A Theory for the Instability of Public Telecommunications Systems

Eli M. Noam
Columbia Institute for Tele-Information, Columbia University, Uris, 116th & Broadway, New York, N.Y. 10027, USA

1. The Concept of Network

Networks are an important concept in society and economy. They abound as *physical* facilities, such as those of electric utilities, communications, and transportation. They also exist as *relational* systems, such as those of "old boys", political supporters, and intelligence agents.

The term "network" goes a long way back; it is used, in the King James translation, by the Supreme Regulator: "And the Lord spake unto Moses, saying..... You shall also make it a grating, a network of brass..." *Exodus* XXVII, V. 4. In Hebrew, the word is "reshet", (net) similarly used today for telecommunications and other networks.

The term is used by most academic disciplines, and with a variety of meanings. Chemists apply it to arrangements of molecules (Zacharisen, 1932). Biologists to cell structures (Knox, 1830). Mathematicians to topology (Klingman, 1981). Electrical engineers to distribution systems (for high voltage), or for circuit configurations of components (for weak voltage) (Karni, 1986).

Operations researchers use a network terminology to solve shortest path problems, maximum flow models, and optimal routing (Elmaghraby, 1970). Computer scientists apply the term for computer interconnections in hardware, and to implementation algorithms in software.

In the social sciences, political scientists use the concept of networks in discussing hierarchies, interactions, gatekeepers, and policy communities (Richardson et al., 1985). For sociologists and social anthropologists, networks are a major way to see the world; a basic point is that the nature of linkage affects behaviour (Barnes, 1954; Bott, 1957). Sociologists speak of network *dyads* — interpersonal linkage between two persons in which each is indebted to the other, and similar in some ways to the exchange the relation of economics.

Among the social science disciplines, economists have probably paid the least attention to networks. There is no body of analysis for the network concept.

Somewhat related is work on market structure by some industrial organization theorists (Baumol et al., 1982). Closer are public choice theories of group formation, discussed in the following section. Other writings on networks are Noam (1988), Heal (1989) and Economides (1989).

It would go beyond the scope of this essay to discuss the factors underlying all network arrangements. But focusing on telecommunications, we can look at its key institution, the shared network, and analyze its dynamics.

2. Theories for the Emergence of Multiple Networks

A number of explanations have been offered — explicitly or implicitly — for the demise of monopoly in telecommunications. There are few major types of theories.

2.1. Technological Explanations

This perspective comes in two variants:

(A) "More powerful technology leads to new transmission options and thereby to competition and the breakdown of monopoly".

(B) "The merging of telecommunications and computing technologies breaks down traditional barriers separating different industries and undermines monopoly power".

These views are typically held by technologists, and they are influential in an engineering-oriented industry such as telecommunications. But they are not sufficient as explanations, or else one would observe a diversity of physical networks also emerging in, say, France, Australia, or Mexico. After all, the same transmission and switching technologies are available anywhere on the globe.[1] Yet their impact on network structure has varied, and provide no evidence for a technological determinism at work. Technology change provides the precondition for change, but it is not a sufficient condition.

2.2. Political Explanations

Here there are three related versions, all using the perspective of countervailing powers:

(C) "In the information age, a telecommunications monopoly becomes too powerful and its scope needs to be limited."

(D) "Government regulation proves incapable of controlling a monopoly, and is therefore replaced by policies encouraging a competitive industry structure."

(E) "Large business users successfully fight restrictions based on natural monopoly."

The problem with these views is that the introduction of a multiplicity of carriers is only one policy option out of several. An alternative response to political power or regulatory inefficiency might well be a stricter or more effective regulation, as would be nationalization or a size-reduction along geographical and/or functional lines while maintaining monopoly. Thus, it is not clear why competition is the necessary remedy to monopoly power. As to "natural monopoly", it is often and incorrectly believed that they led to national monopoly. Maybe a natural monopoly exists for a local exchange area, but the examples of the United States, Canada, Denmark, Finland, and several other countries show that this does not prove that a nationwide horizontal integration of local exchange areas is required. And if it were, why do these economies of scale end exactly at the national frontiers? If we look at the birth of the monopoly system in the sixteenth and seventeenth centuries and the establishment of European postal monopolies, we see that the monopoly was quite unnaturally caused by politics of the revenue needs of the state, rather than by second order conditions of production functions.

2.3. Non-Sustainability Explanation

Another view is that a monopoly, even if efficient across its multiple products, cannot protect itself from entry into some lines of business.

(F) "The diversification of telecommunications makes it difficult for any one provider to serve all sub-markets without competitive entry."

This view is essentially that of an economic non-sustainability theory advanced by Baumol et al. (1982). It can explain the emergence of entrants for new products of a multi-product firm, but it does not adequately cover competition in traditional core markets of a telecommunications monopolist, unless one accepts very restrictive assumptions (Shepard, 1983).

2.4. Market Structure Explanations

There are two variants to this approach, one passive and the other active:

(G) "Monopoly's inefficiency eventually leads to the emergence of competition."

(H) "Competition is a policy chosen to enhance efficiency and technological development."

These views are held by many economists. Theory (G) is a Milton Friedman type view of the impermanence of privileged economic arrangements. Theory (H) expects governments not to simply wait for competitors' entry but to institute proactive and pro-competitive policies. Examples are the United Kingdom and Japan, where competition was introduced from above. These two views have in common the premise of inefficiency of monopoly. In other words, a multi-carrier market structure is believed to be emerging due to some failure of the traditional system. Yet this assumption is at tension with the reality of network performance in those countries where structural changes in networks is most rapid. If inefficiency were the causal force for rival entry, Egypt or Mexico (to use two examples) should have introduced competition long before the US and Japan, which had arguably the most advanced and ubiquitous networks in the world even *before* embarking on their liberalizing policies.

It has always exasperated the proponents of the traditional network system to be told that their problem was inefficiency. This clashed with their observations of economies of scale, benefits of long-term technological planning, and effectiveness of end-to-end responsibility. Thus, explanations based on the inadequacies of the monopoly system are not persuasive.

2.5. A New Approach: The Tipping of Network Coalitions

None of these eight theories for the emergence of multiple networks provides an adequate explanation, though they all contain some truth, and their aggregate holds some explanatory power.

In contrast, this essay advances a ninth and alternative view, that of the dynamics of group formation. The thesis of this essay is not based on the *failure* of the existing system on its success. Changes in technology politics, and cost are merely enabling a more fundamental shift of coalitions.

> (I) "The breakdown of monopoly is due to the very *success* of the traditional system in advancing telephone service and in making it universal and essential. As the system expands, political group dynamics take place, which lead to redistribution and overexpansion. This provides increasing incentives to exit from a sharing coalition, and to an eventual 'tipping' of the network from a stable single coalition to a system of separate sub-coalitions."

This view of success undermining its own foundations is basically Schumpeterian. From the monopoly's perspective, it is deeply pessimistic, because it implies that the harder their efforts and the greater their success, the closer the end to their special status is at hand. Like in a Greek tragedy, their preventive actions only assure their doom.

3. A Model of Networks[2]

Perhaps the best way to look at a network is as a *cost sharing arrangement* between several users. Fixed costs are high, marginal costs low, and a new participant C helps the incumbents A and B to lower their cost. In that it is similar to a swimming pool or national defense, that is, to a "public good". But while there is basically only one national defense system, there are many types of arrangements for swimming pools. A user may want to share the pool with a few dozen families, but not with thousands. A pure public good admits everyone, a pure private good, only one. But there is a wide spectrum between the pure private good and the pure public good (Buchanan, 1965). A telecommunications network is one intermediate example. It is not a private good, yet it does not meet the two main conditions for a public good: non-rival consumptions and non-excludability. In fact, non-excludability has to be established as a legal requirement — the universal service obligation. What has been happening in recent years to telecommunications, and what goes by the more dramatic labels of divestiture and deregulation, is largely a shift in the degree of its intermediate position, a shift toward the direction of private good.

We will now develop, in a stepwise fashion, a model for network evolution and diversification.

3.1. The Basic Model[3]

Let the total cost of a network serving n subscribers be given by

$$TC(n) = F + f(n) \tag{1}$$

where fixed cost $F \geq 0$ and marginal cost $f'(n) \geq 0$. n assumes that users are homogenous. Of course, some network participants are much larger than others, but that poses no problem if we define a large organization to consist of multiple members of type n, e.g., telephone lines or terminals rather than subscriptions. Later, we will drop this assumption. Let an individual's utility be given by u (−P, n), where P is the price for network usage, and n is the number of network

[2]It should be noted that this model can be easily adopted for "standards coalitions" rather than "network coalitions". For the literature on standards, see David, Paul A., "Some new standards for the economists of standardization in the information age", in Dasgupta and Stoneman (1987b).

[3]I will follow the network analysis as developed in Noam, Eli, "The Next Stage in Telecommunications Evolution: The Pluralistic Network", paper presented at the Pacific Telecommunications Conference, Japan, October 1988, and begin in sections 1-4 with some of the methodology of my Columbia colleague

members.[4] We assume network externalities to exist, ($\partial u/\partial n > 0$), though at a declining rate ($\partial^2 u/\partial n^2 < 0$), i.e., a subscriber is better off the more other members there are on the network, *ceteris paribus* (including network performance and price).[5] For simplicity, utility is expressed in monetary units.

$$u = u(-P) + u(n) = -P(n) + u(n). \tag{2}$$

We assume that the network membership is priced at average cost, i.e. that users share costs equally. (This assumption will be dropped later.) This can be shown schematically in Figure 1, where u(n) is steadily increasing, though at a declining rate, and $P = AC = [F + f(n)]/n$ is declining, at least at first. The network, at this stage, is in its cost-sharing phase.

3.2. Critical Mass

Subscribers will find it attractive to join a well-sized network, because total costs are shared by many, making average costs low, while the number of subscribers n adds to utility. This can be seen in Figure 1, where the utility of joining a network rises at first. Conversely, where the network is small, average cost is high, and externalities small. In that range, below a "critical mass" point n_1, a network will not be feasible, unless supported by external sources. We define critical mass to be the smallest number of users such that a user is as well off as a non-user $u(n)=P(n)$.[6]

To reach n_1 requires a subsidy of sorts, either by government or by the network operator's willingness to accept losses in the early growth phases of operations. The strategic problem is to identify in advance a situation in which such a break-even point n_1 will be reached within the range $n < N$, where N=total population. Possibly, such a point does not exist, and subsidies would have to be permanent in order to keep the network from imploding. We will return to the critical mass issue later in subsection 12.

[4]Strictly speaking, income is allocated to telephone and to other consumption:

$$y = cp + [F + f(n)]/n.$$

A subscriber's utility is then given by

$$u(y/p-[F + f(n)]/np;\ n).$$

[5]For convexity, assume $u(c,P,1) > u(c,P,0)$, i.e. the first user has positive benefits even if no one else is on the network.

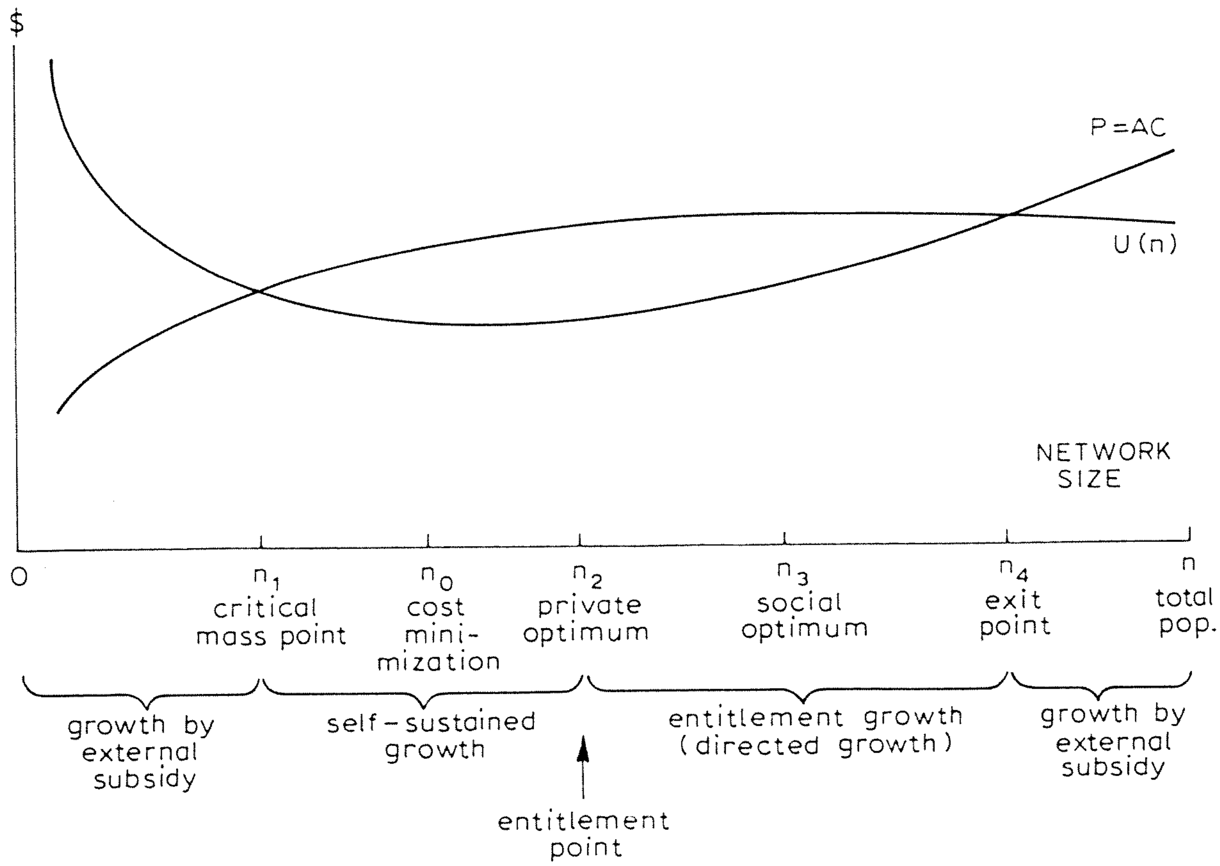

Figure 1. Stages in network expansion

3.3. Private Optimum

Through the cost-sharing phases of network growth, the earlier network users can lower their cost by adding members. However, at some point average cost AC increases.[7]

[7]Over the years, lower-cost subscribers have tended to be added to the network earlier than high-cost subscribers. As the network reaches universality, connecting the last members increases cost. An indicator for rising costs: In the Bell System, the average capital investment cost per new telephone grew steadily (in 1982/3 dollars).

1945:	$1928
1955:	$2050
1965:	$2580
1975:	$3960
1985:	$4624

Sources Telecom Factbook, 1986, FCC, Statistics of Communications Common Carriers, 1945, 55, 65, 75, 85.

$$\frac{dAC}{dn} = \frac{1}{n}[AC - f'(n)]. \tag{3}$$

That is, average cost increases in the range beyond the point n_0 where $AC=f'(n)$.

Beyond n_0 expansion becomes unattractive for cost reasons. But, some further expansion would be accepted by the network members since newcomers beyond the low cost point would still add to utility. This will be up to the point n_2 where the total derivative with respect to the number of users is equal to zero.

$$\frac{du}{dn} = -\frac{\partial u}{\partial c}\frac{\partial AC}{\partial n}\frac{1}{p} + \frac{\partial u}{\partial n} = 0 \tag{4}$$

which holds where the marginal utility of an added subscriber is

$$\frac{\partial u}{\partial n} = \frac{\partial u}{\partial c}\cdot\frac{1}{p}\cdot\frac{\partial AC}{\partial n}. \tag{5}$$

This is the case in the range of increasing AC. Graphically, n_2 would lie where the two derivatives are of equal size, $u'(P)=u'F(n)$. Left to themselves, the existing subscribers of the network would not accept members beyond n_2, the private optimum.

3.4. Social Optimum

From a societal point of view, however, the optimal network size in an equal price system may diverge from the private optimum.

Assume social welfare given by the sums of utilities.

$$W = n[u(-P(n)) + u(n)] = n[(F + f(n))/n + u(n)] \tag{6}$$

so that

$$\frac{dW}{dn} = f'(n) + n\ u'(n) + u(n) = 0. \tag{7}$$

Social welfare increases at n_2, and becomes zero at a point of intersection n_4. Hence, social optimum n_3 is somewhere in between those two points.

What is the implication? Left to itself, the network association will cease growth beyond n_2, at least as long as costs are equally shared. Existing network subscribers would not want to admit newcomers beyond n_2. Latecomers beyond that point add cost, because they raise AC, and add fewer externality benefits. In the words of Alfred Kahn: "People who don't have a phone I don't want to talk to". The socially optimal size n_3 will not be reached by itself, but by some external governmental direction through required expansion, and/or a differentiated pricing scheme, or through some internal politics of expansion that will be described below.

3.5. Entitlement Point and Universal Service Obligation

To understand the politics of government-directed network expansion, let us assume a political decision mechanism in which majorities rule. When private optimum size $n_2 < N/2$, there are more people outside than inside the network, while there are positive net benefits, i.e., $u(n_2) - AC(n_2) > 0$. A majority consisting of $N - n_2$ network outsiders would therefore outvote the n_2 network insiders, and require the opening of the network to additional members. This would be the case up to the point where network size reaches N/2, at which point the network insiders have grown to a majority and will resist further growth. Beyond N/2 then (or where $n_2 \geq N/2$ and a majority against expansion exists from the beginning) a politically directed growth will occur if the coalition of network insiders can be split by aligning the remaining outsiders N/2 with some of the insiders who are offered a more favourable share of cost, i.e., by price-discrimination. It can be shown that this coalition formation will lead to an over-expansion of the network. The dynamics of such price discrimination and its impact are discussed in greater detail in subsections 3.8 and 3.9 that deal with expansion and network tipping.

Politically directed growth beyond private optimum n_2 can be termed an "entitlement growth" because it is based on political arguments of *rights* to participate in the network where average net benefits are positive (encouraging attempts of entry) while marginal net average benefits are negative, leading to attempts at exclusion. When the marginal net benefits are positive, there is no need to resort to the language of entitlements, since growth is self-sustaining and sought by network insiders. It is only beyond that point that entitlements, rights, and universal service rights (i.e. obligations by the network) become an issue. We can thus define n_2 as the "entitlement point".

This way of analyzing entitlements serves to clarify the often-asked question: for which services will universal service be extended? Using the analysis, the answer is to those services that

- (a) have grown beyond minimum critical mass and
- (b) have reached, through self-sustained growth, a private optimum, beyond which further growth is not internally generated because *marginal* average net benefits are zero, but where
- (c) average net benefits are positive (and therefore encourage demand for entry), and
- (d) the number of those excluded is sufficiently large to lead to an opening by political means.

3.6. Exit from the Network

If $n_2 < N$, with N being the total population, some people would be left out of the network. But as discussed in the previous section, a government would require for the network to be open to other users. Yet there may well be a point where the network is expanded to an extent that, given its increasing marginal cost, a user is better off by not participating. We define n_4 as the "exit point", i.e., the largest n such that the indifference exists between dropping off the network and sharing in the cost of supporting the expanded network.

$$u(n) = u(P). \tag{8}$$

It is possible that this exit point lies beyond the total population, $n_4 > N$. But this seems not likely under an average-pricing scheme, because the last subscribers may impose a heavy burden on the rest of subscribers, and the subsequent departure of some subscribers would lead to further reduction in the utility of the remaining members and may induce a secondary exodus. Thus, assuming $n_4 < N$, a government's aim to establish a truly universal service, without resorting to a subsidy mechanism or price discrimination, is likely to be infeasible. In other words, a universal service policy is likely to be dependent on a redistributive policy.

3.7. Political Price Setting and Redistribution

We have so far assumed that universal service is something imposed externally by government. In this section, however, it will be shown that the *internal* dynamics of network members will take the network towards universal service, and towards its own disintegration.

As has been shown, a network will cease to grow on its own after private optimum n_2. But this conclusion was based on a pricing scheme of equal cost shares. Yet there is no reason why such equality of cost shares would persist if they are allocated through a decision mechanism that permits the majority of network users to impose higher cost shares on the minority. (This assumes that no arbitrage is possible.) Unequal prices and a departure from cost could be rationalized benignly as merely "value of service" pricing, i.e. higher prices for the users who value telephone greatly.

Suppose for purposes of the model that decisions are made through voting by all network members.[8] Let us assume at this stage that all users are of equal size (or that voting takes place according to the number of lines a subscriber uses, which is the same thing) and that early network users have lower demand elasticity for

[8]This analysis should not suggest that a voting mechanism is governing in reality (although it exists for telephone cooperatives in Finland and the US) but rather to understand the pressures and dynamics that

network use. The determinative vote is provided by the median voter located at n/2. A majority would not wish to have its benefits diluted by a number of beneficiaries larger than necessary. This is the principle of the "minimal winning coalition". Its size would be n/2 + 1.

A majority will establish itself such that it will benefit maximally from the minority. The minority that can be maximally burdened are the users with less elastic demand for telephone service, which are the early subscribers. But there is a limit to the burden, given by utility curve u(n). If price gets pushed above u(n), subscribers will drop off. Hence, the majority ($n_2/2 + 1$) will burden the minority ($n_2/2 - 1$) with a price up to positive utility, and they will bear the rest of the cost. The minority's price P_1 will be such that

$$P_1 = u(n_2) \tag{9}$$

The majority's price will then be[9]

$$P_2 = (F + f(n_2) - n_2/2 \cdot P_1)/(n_2/2) = 2AC - P_1. \tag{10}$$

Figure 2 shows this relation in a different way. Suppose u(n) gives a subscribers's utility for various network sizes n, expressed in monetary units, and P=AC(n) is the price vector. Private optimum is at n_2. The minority $n_2/2$ can be charged a price P_1, where P is equal to utility $u(n_2)$. Area A constitutes the subsidy; B is total cost; remaining majority is C=B−A, and the price charged to majority subscribers is

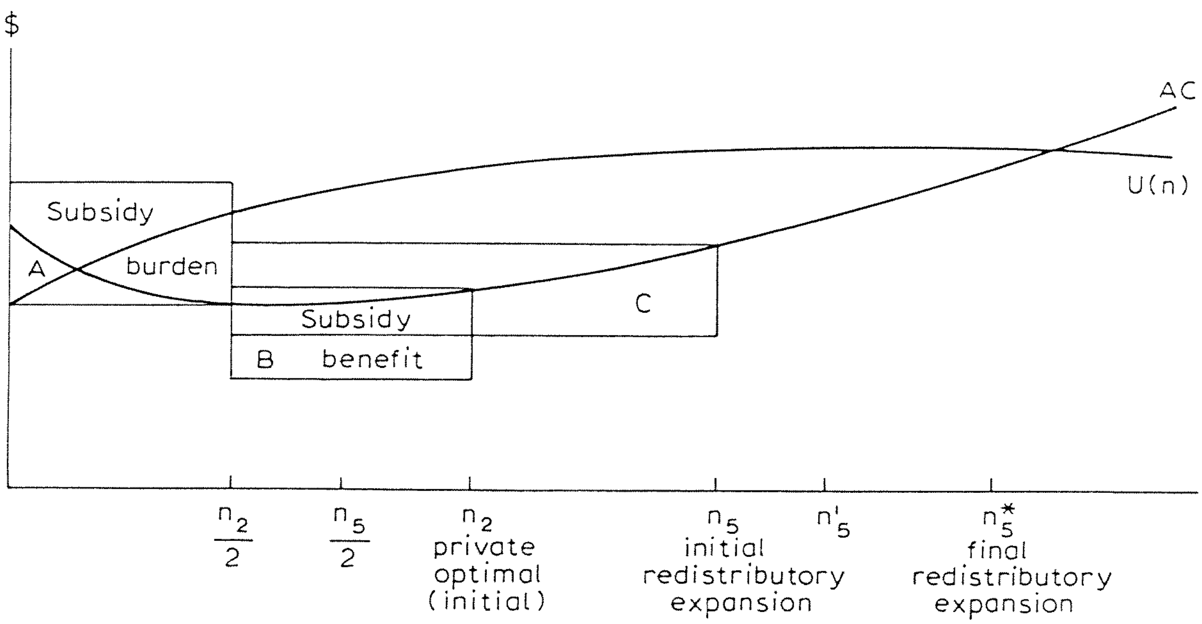

Figure 2. Redistribution in networks

[9]For simplicity, we use in the following n/2 rather than (n/2) − 1.

$$C/(n_2/2), \text{ or } P_2 = [F + f(n_2) - (n_2/2)u(n_2)]/(n_2/2), \tag{11}$$

$$P_2 = 2AC - u(n_2) = 2AC - P_1. \tag{12}$$

This then is the redistributory outcome, assuming no discrimination within majority and minority.

3.8. Monopoly and Expansion

But such redistribution is not a stable equilibrium. Before, network size n_2 was reached (once the critical mass threshold was crossed) by voluntary association. Further members were not admitted, because they lowered utility to the incumbents. But with internal redistribution, several things happen. There are now incentives for the minority network members to exit the network and form a new one in which they would not bear the redistributory burden. This would be possible if the minority were of a size larger than critical mass, $n_2/2 > n_1$. Even where that were not the case, the minority could band together with those beyond network size n_2 who desire telephone service but were previously excluded.

This exit would deprive the majority of the source of its subsidy and is therefore held undesirable. The only way for the majority to prevent this "cream-skimming" or "cherry-picking" is to prohibit the establishment of another network, both by those wanting to leave the original network and similarly by those not admitted to it by being beyond n_2.

Thus, a monopoly system and the prevention of arbitrage become essential to the stability of the system.

At the same time, and importantly, the model predicts that the network must expand beyond n_2. For the majority, there is added utility from added network members, while most of its cost is borne by the minority. They will therefore seek expansion. The cost to the majority is only that the subsidy by the minority must be shared with more network participants. Therefore, the majority would admit new members up to the point n_5 where marginal utility to its members is equal to the marginal price due to the diluted subsidy.

But this is not the end of the story! With expansion to n_5, the majority is now $n_5/2$ rather than $n_2/2$, i.e. larger than before, and it can also tax a larger minority $(n_5/2)$ than before. Hence, the expansion process would take place again, leading to a point $n_5^1 > n_5$. This process would continue, until an equilibrium would be reached at the point where for a minority member

$$\frac{du}{dn} = 0 \tag{13}$$

with

$$\frac{du}{dn} = \frac{\partial u}{\partial n} + \frac{\partial P_2}{\partial n} \cdot \frac{\partial u}{\partial P_2}. \tag{14}$$

$\partial u/\partial P=1$ by assumption that utility is expressed in monetary units. P_2 is given by $u(n_2)-[F+f(n_2)]/n_2$, and therefore

$$\frac{\partial P_2}{\partial n} = \frac{\partial u}{\partial n} + \frac{F}{n_2^2} - \frac{n\,f'(n)+f(n)}{n^2}. \tag{15}$$

So that

$$\frac{du}{dn} = \frac{\partial u}{\partial n} + \frac{\partial u}{\partial n} + \frac{F+f(n)}{n^2} - \frac{f'(n)}{n}, \tag{16}$$

$$\frac{du}{dn} = 2\frac{\partial u}{\partial n} + \frac{1}{n}(AC - MC). \tag{17}$$

The optimizing point n_5^* is where $du/dn = 0$, i.e. where the equation holds

$$\frac{\partial u}{\partial n} = \frac{(MC-AC)}{2n}. \tag{18}$$

n_5^* will be larger, the greater the marginal utility from added network members is, the smaller marginal cost, and the greater fixed cost are.[10]

3.9. Network Tipping[11]

As this process of expansion takes place, the minority is growing, too. The likelihood that its size increases beyond the point of critical mass n_1 is increased, and the utility of its members, given the burden of subsidy, may well be below that of membership in a smaller but non-subsidizing alternative network. We have so far assumed that there is only one network, and that a user's choice is whether to join or not. Suppose there are no legal barriers to the formation of a new network. In that case, a user's choice menu is to stay, to drop off altogether, or to join a new network association. Assume that the new network would have the same cost characteristics as the traditional network has. (In fact, it may well have a lower cost function for each given size if there has been accumulated monopolistic inefficiency in the existing network and rent-seeking behaviour by various associated groups.)

Then, minority coalition members would find themselves to be better off in a

[10] A very similar analysis can be undertaken for a political system in which network-outsiders have a vote, as in a governmentally directed system.

[11] The terminology of "tipping" I owe to Schelling (1978).

new network B, and they would consider such a network, abandoning the old one. The only problem is that of transition discontinuity. A new network, in its early phases, would be a money-losing proposition up to its critical mass point n_1'.[12]

The point where exit becomes possible, given the redistributory burden that keeps utility just balanced with price, is the point n beyond where

$$u(n/2) > AC\ (n/2), \tag{19}$$

which is at the critical mass point. Hence, the majority will strive to exit the redistributory network once the latter's size is more than twice the size of critical mass.

The majority may attempt to alleviate these pressures to exit by reducing the redistributory burden and thus keeping the minority from dropping out. But that means the network size n_7 would not be optimal to the majority anymore, and members would have to be forced out. And this, in turn, would reduce its majority, so that it would have to drop the subsidizing burden from at least some minority members as the n/2 point separating the majority from the minority shifts leftwards.

This means either higher burdens on the shrinking minority — frustrating the purpose of bribing it into staying — or still less benefits for the majority if it wants to keep the network from fragmenting. Such a disequilibrium process will continue up to the point where network size $n=2\ n_1$, i.e. where the minority may be too small to create a self-supporting new network. One might call this the effect of potential exit by the minority, and results in a lessened redistribution of newcomers to the network in order to keep the first entrants inside.

3.10. Unequal User Size

We have assumed so far that network voters are of equal size. In reality however, some users are much larger in terms of lines n than others. The minority's position would be further weakened if voting were governed by a principle of "one subscriber, one vote" rather than the "one line, one vote" previously assumed.

Suppose users are ordered according to size on Figure 1; in other words, the largest users are those that have joined the network first. This is not unrealistic, since users with great needs for telecommunications are likely to have been the first to acquire a telephone, and early subscribers had the longest time to expand usage. Let us further represent the distribution of lines n for a user v by

$$n = Av^{-a} \tag{20}$$

where $A > 0$, $a \geq 1$.

[12] This corresponds to reality; new networks such as MCI and Sprint have lost money for a number of

The median voter (or median account) is v/2 and its preferences govern. But the network size provided by the users arrayed to the left of such median user is larger than those to the right. They are given by

$$n_m = A \int_{n/2}^{n} v^{-a}\, dv = (A/(1 - a))\ (1 - n/2^{1-a}). \tag{21}$$

n_m, the median account, is to the right of n/2 in Figure 1. In other words, the median voter whose preferences govern is at a network size greater than the median point of the network size. The more the distribution of lines is skewed (the larger the coefficients A and a) the further to the right is n_m. And the more skewed the distribution, the more likely is it that the voting minority will reach, by itself, a size beyond the critical mass point.

3.11. Interconnection

The process of unravelling of the existing network would commence even earlier if a new network has the right to interconnect into the previous one, because in that case it would enjoy the externality benefits of a larger reach $n_A + n_B$ while not being subject to redistributory burden.

Would there exist, for any sub-network, internal redistribution based on coalitions? Once the possibility of exit is established, each burdened sub-group could join another network. Thus, internal redistribution will happen only if a network is unique to some extent, and thus if a burdened user will not readily switch into another network.

Network interconnection means that the network still centres around a society-wide concept of interconnected users. But it consists now of *multiple* subnetworks that are linked to each other. Each of these subnetworks has its own cost-sharing arrangements, with some mutual interconnection charges. Interconnection facilitates the emergence of new networks. It lowers entry barriers. But given entry, it may reduce competition by establishing cooperative linkages instead of end-to-end rivalry (Mueller, 1988). Interconnection is a useful concept, because it responds to the often-made claim that a single network is necessary for universal reach. This is clearly incorrect. Interaction does not usually require institutional integrations, and this was one of Adam Smith's major insights. Otherwise, we would have only one large bank for all financial transactions. But as the next section will show, it also may lead to market failure in the establishment of the original network.

3.12. Subsidies for Reaching Critical Mass

We have mentioned before that waiting for demand to materialize prior to the introduction of a network or network service may not be the optimal private or public network policy. Demand is a function of price and benefits, both of which are in turn functions of the size of the network. Hence, early development of a network may require internal or external support in order to reach critical mass.

This suggests the need, in some circumstances, to subsidize the early stages of the network — up to the critical mass point n_1 — when the user externalities are still low but cost shares high. These subsidies could come either from the network provider or its membership as a start-up investment, or from an external source such as a government as an investment in "infrastructure", a concept centred around externalities. The question is how the internal support is affected by the emergence of a system of multiple networks.

The private start-up investment in a new form of network is predicated on an expectation of eventual break-even and subsequent positive net benefits to members. But if one can expect the establishment of additional networks, which would keep network size close to n_1, there would be only small (or no) net benefits realized by the initial entrants to offset their earlier investment. This would be further aggravated by interconnection rights, because a new network could make immediate use of the positive network externalities of the membership of the existing network that were achieved by the latter's investment. Hence, it is less likely that the initial risk would be undertaken if a loss were entirely borne by the initial network participants while the benefits would be shared with other entrants who would be able to interconnect and thus immediately gain the externality benefits of the existing network users, but without contributing to their cost-sharing. The implication is that in an environment of multiple networks which can interconnect, less start-up investment would be undertaken. It pays to be second. A situation of market failure exists.

How could one offset this tendency if it is deemed undesirable? Patents are one solution. If a contemplated new network arrangement is technologically innovative, it might obtain a patent protection for some period. Where a service is innovative but not patentable, one might create a "regulatory patent" for a limited period of protection, or the initial approval (where necessary) might be accelerated. Similarly, interconnection rights might be deferred for a period, or joint introductions be planned that eliminate the first entrant penalty. But these measures would also reduce the usefulness of alternative networks, and could hence lead to the dynamics of political expansion, redistribution, and break-up described in earlier section.

It is possible, moreover, that none of these measures would be as effective in generating the investment support in the way that a monopoly network would that can reap all future benefits. This would mean that the private and social benefits of networks in the range between n_1 and n_4 would not be realized. In such a situation, there may be a role for direct outside support, such as by a government subsidy. This may strike one at first as paradoxical. Shouldn't a competitive system of multiple networks be *less* in need of government involvement than a monopoly? But on second though, there is some economic logic to this. Just as the subsidies [illegible] that were previously *internally* generated by other network users will have to be raised *externally* (through the normal mechanism of taxation and allocation) if at least some users are still to be supported, so might subsidies to the start-up of a network as a whole have to be provided externally, also through taxation and allocation, where network externalities as well as start-up costs are high enough to make the establishment of a network desirable.

3.13. Social Welfare and Multiple Networks

If network associations can control their memberships, stratification is inevitable. They will seek those members who will provide them with the greatest externality benefits — those that have many actual or potential contacts with. Furthermore, they will want to admit low-cost, high volume, good risk customers as club members. Thus, different affinity-group networks and different average costs will emerge.

But what about social welfare in such a differentiated system? The traditional fear is that the loss of some cost-sharing and externalities brought by a second network would reduce social welfare. But the news is not necessarily bad. Where the network was at n_3 or substantially larger than the socially optimal size n_4, the fracture of the network could increase social welfare, depending on the cost and utility functions, if cost closer to n_0 is reached. Where mutual interconnection is assured, one can keep the externalities benefits (and even increase them) while moving down the cost curve towards a lower AC. Furthermore, the cost curves themselves are likely to come down with the ensuing competition.

The welfare implications of the formation of collective consumption and production arrangements is something analyzed by theorists of clubs. (Schelling, 1969; Buchanan, 1965; Tullock, 1971; Rothenberg, 1976; Tiebout, 1956; McGuire, 1972.) The club analysis, applied to networks, can show:

1. Given mobility of choice, different groups will cluster together in different associations according to quality, size, price, interaction, and ease of internal decision-making. The economically optimal association size need not encompass the entire population.[13]

Optimal group size will vary according to the dimension to be optimized. Optimal group size depends on the ratio of marginal utilities for different dimensions, set equal to the ratio of transformation in production, and is in turn related to size (Buchanan, 1965).

But this does not imply that one should keep networks non-ubiquitous and unequal. Financial transfers can be used.

[13]The results discussed would not hold if the marginal costs of new network participants drops continuously more than their marginal benefit to an existing network user. The latter is unlikely since [illegible]

However:

2. It is generally not Pareto-efficient to attempt income transfer by integrating diverse groups and imposing varying cost shares according to some equity criteria. It is more efficient to allow sub-groups to form their own associations and then redistribute by imposing charges on some groups and distribute to others. The set of possible utility distributions among separate groups dominates (weakly) the set of such distributions among integrated groups (McGuire, 1972). User group separation with direct transfer is more efficient than the indirect method of enforced togetherness with different cost shares. In other words, differentiated networks plus taxation or another system of revenue shifting such as access and interconnection charges, is more efficient than monopoly and internal redistribution.

4. Conclusion and Outlook

The theoretically-based analysis of the model show that under its assumptions a network coalition, left to itself under majority-rule principles, would expand beyond the size that would hold under rules of equal treatment of each subscriber. Such an arrangement can be stable only as long as arbitrage is prevented, as long as the minority cannot exercise political power in other ways, and, most importantly, as long as it has no choice but to stay within the burdensome network arrangement.

But beyond that point, the pro-expansion policy creates incentives to form alternative networks. And the more successful network policy is in terms of achieving universal service and "affordable rates," the greater the pressures for fracture of the network. Hence, the very success of network expansion bears the seed of its own demise. This is what I call the "tragedy of the common network," in the Greek drama sense of unavoidable doom, and borrowing from the title of J. Hardin's classic article "The Tragedy of the Commons" (Hardin, 1968) on the depletion of environmental resources.[14] In the case of telecommunications the tragedy is that the breakdown of the common network not caused by the failure of the system but rather from its very success — the spread of service across society and the transformation of a convenience into a necessity.

The model of this essay identified an evolution of network development with three stages:

1. *The cost-sharing network*
   Expansion of the network is based the logic of spreading fixed costs across a large number of participants, beyond a take-off point ("critical mass"), and of increasing the value of network connectivity.

[14]Tragedy is used in the sense of Alfred North Whitehead: "The essence of traumatic tragedy is not unhappiness. It resides in the solemnity of the remorseless working of things."

2. *The redistributory network*
At this stage, the network grows beyond the size that is optimal to its original members through politically directed expansion caused by the formation of internal coalitions that lead to transfers from some users to others.
3. *The pluralistic network*
Beyond a certain point, the cohesion of the unitary network breaks apart because the dynamics of expansion and redistribution leads to a divergence in the interests of its participants that cannot be reconciled anymore within one network. The results are exit, formation of new networks, and the emergence of a federation of subnetworks. The network has progressed to its "tipping point", where its cohesion breaks up and a multi-network system emerges.

These trends have a certain logical progression. At first the network expands because is makes economic and technical sense. Later, because it makes political sense. But as the network provider succeeds in providing full service to every household, it also undermines the foundation of its exclusivity.

Most countries in the world are still in the phases of the cost-sharing or redistributory networks. A few have reached substantial universal penetration of telecommunications for a number of years, and it is here that the transformation of the network system has begun progressing to its third stage.[15] In the United States, whereas in 1975 virtually all of capital equipment in telecommunications was invested by the carriers, in 1986 it was only 2/3. About $15 billion were invested by non-carriers, mostly large users (Crandall, 1988). Such equipment includes PBXs, multiplexers, concentrators, network management equipment, satellite and microwave facilities, etc. Since there is a strong correlation of economic growth and telephone penetration — roughly an additional telephone per $50,000 GNP — countries with high economic growth are likely to progress rapidly through the first two stages towards the third. Eventually, through economic growth and through the instabilities transmitted from the more advanced countries, many other countries will move into a pluralistic network environment, and be faced with the policy issues inherent to such a system.

Where does all this lead to?

It leads, first, to diversification and disintegration. But in a broader sense, it leads to *normalization*. Normalization means that telecommunications network provision will resemble much of the rest of the economy: no monopoly, some redistribution,

[15]Even in highly developed countries the achievement of substantial network penetration ($n \rightarrow N$) is a very recent phenomenon. In Germany, penetration in 1960 was 12 percent. In 1980, it was 75 percent. In France, it was 6 percent in 1967 and 54 percent in 1983. Hence, the imposition of the costs of the last subscribers is a fairly new development, and responses to it are only now beginning to work themselves out. In the US, universal service was achieved some 25 years earlier.

some international provision, a mix of public and private providers, and some regulation.

The network environment is likely to become essentially a pluralistic network of user associations, or network of networks which are part overlapping, part general, part specialized along various dimensions such as geography, price, size, performance, software value-added, ownership status, access rights, specialization, etc. This is not to say that economies of scale and scope will become irrelevant: there will still be broad-based public networks, and powerfully integrated networks with broadband capability. But just as important will be the economies of group specialization and of clustering. These differentiations will permit users with similar needs, or with frequent interaction, to operate on more efficient networks. It will also permit the traditional networks to be more efficient for their members, since they need not be all things to all people. On the other hand, there may well be market failures inherent to a pluralistic network environment that will retard or prevent investments in new focus of networks.

Where does such segmentation leave future policy makers? It would be naive to expect less regulatory tasks. Many disputes become less intraorganizational within the single network and more public in nature. The main regulatory tasks which the pluralistic network system raises for the future would deal especially with "inter" issues: inter-operability, interconnection, integration, inter-media, international. These issues include:

1. Protection of technical compatibility.
2. Protection of physical interconnection and service access.
3. Protection of free-flow of information content such as embodied in common carriage principles.
4. Establishment of new mechanisms of inter-network redistribution as a substitute for the intra-network cost-shifting, in order to assure the viability of a backbone network.
5. International coordination of national policy arrangements to match the global scope of networks.
6. The prevention of oligopolistic behaviour and of cyclical instability.
7. Protection of internetwork service quality and privacy.
8. Consumer protection in a system of multiple and private networks.
9. Assurance of optimal network investment in an environment where market failure may exist.

None of these tasks is beyond our grasp in terms of complexity or political feasibility. But they require us to end the nostalgia for the simplicity of the golden age, and to imagine a very different network environment.

References

Barnes, J.A., 1954, Class and Committees in a Norwegian Island Parish, Human Relations Vol. 7.

Baumol, William J., John C. Panzar, and Robert D. Willig, 1982, Contestable Markets and the Theory of Industry Structure (Harcourt Brace Jovanovich, New York).

Bott, E., 1971, Family and Social Network, 2nd edition (London).

Buchanan, James M., 1965, An Economic Theory of Clubs, Economica, Vol. 32(125), 1-14.

Crandall, Robert W., 1988, Fragmentation of the Telephone Network: Implications for the Policymaker (Brookings Institution, Washington, DC).

Dasgupta, P. and Stoneman, P.L. eds., 1987, Economic policy and technological performance (Cambridge University Press, Cambridge).

Economides, Nicholas, 1989, Desirability of Compatibility in the Absence of Network Externalities, American Economic Review, Vol. 79(5), 1165-1181.

Elmaghraby, Salah E., 1970, Some Network Models in Management Science, (Springer, New York) pp. 1-3.

Hardin, Garrett, 1968, The Tragedy of the Commons, Science, Vol. 162, Dec. 13.

Heal, Geoffrey, 1989, The Economics of Networks, Columbia University CTIS, Unpublished paper.

Karni, Shlomo, 1986, An Analysis of Electrical Networks (Wiley, New York) pp. 1-4.

Klingman, David J. and Mulvey, J. eds., 1981, Network Models and Associated Applications (Elsevier, Amsterdam and New York).

Knox, Robert, 1830, Elements of General Anatonomy (translated from Inst. edition of "Beclard's Anatomy" by D. A. Beclard) (Maclachlan & Stewart, Edinburgh, Scotland) p. 214.

McGuire, Martin, 1972, Private Good Clubs and Public Good Clubs: Economic Models of Group Formation, Swedish Journal of Economics, Vol. 74(1), 84-99.

Mueller, Milton, 1988, Interconnection Policy and Network Economics. Paper presented at Telecommunications Policy Research Conference, Airlie, Virginia, Oct. 31.

Noam, Eli, 1988, The Next Stage in Telecommunications Evolution: The Pluralistic Network. Paper presented at the ITS Conference, MIT. Working Paper #316, Columbia University CTIS.

Noam, Eli, 1983, Telecommunications Today and Tomorrow, (Harcourt Brace Jovanovich, New York).

Richardson, Jeremy John, Gunnel, Gustafson and Art Jordan, as cited in Rhodes, R.A.W., 1985, Power Dependence, Policy Communities and Intergovernmental Networks, Colchester, Essex, Department of Government (University of Essex, Wivenhoe Park) 6-8.

Rothenberg, Jerome, 1976, Inadvertent Distributional Impacts in the Provision of Public Services to Individuals, in: Ronald Grieson, ed., Public and Urban Economics (Lexington Books, Lexington, MA).

Schelling, Thomas, 1978, Micromotives and Macrobehavior (W.W. Norton, New York).

Schelling, Thomas C., 1969, Models of Segregation (Rand, Santa Monica).

Shepard, William, 1983, Concepts of Competition and Efficient Policy in the Telecommunications Sector, in: E. Noam, ed., Telecommunications Today and Tomorrow (Harcourt, Brace, Jovanovich, New York).

Telecom Factbook, 1986, Statistics of Communications Common Carriers, 1945, 55, 65, 75, 85, Federal Communications Commission.

Tiebout, Charles, 1956, A Pure Theory of Local Expenditures, Journal of Political Economy, Vol. 64(5), 414-424.

Tullock, Gordon, 1971, Public Decisions as Public Goods, Journal of Political Economy, Vol. 179(4), July-Aug, 913-918.

Zacharisen, W. M., 1932, The Atomic Arrangement in Glass, Journal of the American Chemical Society, Vol. 54(10), October 5, 38-42.