
Modelamiento de redes sociales múltiples

Modeling of multilayer social networks

Víctor Nicolás Álvarez Monroy.^a
victoralvarez@usantotomas.edu.co

Dr. Juan Camilo Sosa Martínez^b
jcsosam@unal.edu.co

Resumen

En este trabajo se propone un modelo estadístico para caracterizar simultáneamente dos o más redes sociales, donde interactúan el mismo conjunto de actores. Además de investigar las relaciones dentro de las redes, el modelo propuesto permite extrapolar la información entre redes, con el fin de obtener mejores resultados en términos de bondad de ajuste y predicción. Esta propuesta se basa en una extensión jerárquica del modelo de espacio latente de distancias, que asume una posición social “global” para cada actor, lo que permite estudiar de forma parsimoniosa los roles sociales desde varios puntos de vista. Las capacidades del modelo se ilustran utilizando varios conjuntos de datos reales, teniendo en cuenta diferentes tipos de relaciones.

Palabras clave: Cadena de Markov de Monte Carlo, Estadística Bayesiana, Redes Sociales, Modelo de Espacio Latente.

Abstract

In this work, we propose a statistical model to simultaneously characterize two or more social networks, where the same set of actors interact with each other. In addition to investigate ties within networks, the proposed model shares information among networks in order to obtain better results in terms of both goodness-of-fit and prediction. Our proposal is based on a hierarchical extension of the latent space distance model, by assuming a “global” social position for every actor, which allows us to study parsimoniously social roles from several perspectives. The capabilities of the model are illustrated using several real datasets, taking into account different types of relationships.

Keywords: Markov Chain Monte Carlo, Bayesian Statistics, Networks, Latent Space Model.

1. Introducción

El estudio de redes sociales es recurrente en todos los campos del conocimiento, primordialmente donde se investigan las interacciones de un conjunto de individuos bajo diferentes contextos y puntos de vista. Ejemplos abundan en todas las ciencias: en la industria, con la planificación y optimización de procesos operacionales; en la biología, con el estudio de sistemas complejos relacionados con el genoma humano; en la economía, con la investigación de relaciones entre naciones, en términos de cooperación económica, derechos humanos, y acciones militares; entre muchos otros.

^aEstudiante, Maestría en Estadística Aplicada, Universidad Santo Tomás, Bogotá.

^bProfesor Asistente, Departamento de Estadística, Universidad Nacional de Colombia, Bogotá.

Otro campo de aplicación que ilustra tanto la importancia como los retos del estudio de redes es el de las comunicaciones. La forma y la velocidad con que nos conectamos constituyen actualmente un eje importante de desarrollo, debido a que cada individuo establece diferentes formas de relacionarse de acuerdo con su entorno. Por ejemplo, los diferentes roles que desempeñan los empleados de una compañía no son iguales a las relaciones interpersonales que se establecen fuera de su trabajo. Sin embargo, las dos formas de interactuar pueden tener una conexión, a pesar de que estas sean diferentes. Muestra de estas interacciones lo reflejan las redes sociales *online*, tales como LinkedIn y Twitter, las cuales tienen diferentes enfoques, laboral y de opinión, respectivamente, pero con algunas estructuras en común.

Estudiando redes sociales, típicamente es de interés investigar la probabilidad de que dos actores interactúen entre sí junto con su “posición social” en el sistema. Mas aún, si los individuos participan en diferentes redes sociales, y por lo tanto se relacionan de diversas formas en diversos escenarios. Así, se quiere estudiar las dependencias estructurales tanto “dentro” como “entre” relaciones, dado que se dispone de diferentes fuentes de información en los que se encuentran involucrados los mismos actores del sistema.

Un modelo estadístico muy popular en la literatura para una sola red, es el modelo de espacio latente de distancia (Hoff et al., 2002). De acuerdo con este modelo, las probabilidades de interacción dependen de la distancia entre las posiciones de los actores en un “espacio social” latente (i.e., no observado). Esta formulación es muy popular porque permite modelar directamente características que se observan directamente en las redes sociales, como la transitividad (el amigo de un amigo es un amigo) y el balance (el enemigo de un amigo es un enemigo), por ejemplo. Avances significativos en el modelamiento de redes simples por medio de variables latentes se encuentran en Nowicki and Snijders (2001), Hoff et al. (2002), Schweinberger and Snijders (2003), Hoff (2005), Handcock et al. (2007), Linkletter (2007), Krivitsky and Handcock (2008), Hoff (2008), Krivitsky et al. (2009), Hoff (2009), Li et al. (2011), Raftery et al. (2012), Minhas et al. (2019).

En este trabajo se extiende el modelo de distancias de Hoff con el fin de investigar de manera conjunta el proceso generativo de redes múltiples croseccionales (estáticas). En este sentido, Gollini and Murphy (2016) y Salter-Townshend and McCormick (2017) proveen contribuciones muy importantes, permitiendo que los actores desempeñen diferentes roles en diferentes relaciones. Nuestra propuesta construye sobre estos modelos, dado que mantiene la verosimilitud Bernoulli multivariada, pero añade más estructura en la distribución previa correspondiente al posicionamiento social de cada actor, lo cual permite caracterizar parsimoniosamente sus características dentro de cada red y a través de las redes.

Desde la perspectiva de los espacios latentes, además del trabajo pionero de Salter-Townshend and McCormick (2017), otras alternativas para estudiar redes sociales múltiples han surgido durante los dos últimos años. En conectomía cerebral, Durante et al. (2018) proponen un modelo no paramétrico Bayesiano a través de un modelo de mezcla que reduce dimensionalidad e incorpora eficientemente la información de las redes dentro de cada componente de la mezcla; luego, Wang et al. (2019) proponen un modelo para estudiar similitudes y diferencias de las matrices de adyacencia, basado en una descomposición jerárquica de valores singulares. A su vez, D’Angelo y colaboradores extienden los modelos de espacio latente en otros contextos, incorporando efectos específicos para los nodos (D’Angelo et al., 2018) y covariables específicas para los enlaces D’Angelo et al. (2019), y finalmente, introduciendo grupos en el marco de una mezcla infinita de distribuciones D’Angelo et al. (2020). Otros avances importantes desde una perspectiva frecuentista se encuentran disponibles en Zhang (2020).

Trabajos adicionales relacionados con redes múltiples croseccionales incluyen la detección de comunidades (e.g., Han et al., 2015, Reyes and Rodriguez, 2016, Paul et al., 2016, Gao et al., 2019, Paez et al., 2019, y Paul et al., 2020), y la evaluación de percepción cognitiva en estructuras sociales (e.g., Swartz et al., 2015, Sosa, 2017, Sewell et al., 2019). Finalmente, desde el punto de vista dinámico, existen una gran variedad de enfoques para modelar la evolución de un sistema a lo largo del tiempo (e.g., Durante and Dunson, 2014, Hoff, 2015, Sewell and Chen, 2015, Sewell and Chen, 2016, Sewell et al., 2017, Gupta et al., 2018, Kim et al., 2018, Turnbull, 2020).

Este trabajo está estructurado como sigue: en la Sección 2 se presentan varios casos que motivan el estudio de redes sociales múltiples; en la Sección 3 se estudian los modelos de espacio latente de distancia; en la Sección 4 se propone un modelo para la caracterización de redes múltiples, donde además se discuten las propiedades de intercambiabilidad junto con la elicitación de los hiper-parámetros del modelo; en la Sección 5 se consideran todos los aspectos relacionados con la implementación del modelo; en la Sección 6 se presentan los resultados correspondientes a los ejemplos de motivación; y finalmente, en la Sección 7 se discuten los resultados junto con algunas alternativas de investigación futura.

1.1. Redes sociales

Una **red social** (o simplemente red) hace referencia a un conjunto de n actores (elementos o individuos) junto con una variable $y_{i,i'}$ definida entre cada uno de ellos, para todo par $i, i' = 1, \dots, n, i \neq i'$. Esta variable $y_{i,i'}$ caracteriza la interacción de dos individuos i y i' en un sistema particular. En el caso más simple, la variable $y_{i,i'}$ es una variable dicotómica que toma los valores de 0 y 1; 0, si no hay una conexión entre los actores i y i' ; y 1, si efectivamente existe una conexión.

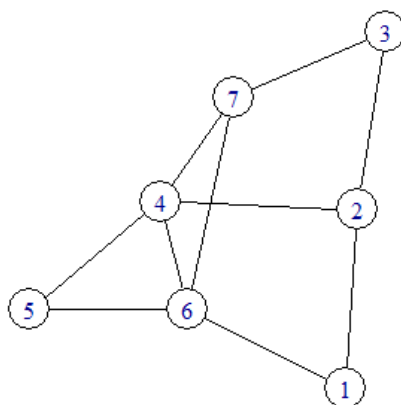


Figura 1: Ejemplo de una red no dirigida con $n = 7$ actores.

Fuente: Elaboración propia

Los datos, se pueden representar por medio de un grafo, el cual es una colección de vértices (o nodos) y aristas (o lazos). En el contexto de las redes sociales, los actores corresponden a los vértices, mientras que las aristas representan las relaciones entre dichos actores.

La Figura 1 ilustra una red para una muestra de $n = 7$ vértices los cuales pueden representar; senadores de un partido político, integrantes en una familia o un conjunto de países de una región determinada.

Una red se denomina **no dirigida** es cuando $y_{i,i'} = y_{i',i}$ para todos par de vértices, en caso contrario se dice que la red es **dirigida**, un ejemplo de red dirigida se ilustra en la Figura 2, donde claramente existe una direccionalidad implícita en las relaciones. Se observa que la red de la Figura 1 es dirigida, dado que las aristas no hacen referencia a ninguna dirección.

Por otra parte, al tener dos vértices i y i' en una red, si existe una conexión entre ellos se denota como $i \sim i'$, lo cual implica que existe una arista entre el vértice i y el vértice i' . Se dice que dos vértices conectados son vecinos, y la vecindad del vértice i , N_i , siendo el conjunto de vértices vecinos al vértice i :

$$N_i = \{i' : i \sim i'\}.$$

Teniendo como resultado un orden matricial, donde cada interacción será representada por un número (en nuestro caso será 1), señalándonos que hay interacción entre dos vértices, generando una matriz asociada a la red (o matriz de adyacencia) $\mathbf{Y}$, siendo una matriz de $n \times n$ elementos $y_{i,i'}$, tal que:

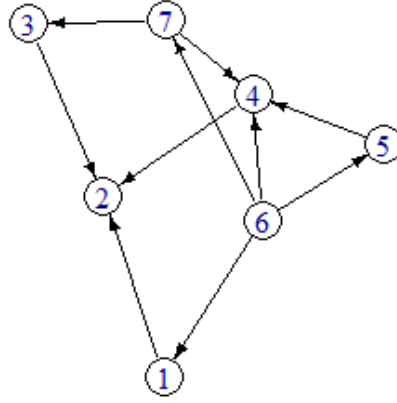


Figura 2: Ejemplo de una red dirigida con siete actores actores.

Fuente: Elaboraci n propia

$$y_{i,i'} = \begin{cases} 1 & \text{si } i \sim i', i \neq i', \\ 0 & \text{de otra manera} \end{cases}$$

Llamaremos **sub-red** de $\mathbf{Y}$, a un sub-conjunto $\mathbf{Y}'$ de la red $\mathbf{Y}$. En una red no dirigida, su matriz asociada $\mathbf{Y}$ es sim trica, pudi ndose representar como un vector $\mathbf{y}$ de $\binom{n}{2}$ componentes. Para ilustrar estos, consideremos el grafico de la Figura 1, cuya matriz de adyacencia es:

$$\mathbf{Y} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

donde $\mathbf{y} = (1, 1, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 1)'$ es la representaci n vectorial de la red, lo cual permite una simplificaci n computacional favorable.

De la misma forma el **grado** de un nodo i , d_i , es el n mero de vecinos (nodos conectados mediante una arista). Esto es:

$$d_i = |\mathcal{N}_i|$$

La **sucesi n gradual** de una red, es una secuencia $(D_0, \dots, D_M)$ donde D_k es el n mero de v rtices en la red con grado k , y M es el grado m ximo observado en la red. La **distribuci n de grados** est  dada por $(D_0/n, \dots, D_M/n)$, donde n es el n mero de nodos de la red.

Llamaremos **grado ampliado** del v rtices i , e_i , estado determinado por:

$$e_i = \sum_{i': i \sim i'} |\mathcal{N}_i / \{i\}|$$

Teniendo en cuenta la red de la Figura 1, los grados de los v rtices son; $d_1 = 2$, $d_2 = 3$, $d_3 = 3$, $d_4 = 4$, $d_5 = 3$, $d_6 = 3$ y $d_7 = 2$. Por lo tanto, la sucesi n de grados para la red es $(0, 2, 4, 1)$. Y por ejemplo el grado ampliado del nodo 3 es $e_3 = \sum_{i' \in \mathcal{N}_3} d_{i'} - d_3 = 4$.

Si se quiere determinar el largo de **trayectoria** para v entre el vértice i y el vértice j es una sucesión de índices $\{i_0, i_1, \dots, i_{v+1}\}$ con $i_0 = i$ y $i_{v+1} = j$, tal que:

$$\prod_{t=0}^v y_{i_t, i_{t+1}} = 1.$$

Se denomina **distancia geodésica mínima** entre el vértice i y i' , $g_{i,i'}$, a la longitud del camino más corto de i a i' en $\mathbf{Y}$. Si no hay un camino, entonces $g_{i,i'} = \infty$.

La **sucesión geodésica mínima** es $(G_0, \dots, G_M)$, donde G_k es el número de nodo pares que tienen la distancia geodésica mínima k , y M es la distancia geodésica más grande. También llamamos **triángulo** a un ciclo de tres longitudes en una red. Tres nodos i, i' , y k forma un triángulo si $i \sim i'$, $i' \sim k$, y $k \sim i$.

La **densidad** de una red no dirigida es la frecuencia de las aristas existentes en relación con las aristas potenciales. Por ejemplo, una red no dirigida G , sin auto-bucles y sin múltiples aristas, la densidad de una sub-red $H = (V_H, E_H)$ es

$$\text{den}(H) = \frac{|E_H|}{|V_H|(|V_H| - 1)/2}$$

si el valor $\text{den}(H)$ se acerca a a cero se dice que la red es dispersa, de lo contrario si se acerca a uno, se dice que la red es densa.

El valor $cl_T(G)$ se denomina **transitividad** de una red, se define como:

$$cl_T(G) = \frac{3\tau_\Delta(G)}{\tau_3(G)}$$

donde $\tau_\Delta(G)$ es el número de triángulos en la red G , y $\tau_3(G)$, el número de triplas conectadas (es decir, una sub-red de tres vértices conectados por dos aristas). Esta cantidad es de interés en la literatura de redes sociales, conocidas también como: “fracción de triples transitivos”. La transitividad es una medida de agrupamiento global, que resume la frecuencia relativa con la que se conectan los triples cercanos para formar triángulos.

Un factor importante es la tendencia a relacionarse con actores con similares características, la cual es denominada **asortatividad**, definida como:

$$r_a = \frac{\sum_i f_{ii} - \sum_i f_{i+} f_{+i}}{1 - \sum_i f_{i+} f_{+i}}$$

donde $f_{ii'}$ es la fracción de aristas en G que unen un vértice en la i -ésima categoría con un vértice en la i' -ésima categoría, tal que f_{i+} y f_{+i} denotan la i -ésima fila y columna marginales, respectivamente, de la matriz resultante f .

El valor r_a se encuentra entre -1 y 1 . Es igual a cero cuando la mezcla en la red no es diferente de la obtenida mediante una asignación aleatoria de aristas que conserva la distribución de grados marginales.

2. Algunos ejemplos de motivación

Los individuos en un entorno social se pueden relacionar de diferentes maneras, generando así un sistema de J redes sociales sobre mismo conjunto de I actores. A continuación, se presentan algunos ejemplos de redes binarias múltiples no dirigidas con el fin de evidenciar escenarios donde este concepto tiene cabida.

2.1. Sala de cableado

Roethlisberger and Dickson (2003)¹ recopilaron datos de $I = 14$ empleados de la sala de cableado de la empresa Western Electric: dos inspectores, tres soldadores y nueve operarios. Los autores recopilaron $J = 4$ tipos de relaciones, a saber, participación en juegos (Juegos), participación en diálogos a través de ventanas abiertas (Diálogos), amistad (Amistad) y comportamiento antagonico (Enemistad). La Figura 3 y la Tabla 1 muestran una descripción de las redes.

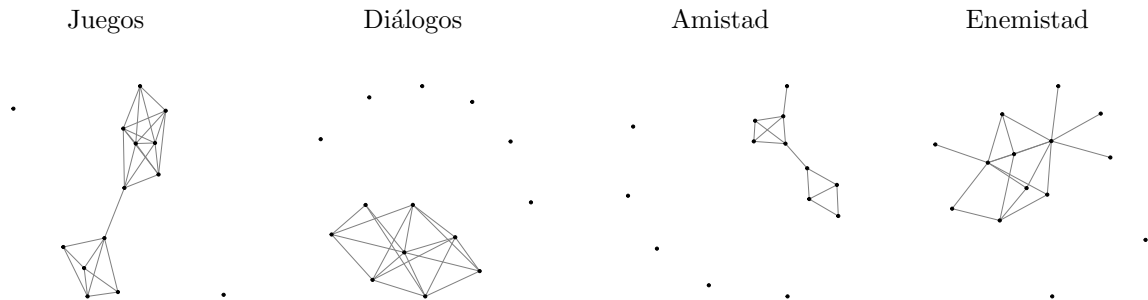


Figura 3: Grafos para el conjunto de datos de la sala de cableado.

Fuente: Elaboración propia

2.2. Microfinanzas

Banerjee et al. (2013)² recopilaron datos demográficos de varias villas en Karnataka, al sur de la India, con el fin de estudiar las micro-finanzas. En este proyecto se realizó un censo de los hogares y sus miembros, a los cuales se le hizo un cuestionario detallado sobre las relaciones que tenían en la aldea. Estos datos se registraron entre $I = 74$ hogares de una villa en particular, en términos de $J = 3$ tipos de relaciones: intercambios de dinero y enseres (Intercambios), recomendaciones personales y médicas (Consejos), y interacciones sociales (Sociales). La Figura 4 y la Tabla 1 muestran una descripción de las redes.

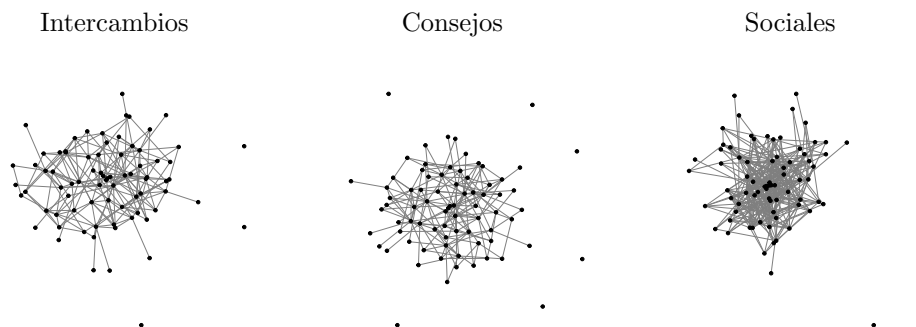


Figura 4: Grafos para el conjunto de datos de micro-finanzas.

Fuente: Elaboración propia

¹La base de datos se tomó en: <http://vlado.fmf.uni-lj.si/pub/networks/data/ucinet/wiring.datt>

²La base de datos se tomó en: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/21538>

2.3. Redes sociales *on-line*

Huawei es una empresa China de gran importancia a nivel global, y en uno de sus procesos de investigación en tecnología, recopiló información acerca de las conexiones entre mil usuarios (anonimizados) que utilizaban $J = 3$ redes sociales de manera frecuente, a saber, Facebook, Twitter, e Instagram, con el fin de direccionar sus procesos de marketing efectivamente. El conjunto de datos se encuentra disponible en <https://www.kaggle.com/andrewlucchi/huawei-social-network-data>. En este caso, se hizo una representación con una muestra aleatoria de $I = 100$ actores. La Figura 5 y la Tabla 1 muestran una descripción de las redes.

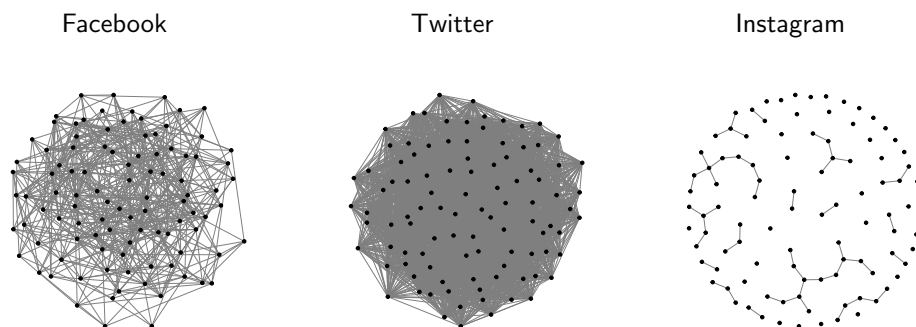


Figura 5: Grafos para el conjunto de datos de redes sociales *on-line*.

Fuente: *Elaboración propia*

2.4. Personal gerencial

Krackhardt (1987)³ recopiló datos en la gerencia de una empresa de fabricación de máquinas de alta tecnología, para evaluar los efectos de una intervención que se llevó a cabo en su administración. A los $I = 21$ empleados se les preguntó acerca de sus relaciones de amistad y aquellas que percibía entre los demás, lo cual produjo $J = 21$ redes, cada una asociada con la percepción de cada empleado. Un conjunto de datos de esta índole se denomina *Cognitive Social Structure* (CSS). La Tabla 1 muestra una descripción de las redes.

Datos	Abreviatura	I	J	Densidad	Transitividad	Asortatividad
Sala de cableado	CABLEADO	14	4	0.217	0.591	-0.187
Hogares al sur de la India	HOGARES	74	3	0.064	0.186	-0.215
Redes sociales <i>on-line</i>	SOCIALES	100	3	0.206	0.204	-0.560
Personal gerencial	GERENCIA	21	21	0.124	0.337	-0.195

Tabla 1: Estadística descriptiva para los ejemplos de motivación. Los valores de la densidad, la transitividad, y la asortatividad corresponden al valor promedio de todas las redes.

Fuente: *Elaboración propia*

3. Modelo de distancia de espacio latente para una red

El uso de efectos aleatorios en el contexto de modelos lineales generalizados es una alternativa popular modelar redes. Específicamente, considere un modelo en el que las $y_{i,i'}$ s son condicionalmente indepen-

³La base de datos se tomó en: <http://vlado.fmf.uni-lj.si/pub/networks/data/ucinet/krackfr.dat>

dientes con probabilidades de interacci n

$$\Pr[y_{i,i'} = 1 \mid \beta, \gamma_{i,i'}, \mathbf{x}_{i,i'}] = g^{-1}(\mathbf{x}_{i,i'}^\top \beta + \gamma_{i,i'}), \quad i < i', \quad (1)$$

donde $\beta = (\beta_1, \dots, \beta_P)$ es un vector (desconocido) de efectos fijos, $\mathbf{x}_{i,i'}^\top \beta = \sum_{p=1}^P \beta_p x_{i,i',p}$ es un predictor lineal que representa los patrones en los datos asociados con las covariables observadas en los actores $\mathbf{x}_{i,i'} = (x_{i,i',1}, \dots, x_{i,i',P})$, $\gamma_{i,i'}$ es un efecto aleatorio espec fico no observado (latente), que representa cualquier patr n adicional en los datos no relacionados con aquellos de los predictores, y $g(\cdot)$ es una funci n de enlace (conocida).

Siguiendo resultados en Aldous (1985) y Hoover (1982) (ver tambi n Hoff, 2008, para m s detalles), se puede mostrar que si la matriz de efectos aleatorios $\Gamma = [\gamma_{i,i'}]$ es intercambiable conjuntamente, entonces existe una funci n sim trica $\alpha(\cdot, \cdot)$ tal que $\gamma_{i,i'} = \alpha(\mathbf{u}_i, \mathbf{u}_{i'})$, donde $\mathbf{u}_1, \dots, \mathbf{u}_I$, es una secuencia de vectores aleatorios independientes no observados. El impacto de tales variables latentes en (1) est  dictado en gran medida por la forma de la funci n $\alpha(\cdot, \cdot)$. Por lo tanto, es principalmente a trav s de $\alpha(\cdot, \cdot)$ que es posible capturar otras caracter sticas relevantes de los datos relacionales. Una serie de formulaciones potenciales para $\alpha(\cdot, \cdot)$ han sido explorados en la literatura hasta la fecha.

En particular, el modelo de distancia de Hoff et al. (2002) supone que cada actor i tiene una posici n $\mathbf{u}_i = (u_{i,1}, \dots, u_{i,K}) \in \mathbb{R}^K$ en un espacio social (espacio de caracter sticas latentes), y que la probabilidad de un enlace entre dos actores disminuye a medida que los actores se “alejan” en el espacio social. Para este fin, los efectos latentes se especifican como $\alpha(\mathbf{u}_i, \mathbf{u}_{i'}) = -\|\mathbf{u}_i - \mathbf{u}_{i'}\|$, donde $\|\cdot\|$ denota la norma euclidiana. Las estructuras latentes basadas en distancias inducen naturalmente la transitividad, la cual es una caracter stica primordial de muchas redes sociales reales. Adem s, se observa que modelar las posiciones como pertenecientes a un espacio euclidiano de baja dimensi n proporciona una alternativa eficaz para representar gr ficamente datos relacionales.

Dada una red binaria no dirigida $\mathbf{Y} = [y_{i,i'}]$, la verosimilitud (sin covariables) bajo un modelo de distancia est  dada por

$$p(\mathbf{Y} \mid \zeta, \{\mathbf{u}_i\}) = \prod_{i,i': i < i'} \text{Ber}(y_{i,i'} \mid \text{expit}(\zeta - \|\mathbf{u}_i - \mathbf{u}_{i'}\|)),$$

donde $\text{expit}(x) = 1/(1 + e^{-x})$ es la inversa de la funci n logit, $\zeta \in \mathbb{R}$ es un efecto fijo que representa la propensi n promedio de observar una conexi n entre dos actores dados, y $\mathbf{u}_1, \dots, \mathbf{u}_I$ representan las posiciones sociales no observadas en $\mathbb{R}^K$ de los actores que conforman el sistema. Se observa que la verosimilitud solo contiene $I(I-1)/2$ t rminos dado que $\mathbf{Y}$ corresponde a la matriz de adyacencia de una red no dirigida.

Para proceder con un an lisis completamente Bayesiano y hacer inferencia sobre los par metros del modelo, se deben especificar las distribuciones previas para ζ y cada $\mathbf{u}_i$. Una opci n que parece funcionar bien en la pr ctica es $\zeta \mid \tau^2 \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \tau^2)$ y $\mathbf{u}_i \mid \sigma^2 \stackrel{\text{iid}}{\sim} \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, donde $\mathbf{I}$ denota la matriz identidad; por lo general, se completa la formulaci n de la distribuci n previa haciendo $\tau^2 \sim \text{IGam}(a_\tau, b_\tau)$ y $\sigma^2 \sim \text{IGam}(a_\sigma, b_\sigma)$. As , el conjunto de par metros del modelo es $\Upsilon = (\zeta, \mathbf{u}_1, \dots, \mathbf{u}_I, \sigma^2, \tau^2)$, el cual tiene un total de $IK + 3$ cantidades desconocidas por estimar, asociados con los hiper-par metros a_τ , b_τ , a_σ , y b_σ . La Figura 6 muestra la representaci n del modelo de distancia para una sola red por medio de un *Directed Acyclic Graph* (DAG).

Una selecci n cuidadosa de los hiper-par metros es fundamental para garantizar el rendimiento adecuado del modelo. Para este fin, se establece $a_\tau = 2$ y $b_\tau = 100$, que coloca una distribuci n previa difusa para ζ . Del mismo modo, imitamos una heur stica dada en Krivitsky and Handcock (2008, Sec. 2.4) haciendo a_σ y b_σ de tal forma que a priori σ^2 est  vagamente concentrada (e.g., $\text{CV}(\sigma^2) = 1$) al rededor de $E[\sigma^2] = I^{2/K}$, i.e., proporcional al volumen bidimensional de una bola euclidiana de radio $I^{1/K}$.

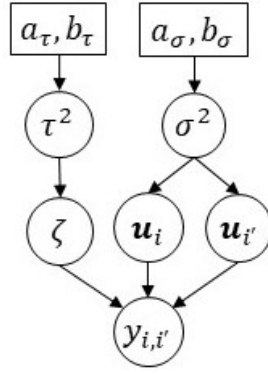


Figura 6: Representación DAG del modelo de distancia para una sola red. Los círculos corresponden a cantidades aleatorias (parámetros), mientras que los cuadrados a cantidades fijas (hiper-parámetros).

Fuente: *Elaboración propia*

4. Modelo de distancia de espacio latente para redes múltiples

En esta sección se presenta nuestra propuesta para modelar simultáneamente un conjunto de $J \geq 2$ redes sociales binarias no dirigidas y no reflexivas $\mathbf{Y}_1, \dots, \mathbf{Y}_J$, observadas sobre mismo conjunto de I actores, donde $\mathbf{Y}_j = [y_{i,i',j}]$. Observe que en este caso $y_{i,i',j} \in \{0, 1\}$ y $y_{i,i',j} = y_{i',i,j}$. Ya que cada red contiene información relevante acerca del sistema social, en lugar de solo ajustar modelos de distancia independientes para cada red, el propósito de nuestro enfoque consiste en extrapolar la información entre redes, con el fin de obtener mejores resultados en términos de bondad de ajuste y predicción dentro de cada red.

4.1. Modelamiento

Nuestra propuesta se basa en una extensión jerárquica constituida por varias etapas por del modelo para una sola red discutido en la Sección 3. Para tal efecto, se asume que las conexiones son condicionalmente independientes con distribución Bernoulli, lo cual resulta en una verosimilitud dada por

$$p(\mathbf{Y} \mid \{\zeta_j\}, \{\mathbf{u}_{i,j}\}) = \prod_j \prod_{i, i': i < i'} \text{Ber}(y_{i,i',j} \mid \text{expit}(\zeta_j - \|\mathbf{u}_{i,j} - \mathbf{u}_{i',j}\|))$$

donde $\zeta_j \in \mathbb{R}$ es un efecto fijo que representa la propensión promedio de observar una conexión entre dos actores dados en la red j , y $\mathbf{u}_{i,j} = (\mathbf{u}_{i,j,1}, \dots, \mathbf{u}_{i,j,K})$ denota la posición social del actor i en la red j , definida en un espacio euclidiano K -dimensional. En este contexto la interpretación de las posiciones sociales no ha cambiado: si las características latentes de los individuos i y i' en la red j se “alejan” en el espacio social, entonces la distancia $\|\mathbf{u}_{i,j} - \mathbf{u}_{i',j}\|$ se incrementa, y por lo tanto la probabilidad de observar un enlace entre los actores i e i' disminuye.

Siguiendo la formulación estándar de los modelos de distancia, a las posiciones sociales $\mathbf{u}_{i,j}$ se les asigna distribuciones normales condicionalmente independientes, pero esta vez con media y varianza sujetas al estado de cada individuo, $\mathbf{u}_{i,j} \mid \boldsymbol{\eta}_i, \sigma_j^2 \stackrel{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\eta}_i, \sigma_j^2 \mathbf{I})$. La media $\boldsymbol{\eta}_i = (\eta_{i,1}, \dots, \eta_{i,K})$ se puede interpretar convenientemente como la posición social “global” del actor i en relación con todas las relaciones que definen el sistema. Ahora, estableciendo que tanto las medias globales $\boldsymbol{\eta}_i$ como los componentes de varianza σ_j^2 tienen una distribución común, $\boldsymbol{\eta}_i \mid \boldsymbol{\theta}, \kappa^2 \stackrel{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\theta}, \kappa^2 \mathbf{I})$ y $\sigma_j^2 \mid \alpha_\sigma, \beta_\sigma \stackrel{\text{iid}}{\sim} \text{IGam}(\alpha_\sigma, \beta_\sigma)$, es posible capturar similitudes entre los actores, y así, compartir información a través de las redes. Para completar con la formulación jerárquica de estos niveles, se hace $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, $\kappa^2 \sim \text{IGam}(a_\kappa, b_\kappa)$, $\alpha_\sigma \sim \text{Unif}(c_\sigma, d_\sigma)$, y $\beta_\sigma \sim \text{Gam}(a_\sigma, b_\sigma)$. Finalmente, siguiendo una idea similar, a los efectos fijos de las probabilidades interacción ζ_j se les asigna una distribución previa jerárquica conjugada, $\zeta_j \mid \mu_\zeta, \tau_\zeta^2 \sim \mathcal{N}(\mu_\zeta, \tau_\zeta^2)$, con $\mu_\zeta \sim \mathcal{N}(\mu_0, \sigma_0^2)$ y $\tau_\zeta^2 \sim \text{IGam}(a_\zeta, b_\zeta)$.

As , el conjunto de par metros del modelo es

$$\Upsilon = (\zeta_1, \dots, \zeta_J, \mathbf{u}_{1,1}, \dots, \mathbf{u}_{I,J}, \boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_I, \sigma_1^2, \dots, \sigma_J^2, \mu_\zeta, \tau_\zeta^2, \boldsymbol{\theta}, \kappa^2, \alpha_\sigma, \beta_\sigma),$$

el cual tiene un total de $IK(J+1) + 2J + K + 5$ cantidades desconocidas por estimar, asociados con los hiper-par metros $\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, a_\kappa, b_\kappa, \mu_0, \sigma_0^2, a_\zeta, b_\zeta, a_\sigma, b_\sigma, c_\sigma, d_\sigma$. La Figura 7 muestra la representaci n del modelo de distancia para una desde m ltiples por medio de un DAG. En esta representaci n son evidentes las tres jerarqu as del modelo.

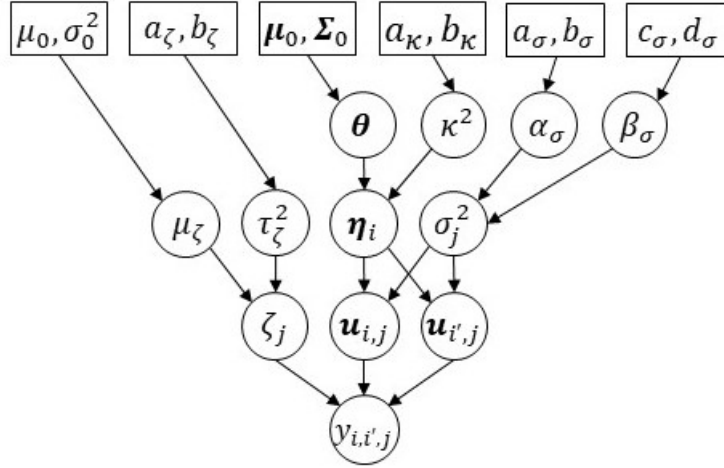


Figura 7: Representaci n DAG del modelo de distancia para redes m ltiples. Los c rculos corresponden a cantidad aleatorias (par metros), mientras que los cuadrados a cantidades fijas (hiper-par metros).

Fuente: *Elaboraci n propia*

4.2. Intercambiabilidad

Aldous (1985) extiende las nociones de intercambiabilidad de De Finetti para matrices de dos dimensiones, mediante la introducci n de los conceptos de intercambiabilidad separada (la distribuci n conjunta de una matriz bidimensional de variables aleatorias es invariante e independiente de permutaciones arbitrarias de filas y columnas) e intercambiabilidad conjunta (la distribuci n conjunta de una matriz bidimensional de variables aleatorias es invariante solo si se aplica la misma permutaci n arbitraria tanto a las filas como a las columnas). La intercambiabilidad conjunta implica intercambiabilidad separada, pero no viceversa. Estas nociones pueden extenderse f cilmente a arreglos de dimensiones superiores, al considerar cada posible par de dimensiones, por ejemplo.

En el contexto de las redes m ltiples, sean π_1 y π_2 dos permutaciones de $\{1, \dots, I\}$, y π_3 una permutaci n de $\{1, \dots, J\}$. Bajo la versi n del modelo que se presenta en la Secci n 4.1, la distribuci n marginal conjunta de las observaciones $\{y_{i,i',j}\}$ es la misma que la distribuci n de $\{y_{\pi_1(i), \pi_2(i'), \pi_3(j)}\}$ solo si $\pi_1 = \pi_2$. En otras palabras, el modelo define una distribuci n de probabilidad intercambiable de manera conjunta para datos provenientes de redes m ltiples. La intercambiabilidad conjunta (en lugar de una forma m s d bil de intercambiabilidad) es particularmente atractiva en este contexto, porque todos los  ndices i e i' (y potencialmente j) se refieren al mismo conjunto de actores.

4.3. Elicitaci n de los hiper-par metros

Como antes, una selecci n cuidadosa de los hiper-par metros es fundamental para garantizar el rendimiento adecuado del modelo. As , para la distribuci n previa de $\boldsymbol{\theta}$, se hace $\mu_0 = \mathbf{0}$ y $\boldsymbol{\Sigma}_0 = 100\mathbf{I}$, lo que a

priori centra en el origen a los promedios globales $\boldsymbol{\eta}_i$ de manera no informativa. En el planteamiento del modelo latente para una red (Sección 3) se indica que la implementación del modelo con hiper-parámetros que generen una distribución previa no informativa, con el fin de evitar la concentración degenerada de posiciones latentes en un solo punto del espacio social. Por tal motivo, se recomienda seguir la heurística dada en Krivitsky and Handcock (2008, Sec. 2.4), para no asumir ninguna concentración a priori en la distribución de las posiciones latentes de los vértices, antes de observar las redes. Este mismo concepto se implementó en el caso de redes múltiples.

Siguiendo el misma idea para los efectos fijos ζ_j , se hace $\mu_0 = 0$ y $\sigma_0^2 = 100$. Luego, para la componente de varianza κ^2 , se establece una distribución previa con media finita $I^{2/K}$, pero varianza infinita, con $a_\kappa = 2$ y $b_\kappa = I^{2/K}$; esto permite variar de manera difusa los $\boldsymbol{\eta}_i$ de acuerdo con la heurística dada en la Sección 3. Similarmente, para la componente de varianza τ^2 , se hace $a_\zeta = 2$ y $b_\zeta = 1$, lo que a priori permite variar de manera razonable los ζ_j en un rango de valores apropiado. Finalmente, con el fin de dejar variar difusamente las características latentes propias de los en cada una de las redes en un espacio razonable, dependiendo de la cantidad de redes, se hace $a_\sigma = 1$, $b_\sigma = J^{-1/K}$, $c_\sigma = 0$ y $d_\sigma = 10$.

5. Cómputo

Para una dimensión latente K fijo, se puede explorar la distribución posterior de los parámetros usando métodos de *Markov chain Monte Carlo* (MCMC; e.g., Gamerman and Lopes, 2006), donde la distribución posterior se aproxima usando muestras dependientes, pero aproximadamente idénticamente distribuidas, $\boldsymbol{\Upsilon}^{(1)}, \dots, \boldsymbol{\Upsilon}^{(B)}$ de la distribución posterior $p(\boldsymbol{\Upsilon} \mid \mathbf{Y}) \propto p(\mathbf{Y} \mid \boldsymbol{\Upsilon})p(\boldsymbol{\Upsilon})$, donde $\boldsymbol{\Upsilon}$ es el conjunto de parámetros del modelo dado en la Sección 4.1. Estimaciones puntuales e intervalos de credibilidad se pueden aproximar a partir de las distribuciones empíricas que proporcionan las muestras.

En nuestro algoritmo, donde sea posible, se toman muestras de las distribuciones condicionales completas como en un muestreador de Gibbs usual; de lo contrario, se usan versiones adaptativas con pasos de Metropolis-Hastings (e.g., Haario et al., 2001). Los detalles sobre el algoritmo implementado aquí se pueden encontrar en el Apéndice A.

5.1. Identificabilidad

Los modelos de distancia son invariables a rotaciones y reflejos del espacio social. De hecho, para cualquier matriz ortogonal $\mathbf{Q}$ de tamaño $K \times K$, la verosimilitud asociada con la reparametrización $\tilde{\mathbf{u}}_{i,j} = \mathbf{Q}\mathbf{u}_{i,j}$ es independiente de $\mathbf{Q}$, dado que $\|\tilde{\mathbf{u}}_{i,j} - \tilde{\mathbf{u}}_{i',j'}\| = \|\mathbf{u}_{i,j} - \mathbf{u}_{i',j'}\|$. En el modelo planteado, los parámetros que no son identificables son las posiciones latentes $\mathbf{u}_{i,j}$, debido a que los modelos de distancia entre conjuntos de puntos de un espacio euclidiano son invariantes bajo movimientos rígidos, ya que asociado con cada tensor de posiciones latentes $\mathbf{U} = [\mathbf{u}_{i,j}]$, existe un número infinito de posiciones que producen la misma verosimilitud respecto al modelo propuesto. Por favor ver la Sección 5.1 para más detalles.

Sin embargo, dado que el predictor lineal $\eta_{i,i',j} = \zeta_j - \|\mathbf{u}_{i,j} - \mathbf{u}_{i',j}\|$ es completamente identificable, y en consecuencia, las probabilidad de interacción $\text{expit}(\eta_{i,i',j})$ siguen siendo identificable gracias a la inyectividad de la función expit , se concluye que el modelo propuesto es completamente identificable. Ver por favor (Hoff et al., 2002, p. 1091) para más detalles.

Este problema se aborda utilizando un enfoque de transformación de parámetros similar a lo descrito en Hoff et al. (2002, Sec. 3). En particular, se restringe la inferencia a una clase particular de posiciones latentes, por medio de una transformación de las muestras posteriores $\mathbf{u}_{i,j}^{(b)}$ a un sistema de coordenadas compartido. Así, para cada muestra $\boldsymbol{\Upsilon}^{(b)}$, se obtiene una matriz de transformación ortogonal $\mathbf{Q}^{(b)}$ minimizando la distancia de Procrustes,

$$\tilde{\mathbf{Q}}^{(b)} = \underset{\mathbf{Q} \in \mathcal{S}^K}{\text{argmin}} \text{tr} \left\{ \left(\mathbf{W}^{(1)} - \mathbf{W}^{(b)}\mathbf{Q} \right)^\top \left(\mathbf{W}^{(1)} - \mathbf{W}^{(b)}\mathbf{Q} \right) \right\},$$

donde $\mathcal{S}^K$ denota el conjunto de matrices ortogonales de tama o $K \times K$ y $\mathbf{W}^{(b)}$ es una matriz rectangular de tama o $I \times K$ cuyas filas corresponden a las medias globales $\boldsymbol{\eta}_1^{(b)}, \dots, \boldsymbol{\eta}_I^{(b)}$. Este problema de optimizaci n se puede resolver f cilmente utilizando una descomposici n de valores singulares (e.g., Borg and Groenen, 2005, Sec. 20.2). Una vez que se obtienen las matrices $\tilde{\mathbf{Q}}^{(1)}, \dots, \tilde{\mathbf{Q}}^{(B)}$, la inferencia posterior para las posiciones latentes se basa directamente en $\tilde{\mathbf{u}}_{i,j}^{(b)} = \tilde{\mathbf{Q}}^{(b)} \mathbf{u}_{i,j}^{(b)}$ y $\tilde{\boldsymbol{\eta}}_i^{(b)} = \tilde{\mathbf{Q}}^{(b)} \boldsymbol{\eta}_i^{(b)}$.

5.2. Selecci n de la dimensi n latente K

Establecer $K = 2$ es popular en la literatura de modelos de espacio latente, ya que simplifica la visualizaci n, y por lo tanto, facilita la descripci n de las relaciones sociales. Sin embargo, nuestro objetivo va m s all  de una mera descripci n de las redes, por lo que el valor de K juega un papel cr tico en el resultados. A continuaci n se discuten algunas metodolog as para seleccionar adecuadamente la dimensi n del espacio social.

La literatura de redes se ha centrado en gran medida en el *Bayesian Information Criterion* (BIC; e.g., Hoff, 2005, Handcock et al., 2007). Sin embargo, el BIC suele ser inapropiado para modelos jer rquicos ya que la estructura jer rquica implica que el n mero efectivo de par metros ser  t picamente menor que el n mero real de par metros en la verosimilitud. Una alternativa al BIC es el *Watanabe-Akaike Information Criterion* (WAIC; Watanabe, 2010, Gelman et al., 2014),

$$\text{WAIC}(K) = -2 \sum_j \sum_{i,i': i < i'} \log \mathbb{E}[p(y_{i,i',j} | \boldsymbol{\Upsilon})] + 2p_{\text{WAIC}},$$

donde

$$p_{\text{WAIC}} = 2 \sum_j \sum_{i,i': i < i'} (\log \mathbb{E}[p(y_{i,i',j} | \boldsymbol{\Upsilon})] - \mathbb{E}[\log p(y_{i,i',j} | \boldsymbol{\Upsilon})]).$$

corresponde a la complejidad del modelo (n mero efectivo de par metros), y $\boldsymbol{\Upsilon}$ es el conjunto de par metros asumiendo que la dimensi n del espacio social es K . Los valores esperados en estas expresiones se calculan respecto a la distribuci n posterior $p(\boldsymbol{\Upsilon} | \mathbf{Y})$, los cuales se pueden aproximar usando las muestras $\boldsymbol{\Upsilon}^{(1)}, \dots, \boldsymbol{\Upsilon}^{(B)}$ provenientes del algoritmo basado en MCMC.

6. Ilustraciones

6.1. Sala de cableado

Aqu  se analizan las redes m ltiples de la sala de cableado introducidas en la Secci n 2, usando nuestra propuesta. Los resultados que se presentan a continuaci n se basan en $B = 20,000$ muestras de la distribuci n posterior obtenidas despu s de reducir la cadenas de Markov originales cada 25 observaciones y un periodo de calentamiento de 20,000 iteraciones. La convergencia fue monitoreada haciendo seguimiento de la variabilidad de la distribuci n conjunta de los datos y los par metros del modelo utilizando el procedimiento del cadenas m ltiples dado en Gelman et al. (1992). Por ejemplo, el panel derecho de la Figura 8 muestra la cadena de la log-verosimilitud asociada con el valor de la dimensi n latente que optimiza el criterio de informaci n WAIC, donde se observa que no hay evidencia aparente de falta de convergencia.

Dimensi n del espacio social

La Figura 8 muestra los valores de WAIC asociados con modelos ajustados con diferente dimensi n K del espacio social. Este criterio favorece una elecci n de $K = 4$ (WAIC = 201.46), que es el valor que se usa en todos los an lisis de este conjunto de datos. As , nuestro modelo utiliza 297 par metros con $K = 4$.

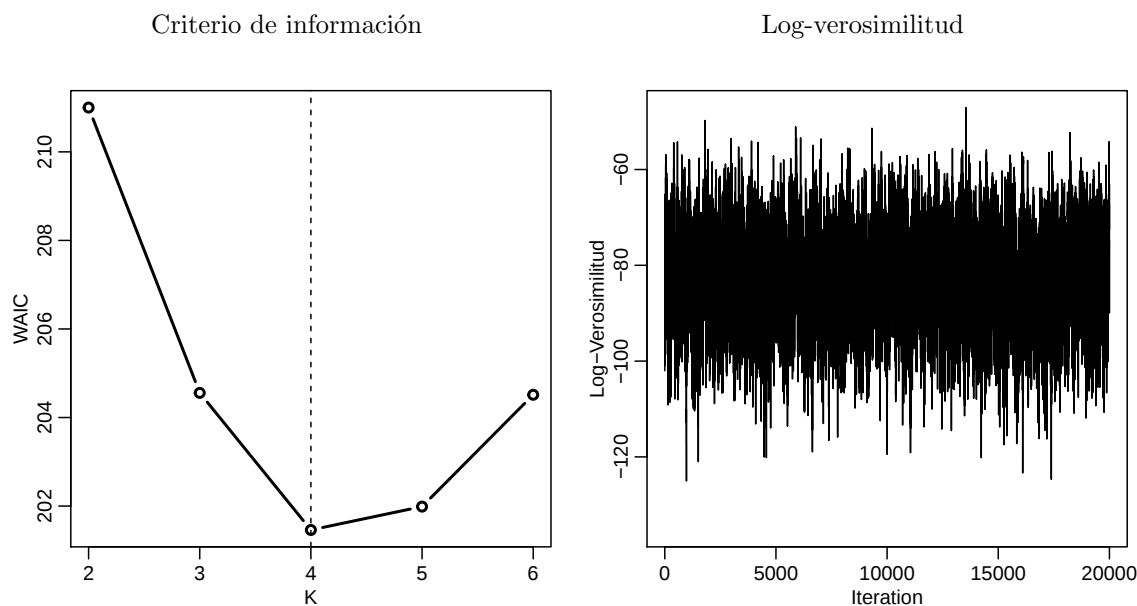


Figura 8: Valores de WAIC para seleccionar la dimensión K del espacio social para el análisis de los datos de la sala de cableado, ajustando el modelo propuesto en la Sección 4, y cadena de Markov de la log-verosimilitud asociada con el valor de K que optimiza el criterio de información WAIC.

Fuente: *Elaboración propia*

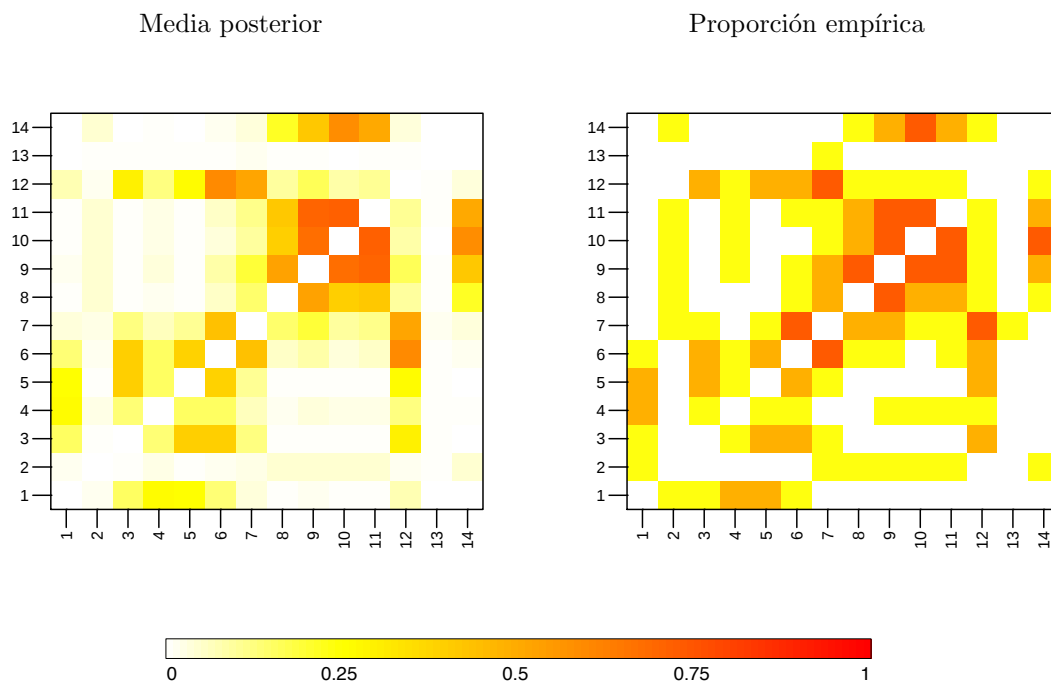


Figura 9: Dos estimaciones de la probabilidad consenso de interacción entre dos actores. El panel izquierdo proporciona la media posterior correspondiente bajo nuestro modelo, mientras que el panel derecho muestra la proporción de enlaces observados a través de la redes.

Fuente: *Elaboración propia*

Red de consenso

Nuestro modelo se puede usar para generar una estimaci n de una red de “consenso”, la cual colapsa la informaci n de todas las redes en una sola red. En particular, la probabilidad consenso de observar un enlace entre los actores i e i' se puede estimar con $\vartheta_{i,i'} = \text{expit}(\mu_\zeta - \|\boldsymbol{\eta}_i - \boldsymbol{\eta}_{i'}\|)$. La Figura 9 presenta la matriz de medias posteriores,

$$\mathbb{E}[\vartheta_{i,i'} \mid \mathbf{Y}] = \frac{1}{B} \sum_{b=1}^B \text{expit} \left(\mu_\zeta^{(b)} - \|\boldsymbol{\eta}_i^{(b)} - \boldsymbol{\eta}_{i'}^{(b)}\| \right),$$

junto con un mapa de calor de la proporci n de enlaces observados a trav s de la redes, $\frac{1}{J} \sum_{j=1}^J y_{i,i',j}$. Aunque la estimaci n que proporciona nuestro modelo es menos “densa” que la proporci n emp rica, se observa que las estimaciones son muy similares. Esto tambi n sugiere que el modelo caracteriza correctamente el proceso generativo de los datos.

Proyecciones en el espacio social

Las posiciones latentes $\mathbf{u}_{i,j}$ en el espacio social proporcionan una herramienta poderosa para describir las interacciones sociales. La primera fila de la Figura 10 muestra las coordenadas a lo largo de las dos dimensiones de mayor variabilidad de las posiciones $\mathbf{u}_{i,j}$ para cada una de las cuatro redes. Se observa que la din mica social de Juegos y Amistad es muy similar, salvo que en Juegos los actores 1 y 4 tienen la misma posici n social; asimismo, en estas dos relaciones se diferencian claramente dos grupos de individuos. La relaci n de Di logos tambi n tiene cierta semejanza con Juegos y Amistad, aunque la diferenciaci n de los grupos es menos notoria, y adem s, las caracter sticas del actor 3 no concuerdan con aquellas de los dem s actores. De otra parte, aquellos actores que tienen posiciones cercanas (lejanas) en Amistad, por lo general tienen posiciones lejanas (cercanas) en Enemistad; este efecto es particularmente claro entre los actores $\{1; 4\}$ y $\{10; 14\}$. La descripci n de los roles sociales tambi n se puede llevar a cabo por medio de las posiciones promedio $\boldsymbol{\eta}_i$.

Como se observ  anteriormente, las posiciones latentes $\mathbf{u}_{i,j}$ tambi n permiten distinguir grupos de actores que cumplen roles sociales semejantes. Con el fin de identificar estos grupos es posible aplicar alguna t cnica no supervisa de agrupamiento (e.g., agrupamiento jer rquico, k -medias), o por el contrario, incluir directamente en el modelo un conjunto de par metros que asignen los actores a los grupos. La  ltima alternativa es preferible dado que permite cuantificar directamente la incertidumbre relacionada con el agrupamiento junto con su relaci n con los dem s par metros del modelo (ver la Secci n 7 para m s detalles).

Ajuste del modelo

Aqu , se eval a el modelo utilizando m tricas dentro de la muestra. Primero, se comparan las matrices de adyacencia observadas, $\mathbf{Y}_j = [y_{i,i',j}]$, con la medias posteriores correspondientes de las probabilidades de interacci n, $\hat{\boldsymbol{\Theta}}_j = [\hat{\theta}_{i,i',j}]$, donde

$$\hat{\theta}_{i,i',j} = \mathbb{E}[\text{expit}(\zeta_j - \|\mathbf{u}_{i,j} - \mathbf{u}_{i',j}\|) \mid \mathbf{Y}] = \frac{1}{B} \sum_{b=1}^B \text{expit} \left(\zeta_j^{(b)} - \|\mathbf{u}_{i,j}^{(b)} - \mathbf{u}_{i',j}^{(b)}\| \right).$$

La Figura 10 muestra las probabilidades de interacci n estimadas. En general, se observa que, las probabilidades estimadas coinciden con las relaciones observadas en todas las redes, lo cual constituye evidencia adicional de que el modelo ajusta correctamente los datos.

Ahora, siguiendo a Gelman et al. (2013, Cap. 6), se explora m s a fondo la bondad de ajuste del modelo generando r plicas del conjunto de datos a partir de los par metros de la cadena de Markov del modelo ajustado, y luego calculando un conjunto de estad sticas de resumen para cada muestra, cuyas distribuciones luego se comparan con los valores observados en la muestra original. La Figura 11 muestra res menes de la distribuci n emp rica de las estad sticas de prueba seleccionadas (densidad, transitividad, asortatividad) para cada red (ver Kolaczyk and Cs rdi, 2014 para m s detalles acerca de las estad sticas de prueba). El modelo captura adecuadamente las caracter sticas estructurales de los

datos dado que todos los valores observados pertenecen a los intervalos de credibilidad. Los resultados con el número de triángulos, la distancia promedio, la eigen-centralidad promedio, el grado promedio, y la desviación estándar del grado (no mostrados aquí) son equivalentes.

6.2. Predicción y bondad de ajuste

Con el fin de evaluar la capacidad del modelo propuesto (CONJUNTO, para abreviar) para predecir enlaces perdidos, en esta sección se evalúa su poder predictivo fuera de la muestra a través de un experimento de validación cruzada, usando los cuatro sistemas de redes múltiples que se presentan la Sección 2 (ver la Tabla 1 para más detalles). Además, como punto de referencia, también se considera el modelo de distancia para una sola red discutido en la Sección 3, ajustado de manera independiente en cada una de las redes (INDEPENDIENTE, para abreviar).

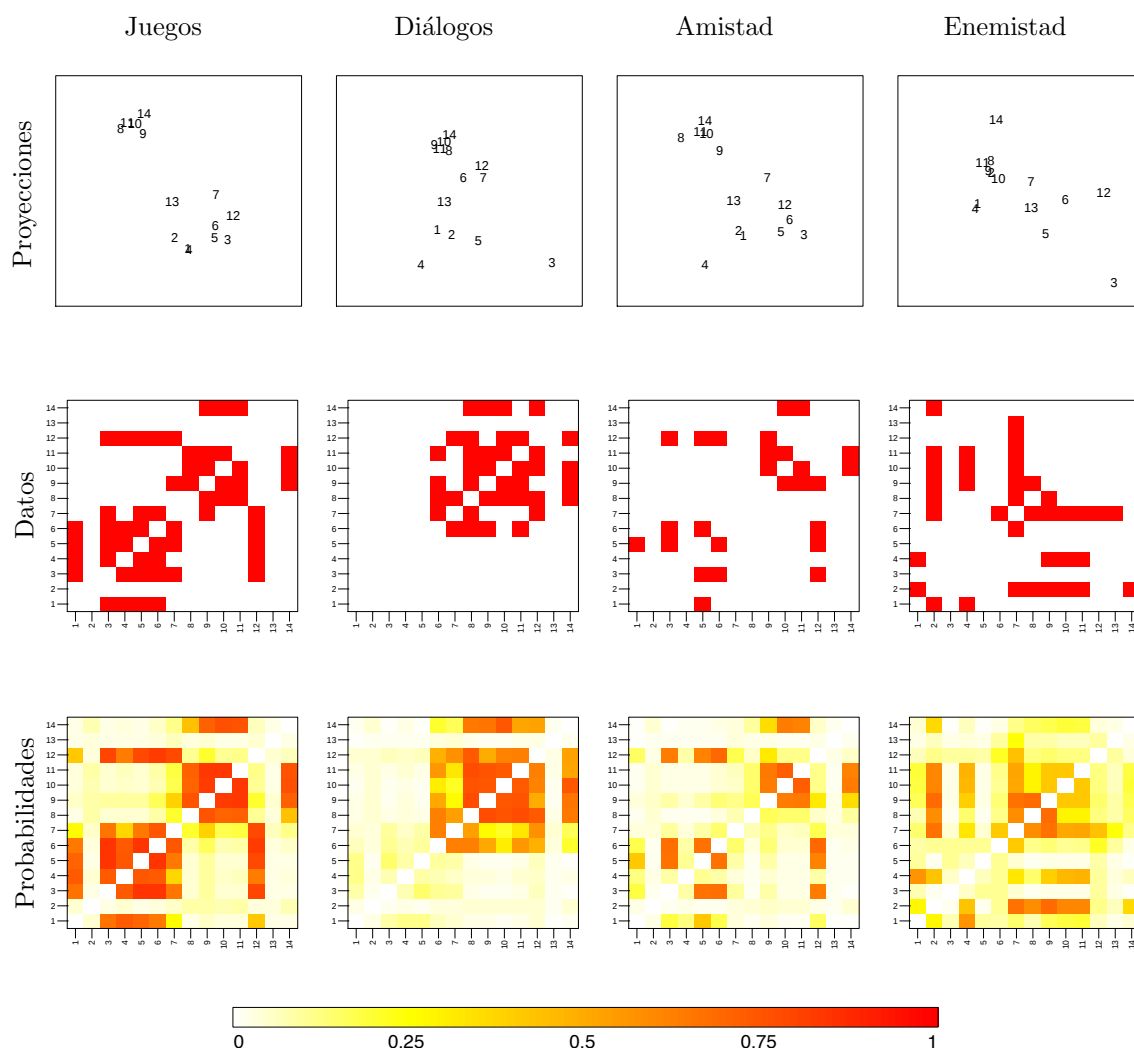


Figura 10: Posiciones latentes $u_{i,j}$ a lo largo de las dos dimensiones con mayor variabilidad, matrices de adyacencia, Y_j , y matrices de probabilidades de interacción estimadas, $\hat{\Theta}_j$, para el conjunto de datos de la sala de cableado.

Fuente: *Elaboración propia*

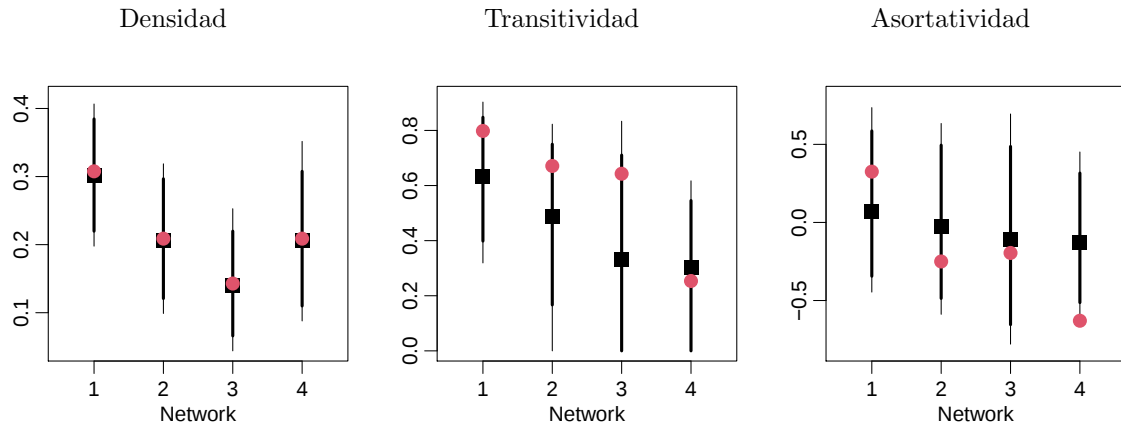


Figura 11: Intervalos de credibilidad al 95% (l nea gruesa) y 99% (l nea delgada), media posterior (cuadrado negro), y valor observado (redondel rojo) de la distribuci n emp rica de la densidad, la transitividad, y la asortatividad de 20,000 r plicas del conjunto de datos.

Fuente: *Elaboraci n propia*

Por lo tanto, para cada combinaci n de, modelo (CONJUNTO, INDEPENDIENTE), conjunto de datos (CABLEADO, HOGARES, SOCIALES, GERENCIA), y dimensi n latente (valor de K en $1, \dots, 6$ que minimiza el WAIC), se ejecuta un experimento de validaci n cruzada de la siguiente manera. Primero, se dividen los datos aleatoriamente en cinco conjuntos de aproximadamente el mismo tama o. A continuaci n, para cada conjunto S , se ajusta el modelo condicional en $D = \{y_{i,i',j} : (i,i',j) \notin S\}$, y para cada y^* en S , se calcula $E[y^* | D]$, la media predictiva posterior de y^* usando todos los datos que no est n en S . Luego, usando tales predicciones, se construye un clasificador binario para obtener la curva *receiver operating characteristic* (ROC) correspondiente. Por  ltimo, se cuantifica la capacidad predictiva de cada curva ROC a trav s del *area under the curve* (AUC). En este contexto, el AUC es una medida de qu  tan capaz es un modelo de predecir enlaces perdidos (cuanto mayor sea el AUC, mejor ser  el modelo para predecir 0's como 0's y 1's como 1's). En todos los casos, las inferencias se basan en muestras de 20,000 muestras de la distribuci n posterior obtenidas despu s de reducir la cadenas de Markov originales cada 25 observaciones y un periodo de calentamiento de 20,000 iteraciones.

Medida	AUC		WAIC		BIC	
Datos	CONJ.	INDE.	CONJ.	INDE.	CONJ.	INDE.
CABLEADO	0.8827	0.8886	199.20	216.48	2,433.00	1,062.82
HOGARES	0.8866	0.7765	2,513.41	3,442.38	19,032.71	22,191.42
SOCIALES	0.8322	0.8340	10,577.20	10,855.31	29,934.11	17,187.44
GERENCIA	0.9036	0.7974	1,505.73	2,267.71	25,299.21	15,116.49

Tabla 2: Valores del AUC, WAIC y del BIC para evaluar respectivamente el poder predictivo y la bondad de ajuste de CONJUNTO e INDEPENDIENTE, utilizando las redes m ltiples de la Tabla 1. El AUC corresponde al valor promedio de los cinco experimentos de validaci n cruzada. Mientras que para el WAIC y BIC corresponde al valor m s peque o de K en $1, \dots, 6$ que minimiza el WAIC.

Fuente: *Elaboraci n propia*

Los resultados de la validaci n cruzada se presentan en la Tabla 2. Estos hallazgos sugieren que en t rminos predictivos, CONJUNTO es igual o m s competitivo que INDEPENDIENTE, lo cual es particularmente evidente en HOGARES y GERENCIA donde la capacidad de predecir enlaces perdidos de CONJUNTO es claramente superior.

Finalmente, para evaluar la bondad de ajuste de cada modelo, se calcula el WAIC (ver Secci n 4 para

más detalles) para todos los conjuntos de datos. La Tabla 2 presenta los resultados para el valor más pequeño de K en $1, \dots, 6$ que minimiza el criterio de información. En este caso, CONJUNTO presenta una bondad de ajuste mejor en todas las instancias, y por lo tanto constituye una mejor opción para caracterizar el proceso aleatorio de da origen a los conjuntos de datos.

7. Discusión

En este trabajo se presenta un enfoque novedoso para modelar un sistema de redes múltiples sobre el mismo conjunto de actores con un método que propicia el flujo de información entre redes, en contraposición a una caracterización independiente de cada una de ellas. Nuestra propuesta se basa en una extensión jerárquica natural de un modelo de distancia de espacio latente, y permite estudiar simultáneamente las posiciones sociales de los actores dentro y entre redes. Además, nuestros experimentos proporcionan evidencia empírica suficiente para establecer que el modelamiento conjunto propuesto es una alternativa muy competitiva en términos de predicción y bondad de ajuste.

El modelo propuesto es susceptible de modificaciones. Por ejemplo, el modelo se puede adaptar fácilmente para incluir los efectos fijos asociados con un conjunto de covariables $\{x_{i,i',j,p}\}$, por medio del predictor lineal $\mathbf{x}_{i,i',j}^\top \boldsymbol{\beta}_j - \|\mathbf{u}_{i,j} - \mathbf{u}_{i',j}\|$, donde $\mathbf{x}_{i,i',j}^\top \boldsymbol{\beta}_j = \sum_{p=1}^P x_{i,i',j,p} \beta_{j,p}$ representa los patrones en los datos asociados con las covariables. Así mismo, en lugar de considerar un modelo de distancia, también es posible hacer una formulación basada en otro tipo de efectos latentes, como en un modelo factorial (Hoff, 2009). Adicionalmente, el modelo propuesto se puede extender para manejar redes no dirigidas por medio de una diferenciación de las posiciones latentes en función de características latentes de “salida”, $\mathbf{u}_{i,j}$, y características latentes de “llegada”, $\mathbf{v}_{i,j}$, de forma que los efectos latentes se puedan expresar como $-\|\mathbf{u}_{i,j} - \mathbf{v}_{i',j}\|$. Finalmente, en mismo espíritu de Green and Hastie (2009), también es posible hacer una especificación trans-dimensional del modelo donde se asigne una distribución previa a la dimensión latente K .

Como se observó en el análisis de las redes de la sala de cableado en la Sección 6, las posiciones sociales son susceptibles de exhibir patrones de agrupamiento, los cuales que se pueden modelar directamente mediante la inclusión de parámetros de asignación a los grupos $\xi_1, \dots, \xi_I$ en la distribución previa. Específicamente, se puede asumir que los actores del sistema están agrupados en H grupos, cada uno ocupando una posición $\boldsymbol{\theta}_h$ en el espacio social; así, a cada uno de los actores se le asigna una posición global $\boldsymbol{\eta}_i$ correspondiente a un desvío normal de la posición del grupo al cual pertenece, de forma que $\boldsymbol{\eta}_i | \{\boldsymbol{\theta}_h\}, \kappa^2, \xi_i \stackrel{\text{ind}}{\sim} \mathcal{N}(\boldsymbol{\theta}_{\xi_i}, \kappa^2 \mathbf{I})$, donde $\xi_i = h$ significa que el actor i hace parte del grupo h .

Gran parte de la literatura de criterios de información se ha centrado en el criterio de información Bayesiano (BIC). Sin embargo, se ha demostrado que el BIC es inapropiado para modelos jerárquicos ya que la estructura jerárquica implica que el número efectivo de parámetros será típicamente menor que el número real de parámetros en la probabilidad. De otra parte, en el contexto de este trabajo no se considera útil el BIC dado que difiere notoriamente del criterio de información de Watanabe-Akaike (WAIC), en que no están motivados por una estimación del ajuste predictivo sino por el objetivo de aproximar la densidad de probabilidad marginal de los datos. Por tal motivo, es completamente posible que el modelo prediga bien y tenga un WAIC bajo, pero, debido a la función de penalización excesiva del BIC (Ver Tabla 2), pueda tener un BIC relativamente alto (es decir, pobre).

Finalmente, se recomienda hacer implementaciones del modelo que permitan estudiar redes “grandes”. Esta es actualmente un área de investigación activa (e.g, Ma and Ma, 2017, Spencer et al., 2020, Aliverti and Russo, 2020).

Referencias

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American statistical Association*, 88(422):669–679.

- Aldous, D. J. (1985). Exchangeability and related topics. In * cole d' t  de Probabilit s de Saint-Flour XIII—1983*, pages 1–198. Springer.
- Aliverti, E. and Russo, M. (2020). Stratified stochastic variational inference for high-dimensional network factor model. *arXiv preprint arXiv:2006.14217*.
- Banerjee, A., Chandrasekhar, A. G., Duflo, E., and Jackson, M. O. (2013). The diffusion of microfinance. *Science*, 341(6144).
- Borg, I. and Groenen, P. J. (2005). *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media.
- D'Angelo, S., Alf , M., and Fop, M. (2020). Model-based clustering for multivariate networks. *arXiv preprint arXiv:2001.05260*.
- D'Angelo, S., Alf , M., and Murphy, T. B. (2018). Node-specific effects in latent space modelling of multidimensional networks. In *49th Scientific meeting of the Italian Statistical Society*.
- Durante, D. and Dunson, D. B. (2014). Nonparametric bayes dynamic modelling of relational data. *Biometrika*, 101(4):883–898.
- Durante, D., Dunson, D. B., et al. (2018). Bayesian inference and testing of group differences in brain networks. *Bayesian Analysis*, 13(1):29–58.
- D'Angelo, S., Murphy, T. B., Alf , M., et al. (2019). Latent space modelling of multidimensional networks with application to the exchange of votes in eurovision song contest. *Annals of Applied Statistics*, 13(2):900–930.
- Gamerman, D. and Lopes, H. F. (2006). *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. CRC Press.
- Gao, L. L., Witten, D., and Bien, J. (2019). Testing for association in multi-view network data. *arXiv preprint arXiv:1909.11640*.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.
- Gelman, A., Hwang, J., and Vehtari, A. (2014). Understanding predictive information criteria for bayesian models. *Statistics and computing*, 24(6):997–1016.
- Gelman, A., Rubin, D. B., et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, 7(4):457–472.
- Gollini, I. and Murphy, T. B. (2016). Joint modeling of multiple network views. *Journal of Computational and Graphical Statistics*, 25(1):246–265.
- Green, P. J. and Hastie, D. I. (2009). Reversible jump mcmc. *Genetics*, 155(3):1391–1403.
- Gupta, S., Sharma, G., and Dukkipati, A. (2018). Evolving latent space model for dynamic networks. *arXiv preprint arXiv:1802.03725*.
- Haario, H., Saksman, E., Tamminen, J., et al. (2001). An adaptive metropolis algorithm. *Bernoulli*, 7(2):223–242.
- Han, Q., Xu, K., and Airoldi, E. (2015). Consistent estimation of dynamic and multi-layer block models. In *International Conference on Machine Learning*, pages 1511–1520.
- Handcock, M. S., Raftery, A. E., and Tantrum, J. M. (2007). Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 170(2):301–354.

- Hoff, P. (2008). Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in neural information processing systems*, pages 657–664.
- Hoff, P. D. (2005). Bilinear mixed-effects models for dyadic data. *Journal of the american Statistical association*, 100(469):286–295.
- Hoff, P. D. (2009). *A first course in Bayesian statistical methods*, volume 580. Springer.
- Hoff, P. D. (2015). Multilinear tensor regression for longitudinal relational data. *The annals of applied statistics*, 9(3):1169.
- Hoff, P. D., Raftery, A. E., and Handcock, M. S. (2002). Latent space approaches to social network analysis. *Journal of the american Statistical association*, 97(460):1090–1098.
- Hoover, D. N. (1982). Row-column exchangeability and a generalized model for probability. *Exchangeability in probability and statistics (Rome, 1981)*, pages 281–291.
- Kim, B., Lee, K. H., Xue, L., and Niu, X. (2018). A review of dynamic network models with latent variables. *Statistics surveys*, 12:105.
- Kolaczyk, E. D. and Csárdi, G. (2014). *Statistical analysis of network data with R*, volume 65. Springer.
- Krackhardt, D. (1987). Cognitive social structures. *Social networks*, 9(2):109–134.
- Krivitsky, P. N. and Handcock, M. S. (2008). Fitting latent cluster models for networks with latentnet. *Journal of Statistical Software*, 24(5).
- Krivitsky, P. N., Handcock, M. S., Raftery, A. E., and Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. *Social networks*, 31(3):204–213.
- Li, W.-J., Yeung, D.-Y., and Zhang, Z. (2011). Generalized latent factor models for social network analysis. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI), Barcelona, Spain*.
- Linkletter, C. D. (2007). *Spatial process models for social network analysis*. PhD thesis, Citeseer.
- Ma, Z. and Ma, Z. (2017). Exploration of large networks with covariates via fast and universal latent space model fitting. *arXiv preprint arXiv:1705.02372*.
- Minhas, S., Hoff, P. D., and Ward, M. D. (2019). Inferential approaches for network analysis: Amen for latent factor models. *Political Analysis*, 27(2):208–222.
- Nowicki, K. and Snijders, T. A. B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American statistical association*, 96(455):1077–1087.
- Paez, M. S., Amini, A. A., and Lin, L. (2019). Hierarchical stochastic block model for community detection in multiplex networks. *arXiv preprint arXiv:1904.05330*.
- Paul, S., Chen, Y., et al. (2016). Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel. *Electronic Journal of Statistics*, 10(2):3807–3870.
- Paul, S., Chen, Y., et al. (2020). Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *The Annals of Statistics*, 48(1):230–250.
- Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using pólya–gamma latent variables. *Journal of the American statistical Association*, 108(504):1339–1349.
- Raftery, A. E., Niu, X., Hoff, P. D., and Yeung, K. Y. (2012). Fast inference for the latent space network model using a case-control approximate likelihood. *Journal of Computational and Graphical Statistics*, 21(4):901–919.

- Reyes, P. and Rodriguez, A. (2016). Stochastic blockmodels for exchangeable collections of networks. *arXiv preprint arXiv:1606.05277*.
- Roethlisberger, F. J. and Dickson, W. J. (2003). *Management and the Worker*, volume 5. Psychology press.
- Salter-Townshend, M. and McCormick, T. H. (2017). Latent space models for multiview network data. *The annals of applied statistics*, 11(3):1217.
- Schweinberger, M. and Snijders, T. (2003). Settings in social networks: A measurement model. *Sociological Methodology*, 33(1):307–341.
- Sewell, D. K. and Chen, Y. (2015). Latent space models for dynamic networks. *Journal of the American Statistical Association*, 110(512):1646–1657.
- Sewell, D. K. and Chen, Y. (2016). Latent space models for dynamic networks with weighted edges. *Social Networks*, 44:105–116.
- Sewell, D. K., Chen, Y., et al. (2017). Latent space approaches to community detection in dynamic networks. *Bayesian Analysis*, 12(2):351–377.
- Sewell, D. K. et al. (2019). Latent space models for network perception data. *Netw. Sci.*, 7(2):160–179.
- Sosa, J. (2017). *A Latent Space Approach for Cognitive Social Structures Modeling and Graphical Record Linkage*. PhD thesis, UC Santa Cruz.
- Spencer, N. A., Junker, B., and Sweet, T. M. (2020). Faster mcmc for gaussian latent position network models. *arXiv preprint arXiv:2006.07687*.
- Swartz, T. B., Gill, P. S., and Muthukumarana, S. (2015). A bayesian approach for the analysis of triadic data in cognitive social structures. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, pages 593–610.
- Turnbull, K. (2020). *Advancements in latent space network modelling*. PhD thesis, Lancaster University.
- Wang, L., Zhang, Z., Dunson, D., et al. (2019). Common and individual structure of brain networks. *The Annals of Applied Statistics*, 13(1):85–112.
- Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11(Dec):3571–3594.
- Zhang, X. (2020). *Statistical Analysis for Network Data using Matrix Variate Models and Latent Space Models*. PhD thesis, The University of Michigan.

A. Algoritmo de MCMC para el modelo de redes m ltiples

Este algoritmo de MCMC itera sobre los par metros del modelo Υ dados en la Secci n 4.1. Donde sea posible, se toman muestras de las distribuciones condicionales completas como en un muestreador de Gibbs usual; de lo contrario, se usan versiones adaptativas con pasos de Metropolis-Hastings (e.g., Haario et al., 2001). Alternativamente, en el mismo esp ritu de Albert and Chib (1993), se pueden introducir variables aleatorias con distribuci n Polya-Gamma para facilitar el c culo (Polson et al., 2013).

En este caso, la distribuci n posterior conjunta est  dada por:

$$p(\Upsilon \mid \mathbf{Y}) = \prod_{j,i < i'} \text{Ber}(y_{i,i',j} \mid \theta_{i,i',j}) \times \prod_{i,j} \text{N}(\mathbf{u}_{i,j} \mid \boldsymbol{\eta}_i, \sigma_j^2 \mathbf{I}) \times \prod_j \text{N}(\zeta_j \mid \mu_\zeta, \tau_\zeta^2) \times \prod_j \text{lGam}(\sigma_j^2 \mid \alpha_\sigma, \beta_\sigma)$$

$$\begin{aligned} & \times \prod_i \mathbf{N}(\boldsymbol{\eta}_i \mid \boldsymbol{\theta}, \kappa^2 \mathbf{I}) \times \mathbf{N}(\boldsymbol{\theta} \mid \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \times \text{IGam}(\kappa^2 \mid \alpha_\kappa, \beta_\kappa) \times \mathbf{N}(\mu_\zeta \mid \mu_0, \sigma_0^2) \times \text{IGam}(\tau_\zeta^2 \mid a_\zeta, b_\zeta) \\ & \times \text{Unif}(\alpha_\sigma \mid c_\sigma, d_\sigma) \times \text{Gam}(\beta_\sigma \mid a_\sigma, b_\sigma), \end{aligned}$$

donde $\theta_{i,i'} = \text{expit}(\zeta_j - \|\mathbf{u}_{i,j} - \mathbf{u}_{i',j}\|)$ denota la probabilidad de una interacción entre los actores i e i' .

Para un conjunto de hiper-parámetros dado, $\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0, a_\kappa, b_\kappa, \mu_0, \sigma_0^2, a_\zeta, b_\zeta, a_\sigma, b_\sigma, c_\sigma, d_\sigma$, el algoritmo genera un nuevo estado $\boldsymbol{\Upsilon}^{(b+1)}$ a partir de un estado actual $\boldsymbol{\Upsilon}^{(b)}$, $b = 1, \dots, B$, de la siguiente manera:

1. Muestrear $\mathbf{u}_{i,j}^{(b+1)}$, $i = 1, \dots, I$, $j = 1, \dots, J$, siguiendo algoritmo de Metropolis-Hastings, considerando la distribución condicional completa:

$$\begin{aligned} p(\mathbf{u}_{i,j} \mid \text{rest}) & \propto \prod_{i': i < i'} \text{Ber}(y_{i,i',j} \mid \text{expit}(\zeta_j - \|\mathbf{u}_{i,j} - \mathbf{u}_{i',j}\|)) \\ & \times \prod_{i': i > i'} \text{Ber}(y_{i',i,j} \mid \text{expit}(\zeta_j - \|\mathbf{u}_{i',j} - \mathbf{u}_{i,j}\|)) \times \mathbf{N}(\mathbf{u}_{i,j} \mid \boldsymbol{\eta}_i, \sigma_j^2 \mathbf{I}). \end{aligned}$$

2. Muestrear $\zeta_j^{(b+1)}$, $j = 1, \dots, J$, siguiendo un algoritmo de Metropolis-Hastings, considerando la distribución condicional completa:

$$p(\zeta_j \mid \text{rest}) \propto \prod_{i,i': i < i'} \text{Ber}(\text{expit}(\zeta_j - \|\mathbf{u}_{i,j} - \mathbf{u}_{i',j}\|)) \times \mathbf{N}(\zeta_j \mid \mu_\zeta, \tau_\zeta^2).$$

3. Muestrear $(\sigma_j^2)^{(b+1)}$, $j = 1, \dots, J$, de una distribución Gamma Inversa:

$$\sigma_j^2 \mid \text{rest} \sim \text{IGam}\left(\alpha_\sigma + \frac{IK}{2}, \beta_\sigma + \frac{1}{2} \sum_i (\mathbf{u}_{i,j} - \boldsymbol{\eta}_i)^\top (\mathbf{u}_{i,j} - \boldsymbol{\eta}_i)\right).$$

4. Muestrear $\boldsymbol{\eta}_i^{(b+1)}$, $i = 1, \dots, I$, de una distribución Normal multivariada:

$$\boldsymbol{\eta}_i \mid \text{rest} \sim \mathbf{N}\left(\left[\frac{1}{\kappa^2} + \sum_j \frac{1}{\sigma_j^2}\right]^{-1} \left[\frac{1}{\kappa^2} \boldsymbol{\theta} + \sum_j \frac{1}{\sigma_j^2} \mathbf{u}_{i,j}\right], \left[\frac{1}{\kappa^2} + \sum_j \frac{1}{\sigma_j^2}\right]^{-1} \mathbf{I}\right).$$

5. Muestrear $\boldsymbol{\theta}$ de una distribución Normal multivariada:

$$\boldsymbol{\theta} \mid \text{rest} \sim \mathbf{N}\left(\left[\boldsymbol{\Sigma}_0^{-1} + \frac{I}{\kappa^2} \mathbf{I}\right]^{-1} \left[\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0 + \frac{1}{\kappa^2} \sum_i \boldsymbol{\eta}_i\right], \left[\boldsymbol{\Sigma}_0^{-1} + \frac{I}{\kappa^2} \mathbf{I}\right]^{-1}\right).$$

6. Muestrear κ^2 de una distribución Gamma Inversa:

$$\kappa_h^2 \mid \text{rest} \sim \text{IGam}\left(\alpha_\kappa + \frac{IK}{2}, \beta_\kappa + \frac{1}{2} \sum_i (\boldsymbol{\eta}_i - \boldsymbol{\theta})^\top (\boldsymbol{\eta}_i - \boldsymbol{\theta})\right).$$

7. Muestrear μ_ζ de una distribución Normal:

$$\mu_\zeta \mid \text{rest} \sim \mathbf{N}\left(\left[\frac{1}{\sigma_0^2} + \frac{J}{\tau_\zeta^2}\right]^{-1} \left[\frac{\mu_0}{\sigma_0^2} + \frac{\sum_j \zeta_j}{\tau_\zeta^2}\right], \left[\frac{1}{\sigma_0^2} + \frac{J}{\tau_\zeta^2}\right]^{-1}\right).$$

8. Muestrear τ_ζ^2 de una distribución Gamma Inversa:

$$\tau_\zeta^2 \mid \text{rest} \sim \text{IGam}\left(a_\zeta + \frac{J}{2}, b_\zeta + \frac{1}{2} \sum_j (\zeta_j - \mu_\zeta)^2\right).$$

9. Muestrear α_σ siguiendo un algoritmo de Metropolis-Hastings, considerando la distribuci3n condicional completa:

$$\log p(\alpha_\sigma \mid \text{rest}) \propto J[\alpha_\sigma \log \beta_\sigma - \log \Gamma(\alpha_\sigma)] - \alpha_\sigma \sum_j \log \sigma_j^2.$$

10. Muestrear β_σ de una distribuci3n Gamma:

$$\beta_\sigma \mid \text{rest} \sim \text{Gam} \left(a_\sigma + J\alpha_\sigma, b_\sigma + \sum_j \frac{1}{\sigma_j^2} \right).$$

B. Notaci3n

El valor absoluto de un ńmero real x se denota con $|x|$ y la funci3n Gamma con $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$. Las matrices y los vectores se escriben con negrillas. Por ejemplo, $\mathbf{x} = (x_1, \dots, x_n)$ denota un vector columna de tama1o $n \times 1$. Adem1s, se usa $\mathbf{0}$ y $\mathbf{1}$ para denotar los vectores columna cuyas entradas son todas iguales a 0 y 1, respectivamente, e $\mathbf{I}$ para denotar la matriz identidad. La transpuesta de un vector $\mathbf{x}$ se denota con $\mathbf{x}^\top$; an1logamente para las matrices. Adem1s, si $\mathbf{X}$ es una matriz cuadrada, se usa $\text{tr}(\mathbf{X})$ para denotar su traza y $\mathbf{X}^{-1}$ para denotar su inversa. La norma de un vector $\mathbf{x}$, dada por $\sqrt{\mathbf{x}^\top \mathbf{x}}$, se denota con $\|\mathbf{x}\|$. De otra parte, $p(\cdot \mid \cdot)$ se usa para denotar una funci3n de densidad de probabilidad condicional; de manera similar con $p(\cdot)$, la cual denota una distribuci3n de probabilidad marginal. La misma notaci3n se usa tanto para funciones de densidad continua como para funciones de masa de probabilidad discreta. Aunque es un abuso de notaci3n, la distribuci3n condicional completa de un par1metro θ dado el resto de los par1metros y los datos se denota con $p(\theta \mid \text{rest})$.