NOTES ON

NETWORK ECONOMICS AND THE “NEW ECONOMY”

by

Prof. Nicholas Economides

Stern School of Business

http://www.stern.nyu.edu/networks/

August 2000

I. NETWORKS AND THE "NEW ECONOMY"

1. In recent years, high technology industries have been playing an even more central role in the U.S. and world economy, exhibiting very fast growth and extremely high valuations of their equity. Many of the high technology industries are based on networks (such as the telecommunications network and the Internet). Other high industries, such as the computer software and hardware industries, exhibit properties that are typically observed in networks. So, to understand the "new economy" we need to understand the economics of networks.

2. Networks are composed of *complementary* nodes and links. The crucial defining feature of networks is the complementarity between the various nodes and links. A service delivered over a network requires the use of two or more network components. Thus, network components are complementary to each other. Figure 1, represents the emerging ***Information Superhighway*** network. Clearly, services demanded by consumers are composed of many complementary components. For example, interactive ordering while browsing in a "department store" as it

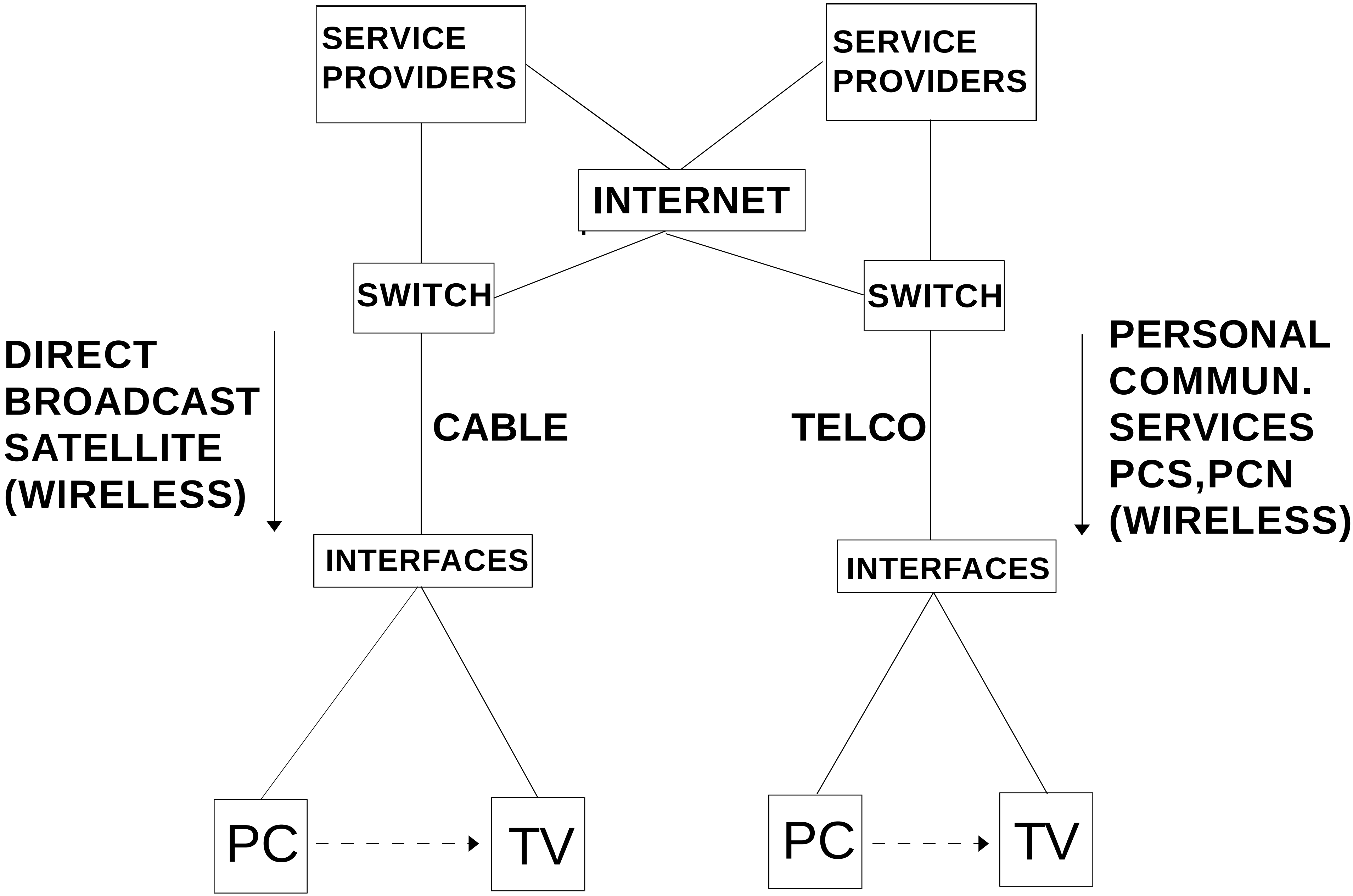
SERVICE PROVIDERS
SERVICE PROVIDERS
INTERNET
SWITCH
SWITCH
DIRECT BROADCAST SATELLITE (WIRELESS)
CABLE
TELCO
PERSONAL COMMUN. SERVICES PCS,PCN (WIRELESS)
INTERFACES
INTERFACES
PC
TV
PC
TV

Figure 1

appears in successive video frames requires a number of components: a database engine at the service provider, transmission of signals, decoding through an interface, display on a TV or computer monitor, etc. Clearly, there are close substitutes for each of these components; for example, transmission can be done through a cable TV line, a fixed telephone line, a wireless satellite, PCN, etc.; the in-home interface may be a TV-top box or an add-on to a PC, etc. It is likely that the combinations of various components will not result in identical services. Thus, the information superhighway will provide substitutes made of complements; this is a typical feature of networks.

3. Figure 2 shows this feature in a simple star telephone network. A phone call from A to B is composed of AS (access to the switch of customer A), BS (access to the switch of customer B), and switching services at S. Despite the fact that goods AS and BS look very similar and have the same industrial classification, they are *complements* and not substitutes.

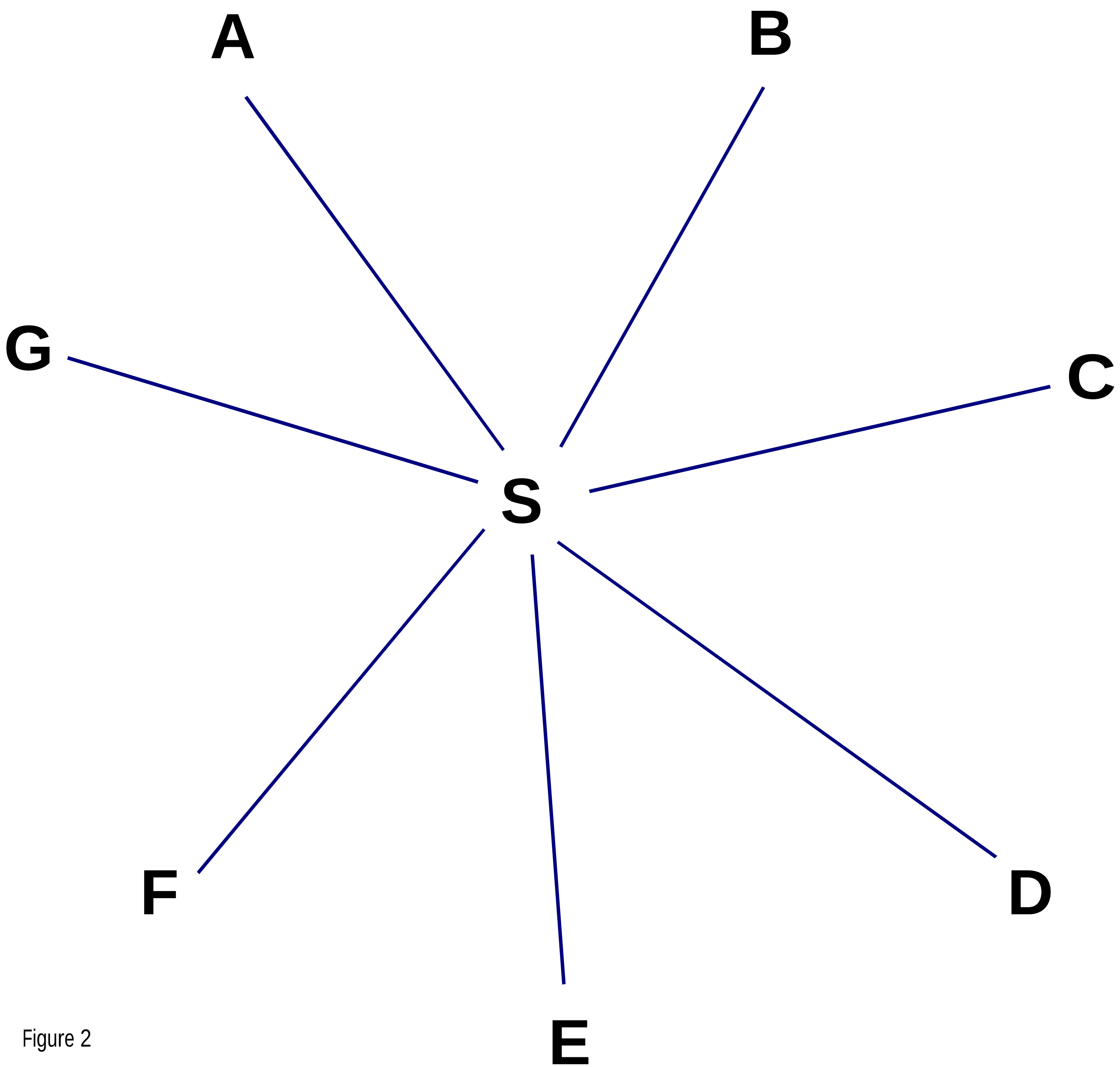

Figure 2

4. **One-way and Two-way Networks.** Networks where services AB and BA are distinct are called "two-way" networks. Two-way networks include railroad, road, and many telecommunications networks. When one of AB or BA is unfeasible, or does not make economic sense, or when there is no sense of direction in the network so that AB and BA are identical, then the network is called a one-way network. In a typical one-way network, there are two types of components, and composite goods are formed only by combining a component of each type, and customers are often not identified with components but instead demand composite goods. For example, broadcasting and paging are one-way networks.

5. The classification in network type (one-way or two-way) is not a function of the topological structure of the network. Rather, it depends on the interpretation of the structure to represent a specific service. For example, the network of Figure 3 can be interpreted as a two-way telephone network where SA represents a local switch in city A, A_i represents a customer in city A, and similarly for SB and B_j. We may

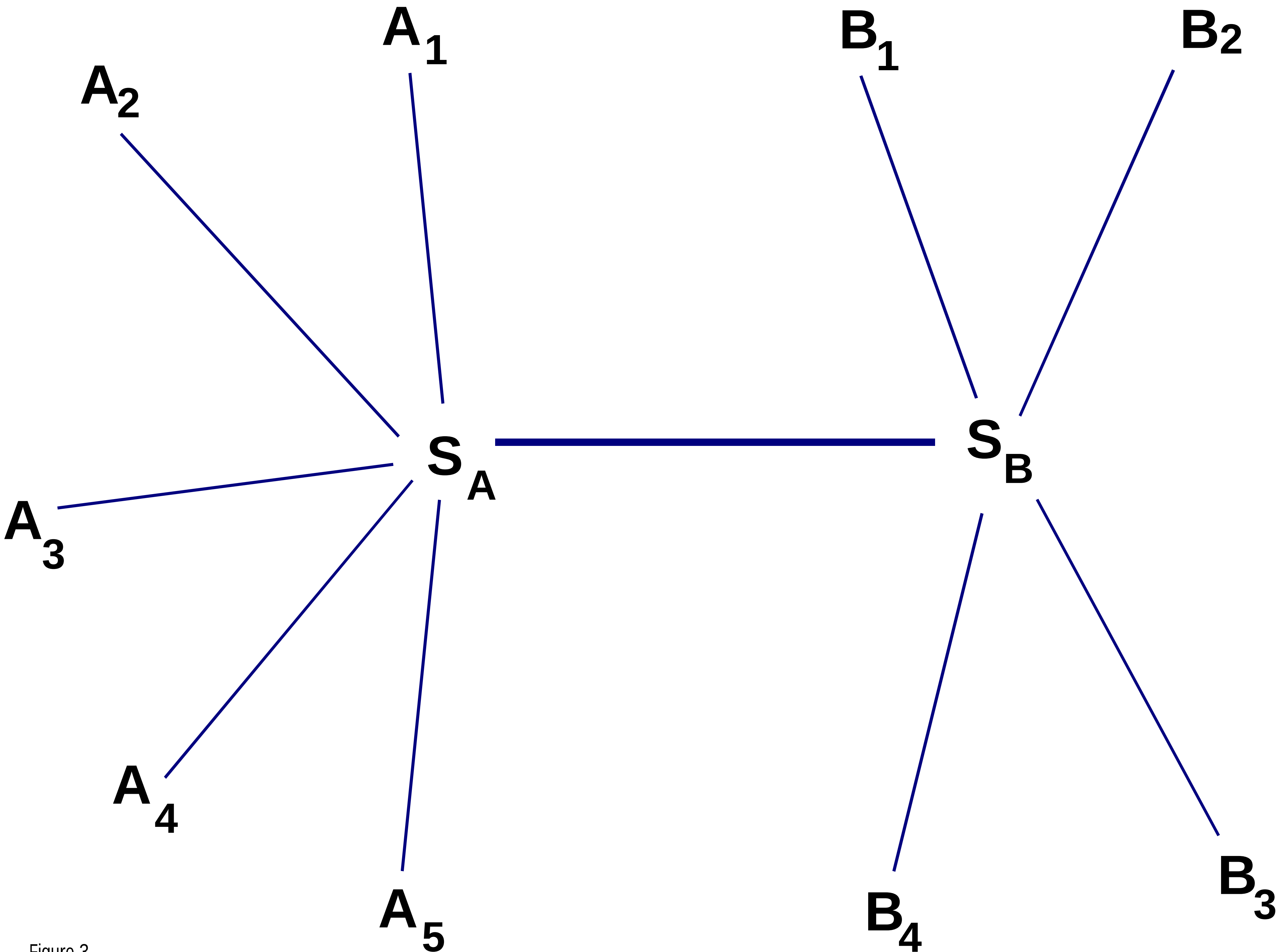

Figure 3

identify end-nodes, such as A_i and B_j, end-links, such as A_iS_A and S_BB_j, the interface or gateway S_AS_B, and switches S_A and S_B. In this network, there are two types of local phone calls $A_iS_AA_k$ and $B_jS_BB_R$, as well as long distance phone call $A_iS_AS_BB_j$.

6. We can also interpret the network of Figure 3 as an Automatic Teller Machine network. Then a transaction (say a withdrawal) from bank B_j from ATM A_i is $A_iS_AS_BB_j$. Connections $A_iS_AA_k$ and $B_jS_BB_R$ may be feasible but there is no demand for them.

7. The crucial relationship in both one-way and two-way networks is the complementarity between the pieces of the network. This crucial economic relationship is also often observed between different classes of goods in non-network industries. Figure 4 can represent two industries of complementary goods D and S, where consumers demand combinations D_iS_j. Notice that this formulation is formally identical to our long-distance network of Figure 3 in the ATM interpretation.

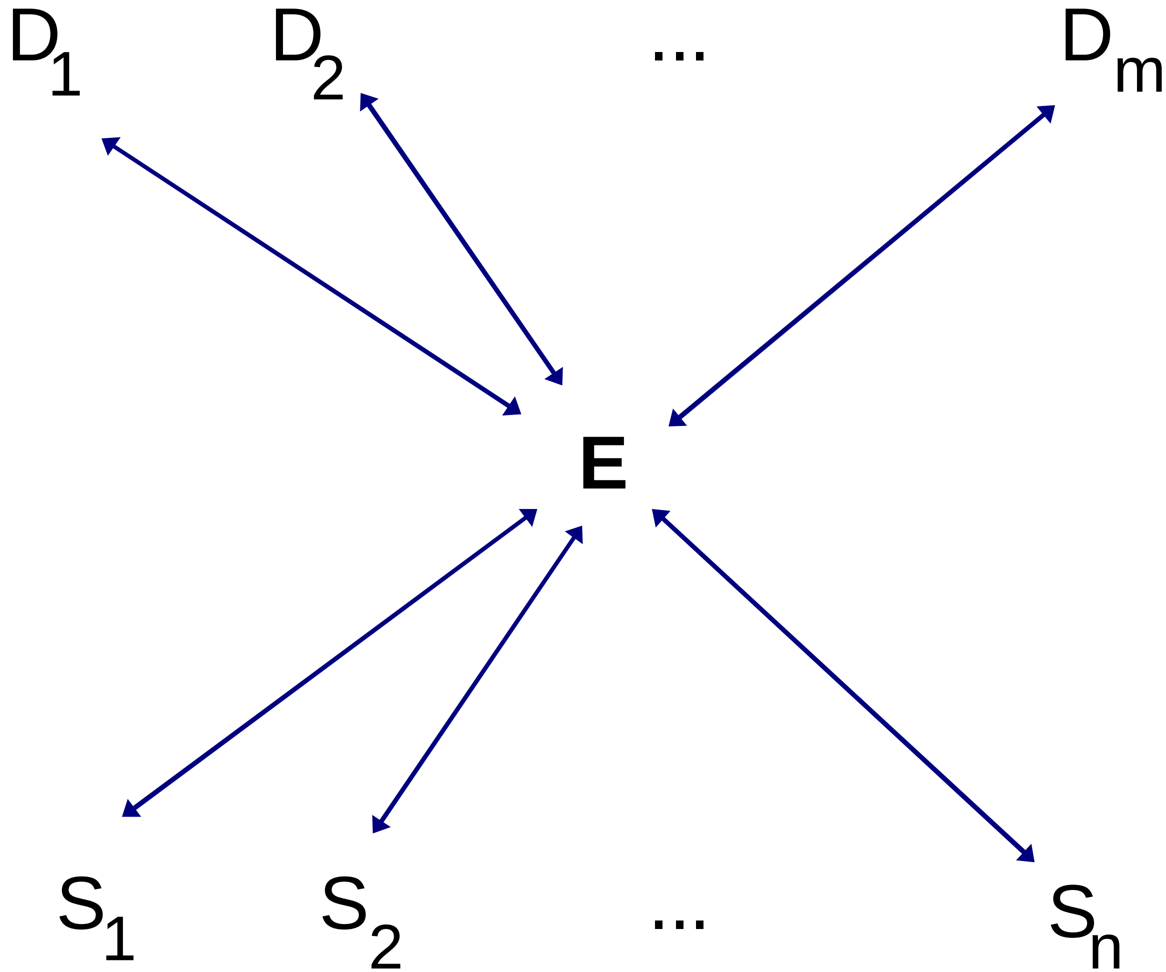

Figure 4

8. **Compatibility**. So far we have assumed *compatibility*, i.e., that various links and nodes on the network are costlessly combinable to produce demanded goods. Two complementary components A and B are compatible when they can be combined to produce a composite good or service. For example, we say that a VHS video player is compatible with a VHS tape. Two substitute components A_1 and A_2 are compatible when each of them can be combined with a complementary good B to produce a composite good or service. For example, two VHS tapes are compatible. Similarly, we say that two VHS video players are compatible.

9. Links on a network are potentially complementary, *but it is compatibility that makes complementarity actual*. Some network goods and some vertically related goods are immediately combinable because of their inherent properties. However, for many complex products, actual complementarity can be achieved only through the adherence to specific technical compatibility standards. Thus, many providers of network or vertically-related goods have the option of making their

products partially or fully incompatible with components produced by other firms. This can be done through the creation of proprietary designs or the outright exclusion or refusal to interconnect with some firms.

10. **Virtual Networks**. A virtual network can be thought of as a collection of compatible goods that share a common technical platform. For example, all VHS video players make up a virtual network. Similarly, all computers running Windows 98 can be thought of as a virtual network. More generally, a virtual network can be thought of a combination of two collections of two types of goods $\{D_1, \ldots, D_m\}$ and $\{S_1, \ldots, S_n\}$ such that (i) each of the D-type good is a substitute to any other D-type good; (ii) each of the S-type good is a substitute to any other S-type good; and (iii) each of the D-type good is a complement to any S-type good. Virtual networks are one-way networks. Examples of virtual networks: computer hardware and software; computer operating systems and software applications, etc. Clearly there are many more virtual networks than traditional networks.

11. We proceed with three more definitions before describing the most important feature of networks. **Installed Base** is the number of users (or units of the product sold) that share a particular software or hardware platform. For example, the installed base of VHS video players is the collection of video players of various manufacturers (JVC, Toshiba, Panasonic, etc.) that play VHS tapes.

12. **Path-dependence** is the dependence of a system or network on past decisions of producers and consumers. For example, the price at which a VHS player can be sold today is path dependent because it depends on the number of VHS players sold earlier (on the installed base of VHS players).

13. **Bottleneck**. A bottleneck is a part of the network for which there is no available substitute in the market. For example, a firm may monopolize a link of a railroad network, like the link AB in Figure 5.

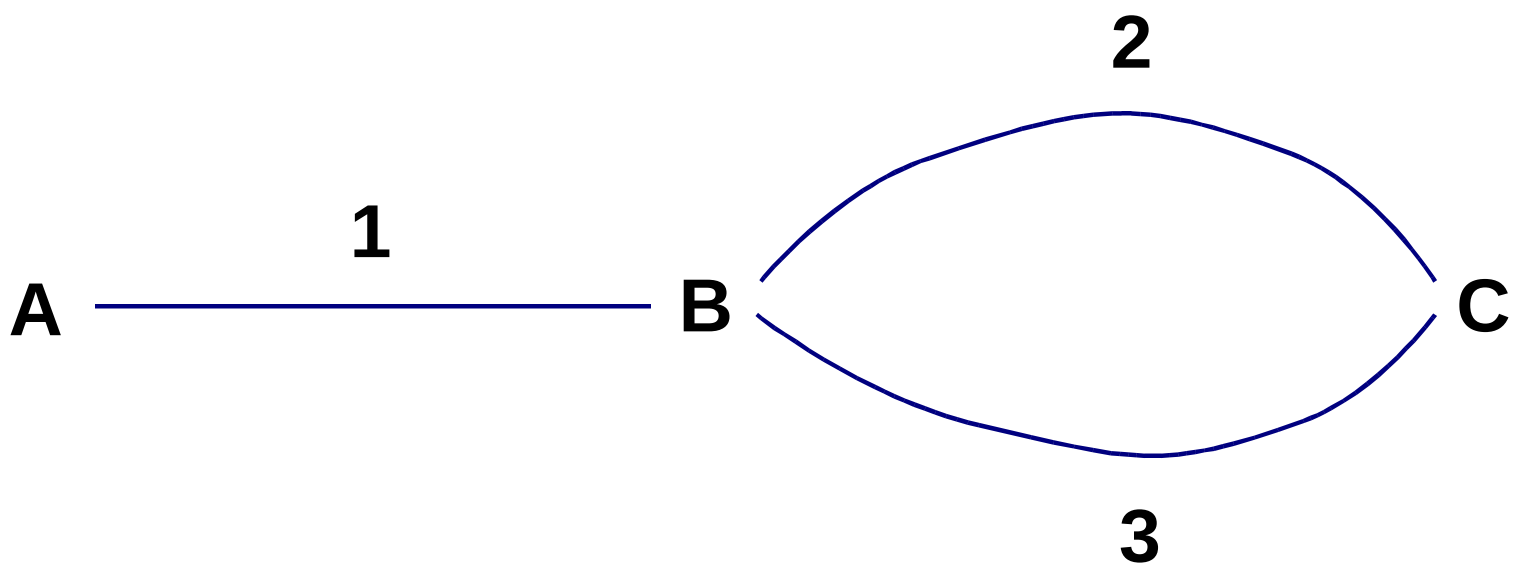

Figure 5

II. NETWORK EXTERNALITIES

14. The existence of network externalities is the key reason for the importance, growth, and profitability of network industries and the new economy. A network exhibits *network effects* or *network externalities* when the value (to a consumer) of a subscription to the network is higher when the network has more subscribers, everything else being equal. In a traditional network (say a telecommunications network), network externalities arise because a typical subscriber can reach more subscribers in a larger network. In a virtual network, network externalities arise because larger sales of component A induce larger availability of complementary components B_1, ..., B_n, thereby increasing the value of component A. The increased value of component A results in further positive feedback. Despite the cycle of positive feedbacks, it is typically expected that the value of component A does not explode to infinity because the *additional* positive feedback is expected to decrease with increases in the size of the network.

15. **Sources of Network Externalities**. The key reason for the appearance of network externalities is the complementarity between the components of a network. Depending on the network, the externality may be direct or indirect. When customers are identified with components, the externality is direct. Consider for example a typical two-way network, such as the local telephone network of Figure 2. In this n-nodes 2-way network, there are 2n(n - 1) potential goods. An additional (n + 1th) customer provides direct externalities to all other customers in the network by adding 2n potential new goods through the provision of a complementary link (say ES) to the existing links.

16. In typical one-way networks, the externality is only indirect. When there are *m* varieties of component D and *n* varieties of component S as in Figure 4 (and all D-type goods are compatible with all S-type), there are mn potential composite goods. An extra customer yields indirect externalities to other customers, by increasing the demand for components of types A and B and thereby (because of the presence of

economies of scale) potentially increasing the number of varieties of each component that are available in the market.

17. Financial exchange networks also exhibit indirect network externalities. There are two ways in which these externalities arise. First, externalities arise in the act of exchanging assets or goods. Second, externalities may arise in the array of vertically related services that compose a financial transaction. These include the services of a broker, of bringing the offer to the floor, matching the offer, etc. The second type of externalities are similar to other vertically-related markets. The first way in which externalities arise in financial markets is more important.

18. The act of exchanging goods or assets brings together a trader who is willing to sell with a trader who is willing to buy. The exchange brings together the two complementary goods, "willingness to sell at price p" (the "offer") and "willingness to buy at price p" (the "counteroffer") and creates a composite good, the "exchange transaction." The two original goods were complementary and each had

no value without the other one. Clearly, the availability of the counteroffer is critical for the exchange to occur. Put in terms commonly used in Finance, minimal *liquidity* is necessary for the transaction to occur.

19. Financial markets also exhibit positive size externalities in the sense that the increasing size (or thickness) of an exchange market increases the expected utility of all participants. Higher participation of traders on both sides of the market (drawn from the same distribution) decreases the variance of the expected market price and increases the expected utility of risk-averse traders. *Ceteris paribus*, higher liquidity increases traders' utility. Thus, financial exchange markets also exhibit network externalities.

20. As we have noted earlier, network externalities arise out of the complementarity of different network pieces. Thus, they arise naturally in both one- and two-way networks, as well as in vertically-related markets. The value of good X increases as more of the complementary good Y is sold, and vice versa. Thus, more of Y is sold as more X is sold. It follows that the value of X increases as more of it is sold. This positive feedback loop seems explosive, and indeed it would be, except for the inherent downward slope of the demand curve.

21. To understand this better, consider a fulfilled expectations formulation of network externalities. Let the willingness to pay for the nth unit of the good when n^e units are expected to be sold be $p(n; n^e)$. [In this formulation n and n^e are normalized so that they represent market coverage, ranging from 0 to 1, rather than absolute quantities.] $p(n; n^e)$ is a decreasing function of its first argument because the demand slopes downward. $p(n; n^e)$ increases in n^e; this captures the network externalities effect, i.e., the good is more valuable when the expected

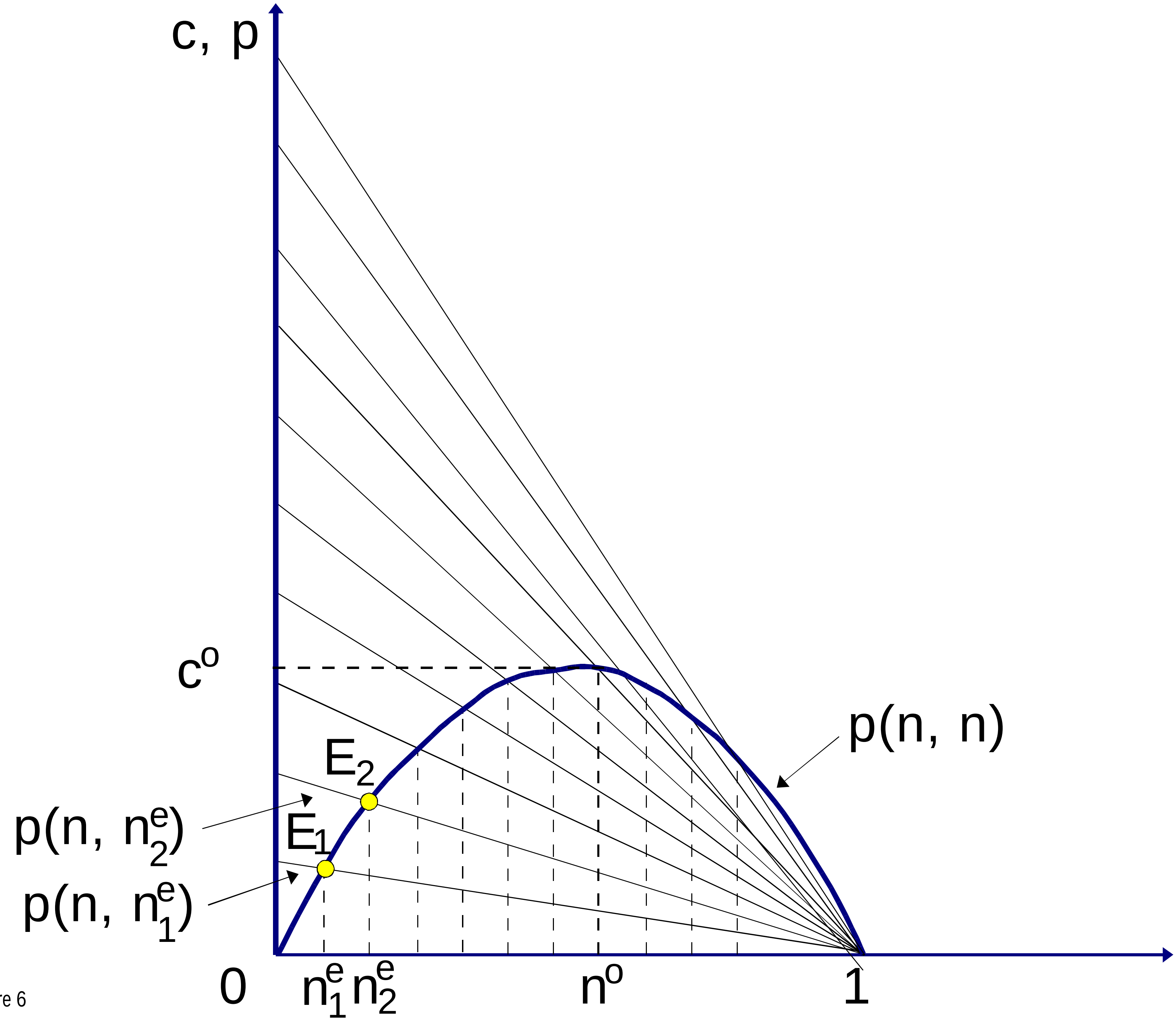

Figure 6

sales n^e are higher. At a market equilibrium of the simple single-period world, expectations are fulfilled, $n = n^e$, thus defining the fulfilled expectations demand $p(n, n)$.

22. Figure 6 shows the construction of a typical fulfilled expectations demand in a network industry. Each willingness-to-pay curve $p(n, n_i^e)$, i = 1, 2, ..., shows the willingness to pay for a varying quantity n, given an expectation of sales $n^e = n_i^e$. At $n = n_i^e$, expectations are fulfilled and the point belongs to $p(n, n)$ as $p(n_i^e, n_i^e)$. Thus $p(n, n)$ is constructed as a collection of points $p(n_i^e, n_i^e)$. It is reasonable to impose the condition $\lim_{n\to 1} p(n, n) = 0$. This means that, as the market is more and more covered, eventually we reach consumers who are willing to pay very little for the good, despite the fact that they are able to reap very large network externalities. It follows that $p(n, n)$ is decreasing for large n. In Figure 6, the fulfilled expectations demand at quantity zero is $p(0, 0) = 0$. This means that consumers think that the good has negligible value when its sales (and network effect) are zero. Although this is true for many network goods, some network goods have

positive inherent value even at zero sales and no network effects. If the good has an inherent value k, p(0, 0) = k, the fulfilled expectations demand curve in Figure 6 starts at (0, k).

23. Economides and Himmelberg (1995) show that the fulfilled expectations demand is increasing for small n if either one of three conditions hold:

(i) the utility of every consumer in a network of zero size is zero; or

(ii) there are immediate and large external benefits to network expansion for very small networks; or

(iii) there is a significant number of high-willingness-to-pay consumers who are just indifferent on joining a network of approximately zero size.

The first condition is straightforward and applies directly to all two-way networks, such as the telecommunications and fax networks where the good has no value unless there is another user to connect to. The other two conditions are a bit more subtle, but commonly observed in networks and vertically-related industries. The second condition holds for networks where the addition of even few users increases significantly the value of the network. A good example of this is a newsgroup on an obscure subject, where the addition of very few users starts a discussion and increases significantly its value. The third condition is most common in software markets. A software application has value to a user even if no one else uses it. The addition of an extra user has a network benefit to other users (because they can share files or find trained workers in the specifics of the application), but this benefit is small. However, when large numbers of users are added, the network benefit can be very significant.

24. **Critical Mass.** When the fulfilled expectations demand increases for small n, we say that ***the network exhibits a positive critical mass under perfect competition***. This means that, if we imagine a constant marginal cost c decreasing as technology improves, the network will start at a positive and significant size n^o (corresponding to marginal cost c^o). For each smaller marginal cost, $c < c^o$, there are three network sizes consistent with marginal cost pricing: a zero size network; an

unstable network size at the first intersection of the horizontal through c with p(n, n); and the Pareto optimal stable network size at the largest intersection of the horizontal with p(n, n). The multiplicity of equilibria is a direct result of the coordination problem that arises naturally in the typical network externalities model. In such a setting, it is natural to assume that the Pareto optimal network size will result.

25. In the presence of network externalities, it is evident that *perfect competition is inefficient*. The marginal social benefit of network expansion is larger than the benefit that accrues to a particular firm under perfect competition. Thus, perfect competition will provide a smaller network than is socially optimal, and, for some relatively high marginal costs, perfect competition will not provide the good while it is socially optimal to provide it.

26. Since perfect competition is inefficient, state subsidization of network industries is beneficial to society. The Internet is a very successful network that was subsidized by the US government for many

years. However, the subsidized Internet was aimed at promoting interaction among military research projects. During the period of its subsidization, almost no one imagined that the Internet would become a ubiquitous commercial network. Despite that fact, the foundation of the Internet on publicly and freely available standards has facilitated its expansion and provided a guarantee that no firm can dominate it.

III. STRATEGIC CHOICES IN NETWORK INDUSTRIES

27. **Standards Wars**. As we have discussed earlier, a network good has higher value because of the existence of network effects. Different firms conforming to the same technical standard can create a larger network effect while still competing with each other in other dimensions (such as quality and price). But even the decision to conform to the same technical standard is a strategic one. A firm can choose to be compatible with a rival and thereby create a larger network effect and share it with the rival. A firm could alternatively choose to be incompatible with the rival, but keep all the network effects it creates to itself. Which way the decision will go depends on a number of factors. First, in some network industries, such as telecommunications, interconnection and compatibility at the level of voice and low capacity data transmission is mandated by law. Second, the decision will depend on the expertise that a firm has on a particular standard (and therefore on the costs that it would incur to conform to it). Third, the choice on compatibility will depend on the relative benefit of keeping all the network effects to itself by choosing

incompatibility versus receiving half of the larger network benefits by choosing compatibility. Fourth, the choice on compatibility depends on the ability of a firm to sustain a dominant position in an ensuing standards war if incompatibility is chosen. Finally, the compatibility choice depends on the ability of firms to leverage any monopoly power that they manage to attain in a regime of incompatibility to new markets.

28. Standards may be defined by the government (as in the case of the beginning of the Internet), a world engineering body (as in the case of the FAX), an industry-wide committee, or just sponsored by one or more firms. Even when industry-wide committees are available, firms have been known to introduce and sponsor their own standards. Our discussion is on the *incentives* of firms to choose to be compatible with others.

29. We first examine the simple case when standardization costs are different and firms play a coordination game. A 2X2 possible version of this game is presented below. Entries represent profits. In this version, there is full compatibility at both non-cooperative equilibria.

		Player 2	
		Standard 1	Standard 2
Player 1	Standard 1	(a, b)	(c, d)
	Standard 2	(e, f)	(g, h)

30. Standard 1 is a non-cooperative equilibrium if $a > e, b > d$. Similarly, standard 2 is an equilibrium if $g > c, h > f$. In this game, we will assume that firm i has higher profits when "its" standard i get adopted, $a > g$, $b < h$. Profits, in case of disagreement, will depend on the particulars of the industry. One standard assumption that captures many industries is that in case of disagreement profits will be lower than those of either standard, $e, c < g$; $d, f < b$. Under these circumstances, the setting of either standard will constitute a non-cooperative equilibrium. There is no guarantee that the highest joint profit standard will be adopted. And, since consumers surplus does not appear in the

matrix, there is no guarantee of maximization of social welfare at equilibrium.

31. The same standards game can have a different version where each side likes its own sponsored standard no matter what the opponent does, i.e., each side has a dominant strategy. This game results in an in incompatibility equilibrium.

		Player 2	
		Standard 1	Standard 2
	Standard 1	↑ (a, b) →	(c, d) ↑
Player 1			
	Standard 2	(e, f) →	(g, h)

32. To understand the relative benefits of the compatibility decision, we need to examine the industry structures that would arise under either choice. When all firms are compatible, one expects equality to the extent that it is the rule in non-network industries. However, in industries exhibiting strong network externalities, in a regime of total

incompatibility (where each firm has its own incompatible standard), we expect to observe extreme inequality in market shares and profits. This is commonly observed in the computer software and hardware industries and in most of the new markets created by the Internet. Sometimes, such extreme inequality is commonly explained in industry circles by attribution to history. Stories abound on who or which company "was at the right place at the right time" and therefore now leads the pack. Traditional economic theory cannot easily explain such extreme inequality and may also resort to "managerial," "entrepreneurship," or "historical" explanations which are brought over in economics only when all else fails. As a last resort, if all else fails to explain a market phenomenon, economists tend to dismiss what they cannot explain as an "aberration" or a temporary phenomenon that will certainly disappear in the long equilibrium! Such explanations are deficient not only because they may be incorrect, but also because they tend to treat situations as isolated events and therefore lose all potential predictive power that is derived from correct modeling of economic phenomena.

33. There is a simple explanation of market structure in network industries without resorting to managerial, entrepreneurship, or historical explanations. The explanation is based on two fundamental features that network industries have and other industries lack: the existence of network externalities and the crucial role of technical compatibility in making the network externalities function.

34. Firms can make a strategic choice on if they are going to be compatible with others, and, sometimes, on if they will allow others to be compatible with them. The ability of a firm to exclude other firms from sharing a technical standard depends on the property rights that a firm has. For example, a firm may have a copyright or a patent on the technical platform or design, and can therefore exclude others from using it.

35. Compatibility with competitors brings higher network externality benefits ("network effect") and therefore is desirable. At the same time, compatibility makes product X a closer substitute to competing products ("competition effect"), and it is therefore undesirable. In making a

choice on compatibility, a firm has to balance these opposing incentives. In a network industry, the traditional decisions of output and price take special importance since higher output can increase the network externalities benefits that a firm can reap.

36. Inequality in market shares and profitability is a natural consequence of incompatibility. Under incompatibility, network externalities act as a *quality feature* that differentiates the products. Firms want to differentiate their products because they want to avoid intense competition.

37. In making the choice between compatibility and incompatibility, firms take into account the intensity of the network externality. The more intense the network externality, the stronger is the incentive for a firm to break away and be incompatible from substitutes. It follows that, in industries with very intense network externalities, firms will choose incompatibility.

38. Incompatibility implies inequality. Inequality is accentuated by output expansion to increase the network externality effect. Moreover, a

firm of higher output has a higher perceived quality, and is therefore able to quote a higher price. Thus, the inequality in profits is even more acute than the inequality of outputs.

39. Suppose that there are $S = \{1, \ldots, S\}$ firms, and potentially $I = \{1, \ldots I\}$ technical platforms. Firms have the option to coordinate to the same platform (full compatibility), have incompatible designs (total incompatibility), or coordinate in groups to compatible platforms that are incompatible with others (partial incompatibility). For a benchmark, assume that all firms produce identical products, except for whatever quality is added to them by network externalities. Also assume that the no firm has any technical advantage in production over any other with respect to any particular platform and that there are no production costs. We consider here only the extreme case of "pure network goods" where there is no value to the good in the absence of network externalities. The summary of the equilibria under total incompatibility (which can be enforced when firms have proprietary standards) is in the following tables. Firm #1 has the largest sales, firm #2 is the second largest, etc.

Table 1: Quantities, Market Coverage, and Prices Under Incompatibility

Number of firms S = I	q_1	q_2	q_3	Market coverage $\Sigma^I_{j=i}\, q_j$	p_1	p_2	p_3	p_{last}
1	0.6666			0.6666	0.222222			2.222e-1
2	0.6357	0.2428		0.8785	0.172604	0.0294		2.948e-2
3	0.6340	0.2326	0.0888	0.9555	0.170007	0.0231	0.0035	3.508e-3
4	0.6339	0.2320	0.0851	0.9837	0.169881	0.0227	0.0030	4.533e-4
5	0.6339	0.2320	0.0849	0.9940	0.169873	0.0227	0.0030	7.086e-5
6	0.6339	0.2320	0.0849	0.9999	0.169873	0.0227	0.0030	9.88e-11
7	0.6339	0.2320	0.0849	0.9999	0.169873	0.0227	0.0030	0

Table 2: Profits, Consumers' and Total Surplus Under Incompatibility

Number of firms S = I	Π_1	Π_2	Π_3	Profits of Last Firm Π_{last}	Total Industry Profits $\Sigma^I_{j=i}\, \Pi_j$	Consumers' surplus CS	Total Surplus TS
1	0.1481			0.1481	0.1481	0.148197	0.29629651
2	0.1097	7.159e-3		7.159e-3	0.1168	0.173219	0.29001881
3	0.1077	5.377e-3	3.508e-4	3.508e-4	0.1135	0.175288	0.28878819
4	0.1077	5.285e-3	3.096e-4	1.474e-5	0.1132	0.175483	0.28868321
5	0.1077	5.281e-3	2.592e-4	8.44e-7	0.1132	0.175478	0.28867817
6	0.1077	5.281e-3	2.589e-4	1.18e-14	0.1132	0.175478	0.28867799
7	0.1077	5.281e-3	2.589e-4	0	0.1132	0.175478	0.28867799

40. The market equilibria exhibit extreme inequality. The ratio of outputs of consecutive firms is over 2.6. The ratio of profits of consecutive firms is even larger. Entry after the third firm has practically no influence on the output, prices, and profits of the top three firms as well as the consumers' and producers' surplus. From the fourth one on, firms are so small that their entry hardly influences the market.

41. Although consumers' surplus is increasing in the number of active firms, total surplus is *decreasing* in the number of firms. That is, the more firms in the market, the *lower* is total welfare. This remarkable result comes from the fact that when there are fewer firms in the market there is more coordination and the network effects are larger. As the number of firms decreases, the positive network effects increase more than the dead weight loss, so that total surplus is maximized at monopoly!

42. Compared to the market equilibrium under compatibility, the incompatibility equilibrium is deficient in many dimensions. Consumers' and total surplus are higher under compatibility; the profits

of all except the highest production firm are lower under incompatibility; and prices are lower under compatibility except for duopoly.

43. The remarkable property of the incompatibility equilibrium is the extreme inequality in market shares and profits that is sustained under conditions of free entry. Antitrust and competition law have placed a tremendous amount of hope on the ability of free entry to spur competition, reduce prices, and ultimately eliminate profits. In network industries, as shown in this paper, free entry brings into the industry an infinity of firms but it fails miserably to reduce or to flatten the distribution of market shares. Entry does not eliminate the profits of the high production firms. And, it is worth noting that, at the equilibrium of this market, there is no anti-competitive behavior. Firms do not reach their high output and market domination by exclusion, coercion, tying, erecting barriers to entry, or any other anti-competitive behavior. The extreme inequality is a natural feature of the market equilibrium.

44. Another feature of the equilibrium discussed earlier is the fact that total surplus is highest at monopoly while consumers' surplus is lowest

at monopoly. This poses an interesting dilemma for antitrust authorities. Should they intervene or not? In non-network industries, both consumers' and total surplus are lowest at monopoly. In this network model, maximizing consumer's surplus would imply minimizing total surplus.

45. Whatever the answer to the previous dilemma, there is an even more difficult problem for antitrust authorities. At the long run equilibrium of this model, free entry is present and an infinity of firms have entered, but the equilibrium is far from competitive. No anti-competitive activity has lead firms to this equilibrium. Traditional antitrust intervention cannot accomplish anything because the conditions such intervention seeks to establish already exist in this market. Unfortunately the desired competitive outcome is not.

46. Can there be an improvement over the market incompatibility equilibrium? Yes, a switch to the compatibility equilibrium which has higher consumers' and total surpluses for any number of firms. Is it within the scope of competition law to impose such a change? It

depends. Firms may have a legally protected intellectual property right that arises from their creation of the design of the platform. Only if anti-competitive behavior was involved, can the antitrust authorities clearly intervene.