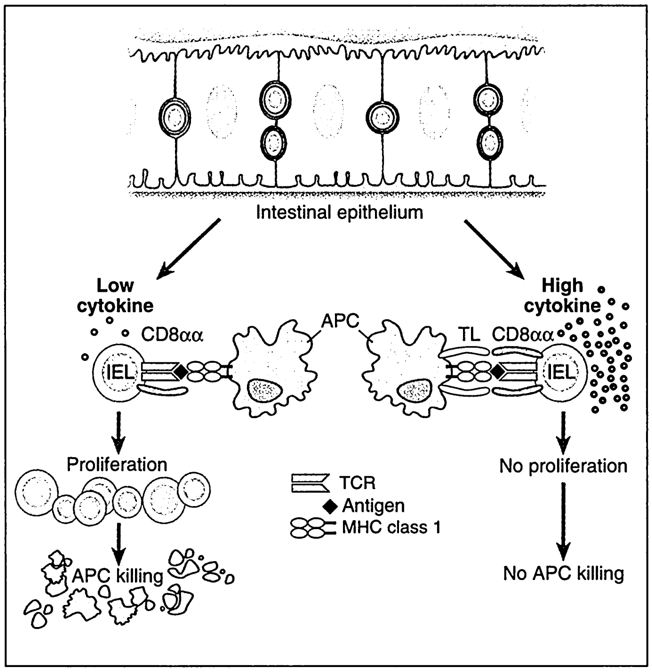
tor that considerably enhances TCR-mediated responses. The CD8 complex of these cells is a heterodimer comprising CD8α and CD8β chains. This heterodimer interacts with classical major histocompatibility (MHC) class I molecules, which are expressed by virtually all cells in the body. These MHC class I molecules present antigenic peptides to CD8αβ T cells. Most IELs, however, express a different CD8 complex—a CD8αα homodimer composed of two α chains. Leishman *et al.* demonstrate a high-affinity interaction between the CD8αα homodimer and an unusual (nonclassical) MHC class I molecule called thymus leukemia antigen (TL). The TL molecule has two interesting characteristics: It does not present antigenic peptides (in contrast to its classical MHC class I relatives), and it is expressed almost exclusively by epithelial cells of the small intestine (*7*). Strong interactions between CD8αα and TL enable IELs to interact directly and locally with the gut epithelium, but independently of antigen recognition and TCR specificity.



TL death us do part. Interactions between CD8αα and TL regulate the behavior of intraepithelial lymphocytes (IELs). Epithelial cells of the small intestine (yellow) express the TL molecule and are overlaid by a layer of mucus (pink). IELs (blue), localized among the gut epithelial cells, express CD8αα (red). **(Bottom, left)** If isolated IELs are stimulated by antigen-presenting cells (APCs) that express antigen but lack TL, they divide and kill target cells but secrete low amounts of cytokines. **(Bottom, right)** If APCs express both antigen and TL, IELs secrete high amounts of cytokines but do not divide and do not kill target cells.

What are the consequences of this interaction? Leishman *et al.* (*2*) compared IEL responses to antigen-presenting cells that did or did not express the TL molecule (see the figure). Surprisingly, they found that CD8αα-TL interactions could either enhance or suppress IEL responses. Such interactions considerably enhance cytokine release by IELs but inhibit their proliferation and cytotoxicity. These apparently paradoxical effects make a lot of sense in the particular environment of the small intestine. By inhibiting proliferation, CD8αα-TL interactions prevent IELs from dividing and disrupting the gut epithelium. In addition, by blocking T cell killer activity, these interactions prevent the elimination of healthy epithelium by self-reactive IELs (*2*). In contrast, by favoring interferon-γ production, the binding of CD8αα to TL may promote turnover of gut epithelium (*1*).

These results indicate that the small intestine and IELs have developed a unique way to control local homeostasis and to ensure continuous epithelial cell renewal. The mechanisms by which CD8αα-TL interactions induce such paradoxical effects on IEL responses remain to be discovered. Hints may come from certain types of inflammatory bowel disease that are associated with a deficiency in regulatory T lymphocytes, or overproduction of the inflammatory cytokine interleukin-10 (*8*). It is possible that in these disorders there is a severing of the interaction between CD8αα and TL. If so, then these diseases may yield valuable information about the maintenance of gut homeostasis.

References

1. D. Guy-Grand *et al.*, *Eur. J. Immunol.* **28**, 730 (1998).
2. A. J. Leishman *et al.*, *Science* **294**, 1936 (2001).
3. H. Saito *et al.*, *Science* **280**, 275 (1998).
4. B. Rocha, P. Vassalli, D. Guy-Grand, *J. Exp. Med.* **173**, 483 (1991).
5. B. Rocha, H. von Boehmer, D. Guy-Grand, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5336 (1992).
6. D. Masopust, V. Vezys, A. L. Marzo, L. Lefrancois, *Science* **291**, 2413 (2001).
7. R. Hershberg *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9727 (1990).
8. K. J. Maloy, F. Powrie, *Nature Immunol.* **2**, 816 (2001).

PERSPECTIVES: NETWORK ANALYSIS

The Structure of the Web

Jon Kleinberg and Steve Lawrence

In the span of a decade, the World Wide Web has grown from a small research project into a vast repository of information and a new medium of communication. Unlike other great networks of the past century—such as the electric power grid, the telephone system, or the highway and rail systems—the Web does not have an engineered architecture. Rather, it is a virtual network of content and hyperlinks, with over a billion interlinked "pages" created by the uncoordinated actions of tens of millions of individuals.

Because of the decentralized nature of its growth, the Web has been widely believed to lack structure and organization as a whole. Recent research, however, shows a great deal of self-organization. Analyses of the Web's network of hyperlinks have revealed an intricate structure that is proving to be valuable for organizing information, improving search methods, and understanding the Web in a broader technological and social context.

A recent study (*1*) indicates that the Web contains a large, strongly connected core in which every page can reach every other by a path of hyperlinks. This core contains most of the prominent sites on the Web. The remaining pages can be characterized by their relation to the core: Upstream nodes can reach the core but cannot be reached from it, downstream nodes can be reached from the core but cannot reach it, and "tendrils" contain nodes that can neither reach nor be reached from the core.

In fairly large snapshots of the Web, these four components—core, upstream, downstream, and tendril regions—have roughly comparable sizes. Moreover, the core is very compact: The shortest path from one page in the core to another involves 16 to 20 links on average, a "small-world" situation in which typical distances

J. Kleinberg is in the Department of Computer Science, Cornell University, Ithaca, NY 14853, USA. E-mail: kleinber@cs.cornell.edu S. Lawrence is in the NEC Research Institute, Princeton, NJ 08540, USA. E-mail: lawrence@research.nj.nec.com

are very small relative to the overall size of the system (*1*–*4*).

Also at a global level, studies have analyzed the distribution of hyperlinks among pages. Several studies have shown that the number of links to and from individual pages is distributed according to a power law over many orders of magnitude (*1*, *5*, *6*); the fraction of pages with n in-links is roughly $n^{-\alpha}$ for $\alpha \sim 2.1$.

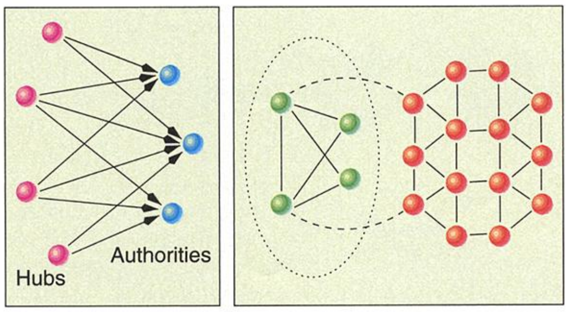
If the processes that drive Web growth are highly decentralized, then the power law must arise from a composite of local behavior. An appealing proposal, suggested independently in different forms (*5*, *7*, *8*), is the mechanism of preferential attachment. In this randomized, "rich-get-richer" process, the network grows by the sequential arrival of new nodes, and the probability that an existing node gains a link is proportional to the number of links it currently has. The result is a power law distribution of links.

It is thus plausible for a power law to arise through a simple mechanism. Nevertheless, we are far from a complete understanding of the processes governing Web growth. Deviations from power-law scaling occur, especially at small numbers of links (*1*). Furthermore, the deviation varies for different categories of pages (*9*). For example, the distribution of links to university home pages diverges strongly from a power law, following a far more uniform distribution. Recent models seek to improve on the accuracy of the original preferential attachment models (*9*, *10*).

At a local level—the scale of small neighborhoods and focused regions of the Web—the structure turns out to be even more intricate and quite nonuniform. Pages and links are created by users with particular interests, and pages on the same topic tend to cluster into natural "community" structures that exhibit an increased density of links.

Turning this observation around leads to a powerful method for analyzing the content of the Web. An unusually high density of links among a small set of pages is an indication that they may be topically related. A characteristic pattern in such communities consists of a collection of "hub" pages—guides and resource lists—linking in a correlated fashion to a collection of "authorities" on a common topic (see the left panel in the figure) (*11*). A related pattern is one in which authorities on a topic link directly to other authorities, again creating a density of links (*12*).

Link analysis as a means of finding authoritative, relevant sources on the Web has proven useful in the design of improved search engines (*12*, *13*). This application of link analysis has clear connections with, as well as interesting contrasts to, citation analysis of scientific literature and the identification of "central" individuals in a social network (*3*, *11*, *14*).



How is the Web organized? (Left) Web pages can be defined as hubs and authorities. A hub is a page that points to many authorities, whereas an authority is a page that is pointed to by many hubs (*11*). Characteristic patterns of hubs and authorities can be used to identify communities of pages on the same topic. **(Right)** An alternate method for identifying communities seeks a set of nodes for which the link density is greater among members than between members and the rest of the network (*15*).

Knowing the characteristic link structures that identify Web communities, one can examine a large snapshot of the Web for all occurrences of the link-based "signature" of a community. Using a signature corresponding to an interlinked collection of hubs and authorities, one large-scale study found over 100,000 coherent community structures; estimates based on sampling suggested that the overwhelming majority covered focused topics (*6*). The list included communities not considered by the creators of popular Web portals (for example, a community of people concerned with oil spills off the coast of Japan), showing that analysis of the Web's structure can help to define topics and social groupings of interest to its denizens.

A community can also be defined as a collection of pages in which each member page has more links to pages within the community than to pages outside the community (see the right panel in the figure) (*15*). This definition may be naturally extended to identify communities with varying levels of cohesiveness. Communities defined in this way are closely related to network flow computations, a powerful combinatorial technique designed for graph partitioning problems. As with the previous approach, this method of searching for communities reveals a remarkable degree of self-organization in the Web's link structure, and textual analysis of the communities shows that the constituent pages are topically related.

Analysis of the Web's structure is leading to improved methods for accessing and understanding the available information, for example, through the design of better search engines, automatically compiled directories, focused search services, and content filtering tools. Although researchers have been surprised at what can be discovered based solely on the structure of the Web, the integration of link- and content-based analysis will typically improve upon either method alone. Beyond these applications, the appearance of an increasing fraction of human knowledge and communication on the Web offers an unprecedented opportunity for charting and analyzing interests and relationships within society, as reflected in the Web's content and hyperlinks.

The migration of communication and commerce to the Web is also altering information flow in the world. We are only beginning to understand how link structure affects the visibility of Web sites. New or niche sites with few links to them may have difficulty competing with highly prominent sites for attention. By favoring more highly linked sites, search tools may increase this effect. But deeper analysis, exposing the structure of communities embedded in the Web, raises the prospect of bringing together individuals with common interests and lowering barriers to communication.

References and Notes

1. A. Broder *et al.*, in *Proceedings of the Ninth International World Wide Web Conference* (Elsevier, Amsterdam, 2000), pp. 309–320.
2. R. Albert, H. Jeong, A.-L. Barabási, *Nature* **401**, 130 (1999).
3. S. Wasserman, K. Faust, *Social Network Analysis* (Cambridge Univ. Press, Cambridge, 1994).
4. D. Watts, S. Strogatz, *Nature* **393**, 440 (1998).
5. A.-L. Barabási, R. Albert, *Science* **286**, 509 (1999).
6. R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, in *Proceedings of the Eighth International World Wide Web Conference* (Elsevier, Amsterdam, 1999), pp. 403–415.
7. B. Huberman, L. Adamic, *Nature* **401**, 131 (1999).
8. R. Kumar, P. Raghavan, S. Rajagopalan, A. Tomkins, in *Proceedings of the IEEE Symposium on Foundations of Computer Science* (IEEE Computer Society Press, Los Alamitos, CA, 2000), pp. 57–65.
9. D. M. Pennock, C. L. Giles, G. W. Flake, S. Lawrence, E. Glover, *Winners Don't Take All: A Model of Web Link Accumulation* (Technical Report 2000-164, NEC Research Institute, Princeton, NJ, 2000).
10. R. Albert, A.-L. Barabási, *Phys. Rev. Lett.* **85**, 5234 (2000).
11. J. Kleinberg, in *Proceedings ACM-SIAM Symposium on Discrete Algorithms* (ACM-SIAM, New York/Philadelphia, 1998), pp. 668–677.
12. S. Brin, L. Page, in *Proceedings of the Seventh International World Wide Web Conference* (Elsevier, Amsterdam, 1998), pp. 107–117.
13. S. Chakrabarti *et al.*, *IEEE Computer* **32**, 60 (1999).
14. L. Egghe, R. Rousseau, *Introduction to Informetrics* (Elsevier, Amsterdam, 1990).
15. G. W. Flake, S. Lawrence, C. L. Giles, F. Coetzee, *IEEE Computer*, in press.

16. J. K. is supported in part by grants from the NSF, the Office of Naval Research, and the Packard Foundation.