h}_{i,\mathcal{C}(i)} \leq \hat{h}_{i,\hat{\mathcal{C}}(i)} \leq h_{i,\mathcal{C}(i)}$, en cuyo caso podemos ver que

$$\|\hat{h}_{i,\bullet} - h_{i,\bullet}\|_2^2 \geq (\hat{h}_{i,\mathcal{C}(i)} - h_{i,\mathcal{C}(i)})^2 + (\hat{h}_{i,\hat{\mathcal{C}}(i)} - h_{i,\hat{\mathcal{C}}(i)})^2 \geq \frac{1}{2} (h_{i,\mathcal{C}(i)} - h_{i,\hat{\mathcal{C}}(i)})^2 \quad (\text{A.150})$$

pues en los valores de $\hat{h}_{i,\mathcal{C}(i)}$ y $\hat{h}_{i,\hat{\mathcal{C}}(i)}$ estudiados, la suma de cuadrados se minimiza cuando $\hat{h}_{i,\hat{\mathcal{C}}(i)} = \hat{h}_{i,\mathcal{C}(i)} = \frac{h_{i,\mathcal{C}(i)} + h_{i,\hat{\mathcal{C}}(i)}}{2}$ y en cuyo caso se deduce lo afirmado.

Notemos además que tenemos que

$$\|h_{\bullet,l}\|_2^2 \lesssim \varepsilon_n n. \quad (\text{A.151})$$

Luego, usando la cota obtenida en (A.146), reordenando los términos y por último usando la Proposición 5.4.7 y la cota (A.151), obtenemos que

$$\begin{aligned} \#(\mathcal{B}(\hat{h})) \delta_n^2 &\lesssim \sum_{i=1}^n \|\hat{h}_{i,\bullet} - h_{i,\bullet}\|_2^2 \\ &= \sum_{l=1}^k \|\hat{h}_{\bullet,l} - h_{\bullet,l}\|_2^2 \\ &\lesssim k \max_{1 \leq l \leq k} \left\{ \|\hat{h}_{\bullet,l} - h_{\bullet,l}\|_2^2 \right\} \\ &\lesssim \frac{1}{\varepsilon_n^2 \log(n)} \max_{1 \leq l \leq k} \left\{ \|h_{\bullet,l}\|_2^2 \right\} \\ &\lesssim \frac{1}{\varepsilon_n \log(n)} n, \end{aligned} \quad (\text{A.152})$$

y despejando llegamos a que

$$\#(\mathcal{B}(\hat{h})) \lesssim \frac{1}{\delta_n^2 \varepsilon_n \log(n)} n = o(n) \quad (\text{A.153})$$

lo cual muestra lo que queríamos. ■