

**INTRODUCCIÓN A LOS MÉTODOS DEL
ANÁLISIS DE REDES SOCIALES.**

CAPÍTULO PRIMERO

**Robert A. Hanneman. Departamento de Sociología de la
Universidad de California Riverside**

NOTA PREVIA

Este documento está traducido para la lista REDES con permiso del autor a partir de la versión electrónica disponible en <http://wizard.ucr.edu/~rhannema/networks/text/textindex.html> [Fecha de consulta: Octubre de 2000].

La traducción se ha realizado por bloques. En este documento se presenta el primer capítulo, traducido por Maria Ángela Petrizzo y revisado por José Luis Molina.

Lazos y relaciones, datos de redes y datos reticulares son utilizados en la traducción como expresiones equivalentes.

SOBRE ESTE LIBRO

Este libro introduce muchos de los enfoques del análisis formal de las redes sociales, proporcionando breves referencias y ejemplos de la mayoría de sus áreas de desarrollo más importantes. El texto se apoya en las obras de Freeman, Borgatti y Everett (los autores del programa UCINET). Los materiales y su organización, están inspirados en el texto de Wasserman y Faust, y en un seminario de postgrado dirigido por el profesor Phillip Bonacich en la UCLA en 1998. Los errores y omisiones, por supuesto, son responsabilidad del autor.

TABLA DE CONTENIDOS

1. Los datos de las redes sociales.
 2. ¿Por qué métodos formales?
 3. Utilizando gráficos para representar relaciones sociales.
 4. Utilizando matrices para representar relaciones sociales.
 5. Propiedades básicas de las redes y los actores
 6. Centralidad y poder.
 7. Camarillas y subgrupos.
 8. Posiciones en la red y roles sociales: el análisis de la equivalencia.
 9. Equivalencia estructural.
 10. Equivalencia automórfica.
 11. Equivalencia regular.
- Bibliografía de trabajos relacionados con (o ejemplos de) métodos de redes sociales.

CAPÍTULO I . LOS DATOS DE LAS REDES SOCIALES.

INTRODUCCIÓN: ¿SON DIFERENTES LOS DATOS DE LAS REDES SOCIALES?

De entrada, no hay nada inusual en los datos utilizados en el análisis de redes sociales. Es cierto que quienes trabajan con redes utilizan un lenguaje especializado para describir la estructura y los contenidos de los conjuntos de observaciones estudiados, pero los datos de redes también pueden describirse y entenderse utilizando las perspectivas y los conceptos propios de métodos más familiares como la investigación sociológica tradicional (*cross-sectional survey research*).

Sin embargo, a menudo los datos que los investigadores en redes sociales estudian resultan muy distintos de la tradicional matriz rectangular tan familiar para los investigadores y el análisis estadístico. Las diferencias son importantes porque nos llevan a observar los datos de una manera distinta –y también a pensar de manera distinta-- sobre cómo aplicar los procedimientos estadísticos.

Los datos sociológicos “convencionales” están compuestos por una matriz rectangular con mediciones. Las filas de la matriz son los casos, sujetos u observaciones. Las columnas son las puntuaciones (cuantitativas o cualitativas) de los atributos, variables o mediciones. Entonces, cada celda de la matriz describe la puntuación de algún actor con respecto a algún atributo. En algunos casos, estas matrices pueden tener una tercera dimensión que representa los cuadros de observaciones de múltiples grupos.

<i>Nombre</i>	<i>Sexo</i>	<i>Edad</i>	<i>Grado de Inclusión</i>
Bob	Hombre	32	2
Carol	Mujer	27	1
Ted	Hombre	29	1
Alice	Mujer	28	3

La estructura de datos fundamental es aquella que permite comparar qué tan parecidos o diferentes son los actores entre sí en base a los atributos (comparando las filas). O quizá más frecuentemente, examinemos qué tan parecidas o diferentes son las variables entre sí (comparando las columnas).

Los datos de la “red” (en su forma más pura), constituyen una matriz cuadrada de mediciones. Las filas de la matriz son los casos, sujetos y observaciones. Las columnas son el mismo conjunto de casos, sujetos y observaciones –allí está la diferencia clave con los datos convencionales. En cada celda de la matriz se describe una relación entre los actores.

¿Quién dice relacionarse con quién?				
	Elección			
Informante	Bob	Carol	Ted	Alice
Bob	---	0	1	1
Carol	1	---	0	1
Ted	0	1	---	1
Alice	1	0	0	---

Podemos observar esta estructura de datos de la misma forma que los datos atributivos. Comparando las filas de la matriz, podemos observar qué actores se parecen a otros en la persona que han escogido. Mirando a las columnas, podemos ver quién es parecido a quién en términos de haber sido escogido por los demás. Estas formas son muy útiles para observar los datos, puesto que sirven para ver aquellos actores que tienen posiciones similares en la red. Este es el primer punto importante del análisis de redes: buscar qué actores está ubicados o insertados en la red.

Pero a un analista de redes sociales también le gusta mirar la estructura de datos de una forma distinta: globalmente. El analista puede notar que en la matriz existen cantidades iguales de ceros y unos. Esto sugiere que hay una “densidad” moderada

de preferencias. También puede comparar las celdas que están por encima y por debajo de la diagonal para mirar si hay reciprocidad entre las opciones (por ejemplo: Bob escogió a Ted pero ... ¿Ted escogió a Bob?). Éste es el segundo punto importante del análisis de redes: buscar cómo la estructura de las opciones individuales se abstrae en patrones más generales.

Es posible pensar en los datos de la red social en los mismos términos que los “datos convencionales”. Uno puede pensar que las filas son simplemente una lista de casos, y las columnas atributos de cada actor (por ejemplo, las relaciones con otros actores podrían verse como “atributos” de cada actor). De hecho, muchas de las técnicas utilizadas por los analistas de redes (como el cálculo de correlaciones y distancias) se aplican exactamente de la misma forma en datos de red y en datos convencionales.

Sin embargo, mientras pueden considerarse los datos de la red como una forma especial de datos convencionales (y de hecho lo son), los analistas de redes observan los datos de redes de forma diferente. Más que pensar en cómo la relación de un actor con los demás describe el atributo de su “ego”, los analistas de redes observan la estructura de conexiones en la que el actor se encuentra involucrado. Los actores se describen a través de sus relaciones, no de sus atributos. Y las relaciones en sí mismas son tan fundamentales como los actores que se conectan a través de ellas.

La mayor diferencia entre los datos convencionales y los reticulares, es que los datos convencionales se centran en actores y atributos mientras los datos de red se centran en actores y relaciones. La diferencia en el énfasis es muy importante para las opciones que el investigador debe tomar en el diseño de la investigación, en el muestreo controlado, el desarrollo de mediciones y en el manejo de los datos resultantes. No se trata que las herramientas de investigación sean distintas entre los analistas de redes y los científicos sociales (a menudo no lo son). Se trata de que los propósitos específicos y el énfasis de la investigación de redes sociales implican consideraciones diferentes.

En este capítulo, estudiaremos algunos de los temas que han surgido en el diseño, muestreo y medición de redes sociales. Nuestra discusión se centrará en los dos aspectos de los datos de la red: los nodos (o actores) y las líneas (o relaciones). Trataremos de mostrar algunas de las formas en las que los datos de redes son similares o diferentes de los datos de actores y sus atributos más familiares. Introduciremos algunos términos nuevos que facilitan la descripción de las características especiales de los datos reticulares. Finalmente, discutiremos con

brevidad cómo las diferencias entre los datos reticulares y los de actor-atributo son importantes en la aplicación de las herramientas estadísticas.

Nodos

Los datos de red se definen por actores y por relaciones (o por nodos y por vínculos). La parte de los datos de la red correspondiente a nodos o actores debería ser bastante precisa. Otros enfoques empíricos en ciencias sociales también trabajan en términos de casos, sujetos o elementos representativos y semejanzas. Sin embargo, existe una diferencia en la mayoría de los datos de redes —de hecho se convierte en una gran diferencia—y es la forma en la que los datos son recogidos comúnmente — y en la que son estudiadas las poblaciones y muestras.

El análisis de redes se centra en las relaciones entre los actores, y no en los actores individuales y sus atributos. Esto significa que los actores a menudo no son muestreados de forma independiente, como en muchos otros tipos de estudios (más frecuentemente en encuestas). Supongamos que estudiamos los vínculos de amistad, por ejemplo. John ha sido escogido para estar en nuestra muestra. Cuando le preguntamos a él, John identifica siete amigos. También necesitamos rastrear a cada uno de esos siete amigos, y preguntarles sobre sus vínculos de amistad. Los siete amigos están en nuestra muestra en tanto que John lo está (y viceversa). Por tanto, los “elementos de la muestra” no son en absoluto “independientes”.

Los nodos o actores incluidos en estudios que no son de redes, tienden a ser el resultado de un muestreo probabilístico independiente. Los estudios de redes intentan incluir todos los actores que se muestran en un límite (que a menudo se muestra de forma natural). Con frecuencia los estudios de redes no utilizan “muestras” en absoluto, al menos en el sentido convencional. Al contrario, intentan incluir todos los actores en una población o poblaciones determinada(s). Por supuesto, las poblaciones incluidas en un estudio de red pueden ser una muestra de un conjunto grande de poblaciones. Por ejemplo, cuando estudiamos los patrones de interacción entre estudiantes en las clases, incluimos todos los niños que están en el salón de clases (es decir, estudiamos la totalidad de la población del aula). El aula en sí misma, debe seleccionarse a través de métodos probabilísticos dentro de una población de aulas (todas las de una escuela)

El uso de poblaciones completas como forma de seleccionar observaciones en muchos estudios de redes, hace que sea importante para el analista tener claros los límites de cada población que vaya a ser estudiada y la forma en la que se seleccionarán, de entre esa población, las unidades individuales de observación. El análisis de redes frecuentemente implica diferentes niveles de análisis, con actores agrupados en un nivel inferior (por ejemplo, la red puede describirse de forma “anidada”).

POBLACIONES, MUESTRAS Y LÍMITES

Como hemos dicho, rara vez los analistas de redes sociales diseñan muestras. Muy a menudo, identifican una población y dirigen un censo (por ejemplo, incluyen todos los elementos de la población como unidades de observación). Un analista puede examinar todos los sustantivos y objetos presentes en un texto, todas las personas en una fiesta de cumpleaños, todos los miembros de un grupo de parientes, de una organización, de un vecindario o una clase social (por ejemplo, los propietarios de tierras en una región, o reino).

Los métodos de investigación empírica a menudo utilizan un enfoque muy distinto para decidir qué nodos estudiar. Se hace una lista de todos los nodos (a veces estratificada o agrupada) y se seleccionan los elementos individuales por métodos probabilísticos. La lógica del método trata a cada individuo como una réplica separada, es decir, en cierto modo intercambiable con cualquier otro.

Debido a que los métodos de redes se centran en las relaciones entre los actores, éstos no pueden ser muestreados independientemente para incluirlos como observaciones. Si un actor llega a ser seleccionado, han de incluirse todos los actores con los cuales éste tiene (o puede tener) relaciones. Como resultado, los enfoques de red intentan estudiar poblaciones totales a través del censo, más que a través de la muestra (discutiremos brevemente algunas excepciones de esto en el apartado de muestreo de relaciones).

Las poblaciones que los analistas de redes estudian son notablemente diversas. En un extremo, pueden constar de símbolos en textos o sonidos en pronunciaciones; en el otro, naciones en el mundo y sistemas de estados pueden constituir la lista de nodos. Quizás las más comunes, por supuesto, son aquellas formadas por personas

individuales. En cada caso, sin embargo, los elementos de la población a ser estudiada se definen a través de algún límite.

Los límites de las poblaciones estudiadas por los analistas de redes pueden ser, básicamente, de dos tipos. El primero de ellos, probablemente el más común, es el que viene creado o impuesto por los propios actores. Todos los miembros de una aula, organización, club, vecindario o comunidad pueden constituir una población de este tipo. Son estos grupos o redes articuladas de manera natural. Por tanto, los estudios de redes sociales trazan a menudo en cierto modo los límites alrededor de una población que se sabe, *a priori*, que es una red. Alternativamente, un analista de redes puede asumir un enfoque más “demográfico” o “ecológico” para definir los límites de la población. Podemos diseñar observaciones contactando toda la gente que se encuentra en un área espacial delimitada, o a quienes poseen un cierto rasgo (por ejemplo, tener un ingreso familiar de alrededor de un millón de dólares al año). En este caso hay motivos para sospechar que la red existe, pero la entidad a ser estudiada es una agregación abstracta impuesta por el investigador –más que un patrón de acción social institucionalizada que haya sido identificado y etiquetado por sus participantes.

Los analistas de redes pueden expandir los límites de sus estudios haciendo réplicas de las poblaciones. Más que estudiar un vecindario, a menudo se estudian varios. Este tipo de diseño (que puede utilizar métodos de muestreo para seleccionar las poblaciones), permite la replicación y la verificación de hipótesis a través de la comparación de poblaciones. Una segunda, y también muy importante forma en la que los estudios de redes expanden su ámbito es a través de la inclusión de múltiples niveles o modalidades de análisis.

MODALIDADES Y NIVELES DE ANÁLISIS

El analista de redes tiende a ver a las personas inmersas en redes de relaciones directas con otras personas. A menudo, estas redes de relaciones interpersonales se convierten en “hechos sociales” y toman vida propia. Una familia, por ejemplo, es una red de relaciones cercanas entre un conjunto de personas. Pero esta red en particular, ha sido institucionalizada y ha recibido un nombre y una realidad más allá de los nodos que la componen. Los individuos en sus relaciones de trabajo pueden ser observados como inmersos en organizaciones; en sus relaciones de ocio pueden estar inmersos en asociaciones voluntarias. Los vecindarios, comunidades y aun las

sociedades son, con variedad de grados, entidades sociales en sí mismas. Y, como entidades sociales, pueden formar lazos con los individuos inmersos con ellas y con otras entidades sociales.

A menudo, los conjuntos de datos de redes describen los nodos y las relaciones entre éstos para una sola población delimitada. Si se estudian los patrones de amistad entre los estudiantes en una aula, estaríamos haciendo un estudio de este tipo. Pero un aula está dentro de una escuela –que puede ser vista como una red de clases relacionadas entre sí y otros actores (directores, administradores, bibliotecarios, etcétera). Y estas escuelas existen en distritos escolares, los cuales pueden verse a su vez como redes de escuelas y otros actores (comités escolares, grupos de investigación, departamentos de personal y adquisiciones). Puede incluso, haber patrones de relaciones entre los distritos escolares (a través de intercambio de escolares, profesores, materiales curriculares, etcétera).

La mayor parte de quienes trabajan con redes ven a las personas como enlazadas a redes, que están enlazadas a redes que a su vez están también enlazadas a redes. Describen tales estructuras como “multimodales”. En nuestra escuela del ejemplo, los estudiantes individuales y los profesores forman un modo, las aulas de clases otro modo, las escuelas un tercer modo, y así sucesivamente. Un conjunto de datos que contiene información sobre dos tipos de entidades sociales (por ejemplo personas y organizaciones) es una red de dos modos.

Por supuesto, este tipo de perspectiva de la naturaleza de las estructuras sociales no es única de los investigadores de redes. El análisis estadístico se ocupa de los mismo temas en sus diseños “jerárquicos” o anidados. Los teóricos hablan de los niveles macro-meso-micro de análisis, o desarrollan esquemas para identificar diferentes niveles de análisis (el más usual en sociología es individual, grupal, organizacional, comunitario, institucional, social y orden global). Una ventaja de la perspectiva de redes es que de una forma natural predispone al analista a focalizar en múltiples niveles de análisis simultáneamente. Esto es, el analista de redes está siempre interesado en cómo el individuo está integrado en una estructura y cómo la estructura emerge de las micro relaciones entre partes individuales. La habilidad de los métodos de redes para representar tal relación multimodal es, al menos potencialmente, un paso adelante en el rigor del análisis.

Habiendo afirmado que los métodos de redes sociales son particularmente adecuados para trabajar con múltiples niveles de análisis y estructuras de datos multimodales,

debe admitirse que los analistas de redes rara vez aprovechan esta característica. La mayoría de los análisis de red no se mueven más allá del simple micro o macro reduccionismo – y es correcto. Pocos análisis en todo caso de redes han intentado trabajar con más de dos modos simultáneamente. Y, aún cuando ha sido así, la estrategia más común ha sido examinarlos de forma más o menos separada (una excepción es el análisis conjunto de redes de dos modos)

RELACIONES

La otra parte de la investigación de datos reticulares tiene que ver con qué lazos o relaciones serán medidas en los nodos seleccionados. En este punto hay dos temas importantes que discutir. En muchos estudios de redes se estudian todos los lazos, o un tipo específico entre todos los nodos seleccionados – es decir, se realiza un censo. Pero, a veces se utilizan diferentes enfoques (porque son menos costosos o porque es necesario generalizar), es decir, un muestreo de relaciones. A veces en los datos de la red también hay otro tipo de muestreo de relaciones. Cualquier conjunto de actores puede estar conectado con diferentes tipos de lazos y relaciones (por ejemplo, los estudiantes en un aula de clases pueden gustar o no de los demás, pueden o no jugar juntos, pueden o no compartir la misma comida, etcétera). Cuando se recogen los datos de la red, con frecuencia se seleccionan, o muestrean, los tipos de relaciones que serán medidos entre un conjunto de tipos de relaciones.

MUESTREO DE RELACIONES

Dado un conjunto de actores o nodos, hay algunas estrategias para decidir cómo actuar en la recolección de medidas de las relaciones existentes. Al final de la gama de enfoques, se encuentran los métodos de “redes completas”. Este enfoque aporta el máximo de información, pero también es costoso y difícil de utilizar y puede ser de difícil generalización. En el otro extremo se encuentran métodos muy parecidos a los utilizados en la investigación empírica convencional. Estos enfoques aportan considerablemente menos información sobre la estructura de la red, pero a menudo son menos costosos y también facilitan la generalización desde las observaciones en la muestra hacia el total de la población. Sin embargo, no existe un método “correcto” para todos los problemas y preguntas de investigación.

Los Métodos de redes completas requieren que se recoja información acerca de los lazos de cada actor con los demás. En esencia, este enfoque utiliza un censo de los lazos en una población de actores –más que un muestreo de éstos. Por ejemplo, podemos recoger datos sobre el comercio de cobre entre pares de naciones-Estado en el sistema mundial, tomándolos de los datos del FMI; podemos examinar las juntas directivas de todas las corporaciones públicas superponiendo a los directivos; podemos examinar el número de vehículos que transitan entre pares de ciudades; podemos mirar el flujo de correos electrónicos entre todos los pares de empleados de una compañía; podemos decirle a cada niño en un grupo de juegos que identifique a sus amigos entre los demás.

Ya que recolectamos información sobre lazos entre pares o díadas, los datos de redes completas aportan una fotografía completa de las relaciones en la población. La mayoría de los enfoques y métodos especiales de análisis de redes que discutiremos a lo largo de este texto, fueron desarrollados para utilizarse con datos de redes completas. Los datos de estas redes son necesarios para definirla apropiadamente y para medir muchos de los conceptos estructurales del análisis de red (por ejemplo el grado de intermediación)

Los datos de redes completas nos conducen a descripciones muy potentes y a análisis de estructuras sociales. Desafortunadamente, pueden también ser muy costosos y difíciles de obtener. Obtener datos de cada miembro de una población y tener cada rango o índice de cada uno de los demás, pueden ser tareas desafiantes para cualquier grupo, excepto para los pequeños. La labor se hace más manejable pidiendo a informantes que identifiquen un número limitado de individuos específicos con los cuales éstos tienen relaciones. Entonces, estas listas se compilan e interconectan. Pero, para grupos grandes (por ejemplo, todas las personas de una ciudad) la labor es prácticamente imposible.

En muchos casos, los problemas no son tan graves como puede pensarse. La mayoría de las personas, grupos y organizaciones, tienden a tener un número limitado de lazos –o al menos un número limitado de lazos fuertes. Esto probablemente se deba a que los actores sociales tienen recursos limitados, energía, tiempo y capacidad cognitiva –y no pueden mantener un gran número de lazos fuertes. También es cierto que las estructuras sociales pueden desarrollar un grado considerable de orden y solidaridad con relativamente pocas conexiones.

Los métodos de “bola de nieve” comienzan focalizando un actor o conjunto de actores, a cada uno de los cuales se les pregunta por algunos de sus lazos con otros actores. Entonces, se toman todos los actores mencionados (que no sean parte de la lista original) y se les pregunta de nuevo por algunos de sus lazos. El proceso continúa hasta que no se identifiquen nuevos actores o hasta que se decida detenerlo (a menudo por motivos de tiempo y recursos, o porque los nuevos actores identificados son muy marginales con respecto del grupo que se intenta estudiar).

El método de bola de nieve puede ser muy útil para el seguimiento de poblaciones “especiales” (a menudo subconjuntos de gente numéricamente pequeños, mezclados con otros más grandes). Las redes de contactos de negocios, élites de la comunidad, subculturas desviadas, filatélicos ávidos, redes de parentesco y muchas otras estructuras pueden ser efectivamente ubicadas y descritas a través de métodos de bola de nieve. A veces no es difícil alcanzar la completitud en “muestreos” de bola de nieve. Las limitaciones en los números de lazos fuertes que tienen la mayoría de los actores, y la tendencia de los lazos a ser recíprocos, a menudo facilitan la búsqueda de límites.

Hay dos grandes restricciones y debilidades potenciales de los métodos de bola de nieve. En primer lugar, los actores que no están conectados (es decir, “aislados”) no se pueden describir a través de este método. Su presencia y cantidad, puede ser una característica muy importante de las poblaciones para algunos propósitos analíticos. El método de bola de nieve tiende a exagerar la “conectividad” y “solidaridad” de las poblaciones de actores. En segundo lugar, no hay ninguna forma garantizada de encontrar todos los individuos conectados dentro de la población. ¿Dónde echar a rodar la bola de nieve? Si comenzamos en el lugar o lugares equivocado(s), podemos perder todo los subconjuntos de actores que se encuentran conectados –pero no anexionados a nuestros puntos de partida.

Los enfoques de bola de nieve pueden reforzarse pensando con cuidado cómo seleccionar los nodos iniciales. En muchos estudios puede haber un punto de partida natural. En estudios sobre el poder en comunidades, por ejemplo, es común comenzar con los jefes ejecutivos de las organizaciones económicas, culturales y políticas más grandes. Mientras este enfoque no pretenda considerar la mayor parte de la comunidad (aquellos individuos que están “aislados” de la élite de la red), el enfoque identificará probablemente la élite social con bastante eficiencia.

Redes ego-céntricas (con conexiones a otros) . En muchos casos no será posible (o necesario) trazar la totalidad de las redes comenzando con nodos focales (como en el método de bola de nieve). Un enfoque alternativo es comenzar con una selección de nodos focales (egos), e identificar los nodos con los que ellos están conectados. Entonces determinamos cuáles de los nodos identificados en la primera fase están conectados con los demás. Esto puede hacerse contactando cada uno de los nodos; a veces podemos pedir a ego que nos informe qué nodos con los que está conectado tienen lazos con otros.

Este tipo de enfoque puede resultar muy efectivo para obtener un formulario de datos relacionales de poblaciones grandes y puede combinarse con enfoques basados en los atributos. Por ejemplo, podemos tomar una muestra simple aleatoria de estudiantes universitarios varones, y pedirles que identifiquen a sus amigos más cercanos y cuáles de éstos conocen a otros. Este tipo de enfoque puede darnos una imagen buena y fiable de los tipos de redes (o al menos de los vecindarios locales) en los cuales se encuentra insertos los individuos. Podemos obtener resultados tales como cuántos nodos de conexiones tienen y hasta qué puntos esos nodos forman núcleos fuertes. Tales datos pueden ser muy útiles para entender las oportunidades y restricciones que tienen los individuos, como resultado de la forma en la que están insertos en sus redes.

El enfoque egocéntrico con conexiones a otros también puede aportar alguna información sobre la red en su totalidad, vista no ya con un enfoque de censo o de bola de nieve. Tales datos son, de hecho, conjuntos de datos de micro redes – muestreos de áreas locales de redes grandes. Muchas propiedades de red –distancia, centralidad y varios tipos de equivalencia posicional-- no pueden evaluarse con datos egocéntricos. Sin embargo, algunas propiedades tales como la densidad de toda la red pueden estimarse de razonablemente con datos egocéntricos, mientras que otras, como la prominencia de lazos recíprocos, las camarillas y otras similares pueden ser estimadas bastante directamente.

Redes Ego-céntricas (sólo individuos)

Los métodos egocéntricos realmente se centran en el individuo más que en la red como un todo. Recolectando información de los vínculos entre los actores conectados con cada individuo, podemos incluso hacer una buena fotografía de las redes “locales” o “vecindarios” de individuos. Esa información es muy útil para entender cómo las

redes afectan a los individuos, y ellos también proporcionan una fotografía (incompleta) del entramado de la red como un todo.

Supongamos, sin embargo, que obtenemos información de los vínculos del individuo con los otros, pero no sobre las conexiones entre esos otros individuos. Datos de este tipo no constituyen realmente datos de una "red" como tal. Es decir, no pueden representarse como una matriz cuadrada de lazos actor * actor. Pero no significa que los datos egocéntricos sin conexiones entre los otros no sirvan para que los analistas utilicen un enfoque estructural o de red para entender a los actores. Podemos saber, por ejemplo, que algunos actores tienen muchos amigos muy cercanos y parientes y que otros tienen pocos. Sabiendo esto, podemos entender algo acerca de las diferencias en la localización de los actores en la estructura social, y hacer algunas predicciones sobre cómo esos puestos limitan su comportamiento. Lo que no podemos saber con ninguna certeza, partiendo de datos egocéntricos, es la naturaleza de la estructura marco de la red en su totalidad.

En redes egocéntricas, los individuos identificados y conectados con cada ego probablemente son un conjunto que no está conectado con aquellos que si lo están a otros egos. Mientras no podamos valorar la densidad total de la conectividad de la población, debemos movernos en un ámbito muy general. Si tenemos alguna razón buena razón teórica para pensar acerca de los otros en términos de sus roles sociales, más que como ocupantes individuales de tales roles, las redes egocéntricas pueden decirnos mucho acerca de las estructuras sociales. Por ejemplo, si identificamos cada uno de los otros conectados a un ego a través de una relación de amistad como "pariente", "compañero de trabajo", "miembro de la misma iglesia", etcétera, podemos construir una fotografía de las redes de las posiciones sociales (más que las redes de individuos) en las cuales se encuentran inmersos los egos. Tal enfoque, por supuesto, asume que categorías como "pariente" son reales e importantes para los patrones de interacción.

RELACIONES MÚLTIPLES

En un conjunto de datos convencional del tipo actor * atributo, cada actor es descrito a través de muchas variables (y cada variable es atribuida a varios actores). En el tipo más común de conjunto de datos de redes sociales sobre actor * relación, sólo se describe un tipo de relación. En tanto estemos interesados en múltiples atributos de

los actores, estaremos interesados en múltiples tipos de relaciones que conectan a éstos en la red.

Observando la red de relaciones en un departamento académico de una Facultad, por ejemplo, estaremos interesados en qué profesores tienen estudiantes en común, prestan servicios a los mismos comités, son amigos fuera del lugar de trabajo, tienen una o más áreas de especialización en común y son coautores de artículos. Las posiciones que tienen los actores en la red de afiliaciones de grupos tienen múltiples facetas. Las posiciones en un conjunto de relaciones pueden reforzar o contradecir las posiciones en otro (uno puede compartir lazos de amistad con un conjunto de personas con los cuales no se trabaja en comités, por ejemplo). Los actores pueden estar enlazados a otros de forma cercana en una red relacional, pero pueden estar muy distantes de otros en una red relacional diferente. Las ubicaciones de los actores en redes multirelacionales y la estructura de redes compuestas de múltiples relaciones, son unas de las áreas más interesantes y aún relativamente poco exploradas, del análisis social .

Cuando recolectamos datos de redes sociales acerca de ciertos tipos de relaciones entre actores, estamos, en cierto modo, haciendo un muestreo de una población de relaciones posibles. A menudo, nuestra teoría y la pregunta de investigación, indican qué tipo de relaciones entre los actores consideramos más importantes para el estudio. En este caso no se hace un muestreo de relaciones, sino una selección de éstas. En un estudio vinculando la dependencia y el crecimiento económicos, por ejemplo, se podrían recoger datos sobre el intercambio de espectáculos de músicos entre naciones --aunque probablemente estos datos no sean los más relevantes.

¿Si no sabemos qué relaciones estudiar, cómo podemos decidirnos? Existen algunos enfoques conceptuales que pueden ayudarnos. La teoría de sistemas, por ejemplo, sugiere dos dominios el material y el informacional.

Las cosas materiales se “mantienen”, en el sentido en que sólo pueden ubicarse en un nodo de la red cada vez. Los movimientos de personas entre organizaciones, intercambio de dinero entre particulares, de automóviles entre ciudades y otros similares, son ejemplos de cosas materiales que se mueven entre nodos y, por ello, establecen una red de relaciones materiales. Lo informacional, para los teóricos de sistemas “no se mantiene”, en tanto que puede estar en más de un lugar al mismo tiempo. Si yo sé algo y lo comparto contigo, ambos lo sabremos. En cierto sentido, el hecho de compartir información puede ser utilizado para establecer lazos entre dos

nodos. Sin embargo, hay que tener la precaución de no confundir la simple posesión de un atributo común (por ejemplo el género) con la presencia de un lazo (por ejemplo el intercambio de puntos de vistas entre dos personas acerca de temas de género).

Las metodologías para trabajar con datos multirelacionales no están tan desarrolladas como aquellas que trabajan con relaciones simples. Muchas áreas de trabajo interesantes tal como la correlación de redes, escalado y agrupación multidimensional, y álgebra de roles se han desarrollado para trabajar con datos multirelacionales. La mayor parte de estas temáticas están más allá del espacio cubierto por este texto y es mejor abordarlos una vez se ha desarrollado el análisis básico de redes de relaciones simples.

ESCALAS DE MEDIDA

Como otros tipos de datos, también la información que se recoge sobre los lazos entre los actores puede ser medida en diferentes “niveles de medición” (podemos por ejemplo asignar puntuaciones a nuestras observaciones). Los diferentes niveles de medición son importantes, ya que limitan los tipos de preguntas que el investigador puede plantearse. Las escalas de medida también son importantes ya que cada tipo de escala tiene diferentes propiedades matemáticas y exige la utilización de algoritmos diferentes para describir patrones y ensayar inferencias sobre ellos.

Generalmente se distingue tres tipos de niveles de medición: nominal, ordinal y de intervalo (el nivel de razón puede, para cualquier propósito práctico, ser agrupado en la categoría intervalo). Es útil, sin embargo, dividir la medida nominal en variaciones binarias y de categoría múltiple; también es útil distinguir entre medidas ordinales de rango completo y medidas ordinales agrupadas. Describiremos brevemente estas variaciones y presentaremos ejemplos de cómo pueden aplicarse en estudios de redes sociales.

Medidas binarias de relaciones. En la mayoría de ocasiones para asignar números a relaciones es fácil distinguir entre relaciones ausentes (codificadas con un cero) y lazos que están presentes (codificados con un uno). Si le pedimos a informantes en una encuesta que comenten “¿con cuáles de estas otras personas de la lista simpatiza?”, estaremos haciendo una medición binaria. Cada persona seleccionada de la lista será codificada con uno. Los no seleccionados estarán codificados con cero.

La mayor parte del desarrollo de la teoría de grafos en matemáticas y muchos algoritmos para la medición de propiedades de actores y redes, han sido desarrollados a través de datos binarios. Los datos binarios son ampliamente utilizados en análisis de redes, así que no es poco frecuente encontrar datos medidos en el nivel “más alto” transformados en puntuaciones binarias antes de proceder al análisis. Para hacer esto, uno simplemente selecciona algún “punto de corte” y vuelve a recodificar los casos por debajo de ese punto (con cero) o por encima de él (con uno). Los datos binarios implican desperdicio de información. El analista necesita considerar qué es relevante (por ejemplo ¿a qué se refiere la hipótesis? ¿se refiere a la presencia de estructuras de relaciones, o a la fortaleza de éstos?), y qué algoritmos deben aplicarse para decidir si es razonable recodificar los datos. Muy a menudo, el poder y simplicidad del análisis de datos binarios compensa la pérdida de información.

Medidas nominales de categoría múltiple de las relaciones En la recolección de datos podemos pedir a nuestros informantes que miren una lista de individuos y nos digan: “para cada persona en la lista, qué categoría que describe mejor su relación con ella: amistad, amor, relación de negocios, parentesco o ninguna relación”. Debemos puntuar cada persona en la lista según tenga una relación de tipo “1”, “2” etcétera. Este tipo de escala es nominal o cualitativa –cada relación de la persona hacia el sujeto es codificada por su tipo, más que por su fortaleza. A diferencia de los datos binarios nominales (verdadero – falso), la medida nominal de categoría múltiple es una selección múltiple.

El enfoque más común para analizar medidas nominales de categorías múltiples, es utilizarlo para crear series de medidas binarias. Es decir, podemos tomar los datos a partir de la pregunta anterior y crear conjuntos separados de puntuaciones por lazos de amistad, por lazos sentimentales, por lazos de parentesco, etcétera. Esto se parece mucho a una “codificación simulada”, como una forma de manejar muchos tipos de medidas de selección múltiple en análisis estadístico. Al examinar los datos resultantes, sin embargo, puede recordarse que, en la mayoría de las redes resultantes, a cada nodo se le permite un solo tipo de relación. Es decir, una persona puede tener un lazo de amistad o uno sentimental –pero no ambos- como resultado de la forma en que hemos realizado la pregunta. Examinando las redes resultantes, las densidades pueden bajar artificialmente, y existirá una inherente correlación negativa entre las matrices.

Medidas ordinales agrupadas de relaciones: Una de las primeras tradiciones en el estudio de redes sociales interroga a los informantes sobre el índice en que otros les

“agradan”, “desagradan” o les son “indiferentes”. El resultado es una escala ordinal agrupada (por ejemplo, puede haber más de una persona que “agrade” y las categorías reflejan una línea importante en el orden de intensidad). A menudo, este tipo de escala de tres puntos ha sido codificada: -1, 0 y +1, para reflejar el agrado negativo, indiferencia y agrado positivo. Cuando se puntúa de esta forma, los signos positivos y negativos, hacen más fácil escribir algoritmos que tendrán en cuenta y describirán diferentes propiedades de la red (por ejemplo el equilibrio estructural del grafo).

Las medidas ordinales agrupadas pueden utilizarse para reflejar una gran cantidad de diferentes aspectos cuantitativos de las relaciones. Los analistas de redes a menudo están involucrados en la descripción de la “fuerza” de los lazos. Pero “fuerza” puede significar muchos tipos de cosas diferentes. Una dimensión es la frecuencia de la interacción - ¿los actores tienen contacto diariamente, semanalmente, mensualmente, etcétera. Otra dimensión es la “intensidad”, lo cual a menudo refleja el grado de implicación emocional asociado con la relación (por ejemplo, los lazos de parentesco pueden ser infrecuentes, pero implican una “carga emocional” debido a las expectativas altamente ritualizadas e institucionalizadas). Las relaciones pueden ser más fuertes si involucran muchos contextos diferentes o tipos de lazos. Totalizando los datos nominales en función la presencia o ausencia de múltiples tipos de relaciones, surge una escala ordinal (de hecho, un intervalo) de una dimensión que mide la fuerza de la relación. Las relaciones recíprocas también parecen ser más fuertes. Normalmente podemos valorar la reciprocidad pidiendo a cada actor de una díada que nos diga sus preferencias sobre los demás. Sin embargo, también se puede preguntar a cada actor acerca de sus percepciones del grado de reciprocidad en una relación: “¿Quiere decir que ustedes no se agradan mutuamente, que a usted le agrada más X que usted a él, que a X le agrada usted más que él a usted, o que se agradan mutuamente?”

Las escalas ordinales de medición contienen más información que las nominales. Es decir, las puntuaciones reflejan sutiles diferencias de la fortaleza del lazo más que una simple “presencia o ausencia”. Esto debe verse como algo positivo. Sin embargo con frecuencia es difícil aprovechar los datos ordinales. Los algoritmos más comúnmente usados para el análisis de redes sociales han sido diseñados para datos binarios. Muchos han sido adaptados para datos continuos – pero para escalas de intervalo más que para escalas ordinales de medición. Los datos ordinales, en consecuencia, a menudo son convertidos en binarios escogiendo algún punto de corte y recodificándolos. De forma alternativa, los datos ordinales a veces son tratados como

si ellos fueran realmente intervalos. La primera estrategia, sin embargo, tiene el riesgo de la adecuada selección de los puntos de corte; la última estrategia también tiene riesgos ya que los intervalos de puntos separados en una escala ordinal pueden resultar muy heterogéneos.

Medidas ordinales del ranking de relaciones: A veces es posible puntuar la fortaleza de todas las relaciones de un actor en un orden de posición desde la más fuerte hasta la más débil. Por ejemplo, puedo pedir a cada informante que escriba un “1” junto al nombre de la persona en la clase que más le agrada, un “2” junto al nombre que le agrada en segundo término, etcétera. La clase de escala que resultaría de esto, sería una “escala total del orden de posiciones”. Tales escalas reflejan las diferencias de grado de intensidad, pero no necesariamente diferencias iguales –es decir, la diferencia entre mi primera y segunda selecciones no necesariamente es la misma que la diferencia entre mi segunda y tercera selecciones. Cada relación, sin embargo, tiene una puntuación única (1era, 2da, 3era, etcétera)

Las medidas ordinales del ranking de relaciones son poco frecuentes en la literatura de análisis de redes sociales, así como también en la mayoría de las demás orientaciones. En consecuencia, hay relativamente pocos métodos, definiciones y algoritmos que se beneficien específica o totalmente de la información de tales escalas. Muy a menudo, las medidas ordinales del ranking de relaciones son tratadas como si fuesen intervalos. Existe la posibilidad de que sea menos arriesgado tratar las medidas ordinales del ranking de relaciones (comparadas con las medidas ordinales agrupadas) como si fuesen un intervalo, aunque tal suposición es discutible. Por supuesto, también es posible agrupar las puntuaciones del ranking de posiciones (por ejemplo produciendo una escala ordinal agrupada), binarizar los datos (por ejemplo las tres primeras elecciones serán tratadas como lazos, y las restantes como ausencia de lazos). En la combinación de información de múltiples tipos de lazos con frecuencia es necesario simplificar las escalas de ranking de posiciones. Pero si tenemos varias escalas de ranking que deseamos combinar para formar una escala (por ejemplo posiciones de los gustos de las personas hacia los demás, frecuencia de interacción , etcétera), la suma de tales escalas en un índice puede ser razonablemente tratado como una verdadera medida de intervalo.

Medidas de intervalo de relaciones. El nivel más “avanzado” de medida nos permite discriminar entre las relaciones informadas de forma tal que puede validarse su estado, por ejemplo, “este lazo es dos veces más fuerte que este otro”. Los lazos son

valorados en escalas en las que la diferencia entre un “1” y un “2” refleja la misma diferencia real que entre “23” y “24”.

Verdaderas medidas de intervalo de la fuerza de muchos tipos de relaciones son relativamente fáciles de construir con un poco de imaginación y perseverancia. Pedir a los informantes que informen de los detalles de la frecuencia o intensidad de lazos a través de métodos como cuestionarios o entrevistas puede ser, sin embargo, menos fiable –particularmente si las relaciones rastreadas son poco destacables e infrecuentes. Más que preguntar si dos personas se comunican, uno puede contar el número de entregas de correo electrónicos, llamadas telefónicas y correo interno entre ellos. Más que preguntar si dos naciones tienen intercambios con otra, debe observarse las estadísticas de la balanza de pagos. En muchos casos, es posible construir niveles de medida de intervalos de la fuerza de relaciones utilizando herramientas de observación (por ejemplo estadísticas recogidas para otros propósitos).

Las medidas continuas de la fuerza de las relaciones permiten la aplicación de un amplio rango de herramientas matemáticas y estadísticas para la exploración y análisis de datos. Muchos de los algoritmos que han sido desarrollados por los analistas de redes sociales, originalmente pensados para datos binarios, han sido ampliados para aprovechar la información disponible en las medidas de intervalos. Donde sea posible, las conexiones deben medirse en el nivel del intervalo –en tanto que siempre podemos movernos más tarde a un enfoque menos refinado; en cambio, si los datos se recogen en el nivel nominal, es mucho más difícil moverse a un nivel más refinado.

Aún cuando es una buena idea medir la intensidad de la relación en el nivel más refinado posible, la mayoría de los análisis de redes no trabajan en este nivel. La comprensión más importante de los análisis de redes, y de muchos de las herramientas matemáticas y gráficas utilizadas por los analistas de redes, fueron desarrolladas para grafos simples (por ejemplo, binarios, indirectos). Muchas de las descripciones de la inserción de los actores en sus redes, y de las redes en sí mismas, son pensadas en la literatura de redes comúnmente en términos discretos. Como resultado, a menudo se busca reducir incluso los datos de intervalo al nivel binario escogiendo un punto de corte y codificando la fortaleza del lazo por encima de él con un “1”, y con un “0” si está por debajo. Desafortunadamente, no hay una forma “correcta” de escoger un punto de corte. La teoría y los propósitos del análisis proporcionan la mejor guía para ello. A veces el examen de los datos puede ayudar

(quizás la distribución de la fuerza de las relaciones es realmente bimodal discreta, y muestra un punto de corte claro; quizás la distribución está fuertemente sesgada y la característica principal es la distinción entre la presencia y la ausencia de relación). Cuando se escoge un punto de corte, es sensato considerar también valores alternativos más altos o más bajos, y repetir el análisis con diferentes puntos de corte para ver si se afecta el sentido de los resultados. Esto puede ser tedioso, pero es muy necesario. De otro modo, uno puede equivocarse al pensar que se ha encontrado un patrón real, cuando sólo se ha observado la consecuencia del punto en que hemos decidido insertar el punto de corte.

UNA NOTA SOBRE LA ESTADÍSTICA Y LOS DATOS DE REDES SOCIALES

El análisis de redes sociales es más una rama de la sociología “matemática” que un “análisis estadístico o cuantitativo”, aún cuando los especialistas en redes de hecho trabajan con ambos enfoques. La distinción entre ambos enfoques no está muy clara. Los enfoques matemáticos del análisis de redes tienden a tomar los datos como “determinísticos”. Es decir, tienden a observar las relaciones medidas y la fortaleza de la relación, como si realmente reflejasen el estatus “real”, “final” o de “equilibrio” de la red. Los enfoques matemáticos también tienden a asumir que las observaciones no son una “muestra” de una población mayor de observaciones posibles, sino que las observaciones son a menudo vistas como la población de interés. Los estadísticos tienden a observar las puntuaciones particulares de la fuerza de las relaciones como resultado de una tendencia real destacable, o de una distribución de probabilidad de la fuerza de las relaciones. Los estadísticos también tienden a pensar en un conjunto particular de datos de red como una “muestra” de una clase o población mayor de tales redes o elementos de red –y tienen cuidado de que los resultados de su estudio pueda reproducir en el estudio “siguiente” resultados similares.

En los capítulos que siguen en este texto, nos vincularemos más con el lado “matemático” del análisis de redes que con el “estadístico” (de nuevo, es importante recordar que expongo las diferencias en esta discusión). Antes de seguir con ello, debemos mencionar un par de puntos importantes sobre la relación entre el material que se estudiará aquí y los enfoques estadísticos principales en sociología.

Por un lado, hay una pequeña diferencia aparente entre los enfoques estadísticos convencionales y los enfoques de redes. Las herramientas estadísticas descriptivas

univariadas, bivariadas o incluso multivariadas, se utilizan comúnmente en la descripción, exploración y modelado de datos de redes sociales. Los datos de redes sociales, como hemos dicho antes, se representan fácilmente como matrices de números –al igual que otros tipos de datos sociológicos. Como resultado, los mismos tipos de operaciones pueden hacerse con datos de redes y con los demás tipos de datos. Los algoritmos de estadística se utilizan con frecuencia para describir las características de las observaciones individuales (por ejemplo, la fortaleza de lazo mediana de un actor X con todos los demás en una red) y la red como un todo (por ejemplo, la media de toda la fuerza de los lazos para todos los actores en la red). Los algoritmos estadísticos se utilizan mucho en la evaluación del grado de similitud entre los actores y si se encuentran patrones en los datos de la red (por ejemplo, análisis factorial, análisis de *clusters*, escalado multidimensional). Incluso las herramientas del modelado predictivo son aplicadas con frecuencia a los datos de redes (por ejemplo, la correlación y la regresión).

Las herramientas estadísticas descriptivas son de hecho algoritmos para resumir las características de las distribuciones de frecuencias. Es decir, son operaciones matemáticas. Donde la estadística es realmente “estadística”, es en el lado inferencial. Es decir, cuando nuestra atención se gira hacia la valoración de la capacidad de reproducción o la probabilidad del patrón que describimos. La estadística inferencial puede ser, y de hecho es, aplicada al análisis de datos reticulares. Pero hay algunas diferencias muy importantes entre la esencia de la estadística inferencial utilizada en los datos de redes y aquella más comúnmente utilizada en los cursos básicos de análisis estadístico en sociología.

Probablemente el interés más común en la aplicación de estadística inferencial a los datos de ciencias sociales consiste en responder a preguntas acerca de la estabilidad, replicación o generalización de los resultados observados en una muestra. La pregunta central es: si se repite el estudio en una muestra diferente (extraída a través del mismo método), ¿qué probabilidad existe que obtenga la misma respuesta acerca de lo que ocurre en la población de la cual he extraído ambas muestras? Es una pregunta realmente importante -ya que ayuda a valorar la confianza (o la falta de ésta) así como la información que puede obtenerse poniendo a prueba nuestras teorías.

Para la mayoría de las observaciones utilizadas en los análisis de redes extraídas de alguna población de actores y/o lazos identificable por métodos de muestreo probabilístico, se aplica el mismo tipo de pregunta sobre la posibilidad de generalización de los resultados. A menudo este tipo de pregunta inferencial es de

menor interés para los investigadores de redes. En muchos casos, se encuentran estudiando una red en particular o un conjunto de éstas, y no tienen interés en generalizar una población de tales redes (bien porque no hay tal población o porque no interesa generalizarla de forma probabilística). En otros casos, podemos tener interés en generalizar, pero la muestra no ha sido extraída por métodos estadísticos. El análisis de redes, a menudo confía en herramientas como la observación directa, experimentos de laboratorio y documentos como archivos de datos – y frecuentemente no hay formas plausibles de identificar poblaciones y extraer muestras a través de métodos de probabilidad.

El otro uso principal de la estadística inferencial en las ciencias sociales es la contrastación de hipótesis. En muchos casos se utilizan herramientas similares o parecidas para la generalización y para la contrastación de hipótesis. La lógica básica de la contrastación de hipótesis es comparar un resultado observado en una muestra, con algún valor de la hipótesis nula, relativa a la variabilidad del resultado de la muestra bajo la suposición de que la hipótesis nula es válida. Si el ejemplo resulta muy distinto de lo que debería observarse bajo la suposición de que la hipótesis nula es válida, entonces la hipótesis nula probablemente no sea válida.

La clave en la cadena inferencial de contrastación de hipótesis, es la estimación de los errores estándar de las estadísticas. Es decir, la estimación del monto estimado en que el valor estadístico puede “variar” de una muestra a la otra, simplemente como resultado de los errores de muestreo. Con poca frecuencia pueden observarse o calcularse directamente tales errores estándar ya que no disponemos de réplicas. Sin embargo, la información de nuestra muestra sirve para estimar la variabilidad de la muestra.

Muchos procedimientos estadísticos comunes permiten estimar los errores estándar mediante aproximaciones muy bien validadas (por ejemplo el error estándar de una media se estima, generalmente, por la desviación estándar de la muestra dividida por la raíz cuadrada del tamaño de la muestra). Estas aproximaciones, sin embargo, se mantienen cuando se extraen observaciones de un muestreo aleatorio independiente. Las observaciones de la red son, en su mayoría, no independientes por definición. En consecuencia, las fórmulas inferenciales convencionales no se aplican a los datos de redes (aunque pueden aplicarse las fórmulas desarrolladas por otros tipos de muestreo dependiente). Es particularmente peligroso asumir que tales fórmulas se pueden aplicar, ya que la no independencia de las observaciones de la red

usualmente resultarán en subestimaciones de la variabilidad real del muestreo, y demasiada confianza en nuestros resultados.

El enfoque de la mayoría de los analistas de redes interesados en inferencia estadística para contrastar hipótesis sobre propiedades de la red, es trabajar directamente las distribuciones de probabilidad de los estadísticos. Se utiliza este enfoque porque: 1) no se han desarrollado aproximaciones para las distribuciones de muestreo de la mayoría de las estadísticas descriptivas utilizadas por los analistas de redes y 2) el interés a menudo se centra en la probabilidad de que un parámetro relativo de alguna teoría inicial (usualmente aleatoria) más que en la probabilidad de que una red dada sea típica en la población de todas las redes.

Supongamos, por ejemplo, que estamos interesados en la proporción de actores dentro de una red que son miembros de grupos o cualquier otro parámetro o estadística de la red. La noción de grupo implica una estructura – conexiones no aleatorias entre los actores. Se tienen los datos sobre una red de diez nodos, en los cuales hay 20 lazos simétricos entre los actores y se observa que hay un grupo que contiene cuatro actores. La pregunta deductiva puede ser: si los lazos entre los actores fueran eventos aleatorios, ¿qué tan probable será que esa red de diez nodos y 20 lazos simétricos muestre grupos de tamaño cuatro o más? Si se concluye que los grupos de tamaño igual o mayor que cuatro en redes de este tamaño y grado son comunes, se debe ser cauteloso en concluir que se ha descubierto una “estructura” de no-aleatoriedad. Si se concluye que tales grupos (aquellos más numerosos o más inclusivos) son muy improbables bajo la suposición de que los lazos son sólo aleatorios, entonces es muy acertado llegar a la conclusión de que existe una estructura social.

Pero, ¿cómo se determina esta probabilidad? Uno de los métodos utilizados es la simulación –y como la mayoría de las simulaciones, se requiere una gran cantidad de recursos y habilidades de programación. En el caso anterior, se puede utilizar una tabla de números aleatorios para distribuir 20 lazos entre 10 actores, y entonces buscar la red resultante para grupos de tamaño cuatro o más. Si no se encuentran grupos, se asigna cero al ensayo; si se consigue un grupo, se asigna un uno. El resto es simple. Sólo repetir el experimento varios cientos de veces y añadir qué proporción de “ensayos” devienen “éxitos”. La probabilidad de un éxito entre estos experimentos de simulación es un buen estimador de la probabilidad de que pueda encontrarse una red de este tamaño y densidad, teniendo grupos de este tamaño “sólo por accidente”

cuando los mecanismos causales no-aleatorios que se cree causa el grupo no operan de hecho.

Esto puede parecer extraño, y es ciertamente un gran trabajo (la mayoría del cual, afortunadamente, puede ser hecho por ordenador). Pero realmente, no es muy diferente de la lógica de la prueba de hipótesis con datos que no sean reticulares. Los datos de redes sociales tienden a diferir de los datos de encuestas más “convencionales”, en algunas formas claves: los datos de redes a menudo no son muestras probabilísticas, y las observaciones de los nodos individuales no son independientes. Estas diferencias son importantes para las preguntas sobre la generalización de los hallazgos y sobre los mecanismos de contrastación de hipótesis. Sin embargo, no hay nada sustancialmente distinto sobre la lógica del uso de la estadística descriptiva e inferencial para datos de redes sociales.

La aplicación de la estadística a los datos de redes sociales es un área interesante y, en el momento de redactar este trabajo, es un campo en desarrollo. En tanto en cuanto este texto se centra en los usos básicos y comunes del análisis de redes, no hemos de decir mucho más acerca de la estadística. Mucho de lo que sigue en el libro puede relacionarse con el lado “descriptivo” de la estadística (desarrollando índices para describir ciertos aspectos de la distribución de lazos relacionales entre actores en redes). Para aquellos interesados en el lado inferencial, un buen punto de partida es la segunda mitad del excelente libro de Wasserman y Faust.