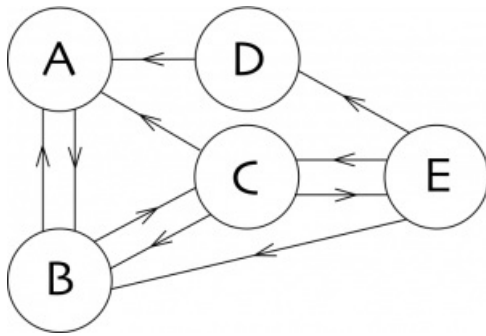


# Cómo funciona Google: cadenas de Markov y valores propios

2012-12-11 22:12:45 Antoine Nectoux



Versión original escrita por Christiane Rousseau.

Desde sus comienzos, Google se convirtió en “el” motor de búsqueda. Esto es debido a la supremacía de su algoritmo de jerarquización: el *algoritmo PageRank*. De hecho, debido a la enorme cantidad de páginas web en la World Wide Web, muchas búsquedas finalizan con miles o millones de resultados. Si estas páginas no estuvieran adecuadamente ordenadas, la búsqueda no sería de ninguna utilidad, ya que nadie es capaz de explorar millones de entradas.

¿Cómo funciona el algoritmo PageRank?

Esto se explicará más adelante, pero antes vamos a hacer una búsqueda en Google. En junio de 2010 se obtuvieron 16.300.000 resultados para *Klein project*, si bien el proyecto estaba comenzando. En esta fecha en concreto, la primera entrada era

<http://www.mathunion.org/icmi/other-activities/klein-project/introduction/>

en lugar de

<http://www.kleinproject.org/>

La primera url es la dirección web de una *página* que está localizada en el sitio web de la Unión Matemática Internacional (International Mathematical Union): <http://www.mathunion.org>. Como este es un organismo importante, su página web oficial aparece la primera cuando se realiza la búsqueda “International Mathematical Union”. Es más, transmite parte de su importancia a todas sus páginas, una de las cuales es

<http://www.mathunion.org/icmi/other-activities/klein-project/introduction/>

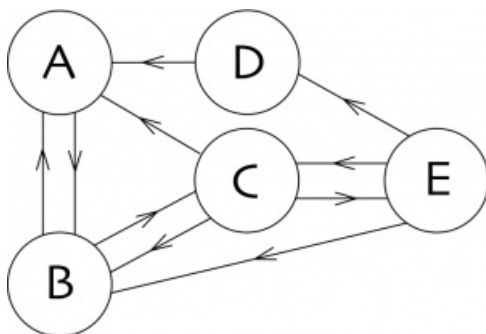
Cabría esperar que dentro de unos pocos meses la página

<http://www.kleinproject.org/>

apareciera la primera en la búsqueda de *Klein project*.

Para explicar el algoritmo se modela la red mediante un grafo orientado. Los vértices son las *páginas* y las aristas orientadas son los *enlaces* entre páginas. Como ya hemos explicado, cada página corresponde a una url diferente. Por tanto, un sitio web puede contener muchas páginas pero este modelo no diferencia entre las páginas individuales de un sitio web y su página principal. Sin embargo, es más probable que el algoritmo dé más valor a la página principal de un sitio web importante.

## Un ejemplo sencillo



Consideremos el ejemplo de la sencilla red de la izquierda, compuesta por cinco páginas llamadas *A*, *B*, *C*, *D* y *E*. Esta red tiene pocos enlaces: en *A* hay solamente un enlace a *B*, mientras que si estamos en *C* encontramos tres enlaces y podemos elegir entre pasar a *A*, *B* o *E*. Notar que hay al menos un enlace desde cada página.

Proponemos un sencillo juego, que consiste en dar un *paseo aleatorio* por el grafo orientado. Comenzando en una página cualquiera, en cada paso elegimos un enlace al azar en la página en la que nos encontramos y lo seguimos. Por ejemplo, en nuestra red, si comenzamos en *B*, entonces podemos elegir entre ir a *A* o a *C* con probabilidad  $1/2$  para cada caso, mientras que si empezamos en *D*, entonces necesariamente tenemos que ir a *A* con probabilidad 1. Si repetimos el juego,

¿dónde estaremos después de  $n$  pasos?

Para automatizar el proceso, resumimos la información de la red en la siguiente matriz  $P$ , donde cada columna representa la página de salida y cada fila es la página de destino.

$$P = \begin{pmatrix} A & B & C & D & E \\ 0 & \frac{1}{2} & \frac{1}{3} & 1 & 0 \\ 1 & 0 & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & 0 & 0 & \frac{1}{3} \\ 0 & 0 & 0 & 0 & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} & 0 & 0 \end{pmatrix} \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix}$$

Notar que la suma de todos los valores de una misma columna de  $P$  es igual a 1 y que todas las entradas de la matriz son mayores o iguales que cero. Una matriz que satisface estas dos propiedades es de un tipo muy especial: es la matriz de un *proceso de cadena de Markov*, también conocida como *matriz de transición del proceso de Markov*. Este tipo de matrices tienen siempre como valor propio 1 y existe un vector propio de valor propio 1 cuyas componentes son todas menores o iguales que 1 y mayores o iguales que 0 y cuya suma es 1. Antes de definir valores y vectores propios, vamos a explorar las ventajas de la representación matricial del grafo de la red.

Consideramos una variable aleatoria  $X_n$  con valores en el conjunto de páginas  $\{A, B, C, D, E\}$ ; este conjunto contiene  $N = 5$  páginas.  $X_n$  representa la página en la que nos encontramos después de  $n$  pasos del camino aleatorio. Por tanto, si  $p_{ij}$  es el elemento de la matriz ubicado en la  $i$ -ésima fila y  $j$ -ésima columna, entonces  $p_{ij}$  representa la probabilidad condicionada de que nos encontremos en la  $i$ -ésima página en el paso  $n + 1$ , supuesto que estábamos en la  $j$ -ésima página en el paso  $n$ :

$$p_{ij} = \text{Prob}(X_{n+1} = i \mid X_n = j).$$

¡Notar que esta probabilidad no depende de  $n$ ! Por ello, se dice que un proceso de cadena de Markov *no tiene memoria*. No es difícil de probar que las probabilidades después de dos pasos pueden resumirse en la matriz  $P^2$ . Probémoslo (el lector puede omitir la demostración, si lo desea). Del teorema de la probabilidad total se deduce que

$$\text{Prob}(X_{n+2} = i \mid X_n = j) = \sum_{k=1}^N \text{Prob}(X_{n+2} = i \text{ y } X_{n+1} = k \mid X_n = j).$$

Por la definición de probabilidad condicionada,

$$\text{Prob}(X_{n+2} = i \mid X_n = j) = \sum_{k=1}^N \frac{\text{Prob}(X_{n+2} = i \text{ y } X_{n+1} = k \text{ y } X_n = j)}{\text{Prob}(X_n = j)}.$$

Usamos ahora un truco muy conocido: multiplicar y dividir por la misma cantidad:

$$\begin{aligned} \text{Prob}(X_{n+2} = i \mid X_n = j) &= \\ \sum_{k=1}^N \frac{\text{Prob}(X_{n+2} = i \text{ y } X_{n+1} = k \text{ y } X_n = j)}{\text{Prob}(X_{n+1} = k \text{ y } X_n = j)} \frac{\text{Prob}(X_{n+1} = k \text{ y } X_n = j)}{\text{Prob}(X_n = j)}. \end{aligned}$$

El primer cociente es igual a

$$\text{Prob}(X_{n+2} = i \mid X_{n+1} = k \text{ y } X_n = j) = \text{Prob}(X_{n+2} = i \mid X_{n+1} = k),$$

ya que los procesos de cadena de Markov no tienen memoria del paso anterior. Por tanto,

$$\begin{aligned} \text{Prob}(X_{n+2} = i \mid X_n = j) &= \sum_{k=1}^N \text{Prob}(X_{n+2} = i \mid X_{n+1} = k) \text{Prob}(X_{n+1} = k \mid X_n = j) \\ &= \sum_{k=1}^N p_{ik} p_{kj} \\ &= (P^2)_{ij}. \end{aligned}$$

En nuestro ejemplo,

$$P^2 = \begin{pmatrix} A & B & C & D & E \\ \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & 0 & \frac{11}{18} \\ 0 & \frac{2}{3} & \frac{4}{9} & 1 & \frac{1}{9} \\ \frac{1}{2} & 0 & \frac{5}{18} & 0 & \frac{1}{6} \\ 0 & 0 & \frac{1}{9} & 0 & 0 \\ 0 & \frac{1}{6} & 0 & 0 & \frac{1}{9} \end{pmatrix} \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix}$$

Iterando esta idea, es evidente que el elemento  $(P^m)_{ij}$  de la matriz  $P^m$  representa la probabilidad  $\text{Prob}(X_{n+m} = i \mid X_n = j)$ . Por ejemplo,

$$P^{32} = \begin{pmatrix} A & B & C & D & E \\ 0.293 & 0.293 & 0.293 & 0.293 & 0.293 \\ 0.390 & 0.390 & 0.390 & 0.390 & 0.390 \\ 0.220 & 0.220 & 0.220 & 0.220 & 0.220 \\ 0.024 & 0.024 & 0.024 & 0.024 & 0.024 \\ 0.073 & 0.073 & 0.073 & 0.073 & 0.073 \end{pmatrix} \begin{matrix} A \\ B \\ C \\ D \\ E \end{matrix}$$

Notar que todas las columnas de  $P^{32}$  son idénticas si trabajamos con 3 decimales de precisión y, además, son iguales a las columnas de  $P^n$  cuando  $n > 32$ . Si elegimos mayor precisión, también encontraremos esta estabilidad, pero para un exponente  $n$  mayor que 32. Por tanto, después de  $n$  pasos, para  $n$  suficientemente grande, ¡la probabilidad de estar en una determinada página no depende de la página en la que comenzamos!

Más aún, si consideramos el vector

$$\pi^t = (0.293, 0.390, 0.220, 0.024, 0.073)$$

(aquí  $\pi$  es un vector vertical, luego su traspuesto  $\pi^t$  es un vector horizontal, se comprueba fácilmente que  $P\pi = \pi$ . Si la  $i$ -ésima coordenada de del vector  $\pi^t$  representa la probabilidad de estar en la página  $i$  en un determinado momento  $n$  y, por tanto,  $\pi^t$  representa la distribución de probabilidad de las páginas en ese momento  $n$ , entonces  $\pi^t$  también representa la distribución de probabilidad en el momento  $n + 1$ . Por esta razón, el vector  $\pi^t$  se dice *distribución estacionaria*. Esta distribución estacionaria permite la ordenación de las páginas. En nuestro ejemplo, ordenamos las páginas de la siguiente manera:  $B, A, C, E, D$ , declarando que  $B$  es la página más importante.

### El caso general

El caso general puede ser abordado de la misma manera que nuestro ejemplo. Representamos la red como un grafo orientado en el cual los  $N$  vértices representan las  $N$  páginas de la red y las aristas orientadas representan enlaces entre páginas. Trasladamos la información del grafo a una matriz  $P$  de tamaño  $N \times N$ , donde la  $j$ -ésima columna representa la  $j$ -ésima página de partida y la  $i$ -ésima fila representa la  $i$ -ésima página de destino. En nuestro ejemplo encontramos un vector  $\pi$  que cumple que  $P\pi = \pi$ . Este vector es, por tanto, un vector propio de valor propio 1. Recordamos en este punto las definiciones de vector propio y valor propio.

**Definición.** Sea  $P$  una matriz de tamaño  $N \times N$ . Un número  $\lambda \in \mathbb{C}$  se dice *valor propio* de  $P$  si existe un vector no nulo  $X \in \mathbb{C}^N$  tal que  $PX = \lambda X$ . Tal vector  $X$  se dice *vector propio* de  $P$ .

Recordamos también el método para encontrar valores y vectores propios.

**Proposición** Sea  $P$  una matriz de tamaño  $N \times N$ . Los valores propios de  $P$  son las raíces polinomio característico  $\det(\lambda I - P) = 0$ , donde  $I$  es la matriz identidad de tamaño  $N \times N$ . Los vectores propios asociados al valor propio  $\lambda$  son las soluciones no nulas del sistema lineal homogéneo  $(\lambda I - P)X = 0$ .

El siguiente teorema debido a Frobenius garantiza que la matriz asociada al grafo de una red siempre tiene una solución estacionaria.

**Teorema** (Teorema de Frobenius) Sea  $P = (p_{ij})$  la matriz de transición de un proceso de cadena de Markov de tamaño  $N \times N$  Markov (esto es,  $p_{ij} \in [0, 1]$  para cualesquiera  $i, j$  y la suma de las componentes de cada columna es igual a 1:  $\sum_{i=1}^N p_{ij} = 1$ ). Se tiene que

1.  $\lambda = 1$  es valor propio de  $P$ .
2. Cualquier valor propio  $\lambda$  de  $P$  verifica  $|\lambda| \leq 1$ .
3. Existe un vector propio  $\pi$  de valor propio 1, cuyas coordenadas son todas mayores o iguales que cero. Sin pérdida de generalidad, podemos asumir que la suma de sus coordenadas es 1.

Ahora es el momento de admirar la potencia de este teorema. Para ello, consideramos la siguiente hipótesis simplificada:

suponemos que la matriz  $P$  tiene una base formada por vectores propios  $\mathcal{B} = \{v_1, \dots, v_n\}$  y que  $v_1$  es el vector  $\pi$  del teorema de Frobenius. Se tiene que para cada  $v_i$  existe  $\lambda_i$  tal que  $Pv_i = \lambda_i v_i$ . Sean  $X$  un vector no nulo y  $X^t = (x_1, \dots, x_N)$ , donde  $x_i \in [0, 1]$  y  $\sum_{i=1}^N x_i = 1$ . Descomponemos  $X$  con respecto a la base  $\mathcal{B}$ :

$$X = \sum_{i=1}^N a_i v_i.$$

Una demostración técnica que omitimos aquí prueba que  $a_1 = 1$ . Ahora calculamos  $PX$ :

$$PX = \sum_{i=1}^N a_i P v_i = \sum_{i=1}^n a_i \lambda_i v_i,$$

ya que  $v_i$  es un vector propio de valor propio  $\lambda_i$ . Iterando, se obtiene que

$$P^n X = \sum_{i=1}^n a_i \lambda_i^n v_i.$$

Si todos los valores  $\lambda_i$  para  $i > 1$  verifican  $|\lambda| < 1$ , entonces

$$\lim_{n \rightarrow \infty} P^n X = a_1 v_1 = \pi.$$

¡Esto es exactamente lo que sucede en nuestro ejemplo!

Sin embargo, conviene notar que el teorema no garantiza que cualquier matriz  $P$  que verifique las hipótesis tiene esta propiedad. Describamos entonces los posibles casos patológicos y sus remedios.

#### Posibles casos patológicos.

- El valor propio 1 puede ser una raíz múltiple del polinomio característico  $\det(\lambda I - A) = 0$ .
- La matriz  $P$  puede tener otros valores propios  $\lambda$  distintos 1 de módulo 1.

¿Qué podemos hacer en estos casos?

Decimos que una matriz de transición de Markov que no tiene patologías es una *matriz de transición de Markov regular*, esto es:

**Definición.** Una matriz de transición de Markov se dice *regular* si

- El valor propio 1 es una raíz simple del polinomio característico  $\det(\lambda I - A) = 0$ .
- Todos los valores propios  $\lambda$  de  $P$  distintos de 1 tienen módulo estrictamente menor que 1.

Notar que la mayoría de las matrices  $P$  son regulares. Por tanto, si la matriz de transición de Markov de una red no es regular, la estrategia a seguir es deformarla ligeramente hasta conseguir una matriz regular.

#### Remedio.

Consideramos la matriz  $Q = (q_{ij})$  de tamaño  $N \times N$  con  $q_{ij} = \frac{1}{N}$  para cualesquiera  $i, j$ . Reemplazamos la matriz  $P$  asociada a la red por la matriz

$$(1) \quad P_\beta = (1 - \beta)P + \beta Q,$$

para un valor pequeño de  $\beta \in [0, 1]$  (el valor  $\beta = 0.15$  es el usado por Google). Notar que todos los elementos de la matriz  $P_\beta$  siguen siendo no negativos y que la suma de los elementos de cada columna es 1, luego es una matriz de transición de Markov. El siguiente teorema garantiza que existe tal valor  $\beta$  que remedia esta patología.

**Teorema** Dada una matriz de transición de Markov  $P$  existe  $\beta$  positivo, tan pequeño como se quiera, tal que la matriz  $P_\beta$  es regular. Sea  $\pi$  el vector propio de valor propio 1 para la matriz  $P_\beta$ , normalizado de manera que la suma de sus coordenadas es 1. Dado un vector no nulo  $X$  con  $X^t = (p_1, \dots, p_N)$ , donde  $p_i \in [0, 1]$  y  $\sum_{i=1}^N p_i = 1$ , se tiene que

$$(2) \quad \lim_{n \rightarrow \infty} (P_\beta)^n X = \pi.$$

#### Relación con el teorema del punto fijo de Banach

En otra viñeta se trata el teorema del punto fijo de Banach, del que el teorema anterior puede ser visto como una aplicación particular. De hecho, se tiene que

#### Teorema

Sea  $P$  una matriz de transición de Markov regular. Consideramos

$\mathcal{S} = \{X \mid X^t = (p_1, \dots, p_N), p_i \in [0, 1], \sum_{i=1}^N p_i = 1\}$ ,

con una cierta distancia  $d(X, Y)$  entre puntos (que depende de  $P$ ). Sea el operador lineal  $L : \mathcal{S} \rightarrow \mathcal{S}$  definido en  $\mathcal{S}$  por  $L(X) = PX$ .

Se tiene que el operador  $L$  es una contracción en  $\mathcal{S}$ , esto es, existe  $c \in [0, 1[$  tal que para cualesquiera  $X, Y \in \mathcal{S}$  se cumple que

$$d(L(X), L(Y)) \leq c d(X, Y).$$

Además, existe un único vector  $\pi \in \mathcal{S}$  tal que  $L(\pi) = \pi$  y dado cualquier  $X_0 \in \mathcal{S}$ , se puede definir por inducción la secuencia  $\{X_n\}$ , donde  $X_{n+1} = L(X_n)$ , y se cumple que  $\lim_{n \rightarrow \infty} X_n = \pi$ .

**Definición of la distancia  $d$ .**

La definición de la distancia  $d$  es un tanto sutil y puede ser omitida. Se incluye aquí por completitud, para el lector que desee insistir en los detalles. Nos limitamos al caso en que la matriz  $P$  es diagonalizable. Sea  $\mathcal{B} = \{v_1 = \pi, v_2, \dots, v_N\}$  una base de vectores propios. Cualesquiera vectores  $X, Y \in \mathcal{S}$  pueden expresarse en la base  $\mathcal{B}$ :

$$X = \sum_{i=1}^N a_i v_i, \quad Y = \sum_{i=1}^N b_i v_i,$$

donde  $a_i, b_i \in \mathbb{C}$ . Definimos entonces

$$d(X, Y) = \sum_{i=1}^n |a_i - b_i|.$$

Con esta definición de distancia  $\mathcal{S}$  es un espacio métrico completo, es decir, toda sucesión de Cauchy es convergente.

Este teorema no solo garantiza la existencia de  $\pi$ , sino que da un método para construirlo como límite de la secuencia  $\{X_n\}$ .

Nuestro sencillo ejemplo ilustra esta convergencia. De hecho, la  $j$ -ésima columna de  $P^n$  es el vector  $P^n e_j$ , donde  $e_j^t = (0, \dots, 0, 1, 0, \dots, 0)$  es el  $j$ -ésimo vector de la base canónica. Por supuesto, en nuestro ejemplo podríamos haber hallado directamente el vector  $\pi$  resolviendo el sistema  $(I - P)X = 0$  que tiene como matriz de coeficientes

$$I - P = \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{3} & -1 & 0 \\ -1 & 1 & -\frac{1}{3} & 0 & -\frac{1}{3} \\ 0 & -\frac{1}{2} & 1 & 0 & -\frac{1}{3} \\ 0 & 0 & 0 & 1 & -\frac{1}{3} \\ 0 & 0 & -\frac{1}{3} & 0 & 1 \end{pmatrix}.$$

Las soluciones de este sistema son todas de la forma  $(4s, \frac{16}{3}s, 3s, \frac{1}{3}s, s)^t$  para  $s \in \mathbb{R}$ . Por tanto, la solución cuya suma de coordenadas es igual a 1 es  $\pi$ , donde

$$\pi^t = \left( \frac{12}{41}, \frac{16}{41}, \frac{9}{41}, \frac{1}{41}, \frac{3}{41} \right).$$

## Cálculo en la práctica de la distribución estacionaria

Hemos identificado la sencilla idea que subyace en el algoritmo. Sin embargo, encontrar la distribución estacionaria  $\pi$  (esto es, un vector propio de valor propio 1 para la matriz  $P_\beta$  en (1)) no es una tarea sencilla cuando la matriz tiene miles de millones de filas y columnas: el tiempo de computación y la memoria física que se requiere para ello representan verdaderos desafíos. El habitual método de eliminación de Gauss no es útil en este caso por el tamaño de los cálculos y porque se requiere dividir por coeficientes pequeños. Un algoritmo más eficiente hace uso de la propiedad (2) de estas matrices (ver [LM]). Este es el vínculo con la viñeta acerca del teorema del punto fijo de Banach, en la que se verá que la demostración del dicho teorema proporciona un algoritmo para construir el punto fijo.

De hecho, se parte de un  $X_0$  tal que

$$X_0^t = \left( \frac{1}{N}, \frac{1}{N}, \dots, \frac{1}{N} \right),$$

y se necesita calcular  $\lim_{n \rightarrow \infty} (P_\beta)^n X_0$ . Habitualmente  $P_\beta^n X_0$  para  $n$  entre 50 y 100 da una aproximación de  $\pi$  bastante buena. Por inducción, se obtiene  $X_{n+1} = P_\beta X_n$ , pero incluso estos cálculos resultan largos. De hecho, debido a su construcción, la matriz  $P_\beta$  en (1) no tiene elementos no nulos, mientras que la mayoría de los elementos de la matriz  $P$  son cero, algo que debemos aprovechar:

$$X_{n+1} = (1 - \beta)PX_n + \beta QX_n.$$

Debido a la especial forma de  $Q$ , es fácil comprobar que si  $X$  es un vector cuyas coordenadas suman 1 entonces  $QX = X_0$ . Por tanto, es suficiente con calcular la secuencia

$$X_{n+1} = (1 - \beta)PX_n + \beta X_0.$$

## Conclusión

En esta viñeta se ha presentado la parte pública del algoritmo PageRank de Google. Ahora el lector puede experimentar con redes sencillas y encontrar trucos para mejorar la posición de su página personal añadiendo enlaces tanto internos como externos de manera adecuada. Algunas de las partes privadas más sofisticadas continúan aún en desarrollo. Algunas de ellas consisten en remplazar la matriz “neutra”  $Q$  en (1) por otras matrices que reflejen los *gustos* del navegante en la red. Otras se aseguran de que el orden de las páginas no sea demasiado sensible a las manipulaciones de aquellos que intentan mejorar la posición de sus páginas.

Como conclusión general, ¿qué hemos observado? Que una idea simple e ingeniosa ha dado lugar a un enorme progreso en la eficiencia de los motores de búsqueda y al nacimiento de un verdadero imperio comercial.

Si bien la implementación es en sí una hazaña computacional, la idea inicial requiere matemáticas “elementales”: álgebra lineal y teoría de probabilidades. Las herramientas utilizadas, especialmente la diagonalización de matrices, son estándar en matemáticas pero han demostrado su verdadera potencia cuando se han usado fuera de su contexto “habitual”. También hemos incidido en las ideas unificadoras dentro de la ciencia cuando hemos considerado el teorema del punto fijo de Banach y sus aplicaciones a situaciones tan alejadas de su origen.

## Bibliografía

[E] M. Eisermann, *Comment Google classe les pages webb*, <http://images.math.cnrs.fr/Comment-Google-classe-les-pages.html>, 2009.

[LM] A. N. Langville and C. D. Meyer, A Survey of Eigenvector Methods for Web Information Retrieval, *SIAM Review*, Volume 47, Issue 1, (2005), pp. 135–161.

[RS] C. Rousseau and Y. Saint-Aubin, *Mathematics and technology*, SUMAT Series, Springer-Verlag, 2008 (A French version of the book exists, published in the same series.)

## Share this:

- [Email](#)
- [Print](#)
- [Facebook](#)
- [Twitter](#)
- 

Este mensaje está disponible en: [Inglés](#), [Francés](#), [Alemán](#), [Italiano](#), [Árabe](#), [Khmer](#)