

Ricardo Mansilla Corona* y Rosa María Mendoza Rosas*

De las redes complejas a las epidemias

Resumen | La Teoría de Redes Complejas es uno de los cuerpos teóricos que ha servido de sustento a las nuevas propuestas de modelación epidémica y el cual tomamos como base para desarrollar nuestro trabajo. Las ciudades de un país conectadas por medio de carreteras, líneas aéreas o rutas de cabotaje constituyen una red a través de cuyos vértices se transmiten las enfermedades, por eso es muy importante comprender primero la topología de estas redes, para posteriormente entender el proceso difusivo de estas afecciones. El artículo describe los resultados obtenidos del estudio de la red de ciudades mexicanas con más de 5000 habitantes. Se presentan las conexiones que se dan entre las diversas localidades del país. Por último se analizan los resultados y su pertinencia para la elaboración de políticas públicas.

From Complex Networks to Epidemics

Abstract | Complex Networks Theory is one of the theoretical bodies which has served as a platform for new propositions for epidemic modelling, and which we have adopted as a basis for the development of our work. The towns of a country, connected by highways, airlines, or short haul routes make up a network the nodes of which serve as channels for the transmission of diseases. Therefore, it is important, as a first measure, to understand the topology of this network, before proceeding to the study of the diffusive processes of the disease in question. This article describes the results obtained by the study of Mexican towns with more than 5,000 inhabitants. We present the links between the different settlements in this country. Finally, we analyze the results and their relevance for the formulation of public policy.

Palabras clave | teoría de redes complejas – sistema complejo – propiedad emergente – mundo pequeño – epidemias

Keywords | theory of complex networks – complex system – emergent property – small world – epidemics

Introducción

A LO LARGO DE TODA la historia de nuestra civilización han quedado documentadas muchas evidencias de la proliferación de enfermedades contagiosas a través de los asentamientos humanos. Constancia de estos fenómenos pueden

* Centro de Investigaciones Interdisciplinarias en Ciencias y Humanidades-Universidad Nacional Autónoma de México. **Correos electrónicos:** mansy@unam.mx y romanero@unam.mx

encontrarse en la Biblia, el Talmud o el Corán.¹ Sin duda el más conocido de estos sucesos fue el que obligó a Mineptah, el faraón egipcio que sucedió a Ramses II, a permitir la salida de Egipto de los judíos alrededor del 1224 a.C. Las epidemias se encuentran sin duda alguna entre los más fieles compañeros de viaje de la humanidad.

Una vez rebasada la larga noche de la Edad Media con sus interpretaciones expiatorias de la existencia de las epidemias, el Renacimiento volvió a posicionar al ser humano en el centro de la discusión epistemológica. De esta forma, el avance en la comprensión de las enfermedades contagiosas y sus manifestaciones colectivas se sustentó en la observación experimental del comportamiento de estas epidemias. La obra de G. Fracastoro en 1546 *De contagione et contagiosis morbis et eorum curatione* inicia una época en los estudios de este tipo de procesos. El trabajo de D. Bernoulli de 1760 marca un hito al integrar las técnicas matemáticas al análisis de la influencia de la vacunación sobre la difusión de la viruela en Francia. El éxito que había obtenido el cálculo de I. Newton en la formulación de la Mecánica Clásica, convirtió a las ecuaciones diferenciales en las herramientas idóneas para la descripción de las epidemias como los procesos evolutivos que realmente son. Esta etapa culmina con los trabajos de W. O. Kermack y A. G. McKendrick (Kermack, McKendrick 1927), y la línea de pensamiento por ellos iniciada. La aparición de las computadoras digitales en la segunda mitad del siglo XX hizo posible la simulación numérica de aquellos modelos en ecuaciones diferenciales para los cuales sólo se tenía un conocimiento limitado de sus soluciones. Fue posible además desarrollar modelos de propagación epidémica basados en autómatas celulares (Boccaro y Cheong 1992) y más tarde por medio de los modelos de agentes múltiples

*La Teoría de Redes
Complejas es uno de los
cuerpos teóricos que ha
servido de sustento a las
nuevas propuestas de
modelación epidémica.
Las ciudades de un país,
conectadas por medio de
carreteras, líneas aéreas o
rutas de cabotaje constituyen
una red a través de cuyos
vértices se transmiten las
enfermedades. Entender la
topología de estas redes es
el primer paso para explicar
el proceso difusivo*

ras digitales en la segunda mitad del siglo XX hizo posible la simulación numérica de aquellos modelos en ecuaciones diferenciales para los cuales sólo se tenía un conocimiento limitado de sus soluciones. Fue posible además desarrollar modelos de propagación epidémica basados en autómatas celulares (Boccaro y Cheong 1992) y más tarde por medio de los modelos de agentes múltiples

¹ El nombre tenía una connotación mística. La palabra epidemia proviene del griego *epi* que quiere decir "encima" y *demos* que quiere decir "pueblo".

(Mansilla y Gutiérrez 2001). Las capacidades de las computadoras digitales, que al decir de Pagels (1989) se han convertido en los instrumentos básicos de la Teoría de la Complejidad, influyeron de forma decisiva en nuestra comprensión de la propagación de enfermedades contagiosas.

El paradigma basado en la modelación por medio de ecuaciones diferenciales comenzó a agotarse a principios del siglo XXI debido a la revolución en los medios de transporte y comunicaciones. Las hipótesis sobre las cuales están basados esos modelos dejaron de ser realistas. Para poner en perspectiva este hecho baste decir que más del 65% del contagio de la gripe aviar (H5N1) iniciada en China en 2004 ocurrió a través de los aeropuertos (Mansilla 2015). Se imponía un nuevo arquetipo para la modelación de los procesos de contagio.

La Teoría de Redes Complejas es uno de los cuerpos teóricos que ha servido de sustento a las nuevas propuestas de modelación epidémica. Las ciudades de un país, conectadas por medio de carreteras, líneas aéreas o rutas de cabotaje constituyen una red a través de cuyos vértices se transmiten las enfermedades. Entender la topología de estas redes es el primer paso para explicar el proceso difusivo.

En esta contribución pretendemos mostrar los resultados que hemos obtenido del estudio de la red de ciudades mexicanas con más de 5 mil habitantes. Comenzaremos ofreciendo en el epígrafe siguiente un panorama de los aspectos más importantes de la Teoría de Redes Complejas para, a partir de allí, describir los resultados obtenidos en la red de ciudades mexicanas. Culminamos este trabajo con una sección de conclusiones donde analizamos los resultados obtenidos y su pertinencia para la elaboración de políticas públicas.

Origen de los estudios sobre redes complejas

Contrariamente a lo que suponen muchas personas, los estudios sobre redes tiene una larga historia. Sin duda el primer estudio documentado sobre redes está basado en el Problema de los Siete Puentes de Königsberg, el cual fue resuelto de manera brillante por Leonard Euler y su solución fue publicada en 1741. La ciudad de Königsberg, se encuentra actualmente en territorio ruso y se conoce por el nombre de Kaliningrado. Está situada en ambos márgenes del río Pregel e incluye dos grandes islas, que se encuentran conectadas al resto de la ciudad a través de siete puentes, como se muestra en la figura 1.

El problema consistía en encontrar un recorrido en la ciudad de manera tal que cada puente fuera cruzado solo una vez. Euler observó con gran ingenio, que el problema se resolvía si se encontraba una sucesión de cruces de puentes. Esta última idea le permitió reescribir el problema utilizando el concepto de red, como puede verse en la figura 2. Así, las diferentes masas de tierra se colapsarían en los nodos de la red y los puentes serían los vértices de la misma,



Figura 1. Los siete puentes de Königsberg.



Figura 2. Los siete puentes de Königsberg traducidos en un grafo.

como se muestra en la figura 2. Por último, Euler hizo notar que cada vez que una masa de tierra (nodo) es visitada, debe abandonarse por otro puente (arista). De esta forma, la cantidad de puentes que tocan una masa de tierra debe ser un número par siempre. Puesto que a las cuatro masas de tierra arriban tres puentes, no existe solución para este problema. Este trabajo de Euler es el origen de la moderna Teoría de Grafos y de la Topología.

Es también notable que los primeros trabajos dedicados a la aplicación de redes a los fenómenos sociales se desarrollaron a finales del siglo XIX e incluyen las contribuciones de Durkheim (1894 [2007]) y Tönnies (1912). Este último argumentó que los miembros de los grupos sociales pueden establecer lazos personales y directos que los vinculan con quienes comparten valores y creencias, y que también pueden existir vínculos sociales formales e instrumentales. Durkheim aportó una explicación no individualista de los fenómenos sociales al argumentar que los mismos surgen cuando los individuos que interactúan, “constituyen una realidad que ya no puede explicarse en términos de los atributos de los actores individuales”.² Cabe llamar la atención de que Durkheim manifestó aquí claramente la existencia de una propiedad emergente en el sentido de la moderna teoría de los sistemas complejos. Tanto Durkheim como Tönnies hicieron un uso intuitivo de este concepto.

A principio de la década de 1930 se iniciaron dos líneas de trabajo diferentes en esta área. Jacob L. Moreno fue pionero en el registro sistemático de datos y en el análisis de la interacción social de pequeños grupos, en especial las aulas y grupos de trabajo (Moreno 1951). Es el creador del concepto de la sociometría. Mientras, un grupo de Harvard liderado por W. Lloyd Warner y Elton Mayo exploró las relaciones interpersonales en las actividades laborales. Este tipo de técnicas adquirieron un notable reconocimiento en la comunidad de científicos sociales de la época. Tal es así que en 1940, en un discurso dirigido a los antropólogos británicos, A.R. Radcliffe-Brown instó al estudio sistemático

Fue Georg Simmel en la segunda mitad del siglo XIX, el primer estudioso que pensó directamente en términos de red social. Sus ensayos apuntan a la naturaleza del tamaño de la red y a la probabilidad de interacción en redes ramificadas

² La afirmación anotada de Durkheim se refiere, sin hacerlo explícito a una *propiedad emergente* de los grupos humanos. Esta es una de las propiedades más comunes de los sistemas complejos.

de las redes (Radcliffe-Brown 1940). Fue Georg Simmel (1908) en la segunda mitad del siglo XX, el primer estudioso que pensó directamente en términos de red social. Sus ensayos apuntan a la naturaleza del tamaño de la red y a la probabilidad de interacción en redes ramificadas.

Otro aspecto que recibió tempranamente atención de los investigadores fue la propagación de impulsos en una red. En un trabajo publicado en 1951, R. Solomonoff y A. Rapoport desarrollaron un modelo teórico para la difusión de impulsos sobre una red en la cual sus nodos tenían una cantidad fija de conexiones (Solomonoff y Rapoport 1951). La inspiración de este trabajo provenía de las características de la difusión de impulsos en redes de neuronas, la difusión de epidemias en una población y la interacción de los componentes de una red genética. El modelo no era muy realista. Posteriormente A. Rapoport publicó un par de trabajos donde suavizaba un poco las hipótesis de su anterior contribución (Rapoport 1953; 1953a) permitiendo que los miembros de la población tuvieran diferente número de conexiones y que hubiera traslape dentro de los conocidos de diferentes miembros de la población.

El aspecto de los estudios sobre redes sociales que ha tenido mayor impacto en el público en general es sin duda la propiedad de “mundo pequeño” (*small world*). La idea original fue inicialmente propuesta en 1930 por el escritor húngaro Frigyes Karinthy en un cuento llamado *Chains*, en el cual se sugería que el número de conocidos crecía exponencialmente con el número de enlaces en la cadena, y en consecuencia sólo un pequeño número de enlaces eran necesarios para que el conjunto de conocidos se convirtiera en la población humana entera. El primer trabajo académico sobre el tema lo iniciaron Manfred Kochen e Ithiel de Sola Pool a principio de los años 1960, aunque los resultados de su investigación aparecieron un poco más tarde (Pool 1978). En 1960 apareció el trabajo seminal de P. Erdős y A. Rényi sobre redes aleatorias (Erdős y Rényi 1960).

Más adelante el psicólogo Stanley Milgram, ya famoso por su contribución al estudio de la relación entre la sumisión de los valores éticos y la autoridad (Milgram 1963), publicó un trabajo (Milgram 1967) basado en una investigación empírica sobre la propiedad de mundo pequeño. El mismo consistía en proponer a un grupo de personas, elegidas en un par de poblados de Kansas y Nebraska, que enviaran por sus propios medios una carta a una persona que vivía en Boston. El envío debía hacerse usando conocidos de las personas involucradas sin hacer uso del sistema de correos convencional.

Milgram encontró que para hacer llegar la carta a su destinatario en Boston se requirió, en promedio, de seis intermediarios. Este resultado fue objeto de debate durante largo tiempo. Recientemente un estudio mucho más abarcador (Leskovec y Horvitz 2007) obtuvo como resultado un promedio de 6.6 intermediarios.

La irrupción de las modernas computadoras digitales cambió drásticamente

la manera de investigar este tipo de fenómenos sociales. Como lo vaticinó H. Pagels (1989), las computadoras son las herramientas fundamentales para la investigación de los sistemas complejos. De la misma forma que el microscopio inventado por Zacharias Janssen cambió la investigación en biología y el telescopio de Galileo transformó drásticamente nuestra concepción del mundo, las computadoras han abierto una ventana de nuestra realidad desde donde hemos podido acceder a áreas de investigación impensables antes de ellas. De la franca escasez de datos empíricos se ha pasado al *tsunami* incontrolable de los mismos. En este proceso, el protagonista más importante ha sido sin duda internet.

Su historia comienza a principio de los años 1960, en plena Guerra Fría cuando la ARPA, una agencia militar de EEUU, comenzó a financiar proyectos de investigación en universidades para la creación de una red de computadoras. El primer intento exitoso de construcción de una tal red fue hecho por L. Roberts del MIT y fue bautizada con el nombre de ARPANET por razones obvias. No obstante, el protocolo de transmisión de datos de la misma no poseía las propiedades idóneas para el funcionamiento sobre una arquitectura abierta. Fueron R. Kahn y V. Cerf de la Universidad de Stanford quienes finalmente crearon el TCP/IP (*Transmission Control Protocol/Internet Protocol*) que es el protocolo de transmisión de datos en la internet actual. Este último avance se desarrolló fuera de la égida de los militares.

Como ocurre siempre con los proyectos científicos que no tienen apoyo financiero de la industria o de los militares, a finales de la década de los ochenta, nadie sabía aún cómo obtener financiamiento para seguir desarrollando una tecnología que nadie sabía tampoco cómo utilizar comercialmente. El tema del acceso a los recursos se convirtió en un aspecto vital en el desarrollo de las redes de computadoras en la comunidad académica y científica. La cantidad de información crecía rápidamente, no así los recursos para almacenarla.

Las computadoras son las herramientas fundamentales para la investigación de los sistemas complejos. De la misma forma que el microscopio inventado por Zacharias Janssen cambió la investigación en biología y el telescopio de Galileo transformó drásticamente nuestra concepción del mundo, las computadoras han abierto una ventana de nuestra realidad desde donde hemos podido acceder a áreas de investigación impensables antes de ellas

Un descubrimiento cambió el curso de los acontecimientos. De la misma forma en que J. Gutenberg transformó la manera de producir libros, el lenguaje HTML (*HyperText Markup Language*) revolucionó la presentación de la información en la web. A partir de su creación el concepto de *página web* incorporaría textos, datos, imágenes y sonido de una manera integrada. Así, en el año 1990 algunas empresas propusieron proyectos para participar en internet. Este proceso se aceleró hacia 1993. El modelo de negocios consistía en principio en promocionar artículos o servicios comerciales en ciertas áreas de las páginas web. El precio de esta promoción era proporcional al tamaño del espacio seleccionado y al número de visitas que recibiera la página. Comenzó de esta forma una lucha por la popularidad virtual.

Los iniciadores de estos emprendimientos nunca imaginaron que los sitios más exitosos no serían aquellos que brindaban noticias, modas, información financiera o meteorológica, o bien recetas de cocina. Los vencedores de esta batalla por la popularidad serían aquellos sitios web que apostaron a proveer algo bien intrínseco de la condición humana: la comunicación interpersonal.

El origen de las redes sociales en la web se remonta, por lo menos al año 1995, cuando Randy Conrads crea el sitio web www.classmates.com. Con esta red social se pretendía que los usuarios pudieran recuperar o mantener el contacto con antiguos compañeros de colegio, preparatoria, universidad, etcétera.

En 2002 comienzan a aparecer sitios web promocionando los círculos de amigos en línea. Este último término se empleaba para describir las relaciones en las comunidades virtuales, y se hizo popular en 2003 con la llegada de sitios tales como MySpace o Xing. El tráfico en estos sitios se hizo muy atractivo a los comerciantes. Debido a esto, grandes compañías entraron en el terreno de las redes sociales en internet. Sin embargo, como se sabe, los sitios más exitosos tuvieron en general comienzos muy humildes.

En el año 2004, M. Zuckerberg fundó *Facebook*, un sitio de intercambio de información personal, sin duda el más exitoso de la historia. Más adelante en 2005, Chad Hurley, Steve Chen y Jawed Karim crearon www.youtube.com, un sitio para almacenar y compartir videos. Por último en 2006 E. Williams, N. Glass y J. Dorsey crearon *Twitter*,³ un sitio que permite el envío de mensajes a la comunidad de seguidores del usuario. Queremos hacer énfasis en un aspecto distintivo de estos portales: en todos ellos se intercambia información con el conjunto de amigos o seguidores de la persona que la genera. El concepto de red es pues intrínseco a los mismos.

Estas redes virtuales son los otros protagonistas esenciales del proceso de

3 Al lector interesado en el tormentoso proceso de fundación y desarrollo de este sitio, se le recomienda Bilton 2013.

creación de conocimiento científico sobre redes sociales, entendidas estas últimas como el conjunto de individuos pertenecientes a una población conectados entre sí por algún criterio específico, como puede ser amistad, interés sobre un tema en particular, nacionalidad, afición a algún deporte, etc. Vale la pena señalar aquí que el sustrato material de una tal red no tiene porqué ser internet. Todos participamos en nuestra vida cotidiana de varias redes. En el mejor de los casos, internet le ha brindado a las redes sociales el sustrato físico que les permite un desempeño dinámico y la capacidad de compartir grandes volúmenes de información a unas velocidades nunca antes vistas. Se abrió con ello una suerte de Caja de Pandora de la era digital. Un ejemplo paradigmático de esto es el llamado Juego de Kevin Bacon.⁴ Este actor norteamericano se caracteriza por haber participado en una gran cantidad de películas, siendo siempre actor de reparto. Recientemente Brett Tjaden, un científico de la computación en la Universidad de Virginia creó un sitio donde se calculó la red de todos los artistas que han actuado con Kevin Bacon, así como las conexiones entre los mismos. Curiosamente esta red de actores poseía también propiedades de mundo pequeño como las encontradas por Milgram.

Los estudios de redes se han extendido además a redes metabólicas (Watts y Strogatz 1998), coautoría de trabajos científicos (Newman 2001; 2001a; 2001b), cadenas tróficas en ambientes naturales (Montoya y Solé 2000) y a redes de ciudades (Mendoza y Mansilla 2010). A este último tipo de redes está dedicado este trabajo.

Características cuantitativas fundamentales de las redes complejas

La información más básica que debe tenerse de una red, es la colección de vínculos entre los diferentes miembros de la misma. Una manera muy efectiva de lograrlo es por medio de la llamada *matriz de adyacencia*. Sean $\{M_1, \dots, M_n\}$ el conjunto de los miembros de una red. Denominamos matriz de adyacencia a un arreglo numérico del tipo:

$$A = (a_{ij}) \quad ; \quad a_{ij} = \begin{cases} 1 & \text{si } M_j \text{ está conectado con } M_i \\ 0 & \text{en otro caso} \end{cases}$$

Si tomamos una fila cualquiera de esa matriz (la i -sima por ejemplo) los valores iguales a 1 que encontremos en la misma nos informan de los miembros (también llamados nodos) de la red que están conectados con M_i . Consideraremos

⁴ Ver: <https://oracleofbacon.org/index.php>

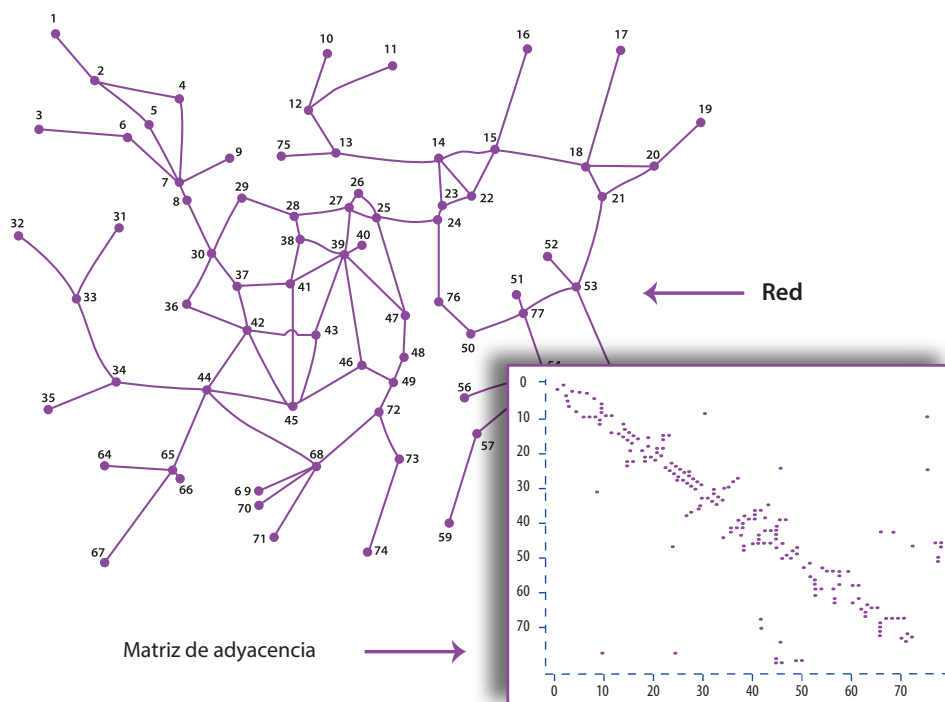


Figura 3. Gráfico de una red y su matriz de adyacencia.

redes simétricas a lo largo de este trabajo, es decir si M_j está conectado con M_i también ocurre en sentido contrario.

Un ejemplo de una red y su matriz de adyacencia aparece en la figura 3. Tomemos por ejemplo el nodo 39, el cual tiene siete conexiones.

Si observamos la fila 39 de la matriz de adyacencia, veremos que existen siete puntos que representan los siete miembros del conjunto de nodos con los cuales M_{39} está conectado.

Otra característica de las redes que ha sido muy estudiada es el comportamiento del número de conexiones que los diferentes nodos tienen. El enfoque al estudio de este problema ha sido probabilístico desde el seminal trabajo de Erdős y Rényi (1960). Estos matemáticos se dedicaron a estudiar las llamadas *redes aleatorias*. Precisemos qué entendemos por este concepto.

La construcción de una red aleatoria se realiza a partir del siguiente procedimiento:

- a) Se toman dos nodos al azar que no se encuentren previamente conectados.
- b) Se genera un número aleatorio $0 < r < 1$. Si $r < p_c$, donde p_c es una probabi-

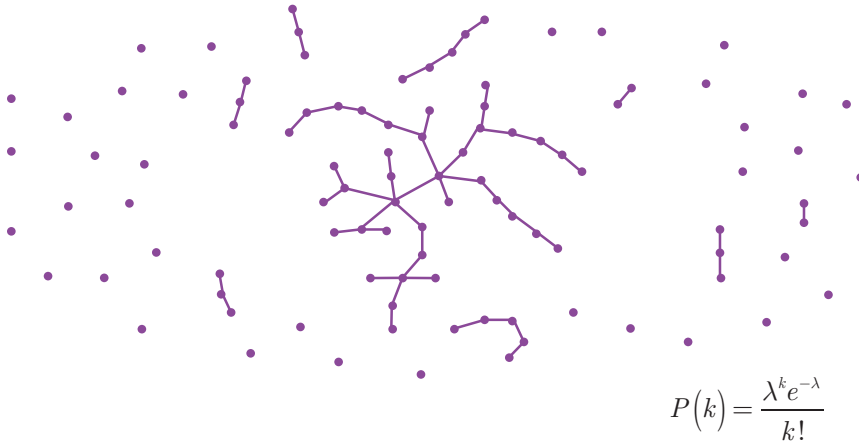


Figura 4. Instancia de la generación de una red aleatoria y su función de densidad de probabilidades de conexiones.

lidad umbral, entonces estos dos nodos se conectan entre sí. En caso contrario se quedan como antes.

c) Se repiten los puntos a) y b).

En la figura 4 se muestra una instancia de la construcción de una red aleatoria. Se puede observar además la función de densidad de probabilidades de las conexiones de los nodos, que tiene la forma de una distribución de Poisson. El parámetro es una función de la probabilidad p_c de conexión.

Precisemos el concepto de función de densidad de probabilidades de las conexiones de los nodos. El sentido de $P(k)$ es el que sigue: es la probabilidad de tomar un nodo al azar en la red y que éste tenga exactamente k conexiones con otros nodos.

En el caso de que la red se construya de manera aleatoria, según el esquema de Erdős y Rényi, entonces la probabilidad $P(k)$ debería ser binomial:

$$P(k) = \frac{n!}{k!(n-k)!} p_c^k (1-p_c)^{n-k}$$

donde n es el número de nodos (conectados o no) de la red original. Como es conocido cuando $n \rightarrow \infty$ se tiene:

$$P(k) \approx \frac{(np)^k e^{-np}}{k!}$$

que es la aproximación a la binomial y que coincide con la distribución de Poisson obtenida en Erdős y Rényi (1960).

Sin embargo una gran sorpresa fue el descubrimiento de que, para una abrumadora cantidad de redes reales, la función de densidad de probabilidades de las conexiones no seguía el modelo propuesto por Erdős y Rényi. Los investigadores encontraron otro tipo de funciones de distribución, bien conocidas por los físicos, las llamadas leyes de potencia:

$$P(k) \sim \frac{1}{k^\alpha}$$

Las leyes de potencia son ubicuas en la naturaleza y la sociedad. G. K. Zipf las encontró en las diferentes frecuencias de utilización de las palabras, así como en la distribución del tamaño de las ciudades en un país (Zipf 1966). W. Pareto las encontró en la forma en que se distribuían los ingresos o las riquezas de los integrantes de una sociedad (Pareto 1896).

¿Cuál es el mecanismo que genera una distribución del tipo ley de potencia? El primer intento de explicación de este fenómeno es debido a A. Barabasi y R. Albert (1999). El mismo propone el concepto de “conexión preferencial” (*preferential attachment*) que consiste en que la probabilidad de conexión a un nodo depende del propio grado del nodo. Por ejemplo, es más probable que una página web incluya hipervínculos a otras páginas más populares, con ello tales páginas altamente conectadas son más fáciles de encontrar y por lo tanto se conocen bien. Un nuevo documento es más probable que sea poco conocido y por lo tanto menos citado. Implícitamente, el modelo de Barabasi y Albert supone que las redes crecen a partir de un pequeño número de nodos con el mecanismo antes explicado. Es por tanto un modelo dinámico de la evolución de las redes libres de escala. Debemos señalar aquí que el modelo de Barabasi y Albert es un mecanismo de explicación de una propiedad conocida de las redes reales. Existen sin embargo varios enfoques teóricos para la explicación de este comportamiento⁵ (Dorogovtsev *et al.* 2000; Krapivsky *et al.* 2000).

Similar situación ocurre con las redes que poseen la propiedad de mundo pequeño, que fue estudiada inicialmente por Kochen, Sola Pool y Milgram. La primera propuesta de un mecanismo que generara tal propiedad en una red fue ofrecida en Watts, Strogatz (1998). En este trabajo los autores, partiendo de una red simétrica y homogénea en el número de conexiones, introducen nuevos vínculos de manera aleatoria, eliminando de la misma manera otros y generan así propiedades de mundo pequeño en la red original.

5 Una simulación del proceso de conexión preferencial puede verse en: <http://markov.uc3m.es/2012/11/temporal-networks-with-igraph-and-r-with-20-lines-of-code/>

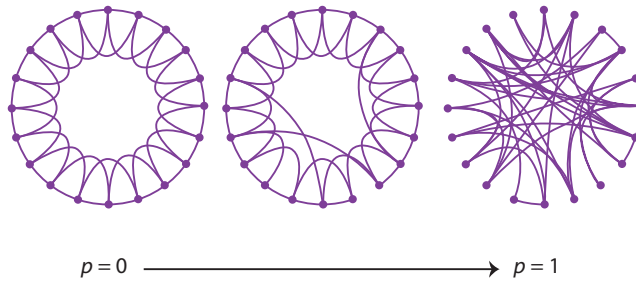


Figura 5. Proceso de generación de una red de mundo pequeño. En la medida en que el valor de la probabilidad p aumenta, más frecuentes son los vínculos aleatorios agregados. Si $p \rightarrow 1$ la red se convierte en aleatoria.

El proceso de generación de una red con la propiedad de mundo pequeño aparece explicado en la figura 5. Partiendo de una red ordenada donde cada nodo tiene solo vínculos con sus dos vecinos a la derecha y a la izquierda, se empiezan a crear vínculos aleatorios con una probabilidad p . Si el valor de esta probabilidad es elegido adecuadamente se genera una red con la característica de mundo pequeño. Si p es demasiado grande la red resultante es aleatoria.

Otro aspecto importante en el estudio de una red es su “diámetro”, entendido este último como la distancia máxima entre un par de nodos de la misma. Aquí implícitamente estamos asumiendo que a los vértices de nuestra red se les asignan números que pueden ser interpretados como longitudes. Por tanto el diámetro de una red es la longitud máxima de un camino que une a dos nodos de la misma. En caso de que una red sea aleatoria se conoce el siguiente resultado acerca de su diámetro (Chung y Lu 2001):

$$d = \frac{\ln(N)}{\ln(\langle k \rangle)}$$

Aquí N simboliza el número total de nodos de la red y $\langle k \rangle$ el promedio de la variable aleatoria que representa al número de conexiones de los nodos. En el antes mencionado trabajo se obtienen las siguientes propiedades del comportamiento de las redes en términos de p y $\langle k \rangle$:

- Si $\langle k \rangle = pN < 1$; las redes están compuestas por árboles aislados.
- Si $\langle k \rangle = pN = 1$; una componente gigante aparece en la red.
- Si $\langle k \rangle \geq \ln(N)$; prácticamente cualquier red es totalmente conexa.

Como vemos de los resultados anteriores, la aglomeración de los nodos en una red depende de la probabilidad p de conexión entre los mismos si esta red es aleatoria. En general las redes complejas exhiben un alto grado de aglomeración entre sus nodos. Es conveniente tener una medida cuantitativa de esta propiedad. Veamos el caso inicialmente de una red aleatoria. Si consideramos un nodo en la misma junto a sus vecinos, la probabilidad de que dos de estos últimos estén conectados es igual (dado que la red es aleatoria) a la probabilidad p de que dos nodos elegidos al azar se conecten. Por tanto una definición natural para un coeficiente de aglomeración sería:

$$C = \frac{\langle k \rangle}{N}$$

nótese que en el caso particular de que la red sea aleatoria tendríamos $C = p$.

Como ya hemos visto, una red puede ser descrita por su matriz de adyacencia. Al conjunto de los valores propios de esta matriz se le conoce como su *espectro*. Por asociación, con frecuencia se le llama espectro de la red al conjunto de los valores propios de su matriz de adyacencia. Si asumimos que los vértices de la red pueden ser recorridos en ambas direcciones, esto es, que la matriz de adyacencia es simétrica, entonces su espectro tiene que estar constituido por números reales solamente. En tal caso podemos definir la función de densidad espectral:

$$\rho(\lambda) = \frac{1}{N} \sum_{i=1}^N \delta(\lambda - \lambda_i)$$

Aquí N representa la cantidad de nodos de la red, λ_i ; $i = 1, \dots, N$ son los valores propios de la matriz de adyacencia y $\delta(\lambda)$ es la función delta de Dirac.

La función de densidad espectral representa la frecuencia de aparición de cada valor propio dentro del conjunto total. El interés en la función de densidad espectral se debe a que muchas de las propiedades topológicas de las redes pueden ser deducidas de esta función (Metha 1991; Crisanti 1993). En particular los momentos de esta distribución (Albert y Barabasi 2002):

$$\sum_{i=1}^n \rho(\lambda_i) \lambda_i^k = \frac{1}{N} \sum_{i=1}^N \lambda_i^k = \sum_{i_1, \dots, i_k} a_{i_1 i_2} a_{i_2 i_3} \dots a_{i_k i_1}$$

representan la cantidad de caminos sobre la red de longitud k que regresan al mismo punto. Existen resultados de la teoría de matrices aleatorias (Wigner 1955; 1957; 1958) que garantizan bajo ciertas condiciones una forma especial para la función de densidad espectral. Por ejemplo, supongamos que una red crece con una probabilidad de conexión dependiente del número de nodos que posee según la siguiente relación $p \sim cN^{-\alpha}$. Entonces la densidad espectral tiene la forma:

$$\rho(\lambda) = \begin{cases} \frac{\sqrt{4Np(1-p) - \lambda^2}}{2\pi Np(1-p)} & |\lambda| < 2\sqrt{Np(1-p)} \\ 0 & \text{en otro caso} \end{cases}$$

lo cual se conoce como la ley de Wigner.

A continuación estudiaremos las propiedades fundamentales de la red de ciudades de México a la luz de los resultados teóricos aquí expuestos.

Propiedades de la red de ciudades de México

Matriz de adyacencia

Como vimos en la sección anterior, una de las herramientas más útiles para entender las propiedades de una red es su matriz de adyacencia. Como ya hemos comentado, en la base de datos del Instituto Nacional de Estadística y Geografía (INEGI) cada ciudad tiene un número de código. Este número es asignado en base a dos criterios. El primero es el orden alfabético del estado de la federación a la cual pertenece la ciudad, y el segundo criterio es la proximidad geográfica entre las ciudades. Este código caracteriza unívocamente a las ciudades. La construcción de la matriz de adyacencia se hizo a partir de dichos códigos. Este procedimiento tiene la ventaja de que ciudades geográficamente cercanas tienen en general filas y columnas cercanas en la matriz.

En la figura 6 se muestra una representación de la matriz de adyacencia de las ciudades consideradas. Se han insertado en la figura algunas etiquetas para resaltar algunos estados de la federación. La línea horizontal que posee la etiqueta con el número 9 se refiere al Distrito Federal, el cual tiene 65 conexiones directas con otras ciudades y municipios del país.

Puesto que la matriz de adyacencia es simétrica, existe también una línea vertical correspondiente a las conexiones del Distrito Federal. Como veremos más adelante, esta línea etiquetada con el 9 está relacionada con el vector propio dominante de la matriz de adyacencia.

Propiedades espectrales de la matriz de adyacencia

La matriz de adyacencia tiene $N = 5823$ valores propios reales. Como hemos discutido antes, la densidad espectral se define como:

$$\rho(\lambda) = \frac{1}{N} \sum_{i=1}^N \delta(\lambda - \lambda_i)$$

Obviamente, cuando $N \rightarrow +\infty$ la densidad espectral $\rho(\lambda)$ tiende a una función continua. El estudio de las propiedades del espectro de las matrices de adyacencia

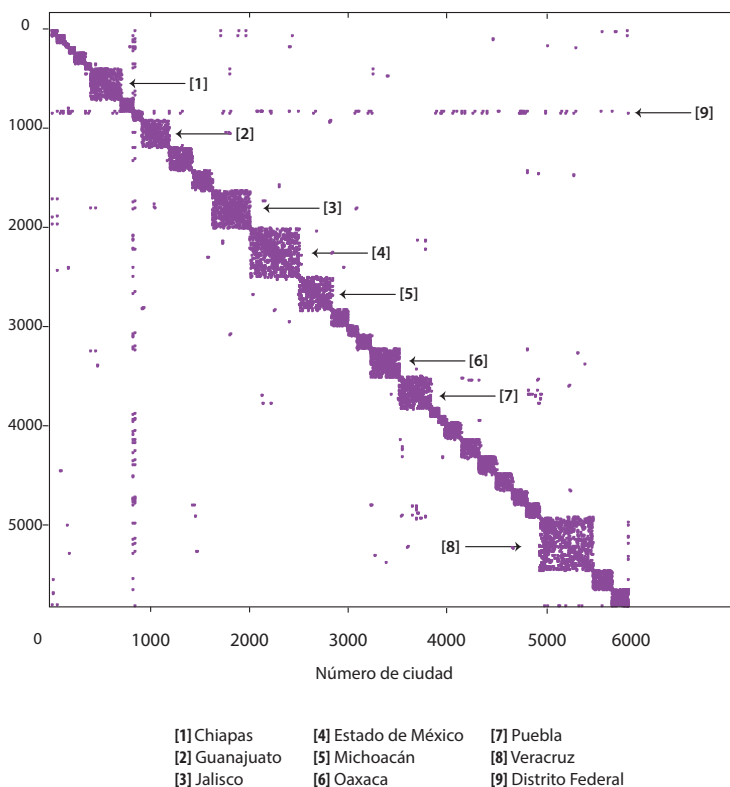


Figura 6. Representación de la matriz de adyacencia de la red de ciudades mexicanas. Los puntos morados representan caminos entre las diferentes ciudades. Se insertaron etiquetas para algunos estados.

ha recibido una atención muy amplia en la literatura reciente. Es bien conocido que ciertas propiedades del espectro están fuertemente relacionadas con la topología de la red representada por la matriz de adyacencia (Metha 1991; Crisanti *et al.* 1993). Por ejemplo, como hemos mencionado en la sección anterior, si una red crece con una probabilidad de conexión dependiente del número de nodos a partir de la siguiente relación $p \sim cN^{-\alpha}$, entonces la densidad espectral tiene la forma:

$$\rho(\lambda) = \begin{cases} \frac{\sqrt{4Np(1-p) - \lambda^2}}{2\pi Np(1-p)} & |\lambda| < 2\sqrt{Np(1-p)} \\ 0 & \text{en otro caso} \end{cases}$$

que, como ya mencionamos, se conoce como la ley de Wigner.

La densidad espectral de la matriz de adyacencia puede verse en la figura 7a. El eje horizontal de la figura ha sido reescalado como $\lambda/\sqrt{Np(1-p)}$ mientras que el eje vertical se reescaló como $\rho/\sqrt{Np(1-p)}$. El parámetro p es la *probabilidad de conexión implicada* que se calcula como sigue:

La desviación estándar de la centralidad de los grados de conexión k es $\sigma_k = 3.3963$. Este resultado se verificará. Por otra parte es conocido que $\sigma_k = \sqrt{Np(1-p)}$. Usando esta última relación calculamos la probabilidad implicada de conexión $p = 0.0020$. Como se muestra en Farkas *et al.* (2001), el valor propio dominante de la matriz de adyacencia λ_1 debe satisfacer la condición:

$$\lim_{N \rightarrow +\infty} \frac{\lambda_1}{N} = p$$

En nuestro caso hemos obtenido de manera aproximada esta relación pues $\lambda_1 = 9.7286$, lo cual nos brinda el siguiente resultado:

$$\frac{\lambda_1}{N} = \frac{9.7286}{5823} = 0.0017$$

con un buen acuerdo con el valor $p = 0.0020$ antes obtenido.

En la figura 7b se hace una comparación de la densidad espectral de la matriz de adyacencia con la ley del semicírculo de Wigner. Recuérdese que los ejes han sido reescalados convenientemente. Es importante notar la diferencia entre la figura 7b de este trabajo y la figura 3 de Farkas *et al.* (2001), donde la densidad espectral de redes de tipo mundo pequeño es comparada con la ley del semicírculo. Como puede verse, no existe ninguna asimetría observable en la densidad espectral de la matriz de adyacencia. No obstante, como veremos más adelante en este trabajo, nuestra red tiene propiedades de mundo pequeño.

Otro aspecto importante de la figura 7a es la comparación de las densidades espectrales de la matriz de adyacencia de toda la red y de la matriz de adyacencia de la red sin las conexiones aéreas. Nótese que ambas densidades espectrales son prácticamente iguales, lo cual quiere decir que las conexiones aéreas entre ciudades no influyen prácticamente en la topología de la red de ciudades mexicanas.

Una medida más sofisticada de la importancia de los nodos dentro de una red es la llamada *centralidad de vector propio* (Newman 2007; Goh *et al.* 2001). Esta medida distingue la importancia de los nodos dentro de la red en términos del nivel de conexiones que cada uno de ellos tiene. Obviamente, las conexiones a nodos altamente conectados son más importantes que aquéllas con nodos menos conectados.

Como es conocido —ver por ejemplo Newman (2007)—, si denotamos la

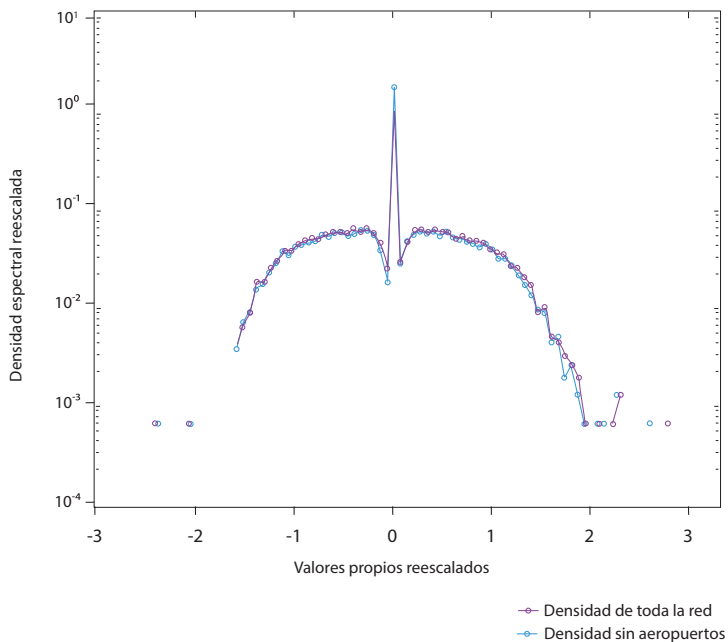


Figura 7a. Densidad espectral de la matriz de adyacencia. Como se explica en el texto, los ejes están reescalados. El gráfico en morado corresponde a toda la red en azul eliminando las conexiones aéreas.

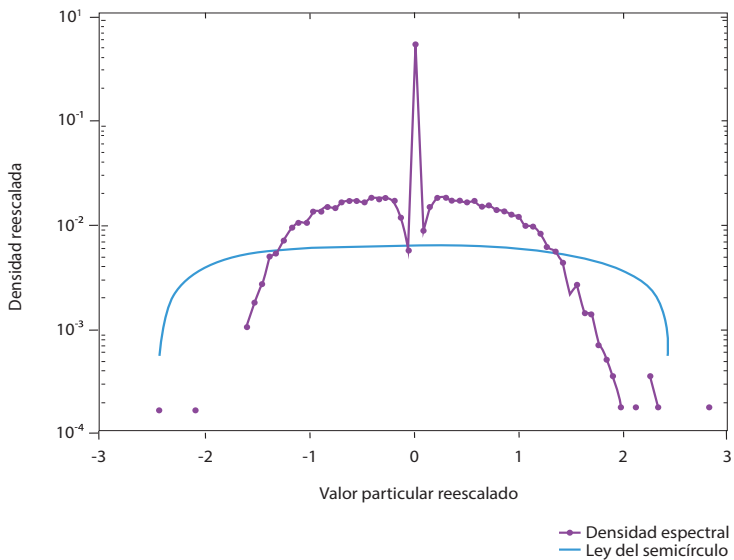


Figura 7b. Comparación de la densidad espectral de la matriz de adyacencia con la ley del semicírculo de Wigner.

centralidad de vector propio del nodo i -simo por x_i , entonces el vector $\mathbf{x} = (x_1, \dots, x_N)$ satisface $\mathbf{Ax} = \lambda_1 \mathbf{x}$ donde λ_1 es el vector propio dominante de la matriz de adyacencia \mathbf{A} .

En la figura 8 se muestran las componentes de vector propio de las centralidades \mathbf{x} . A diferencia de lo expuesto en Goh *et al.* (2001) en nuestra figura sólo se muestran las componentes de este vector y no sus cuadrados. De acuerdo con el Teorema de Perron Fobrenius, todas las componentes de este vector deben ser no negativas.

Se agregaron algunas etiquetas para designar algunos nodos importantes en la red de ciudades. Obviamente el de mayor importancia es el referido al Distrito Federal, aunque algunos otros (como Durango) también son muy relevantes.

Propiedades de mundo pequeño de la red de ciudades

Existen dos tipos de topologías que se han estudiado extensivamente en la literatura reciente: las redes ordenadas y las aleatorias. Pero existe además una fuerte evidencia empírica de que muchas redes sociales, tecnológicas y biológicas no pertenecen a estos dos grupos extremos. Más bien se encuentran en una situación intermedia. Como mostraremos en esta sección, la red de ciudades estudiadas tiene propiedades de mundo pequeño, lo cual es muy relevante para el estudio de la difusión de epidemias sobre la misma.

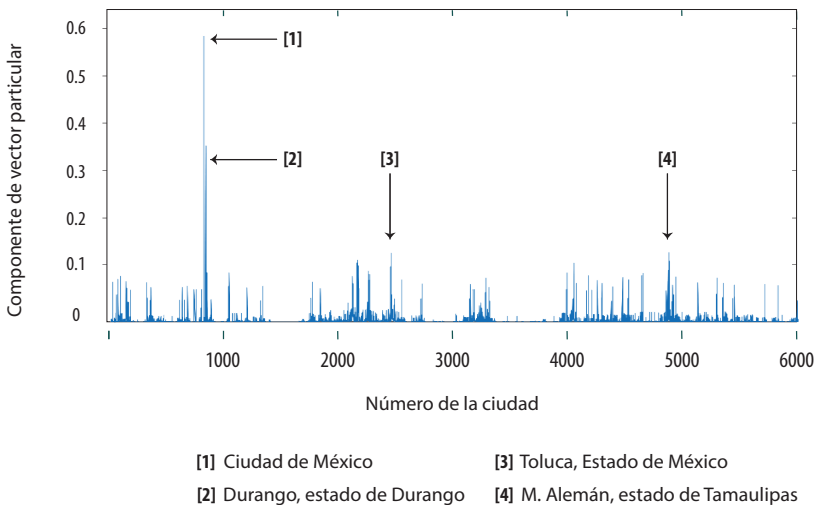


Figura 8. Componentes del vector de centralidades de los nodos de la red.

Las características básicas que distinguen a una red de mundo pequeño son las siguientes: un par de nodos puede ser conectado a través de una cadena de nodos intermedios de corta longitud. Otra propiedad está relacionada con la aglomeración local: dos nodos conectados ambos a un tercero, con una alta probabilidad están conectados entre sí (ver por ejemplo la definición 4.3.1 de Watts 1999; Barrat y Weigt 2000; Klemm y Eguiluz 2002).

En la tabla I se muestra la distribución de los caminos más cortos. En la primera columna se muestran las diferentes longitudes de los caminos, en la segunda columna se muestran el número de caminos de esa longitud que existen en la red estudiada y en la tercera columna se muestra la fracción de la totalidad de posibles caminos que esa cantidad representa. Obviamente, la tercera columna puede ser interpretada como la probabilidad de elegir un camino al azar del total de caminos posibles y que tenga la longitud referida.

El número total de posibles caminos en la red es $N(N-1)/2 = 16950753$. La cantidad real de los mismos es 16,649,481. La matriz A tiene solamente 10,312 coeficientes distintos de cero, por tanto, ésa es la cantidad de caminos de longitud 1 sobre la red. Como puede verse el diámetro de la misma es 27.

Denotemos por l la variable aleatoria que representa la longitud de un camino elegido al azar. Usando los datos de la tabla I tenemos $\langle l \rangle = 8.7146$. Por otra parte se tiene que $\ln N = \ln 5823 = 8.6696$. Ambos valores difieren en menos del 0.52%. Es bien conocido (ver Watts y Strogatz 1998; Newman 2003; Barthelemy y Amaral 1999; Bollobas 2001) que $\langle l \rangle = \ln N$ es una característica distintiva de las redes de mundo pequeño.

La distribución de los grados de los nodos en la red

La función de densidad de probabilidades de los nodos de una red es sin duda la propiedad más estudiada de las mismas. El interés por esta propiedad parece haberse desarrollado a partir de la comprobación de que muchas redes del mundo real poseen la propiedad de invariancia de escalas (Faloutsos *et al.* 1999; Chen *et al.* 2002). Por otra parte, como ya mencionamos con anterioridad, las redes aleatorias tipo Erdős-Rényi tienen función de distribución de sus grados de tipo Poisson (Erdős y Rényi 1960).

En la figura 9 se muestra la función cumulativa de probabilidades de la distribución de los grados en la red:

$$P(k > z) = \sum_{k>z} p_k$$

donde p_k es la probabilidad de observar en la red un nodo con k conexiones. Para esta distribución empírica se encontró un buen ajuste a una distribución tipo Beta (Mansilla *et al.* 2007; Martinez-Mekler *et al.* 2009):

Tabla I. Distribución de los caminos más cortos para cada longitud.

Longitud del camino más corto	Número de pares de ciudades	Probabilidad
1	10,312	0.0006193640767034
2	47,518	0.0028540479244366
3	146,139	0.0087774676886494
4	473,881	0.0284624581101888
5	1,017,893	0.0611371565290746
6	1,685,940	0.1012617020439560
7	2,262,213	0.1358740754510620
8	2,537,007	0.1523788787960600
9	2,443,484	0.1467616574475800
10	2,071,338	0.1244096535987770
11	1,554,809	0.0933856517392436
12	1,045,742	0.0628098359484027
13	639,825	0.0384294627983640
14	356,848	0.0214331683517690
15	187,519	0.0112628522400444
16	92,055	0.0055290496587401
17	43,260	0.0025983019742230
18	20,010	0.0012018498036108
19	8,534	0.0005125730246884
20	3,272	0.0001965243656879
21	1,227	0.0000736966371330
22	340	0.0000204212360434
23	119	0.0000071474326152
24	30	0.0000018018737685
25	17	0.0000010210618022
26	2	0.0000001201249179
27	1	0.0000000600624590

$$P(k > z) = 0.0323 \frac{(24 - z)^{1.4541}}{z^{1.513}}$$

La identificación de estos coeficientes se hizo por medio de una regresión lineal en las variables logarítmicas. El ajuste del modelo fue $R^2 = 0.9545$.

Por último queremos señalar que el valor esperado de los grados es $\langle l \rangle = 3.5732$ con una desviación estándar de $\sigma_k = 3.3963$. Este último número fue utilizado con anterioridad en este trabajo cuando se definía la probabilidad implicada de conexión.

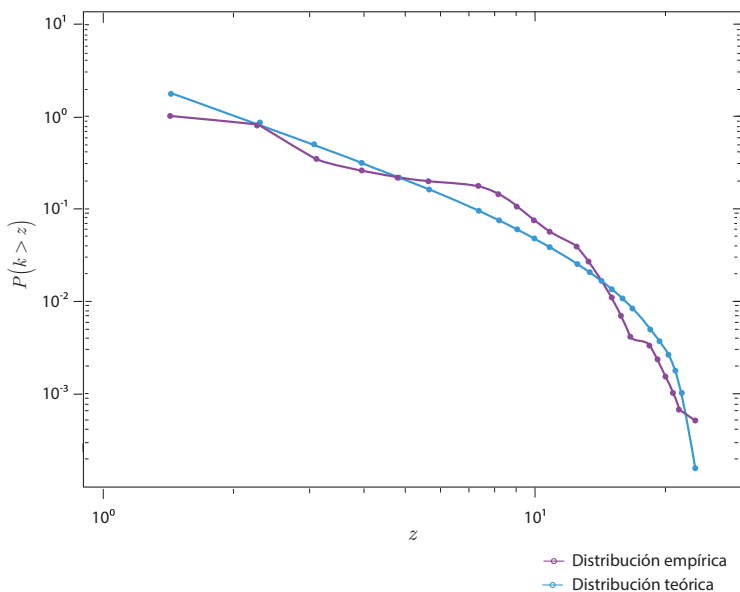


Figura 9. Función acumulativa de la distribución de los grados.

Conclusiones

Los resultados de esta investigación permiten establecer varios lineamientos de políticas públicas referidas al sector de la salud: la importancia de los aeropuertos en la transmisión de enfermedades; como hemos comentado con anterioridad, si bien el 65% del contagio de la gripe aviar (H5N1) iniciada en China en 2004 ocurrió a través de los aeropuertos (Mansilla 2015), en lo que a México se refiere, esta amenaza no es tan severa, toda vez que las propiedades topológicas de la red de ciudades no cambian drásticamente si las conexiones aéreas no son tenidas en cuenta, como se muestra por ejemplo en la figura 7a. En lo referente a la República Mexicana, el transporte por carretera ofrece un nivel de riesgo esencial en la transmisión de enfermedades contagiosas. Otro aspecto a tener en cuenta es la propiedad de mundo pequeño que exhibe la red de ciudades estudiadas. Como se mostró en este trabajo, la longitud del camino promedio para unir dos ciudades es de $\langle l \rangle = 8.7146$. Esto pone ciertas cotas a la vigilancia de los focos infecciosos. Otra característica importante de la red de ciudades es que, a diferencia de lo estudiado por otros autores, la distribución de los grados de esta red de ciudades no sigue una ley de potencias. Esto está fuertemente relacionado con la distribución de las centralidades de los nodos de la red mostrada en la figura 8, lo cual se convierte en una herramienta esencial en la elaboración de campañas de salud en el país, toda vez que han quedado

identificadas un grupo de ciudades con la propiedad de que si la epidemia las alcanza, se expandirá por todo el territorio nacional. ■

Referencias

- Albert, R. y A. Barabasi. «Statistical mechanics of complex networks.» *Reviews of Modern Physics* 74 (2002): 47.
- Barabasi, A. y R. Albert. «Emergence of Scaling in Random Networks.» *Science* 286 (1999): 509-512.
- Barrat, A. y M. Weigt. «On the properties of small-world network models.» *European Physical Journal B* 13 (2000): 547-560.
- Barthelemy, M. y L. Amaral. «Small world networks: Evidence for a crossover picture.» *Physical Review Letters* 82 (1999): 3180.
- Bilton, N. *Hatching Twitter*. Hodder & Stoughton, 2013.
- Boccara, N. y K. Cheong. «Automata network SIR models for the spread of infectious diseases in populations of moving individuals.» *Journal of Physics A* 25 (1992): 2447-2461.
- Bollobas, B. *Random Graphs*. Segunda edición. Academic Press, 2001.
- Chen, Q. et al. «The origin of power laws in Internet topologies revisited.» *Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communication Society*. IEEE Computer Society, 2002.
- Chung, F. y L. Lu. «The diameter of sparse random graphs.» *Advances in Applied Mathematics*. 26 (2001): 257-279.
- Crisanti, A. et al. «Products of random matrices in Statistical Physics.» *Springer Series in Solid-State Sciences* 104 (1993).
- Dorogovtsev, S. et al. *Phys. Rev. Lett* 85 (2000): 4633.
- Durkheim, E. *Les règles de la méthode sociologique (1894)*. 13ª edición, París, Presses Universitaires de France, 2007.
- Erdos, P. y A. Rényi. «On the evolution of random graphs.» *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5 (1960): 17.
- Euler, L. «Solutio problematis ad geometriam situs pertinentis.» *Commentarii academiae scientiarum imperialis Petropolitanae*, nº 8 (1741): 128-140.
- Faloutsos, M. et al. «On the power-law relationship of internet topology.» *Computer Communications Review* 29 (1999): 251-262.
- Farkas, I. et al. «Spectra of real world graph: Beyond the semicircle law.» *Physical Review E* 64, (2001).
- Goh, K. et al. «Spectra and eigenvectors of scale free network.» *Physical Review E*, nº 64, (2001).
- Kermack, W. y McKendrick, A. «Contributions to the mathematical theory of epidemics.» *Bulletin of Mathematical Biology* 53, 33-55, 1927.

- Klemm, K. y V. Eguiluz. «Growing scale-free networks with small-world behavior.» *Physical Review E* 65, (2002).
- Krapivsky, P. et al. *Phys. Rev. Lett* 85 (2000): 4629.
- Leskovec, J. y E. Horvitz. «Planetary scale views on instant-messaging networks.» *Microsoft Research Technical Report*. 2007. <http://arxiv.org/pdf/0803.0939v1.pdf>.
- Mansilla, R. «Nuevos paradigmas en la modelación de epidemias.» En *Las Ciencias de la Complejidad y la Innovación Médica. Aplicaciones*. Academia Nacional de Medicina de México, 2015.
- et al. «On the behavior of journal impact factor rank order distribution.» *Journal of Informetrics* 1 (2007): 155-160.
- y J. Gutiérrez. «Deterministic site exchange cellular automata models for the spread of diseases in human settlements.» *Complex Systems* 13 (2001): 143-159.
- Martinez-Mekler, G. et al. «Universality of rank-ordering distribution in the Arts and Science.» *PLoS ONE* 4, E4791 (2009): 1-7.
- Mendoza, R. y R. Mansilla. *The Network of Mexican Cities*. 2010. <http://arxiv.org/abs/1006.3887>.
- Metha, M. *Random matrices*. Academic Press, 1991.
- Milgram, S. «Behavioral Study of Obedience.» *Journal of Abnormal and Social Psychology* 67 (1963): 371.
- . «The Small World Problem.» *Psychology Today* 1 (1967): 60.
- Montoya, J. y J. Solé. *preprint cond-mat/0011195*. 2000.
- Moreno, J. *Sociometría, un método experimental y ciencia de la sociedad. Un acercamiento a una nueva orientación política*. Nueva York: Beacon House, 1951.
- Newman, M. *Proc. Natl. Acad. Sci* 98 (2001): 404.
- . *Phys. Rev. E* 64, 016131 (2001a).
- . *Phys. Rev. E* 64, 016132 (2001b).
- . «The Mathematics of networks.» En *The New Palgrave Encyclopedia of Economics*. 2007.
- . «The structure and function of complex networks.» *SIAM Review* 45 (2003): 167-256.
- Pagels, H. R. *The Dreams of Reason: The Computer and the Rise of the Sciences of Complexity*. Simon & Schuster, 1989.
- Pareto, W. *Cours d'Economie Politique*. Ginebra: Droz, 1896.
- Pool, I. y M. Kochen. «Contacts and Influence.» *Social Networks* 1 (1978): 1.
- Radcliffe-Brown, A. «On Social Structure.» *Journal of the Royal Anthropological Institute* 70 (1940): 1.
- Rapoport, A. «Spread of information through a population with socio-structural

- bias. Assumption of transitivity.» *Bulletin of Mathematical Biology* 15 (1953): 523.
- . «Spread of information through a population with socio-structural bias. Various models with partial transitivity.» *Bulletin of Mathematical Biology* 15 (1953a): 535.
- Simmel, G. *Sociología: estudios sobre las formas de socialización*. Original de 1908. Madrid: Revista de Occidente, 1977.
- Solomonoff, R. y A. Rapoport. «Connectivity in random nets.» *Bulletin of Mathematical Biology* 13 (1951): 107.
- Tönnies, E. *Gemeinschaft und Gesellschaft. Grundbegriffe der reinen Soziologie*. Darmstadt: Wissenschaftliche Buchgesellschaft, 1912.
- Watts, D. *Small worlds*. Princeton University Press, 1999.
- y S. Strogatz. *Nature* 393 (1998): 440.
- Wigner, E. *Ann. Math* 62 (1955): 548.
- . *Ann. Math* 65 (1957): 203.
- . *Ann. Math* 67 (1958): 325.
- Zipf, G. *Human Behaviour and the Principle of Least Effort*. Hafner Publishing, 1966.

