Notes on Network Economics

for the MBA class

"Networks, Telecommunications Economics,

and Digital Convergence"

Prof. Nicholas Economides

Spring 2009

Notes on Network Economics

Contents

1. Introduction

This course is about the economics of network industries. The crucial defining feature of a network industry is the existence of *network effects*. An market has network effects when, everything else being equal, the willingness to pay of a consumer increases with the number of units sold or expected to be sold. In the presence of *full compatibility*, each firm is able to reap the network effects from the whole industry. In the absence of full compatibility, a firm is able to reap the network effects from a set of firms which are compatible with it, which may even consist of only itself.

The economics (market structure and shares, pricing, investment) of a network industry are fundamentally different depending on whether a firm is able to individually define and alter technical standards so that its products are compatible or incompatible with other firms. When firms are able to lawfully alter technical standards and be incompatible with other firms, there are fundamentally new and interesting features of market structure. Moreover, the decision whether to be compatible with other firms, and to what extent, is a fundamental strategic decision of firms in these markets. We will study this class of markets in the first part of the course.

On the other hand, in a regulated industry, and most notably in many telecommunications markets, firms are required to be compatible with competitors. In telecommunications, this is guaranteed by mandatory interconnection among all service providers. Of course, this maximizes network effects, and simultaneously eliminates the possibility of choosing incompatibility. Thus, in a regulated network industry the issues are usually confined to the special case of full compatibility. Regulation also raises a number of new issues. In practice, only some telecommunications markets are regulated and many of these are progressively deregulated. This raises the issue of which telecommunications markets to deregulate and how to shield markets from potentially anti-competitive effects. We will study the telecommunications sector in the second part of the course.

2. Fundamentals of Network Industries

Network industries are a large part of the world economy. A key network industry is telecommunications, providing voice and data services, including the Internet and the world wide web. Another key network industry is computer software and hardware. These two sectors, telecommunications and computers, have been the engines of fast growth of the world economy. In the news and entertainment sector, network industries include broadcasting and cable television, which in recent years are reaching into traditional telecommunications services. In transportation, networks include airlines, railroads, roads, and shipping, and the delivery services that "live" on these, such as the postal service and its competitors. In the financial sector, networks include traditional financial exchanges for bonds, equities, and derivatives, clearing houses, B2B and B2C exchanges, credit and debit card networks, as well as automated transactions banking networks, such as ATM networks.

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 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, as will be explained in detail. So, to understand the "new economy" we need to understand the economics of networks. Many other traditional industries, such as railroads and electricity are also network industries.

Besides network industries where the network is immediately apparent, many of the features of networks apply to *virtual networks*. A virtual network is 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 can be thought of as a virtual network. Compatible computer software and hardware make up a network, and so do computer operating systems and compatible applications. More generally, networks are composed of complementary components, so they also encompass wholesale and retail networks, as well as information networks and servers such as telephone yellow pages, Yahoo, Google, etc.

Many network industries exhibit increasing returns to scale in production: unit (average) cost decreases with increasing scale of production, such in production of microchips and computer hardware. Often incremental cost is negligible (for example in software). However, these are also features of non-network industries and are *not* the defining feature of network industries. Thus, increasing returns to scale in production is also not the defining feature of the competition policy issues that are rooted in the existence of networks.

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 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 substitutes for each of these components; for example, transmission can be done through a cable TV line, a fixed telephone line, a wireless satellite, PCS, 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 substitute but not identical services. Thus, the information superhighway will provide substitutes made of complements; this is a typical feature of networks.

A common and defining feature of network industries is the fact that they exhibit increasing returns to scale in consumption, commonly called network effects. The existence of network externalities is the key reason for the importance, growth, and profitability of network industries and the “new economy.” A market exhibits

network effects (or network externalities) when the value to a buyer of an extra unit is higher when more units are sold, everything else being equal.

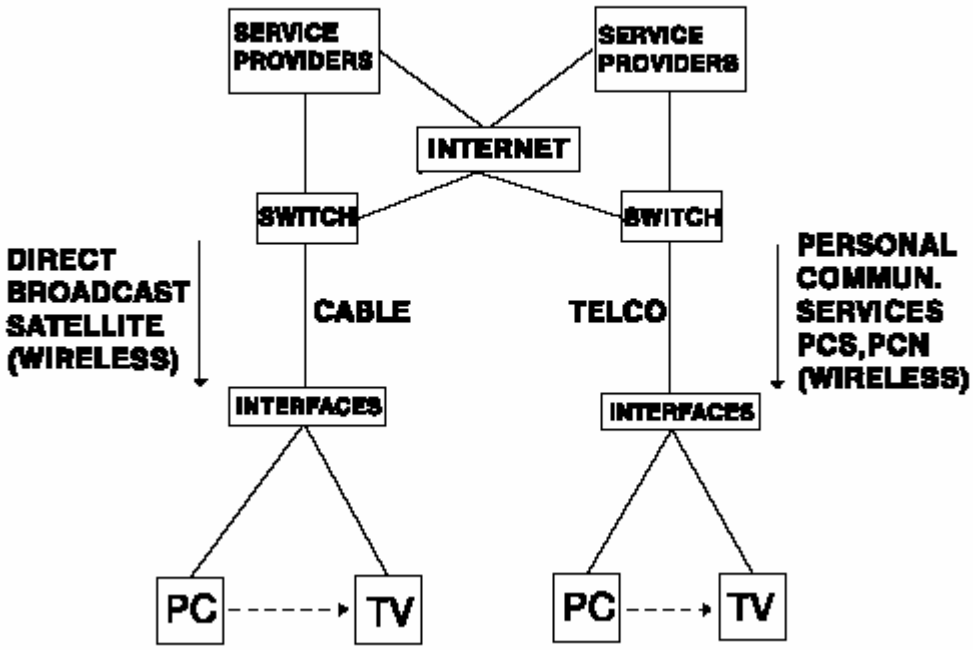


Figure 1: The Information Superhighway

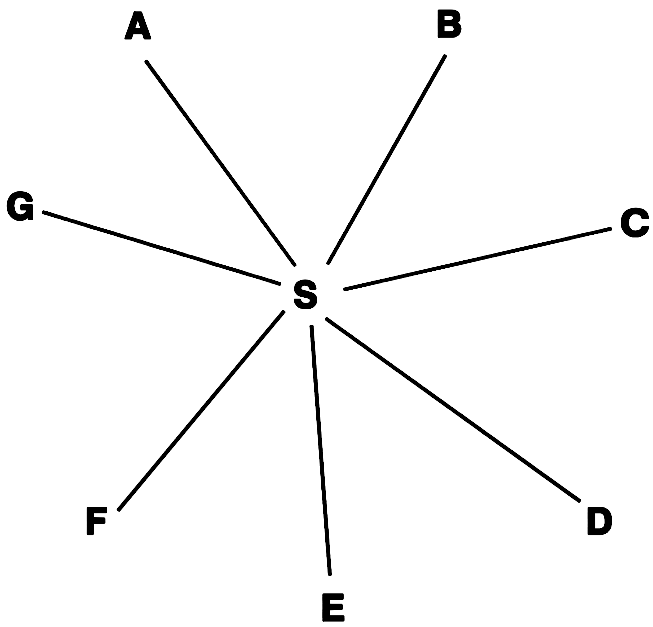


Figure 2: A star network

Network effects arise because of complementarities. In a traditional network, network externalities arise because a typical subscriber can reach more subscribers in a larger network. See Figure 2 which depicts a traditional telecommunications network where customers A, B, …, G are connected to a switch at S. Although goods "access to the switch" AS, BS, …, GS have the same industrial classification and traditional economics would classify them as substitutes, they are used as *complements*. In particular, when customer A makes a phone call to customer B, he uses *both* AS and BS.

2.1 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, radio and TV broadcasting and early paging networks are one-way networks.

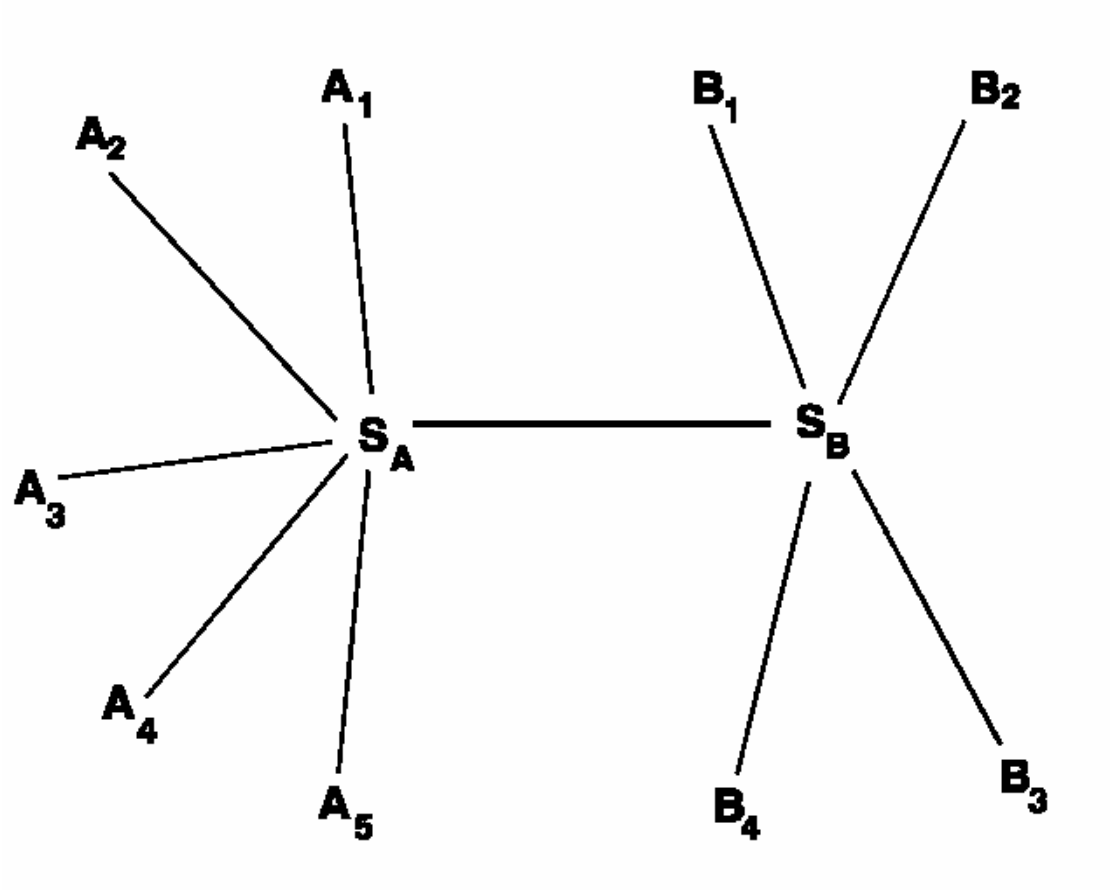
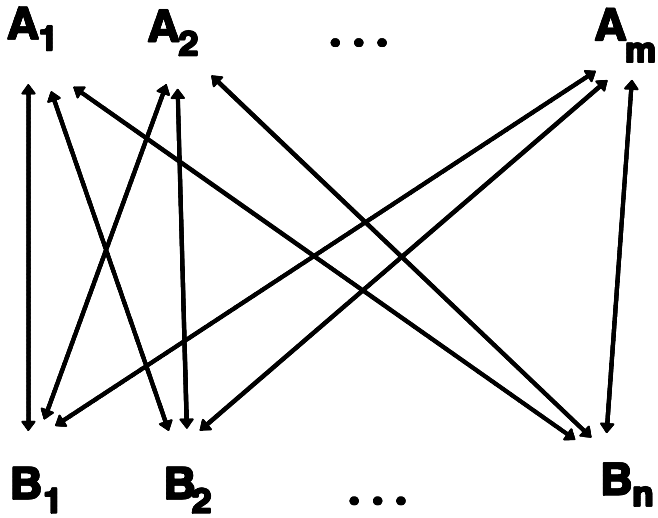


Figure 3: A long distance network

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 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 $A_iS_AS_BB_j$.

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.

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 A and B, 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.



<u>Figure 4</u>: A virtual network of complementary goods

2.2 <u>Compatibility</u>

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-format video player is compatible with a VHS-format 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. Similarly we say that two software products are compatible (more precisely *two-way* compatible) when they each can read and write files in a common format. Clearly, compatibility may be *one-way* when one the files of format B_1 of software A_1 can be read by software A_2 , but the files format B_2 of software A_2 cannot be read by software A_1. Moreover, compatibility may be only partial in the sense that software A_1 is able to read files of format B_2 but unable to write files in that format.

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. As we will see, it is not always in the best interests of a firm to allow full compatibility of its products with those of its competitors. The extent to which a firm is compatible with the products of other firms is an important strategic decision for a firm, and will be discussed in detail further on.

2.3 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 XP 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 $\{A_1, \ldots, A_m\}$ and $\{B_1, \ldots, B_n\}$ such that (i) each of the A-type good is a substitute to any other A-type good; (ii) each of the B-type good is a substitute to any other B-type good; and (iii) each of the A-type good is a complement to any A-type good. Virtual networks are one-way networks. Examples of virtual networks: computer hardware and software; computer operating systems and software applications, cameras and compatible film, razors and compatible blades etc. Clearly there are many more virtual networks than traditional networks.

In a *virtual network,* externalities arise because larger sales of components of type A induce larger availability of complementary components B1, ..., Bn, thereby increasing the value of components of type A. See Figure 4. 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.

3. Network Effects

3.1 Sources of Network Effects

In traditional non-network industries, the willingness to pay for the last unit of a good decreases with the number of units sold. This is called *the law of demand*, and is traditionally considered to hold for almost all goods. 1 However, the existence of network effects implies that, as more units are sold, the willingness to pay for the last unit may be higher. This means that for network goods, the fundamental law of demand is violated: for network goods, some portions of the curve demand can slope upwards. This means that, for some portions of the demand curve, as sales expand, people are willing to pay more for the last unit.

The law of demand is still correct if one disregards the effects of the expansion of sales on complementary goods. But, as increased sales of a network good imply an expansion in the sales of complementary goods, the value of the last unit increases. Combining the traditional downward slopping effect with the positive effect due to network expansion can result in a demand curve that has an upward-slopping part.

The key reason for the appearance of network externalities is the complementarity between network components. Depending on the network, the network effect 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 1. 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.

In typical one-way networks, the network effect is only indirect. When there are m varieties of component A and n varieties of component B as in Figure 2 (and all A-type goods are compatible with all of B-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. In the presence of economies of scale in production, the increase in demand may potentially increase the number of varieties of each component that are available in the market.

Exchange networks (financial networks such as the NYSE and NASDAQ, commodities, futures, and options exchanges as well as business to business "B2B" exchanges) 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, 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.

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.

Financial and business-to-business exchanges 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.

3.2 Network Effects Under Compatibility and Perfect Competition

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.

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). In this formulation, n and n_e are normalized so that they represent market coverage, ranging from 0 to 1, rather than absolute quantities. Willingness to pay $p(n; n^e)$ 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., that the good is more valuable when the expected sales n^e are higher. At a market equilibrium of the simple single-period world, expectations are fulfilled, n = ne, thus defining the fulfilled expectations demand p(n, n).

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).

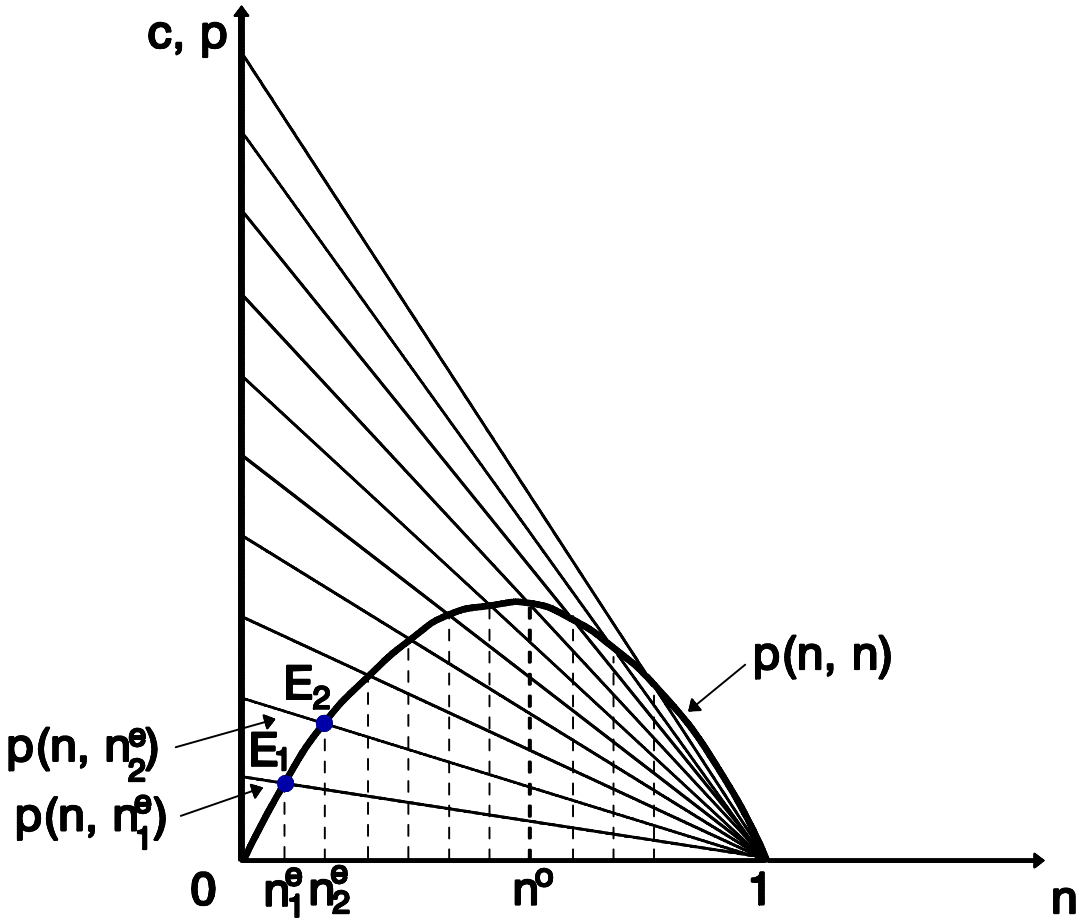


Figure 6: The fulfilled expectations demand and critical mass

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.

3.2.1 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.

3.2.2 Multiplicity of Equilibria

The existence of an upward slopping part of the demand curve and the multiplicity of equilibria even under perfect competition also allows for a network to start with a small size and then expand significantly. Suppose, for example, that marginal cost is at $c < c_o$ and a new invention creates a new product with significant network effects. Then, it is possible that the industry starts at the left intersection of the horizontal at c with $p(n, n)$ as expectations are originally low, and later on advances suddenly and quickly to the right intersection of the horizontal at c with $p(n, n)$. Thus, the multiplicity of equilibria in network industries can lead to sudden significant expansions of network size.

3.2.3 Efficiency

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.

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.

4. General Features of Markets with Network Effects

4.1 Ability to Charge Prices on Both Sides of a Network

There are a number of fundamental properties of network industries that arise out of the existence of network effects. First, a firm can make money from either side of the network. For example, a telecommunications services provider can charge subscribers when they originate calls or when they receive calls or for both. When a network consists of software clients and servers, both provided by the same firm, the firm can use the prices of the client and server software to maximize the network effect and its profits. For example it can distribute the client software at marginal cost (free) and make all its profits from the server software. In a similar vein, Adobe distributes the "Acrobat Reader" free while it makes its profits from the "Acrobat Distiller" product that allows the creation of files that can be read by the Acrobat Reader. The availability of prices on both sides of the network allows for complex pricing strategies, and, depending on the dynamics and market shares on the two sides of the market, can be used strategically to enhance and leverage a firm's strong strategic position on one side of the network. Of course, this is not confined to high technology or software industries, and applies wherever complementary components are present.

4.2 Externalities Internalized or Not

In network industries, often the additional subscriber /user is not rewarded for the benefit that he/she brings to others by subscribing. Hence typically there are "externalities," i.e., benefits not fully intermediated by the market. However, firms can use price discrimination to provide favorable terms to large users to maximize their network effect contribution to the market. For example, a large customer in a financial market can be given a very low price to be compensated for the positive network effect it brings to the market. It is anecdotally known that Cantor Fitzgerald, which has a 70% market share in the secondary market for US government 30-year bonds, offered to Salomon (the largest "primary dealer" and trader of US bonds) prices equal to 1/10 to 1/5 of those charged to small traders. This is consistent with profit maximization by Cantor Fitzgerald because of the liquidity (network effect) brought to the market by Salomon, which is by far the largest buyer ("primary dealer") in the auctions of US government bonds.

4.3 Fast Network Expansion

Generally, the pace of market penetration (network expansion) is much faster in network industries than in non-network industries. In the earlier discussion on critical mass, we saw that, in a one-period model, as unit cost decreases, the network starts with significant market coverage. In the presence of frictions and not perfectly elastic supply, the network expansion is not instantaneous from 0 to n^o but rather is a rapid expansion following an S-shaped curve, as seen in Figure 7. This figure compares the market share expansion of a new good (diffusion) in presence (delta = 1) and absence (delta = 0) of network effects as a function of time. The self-reinforcing nature of network effects leads to a much faster expansion when they are present. For a detailed discussion, see Economides and Himmelberg (1995).

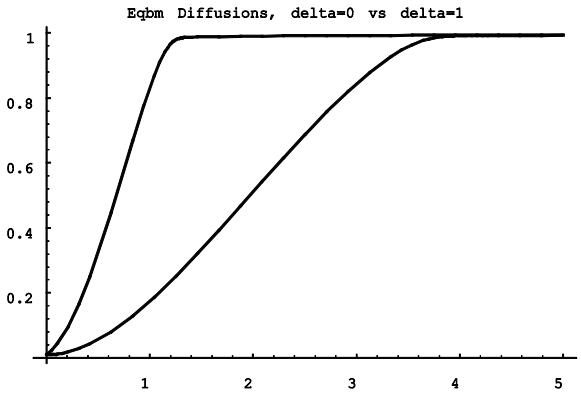


Figure 7: Equilibrium diffusion with and without network effects

5. Features of Markets with Network Effects Under Incompatibility and Imperfect Competition

5.1 Strategic Choices of Technical Standards and Compatibility In Network Industries

Standards Wars. A key strategic decision for a firm is the extent to which it will be compatible with other firms. 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.

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.

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)

<u>Figure 8</u>: Standards war leading to compatibility

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.

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

<u>Figure 9</u>: Standards war leading to incompatibility

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.

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.

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.

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.

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.

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 are more likely to choose incompatibility. As we will see in detail below, 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.

5.2 Inequality of Market Shares, Prices, and Profits

Markets with strong network effects where firms can chose their own technical standards are "winner-take-most" markets. In these markets, there is extreme market share and profits inequality. The market share of the largest firm can easily be a multiple of the market share of the second largest, the second largest firm's market share can be a multiple of the market share of the third, and so on. This geometric sequence of market shares implies that, even for a small number of firms n, the nth firm's market share is tiny. In equilibrium, there is extreme market share and profits inequality.

The reason for the inequality is straightforward. A firm with a large market share has higher sales of complementary goods and therefore its good is more valuable to consumers. This feeds back resulting in even higher sales. Conversely, a firm with small market share has lower sales of complementary goods, and the feedback results in even lower sales. However, the low sales firm is not necessarily driven out of business because that would require too low a price by the high sales firm. In the absence of fixed costs, an infinite number of firms can survive, but there is tremendous inequality in market shares, prices, and profits among them. Good examples of this market structure are the PC operating systems market and many software applications markets.

To understand the extent of market share, price, and profits inequality in network industries, we provide results from Economides and Flyer (1998). As a benchmark, they assume that all firms produce identical products, except for whatever quality is added to them by network externalities. Also assume that 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. The ith firm sells quantity q_i at price p_i, and firms are ordered in decreasing quantity so that $q_1 > q_2 > q_3$, i.e., firm #1 has the largest sales, firm #2 is the second largest, etc. The maximum potential sales is normalized to equal 1. At equilibrium not all consumers buy the good, that is, total sales = $\Sigma^{I}_{j=i}\, q_j < 1$.

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

Number of firms I	Sales of largest firm q_1	Sales of second firm q_2	Sales of third firm q_3	Market coverage $\Sigma^I_{j=i}\, q_j$	Price of largest firm p_1	Price of second firm P_2	Price of third firm p_3	Price of smallest firm p_i
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

Notice that even with no fixed costs and an *infinite* number of firms, the Herfindahl-Hirschman index at the equilibrium of this market is HHI = 0.464, which corresponds to between two and three firms of equal size.

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

Number of firms I	Π_1	Π_2	Π_3	Profits of Last Firm Π_I	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

The market equilibria exhibit extreme inequality. The ratio of outputs of consecutive firms is over 2.6. Ratios of prices of consecutive firms is at least 7. The ratio of profits of consecutive firms is about 20. This means that a firm that has about 38% of the sales of the immediately larger firm, can charge only 15% of the price of the next larger firm, and receives only 5% of the profits of the immediately larger firm. 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.

5.3 Monopoly May Maximize Total Surplus

In industries with significant network externalities, under conditions of incompatibility between competing platforms, monopoly may maximize social surplus. This is because, when strong network effects are present, a very large market

share of one platform creates significant network benefits for this platform, which contribute to large consumers' and producers' surpluses. It is possible to have situations where a breakup of a monopoly into two competing firms of incompatible standards *reduces* rather than increases social surplus because network externalities benefits are reduced. This is because *de facto* standardization is valuable, even if done by a monopolist.

In the Economides-Flyer model, 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 in a monopoly! Total surplus is highest while consumers' surplus is lowest in a monopoly. This poses an interesting dilemma for antitrust authorities. Should they intervene or not? In non-network industries, typically both consumers' and total surplus are lowest in a monopoly. In this network model, maximizing consumer's surplus would imply minimizing total surplus.

Compared to the market equilibrium under compatibility, the incompatibility equilibrium is deficient along many dimensions. Consumers' and total surplus are higher under compatibility; the profits of all except the highest production firm are higher under incompatibility; and prices are lower under compatibility except possibly in a duopoly.

5.4 No Anti-Competitive Acts are *Necessary* to Create Market Inequality

Because inequality is natural in the market structure of network industries, there should be no presumption that anti-competitive actions are responsible for the creation of market share inequality or very high profitability of a top firm. Thus, no anti-competitive acts are *necessary* to create this inequality. The "but for" benchmark against which anti-competitive actions in network industries are to be judged should not be "perfect competition" but an environment of significant inequality and profits.

5.5 In Network Industries, Free Entry Does Not Lead to Perfect Competition

In network industries, free entry does not lead to perfect competition. In a market with strong network effects, once few firms are in operation, the addition of new competitors, even under conditions of free entry, does not change the market structure in any significant way. Although eliminating barriers to entry can encourage competition, the resulting competition may not significantly affect market structure. This implies that, in markets with strong network effects, antitrust authorities may not be able to significantly affect market *structure* by eliminating barriers to entry. See the earlier example where the addition of the fifth firm hardly changes the output of the first four firms.

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, free entry brings into the industry an infinity of firms, but it fails miserably to reduce inequality in market shares, prices and profits. 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.

At the long run equilibrium of this model with free entry, an infinity of firms have entered yet the equilibrium is far from competitive. No anti-competitive activity has led firms to this equilibrium. Traditional antitrust intervention cannot accomplish anything because the conditions that such intervention seeks to establish already exist in this market. Unfortunately, the desired competitive outcome is not.

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.

5.6 Imposing a "Competitive" Market Structure is Likely to Be Counterproductive

In network markets, antitrust interventions may be futile. Because "winner takes most" is the natural equilibrium in these markets, attempting to superimpose a different market structure (say, one in which all firms have approximately equal market shares) may be both futile and counterproductive.

5.7 Nature of Competition is Different in Network Industries

In network industries, often competition *for the market* takes precedence over competition *in the market.* The fact that the natural equilibrium market structure in network industries is winner-take-most with very significant market inequality does not imply that competition is weak. Competition on which firm will create the predominant (top) platform and reap most of the benefits is, in fact, often intense. In network industries, there is typically an intense race to be the dominant firm. In network industries, we often observe Schumpeterian races for market dominance.

A good recent example of Schumpeterian competition is the competition among dot-coms in 1999-2000. As explained earlier, economic models imply a high valuation of the dominant firm compared to other firms in the same network industry. The same perception prevailed on Wall Street. During that period, dot-com firms advertised very intensely and subsidized consumers so as to be able to achieve the coveted dominant position in the market. The easy availability of capital for dot-coms at the time facilitated this behavior as firms "burned" almost all the cash they had in

their attempts to get the top market share. Many of the dot-coms failed because demand for their services was much lower than predicted or because of flaws in their business models. However, all the successful dot-coms, such as eBay, Amazon, and Yahoo, also followed this strategy.

Generally, in network industries, the costs of entry may be higher, but the rewards of success may also be higher compared to non-network industries.

5.8 Path Dependence

The existence of network effects underlines the importance of *path-dependence*. 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 (the installed base of VHS players). The existence of an installed base of consumers favors an incumbent. However, competitors with significant product advantages or a better pricing strategy can overcome the advantage of an installed base.

For example, in the market for video players, VHS overcame Beta after six years of higher installed base by Beta. This was an implication of

(i) Sony's mistake in disregarding network externalities and not licensing the Beta format;
(ii) Matsushita's widespread licensing of VHS;
(iii) The fact that one low-end, low-priced VHS player can contribute as much to the network effect as a high-end, high-priced Beta player.

In the Beta/VHS case, it is clear that Sony mistakenly ignored the network effects that arose from the availability of rental tapes of pre-recorded movies. The main function of video recorders was originally thought to be "time delay" in watching material recorded from the TV. The pre-recorded market emerged later, first as a market where movies were sold, and later as a movies rental market. The emergence of markets for "movies for sale" and "movies for rent," which had to be recorded in a particular format, created a significant complementary good for Beta and VHS players. The significant cost of physical distribution of tapes throughout the country and the costs of carrying a significant inventory of titles made the choice of what movies to bring and in what format crucially depended on present and forecasted demand. This forecast was highly correlated with the present and forecast installed base of video players in each format. Thus, although network effects and path dependence played a crucial role in determining the fate of Beta, the outcome was far from predetermined. Early, more aggressive licensing of the Beta format by Sony or the early promotion of low-end Beta players could have reversed the demise of the Beta format.

An often cited example on path dependence is the prevalence of the QWERTY keyboard despite claims of more efficient function by the alternative Dvorak keyboard. However, for many business applications, and for antitrust purposes, the QWERTY example is not crucial because there was no significant strategic business

interest in the success of either design. There is also a factual dispute on whether the Dvorak keyboard was significantly more efficient than the QWERTY one.

6. Competition Policy Issues in Network Industries

6.1 One-sided Bottlenecks

Interconnection issues in telecommunications, railroads, airline, and other transportation networks are very common. Often one company controls exclusively a part of the network, which is required by others to provide services. We call this network part "a bottleneck." Generally, bottlenecks can be divided into two categories: one-sided and two-sided. A one-sided bottleneck is monopolized by a firm and this firm does not require the use of a different bottleneck. Such a bottleneck is shown as link AB in Figure 10. An example of such a bottleneck is the connection of local service telecommunications subscribers to a switch. This is typically called "the last mile," and often called "the local loop." After the 1984 breakup of AT&T, the local loop has been monopolized by the local exchange carrier, typically a Regional Bell Operating Company ("RBOC") or GTE. Excluding cellular phones, the local loop is a required input in the production of long distance services, and typically long distance companies do not have a comparable local loop. Similarly, such a one-way bottleneck can arise when a firm monopolizes a railroad track such as AB. In telecommunications, the local exchange bottleneck has traditionally resulted in high prices for use of the bottleneck to originate ("access origination") or terminate calls ("access termination").

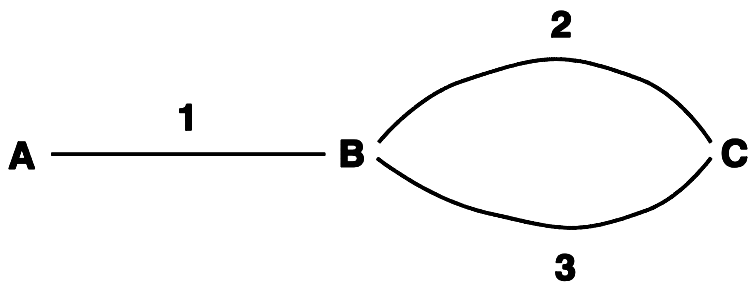


Figure 10: A one-sided bottleneck

The potential anti-competitive consequences of a one-sided bottleneck are obvious, and have been understood since the early days of the telecommunications network when AT&T enjoyed a monopoly in long distance (say here AB) but faced competition in local markets. In the context of Figure 10, the early AT&T was in possession of links 1 (long distance) and 2 (local), but did not allow an independent firm which possessed link 3 to interconnect at B and provide part of the long distance service CBA. For over two decades in the beginning of the twentieth century, AT&T refused to interconnect independent local telecommunications companies to its long distance network, unless they became part of the Bell System, which essentially meant unless they were acquired.

AT&T claimed that the main reason for its refusal to interconnect was the low technical standards of the independents, as well as incompatibilities, that would jeopardize AT&T's network after interconnection. While there is some truth to those claims, it is unlikely that they applied to all independents. Moreover, once acquired by AT&T, independents were interconnected with AT&T's network, after some modifications. This shows that the refusal to interconnect was mainly a strategic and commercial decision rather than a technical one.

The early AT&T foreclosure of independents through a refusal to interconnect shows the importance of complementarities in networks and the way that companies can leverage dominance in one market to create dominance in a market for complementary goods, especially when the complementary good requires the monopolized input to provide a final service. In this case, AT&T monopolized long distance and was able to leverage its position in long distance (through the refusal to interconnect with independent locals) and gain a dominant position in local telecommunications markets throughout the country.

The continued foreclosure of the independents by AT&T and its "refusal to deal" with them caused regulation to be established at the State and Federal levels in the 1930's. The 1934 Federal Communications Act ("1934 Act") imposed mandatory interconnection in an attempt to stop the foreclosure of independents and stabilized the market share of local lines held by AT&T. However, at that point AT&T's market share of local lines had already reached close to the 89% that AT&T had in 1981, prior to the 1982 agreement to be broken up.

A major revision of the 1934 Act, the Telecommunications Act of 1996 ("1996 Act"), mandates interconnection of all public switched telecommunications networks at any technically feasible point. The 1996 Act and similar European Union regulations attempt to solve the problem of the monopolization of the key parts of local telecommunications network. They impose unbundling of the network and forced leasing to entrants of some of the monopolized parts of the network, including the local loop. The goal is to make "mix and match" entry strategies feasible for local voice telephone service as well as broadband Internet access through digital subscriber lines ("DSL") that utilizes high frequencies transmission though copper local loops. Thus, they mandate access prices for unbundled parts of the network (unbundled network elements or "UNEs") at cost-based prices. The FCC and state PUCs accepted the view that lease prices should be based on *forward-looking* costs rather than on historical, accounting, or embedded costs (which was favored by RBOCs). In setting prices for unbundled network elements, the FCC and state public utility commissions ("PUCs") also rejected the relevance of prices based on *private* opportunity cost, such as the "Efficient Components Pricing Rule," ("ECPR"). Such rules derive prices for components from the monopoly prices of end-to-end services. Thus, the ECPR and its varieties would guarantee the monopolist's profits despite market structure changes in the markets for components that are used to create final services.[1] To prevent anti-competitive actions in telecommunications, the 1996 Act also imposes a number of rules, such as number portability, mandatory resale of services, transparency, non-discrimination, etc. A full discussion of these rules can be found at Economides (1999). Still, the 1996 Act missed opportunities to define

[1] See Economides and White (1995), and Economides (2003).

technical standards and require technical compatibility of telecommunications equipment.

Unfortunately, legal maneuvers by the incumbent local exchange carriers and high prices for the unbundled network elements considerably delayed very significant entry in local telecommunications markets.

6.2 Two-Sided Bottlenecks

In a two-sided bottleneck, each of two firms is a monopolist, each with a different bottleneck, and each firm requires the other's bottleneck to produce its output. For example, suppose there are two local telephone companies, each customer subscribes only to one local telephone company, and each company requires the other's network to complete calls. This could be represented in Figure 5 with the second link BC (number 3) removed, and considering AB to belong to firm 1, BC to belong to firm 2, firm 1 selling service ABC and firm 2 selling service CBA. In the context of this example, each of firms 1 and 2 buys *access termination* from the other. If each firm i =1, 2, sells both services ABC and CBA, then each firms buys both *access origination* and access termination from the other.

Many of the issues of traditional bottlenecks have been dealt by regulation in the United States and the European Union. In monopolized one-way bottlenecks, such as access origination and termination used in the creation of long distance calls, access prices are regulated and there has been a tendency to decrease the regulated prices. However, prices are still high. In the two-way bottleneck of access used in the creation of local calls by competing local exchange carriers, the Telecommunications Act of 1996 ("1996 Act") imposes cost-based reciprocal fees and allows the possibility of "bill and keep," i.e., zero prices. If cost- based reciprocal compensation were not the rule, and firms were able to set termination at profit maximizing levels, Economides, Lopomo and Woroch (1996a,b), have shown that a large network will try to impose very high termination charges on an opponent's small network so that no calls terminate from the small network to the large one. Without the possibility of such across-networks calls, a small network will only be able to provide within-network calls, and, being small, will be of little value to potential subscribers. As a consequence, large networks are able to provide more value to subscribers, and the small network is foreclosed. Starting from a regime of a large local incumbent and a small potential entrant, the large incumbent can set up termination access fees so that the entrant is kept out of the market.[2]

In summary, in the absence of specific regulatory rules, two-sided bottlenecks can lead to foreclosure of competitors, even when each firm requires use of the bottleneck of the other to complete calls or provide a service.

---

[2] This is not just a theoretical possibility. Telecom New Zealand ("TNZ"), operating in an environment of weak antitrust and regulatory intervention (so called "light-handed regulation"), offered such high termination fees that the first entrant into local telecommunications, Clear, survives only by refusing to pay interconnection fees to TNZ, while the second entrant, BellSouth New Zealand exited the local telecommunications market.

6.3 Market Power Creation Specific to Networks: The Importance of Technical Standards

The example of early AT&T's refusal to deal (interconnect) with independents (and with interconnected networks of independents) can also arise in milder terms when a firm X that has a significant position in its industry insists that firms that provide complementary products Y do not also provide them to X's competitors. For example, in the mid-1980s Nintendo refused to allow third party games (software) to play on its game console (hardware) unless the software manufacturers agreed not to write a similar game for competing game systems for two years. Faced with this condition imposed by Nintendo, software developers had to make a choice to either write a game for Nintendo or for the competing platforms of Atari and Sega. Clearly this restriction reduced the potential revenue of a game developer who would like, for a small additional cost, to port its game to the alternative systems. But, also, more importantly, the restriction forced developers to predict which game system would have higher sales, and create software just for this system. Thus, Nintendo used its dominance of the game market at that point in time to coerce developers to write software just for its platform, and thereby increased the value of the Nintendo virtual network (of hardware and software). Nintendo abandoned this requirement under antitrust challenge.

Because of the extreme inequality of market shares, prices, and profits in a network industry, restriction of the installed base of a firm in a network industry can be very detrimental. It can push a firm to a lower rank with significantly lower profits, or, in extreme cases, push a firm to such a low market share that it has to close down because it cannot recover its fixed costs.

Another example from the computing industry illustrates a situation of market power creation specific to networks. Suppose that firm A chooses to make its product A incompatible with the products of other firms that perform similar functions, and it also subsidizes firms that produce complementary goods B to its product A.[3] Alternatively, we may assume that firm A subsidizes its own division that sells complementary goods B. As a result

(i) the value of firm A's product increases;
(ii) the entry hurdle of firm A's rivals increases;
(iii) there is possible creation of market power.

Firm A's defense will be that its actions are pro-competitive since their primary cause is the enhancement of the value of product A. From the point of view of A's competitors, the actions of A look very much like anti-competitive behavior since the abundance of complementary goods B for product A puts them at a competitive disadvantage.

Note that the existence of incompatibility is a *necessary* condition for possible creation of market power. Moreover, the key to increasing social welfare is to move to compatibility. That is, assuming that innovation and product availability would not be reduced, the best of all worlds is to have public standards and full compatibility.

---

[3] For example, one can think of A as a computer operating system ("OS"), and B as an application. OS manufacturers can and do embed software routines that are useful to application software developers since they reduce the cost of writing applications.

However, it is very difficult for US antitrust authorities to intervene and/or define standards.

Besides the use of technical standards, firms can also use bundling and other pricing strategies as well as non-price discrimination strategies to leverage market power across markets.

6.4 Vertical Integration and Vertical Control Issues in Network Industries

In networks, as in other settings, there are potentially anti-competitive issues arising from the possibility of vertical integration and the behavior of vertically integrated firms. These may include, the bundling of components through vertical integration, contract, or manipulation of technical standards so that an entrant must enter both components markets even if it desires to enter only one of the markets. Often firms have expertise or a technical advantage in only one component, and would like to enter only in the market for that component. An incumbent can strategically alter the market environment through acquisition or contract so that the entrant can only be successful if it enters more than one markets. This increases the financial hurdle for an entrant, and it also forces it to sell components where it does not have expertise. Thus, it makes it more likely that entry will not occur.

A vertically integrated firm can also use discrimination in price charged to a subsidiary compared to the price charged to a downstream competitor, or discrimination in quality provided to subsidiary compared to quality provided to a downstream competitor, that is, raising rivals' costs. These issues are discussed in more detail in Economides (1996b).

Firms in network industries can also use a variety of way to manipulate technical standards in joint ventures to achieve market power. The issue of market power also arises in "aftermarkets," where consumers are "locked in" to a service that arises out of commitments of a durable nature. For example, in an important case that reached the Supreme Court, Kodak refused to supply parts to independent firms that serviced Kodak photocopiers. Although one could argue that there was significant competition in the market for new photocopiers and Kodak did not have a dominant position in that market, once customers had bought a Kodak photocopier, they were "locked in," and faced significant costs to buy a new photocopier of a different brand. So, Kodak's actions could be anti-competitive in the "aftermarket" for repair services of consumers who have already bought Kodak photocopiers. A similar case of anti-competitive actions can be made in aftermarkets where consumers are locked-in by having made an investment in a durable good that is incompatible with other comparable durable goods, or are locked in other ways. For example, consumers without number portability in wireless cellular and PCS markets may be "locked in" to the service of a particular provider or network. Similarly consumers can be "locked in" to the e-mail service of an Internet Service Provider ("ISP") since there is no portability of e-mail addresses and many ISPs do not offer forwarding of incoming e-mail messages to another ISP.

6.5 B2B and Other Exchanges Issues

The world of business to business exchanges lacks the regulation of traditional financial and commodity exchanges. Many proposed B2B exchanges are run by the firms that also are trading. For example, ENRON was proud of the fact that it was participating as a trading party in B2B exchanges that it organized and ran. Such a situation would be strictly prohibited in traditional financial and commodity exchanges because of the possibility that the organizer of the exchange would take advantage of the information created in the trading process to fashion privately beneficial trades. In another example, COVISINT, an exchange for automobile parts organized by automobile manufacturers has been accused of acting to consolidate the monopsony power of car manufacturers. In general, B2B exchanges can provide substantial benefits by consolidating trades, increasing market liquidity, improving standardization, and reducing search costs.[4] But B2B exchanges also have the potential of creating significant antitrust issues.

6.6 Dynamic Efficiency Issues

The world of networks and dynamic effects brings to the forefront the fact that behavior that exhibits static efficiency may lack dynamic inter-temporal efficiency. The possibility exists of a lock-in to a technology and a path which, when decisions are taken in every period, looks optimal given past decisions, but is sub-optimal if earlier investment decisions had been delayed and all the decisions were taken at once. In a world with network effects a "lock in" to an inferior technology can easily occur as firms (and countries) find it more desirable to invest in the technology in which they already invested. This can occur under perfect competition, but the problem can easily become much more important under oligopoly, as firms race to become dominant, given the importance of dominance in a network industry.

6.7 Innovation Issues

An important antitrust issue is the speed of innovation in a network industry as affected by strategic decisions of firms and potentially anti-competitive actions. The effects of actions on innovation are important because innovation affects the welfare of future consumers, and this should be taken into consideration in an antitrust decisions. The difficulty in dealing with innovation issues in antitrust arises from the fact that the efficiency and intensity of innovation in monopoly compared to perfect competition and oligopoly are open questions in economics. Thus, it is very hard to make general statements on innovation in an antitrust context.

6.8 Criteria to be Used for Antitrust Intervention in Network Industries

When an antitrust intervention is considered in a network industry, a number of considerations that arise out of the nature of network industries have to be taken account. These are explained in detail in earlier sections, primarily in section 3. First,

---

[4] See Economides and Siow (1988) for a discussion of the benefits of B2B and other exchanges.

the benchmark of the "but for" world that should be considered should be a network industries equilibrium with significant inequality, rather than a perfectly competitive equilibrium. Second, competitors' harm should not be a sufficient reason for intervention. The right question is, "were consumers (past, present, future) harmed by specific actions?" Third, uncertainty should be taken into account, and caution should be used in guessing how a high tech industry would have evolved but for the anti-competitive action(s). Fourth, it is possible that monopoly may maximize total surplus. Fifth, it will not be possible to sustain a long-term equilibrium with equal market shares, and a short term equilibrium with equal market shares may have low total surplus. Sixth, path dependence and the value of installed base are limited by Schumpeterian competition, and upheavals are not uncommon in network industries. Seventh, especially in software industries, the extent and functionality of products is flexible. This can help an incumbent because it can expand the functionality of its product, but can also help its rivals as they may incorporate functionalities of the incumbent's product in theirs.

6.9 Criteria to be Used for Remedies

When a remedies phase is reached, a liability finding has already been made. The objective of remedies is to stop practices that were found to be illegal, prevent the recurrence of such practices, and restore any recurring threat posed by such practices.

Any intervention by antitrust authorities creates a disruption in the workings of markets. The objective of the remedial relief is to accomplish the objectives mentioned in the previous paragraph without damaging efficient production and competition in the market. The potential damage that antitrust intervention can produce is larger when it is applied to an industry with fast technological change, where leaps to new and more efficient technologies are expected, while the specific nature of the future winning technology is unknown. Often, it is difficult to predict future winning technologies, and therefore it is hard to fashion an antitrust remedy with an accurate prediction of its effect on industry structure and competition a few years down the road. Of course, this uncertainty is multiplied when the remedy creates a significant intervention in the industry. Therefore, lacking the knowledge of the effects of their actions, it is in the public interest that antitrust authorities and courts avoid extensive intervention in industries with fast technological change. It is best to intervene only to the extent that (i) intervention reverses the effects of actions for which liability was established; and (ii) the effects of the intervention are predictable.

In markets with network effects, as I have explained in detail above, the existence of network effects has crucial implications on market structure and the ability of antitrust authorities to affect it. In markets with strong network effects, even in the absence of anti-competitive acts, the existence of network effects in markets, results in significant inequalities in market shares and profits. The resulting equilibrium market structure can be called a "*natural oligopoly*" where very few firms dominate the market. The structural features of natural oligopoly for a software market cannot be altered by antitrust intervention without significant losses for society. The very nature of markets with network effects implies that the ability of

antitrust authorities to alter market structure in such industries is limited, as discussed above.

As an alternative to antitrust and competition law, economic regulation can and has been established in three exceptional case: (i) for those markets where it is clear that competition cannot be achieved by market forces; (ii) where deviation from efficiency is deemed socially desirable; and (iii) where the social and private benefits are clearly different, since in each of these cases, it is clear that a market without intervention will not result in the desired outcome. I will leave case (ii) aside, since a discussion of it would lead us to a detailed discussion of specific industries. The requirements for case (iii) are typically met in many network industries, since expansion of the network creates network effects that are typically not fully internalized by markets. However, it would be foolish to advocate regulation as the standard solution in network industries because of the existence of network effects. Often, a much smaller intervention, such as subsidization of the network to help network effects will be enough.

In case (i), where it is clear that competition and its benefits cannot be achieved by market forces, regulation may be a solution. The significant advantage of industry-specific regulation is that it can be tailored to the specifics of the industry, and specific rules on pricing and availability of particular products and services. Regulators, such as the FCC also have staffs that can provide impartial technical advice that would be unavailable to a court.

However, regulation has a number of drawbacks. First, it is best suited for industries with well-defined and slow-changing products and services. With stable product definitions, rules can be devised and specific pricing can be implemented if necessary. Second, as a corollary to the first observation, regulation is not well suited in industries with rapid technological change and frequently changing product definitions. Moreover, in an industry with fast technical progress, regulation can be used by the regulated companies to keep prices relatively high, as exemplified by telecommunications regulation. Third, often regulators are very close to the interests of the regulated parties rather than to the interests of the public. Fourth, experience has shown that often regulators are not well informed about key variables as well as changes in the industry. Fifth, regulators at both the state and federal levels are under pressure and influence by both the executive and the legislative part of government, and cannot be as impartial as a court. Sixth, there is a tendency for regulators to expand their reach into related and new markets. For example, the California Public Service Commission has recently asserted its authority over the Internet! These drawbacks can create significant surplus loss due to regulation.

In summary, regulation should be used sparingly in industries with stable products, if it is clear that antitrust has failed, and keeping in mind that regulation can also cause a significant surplus loss.

7. References

Economides, Nicholas, (1996a), "The Economics of Networks," *International Journal of Industrial Organization*, vol. 16, no. 4, pp. 675-699, pre-publication copy at http://www.stern.nyu.edu/networks/94-24.pdf.

Economides, Nicholas, (1996b), "The Incentive for Non-Price Discrimination by an Input Monopolist," *International Journal of Industrial Organization*, vol. 16 (May 1998), pp. 271-284, pre-publication copy at http://www.stern.nyu.edu/networks/1136.pdf.

Economides, Nicholas, (1999), "The Telecommunications Act of 1996 and its Impact," *Japan and the World Economy,* vol. 11 (1999), pp. 455-483, pre-publication copy at http://www.stern.nyu.edu/networks/98-08.pdf.

Economides, Nicholas, (2001a), "The Microsoft Antitrust Case," *Journal of Industry, Competition and Trade: From Theory to Policy*, pre-publication copy at http://www.stern.nyu.edu/networks/Microsoft_Antitrust.final.pdf.

Economides, Nicholas, (2001b), "The Microsoft Antitrust Case: Rejoinder," *Journal of Industry, Competition and Trade: From Theory to Policy* (August 2001), vol. 1, no. 1, pp. 71-79, pre-publication copy at http://www.stern.nyu.edu/networks/Microsoft_Antitrust.Rejoinder.pdf.

Economides, Nicholas, (2002), "Amicus Brief of Nicholas S. Economides on the Revised Proposed Final Judgment in the Microsoft case," submitted to the United States Department of Justice, available at http://www.usdoj.gov/atr/cases/ms_tuncom/major/mtc-00022465.htm.

Economides, Nicholas, (2003), "The Tragic Inefficiency of M-ECPR," (Ed) Shampine, A. (2003), *Down to the Wire: Studies in the Diffusion and Regulation of Telecommunications Technologies.* Nova Science Publishers, Inc., New York, NY, pre-publication copy at http://www.stern.nyu.edu/networks/tragic.pdf.

Economides, Nicholas and Fredrick Flyer, (1998), "Compatibility and Market Structure for Network Goods," Discussion Paper EC-98-02, Stern School of Business, NYU, available at http://www.stern.nyu.edu/networks/98-02.pdf.

Economides, Nicholas and Charles Himmelberg, (1995), "Critical Mass and Network Evolution in Telecommunications," in *Toward a Competitive Telecommunications Industry: Selected Papers from the 1994 Telecommunications Policy Research Conference*, Gerard Brock (ed.), 1995, pre-publication copy at http://www.stern.nyu.edu/networks/tprc.pdf.

Economides, Nicholas, Giuseppe Lopomo, and Glenn Woroch, (1996a), "Regulatory Pricing Policies to Neutralize Network Dominance," *Industrial and Corporate Change,* vol. 5, no. 4, pp. 1013-1028, pre-publication copy at http://www.stern.nyu.edu/networks/96-14.pdf.

Economides, Nicholas, Giuseppe Lopomo, and Glenn Woroch, (1996b), "Strategic Commitments and the Principle of Reciprocity in Interconnection Pricing," Discussion Paper EC-96-13, Stern School of Business, at http://www.stern.nyu.edu/networks/96-13.pdf.

Economides, Nicholas and Robert Schwartz, (1995a), "Electronic Call Market Trading," *Journal of Portfolio Management*, vol. 21, no. 3, pp. 10-18, pre-publication copy at http://www.stern.nyu.edu/networks/Electronic_Call_Market_Trading.pdf.

Economides, Nicholas and Robert Schwartz, (1995b), "Equity Trading Practices and Market Structure: Assessing Asset Managers' Demand for Immediacy," *Financial Markets, Institutions & Instruments*, vol. 4, no. 4, (November 1995) pp. 1-46, pre-publication copy at http://www.stern.nyu.edu/networks/Equity_Trading_Practices.pdf.

Economides, Nicholas, and Aloysius Siow, (1988), "The Division of Markets is Limited by the Extent of Liquidity," *American Economic Review,* vol. 78, no. 1 (March 1988), pp. 108-121, at http://www.stern.nyu.edu/networks/aer88.pdf.

Economides, Nicholas and Lawrence J. White, (1994), "Networks and Compatibility: Implications for Antitrust," *European Economic Review*, vol. 38, pp. 651-662.

Economides, Nicholas and Lawrence J. White, (1995), "Access and Interconnection Pricing: How Efficient is the Efficient Component Pricing Rule?" *The Antitrust Bulletin*, vol. XL, no. 3, pp. 557-579, pre-publication copy at http://www.stern.nyu.edu/networks/95-04.pdf.

Economides, Nicholas and Lawrence J. White, (1998), "The Inefficiency of the ECPR Yet Again: A Reply to Larson," *The Antitrust Bulletin* vol. XLIII (1998), no. 2, pp. 429-444, pre-publication copy at http://www.stern.nyu.edu/networks/96-07.pdf.

Posner, Richard A., (2000), "Antitrust In The New Economy," *Tech Law Journal*, at http://www.techlawjournal.com/atr/20000914posner.asp.