Afroamericana, litografía, 37 x 55 cm, 2007

Algunos aspectos de la modelación matemática de redes reales

◆ Gabriela Hinojosa

En una amplia variedad de "sistemas" encontramos lo que se conoce como "estructura de red real". El término "red real" hace referencia a redes en sentido matemático, las cuales representan procesos o sistemas que forman parte de nuestra vida. Por ejemplo, una gran parte de la física consiste en estudiar un amplio número de partículas (como las moléculas que constituyen la atmósfera) y la forma en que estas interactúan. La célula se puede describir como una red de componentes químicos conectados por reacciones químicas.

También encontramos redes reales sociales, como la que forma la colaboración científica, donde los nodos (vértices) son los científicos y dos nodos están conectados por un segmento de línea, si ambos escribieron un artículo juntos. Las llamadas telefónicas de larga distancia es otro ejemplo, cuyos nodos son los números de teléfono y cada llamada de larga distancia concluida representa una conexión entre dos de ellos. De la misma manera podemos hablar de redes ecológicas, comerciales, entre otras.

El internet es una enorme red real de computadoras y servidores, conectadas de forma física e inalámbrica, que afecta muchos aspectos de nuestra vida cotidiana, como la manera en que guardamos y recuperamos información, realizamos compras, operaciones bancarias y, en general, la manera como nos comunicamos. Por ejemplo, actualmente la información no se encuentra únicamente en forma impresa, sino también se puede encontrar *on line*, es decir, a través de un complejo conjunto de páginas web interconectadas.

En este artículo nos enfocaremos en el estudio de lo que llamamos "gráfica web", denotada por *W*. Para comenzar, como su nombre lo indica, *W* es una gráfica en términos matemáticos, que consiste de vértices (puntos) que representan las páginas web y aristas (segmentos de recta) que los unen, los cuales corresponden a los enlaces (*links*) entre ellas.

En los últimos años, la gráfica web ha sido un campo muy activo de estudio, tanto teórica como experimentalmente. La web no solo es fascinante por sí misma, sino que además nos brinda un mejor entendimiento acerca de las redes reales en general, así como su evolución. La complejidad de *W* radica en que es una estructura evolutiva con páginas y enlaces que aparecen y desaparecen continuamente en el tiempo, por lo que la primera pregunta que surge es qué tan grande es la web.

Por supuesto que todos sabemos que existe un número muy grande de páginas web y enlaces entre ellas, por lo que es muy difícil obtener una

◆ Profesora e investigadora, Facultad de Ciencias, UAEM

cifra exacta del número de páginas. En 2005, un estudio de Hirate encontró 53.7 billones de páginas web, pero dicho número está cambiando constantemente.[1]

Diversos experimentos en la web han demostrado que su estructura local en una escala microscópica luce muy similar a una célula biológica, pero su estructura global es muy distinta.[2] Se dice que *W* tiene estructura en forma de corbata de moño. El nudo de la *corbata* consiste de un componente llamado *núcleo central*, el cual está formado por páginas tales que cualesquiera dos de ellas están enlazadas. Este enorme componente es el corazón de la web y se denota por SCC (*giant strongly connected component*), por sus siglas en inglés (Fig. 1).

El segundo componente (IN) es el lado izquierdo de la *corbata* y consiste en páginas que se enlazan al SCC, aunque ninguna página del núcleo se enlaza con alguna de ellas, por lo que estas páginas posiblemente son nuevos sitios todavía no descubiertos por los usuarios y, por lo tanto, no se han enlazado con ellas. El tercer componente (OUT) es el lado derecho de la *corbata* y está formado por páginas que son accesibles desde el SCC, pero que no están enlazadas con él. Finalmente, las páginas restantes están aisladas o forman pendientes y únicamente se enlazan a páginas que no pertenecen al SCC.

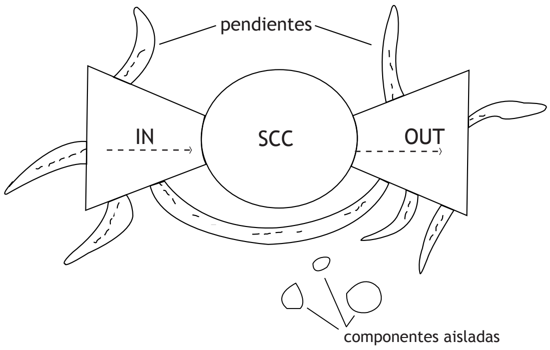


Fig. 1. Representación gráfica (*W*) de la estructura de la web

Se han hecho varias estimaciones del tamaño de las distintas regiones. Así, mientras en el año 2000 se encontró que aproximadamente la tercera parte de las páginas estaban contenidas en el núcleo, un estimado más reciente muestra que más de las dos terceras partes de todas las páginas web pertenecen al SCC.[3] Este crecimiento en el tamaño del núcleo refleja el incremento de la conectividad entre las páginas web, lo que nos indica que la gráfica web está cambiando constantemente. Sin embargo, conviene preguntarse cómo se está llevando a cabo este cambio, es decir, si *W* tiene propiedades que la caracterizan; si acaso existen buenos y rigurosos modelos matemáticos que describen estas propiedades y si nos sirve la estructura gráfica para obtener información de *W*. La respuesta a varias de estas preguntas es que sí.

[1] Y. Hirate, S. Kato, H. Yamana, "Web structure in 2005", en *Proceedings of the 4th Workshop on Algorithms and Models for the Web-Graph*, 2006, pp. 36-46.

[2] A. Broder *et al.*, *Graph structure of the web*, en 9th International Word Wide Web Conference, http://www9.org/w9cdrom/160/160.html, consultado en marzo de 2011.

[3] Y. Hirate, S. Kato, H. Yamana, "Web structure…", *op. cit.*

El objetivo de este artículo es presentar una introducción al fascinante tema de la modelación matemática de la web. Para ello se darán las propiedades que se consideran más sobresalientes, aunque esto no significa que no existan otras. Deseo enfatizar que no existe un modelo que describa completamente *W*; de hecho existen varios modelos para ello, algunos de los cuales explican mejor que otros dichas propiedades. Esto se debe a que la gráfica web es sumamente compleja. Muchos trabajos recientes han aplicado algoritmos de gráficas en *W* con muy buenos resultados. A continuación introduciré algunos de estos conceptos teóricos básicos.

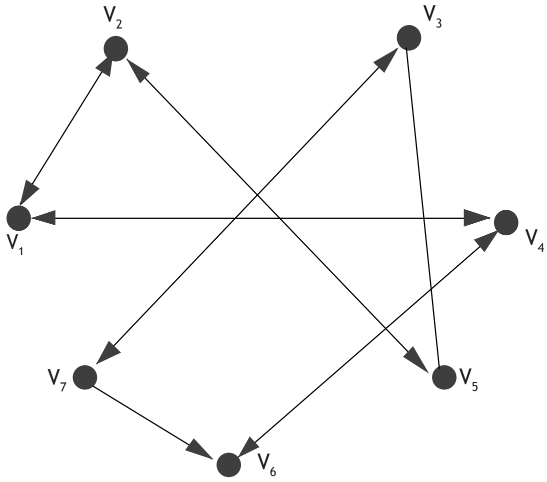


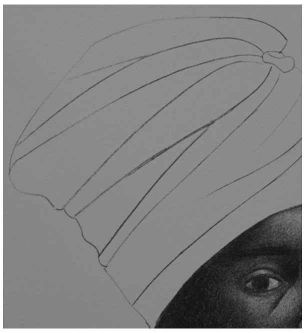
Fig. 2. Diagrama de puntos y líneas que representan vértices y aristas

Teoría de gráficas

La teoría de gráficas es una de las áreas más antiguas de las matemáticas. Una de sus principales aplicaciones sirve para la modelación de redes en el mundo real. Una *gráfica dirigida G* consiste de un conjunto no vacío de vértices *V(G)* y en un conjunto de aristas *E(G)*. Cada arista es un par orientado de vértices *(u,v)* que representan una conexión directa de *u* a *v*. Típicamente, una gráfica se dibuja como un diagrama de puntos, que representan los vértices, y de líneas entre ellos, que representan las aristas (figura 2). Aquí se está excluyendo el caso en que *V(G)* sea vacío, ya que así, al no tener vértices, tampoco hay aristas, por lo que no tenemos una gráfica. Sin embargo *E(G)* sí puede ser vacío.

Como ya se mencionó, en el caso de la gráfica web *W* los nodos representan páginas web, y las aristas, enlaces entre ellas. La gráfica *W* frecuentemente se considera como una gráfica dirigida, ya que sus aristas tienen una orientación que indica qué página se conecta con cuál.

El *grado de salida* de un vértice *u* es el número de distintas aristas *(u,v1), ..., (u,vk)*; es decir, es el número de enlaces de *u*. Si *u* es una página web, entonces *u* establece enlaces con otras páginas, y ese número es su grado de salida. El *grado de entrada* es el número de distintas aristas *(v1,u), …, (vk,u)*; en otras palabras, es el número de enlaces *a u*. En una página web *u* será el número de páginas que se enlazan con ella. El *grado de un vértice u*, denotado por *grad(u)*, es la suma de su grado de salida más su grado de entrada. Por ejemplo, en la figura anterior, el grado de salida del vértice *v6* es 2 y el grado total grad(v_6) es 4.



Una *trayectoria* de un vértice *u* a un vértice *v* es una secuencia de aristas *(u,u1), (u1,u2), …, (uk,v)*. Observemos que una trayectoria de *u* a *v* no implica una trayectoria de *v* a *u*. Definimos la *longitud de una trayectoria* como el número de aristas que la conforman. Así, en el ejemplo anterior hay dos trayectorias de v_2 a v_3, a saber *y* que consiste de las aristas (v_2,v_5) y (v_5,v_3), y *α* dada por las aristas (v_2,v_1), (v_1,v_4), (v_4,v_6), (v_6,v_7) y (v_7,v_3).

La *distancia* de *u* a *v*, denotada por *d*(*u*,*v*), es el más pequeño entero *k*, tal que existe una trayectoria de longitud *k* entre *u* y *v*. Si no existe una trayectoria de *u* a *v*, decimos que la distancia de *u* a *v* es infinita. Si (*u*,*v*) es una arista, entonces *d*(*u*, *v*)=1. Retomando nuestro ejemplo, *d*(*v2*, *v3*)=2.

En el caso de la gráfica web, se ha encontrado que, en promedio, el número de enlaces requeridos para ir de una página a otra es menor a 16.[4] Esto, como explicaremos más adelante, muestra que aunque *W* es una gráfica enorme, en realidad parece ser pequeña.

Propiedades de la web

Aunque este artículo se enfoca en la gráfica *W*, las siguientes propiedades caracterizan a otras gráficas del mundo real, como redes biológicas, sociales y tecnológicas. Recordemos que *W* tiene una enorme cantidad de nodos, por lo que para su modelación estaremos siempre considerando promedios.

Dispersión. Un rasgo importante de *W* es que es *dispersa*; esto es, el grado promedio de una página es muy pequeño en relación con el número total de vértices, denotado por |*V(W)*|. Matemáticamente hablando, la frase anterior se expresa como: el grado promedio de *W* es menor o igual al producto e|*V(W)*|, donde *e* es un número positivo menor a 1.

Distribución de la ley de potencia del grado. Esta es una de las principales características observadas en *W* y se refiere a lo siguiente: dado un entero positivo *k*, definimos *Nk* como el número de vértices de la gráfica web que tienen grado *k*. Así, la distribución del grado de *W* satisface lo que se conoce como la ley de potencia, esto es, para cada grado *k*, el promedio de páginas de *W* con grado *k*, entre el tiempo *t*, *Nk* ÷ *t*, *se aproxima asintóticamente a* k^{-b}, para una constante real *b* >*1*. En otras palabras, el cociente *Nk* ÷ *t* es muy cercano al valor k^{-b} cuando *t* es "muy grande".

La distribución de la ley de potencia del grado refleja un cierto aspecto "antidemocrático" de *W*, ya que mientras la mayoría de las páginas tiene un pequeño número de enlaces, algunas tienen un gran número. Esto no es sorprendente, ya que la elección de los enlaces de las nuevas páginas a las ya existentes está determinada por los intereses de los usuarios.

Propiedad del mundo pequeño. Debido a los avances tecnológicos como el correo electrónico y los teléfonos celulares, el mundo parece un lugar más pequeño. Una medida global muy conocida es lo que llamamos el diámetro de *W*, *diam(W)*, que es igual al promedio de las distancias *d(u,v)* entre

---

[4] *Ibid.*

cualesquiera dos vértices *u*, *v*, con la propiedad de que *d(u,v)* es finita. Así, la propiedad del mundo pequeño indica que *diam(W)* es mucho más pequeño que el número de vértices de *W*, lo cual a groso modo nos dice que, en promedio, a través de un número pequeño de enlaces podemos ir de una página a otra.

Estructura de comunidad. La web contiene muchas *comunidades*, es decir, conjuntos de páginas que comparten un interés o un tema común. Todavía no existe un consenso para una definición precisa de comunidad en *W*. Sin embargo, en trabajos recientes el término comunidad se refiere a componentes conexos locales de la gráfica, los cuales contienen una estructura muy rica.

Características de la modelación

Como ya habíamos comentado, modelar *W* no es tarea fácil. Para esto, es necesario reducir el número de parámetros que intervienen, así como simplificar su inherente complejidad. Los buenos modelos para *W* comúnmente utilizan unos cuantos parámetros, por lo que deben captar los rasgos más sobresalientes de *W*. En general, el equilibrio entre la simplicidad de diseño y las dificultades de analizar el modelo hacen de esta materia todo un desafío.

En un principio, los modelos de *W* fueron planteados y analizados de una forma poco estricta. El reto actual es diseñar modelos matemáticamente rigurosos que simulen una o más de las características observadas en *W*. En la actualidad, los modelos más comunes son los llamados *modelos estocásticos*; esto es, modelos en los cuales las gráficas se elaboran sobre una sucesión infinita de tiempo-escalón, vía ciertas reglas probabilísticas. Para ello, se utilizan como herramienta las *gráficas random o gráficas aleatorias*.

La teoría de gráficas aleatorias estudia las propiedades del espacio de probabilidad asociado a las gráficas con N vértices, cuando N tiende a infinito. Las gráficas aleatorias fueron estudiadas primero por los matemáticos húngaros Paul Erdös y Alfréd Rényi en la década de 1950. En el modelo de Erdös-Rényi se empieza con N vértices y cada par de ellos está conectado con probabilidad p, donde $0<p<1$, lo que crea un espacio de gráficas tales que sus aristas están distribuidas aleatoriamente. Dicho espacio se denota por $G(n,p)$.

Muchas propiedades de las gráficas aleatorias se demuestran usando argumentos probabilísticos. Al respecto, Erdös y Rényi usaron la definición probabilística: toda gráfica aleatoria de N vértices tiene una propiedad Q, si la probabilidad de tener Q se aproxima a 1 cuando N tiende a infinito.

En la última década, se ha propuesto un gran número de modelos rigurosos para la gráfica *W* y cada uno de ellos nos ayuda a profundizar en la comprensión de los mecanismos que operan en *W*, así como en su evolución. Actualmente, el enfoque que busca entender el comportamiento de la red como un todo está tomando un papel más activo. En esta dirección, el entendimiento de la topología de la red es inevitable, por lo que existen muchas preguntas que todavía no tienen respuesta.