

Modelos matemáticos de la sociedad y aplicaciones. Crecimiento de las redes sociales.

Nelia Tello¹

*Escuela Nacional de Trabajo Social, Universidad Nacional Autónoma de México,
México.*

José Antonio de la Peña²

*Instituto de Matemáticas, Universidad Nacional Autónoma de México y Consejo
Nacional de Ciencia y Tecnología, México.*

Resumen

Revisamos varios modelos de crecimiento que han sido propuestos para redes: el crecimiento aleatorio de Erdős y Rányi, el crecimiento libre de escala de Barábasi propuesto originalmente para el internet, el modelo α de crecimiento de Watts y Strogatz. Una aplicación del Teorema de Turán muestra que la optimización del α -crecimiento implica la aparición de clanes en la red. Nuestro modelo de crecimiento propone una combinación del crecimiento del tipo libre de escala y el α -crecimiento, una vez satisfecho el principio de acotamiento del número de vecinos $c(i) \leq D$, para toda i , por un número de Dunbar D fijado a priori. Alternativamente al establecimiento de relaciones de acuerdo al principio de preferencia de los más populares que da lugar, en algunos casos, al crecimiento libre de escala, proponemos el establecimiento de nuevas relaciones de acuerdo al principio de preferencia de los más poderosos.

Palabras clave: crecimiento libre de escala, modelo α de crecimiento, número de Dunbar, teorema de Turán.

Abstract

We review several models of growth of networks: random growth as proposed by Erdős and Rányi, the scale-free networks as proposed by Barábasi for the growth of the internet, the α -model of Watts and Strogatz. We apply Turán's theorem to show that the optimization of the α -model implies the appearance of clans in a social network. The growth model we propose for social networks combines the properties of scale-free type networks and the α -model once the bound $c(i) \leq D$, for all vertices i , is satisfied for an a priori constant D called the Dunbar number. Alternative to the establishment of new relations according to the preference of the most popular, we propose the establishment of new relations preferring the most powerful.

Key words: growth of a network, model α , Dunbar number, Turán theorem.

Hace 2000 años, la población mundial era de 300 millones de habitantes. Transcurrieron más de 1600 años para que la población del mundo se duplicara

¹E-mail: neliatello@me.com

² E-mail: jap@matem.unam.mx

para llegar a los 600 millones de habitantes. Alrededor de 1800 se alcanzaron los primeros 1000 millones de habitantes y alrededor de 1927, los 2 mil millones de habitantes. Actualmente se estima la población mundial en 6,300 millones de habitantes.

El modelo más simple que expresa el crecimiento de la población es un modelo exponencial de la forma

$$P(t) = P(t_0)e^{r(t-t_0)}$$

Donde $P(t)$ es la población medida en el tiempo t , el tiempo de referencia a partir del que se calcula es t_0 , con los tiempos medidos en años y r es la tasa de crecimiento anual en el año t_0 . Por ejemplo, la tasa de crecimiento actual de la población es de 1.7%, es decir se incorporan 17 personas por cada mil cada año. Con esta tasa de crecimiento la población tardaría en duplicarse

$$2 = e^{0.017 t}, \text{ o sea, } t = \frac{\ln 2}{0.017} = 40.7,$$

esto es, poco más de 40 años.

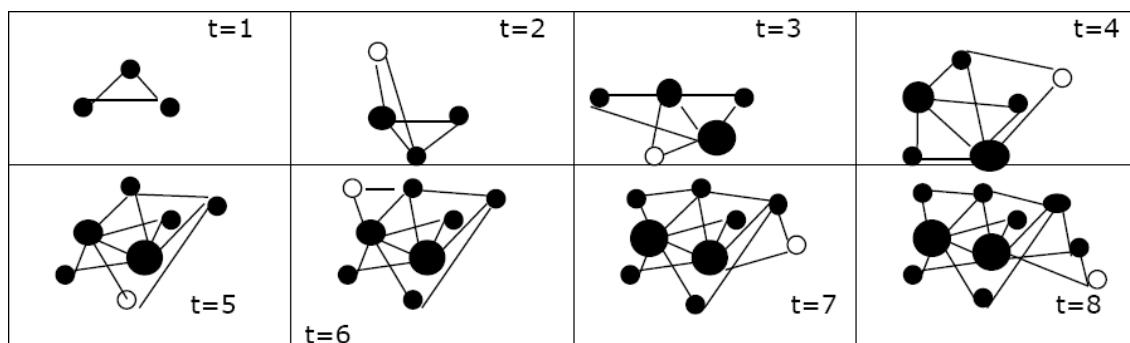
Las importantes variaciones en el crecimiento de la población en el siglo XX se debieron a variados factores. Por una parte, campañas de higiene y vacunación tuvieron efectos dramáticos en la disminución de la mortandad infantil, por otro lado, la población se fue transformando en más urbanizada y por ello menos dependiente de las grandes familias, lo que hizo descender las tasas de crecimiento en los países desarrollados.

Modelos de crecimiento de redes

Una red social S , como los habitantes de un país o una región de la tierra, al crecer, no sólo lo hace como una agregación de nuevos miembros, sino también como una adición de nuevas relaciones, que ligan tanto a los nuevos individuos de la red con individuos ya presentes, como individuos ya presentes entre sí. ¿De qué manera se agregan estas nuevas relaciones?

Por muchos años se estudió el modelo de las *redes aleatorias* introducido por Erdős y Ranyi en 1959, como el paradigma del comportamiento de una red con el paso del tiempo. Sin embargo, la idea es poco atractiva, pues las relaciones humanas no se establecen al azar, sino siguiendo una controlada toma de decisiones. Hace 10 años, Barábasi y Albert (1999) propusieron un modelo alternativo de crecimiento al

observar ciertas propiedades de las redes de internet. Así, sugirieron que un elemento nuevo que se une a la red establece sus lazos de *relación preferencial* con individuos que tienen alto número de relaciones ya establecidas, como en el ejemplo siguiente.



En cada paso un nuevo nodo se adiciona a la red y las relaciones preferenciales se establecen con individuos en la red que tienen ya alto número de vecinos. Así, los nodos con mayor número de vecinos tienden a tener más vecinos en el futuro, situación que se conoce como el efecto San Mateo. En las gráficas, el tamaño de los nodos crece con el número de vecinos. Si se agregan dos relaciones cada vez que se agrega un nuevo nodo, puede demostrarse que el número de nodos que tienen k vecinos se aproxima a $1/k^3$.

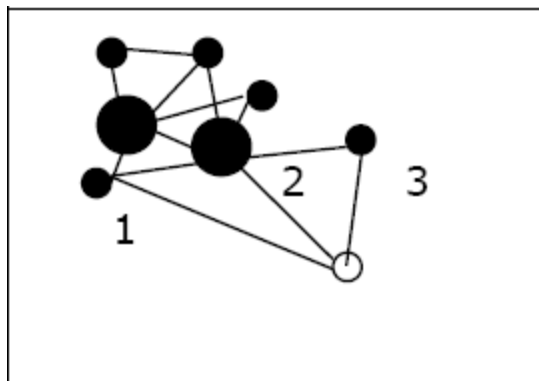
Mientras el modelo de crecimiento aleatorio predice una distribución normal para la función $C(k)$ que cuenta el número de nodos en la red S que tienen exactamente k vecinos (esto es, los valores $C(k)$ se toman sobre una campana de Gauss), en el modelo de crecimiento con relaciones preferenciales $C(k) = a k^{-g}$ es una función potencial con g una constante positiva que depende de la red. Estas redes se llaman *libres de escala* porque la arquitectura de la red se ve igual ya sea la red pequeña o muy grande de tamaño. Esta es una diferencia esencial que resulta muy adecuada para distinguir el comportamiento de redes aleatorias de las redes de internet, según observaron Barábasi y Albert (1999). Muchas redes en la naturaleza se comportan como redes libres de escala, fenómeno que ha atraído la atención de numerosos científicos de diferentes campos de conocimiento, ver por ejemplo, la reciente revisión de Barábasi (2009).

En el modelo de crecimiento preferencial el grado $c_t(x)$ de un nodo x crece con el tiempo t , es decir, con la edad del nodo. Este fenómeno no se observa en las páginas de la *www*, por ello se argumenta que la red de redes no tiene crecimiento preferencial a pesar de tener un crecimiento libre de escala. Para más de esta discusión ver Newman (2002).

En una red muy grande como internet, o Twitter, el número de contactos que establece un nodo puede ser arbitrariamente grande, en efecto, la función $C(k) = ak^g > 1$ tiene solución para k suficientemente grandes. Por otra parte, recordemos que esto no sucede así en el caso de las redes sociales que estamos considerando, pues según hemos discutido desde el capítulo 1, el número de vecinos de cada nodo está acotado por arriba por el número de Dunbar D , que en general satisface $D \leq 150$. Esto es, nuestras *redes sociales no son libres de escala*.

α -crecimiento y el teorema de Turán.

En las redes sociales otro tipo de fenómeno debe considerarse. Propuesto por Watts y Strogatz (1998), el *modelo α de crecimiento* de una red propone que las relaciones de un nodo nuevo x en la red se construyen dando preferencia a enlaces con nodos que están relacionados con nodos que ya tienen relación con x . Esto lo ilustramos en la siguiente red:

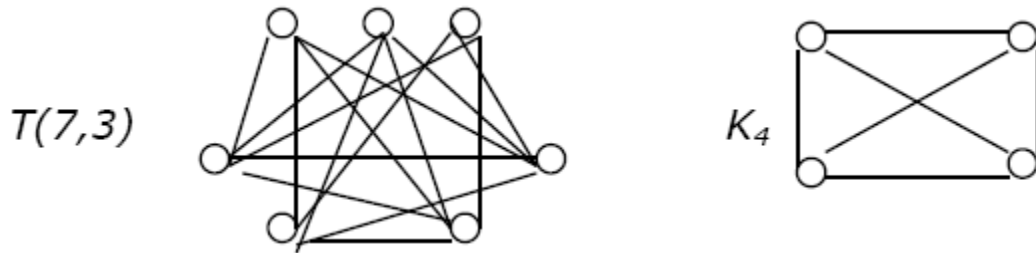


Modelo α de crecimiento: una vez establecida la relación del nodo nuevo con 2, es más probable que se relacione con 1 o con 3 que con cualquiera de los otros vértices.

Si la red S sigue el modelo α de crecimiento entonces los nodos de S tienen una tendencia a formar clanes, es decir, conforme el tiempo pasa y se van agregando nuevos nodos, el número de clanes $Cl(k)$, con k nodos, crece. Este resultado se puede derivar como consecuencia del Teorema de Turán en combinatoria.

En 1941, Paul Turán inició el estudio de los problemas extremos en gráficas mostrando el siguiente resultado. Dados dos números $s \leq n$ y una partición (n_1, n_2, \dots, n_s) de n con $n_1 \leq n_2 \leq \dots \leq n_s \leq n_1 + 1$, la gráfica $T(n, s)$ obtenida uniendo cada vértice de la clase i con cada vértice de la clase j , para todos $1 \leq i < j \leq s$, es maximal con la propiedad de no contener una subgráfica completa K_{s+1} . Ilustramos esto con la gráfica $T(7, 3)$ que es maximal sin contener a K_4 , esto es,

$T(7,3)$ no contiene clanes con 4 nodos pero agregando cualquier arista, la gráfica obtenida contiene algún clan con 4 nodos.



Supongamos que G es una gráfica con n nodos y e aristas, el Teorema de Turán (1941) afirma que si la desigualdad

$$e > \left(1 - \frac{1}{r-1}\right) \frac{n^2}{2}$$

es válida, entonces G contiene una subgráfica completa K_r . Un resultado relacionado y, en cierto sentido, más preciso utiliza la matriz A de la red y las trazas de potencias de A .

Teorema [de la Peña (2010)]: Sea G una gráfica con número de Dunbar D . Consideremos $A=A(G)$ la matriz de la red. Si para $r \geq 3$ se satisface

$$tr(A^3) > \left(1 - \frac{1}{r-2}\right) \sum_{i=1}^n c(i)^2$$

entonces

- (a) G contiene una subgráfica completa K_r ;
- (b) Si $r \geq 3$, al menos $\frac{tr(A^3)}{6}$ nodos de G pertenecen a triángulos;
- (c) Si $r \geq 4$, al menos $\frac{n}{4(1+D^2)}$ nodos de G pertenecen a clanes con r nodos.



Ejemplos: Para la red K_4 la traza de A^3 es 24 y $c(x)=3$ para todo nodo x , entonces el Teorema es válido con $r=4$. En la red de arriba con $2m$ nodos ($m \geq 3$) se identifican la primera y última columnas. La traza de A^3 es $24m$ y $c(x)=5$ para todo nodo x y el Teorema es válido con $r=3$. Para $m > 3$ el teorema de Turán no da ninguna información.

Recordamos que $c(i)$ es el número de vecinos del nodo i en G . En general, la cantidad $\Delta(G) = \frac{tr(A^3)}{6}$ cuenta el número de triángulos en G . Si G' se obtiene de G

agregando un nuevo nodo $*$, entonces $\Delta(G') - \Delta(G)$ es el número de triángulos que pasan por $*$. Si $A(G') = (a_{xy})$ entonces

$$\Delta(G') - \Delta(G) = \sum_{1 \leq i, j \leq n} a_{*i} a_{ij} a_{j*}$$

de donde se sigue que la optimización del α -crecimiento en una red G es equivalente a la creación de un máximo de triángulos, a su vez, según el Teorema, esto produce un aumento del número de clanes en la red.

Otra cantidad medida con frecuencia en una red S con matriz de red $A = A(S) = (a_{ij})$ es el cociente

$$\text{trans}(S) = \frac{6\Delta(S)}{\sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n a_{ik} a_{kj}}$$

conocido como el *índice de transitividad* (o el *número de acumulación*³) este número mide la fracción de caminos de longitud dos en S que se terminan cerrando para formar un triángulo. En otros términos, $\text{trans}(S)$ es la probabilidad de que el amigo de un amigo sea mi amigo también.

El índice de transitividad ha sido ampliamente medido en redes sociales, por ejemplo, la página UCInet incluye un programa para calcular este índice para cualquier red. La red de actores de cine (donde las relaciones se establecen por haber participado en la misma película) tiene un índice de transitividad de 0.2, la red de directores de películas (donde la relación se establece por haber dirigido para un mismo estudio) tiene un índice de 0.6, la red de coautoría de artículos matemáticos tiene un índice de 0.15.

En términos matriciales, es claro que el índice de transitividad puede reescribirse como

$$\text{trans}(S) = \frac{\text{tr}(A^3)}{\sum_{i=1}^n c(i)^2}.$$

Con lo cual nuestro teorema anterior se puede reescribir como:

Corolario: Si $\text{trans}(S) > (1 - \frac{1}{r-2})$, entonces S tiene un clan K_r .

³ La terminología en inglés es mejor conocida, el *clustering number*, lo que lleva a confusiones pues así también es llamada otra cantidad asociada a las gráficas.

Los principios de crecimiento de una red social

Consideremos una serie de redes sociales $S_1, S_2, \dots, S_n = S$ de manera que S_{t+1} se obtiene de S_t añadiendo el nodo $t+1$. La matriz de la red S_t es $A_t = (a_{xy}^{(t)})$ y $c_t(x)$ es el número de vecinos del nodo x en la red S_t . Sea r_t el radio espectral de A_t y u_t el vector de Perron correspondiente satisfaciendo $\max_{1 \leq x \leq t} u_t(x) = 1$. Consideremos los siguientes principios:

(1.t) Para el tiempo t ,

$$c_{t+1}(x) \leq D, \text{ para } 1 \leq x \leq t \text{ y } c_{t+1}(t+1) \leq d;$$

donde $d < D$ son constantes independientes de t ;

(2.t) Para el tiempo t , la cantidad

$$\sum_{1 \leq x \leq t} a_{x,t+1}^{(t+1)} c_t(x)$$

es máxima posible;

(2'.t) Para el tiempo t , la cantidad

$$\sum_{1 \leq x \leq t} a_{x,t+1}^{(t+1)} u_t(x)$$

es máxima posible;

(3.t) Para el tiempo t , la cantidad

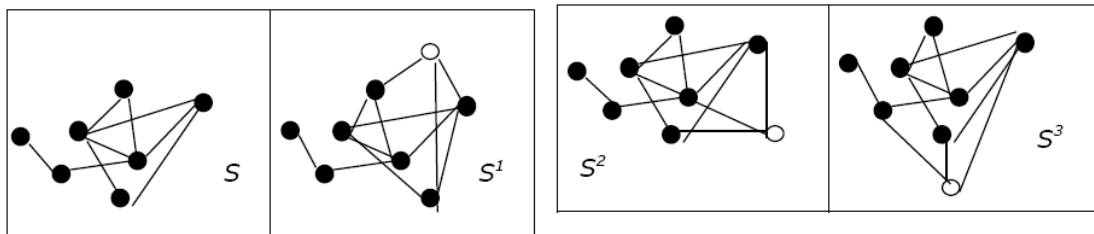
$$\Delta(S_{t+1}) - \Delta(S_t)$$

es máxima posible.

En otras palabras, si la red S cumple los tres principios (1.t), (2.t) y (3.t), para todo tiempo t , entonces S satisface el acotamiento de Dunbar, el modelo de crecimiento preferencial de Barábasi y el modelo de α -crecimiento de Watts-Strogatz. Si existe un número $\frac{1}{2} < p \leq 1$, tal que (1.t), (2.t) y (3.t) se cumplen con probabilidad p , entonces diremos que la red social S tiene *crecimiento orientado hacia la popularidad*⁴.

⁴ Por supuesto, puede pasar que el máximo del número $\sum_{1 \leq x \leq t} a_{x,t+1}^{(t+1)} c_t(x)$ se alcance construyendo una extensión S' de S que viole el principio (1), o bien que el máximo de $\Delta(S_{t+1}) - \Delta(S_t)$ se alcance construyendo una extensión S' de S que viole (2), estos casos deben excluirse. Esto es, se debe buscar que se satisfaga (2) de entre las extensiones de S que satisfacen (1) y buscar que se satisfaga (3) de entre las extensiones de S que satisfacen (1) y (2).

Modelos de crecimiento de redes sociales basados en principios similares han sido propuestos antes. Por ejemplo, Jin, Girvan y Newman (2001) estudian redes cuyo crecimiento se basa en encuentros de individuos regidos por los siguientes principios generales: (1) los encuentros son más frecuentes entre individuos que tienen un conocido común o más, (2) los encuentros entre individuos que se encuentran poco disminuyen rápidamente, (3) hay una cota común al número de amistades que un individuo puede sostener. Usando simulaciones por computadora, estos investigadores mostraron que modelos con estas características reproducen muchos de los fenómenos observados en las redes sociales reales, por ejemplo, los altos niveles de concentración en algunos nodos, la fuerte estructura comunitaria en que los individuos tienen más enlaces con otros individuos en su comunidad que con otros fuera de ella.



Las redes S^j se obtienen de la red S por extensión en un nodo de manera que se cumplan los principios (1), (2) y (3) con $3=d < D=4$. El máximo valor posible para $\sum_{1 \leq x \leq t} a_{x,t+1}^{(t+1)} c_t(x)$ es 7 que se puede obtener como suma de tres números $3+2+2$, las redes construidas corresponden a algunos de estos casos. El valor de $\Delta(S^j) - \Delta(S)$ es 1, 2, 1, según $j=1,2,3$. Todos satisfacen (2) una vez que se han elegido las relaciones del nuevo nodo.

Si una red S tiene un crecimiento orientado hacia la popularidad, entonces en cada paso t , los enlaces del nodo $t+1$ se buscan de manera que se maximice su *popularidad* (número de vecinos). Por otra parte, si se cumple (2'.t), los enlaces del nodo $t+1$ se buscan de manera que se maximice su *poder*. Si existe un número $\frac{1}{2} < p \leq 1$, tal que (1.t), (2'.t) y (3.t) se cumplen con probabilidad p , entonces diremos que la red social S tiene *crecimiento orientado hacia el poder*.

Sea S una red social. Consideremos una secuencia $S_1, S_2, \dots, S_n=S$ de manera que S_{t+1} se obtiene de S_t añadiendo el nodo $t+1$ de manera que S tiene crecimiento pop-orientado (con $p=1$). Sea $A=A(S)=(a_{xy})$ la matriz de la red con radio espectral r . Entonces las siguientes afirmaciones son válidas:

(a) Para $1 \leq t \leq d + 1$, la red S_t es una gráfica completa K_t .

(b) La red S puede seguir creciendo, esto es, existe otra red S^1 obtenida de S añadiendo un nodo $n+1$ de forma que S^1 tiene crecimiento pop-orientado.

(c) Para cada $d+1 \leq t \leq n$, el promedio c_t del valor de $c_t(x)$ con $1 \leq x \leq t$, satisface

$$d \leq c_t \leq r < D.$$

(d) Los clanes más grandes en S son de la forma K_{d+1} .

(e) Sea S^1 una red obtenida de S añadiendo un nodo de forma que S^1 tiene crecimiento orientado hacia la popularidad y S^2 otra red obtenida de S añadiendo un nodo de forma que S^2 tiene crecimiento orientado hacia el poder. Sea r_k el radio espectral de la matriz de la red $A_k=A(S^k)$, $k=1,2$. Entonces $r_1 \leq r_2$.

Demostración (e): El radio espectral r_1 de $A_1=(a'_{xy})$ es el máximo del cociente de Rayleigh

$$R(A_1) = R_1 = \frac{\sum_{i=1}^{n+1} \sum_{j=1}^{n+1} a'_{ij} x_i x_j}{\sum_{i=1}^{n+1} x_i^2}$$

tomando x_i números reales, no todos cero. Similarmente, el radio espectral r_2 de $A_2=(a''_{xy})$ es el máximo del correspondiente cociente de Rayleigh $R_2=R(A_2)$. Será suficiente mostrar que para las formas $R_1(x)$ y $R_2(x)$ se tiene

$$\max_x R_1(x) \leq \max_x R_2(x),$$

o bien, que las formas cuadráticas $\sum_{i=1}^{n+1} \sum_{j=1}^{n+1} a'_{ij} x_i x_j$ y $\sum_{i=1}^{n+1} \sum_{j=1}^{n+1} a''_{ij} x_i x_j$ satisfacen

$$\max_{||x||=1} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} a'_{ij} x_i x_j \leq \max_{||x||=1} \sum_{i=1}^{n+1} \sum_{j=1}^{n+1} a''_{ij} x_i x_j.$$

Como ambas matrices A_1 y A_2 son extensiones de A , basta mostrar que

$$\max_{||x||=1} \sum_{i=1}^{n+1} a'_{i,n+1} x_i \leq \max_{||x||=1} \sum_{i=1}^{n+1} a''_{i,n+1} x_i.$$

Para $c(i)$, el número de vecinos de i en S , $y_k = a'_{k,n+1}$ se tiene,

$$\begin{aligned} \max_{||x||=1} \sum_{i=1}^{n+1} a'_{i,n+1} x_i &\leq (1) \frac{\sum_{i=1}^n a'_{i,n+1} c(i)}{\sum_{i=1}^n c(i)^2} = \frac{\sum_{i=1}^n \sum_{k=1}^n a_{ik} y_k}{\sum_{i=1}^n c(i)^2} \leq (2) r \frac{\sum_{i=1}^n y_k}{\sum_{i=1}^n c(i)^2} = \\ &= d \frac{\sum_{i=1}^n a_{ik} u_k}{\sum_{i=1}^n c(i)^2} \leq (3) \frac{\sum_{i=1}^n \sum_{k=1}^n a_{ik} a''_{i,n+1}}{\sum_{i=1}^n c(i)^2} = \frac{\sum_{i=1}^n a''_{i,n+1} c(i)}{\sum_{i=1}^n c(i)^2} \leq \max_{||x||=1} \sum_{i=1}^{n+1} a''_{i,n+1} x_i \end{aligned}$$

donde la primera desigualdad (1) se debe a la definición de A_1 permitiendo la variación del vector $(a'_{k,n+1})$, la desigualdad (2) se debe a que $r=R(A)$, la desigualdad (3) se debe a la definición de u y $\sum_{i=1}^n a'_{i,n+1} = d = \sum_{i=1}^n a''_{i,n+1}$ QED

El resultado obtenido pone de manifiesto que el crecimiento orientado hacia el poder genera individuos más poderosos, lo que en principio es deseable para los nuevos individuos. Sin embargo, en una red real, en general, es más difícil leer el poder relativo de los individuos de la red que su popularidad (que se obtiene

sencillamente del número de vecinos). En conclusión, aunque como método de crecimiento es más *efectivo* el orientado al poder, es más difícil llevarlo a cabo, por lo que en la práctica es más *eficiente* orientarse por la popularidad.

Referencias bibliográficas

- Barábasi, A.L. and Albert, R.(1999) *Emergence of scaling in random networks. Science*, 286:509-512, October 15.
- Barábasi, A.L (2009) *Scale free networks: a decade and beyond. Science* 325, 412-413.
- Cvetkovic, D., Doob, M. and Sachs, H. (1980) *Spectra of Graphs -- Theory and applications*. Academic Press.
- Gantmacher, F.R. (1974) *The theory of matrices*.Vol II. Chelsea, New York.
- Dunbar, R.I.M. (1992) *Neocortex size as a constraint on group size in primates, Journal of Human Evolution*, vol. 20, pp. 469-493.
- Jin, E., Girvan,M. and Newman, M.E.J.(2001). *Structure of growing social networks. Phys. Rev. E* 64.
- Newman,M.E.J. (2003) *The structure and function of complex networks. ArXiv* 03035.
- De la Peña, J.A. (1996), *Algebra Lineal Avanzada*. Fondo de Cultura Económica.
- De la Peña, J.A. (2010) *A remark on Turán's theorem*. To appear.
- Watts, D. J. & Strogatz, S. H. (1998). *Collective dynamics of 'small-world' networks. Nature* 393, 440-442.