Abstract

A network exists when a product’s value to the user increases as the number of users of the product grows. Each new user of the product derives private benefits, but also confers external benefits (network externalities) on existing users. Network externalities may cause markets to fail. Networks may not reach optimal size, because users fail to take account of external benefits. Markets in which incompatible standards compete may ‘tip’ in the direction of a standard that gains an early advantage, even if that standard is objectively inferior. Suppliers of network goods may compete to become the de facto standard or may attempt to make their products compatible. The theory of network externalities has been applied in numerous legal areas including antitrust (monopolization and cooperative standard-setting); intellectual property (the scope of protection for dominant software programs); and corporate law (the selection of contract terms).

JEL classification: K00

Keywords: Increasing Returns, Positive Feedback, Lock-In, Standardization

1. Introduction

The subject of network externalities has become popular in the economic and legal literature since the mid-1980s. The fundamental idea is that the act of joining a network confers a benefit on all other participants in the network. According to proponents, this seemingly simple phenomenon strongly influences competitive strategies and market outcomes in network markets. Network externalities can, for example, influence consumers’ decisions whether to adopt a new technology and producers’ decisions whether to standardize their products. Ultimately, the literature suggests, network externalities can cause markets to fail: ‘equilibrium may not exist, or multiple equilibria may exist’ and ‘the fundamental theorems of welfare economics may not apply’.

Network Externalities

Networks may not reach optimal size, because purchasers do not take account of social benefits of their purchases, and markets may converge on an inferior standard that effectively excludes better products.

The theory of network externalities is well developed, especially for such a new field of inquiry. Its policy implications are less clear. Some economists claim that the inefficiencies associated with network externalities are common in some important markets, including computer hardware and software, transportation and telecommunications. And some argue that network externalities justify changes in the law of antitrust, corporations, economic regulation and intellectual property. Some of the proposed policy changes would involve greater government intervention in the market, some would involve less. Other scholars suggest that courts and policymakers should be cautious in relying on network externalities. They argue that the theoretical market failures associated with network externalities would only occur under rare conditions. They also challenge the empirical evidence that network externalities have actually thwarted the adoption of better technologies. Any real market failures in network markets, they argue, probably stem from familiar economic phenomena (like natural monopoly) that are better analyzed using conventional tools. Even if the market has failed, government intervention is inappropriate because network externalities cannot be identified in practice or government action is likely to be more costly than any benefits it may produce. In the following discussion, we will address both theoretical and policy issues, trying wherever possible to keep them distinct.

A. Types of Network Externality

The users of certain products can be thought of as forming networks, either because they are physically connected or because they have close market relationships. A network externality is a benefit conferred on users of such a product by another’s purchase of the product. (A purchase may also impose external costs, but the literature focusses primarily on the sources and consequences of external benefits, or positive consumption externalities.) Some analysts distinguish between direct externalities, exemplified by communications networks, and indirect externalities, exemplified by the ‘hardware/software paradigm’ (Katz and Shapiro, 1985, 1994; Church and Gandal, 1992b). Note that the distinction between direct and indirect externalities refers to the source of benefit to participants in the network, not necessarily to the magnitude of the network effect.
2. Direct Externalities

Direct externalities typically occur in a physical, two-way communications network (Rohlfs, 1974). My purchase of a fax machine, for example, directly benefits existing fax machine owners, who now have an additional person with whom they may communicate. If there are \( n \) fax machines in the network, each owner has \( n(n-1) \) potential interlocutors; an additional fax machine adds \( 2n \) (total) potential communications within the system, and thus enhances the value of membership, assuming that each owner may at some point wish to communicate with every other owner (Economides, 1996). Similar reasoning applies to telephones and internet access software, each of which expands the number of people who can communicate over a physical network. Users of the network thus receive increasing returns in consumption. (Notice that direct benefits do not accrue to users of a one-way physical network, like paging or electric power transmission, because the existing users of the networked goods may not interact with new users any more easily than those who are not on the network (Economides and White, 1994; Economides, 1996).) Producers of network goods may also receive increasing returns to scale in production, at least up to some critical mass.

The ‘positive feedback effect’ (Arthur, 1989, 1990) of increased network size makes the larger network that much more attractive to new purchasers, and the goods that permit access to the network that much more valuable. The value of the good to the purchaser also depends on the purchaser’s expectations about the future size of the network. The demand for a fax machine is thus a function not only of the price of the product, but also of the expected size of the network to which the fax machine will be connected. This last point resolves the apparent paradox that, despite the downward slope of the demand curve, the marginal purchase of a good can yield a higher value than inframarginal goods: the value contributed by the expected size of the network offsets the reduction in value from the purchase of a marginal unit (Economides, 1996).

Direct network externalities might arise even in the absence of a physical network. Those who speak a language, for example, constitute a network. The value of knowing a language depends in part on how many others speak it. A student’s decision to learn a language is influenced by the expected size of the network of speakers; and the student’s learning the language directly benefits other speakers. Similarly, the value to an individual of a particular word processing program, say WordPerfect, likely will depend in part on the number of others who select WordPerfect and with whom the individual expects to exchange files. This effect is diminished to the extent that conversion between programs is possible, but, so long as conversion is imperfect or costly, the effect persists.
3. Indirect Externalities

Economists have also identified indirect externalities in the ‘network’ of users of systems of compatible devices, even if the devices owned by different users are not physically connected. A system can be any combination of a durable good and associated goods or services that perform some desired function. Katz and Shapiro (1994) illustrate this type of system with the hardware/software paradigm, which includes not only computer hardware and software, but many other product combinations, including cameras and film, phonographs and records, and television sets and programming. More broadly, a typewriter keyboard could be considered hardware, and the capability and experience of using that keyboard could be considered software. Owners of compatible hardware and software systems constitute a ‘virtual network’ (Katz and Shapiro, 1985, 1994; Arthur, 1989, 1990). At the extreme, any combination of complementary products can be described as a system, and those who purchase the system can be said to form a virtual network. Thus, a cappuccino machine, a coffee grinder, espresso coffee beans and milk form a system, and those who drink home-made cappuccino form a network.

Indirect externalities can arise in these markets only when, as is typical, the components are purchased at different times. For example, applications software programs are often purchased at various times over the useful life of a computer. In these circumstances, adoption of the hardware by one purchaser confers external benefits on other users of the same hardware, because it expands the installed base of the hardware, stimulating demand for compatible software. Suppliers may therefore take advantage of scale economies and provide more varieties of software. The availability of more diverse and inexpensive software enhances the value of the existing users’ hardware. These indirect benefits stemming from strong complementarities are said to cause positive feedback in much the same way as direct benefits of expanded communication in physical networks.

Indirect network externalities may arise in many contexts. Participants in physical networks, both one-way and two-way, may receive indirect external benefits, if the increased size of the network results in more options and better service. Traders on a financial exchange may also receive indirect benefits from the array of services provided by the exchange (Economides, 1993a, 1996). And users of Discover cards benefit when more people carry Discover cards, because more merchants will accept a widely-adopted card (Evans and Schmalensee, 1996).
4. Network Effects versus Network Externalities

Liebowitz and Margolis (1994, 1995a) distinguish network externalities from network effects. A network effect exists when ‘the net value of an action ... is affected by the number of agents taking equivalent actions’ (Liebowitz and Margolis, 1994, p. 135). Network effects, so defined, are ubiquitous in the economy. Purchases of a good by one group of consumers may bid up its price and thereby affect other consumers of that good as well as consumers of complementary and substitute products. Liebowitz and Margolis point out that economists at one time misinterpreted these interactions as inefficiencies, but now recognize them to be pecuniary external economies and diseconomies, which the price system internalizes as wealth transfers between, say, purchasers and suppliers. Liebowitz and Margolis (1994, p. 135) would limit the term network externality to those specific network effects in which ‘the equilibrium exhibits unexploited gains from trade regarding network participation’. Network externalities, so defined, do cause market failure, but are far less common than network effects generally.

Liebowitz and Margolis (1995a) accept the distinction between direct and indirect network effects, but argue that they are fundamentally different in their consequences for efficiency. They recognize that direct network effects in physical networks may (in limited circumstances) be true externalities. But they challenge the notion that indirect network externalities arise whenever complementary goods become more plentiful and cheaper as the number of users of the related product increases. They contend that much of what the literature calls indirect network externalities are merely positive pecuniary externalities that result in wealth transfers.

These theoretical points lead to very different policy conclusions. If the price of complementary goods decreases as a network grows because rents are transferred from input suppliers or producers to consumers, the market does not fail, and no state-sponsored remediation is necessary. If instead price falls due to positive technological externalities, remediation may be required, but only because of a conventional market failure in an upstream or downstream market, not network externalities.

Katz and Shapiro (1994), leading scholars in the field, have adopted the distinction between network effects and network externalities in theory, though they disagree in its application, and generally argue that true network externalities are more common than Liebowitz and Margolis suggest. Other scholars continue to use the term network externalities to encompass all network effects. (Economides, 1996; Klausner, 1995).
B. Effect of Externalities on Network Size

5. Physical Networks/Direct Externalities

The existence of positive network externalities in adoption of a technology in
a physical network implies that the private marginal benefit of acquiring a
network technology is lower than the social benefit. In deciding whether to buy
a network good, the individual compares the price only with his private benefit,
not the benefit that his purchase confers on other users. Consequently, the
equilibrium size of a physical network under perfect competition, with direct
network externalities, may be smaller than the social optimum (Katz and
Shapiro 1994; Economides, 1996).

The size of a physical network is also affected by consumer expectations.
Because information about other potential purchasers’ future actions is always
imperfect, consumers’ expectations may result in a nonoptimal network size.
For example, if all consumers expect no one else to purchase, then the network
size will be zero, even if all consumers would benefit by joining the network;
if all expect everyone to purchase, the size will be large (Katz and Shapiro,
1994). The importance of consumer expectations gives producers an incentive
to convince consumers through a variety of practices that their networks will
attract many users.

Some physical networks may be owned by a single firm. Firms that own a
network may use their property rights to internalize consumption externalities.
This result is most apparent in entirely internal networks. Fax machines, for
example, faced an initial externality problem in that purchasers had no
incentive to buy one unless others did also. If there were a high enough degree
of uncertainty, the market might not have developed at all. But large firms
apparently had enough incentives to purchase fax machines for purely internal
communications to create a critical mass (Katz and Shapiro, 1994, p. 97, n. 4).

A single firm’s ownership or sponsorship of a physical network with
individual subscribers can also internalize network externalities. Consumer
surplus is maximized if the sum of marginal private and social benefits of a
purchaser equals marginal cost (Liebowitz and Margolis 1994; Katz and
Shapiro, 1994). Liebowitz and Margolis argue that ownership of such a
network can usually solve the problem of a suboptimal network size that flows
from direct network externalities. The owner will set access prices that reflect
the fact that additional users provide benefits to all other users. If marginal
costs are assumed to increase, a network can reach an optimal size. Liebowitz
and Margolis believe that the assumption of increasing marginal cost, typical
to economic models, is appropriate here because economies of scale are
exhaustible for many network commodities. In these circumstances, more than
one network can coexist, so competition is possible. If, on the other hand,
economies of scale are inexhaustible, the market is a natural monopoly, which may be inefficient but not because of something unique to the network character of the good.

Suppose that ownership of the network does imply a monopoly. Economides and Himmelberg (1995) argue that the incentive of the monopolist to restrict output overwhelms its incentive in increase consumer demand by influencing expectations about network size. The resulting monopoly equilibrium is at a lower network size than under competition. Katz and Shapiro (1994) argue that, if a monopolist charges a single access fee above marginal cost, external benefits will not be internalized. But price discrimination ‘may well allow the network to internalize the adoption externalities: for example by setting access fees at or below cost and earning profits on usage fees’ (p. 101). The network owner thus addresses the problem of buyer expectations about the future size of the network.

Liebowitz and Margolis (1994) recognize that not all networks that generate direct network externalities can be owned. For example, the network of speakers of Esperanto is not ownable and so may fail due to direct network externalities. But other market mechanisms may minimize any social cost in such cases. Intermediary organizations may, for example, facilitate communication among groups within the population. Network effects can also be internalized by the direct interaction of participants. For example, computer programmers working on a common project can agree to adopt the same language. These interactions will tend to internalize potential externalities when transactions are easy. Liebowitz and Margolis believe, therefore, that network externalities will not often result in a suboptimal size of a network, and to the extent that a network is inefficiently small, the reason can be traced to the conventional market failure associated with natural monopoly.

6. Virtual Networks/Indirect Externalities

The analysis of the effect of indirect externalities on the size of virtual networks is somewhat different. In these markets, externalities may arise because of the durable nature of the hardware component in the system. The purchaser of hardware knows he will be ‘locked in’ to the product for a time, because the cost of switching to different hardware will be substantial. (On the effects of switching costs, see Klemperer, 1987a, 1987b, 1989). Consumers then may be uncertain about the future availability and price of software for the product. If this uncertainty can be addressed, then the network externality problem disappears. If, for example, all of the components of the system are competitively supplied by firms with U-shaped average costs, the market will reach an efficient competitive equilibrium. The market is indistinguishable from any competitive market with complementary products (Katz and Shapiro,
Note that this outcome differs from the inefficient outcome in physical network markets under competition discussed in the previous section.

Katz and Shapiro (1994) argue that network externalities can arise if hardware is supplied competitively at marginal cost, but differentiated software is provided at a price above marginal cost by firms subject to scale economies. Apparently responding to Liebowitz and Margolis, Katz and Shapiro (p. 100) state that ‘if all goods were priced at marginal cost, these network externalities would be merely be pecuniary externalities, and market equilibrium in hardware/software markets would be efficient’. If, however, software is not priced at marginal cost, a suboptimal variety of software may result or software may be produced at an inefficiently high cost, and the resulting networks may be smaller than optimal because the marginal social benefit of additional sales is greater than the private benefit to the purchaser. In such circumstances, Katz and Shapiro argue, a subsidy to hardware suppliers (or software suppliers) can increase consumer welfare by expanding the supply and variety of software. Liebowitz and Margolis presumably would respond that the subsidy in this case is (theoretically) justified because of imperfect competition in software, rather than because of any indirect network externality.

If a monopolist supplies both hardware and software, consumers may fear that, once they are locked in to the durable good, the monopolist will exploit them by increasing prices of the software. (Farrell and Shapiro, 1988; Katz and Shapiro, 1994; Shapiro and Teece, 1994). Sponsorship arrangements by the monopolist may provide a solution. Thus, network externalities can be avoided if the seller can somehow commit to software prices in advance of purchase. If such commitment is possible, consumers would have no reason to fear lock-in specifically. They would, of course, be subject to monopoly pricing (either single-price or discriminatory), but any inefficiency would then be attributable to the monopoly, not to network externalities. (Katz and Shapiro, 1985, 1994).

The monopolist may attempt in various ways to approximate a commitment to future supply of compatible goods at prices consumers expect to pay. It may, for example, attempt to guarantee a competitive supply of compatible software, even inviting competitive firms to enter the market (Economides, 1996; Farrell and Gallini, 1988; Katz and Shapiro, 1994). It may also adopt a policy of leasing hardware, thereby assuring that customers would not be locked in, were it to raise software prices. Or it might engage in ‘penetration pricing’ of hardware, which would mean that, although consumers may pay high software prices, they will be compensated by lower hardware prices. Finally, it may rely on its reputation as a bond securing purchasers against exploitation (Katz and Shapiro, 1994). If it were to exploit purchasers, it would damage its ability to sell later generations of its hardware.
C. Competition Among Networks

7. Tipping and Lock-In

One of the most striking consequences of network externalities is their effect on the nature of competition between sellers of products embodying different, incompatible standards. The literature cites many examples: VHS and Beta standards for VCRs; phonographs and compact disc players; conventional versus high-definition television; and so forth. In such markets, network externalities may favor the market’s adoption of a single seller’s product as the de facto standard, such as VHS rather than Beta as a videotaping format. In corresponding theoretical models, one technology gets an early advantage (for whatever reason), and positive feedback of the larger network size leads new users to adopt that technology. It is suggested that the first mover in markets for information-based software receive continuously increasing returns to scale, reinforcing early successes and aggravating early defeats (Arthur, 1989, 1990, 1994, 1996). Ultimately, competing technology leaves the market. A market that settles on a single standard is said to have ‘tipped’.

In other markets, network externalities and scale economies may be exhausted at a smaller network size, so that the market can accommodate more than one network. Moreover, consumers’ heterogeneous preferences can result in multiple standards, such as the IBM-PC and the Macintosh computer (Liebowitz and Margolis, 1995a; Katz and Shapiro, 1994, p. 106). Firms, therefore, might prefer to promote their own incompatible networks regardless of the resulting size. Or each firm might prefer its own standard, yet prefer compatibility with a rival’s standard if only one standard can prevail.

The tendency of network markets to tip leads to particularly intense competition early in the market’s existence (Farrell and Shapiro 1988). Competitors may employ aggressive strategies like ‘penetration pricing’ - in some instances giving the product away - in order to become the de facto standard. Katz and Shapiro (1994, p. 107) observe that in such cases firms are ‘bidding for future monopoly profits’. After achieving dominance, the firm may recoup some or all of the expenses it incurred in the early competition. In such cases, what appears to be monopoly profit earned by the firm in later stages of production is actually recovery of the earlier expenditures. Another possibility is that a firm with an early lead may have an incentive to deter new entry by low pricing in order to build on its advantage (Farrell and Saloner, 1986a).

One implication of the network externalities literature is that the market may settle on a good with a lower social valuation. The literature cites numerous examples: the QWERTY typewriter keyboard (David, 1985); VHS and Beta videocassette formats (Arthur, 1990); and AM stereo (Besen and Johnson, 1986). Proponents of network externalities point out that a good’s
value to the user depends upon its inherent benefit (the value of the good to the purchaser even if no one else adopted it) plus its network benefit. The network benefits increase with the size of the network. Initial adopters will primarily take account of the good's (apparent) inherent benefit to them. If that inherent benefit is greater for good A than for good B, the initial adopters will choose A. As the size of network increases, it becomes still more advantageous to choose A over B, and all subsequent adopters will do so. But suppose that, although B offers lower inherent benefits, its network benefits increase at a higher rate than those of good A, and at some network size actually exceed those of A. In theory, B would offer greater social benefits if it were adopted by the market as a whole; but the market nevertheless chooses A (Farrell and Saloner, 1986b; Klausner, 1995). Katz and Shapiro (1994, p. 106) state that 'standardizing on a single system can be very costly if the system turns out to be inferior to another system'. This type of standardization may also affect innovation incentives in the market.

Once the market tips toward a single standard, it may remain on that standard and its successors for a long time. Markets may exhibit 'excess inertia' and remain locked into a standard, even though an objectively 'better' standard is available (Katz and Shapiro, 1985; Farrell and Saloner, 1986a). Present users face substantial switching costs; even though all users would be better off with the new standard, those benefits do not accrue to the present users who must pay for switching. New purchasers also may opt for the established standard because of the immediate benefit that the established network offers; they do not take account of the benefit that purchasing the new technology would confer on later purchasers. Even if they anticipate the new technology would be widely adopted, the benefits of that adoption to the purchaser may be so far in the future that they are substantially discounted (Farrell and Saloner, 1986b, p. 947, n. 14).

Theoretical models, however, demonstrate no inevitable tendency of markets to lock-in on inferior products. Changes in the assumptions underlying the models (for example, concerning communication and information in the market) may eliminate the outcome of excess inertia (Farrell and Saloner, 1985, 1986a; Liebowitz and Margolis, 1990). Moreover, markets may actually exhibit 'insufficient friction' (Katz and Shapiro, 1986b, 1994) or 'excess momentu' (Farrell and Saloner, 1986b), tipping suddenly and inefficiently to new technologies. Insufficient friction may be inefficient because consumers do not take account of the network benefits their purchase of the established product would confer on existing users. Given this external benefit, it may be socially preferable to stay with the established technology. Nevertheless, consumers, fearing that they will be stranded with an obsolete, unsupported technology (the 'penguin effect', Farrell and Saloner 1987), may jump to the new technology. Katz and Shapiro (1986b, 1992) suggest that sponsorship of proprietary
technologies may lead to inefficient adoption of one technology over another through excess momentum. For example, if the incumbent technology is supplied competitively under a shared standard, a new, proprietary technology might be offered at ‘penetration’ prices in order to attract an installed base. The incumbent competitive suppliers would be unable to price below their marginal cost, because they can have no expectation of recoupment.

8. Criticism of the Tipping Scenario

Liebowitz and Margolis (1995a) dispute the suggestion that increasing returns to scale in high-technology markets lead to market failure. They point out that the frequently-observed relationship between increasing participation in a market with indirect network effects and falling prices is ambiguous. Such a relationship may result from scale economies, but it may also reflect technological progress. In the latter case, prices drop because advances in technology reduce industry costs, not because of a network effect.

Liebowitz and Margolis (1994) also criticize the suggestion that markets fail because the ‘wrong’ network is chosen, so that total net benefits are lower that they would have been under the losing technology. This kind of asserted failure is a function of models that assume inexhaustible scale economies, for only then does a single network survive. Liebowitz and Margolis argue that the marginal gains of network size are often exhausted at sizes above a critical mass that is small relative to the total market. If the benefits of size are exhausted, multiple networks can exist. In that event, not only are monopoly problems lessened, but the theoretical question shifts to whether the best set of networks emerge. And the inframarginal externality that may affect the discrete choice of a network ‘is not different from other coordination problems that exist in many other market choices’ (Liebowitz and Margolis, 1994, p. 141).

Liebowitz and Margolis are especially critical of claims that network externalities lead consumers to make discrete choices among networks, choices that do not maximize social welfare. Such externalities cannot be addressed by taxes and subsidies, because those policies can only affect the scale of a network. Rather, the assumed market failures occur because network externalities supposedly prevent value-increasing transitions from one technology to a superior one. At the extreme, the literature seems to suggest that economic actors may easily become forever stuck in technologies that are widely recognized as inferior to available alternatives. Liebowitz and Margolis (1994, p. 145) colorfully call this the ‘Chicken Little view of market transitions’. They contend that economics has generally not done well in explaining transition, which affects all components of the economy, not just networks. Nevertheless, value-increasing transitions do occur, even if the
process is not well understood. Evans and Schmalensee (1996) note that, for example, the computer software market ‘exemplifies Schumpeter’s (1942) view of modern competition - one in which firms and industries are constantly created and destroyed through the process of innovation’.

Whatever the theoretical possibility of market failure brought about by network externalities, critics complain that empirical support has not been adduced. Proponents of network externalities, for example, have often pointed to the conventional QWERTY keyboard as an illustration of lock-in. The keyboard arrangement was chosen, it is said, to slow down the typist to avoid mechanical jamming; other arrangements permit faster typing. Even though technological improvements have eliminated the mechanical problems, users are locked-in to the inferior QWERTY arrangement - consumers wish to avoid the costs of learning a different arrangement, and producers will not supply alternative products because not enough consumers would purchase them. Liebowitz and Margolis (1990) argue forcefully, however, that the story is factually incorrect, for the QWERTY keyboard is not demonstrably inferior to the leading alternative. Similarly, Liebowitz and Margolis (1994) find evidence that the Beta videotaping format is no better than VHS. In all, Liebowitz and Margolis believe that real indirect network externalities are of limited theoretical importance and have not been established empirically. For a similar discussion of the complexity of the issue of objective ‘superiority’ of one technology to another, see Van Vleck (1995).

D. Compatibility and Standardization Among Network Goods

9. Compatibility and Standardization Among Network Goods

Many of the private and social consequences of network externalities depend upon whether products competing in a network market are compatible in a relevant sense. The term compatibility is used in the literature to refer to different relationships depending upon the context. For example, physical networks are said to be compatible if they allow direct interconnection; virtual networks are said to be compatible if various components of the system are ‘interoperable’, that is, they can work together; and various products, such as computer software, may be said to be compatible if consumers can use them without significant retraining (Lemley and O’Brien, 1997). For useful surveys of the issues surrounding compatibility and standardization in markets with network externalities, see Cohen (1996) and Katz and Shapiro (1994).

Whether systems are made compatible depends on a host of factors that may influence decisions of firms on questions of design and contracting with other producers. Compatibility may arise as a result of a single firm’s product
becoming the de facto standard in the market; or it may arise through collective standard-setting mechanisms. In addition, converters or adapters may make products compatible to varying degrees (Farrell and Saloner, 1992). In some instances, it may be to a firm’s advantage to try to prevent compatibility with existing systems and compete with them to become the de facto standard for at least a part of the market. In other instances, firms may choose to make their products compatible with other products by licensing or by participation in coalitions or standard-setting organizations (Katz and Shapiro, 1994).

Compatibility, by expanding the size of networks, offers both private benefits and private costs to producers and consumers of network goods. Participants in compatible physical networks receive the direct external benefits of communicating with a larger number of consumers, and save the costs of owning two sets of hardware. Participants in compatible virtual networks receive the indirect benefits of a larger network, including a greater variety of components that may be mixed and matched to achieve an optimal system (Matutes and Regibeau, 1988), and a reduced risk of being stranded with obsolete technology. Producers receive the benefits associated with larger scale. Compatibility may impose costs, however, depending upon how it is achieved. If it is achieved by standardization, then there may be a reduction in the variety of systems and products available to first-time consumers (Gilbert, 1992). If it is achieved by adapters, the adapters themselves impose costs (Katz and Shapiro, 1994).

Whether firms have an incentive to create compatibility depends upon how the resulting changes in the terms of competition affect them. Compatibility reduces the risk that the market will tip to a rival’s sponsored standard, and so may reduce the intensity of competition, at least in early stages when incompatible systems would be competing to become the market’s de facto standard (Katz and Shapiro, 1986b). This tendency suggests that some firms (or coalitions of firms) may have an incentive to agree to adopt a compatible standard (Besen and Farrell 1994; Katz and Shapiro, 1994). But if systems are compatible, then the market may accommodate more firms, and so compatibility may actually increase competition over the long term (Katz and Shapiro, 1994). Firms that standardize reduce their chances of becoming the dominant firm.

When networks are incompatible, then competition is between physical networks or between systems of goods that comprise virtual networks. When virtual networks are compatible, however, competition is among the components of the system (Matutes and Regibeau, 1988; Economides 1988, 1996; Katz and Shapiro, 1994). In such situations, the firm’s preference for compatibility depends upon the tradeoff between the increased demand that compatibility creates and the increased competition that it entails. Firms that believe they offer a better overall system (like Apple Computer in its earlier years, perhaps) may opt for incompatibility; firms that believe they offer a better individual component may opt for compatibility. In general, the ‘winner
take most’ nature of competition between incompatible systems may lead firms with very promising (or highly regarded) technology to oppose compatibility, even when compatibility is socially preferable (Katz and Shapiro, 1994). The social optimum is not achieved because firms individually do not take account of the effects of compatibility on others.

E. Policy Implications

10. Policy Implications: General Remark

The many inefficiencies identified in the network externalities literature have led some to propose changes in government policies to accommodate the theory. But most economists have counselled caution in using the network externalities literature to justify greater government intervention. Katz and Shapiro (1994), for example, note that private market responses may correct any inefficiency. Furthermore, they point out that interest group pressures make government responses less likely to favor efficiency than powerful incumbent producers. And there is little reason to think the government has the ability to identify the correct market outcome better than market mechanisms (Arrow, 1995; Vita and Wellford, 1994; Lopatka and Page, 1995a, 1995b). Liebowitz and Margolis (1994) are even more insistent that the network externalities literature is far too limited in its theoretical implications and its empirical support to justify a more interventionist government policy.

Nevertheless, legal literature and to a lesser extent the case law have begun to take account of network externalities. A search of the WESTLAW ‘Journals and Law Reviews’ database reveals over references to ‘network externalities’ in diverse areas of law. And the concept has begun to be accepted by some policymakers and courts. The following sections will address some of these issues.

11. Antitrust: Monopolization

Because it predicts that a single firm may dominate a market, network externalities theory has been recruited to support claims of monopolization. It implies that a firm may have an incentive to make its network incompatible, to achieve higher sales volumes than its competitors during early stages of production, and to convince potential users that its standard eventually will prevail. The practices that a firm uses to influence the direction in which the market tips are likely to be attacked as exclusionary. For example, in *Eastman Kodak Co. v. Image Technical Services, Inc.* (1992), the Supreme Court held
that an equipment manufacturer (without monopoly power in the equipment market) could have monopolized by requiring consumers to purchase its repair services as a condition of obtaining its replacement parts. The Court did not mention network externalities explicitly, but did treat the equipment, services, and parts as a sort of system. Recall that Katz and Shapiro (1994) specifically refer to ‘durable equipment and repair services’ as an example of the hardware/software paradigm, or an indirect network. The Court’s reasoning closely resembles a network externalities analysis. It stated that purchasers were locked-in to Kodak equipment and would find it costly to predict the prices of aftermarket services and products at the time of equipment purchase. High switching costs and imperfect information are important features of network externalities theory.

The United States, in its amicus brief in *Kodak*, appeared to reject network externalities theory by arguing that Kodak could not extract monopoly profits in aftermarket because it lacked monopoly power in the equipment market. But the Antitrust Division under the Clinton administration explicitly invoked network externalities theory in suing Microsoft for monopolizing the market in personal computer operating systems, although it limited the scope of its complaint to restraints in Microsoft’s distribution contracts (*United States v. Microsoft*, 1995). In the litigation challenging the consent decree in that case, all sides (and the District Court) agreed that network externalities in the hardware and software markets had important antitrust implications and might justify stricter application of monopolization standards. A group of anonymous amici curiae filed a brief that was co-authored by Brian Arthur and Garth Saloner, both prominent theorists of network externalities. For a full analysis of the brief and the district court’s opinion, see Lopatka and Page (1995a).

Network externalities theory has limitations as a policy guide in monopolization cases. (Lopatka and Page, 1995a, 1995b; Evans and Schmalensee, 1996; Gifford, 1996). It does not, for example, support breakup of a firm that achieves dominance. The theory predicts that networks will often grow large for efficiency reasons. Tipping does not necessarily imply that a producer is forever locked in, because new technology (and creative new producers) can benefit from the same phenomenon to leapfrog the old dominant firm. The very possibility of tipping is an important incentive to innovation and novel marketing strategies. And history contains enough examples of technological change sweeping away previously dominant firms to caution against government action to break up a market leader.

Lock-in may not even signal market failure. A firm may dominate because it in fact offers the best product. An antitrust court could not confidently declare that the wrong technology had triumphed (Vita and Wellford, 1994). While excess inertia is possible, theory does not tell us when it is present. Thus there is no justification for encouraging displacement of a dominant firm by weakening its locked-in position. Nor is there justification for detailed
supervision of the dominant firm’s practices. If the market is a natural monopoly, it is not well suited to antitrust regulation, regardless of network externalities.

For similar reasons, network externalities theory does not offer much assistance in identifying ‘exclusionary’ conduct that creates monopoly power. Some models suggest that markets can be tipped by trivial events or an early advantage, so that the development of the market is path dependent (Arthur, 1989; David, 1985). Microsoft’s MS-DOS operating system may have achieved dominance because IBM selected it long ago for the IBM PC (Besen and Farrell, 1994). If historic accident determines market outcomes, however, none of Microsoft’s subsequent competitive acts can be blamed for its dominant position. Some practices may be theoretically inefficient, but courts lack the knowledge to identify when they are so. Firms may, for example, use advertising to manipulate consumers’ expectations about which standard will prevail (Besen and Farrell, 1994; Katz and Shapiro, 1994). Or they may use product preannouncements to discourage existing customers from switching to another supplier and to encourage prospective purchasers to wait (Farrell and Saloner, 1986b). But there is no effective way to distinguish these practices from efficient dissemination of information (Ordover and Willig, 1981).

Network externalities theory may actually provide efficiency explanations for apparently exclusionary practices. For example, vertical integration or exclusive contracts may allow a producer to establish a credible commitment to provide a supply of complementary goods. Similarly, a lease-only policy (sometimes suspect in antitrust law) may be an effort to assure consumers that they will not be locked in to a dying network.

Predatory pricing issues may also be influenced by network externalities. Katz and Shapiro (1994, p. 104), describe ‘penetration pricing’: ‘by selling hardware below cost early on, the network sponsor is stimulating the demand for software, which may lead to a lower price of software if software is produced according to economies of scale or if the elasticity of demand for software is higher for marginal consumers than for the average hardware consumer’ (see also Besen and Farrell, 1994). This pricing strategy may resemble predatory pricing; but its goal is to overcome network externalities in building an installed base. Below-cost pricing or even giveaways of some products, like computer software, may be a rational means of establishing an installed base of complementary product sold by the same firm. Farrell (1989) suggests that this sort of competitive below-cost pricing by sponsors of proprietary technologies will typically lead to the better technology being adopted. Thus, Lemley (1996) concludes that such a price war raises concerns only if it is asymmetric, that is, if one of the combatants has greater staying power.

Although network externality theory is susceptible to misuse in the analysis of monopolization, it may sometimes provide a helpful context for examining allegedly exclusionary conduct. For example, Lemley (1996) suggests that
concealment of a proprietary claim to a market standard during a period in which that standard is being promoted for acceptance in the market or by standard-setting organizations may be monopolistic.

12. Antitrust: Horizontal and Vertical Agreements

Antitrust’s traditional hostility to horizontal agreements may cast suspicion on some agreements in network industries. But antitrust has also recognized that some industries require cooperation in order to exist. (See, for example, National Collegiate Athletic Ass’n v. Board of Regents, 1984). Network industries often fall in this category (Carlton and Klamer, 1983; Evans and Schmalensee, 1996). Perfect competition may be inefficient (in physical networks) or entirely impractical (in virtual networks). Network externalities can often be internalized only by contract or joint ownership. Bank credit cards require cooperation among banks in order to compete with cards issued by a single firm (Carlton and Frankel, 1995). Telecommunications networks require agreements on interconnection and sharing of joint costs and revenues. Producers of both hardware and software must settle on standards to assure compatibility (Brown, 1993). Lemley (1996) mentions the Internet Engineering Task Force as an example of a necessary standard-setting organization. While these points suggest procompetitive explanations for some types of agreements, they do not exclude the possibility of such an organization being used as a cartel or an exclusionary device. Some network joint ventures may thus require continuing antitrust controls. Real estate multilisting services offer network benefits, but do not justify price fixing by participating realtors. And network externalities do not imply that standard-setting joint ventures should be permitted to exclude competitors (Anton and Yao, 1995).

Mergers and joint ventures have become common in network industries. Network externality theory offers a number of efficiency justifications for these arrangements. Physical networks may offer greater value to subscribers by merging, and may benefit from scale economies and greater information. Research and development joint ventures may permit firms to pool information and resources to create more advanced technology. These considerations do not, however, imply that antitrust scrutiny is inappropriate. A merger creating a network that is larger than is justified by scale economies may create an unnecessary danger of monopoly. In Money Station, Inc. v. Board of Governors of Federal Reserve System, the Board argued to the court that an acquisition of a small ATM network by a bank controlling the dominant ATM network was less likely to lead to anticompetitive effects because (p. 1133):
Network externalities, such as the economies of ubiquity, tend to promote the consolidation of regional ATM networks. As a result, in various geographic areas, like the Mideast region, dominant ATM networks have been emerging throughout the EFT industry. One recent study indicates that the ten largest regional networks now account for 80 percent of all regional ATM transactions in the United States. In this light, the Board believes that, as a result of economic and market structure conditions, regions are likely to have one dominant ATM network.

The court properly questioned this reasoning as a way of avoiding competitive concerns about the merger. It is not clear that network externalities require dominance of a single firm.

13. Intellectual Protection for Proprietary Standards

The theory of network externalities suggests that a single product may emerge as the de facto standard in the market. Should the presence of network externalities, and the consequent danger of lock-in, affect the intellectual property protection given to a proprietary de facto standard? Menell (1987, 1989), relying in part on an analogy to the QWERTY story, has argued that network externalities justify limiting copyright protection to computer software that has become the industry standard. In *Lotus Development Corp. v. Borland International* (1995), the court adopted this reasoning, holding that Lotus 1-2-3’s user interface was a ‘method of operation’, like the buttons on a VCR, and therefore not protectable expression under copyright law. One of the reasons the court offered was the need for compatibility. The court found it ‘absurd’ to suggest that ‘if a user uses several different programs, he or she must learn how to perform the same operation in a different way for each program used’. A concurring judge added that ‘if Lotus is granted a monopoly on this pattern, users who have learned the command structure of Lotus 1-2-3 or devised their own macros are locked into Lotus ...’ Consequently, a competitor (Borland) should be allowed to copy Lotus’s menu command structure.

This decision adopts a network externality theory raised by Borland and amici curiae in their briefs on appeal. It has been questioned by Dam (1995), who points out that ‘compatibility’ in this case is not the interoperability that allows use of competitive software or the interconnection that allows communication across a physical network. Rather, compatibility in this case means only that a user may costlessly adopt a competing product. The argument is that the menu structure has become the de facto industry standard, and so free copying should be allowed to assure adequate competition. But it is questionable whether network externalities are at issue here. The users of Lotus
Network Externalities

1-2-3 do not constitute a physical network, so there are no direct external benefits from expansion of the installed base. Furthermore, the case does not involve complementary products, so there are no apparent indirect network externalities from greater availability of compatible goods. There are at best so-called convenience externalities. There are switching costs between spreadsheets, but not evidently more than between products generally.

One difficulty with the use of network externalities to limit copyright protection is that it is difficult to limit the principle to computer software. It has been suggested that many popular copyrighted works, such as *Gone With the Wind*, create a network of users that may be exploited in spinoffs and sequels. Thus, the case of computer software does not appear to be unique among copyrighted products (Lunney, 1996). More generally, the application of the network externalities concept to negate copyright protection for popular software is similar to the antitrust argument that the successful firm has monopolized, and is questionable for the same reasons. Weakening copyright protection allows greater competition by clones, but reduces the payoff for innovators, by threatening loss of protection for successful firms.

14. Contract Terms and Other Norms

Some scholars have argued that social norms have network aspects. Conventional rules like those governing right of way for drivers at an intersection have greater value the more people adopt them. This extension of the concept of network externalities has been most fully developed by Klausner and Kahan (Klausner, 1995; Kahan and Klausner, 1996), who have argued that the adoption of corporate contract terms can create indirect network externalities. Common use and judicial interpretation of contract terms benefit the ‘network’ of firms adopting those terms by, for example, clarifying the meaning of the set of legal rules governing the firm. Larger network size generates positive feedback because it increases the stock of judicial precedent and the common understanding of the relevant rules. Network externalities thus modify the conception of the firm as a nexus of contracts. They imply that contract terms that maximize the value of the individual firm may not maximize social wealth. State corporation laws, by providing default rules for contractual provisions, are analogous to industry standards in physical networks like telecommunications, because they can facilitate the creation of networks. Their goal should be to promote the creation of an optimal mix of uniformity and diversity by providing ‘open-ended’ default rules menus of alternative provisions. Network externalities may also account for the dominance of Delaware corporation law. Some scholars have suggested that states compete in a race to the top for corporate charters by offering statutes that maximize firm value. But Delaware’s dominance (despite the prevalence of similar laws)
may reflect the positive feedback effect of its large network of incorporated firms. This result implies that Delaware law has become popular not because it offers optimal terms, but because it has become locked in.

Bibliography on Network Externalities (0760)


Bolton, Walter et al. (1990), Telecommunications Policy for the 1990s and Beyond, New York, M.E. Sharpe, Inc.


Cabral, Luis (1990), ‘On the Adoption of Innovations with Network Externalities’, 19 Mathematical Social Sciences, 229-308.


Economides, Nicholas and Salop, Steven (1992), ‘Competition and Integration among Complements, and Network Market Structure’, 40 Journal of Industrial Economics, 105-123.


Other References


Cases