NETWORK INDUSTRIES AND NETWORK GOODS

(Introductory chapter to a study for the European Commission)

Claude Crampes, IDEI

September 1997

I. NETWORK INDUSTRIES AND NETWORK GOODS

Introduction

Are network-industries extraordinary economic entities, characterised by features so specific that they escape traditional analysis? Or are they just the meeting place of complex but classical economic problems? The answer to this question is not purely academic. There is a strong policy concern in the idea that a network industry is analytically inextricable because it cannot be dismantled in any way without losing its intrinsic qualities. The observation that compatibility and interconnection produce substantial benefits to the consumer might suggest a rationale for monopolisation, vertical integration and regulation. In practice, however, various segments can be served by non integrated firms and exposed to some form of competition. For instance, in the water, gas or electricity industries, there is no evidence that efficiency is impaired when upstream stages of the production process (collection, extraction, generation...) and the downstream stages (distribution, supply) are fulfilled by different players.

Modern economic analysis is now well equipped for the analysis of network industries. New tools have recently been developed that shed light on alternative policies regarding investment, production, pricing and regulation of these industries.

To obtain some insight into the economic problems raised by large network industries, we begin by giving definitions and illustrations of the concept of a network (section 1). Then, we exhibit purely technological costs and benefits of network activities in section 2. We focus on gains from horizontal and vertical integration which are essential in network activities, even if they are not specific to these activities. The third section is devoted to a discussion of the economic advantages and drawbacks specifically due to a network organisation: density economies, on the one hand, and club and congestion externalities, on the other hand. In section 4, we give some intuitions on how networks can tolerate variable doses of static and dynamic competition without losing their efficiency. Finally, in section 5 we present a typology of network industries, based on the characteristics discussed in the previous sections which is helpful for policy oriented decisions.

1. The concept of network

1.1. Definitions

There is a strong heterogeneity in the population of the so-called network industries and the frontier with other types of industries is somewhat fuzzy. That electricity, telecommunications or postal services must be studied as networks is not questionable. In contrast, food-retails, newspapers or insurance services, although they possess some similar features (e.g. multi-point distribution), often are not called networks.

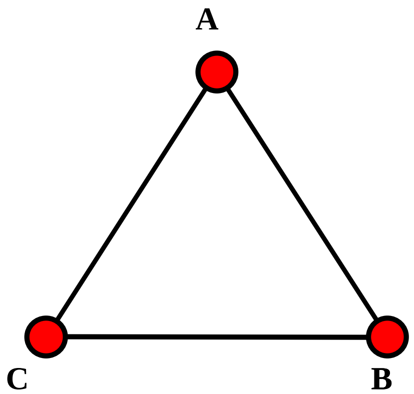
In its simplest economic definition, a network is a set of points (or nodes) and interconnecting lines (or edges) organised with the object of transmitting flows of energy (electricity, heat), information (sound, data, pictures) or material (water, freight, passengers, etc.). Each point can be an *initial node* from which the flow is emitted, a *terminal node*, receiving the flow, or a node that plays an *intermediary* role of transmission, storage, amplification, co-ordination, dispatching and so on. Some networks are one-way, like gas, cable TV and water delivery while others are two-way, like passenger transportation or telephone.

The essence of a network is that (almost) every pair of nodes can be linked by more than one line. Consequently, the path between an initial node and a terminal node is generally not unique. This allows important organisational advantages, but it may also imply some costly drawbacks when the flows cannot be perfectly controlled. This characteristic of multiple potential links explains why many economists enlarge the notion of network to informal organisations, such as the set of the users of VHS video-recorders or the users of PC-type computers. Their common feature is that flows of products or services can be easily transmitted between any two points in the network, with or without the help of a material infrastructure.

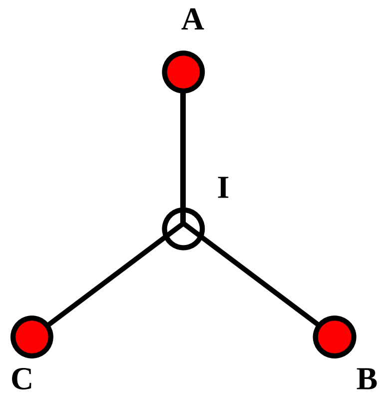
Given the flow transfers to be performed, the best network is the one that minimises total costs. Consequently, the design of a network results from a trade-off between building costs and operating costs. When building costs are very high as compared with operating costs, the best network contains few lines, which may require the creation of interconnection nodes. On the contrary, when operating costs are dominant, the best solution is a dense network with most final nodes directly connected to each other, without any intermediary. (see Box 1)

BOX 1: Dense *vs.* minimal network

Suppose that three agents A, B and C are to be connected within a network. The first solution is a « dense network » like network α that connects directly any couple of the three agents. It is clearly very costly in terms of infrastructure investment. On the contrary, network β is a « minimal network ». In this second type of connection, one needs shorter lines but it is necessary to create an intermediary node I, which any flow will have to cross, incurring operating costs for screening and dispatching.

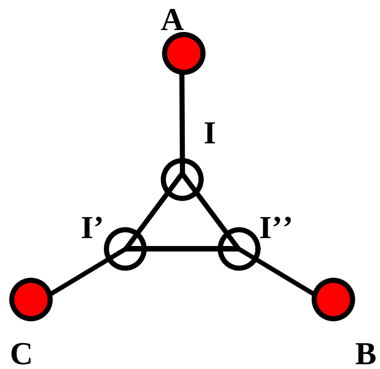


Network α



Network β

Most networks have an intermediary configuration like network γ, reflecting the trade-off between operating costs and building costs. Of course, a final node can be used as an intermediary node for storing, translating or rerouting the flow.



Network γ

1.2. Some illustrations

At this very simple level of definition, we can give examples of network configurations and begin to present some of the economic and political problems they raise. For air transport of passengers, the essential determinants in the configuration of commercial networks are the intensity of the administrative regulation and the development of computerisation in airport management, together with the economic cost of buying (or renting) and operating planes. Since aerial routes are free and since the companies of air transport can lease almost anything (airport slots, aeroplanes, crews, reservation systems, etc.), the design of their commercial network is actually very sensitive to the public authorisations or obligations to operate airlines. While most domestic airlines-networks have long been close to the model of configuration α in Box 1 under regulated access, in countries where competition is at work airline-companies have adopted an « hub-and-spoke » system, that is a minimal configuration β. In most cases, the intermediary point chosen (the hub) was an airport that already existed: for instance point I was located in A which resulted in the disappearance of the direct line between B and C (see also Box 3). Nevertheless, the new design of the networks has resulted in costly arrangements of airports and of company organisations to create hubs, which show that the suppression of lines entails high connection costs.

An illustration of a dense network α is given by the citizen-band addicts who can call each other within a given area using free airwaves. Clearly, the building cost of lines is zero and they do not need any intermediary node for dispatching their calls. Each member can send a message to all others and can receive their messages, provided they are simultaneously present on the same wavelength. If two members want to communicate privately, they have to agree on unusual particular timetables and wavelengths, or to adopt a cipher.

In railway transportation, both building and operating costs are very high. Competition by aeroplanes, cars and trucks has resulted in a loss of density of the rail networks as compared with the first half of the century. The setting of new railnets is highly expensive because of the price of land and the requirements of fixed equipment. This, together with the maintenance costs, explains configurations with a minimised length of lines. But the shorter the total length of lines the longer the average duration of travel, which increases the operating cost. As the designers of the network have to take into account the total cost, the final layout is always a compromise. As an illustration, network γ in Box 1 is not very different from the map of the High Speed Train planned for Texas in 1992, with A = Dallas, B = Houston and C = San Antonio.

The identification of the final points of a network (and consequently its size) is exogenous or endogenous depending on the nature of the flow transported. When the product is not storable, as is the case for electricity, the network necessarily connects the production nodes to the final consumption nodes. On the contrary, for storable goods like letters, the network can be arbitrarily shortened upstream as well as downstream, each user sending or receiving mail at home, or at collective boxes, or at the central post-office. The same is true for water, but because of its permanent use and of public health requirements, the best solution is a direct connection of the drinkable water pipes to the taps of the final user and of the final user's draining system to the sanitation network, except if a cheaper local alternative is possible (e.g. sinking a well). Similarly for gas, each user has the choice between the connection to the distribution network and repeated purchases of gas cylinders. In areas with a very low density of population, the connection costs to a distribution network are too high and the use of cylinders will be preferred.

Also, the design and the operation of a network strongly depend on the homogeneity of the product. A kilowatt-hour of electricity, a cubic meter of drinkable water or a cubic meter of domestic gas are completely standardised products. When they are injected at one point of the network, their destination does not matter since each unit is a perfect substitute for any other unit, which highly simplifies the dispatching. Clearly, this cannot be true for personal letters or telephone service, nor for passengers transportation. In the latter cases, each unit injected into the network is identified from the very beginning and no substitution is allowed. Sorting, bulking and unbulking are essential and costly stages in the operation of the network. The identification of units circulating through the network is also essential for measuring and billing individual consumption. For electricity, gas or water this can only be done at the very final level of consumption while telephone meters can be grouped at a central level since calls are totally identified by the phone numbers of the agents on both sides of the line.

The path followed by a unit of product in the network is more or less under the control of the operators. Actually, in almost all networks an analogue of the electric Kirchhoff laws is at work, that is flows have a natural tendency to take the path of least resistance. When the flow is circulating slowly enough (e.g. water or gas) its path can be approximately controlled by switching on or off the commutation nodes. But in the telecommunications and electricity industries, the high speed of flows limits the permanent control of the paths. In network α of Box 1 for instance, if agent A wants to send energy to agent B at a time when C is connected, a part of the total injection will necessarily pass in transit through C. This transit can result in inconvenience and even in damages for agent C.

Most networks can be interpreted as a set of interconnected sub-networks. For instance, network γ can be seen as the four-point network {C, I', I'', B} connected to the two-point network {A, I}. Historically, most national or international networks have been developed by progressive interconnection of small local organisations. Interconnection raises problems of compatibility in the characteristics of the flowing product and in the operating procedures. Interconnecting nodes and lines are critical components of the entire architecture. The possibility to disconnect very rapidly some subsets from the global infrastructure is of great importance for any network. Actually, such a possibility is essential for the integrity and survival of some activities, for example in the distribution of drinkable water to prevent contamination and, in the electricity sector, to prevent total collapse, like the one that resulted in France on December 19, 1978 from an unusually high demand.

1.3. Operators and users

A wide variety of agents are concerned by networks: designers, builders, owners, users, customers, operators and regulators. Several questions result from this variety and from the nature of network activities.

First, as network infrastructures are generally installed for a very long time, the private or public nature of designers and builders is not neutral in so far as the decision horizon is shorter for private institutions (omitting the private agenda of the public operators). Short-sighted decisions will result in under-investment both quantitatively and qualitatively. The same kind of problem appears for the use of the geographical space. Absent any private incentives to take account of the environmental damages, one can think that the infrastructure installed will not be exactly the same when initiated by a public or by a private agent[1]. This is one of the reasons why the utilisation of public property is to be tightly monitored. For instance, the public authority can keep the property rights on the network industry and exclusively concede its operation to a private firm after a selection process, like in the water activity or for urban transportation. The operator's behaviour is restricted by quality or environmental qualifications, as well as by Universal Service Obligations. The main drawback of the concession system is that since the government keeps the control of access to the industry, when the operation requires sunk investments, the firm will not invest at the optimal level, fearing a potential « expropriation ». When investments can be recovered, one can use alternative ways to regulate private operators. For instance in telecommunications, the operator is the owner of the infrastructure and, as a residual claimant, his investment can be supposed to be more efficient[2].

The infrastructure is used by two types of agents: service providers and service customers. In simple activities, like the first telephone networks, only final users are present. They are directly in contact with the infrastructure operator who interconnects them on a demand basis and charges them correspondingly. When the complexity of the operation increases, new agents can be necessary to satisfy and to stimulate demand. The supply of enhanced services (e.g. inquiries and interconnections in telecommunications, reservation and catering in passengers transportation) can be performed by the infrastructure operator or by new distinct agents or by both. If separate suppliers of services are present, their activity is strongly dependent on the infrastructure operator. Either a provider of final services or not, the operator is very powerful since he controls the access to final users. To prevent risks of abuse, explicit rules and tariffs of access should be established on an efficiency basis.

[1] It does not mean that public firms are necessarily environmental friendly. Some are big polluters.
[2] Unless a tight price regulation is expected, so that the private operator anticipates an expropriation of profits through very low prices. Then, although the owner, he is a residual claimant on nothing, and this drives his investment incentives to zero.

The ownership of the infrastructure is a sensitive problem in any network activity since the infrastructure clearly appears to be a « public good », that is a good that can be used simultaneously and/or successively by several agents (and often by a large number of agents) without any loss in the quality of use. Additionally, the use of the network by one agent can create indirect beneficial or detrimental effects on the other agents. The control of and payment for these externalities need the presence of a « super-agent » able to rationalise the behaviour of small independent users. This agent can be a private, a public or a mixed entity, but it needs some collective concern to handle the maintenance and development of the infrastructure.

2. Costs and benefits from integration

Because of their complexity, one could think that networks cannot be analysed with traditional economic methods. This sometimes alleged inability would be the mere reflect of the intrinsic integrity of networks, with the political conclusion that large integrated firms are the only suitable solution for an efficient organisation of network activities. Actually, as we show in this section, modern economic analysis has numerous high performance tools in hand, and keeps on increasing its toolbox. Some features of networks are already well-known, as they are shared with many other economic activities; examples are economies of scale or economies of scope. Others are more recent fields of knowledge, and deserve more research, such as the gains from rerouting or the economies of clubs. Thanks to these instruments, economic analysis is able to determine without dogmatism when the need for co-ordination of activities is so intense that it entails some degree of integration,. Moreover, technical integration and economic integration are not synonymous, and economic integration is not necessarily the best way to reach efficient co-ordination.

2.1. Economies of scale and economies of scope

In any economic activity, economies of scale become apparent in the decrease of the average cost of production when output is expanded. Economies of scale result from large fixed costs and/or weakly increasing variable costs. In all the networks where a physical and/or logical infrastructure[3] is essential, strong economies of scale are at work since once the equipment is installed, any additional unit of output lowers the fixed cost per unit of product. This decrease in unit cost is effective up to the technical limit of the equipment, but some gains from « squeeze » are sometimes still available for a given saturated equipment as we will see later (cf. *infra* 3.1).

Now, when designing and installing the infrastructure, one can rely on the so-called « surface/volume effect » to lower the unit cost when output is an increasing function of the plant volume while total cost increases with the envelope of this volume (see Box 2). This effect leads to the installation of very large equipment to obtain very small average costs. It is at work in the design of equipment for storage and transportation of oil, gas, freight, parcel post. The surface/volume effect is also essential in the evolution of passengers' transportation by plane (wide-bodied jets), bus or train. For trains, the size is obviously limited by the size of bridges and tunnels as well as by the width of railways. Also note that when the size of the equipment is increased, the operating costs increase less than proportionally. For example, in the transport of freight, the crew of planes or boats is almost independent of the size of the planes or boats; similarly, it takes one person to drive a truck, whatever the truck.

The variable-cost component of economies of scale is less obvious in network activities. First because the definition of the activity can be imprecise due to vertical integration (see *infra* 2.2). For instance, generation costs are part of the total cost in an integrated power system while manufacturing costs are obviously not included in the cost of a freight transportation network. For this reason, the successive stages of the activity should be disconnected, at least from the accounting point of view, before any attempt to quantify the economies of scale. A second reason is that in most networks, output has an essential geographical characteristic. For instance, in freight transportation output is measured in tonnes-kilometres. Consequently, output can increase because of an increase in the quantity of commodity transported, because of an increase in the delivery area or because of both, with obvious divergent effects on operating costs. With a given equipment, operation costs are more or less proportional to the quantity produced, but increase more than proportionally with distance, except in telecommunications.

The more standardised the product, the stronger the economies of scale. But for heterogeneous products, economies of scope can still be at work. In networks, when the infrastructure of storage and transport is not totally dedicated to one specific activity, there still remain possibilities of gains from diversification of activities. The reason is that the infrastructure has a nature of public good which allows to incur less than the stand-alone cost when one adds a new activity to the other activities.

Definitely, a pipeline cannot transport water if it is currently used to transport gas or an oil tanker cannot be occasionally devoted to the transport of food products. But on the other hand, in a warehouse one can store a wide variety of commodities and a distribution network can deliver very heterogeneous parcels, including passengers, with minor additional costs. For instance, some planes have two lives: during the day they transport passengers and at night, seats are removed and they are used as postal planes.

Telecommunications are the industry where recent changes in technologies have dramatically focused on the gains from economies of scope. Thanks to the use of optical fibres and of digital signals, it has now become indifferent to transport data, sound or pictures. Consequently, a telephone company is a good candidate for data-base transfers, and a cable-TV network can enter the telephone market at a reasonable cost[4]. Moreover, when wires are to be used, the control or ownership of a part of the public property is essential. This explains that some potential challengers in telecommunications (broadly defined so as to include transfers of data), such as power, gas, water and rail utilities are companies with access to the public property.

[3] By logical infrastructure, we mean the whole operating system of the network.
[4] The cable operator will have to upgrade its one-way network to make it a two-way network.

BOX 2: The gains from large scale operation

Let $TC(q) = FC + VC(q)$ denote the total cost of production TC with a fixed component FC and a component VC increasing with the level of output q. There exist **economies of scale** when the ratio TC/q is a decreasing function of q. The higher the output, the lower its total unit cost. In most networks, the costs of infrastructure are very high so that the unit fixed cost FC/q is an essential part of the total average cost and economies of scale are important on this side of the saturation threshold of the equipment. For VC/q to decrease, the variable cost has to increase less than proportionally to output, which can be explained by standardisation and learning.

The **surface/volume effect** is a strong determinant of the large scale of many industrial, storage or transportation equipment. In many industries, building and maintenance costs are directly dependent on the surface of the physical envelope of the equipment (walls and roof of warehouses, tubular steel of pipelines, hull of tankers...) while the output can be measured by the volume of the equipment (commodities stored, gas injected, oil transported...). As surface increases with the square of the equipment scale while volume increases with its cube, the ratio surface/volume, that is the average cost, is a decreasing function of the scale. This effect leads to the building of very large equipment, limited by the resistance of materials and by environmental considerations.

Stochastic gains result from the decreasing probability that independent events occur simultaneously when the number of events increase. Then the reliability of a production or a transmission node increases with the number of substitute pieces of equipment installed, while the maintenance cost (spare parts, repairers) does not increase in proportion. In telecommunications, for a given quality of service (a given probability that a call can be transmitted), the ratio of the entrant traffic to the number of network gears is increasing (Erlang law).

BOX 2: **The gains from large scale operation** *(continued)*

Economies of scope denote the ability of firms to produce several goods at a total cost smaller than the sum of the costs of these activities isolated one from the other. If we consider the production of quantities q_a of good a and q_b of good b, one simple definition of scope economies is $C(q_a, q_b) < C(q_a, 0) + C(0 , q_b)$ where the costs on the right-hand side of the inequality represent *the stand-alone costs*. These economies result from the ability to use some pieces of equipment and a part of the workforce for producing goods or services that are not too different.

An extreme case of scope economies is represented by inputs that possess the nature of « public goods », i.e. that are not destroyed by use, like informational inputs. For instance, the data base made of the clients of a bank can be used without any additional cost to propose insurance services. Then, economies of scope are a strong incentive for the diversification of activities.

Vertical integration refers to the performance within a single firm of a number of successive distinct operations needed to produce a given commodity. From a public point of view, vertical integration can be justified in so far as it allows to decrease costs for producing the final product. This decrease can proceed from technical complementarities, for instance through a better co-ordination between the successive stages of a production process. Also, the social gains can result from the removal of successive private profit-margins obtained by intermediary firms with some market power. But upstream vertical integration is also a device for increasing the control a firm already has on its suppliers, clients and competitors. By controlling the delivery of certain strategic inputs, a firm can win a dominant position on its final market. In network activities, the most significant example is the control of the access to transport infrastructure by a firm that is also a user of this infrastructure.

2.2. Vertical integration

Any production process requires a sequence of elementary operations that can be performed either within the same firm or by independent firms. A vertically integrated firm controls all the successive operations until the final consumer. For instance, in the electricity industry vertically integrated firms providing generation, transmission and supply are common. The drawback of this arrangement is the lack of competitive incentive at the intermediary stages of the production process. The upstream units are too far from the end of the process to be influenced by market reactions, and they are certain that their output, even if its quality is low, will be accepted by (inside) users. For this reason their level of effort will probably be far from efficient[5]. But vertical integration has strong advantages in terms of access control and in terms of co-ordination.

Rapidity and reliability are necessary qualities in the operation of networks. When an activity is not vertically integrated, each part of the production process is under the control of separate entities with potentially divergent objectives[6]. This can result in a very low performance as compared with an operation under the control of only one supervisor. Suppose a passenger takes a plane to go from A to C and there is a stop at an intermediary point B. When segments AB and BC are operated by distinct carriers, it is less likely that arrival at and departure from B will take place at the same terminal, or that the timetables will be compatible, or that luggage will not need a new registration, etc. In all the transport activities, vertical integration can facilitate the resolution of problems at the nodes, particularly the transhipment of cargo in multimodal transport. In electricity, the need for co-ordination is obvious as the flow is not storable and moves at the speed of light, but technical co-ordination can be performed by an independent firm.

The social drawback of vertical integration in network activities is the potential abuse of power in the control of access to infrastructure. In any economic activity, the control of an essential input is a major cause for upstream integration, and conversely the owner of an

[5] To foster efficiency, one can try to mimic market incentives by using complex reward schemes.
[6] Objectives are also divergent within an integrated organisation, but it is the very rationale of the organisation to limit the degree of divergence.

essential input has a strong incentive to integrate the downstream activity. Thanks to this control, a firm becomes responsible for the quality of the input in question and for the regularity of deliveries. But it can also take unfair advantage on its use, which is bad from a collective point of view.

In network activities, it appears that the essential input is the transmission infrastructure. For example in electricity, it includes the transport infrastructure for high-voltage transmission and the distribution lines for the electricity under low-voltage. As the good or service has to be delivered to every final user and as it would be too costly to install more than one delivery system, the one who controls the transmission system clearly controls the access of any upstream provider to the final market. And if the infrastructure owner is also a final-service provider, a fair competition is problematic[7]. For this reason, vertical integration is not necessarily the best economic answer to the need for technical vertical co-ordination. For example, in most countries, airlines companies are not the owners of airport installations. On the other hand, some have got the control of reservation networks which also are essential inputs for this activity. In the British experiment for rail traffic, train operators will be separated from the ways and stations operators.

For competitors, the alternative is to bypass the infrastructure, either by installing a new one or by developing a new technology. The duplication of an infrastructure is bad on pure technological grounds, because economies of scale from each network will not be totally exploited. This occurred at the beginning of the service-network era with the duplication of gas first, then electricity networks in most countries. It still exists in North America for some railroads and in some US towns for Cable TV. In telecommunications, wireless technologies allow new entries without having recourse, at least partially, to the installed infrastructure. In the same way, TV transmission by satellite is a complete bypass of cable operators.

[7] In the US gas industry, before the 1992 Order, entrants used to complain against the transportation firms for interrupting their service when the operators needed the pipeline capacity for the sales service of their own firm.

2.3. Cross subsidies

Vertical or horizontal relations between production segments open the possibility, and sometimes the necessity of cross-subsidisation. This means that a profitable production segment feeds another one by transfers or biased sharing of common costs, either to make it economically feasible (which is good for the economy if the social utility from the mix of products is larger than its cost) or to flatten a competitor by an artificial price cut (which is obviously bad).

In the industries where vertical or/and horizontal economies of scope are strong, competition is not a very effective barrier against this practice. When an integrated firm sells simultaneously on several segments, actual competitors selling on separate segments cannot take advantage of the synergy between activities. Of course, actual competition as well as potential competition prevent the integrated firm from exaggerating the subsidy amount. On the market where the firm involved levies its resources in order to support the other activity, the price cannot exceed the average stand-alone cost (see Box 2) without attracting the entry of challengers. But the administrative limitation of price cuts (or the drastic solution of disintegration) remains a necessity to keep open a minimal competition on the subsidised segment. An example is the electricity sector where it is quite easy for a distributor to sell a cheap energy subsidised by high connection fees levied on captive users.

Now, in the sectors where there is little gain from integration, public intervention is not necessary to prevent cross-subsidies. Competitive mechanisms are quite sufficient. On the contrary, the government must intervene ... to restrict entry in the « rich » segment if the support of the « poor » segment is a political or social necessity. This is the case of any clause of « Universal Service Obligation » imposed to an operator (see *infra* Part three). In the postal sector, to fix a (low) uniform price for short distance and long distance letters is not feasible if the operator is not protected against entry in the intra-muros segment.

3. Density, externalities and interconnection

In this section, we present economic features that are more specific to network activities and that are to be taken into account in the appraisal of the performance of operators. First, we introduce the costs and gains from « compression » that are present in all the network activities even if their importance is not the same everywhere. Then, we show how the positive and negative externalities created by the use of the network can change the behaviour of the agents and result in a wide variety of alternative equilibrium configurations.

3.1. The economies of density

Since one of the objectives of network industries is to interconnect isolated agents, one can think that the more numerous the members of the network, the more efficient the industry because the infrastructure cost is divided between a larger number of members. Nevertheless, networks are strongly dependent on geographical characteristics and the number of agents can be meaningless without any reference to the area affected. In the same way, since networks are used for transferring flows, one could say that their performance is increasing with the global quantity of flow processed. Actually, a given flow passes in transit through lines at the same time as other flows or in a specified order, crosses other lines that transport other flows, is stored, transformed, re-routed at certain nodes where other flows are processed, etc. Consequently, a measure of performance is highly dependent on the physical and organisational characteristics of each network, more precisely on its ability to transfer several flows, simultaneously or sequentially, with a high quality of service. So, it appears that density (the density of connected agents as well as the density of goods or services processed) is a feature indispensable to estimate the performance of any network.

Consider first the geographical density of agents to be interrelated within a network. Clearly the cost for connecting a given number of people is increasing with any measure of the distance between them. Pipes, wires, ways or fibres are expensive to produce and to install. Moreover, operating costs are increasing with distance because of the equipment and the quantity of energy needed for transport or emission. If agents are too much scattered, one single interconnected network can be a very poor arrangement. It would be less expensive to organise small local, mostly non-connected networks.

In the case of water for instance, most distribution networks are organised under pure geographical considerations using gravitational forces. The essential constraints are the natural availability of water at each point (waterway, ground water...) and the slope between the taking nodes and the supply nodes. Historically, urbanisation has increased the need for networks of drinkable water and sanitation, as Roman vestiges show. The present outcome is a large number of small networks without any permanent interconnection within a super-network, except when there is a structural shortage in a region.

Consider now the case of electricity. It can be produced with very different technologies, from small to vast, so that power plants can be installed almost anywhere. Physically, one can imagine electricity without any network, each user generating just enough to satisfy his own needs. However, networks have been installed from the very beginning of the electricity era. Large scale economies are obvious pieces of explanation but the main reason was the need for light in high density urban zones. For interconnections of local networks, the main drawback is that power losses are increasing with distance and are proportional to the square of the current transmitted. But the use of alternating current and high-voltage transmission allows to limit these losses because alternating current may be easily converted to higher or lower voltages by means of transformers and because the greater the voltage on the line, the smaller the quantity of current. Additionally, on a given line the flow of power can go in any direction. Consequently, networking appears as a low-cost solution for providing a non storable good to people with non-simultaneous consumption except for remote and/or non permanent locations. The additional cost of large networks is that they need an organisation able to co-ordinate numerous decentralised decision nodes. In the electricity industry, dispatchers have been playing a central role from the very beginning.

High density of users does not entail only advantages. Crowding can be the source of serious drawbacks mainly in transport activities. When people want to trade more material commodities or decide to travel more, the higher the density the more likely their lines cross each other. Costly works are necessary to prevent congestion of lines and of storing nodes as well as collisions at crossing points. For instance to regulate the traffic at cross-roads, some cheap « software » solution can improve the performance (traffic lights, no U-turn or no left turn signals). But if traffic is very dense, « hardware » changes will be necessary, for instance road broadening or subway digging. An intermediary option lies in arranging the cross-roads with a roundabout.

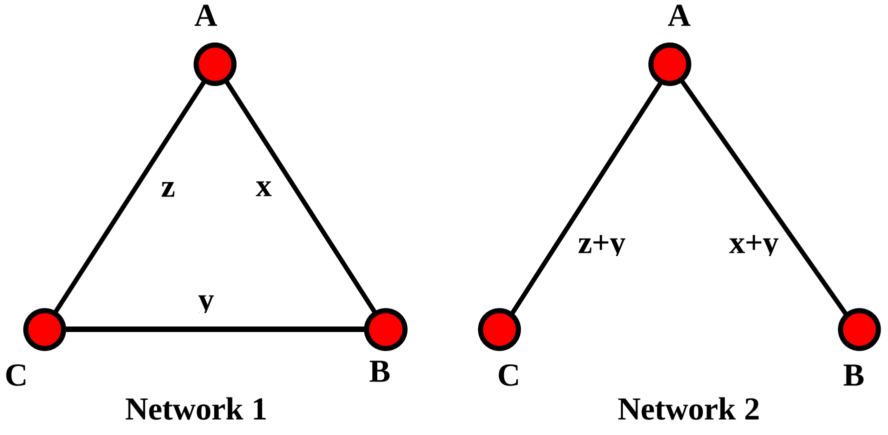
To increase the flow of commodities, passengers, information or power through an installed network at low cost is the dream of all operators. In some activities like gas or water this can result from higher pressure, at the cost of more fuel and sometimes new pipes. In electricity, better performance is obtained thanks to the use of very high-voltage transportation, more efficient transformers and improved isolation of lines. During the last two decades, outstanding progress in the transport of goods has been made due to the standardisation of containers which allow to shorten the delay for transhipment. Also, the development of several types of coding (bar code, postal code) has greatly facilitated the regulation of flows and inventories of goods.

For the next decade, one can expect that the industry most affected by these « squeeze gains » will be telecommunications. As far back as 1875, Edison's quadruplex boosted the development of telegraph because it allowed the simultaneous use of the same line by two operators on one end of a line and two operators on the other end. So Edison succeeded in multiplying by four the capacity of each telegraphic line. Similarly, the use of digital signals instead of electromagnetic signals allows a fast growth in all the forms of communication. Digital data can be generated directly in a binary code by a computer as well as they can be produced from a voice or visual signal by an encoding process. Then these coded data can be easily compressed before they are transmitted, on the condition that the receiver possesses the device to decode them. The result is for example that on a given TV channel, one can transport eight times more digital programs than analogical programs. Due to this increased density in the transport of signals, there is an enlarged variety of choice for any individual user (helped by computers) and/or a cheap possibility of transforming any one-way line into a two-way line (like telephone). Then, all the economic operations that need interactivity between agents could be dramatically changed. Many fields will be and already are concerned by digitalisation: telephone and TV, but also many forms of entertainment and education as well as many service sectors (banking) and mail-order shopping.

Of course squeezing is not quite safe. First because a network with a high degree of compression is likely to have a low quality of service so that clients can be discouraged to use it. But one can think that the operators, or at least the regulator of the sector, will limit the level of compression. Second because compression is obtained at the cost of a low redundancy which means less security. Take the example of a firm operating a water network. When a new group of customers appear, the operator has several possibilities to supply them. First, a direct independent connection to a take node, which is not the best solution in a high-density population area. Nevertheless this solution is very secure since the risks of cuts are divided between two independent infrastructures. Second, a simple connection to the existing network, which is very cheap but can provoke important pressure drops, resulting in a high dissatisfaction of old and new customers. Third, a connection to the existing network, combined with an increase in the injection pressure. Clearly, now the risk is that old pipes cannot support this new constraint and that leakage increases, lowering the overall performance. A fourth alternative is to let the users install and operate their own equipment, for example to pump up water and store it in tanks. Its main advantage is to fit the individual needs. Also, pressure is not increased in all the pipes. But it has the social drawbacks to be more costly than a centralised "booster" and to decrease pressure in other parts of the network.

<u>BOX 3</u>: <u>The gains from rerouting</u>

In a network, the optimal way from one point to another is not necessarily the shortest path. When designing the infrastructure as well as when operating it, one has to take account of the traffic on all the neighbouring edges and through all the neighbouring nodes. As there exist several possible paths between two points, one portion or the whole traffic shall be diverted to minimise total cost.



Consider for instance the traffic forecasts x, y and z between nodes A and B, B and C, C and A respectively. If $D_{ac}(z + y) + D_{ab}(x + y) < D_{ab}(x) + D_{bc}(y) + D_{ac}(z)$ where $D_{ij}(.)$ stands for the total cost function for traffic between i and j, it is worth to constraint all the flows between B and C to transit through node A. The solution will be minimal like Network 2 instead of the dense Network 1. The drastic solution is to never build line CB.

<u>BOX 3</u>: <u>The gains from rerouting</u> *(continued)*

Actually the solution can be intermediary, that is line CB is installed but a portion of traffic between C and B is transferred through point A. Consider the following illustration:

line	**AB**	**BC**	**CA**	total
traffic forecast	x	y	z	
night	1	3	1	5
day	2	1	2	5
capacity				
independent solutions	2	3	2	7
optimised solution	2	2	2	6

On lines AB and AC, traffic is lower at night than during the day and conversely on line BC. In a sub-optimal solution, one installs on each line a capacity equal to the maximal flow to transfer, which means a total capacity of 7 for the whole infrastructure. A better solution consists in installing less capacity on BC and, at night, to transfer one unit through the lines AB and AC. The resulting network is saturated during the night and almost saturated during the day.

3.2. Clubs and networks

In a network, the interdependency of agents generates strong externalities, that is effects of the decisions by each member on the level of satisfaction of other members. These externalities can be positive, which means that an action by a member increases the utility of the others: we will refer to this type as « club externalities ». On the contrary when they affect negatively the utility of other agents we will speak of « congestion externalities ». In many cases, the size and composition of a network are the result of a trade-off between club benefits and congestion costs.

Positive externalities enhance the utility from being a member of a network and the larger the network, the stronger the club effect. In networks where final users are the main stimulating persons of the service, like telephone or electronic mail, club effects are direct, which means that the gain from being a member is directly increased by the entry of new persons. There exist indirect club effects when one benefits from the large number of members only through some equipment that possesses the nature to be a public good. For instance, the holders of a specific credit card are better off when they become more numerous because more automatic teller-machines are installed and more shops accept this mean of payment. The same kind of externality exists in informal networks like the users of computers and video recorders: the more numerous they are using one specific electronic device, the richer the variety of programs or films the producers will make available[8].

At a given access price, when the composition of the network changes qualitatively or quantitatively, some members can be induced to leave while some non members would like to be registered. This means that the number and/or identity of those who *wish to be in* depend on the number and/or identity of those who *are in*. An equilibrium is reached when demand is compatible with itself, that is when the number of actual members (and possibly their identity) is exactly the same as the number (and identity) of the candidates (see Box 4). But because of the club effect, at a given price there generally exist several equilibria.

---

[8] Actually, these « indirect club-externalities » result from economies of scale and/or of scope: when the number of customers increase the network operator can decrease its unit cost and/or propose to sell a wider variety of services. This effect is at work in any industry.

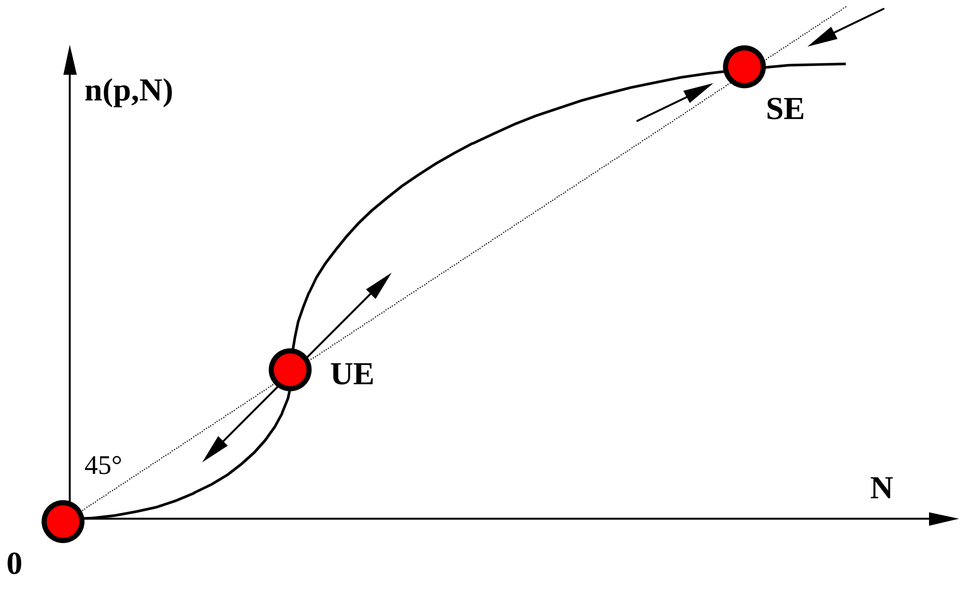
Now, all the potential equilibria are not identical. Some are stable because any slight exogenous shock like accidental entry or exit will be counterbalanced by a spontaneous exit or entry letting the final configuration unchanged. Others are quite unstable: a small initial shock in one direction or the other develops into a « snowball effect ». For instance, as some members leave, the club loses its attractiveness and others will follow out. If this second wave of resignation is larger than the first one, it is the beginning of a positive feedback which can end only with the complete vanishing of the network. The smallest unstable equilibrium is the « critical size »[9]. It is essential to identify it before trying to launch a new network. If the operator is not able to reach this minimal equilibrium in one jump, he will never succeed. Many examples of this failure can be found in the telecommunication industry. In France, a service of data transfer by radiowave named Mobipac was interrupted at the end of 1996 only three years after its beginning. In the same way, Bi-Bop, a mobile telephone service will soon disappear because of the disaffection of users. In England the « Telepoint » mobile phone started in 1989 and disappeared in 1992.

Of course, the equilibrium size of a network is also highly dependent on the state of the infrastructure at each moment. If too many people want to use a telephone network at the same time, they can provoke a complete breakdown. Why do networks usually break down? If we except a simultaneous shock on several essential parts (flood, storm, etc.), the main reason is that a line is broken and the other lines cannot stand the (automatic or semi-controlled) rearrangement of flows. For instance, thermal constraints become suddenly too high in an electric grid, like it happened in California at the beginning of August 1996. Alternatively, an isolated problem in one site can spread through the network and can result in a failure of the whole system if no safeguard device has been installed. In computer networks for instance, if disconnection systems are not installed, a small initial incident on one computer can provoke the chain interruption of all the programs, computer after computer. For example, on August 7 of 1996, America On Line, a gateway provider to Internet, suffered a complete collapse during the whole day.

[9] It should be stressed at this stage that the equilibria depend on prices, the access price as well as the price of use. So actually there is not one unique « critical size » but one for each level of the overall price. The one we refer to here is the critical size corresponding to the lowest price the operator can charge without losing money.

<u>BOX 4</u>: <u>Network externalities and multiple equilibria</u>

In order to illustrate positive externalities that appear in networks, denote n the number of people who *would like to be members* of a club and N the number of *actual members*. If p is the entry fee, n(p,N) is decreasing in p. But it is likely to be increasing in N because a large club means a high quality of service and many opportunities to meet people and to establish relations. For a given p, if n(p , N) > N like in the central part of the figure below, the actual network contains fewer members than the number of those who would like to be members. A new entrant transforms the size of the network from N to N+1, creating an additional externality. If this externality is very strong, it creates a «snowball effect» that creates a large network. When this positive feedback is exhausted, one reaches an equilibrium -that is a size N- such that n(p, N) = N. For such a value of N, no outside agent wants to enter and no inside agent wants to exit. In the figure, equilibrium is reached when the curve n(p, N) crosses the 45° line. Conversely, when n(p , N) < N like on the left and right parts of the figure, a process of downsizing is at work.



<u>BOX 4</u>: <u>Network externalities and multiple equilibria</u> *(continued)*

Depending on the shape of the function n(.,.) and on the value of p, the « membership curve » can cross the 45° line only for N = 0 or many times, which means that there can exist a large number of equilibria. Some, like SE, are stable equilibria: if a member leaves for any exogenous reason, he is replaced by a new one and an entrant expels an old member. Others like UE are unstable because if one member leaves, several others follow him and if a new member subscribes, he incites many others to do the same. For a given price, the highest stable equilibrium is the saturation size of the network. The smallest unstable equilibrium is the critical size. If the operator is not able to reach this threshold, the network cannot exist at this price level. Put differently, for this price the equilibrium size will be the smallest stable one, that is N=0.

One can find a simple example of repeated network externalities in public or private lotteries. The number of participants n is a decreasing function of the ticket price p and an increasing function of the jackpot J (most players are not aware that their probability of winning decreases with the number of participants). Assuming that all the money collected is included in the jackpot, there is an equilibrium when $p \times n(p,J) = J$. Clearly, several equilibria are possible. A small J will attract little people which means little money to be included in the prize. On the contrary, a large jackpot should be attractive enough to collect the money necessary to pay it.

4. Networks and competition

When we speak of competition in network activities, one must specify whether competition takes place *ex ante* or *ex post*. The reason is that even in the activities where economies of technical integration are so strong that a monopoly is the most-efficient social solution (e.g. very-high-voltage-electricity transportation), *ex ante* competition can still be used as a regulation device[10]. For instance, a natural-monopoly position can be allocated as a temporary franchise through an auction process, like the sealed-bid auctioning of the water concession in Buenos Aires. If the process is opened to all specialised firms and designed to be collusion-proof, it will result in the allocation of the franchise to the most efficient candidate. And if the incumbent is threatened to lose its position at the end of the concession period when a new bidding procedure is opened, it will be stimulated by strong incentives to manage efficiently the industry it has in charge.

In this section, we focus on *ex post* or actual competition, that is competition between two or more firms already active in a market. We first show how network specificity modifies the model of competition used in traditional Industrial Organisation (4.1). Then we examine the problem of dynamic competition between network firms (4.2). The features of *ex ante* competition are essentially those of auctions mechanisms, which are beyond the scope of this introductory chapter. They are sometimes used to allocate franchises that are briefly presented in section 4.3.

4.1. Competition under network externalities

Probably the most specific characteristic of competition between networks operators is the very large number of potential equilibrium-configurations for a given set of competing firms. This multiplicity results from the club effect described formerly in Box 4. Because of the club effect, the attractiveness of a given network is closely dependent on what each potential customer thinks about the total number (or the identity) of customers in each competing network. The point is that the quality of the service offered by each firm is

[10] The « natural regulation » of industries with natural-monopoly characteristics through potential competition is presented in section II: Government intervention in network industries.

endogenously determined by the participation of clients. And it is the expected quality that matters more than the true quality, if any. This expectation effect is a decisive determinant in the dynamics of competition between network operators as we will see in the next section. In any case, competitors cannot survive if they do not reach the critical size corresponding to the price they have chosen (see Box 4). This « market threshold » is an additional limitation to the number of active operators in a given network industry, the first limitation being the usual technological threshold: if firms are too numerous they cannot pay back their operating costs. For all observers of the English mobile-phone experience (Telepoint, 1989-1992), the failure resulted from the inability of the four franchisees to set up a base large enough for their network to be attractive to new customers. On the contrary, at the beginning of the French Minitel experimentation, France Telecom had chosen to distribute the terminals for free, despite its monopolist position. Later, after the critical threshold was cleared, France Telecom organised the sale of the terminals in its local agencies.

A second key feature of network competition is that it can be transformed into some form of co-operation by decisions of compatibility, sometimes unilateral decisions. When the utility of goods depends on the number of their users, a decision of compatibility between several products or services dramatically changes the size of individual customer-networks. For instance, when the two French credit-card networks decided to make their cards compatible in 1984, the utility of being a user was instantaneously increased. Compatibility can be achieved by standardisation. The most striking example is the electricity industry, where there is a long tradition of co-operation at the national and international levels, with frequent power exchanges through interconnections. The reason why producers may prefer not to standardise is that standardisation makes products more homogenous. Consequently, it can enhance competition and reduce the profit margin of sellers. For each producer, the trade-off is between a small weakly competitive market (few clients multiplied by a high mark-up) and a large highly competitive market (many clients multiplied by a low mark-up). Standardisation can alternatively be achieved through the use of gateways. In many cases, it is a cheaper solution and in some cases, it can be decided unilaterally. For example in the activities of freight or passengers transport, an operator can prefer to use buses or trucks on given segments of his own network or his competitors' networks rather than to install railways or to operate airlines. Multimodal platforms that many regions are installing to improve the interconnection between their transport networks are gateways that allow to take advantage of complementarities of heterogeneous systems. But they also stimulate competition on some parts of the networks.

From the social point of view, standards and gateways are not equivalent because of their effect on the product or service variety. While achieving compatibility through gateways increases the possibility of choice of the customers, standardisation on the contrary reduces the variety offered on the market.

Pricing is an essential element of the competition among firms in any market. This holds in particular when these firms are network operators. When they propose very similar services, price competition is fierce, because price is the only element that differentiates sellers in the opinion of buyers. The best known example is the competition between airlines companies on a given destination (even with railways companies when the distance is short enough). To avoid this competition, operators try to make their services heterogeneous, by improving the provided quality or by specialising on some market segments. But obtaining these segments can be a source of competition by itself. For example, all airlines companies would like to get the most profitable take-off and landing slots, early in the morning and late in the afternoon. The allocation of these slots often remains traditional: the oldest firms are privileged by « grandfathering » practices. But in the USA, some airport authorities have begun to allocate them by means of auctions, and in some countries there exists a grey market where slots are reallocated.

In networks where a physical or an electronic connection is needed before the service can be consumed, pricing competition is more complicated because the service providers can combine access fee and service fee, in such a way that it is very difficult for a given user to know where the cheapest offer is. For instance, in the French mobile telephone, three operators are currently proposing free or expensive connection (to get the telephone set), monthly or quarterly subscriptions (to get the right to use the telephone set) and linear or multilinear prices for the utilisation. The use of complex multipart tariffs has several advantages for the sellers. First, as we have just mentioned, it makes the tariff grid obscure for the users so that it relaxes competition. Second, it is a way to win the loyalty of customers. When customers have to pay a two-part tariff with a low variable part, the cost of switching to a new seller before the expiration of the contract is represented by the high fixed-part. This is an essential feature of the competition between the satellite and cable operators for TV programming as well as between telephone operators. Third, it is a legal method to discriminate between customers. By proposing to all users several combinations of fixed and variable prices, the operators lead each customer to select an average price different from the price paid by others. This discrimination is authorised despite the usual uniform-price obligations, because the same grid is proposed to everybody and discrimination results from a process of self selection.

As we have seen formerly, some pieces of network infrastructures are not totally devoted to a specific activity. When an infrastructure admits alternative uses, the operator can change very quickly its activity or he can diversify, which means to be present simultaneously on several markets. In the case of quick moves from one market to another, competition is more intense since the number of sellers can change very rapidly. For example, an airline company can easily modify the allocation of its planes and crews if it appears that some lines are more profitable than others or if the access to some protected lines is now authorised. In freight transport, if trucks are not dedicated to a given product, the producers can adapt their behaviour to sudden variations in demand. Therefore, alternative but not simultaneous uses increase competition. On the contrary, when a given piece of equipment can be simultaneously used for several activities, the operator can run it as a mechanism of cross-subsidisation. Then, there is a possibility of unfair competition on some markets. For instance, with the same telecommunication infrastructure it is now possible to send sound, pictures or data. Consequently, cable-TV operators who possess a strong position for delivering entertainment or sport programs can collect resources from their TV activity and compete fiercely in the provision of telephone services.

Finally, we can evoke the access problem that will be developed further in this report. How to organise competition around an essential infrastructure efficiently ? In the activities where a given equipment is necessary to deliver a good or service to final customers, if it is too costly to duplicate this equipment, the firm that controls it has the ability to discriminate between the service providers. The classic example is the transport and distribution infrastructure for electricity generators. With this power in hands, the infrastructure manager can spoil the competition between generators by mimicking the behaviour of a vertically integrated firm. The problem can be worse if the firm that operates the infrastructure is also allowed to sell the final service, like in the British telephone sector. The operator is then simultaneously a provider and a competitor for the final service producers. In any case, when there exists a natural monopoly position somewhere in the successive stages of an industry, it is difficult to imagine true competition in other stages of this industry without an economic regulation of the monopolised segment.

4.2. Investment and dynamic competition

The necessity to reach a critical size to survive, and the endogenous quality of networks modify the usual dynamics of competition between firms. First, because the initial decisions can be critical for the success or the failure of a given service provider. When the decision to subscribe to a provider is costly, customers want to be sure of the quality of each provider before they take their decision, because their expenses are unrecoverable. Then, each competitor tries to persuade the potential users that his network will be the best one. And if enough customers can be persuaded that a given network will be the best one, it will effectively be so, because a large number decide to subscribe and they enhance the quality of the network. In these activities, competition is then likely to be fierce at the earlier stage. Firms have to spend large amounts of money in advertising and in investment to gain the credibility of clients. And to obtain their subscription, they have to propose low introductory prices. These three initial weapons (large investment, costly advertising and low initial prices) were used by JVC (Matsushita) to win the domestic-video battle against Sony and Philips in the early 80s. On the contrary, in England, none of the first franchised operators of Telepoint (Callpoint, Zonephone and Phonepoint) followed this strategy and they all missed the set-up of a mobile telephone network in London. Note that in some cases, the identity of users is more important than their number. Consequently, the network operators have an incentive to quickly attract some particular pioneers, for instance large firms for a data-exchange network.

Because of switching costs, users are locked-in and competition between the operators that survive the set-up phase is softer. Then, profits are high for survivors, and these expected high profits are a stimulus for competing at the initial stage. When connection is not needed, the sellers can still impair the mobility of customers through some premium like the frequent-flyer programs of the airlines carriers.

For the aforementioned reasons, network firms have a strong incentive to invest in infrastructure at the very beginning of their activity, trying to create a momentum in their favour. When the industry becomes mature, and each surviving firm has constructed a large installed base, competition becomes weaker and the incentives to invest and to maintain the quality of service decrease. Then, technological competition is not very intensive in mature network-industries, unless an innovation creates some dramatic change, like wireless telephone or data compression did. Without a drastic technological improvement, the entry of a challenger into a mature network industry is almost impossible. It would need a strong price-cut to compensate the club advantage of the incumbents.

The dynamic regulation of the industry is an essential determinant of the investment policy of competitors. For example, in sectors such as mobile telephone where access is restricted, the licensees adapt their investment decisions both to the present and to the expected intensity of regulation. They can have an incentive to overinvest if they expect an increase in the number of licenses for the use of the same technology. The aim of the large initial investment is to create a large installed base, hoping that the induced club externality will deter future entry. In the same way, the present regulation of prices restricts the future benefits or the future revenues from investment according to the type of regulation. Then, it directs the nature and the size of investments.

4.3. Concession contracts for network activities

Concessions to private operators are increasingly common in network industries. One reason is that for their transmission segments, these industries often require the use of public assets such as water, air or land, and large shares of the population view these as assets that can only belong to the national patrimony. The concession arrangement is then a useful compromise between purely public and purely private organisations of a network industry. The operator receives from the (public) owner a right to use for a limited time, but not a property right on, the waves, subsoil or aerial locations needed for the service provision.

The specific design of the concession system, its regulatory regime, the tariff regime or even the way the concession is awarded reveal the priorities of the government.[11]

The three basic methods for allocating a concession are administrative decisions, lotteries and auctions. An auction offers two advantages over the former alternatives: it can

[11] To raise fiscal revenue, to promote efficient static and dynamic choices, to modify the distribution of income, to increase the quality and safety of products, to protect the environment, etc.

raise revenue, which is good for the seller, but it can also be designed to identify the firms with the highest use-values for the market to be conceded, which is good for efficiency.

Clearly, auctions are not specific to the assignment of concessions. They are just a way to organize competition in certain circumstances where there are not too many potential buyers and where the seller thinks he will be better off under this kind of "collective bargaining" than under any succession of bilateral bargaining rounds. Many exchanges are performed under auctioning. For each item, two elements need to be determined by the auction: How is the winner selected? and How much does he/she have to pay?[12]

In practice, the bidding process generally starts with some type of prequalification of potential bidders based on technical and financial criteria. This process reduces the number of bidders, which is bad for competition, but it also reduces the risks of non-compliance by unreliable bidders. The next relevant institutional feature is the way in which the winner is selected. Here also, there are many options. Until recently it was common to rate various aspects of the technical and financial proposal and add up the results in a weighted or unweighted average. This process lacks the transparency and hence the efficiency that many potential bidders would like to see. So the two most common options in the award concession contracts boil down to this: (i) the winner is the highest payer for the right to provide the service or (ii) the winner is the bidder offering the lowest price to be paid by the consumers (in both cases for a set level of investment and quality requirements). The first criterium favors the fiscal objectives and aims at maximizing the revenue to the government. The payments are typically made as a lump sum at the beginning of the concession period or as an annual payment (which often boils down to a rental fee for the use of the existing infrastructure made available by the government). The second model focuses on the interest of users and ensure price minimization rather than revenue maximization.

One of the drawbacks of the auctionning procedure is the risk of collusion between candidates. And collusion may be stimulated by the design of the auctions. If the government imposes an upper limit of the shares anyone bidder can own of the deal, it is very likely that various bidders will have a strong incentive to collude. In Brazil’s railways concessions for

[12] Note that in an auction, the bids are not necessarily the price that the winner will pay to the seller. It can be any characteristics of the exchange as far as it can be measured and verified: quantity to produce, sale price, quality of service etc. Argentina's 1993 cellular license auction illustrates the variety of public policy purposes to which auctions can be dedicated. Competition was not over price but over which bidder could offer to set up cellular telephone service in the fastest time. A consortium including GTE and AT&T won by promising to provide cellular service across a vast area of Argentina's countryside in only one month.

instance, no single operator could own more than 20% of the shares. There were never more than 6 or 7 potential candidates to the concessions so that it was clear that they had a strong incentive to create consortia together and to figure out how to share profits after winning the bid as a single candidate. Similar issues can arise when the government is trying to impose redistributive criteria on the design of the auction by requiring the participation of minority owned or locally owned firms, as part of international consortia.

The final design of an auction procedure should take account of the quantity of information the auctioneer wants to give to candidates. For a unique concession assignment, the more the candidates know on each other, the greater the chances the prize to be attributed to the highest valuation bidder. On the other hand, in a multiple-round auction, with a good knowledge of other's valuation, candidates can collude to organize alternate assignments of prizes. If there are several items to allocate, like the concessions for electricity distribution or various segments of highways, the decision to auction the concessions in sequence rather than in a large simultaneous auction depends on a delicate trade-off. On the one hand, a sequence of elementary auctions has the advantages of administrative simplicity and of instantaneous plan revision: at the end of each sale, every bidder knows exactly what everybody got and what remains to sell. But on the other hand, this process impedes aggregation of prizes by eliminating backup strategies. Also, predatory bidding can occur. These drawbacks can be avoided using large simultaneous auctions or, to keep some of the advantages of sequencing, a simultaneous auction but with multiple rounds. In this case, the problem is to define clear stopping rules: the auction procedure should have a finite duration with aggregation possibilities until the deadline. For instance, one can decide that after a certain duration only active bidders can still offer new bids and/or that the increment to announce for winning the prize is a decreasing function of the number of active bidders.

For the water concession of Buenos Aires in April 1993, the candidates had to compete by proposing a discount coefficient to multiply the computed "public tariffs".

5. A typology of network industries.

The wide variety of network industries suggests several alternative typologies depending on the objectives of the classification. For instance for technical or legal reasons, the analytic key can be the nature of the flow transmitted. Thus we could oppose energy networks (electricity, heat) to material networks (passengers, water, gas, postal services) and to information networks (telecom, TV, electronic money). Alternatively, for medical, legal or/and ethical reasons, one can privilege a distinction between the transport of persons (air, train, bus) on the one hand and all the other types of networks on the other hand.

From an economic policy point of view, the problem is to know whether competition can be used as a regulatory mechanism in some network activities or administrative regulation is necessary. A broader discussion of the justification for public intervention is presented in the next chapter. Here we only consider the efficiency argument. With this restricted approach, competition should be impeded only when it could limit the beneficial effects from vertical and/or horizontal integration or sectorial agreements.

What are the risks of efficiency loss in network industries? As compared with other industries, it results from our previous developments that the main distinctive features of a network are, first, the existence of strong externalities (positive club externalities and negative congestion externalities) and, second, their node-edge structure. The cost of installing and operating such an infrastructure, as well as the loss of club benefits, can be strong reasons to blockade the entry of any competitor.

In the next page, we present a table where we try to summarise the main characteristics of the industries analysed in the first part using the two keys mentioned above: the node-edge structure and the existence of externalities. We have classified the industries in three groups. First, we think that local networks deserve a special treatment because of the narrowness and the specificity of the markets concerned. In this first category, we find the water sector as well as the Urban Public Transport activities (upt). Other industries not studied in the first part, for example garbage collection or urban heating, belong to this category as well. The other industries, with a national or an international concern are classified according to their « one-way » or « two-way » nature. The motivation for this distinction is that the nodes, which are essential points of control in any network, are more complicated and harder to manage when they are transit points in two directions.

MAIN FEATURES	one way electricity, gas	two-way trains	 telecom, postal service, airlines	local networks water, urban public transport (upt)	 Notes
1. Lines **1.1. Infrastructure**	■ no duplication	■ no duplication	■ duplication possibilities ■ competition (postal service, airlines) imperfect competition (telecom)	■ no duplication	**duplication can be justified by** * yardstick competition * differentiation * dynamics * security
1.2. Substitutes	■ partial substitution	■ strong substitution	■ low substitutability for telecom (but Internet...) ■ strong for postal services (fax)	■ no substitute for water ■ strong substitutes for u.p.t.	pollution and congestion in urban areas.
2. Nodes	■ gas is storable at nodes ■ electricity not storable	■ stations are monopolies	■ airports are local monopolies	■ depollution units have Minimal Optimal Scale	terminal nodes and intermediary nodes can have a very different status with respect to competition
2.1. Upstream	■ competition in production for electricity ■ gas storage can be competitive	■ multimodal competition in dispatching	■ multimodal competition	■ multimodal competition for u.p.t.	
2.2. Downstream	■ distribution monopolies	■ multimodal competition	■ telecommunications and postal services are switching from monopoly to multitechnological competition	■ water : competition in treatment ■ natural monopoly for water distribution	
3. Externalities	■ Kirchhoff law ■ need for dispatching	■ congestion ■ need for co-ordination	■ club externalities in telecom, ■ congestion	■ congestion, ■ quality variations	* indirect club externalities everywhere * problem of compatibility

Electricity and gas use one-way networks where it appears that the critical element is the transport infrastructure. While the generation equipment can be duplicated on different scales and with various technologies, the duplication of electric lines would be too costly as compared with the gains from competition. The same is true in the gas industry with the storage installation on the one hand and the pipelines on the other hand. This is also true for water distribution but the problem is to be solved at the local level.

For railways, stations (nodes) and rails (edges) could hardly be duplicated at low cost. But the rolling equipment may be owned and operated by distinct agents and there exists a strong intermodal competition. In air transport, airports play the same role as stations for trains: they are obvious natural monopolies. But the lines can be opened to a large number of competitors, provided their flight programs are co-ordinated by an independent agency. In all these transport activities, the other transport modes are both substitutes and complements. The interconnection between the different modes enhance the efficiency of the whole transportation network.

Telecommunications are the sector most affected by recent technological changes. While some decades ago, it was unthinkable to connect telephone users without a network of wires and manually operated switchboards, which means a non duplicable infrastructure, nowadays the technical progress both in hardware and software allows users to chose their long distance operator. And within the next decade, they will probably be able to bypass the local loop. Therefore, the telecommunications industry is the one that can be widely opened to the regulation by market mechanisms.

Finally, from an economic policy point of view, it appears that the key problem is to know whether, in a given network, the nodes and the lines can or cannot be duplicated at a reasonable cost. And, if duplication is too costly, can nodes and lines be bypassed at a reasonable cost by potential entrants? This approach allows to identify bottlenecks, if any, and to decide when an administrative regulation is to be installed. The table in the next page presents seven network industries focusing on their ability to support competition, `monopoly,` or a *mix* of free market and planned organisation.

Network Taxonomy	**Entry Mode**	**Lines and Nodes**			**Exit Mode**
Electricity	generation	transformation	transport **(dispatching)**	transformation	local distribution
Gas	generation	storage	*pipeline* **(dispatching)**	storage	local distribution
Water	*pumping, treatment*		distribution/collect **(basin regulation)**		treatment,dumping
Postal Service	clearance	sorting	transport **(international co-ordination)**	sorting	local distribution
Telecom	*local loop*	*switch*	long distance **(allocation of audio-frequencies + co-ordination)**	*switch*	*local loop*
Airlines	intermodal	airport	transport **(traffic control)**	airport	intermodal
Railways	intermodal	station	*transport* **(traffic control)**	station	intermodal

Key : competitive activity
mixed activity
monopolistic activity

This type of classification is very sensitive to the definition of each network. For instance, an electric network is not usually restricted to the transportation and distribution of power. It is defined as including the generation nodes. On the contrary, a railway network is limited to the transportation between stations. The upstream grouping of passengers and freight as well as the downstream degrouping are realised by other modes of transportation. Also, in terms of possibility of duplication, it appears that the notion of node, without additional qualification, is not totally pertinent. Indeed, while entry nodes can be easily identified in the power industry (generation stations), they are not in the postal sector, unless we accept that each household or firm can be considered as an entry node. For this reason, it is more convenient to discuss the possibility to introduce some form of competition into networks in terms of entry and exit **Modes** rather than in terms of **Nodes**. This is how we have labelled the first and the third columns in the preceding table.

With this presentation, it clearly appears that the entry mode is able to accept high degrees of competition in most industries. On the contrary, the network activity *per se* is intrinsically a natural monopoly position in at least one of its essential parts: nodes for airlines, railways and postal service; lines for water and maybe for gas, both for electricity. The telecommunications sector is quite atypical in this set. Finally, with the exception of transport (because of intermodality) and telecommunications (by anticipation to the bypass of the local loop), the exit mode is also a natural monopoly position, that is an activity where some central regulation is needed, particularly restrictions in terms of universal service obligations.