

PAUL BARAN, NETWORK THEORY, AND THE PAST, PRESENT, AND FUTURE OF THE INTERNET

CHRISTOPHER S. YOO*

Paul Baran's seminal 1964 article "On Distributed Communications Networks" that first proposed packet switching also advanced an underappreciated vision of network architecture: a lattice-like, distributed network, in which each node of the Internet would be homogeneous and equal in status to all other nodes. Scholars who have embraced the concept of a flat network have largely overlooked the extent to which it is inconsistent with network theory, which emphasizes the importance of short cuts and hubs in enabling networks to scale. The flat network is also inconsistent with the actual way the Internet was deployed, which relied on a three-tiered, hierarchical architecture resembling what Baran called a decentralized network. However, empirical studies reveal that the Internet's architecture is changing; as large content providers build extensive wide area networks and undersea cables to connect directly to last-mile networks, it is in the process of becoming flatter and less hierarchical. This change is making the network become more centralized rather than more distributed. As a result, this article suggests that the standard reference model that places backbones at the center of the architecture may need to be replaced with a radically different vision: a stack of centralized star networks, each centered on one of the leading content providers.

INTRODUCTION.....	162
I. BARAN'S UNREALIZED VISION OF A DISTRIBUTED NETWORKS	164
A. <i>Baran's Vision of a Distributed Network</i>	164
B. <i>The Partial Reception of Baran's Vision</i>	167
II. NETWORK SCIENCE'S INSIGHTS INTO THE BENEFITS OF NETWORKS.....	172
A. <i>Duncan Watts and Small-World Networks</i>	173
B. <i>Albert-László Barabási and Scale-Free Networks</i>	176

* John H. Chestnut Professor of Law, Communication, and Computer & Information Science and Founding Director of the Center for Technology, Innovation and Competition, University of Pennsylvania.

C. <i>Implications for the Distributed Network</i>	178
III. THE EVOLUTION OF THE MODERN INTERNET	179
A. <i>The Rise of the Client-Server Architecture and the Decline of Peer-to-Peer</i>	180
B. <i>The Flattening of the Network Architecture</i>	182
CONCLUSION	185

INTRODUCTION

Paul Baran’s article “On Distributed Communications Networks”¹ represents a landmark in the history of the Internet that remains underappreciated by the legal academy.² It articulated a bold new approach to communications networking that would serve as one of the foundations for the Internet. Baran’s proposal was complex, but its essence can be distilled into three basic principles. First, to make the system robust against physical attack, all nodes would be connected to the network by multiple links.³ Second, communications would be divided into individual packets that would travel independently through the network and be reassembled once they reached their destinations.⁴ Third, to take advantage of multiple links and to eliminate the vulnerability associated with a single centralized point of control, each node

1. Baran first presented his ideas to the Air Force in RAND Briefing B-265 in 1961. He formalized his ideas in a paper in 1962. PAUL BARAN, ON DISTRIBUTED COMMUNICATIONS NETWORKS (1962) [hereinafter BARAN RAND Paper], <https://www.rand.org/pubs/papers/P2626.html> [<https://perma.cc/6NRR-686K>]. This initial paper was expanded into a series of eleven research memoranda detailing Baran’s proposed system. The first research memorandum was published in 1964. Paul Baran, *On Distributed Communications Networks*, 12 IEEE TRANSACTIONS ON COMM. SYS. 1 (1964). All eleven research memoranda are available from the RAND Corporation’s website. *Paul Baran*, RAND CORP., https://www.rand.org/pubs/authors/b/baran_paul.html [<https://perma.cc/A9B4-CGJ4>].

2. A Westlaw search reveals only five citations to Baran’s article. I am aware of only two law review articles that reproduce Baran’s classic figure; of these, one applies the principles to securities law, and the other deletes the decentralized network that ends up being the most important of the three network architectures. See Patrick S. Ryan, *Wireless Communications and Computing at a Crossroads: New Paradigms and Their Impact on Theories Governing the Public’s Right to Spectrum Access*, 3 J. ON TELECOMM. & HIGH TECH. L. 239, 262 fig.1 (2005); Will Schildknecht, Comment, *Designing for Robustness: Overcoming Systemic Risk in the Political Branches*, 103 CALIF. L. REV. 433, 440 fig.1 (2015). Baran’s diagram is occasionally reproduced in histories of the Internet. *E.g.*, KATIE HAFNER & MATTHEW LYON, *WHERE WIZARDS STAY UP LATE: THE ORIGINS OF THE INTERNET* 58 (1996). It also appears in leading books on network science. *E.g.*, ALBERT-LÁSZLÓ BARABÁSI, *LINKED: HOW EVERYTHING IS CONNECTED TO EVERYTHING ELSE AND WHAT IT MEANS FOR BUSINESS, SCIENCE, AND EVERYDAY LIFE* 145 fig.11.1 (2002); MARK BUCHANAN, *NEXUS: SMALL WORLDS AND THE GROUNDBREAKING THEORY OF NETWORKS* 79 fig.7 (2002).

3. BARAN RAND Paper, *supra* note 1, at 2.

4. *Id.* at 23–27.

would route traffic independently on a simple store-and-forward basis.⁵

Though controversial when initially proposed, the second and third principles ultimately laid the basis for the packet switching method that would eventually become one of the Internet's foundational technologies.⁶ The first principle, on the other hand, proved to be less influential—it was neither incorporated into the design of the Advanced Research Projects Agency Network (ARPANET) nor the subsequent Internet.⁷

The misconception that the Internet represents a lattice-like distributed network continues to resonate in many quarters. It is sometimes erroneously presented as the architecture of the current Internet.⁸ More recently, the supposed historical provenance of the distributed architecture has sometimes been invoked as validation for wireless mesh networks and other technologies that employ similar topologies.⁹ Even more problematically, emerging research investigating whether the Internet's architecture is flattening often mischaracterizes the networks as becoming more mesh-like.¹⁰

This Article aims to give Baran's first principle its proper due by acknowledging both its contributions and its limitations. First, this Article rebuts claims that the Internet, at its inception, incorporated a distributed topology by analyzing those arguments in a historical context. Second, this Article draws on burgeoning research in network science to explain the critical role that hierarchical network structures play in ensuring that the network can operate in a scalable manner. Third, this Article reviews recent developments in the Internet's architecture to demonstrate how the

5. *Id.* at 27–30.

6. JANET ABBATE, *INVENTING THE INTERNET* 39, 226 n.38 (1999).

7. *Id.* at 39–40 (noting that elements of Baran's system “that were specifically adapted to the Cold War Threat, including very high levels of redundancy” were not adopted by Davies or by the ARPANET, while other features, such as “adaptive routing, and efficient packet switching” were added to later systems); *see also* BARABÁSI, *supra* note 2, at 144, 147 (noting that “the topology of today's Internet has little to do with [Baran's distributed] vision” and that “[i]ronically, the principles directing today's Internet match Baran's original vision in every respect except the guiding principle that motivated his work: undercutting vulnerability to attacks”).

8. *See* BARRIE SOSINSKY, *NETWORKING BIBLE* 456 (2009) (“The Internet was designed to be a highly redundant mesh structure that could survive any outages to a substantial portion of the network and still be operable.”); *see also* Ryan, *supra* note 2, at 264 (“On the right side of Figure 1, the ‘distributed’ graphic depicts the way in which the Internet operates today.”).

9. *See* Yochai Benkler, *Some Economics of Wireless Communications*, 16 *HARV. J.L. & TECH.* 25, 75 (2002) (“The basic characteristic of the Internet's robustness—its redundancy and decentralized architecture—is replicated in open wireless networks at the physical layer of the communications infrastructure.”).

10. *See, e.g.*, Stanley M. Besen & Mark A. Israel, *The Evolution of Internet Interconnection from Hierarchy to “Mesh”: Implications for Government Regulation*, 25 *INFO. ECON. & POL'Y* 235, 235 (2013); Amogh Dhamdhere & Constantine Dovrolis, *The Internet Is Flat: Modeling the Transition from a Transit Hierarchy to a Peering Mesh*, 6 *ACM CONEXT* art. 21, at 2 (2010).

network is flattening.¹¹ A careful analysis of the evidence about average path length reveals that the network is becoming increasingly centralized rather than more distributed.

I. BARAN'S UNREALIZED VISION OF A DISTRIBUTED NETWORKS

This Part describes Baran's innovative reconceptualization of how to design a communications network and analyzes the extent to which his ideas were received. Section A lays out the motivation driving Baran's vision and the principles underlying it. Section B explains why those ideas were not incorporated into the actual Internet.

A. Baran's Vision of a Distributed Network

Baran's motivation for creating his revolutionary approach to communications networking was driven by the dominant issue of the 1960s—the Cold War and the overhanging threat of nuclear annihilation.¹² America's nuclear capability depended on maintaining what the strategic defense community called “minimum essential communication,” which was the amount of connectivity needed for the U.S. to credibly maintain the threat of mutually assured destruction, which most observers believed was essential to deterring the Soviet Union from attempting a preemptive strike.¹³ Unfortunately, the existing communications infrastructure in the U.S. was dangerously fragile. Military leaders and analysts were concerned that an attack could cripple the country's communications capabilities to the point where the President could not order a counterstrike.¹⁴

Accordingly, Baran focused his efforts on improving the survivability of the communications network.¹⁵ His principal mechanism for doing so was redundancy.¹⁶ The choice, as Baran framed it, came down to three possible network architectures: a *centralized network*, a *decentralized network*, and a *distributed network*, as depicted in Figure 1.¹⁷

11. This analysis extends earlier research published in the formerly named version of this journal. Christopher S. Yoo, *Innovations in the Internet's Architecture that Challenge the Status Quo*, 8 J. ON TELECOMM. & HIGH TECH. L. 79 (2010).

12. HAFNER & LYON, *supra* note 2, at 55–56 (noting that Baran felt “the problem of building a more stable communications infrastructure . . . was the most important work he could be doing” and quoting Baran as saying that his work “was done in response to the most dangerous situation that ever existed”).

13. *Id.* at 58.

14. *Id.* at 55.

15. *Id.* at 54–57; ABBATE, *supra* note 6, at 10–11.

16. ABBATE, *supra* note 6, at 11.

17. See *infra* Figure 1. Interestingly, some discussions omit the intermediate decentralized example. See, e.g., Ryan, *supra* note 2, at 262 fig.1.

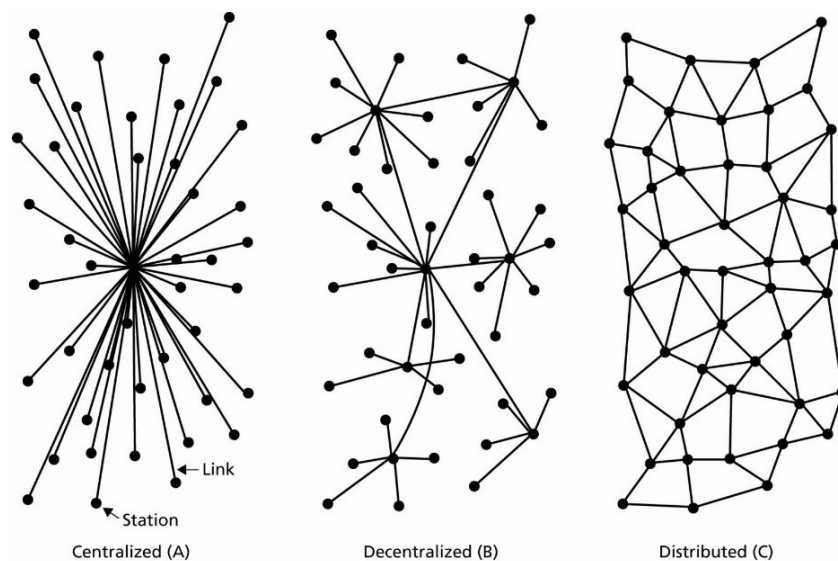


FIGURE 1: BARAN'S CENTRALIZED, DECENTRALIZED, AND DISTRIBUTED NETWORKS¹⁸

The centralized network is a classic star network, in which the peripheral nodes connect directly to a central node.¹⁹ The fact that destruction of the central node would disconnect every node in the system leaves this network extremely vulnerable to attack.²⁰ The decentralized network is comprised of a number of smaller stars connected to form a larger star, with an added link connecting the two stars at the top of the graph to form a loop.²¹ This network architecture represents the then-current structure of the U.S. telephone network.²² Although more resilient than a centralized network to attack, a decentralized network can still be disabled by an attack on a small number of nodes.²³ The distributed network, on the other hand, is more resilient. It is a lattice-like architecture in which the nodes are undifferentiated—they are all functionally similar and have a similar number of links to other nodes. Or, to use the language that will emerge later in this Article, the nodes have a similar enough number of connections to other nodes so that the network has a characteristic scale.²⁴

18. BARAN RAND Paper, *supra* note 1, at 4 fig.1.

19. *See supra* Figure 1, illus. (A).

20. BARAN RAND Paper, *supra* note 1, at 3.

21. *See supra* Figure 1, illus. (B).

22. Juan D. Rogers, *Internetworking and the Politics of Science: NSFNET in Internet History*, 14 INFO. SOC'Y 213, 219 (1998).

23. BARAN RAND Paper, *supra* note 1, at 3.

24. BARABÁSI, *supra* note 2, at 70.

For Baran, distributed networks rely on redundancy to increase the likelihood that two particular nodes could maintain contact with one another should other nodes in the network fail.²⁵ Baran's model indicates that connecting each node to only three or four others would be sufficient for a distributed network to approach the maximum theoretical level of survivability.²⁶

Baran's other principles followed as corollaries from his basic commitment to redundancy. For example, the fact that communications in a distributed network have to traverse an arbitrary number of hops meant that the network must necessarily be digital, since analog signals degrade every time they pass from one link to another.²⁷ Furthermore, redundant links would not make the network more resilient unless each of the nodes were able to make independent routing decisions and reroute traffic autonomously in the face of partial network failure.²⁸ That his design depended on nodes having sufficient intelligence to make their own routing decision meant that each would have to consist of a low-cost, unmanned computer.²⁹

Baran's was not the first design to use a distributed architecture to increase survivability, but it was the first to use both unmanned, decentralized switching and an all-digital network with substantial intelligence concentrated in the nodes.³⁰ Rather than attempting to maximize the reliability of each network element, Baran's design relied on redundancy to compensate for any failures. Specifically, it used a large number of cheap, autonomous nodes to reinforce network survivability.³¹

Baran's design represented a radical departure from the traditional approach to networking, which relied on centralized control and hardening of key locations to ensure reliability and survivability.³² In the words of RAND's history commemorating its work with the Air Force, Baran's network was "unusual" in that each node had "equal status" and was capable of receiving, routing, and transmitting information autonomously.³³

Although Baran's proposal yielded substantial economic benefits, enhancing survivability represented the heart of his

25. BARAN RAND Paper, *supra* note 1, at 3–14.

26. *Id.* at 3–5.

27. ABBATE, *supra* note 6, at 16; HAFNER & LYON, *supra* note 2, at 57.

28. ABBATE, *supra* note 6, at 13.

29. *Id.* at 16; HAFNER & LYON, *supra* note 2, at 61.

30. ABBATE, *supra* note 6, at 15.

31. *Id.* at 17.

32. *Id.*

33. Willis Ware, *Computing*, in 50TH PROJECT AIR FORCE 1946–1996, at 37 (1996), <https://www.rand.org/content/dam/rand/www/external/publications/PAFbook.pdf> [<https://perma.cc/79WG-BWV2>].

approach.³⁴ Indeed, other aspects of Baran's design underscore his preference for survivability over economic benefits. For example, Baran proposed that the nodes be located far away from population centers and be built with substantial excess capacity.³⁵ He also advocated that the network incorporate both cryptography to ensure security, and a priority system to allow messages from high-level users to take precedence over those from lower-level users.³⁶

B. *The Partial Reception of Baran's Vision*

Despite Baran's best efforts, his network was never built.³⁷ Although in August 1965 he convinced the Air Force to construct a system along the lines he proposed, the Department of Defense (DOD) reassigned responsibility for building and running the network to the Defense Communications Agency, which lacked any expertise in digital technology.³⁸ Rather than see such inexperienced hands mismanage the network, Baran and his Air Force supporters opted to kill the project, and Baran moved on to other matters.³⁹ The eventual construction of a network based on Baran's principles would have to wait until 1969, when the DOD created the ARPANET.⁴⁰

Meanwhile, across the Atlantic, UK researcher Donald Davies independently developed the idea of packet switching shortly after Baran.⁴¹ Although Davies retrospectively acknowledged Baran's contributions,⁴² his primary motivation differed from Baran's. Davies sought to improve the economic efficiency of the network, not enhance its survivability in the face of a nuclear attack.⁴³ Toward that end, he rejected the high levels of redundancy that Baran envisioned in favor of a more modest approach that integrated, among other things, a time-sharing principle.⁴⁴

Although Davies's attempt to construct his network also failed for lack of funding,⁴⁵ he was able to introduce both his and Baran's work to Lawrence Roberts, who was in charge of building the ARPANET and had the budget to construct a network that

34. See ABBATE, *supra* note 6, at 18.

35. *Id.* at 11.

36. *Id.*

37. *Id.* at 21.

38. *Id.* at 20.

39. *Id.* at 20–21.

40. See *id.* at 21; HAFNER & LYON, *supra* note 2, at 9.

41. ABBATE, *supra* note 6, at 8.

42. Interestingly, it was Davies who coined the terms *packets* and *packet switching*. Baran instead referred to *message blocks* and *distributed adaptive message block switching*. Davies would later say to Baran, “[w]ell, you may have got there first, but I got the name.” HAFNER & LYON, *supra* note 2, at 67.

43. *Id.* at 8–10, 27.

44. ABBATE, *supra* note 6, at 27–29.

45. *Id.* at 29–33.

incorporated new principles.⁴⁶ However, Roberts's goals, like Davies', differed from Baran's, which caused him to adopt only part of Baran's design.⁴⁷ Most importantly, Roberts, like Davies, did not share Baran's focus on survivability,⁴⁸ as that was much less of a concern in nonmilitary contexts, such as research and commercial deployment.⁴⁹ Instead, Roberts's primary goal was to interconnect the several major research computing centers and the time-sharing computers that the Advanced Research Projects Agency (ARPA) was funding.⁵⁰ Thus, Baran's ideas about survivability "were all but ignored by the military . . . [and] the topology of today's Internet has little to do with his vision."⁵¹

Although Roberts rejected the high levels of redundancy, reliability, and excess capacity that Baran's distributed network required, he did adopt Baran's emphasis on decentralized autonomous routing and packet switching.⁵² Indeed, the primary goals underlying the ARPANET's design were to keep average delay time below 0.2 seconds and decrease cost; reliability, along with flexibility and expandability, were secondary concerns.⁵³ In addition, under the ARPANET's design, reliability was defined as requiring at just two nodes to fail before the network would become disconnected⁵⁴—a stark contrast to Baran's focus on designing a network that could continue to operate even if as many as half of its nodes were destroyed.

Figure 2, a historical representation of the ARPANET design from 1969 to 1977, depicts the manifestation of Robert's work when it had reached maturity.⁵⁵ A casual look at Figure 2 shows that the vast majority of nodes were connected by only two links,⁵⁶ and a more in depth historical analysis confirms that between 1969 and

46. *Id.* at 37, 39–41.

47. HAFNER & LYON, *supra* note 2, at 79–80.

48. *Id.* at 77.

49. See ABBATE, *supra* note 6, at 27; see also David D. Clark, *The Design Philosophy of the DARPA Internet Protocols*, 18 *COMPUTER COMM. REV.* 106, 106 (1988) (explaining that "[a]n architecture primarily for commercial deployment," unlike an architecture designed to operate in a military context, would make accountability a top priority and survivability a much lower priority).

50. See ABBATE, *supra* note 6, at 44–46.

51. BARABÁSI, *supra* note 2, at 144; accord CHRISTOS J.P. MOSCHOVITIS ET AL., *HISTORY OF THE INTERNET: A CHRONOLOGY, 1843 TO THE PRESENT* 35 (1999) ("Although Baran's work was extremely influential on the ARPAnet founders, his imagined network never came to fruition.").

52. ABBATE, *supra* note 6, at 39, 226 n.38.

53. H. Frank et al., *Topological Considerations in the Design of the ARPA Computer Network*, in 36 *AFIPS CONFERENCE PROCEEDINGS: 1970 SPRING JOINT COMPUTER CONFERENCE* 581, 581 (1970).

54. *Id.*; see also ABBATE, *supra* note 6, at 58 (explaining "the performance constraints specified by ARPA . . . [included] a minimum of two links per IMP (for reliability)").

55. See *infra* Figure 2.

56. *Id.*

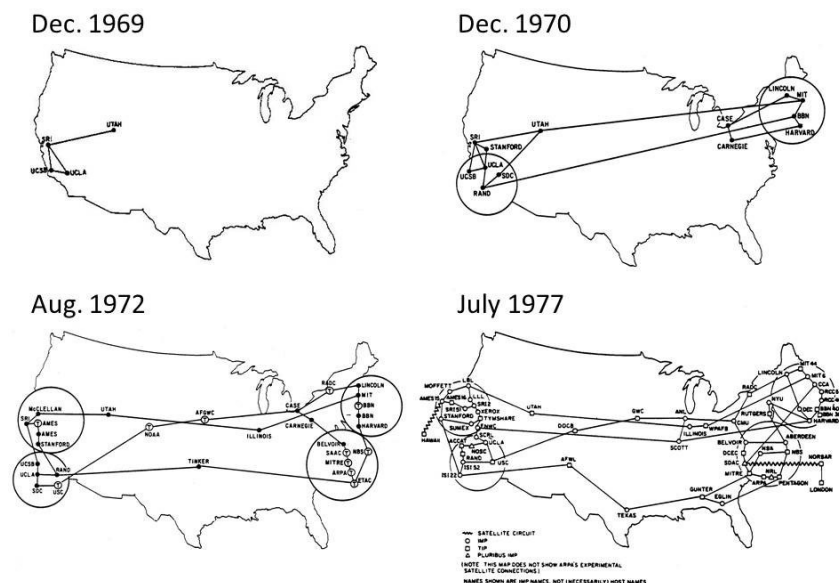


FIGURE 2: TOPOLOGY OF THE ARPANET⁵⁷

1977, the average number of links connected to the ARPANET nodes ranged from only 2.00 to 2.46,⁵⁸ a far cry from Baran's vision of highly redundant connections that would permit the network to operate effectively, even if it lost as many as half of its nodes.⁵⁹

In 1986, the National Science Foundation (NSF) began to deploy the NSFNET in order to provide remote access to supercomputing centers.⁶⁰ This time, however, the federal government funded the project to fend off the danger of losing out to Japan in supercomputing leadership.⁶¹ By the end of 1989, the NSFNET had replaced the ARPANET.⁶²

Unlike the ARPANET, university consortia or state-university partnerships often helped operate the NSFNET's network through regional connections, which provided connectivity from distant locations to the NSFNET backbone's interconnection points.⁶³ In

57. F. HEART ET AL., A HISTORY OF THE ARPANET: THE FIRST DECADE, at III-79 fig.5, III-81 fig.7, III-84 fig.10, III-89 fig.15 (1978), <http://www.walden-family.com/bbn/arpamet-completion-report.pdf> [<https://perma.cc/B9V5-DYFV>].

58. *Id.* at III-91 fig. 16.

59. PAUL BARAN, SOME REMARKS ON DIGITAL DISTRIBUTED COMMUNICATIONS NETWORKS 5 (1967), <https://www.rand.org/pubs/papers/P3536.html> [<https://perma.cc/M5BX-KXLF>].

60. Rogers, *supra* note 22, at 217.

61. *Id.*

62. HAFNER & LYON, *supra* note 2, at 255–56.

63. CHRISTOPHER S. YOO, THE DYNAMIC INTERNET: HOW TECHNOLOGY, USERS, AND BUSINESSES ARE TRANSFORMING THE NETWORK 57 (2012).

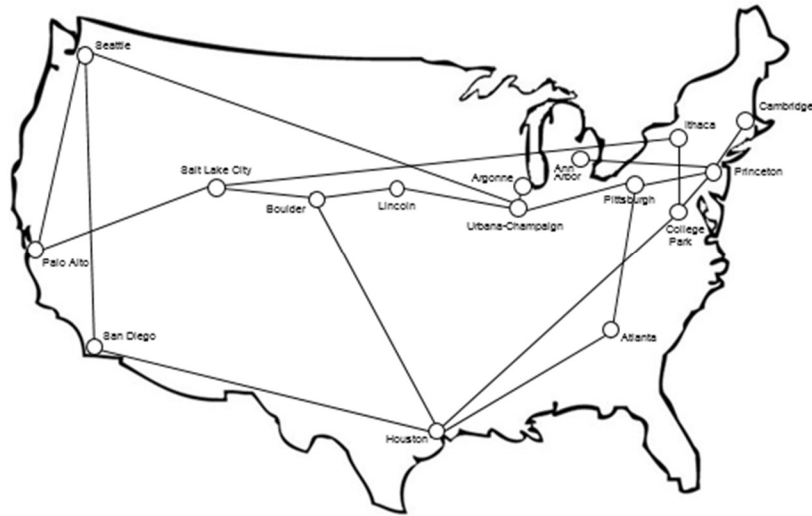


FIGURE 3: THE TOPOLOGY OF THE NSFNET BACKBONE⁶⁴

in addition to being linked together, each of the supercomputing centers became the hub for a web connecting local campus networks to other universities.⁶⁵ Notwithstanding the operational differences, however, the topology of the NSFNET's network ultimately looked similar to that of the ARPANET's.⁶⁶ The average number of links connected to each supercomputer center was only 2.38.⁶⁷ Furthermore, the NSFNET backbone, portrayed in Figure 3, is functionally equivalent to the network of nodes connected by thicker white lines that appear at the top of Figure 4. The main difference is that they are placed with only a slight slant backward. The links to local campuses are depicted by the thinner white lines emanating from each of those nodes.

The NSFNET's topology required every packet traveling through the network to traverse a series of three-level hierarchies.⁶⁸ Specifically, a campus network would hand off the traffic it was originating to its regional network.⁶⁹ The regional network would then hand the traffic off to the NSFNET backbone.⁷⁰ The backbone

64. Yoo, *supra* note 11, at 82 fig.1.

65. *Id.*

66. Compare *supra* Figure 2, with *supra* Figure 3.

67. Calculated from Figure 3, *supra*.

68. See *infra* Figure 5.

69. YOO, *supra* note 63.

70. *Id.*

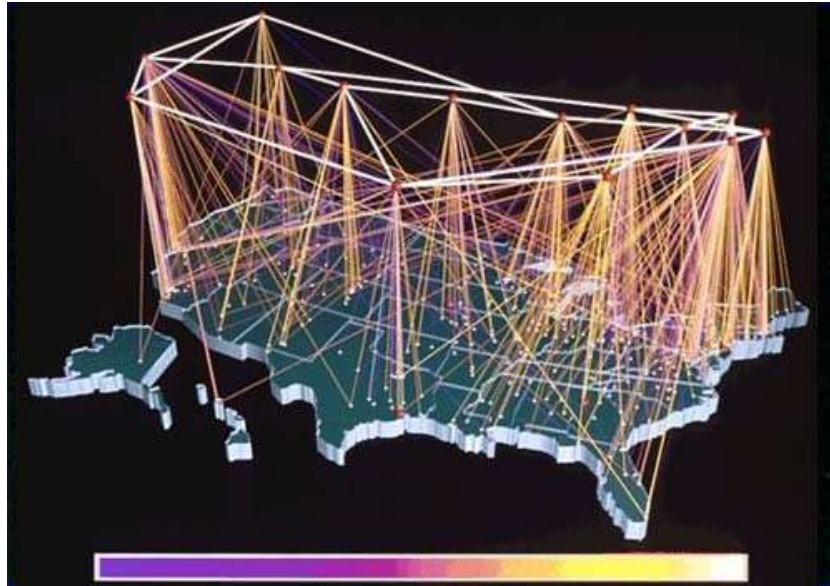


FIGURE 4: THE TOPOLOGY OF THE NSFNET'S REGIONAL NETWORKS⁷¹

would then route the traffic to the appropriate regional network.⁷² That regional network would then pass the packets along to the campus network that was the ultimate destination.⁷³ This routing architecture permitted each campus network to connect to only a single regional network.⁷⁴

The federal government retained this three-tiered hierarchy when it privatized the NSFNET in the mid-1990s.⁷⁵ The result: the NSFNET backbone was replaced by a group of Tier-1 Internet Service Providers (ISPs)—also called backbones—interconnecting four network access points.⁷⁶ The regional networks were replaced by Tier-2 ISPs—also called regional ISPs.⁷⁷ The campus networks were replaced by Tier-3 ISPs—also called last-mile providers.⁷⁸ Figure 6 shows that the privatized network can more effectively be illustrated through concentric circles, where the core of the network is an amalgamation of interconnected backbones, rather than through the more vertical configuration that represented the NSFNET's three-tiered network, as depicted in Figure 5.

71. KAREN D. FRAZER, NSFNET: A PARTNERSHIP FOR HIGH-SPEED NETWORKING, FINAL REPORT 1987–1995, at 4 fig.1 (1996).

72. YOO, *supra* note 63.

73. *Id.*

74. *Id.*

75. *Id.* at 82.

76. *Id.*

77. *Id.*

78. *Id.*; see also *infra* Figure 6.

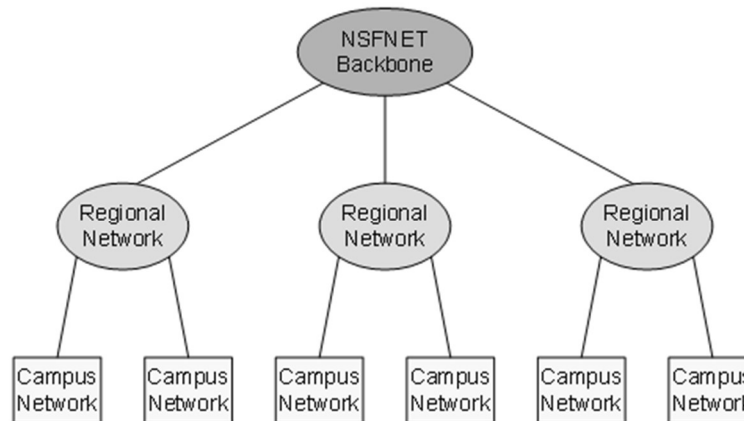


FIGURE 5: THE NSFNET THREE-TIERED NETWORK ARCHITECTURE⁷⁹

Comparing the concentric network in Figure 6 to the three networks that Baran laid out demonstrates that the architecture of the NSFNET's privatized network bears greater similarity to the decentralized network that Baran disfavored rather than the distributed network that he preferred.⁸⁰ This three-tiered hierarchy has become the standard reference model taught in every computer networking textbook.⁸¹

II. NETWORK SCIENCE'S INSIGHTS INTO THE BENEFITS OF NETWORKS

As a historical matter, neither the ARPANET nor the NSFNET embraced Baran's vision of a distributed, lattice-like network.⁸² Consequently, neither did the Internet.⁸³ Far from being a mere matter of path dependence, the burgeoning field of network science, pioneered by scholars such as Duncan Watts, Albert-László Barabási, and others, has instead provided a present-day justification for adopting the decentralized, hierarchical Internet structure.

79. Yoo, *supra* note 11, at 83 fig.2.

80. Compare *infra* Figure 6, with *supra* Figure 1, illus. (B).

81. See, e.g., Dhamdhere & Dovrolis, *supra* note 10, at 1.

82. See *supra* notes 43–44 and 47–49 and accompanying text.

83. See *supra* notes 75–81 and accompanying text.

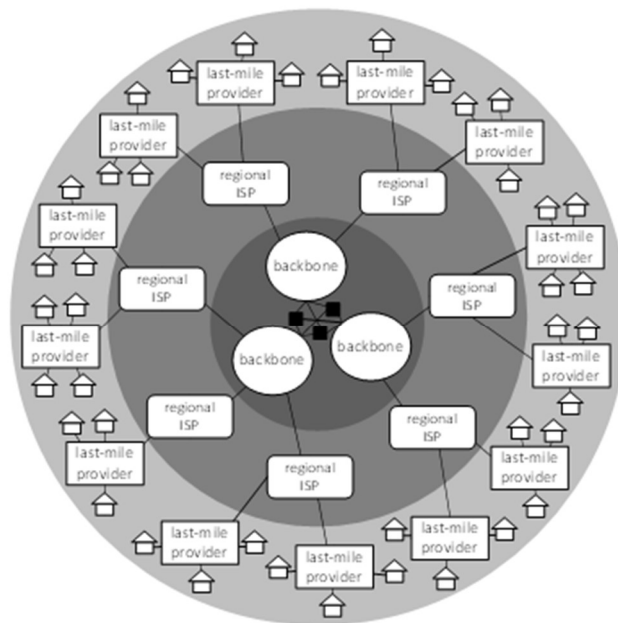


FIGURE 6: THE NETWORK HIERARCHY DEPICTED AS A SERIES OF CONCENTRIC RINGS⁸⁴

A. Duncan Watts and Small-World Networks

One of the most enigmatic principles in network science is commonly known as the small-world phenomenon.⁸⁵ Specifically, common sense would suggest that as the size of a network increases, the number of links that one would need to traverse in order to connect two arbitrary nodes should increase as well. However, the works of Stanley Milgram and others have suggested otherwise. In his famous study, Milgram tested how many links were necessary to connect a randomly selected person—or “node”—in Kansas or Nebraska to a specific person in the Boston area.⁸⁶ Despite speculation that such a connection might require 100 or more links,⁸⁷ Milgram found that the chains connecting these two

84. YOO, *supra* note 63, at 59 fig.4-4.

85. Stanley Milgram, *The Small-World Problem*, PSYCHOL. TODAY, May 1967, at 61.

86. Milgram’s initial experiment measured how many links it took to connect volunteers in Wichita, Kansas, to the wife of a divinity school student living in Cambridge, Massachusetts. *Id.* at 64. Milgram’s second experiment, conducted with Jeffrey Travers, measured how many links it took to connect volunteers in Omaha, Nebraska, to a stockbroker working in Boston and living in Sharon, Massachusetts. *Id.*; see also Jeffrey Travers & Stanley Milgram, *An Experimental Study of the Small World Problem*, 32 SOCIOLOGY 425, 429–30 (1969).

87. Milgram, *supra* note 85, at 65.

people traversed an average of only 5.7 links,⁸⁸ a finding that was later popularized by the play, *Six Degrees of Separation*, which explored the notion that any two people in the world can be connected through fewer than six other people.⁸⁹ In addition, Milgram found that certain key intermediaries—most notably in the case of a clothing merchant—played an outsized role in establishing the connection between Nebraska and Boston.⁹⁰

Duncan Watts, initially with Steven Strogatz in their landmark article, *Nature*, and later in his book, *Six Degrees*, sought to further synthesize the principles underlying the small-world phenomenon by providing an explanation of how the random introduction of a small number of long-distance connections can allow highly clustered networks to still retain short path lengths.⁹¹ Taking a cue from Paul Erdős and Alfred Rényi, Watts considered at one extreme those networks where the links between the nodes are completely random.⁹² Random networks are characterized by two notable qualities. First, because long-distance and short distance links appear with the same probability, random networks exhibit very low levels of local clustering.⁹³ Second, the presence of long-distance links means that the typical path will grow relatively slowly with the size of the network, with the average path length rising logarithmically with the number of nodes.⁹⁴

These random networks stand in stark contrast to both Baran's distributed network and other lattice-like networks, in which every node is connected to each of its immediate neighbors. Indeed, highly ordered, lattice-like networks exhibit different qualities than random networks. First, because their nodes are connected exclusively by short distance links to their immediate neighbors, traditional lattice-like networks exhibit a high level of clustering.⁹⁵ Second, the absence of long-distance links means that the average path length will grow linearly with the size of the network, which is much faster than logarithmic growth.⁹⁶ Finally, the longer average path length associated with lattice-like networks makes them inherently *large-world networks*; specifically, the number of links needed to travel between two randomly chosen nodes in the

88. Travers & Milgram, *supra* note 86, at 436–38.

89. JOHN GUARE, *SIX DEGREES OF SEPARATION* (1990).

90. Milgram, *supra* note 85, at 66; Travers & Milgram, *supra* note 86, at 436–39.

91. Duncan J. Watts & Steven H. Strogatz, *Collective Dynamics of "Small-World" Networks*, 393 *NATURE* 440, 441 (1998); *see also* DUNCAN J. WATTS, *SIX DEGREES: THE SCIENCE OF A CONNECTED AGE* (2003).

92. *See* P. Erdős & A. Rényi, *On Random Graphs I*, 6 *PUBLICATIONES MATHEMATICAE* 290 (1959); P. Erdős & A. Rényi, *On the Evolution of Random Graphs*, 5 *PUBL'NS MATH. INST. HUNGARIAN ACAD. SCI.* 17 (1960).

93. *See* Watts & Strogatz, *supra* note 91, at 440.

94. *Id.*

95. *Id.* at 441.

96. *Id.*

network tends to be rather large.⁹⁷ The shorter path length associated with random networks, in contrast, makes them inherently *small-world networks*, in that the number of links needed to travel between two randomly selected nodes tends to be somewhat small.⁹⁸

To demonstrate this difference, Watts and Strogatz compared the behavior of a lattice-like network with that of a random network.⁹⁹ To avoid complications created by nodes located on the boundary of the graph, which necessarily have fewer interactions than other nodes, Watts and Strogatz modeled the lattice-like network as a circle, with each node connected to its two nearest neighbors and not to any distant nodes.¹⁰⁰ To be sure, this conception of the lattice-like network differs from Baran's distributed network, in which each node should be connected to four neighboring nodes.¹⁰¹ Watts and Strogatz then ran an experiment where each link could be reassigned with probability p to another randomly chosen node in the circle.¹⁰² If the probability was zero, no links would ever be reassigned.¹⁰³ Under that condition, the network should remain both strict-lattice and a large-world.¹⁰⁴ If the probability was 100 percent, all links would be reassigned, and the network should become both an Erdős-type random network and a small-world network.¹⁰⁵

A relevant takeaway from Watts and Strogatz' work is what happens when only a small number of links is reassigned, as depicted in the middle graph of Figure 7. Specifically, Watts and Strogatz found that introducing a small number of long-distance links into the network radically reduced the average path length, while still retaining a highly clustered structure.¹⁰⁶ This showed how short cuts, as Watts and Strogatz called them, can allow networks to be both highly clustered *and* small-world networks.¹⁰⁷ Indeed, the introduction of five such short cuts reduced the average path length by half.¹⁰⁸ The effect does decay rather rapidly, however. A further 50 percent reduction in path length would require introducing an additional 50 short cuts.¹⁰⁹ Watts and

97. *See id.*

98. *See* Watts & Strogatz, *supra* note 91, at 440.

99. *See id.*

100. *See id.* at 441 fig.1; *see also infra* Figure 7 (left-most graph).

101. *See supra* note 26 and accompanying text.

102. Watts & Strogatz, *supra* note 91, at 440–41.

103. *See id.*

104. *See id.* at 441 fig.1; *see also infra* Figure 7 (left-most graph).

105. *See* Watts & Strogatz, *supra* note 91, at 441 fig.1; *infra* Figure 7 (right-most graph).

106. *See* Watts & Strogatz, *supra* note 91.

107. *Id.* at 441 fig.1, fig.2.

108. WATTS, *supra* note 91, at 89.

109. *Id.*

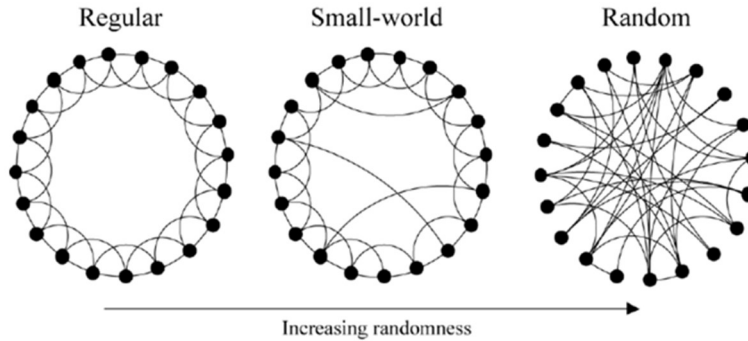


FIGURE 7: WATTS AND STROGATZ ON HOW RANDOM LINKS CAN CREATE SMALL WORLDS¹¹⁰

Strogatz found short cuts to be empirically present in a wide range of networks, including those of film actors, the electrical power grid, and the neural network of a particular nematode worm.¹¹¹

The implications of these findings are profound. Watts and Strogatz showed that adding a limited number of random shortcuts to a strict lattice-like structure can play a key role in permitting Baran's distributed networks to scale.¹¹² Under this model, each node is no longer undifferentiated in terms of connectivity. Instead, there are a handful of nodes with more important levels of connectivity that keep the network small-world, which enables it to scale, a characteristic that had eluded Baran's distributed network. A notable caveat, however, is that these nodes start to increasingly resemble the centers of the smaller stars in Baran's decentralized network.

B. Albert-László Barabási and Scale-Free Networks

Watts's work was extended by scholars including Albert-László Barabási, who noted that all of the nodes in the random networks in the Erdős-Rényi and Watts-Strogatz studies had a functionally equivalent number of links.¹¹³ In other words, the number of links connected to the typical node in a random network has a given *scale*, a specific characteristic that is represented across all nodes in the network.¹¹⁴ In random networks, the number of links associated

110. Watts & Strogatz, *supra* note 91, at 441 fig.1.

111. *Id.*

112. *See supra* note 106–107 and accompanying text.

113. *See* Réka Albert et al., *Diameter of the World-Wide Web*, 401 NATURE 130, 130 (1999); *see also* Albert-László Barabási et al., *Mean-Field Theory for Scale-Free Random Networks*, 272 PHYSICA A 173, 175–78 (1999).

114. Barabási et al., *supra* note 113, at 178.

with each node should thus follow a Poisson distribution, which should occur when all nodes have roughly the same number of links.¹¹⁵ However, Barabási observed that the Internet evinced a very different pattern.

Instead of following a Poisson distribution, the empirical literature indicated the number of links connected to each node exhibited a power-law distribution, which occurs when there are a small number of nodes with a disproportionately large number of links.¹¹⁶ The heterogeneity in the number of links attached to each node in such networks meant that the network did not have an inherent characteristic of scale in terms of the number of links.¹¹⁷ As such, Barabási called such networks *scale-free* networks.¹¹⁸

Barabási proceeded to develop a dynamic theory to explain why scale-free networks would develop under the Watts-Strogatz and other similar models. Specifically, Barabási observed that those models assumed that the number of nodes in a given network is static, and network connections form randomly among nodes.¹¹⁹ With that observation, Barabási created a model with a dynamic network, where new nodes would be free to decide where to attach to the network.¹²⁰ Under Barabási's model, these nodes would follow the practice of *preferential attachment*: they would be more likely to attach to nodes that were already highly connected.¹²¹ Barabási's result: a network where the majority of nodes have only one or two links while a handful of other nodes—those that attracted the new nodes—have a large number of links.¹²²

The presence of these *hubs* radically reduces the shortest path between two nodes.¹²³ Moreover, in a random network the average path length should grow linearly with the number of nodes in the network, but Barabási's work suggested that the average path length would grow logarithmically—and thus considerably slower—in scale-free networks.¹²⁴ He noted that a wide range of networks had empirically followed this pattern, including the Internet, the World Wide Web, film actor collaborations, scientific

115. Albert-László Barabási & Réka Albert, *Emergence of Scaling in Random Networks*, 286 SCI. 509, 510–11 (1999).

116. See Réka Albert & Albert-László Barabási, *Statistical Mechanics of Complex Networks*, 74 REV. MOD. PHYSICS 47, 50–52 (2002); H. Jeong et al., *Measuring Preferential Attachment for Evolving Networks*, 61 EUROPHYSICS LETTERS 567 (2003).

117. See Barabási et al., *supra* note 113, at 178 (explaining that “many systems in nature have the common property that $P(k)$ is free of scale, following a power-law distribution over many orders of magnitude”)

118. Barabási & Albert, *supra* note 115, at 511.

119. *Id.* at 510–11.

120. *Id.* at 511.

121. *Id.*

122. *Id.*

123. Albert-László Barabási, *The Physics of the Web*, PHYSICS WORLD, July 2001, at 33, 34 fig.1 (2001).

124. Albert et al., *supra* note 113, at 130.

collaborations, human sexual contacts, cellular networks, ecological networks, phone call networks, citation networks, linguistic networks, power networks, neural networks, and protein folding.¹²⁵

Later empirical studies refined Barabási's, particularly as it applied to the Internet. Specifically, these studies found that the Internet's architecture did not resemble a lattice-like grid and that the preferential attachment rates fell slightly below those identified in Barabási's hypothesis.¹²⁶ Other studies also found preferential attachment but concluded that it occurred at rates above those which Barabási predicted.¹²⁷

Germane to Baran's work, Barabási's observations pointed out that decentralized networks with a limited number of hubs are actually *more* resilient against incidental node and link failures; indeed, random failure will almost certainly affect small nodes, which have little to no impact on the network's overall connectivity.¹²⁸ He did note, however, that the presence of key focal points makes decentralized networks more vulnerable to *deliberate attacks*, which would undoubtedly target the hubs.¹²⁹

C. Implications for the Distributed Network

The developments in network theory provide insight into why the Internet did not incorporate the topology that Baran envisioned. The fact that average path length in traditional lattice-like networks grows linearly with the number of nodes means that such networks are difficult to scale.¹³⁰ The introduction of long-distance shortcuts and hubs allows average path length to grow at a slower, logarithmic rate, as depicted in Figure 8.

Hubs thus enable shortcuts to play a critical role in giving networks the small-world characteristics in which average path length remains intact as the network grows. These features deviate from Baran's distributed structure—in which all nodes are undifferentiated and of equal importance—and instead make certain nodes more important. Network theory thus provides an explanation of why the Internet mimicked a more decentralized structure of satellite hubs instead of Baran's distributed structure. Some commentators have raised the concern that the increased

125. Albert & Barabási, *supra* note 116, at 49–54.

126. See Georgos Siganos et al., The Evolution of the Internet: Topology and Routing 8–9 (2002) (unpublished manuscript), www.cs.ucr.edu/~siganos/papers/tech1.pdf [<https://perma.cc/Y5HC-BPEN>]; see also S. Shakkottai et al., *Evolution of the Internet AS-Level Ecosystem*, 74 EUR. PHYSICAL J. B 271 (2010).

127. See Qian Chen et al., *The Origin of Power Laws in Internet Topologies Revisited*, 2 PROC. IEEE INFOCOM 608 (2002).

128. See BARABÁSI, *supra* note 2, at 114; see also Réka Albert et al., *Error and Attack Tolerance of Complex Networks*, 406 NATURE 378, 380 (2000).

129. BARABÁSI, *supra* note 2, at 118, 120; Albert et al., *supra* note 128, at 380.

130. See *supra* note 96 and accompanying text.

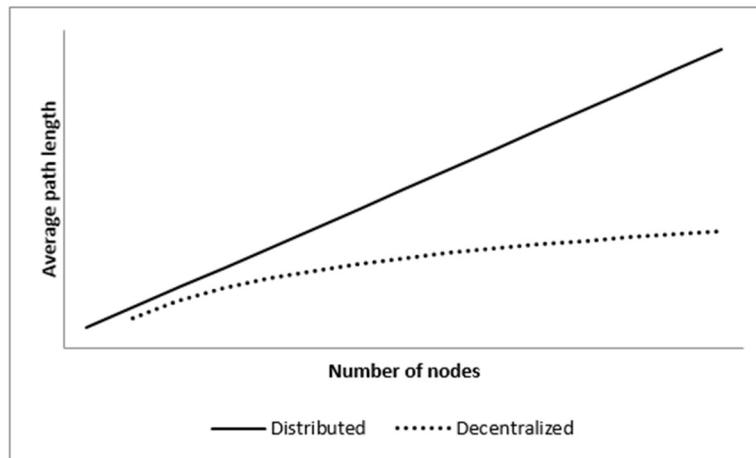


FIGURE 8: AVERAGE PATH LENGTH GROWTH FOR DISTRIBUTED AND DECENTRALIZED NETWORKS

importance of some hubs may give them a competitive advantage.¹³¹ But such concerns can be alleviated by adding competition from new hubs so long as the costs of network reconfiguration are manageable.

III. THE EVOLUTION OF THE MODERN INTERNET

The conventional wisdom that the Internet's foundational topology adhered to a three-tier hierarchy was so well established that it served as the reference model in every textbook on computer networking.¹³² As noted above, the resulting architecture bore a far stronger resemblance to what Baran called a decentralized network than it did to his distributed network.

More recently, two forces have pushed the network even further away from Baran's distributed model. The first is the triumph of the client-server architecture over the peer-to-peer model, as discussed in Part A. The second is the emergence of topologies that deviate from the three-tiered hierarchical model, as discussed in Part B. Although changes such as secondary peering and multihoming may make the network more mesh-like, other features make the hierarchy even flatter (such as content delivery networks and direct interconnection between content providers and last-mile networks). Interestingly, these changes have made the

131. BARABÁSI, *supra* note 2, at 58; see also Konstantinos Stylianou, *Exclusion in Digital Market*, 24 MICH. TELECOMM. & TECH. L. REV. 181, 230 (2018); Kevin Werbach, *The Centripetal Network: How the Internet Holds Itself Together, and the Forces Tearing It Apart*, 42 U.C. DAVIS L. REV. 343 (2008).

132. See *supra* note 81 and accompanying text.

architecture begin to resemble the star-like network that Baran described as centralized.

A. The Rise of the Client-Server Architecture and the Decline of Peer-to-Peer

During the Internet's early days, before it became a mass-market phenomenon in the mid-1990s, it was used primarily by tech-savvy researchers to send email, share files, and view bulletin boards.¹³³ When that was the case, the network could be characterized as a "peer-to-peer any-to-any flat mesh" in which nodes would typically request and send data.¹³⁴ Using the early network in that way was consistent with the assumption of Baran's distributed network that nodes were generally homogeneous and of equal status.

The advent of the World Wide Web during the mid-1990s upended this dynamic.¹³⁵ The Internet began to embrace a client-server architecture, which was inherently asymmetric—nodes were divided into the end users who requested content and the content providers who furnished it.¹³⁶ Although file sharing slowed this transformation for a time, peer-to-peer traffic has been more or less steadily declining as a percentage of traffic since at least 2005,¹³⁷ and has been declining in absolute terms since 2012.¹³⁸ Cisco forecasts that it will continue to decline, sinking to a mere 2% by 2021.¹³⁹

The asymmetry inherent in the client-server architecture is inconsistent with Baran's vision of a distributed model with homogeneous nodes.¹⁴⁰ The effect of this asymmetric architecture has been compounded as traffic has become increasingly concentrated among a small number of content providers.¹⁴¹ In such an environment, it no longer makes business sense for content providers to reside at the edge of the network and rely on a series

133. Christopher S. Yoo, *Free Speech and the Myth of the Internet as an Unintermediated Experience*, 78 GEO. WASH. L. REV. 697, 703 (2010).

134. Geoff Huston, *The Death of Transit?*, ISP COLUMN, Oct. 2016, at 4, <http://www.potaroo.net/ispcol/2016-10/xtransit.pdf> [<https://perma.cc/9AU7-83ZL>].

135. Yoo, *supra* note 133.

136. *Id.*; Huston, *supra* note 134.

137. See YOO, *supra* note 63, at 30–31; *infra* Figure 9.

138. See Christopher S. Yoo, *Possible Paradigm Shifts in Broadband Policy*, 9 I/S: J.L. & POLY INFO. SOC'Y 367, 378 (2014).

139. CISCO, CISCO VISUAL NETWORKING INDEX: FORECAST AND METHODOLOGY, 2013–2018, at 10 tbl.10 (2014), http://www.anatel.org.mx/docs/interes/Cisco_VNI_Forecast_and_Methodology.pdf [<https://perma.cc/ZPH7-QW7G>].

140. Huston, *supra* note 134.

141. SANDVINE, 2016 GLOBAL INTERNET PHENOMENA: LATIN AMERICA & NORTH AMERICA 4 tbl.1 (2016), <http://www.sandvine.com/hubfs/downloads/archive/2016-global-internet-phenomena-report-latin-america-and-north-america.pdf> [<https://perma.cc/K48W-VDWY>].

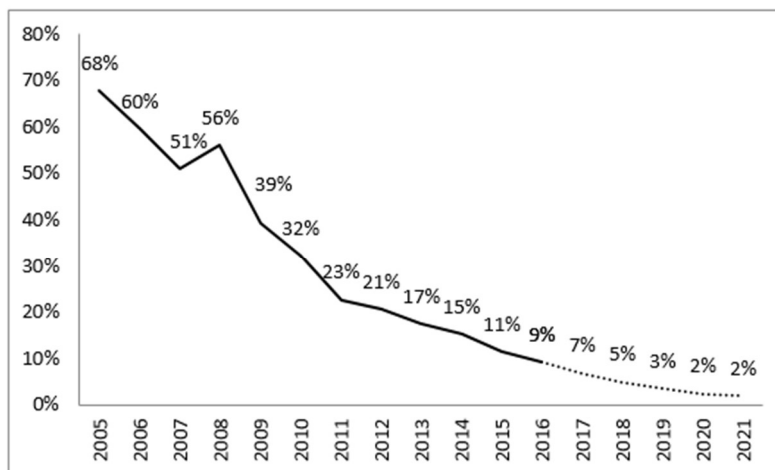


FIGURE 9: PEER-TO-PEER TRAFFIC AS A PERCENTAGE OF CONSUMER INTERNET TRAFFIC¹⁴²

of public networks to transmit their data to end users, particularly when there are delays associated with DNS resolution, TCP and TLS handshakes, and long round-trip times.¹⁴³ In addition, content providers can avoid the uncertainties inherent in sharing bandwidth with others, and the concomitant possibilities of delay and security risks, by building their own private networks out to local caches.¹⁴⁴ When that is the case, it makes sense for content providers to bypass the network and host their content as close to the end users as possible.¹⁴⁵

This dynamic explains why large content providers have begun to connect directly to last-mile providers. Indeed, in 2008 Google had already begun to construct a wide area network that covered most of the U.S. and connected to Europe, Asia, and South America.¹⁴⁶ Yahoo! and Microsoft were likewise constructing wide area works covering the U.S.¹⁴⁷ More recent press reports have covered how Google, Facebook, Microsoft, and Amazon are

142. See *supra* notes 137–139 and accompanying text.

143. Huston, *supra* note 134, at 3.

144. *Id.*

145. *Id.*

146. Phillipa Gill et al., *The Flattening Internet Topology: Natural Evolution, Unsilently Barnacles or Contrived Collapse?*, in *PASSIVE AND ACTIVE NETWORK MEASUREMENT: 9TH INTERNATIONAL CONFERENCE, PAM 2008*, at 1, 1–2 (Mark Claypool & Steve Uhlig eds., 2008). The paper was motivated by a 2005 press report about the growth of Google's private network. R. Scott Raynovich, *Google's Own Private Internet*, *LIGHT READING* (Sept. 20, 2005), <https://www.lightreading.com/ethernet-ip/googles-own-private-internet/d/d-id/618122> [<https://perma.cc/8TD9-Y949>].

147. Gill et al., *supra* note 146.

currently constructing large networks of undersea cables to interconnect proprietary data centers with last mile providers located ever closer to end users.¹⁴⁸ In fact, the international capacity of these companies increased tenfold from 2013 to 2017.¹⁴⁹ The upshot: rather than waiting for users to reach out for content, content providers are increasingly bringing content directly to end users.

B. *The Flattening of the Network Architecture*

Network architecture has evolved considerably over the past three decades.¹⁵⁰ In the words of the emerging empirical literature, the Internet's architecture has become *flatter*; traffic is no longer passing through all three layers of the Internet's traditional three-layer hierarchy.¹⁵¹

This flattening of network architecture is likely the result of one of two phenomena. First, it may be caused by increased density of connections in the core, either from Tier 3 ISPs bypassing Tier 2 ISPs and connecting directly to Tier 1 ISPs or from Tier 2 ISPs bypassing Tier 1 ISPs by engaging in secondary peering.¹⁵² If this is the case, the network can be said to be becoming more mesh-like. Alternatively, large content providers could be bypassing both Tier 1 and Tier 2 ISPs altogether and using content distribution networks and data centers to connect directly to Tier 3 ISPs.

A key indicator of a network's true underlying architecture is the rate of change in the average path length over time.¹⁵³ Specifically, and as noted earlier, if the network is distributed, average path length should grow linearly with the number of nodes.¹⁵⁴ If the network is decentralized, the average path length should grow logarithmically with the number of nodes.¹⁵⁵ Decreases in average path length are associated with centralized networks. In the limit, the average path length of a centralized star network is two.

Empirical studies of changes in the average path length thus provide important insights into how network architecture is changing. If the network is decentralized or distributed, the fact

148. Alan Mauldin, *A Complete List of Content Providers' Submarine Cable Holdings*, TELEGEOGRAPHY (Nov. 9, 2017), <https://blog.telegeography.com/telegeographys-content-providers-submarine-cable-holdings-list> [<https://perma.cc/D6HS-D6KL>].

149. *Id.*

150. See Yoo, *supra* note 11, at 85–90 (discussing how the architecture of the Internet has evolved).

151. See Gill et al., *supra* note 146, at 2; see also Besen & Israel, *supra* note 10; Dhamdhere & Dovrolis, *supra* note 10.

152. YOO, *supra* note 63, at 64–68.

153. See *supra* notes 91, 96, 106–108, 124 and accompanying text.

154. See *supra* note 98 and accompanying text.

155. See *supra* note 124 and accompanying text.

that the number of nodes in the network has been constantly increasing implies that the average path length should be growing. However, and somewhat surprisingly, empirical studies have found the contrary: the average path length in the Internet has been either flat¹⁵⁶ or decreasing,¹⁵⁷ despite growth in the number of nodes. This phenomenon suggests that the network is not becoming more mesh-like; it is instead becoming more centralized. Under this scenario, the flattening of the Internet's architecture is more associated with content providers connecting directly with Tier 3 ISPs on the edge of the network rather than the result of denser peering relationships in the core of the network. In other words, the flattening of the architecture begins to more closely reflect Baran's centralized network, where the average path length is 1.

Studies have attempted to move beyond merely studying average path length by examining directly whether the flattening of the network is associated more with the densification of the core or with the ISP's bypassing the core altogether. The earliest studies observed that traffic from the largest content providers traversed fewer hops on Tier 1 networks than did traffic from smaller content providers.¹⁵⁸ Although a 2011 study found it plausible that an increase in multihoming, rather than content providers bypassing Tier 1 ISPs,¹⁵⁹ was the cause of the lack of growth in average path length, a second 2011 study—one that actually evaluated the network empirically—found otherwise; it reinforced the conclusion in the earlier studies: the flattening of the network has stemmed from content provider forging routes that bypass Tier 1 ISPs rather than form multihoming.¹⁶⁰ What is more, a 2012 study implicitly minimized the impact of increased connectivity among Tier 1 and Tier 2 ISPs when it concluded that although the core of the network was densifying, the interconnectedness of the overall network remained unchanged.¹⁶¹

Together these studies contradict suggestions that the flattening of the Internet means that the network is becoming more mesh-like, and instead indicate that content providers are increasingly bypassing the other tiers and connecting directly to

156. Amogh Dhamdhere & Constantine Dovrolis, *Twelve Years in the Evolution of the Internet Ecosystem*, 19 IEEE/ACM TRANSACTIONS ON NETWORKING 1420, 1424–25 (2011); Siganos et al., *supra* note 126.

157. Jure Leskovec et al., *Graph Evolution: Densification and Shrinking Diameters*, 1 ACM TRANSACTIONS ON KNOWLEDGE DISCOVERY FROM DATA, art. 2, at 14–19 (2007); Miao Li et al., *Flattening and Preferential Attachment in the Internet Evolution*, 14 PROC. ASIA-PACIFIC NETWORK OPERATIONS & MGMT. SYMP., at 1 (2012); Yang Xiang et al., *Internet Flattening: Monitoring and Analysis of Inter-Domain Routing*, 2011 PROC. IEEE INT'L CONF. ON COMM., at 4–5 (2011).

158. Gill et al., *supra* note 146, at 6; Craig Labovitz et al., *Internet Inter-Domain Traffic*, 40 ACM SIGCOMM COMPUTER COMM. REV. 75, 79 (2010).

159. Dhamdhere & Dovrolis, *supra* note 156.

160. Xiang et al., *supra* note 157, at 3–4.

161. Li et al., *supra* note 157, at 5.

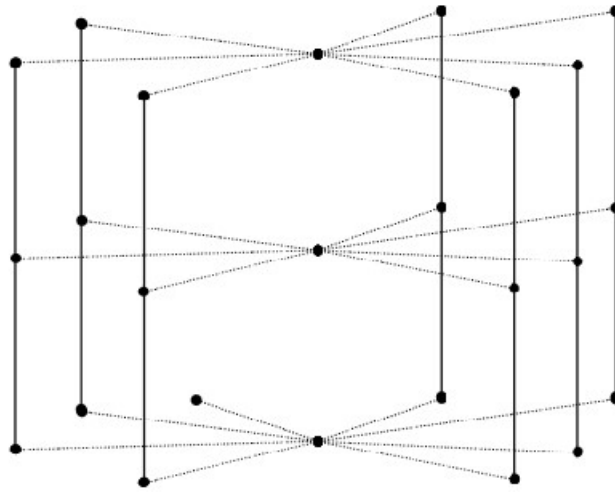


FIGURE 10: NETWORK ARCHITECTURE RECONCEPTUALIZATION AS A STACK OF STARS CENTERED ON CONTENT PROVIDERS

last-mile ISPs. This, in turn, suggests the possibility that the standard reference model should be replaced with a radically different vision—a stack of star networks as depicted in Figure 10, each centered on one of the leading content providers and connecting by private links (represented by dotted lines) to the same public locations at the edge (represented by solid lines). Rather than resembling a mesh network, this emerging architecture bears greater similarity to Baran’s centralized network.

This reconceptualization has profound implications. Investment in network capacity helps meet users’ ever-growing demand for bandwidth. Moreover, content providers and their customers can benefit from improved risk management and cost control, support for the deployment of new applications, and improvements in quality of service.¹⁶² At the same time, as Geoff Huston has noted, the resulting topology would resemble the historical broadcast television network, in which local distribution networks that carried content from multiple sources were fed by proprietary satellite networks.¹⁶³ The resulting diversion of long-haul traffic from public to private networks may make backbones more difficult to sustain, which in turn may adversely affect smaller

162. Yoo, *supra* note 63, at 68; Gill et al., *supra* note 146, at 2–3, 7–8.

163. Geoff Huston, *The Death of Transit and Beyond*, Remarks at the Asia Pacific Regional Conference on Operational Technologies (APRICOT 2017) 22 (Feb. 28, 2017), https://2017.apricot.net/assets/files/APIC674/2017-02-28-death-of-transit_1488193223.pdf [<https://perma.cc/G2EK-N2CS>].

content providers that are unable to pursue the same strategy. Increasing centralization may increase the ability of some actors to serve as gatekeepers.

CONCLUSION

The seminal work of Paul Baran has never received its full due. Although his breakthrough research established the foundation upon which modern packet switching was built, Baran's three-part taxonomy of centralized, decentralized, and distributed networks has been overlooked—the reception of his ideas remains incomplete. Specifically, the early Internet architects were less concerned with survivability than was Baran while network theorists such as Watts and Barabási showed hubs, shortcuts, and other modifications to the distributed architecture can allow networks to scale. The result was a three-tiered hierarchy of ISPs that formed what Baran would define as a decentralized network, not the distributed network that he preferred.

The dramatic changes to the network hierarchy over recent years suggests that we should reconceptualize the architecture still further. Specifically, the fact that the average path length is shrinking while routes are increasingly bypassing Tier 1 ISPS to connect directly to the last mile compels us to consider whether the standard reference model should be replaced with a radically different vision: a stack of star networks, each centered on one of the leading content providers, which would functionally bear a stronger resemble Baran's centralized network.

The expansion of content providers into wide area networking may yield a number of substantial benefits. Greater investment in transmission capacity is essential to meet users' ever-growing demand for bandwidth. However, the shift from public to private backbones may have certain implications that need to be thought out.

Such concerns must be analyzed in light of the full range of options instead of being simply asserted. The full impact of the change can only be understood once the full choice set is endogenized, including multihoming, the creation of alternative hubs, cooperative investments in competing technology, and other institutional forms, to name a just few.

