Análisis de grafos bipartitos en función de múltiples proyecciones: aplicación a un caso de estudio histórico

Rodrigo René Cura, Julio E. Vezub

CENPAT - CONICET, UNPSJB

Abstract. Los grafos bipartitos son una representación natural en muchos sistemas complejos al describir relaciones entre dos tipos diferentes de actores o entidades. Una forma común de estudiarlos es a partir de la proyección sobre uno de los conjuntos de nodos para aplicar técnicas de Análisis de Redes Sociales (ARS). Es sabido que este tipo de proyección siempre contiene menos información que la red bipartita original y se han realizado muchos esfuerzos en la búsqueda de un método ideal que sintetice el grafo en la forma más fiel posible, derivando en incontables maneras de hacerlo, cada una con diferentes características. En este artículo presentamos una estrategia de análisis de un grafo bipartito a partir del uso de múltiples proyecciones, con el fin de obtener un conjunto de vistas con variados enfoques orientados a hacer más evidentes algunas propiedades subyacentes de la estructura de la red original. Aquí mostramos los resultados de aplicar esta técnica al estudio de las redes epistolares entre caciques patagónicos, comerciantes, hacendados, religiosos y autoridades político-militares argentinas y chilenas de la segunda mitad del siglo XIX. A su vez, comparamos la información obtenida en las etapas previas del trabajo con la misma fuente de datos, que estuvo basada en la representación de la red bajo la forma de grafo tradicional.

1 Introducción

En el marco del proyecto "Redes Sociales Indígenas y Formación del estado en Cuyo, Pampa y Patagonia (1850-1900)"[1] se inició durante 2012 la construcción de una base de datos unificada para las distintas regiones que abarca la investigación, metodológicamente orientada al almacenamiento, sistematización y análisis de información relativa a los actores sociopolíticos relevantes, las relaciones y los acontecimientos que protagonizaron en las tres regiones durante la segunda mitad del siglo XIX. Ello se planteó con el objetivo de aportar visiones macro, o de conjunto, que permitan reconstruir la dinámica relacional entre indígenas, criollos e "indígeno-criollos"[2–4] en aquello que uno de los pioneros del análisis de redes, Maurizio Gribaudi[5], ha definido como la reconstrucción de la morfología social en una "escala 1:1", la que se configura a partir de datos micro como los del parentesco, la economía, la circulación de información, los conflictos, las alianzas y las jerarquías políticas en el contexto del proceso de formación

y expansión del Estado nacional argentino hacia las periferias regionales y provinciales. Dicho más claramente, nos propusimos describir y analizar de la manera más vasta que sea posible el conjunto de lazos estructurantes entre las poblaciones que habitaban las tolderías, ranchos, estancias y fortines del siglo XIX. La comprensión de los fenómenos sociales como "configuraciones"[1] en la dirección planteada por Norbert Élias[6], y la construcción de una base de datos a partir del registro de fuentes históricas, nos permitirá a su vez la identificación de los "trending topics" presentes en los documentos, editados o inéditos, mediante la aplicación de minería de textos.

En una etapa más modesta de la construcción de la base se decidió comenzar por la correspondencia que se enviaban entre sí los actores de las fronteras cuyanas, pampeanas y patagónicas. Esto fue así por dos motivos, el primero porque la circulación epistolar del período articulaba redes de información que deben ser concebidas como un condensado comunicacional de redes sociales en un sentido más amplio[7] [2]. El segundo motivo es que las cartas suministran metadatos sencillos que pueden ser rápidamente organizados, como ser la fecha, el remitente y destinatario, los lugares de remisión y destino, membretes, tratamientos formales entre los corresponsales, secretarios o escribientes, seguimiento de la secuencia epistolar, etc. Esto facilitó el diseño experimental de la base mediante una etapa previa a la minería de textos más compleja sobre el conjunto del contenido de las cartas que los actores intercambiaban.

Llamativamente, si bien el ARS tiene antecedentes muy importantes en las Ciencias Sociales, al igual que la colaboración interdisciplinaria entre estas y la Informática, su aplicación para el análisis de redes de correspondencia ha sido menos explorado[8–12]. En el caso del período y las regiones que nos ocupan, la práctica de la escritura entre los caciques que habían sido alfabetizados en las misiones religiosas fronterizas desde el siglo XVIII, o que contaban en su entorno con amanuenses europeos y criollos, ha pasado mayormente inadvertida hasta muy recientemente, principalmente por los prejuicios de historiadores y antropológos que estigmatizaron durante décadas a las poblaciones indígenas del siglo XIX por su carácter supuestamente ágrafo, condenándolas a la sempiterna oralidad del bárbaro y el salvaje.

Fue entonces a partir de la advertencia de esta evidencia histórica de las prácticas escriturales y de archivo de los caciques, que en trabajos previos contribuimos a dilucidar[13, 14], cuando se planteó la necesidad

[1] Élias (1991) definió la categoría de la "configuración", para resaltar el carácter relacional e interdependiente de los actos humanos en cualquier.

[2] Para ello nos valemos de Bourdieu y su noción de "capital" relacionada con los fenómenos característicos del poder del Estado, incluido el "capital informacional": "El Estado es el resultado de un proceso de concentración de los diferentes tipos de capital, capital de fuerza física o de instrumentos de coerción (ejército, policía), capital económico, capital cultural o, mejor dicho informacional, capital simbólico, concentración que, en tanto que tal, convierte al Estado en poseedor de una especie de metacapital, otorgando poder sobre las demás clases de capital y sobre sus poseedores"

de recurrir al intercambio entre científicos sociales e informáticos para profundizar el conocimiento del pasado mediante métodos que combinaran las técnicas heurísticas clásicas de los documentos con innovaciones cualitativas y cuantitativas para la interpretación histórica. Dentro de este esquema las cartas intercambiadas por los actores indígenas, criollos e indígeno-criollos fueron las fuentes privilegiadas para la conformación de la plataforma de datos y el comienzo del primer diseño de un modelo de base. Las cartas funcionan como los componentes primordiales para la reconstrucción de la red social, conectando diferentes tipos de entidades como actores, lugares y eventos, proporcionando como ya se dijo una visión panorámica de las prácticas y los lazos sociales.

En la primera instancia, orientada a aplicar técnicas de análisis de redes sociales, y por la naturaleza de las cartas como objeto de relación entre las entidades, se diseño un primer modelo que representaba a las mismas como aristas en una red de entidades como que a su vez fueron definidas como nodos, siguiendo lo aplicado en los casos afines que habían sido planteados por los antecedentes[8, 11, 12]. Esta estrategia funcionó para establecer vínculos entre actores y determinar la intensidad de los mismos en función del volumen de correspondencia entre dichos actores, o para observar la distribución geográfica de los remitentes y destinatarios de un actor en particular entre otros casos. No obstante estos progresos, la simplificación del modelo hizo perder mucha información subyacente de las relaciones entre las diferentes entidades que una misma carta congrega, por ejemplo, su respectiva localización espacial, como se verá de inmediato.

En virtud de esas dificultades iniciales, proponemos en el presente trabajo una redefinición del modelo utilizado hasta el momento, y la formalización del mismo con la intención de obtener un nuevo modelo general para ser aplicado en casos similares o al extender los volúmenes de la base. Le concedemos un papel decisivo a las cartas como concentradoras de información sobre las entidades, e intentamos maximizar los datos que pueden recolectarse de ellas mediante diferentes vistas de la gran red, denominadas proyecciones. Finalmente, en base a nuestro caso de estudio planteamos algunas proyecciones particulares y evaluamos la respuesta del modelo hacia ellas.

El resto del artículo está organizado como sigue. En la sección 2 presentaremos el primer enfoque que habíamos dado a la construcción de un grafo en función de la red de correspondencia, en la sección 3 se explicará la reestructuración del modelo como un grafo bipartito, la sección 4 definirá las proyecciones, y dentro de estas las que utilizamos en nuestro caso de estudio. Por último, en la sección 5 se discuten las conclusiones y se establece el lineamiento a seguir en los trabajos futuros.

2 Primer enfoque: Grafo de modo uno

El modelo desarrollado originalmente para la construcción de la red de correspondencia otorgaba a las cartas el rol de relación entre actores, lugares

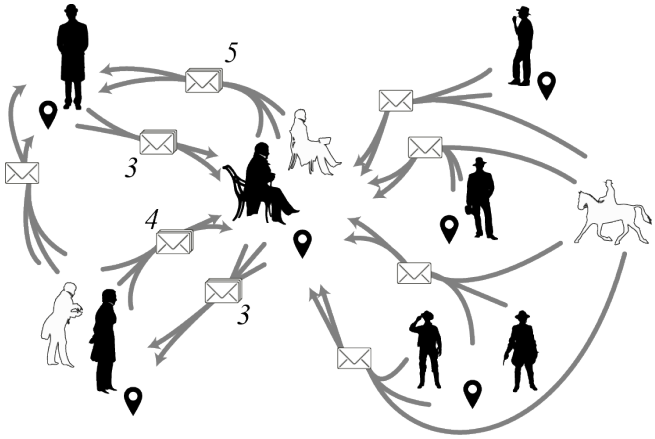

Fig. 1. Una red de ejemplo reducida, construida en base a las características observadas en el corpus documental. En negro los actores remitentes y destinatarios, en blanco los escribientes. El punto negro indica la locación geográfica de remitentes y destinatarios según lo registrado en la carta.

y/o eventos, lo que derivó en una representación como un grafo dirigido con pesos.

Definición 1 (Modelo 1). *Sea G' un grafo de la forma*

$$G' = \langle X, A', w \rangle$$

donde X es el conjunto de nodos, A' el conjunto de aristas, tal que $A' \subseteq (X \times X)$ y la función $w(a'_k) : A' \to \mathbb{R}$ es el peso de $a'_k \in A'$.

En este modelo, una carta no está conceptualmente definida, sino que está representada en forma factorizada en un conjunto de relaciones uno a uno entre las entidades que conecta, apreciable en la figura 2.

Con esta estrategia no existe forma de determinar que conjunto de aristas representan a una misma carta y limita al modelo en la identificación de las relaciones subyacentes entre las entidades más allá de la relación de correspondencia.

3 Re-estructuración: Grafo bipartito

En consecuencia a la perdida de información que surge de aplicar el modelo 1, se presenta a continuación la definición de un segundo modelo para la representación de la red como un grafo bipartito dirigido.

Definición 2 (Modelo 2). *Sea G un grafo bipartito de la forma*

$$G = \langle X \cup Y, A \rangle$$

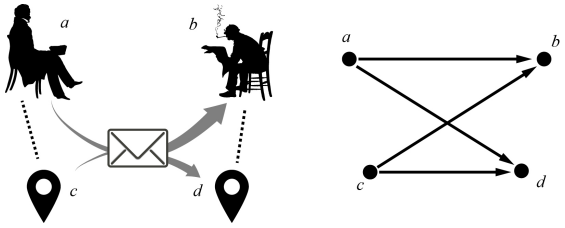

Fig. 2. El actor a desde la locación c envía una carta al actor b en la locación d (izquierda). Grafo que representa la red de correspondencia según el modelo 1 (derecha). Se puede observar como el grafo pierde las relaciones actor/locación indicadas por líneas punteadas y factoriza la carta en cuatro aristas.

con X e Y los conjuntos disjuntos de nodos y A el conjunto de aristas dirigidas tal que $A \subseteq (X \times Y) \cup (Y \times X)$.

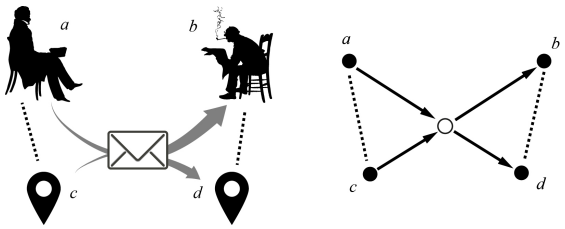

Fig. 3. Las líneas punteadas indican las relaciones indirectas entre entidades que son parte del origen o destino de una misma carta.

Dar a la carta la forma de nodo hace posible ver con mayor claridad el flujo de información y facilita inferir las relaciones indirectas entre las entidades que agrupa. En la figura 3 se observa el caso anterior, pero con su respectiva representación según el modelo 2.

3.1 Efecto de la re-estructuración aplicada al ejemplo

Para la clarificación de lo presentado hasta el momento tomaremos la red presentada en el ejemplo en la figura 1 y aplicaremos los modelos definidos.

Según el modelo 1 los actores y entidades de la red son representados como nodos, y las cartas como el conjunto de aristas obtenidas del producto cartesiano de los nodos origen (remitentes) y destino (destinatarios). En la figura 4 se observa la representación gráfica del resultado. Aquí se pueden ver las relaciones entre los remitentes y destinatarios por su distribución espacial, pero se debe tener en cuenta que esta distribución fue dada manualmente para conseguir una

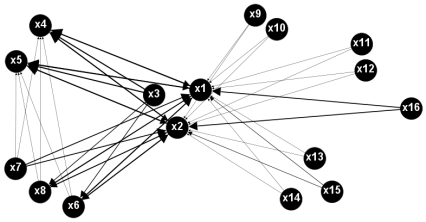

Fig. 4. Grafo resultante del caso de ejemplo en base al modelo 1.

mayor claridad en la asociación del modelo con la red de origen de la figura 1 y la posibilidad de calcularlas matemáticamente se perdió.

Al aplicar el modelo 2 las cartas aparecen como nodos concentradores de entidades donde, de la misma forma que con el primer modelo, se puede observar el flujo de la información por las entidades de la red, pero además, es fácil inferir las relaciones entre remitentes y destinatarios de una misma carta, información perdida al usar el modelo 1.

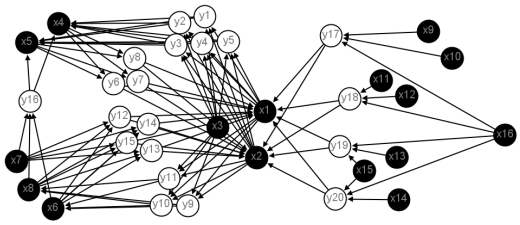

Fig. 5. Grafo resultante del caso de ejemplo en base al modelo 2. En negro las entidades y en blanco las cartas.

4 Proyecciones

Una proyección puede verse como una función que transforma un grafo bipartito en un grafo de modo uno, ya sea dirigido o no, mediante la aplicación de algún tipo de estrategia particular. Aquí presentamos la definición formal que utilizaremos en este trabajo.

Definición 3 (Proyección). *Sea* $G' = \langle X, A', w \rangle$ *un grafo de modo uno y* $G = \langle X \cup Y, A \rangle$ *un grafo bipartito, según lo definido en los modelos 1 y 2 respectivamente. Se define la proyección sobre* X *de* G *como*

$$P_X : G \to G'$$

con $w(a'_k)$ calculada en función de los elementos de Y adyacentes a los elementos de X conectados por la arista $a'_k \in A'$.

Según la estrategia utilizada y la definición de la función w un grafo puede tener un conjunto de proyecciones diferentes y cada una de ellas reflejar variados aspectos de la información contenida en él. Así mismo, el grafo resultante de aplicar la proyección puede variar en sus características en múltiples formas, por ejemplo, ser dirigido o no dirigido, con pesos o sin ellos si se ignora la salida de w, etc.

Para el caso concreto de la red de correspondencia planteamos dos proyecciones en particular, ambas calculadas en función de la matriz de adyacencia del grafo origen. Antes de pasar a la descripción de las mismas mostramos la formalización de la matriz correspondiente al grafo bipartito para presentar sus propiedades.

Definición 4 (Matriz de adyacencia). *Sea G un grafo bipartito se define su matriz de adyacencia como*

$$\boldsymbol{M_G} = \begin{bmatrix} \mathbf{0} & \boldsymbol{M_X} \\ \boldsymbol{M_Y} & \mathbf{0} \end{bmatrix}_{p \times p}$$

tal que $\boldsymbol{M_X} = [mx_{ij}]_{n \times m}$ sea la matriz de adyacencia que contiene las aristas de X hacia Y, $\boldsymbol{M_Y} = [my_{ij}]_{m \times n}$ sea la matriz de adyacencia que contiene las aristas de Y hacia X y $n + m = p$.

4.1 Proyecciones aplicadas al caso de estudio

En nuestro caso de estudio las proyecciones experimentales se definieron sobre el conjunto X del grafo G [3], por representar este a las entidades que se desea estudiar. A continuación se define, se explica el objetivo y se muestra el resultado sobre el caso de ejemplo de cada una de ellas.

Definición 5. *Sea G un grafo bipartito y M_G su matriz de adyacencia según se definió anteriormente, se define a la proyección Pa_X sobre X de G como*

$$Pa_X(G) = M_X M_Y$$

donde el valor en cada una de las posiciones de la matriz es el peso de la arista correspondiente.

La intención de la proyección Pa_X es sintetizar las relaciones epistolares entre las entidades mediante el trazado de aristas directas entre remitentes y destinatarios. El peso de las aristas representa la intensidad del enlace de comunicación en base a la cantidad de cartas de una entidad a otra. Se puede observar en la estructura de la matriz, al no ser simétrica, que el grafo resultante

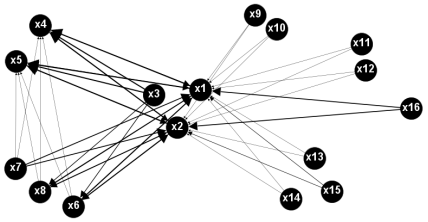

Fig. 6. Grafo resultante de aplicar Pa_X a G.

es dirigido. En la figura 6 se puede observar el resultado de la proyección del grafo de ejemplo.

Es fácil notar que el grafo resultante de esta primera proyección es isomorfo al grafo de la red de correspondencia obtenido en función del modelo 1 y, como se mencionó anteriormente, se sabe que una proyección siempre contiene menos información que el grafo bipartito original. A partir de esto, es posible concluir que el modelo 2 contiene más información de la red que el 1.

Definición 6. *Sea G un grafo bipartito y M_G su matriz de adyacencia según se definió anteriormente, se define a la proyección Pb_X sobre X de G como*

$$Pb_X(G) = \begin{cases} 0, & i = j \\ pb_{ij}, & i \neq j \end{cases} : pb_{ij} \in M_X M_X^\top + M_Y^\top M_Y$$

con el peso de las aristas en cada una de las posiciones de la matriz.

La proyección Pb_X se enfoca en extraer las comunidades de nodos asociadas como remitentes o destinatarios comunes. Se puede apreciar los fuertes vínculos entre los actores políticos con sus respectivas ubicaciones geográficas y sus escribientes, en el caso de contar con alguno. En el caso particular del nodo x_{16}, presenta una estructura peculiar al relacionarse con varios pares de nodos. Según los antecedentes, ese nodo coincide con un escribiente itinerante que recolectaba correspondencia a lo largo de su ruta para entregarla a x_1. En la figura 7 se muestra el grafo resultante de esta proyección.

5 Conclusiones y trabajo futuro

Entre las varias conclusiones que se pueden obtener a partir de los resultados obtenidos en este trabajo, queremos destacar que el uso de una grafo bipartito para el estudio de redes de correspondencia (o cualquier fenómeno

[3] La forma en que definimos las proyecciones tienen fines únicamente experimentales, este trabajo no tiene la intención de tratar de definir un algoritmo ideal para proyectar grafos bipartitos, por esta razón se le concedió primacía a la simplicidad de cálculo y perspectiva de la información obtenida de las mismas.

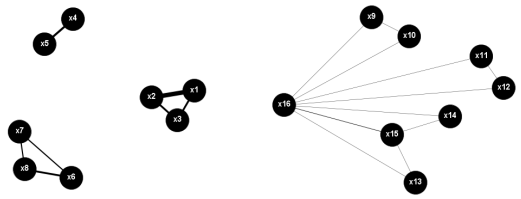

Fig. 7. Grafo resultante de aplicar Pb_X a G.

comunicacional) es más adecuado que la representación como un grafo simple por varias razones: 1) se trata de un modelo más fiel al original; 2) un diseño de esta forma requiere menos esfuerzo en la construcción de la abstracción, 3) se puede visualizar que contiene más información al obtener una representación isomorfa al grafo modo uno original mediante la aplicación de una proyección y, finalmente, 4) que en base a lo último es sencillo identificar que al formalizar una red que naturalmente responde a una estructura bipartita como un grafo de modo uno, se está realizando implícitamente una proyección arbitraria de la misma en la que se producía una perdida de información.

Se puede entonces apreciar, que la estrategia de estudiar un grafo bipartito por medio de múltiples proyecciones, hace posible poner un lente sobre diferentes características subyacentes a la red bipartita original, donde cada una de ellas revela diferentes conjuntos de información, en ocasiones oculta a una primera mirada.

Como trabajo futuro, se planea profundizar en este caso de estudio de las redes de correspondencia, avanzando en la búsqueda de un modelo general y sus aplicaciones. Además, se espera desarrollar un conjunto de proyecciones en base a indicadores primordiales de este tipo de redes sociales.

Referencias

1. Escolar, D., Salomón Tarquini, C. y Vezub, J. E.: Redes sociales indígenas y formación del estado en Cuyo, Pampa y Patagonia (1850-1900), Proyecto PICT 2011-1457, Agencia de Promoción Científica y Tecnológica. (2011, ms.).
2. Vezub, J. E. y Salomón Tarquini, C: Cartografía de redes indígeno-criollas y guerras de producción de soberanía estatal en Pampa y Patagonia (1870-1890). CONICET, PIP - Grupo de Investigación 2012-2014, Cód. Nro. 11220110100937. Ms.
3. Escolar, D.: El estado montonero: archivos indigeno-criollos, legalidad y rebelión en Cuyo siglos XIX-XX. Coloquio internacional: Les sociétés indiennes entre classifications ethnographiques et mémoire collective. Université Rennes 2, Collectif Histoire et Anthropologie sur L'Amérique Latin, Centre de Recherches Historiques de L'Ouest, 21, 22 y 23 de marzo del 2013.
4. Vezub, J. E.: El proceso de popularización indígeno-criollo en Pampa y Patagonia del siglo XIX. Fradkin R. y Di Meglio, G. (eds.), Hacer política. La participación popular en el siglo XIX rioplatense, Buenos Aires, Prometeo Libros, 2013.

5. Acha, O. y Vezub, J. E.: Entre la microhistoria y la morfología. Mundo académico y recorridos personales. Una entrevista con Maurizio Gribaudi. Revista de Historia, Departamento de Historia de la Universidad Nacional de Mar del Plata, 2005. pp. 187-213.
6. Élias, N.: Qu'est-ce que la sociologie? Paris, Éditions de l'aube, 1991.
7. Bourdieu, P.: Razones prácticas sobre la teoría de la acción. Barcelona, Anagrama (1999).
8. Barabási, A.L.: Linked: The New Science of Networks. Perseus publishing, Cambridge (2002).
9. Hanneman, R. A. & Mark Riddle: Introduction to social network methods. Riverside, CA: University of California, Riverside (published in digital form at http://faculty.ucr.edu/ hanneman/), (2005).
10. Miceli, J. y Guerrero S.: Redes libres de escala y su uso en el análisis de datos etnográficos: El caso de la comunidad tehuelche de Chalía, en Grupo Antropocaos / Diego Díaz et.al., Exploraciones en antropología y complejidad, Buenos Aires, Editorial SB (2007).
11. Reynoso, C.: Redes sociales y complejidad. Modelos interdisciplinarios en la gestión sostenible de la sociedad y la cultura, Buenos Aires, Editorial SB (2011).
12. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press (1994).
13. Vezub, J. E.: Valentín Saygüeque y la "Gobernación Indígena de las Manzanas". Poder y etnicidad en la Patagonia Septentrional (1860-1881). Buenos Aires, Prometeo Libros (2013).
14. Vezub, J. E.: Mapuche-Tehuelche Spanish Writing and the Argentinian-Chilean Expansion during the Nineteenth Century. Delmas A. & Penn, N. (eds.), Written Culture in a Colonial Context. Africa and the Americas, 1500-1900. Cape Town University Press (2011).