2. Economie des réseaux

Fabrice LEQUEUX

Maître de conférences à l'Université Paris Sud XI

Université Paris I Panthéon-Sorbonne

ENSTA IP Paris



Introduction



Katz et Shapiro [1985] ont été à l'origine d'une branche particulière de l'économie industrielle : **l'économie des réseaux** (*economics of networks*).



Michael KatzUniversity of California at Berkeley



Carl ShapiroUniversity of California at Berkeley

<u>Introduction</u>



Cette branche de l'analyse économique s'est développée à partir de la fin des années 1970, au moment où une grande vague de déréglementation frappait aux Etats-Unis certaines activités organisée en réseau, notamment le transport aérien,

les télécommunications, le gaz et l'électricité.

L'une des particularités de ces activités est l'importance des **coûts fixes** (liés aux infrastructures) et l'existence d'**externalités**, susceptibles de provoquer certains dysfonctionnements du marché (antitrust).

L'économie des réseaux plonge donc ses racines dans cette problématique, et a bénéficié de nombreux développements, notamment avec les travaux de Economides [1989, 1991, 1996].

Nicholas Economides Stern School of Business, New York University



1. Qu'est-ce qu'un réseau?

Nous appelons « réseau » un ensemble d'éléments unis par des liens univoques ou multivoques de nature variée, dont une partie au moins nécessite d'être mobilisée pour la réalisation d'une production quelconque.

Les éléments qui constituent le réseau sont donc complémentaires entre eux.

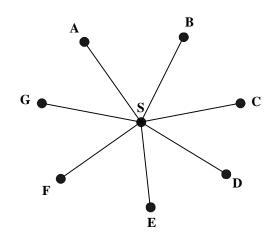
Parfois, ils peuvent également se substituer les uns aux autres dans la mesure où une production équivalente peut être obtenue à partir de différentes combinaisons des éléments du réseau.



Réseau classique « en étoile » hub & spoke network

Cette figure montre un réseau classique en étoile, comme celui du réseau téléphonique local ou du transport aérien.

Dans ce cas, les liaisons de point à point (disons de A vers B) passent par un élément central S (un commutateur ou un hub), chargé de rediriger l'appel, ou le voyageur, vers le pôle de destination.

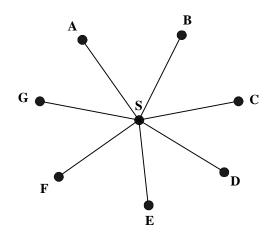




Réseau classique « en étoile » hub & spoke network

2 types de réseau se distinguent alors selon l'arrangement topologique des éléments constituant le réseau :

- Lorsque la liaison AB est distincte de la liaison BA, le réseau est appelé « réseau à double voie » (two-way network) ou « réseau de communication ». Dans ce cas, la communication entre le point A et le point B peut s'effectuer dans les deux sens. Le réseau aérien, autoroutier ou ferroviaire, ainsi que le réseau téléphonique, sont des exemples de réseaux de communication à double voie.
- Lorsque l'une des 2 liaisons n'existe pas, ou que la liaison AB est identique à la liaison BA, le réseau est appelé « réseau à simple voie » (one-way network) ou « réseau de distribution/diffusion ». Dans ce type de réseau, la communication de point à point s'effectue en sens unique. Le réseau électrique, de télévision hertzienne ou encore de radio sont des exemples de réseaux de diffusion à simple voie.





Les biens-systèmes Katz et Shapiro [1994]

Plus largement, un réseau à simple voie peut aussi être constitué de deux types de biens (ou plus), dont l'usage nécessite qu'ils soient combinés entre eux.

Les consommateurs ne dissocient alors pas chaque composant, mais émettent une demande pour le produit composite.





2. Externalités de réseau

De manière générale, il y a **externalité** lorsque la production ou la consommation d'un agent économique agit sur le bienêtre ou la satisfaction d'un ou plusieurs autres agents.

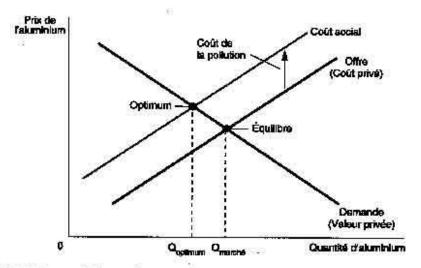


FIG. 10.2 La pollution et l'optimum social. En présence d'une externalité négative à la production, le coût social de production de l'aluminium est supérieur au coût privé. La quantité optimale d'aluminium Q_{optimum} est donc inférieure à la quantité d'équilibre du marché,



Dans le cas des **externalités de réseau**, cela signifie que la valeur d'un bien augmente (dans le cas d'externalités positives) ou diminue (dans le cas d'externalités négatives) en fonction du nombre d'unités de ce bien qui composent le réseau.





On peut alors distinguer 2 types d'externalités de réseau [Katz et Shapiro, 1985].

 Les externalités directes de réseau interviennent lorsque l'accroissement du nombre d'utilisateurs d'un produit, d'un service ou d'une technologie a un effet direct sur l'utilité que retirent les agents économiques de ce produit, service ou technologie.

Le réseau téléphonique est un exemple typique de réseau soumis à des externalités directes : plus le nombre d'abonnés au réseau est élevé, et plus l'utilité du service offert à chaque utilisateur est importante. **L'effet de club** est le phénomène par lequel l'utilité réelle d'une technique ou d'un B&S dépend du nombre de ses utilisateurs.



La raison essentielle de l'apparition des externalités de réseau réside dans la complémentarité entre les éléments qui composent le réseau.

Lorsque chaque utilisateur *détient* un composant du réseau, les externalités sont directes. Considérons **un réseau à double voie** (un réseau téléphonique, par exemple) composé de n éléments. Au sein de ce réseau, il existe donc n(n-1) communications possibles. Chaque utilisateur supplémentaire du réseau agit sur le nombre total de communications, en ajoutant 2n liaisons complémentaires. Il en résulte des externalités directes de réseau qui accroissent à la fois l'utilité du service rendu par le réseau, et la satisfaction de chaque utilisateur.



 Les externalités indirectes de réseau se produisent lorsque l'accroissement du nombre d'utilisateurs d'un produit, d'un service ou d'une technologie engendre une amélioration des caractéristiques de l'offre de ce produit ou de cette technologie, ou de l'offre de biens et services complémentaires.

Les effets de club ne sont pas nécessairement propres aux réseaux physiques et se retrouvent sur de nombreux marchés.

L'exemple-type d'externalités indirectes de réseau est l'accroissement de la qualité et de la variété des programmes logiciels destinés à une console, ou à une machine, qui bénéficie d'une large « base installée ».



Lorsque l'on a affaire à un réseau à simple voie, les externalités ne peuvent être qu'indirectes.

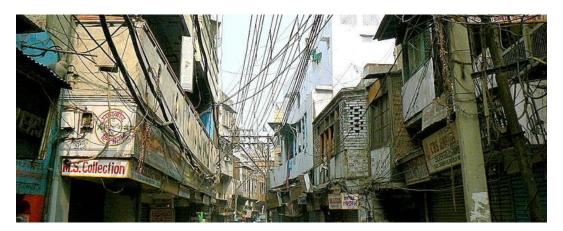
Imaginons un réseau constitué de deux biens A et B, dont l'usage nécessite qu'ils soient combinés dans des proportions équivalentes. S'il existe n variétés de bien A et m variétés de bien B, les consommateurs peuvent acquérir nm biens composites ou « biens-systèmes ».

Dans ce cas, un consommateur supplémentaire ne viendra pas accroître directement la valeur du réseau. Les externalités ne pourront être qu'indirectes dans la mesure où, en augmentant la demande de chaque variété de biens, le nouveau consommateur participera à l'accroissement de la diversité des produits offerts et éventuellement, par le jeu des économies d'échelle, à la réduction de leur prix.



Si la première spécificité de l'économie des réseaux réside dans l'existence d'externalités (**effets de réseau**), la seconde tient certainement à la structure particulière des **coûts de production**.

Il est alors nécessaire de distinguer les coûts de l'infrastructure (la « couche basse » du réseau) des coûts des services transitant par l'infrastructure (la « couche haute » du réseau).





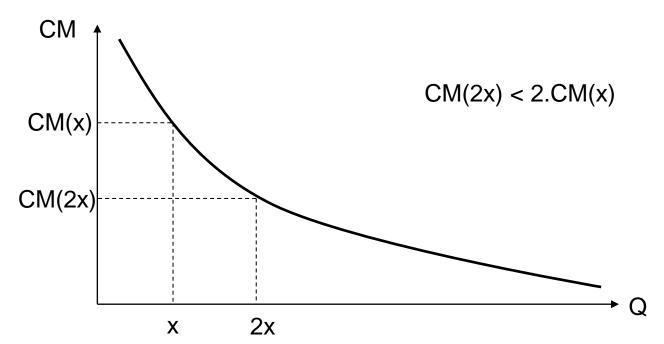
Les **coûts de l'infrastructure** dépendent directement de la « dimension du réseau », c'est-à-dire de son étendue et de sa capacité (couverture géographique et quantité de services que le réseau est techniquement capable d'acheminer simultanément).

Ces services fondés sur des réseaux exigent d'énormes investissements au départ pour fournir le service au premier client, mais une fois l'investissement réalisé, la fourniture du service à un client supplémentaire ne coûte quasiment plus rien (le coût marginal est faible).

Il en résulte que les **coûts fixes** liés à l'infrastructure sont colossaux, et les économies d'échelle sont très importantes, ce qui limite nécessairement les possibilités de concurrence (monopole naturel).



On dit qu'il y a **monopole naturel** sur un marché lorsque la fonction de coût est *sous-additive*, c'est-à-dire que pour tout niveau de production, le coût des facteurs utilisés est minimal lorsque la production est réalisée par une seule entreprise.





Un monopole naturel s'établit donc lorsque la production se caractérise par des **rendements d'échelle croissants** (toute augmentation des deux facteurs de production engendre une augmentation proportionnellement plus forte de la production).

Les rendements d'échelle croissants conduisent à une baisse du coût unitaire de production (économie d'échelle), ce qui permet la baisse des prix et l'élimination des concurrents.

Chaque entreprise cherche à attirer toujours plus de clients pour renforcer les économies d'échelle, et baisser ses prix, ce qui lui procure de nouveaux clients et de nouvelles baisses de prix, etc,



Le monopole naturel apparaît quand les capacités de production de l'opérateur dominant sont suffisantes pour alimenter tout le marché, ce qui signifie que des entreprises concurrentes sont économiquement inutiles : il faut éviter le gaspillage de ressources nécessaires à la constitution d'un réseau concurrent moins grand qui ne serait pas compétitif, et qui disparaitrait à terme.

Cependant, les **coûts variables** liés à l'exploitation du réseau sont relativement faibles, surtout lorsque le taux d'utilisation du réseau n'est pas maximal. Mais plus on s'approche de la saturation des équipements, et plus il y a de risques de pannes et d'incidents susceptibles d'occasionner des coûts supplémentaires.



Le problème est que les prix de monopoles sont généralement plus élevés que les prix concurrentiels; et que cet équilibre de monopole nuit à l'intérêt général.

En situation de CPP, l'entreprise produit tant que la Rm>Cm. Mais l'entreprise étant « price taker », sa Rm=RM=p. Nous avons donc l'équilibre p=Cm.

Mais en monopole, la Rm n'est pas fixe. Elle est fonction du prix.

Vendre plus cher = vendre moins.

Le monopole va donc égaliser sa Rm à son Cm pour déterminer son volume de production Mais pour cette quantité produite, la courbe de demande lui donnera la disposition à payer des consommateurs,.

Léon Walras (1834-1910)



Le problème est que les prix de monopoles sont généralement plus élevés que les prix concurrentiels; et que cet équilibre de monopole nuit à l'intérêt général.

En situation de CPP, l'entreprise produit tant que la Rm>Cm. Mais l'entreprise étant « price taker », sa Rm=RM=p.

Nous avons donc l'équilibre p=Cm.

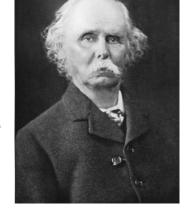
Mais en monopole, la Rm n'est pas fixe.

Elle est fonction du prix.

Vendre plus cher = vendre moins.

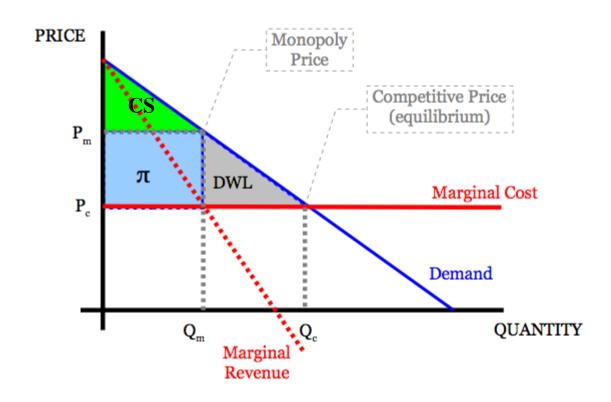
Le monopole va donc égaliser sa Rm à son Cm pour déterminer son volume de production Mais pour cette quantité produite, la courbe de

demande lui donnera la disposition à payer des consommateurs,.



Alfred Mashall (1842-1924)







A ce prix de monopole, beaucoup de clients ne peuvent pas s'offrir le service, alors même qu'ils sont intéressés. Ces personnes sont représentées par le triangle gris, le « deadweight loss » (poids mort perdu). « Perdu », car en situation concurrentielle ils auraient été servis.

En effet, ils peuvent payer un prix supérieur au coût marginal de production. C'est-à-dire qu'en vendant au prix qu'ils sont prêts à payer, le monopole continuerait à gagner de l'argent. Mais il en gagnerait un peu moins qu'avec les clients précédents.

De la même façon, certains clients sont prêts à payer plus cher que le prix de monopole : ils réalisent un « surplus de consommateur » (triangle vert).



2 situations se présentent alors :

- 1. Le prix pratiqué est identique pour tous les consommateurs et, dans ce cas, servir des clients supplémentaires (produire plus) revient à baisser les prix (et donc le profit global). L'introduction de la concurrence fait baisser ces prix.
- 2. Il faut pratiquer des prix différents en fonction de la disposition à payer des consommateurs. C'est le principe de la discrimination tarifaire.



Discrimination du 1er degré:

Le prix est fixé en fonction de la disposition à payer du client. C'est un cas d'école car en réalité, le client n'a jamais intérêt à révéler sa disposition à payer...





Discrimination du 2^{ème} degré:

Le client révèle sa disposition à payer en choisissant par une formule, un menu, une version (stratégie de *versionning* : edition Pro, Home, Premium, Gold, Light...)







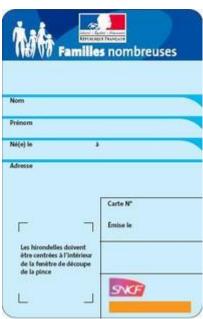


Discrimination du 3^{ème} degré:

Le client est placé astucieusement dans une catégorie en fonction de ses caractéristiques (étudiant ou -25 ans, senior, famille nombreuses...),









L'apparition de la **numérisation** dans les services liés aux activités de réseau a renforcé de degré des économies d'échelle.

En effet, les services ou les applications qui utilisent l'information (numérisée) bénéficient de **rendements d'échelle croissants**. Les coûts de développement sont très élevés mais la numérisation réduit substantiellement les coûts de reproduction qui deviennent alors quasiment nuls.

Il en résulte que le coût moyen de production de ces services informationnels offerts en réseau est décroissant avec le nombre d'utilisateurs. L'objectif des opérateurs de réseau est alors multiplier les exemplaires ou d'offrir leurs services au plus grand nombre de clients dans la limite des capacités du réseau.



Les externalités de réseau directes et indirectes ont largement été mobilisées afin d'expliquer comment un service ou une technologie peut se **propager au sein d'une structure réticulaire**.

Pour Arthur [1988a, 1988b, 1989], la diffusion technologique

obéit à un processus dynamique d'autorenforcement, dont le moteur réside dans l'action même d'adopter.



Brian Arthur
Santa Fe Institute Page 28



Le modèle de base considère deux technologies A et B, qui ne sont pas « sponsorisées » . Par ailleurs, on fait l'hypothèse qu'il existe 2 types d'agents R et S (en nombre équivalent) qui se distinguent par leur préférence naturelle, respectivement pour la technologie A et B.

A chaque période, chaque agent choisit donc entre la technologie A et la technologie B. La valeur nette de la technologie A ou B dépend :

- du nombre d'utilisateurs actuels nA et mB, à la période où il fait son choix;
- des rendements d'adoption croissants, décroissants ou constants donnés par r et s, respectivement positif, négatif ou nul;
- de la préférence naturelle des agents pour l'une ou l'autre des technologies. On a donc aR > bR et aS < bS pour signifier que les agents de type R ont une préférence pour la technologie A, tandis que les agents de type S préfèrent la technologie B.



Rendements associés à l'adoption des technologies dans le modèle de Arthur

	Technologie A	Technologie B
Agent R	aR + rnA	bR + rmB
Agent S	aS + snA	bS + smB



B. Arthur établit la notion de **rendements croissants d'adoption** (RCA). Afin de mettre en évidence le rôle des externalités de réseau dans la diffusion des technologies, le modèle d'Arthur montre que plus une technologie est adoptée, plus elle accroît son « rendement », devenant par là plus attractive pour les utilisateurs potentiels.

Ce schéma repose donc en partie sur une dynamique héréditaire (représentée par des mécanismes d'auto-renforcement), qui accorde une importance particulière aux choix effectués par les premiers utilisateurs.

C'est ce que Arthur [1988a] appelle la propriété de pathdependence . Liebowitz et Margolis [1998a] résument cette propriété en ces termes : « Where we are today is a result of what has happened in the past ». Ce sont donc les premiers moments de la compétition qui en déterminent l'issue finale.



1. La propriété de path-dependence

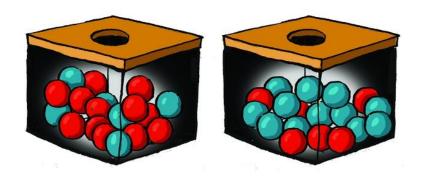
La propriété de dépendance du sentier indique que ce sont les tout premiers moments d'un évènement qui influence son issue.

Pólya et Eggenberger [1923] ont illustré ce schéma dans le célèbre exemple de l'urne . Dans une urne de capacité infinie, une boule rouge et une boule blanche ont été préalablement placées. A chaque période, on tire aléatoirement une seule boule puis, après l'avoir replacée dans l'urne, on y ajoute une boule supplémentaire de couleur identique. A chaque période t, la probabilité de tirer une boule rouge (ou blanche) est alors strictement égale à la proportion de boules rouges (blanches) qui ont été placées dans l'urne. Ce processus est donc pathdependent.



La question qui se pose alors est de savoir comment évolue la distribution des probabilités (des proportions) sur le long terme. Ces probabilités varient-elles inexorablement entre 0 et 1, ou tendent-elles vers une limite, constituant ainsi l'émergence d'une « structure » ?

Reprenant les travaux de Pólya et Eggenberger [1923], Arthur, Ermoliev et Kaniovski [1987] ont montré que, de période en période, la distribution des probabilités tendait à se stabiliser, exprimant ainsi l'apparition d'une « structure ».



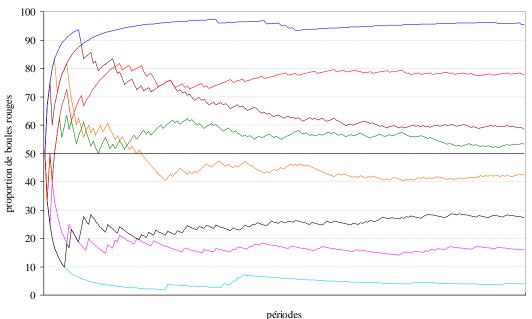


Les simulations montrent clairement ce phénomène. Pour chaque série, la proportion de boules rouges (la probabilité de tirer une boule rouge) part de 50% à la période 0 et fluctue autour <u>d'un point fixe qui constitue sa limite de long terme</u>.

Bien entendu, il existe des perturbations issues du tirage aléatoire des boules. Cependant, l'addition d'une boule supplémentaire à chaque période constitue une perturbation de plus en plus faible compte tenu de l'accroissement du nombre total de boules dans l'urne. Ainsi, les déviations s'amenuisent de période en période et la distribution des probabilités tend vers sa limite de long terme.



Cette limite varie aléatoirement à chaque lancement du processus de Pólya-Eggenberger, en fonction des premiers mouvements aléatoires. Les proportions de boules rouges et blanches se stabilisent, et la structure qui en résulte est donc parfaitement aléatoire.



Page 35



2. La propriété d'inflexibilité (lock-in)

De manière générale, les petits événements, qui interviennent dans les processus path-dependent, ont <u>des conséquences</u> <u>cumulatives</u> à l'origine de fortes inerties au retour.

L'inflexibilité (lock-in) constitue alors une propriété fondamentale du modèle de Arthur [1988a, 1988b, 1989].

Une situation de lock-in (verrouillage technologique) émerge au moment où l'importance de la base installée de l'une des deux technologies l'emporte sur les préférences initiales des agents, qui sont amenés à renverser leur choix.



En d'autres termes, l'écart qui se creuse entre les niveaux d'adoption procure une avance suffisante dans le rendement d'une des deux technologies pour que les agents l'adoptent massivement, quelles que soient leurs préférences naturelles.

Il existe donc un seuil à partir duquel le différentiel dans les niveaux d'adoption conduit une technologie à gagner la compétition.



La question qui se pose est de savoir quel est le nombre nécessaire d'adopteurs de la technologie A (comparativement à celui de la technologie B), pour que le lock-in se fixe sur cette première technologie ?

Ce seuil ressort aisément des équations correspondant aux rendements associés à l'adoption des technologies (<u>slide 29</u>). La technologie A l'emporte dès que :

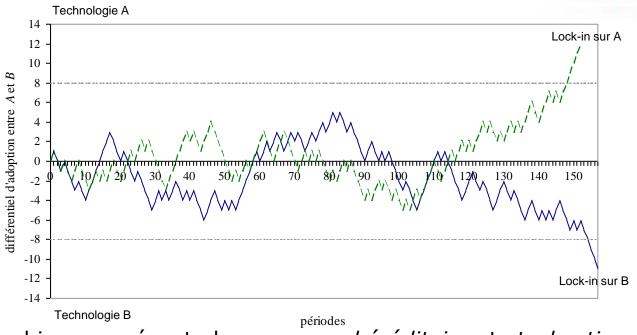
$$a_S + sn_A > b_S + sn_B$$

 $\Leftrightarrow n_A - n_B > \frac{b_S - a_S}{s}$

De manière équivalente, le seuil nécessaire (en termes de nombre relatif d'adoptions) pour que le lock-in se fixe sur la technologie B est donné par :

$$n_B - n_A > \frac{a_R - b_R}{r}$$





Ce graphique représente le processus *héréditaire* et *stochastique* qui conduit à l'émergence d'une situation de lock-in sur l'une des deux technologies. Les paramètres retenus pour refléter les préférences initiales des utilisateurs sont les suivants: aR = 3; aS = 1; bR = 1 et bS = 3. En d'autres termes, les agents de type R ont une préférence 3 fois supérieure pour la technologie A que pour la technologie B (et inversement pour les agents de type S). Les rendements d'adoption sont croissants pour les deux technologies. Ils sont fixés à r = s = 0,25.

Page 39



Dans ce cas, le seuil à partir duquel le nombre relatif d'adoptions conduit à une situation de lock-in correspond à un différentiel dans la base installée des deux technologies de huit (millions d'unités, par exemple).

Ainsi, dès que l'écart entre le nombre d'adopteurs de la technologie A (ou de la technologie B) a atteint 8, la technologie la plus avancée (en termes de base installée) gagne la compétition. Son rendement s'est suffisamment accru pour qu'elle profite d'un effet de « verrouillage technologique », qui conduit les usagers potentiels à adopter massivement cette technologie, quelles que soient leurs préférences initiales. Le modèle produit alors de fortes inerties au retour de sorte que, lorsqu'une solution technologique s'est imposée, il devient très difficile de la quitter.



3. Propriété de non prédictibilité

Une autre propriété du modèle concerne la non-prédictibilité de l'issue finale de la compétition au début du processus.

En effet, si la dynamique de ce système se caractérise par une dimension <u>fortement héréditaire</u> (mécanismes d'auto-renforcement), elle fait également jouer des éléments de <u>dynamique stochastique</u>.

Ainsi, la trajectoire d'évolution est parfaitement imprévisible et, tant que l'un des seuils de lock-in n'a pas été atteint, toute bifurcation reste encore possible. Les simulations du <u>slide 34</u> montrent distinctement les aspects chaotiques de ce système dynamique à la Pólya-Eggenberger (à la fois héréditaire et stochastique).



4. Propriété d'inefficience économique

Dans la mesure où la dynamique de ces systèmes intègre des éléments stochastiques (« petits événements inconnus »), rien ne permet d'assurer que l'équilibre qui s'établira sera économiquement efficient.

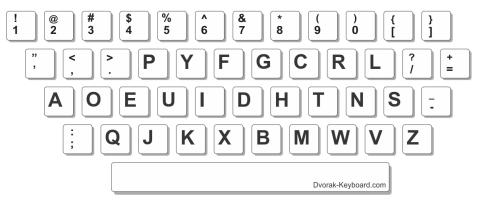
Ainsi, David [1985, 1986] a montré comment certains « accidents historiques », dont les conséquences sont amplifiées par les rendements croissants d'adoption, ont finalement conduit à l'établissement du standard de clavier Qwerty, bien que d'autres configurations (notamment le clavier Dvorak) eurent été plus efficientes.

De la même façon, la concurrence entre les formats de cassette vidéo VHS et Betamax a été abondamment utilisée pour illustrer l'apparition d'inefficiences économiques sur les marchés, dans le cadre d'une compétition entre standards.



La disposition Dvorak (ou DSK pour Dvorak Simplified Keyboard) est une disposition des touches de clavier **optimisée pour la saisie de l'anglais** et mise au point par August Dvorak et William Dealey dans les années 1930.

L'échec de la généralisation de ce clavier a été utilisé comme illustration de la théorie de la dépendance au sentier, laquelle montre comment des décisions passées de l'histoire des techniques s'enracinent et créent des effets irréversibles malgré leur manque de rationalité : la plupart des personnes ayant déjà une grande aisance avec le QWERTY refusent d'en apprendre une nouvelle (switching costs).





La disposition qwerty, et ses dérivées azerty et qwertz, sont un héritage des contraintes de conception des premières machines à écrire. Elles permettent de pallier des insuffisances mécaniques : les touches des digrammes fréquemment tapés ont été éloignées au maximum afin d'éviter que les percuteurs se bloquent lors d'une frappe rapide (et donc ce choix avait malgré tout pour effet d'accélérer la frappe sur les machines à écrire).





Dans un contexte ultra concurrentiel, il ne faut pas laisser le hasard décider de quel type d'agents R ou S apparaîtront successivement sur le marché.

Les industriels vont développer des stratégies destinées à « sponsoriser » leur technologie afin qu'elle s'impose comme standard.

SONY TOSHIBA



Dans un conte hasard décide successiveme

Les industriels « sponsoriser standard.

SOI

TOSHIDA

pas laisser le oparaîtront

destinées à mpose comme





Les tout premiers moments de la compétition technologique sont décisifs car les premiers adopteurs (technophiles) améliorent le rendement de la technologie par **effet de réseau direct**, et entraînent la localisation du progrès technique sur elle (**effet de réseau indirect**).

Il est donc décisif de :

 proposer d'emblée une vaste gamme de biens complémentaires (stratégie de JVC pour imposer le standard VHS)

accroitre rapidement le nombre d'adopteurs (externalités directes), quitte à baisser rapidement le prix du bien principal.

IV. La guerre des 2 mondes



- Les marchés de réseaux sont habituellement sélectionnistes et les externalités directes et indirectes, les rendements croissants et le processus cumulatifs aboutissent à la sélection d'un seul et unique réseau.
- Mais cela n'a pourtant jamais été le cas dans la guerre concurrentielle et technologique entre Apple et Microsoft.
- Pourquoi ?
- « Internet, Nouvelle Guerres », https://youtu.be/Z-304SCbH80
- « Apple et la tyranie du cool »
- « Steve Job vs Bill Gates, le hippie et le geek »