

Internet: La imprenta del Siglo XXI

Por
Alejandro Piscitelli.



Editorial Gedisa.

Barcelona.

**Primera edición:
2005.**

**Este material es de
uso exclusivamente
didáctico.**

<i>Agradecimientos</i>	9
------------------------------	---

<i>Introducción</i>	13
---------------------------	----

PARTE I ECOLOGÍA DE LA RED

1. La dinámica de la web.....	21
2. Las «power laws».....	33

PARTE II ESCRIBIR, LEER y ESTUDIAR EN RED

3. Los weblogs ponen la red al servicio de todas las voces.....	51
4. Sindicación de contenidos.....	67
5. Distribución de contenidos y escritura colaborativa.....	81
6. Computadoras y educación.....	93
7. El diseño empático en el <i>e-learning</i>	109

PARTE III PENSAR EN RED

8. La mirada de Marshal McLuhan.....	119
9. Contra el determinismo tecnológico.....	133

PARTE IV DISEÑAR EN RED

10. Internet como sistema emergente.....	147
11. El futuro de la red: la web semántica.....	157

<i>Conclusiones</i>	171
---------------------------	-----

<i>Bibliografía</i>	179
---------------------------	-----

1

La dinámica de la web

¿Por qué ocuparnos de Internet? Lo cierto es que estos 10 años de existencia de la web finalmente nos permiten empezar a concebir la pantalla (y avizorar así su futuro) como un nuevo soporte intelectual. Un soporte intelectual en el sentido en que lo fueron en su momento la inscripción en piedra, el papiro, el papel, y recientemente los soportes electromagnéticos. Lo que sucedió en Internet -que fue en un momento el espacio privilegiado del texto y del blanco y negro, pero terminó estallando en una paleta de nuevas formas expresivas y de nuevas formas de enseñar, aprender, negociar, vender y entretener-, es similar a lo que sucedió con todas las tecnologías de los medios.

Operan en esta transformación dos principios básicos: el primero de ellos es que cada nuevo avance tecnológico implica ganancias pero también pérdidas. El segundo es que los nuevos medios generalmente recuperan elementos de la comunicación natural (biológica) que habían sido eclipsados por los medios anteriores (Grusin y Bolter, 2000). Sin embargo, es imposible entender estas paradojas si antes no se entiende cómo se producen. Para ello hay que investigar la infraestructura que hace posible el nuevo medio y sobre todo las reglas que regulan su emergencia, crecimiento, expansión y decadencia.¹ A ello nos dedicaremos en este libro.

En el caso específico de Internet, sus aspectos más interesantes han sido sistemáticamente oscurecidos por las narrativas que han dominado el campo de su interpretación, que han oscilado entre el tecnicismo a ultranza, el discurso ingenieril y más recientemente -en la última década- por los mitos filoeconómicos que la transformaron en el Rey Midas o en la gallina de los huevos de oro. Con la mínima distancia que nos permite el haberla recorrido en múltiples direcciones (Piscitelli, 2001 y 2002), y con un creciente escepticismo acerca de su capacidad de anulación de lecturas alternativas, un nuevo recorrido por sus aspectos sociales y tecnológicos ayuda a poner las cosas en una perspectiva más moderada y útil.²

Si bien a lo largo de este libro se profundizará en estos aspectos, es interesante comprender aquí que en Internet se produce la emergencia de complejidad: la interacción de elementos simples desprovistos de toda inteligencia acaba por generar un cerebro y una piel globales,

¹ Las últimas décadas han sido fructíferas, al menos en lo que hace a la epistemología, al renunciar a ciertos determinismos tecnológicos y reduccionismos de distinto orden. Así, hemos comenzado por fin a entender con cierta elegancia, economía de conceptos y precisión en qué debe consistir una ecología de los medios. En este proceso hemos debido renunciar a cualquier intento totalizador que trate de encapsular la complejidad de la comunicación humana bajo un solo enfoque o precepto. Como modelo de lo que no necesitamos hacer está, por ejemplo, la desahuciada semántica general de Alfred Korzybski. Entre los varios puntos de partida que pueden llevarnos a mejorar nuestra comprensión de esta ecología están las entradas que nos abren los siguientes enfoques: la enunciación de la problemática y algunos estudios de campo de Gregory Bateson (1993); las leyes de la forma de Marshall McLuhan (1988); la mediamorfosis de Roger Fiedler (1998); la mediología de Régis Oebray (1993); la remediación de Richard Grusin y Jay Bolter (2000); la simplexología de Frank Zingrone (2001); así como muchos otros intentos más o menos felices de entender esta endiablada problemática.

² Si bien falta a nuestro juicio un buen análisis *à la* Latour, existen ya varios análisis complejos de la red que merecen nuestro interés. No nos referimos aquí a las monografías que han entendido los casos exitosos de exploración/explotación de la red con fines comerciales (como *The Perfect Store. Inside eBay* de Adam Cohen o *Amazon.com: Get Big Fa.net* de Robert Spector), ni mucho menos a las lecturas más o menos serias acerca de la evolución de la web y de su naturaleza esquizofrénica de paraíso-infierno (como *Dot.com: How America Lost Its Mind and Money in the Internet Era* de John Cassidy o *Burn Rate: How I Survived the Gold Rush Years on the Internet* de Michael Wolff), sino a los enfoques más sazonados -generalmente provenientes del campo europeo- como los de Hervé Fischer (*El choque digital*) o de Pierre Levy (*CiberCultura*), incluyendo obviamente *Escape Velocity: cyber Culture at the End of the Century*, de Mark Dery.

inesperados y superpoderosos. Se trata, por otro lado, de una red libre de escala, es decir, una estructura en la cual no se verifica una distribución normal (gaussiana) del crecimiento. Por el contrario, la red se encuentra regida por leyes exponenciales que hacen que cuatro sitios mayoritarios concentren el 50% de las visitas mientras que la mayoría no convoca prácticamente a nadie (Barabási et al., 2001). Esta característica, llamada de diferentes formas y revisitada varias veces a lo largo de este libro, constituye uno de sus aspectos actuales más significativos. Por ello, antes de adentrarnos en la red propiamente dicha, habrá que examinar una familia de palabras entre las que se incluyen: mundos pequeños, redología, redes libres de escala, regla de Pareto, reglas de los medios, etc., que forman su contexto y hacen posible su comprensión. En los capítulos que siguen se abordarán estos temas.

Ahora bien, ¿cómo empezó todo? Cada vez que se habla de los orígenes de Internet, un personaje mítico reaparece en el relato. Se trata nada más y nada menos que de Paul Baran, quien con apenas 30 años aceptó un trabajo nuevo en la Rand Corporation, un *think tank* creado por el gobierno de Estados Unidos en 1946 (Abbate, 1999). Corría el lejísimo año 1959 cuando a Baran le ofrecieron una tarea por demás ambiciosa: diseñar un sistema de comunicaciones capaz de resistir un presunto ataque nuclear soviético. En efecto, en esa época, la Rand Corporation estaba asociada a las elucubraciones bélicas de personajes como Herman Kahn, una de cuyas debilidades fue dedicarse a los juegos de guerra y calcular, por ejemplo, la muerte eventual de millones de personas como resultado de un ataque nuclear. A Baran, tales asuntos no le preocupaban demasiado, de modo que miró hacia otro lado y se dispuso a crear un sistema de comunicaciones totalmente distinto al existente, algo que plasmó publicando más tarde una serie de 12 volúmenes que harían historia, aunque con unos cuantos zigzags en el medio.

Para él, la vulnerabilidad de la red de comunicaciones existente derivaba obviamente del sistema de control y comando militar sobre la que había estado basada durante casi un siglo. Al analizar los sistemas dominantes en la época, encontró tres modelos de organización de redes a los que denominó *centralizado*, *descentralizado* y *distribuido*. Naturalmente, descartó de un plumazo las dos primeras configuraciones, dado el altísimo grado de vulnerabilidad que exhibirían ante un ataque nuclear. Se trataba de hallar un sistema que pudiera preservar la red entera si se destruían nodos por un impacto.

Para Baran, una estructura invulnerable a esos presuntos ataques habría de ser una red distribuida en forma de retícula (muy parecida aun sistema de autopistas) con tan alto grado de redundancia que, aún cuando muchos de los nodos fueran borrados del mapa por un ataque, caminos alternativos mantendrían la conexión entre el resto de los nodos sobrevivientes.

Este es un primer punto interesante: los militares le pidieron a Baran que diseñara la red alternativa: la fantasía o mito acerca de un origen militar de la red tiene un punto a su favor. Pero lo que esta narrativa olvida u oculta es que finalmente los militares no hicieron a Paul Baran el menor caso. Sin embargo, quienes finalmente la diseñaron llegaron a un idéntico fin a través de numerosos caminos alternativos.

La propuesta de Baran implicaba descomponer los mensajes en pequeños paquetes de información (*packet-switching*), capaces de viajar independientemente unos de los otros por la red. Para lograr ese (entonces) delirio, hacía falta un sistema de comunicación inexistente en ese momento, por lo cual Baran abogó por la creación de un sistema de comunicación digital. Sin embargo, se trataba de una demanda que difícilmente podía ser sostenida o cumplida por el monopolio telefónico AT&T, que reinó omnímodo en Estados Unidos desde la década de 1910 hasta su desmembramiento en 1982. Quien era entonces su presidente, Jack Ostermani, rechazó la propuesta de Baran, sosteniendo que era imposible de lograr y que, en el caso en que funcionara, eso implicaría crearle una competencia interna a AT&T.

Las ideas de Baran naufragaron, pues, ante la oposición combinada del pensamiento centralizado de los militares y las prácticas monopólicas -no menos centralizadas- de la industria. Solo serían valoradas casi una década más tarde, cuando la Defense Advanced Research Project Agency (Darpa) las desarrolló por su propia cuenta, (re)descubriendo imprevistamente que habían llegado 10 años más tarde al mismo lugar. Entonces, todo fue diferente en el escenario internacional y local. Lo que hasta ese entonces se concebía inviable se reveló ahora necesario.

En pleno corazón de la Guerra Fría, el durísimo golpe narcisista que supuso para los estadounidenses el lanzamiento de la perra Laica al espacio motivó que el presidente y general Dwight Eisenhower creara Darpa. A los pocos años -y sobre todo a partir de la creación posterior de la National Aeronautics and Space Administration (NASA)-, Darpa se convirtió en una usina estratégica de proyectos paramilitares y pasó a ser sencillamente ARPA. Así Internet entró en el radar de sus preocupaciones entre 1965 y 1966, cuando Bob Taylor -el director de los programas de computación de ARPA -se dio cuenta, en un día de epifanía y revelación, de algo trivial y catastrófico: las distintas redes de computadoras entonces existentes (máquinas que costaban entonces millones de dólares) eran incapaces de comunicarse entre sí. Peor aún, dos máquinas pertenecientes a distintas redes, cohabitantes en la misma sala, no podían *hablar* entre ellas.

Distinguido el problema, se trataba entonces de crear la solución. Esta consistiría en hallar algún protocolo capaz de interconectar máquinas hasta ese momento mudas, competentes exclusivamente en dialectos que nadie podía traducir. Con un millón de poderosos y sustanciales dólares de la época, ARPA se puso a trabajar en serio, al mismo tiempo que un grupo de investigadores ingleses comandados por Donald Davis, del Laboratorio Nacional de Física (NPL) de Inglaterra, redescubrían las principales ideas y enseñanzas de Baran por su cuenta, testimoniando el peso de los descubrimientos múltiples y simultáneos.

En un simposio que tuvo lugar en Texas en 1967, estas ideas empezaron a circular ampliamente y lo que había sido anatema una década atrás se convertiría en la ciencia normal de la época. El *packet-switching* que Baran había imaginado se convertía en el abc de la red, e Internet estaba lista para dar sus primeros pasos.

De todos modos, ahora es fundamental centrar nuestra atención en la topología actual -es decir, en la arquitectura espacial- de la red, así como tratar de desentrañar los principios de crecimiento y construcción -si es que son detectables- que dan cuenta de la forma actual de Internet y de su semejanza con otras redes, e intentar deducir a partir de todo ello su posible futuro. Necesitamos entender la topología si queremos diseñar herramientas y ofrecer servicios que corran eficazmente sobre ella. Una de sus características extraordinarias es que, aunque es un producto 100% humano, la red no ha sido diseñada de forma centralizada: desde el punto de vista estructural, se parece mucho más aun ecosistema que a un reloj suizo o a cualquier otro aparato que haya salido de un plano de un ingeniero y se haya concretado en el espacio físico.

Lo cierto es que existen en Internet innumerables mapas que buscan delinear la red o recorrer sus principales configuraciones (Dodge, y Kitchin, 2000,2001), pero no existe *el* mapa de la red. Si visitamos el Sitio www.peacockmaps.COM encontraremos unos mapas sugestivos realizados por Bill Cheswick y Hal Burch: se trata de unos llamativos entramados, densos en nodos y enlaces, que solo encuentran alguna semejanza en las tomografías computadas o las resonancias nucleares del cerebro. Sin embargo, hay una importante distinción: mientras que el cerebro hace milenios que se ha estabilizado anatómicamente, la red crece sin parar y multiplica diariamente su hiperconectividad y su densidad.

¿Por qué es tan importante esta cuestión de los mapas? En todo caso, resulta sumamente interesante que, dado que en general el camino va de la idea al objeto, y no a la inversa, en un proceso que comienza con detallados inventarios y topografías, meticulosos diagramas y visualizaciones, Internet esté recorriendo el camino inverso. De hecho, nadie sabe exactamente qué forma tiene la red. Aunque se la «patrulle» permanentemente -se la fotografíe, se la ausculte y se la mida- sus contornos son borrosos y sus centros son difícilmente localizables y discernibles. Esta dificultad está profundamente ligada al carácter anárquico, autoorganizado, fundamentalmente local y distribuido que subyace a la evolución de la red.

Con un mínimo de requisitos cualquier persona puede conectarse a la red. Si hubiera alguien que decidiera *cerrarla* lo único que lograría es aislar una porción minúscula de ella. Inmediatamente la información se «rerrutearía», es decir, establecería una nueva serie de vínculos y reenvíos y, a los efectos prácticos, la caída sería invisible o insignificante. Es precisamente la naturaleza distribuida, descentralizada y localmente aislada de la red la que la vuelve prácticamente inmune a cualquier ataque pero, al mismo tiempo, lo que hace tan difícil retratarla y aislarla.

Navegar en el mar de los sargazos

Insistimos: si estamos obsesionados por establecer uno o varios mapas de la red, ello obedece a que el diseño de servicios nuevos y más eficientes requiere obligadamente que tengamos un mucho mejor conocimiento de su topología. Por otra parte, cuando Internet se inventó, nadie imaginó que podría llegar a tener más de un millón de usuarios, o que sus usos trascenderían primero a los militares y luego a los investigadores. Nadie lo imaginó en 1970 ni en 1980, pero tampoco en el mucho más cercano 1990. A nadie se le ocurrió jamás que existiría una Internet comercial y mucho menos una red de comunicación universal (que inclusive llegaría a Marte y que en noviembre de 2004 tendría más de 900 millones de usuarios). Incluso el correo electrónico -best seller de aplicaciones en la red- emergió en forma inesperada de una pequeña adaptación para la transferencia de archivos que hizo Rega Tomlinson, uno de los propios inventores de la red, en ARPA.³

La web es uno de los mejores ejemplos que hoy pueden mostrarse de desastre exitoso. Es decir, de una idea que, fugitiva de la mesa del dibujante, es abrazada y utilizada por una cantidad impresionante de personas y -antes de que su diseño o funcionalidad estén estabilizados- se autoconfigura de pronto de un modo creativo y poderoso en su dinámica, pese a que resulta altamente ineficiente medido con parámetros ingenieriles y objetivos. Porque, de hecho, si sus creadores, Tim Berners-Lee y Robert Caillou, hubieran podido imaginarse algunas de las consecuencias de lo que pergeñaban al delinear los primeros esbozos de la web afines de la década de 1980, seguramente la experiencia de los usuarios sería hoy totalmente distinta y la forma en que usamos la web también sería diferente.

Lamentablemente, tal como existe hoy, la red difícilmente se adapta a nuestras necesidades. No es casual que la revolución de los weblogs (de lo que hablaremos en próximos capítulos) haya implicado un cambio brutal en su uso y haya implicado una suerte de reapropiación por parte de los usuarios comunes y silvestres. Si los weblogs son tan exitosos, ello se debe justamente a los fallos estructurales en la concepción y el diseño de la red.

Mientras muchos investigadores y programadores insisten en crear nuevas aplicaciones y servicios, otros hemos empezado a hacernos algunas preguntas clave: ¿Qué es exactamente lo que hemos inventado? ¿Qué tipo de criatura hemos liberado en el ciberespacio? ¿En qué puede (o podemos) convertirnos a partir de su mutación y evolución? ¿Hasta qué punto los hábitos de lectura y escritura, que sabemos constitutivos de nuestra identidad y subjetividad, se verán transformados y modificados por su emergencia?

Internet tiene una vida propia. A poco que la usamos y analizamos, exhibe todas las propiedades que son específicas de un sistema evolutivo complejo, lo que la vuelve mucho más parecida a una célula que aun chip. Es por ello que quienes usamos e investigamos la red hemos devenido, en tiempo récord, de exploradores en diseñadores. Todos los internetólogos nos asemejamos cada día más a los biólogos y a los ecólogos (de la información) -más que a los físicos y a los matemáticos-, es decir, lidiamos con un sistema increíblemente complejo que -a los efectos prácticos- existe con anterioridad a nosotros y es independiente de nosotros por completo. Pero sí, por un lado, estamos en mucho mejores condiciones que los biólogos para entender la red (porque, después de todo, nosotros la hemos diseñado y conocemos cada uno de los componentes que la conforman); por el otro, concordamos con ellos en no tener gran idea acerca de lo que pasa (qué tipo de estructura es la que emerge) cuando el sistema se pone en movimiento y se auto organiza.

Antes del año 2000 se sabía poco y nada de la topología de Internet, pero gracias al trabajo seminal de los hermanos Petros, Christos y Michalis Faloutsos, «On Power-Law Relationship of

³ En 1971, Ray Tomlinson inventó un programa de correo electrónico para mandar mensajes a través de una red distribuida. Envío el primer e-mail con un mensaje que decía «Testing 1-2-3» dirigido a sí mismo. El segundo mensaje de correo electrónico fue mucho más importante, se dirigió a todos los usuarios de Arpanet y consistió en las instrucciones y convenciones del recién inventado correo electrónico. En 1972, Ray Tomlinson modificó el programa de correo, eligió el signo @ para denotar «en» y su sistema resultó tan popular que solo un año más tarde el 75% del tráfico de Arpanet era correo electrónico. No se trata de meras anécdotas. Como siempre pasa con las invenciones, Tomlinson no tenía la menor idea del poder de su engendro y nunca pudo imaginar que con esas simples medidas cambiaría la historia del mundo.

the Internet Topology», sabemos ahora que la red de *routers* que corona Internet es una red libre de escala, como decíamos más arriba, es decir, profundamente asimétrica.⁴ Dicho en romance, todos los modelos de la red que imaginaban -hasta bien entrado 1999- que esta crecía en forma gaussiana y azarosa estaban definitivamente equivocados. Esto es lo primero que debemos considerar para analizar su topología y lo llamaremos «reconocimiento».

El descubrimiento de los hermanos Faloutsos no es menor porque está en la base del comportamiento de la red hoy. Sugerimos revisar numerosos gráficos y cuadros que testimonian el carácter libre de escala de la red y su crecimiento exponencial (pueden verse los siguientes sitios: Web Characterization; Self-organized networks; Small World Project; Network dynamics; How big is the web). Sin embargo, nada de esto alcanzará para entender Internet hoy. Necesitamos tres principios adicionales, aparte del reconocimiento.

Debemos considerar entonces la conexión preferencial (*preferential attachment*): la lógica indica que todos deberíamos conectar nuestra computadora al *router* más cercano. Sin embargo, el elegido no será el *router* más cercano sino el más potente. En efecto, lo que realmente interesa (cuando se trata de abaratar la conexión) no es la distancia al nodo, sino el ancho de banda. No siempre es más económico conectarnos a un *router* cercano pero pobre en cantidad de bits por segundo (velocidad de transmisión), que a uno lejano pero más rico en ellos.

Esto se traduce en algo evidente: los nodos más ricos (porque proveen más ancho de banda) son ipso facto los que más conexiones atraen, conformando entonces un efecto bola de nieve o una estructura del tipo «el ganador se queda con todo», base de la distribución asimétrica de Internet.

Pero no alcanza aún con estos dos criterios para entender cómo funciona la red. Porque aunque no sea el criterio definitivo, la distancia también importa. Por un lado, el kilometraje que debe recorrer la fibra óptica es un factor considerable en el precio de la conexión; por otro, los nodos no aparecen al azar en los mapas. Hay más *routers* donde hay más demanda de ellos. Además, la correlación pertinente es aquella que pone en correspondencia la densidad de la población y la densidad de los nodos de Internet: para nuestra sorpresa, la distribución de los *routers* sobre el territorio de Estados Unidos conforma un conjunto fractal.⁵ El hecho de que Internet y el tráfico que la caracteriza sea fractal, o auto-similar significa que independientemente del corte temporal que se tome -un día, una hora, o un segundo- se verá exactamente igual. En síntesis, para entender la dinámica de la red debemos tener en cuenta estas cuatro dimensiones o variables: reconocimiento, conexión preferencial, dependencia de la distancia y estructura fractal subyacente.

Si cada una de estas dimensiones tomara rumbo propio, podría destruir la red. Así si la longitud del cable fuera el único factor a considerar, la red podría fácilmente recaer en una estructura semejante a la de las autopistas. Pero increíblemente (¿mágicamente?) las cuatro dimensiones interactúan entre sí eliminando cualquier posible inestabilidad y haciendo posible una red libre de escala.

Sin embargo, esto no debe hacernos olvidar que tremenda labilidad representa un motivo de sorpresa y alegría pero, simultáneamente, una condición posible para el eventual colapso de la red. Un episodio ocurrido el 25 de abril de 1997 ofreció en efecto un indicio de lo que podría suceder: un técnico de MAI Network Services cambió por error dos líneas de código y permitió así que un *router* indicara cuál de todos ellos proveía el mejor camino a toda la red. Así, decenas de miles de *routers* eligieron ese pequeño tubo como camino privilegiado para enviar mensajes en la red. En pocos minutos centenares de miles de *routers* cayeron en ese agujero negro y se

⁴ Además de los hermanos Faloutsos, Albert Barabási reunió y encapsuló datos sueltos y deshilachados en las fronteras de muchas disciplinas para hacerlos converger en su obra seminal *Linked. How Everything is Connected to Everything Else and What It Mean*. abriendo así las puertas a una redología o teoría de las redes. Existen al menos cuatro libros no menos valiosos que enlazan temas semejantes con maestría, sino igual al menos digna de destacar. Se trata de las obras de David Weinberger *Small Piece Loosely Joined. A Unified Theory of the Web*; de Mark Buchanan, *Nexus. Small Words and the Groundbreaking Theory of Networks*; de Duncan J. Watts, *Six Degrees. The Science of a Connected Age* y de Bernardo Huberman, *The Laws of the Web: Patterns in the Ecology of Information*.

⁵ Es decir, que tiene al menos una de las siguientes características: detalle en escalas arbitrariamente grandes o pequeñas, es demasiado irregular para ser descrito en términos geométricos tradicionales, tiene auto-similitud exacta o estadística, su dimensión es mayor que su dimensión topológica, o es definido recursivamente. Esta definición, como muchas que utilizaremos a lo largo de este libro, proviene de la Wikipedia.

perdió todo el tráfico de la red. Este ejemplo -y hay muchos más ligados a la proliferación de virus y otras intoxicaciones en la red- mostró la vulnerabilidad de Internet en cuanto a la propagación de errores.

Es claro que Paul Baran jamás imaginó que los enemigos de la red podrían estar durmiendo en su seno. Tampoco pensó que no serían los rusos sino los *hackers* quienes, con una sorprendente facilidad, podrían acabar con la red en tiempo récord. Esto puede hacerse apropiándose de los *routers* clave en los servidores de nombres de raíz que orientan todo el tráfico (hay apenas 13 de este tipo en toda la red),⁶ ya sea lanzando marejadas de ataques de denegación de servicio contra los nodos más activos (tal como aconteció en febrero de 2000 contra Yahoo, Amazon, CNN y eBay) o a través del ataque de un virus, como fue el del gusano Código Rojo.

Pero ¿qué es efectivamente Internet? ¿Podría, como se le ocurrió a alguien, ser una computadora (aunque no fuera más que en clave metafórica)? Es obvio que la red está compuesta por computadoras que intercambian páginas y mensajes de correo electrónico, pero esto por sí mismo de ningún modo convierte a la red en una supercomputadora. A menos que puedan inventarse programas, gusanos, troyanos o algún tipo de dispositivos de software que tomaran por ataque a todas las computadoras del mundo y las hicieran trabajar para alguien, Internet seguirá siendo una red no centralizada.

Sin embargo, esta idea de hacer trabajar a toda la red para una misma cabeza no era descabellada y pudo comprobarse en agosto de 2001, con la aparición de una nota publicada en *Nature* por Albert-László Barabási y otros autores acerca de la computación parásita. Lo que hizo el equipo comandado por Barabási fue disfrazar verdaderas proezas computacionales por simples pedidos de armado de páginas. Así, cuando una computadora cualquiera recibía un mensaje de pedido hacía, aparentemente, un chequeo rutinario para comprobar que el mensaje no se había corrompido, pero, en esa operación trivial el equipo estaba en realidad resolviendo un complejo problema computacional diseñado por los investigadores autores de la nota, que usaron así recursos computacionales ajenos sin que los dueños de los equipos se dieran cuenta (*Nature*, vol. 412, agosto, 2001). Es interesante comentar también que el artículo del equipo de Barabási fue respondido por decenas de miles de mensajes enviados desde todos los rincones del mundo, que amenazaban con represalias sin fin a los intrépidos que habían osado invadir sus sacrosantas máquinas.

De todos modos, lo que los investigadores descubrieron fue la posibilidad de esclavizar máquinas remotas, y abrieron así el abanico de preguntas y problemas computacionales, éticos y legales que ni siquiera hoy están bien planteados y mucho menos resueltos. ¿Que pasaría si alguien perfeccionara la técnica y pudiera generar un sistema de concentración de recursos ajenos a gran escala? ¿De quiénes serían los recursos disponibles en la red que de pronto son esclavizados o apropiados por terceros? ¿Constituiría esto el inicio de Internet como supercomputadora? ¿Podría emerger un ser consciente e inteligente como resultado de estos devaneos y operaciones? De lo que no hay duda es que en un futuro sumamente cercano las computadoras empezarán a intercambiar servicios y esfuerzos sobre la base de una demanda ad hoc. Si hoy los chips funcionan mucho más rápido que las redes, en el futuro ambas velocidades se equiparán y problemas inverosímilmente complejos se resolverán a manos de una computadora única, o de un grupo de investigadores aislado.

⁶ El nivel superior de la jerarquía de servicios de nombre de dominio (DNS), también llamado nivel de raíz, es gestionado por un conjunto de 13 servidores llamados servidores de nombre de raíz. Estos servidores son la columna vertebral del funcionamiento de la red. Son propiedad de distintas agencias independientes y los servidores se llaman austeramente A, B, C hasta llegar al M. Diez de ellos están en Estados Unidos y los restantes tres en Tokio, Londres, y Estocolmo. En octubre de 2002 un ataque masivo de denegación de servicio distribuido, de origen desconocido, interrumpió durante una hora el servicio web en nueve de los 13 servidores raíces que controlan la red. Aunque los usuarios finales no experimentaron las consecuencias, y muchos expertos minimizaron a posteriori el episodio, otros analistas insisten en que de haberse prolongado el ataque durante varias horas más, este hubiese supuesto el colapso de la red. Los contraterroristas insisten en que los atacantes o bien no sabían cuánto tiempo hacía falta para bajar a los servidores, o tal vez estaban disfrazando su ataque porque su interés real era probar su red DDoS con vistas a futuros ataques más puntuales. Para que el ataque se propagara al resto de la red se hubiese necesitado al menos cuatro horas de ataque continuado. Durante este intento frustrado de los ciberterroristas se perdió entre el 16% y el 10% de los sitios buscados.

Ya tenemos ejemplos de este tipo discutidos en esa excelente obra sobre la comunicación par a par compilada por Andy Oram, *Peer-to-peer: Harnessing the Power of Disruptive Technologies* (2001). De hecho, experiencias como SETI (Search for Extraterrestrial Intelligence) acudieron a este esquema aunque en una escala infinitamente menor, pero sobre todo pidiendo permiso (algo que Barabási y sus colegas evitaron graciosamente hacer).

Pero si el procesamiento distribuido fuera norma, una infinidad de recursos actualmente disponibles (las horas, semanas o meses que nuestra máquina permanece conectada al ADSL indolentemente, por ejemplo) podrían servir para enormes y sobre todo ajenos fines, sin que nosotros necesariamente nos enterásemos.

Es necesario decirlo una vez más: las tecnologías van mucho más rápido que nuestra capacidad de investigarlas, analizarlas, comprenderlas y sobre todo de establecer normas para su funcionamiento. Es por eso que Internet es un espacio que necesita ser mapeado y analizado casi diariamente.

En síntesis, a pesar de su crecimiento aparentemente azaroso, la web posee sorprendentes regularidades que pueden encontrarse en la organización de los links o en los patrones de comportamiento de sus centenares de millones de usuarios. Muchas de estas regularidades han sido anticipadas por la mecánica estadística, una de las ramas de la física, al considerar que el comportamiento de los usuarios puede tratarse o medirse del mismo modo que las moléculas de un gas o una colonia de insectos. Si en efecto estas metáforas son válidas, será posible diseñar de forma más eficiente el uso de la red ya que, visto de este modo, se trata de un ecosistema informacional gigantesco que puede ser utilizado como laboratorio físico (y metafísico) para revisar y analizar el comportamiento humano y la interacción social. Porque contrariamente a lo que podríamos esperar -y a pesar de su enorme tamaño-, las regularidades y patrones que la red exhibe nos permiten imaginar usos más inteligentes de recuperación de datos, así como la creación de nuevos formatos y narrativas que la utilicen -y la reinventen- de una forma mucho más inteligente y apropiada de lo que hemos podido o sabido hacer hasta ahora.

Sitios web relacionados con los contenidos de este capítulo

- Paper original de Paul Baran de 1959 acerca del packets-witching:
<http://www.ibiblio.org/pioneers/baran.html>
- Acompañante visual del libro de Barabási *Linked. The new science of networks*:
<http://www.nd.edu/networks/linked/down.html>
- Web Characterization. Tendencias en la evolución de la web pública:
<http://www.dlib.org/dlib/april03/lavoie/041avoie.html>
- Self-organized networks. Sección de redes auto organizadas de Barabási en la Universidad de Notre Dame:
<http://www.nd.edu!/networks/>
- Small World Project. Uso del correo electrónico para testear la hipótesis de Stanley Milgran acerca de los grados de separación:
<http://smallworld.columbia.edu/project.html>
- Network dynamics. Sección del Instituto de Santa Fe especializada hace 20 años en ciencias de la complejidad:
<http://www.santafe.edu/research/networkDynamics.php>
- How big is the web? Estudio canónico de Steve Lawrence y C. Lee Giles de 1997 acerca del tamaño de la web:
<http://www.neci.nj.nec.com/homepages/lawrence/websize.html>
- Fractales en la Wikipedia. Definición del término en la más grande enciclopedia colaborativa y auto organizada el mundo:
<http://es.wikipedia.org/wiki/Fractal>
- 13 Routers Maestros (Niveles de raíz). Bases de datos que hacen posible el envío de mensajes en la red:
<http://www.root-servers.org/>
- Mapas de Internet:

<http://www.peacockmaps.com>

- Atlas del ciberespacio:

<http://www.cybergeography.org/atlas/atlas.html>

2

Las «power laws»

Existe un gran debate acerca de la tasa de crecimiento de Internet. Lo que puede precisarse, sin embargo, es que crece a una velocidad mucho menor que cuando saltó a la notoriedad pública, a fines de 1994, y que su tamaño es hoy enorme, tres veces mayor que el que tenía en 1999. En este capítulo realizaremos una suerte de balizamiento de su tamaño efectivo, ya que los datos que en la actualidad existen acerca de su crecimiento, junto con otros acerca de su configuración, nos permiten entender mucho mejor que hace un lustro su morfología y dinámica. En el interin deberemos dar cuenta de los desafíos que plantean la banda ancha y la Internet inalámbrica.

Para comenzar, es importante distinguir entre stock y flujo de la información. Es más o menos fácil sumar la cantidad de información contenida en los sitios web o en los mensajes de correo electrónico almacenados. Pero no es tan sencillo hacerlo con continentes dinámicos, como todo lo que circula en los chats o en el correo electrónico, que solo existen como flujos. Precisamente, uno de los rasgos salientes que han vuelto a Internet tan exitosa es que pertenece al haz de medios (como la radio y la TV) en donde una sola unidad de almacenamiento puede generar un flujo de información cientos de veces mayor, mientras que cada ejemplar de un libro suele ser leído por una o dos personas, lo cual implica un flujo de información sumamente bajo por cada tirada.

Por otro lado, podemos hacer una segunda distinción: existen dos grupos de contenidos en la web. Al primero podemos llamarlo web de superficie y se corresponde con la versión tradicional de la web. Se trata de las páginas estáticas, disponibles de forma pública y abierta y es, en rigor, la parte más pequeña de la web conocida.

En contraposición, tenemos la web profunda, que consiste en las bases de datos y sitios dinámicos desconocidos por la mayoría de quienes navegan, aunque curiosamente la información existente en estas direcciones es entre 400 y 550 veces mayor que la información disponible en la web superficial.

Para contemplar estos datos en términos de crecimiento, hay que considerar que a fines del año 2000 la web superficial tenía alrededor de 2.500 millones de documentos, y que, cuatro años antes, contaba con menos de la mitad, 1.000 millones. Su tasa de crecimiento era en esos años de 7,3 millones de páginas por día (0,1 terabytes).⁷ El tamaño promedio de las páginas oscilaba entre los 10 y los 20 kb. Así, el tamaño total de la web de superficie oscilaba entre los 25 y los 50 terabytes de información, de los cuales correspondían a información textual entre 10 y 20.

Por otro lado, con respecto a la web profunda, si tomamos como base de medida todas las bases de datos conectadas, las páginas dinámicas y las intranets de las organizaciones, entonces debemos hablar de cerca de 550.000 millones de documentos web, con una página promedio de 14 kb. Si quisiéramos guardar toda esta información en algún lugar estaríamos hablando de la impresionante cantidad de 7.500 terabytes.

Dos de los sitios profundos más grandes del mundo son las bases de datos del National Climatic Data Center y de la NASA. Entre las dos, contienen un total de 585 terabytes de

⁷ Un terabyte (tb) equivale a 1.000 millones de bytes. Para dar una idea más concreta de lo que esto implica, puede servir el siguiente ejemplo: una biblioteca completa de investigación académica equivale aproximadamente a solo dos terabytes.

información, prácticamente el 7,8% de la web profunda. Solo 60 de los sitios web más grandes almacenan 750 terabytes de información, lo que supone el 10% de la web profunda.

Ahora bien, si consideramos el flujo de información, las cifras llegan a niveles impensables. El correo electrónico se ha convertido en una de las formas de comunicación más extendidas y utilizadas de todas las épocas: un empleado de oficina estándar recibe unos 40 mensajes diarios. Aunque los números varían, se calcula que durante el año 2000 se enviaron entre 610.000 millones y 1,1 billones de mensajes. Dado que el tamaño promedio de un e-mail es de 18k, el flujo puede llegar a 11.285 y 20.350 terabytes. Por supuesto que una cantidad ínfima de estos e-mails son almacenados. Se calcula que solo se guarda uno de cada 17 mensajes que se reciben.⁸

En cuanto a las listas de distribución, podemos hablar de cerca de 30 millones de mensajes diarios distribuidos en 150.000 listas de distribución. Como las listas de distribución representan el 30% de los mensajes enviados, el número total anual de todos ellos es de alrededor de 36.500 millones, con un volumen agregado de 675 terabytes.

Mucho más difícil es medir el flujo de información en los grupos de noticias (o Usenet). Cidera, el proveedor número 14 de distribución de noticias, recibe 0,2 terabytes de información por día, lo que lleva a una cantidad anual de 73 terabytes, redistribuidos a través de los distribuidores locales y los servidores de noticias un número indefinido de veces.

En cuanto a la circulación de archivos (ftp), es casi imposible determinar las cifras. Lo mismo pasa con los chats y otro tipo de información dinámica: en 2000 había más de 115 millones de usuarios de servicios de mensajería instantánea (distribuidos principalmente entre ICQ y AOL Instant Messenger), de modo que el caudal diario de información circulante es tremendamente difícil de precisar.

Aun así, es posible brindar algunos números globales tentativos, bastante aproximados al uso real total, como dijimos al principio. Pero podemos hacer una aproximación más macro aún, identificando cómo se inscribe este torbellino y terremoto de información dentro de la ecología más amplia de la información: sabemos que, hasta 1999, el mundo producía entre 1 y 2 exabytes (1018 bytes o 1.000 millones de megabytes) de información única por año, lo que equivalía aproximadamente a 250 megabytes por cada hombre, mujer y niño en un mundo de 6.300 millones de habitantes. Sin embargo, los documentos impresos de cualquier tipo apenas llegan al 0,03% de ese total.

Por suerte, se puede acceder a todos estos datos en un exquisito informe producido por la School of Information Management and Systems (SIMS) de Berkeley, en el que se ha estimado la producción anual de información en el mundo, en cualquiera de los formatos existentes.⁹ Allí mismo pueden encontrarse datos para el año 2002¹⁰. Entre ellos, es interesante precisar los siguientes, de orden general: la web superficial llegaba en ese año a 167 terabytes, mientras que la web profunda alcanzaba los 91.850 terabytes. Los mensajes originales reunían una cifra de 440 terabytes y la mensajería instantánea, 274 terabytes. El total llega así a los 532.897 terabytes de información, lo cual demuestra hasta qué punto es evanescente la web.

⁸ Por ejemplo, Mail.com contaba en 2000 con 14,5 millones de casillas de correo electrónico y utilizaba solo 27 terabytes de almacenamiento, cuando el espacio requerido podría involucrar 900 terabytes (contaba con cerca de 500 millones de casillas de e-mail en el mundo).

⁹ Puede visitarse el sitio de la SIMS y leer el informe en <http://sims.berkeley.edu>. El trabajo que citamos está basado en una serie de excepcionales aportes. Entre estos, cabe mencionar los artículos de Jim Gray y Prashant Shenoy: "Rules offhumb in Data Engineerin", 2000. Así como: los trabajos no menos notables de Michael Lesk, «How Much Infonation is There in the World?», 1997; el informe de Andrew Odlyzko, "Content is not Kin", 2000; y los famosos estudios de Ithiel De Sola Pool, Hiroshi Inose, Nozomu Takasaki, Roger Hurwitz "Communications Flows: a Census in the United States and Japan", 1984. También el artículo de este último: "Tracking the Flow of Information". De todo este material lo más interesante son los trabajos de Jim Gray y sobre todo los de Andrew Odlyzko, que aportan una fantástica relectura acerca del pasado y el futuro de la comunicación científica.

¹⁰ Aunque podríamos revisar y actualizar muchos de los números anteriores multiplicándolos (dada la duplicación del almacenamiento de información que pasó de 2 a 5 exabytes, de 1999 a 2002), una muestra de esos números actualizados revela dónde estamos parados hoy (mejor dicho, a mediados de 2003, cuando se terminó el estudio). Si bien, en líneas generales, la web no crece como lo hizo en el período dorado de 1994 a 1996, su tamaño y tasa de aceleración es impresionante. Citemos simplemente uno de los ejes conductores de nuestros análisis: los weblogs. Según el informe de SIMS -insistimos, una foto tomada a mediados de 2003 en este caso- su total era en esa fecha de 1.600.000. Los datos superactualizados de *Technorati* revelan que, a fines de 2004, contábamos con 5.300.000 *weblogs* (véase el capítulo 3 de este libro).

Por otro lado, es necesario entender no tanto la estática, sino la dinámica de la información. Ya es hora de que dejemos atrás los análisis descriptivos y empecemos a teorizar con cierto rigor acerca de las leyes que organizan las redes en general, y la web en particular.

Una web muy fragmentada

Albert Barabási¹¹ y su equipo realizaron en 1999 una serie de análisis -que abarcaron cerca de 200 de los 8.000 millones de páginas censadas actualmente por Google-, que terminaron revelando que Internet es un *caos*. En efecto, el trabajo demostró que existe un núcleo organizado en el centro de la red, compuesto tan solo por el 30 % de los sitios existentes, pero que el resto de los sitios web es, en muchos casos, inaccesible. Pero lo más interesante no es esta afirmación sino el nudo que vincula este trabajo a muchos otros sobre redes de distinto tipo: gracias a Internet, en los últimos años adquirimos un enorme conocimiento acerca de cómo funcionan las redes y, sobre todo, de cuál es su topología. En efecto, la topología es invariante para todas las redes, así se trate de células, cadenas tróficas, citas bibliográficas o de la misma red.¹²

Al analizar la estructura de la web, Barabási constató y explicó en su trabajo que, además de los nodos y enlaces que constituyen el meollo de Internet, existen agrupamientos de nodos (*clusters* o racimos) y un pequeño número de nodos con una enorme cantidad de enlaces a otros nodos (los *hubs* o concentradores). Aunque suene a ejercicio de ingenuo antropocentrismo, la tendencia humana al agrupamiento -que responde sin duda a la búsqueda de seguridad- se encuentra en todas las redes. Según Barabási, los *hubs* pueden resultar más importantes todavía, ya que mantienen la cohesión de la mayoría de las redes.

Si algo parecía claro de la web es que se trata de un territorio inmenso. Sin embargo, luego del trabajo de Barabási descubrimos que las distancias que separan a los habitantes de ese mundo (aparentemente) ilimitado son relativamente cortas. Barabási se apoyó para desarrollar esta hipótesis en una idea acuñada por John Guare en su obra teatral *Six Degrees of Separation*: la distancia que existe entre dos personas es de seis intermediarios. En el caso de Internet, *distancia* es el número de saltos necesarios para ir de un nodo a cualquier otro, siguiendo el camino de los enlaces. En este sentido, una vez que Barabási trabajó e investigó esta idea, la impresión común según la cual Internet contribuye al desmoronamiento de las distancias encontró su base matemática: aun en las redes más complejas, la distancia entre nodos es más bien corta. Ahora sabemos que si la distancia entre un par de humanos conocidos es de 6, entre un par de páginas de la web es de 19.

Los humanos constituimos una red de 6.300 millones de nodos, y la web tiene más de 8.000 millones. Sin embargo, esas redes son *pequeñas*, entendiéndolo por tal que existe un camino corto entre cualquier par de nodos. Pero, antes de que Stanley Milgram lo descubriera afines de la década de 1960 para las redes de personas y Barabási recientemente para los sitios web, esa cifra era inimaginable... Por lo pequeña.

¹¹ Albert-László Barabási, profesor de física en la Universidad de Notre Dame, es autor de *Linked, The New Science of Networks* [Enlazados, la nueva ciencia de las redes], que está dando mucho que hablar y esperamos que pronto esté traducido al castellano. No solo porque está lleno de anécdotas y de antecedentes fascinantes, que muestran cómo a partir de lo local se puede llegar muy rápido a lo global; sino porque Barabási, con gran astucia, une lo conocido con lo desconocido, lo archisabido con lo sorprendente, lo previsible con lo inesperado. Los primeros capítulos del libro son fascinantes por su entramado: el autor parte de los lugares más insólitos para sacar conclusiones de gran alcance, y por demás convincentes, acerca de la naturaleza y futuro de las redes. Se puede consultar un acompañamiento visual del libro en su sitio web. Como bien dice Pisani, lo que falta en el libro de Barabási es una explicación de por qué las redes se multiplican hoy, por qué parecen ganar en fuerza frente a las demás formas de organización. De esto se ocupan John Arquilla y David Ronfeldt en trabajos sobre la Netwar, o guerra de redes. (La obra que editaron, *Networks and Netwars: The Future of Terror; Crime and Militancy*, se puede encontrar completa en formato PDF, en la red.)

¹² Como bien revelaba Francis Pisani en «¿Por qué el mundo es pequeño y por qué los ricos se vuelven más ricos?», nota periodística publicada por *Latinotek*, un gran número de punto coms que se lanzaron al ciberespacio pensando que bastaba con llegar rápido (siguiendo el axioma de *Amazon.com* «Hagámonos grande bien rápido») para que los internautas acudieran en masa, hubieran podido escoger una estrategia más apropiada. O mejor aún, no haber invertido nunca un dólar en una tarea que era claramente imposible.

Lo que más sorprendió a Barabási fue la ausencia total de democracia, equidad y acceso igualitario a estos nodos. La topología de la web nos impide ver *todo*, a excepción de un puñado de los miles de millones de documentos que resultan indexables. De hecho, los nodos no se conectan entre sí al azar, sino que, contrariamente y en un ambiente de constante competencia, buscan a otros no dos más atractivos. Barabási propone la existencia de cierta *fitness* (aptitud, en el sentido evolucionista), lo que lo lleva a afirmar que la noción según la cual los ricos se vuelven más ricos es el mecanismo clave.¹³

Así, Albert-László Barabási y su equipo se dieron cuenta muy pronto de que la muestra de 300.000 páginas alcanzada por su pequeño robot buscador era lo máximo que podían aspirar a analizar. Sin embargo, aún en su minúscula dimensión o a causa de ella, les había revelado el secreto de la web que ya hemos enunciado: se trata de una red libre de escala.

Otro grupo de investigadores que se animó a seguir el experimento, y analizó bloques mucho más numerosos de páginas, llegó a la imprevista conclusión de que la web se encuentra fragmentada en continentes y comunidades que determinan nuestro comportamiento on line. Correlativamente, descubrieron que hay una *terra incógnita* donde nadie se ha aventurado aún. Y por si todos esos resultados no fueran por sí mismos increíblemente interesantes, lo que el estudio de la dinámica de la web ha demostrado es que su estructura tiene impacto en todo: desde la navegación hasta la democracia.

Los robots buscadores en una red libre de escala

Pese a las muchas expectativas, los robots no aspiran a controlar al mundo. Por lo menos, no el tipo de robots que nosotros pensábamos. Existe sin embargo otra clase de robots que aspira a colonizarlo todo. En una escala, claro, microscópica o en todo caso invisible. Porque los robots del siglo XXI son invisibles e inmateriales. No aparecerán flotando en la pantalla o fisgoneando nuestros escritorios -aunque también estén allí-. Una cuidadosa lectura de los archivos de cualquiera de nuestras máquinas mostrará que estos robots se dedican a hacer algo clave: indexar las páginas que visitamos. Se trata de los *buscadores*.

Hoy la situación es muy diferente de la que vivíamos apenas cuatro o cinco años atrás. En ese momento, nuestra fe ciega en los robots buscadores, y el carácter supuestamente isotrópico de la red, nos hacía suponer que si buscadores como *Altavista* o *Hotbot* no encontraban aquello que buscábamos, era sencillamente porque tal información no existía en la red.

Sin embargo, todo cambió en abril de 1998 cuando el *paper* «Accessibility of Information on the Web» [Accesibilidad de la información en la web], publicado por Steve Lawrence y Lee Giles en la prestigiosa revista *Science*, saludó la importancia de la calidad por encima de la cantidad. Esto determinó que era preferible indexar sitios de calidad, antes que apilar meramente la mayor cantidad de sitios posibles, acudiendo a los buscadores. Correlativamente, lo que este enunciado decía era que el universo web estaba lleno de páginas que no valía la pena visitar nunca. Como sucede muchas veces con las investigaciones científicas, nada estaba más lejos de los intereses de Steve Lawrence y Lee Giles, coautores del *paper*, que poner en tela de juicio a los buscadores entonces existentes. Por el contrario, sus preocupaciones eran muy diferentes. Lo que les interesaba era el área del aprendizaje robótico: por ello habían construido un robot llamado *Inquirus*, capaz de interrogar a otros robots acerca de la existencia de documentos que cumplieran con cierta estructura de búsqueda.

Embarcados en este interesante proyecto, de pronto se dieron cuenta de que el robot podía arrojar un beneficio secundario más valioso que el buscado originalmente estimando el tamaño real de la red, un número que en ese momento nadie conocía a ciencia cierta. Entre los resultados que cosechó el *Inquirus* aplicado al buscador *Hotbot*, fue descubrir en 1997 que la web contaba con alrededor de 320 millones de documentos (por lo menos el doble de lo que se creía entonces). Y no solo eso: *Hotbot* se preciaba de ser el más exitoso y exigente de los robots en aquella época pero, de pronto, se vio devaluado al descubrirse que solo indexaba el 34% de toda la web. Como premio de consuelo, pudo jactarse de que a los otros robots les iba aún peor:

¹³ Algo que ya había sido adelantado, en 1996, por Robert H. Frank y Philip Cook en su anticipadora obra *The Winner Takes and Society. When the Few at the Top Gets Much More than the Rest of Us*.

Altavista solo cubría un 28%, y otros buscadores -como Lycos, que pronto caería en manos de Terra y Telefónica- apenas cubrían un 2% de la red.

Como ejemplo del divorcio que suele existir entre la investigación y los negocios, la mayoría de las empresas propietarias de los robots no se incomodaron ni se dieron por aludidos por el irritante resultado y siguieron trabajando como hasta entonces. Lo que quedó atestiguado en febrero de 1999, cuando se repitió el mismo ejercicio los investigadores encontraron que la red había crecido (tenía 800 millones de documentos) pero la capacidad de indexación de sitios a manos de los robots curiosamente había empeorado.

Un excelente buscador de la época, Northern Light, ocupó entonces la *pole position* cubriendo el 16% de la web, pero Altavista había bajado al 15% y Hotbot reseñaba apenas el 11% de las páginas existentes. Mientras tanto Google, que era un benjamín entre los pesos pesados, apenas veía en ese entonces un 7,8% de las páginas estimadas. Ese resultado fue bastante más importante que el aparente encogimiento de hombros de las empresas involucradas. En junio de 2001 Google cubrió por primera vez 1.000 millones de documentos, seguido de cerca por All the web. Si echamos una rápida ojeada hoy, encontramos que Google está próximo a alcanzar los 9.000 millones de documentos.

Sin embargo, esta carrera se asemeja a la de Aquiles contra la tortuga. Por muy impresionante que sea la capacidad de indexación de los motores, la web crece infinitamente más rápido que la posibilidad de estos de analizarla. Además, existe una web profunda que es por lo menos 550 veces más grande que la que los robots alcanzan, por lo que la asimetría entre lo visible y lo existente se ahonda muchísimo más.

De todos modos, cuando pasamos de una cantidad finita y accesible de páginas a los continentes de (des)información que tenemos hoy, mucho más importante que contar con un buscador que encuentre *todas* las páginas es otro que encuentre las *mejores* páginas. Sin embargo, lo que estas investigaciones dejaron en claro es que no solo cuentan los incentivos económicos a la hora de recorrer la web, sino que -por su propia topología- la red tiene una estructura que enfatiza tanto zonas de visibilidad como de invisibilidad. Esto significa que la red está dominada por *hubs* y nodos, con una gran cantidad de links. Pero al mismo tiempo esta topología de gran escala coexiste con numerosas estructuras de pequeña escala, que militan lo que en efecto podemos explorar haciendo simplemente clic aquí y allá.

La asimetría de la red: ley de Zipf

Aunque en la web hay miles de millones de páginas -contando solo la web superficial-, las investigaciones de Albert-László Barabási demostraron que el grado de separación entre dos sitios web, a diferencia de lo que sucede entre dos personas físicas, es de apenas 19 saltos. Esto significa que, a pesar de su vastedad, esta inmensidad de información es navegable. Sin embargo, describir así, de un modo algo simplista, la web no dice mucho. Porque si bien es cierto que si existe un *link* entre dos nodos, este es generalmente corto, no todas las páginas están conectadas entre sí. De hecho, partiendo de cualquier página en particular, solo podemos llegar a un 24% de todos los documentos que hay en la red. El principal responsable de esta enorme limitación en el acceso es la asimetría de los links. Veamos por qué: todos los links en la web son unidireccionales. O sea, todos están direccionados. Cuando no hay un *link* directo entre varios sitios, se pueden encontrar paradas intermedias, pero no existe una simetría entre las idas y las vueltas. Para volver aun sitio, a menos que usemos el comando *atrás*, probablemente tengamos que dar decenas de pasos. La web está llena de estas desconexiones, y eso es lo que determina el límite de su navegabilidad. En un sentido profundo, la web es como una cadena trófica: todos los ingredientes están interconectados pero en una dirección única. Es esta direccionalidad la que determina de forma esencial la topología de la red.

Uno de los primeros trabajos que ilustraron esta situación fue emprendido por Andrei Broder y sus colegas de Altavista, quienes estudiaron una muestra de 200 millones de no dos (cerca de un quinto de todas las páginas que existían en 1999). La principal conclusión de su trabajo, sintetizado en «Graph Structure in the Web» [Estructura gráfica en la web] es que la direccionalidad de Internet no conforma una red única homogénea. Por el contrario, un mapa

mostraría que está dividida en cuatro grandes estratos, cada uno de los cuales nos obliga a obedecer a diferentes reglas de tráfico cuando queremos navegar por ella.

El primero de estos continentes contiene cerca de *un* 25% de todas las páginas web. Denominado a menudo como «núcleo central», es el espacio donde se alojan todos los sitios conocidos, desde Amazon hasta eBay. Su principal rasgo es su fácil navegabilidad, dado que siempre existe un link que vincula cualquiera de los nodos que lo conforman a otro. Esto no significa que esos links existan siempre; pero sí que, saltando una cantidad de veces, es posible llegar a todos los nodos de esa partición.

Los dos continentes siguientes, denominados «adentro» y «afuera», son tan grandes como el primero. Aunque también mucho más difíciles de navegar. Se puede ir, desde los nodos del continente «adentro», al «núcleo central». Pero no hay links reversibles que nos lleven del núcleo central al continente «adentro». Lo mismo sucede con el continente «afuera», ya que numerosísimos links van del núcleo central a él, pero no a la inversa. Un ejemplo son los sitios corporativos, a los que se puede acceder fácilmente desde «afuera». Pero, cuando uno se halla en el «núcleo central», no hay forma de salir y hacer el camino inverso.

El cuarto continente está conformado por rulos e islas desconectadas, grupos aislados de páginas interconectadas que no se pueden alcanzar desde el núcleo central, y desde los cuales tampoco hay links que permitan volver a él. Cada uno de estos sitios aislados puede ser muy grande y contener miles de documentos web.

La existencia de estos cuatro continentes muestra lo difícil que es navegar por la web.¹⁴ Aparentemente, con un solo clic o dos podríamos alcanzar cualquier página. Así lo creíamos hasta que estos estudios vieron la luz. Pero lo cierto es que a dónde lleguemos dependerá de dónde partamos. De lo que ya no cabe duda es de la futilidad de los buscadores cuando de mapear el continente íntegro de la web se trata. Los buscadores tienen relativo éxito cuando pretenden indexar el núcleo central y el área «afuera». En ambos casos, los buscadores pueden hacer pie en sitios determinados desde donde iniciar la búsqueda. Pero la otra mitad de la red -trátese del continente «adentro» o de los islotes y rulos- es absolutamente invisible a los buscadores.

Una buena analogía de esta descripción es el sistema de citas científicas. Todas las publicaciones científicas forman parte de una red, en la cual los nodos son las publicaciones conectadas por estas citas. Esas conexiones están dirigidas. Seguir las referencias de un libro recién publicado en cualquiera de los *papers* de la red es perfectamente factible y deseable. La inversa es imposible, porque ninguna de esas referencias son anteriores a la publicación de los libros recientes.

La red de citas es bastante peculiar, porque su núcleo central es generado de forma cronológica por los *papers* preexistentes. Pero, en general, es muy poco lo que contienen y la mayoría de sus nodos están en los márgenes, en las islas o en los rulos. De hecho, podrían darse ejemplos semejantes con las cadenas tróficas: los leones se comen a los antílopes, pero los antílopes no pueden hacer lo mismo con los leones.

La pregunta obligada es si esta laxitud y falta de conectividad se mantendrá indefinidamente; si solo emana de deficiencias en los algoritmos de búsqueda o de estrategias equivocadas de recuperación de la información. Contrariamente a muchas suposiciones, mientras los links sigan siendo unidireccionales esa homogeneización no se logrará nunca.

Si bien todo lo anterior puede parecer demasiado abstracto y formal (fundado como está en rigurosas nociones matemáticas y físicas), por suerte ha habido científicos sociales que le encontraron rápidamente su correlato sociológico. Así, en junio de 2000, Cass Sunstein, autor de la notable obra *Republica.com* y profesor de derecho en la Universidad de Chicago, llevó a cabo un análisis de 60 sitios políticos. Descubrió, como era de prever, que apenas el 15% remitía a sitios con opiniones contrarias. A la inversa, el 60% mantenía links hacia sitios afines. Lo mismo se corroboró en otro estudio dedicado a analizar el discurso democrático en la red.

¹⁴ En su impresionante trabajo, *Evolutions of Networks. From biological nets to de Internet and WW*, S. N. Dorogovstev y J. F. F. Mendes demostraron que el tamaño y la estructura de estos continentes puede predecirse analíticamente. El tamaño depende de la estructura de la red; pero en todos los casos no hay unificación de sus divisiones, tal como se comentó antes.

La rápida conclusión que sacó Sunstein es que, al limitar las remisiones a opiniones contradictorias y al focalizarse de modo autista en el reforzamiento de las opiniones propias, la topología de la red termina facilitando la segregación y fragmentación social. En efecto, se trata de mecanismos de refuerzo y consolidación de ideas afines. Solo que en la web, a diferencia del mundo real, las posibilidades de generar microcomunidades -con identidades y capacidad de autojustificar indefinidamente sus principios- es altísima, un fenómeno inexistente en el mundo real.

Sin embargo, cuando tenemos un universo ocupado por miles de millones de páginas, ¿es posible todavía hablar de comunidades? ¿Cómo podemos definir, entonces, una comunidad en la red?

Gary Flake, Steve Lawrence y Lee Giles, en su *paper* «Efficient Identification of Web Communities» [Identificación eficiente de comunidades virtuales], sostuvieron que los documentos pertenecen a una comunidad solo cuando mantienen más links entre sí que con otros documentos fuera de la comunidad. Aunque esta definición es lo suficientemente precisa, encontrar esas comunidades es difícil e infrecuente. Barabási sostiene que este tipo de problemas se parece a los problemas completos NP, es decir, esa clase de problemas computacionales que son resolubles en principio, o de manera teórica, pero no de hecho: aunque es posible identificar a dichas comunidades, en la práctica no existen algoritmos para hacerlo efectivamente. Los ejemplos acerca de los sitios proaborto y antiaborto que menciona Sunstein no son irrelevantes. Muestran que en la web, como en la vida real, las posturas conservadoras, aristocráticas, bien financiadas y manejadas por los intereses más convencionales le sacan leguas de ventaja a las otras.

En el año 2000, 6 de cada 10 páginas no habían sido visitadas nunca. Hoy los resultados deben alcanzar cifras de entre 8 y 9 de cada 10. Pero la situación es inevitable, como en la carrera que Aquiles corría contra la tortuga. En ausencia de los buscadores, en vez de una jungla la web habría resultado un agujero negro, invisible y entrópico. La función de los buscadores es impedir que la web se colapse como una ballena ahogada por su peso, apretándose los pulmones en tierra. Lo sorprendente es que, dado que nuestra vida depende cada vez más de la web, nosotros sepamos poco o nada de todas estas leyes, reglas, coordenadas, algoritmos y estructuras profundas, amparados en la comodidad de sobrevolar su estructura de superficie.

Sin embargo, lo más rico de todo este periplo es haber descubierto que la ecología de la información que implica la web en sí misma está bajo el mismo tipo de leyes que rigen las células y los ecosistemas, lo que confirma nuestras presunciones acerca de la naturaleza autoorganizada de los sistemas complejos. Pero, en el caso de la red, dada su naturaleza digital, se trata de un sistema modelo cuyos detalles pueden investigarse con sumo detalle. Del mismo modo que la nave Huygens-Cassini entró a mediados de 2004 por primera vez en la órbita de Saturno y nos reveló secretos desconocidos del planeta de los anillos, nuestros análisis de la web -en estos enfoques de modelos y metáforas- van creando un mundo nuevo, habitado desde hace una década pero cuyo mapeo recién hoy empieza.¹⁵ Todo esto debe llevar a analizar los problemas del código que impera en la web, es decir, del software que acaba por regular los grados de libertad de la red. ¿Qué quiere decir exactamente esto? En sus orígenes, Internet fue diseñada con una arquitectura *end to end*, es decir, un tipo de red que manifiesta su inteligencia en muchas pequeñas funciones ubicadas en sus márgenes. Así, redes simples generan aplicaciones inteligentes. La base de su diseño hizo que, más rápido de lo que podía imaginarse, la red renegara de los controles centralizados y apuntase a una forma de gobierno estrictamente

¹⁵ Uno de los proyectos derivados de esta redología es una investigación acerca de cómo el correo electrónico conecta a la gente a escala mundial. De esto se ocupa James Moody, director de *The Electronic Small World Project*, quien busca entender cómo fluye la información a través del e-mail, cómo se conectan diferentes personas y cuán pequeño es realmente el mundo social en el que vivimos. Mientras que los fundamentalistas digitales insisten en que las relaciones on line pueden romper barreras raciales, sexuales y económicas (porque estas características no son obvias en Internet), cabe dudar mucho de que estas promesas se hayan cumplido. En cualquier caso, es presumible que el correo electrónico esté empequeñeciendo el mundo en algunos sentidos, dice Moody. Detrás de la empresa hay preguntas que nos son muy caras: cómo se diseminan los virus y cómo prevenir sus efectos destructivos; la búsqueda de mejores maneras de compartir y encontrar información, y sobre todo de descubrir el papel que el correo electrónico ha tenido en la transformación de la comunicación.

auto organizado. En este sentido, se convirtió en un «vivisistema» (en el sentido en que Kevin Kelly utilizó este término para definir aquellas entidades que no son del orden de lo construido, un artefacto, sino de lo nacido, lo viviente) y, como tal, su destino está atado a los diseños de origen, pero su evolución está determinada centralmente por lo que hagan de ella los usuarios finales. Como bien lo explicó en su momento Lawrence Lessig (2001), Internet se divide analíticamente en tres capas: la primera de ellas, en orden ascendente, es la capa de la infraestructura (computadoras, satélites, cables y hardware). Le sigue la capa lógica, o capa del código, aquella capa intermedia que concentra los componentes más blandos, o sea, el software y las aplicaciones que hacen funcionar una red de redes como Internet. Históricamente esta capa fue abierta, transparente y flexible, pero desde la década de 1980 su cierre parcial modificó la ecología informacional (al generar un sistema de códigos abiertos *versus* códigos cerrados). La capa superior de Internet está compuesta por los contenidos que circulan por las redes informacionales. Esta capa de contenidos, a diferencia de la capa lógica, ha comenzado un proceso de cierre a escala global desde mediados de la década de 1990. Lo que se promociona como lucha contra la piratería o defensa de los derechos de las discográficas, en rigor no es sino una discusión mucho más profunda acerca del acceso a la información como derecho universal, con múltiples variantes y consecuencias.

En lo que se refiere a las redes, poco importa que la riqueza inicial esté distribuida de forma equitativa. A poco de andar, las desigualdades se disparan y en un tiempo más corto que largo nos encontramos con distribuciones desiguales, tan o más fuertes que las que queríamos evitar anticipadamente con medidas radicales de redistribución. La responsabilidad de la creación de estas asimetrías no debe atribuirse a la psicología (más hobbesiana que rousseauiana), ni a los genes (más egoístas que altruistas), ni a la economía (más compensadora del *entrepreneurship* que del seguimiento de la manada). Se trata de otra cosa, y recién ahora empezamos a entender de qué se trata.

Más allá de las metáforas, y apelando a la memoria histórica, tenemos antecedentes conocidos donde estas distribuciones inequitativas están bien presentes: a fines del año 2004 existían en el mundo 63 millones de direcciones web. Sabemos que una cantidad mínima de ese vasto universo concentra innumerables visitas y que, a la inversa, la mayoría parece no existir. A esto le llamamos red libre de escala. Para avanzar en la explicación es interesante comentar que en una fecha casi prehistórica, abril de 1997, en el ya famoso sitio www.useit.com, Jakob Nielsen escribió un interesante trabajo, «Do Websites Have Increasing Returns?», anunciando que esto sucedería. Allí, Nielsen lo explicaba a partir de un tecnicismo: la ley de Zipf. Esta ley dice que las curvas se convierten en una línea recta cuando se las re- presenta en un diagrama logarítmico en ambos ejes.

Pero ¿qué es exactamente la ley de Zipf? Se trata de lo que ya analizamos en este libro como asimetría y que también suele llamarse *power law* o ley de Pareto. Los tres términos se usan para describir situaciones en las que los fenómenos extraordinarios son escasos, y donde los comunes abundan. George Kingsley Zipf, un lingüista de Harvard, buscó determinar la frecuencia de uso de la tercera u octava de las palabras más comunes en un texto en inglés. La ley de Zipf dice que el tamaño de la *n*-ésima aparición más grande del evento es inversamente proporcional a su rango. En síntesis, una descripción que sigue una distribución de Zipf supone:

- pocos elementos con un alto valor (la cola de la izquierda en los diagramas de campana);
- un número medio de elementos con valores intermedios (la parte media del diagrama);
- una gran cantidad de elementos que tienen un ranking bajísimo (la cola de la derecha en los diagramas).

Las distribuciones de Zipf están muy presentes en el lenguaje natural, pero las investigaciones de Bernardo Huberman y de Albert-László Barabási -además de los trabajos de pensadores originales Como Granovetter o de los hermanos Faloutsos, que generaron las líneas básicas para pensar las redes libres de escala- han demostrado que la web sigue una distribución de Zipf. Es decir, existe una distribución de Zipf para la popularidad de los sitios web y está expresada en el volumen de tráfico y en los links entrantes. Otra vez la conclusión es obvia

pero la demostración matemática de cómo y por qué es así tardó bastante en ser hallada. De hecho, era fácil presuponer que los sitios grandes movilizan un tráfico infinitamente mayor que los sitios pequeños. Un sitio que esté en el puesto número 100 en el ranking tendrá 10 veces más tráfico que un sitio que está en el puesto número 1.000 (en general, un sitio N obtiene M/N partes del tráfico del sitio M).

Tomemos como ejemplo un sitio especializado hipotético. Para simplificar el cálculo digamos que el total de sitios es de 40 millones, y que hay un sitio especializado por cada 20.000 sitios generalistas. A partir de estos números, se trataría de 2.000 sitios sobre una base reducida de 40 millones. Si esos 2.000 sitios están distribuidos equitativamente entre los 40 millones, el tráfico de los cinco sitios principales sería:

Rango del sitio en toda la web	Visitas (pageviews)	Rango del sitio en el tema especializado
# 20.000	10.000.000	1º más grande
# 40.000	5.000.000	2º más grande
# 60.000	3.333.333	3º más grande
# 80.000	2.500.000	4º más grande
# 100.000	2.000.000	5º más grande

Como el mismo Nielsen dice, su propio sitio Useit.com alcanza la considerable cifra de 10 millones de visitas por año, y se convierte así en el más popular de los sitios web del mundo en cuanto a su usabilidad. Dado que probablemente haya 2.000 sitios dedicados al tema, la tabla anterior exhibe tanto el consumo de atención sobre estos sitios como sobre cualquier otro grupo de sitios especializados. Ahora bien, la web en su conjunto recibirá este año aproximadamente unos 4 billones de visitas, de modo que la cantidad de visitas a cada sitio parece irrelevante. Sin embargo, en su nicho no se trata de poca cosa, ya que un sitio que en el conjunto de la web apenas alcanzaría el lugar 1.000.000, sería no obstante el quinto de su categoría, un lugar lo bastante destacado como para que mucha gente lo tuviera en cuenta.

Es claro que no se trata de números irrelevantes: lo que está en juego es saber cuánto más centralizada y concentrada se está volviendo la web. En definitiva, se trata de saber si alguien puede manejarla, controlarla u orientar las visitas (manipulando inclusive a buscadores como Google en esa dirección). Por eso, el conientario de Matthew Hindman y Kenneth Neil Cukier en el editorial abierto de *The New York Times*, «More News, Less Diversity», lleva a que miremos *con* mucho cuidado el supuesto y siempre bien declamado carácter libertario de la red.

Sin embargo, Nielsen es más optimista que estos críticos, ya que sostiene que al tomar la web en su conjunto, la diversidad puede estar asegurada. Insiste, con buen criterio, en que lo importante no es que existan sitios mucho más grandes que otros sobre ciertos temas, ya que esto siempre será así, por la plena vigencia de las leyes de fuerza y de los subconjuntos de Zipf. La pregunta que Nielsen nos invita a hacer es si siempre serán un puñado de sitios los que tendrán el dominio. Según él, esto no ocurre ahora y tampoco sucederá en el futuro.

Para reforzar su planteamiento, Nielsen realizó una breve búsqueda sobre siete temas diferentes. Identificó 59 sitios distintos entre las 70 entradas de los listados de la primera página mostrada por el buscador. Solo en el 16% de los casos se trataba de listados múltiples del mismo sitio, lo que nos indicaría que pocas fuentes estarían monopolizando el debate en Internet. Según él, aunque haya sitios inmensos, y aunque la mayoría de la gente nunca pase de la primera página de resultados del buscador, las preguntas que la gente se hace son tan variadas que en los listados habría poco solapamiento entre distintos temas.

Sin embargo, lo que sostiene Nielsen sólo es cierto hasta cierto punto. En efecto, está mezclando aquí cosas muy diferentes, es decir, dos públicos que no tienen nada que ver entre sí: sitios y públicos especializados con sitios y públicos generalistas. Para los primeros, lo que él dice es rigurosamente cierto y esta es una de las maravillas de la web.¹⁶ Según Nielsen ¿quién

¹⁶ Sugerimos leer al genial Clay Shirky en su excelente nota “Social Software and the Politics of Groups”, donde sostiene que Internet es la primera mesa redonda virtual de la historia. Un agregador de opiniones y personas distintas

no podría concordar con él sobre este punto?-, los sitios pequeños tienen dos ventajas sobre los grandes: son muchísimos y, al ser más especializados, concitan una atención más localizada. Los sitios pequeños le hablan a una comunidad convencida y esclarecida, por lo tanto las visitas que reciben son mucho más valiosas y ajustadas. Por ello, aunque los grandes sitios puedan ser *muy* grandes, los más pequeños siempre tendrán más hits preferenciales (en términos de conexión con los usuarios y de eventual valor comercial).

Puede ser. Pero esto no revierte el meollo del asunto, esto es: 1) la concentración de visitas generalistas en pocos sitios; 2) la imposibilidad de los sitios pequeños de tener un peso significativo en la construcción o desvío de las tendencias (en los sitios de interés general) y 3) al estar concentrados y permanecer cada vez más en manos de empresas comerciales de noticias (orientadas), hacen que la web pierda su valor supuestamente democratizador y revelan el carácter perverso de las leyes de fuerza.

Han pasado ya más de seis años desde que Nielsen hizo sus observaciones generales, ahora sabemos que esto también se aplica a los *weblogs*, de los cuales nos ocuparemos muy especialmente en los capítulos que siguen y que constituyen el centro de nuestro interés en este libro. Pero no debería sorprender a nadie que esto suceda, ya que a partir de un conjunto de Zipf se obtiene otro subconjunto que exhibe las mismas propiedades que el original.

Insistimos: aunque estos análisis y observaciones parecen abstrusos y teóricos, sus consecuencias prácticas y la especial conformación que le dan a la red los convierten en un chaleco de fuerza que condicionará para siempre la vida en la, pantalla.

Sitios web relacionados con los contenidos de este capítulo

- Tasas crecimiento Internet (se agregan por día 7,3 millones de páginas):
http://www.cyveillance.com/web/forms/request.asp?form-type=wp&wp_name=sizin~internet
- Tasas de crecimiento (por día se envían más de 31.000 millones de e-mails):
<http://www.spamfilterreview.com/spam-statistics.html>
- Deep web research: gran variedad de datos acerca de la web profunda:
<http://www.deepwebresearch.info>
- Los 1000 sitios más concurridos de alimentadores de noticias de Usenet:
<http://www.freenix.org/lreseaultop1000/>
- Resumen del estudio acerca de la producción mundial de información de la JSIMS (School of Information Management and Systems) de Berkeley:
<http://www.sims.berkeley.edu/how-much-info>
- Technorati (directorio con referencias a más de 5,3 millones de weblogs):
<http://www.technorati.com>
- The Electronic Small World Project. Mapeo de conexiones de personas a través de e-mails:
<http://smallworld.sociology.ohio-state.edu/html/homepage.html>
- Distribución paretiana de las palabras:
http://web.archive.org/web/120000818062828/http://hobart.cs.umass.edu/allan/cs646-f97_I_char_of_text.html .

sin parangón alguno con las innovaciones anteriores (por eso la importancia del software social y colaborativo que irá en aumento).