Networks and Spatial Economics (2020) 20:833–872
https://doi.org/10.1007/s11067-020-09499-1

On the Net Neutrality Efficiency under Congestion Price Discrimination

Sahar Fekih Romdhane[1] · **Chokri Aloui**[1] · **Khaïreddine Jebsi**[1]

Published online: 28 July 2020

Abstract

In this paper, we study the extent to which net neutrality, defined as price non-discrimination, is welfare improving in comparison to non-net-neutrality. We consider a two-sided congested internet service provider (ISP) that acts as a monopoly platform. The congestion is basically caused by the overuse of the fixed ISP's bandwidth by content providers. Unlike end-users, we allow content providers to be heterogeneous in their sensitivity to congestion. The analysis reveals that the ISP monopolist, by departing from the net neutrality regime, price-favors the most congestion sensitive providers. We argue that these providers play a crucial role in creating traffic and generating profit for the ISP platform. In our paper, whether net neutrality improves or harms social welfare depends on a critical threshold of the platform equilibrium congestion level. This threshold is an indicator or a proxy that indicates for a planner whether or not net neutrality rules should be repealed. When the platform congestion level lies below the threshold, we show that non-net-neutrality makes the society better-off. Exceeding the threshold, two effects are identified: profit-increase effect and consumers' surplus-reduction effect. If the latter outweighs the former, net neutrality increases social welfare when compared to non-net-neutrality.

Keywords Net neutrality · Congestion · Third-degree price discrimination · Two-sided markets

JEL Classification D42 · L12 · D62 · C63

✉ Sahar Fekih Romdhane
fkihromdhane.sahar@gmail.com

Chokri Aloui
alouichokry2004@yahoo.fr

Khaïreddine Jebsi
khairy.jebsi@topnet.tn

[1] Faculty of Economics and Management Science and Laboratoire de Management Stratégique de l'Innovation et Intégration Régionale (LMASIIR), Sousse University, Sousse, Tunisia

1 Introduction

Under new leadership, the Federal Communications Commission (FCC) officially repealed in June 2018 the net neutrality rules enacted during the Obama era.[1] The new proposal will grant Internet Service Providers (ISPs) the power to offer fast and slow lanes, to block some websites and to manage the Internet traffic in their best interests. This decision has been in full debate during the last few months. The “Open Internet” supporters are fighting to push the House of Representatives to overrule the FCC’s decision. The Democrats pledge to restore new strong net neutrality rules. Our paper adds to this recent strident debate. Its main purpose is to study the extent to which net neutrality, essentially considered as price non-discrimination, makes the connecting Internet parties (ISP, content providers, and end-users) better-off when congestion matters. Particularly, we seek to sort out a congestion-related indicator (or a proxy) that indicates for the regulating authority when it is beneficial to enact or abolish the net neutrality rules.

Over the last two decades, academicians, practitioners and politicians have extensively debated the net neutrality rules. These rules prohibit ISPs from discriminating between content providers. In other words, all contents should be treated equally regardless of their sources or destinations. On the one hand, net neutrality proponents insist that it is indispensable for the increase in the number of innovating Internet startups. It also prevents the ISPs from censoring contents. On the other hand, opponents argue that net neutrality reduces the existing ISPs’ potential to upgrade the broadband capacity. This is due to the fact that recovering the investment cost will not be readily ensured. In addition, the overuse of the limited transmission capacity will bring about congestion; and therefore the Internet service quality will be continuously degraded. In spite of this stuck debate, the FCC’s position, during the Obama era, was to keep the Internet neutral. It therefore established the main aspects of net neutrality that are: no blocking, no throttling and no paid prioritization.[2]

To address our theoretical paper’s issue, we consider an ISP that acts as a two-sided monopoly platform.[3] It provides a limited amount of a last-mile bandwidth to distinct but interdependent customer groups (sides): end-users and content providers (web companies). In addition to the fact that both sides extract values from each other’s participation, they are also assumed to suffer from last-mile congestion. This latter, interpreted as a traffic-related delay, is assumed to be basically caused by the content providers’ overuse of the ISP’s fixed transmission capacity. Indeed, an extra web company joining the ISP platform imposes a congestion cost not only on the other web companies but also on the end-users.[4] The congestion cost that each side bears takes the form of a delay cost as each customer group must endure more time to receive the ISP’s service.

[1] For more details, see https://apps.fcc.gov/edocs_public/attachmatch/DOC-347927A1.pdf.

[2] See, https://blogs.lse.ac.uk/medialse/2011/03/30/net-neutrality-the-first-amendment-of-the-internet/.

[3] On two-sided markets, see among others Aloui and Jebsi (2016), Rochet and Tirole (2003, 2006), Armstrong (2006), Rysman (2009), Rosen (2005) and Weyl (2010).

[4] For simplification reasons, we ignore the fact that congestion can also be caused by the end-users. In accordance with Economides and Tåg (2012), we abstract from the network intermediaries that can lie between the ISP and either customer groups or both. For more details about these intermediaries, see Economides (2005, 2007). On the Internet traffic managements, see Smirnov and Crowcroft (2003).

We compare- in terms of consumer surplus, platform's profit and social welfare- net neutrality with non-net-neutrality. In this paper, we presume that content providers value time heterogeneously insomuch as some of them are more sensitive to the delay than others. As for end-users, these are assumed to be homogenous with respect to their sensitivity to congestion. We compare two pricing scenarios. The first is the uniform pricing scenario (see Peitz and Schuett (2016) and Bourreau et al. (2015) among others).[5] This price non-discrimination corresponds, in our paper, with net neutrality. The second scenario stands for the non-net-neutrality regime and therefore the ISP monopoly is allowed to third-degree price discriminate between content providers depending on their willingness to pay to avoid congestion. This heterogeneity feature is observed in practice allowing the ISP to divide the content providers into subgroups. The giant and leading web companies, such as Google and Facebook, deliver larger amounts of data (or packets) to the numerous end-users that they interact with. Accordingly, these heavy providers have a tendency to be willing to pay more to obtain extra or larger lanes across the bandwidth. In contrast, light content providers such as upstarts and small e-commerce businesses are typically less willing to pay to avoid congestion.[6] In fact, the scale of the data exchanged with end-users can be regarded as a signal that helps the ISP to distinguish between congestion sensitive and congestion insensitive content providers. The magnitude of the sensitivity to congestion of one content provider can therefore be captured through the traffic it terminates. To the best of our knowledge, this paper is the first to deal with (intra-group) third-degree price discrimination in a two-sided monopoly setting.[7] As for end-users, price discrimination is not applied on their side.

Under each pricing scenario, we solve a two-stage program. In the first stage, the ISP monopolist sets the optimal price on each side. In the second, consumers (end-users and content providers) decide whether to join the ISP platform or not. This program is solved backwardly.

1.1 Related Literature

To find out the socially appropriate regime that the ISPs should adopt, the economics literature on net neutrality is quickly growing.[8] We particularly focus on the closest theoretical frameworks to our paper's issue. Economides and Tåg (2012) relate net neutrality to the fact that a zero-price is enforced. These authors study the net neutrality efficiency by considering a two-sided ISP platform without taking into consideration

[5] Peitz and Schuett (2016) argue that the uniform pricing is typically possible due to the two-sidedness feature of the Internet market. Bourreau et al. (2015) examine, in a supplementary annex, the case where competing ISPs charge content providers for terminating their traffic. Economides and Tåg (2012) define non-net-neutrality as charging content providers non-zero uniform prices.

[6] In this setting, Krämer and Wiewiorra (2012) claim, on page 1303, that: "*A CP (content provider) that offers an Internet telephony service, for example, is certainly more sensitive to network congestion than a simple email service provider*". Moreover, Peitz and Schuett (2016) note, on page 17, that "*Time-sensitive content includes voice and video telephony, online games, real-time video streaming, and certain cloud services; less time-sensitive content includes email, web browsing, and file sharing, where modest delays in transmission do not matter much*".

[7] Unlike us, Economides and Hermalin (2012) define third-degree time discrimination in the sense that the content providers are charged the same access price but face different delay times.

[8] For a literature review on net neutrality, see Krämer et al. (2013).

the congestion effect. Regarding the monopoly market structure, they prove that imposing net neutrality can be socially desirable. Choi and Kim (2010) analyze the impact of net neutrality, defined as a no-paid prioritization aspect, on the investment incentives of a two-sided congested ISP monopoly. The congestion is modeled following the standard M/M/1 queuing formulation. In their short run analysis, whether net neutrality improves or hurts social welfare depends crucially on the relative magnitudes of both the cost-quality asymmetries and the degree of product differentiation. Cheng et al. (2011) define net neutrality in accordance with the fact that a congested ISP monopolist is not allowed to favor one horizontally differentiated content provider over its competitor in terms of preferential delivery service. These authors assume that congestion, formulated according to the M/M/1 approach, is one-sided as it negatively affects end-users and not content providers. Cheng et al. (2011) demonstrate that net neutrality can improve social welfare depending on the model's parameter values. Drawing on the queuing theory, Krämer and Wiewiorra (2012) link net neutrality regulation to prohibiting a congested ISP monopolist from prioritizing the congestion sensitive content providers that are assumed to be more willing to pay for priority access. Furthermore, under net neutrality, content providers are allowed free access to the ISP's terminating service. Regarding its short-run welfare implication, they show that net neutrality can decrease welfare. Reggiani and Valletti (2016) use a two-sided market model to investigate the impact of net neutrality, defined as a non-prioritization aspect, on the ability of content providers to innovate. The congested monopoly ISP serves a large content provider and a fringe composed of small ones. These authors assume, like us, that content providers (resp. end-users) are also heterogeneous (resp. homogenous) with respect to their sensitivity to congestion. They show that whether net neutrality improves or harms social welfare depends on the advertising rate level.

In contrast with these pre-cited works, except Economides and Tåg (2012), Economides and Hermalin (2012) utilize a congestion micro-foundation which is close to ours. They assume that a one-sided ISP monopoly platform divides a bandwidth into a number of sub-bandwidths. Each fixed bandwidth portion is assigned to one class of content providers. The congestion level within each class is measured by the total demand of packets relative to the dedicated fixed sub-bandwidth. As in our model, the congestion that end-users suffer from is correlated to the congestion-in-transmission provoked by the content providers' class they are connected with. Net neutrality in the Economides and Hermalin (2012) model stands for preventing the ISP monopolist from discriminating, in terms of transmission quality, between content providers. These authors argue that net neutrality can be welfare improving depending on the magnitude of the content demand elasticity with respect to the end-users' sensitivity to congestion. Recently, Peitz and Schuett (2016) have examined the efficiency of five regulatory regimes when traffic inflation matters. This means that some congestion control and compression techniques can play a role in inflating the aggregate traffic. They assume, like us, that there are two types of content (providers): time-sensitive and time-insensitive. The congestion they define is related to the probability of reaching an end-user. This way of modeling congestion is somewhat similar to ours in that the congestion function we postulate is simply the inverse of the probability-of-reaching while assuming that each content provider sends one packet once. Among the five regulatory regimes, the second fits with our definition of net neutrality. In this specific context, Peitz and Schuett (2016) prove that net neutrality can be welfare improving.

The frameworks cited above are mainly interested in dealing with the case of an ISP in a monopoly position. Considering competing horizontally differentiated and congested ISPs, Bourreau et al. (2015) utilize a queuing model to examine the effect of net neutrality on both the edge and core innovation.[9] They assume that content providers are heterogeneous in their sensitivity to congestion and that end-users are homogenous in their connectivity speed preference. They suppose that an increase in the traffic running between content providers and end-users through each platform results in congestion. Among other findings, Bourreau et al. prove that the switch from net neutrality is welfare reducing.

The scope of our paper is to find out a congestion-related proxy that indicates for an Internet regulatory maker when net-neutrality, defined as price non-discrimination, is socially desirable to be enacted. This issue has not previously been touched upon in the net neutrality literature, so we attempt to fill this gap.

1.2 Main Results

Our paper derives some interesting results whose interpretations are relatively new in comparison to the existing literature on net neutrality. We show that the ISP's optimal pricing strategy under non-net-neutrality incarnates the fact that the most congestion sensitive content providers represent a source of profit for their platform. In this regard, the ISP favors them by charging lower price on their side. By doing so, the monopolist re-congests the platform by expanding the demand on the content providers' side. This demand expansion contributes, through cross-externalities,[10] to expand the demand on the end-users' side. This outcome occurs even if the monopolist can either increase or decrease the end-users' access price. We relate this to the fact that the additional most congestion sensitive content providers, without net neutrality, are responsible for the emergence of two countervailing effects: a network-benefit effect and a congestion effect. In the case where the former (resp. latter) outweighs the latter (resp. former), the ISP monopolist decreases (resp. increases) the end-users' price. Thus, by departing from net-neutrality, the ISP platform adopts the standard Caillaud and Jullien (2003) divide-and-conquer pricing strategy.[11] In practice, the privileged web companies like Google, Facebook, Instagram and the web-video services are powerfully attractive for web-surfers. Indeed, what our paper recommends is that any ISP connecting with these giant providers should be allowed lower access fees due to their effective role in its profitability. In accordance with Bourreau et al. (2015), we further argue that the transition from net neutrality to non-net-neutrality is accompanied by a pricing policy that induces extra innovation in services and applications by the big bandwidth *hogs*. However, such a transition results in a reduction in the small upstarts' innovation.

Compared with net neutrality, we show that the non-net-neutrality regime can improve or reduce the surplus of some content providers. This depends crucially on the magnitude of the platform equilibrium congestion level under both pricing

[9] The edge of the Internet network is the part where the applications and the content layers are concentrated. The physical and logical layers form the core of the Internet network.

[10] In a two-sided market context cross-externality or indirect network externality means that the welfare of one customer group (side) increases or decreases in the demand of the other.

[11] This strategy teaches that the side that needs more the other side should pay a higher price while the needed-more should be subsidized. Rochet and Tirole (2003, 2006) term this as '*skewed pricing*'.

scenarios. We prove that there exists a critical threshold of this level, regarded as an indicator or a proxy, through which any Internet regulation authority can decide whether to approve or to discard the net neutrality rules. This indicator depends on two main factors: the participating content providers and the network capacity. Above the threshold, we prove that the deviation from net neutrality deteriorates the surplus of both end-users and content providers. The analysis also reveals that the ISP monopoly is unambiguously better off under non-net-neutrality in comparison to net neutrality. This result is reminiscent of the traditional third-degree price discrimination rule.[12] Importantly, the increase in the ISP's profit is caused, in our case, not only by the demand expansion on the content providers' side where price discrimination matters but also by that on the end-users'. Due to the demand interdependency, we notice that the demand reallocation on the content providers' side, as a result of price discrimination, reallocates the demand on the end-users' side as well. This double reallocation contributes to increasing the ISP's profit when net neutrality is abolished. When the platform equilibrium congestion level under each regime is sufficiently high (i.e. exceeds the aforementioned threshold or indicator), we prove that net neutrality can be preferable to non-net-neutrality in terms of social welfare. In this regard, there are two effects that have to be distinguished: a *profit-increase* effect and a *surplus-reduction* effect. In the case where the latter outweighs the former, we show that net neutrality is welfare improving. When the equilibrium congestion level under each regime is relatively low (i.e. lies below the threshold), we demonstrate that it is undesirable for a regulatory agency to intervene to enforce the net-neutrality rules. The reason behind this argument is that all the concerned parties are better off under non-net-neutrality. In this circumstance, it is preferable to leave the market act freely.

The remainder of the paper is organized as follows. Section 2 defines the model. Section 3 studies the ISP net neutrality equilibrium. Section 4 investigates the non-net-neutrality equilibrium. Section 5 compares the two pricing scenarios. Section 6 concludes.

2 The Model

We consider a nonintegrated ISP that acts as a monopoly[13] access platform and enables the interactions between two distinct but interdependent customer groups (sides): legal content providers (superscripted by C) and Internet end-users or households (hereafter simply called end-users and superscripted by E). Content providers are assumed to capture profits from selling "clicks" to advertisers. The ads volume is assumed to be increasing in the number of end-users.[14] The network infrastructure or facility (bandwidth, say) that the ISP offers to enable the network traffic between both sides is prone to congestion. Figure 1 gives a simple schematic representation for our paper's model.

[12] Tirole (1988) argues, on page 137, that: "*The monopolist is better off under price discrimination, because "at worst" he can always charge the uniform price in each market*".

[13] This is reasonable for some regions where the ISPs monopolize the Internet service provision. For justifications, see Economides and Tåg (2012), Krämer and Wiewiorra (2012) and Reggiani and Valletti (2016).

[14] We assume, as in Economides and Tåg (2012), that these profits take the form of network benefits. On network externality, see for example Lambertini and Orsini (2010).

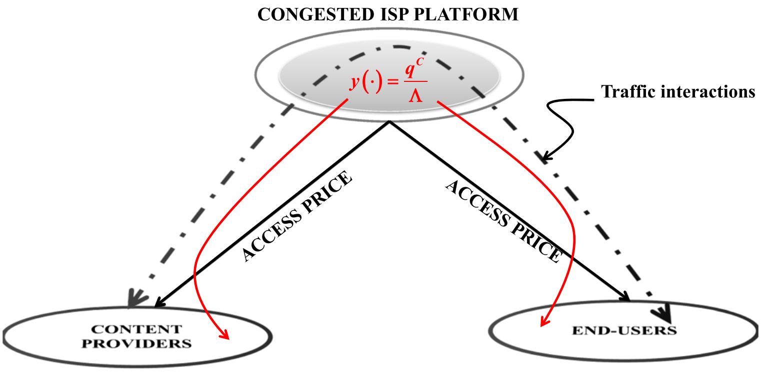

Fig. 1 Model's schematic representation

It shows the congested ISP which acts as a gatekeeper that facilitates the traffic-running between content providers and end-users. The traffic interactions must pass through a fixed capacity bandwidth. Figure 1 also depicts the fact that both sides are charged only for their access and not for their usage.

2.1 Defining the Platform Congestion Function

The congestion our paper deals with is interpreted as a traffic-related delay. It occurs due to the overuse of the ISP's limited network capacity which is denoted by $\Lambda > 0$. In fact, the tremendous traffic that runs through the last-mile bandwidth is essentially caused by the increasing number of content providers. These providers join the ISP platform to interact with end-users through running, storing and transmitting their data. Thus, it is them who are fundamentally responsible for the negative congestion effect that any ISP and its potential customers face. Ignoring the requests for contents incoming from end-users, the platform congestion function giving its congestion level is[15]

$$y(\cdot) = \frac{q^C}{\Lambda} \tag{1}$$

with $q^C \geq 0$ is the (rationally expected) network size of side C. According to Akiene and Kabari (2015) "*A packet is a unit of data that is transmitted across a packet-switched network*".[16] Thus, the demand of side C, in our paper, designates the total number of contents (or number of packets) transmitted through the ISP's bandwidth. As both customer groups share the same transmission capacity, each one of them incurs then a congestion cost which takes the form of a delay cost. This congestion micro-foundation teaches that an extra content provider negatively affects the quality of the traffic transmission between both sides.

[15] We assume a linear congestion level so as to simplify the analysis. Adopting a convex congestion delay function would qualitatively change the main results. The model developed here assumes a given time during which content providers simultaneously join and interact with end-users. See Reitman (1991) for the formulation of different forms of congestion functions in particular for the process sharing. Note that this modeling of congestion is usually adopted in transportation economics for fixed capacity setups (e.g., Basso and Zhang 2007).

[16] See Akiene and Kabari (2015), page 1384.

It is certainly interesting to assume that both customer groups are simultaneously responsible for the platform congestion. In this vein, one can assume that the congestion function is $y(\cdot) = \frac{q^E + q^C}{\Lambda}$ or $y(\cdot) = \frac{q^E \cdot q^C}{\Lambda}$ where q^k is the expected network size of side $k = E, C$. However, we choose to adopt the congestion formulation stated in (1) to make the model more tractable and to avoid complicating calculations.[17]

As aforementioned, the majority of the frameworks studying congestion and net neutrality adopt the standard M/M/1 queuing approach. Yet, the congestion modelling closer to ours is that defined by Economides and Hermalin (2012). The main differences with these authors are:[18]

- To model priority, which is not our paper's main issue, Economides and Hermalin (2012) assume that the ISP monopoly platform divides the total bandwidth capacity to sub-bandwidths so that there are many content providers classes. Each class is therefore characterised by its proper congestion level. However, we consider in our paper two content providers subgroups that share the same fixed bandwidth capacity. We also implicitly suppose that each content provider delivers one unit of content so that q^C fits with the potential number of units delivered across the bandwidth.
- The congestion that end-users suffer from in the Economides and Hermalin (2012) model is purely related to transmission, (i.e. how fast they reach contents). In this setting, the transmission-related congestion on the end-users' side is assumed to be dependent on the time that one content provider needs to transmit the required contents. In our paper, on the other hand, the congestion that end-users face is related to the number of active content providers (or potential number of units sent) and not to the intensity of the end-users' contents consumption. Figure 1 above illustrates the platform congestion level.

2.2 Defining Inverse Demand and Congestion Cost Functions on the End-Users' Side

The subscription fee that the end-users should pay to join the ISP platform is denoted by p^E. We suppose that end-users do not pay for accessing the content providers' websites.[19] In the same line with Bourreau et al. (2015) and Reggiani and Valletti (2016), we assume that end-users are homogenous with respect to their sensitivity to congestion. In other words, their willingness to pay to avoid congestion is supposed to be the same, common knowledge and is denoted by γ^E. The latter can be regarded as

[17] Bourreau et al. (2015) and Reggiani and Valletti (2016) suppose that the total traffic is given by $\lambda \cdot q^C \cdot q^E$ where λ designates the number of visits per end-user. In accordance with this formulation, we argue that the total traffic in our model is simply q^C.

[18] The congestion function we adopt can also be interpreted as the inverse of Peitz and Schuett (2016)'s probability-of-reaching an end-user. The number q^C fits with the total volume of traffic while assuming that each member on side C delivers the content once. At a first glance one can deduce that the higher such a probability, the lower the platform congestion level.

[19] This assumption is also adopted, for example, in Economides and Tåg (2012). It is relaxed in Economides and Hermalin (2012), among others.

the amount one end-user spends so as to install a software application allowing speeding up the Internet connection.

Given the platform congestion level, the congestion cost inflicted on side E is therefore given by $\Omega^E(\cdot) = \gamma^E y(\cdot)$. Indeed, the marginal congestion cost that the platform imposes on the participating end-users is $\theta^E = \frac{\gamma^E}{\Lambda}$ which is assumed to be in [0,1]. The congestion cost on side E is then $\Omega^E(\cdot) = \theta^E q^C$.

The inverse demand function of side E is[20]

$$p^E = a^E - q^E + \alpha^E q^C - \theta^E q^C \tag{2}$$

The parameter $\alpha^E \in [0, 1[$ denotes the marginal network benefits that side C brings to side E. It can be, for instance, considered as a search or transportation cost saved. In this circumstance, by visiting an online store, any end-user can identify the product or the service to buy without moving to the crowded city and wasting time and effort while seeking for it. Irrespective of all kind of network externalities, the parameter a^E stands for the gross-strength or the market extent of side E (Layson 1994, 1998). Taking into account the existing network externalities, we define the net market strength of side E as $A^E = a^E + (\alpha^E - \theta^E)q^C$. Indeed, these externalities strengthen or weaken the extent of market E depending on the sign of $\alpha^E - \theta^E$.

The inverse demand function stated in (2) can be re-written as

$$p^E = a^E - q^E + \phi^E q^C \tag{3}$$

where $\phi^E = \alpha^E - \theta^E$ which is assumed to be in $[-1, 1[$ and defined as the net marginal indirect network externalities that side C exerts on side E. If $\phi^E \in]0, 1[$, these externalities are interpreted as positive network benefits. They stand for congestion when $\phi^E \in [-1, 0]$. For simplifying reasons and in accordance with the literature on two-sided markets, we assume that end-users enjoy (or dislike) an extra content provider, large or small alike.

2.3 Defining Inverse Demand and Congestion Cost Functions on the Content Providers' Side

Unlike end-users, we assume that the content providers are heterogeneous with respect to their sensitivity to congestion. Indeed, there exist two subgroups on their side: subscripted 1 and 2. We suppose that subgroup 1, hereafter labeled type-1, is the most willing to pay to avoid congestion. On the other hand, subgroup 2, hereafter labeled type-2, is the least willing to pay to avoid congestion. In other words, subgroup 1 includes the most congestion sensitive content providers. We emphasize that the content providers being affiliated to subgroup 1 have a tendency to be bigger bandwidth consumers. Indeed, these providers may be willing to pay more to capture extra routes (or lines) across the bandwidth. As mentioned above, this can be warranted by the intensive use of data by the giant Internet platforms such as Google, Facebook and Twitter. These providers, known as bandwidth hogs, deliver voluminous data to end-users and therefore they typically value time more. Thus, the

[20] In the Appendix, we define the utility functions from which the inverse demand functions are derived.

volume of the data exchanged with end-users can be regarded as a signal for the ISP to distinguish the most congestion sensitive from the connecting content providers.

To maintain the paper's focus on pricing, we assume that the monopoly ISP does not discriminate between content providers in terms of capacity or service quality, (i.e. prioritization).[21] As noted above, the uniform pricing regime stands for net neutrality insofar as the ISP monopolist is prohibited from price discriminating between content providers (e.g. Alexandrov and Deb 2012) depending on their sensitivity to congestion. On the other hand, third-degree price discrimination, for fixed amount of bandwidth capacity, fits with non-net-neutrality.[22]

We presume that all content providers, whether being affiliated to subgroups 1 or 2, have identical expectations about the end-users' potential participation. Subgroup i, $i = 1, 2$, is willing to pay γ_i^C to avoid congestion. Its congestion cost is therefore given by $\Omega_i^C(\cdot) = \gamma_i^C y(\cdot)$ with $\gamma_1^C > \gamma_2^C$.[23] The parameter γ_i^C can be thought as the amount of money that a type-i content provider is willing to expend to hire engineers and designers that create the tools required for better content fluidity.[24] These expenditures are typically observed by the ISP as it manages the traffic. The marginal congestion cost on subgroup i is $\theta_i^C = \frac{\gamma_i^C}{\Lambda}$ so that $\theta_1^C > \theta_2^C > 0$. Indeed, the congestion cost of subgroup i is $\Omega_i^C(\cdot) = \theta_i^C q^C$. Throughout the paper, we assume that $\theta_1^C \in [0, 1[$.

The inverse demand function of the content providers subgroup i is

$$p_i^C = a_i^C - q_i^C + \alpha^C q^E - \theta_i^C q^C \tag{4}$$

Conditional on the fact that $q^C = q_1^C + q_2^C$, for $i, j = 1, 2$ and $i \neq j$, Eq. (4) can thus be re-written as follows:

$$p_i^C = a_i^C - \left(1 + \theta_i^C\right) q_i^C + \alpha^C q^E - \theta_i^C q_j^C \tag{5}$$

where $\left(1 + \theta_i^C\right)$ is the price-sensitivity term which involves the marginal congestion cost that subgroup-i content providers inflict on one another. The access fee that subgroup i pays to join the ISP platform is p_i^C. The access price levied on side C when the ISP does not discriminate between both subgroups is denoted $p^C = p_i^C \forall i = 1, 2$. The parameter a_i^C designates the maximum that subgroup i is willing to pay to join the platform irrespective of the existing network externalities. It is then the gross-strength of subgroup i. We assume that $a_i^C = a^C$ and let $\mu^E = \frac{a^E}{a^C}$ be the relative gross-market strength of side E. Further, we emphasize that the net market strength of subgroup i is denoted by $A_i^C = a^C - \theta_i^C q^C + \alpha^C q^E$. This equation shows that subgroup 2 (resp. 1) is stronger (resp. weaker) in the sense of Layson (1994, 1998) since

[21] For more details about prioritization, see for example Krämer and Wiewiorra (2012) and Economides and Hermalin (2012).

[22] In accordance with the net neutrality debate, Peitz and Schuett (2016) define five regulatory regimes. Relating net neutrality to applying uniform pricing on side C corresponds with regime 2.

[23] Since content providers pay an access fee, the willingness to pay to reduce congestion is a function of total delay rather than delay per packet. We are grateful to an anonymous referee for pointing this out.

[24] For more information about this topic see https://www.smashingmagazine.com/2011/11/fluidity-content-design-learning-where-wild-things/.

$$A_2^C > A_1^C \tag{6}$$

Inequality (6) is useful in the forthcoming analysis, particularly when we compare the platform's pricing schemes.

The marginal benefit that all content providers (on either subgroup) reap from the participation of end-users is measured by the parameter α^C which is assumed to be in [0,1[. This parameter captures the marginal benefit that one participating end-user brings to each content provider. The content providers are then homogenous in their marginal valuation to side E. For tractability reason, we suppose that $\phi^E + \alpha^C < 1$. Hereafter, side E (resp. C) is the needed-more side if $\alpha^C > \phi^E$ (resp. $\phi^E > \alpha^C > 0$).[25] The ISP's marginal membership cost is assumed to be equal to zero.

Before proceeding to the analysis that follows, it is worth noting that the marginal congestion costs do not depend on the nature of the additional content transmitted (large or small).

2.4 Defining Demand and Platform Profit Functions

2.4.1 Under Net Neutrality

When the ISP monopolist charges content providers a uniform price, its profit function is given by.

$$\Pi(\cdot) = p^E q^E(\cdot) + p^C q^C(\cdot) \tag{7}$$

such that

$$q^E(\cdot) = \frac{\left(\theta_1^C + \theta_2^C + 1\right)\left(a^E - p^E\right) + 2\phi^E\left(a^C - p^C\right)}{\Xi(\cdot)} \tag{8}$$

and

$$q^C(\cdot) = 2\,\frac{\alpha^C\left(a^E - p^E\right) + a^C - p^C}{\Xi(\cdot)} \tag{9}$$

where $\Xi(\cdot) = \theta_1^C + \theta_2^C + 1 - 2\phi^E\alpha^C > 0$ given the model's assumptions.

In addition, the demand function of subgroup i under the uniform pricing regime is

$$q_i^C(\cdot) = \frac{\left(1 - \theta_i^C + \theta_j^C\right)\left(\alpha^C\left(a^E - p^E\right) + a^C - p^C\right)}{\Xi(\cdot)} \tag{10}$$

[25] The needed-more side is the side that brings more marginal value to the network.

2.4.2 Under Non-net-neutrality

Using (3) and (5), we can determine the demand functions under non-net-neutrality. We get

$$q^E(\cdot) = \frac{\left(\theta_1^C + \theta_2^C + 1\right)\left(a^E - p^E\right) + \phi^E\left(2a^C - p_1^C - p_2^C\right)}{\Xi(\cdot)} \tag{11}$$

$$q_i^C(\cdot) = \frac{\left(\theta_j^C + 1 - \theta_i^C\right)\left(\alpha^C\left(a^E - p^E\right) + a^C\right) - \left(\theta_j^C + 1 - \phi^E\alpha^C\right)p_i^C + \left(\theta_i^C - \phi^E\alpha^C\right)p_j^C}{\Xi(\cdot)} \tag{12}$$

and

$$q^C(\cdot) = q_1^C(\cdot) + q_2^C(\cdot) = \frac{2\alpha^C\left(a^E - p^E\right) + 2a^C - p_1^C - p_2^C}{\Xi(\cdot)} \tag{13}$$

The ISP's profit function when it price discriminates between content providers is.

$$\Pi(\cdot) = p^E q^E(\cdot) + \sum_{i=1}^{2} p_i^C q_i^C(\cdot) \tag{14}$$

with $q^E(\cdot)$ and $q_i^C(\cdot)$ are the demand functions stated in (11) and (12).

Under price discrimination, we particularly note the following:

$$\text{For } \phi^E \in]0, 1 - \alpha^C\left[\text{ we have sign}\left(\frac{\partial q_i^C}{\partial p_j^C}\right) = \text{sign}\left(\theta_i^C - \widehat{\theta}_i^C\right) \text{ with } \widehat{\theta}_i^C = \phi^E\alpha^C \in \right] 0, 1[$$

All else being equal and for positive ϕ^E, the impact of the price charged to subgroup j on the demand of the other subgroup i, $j = 1, 2$ and $i \neq j$ depends on a critical value of θ_i^C. This value, which we denote $\widehat{\theta}_i^C$, is increasing with respect to ϕ^E and α^C. We emphasize here that there are two effects on the content providers' side: a network-benefit effect and a congestion effect.

- In the case where the marginal congestion cost of type i is low, lying below $\widehat{\theta}_i^C$, we argue that the network-benefit effect dominates the congestion effect. An increase in p_j^C will induce some type-j's content providers to leave the platform. Due to cross-externalities, some end-users will leave the platform and therefore there will be less network value for type-i content providers. As a result, some type-i providers will also choose to quit the platform accordingly. In this setting, it is important to notice that the capacity-related rivalry between both types is weak.

- In the case where the marginal congestion cost of subgroup i is high, exceeding $\widehat{\theta}_i^C$, it is easy to see that the congestion effect outweighs the network benefit effect. An increase in p_j^C will reduce the demand of subgroup j. Correspondingly, there will be extra type-i content providers that will join the platform because it is less congested. This is linked to the fact that the capacity-related rivalry on side C is tough. Indeed, the exclusion of some type-j participants, because of the increase in p_j^C, will contribute to alleviate the platform congestion level and reduce the type-i's congestion cost. Therefore, it will further incentivize more type-i members to join.

2.5 Defining the Consumer Surplus and Social Welfare Functions

In order to study the efficiency of each pricing scenario s, $s = D$, U (hereafter, the subscripts D and U stand for non-net-neutrality and net neutrality respectively), we define the consumer surplus of side E, the consumer surplus of the content providers' subgroup i, the consumer surplus of side C, the total consumer surplus and the social surplus (or social welfare). They are respectively given by:

$$CS_s^E(\cdot) = \int_0^{q^E} p^E(t)dt - p^E(\cdot)q^E = \frac{1}{2}\left(q_s^E\right)^2 \tag{15}$$

$$CS_{is}^C(\cdot) = \int_0^{q_i^C} p_i^C(t)dt - p_i^C(\cdot)q_i^C = \frac{1+\theta_i^C}{2}\left(q_{is}^C\right)^2 \tag{16}$$

$$CS_s^C(\cdot) = \sum_{i=1}^{2} CS_{is}^C = \frac{1}{2}\sum_{i=1}^{2}\left(1+\theta_i^C\right)\left(q_{is}^C\right)^2 \tag{17}$$

$$CS_s(\cdot) = CS_s^E(\cdot) + CS_s^C(\cdot) \tag{18}$$

and

$$W_s(\cdot) = CS_s(\cdot) + \Pi_s(\cdot) \tag{19}$$

In Section 3, we study the ISP's optimal uniform pricing policy.

3 The ISP Net Neutrality Equilibrium

We allow the platform to adopt the net neutrality rules as defined above. It is therefore prohibited from price discriminating between content providers.

Correspondingly, it charges both subgroups the same price $\left(p_i^C = p^C\right)$. In this setting, its optimization object is:

$$\max_{p^E, p^C} \Pi\left(p^E, p^C\right)$$

where $\Pi(p^E, p^C)$ is the profit function defined in (7). The following proposition summarizes the main properties of the ISP equilibrium under net neutrality. We give further analyses of these arguments in the corollaries that follow.

3.1 Proposition 1

Under net neutrality, the optimal prices of side E and side C are respectively given by:

$$p_U^E = q_U^E - \alpha^C q_U^C \tag{20}$$

and

$$p_U^C = \frac{q_U^C}{2} + \theta_1^C \frac{q_U^C}{2} + \theta_2^C \frac{q_U^C}{2} - \phi^E q_U^E \tag{21}$$

where

$$q_U^E = \frac{\left(1 + \theta_1^C + \theta_2^C\right) a^E + \left(\phi^E + \alpha^C\right) a^C}{\psi(\cdot)} \tag{22}$$

and

$$q_U^C = \frac{\left(\phi^E + \alpha^C\right) a^E + 2a^C}{\psi(\cdot)} = \frac{2}{1 - \theta_i^C + \theta_j^C} q_{iU}^C \tag{23}$$

with $\psi(\cdot) = 2\left(\theta_1^C + \theta_2^C + 1\right) - \left(\phi^E + \alpha^C\right)^2 > 0$ given the model's assumptions.

Proof See Appendix 1. ■

Under net neutrality, the optimal price on side E consists of two terms. The first term is the traditional markup which measures the monopolist's power on the end-users' market. The second term is a markdown reflecting the network benefits that side E exerts on side C.

The price charged to one subgroup on side C is composed of four components. The first is the monopolist's markup on such a subgroup. The second and third are the congestion costs that one type-i member inflicts on his-type members and those of the other type, respectively. The last term reflects the network externalities that side C exerts on side E. For $\phi^E > 0$, side E reaps marginal network benefits from side C. Conversely, $\phi^E \leq 0$ is the marginal congestion cost the platform inflicts on end-users.

The next two corollaries analyze the ISP's equilibrium under the net neutrality regime. Corollary 1 examines the condition under which the monopolist voluntarily charges content providers zero access price. In corollary 2, we compare the consumer surplus of the two content providers' subgroups.

3.2 Corollary 1

For $\phi^E \in]\alpha^C, 1-\alpha^C[$ and $\alpha^C \in \left[0, \frac{1}{2}\right[$, it is optimal for the monopolist ISP to charge content providers zero access price, (i.e. $p_U^C = 0$) if

$$\phi^E q_U^E = \frac{1+\theta_1^C+\theta_2^C}{2} q_U^C$$

Under this rule, end-users are charged the price p_z^E which is given by

$$p_z^E = \frac{1+\theta_1^C+\theta_2^C - 2\phi^E\alpha^C}{2\phi^E} q_U^C > 0$$

Proof It is easily obtained using Eqs. (20) and (21). ■

When the potential network benefits that side E reaps from side C, $\phi^E q_U^E$, reach a certain critical value $\frac{1+\theta_1^C+\theta_2^C}{2} q_U^C$, it is profit-maximizing for the ISP monopolist to voluntarily allow content providers to join her platform for free. This is accompanied by charging end-users a positive price which we denote p_z^E. It could be important to argue here that some regulating agencies in some countries might not be urged to intervene and enforce the zero-price net neutrality rule. This holds when side C is the needed-more side (i.e. $\phi^E > \alpha^C > 0$). In this context, it seems logical that allowing content providers free access will attract potentially higher demand on their side. Such a higher demand will boost the participation of end-users from which the ISP can extract compensating profits. This can be warranted by the fact that p_z^E increases in q_U^C. Here, we argue that, in some cases, the market forces may lead to autonomously implement the equilibrium that would be required by a regulatory agency.

Comparing the surplus of the content providers' subgroups under net-neutrality, we obtain the following corollary.

3.3 Corollary 2

$$CS_{2U}^C > CS_{1U}^C \text{ given } \theta_1^C > \theta_2^C$$

Proof Using (16) and (23), we easily obtain the result stated in corollary 2. ■

Under net neutrality, we prove that the least congestion sensitive content providers, belonging to subgroup 2, receive the highest consumer surplus. This finding is related to the fact that this subgroup incurs a lower congestion cost in comparison to subgroup 1. It is also linked to the fact that subgroup 2 is stronger insofar as $A_2^C > A_1^C$.

4 The ISP Non-net-neutrality Equilibrium

Under the non-net-neutrality regime, the ISP monopoly platform is allowed to third-degree price discriminate between content providers. Based on this, it takes into account the congestion sensitivity of each subgroup so as to manage its pricing policy on side C. The degree of the sensitivity to congestion of one content provider is incarnated in the scale of the data it exchanges with end-users.

Under this regime, the ISP charges both subgroups different prices. Its object is:

$$\max_{p_1^C, p_2^C, p^E} \Pi\left(p_1^C, p_2^C, p^E\right)$$

where $\prod(\cdot)$ is the profit function defined in (14).

The following proposition characterizes the ISP equilibrium under non-net-neutrality.

4.1 Proposition 2

Under the non-net-neutrality regime, the optimal prices of side E and side-C's subgroup $i \neq j$ are respectively given by:

$$p_D^E = q_D^E - \alpha^C q_D^C \tag{24}$$

and

$$p_{iD}^C = q_{iD}^C + \theta_i^C q_{iD}^C + \theta_j^C q_{jD}^C - \phi^E q_D^E \tag{25}$$

where the demand levels are:

$$q_D^E = \frac{\left(4\left(\theta_1^C + \theta_2^C + 1\right) - \left(\theta_1^C - \theta_2^C\right)^2\right) a^E + 4\left(\phi^E + \alpha^C\right) a^C}{2\Theta(\cdot)} \tag{26}$$

$$q_{iD}^C = \left(2 - \theta_i^C + \theta_j^C\right) \frac{\left(\phi^E + \alpha^C\right) a^E + 2a^C}{2\Theta(\cdot)} = \frac{2 - \theta_i^C + \theta_j^C}{4} q_D^C \tag{27}$$

and

$$q_D^C = 2 \frac{\left(\phi^E + \alpha^C\right) a^E + 2a^C}{\Theta(\cdot)}$$

with $\Theta(\cdot) = 4\left(\theta_1^C + \theta_2^C + 1\right) - \left(\theta_1^C - \theta_2^C\right)^2 - 2\left(\phi^E + \alpha^C\right)^2 > 0$ given the model's assumptions.

Proof See Appendix 1. ■

From the pricing rules stated in proposition 2, we deduce the following:

- We learn from Eq. (24) that the optimal price of side E is made up by two terms. The first, q_D^E, is the standard positive markup that measures the ISP market power on side E. The second term, $\alpha^C q_D^C$, teaches that the ISP monopolist reduces side-E's price by the network benefits that content providers reap from the end-users' potential participation.
- Equation (25) exhibits the ISP pricing rule on side-C's subgroup i. It is composed of four terms. The first one, q_{iD}^C, stands for the positive markup the ISP holds while servicing subgroup i. The second term, $\theta_i^C q_{iD}^C$, is the congestion cost that content providers in subgroup i impose on each other. The novelty that Eq. (25) entails is embodied in the third term, $\theta_j^C q_{jD}^C$: each content provider in subgroup i should pay the congestion cost of subgroup $j \neq i$. The intuition behind this finding is that the monopoly ISP allocates the congestion costs between both subgroups in order to guarantee an efficient use of the platform's fixed capacity. It is in fact merely related to the matter that both subgroups share the same bandwidth and are responsible for the platform congestion. The last term is the network benefits or losses (depending on the scale of ϕ^E) that subgroup i exerts on the participating end-users.

In order to go in depth into the analysis of the ISP pricing equilibrium under non-net-neutrality, we give the following corollaries. Corollary 3 studies the sign of the differential $\left(p_{1D}^C - p_{2D}^C\right)$. Corollary 4 compares the consumer surpluses on both subgroups to each other.

4.2 Corollary 3

Under the non-net-neutrality regime, we have

$$p_{2D}^C > p_{1D}^C \text{ given } \theta_1^C > \theta_2^C$$

Proof Using (25) and (27), we easily obtain corollary 3. ■

Corollary 3 shows that the ISP platform charges subgroup 1, the most congestion sensitive, lower price than subgroup 2. We also readily check that the ISP platform does so even if subgroup 1 incurs a higher congestion cost since $\theta_1^C q_D^C > \theta_2^C q_D^C$. This finding contradicts the traditional textbook congestion pricing according to which the consumers who are willing to pay more to avoid congestion should be charged more.[26] There are two explanations for this finding.[27]

[26] The traditional textbook model does call for consumers who value congestion highly to pay more. However, they are also provided with higher quality service. High Occupancy Toll (HOT) lanes are a good example.

[27] Following the literature on airport congestion pricing (e.g. Brueckner 2002), our result can also be justified by the fact that larger content providers could internalize the congestion they impose. Therefore, the ISP could charge them lower price than small content providers. This observation is in fact criticized by Joseph Daniel (and his coauthors) who argues that the congestion-reducing efforts by larger users allow the small ones to fill the gap; keeping then the congestion level unchanged (see, for instance, Daniel and Harback 2008). This argument can also be observed in the non-atomistic Internet market. Larger content providers can utilize compression techniques (Peitz and Schuett 2016) to reduce the traffic volume. As a result, the bandwidth space they liberate would be exploited by (extra) small contents; keeping then the platform congestion level unchanged.

- The first is reminiscent of the traditional third-degree price discrimination rule which teaches that the strongest market, subgroup 2 in our case (see (6)), should be charged a higher price (Layson 1998).[28]
- The second could be related to the demand interdependencies the ISP platform faces. In this setting, we note that the monopolist cannot only subsidize one side with the aim to exploit the other side depending on the cross-externalities' scale but can do so within one side. The ISP manager utilizes her pricing policy on side C to attract more members on subgroup 1. The latter is then regarded as a source of profit for the ISP business platform. Therefore, it is subsidized to the detriment of subgroup 2.

We notice that the monopolist is aware of the fact that the most congestion sensitive content providers exhibit strong attractiveness forces. These forces are embedded in the highly valued contents they exchange with end-users and hence the intense traffic they create. This traffic generates bigger benefits to the ISP. The phenomenon can be observed in practice. Some leading websites such as YouTube, Facebook, Google and Twitter are largely congestion sensitive. They also supply huge contents like streaming videos, voice telephony services and media products. These contents contribute to create an intense traffic and bring substantial profits to the platform. Indeed, it may be profitable for the ISP manager to attract them by charging lower price for their access.

We draw the attention of an ISP manager that it may not be enough to look for the market side that brings more profit to her platform. Consideration should also be given to the fact that there may be one subgroup among the content providers to favor over the other subgroup(s). The favored subgroup is the one which is characterized by an overwhelming attractiveness. We envisage providing a deep analysis of this fact in Section 5.

Comparing the surplus of the content providers' subgroups under non-net-neutrality, we obtain corollary 4.

4.3 Corollary 4

$$CS_{2D}^{C} > CS_{1D}^{C} \text{ given } \theta_1^C > \theta_2^C$$

Proof Using (16) and (25), we straightforwardly obtain corollary 4. ■

It is worth noticing that subgroup 1 bears the greater congestion cost in comparison with subgroup 2. In this setting, we argue that for equal network benefits ($\alpha^E q_D^E$) and

[28] Layson (1998) emphasizes, on page 517, that "*There are two interesting results concerning the direction of the price changes for the linear interdependent demand case: (1) regardless of how large the cross-price effects are, if marginal cost is constant or rising, price must rise in the strong market and fall in the weak market and (2) regardless of how sharply marginal cost falls, if demands are linear with symmetric cross-price effects, price must rise in the strong market and fall in the weak market*".

for equal submarket gross-strengths ($a_i^C = a^C$) it is the greater congestion cost that causes subgroup 1 to extract lower surplus from joining the platform. It is also important to mention that this finding is related to the fact that subgroup 2 is stronger than subgroup 1 insofar as $A_2^C > A_1^C$.

5 Net Neutrality Versus Non-net Neutrality

Before presenting the surplus and welfare comparisons in propositions 3 and 4 respectively, we firstly give the following lemma that studies the price and demand comparisons.

5.1 Lemma: Price and Demand Comparisons

- **Price comparison**

– **Side *C***

$$p_{1D}^C < p_U^C < p_{2D}^C \tag{28}$$

– **Side *E***

$$\mathrm{sign}\left(p_D^E - p_U^E\right) = \mathrm{sign}\left(\phi^E - \alpha^C\right) \tag{29}$$

- **Demand comparison**

– **Side *C***

$$q_D^C > q_U^C \tag{30}$$

and

$$q_{1U}^C < q_{1D}^C \text{ and } q_{2D}^C > q_{2U}^C \tag{31}$$

– **Side *E***

$$q_D^E > q_U^E \tag{32}$$

Proof See Appendix 1. ■

Inequality (28) teaches that when the ISP switches from net-neutrality to non-net-neutrality, it reduces the type-1's price and increases that of type 2. By doing so, it raises the former's demand and lowers the latter's (see (31)). We shall interpret the differentials $q_{1D}^C - q_{1U}^C$ and $q_{2U}^C - q_{2D}^C$ as the number of the extra-comers to subgroup 1 and the number of the leavers from subgroup 2 respectively. We easily check in Appendix 1 that $q_{1D}^{C}-q_{1U}^{C} > q_{2U}^{C}-q_{2D}^{C}$ and therefore the overall demand on side C increases under non-net-neutrality by the differential $q_{D}^{C}-q_{U}^{C}$. The latter mass consists uniquely of type-1 content providers. Due to cross-externalities, the extra participants on side C under non-net-neutrality contribute to expand the demand on side E by $q_{D}^{E}-q_{U}^{E}$. Bourreau et al. (2015) assimilate the entry of new providers, due to non-net-neutrality, to content innovation. In our context, we interpret $q_{D}^{C}-q_{U}^{C}$ as content innovation that comes only from big bandwidth *hogs*. However, the content innovation of the small startups decreases under non-net-neutrality. We can also argue that the ISP monopoly tends to "*sabotage*" the type-2 content providers by charging them higher price in comparison with type-1 (Bourreau et al. 2015).

The demand expansion on side E occurs even if the ISP monopolist can enlarge or curtail p^E depending on the magnitude of the network externalities (see (29)). In this setting, it happens that the extra type-1 content providers (or the extra innovation) may have the potential to lure additional end-users even if they are charged higher price in comparison to net neutrality. To give further explanations, we emphasize that the network-benefit and congestion effects play a key role. We note the following:

i) $p_{D}^{E} < p_{U}^{E}$ for $\phi^E \in [-1, \alpha^C[$ and $\alpha^C \in [0, 1[$: While departing from net neutrality to non-net-neutrality, the ISP monopolist downwardly adjusts p^E when side E is the needed-more side ($\alpha^C > \phi^E$). The reasoning is that the congestion effect reaching side E outweighs the network benefit it reaps from side C. By decreasing p^E, the ISP monopoly tends to compensate the congestion effect it imposes on end-users.
ii) $p_{D}^{E} > p_{U}^{E}$ for $\phi^E \in]\alpha^C, 1-\alpha^C[$ and $\alpha^C \in \left[0, \frac{1}{2}\right[$: In this case side C is the needed-more side ($\alpha^C > \phi^E$), the ISP monopolist upwardly adjusts the price charged on side E when she moves from net neutrality to non-net-neutrality. Unlike case *i*), we emphasize here that the network benefits gained by side E overcome the congestion effect. Since side C is the needed-more side, end-users will then reap higher network benefits from the demand expansion on it.
iii) $p_{D}^{E} = p_{U}^{E}$ for $\phi^E = \alpha^C$ and $\alpha^C \in [0, 1[$. In this setting, we note that the pricing regime the monopolist undertakes on the content providers' side has no impact on the end-users'. Accordingly, it is immediate here to mention that the two effects defined above offset each other.

The most congestion sensitive content providers can be considered to be responsible at the same time for provoking congestion and bringing network benefits. The ISP monopolist manages then the pricing policy on side E bearing this fact in mind. When the network benefits that subgroup-1 members exhibit overcome the congestion they entail, the ISP increases p^E. On the other hand, when congestion dominates the network benefits, side E is charged a lower price.

While deviating from net neutrality to non-net-neutrality, the monopoly re-internalizes, through its pricing policy, the network externalities it faces. By doing so, it re-congests its platform by increasing the participation of type-1 content providers and reducing that of type-2 with asymmetric magnitudes. This implies, then, a demand expansion (or innovation) on side C by some type-1 members that contribute to expand, via cross-externalities, the demand on side E.

The pricing management without net neutrality shows that the most congestion sensitive content providers are a source of profit for the ISP platform. The platform tends to favor them over the less congestion-sensitive ones by charging them a lower price. This way of internalization arises two effects on the end-users' side: a network-benefit effect and a congestion effect. If the former outweighs the latter, the ISP intervenes by charging end-users more under non-net neutrality in comparison with net neutrality. If the congestion effect dominates the network-benefit effect, the ISP charges end-users less.

In practice, the privileged web companies like Google, Facebook and YouTube are overwhelmingly attractive for web-surfers. This is supported by Robert McMillan's claim that "…*today, half of the Internet's traffic comes from just 30 outfits, including Google, Facebook and Netflix*".[29] Indeed, what the above lemma recommends is that any ISP connecting with these bandwidth *hogs* should allow them to access for lower charges due to their effective role in its profitability.

From the lemma's analysis, we draw an interesting new result compared to the literature on two-sided markets and net neutrality. The platform could distinguish and therefore subsidize the subgroup-not only the side- that represents a source of profit for it.

We show below that these demand expansions and the network values they incarnate will play a crucial role in the consumer surplus, profit and welfare comparisons.

The following proposition investigates the surplus and welfare implications of (not) abolishing net neutrality.

5.2 Proposition 3

Comparing the consumer surplus, profit and welfare levels under both pricing regimes, we have:

i) **Consumers' surplus comparison**

– **Side *E***

$$CS_D^E > CS_U^E \tag{33}$$

– **Side *C***

$$CS_{1D}^C > CS_{1U}^C \tag{34}$$

$$CS_{2D}^C < CS_{2U}^C \tag{35}$$

[29] See https://www.wired.com/2014/06/net_neutrality_missing/.

ii) **Profit comparison**

$$\Pi_D > \Pi_U \tag{36}$$

iii) **Total consumer surplus comparison**

$$CS_D \geq CS_U \text{ for } \begin{cases} y_D \in \left]0, \widehat{y}_D\right] \text{ and } y_U \in]0, y_D[\\ \text{or} \\ y_D \in \left]\widehat{y}_D, +\infty\right[\text{ and } y_U \in \left]0, \widehat{y}_U\right] \end{cases} \tag{37}$$

and

$$CS_D < CS_U \text{ for } y_D \in \left]\widehat{y}_D, +\infty\right[\text{ and } y_U \in \left]\widehat{y}_U, y_D\right[\tag{38}$$

where

$$\widehat{y}_U = \sqrt{\frac{16\left(\frac{\left(q_D^E\right)^2 - \left(q_U^E\right)^2}{\Lambda^2 y_D^2}\right) + \left(\left(1+\theta_1^C\right)\left(2-\theta_1^C+\theta_2^C\right)^2 + \left(1+\theta_2^C\right)\left(2-\theta_2^C+\theta_1^C\right)^2\right)}{4\left(1+\theta_1^C\right)\left(1-\theta_1^C+\theta_2^C\right)^2 + \left(1+\theta_2^C\right)\left(1-\theta_2^C+\theta_1^C\right)^2}}\, y_D$$

and

$$\hat{y}_D = \frac{4}{\left(\theta_1^C - \theta_2^C\right)\Lambda}\sqrt{\frac{\left(q_D^E\right)^2 - \left(q_U^E\right)^2}{3\left(\theta_1^C + \theta_2^C\right) + 2}} > 0$$

are critical values or thresholds of the platform equilibrium congestion level under net neutrality and non-net-neutrality, respectively.

Proof See Appendix 1. ■

- **End-users surplus comparison**

Equation (33) proves that end-users are better off without net neutrality. As abovementioned, it is the additional type-1 content providers who play a crucial role in allowing end-users to derive larger surplus under non-net-neutrality. This surplus is reinforced by the ISP's pricing policy when congestion effect matters.

Depending on side-*E*'s net market strengths under both regimes, we easily verify that:

$$sign\left(A_D^E - A_U^E\right) = sign\left(\phi^E\right)$$

We notice that side E is much stronger under non-net-neutrality when the marginal network benefit that end-users extract from content providers exceeds the marginal congestion cost they bear, (i.e. $\phi^E > 0$). This largeness of the end-users' net market strength warrants the superiority of their surplus when net neutrality is abolished. Conversely, in the case where the marginal congestion cost surpasses the marginal network benefit, we check that side E is weaker under non-net-neutrality. In this case, the greater end-users' consumer surplus is caused by the lower access price they are charged.

We can also notice that end-users gain higher consumer surplus under non-net-neutrality even if they incur a higher congestion cost as $\Omega_D^E > \Omega_U^E$. We argue here that the congestion cost the platform imposes on side E is offset not only by its pricing policy but also by the network benefits they reap due to the demand expansion on side C.

- **Content providers surplus comparison**

On the content providers' side, the ISP favors subgroup 1 over subgroup 2 by charging it lower access price under non-net-neutrality. As a result, the most congestion sensitive content providers reap the highest consumer surplus. This result holds even with a higher congestion cost of heavy content providers.

- **Platform profit comparison**

Based on (36), we assert that the ISP monopoly is unambiguously better off under non-net-neutrality. This result goes in line with the traditional third-degree price discrimination outcome (Tirole 1988). In our case, the increase in the ISP's profit is caused not only by the demand expansion on the content providers' side where third-price discrimination matters but also by the demand expansion on the end-users' side. Due to cross-externalities, we emphasize here that the demand reallocation on side C, as a result of the price discrimination therein, reallocates the demand on side E as well.

- **Total consumer surplus comparison**

We learn from conditions (37) and (38) that there exists a critical value or threshold of the platform equilibrium congestion level under regime k, denoted $\widehat{y}_k, k = U, D$, that indicates the preferable pricing scenario for the platform's consumers. We regard such a threshold as an indicator or a proxy that identifies for a planner the regime to enact: net neutrality or non-net-neutrality. It is also tempting to note that the pre-mentioned indicator depends on the participating content providers (or the traffic in our context) and the platform's fixed transmission capacity. Depending on these factors, the regulator may evaluate the effect of (not) ending net neutrality rules.

As price discrimination on side C leads to demand expansions on both sides, it would seem that the total consumer surplus increases accordingly, as is the case under condition (37). However, condition (38) proves that this fact is not always true. In the case where the equilibrium congestion level under both regimes is sufficiently high (i.e. exceeds the threshold $\widehat{y}_k$), we deduce that non-net-neutrality renders the ISP's consumers as a whole worse-off. The intuition behind this finding is that, without net neutrality, the negative congestion effect strongly dominates the network-benefit effect. As a result, the decrease in the surplus of type-2 content providers outweighs the increase in the surpluses of both end-users and type-1 content providers. In this setting and under condition (38), we straightforwardly verify the following:

$$CS_U > CS_D \text{ for } \underbrace{\left(CS_{2U}^C - CS_{2D}^C\right)}_{\text{Type-2's surplus decrease}} > \underbrace{\left(CS_D^E - CS_U^E\right)}_{\text{End-users' surplus increase}} + \underbrace{\left(CS_{1D}^C - CS_{1U}^C\right)}_{\text{Type-1's surplus increase}}$$

When the platform congestion level on either regime is high enough $(y_k > \widehat{y}_k)$, it is insufficient to ensure a higher total consumer surplus by price-favoring the most congestion sensitive subgroup. By doing so, the ISP monopoly can hurt type-2 more than it can please both type-1 and end-users. Under condition (38) the consumers are thus losing out when the ISP monopolist moves from net neutrality to non-net-neutrality.

Now, we are ready to give the proposition that deals with the social welfare comparison.

5.3 Proposition 4

– **Welfare comparison**

Under (37), we get

$$W_D > W_U \tag{39}$$

Under (38), we get

$$sign(W_D - W_U) = sign(\varsigma - \rho) \tag{40}$$

with

$\varsigma = \Pi_D - \Pi_U$ is the profit-increase effect
and

$\rho = CS_U - CS_D$ is the surplus-reduction effect.

Proof See Appendix 1.

In the case where the platform equilibrium congestion level under each regime is sufficiently high $(y_k > \widehat{y}_k)$ (condition (38)), we prove that net neutrality can be preferable to non-net-neutrality in terms of social welfare. In this regard, there are two opposite effects that must be mentioned. The first is denoted $\varsigma = \Pi_D - \Pi_U$ and defined as the *profit-increase* effect. The second is denoted $\rho = CS_U - CS_D$

and defined as the *surplus-reduction* effect.[30] In the case where the latter weakly dominates the former, (i.e. $\varsigma \leq \rho$) and condition (38) holds, we easily check that net neutrality is welfare improving. It should be noted here that the demand expansion on side C under non-net-neutrality provokes an intense congestion. This latter has two possible implications. Firstly, it tremendously offsets the network-benefit effect. Secondly, it induces a greater reduction in the consumer surplus of the least congestion sensitive subgroup. Such a reduction strongly dominates the sum of the respective increases in the consumer surplus of subgroup 1, the consumer surplus of end-users and the ISP's profit. In this context, we readily check that:

$$W_U > W_D \text{ for } \underbrace{\left(CS_{2U}^C - CS_{2D}^C\right)}_{\text{Type-2's surplus decrease}} > \underbrace{\left(CS_D^E - CS_U^E\right)}_{\text{End-users' surplus increase}} + \underbrace{\left(CS_{1D}^C - CS_{1U}^C\right)}_{\text{Type-1's surplus increase}} + \underbrace{\left(\Pi_D - \Pi_U\right)}_{\text{ISP's profit increase}}$$

Under condition (37) and/or under condition (38) but with $\varsigma > \rho$, we can conclude that it is undesirable for a regulatory agency to intervene in order to impose the net neutrality rules. It is therefore preferable to leave the market act freely as it moves towards establishing the welfare-superiority of the non-net-neutrality regime.[31]

The analysis behind our paper generates some results whose interpretations are interesting and deserve the attention of the net neutrality concerned parties. We show that the traditional redistributive effect on the content providers' side, due to price discrimination, implies also a redistributive effect on the end-users'; the side where third-degree price discrimination does not occur. This is linked to the demand interdependencies the ISP manager faces. We also prove that the ISP monopolist favors, in terms of pricing policy, the most congestion sensitive subgroup. We argue that this pricing behavior is the sort of the traditional Caillaud and Jullien (2003) divide-and-conquer pricing strategy.

The results also reveal that there exists a threshold of the platform equilibrium congestion level under each regime that indicates for a planner when net neutrality is beneficial for the society. This threshold is regarded as an indicator or proxy on whom any Internet regulating agency should base its decision on repealing or not net neutrality protections. In the case where this platform equilibrium congestion level is relatively low, we emphasize that non-net-neutrality dominates net neutrality to all the concerned parties, viz. ISP platform, content providers and end-users. In this setting, we note that it is preferable that any Internet regulatory agency does not intervene to enforce net neutrality rules. The free-functioning of the market leads to the fact that non-net-neutrality is more socially desirable. When the platform equilibrium congestion level is high enough, net neutrality can improve welfare. In this frame, two countervailing effects are identified: a profit-increase effect and a surplus-reduction effect. When the latter dominates the former, we show that net neutrality makes the society better-off.

[30] Through this angle, we draw the attention of a regulator that it would not always be sufficient to give primary, even exclusively, attention to changes in consumers' surplus. Changes in profits could also play a key role.

[31] In Appendix 2, we adopt some numerical values to illustrate the main findings of the welfare comparison.

6 Conclusion

In this paper, we have studied the extent to which net neutrality, defined as price non-discrimination, is welfare-superior in comparison to non-net-neutrality. We have examined a model where a congested ISP monopoly acts as a two-sided platform and supplies access to distinct and interdependent sides which are end-users and content providers. We have assumed that congestion is particularly related to the traffic and affects both sides. It is supposed to be essentially caused by the overuse of the fixed bandwidth by content providers. Unlike end-users, we have allowed content providers to be heterogeneous in their sensitivity to congestion. Indeed, there are two subgroups of content providers; one of which, called heavy content providers, is the most sensitive to congestion or the most willing to pay to avoid congestion. We have shown that the ISP monopolist, by switching from net neutrality, favors the heavy content providers over the light ones. A possible explanation is that the ISP platform adopts the standard divide-and-conquer pricing strategy but within the content providers' side. This is due to the fact that the heavy content providers play a primary role in attracting intense traffic and therefore generating higher profit for the ISP monopoly. The results have also revealed that when the network benefits of the end-users' side reach a critical value, it is profit-maximizing for the ISP to voluntarily allow content providers free access. This zero-price is warranted by the fact that the content providers' side is the needed-more side as it brings more value to the network. In this setting, we have argued that some regulating agencies in some countries might not be urged to intervene and enforce the zero-price net neutrality rule.

We have also proven that the net neutrality is more welfare-improving than non-net-neutrality cannot always occur. We have determined, in our paper, a critical threshold of the platform equilibrium congestion level under each regime that indicates when net-neutrality is more socially desirable and when it is not. This threshold is interpreted as an indicator or a proxy that indicates for any Internet agency the socially preferable regime it should enact: net neutrality or non-net-neutrality. When the platform equilibrium congestion level is relatively low, we have demonstrated that non-net-neutrality, compared to net neutrality, improves welfare. In this context, we have recommended that it is not advisable that a regulatory agency intervenes to enact net neutrality. In the case where the platform equilibrium congestion level is sufficiently high, two effects are defined: a profit-increase effect and a surplus-reduction effect. If the latter overcomes the former, net neutrality increases social welfare in comparison to non-net-neutrality.

Appendix 1

Proofs

Before giving the sketch of the required proofs, we present the primitives from which the inverse demand functions are derived. The inverse demand function of side E, defined in (2), can be obtained by maximizing $U^E(\cdot) - p^E q^E$ where $U^E(\cdot) = a^E q^E - \frac{1}{2}\left(q^E\right)^2 + \left(\alpha^E - \theta^E\right) q^E q^C$. Likewise, the inverse demand function of the content providers' subgroup i, defined in (4), can be deduced from the maximization of $U_i^C(\cdot) - p_i^C q_i^C$ where $U_i^C(\cdot) = a_i^C q_i^C - \frac{1}{2}\left(q_i^C\right)^2 + \alpha^C q_i^C q^E - \frac{\theta_i^C}{2}\left(q^C\right)^2$. In fact, $U^E(\cdot)$ and $U_i^C(\cdot)$ are à-la Dixit (1979) quadratic-concave utility functions.

Proof of Proposition 1

Under the net neutrality regime, the object of the ISP monopolist is

$$\max_{p^E, p^C} \Pi\left(p^E, p^C\right)$$

where $\Pi(p^E, p^C)$ is the profit function defined in (7).

The corresponding first order conditions are:

$$\frac{\partial \Pi}{\partial p^E} = 0$$
$$\frac{\partial \Pi}{\partial p^C} = 0$$

that are equivalent to

$$\begin{cases} \left(1 + \theta_1^C + \theta_2^C\right)p^E + \left(\phi^E + \alpha^C\right)p^C = \dfrac{1 + \theta_1^C + \theta_2^C}{2} a^E + \phi^E a^C \\ \left(\phi^E + \alpha^C\right)p^E + 2p^C = \alpha^C a^E + a^C \end{cases}$$

Using Cramer's rule we easily obtain the ISP's optimal access prices under net neutrality. They are given by (as mentioned above, the subscripts U and D stand for net neutrality and non-net-neutrality respectively):

$$p_U^E = \frac{\left(\theta_1^C + \theta_2^C + 1 - \alpha^C\left(\phi^E + \alpha^C\right)\right)a^E + \left(\phi^E - \alpha^C\right)a^C}{\psi(\cdot)} \tag{41}$$

and

$$p_U^C = \frac{\left(1 + \theta_1^C + \theta_2^C\right)\left(\alpha^C - \phi^E\right)a^E + 2\left(1 + \theta_1^C + \theta_2^C - \phi^E\left(\phi^E + \alpha^C\right)\right)a^C}{2\psi(\cdot)} \tag{42}$$

with $\psi(\cdot) = 2\left(\theta_1^C + \theta_2^C + 1\right) - \left(\phi^E + \alpha^C\right)^2 > 0$ given the model's assumptions.

Inserting (41) and (42) in (8), (9) and (10), we get:

$$q_U^E = \frac{\left(\theta_1^C + \theta_2^C + 1\right)a^E + \left(\phi^E + \alpha^C\right)a^C}{\psi(\cdot)} \tag{43}$$

$$q_U^C = \frac{(\phi^E + \alpha^C)a^E + 2a^C}{\psi(\cdot)} \tag{44}$$

and

$$q_{iU}^C = \frac{\left(1 - \theta_i^C + \theta_j^C\right)}{2} \frac{(\phi^E + \alpha^C)a^E + 2a^C}{\psi(\cdot)} \tag{45}$$

In the case where $\phi^E \in [-1, -\alpha^C[$, μ^E must be in $\left]\frac{-\phi^E - \alpha^C}{1 + \theta_1^C + \theta_2^C}, \frac{2}{-\phi^E - \alpha^C}\right[$ so as to guarantee that $q_U^k > 0$, $k = C, E$.

We deduce from Eqs. (44) and (45) that

$$q_{iU}^C = \frac{1 - \theta_i^C + \theta_j^C}{2} q_U^C \tag{46}$$

Compared to $\frac{q_U^C}{2}$, we learn from Eq. (46) the following:

$$q_{1U}^C = \frac{q_U^C}{2} - \omega_U^C \tag{47}$$

and

$$q_{2U}^C = \frac{q_U^C}{2} + \omega_U^C \tag{48}$$

with $\omega_U^C = \left(\theta_1^C - \theta_2^C\right) \frac{q_U^C}{2} > 0$ is a demand adjusting factor. Equations (47) and (48) imply that $q_{2U}^C > q_{1U}^C$.

Using expressions (41) to (44), we easily check that

$$p_U^E = q_U^E - \alpha^C q_U^C \tag{49}$$

and

$$p_U^C = \frac{q_U^C}{2} + \theta_1^C \frac{q_U^C}{2} + \theta_2^C \frac{q_U^C}{2} - \phi^E q_U^C \tag{50}$$

The ISP's equilibrium profit level under net neutrality is easily obtained using expressions (41) to (44). It is explicitly given by:

$$\Pi_U = \frac{\left(1 + \theta_1^C + \theta_2^C\right)(a^E)^2 + 2\left(\phi^E + \alpha^C\right)a^E a^C + 2(a^C)^2}{2\psi(\cdot)} \tag{51}$$

Based on expressions (15) to (19), the equilibrium consumer surplus and welfare levels under net neutrality are respectively given by:

$$CS_U^E = \frac{1}{2}\left(q_U^E\right)^2 \tag{52}$$

$$CS_{iU}^C = \frac{\left(1+\theta_i^C\right)}{2}\left(q_{iU}^C\right)^2 \tag{53}$$

$$CS_U^C = \sum_{i=1}^{2} CS_{iU}^C \tag{54}$$

$$CS_U = CS_U^E + CS_U^C \tag{55}$$

and

$$W_U = CS_U + \Pi_U \tag{56}$$

To verify the second order conditions for the monopolist's equilibrium under net neutrality, we should verify that the following conditions simultaneously hold:

- $\frac{\partial^2 \Pi}{\partial (p^k)^2} < 0$ for $k = E, C$
  - det $H \geq 0$

with

$$H = \begin{pmatrix} \frac{\partial^2 \Pi}{\partial (p^E)^2} & \frac{\partial^2 \Pi}{\partial p^E \partial p^C} \\ \frac{\partial^2 \Pi}{\partial p^C \partial p^E} & \frac{\partial^2 \Pi}{\partial (p^C)^2} \end{pmatrix}$$

is the Hessian matrix.

We straightforwardly check that

$$\frac{\partial^2 \Pi}{\partial (p^E)^2} = -2\frac{1+\theta_1^C+\theta_2^c}{\Xi(\cdot)} < 0 \tag{57}$$

$$\frac{\partial^2 \Pi}{\partial (p^C)^2} = -\frac{4}{\Xi(\cdot)} < 0 \tag{58}$$

and

$$\det H = \frac{4\psi(\cdot)}{(\Xi(\cdot))^2} > 0 \tag{59}$$

Indeed, inequalities (57), (58) and (59) allow us to deduce that the ISP's profit function is concave with respect to (p^E, p^C). ■

Proof of Proposition 2

Under the non-net-neutrality regime, the object of the ISP monopolist is

$$\max_{p^E, p_1^C, p_2^C} \Pi\left(p^E, p_1^C, p_2^C\right)$$

where $\prod\left(p^E, p_1^C, p_2^C\right)$ is the profit function defined in (14).

The corresponding first order conditions are $(i = 1, 2)$

$$\frac{\partial \prod}{\partial p^E} = 0$$
$$\frac{\partial \prod}{\partial p_i^C} = 0$$

that imply

$$\begin{cases} 2(\theta_1^C + \theta_2^C + 1)p^E + (\phi^E + \alpha^C(1 - \theta_1^C + \theta_2^C))p_1^C + (\phi^E + \alpha^C(1 - \theta_2^C + \theta_1^C))p_2^C = (\theta_1^C + \theta_2^C + 1)a^E + 2\phi^E a^C \\ \left(\phi^E + \left(1 - \theta_i^C + \theta_j^C\right)\alpha^C\right)p^E + 2\left(1 + \theta_j^C - \phi^E\alpha^C\right)p_i^C + \left(2\phi^E\alpha^C - \left(\theta_i^C + \theta_j^C\right)\right)p_j^C = \left(1 - \theta_i^C + \theta_j^C\right)\left(\alpha^C a^E + a^C\right) \end{cases}$$

Using Cramer's rule, we obtain the optimal sides' price under non-net-neutrality. They are given by

$$p_D^E = \frac{\left(\left(4\left(\theta_1^C + \theta_2^C + 1\right) - \left(\theta_1^C - \theta_2^C\right)^2\right) - 4\alpha^C\left(\phi^E + \alpha^C\right)\right)a^E + \left(\phi^E - \alpha^C\right)a^C}{2\Theta(\cdot)} \tag{60}$$

and

$$p_{iD}^C = \frac{\left[\begin{array}{l} \left(2\left(\alpha^C - \phi^E\right) + \left(\alpha^C - 3\phi^E\right)\theta_i^C + \left(3\alpha^C - \phi^E\right)\theta_j^C - \alpha^C\left(\theta_i^C - \theta_j^C\right)^2\right)a^E \\ + \left(2\left(\theta_i^C + 3\theta_j^C + 2 - \left(\theta_i^C - \theta_j^C\right)^2\right) - 2\phi^E\left(\alpha^C + \phi^E\right)\right)a^C \end{array}\right]}{2\Theta(\cdot)} \tag{61}$$

with $\Theta(\cdot) = 4\left(\theta_1^C + \theta_2^C + 1\right) - \left(\theta_1^C - \theta_2^C\right)^2 - 2\left(\phi^E + \alpha^C\right)^2 > 0$ given the model's assumptions.

Plugging (60) and (61) into (11), (12) and (13), we obtain

$$q_D^E = \frac{\left(4\left(\theta_1^C + \theta_2^C + 1\right) - \left(\theta_1^C - \theta_2^C\right)^2\right)a^E + 4\left(\phi^E + \alpha^C\right)a^c}{2\Theta(\cdot)} \tag{62}$$

$$q_{iD}^C = \left(2 - \theta_i^C + \theta_j^C\right)\frac{\left(\phi^E + \alpha^C\right)a^E + 2a^C}{2\Theta(\cdot)} \tag{63}$$

and

$$q_D^C = 2\frac{\left(\phi^E + \alpha^C\right)a^E + 2a^C}{\Theta(\cdot)} \tag{64}$$

In the case where $\phi^E \in [-1, -\alpha^C[$, μ^E must be in $\left]\frac{4\left(-\phi^E - \alpha^C\right)}{4\left(1 + \theta_1^C + \theta_2^C\right) - \left(\theta_1^c - \theta_2^C\right)^2}, \frac{2}{-\phi^E - \alpha^C}\right[$ so as to guarantee that $q_D^k > 0$, $k = C, E$.

Equations (63) and (64) allow us to readily verify that

$$q_{iD}^C = \frac{2 - \theta_i^C + \theta_j^C}{4} q_D^C \tag{65}$$

As shown in the net-neutrality case, we compare q_{iD}^C to $\frac{q_D^C}{2}$. We verify that.

$$q_{1D}^C = \frac{q_D^C}{2} - \omega_D^C \tag{66}$$

and

$$q_{2D}^C = \frac{q_D^C}{2} + \omega_D^C \tag{67}$$

with $\omega_D^C = \left(\theta_1^C - \theta_2^C\right)\frac{q_U^C}{4} > 0$ is a demand adjusting factor under non-net-neutrality. Equations (66) and (67) imply that $q_{2D}^C > q_{1D}^C$.

Managing all the equations from (60) to (65), we deduce that

$$p_D^E = q_D^E - \alpha^C q_D^C \tag{68}$$

and (for $i, j = 1, 2$ and $i \neq j$)

$$p_{iD}^C = q_{iD}^C + \theta_i^C q_{iD}^C + \theta_j^C q_{jD}^C - \alpha^E q_D^E = \frac{2 - \left(\theta_i^C - \theta_j^C\right)^2 + \theta_i^C + 3\theta_j^C}{4} q_D^C - \alpha^E q_D^E \tag{69}$$

Using expressions from (60) to (63) we determine the equilibrium profit level. It is given by:

$$\Pi_D = \frac{\left(4\left(\theta_1^C + \theta_2^C\right) + 4 - \left(\theta_1^C - \theta_2^C\right)^2\right)\left(a^E\right)^2 + 8\left(\phi^E + \alpha^C\right)a^E a^C + 8\left(a^C\right)^2}{4\Theta(\cdot)} \tag{70}$$

Using expressions (15) to (19), we deduce that:

$$CS_D^E = \frac{1}{2}\left(q_D^E\right)^2 \tag{71}$$

$$CS_{iD}^C = \frac{\left(1 + \theta_i^C\right)}{2}\left(q_{iD}^C\right)^2 \tag{72}$$

$$CS_D^C = \sum_{i=1}^{2} CS_{iD}^C \tag{73}$$

$$CS_D = CS_D^E + CS_D^C \tag{74}$$

and

$$W_D = CS_D + \Pi_D \tag{75}$$

In order to verify the second order conditions for the monopolist's equilibrium under non-net-neutrality, we define first the Hessian matrix. It is given by:

$$H = \begin{pmatrix} \frac{\partial^2 \Pi}{\partial (p^E)^2} & \frac{\partial^2 \Pi}{\partial p^E \partial p_1^C} & \frac{\partial^2 \Pi}{\partial p^E \partial p_2^C} \\ \frac{\partial^2 \Pi}{\partial p_1^C \partial p^E} & \frac{\partial^2 \Pi}{\partial \left(p_1^C\right)^2} & \frac{\partial^2 \Pi}{\partial p_1^C \partial p_2^C} \\ \frac{\partial^2 \Pi}{\partial p_2^C \partial p^E} & \frac{\partial^2 \Pi}{\partial p_2^C \partial p_1^C} & \frac{\partial^2 \Pi}{\partial \left(p_2^C\right)^2} \end{pmatrix}$$

Afterwards, we should ensure that the following conditions hold:

- $\frac{\partial^2 \Pi}{\partial (p^E)^2} < 0$ and $\frac{\partial^2 \Pi}{\partial \left(p_i^C\right)^2} < 0$, that is all diagonal elements are negative,
- the determinant of the second leading term, which we denote H_2, is positive and
- the determinant of the Hessian matrix H is negative

with

$$H_2 = \begin{pmatrix} \dfrac{\partial^2 \Pi}{\partial (p^E)^2} & \dfrac{\partial^2 \Pi}{\partial p^E \partial p_1^C} \\ \dfrac{\partial^2 \Pi}{\partial p_1^C \partial p^E} & \dfrac{\partial^2 \Pi}{\partial \left(p_1^C\right)^2} \end{pmatrix}$$

We show that[32]

$$\frac{\partial^2 \Pi}{\partial (p^E)^2} = -2\frac{1+\theta_1^C+\theta_2^C}{\Xi(\cdot)} < 0 \tag{76}$$

$$\frac{\partial^2 \Pi}{\partial (p_i^C)^2} = -2\frac{1+\theta_j^C - \phi^E \alpha^C}{\Xi(\cdot)} < 0 \tag{77}$$

$$\det H_2 = \frac{\begin{bmatrix} -(\alpha^C)^2 (\theta_1^C)^2 \\ +2\left(2+\alpha^C(\alpha^C - \phi^E) + \left(2+(\alpha^C)^2\right)\theta_2^C\right)\theta_1^C \\ +\left(4-(\alpha^C)^2\right)(\theta_2^C)^2 + 2\left(4-\alpha^C(3\phi^E+\alpha^C)\right)\theta_2^C + 4-\left((\phi^E)^2+(\alpha^C)^2+6\phi^E\alpha^C\right) \end{bmatrix}}{[\Xi(\cdot)]^2} > 0 \tag{78}$$

and

$$\det H = -\frac{\Theta(\cdot)}{(\Xi(\cdot))^2} < 0 \tag{79}$$

Inequalities (76) to (79) ensure the concavity of $\Pi\left(p^E, p_1^C, p_2^C\right)$. ■

Proof of the Lemma

- **Price comparison**

Equations (41) and (60) allow us to show that

$$p_D^E - p_U^E = \frac{\left(\theta_1^C - \theta_2^C\right)^2 \left(\alpha^C - \phi^E\right)}{2\Theta(\cdot)} q_U^C \tag{80}$$

[32] As for the sign of *det* H_2, we easily check that equalizing the numerator to zero implies two distinct θ_1^C-roots which are $\overline{\theta}_1^C > 1$ and $\theta_{-1}^{\ C} < 0$.

Furthermore, we deduce from Eqs. (41) and (69) that.

$$p_{1D}^C - p_U^C = \frac{\left(\theta_1^C - \theta_2^C\right) q_U^C}{2\Theta(\cdot)} \Psi(\cdot) \tag{81}$$[33]

with[33]

$$\Psi(\cdot) = \left(\theta_1^C + 1 - \theta_2^C\right)\left(\alpha^C\right)^2 + \phi^E\left(\theta_1^C + 2 - \theta_2^C\right)\alpha^C + \left(\phi^E\right)^2 - \left(\theta_1^C + 2 - \theta_2^C\right)\left(\theta_1^C + \theta_2^C + 1\right) < 0$$

Indeed,

$$p_{1D}^C < p_U^C \tag{82}$$

Likewise, we prove that

$$p_{2D}^C - p_U^C = \frac{\left(\theta_1^C - \theta_2^C\right) q_U^C}{2\Theta(\cdot)} \Gamma(\cdot) \tag{83}$$[34]

with[34]

$$\Gamma(\cdot) = -\left(1 - \theta_1^C + \theta_2^C\right)\left(\alpha^C\right)^2 - \phi^E\left(2 - \theta_1^C + \theta_2^C\right)\alpha^C - \left(\phi^E\right)^2 + \left(2 - \theta_1^C + \theta_2^C\right)\left(\theta_1^C + \theta_2^C + 1\right) > 0$$

Accordingly, we deduce that

$$p_{2D}^C > p_U^C \tag{84}$$

We easily verify that

$$p_{1D}^C - p_{2D}^C = \frac{\left(\theta_2^C - \theta_1^C\right)}{2} q_D^C < 0 \tag{85}$$

At last, inequalities (82), (84) and (85) imply that.

$$p_{1D}^C < p_U^C < p_{2D}^C \tag{86}$$

- **Demand comparison**

We show that

$$q_D^E - q_U^E = \frac{\left(\theta_1^C - \theta_2^C\right)^2\left(\phi^E + \alpha^C\right)}{2\Theta(\cdot)} q_U^C > 0 \tag{87}$$

[33] The equation $\Psi(\cdot) = 0$ has two α^C-roots which are $\alpha_0^C < 0$ and $\alpha_1^C > 1 - \phi^E$.
[34] The equation $\Gamma(\cdot) = 0$ has two α^C-roots which are $\alpha_2^C < 1 - \phi^E$ and $\alpha_3^C < 0$.

$$q_{iD}^{C}-q_{iU}^{C}=\frac{\left(\theta_{i}^{C}-\theta_{j}^{C}\right)q_{U}^{C}}{2\Theta(\cdot)}\lambda_{i}(\cdot) \tag{88}$$

with $\lambda_{i}(\cdot)=3\theta_{i}^{C}+\theta_{j}^{C}-\left(\theta_{i}^{C}-\theta_{j}^{C}\right)^{2}-\left(\phi^{E}+\alpha^{C}\right)^{2}>0$ given the model's assumptions (for $i, j=1, 2$ and $i\neq j$). We straightforwardly check that $\lambda_1(\cdot)>\lambda_2(\cdot)$.

Equation (88) implies that:

$$q_{1U}^{C}<q_{1D}^{C}\text{ and }q_{2D}^{C}<q_{2U}^{C} \tag{89}$$

Using (87) and (88), we verify that

$$sign\left(\left(q_{1D}^{C}-q_{1U}^{C}\right)-\left(q_{2U}^{C}-q_{2D}^{C}\right)\right)=sign\left(\lambda_{1}(\cdot)-\lambda_{2}(\cdot)\right)$$

and therefore

$$\left(q_{1D}^{C}-q_{1U}^{C}\right)>\left(q_{2U}^{C}-q_{2D}^{C}\right)$$

It is also important to mention that:

$$q_{D}^{C}-q_{U}^{C}=\frac{\left(\theta_{1}^{C}-\theta_{2}^{C}\right)^{2}}{\Theta(\cdot)}q_{U}^{C}>0 \tag{90}$$

■

Proof of Proposition 3

- **Consumer surplus comparison**

It is straightforward to check that

$$sign\left(CS_{D}^{E}-CS_{U}^{E}\right)=sign\left(q_{D}^{E}-q_{U}^{E}\right)$$

and therefore (based on (87))

$$CS_{D}^{E}>CS_{U}^{E} \tag{91}$$

Furthermore, we learn from (53), (72) and (89) that:

$$CS_{1U}^{C}>CS_{1D}^{C}\text{ and }CS_{2D}^{C}<CS_{2U}^{C} \tag{92}$$

The total consumer surplus under regime s, $s=U, D$

$$CS_{s}=CS_{s}^{E}+CS_{1s}^{C}+CS_{2s}^{C}$$

Equations (46), (52) and (53) allow us to show that

$$CS_U = \frac{\left(q_U^E\right)^2}{2} + \frac{\left(\theta_1^C+1\right)\left(1-\theta_1^C+\theta_2^C\right)^2+\left(\theta_2^C+1\right)\left(1-\theta_2^C+\theta_1^C\right)^2}{8}\left(q_U^C\right)^2 \quad (93)$$

In addition, Eqs. (65), (71) and (72) imply that

$$CS_D = \frac{\left(q_D^E\right)^2}{2} + \frac{\left(\theta_1^C+1\right)\left(2-\theta_1^C+\theta_2^C\right)^2+\left(\theta_2^C+1\right)\left(2-\theta_2^C+\theta_1^C\right)^2}{32}\left(q_D^C\right)^2 \quad (94)$$

Using Eqs. (93) and (94) and rearranging the terms, we prove that

$$\operatorname{sign}(CS_D-CS_U) = \operatorname{sign}\left(\frac{16\left(\left(q_D^E\right)^2-\left(q_U^E\right)^2\right)+\left(\left(\theta_1^C+1\right)\left(2-\theta_1^C+\theta_2^C\right)^2+\left(\theta_2^C+1\right)\left(2-\theta_2^C+\theta_1^C\right)\right.}{4\left(\left(\theta_1^C+1\right)\left(1-\theta_1^C+\theta_2^C\right)^2+\left(\theta_2^C+1\right)\left(1-\theta_2^C+\theta_1^C\right)^2\right)}\right.$$

which is equivalent to

$$\begin{aligned} sign(CS_D-CS_U) &= sign\left(\sqrt{\frac{16\frac{\left(\left(q_D^E\right)^2-\left(q_U^E\right)^2\right)}{\left(q_D^C\right)^2}+\left(\left(\theta_1^C+1\right)\left(2-\theta_1^C+\theta_2^C\right)^2+\left(\theta_2^C+1\right)\left(2-\theta_2^C+\theta_1^C\right)^2\right)}{4\left(\left(\theta_1^C+1\right)\left(1-\theta_1^C+\theta_2^C\right)^2+\left(\theta_2^C+1\right)\left(1-\theta_2^C+\theta_1^C\right)^2\right)}}\left(\frac{q_D^C}{\Lambda}\right)-\frac{q_U^C}{\Lambda}\right) \\ &= sign\left(\sqrt{\frac{16\left(\frac{\left(q_D^E\right)^2-\left(q_U^E\right)^2}{\Lambda^2 y_D^2}\right)+\left(\left(1+\theta_1^C\right)\left(2-\theta_1^C+\theta_2^C\right)^2+\left(1+\theta_2^C\right)\left(2-\theta_2^C+\theta_1^C\right)^2\right)}{4\left(1+\theta_1^C\right)\left(1-\theta_1^C+\theta_2^C\right)^2+\left(1+\theta_2^C\right)\left(1-\theta_2^C+\theta_1^C\right)^2}}\cdot y_D-y_U\right) \\ &= sign\left(\widehat{y}_U-y_U\right) \end{aligned} \quad (95)$$

where

$$\widehat{y}_U = \sqrt{\frac{16\left(\frac{\left(q_D^E\right)^2-\left(q_U^E\right)^2}{\Lambda^2 y_D^2}\right)+\left(\left(1+\theta_1^C\right)\left(2-\theta_1^C+\theta_2^C\right)^2+\left(1+\theta_2^C\right)\left(2-\theta_2^C+\theta_1^C\right)^2\right)}{4\left(1+\theta_1^C\right)\left(1-\theta_1^C+\theta_2^C\right)^2+\left(1+\theta_2^C\right)\left(1-\theta_2^C+\theta_1^C\right)^2}}\cdot y_D$$

is a threshold of the platform congestion level under the net-neutrality regime.

We learn from (90) that

$$y_U \in]0, y_D[$$

Indeed, we should compare $\widehat{y}_U$ with y_D. We demonstrate that

$$\begin{aligned} sign\left(\widehat{y}_U - y_U\right) &= sign\left(\sqrt{\frac{16\left(\frac{\left(q_D^E\right)^2-\left(q_U^E\right)^2}{\Lambda^2 y_D^2}\right)+\left(\left(1+\theta_1^C\right)\left(2-\theta_1^C+\theta_2^C\right)^2+\left(1+\theta_2^C\right)\left(2-\theta_2^C+\theta_1^C\right)^2\right)}{4\left(1+\theta_1^C\right)\left(1-\theta_1^C+\theta_2^C\right)^2+\left(1+\theta_2^C\right)\left(1-\theta_2^C+\theta_1^C\right)^2}}\cdot y_D - y_D\right) \\ &= sign\left(\frac{16}{\Lambda^2}\frac{\left(\left(q_D^E\right)^2-\left(q_U^E\right)^2\right)}{y_D^2} - \left(3\left(\theta_1^C+\theta_2^C\right)+2\right)\left(\theta_1^C-\theta_2^C\right)^2\right) \\ &= sign\left(\frac{4}{\left(\theta_1^C-\theta_2^C\right)\Lambda}\sqrt{\frac{\left(q_D^E\right)^2-\left(q_U^E\right)^2}{3\left(\theta_1^C+\theta_2^C\right)+2}} - y_D\right) \\ &= sign\left(\widehat{y}_D - y_D\right) \end{aligned} \tag{96}$$

with

$\widehat{y}_D = \frac{4}{\left(\theta_1^C-\theta_2^C\right)\Lambda}\sqrt{\frac{\left(q_D^E\right)^2-\left(q_U^E\right)^2}{3\left(\theta_1^C+\theta_2^C\right)+2}} > 0$ is a threshold of the platform equilibrium congestion level under non-net-neutrality.

Combining the analyses of (95) and (96), we straightforwardly deduce the following conditions:

$$CS_D \geq CS_U \text{ for } \begin{cases} y_D \in \left]0, \widehat{y}_D\right] \text{ and } y_U \in]0, y_D[\\ \text{or} \\ y_D \in \left]\widehat{y}_D, +\infty\right[\text{ and } y_U \in \left]0, \widehat{y}_U\right] \end{cases} \tag{97}$$

and

$$CS_D < CS_U \text{ for } y_D \in \left]\widehat{y}_D, +\infty\right[\text{ and } y_U \in \left]\widehat{y}_U, y_D\right[\tag{98}$$

These are conditions (37) and (38) stated in proposition 3.

- **Profit comparison**

Using (51) and (70), we show that

$$\Pi_D - \Pi_U = \frac{\left(\theta_1^C-\theta_2^C\right)^2}{8} q_D^C q_U^C > 0$$

and therefore

$$\Pi_D > \Pi_U \tag{99}$$

■

Proof of Proposition 4

- **Welfare comparison**

Under conditions (97) and (99), we easily check that:

$$W_D > W_U$$

However, under condition (98) and (99), we show that

$$\begin{aligned}\operatorname{sign}(W_D - W_U) &= \operatorname{sign}(\Pi_D - \Pi_U) - (CS_U - CS_D)\\ &= \operatorname{sign}(\varsigma - \rho)\end{aligned}$$

with

$\varsigma = \Pi_D - \Pi_U > 0$ is defined as the *profit-increase* effect.
and

$\rho = CS_U - CS_D > 0$ is defined as the *surplus-reduction* effect.

Under (98), (99) and $\varsigma > \rho$, we deduce that

$$W_D > W_U$$

On the other hand, under (98), (99) and $\varsigma \leq \rho$, we show that

$$W_D \leq W_U \qquad \blacksquare$$

Appendix 2

Numerical Simulations

Adopting some parameter values, we derive Fig. 2 which shows that non-net-neutrality is preferable from a social point of view until reaching a certain threshold of the marginal congestion cost of type 1 denoted $\widetilde{\theta}_1^C = 0.985$. This latter implies a platform equilibrium congestion level under each regime that satisfies condition (38); the condition under which net neutrality can be welfare-superior.

Using arbitrary values of θ_1^C, in addition to $\widetilde{\theta}_1^C$, we validate, in Table 1, the shape of the curve in Fig. 2 and the results stated in proposition 4.

- **At point A**: For $\theta_1^C = 0.901 < \widetilde{\theta}_1^C$, a value that maximizes the differential $W_D - W_U$, we check that the thresholds of the equilibrium congestion level under non-net-neutrality and net neutrality are respectively given by $\widehat{y}_D$ (which is close to 0.094) and $\widehat{y}_U$ (which is close to 0.411). These thresholds satisfy condition (80). In this case, it is easy to verify that $\varsigma > \rho$ and therefore $W_D > W_U$;

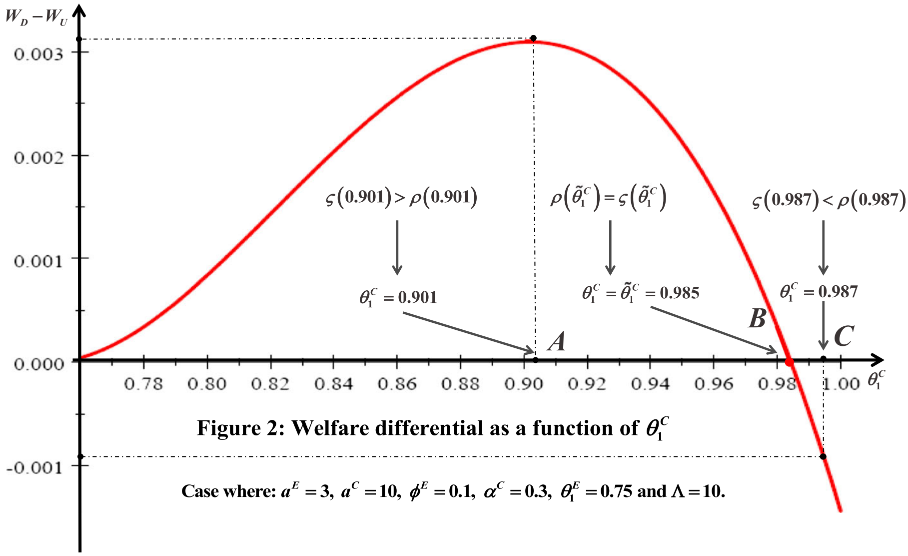

Fig. 2 Welfare differential as a function of θ_1^C

- **At point B**: For $\theta_1^C = \widetilde{\theta}_1^C$, we verify that the thresholds of the equilibrium congestion level under the two regimes are given by $\widehat{y}_D$ (which is close to 0.087) and $\widehat{y}_U$ (which is close to 0.396). They also satisfy condition (38). In this case, we can check that $\rho = \varsigma$ and therefore $W_D = W_U$;
- **At point C**: For $\theta_1^C = 0.987 > \widetilde{\theta}_1^C$, we verify that the thresholds of the equilibrium congestion level are $\widehat{y}_D$ (which is close to 0.887) and $\widehat{y}_U$ (which is close to 0.395) and satisfy condition (38). Exceeding $\widetilde{\theta}_1^C$, we show that $\varsigma < \rho$ and therefore $W_D < W_U$.

Below the threshold $\widetilde{\theta}_1^C$, it is useful to leave the market self-regulation mechanisms which lead to a superiority of the non-net-neutrality regime for all the concerned parties. Beyond the threshold, the net neutrality regime becomes necessary and therefore the intervention of a regulatory agency so as to impose it is strongly required. We

Table 1 Numerical illustrations

	A	B	C
θ_1^C	0.901	0.985	0.987
y_s			
y_D	0.41313	0.40126	0.40107
y_U	0.41221	0.39917	0.39895
$\hat{y}_D$	0.09412	0.08793	0.88736
$\hat{y}_U$	0.41101	0.39625	0.39599
ς	0.04885	0.111	0.11234
ρ	0.04575	0.111	0.11257

can also learn from Fig. 2 that the society is better off under net neutrality in the case where the marginal congestion cost of subgroup-1 is sufficiently large. This justifies the dominance of the congestion effect over the network-benefit effect.

References

Akiene PTK, Kabari LG (2015) Optimization of data packet transmission in a congested network. International Journal of Computer Networks and Security 25:1383–1389

Alexandrov A, Deb J (2012) Price discrimination and investment incentives. Int J Ind Organ 30:615–623

Aloui C, Jebsi K (2016) Platform optimal capacity sharing: willing to pay more does not guarantee a larger capacity share. Econ Model 54:276–288

Armstrong M (2006) Competition in two-sided markets. RAND J Econ 37:668–691

Basso LJ, Zhang A (2007) Congestible facility rivalry in vertical structures. J Urban Econ 61:218–237

Bourreau M, Kourandi F, Valletti T (2015) Net neutrality with competing internet platforms. J Ind Econ 63: 30–73

Brueckner JK (2002) Airport congestion when carriers have market power. Am Econ Rev 92:1357–1375

Caillaud B, Jullien B (2003) Chicken and egg: competition among intermediation service providers. RAND J Econ 34:306–328

Cheng HK, Bandyopadhyay S, Guo H (2011) The debate on net neutrality: a policy perspective. Inf Syst Res 22:60–82

Choi JP, Kim BC (2010) Net neutrality and investment incentives. RAND J Econ 41:446–471

Daniel JI, Harback KT (2008) (When) do hub airlines internalize their self-imposed congestion delays? J Urban Econ 63:583–612

Dixit A (1979) A model of duopoly suggesting a theory of entry barriers. Bell Journal of Economics, the RAND Corporation 10:20–32

Economides N (2005) The economics of the internet backbone. In: Vogelsang I (ed) Handbook of telecommunications. Elsevier Publishers, Amsterdam

Economides N (2007) The economics of the internet, the new Palgrave dictionary of economics. Macmillan, London

Economides N, Hermalin B (2012) The economics of network neutrality. RAND J Econ 43:602–629

Economides N, Tåg J (2012) Net neutrality on the internet: a two-sided market analysis. Inf Econ Policy 24: 91–104

Krämer J, Wiewiorra L (2012) Network neutrality and congestion sensitive content providers: implications for content variety, broadband investment and regulation. Inf Syst Res 23:1303–1321

Krämer J, Wiewiorra L, Weinhardt C (2013) Net neutrality: a progress report. Telecommun Policy 37:794–813

Lambertini L, Orsini R (2010) R&D for quality improvement and network externalities. Netw Spat Econ 10: 113–124

Layson SK (1994) Market opening under third-degree price discrimination. J Ind Econ 42:335–340

Layson SK (1998) Third-degree price discrimination with interdependent demands. J Ind Econ 46:511–524

Peitz M, Schuett F (2016) Net neutrality and inflation traffic. Int J Ind Organ 46:16–62

Reggiani C, Valletti T (2016) Net neutrality and innovation at the core and at the edge. Int J Ind Organ 45:16–27

Reitman D (1991) Endogenous quality differentiation in congested markets. J Ind Econ 39:621–647

Rochet JC, Tirole J (2003) Platform competition in two-sided markets. J Eur Econ Assoc 1:990–1029

Rochet JC, Tirole J (2006) Two-sided markets: a progress report. RAND J Econ 37:645–667

Rosen R (2005) Auctions in a two-sided network: the market for meal voucher service. Netw Spat Econ 5: 339–350

Rysman M (2009) The economics of two-sided markets. J Econ Perspect 23:125–143

Smirnov M, Crowcroft J (2003) Quality of future internet services. Springer, Berlin

Tirole J (1988) The theory of industrial organization. The MIT Press, Cambridge

Weyl G (2010) A price theory of multi-sided platforms. Am Econ Rev 100:1642–1672