Interconnection Discrimination: A Two-Sided Markets Perspective

P. Faratin†, T. Wilkening††

† Computer Science and AI Lab, †† Department of Economics, M.I.T

Abstract

We first motivate a methodological approach to network economics research using tools from Industrial Organization (IO). We then demonstrate the utility of such models in light of an increasingly important problem of interconnection failures that includes network neutrality problem. In particular, we focus on a subset of the problem by focusing on direct interconnection failures when ISPs price discriminate between content providers and content users. We informally show how viewing the problem as a Two-Sided Market can lead to a discrimination rule which highlights the structural/distributional nature of the interconnection problem. Structural consideration has not only policy and antitrust consequences but can also be highly informative to engineering choices.

1. Introduction

The networking research community has used game theory and mechanism design tools to explain and design against various strategic behavior of entities in forwarding, routing, congestion control, Ad Hoc networking, caching and P2P problems. The methodology of the majority of these applications has been to focus on short-term strategic behavior of individual ASs or edge nodes in a large network. However, although useful this methodology does not capture the important *market* variables that determine these short-term behaviors. AS's routing and capacity expansion decisions for example are based on long-term exogenous and endogenous factors such as (threat of) competition and regulation, investment "holdups" and general business risks and uncertainty. The high level goal of this paper is to introduce the community to another branch of economics called Industrial Organization (IO) which takes the *firm* as the basic unit of economic analysis, whose strategic actions can have large and sometimes undesirable (anti-competitive) impact on the network on both the short and long term.

We believe IO is the right tool to consider for networking research for both methodological and design reasons. Methodologically IO models take the firm as the basic unit of analysis and typically include primitives such as the structure of the market (monopoly, duopoly, oligopoly), endogenous and exogenous production and cost structures, nature of demand (elasticities), nature of the goods (complements, substitutes, storability) and the actions available to the agents (choice of mechanism designer, price versus quantity selection, regulatory environment). These endogenous variables have first-order effects on the behavior of an ISP. Current network economics models seriously run the risk of "throwing away the baby out with the bath water" by abstracting away these important variables. Li et.al [Li04] demonstrated an equivalent point that statistical (power-law distribution) models of a network topology do not in fact have much explanatory power because they often lack finer micro-details of engineering constraints that better explain and predict network structure. In a similar manner the micro-details of the market critically matter in forming incentives and network economics models must address these details in the design phase and resist assuming separability of economic and technical design by "over the wall" design.

The second rationale for adopting IO models is constructive. IO models may provide us with information that is useful in thinking about operational as well as architecture and protocol design choices. Internet researchers often make certain structural and behavioral assumptions when designing for the Internet. AS level source routing for example assumes multiple layers of hierarchy. Yet, the Internet is actually becoming increasingly flat as access networks grow and interconnect with backbones through a single hop transit. Why is the Internet becoming flatter? Could a theory explain and predict such structural changes as flattening of the hierarchical peering/transit relationships? IO models of other industries shows why firms may at times have very strong economic incentives to vertically integrate with upstream providers and scale so as to recover high fixed costs, lower variable transit costs and reduce hazards in investments [TIR88]. Additionally, could a theory also explain the long-run *behavioral* rationales of ISPs such as their discrimination strategies (through economic instruments such as vertical integration or foreclosure, exclusive dealing and bundling using pricing and technology)? Such a theory can be immensely valuable for evaluating not only current but potential future (v. GENI and FIND) operational and engineering choices. IO models have indeed been used for analysis by practitioners in policy, legal and anti-trust cases to frame problems such as structural separation of transport from information services, predatory pricing and anti-competitive behaviors in not just Internet but many other industries. Can IO models also serve to inform operators and engineers who operate and design for the Internet and who maybe able to intelligently *shape* the outcomes through design of protocols that change the rules and strategy of the game that can be played by ISPs? Note, such a design-evaluate constructive methodology is a departure from the specify-design methodology of majority of current network economics models (viz. mechanism design). In the latter case engineering constraints are often added *after* the economic properties (incentive compatibility, efficiency, individual rationality, budget balance) are satisfied. For example, a smart market mechanism has many desirable economical properties but is technologically infeasible in the current Internet architecture (every packet has to carry a bid). A design-evaluate (closed-loop) construction methodology on the other hand lets engineers *first* design for technical constraints and only then evaluates the economic effects (by considering the game and the equilibria that is induced by a technical design).

We believe these methodological and constructive goals constitute a novel and interesting long-term networking research agenda. In this paper we demonstrate the usefulness of IO models and methodology by focusing on the concrete problem of discriminatory *price* behavior of ISPs toward content providers under two different structural setups (single and multi-homed). This problem is an instance of a more general interconnection breakdown that arises due to discrimination by ISPs and is a subset of problems in the network neutrality debate. The central contribution of this paper is to show how a new body of IO literature, called Two-Sided Markets (TSM), is providing a testable causal model of discrimination by showing that discrimination may in fact have a rational basis when demand interdependencies *across* markets, multi-homing and non-linear tariffs are taken into considerations. The demonstration of full richness and applicability of these models in the Internet is beyond the scope of this paper. Therefore the goal of this paper is to only informally show how such models can address discrimination concerns and can even be useful when forming operational and engineering expectations over interconnections.

2. The Problem: Interconnection Discrimination

Failures to coordinate and implement differentiated End-to-End QoS, increased competition and lack of innovations are increasingly making the Internet transport service a commodity service. This together with decreasing marginal uptake of broadband services is lowering the margins of transport providers whose incentive to (further) discriminate is increasing. Current network neutrality debates are centered round a number of such discrimination behaviors ranging from non-priced based discrimination practices such as structural (when ISP vertically integrates into content and/or applications), bit and packet level discriminations. In this paper we focus on neutrality problems involving interconnection breakdowns on the *access links* when Access Providers (APs) price discriminate, a problem that is increasing in frequency, most notably when BellSouth threatened Google with higher charges in 2006. We restrict ourselves to this class of discrimination problem by focusing on two markets, content consumers and providers, which seek *direct* "on-net" interconnection through the AP transport services. ISPs, specially smaller scale ones who cannot enjoy scale economies, have a strong incentive to serve content "on-net" so as not only reduce usage-based transit costs but also increase revenues through payments by content providers. Entry by third-party content distribution overlay networks such as Akamai is a strong market signal of the cost-minimization importance of this class of interconnections (see Clark et.al. for an exposition, [CLA05]). We therefore do not consider ISP interconnections, when requested content must be served "off-net" in an end-to-end manner. The nature and extent of economic interactions that can occur between on-net and off-net interconnections is beyond the scope of this paper, and the interested reader is referred to Laffont et.al for an in-depth IO model of the economic (efficiency and welfare) role of interconnection charges on perceived costs of ISPs with a mixture of on and off net traffic [LAF03]. Specifically, we make the following simplifying assumptions. Firstly, APs compete for the consumers *before* the user commits to a single AP. But after the consumer has committed (is single homed) that AP is a monopolist, at least as long as there exists substantial switching costs. Secondly, the market share of the consumers is constant. Third, we assume traffic is "on-net" and there is no other indirect (peering) path to the consumers other than through the monopolist; the AP is a "competitive bottleneck" [ARM05]. Fourthly, because of previous assumption and because the content provider wants to be in contact with the widest population of consumers, content providers multi-home to a number of APs.

The network neutrality debate in this restricted setting often centers around tariff uniformity which states that the interconnection settlement tariffs should be identical for the best-effort class of service, because in absence of QoS, transport is an undifferentiated good and so there is no economic basis for an ISP to price transit contracts differently between, say, Google and a small content provider. In other words a network should offer uniform tariffs and not discriminate against the type/label of the customer (i.e. no third-degree price discrimination [TIR88]). Opponents of uniformity claim this is an over-simplification because what constitutes discrimination is in fact a difficult problem from an economic perspective (indeed, competitive equilibrium severely restricts the set of potential discriminations because goods sold in different times at different state of nature are in fact *different* goods). For example, is an ISP price discriminating when it offers volume discount transit contracts? It is offering the same service independent of volume. The answer is not clear and depends on what cost perspective is adopted. If aggregate traffic is below capacity then such behavior can be deemed undesirable from a usage cost perspective. However, under an opportunity cost perspective discrimination is rational, since traffic is being priced according to the use the ISP could have otherwise put the consumed bandwidth to. Equivalently, whether interconnection prices are discriminatory is equally dependent on the perspective adopted. Below we will show that adopting a perspective that content users and providers are two non-separable and interdependent markets can result in an interconnection discrimination rule that can be qualitatively very different than if a single-sided market perspective was adopted. A prescriptive result from the theory useful for construction by an engineer is that in such (two-sided) markets multi-homing should not be considered only from a resilience perspective. Multi-homing in fact has major economic impact on the prices and ultimately on connectivity of Internet.

3. Two-Sided Markets: Evidence

Classic IO and multi-product literatures show that firms have an incentive to discriminate through both prices (first, second and third degree pricing) and non-price (vertical foreclosure, bundling, exclusive contracts) instruments so as to maximize the amount of surplus they can extract from consumers [TIR88]. The goal in these discriminatory mechanisms is to *capture* (rather than create) maximum surplus ("rent seeking" behavior). Use of coupons, group discounts and bundling are some examples of commonly practiced discrimination strategies in many markets which is also prevalent in the Internet retail sector. For example, retail peak-rate pricing tariff is (second-degree) price discrimination, where the (monopolist, at least during the contract period) operator presents a menu of tariffs that support different peak-rates and customers choose a tariff, based on their private preferences (or "type"). This type of "one-sided" discrimination, where the firm contracts directly with only the consumers in one market, has been well understood and is commonly practiced in the communication as well as many other industries.

However, the central intuition of this paper is that such one-sided discrimination arguments are inherently misguided in interconnection debates because interconnection requires coordination of value-flows *across* a number of markets, not just one. Data and communication services require the coordination of multiple markets including content providers and content users and callers and receivers respectively. The key feature of these interconnection problems is that an ISP has to solve the "chicken-and-egg" problem of bringing "onboard" both content users and content providers. Furthermore, these problems are continual and dynamic since customers are poached by or switch to other providers. For instance, Akamai can be seen to be capturing content providers who were previously served by backbones and who have a high willingness to pay for high quality low latency transportation, a service that is currently unavailable in an end-to-end manner and involving transit across peering points which do not permit marginal payments for higher quality service (peering points are "money insulators", [CLA05]). The chicken-and-egg problem is in fact a prominent problem in a diverse set of "platform" industries ranging from operating systems to search engines, nightclubs, computer games console, auctions, exchanges, credit cards, money and real estate agencies, to name a few, where there is strong cross-market externality/spillover and the volume of demand in one market is highly dependent on volume of demand in the other market [ROC05,EVA03]. Informally, nightclubs need to ensure both men and women are "onboard" for the "platform" to be profitable. Operating systems can be seen as a platform where application developers (one market) develop application and consumers (the other market) pay a transaction independent price for use of the platform and associated applications. Auctions houses such as eBay or Christies, facilitate exchange between buyers and sellers through lowering transaction and information costs. Credit cards such as Visa connect buyers with merchants. Even intangible services such as money can be viewed as a platform that through mandating a standard facilitates exchange. Similarly, an ISP can be viewed as a platform in a two-sided market, interconnecting content provider and content user markets (see [ROC05] for a formal definition of a TSM).

The critical feature of all these diverse examples of platforms is indirect value flow (a.k.a. *indirect network externalities* [ECO92]); that one/both sides of the market benefit from increasing adoption and/or consumption of the other side. Advertizers are attracted more to Google than other search engine because of scale of Google's searchers. More buyers transact over Ebay than other competing auction platforms because there are more sellers on eBay, who in turn are attracted to the platform because of buyer market size. Men value nightclub venues with more women. Developers are attracted to Microsoft platform because more consumers have adopted Windows. ISPs with many "eyeballs" want to interconnect with Cogent who hosts most of the porn on the Internet. The key problem for platforms, such as ISPs, in such externalities is how to maximize profits by recognizing, managing and harnessing these cross-market value flows (or externalities) that lead to increased adoption and/or transaction volumes. Bilateral contracting and transactions between the platform and each side of the market independently of the other side (such as peak-rate pricing) can in fact lead to less than optimal profits for the platform [ROC05].

In practice pricing in two-sided markets often takes the form where one side is subsidized, possibly even below marginal cost to the platform. Such "loss leader" markets are very prevalent. For example, the loss leader on Google's search platform (as well as many other media platforms such as magazines, TV, radio, etc) is the searcher/viewer market. The platform charges advertisers instead. Real estate agencies only charge the seller market a fixed membership charge, providing services to the buyers for free. Credit cards such as Visa charge only a yearly membership charge (which is sometimes even below marginal costs, that is a gift) to the consumer and a transaction charge to the merchants in the form of merchant discount. Nightclubs do not charge women (who in fact may even be subsidized with free drinks), whereas men can be charged both entrance and usage fee. Auctions mostly charge sellers and not buyers a transaction-independent price. Interestingly most operating systems and computer games adopt the reverse strategy, charging the consumers a fixed price (through licensing) while subsidizing the developer markets below cost. Pricing discrimination is apparent in the communication and Internet sectors too. Akamai charges content providers and not access networks for its services (unlike Inktomi who instead charged ISPs and not content providers, and who finally exited the market). ISPs often practice "double billing", charging both unaffiliated content providers and content users. EU mobile operators charge callers and not receivers. Some networks that have large number of "eyeballs" on their network charge below cost transit to "popular" content providers relative to comparable content providers who have equal content volumes but fewer eyeballs.

4. Two-Sided Markets: Theory

In most industries such discriminatory pricing practices are often interpreted as predatory by lawyers and antitrust authorities, because the incumbent charges below cost to capture the market and force exit of competitors, after which can behave as a monopolist and extract more of consumer surplus. Areeda and Turner, two lawyers, laid the foundations of much of current thinking about predatory pricing used by anti-trust authorities. But Areeda and Turner's discrimination rule (if prices are below marginal cost then the firm is behaving anti-competitively), and even market definition, is erroneous if the chicken-and-egg and cross-externalities problem facing the firm is admitted into considerations. The explanatory power of TSM theory is that it not only describes the common structure of divergent industries but more importantly defines a relevant and rational basis for platform discrimination between markets. Recall that price discrimination is a concern in network neutrality because of increased likelihood of interconnection breakdowns. But in a TSM one side is subsidized, possibly even below marginal cost to the platform, to induce growth in that market, with the expectation that the growth in one-market due to lower prices will, through a positive feedback loop, induce positive growth, increased prices and supra-profits in the other market [EVA03,PAR05,ROC05]. In fact, even a monopolist will have an incentive to cross-subsidize markets. Therefore in TSMs the relevant measure is not the level of prices (as in traditional methodology) but rather the *structure* of prices. Distributional considerations, usually concern of anti-trust authorities, were also central to structural reforms of the Telecommunication regulation, where rate of return regulations allowed the incumbent operators to adjust prices on different lines of business according to the elasticities of demand for each product, so as to recover fixed costs of network investment and universal services [NUE05]. Both TSMs and line of business regulation therefore are concerned about structural problems, resulting in the famous Ramsey program [LAF02]. However, the difference is that the concern in telecommunication domain was over pricing structure of multiple products (line of business) whereas in TSM the concern extends *across* multiple markets that have cross externalities. It is in this sense that theory of TSMs unifies multi-product (first formalized under the work of W. Baumol) and network externalities literatures. The exact nature and magnitude of such cross-subsidies is generally dependent on not only the sensitivity of demand in each market to prices (price elasticity of demand) but also: i) the degree of cross-market elasticities, ii) extent of multi-homing and iii) the degree of membership and usage externalities [ARM05,ROC05]. Below we will briefly cover the first two dependencies.

Figure 1 shows the geometry of the pricing problem of a monopolist ISP who intermediates two (single-homed) markets i and j. Let the i and j markets be the content user and producer markets facing usage prices p_i and p_j from the ISP respectively. The problem of the ISP is to determine the structure of profit maximizing prices. Let q_k denote the total consumption of network transport services in market $k \in \{i,j\}$. One potential additive demand function is $q_i = D_i(P_i) + e_{ji}D_j(P_j)$, where $D_i(P_i)$ is the "native" demand in the i market at price P_i, and the additive term is the effect of the consumption in the other market (see [PAR05] for nonlinear demand). The constant $e_{ji} = \partial q_i / \partial q_j$, measures the marginal change in consumption in market i with a marginal increase of consumption in the j market and represents the externality/spill-over effect market j consumption has on market i demand. Figure 1, shows the *benchmark* pricing structure case where there is no cross-market effects, $e_{ji} = e_{ij} = 0$. The solid lines represent the pricing reaction curves $p_i(p_j)$ and $p_j(p_i)$, representing the optimal prices in the i market given prices in the j market and, conversely, the optimal prices in the j market given prices in the i market respectively. Specifically, $p_i(p_j)$ is computed as the solution to the following maximization problem with p_j fixed:

$$p_i^*(p_j) = \arg\max_{p_i} (p_i q_i) + (p_j q_j).$$

When there is no cross-market effects, $e_{ji} = e_{ij} = 0$, the equilibrium set of prices in each market lies on the 45^o line, corresponding to classic monopoly prices. That is, when markets are independent then both markets are priced positively (the level of which is captured by the Lerner index and regulated by the degree of elasticity of demand in each market, [TIR88]).

As mentioned above the majority of ISPs charge positive prices for both content requestors and servers ("double billing", quadrant I). However, as also mentioned above, there is considerable heterogeneity in pricing in the Internet where ISPs also charge considerably less (even below cost) to content providers whose content is much in demand. Figure 2 shows how the pricing structure can diverge from the benchmark independent markets when the cross-market effect from market i (content users/eyeballs) is constant, and cross-market effect from market j to market i increases [PAR05]. That is, as e_{ji} increases (the demand of content users increases as demand for transport by content provider increases) then prices to the content providers decrease, to the point (e_{ji} = 11/10) that the content providers may in fact be subsidized by the platform (because users value content).

The above reasoning is a plausible model of observed pricing structure between "valuable" content providers and users. However, pricing structure may also be skewed and perceived to be discriminatory when one or both sides of the market are multi-homed to competing platforms [ARM05,ROC05, HER06]. Multi-homing is typically viewed from an operational and engineering perspective as a resilience mechanism. Routing overlays can also be seen to be multi-homed across both the layer 3 and the overlay so to choose the best quality routing path to solve the triangular inequality problem. However, there are other endogenous economic reasons for multi-homing (from both the ISP and the node perspectives) other than resilience. Firstly, competition among ISPs must be accounted for in the reasoning when multi-homing. ISPs are in strong competition with one another, differentiating their services so as to increase prices and maximize "rents" that can be extracted. Callers have a choice of mobile or fixed line telephony. Cable competes with DSL along number of service dimensions including speed, customer support, available content, etc. At times firms have an incentive to be exclusive with one side of the market in order to build market share. Incumbent Real Media, for example, tied content providers to propriety Real content format for its players through exclusive contracts so as to compete with the entrant Windows media player, forcing consumers to "multi-home" to both Media and Real players. In general increasing platform differentiation forces one or both side of the market to have an economic incentive to multi-home. However, the TSM literature is beginning to show that the platform in fact would prefer unilateral multi-homing on only one side of the market. When *both* sides multi-home then the pricing structure is closer to cost, lowering potential monopoly profits. For instance, in credit card industry (which incidentally shares many interconnection features as the Internet – see [ROC05]), credit card companies court multi-homed merchants (merchants who accept a number of cards) much more aggressively (through lower merchant charges) when cardholders are also multi-homed (carry multiple cards). The reader is referred to [HER06] for an in-depth analysis when both sides of the market multi-home (a literature we believe can give insights into economics of source-routing). However, when only one side of the market is multi-homed then the pricing structure of the ISP can be significantly above cost to the multi-homed side. To see this we continue to restrict ourselves to the case where the platform is a monopolist on the content users side (who are single homed at least for the duration of

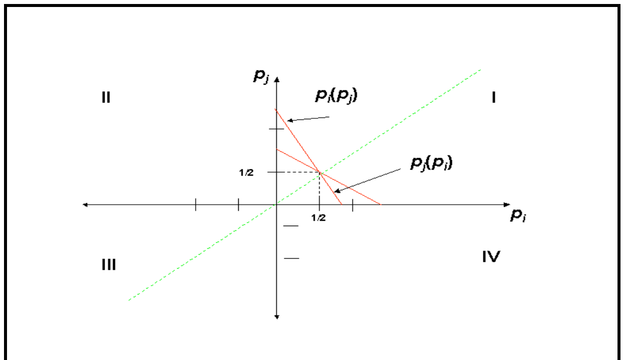

Figure 1: Geometry of the Pricing Problem

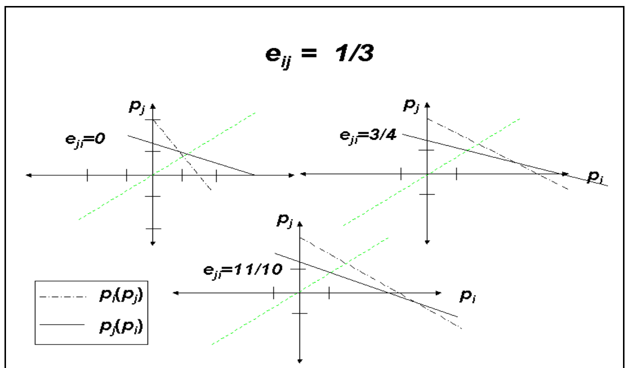

Figure 2: Pricing with Indirect Externalities

contract), but now the content providers are multi-homed, better representing the problem between Google and BellSouth. Armstrong shows the equilibrium price structure in an equivalent market setup between mobile users (content users) and fixed line users (content producers) who wish to call the mobile users [ARM05]. He shows that the "competitive bottleneck's" equilibrium price structure is to set low (in his model subscription) charges to single-homed side and high termination charges to the multi-homed side. The model also predicts that the high prices made on the higher termination charges to the multi-homed side are passed onto the single-homed subscribers in the form of subsidized services.

Multi-homing can therefore dramatically influence the pricing structure across markets. These consequences can be either positive (as in subsidies to single-homed users) or negative (as in above cost markup charges), *if we adopt a single sided perspective*. One prescription of the theory to an operator is to make sure multi or single homing decisions are not made myopically based solely on resilience criterion, but rather consider the connectivity and the cross-market benefits. Carefully consider the benefits of multi-homing expecting higher costs, if the other side is single-homed. A stronger prescription from the theory would be to engineer protocols and architect into the network that if multi-homing is a choice then it is enforced by default. Such configurations will result in lowering price distortion incentives by the ISP. However care should be taken because under a TSM view there *can* be flows of cross-subsidies whose benefits maybe on a longer time scale. Higher costs to Google, for example, today means lower prices for users which in turn may result in increase in demand for the ISP services which will in turn result in more eyeballs for Google's advertisers tomorrow.

5. Conclusion

We have presented an Industrial Organization framework to model on-net interconnection breakdowns that can result from price discrimination and showed that under some circumstances discrimination is in fact rational and both businesses *and* engineers would be prudent to consider these cross-market effects; interconnections occurs not only at layers 2 & 3, but also at the level of markets. The long-term goal of our research agenda is to use IO economic theory to describe economics of value flow in End-to-End off-net interconnection and "coopetition" behavior of ISPs, content providers and content overlay networks. The credit card network, where competing banks form a cooperative to allow customers to seamlessly transact with any merchant, is an interesting industry that shares many of 'coopetition" and "routing-money" problems as the Internet. Models developed there may in fact be useful. We believe such a long-term research agenda is methodologically closer to networking research community goals who view engineering constraints as first-order constraints, followed by other economic and social constraints. Considering, rather than satisfying, economic constraints early on in the design stage is important if innovations are to enter the market. IO, as a maturing discipline, is beginning to provide such analysis tools.

Bibliography

[ARM05] M. Armstrong (2005): *Competition in Two-Sided Markets*, forthcoming, Rand Journal of Economics, forthcoming

[CLA05] David Clark, William Lehr, P. Faratin, S. Bauer and J. Wroclawski (2005): *The Growth of Internet Overlay Networks: Implications for Architecture, Industry Structure and Policy* In the proceedings of Telecommunications Policy Research Conference (TPRC-05), Washington, DC. 2005.

[ECO92] N. Economides and Salop (1992): *Competition and Integration Among Complements, and Network Market Structure*, in Journal of Industrial Economics, vol. XL, No. 1, March 1992, pp. 105-123

[EVA03] D. Evans (2003) *The Antitrust Economics of Two-Sided Markets*. Yale Journal on Regulation: 2003, 20(2), pp: 325—381

[HER06] B. Hermalin and M. Katz (2006), *Your network or mine? The economics of routing rules*, RAND Journal of Economics, forthcoming

[LAF02] J.J. Laffont and J. Tirole (2002): *Competition in Telecommunication*, MIT Press, Cambridge, MA, US, 2002.

[LAF03] J.J Laffont, S. Marcus, P.Rey and J. Tirole (2003): *Internet Interconnection and the Off-Net Pricing Principle*, RAND Journal of Economics, (34), 2, 2003, pp. 370-390.

[LI04] L. Li, D. Alderson, W.Willinger and J. Doyle (2004): *A First-Principles Approach to Understanding the Internet's Router-Level Topology*, SIGCOMM 2004, Portland, Oregon, USA, pp. 3-15.

[NUE05] H. E. Nuechterlein and P.J. Weiser (2005) *Digital Crossroads: American Telecommunications Policy in the Internet Age*, MIT Press, Cambridge, MA, US, 2005

[PAR05] G.G.Parker and M. Van Alstyne (2005): *Two-Sided Network Effects: A Theory of Information Product Design*, Management Science, (51), 10, 2005, pp. 1494-1504.

[ROC05] J. Rochet and J. Tirole (2005): *Two sided Markets*: A Progress Report, forthcoming, Rand Journal of Economics. (http://idei.fr/vitae.php?i=51)

[TIR88] J. Tirole (1988): *Industrial Organization*, MIT Press, Cambridge, MA, US