Network Neutrality in the Internet with Congestion Sensitive End Users and Content Providers

Özgür Kaplanlıoğlu[1]

Sencer Ecer[2]

Abstract

In this paper, we envision a two-sided market mediated by a monopolistic internet service provider, ISP. The ISP provides end-users internet access and carries content providers' (CPs') data packages on its network. We compare the case where network neutrality is strictly practiced with the case where the ISP can "throttle" the traffic of certain content providers. In the model, for simplicity, a single CP is exposed to throttling, while the other CPs, which are part of a continuum, are not. We then study the implications of the violation of network neutrality on total data consumption, congestion, and capacity investment. We show that under discrimination, the ISP charges a lower price to end-users. Paradoxically, this lower price leads to lower data volume in some cases because while the lower price for end-users has an increasing effect on demand for data, lower capacity investment causes more congestion and decreases the demand for data.

JEL Codes: L960, L90

Keywords: Network neutrality, Two-sided markets, Telecommunication

[1]Department of Economics, Istanbul Technical University, 34367, Istanbul, Turkey. E-mail: ozgur_kaplanlioglu@mckinsey.com, ORCID: 0000-0001-6771-2529

[2]Department of Economics, Istanbul Technical University, 34367, Istanbul, Turkey. E-mail: ecer@itu.edu.tr, ORCID: 0000-0003-2956-2990

Boğaziçi Journal Review of Social, Economic and Administrative Studies, Vol. 36, no. 2 (2022), pp. 40-79, doi:10.21773/boun.36.2.1 Research Article Received: 12.10.2022 - Accepted: 15.12.2022

Yoğunluğa Duyarlı Son Kullanıcılar ve İçerik Sağlayıcılar ile İnternette Ağ Tarafsızlığı

Özet

Bu makalede, monopol bir internet servis sağlayıcı (ISP) tarafından aracılık yapılan iki-taraflı bir piyasa tasarlanmıştır. ISP son kullanıcılara internet erişimi sağlamakta ve içerik sağlayıcılarının veri paketlerini kendi ağında taşımaktadır. Çalışmada ağ tarafsızlığının katı bir şekilde uygulandığı durumla, ISP'nın içerik sağlayıcılarının (CP) trafiğini "kısabildiği" durum karşılaştırılmıştır. Modelimizde, basitliği sağlamak adına sadece bir adet kısma uygulamasına maruz bırakılmış sürekliliğin bir parçası olan diğerleri serbest bırakılmıştır. Ağ tarafsızlığı ilkesinden sapılmasının toplam veri tüketimi, ağ yoğunluğu ve kapasite yatırımının denge büyüklükleri üzerindeki etkileri incelenmiştir. Bulgularımıza göre, ağ tarafsızlığından sapıldığı durumda son ISP kullanıcılardan istediği fiyatı düşürmektedir. Paradoksal olarak bu düşük fiyat bazı durumlarda veri için olan talebi artırmaktadır, çünkü düşük kapasite yatırımı yoğunluğu artırmakta bu da talebi düşürmektedir.

JEL Sınıflandırması: L960, L90

Anahtar Kelimeler: Ağ tarafsızlığı, İki-taraflı piyasalar, Telekomünikasyon

1. Introduction

The data consumed through a few social media applications and video streaming applications like Facebook, Instagram, YouTube, and Netflix constitute a greater percentage of all the data carried in networks throughout the world. What would it be like if the ISPs were able to charge a price to these OTTs (Over The Tops) to use their infrastructure? How would consumer welfare be affected if these OTTs accepted to pay the price? Such questions are entirely related to the current network neutrality debate.

Network Neutrality reflects the philosophy that the internet is built over. It is what makes the internet open and unbiased for everybody. However, its economic value and rigor have become questionable in the last decade or so. Some of the reasons include applications and services that flow through the internet becoming more and more data-intensive (such as video streaming services, social media services); excessive congestion problems due to these demanding applications; and increases in the cost of capacity investment.

The debate on whether network neutrality should be relaxed has quickly made itself into the regulator's agenda in the US and the EU. In 2005, the US FCC (Federal Communications Commission) changed the Internet's status from "telecommunication services" to "information services." With this change, the ISPs are no longer exposed to the explicit network neutrality constraints (Economides, 2015b). Following this change, major ISPs and network owners in the US started discriminating on the upstream side of the market. For instance, in 2007, it turned out that Comcast had throttled Bit-Torrent traffic. Comcast did not deny and defended itself, arguing it was just "reasonable network management," which resulted in a dispute between the FCC and Comcast and ended up in court. Similar conflicts occurred between large ISPs, Netflix, and Google a great deal of which ended up with an agreement between the CP (content provider) and the ISP (Greenstein et al., 2016).

From an economics perspective, there is no consensus on the definition of network neutrality. Therefore, it is best to identify the situations that can be considered a clear violation of network neutrality and are of economic interest. As stated before, the network neutrality principle requires all content to be treated equally regardless of its origin. In this respect, two types of violations come forward. "Throttling" of certain content by ISPs (this could be in the form of blocking access entirely or slowing it down), counts as a clear deviation from network neutrality. A deviation of this type has a significant economic interest (as practice of the third-degree price discrimination situation). Likewise, ISP's division of its bandwidth into predetermined partitions for exclusive usage by content providers also counts as a clear violation of network neutrality. Such a tiered service consisting of "slow lanes" and "fast lanes" depicts a second-degree price discrimination setting and again has an economic interest (Economides, 2015b).

Existing studies analyze different aspects of the above-mentioned modes of violation of network neutrality. Proponents of network neutrality claim that it guarantees the openness of the internet. Without network neutrality, large and financially powerful content providers (like Google, Amazon, Netflix, Yahoo, etc.) dominate the internet. This would hinder innovation on the edge and prevent new startups with innovative and revolutionary ideas from coming into existence (Economides, 2015a). On the other hand, opponents argue that with the abolition of the principle, ISPs and Network Operators will internalize a part of content providers' surplus and consequently be incentivized to invest in capacity infrastructure and reduce the prices for the end-user. Below, we review some of the papers that are relevant to our study.

Choi et al. (2010) consider a broadband network with a monopolistic ISP. They build a model where the internet consists of two symmetric CPs. There is a continuum of end-users with a unit mass, two content providers, and a monopolistic ISP. Market shares of content providers are determined by the classical Hotelling approach (Hotelling, 1929). End-users are homogeneous in the sense that each has the same demand for content. Content providers can only charge the advertisers. The ISP, a monopoly, could discriminate via prioritization. In other words, one CP gets the "high speed" service level and the other one gets the "best-effort" service level depending on their willingness to pay for prioritized service. The magnitude of the two profit margins and the difference between them critically affects the equilibrium outcome of the model. For instance, the ISP doesn't choose discrimination against network neutrality unless both margins are above a certain threshold. Likewise, discrimination is socially superior to network neutrality only if relative magnitudes of margins are large enough. Choi et al. (2010) show under discrimination that the network access fee is unambiguously lower than it is under network neutrality. (Although we utilize a linear pricing scheme instead of a constant network access fee for end-users; in our model, we also show that the price for end-users is strictly lower under discrimination.) In Choi et al.'s (2010) model, the ISP's incentives to practice the discriminatory regime depends on its bargaining power and the magnitudes of the CPs' profit margins. Our model shows that the ISP always prefers to discriminate if the discriminated CP's business is profitable. However, our model differs from Choi et al.'s (2010) in the sense that no competition is assumed between CPs and the ISP charges a fee to the discriminated CP based on unit bandwidth rather than offering a prioritized service. In Choi et al.'s (2010) model the non-prioritized content provider is always worse off under discrimination in terms of profit, whereas the prioritized content provider may be better off under discrimination if relative magnitudes of CPs margins are large enough and ISP's bargaining power is low. As a result, the overall effect of discrimination on social welfare depends on the mentioned factors and is ambiguous. Choi et al. (2010) do not explicitly derive the optimal capacity for the two regimes. Instead, they compare relative incentives to invest in network neutrality and discrimination. The conditions where investment incentive is higher for the discriminatory regime than the network neutrality regime is found to be ambiguous. Our model, on the other hand, suggests that under a discriminatory regime the monopolistic ISP always short supply the bandwidth capacity.

In a monopolistic residential broadband internet market with end-users and CPs, Economides and Hermalin (2012) compare network neutrality with discrimination in terms of either practicing tiered service or charging an access fee to CPs. Their model differs from Choi et al.'s (2010). Unlike Choi et al. (2010), they don't assume constant data traffic. The consumption decision is given endogenously by homogeneous households (end-users). In the model, the more congestion end-users are exposed to, or the more expensive the content is, the less content is consumed. Unlike Choi and Kim, Economides and Hermalin (2012) allow CPs to directly sell their content to end-users. This adds a second source of revenue to their revenue stream. In our model, though, we didn't allow such trade between end-users and CPs in order to focus on the ISP's mediation of the two-sided market.

Economides and Hermalin (2012) analyze the problem in three parts. First, in a static setting (i.e., taking bandwidth constant), they present the conditions for optimal bandwidth allocation which maximizes total welfare. Then, they allow the ISP to charge the CPs. Lastly, they switch to a dynamic setting and try to understand the investment incentives of the monopolistic ISP. The principal result is that the total welfare is proportional to the amount of content (data) being carried on the network. In other words, discrimination is welfare superior to network neutrality only if it causes more content to be consumed by end-users. Another result is that blocking access to a positive measure of CPs necessarily reduces the total welfare. They also show, if the elasticity of demand against transmission time (i.e., congestion) is monotone in content type, then there could be bandwidth allocations where discrimination is welfare superior to network neutrality. When it comes to the ISP's incentives to charge the CPs, it is shown that such pricing schemes are welfare inferior. However, it is also shown that when ISPs charge the CPs network access fees are smaller on the end-user's side. Economides and Hermalin (2012) reveal that if the adjustment function is multiplicatively separable in delay time (congestion) and content type, then the ISP will always prefer discrimination vis-a-vis network neutrality. To summarize, when the bandwidth is taken as constant (static setting), there would be an inefficiency caused by preferring a tiering regime (discrimination) vis-a-vis network neutrality. Yet, Economides and Hermalin (2012) also find that, under a dynamic setting, ISP will always install more bandwidth in discrimination vis-à-vis network neutrality, which enhances total welfare. Considering these two opposing effects, under a dynamic setting, the overall effect of deviating from network neutrality on total welfare is ambiguous.

Krämer and Wiewiorra (2012) study short-run and long-run effects of discrimination on content variety, welfare, and capacity investment. Their model of discrimination is practicing two service tiers to CPs as priority class and best-effort class. They found that under discrimination, content variety may or may not be higher than it is under neutrality. They also show that the equilibrium content variety is closely related to how CPs are distributed. If the CPs are distributed uniformly, then content variety does not change, and prioritization does not affect content variety. If congestion insensitive CPs are larger in number (CP distribution is left-skewed), then prioritization causes more CPs to be active in the equilibrium. When CP distribution is right-skewed, the prioritization causes fewer CPs to be active in equilibrium compared to network neutrality in the short run. Sticking to the assumption that CPs are uniformly distributed, the ISP always prefers prioritization to neutrality because of higher profits. Even though the CPs are worse off under discrimination, it is shown that total welfare is unambiguously higher since the end-users and the ISP are each better off. Kramer and Wiewiorra (2012) also investigate the investment incentives of the ISP in the long run. They reveal that under the assumption of uniform distribution of CPs, the ISP's investment incentive is higher. However, for non-uniform distributions of CPs, the result is ambiguous.

Bourreau et al. (2015) consider a similar model to Krämer and Wiewiorra's. The major distinction is that Bourreau et al. (2015) model a duopoly setting where the two ISPs compete for market share in a residential broadband internet market. They claim that lack of competition leads the ISP to prefer discrimination against network neutrality. In such a setting, discrimination leads to various distortions that reduce total welfare. In a duopoly setting, on the other hand, stiff competition between the ISPs softens the welfare distortion caused by the monopolistic ISP and might even reverse it. For this reason, they study discrimination in a duopolistic market. Bourreau et al. (2015) compare the equilibrium under network neutrality and discrimination. In both cases, the equilibrium is the solution of a two-stage game. Under network neutrality, at the second stage, end-users and CPs decide which ISP to join (End-users are single-homing. CPs are multi-homing). At the first stage, the ISPs determine the subscription fees for the end-users and also network capacities. Likewise, under discrimination, at the second stage, end-users and CPs decide which

ISP to join and, CPs decide also whether to pay for the prioritization. At the first stage, ISPs decide subscription fees for the end-users, network capacities and additionally priority fees for the CPs. One of the major results is that content variety and capacity investment are unambiguously higher in discrimination than in neutrality. Average congestion depends on two factors: total traffic and capacity. As claimed, both total traffic and capacity are higher in discrimination. Therefore, whether the average congestion is lower or higher in discrimination depends on which of the two effects dominates. Bourreau et al. show that the capacity effect always dominates and conclude that congestion would be lower under discrimination. Providing a priority service in exchange for a certain fee and increasing the capacity becomes profitable in Bourreau et al. (2015) because of the infinitely many potential entrants. In our model, on the other hand, the number of CPs are fixed, and they are all in the market. The model in our paper better fits non-revolutionary changes in capacity such as investing in the capacity with the existing technology. Their paper's assumption of infinitely many CPs where the potential entrants are highly congestion sensitive may be more relevant in the very long-run. When it comes to the subscription fee for the end-users, ambiguity arises. Their model doesn't lead to lower network access fees in all cases. Their assumption that leads to this result is that content variety (the number of CPs that are in the market) is a component of end-users' utility function. If end-users value content variety sufficiently highly, then they are willing to pay a higher price. Otherwise, the price is lower than the neutrality case in their paper. Specifically, if end-users' preferences for speed and variety is sufficiently high and if advertising rate is sufficiently low, the network access fee is higher under discrimination than under network neutrality. Similar ambiguity exists for ISPs' profits and CPs' profits too. Under certain conditions, network neutrality is more profitable for ISPs. Unless network capacity under discrimination is sufficiently greater than it is under network neutrality, discrimination is not profitable for CPs. Despite the ambiguities in the network access fee, end-user surpluses, and end-user profits, Bourreau et al. (2015) show that total welfare is unambiguously higher in discrimination. As mentioned, discrimination is not always profitable for ISPs. However, they show that competition drives the two ISPs to discriminate even if it is not profitable. In other words, competition may drag ISPs into a "prisoner's dilemma" type of a situation.

The novel contribution of our paper is that we consider a fundamentally different mode of violating network neutrality. In the existing literature, prioritization and blocking access are widely studied. In our study, on the other hand, we define a unit price for bandwidth, and the share of bandwidth bought by the CP is imposed as a bandwidth cap for that CP. Imposing such a cap is already a common practice implemented by mobile network operators. For example, many mobile network operators these days cap YouTube with 480p quality at peak times. A typical network access fee option does not fully address the tradeoff between keeping a CP in the market and avoiding too much congestion, whereas the setting proposed in our model both keeps YouTube in the network and eases the congestion problem it creates. Also, a partitioning of the bandwidth to mutually exclusive shares is not an efficient way of addressing congestion. For these reasons and in line with current practices, we adopt our setting with a cap on bandwidth.

We organize the paper as follows. In the Model section, we explain how the model specifies End Users, ISP, and CPs in detail. In the following two sections, we derive the equilibrium for the network neutrality and the discrimination cases respectively. How discrimination affects the equilibrium outcomes on different aspects such as data volume, congestion, etc. is discussed in the Network Neutrality vs Discrimination section. Concluding remarks are presented in the final section.

2. Model

We study a monopoly Internet Service Provider (ISP) who mediates a two-sided internet market consisting of end-users and content providers (CPs).[3] We'll discuss each of these market players one by one. We then discuss the equilibrium properties under two different cases. First, when network neutrality is strictly practiced, and second, when the ISP violates network neutrality by discriminating a CP.

End-users

In this paper, CPs are differentiated, and end-users have different congestion sensitivities for CPs where θ represents an arbitrary CP. CPs are websites and web applications like YouTube, Facebook, Gmail, cnn.com *etc*. End-users pay a unit price p_x to the ISP for their data consumption and they reduce their data consumption under congestion in proportion to their congestion sensitivity to a given CP. We define data demand through an arbitrary CP θ as follows.

$$x_\theta = x_\theta^0 - d_\theta \left(\frac{x^0}{B}\right) \quad where; x^0 = \sum_\theta x_\theta^0 \tag{1}$$

In the above expression, x_θ^0 is the aggregate data demand for the CP θ under zero-congestion whereas, x_θ is the data demand for the CP θ under nonzero-congestion. We define the congestion as the ratio of total data demand under zero-congestion (x^0) to available bandwidth (B). In the expression, the parameter d_θ denotes the end-users' congestion sensitivity to CP θ. The higher d_θ is, the more data consumption reduces due to the congestion. In defining congestion, $\left(\frac{x^0}{B}\right)$, we consider using the total data demand under zero-congestion across all CPs considering that the network bandwidth (B) is a shared resource and end-users experience the same congestion but with different congestion sensitivities to different CPs. The congestion causes disutility to end-users, which makes data demand under zero-congestion to always be greater than nonzero-congestion ($x_\theta \leq x_\theta^0$).

Next, we spell out the demand functions under network neutrality and under discrimination.

Under Network Neutrality

Data Demand for an Arbitrary CP (θ). We consider a linear demand function under zero-congestion which effectively leads to the following demand function for the CP θ under nonzero-congestion.

---

[3] We study a monopolistic ISP because most broadband internet service providers have high market power, for example, in the US and Europe. Also, even if the CPs may have to have relationships with several ISPs, these ISPs likely have terminating monopolies, so "it is still insightful to investigate the relationship between CPs and a single ISP, particularly if that ISP is thought to be large. For example, it would certainly have a substantial impact on CPs' business model if they would not have access to customers' on AT&T's network (Kramer and Wiewiorra, 2012)."

$$x_\theta(p_x, B) = (a_\theta - b_\theta\, p_x) - d_\theta \left[\frac{(a - b\, p_x)}{B}\right] \tag{2}$$

In the above demand specification, p_x is the unit price of data whereas a and b are shape parameters such that $a = \sum_\theta a_\theta$ and $b = \sum_\theta b_\theta$ and the congestion is rewritten as $\left[\frac{(a-b\,p_x)}{B}\right]$. Note that $x_\theta^0 = (a_\theta - b_\theta\, p_x)$ and $x^0 = (a - b\, p_x)$.

We assume that the elasticity of demand with respect to all non-price variables is constant, which leads demand changes to be rotational in these variables as shown by Graves and Sexton (2006). With this assumption, we particularly have in mind the bandwidth variable that we explicitly model. Having constant elasticity for this variable is reasonable for the average household that we are modeling. To satisfy the constant elasticity property, we assume that $\frac{a_\theta}{b_{\theta,}}$ is constant for all θs.[4]

Considering user-time is a limited resource, we also assume that $a < \bar{a}$ where $\bar{a}$ is finite and positive.

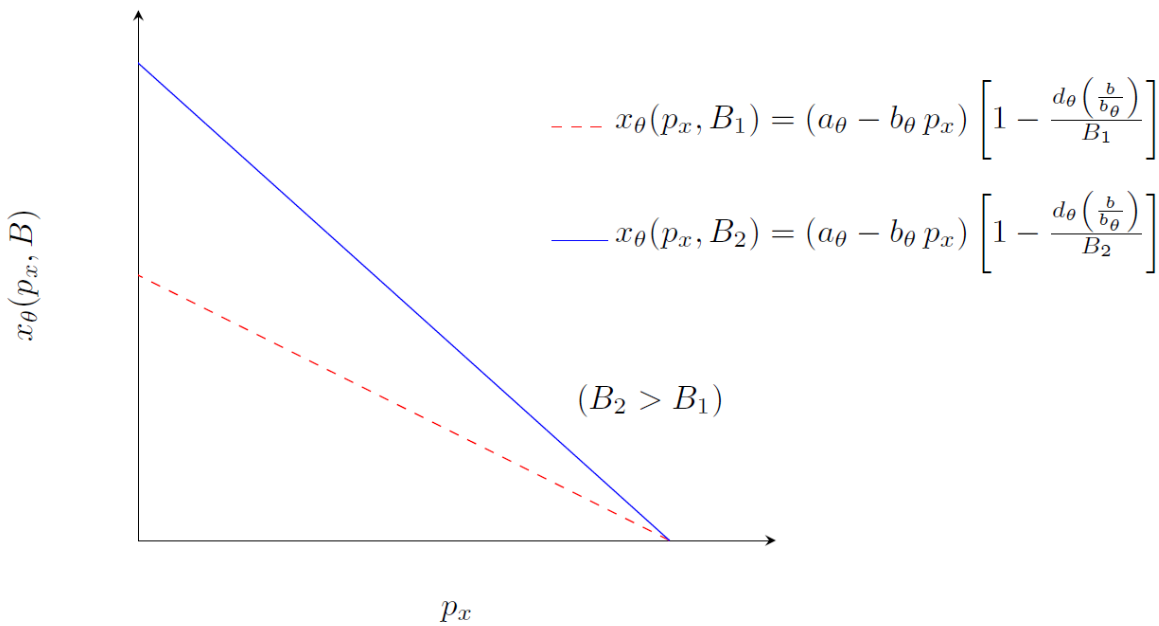
Rearranging (2) and using constant elasticity assumption with respect to the bandwidth we have the following expression for demand.

$$x_\theta(p_x, B) = (a_\theta - b_\theta\, p_x)\left[1 - \frac{d_\theta\left(\frac{b}{b_\theta}\right)}{B}\right] \tag{3}$$

As seen in (3), the congestion (through the bandwidth) has a counterclockwise rotational effect on the demand (Figure 1).

[4] We can readily show that the elasticity is indeed constant. The elasticity is given by $-\frac{dx_\theta}{dB}\frac{B}{x_\theta} = \frac{-\frac{d_\theta b}{B}\left(\frac{a}{b}-p_x\right)}{b_\theta\left(\frac{a_\theta}{b_\theta}-p_x\right)-\frac{d_\theta b}{B}\left(\frac{a}{b}-p_x\right)}$. Assuming $\frac{a_\theta}{b_\theta}$ is constant across all θs, the elasticity expression becomes $-\frac{dx_\theta}{dB}\frac{B}{x_\theta} = \frac{-\frac{d_\theta b}{B}}{b_\theta-\frac{d_\theta b}{B}}$ which is constant across all p_x

Figure 1



Total Data Demand. Then, the aggregate data demand under network neutrality is given by,

$$\sum_\theta x_\theta\,(p_x, B) = x(p_x, B) = (a - b\,p_x) - d\,\frac{(a - b\,p_x)}{B} \tag{4}$$

where $x = \sum_\theta x_\theta$, $a = \sum_\theta a_\theta$, $b = \sum_\theta b_\theta$ and $d = \sum_\theta d_\theta$

Under Discrimination

Under discrimination we assume that a single CP is discriminated by the ISP. We denote the discriminated CP as Θ in the rest of this paper, whereas the collection of all the remaining CPs is denoted as Θ'. Possible candidates for the discriminated CP could be large OTTs like YouTube, Instagram, Facebook, Netflix *etc.* The discriminated CP pays a unit price for the bandwidth to the ISP. We denote this price as p_B. For instance, if the discriminated CP chooses to buy B_Θ of bandwidth, it needs to pay $B_\Theta p_B$ to the ISP where $B_\Theta \leq B$ by definition. However, in our model B_Θ is not for the exclusive usage of the discriminated CP. It solely acts as an upper bandwidth limit. Then, the discriminated CP is capped by B_Θ, but the remaining CPs can also use B_Θ when the network is overly utilized. This setting is more realistic than partitioning the network into mutually exclusive subsets. In this respect, p_B can be thought as a network access fee rather than a unit price for a commodity (bandwidth).

Data Demand for the Discriminated CP (Θ). Under discrimination the demand for CP Θ is given by

$$x_\Theta(p_x, B, p_B) = (a_\Theta - b_\Theta\,p_x) - d_\Theta\left[\frac{(a_\Theta - b_\Theta\,p_x) + \frac{B_\Theta}{B}(a_{\Theta'} - b_{\Theta'}\,p_x)}{B_\Theta}\right] \tag{5}$$

where $(a_{\Theta'} = a - a_{\Theta})$, $(b_{\Theta'} = b - b_{\Theta})$, and $(d_{\Theta'} = d - d_{\Theta})$. As can be seen, under discrimination, congestion is defined as the term in brackets above:

$$\left[\frac{(a_{\Theta} - b_{\Theta}\, p_x) + \frac{B_{\Theta}}{B}(a_{\Theta'} - b_{\Theta'}\, p_x)}{B_{\Theta}}\right]$$

The intuition behind this definition is as follows. Since B_{Θ} is not for the exclusive usage of the CP Θ (it only acts as an upper bandwidth limit for Θ) there will always be instances where data consumption from the remaining CPs causes congestion on the CP Θ. As B gets larger relative to B_{Θ} this effect weakens (as if CP Θ and the remaining CPs, Θ', flow through two seperate networks) and the congestion term converges to the following.

$$\left[\frac{(a_{\Theta} - b_{\Theta}\, p_x) + \frac{B_{\Theta}}{B}(a_{\Theta'} - b_{\Theta'}\, p_x)}{B_{\Theta}}\right] \to \left[\frac{(a_{\Theta} - b_{\Theta}\, p_x)}{B_{\Theta}}\right]$$

So, it is as if the purchased part of the bandwidth is exclusive to Θ. Conversely, as B_{Θ} gets closer to B, congestion expression converges to what it is under neutrality since $(a_{\Theta} - b_{\Theta}\, p_x) + (a_{\Theta'} - b_{\Theta'}\, p_x) = (a - b\, p_x)$ (Figure 2b).

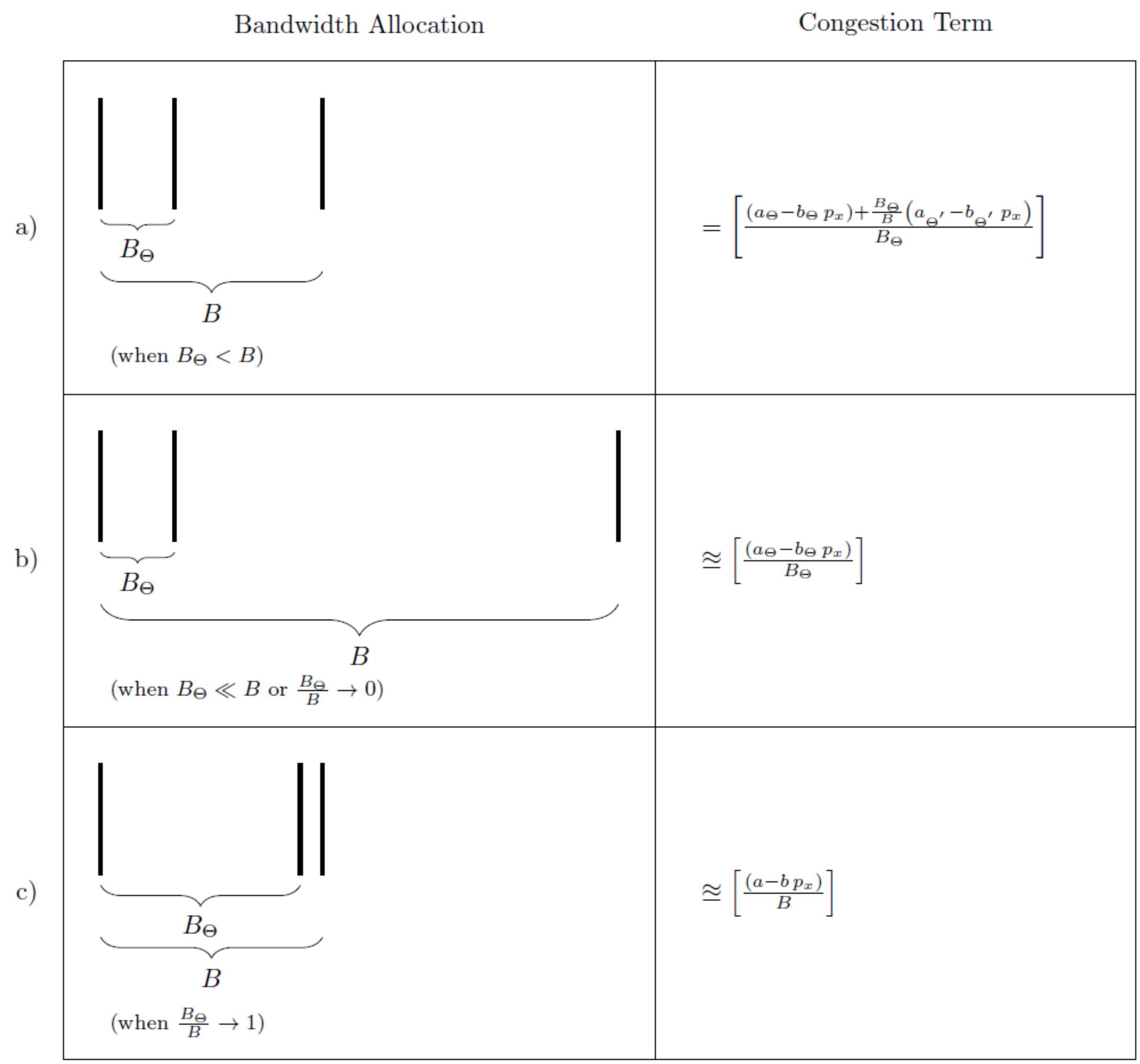
$$\left[\frac{(a_{\Theta} - b_{\Theta}\, p_x) + \frac{B_{\Theta}}{B}(a_{\Theta'} - b_{\Theta'}\, p_x)}{B_{\Theta}}\right] \to \left[\frac{(a - b\, p_x)}{B}\right]$$

We summarize the above ideas in Figure 2. It is insightful to compare CP Θ's demand under discrimination with the demand under neutrality. Rearranging (5), we have,

$$x_{\Theta}(p_x, B, p_B) = (a_{\Theta} - b_{\Theta}\, p_x) - d_{\Theta}\left[\frac{a - b\, p_x}{B}\right] - d_{\Theta}\left[\frac{\left(\frac{B}{B_{\Theta}} - 1\right)(a_{\Theta} - b_{\Theta}\, p_x)}{B}\right] \qquad (6)$$

Comparing (2) and (6), the first two terms in the CP Θ's demand function are the same as the ones in the demand function under neutrality. The third term reflects the effect of the discrimination. If $B_{\Theta} < B$, all else the same, the demand under discrimination is always less than the demand under neutrality. When $B_{\Theta} = B$, (6) converges to (2).

Figure 2



<u>Data Demand for the Remaining CPs (Θ').</u> Under discrimination, the overall data demand for the remaining CPs is given by

$$x_{\Theta'}(p_x, B, p_B) = (a_{\Theta'} - b_{\Theta'}\, p_x) - d_{\Theta'}\left[\frac{(a_{\Theta'} - b_{\Theta'}\, p_x) + \frac{B_\Theta}{B}(a_\Theta - b_\Theta\, p_x)}{B}\right] \tag{7}$$

where the congestion is defined as,

$$\left[\frac{(a_{\Theta'} - b_{\Theta'}\, p_x) + \frac{B_\Theta}{B}(a_\Theta - b_\Theta\, p_x)}{B}\right]$$

Like the previous case, as B becomes greater relative to B_{Θ} it is as if all bandwidth is exclusively used by Θ'.

$$\left[\frac{(a_{\Theta'} - b_{\Theta'}\, p_x) + \frac{B_{\Theta}}{B}(a_{\Theta} - b_{\Theta}\, p_x)}{B}\right] \rightarrow \left[\frac{(a_{\Theta'} - b_{\Theta'}\, p_x)}{B}\right]$$

As B_{Θ} gets closer to B, it is as if network neutrality is imposed (Figure 3b).

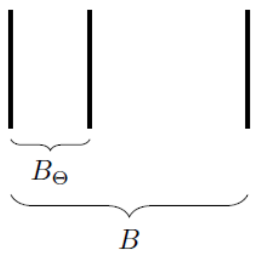
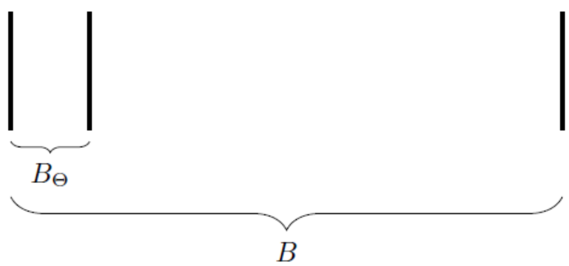
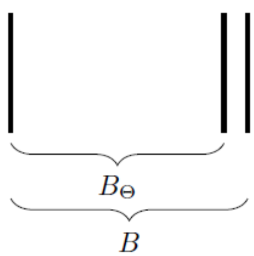
$$\left[\frac{(a_{\Theta'} - b_{\Theta'}\, p_x) + \frac{B_{\Theta}}{B}(a_{\Theta} - b_{\Theta}\, p_x)}{B}\right] \rightarrow \left[\frac{(a - b\, p_x)}{B}\right]$$

We can also compare the remaining CPs' demand under discrimination with the demand under neutrality. Rearranging (7), we have,

$$x_{\Theta'}(p_x, B, p_B) = (a_{\Theta'} - b_{\Theta'}\, p_x) - d_{\Theta'}\left(\frac{a - b\, p_x}{B}\right) + d_{\Theta'}\left[\frac{\left(1 - \frac{B_{\Theta}}{B}\right)(a_{\Theta} - b_{\Theta}\, p_x)}{B}\right] \tag{8}$$

Comparing (2) and (8), the first two terms reflect the situation under neutrality. The third term, on the other hand, is positive and reflects the effect of discrimination. If $B_{\Theta} < B$, all else the same, under discrimination, the demand for remaining contents Θ' is always greater than it is under neutrality. The reason is that when $B_{\Theta} < B$, less congestion is experienced by the remaining CPs.

Figure 3

	Bandwidth Allocation	Congestion Term
a)	 (when $B_\Theta < B$)	$= \left[\frac{(a_{\Theta'} - b_{\Theta'}\, p_x) + \frac{B_\Theta}{B}(a_\Theta - b_\Theta\, p_x)}{B}\right]$
b)	 (when $B_\Theta \ll B$ or $\frac{B_\Theta}{B} \to 0$)	$\approx \left[\frac{(a_{\Theta'} - b_{\Theta'}\, p_x)}{B}\right]$
c)	 (when $\frac{B_\Theta}{B} \to 1$)	$\approx \left[\frac{(a - b\, p_x)}{B}\right]$

<u>Total Data Demand $(\Theta + \Theta')$.</u> Combining (6) and (8), the total data demand under discrimination is given by,

$$x(p_x, B, p_B) = \left(1 - \frac{d}{B}\right)(a - b\,p_x) - \left[\frac{d_\Theta\left(\frac{B}{B_\Theta} - 1\right) - d_{\Theta'}\left(1 - \frac{B_\Theta}{B}\right)}{B}\right](a_\Theta - b_\Theta\,p_x) \qquad (9)$$

where $(a_\Theta - b_\Theta\,p_x) + (a_{\Theta'} - b_{\Theta'}\,p_x) = (a - b\,p_x)$, and $d_\Theta + d_{\Theta'} = d$.

Henceforth we assume that $d_\Theta > d_{\Theta'}$. That is, the demand for the discriminated CP is so sensitive that the congestion parameter is higher than the sum of the congestion parameters for the rest of the market. This assumption could also be considered as the criterion for third-degree price discrimination, which is intuitive in the sense that a CP is willing to pay for bandwidth only if its demand is highly sensitive to congestion, *e.g.*, big OTTs like YouTube, Instagram, Facebook and Netflix.

Comparing (4) and (9), the overall data demand is always less than the overall demand under neutrality whenever $B_\Theta < B$ (since $d_\Theta > d_{\Theta'}$). The decrease in the demand for content Θ compensates for the increase in the demand for contents Θ'. Thus, the overall effect is negative.

Rotational Demand Property Under Discrimination. Under neutrality, the assumption of constant elasticity in bandwidth ensures that the demand is always rotational with a common maximum willingness to pay. By the same assumption, the rotational demand property is still valid under discrimination as shown below.

Rearranging (5)

$$x_\Theta(p_x, B, p_B) = (a_\Theta - b_\Theta\, p_x)\left[1 - \frac{d_\Theta\left(\frac{b_\Theta + b_{\Theta'}\left(\frac{B_\Theta}{B}\right)}{b_\Theta}\right)}{B_\Theta}\right] \tag{10}$$

Rearranging (7)

$$x_{\Theta'}(p_x, B, p_B) = (a_{\Theta'} - b_{\Theta'}\, p_x)\left[1 - \frac{d_{\Theta'}\left(\frac{b_{\Theta'} + b_\Theta\left(\frac{B_\Theta}{B}\right)}{b_{\Theta'}}\right)}{B}\right] \tag{11}$$

The demand functions specified above in (10) and (11) are counterparts of (3) under discrimination. It is easy to see that as $B_\Theta \to B$, (10) and (11) are both converge to (3).

3. Content Providers

The CPs are different in size and in congestion sensitivity. The CPs' revenues come from advertisements, as they don't directly trade with end-users. We assume that CPs' advertisement revenue depends on the total data that is consumed through them. Therefore, there is a rate r_θ, for each CP representing the advertisement revenue per unit data. The CPs are harmed by congestion because it leads to a decrease in the data consumed through them, and, consequently, a decrease in their advertisement revenue. In this manner, the CPs have and incentive to purchase network bandwidth.

Under Network Neutrality

Profit of an Arbitrary CP (θ). Under network neutrality, an arbitrary CP θ profits from advertisement at a unit profit rate of r_θ, which leads to the following profit function.

$$\Pi_\theta = x_\theta(p_x, B) r_\theta$$

Substituting $x_\theta(p_x, B)$ in (2) we have,

$$\Pi_\theta = \left[(a_\theta - b_\theta\, p_x) - d_\theta \frac{(a - b\, p_x)}{B}\right] r_\theta \tag{12}$$

where r_θ is advertisement revenue per unit data.

Under Discrimination

Discriminated CP's (Θ) Bandwidth Demand. Under discrimination, the discriminated CP, Θ, must pay p_B for unit bandwidth. The demand for bandwidth (B_Θ) is determined by solving its profit maximization problem, where the demand is determined by using (5).

$$\hat{x}(p_x, B, p_B) = (a_\Theta - b_\Theta\, p_x) - d_\Theta \left[\frac{(a_\Theta - b_\Theta\, p_x) + \frac{\hat{B}}{B}(a_{\Theta'} - b_{\Theta'}\, p_x)}{\hat{B}}\right]$$

Then, the profit maximization problem is set up as,

$$\begin{aligned} \max_{\hat{B}} \quad & \hat{\Pi} = \hat{x}(p_x, B, p_B) r_\Theta - \hat{B}\, p_B \\ \text{s.t.} \quad & \hat{B} \leq B, \\ & \hat{B} \geq 0 \end{aligned} \tag{13}$$

where B_Θ solves $\max_{\hat{B}} \hat{\Pi}$. We define the profit under optimality as Π_Θ.

$$\frac{\partial \hat{\Pi}}{\partial \hat{B}} = \lambda_1$$
$(\lambda_1 \geq 0$ and $B_\Theta = B)$
$$\frac{\partial \hat{\Pi}}{\partial \hat{B}} = -\lambda_2$$
$(\lambda_2 \geq 0$ and $B_\Theta = 0)$
$$\frac{\partial \hat{\Pi}}{\partial \hat{B}} = 0$$
$(0 < B_\Theta < B)$

Then,

$$B_\Theta = \sqrt{\frac{d_\Theta}{p_B + \lambda_1}(a_\Theta - b_\Theta\, p_x) r_\Theta}$$
$(\lambda_1 \geq 0$ and $B_\Theta = B)$
$$B_\Theta = \sqrt{\frac{d_\Theta}{p_B - \lambda_2}(a_\Theta - b_\Theta\, p_x) r_\Theta} \tag{14}$$
$(\lambda_2 \geq 0$ and $B_\Theta = 0)$
$$B_\Theta = \sqrt{\frac{d_\Theta}{p_B}(a_\Theta - b_\Theta\, p_x) r_\Theta}$$
$(0 < B_\Theta < B)$

Profit of the Discriminated CP (Θ). Like in the network neutrality case, under discrimination, CP Θ profits from the unit data consumed through it at the rate of r_Θ. Unlike network neutrality, it has to pay p_B to the ISP for the unit bandwidth that it uses, which leads to the following profit.

$$\Pi_\Theta = x_\Theta(p_x, B, p_B) r_\Theta - B_\Theta\, p_B$$

Substituting $x_\Theta(p_x, B, p_B)$ in (5) we have,

$$\Pi_\Theta = \left[(a_\Theta - b_\Theta\, p_x) - d_\Theta\left(\frac{a - b\, p_x}{B}\right)\right] r_\Theta - d_\Theta\left[\frac{\left(\frac{B}{B_\Theta} - 1\right)(a_\Theta - b_\Theta\, p_x)}{B}\right] r_\Theta - B_\Theta\, p_B \tag{15}$$

where B_Θ is the bandwidth purchased by CP Θ. The CPs other than Θ has the same profit as before.

$$\Pi_\theta = x_\theta(p_x, B, p_B) r_\theta$$

4. The Internet Service Provider (ISP)

In this study, the ISP is modeled as a monopoly. It mediates the two-sided market, where end-users constitute one side and CPs the other. Under the network neutrality constraints, the ISP only charges end-users based on unit consumption. Thus, the ISP's pricing is linear. On the other hand, under discrimination, the ISP charges the discriminated CP, Θ for its access to network bandwidth as well. Here also the pricing is linear and per unit bandwidth.

Under Network Neutrality

Under neutrality, the ISP's profit is

$$\Pi_{ISP} = x(p_x, B)(p_x - c_x) - B\, c_B$$

Substituting $x(p_x, B)$ with (4) we have

$$\Pi_{ISP} = \left(1 - \frac{d}{B}\right)(a - b\,p_x)(p_x - c_x) - B\,c_B \tag{16}$$

where $x(p_x, B)$ is the demand for data, p_x is the price charged to end-users per unit data, c_x is ISP's cost of supplying a unit data (constant marginal cost is assumed), B is the network bandwidth and c_B is the cost of installing unit bandwidth.

Under Discrimination

Under discrimination, the ISP charges both sides of the market, and it gets the following profit.

$$\Pi_{ISP} = x(p_x, B, p_B)(p_x - c_x) + B_\Theta\,p_B - B\,c_B$$

Substituting $x(p_x, B, p_B)$ in (9) we have,

$$\begin{aligned}\Pi_{ISP} = \; & \left(1 - \frac{d}{B}\right)(a - b\,p_x)(p_x - c_x) \\ & - \left[\frac{d_\Theta\left(\frac{B}{B_\Theta} - 1\right) - d_{\Theta'}\left(1 - \frac{B_\Theta}{B}\right)}{B}\right](a_\Theta - b_\Theta\,p_x)(p_x - c_x) \\ & - B\,c_B \\ & + B_\Theta\,p_B\end{aligned} \tag{17}$$

where p_B is the price charged to CP Θ for bandwidth and B_Θ is the bandwidth demanded by Θ.

5. Equilibrium Under Network Neutrality

We first study the equilibrium under network neutrality, which is the subgame perfect Nash equilibrium (SPNE) of the following two-stage game.

1. The ISP determines the total bandwidth to install, B^*, and the data price for end-users, p_x^*.
2. End-users decide how much data to consume, $x(p_x^*, B^*)$. Since CPs do not make any decision under network neutrality, this stage is a trivial stage.

<u>Stage 1.</u> In the first stage, the ISP chooses the optimal values for p_x and B by solving the following profit maximization problem using the profit function defined in (16).

$$\max_{p_x,B} \quad \Pi_{ISP} = x(p_x, B)(p_x - c_x) - B\, c_B$$

where $x(p_x, B)$ is the end-users' best response which comes from the "Stage 2" given p_x and B. The first order conditions of the problem stated above is the following.

$$\frac{\partial \Pi_{ISP}}{\partial p_x} \leq 0 \quad \text{(with equality when } p_x^* > 0)$$
$$\frac{\partial \Pi_{ISP}}{\partial B} \leq 0 \quad \text{(with equality when } B^* > 0)$$

For interior solution we have[5]

$$\frac{\partial \Pi_{ISP}}{\partial p_x} = 0 = \frac{\partial x(p_x^*, B^*)}{\partial p_x}(p_x^* - c_x) + x(p_x^*, B^*)$$
$$\frac{\partial \Pi_{ISP}}{\partial B} = 0 = \frac{\partial x(p_x^*, B^*)}{\partial B}(p_x^* - c_x) - c_B$$

We solve the equilibrium and obtain the following equilibrium values.[6]

Proposition 1 (Equilibrium Quantities Under Neutrality). *Under network neutrality, the equilibrium price, p^* and the bandwidth, B^* are the following,*

$$p_x^* = \frac{1}{2}\left(\frac{a}{b} + c_x\right) \tag{18}$$

$$B^* = \sqrt{\frac{d}{c_B}(a - b\, p_x^*)(p_x^* - c_x)} \tag{19}$$

The monopoly price p_x^* is not related to the equilibrium bandwidth B^* or aggregate congestion sensitivity d or congestion. This is because these terms are added to the demand function in a multiplicative manner. Recall that congestion has a rotational effect on the demand.

---

[5] The second order conditions are shown to be satisfied in the appendix.

[6] See the Appendix for the proof.

We know that when demand is linear (i.e., is in the form of $k_0 - k_1\, p$) and the marginal cost is constant at c, then equilibrium monopoly price is given by $\frac{1}{2}\left(\frac{k_0}{k_1} + c\right)$. In our case, the demand is $\left(1 - \frac{d}{B}\right)(a - b\, p_x)$. Thus, the equilibrium monopoly price does not depend on $\left(1 - \frac{d}{B}\right)$.

Let us consider capacity as a quality property. Note that the pricing decision does not depend on capacity when congestion has only rotational effect on the data demand. Then, using the seminal result by Spence (1975), given any $x(p_x, B)$ the profit maximizer ISP oversupplies the capacity (quality) to be able to sell the data at a higher price. However, from the regulator's perspective (in which we assume that the regulator is welfare maximizer), we see that the ISP under-supplies the capacity: the welfare maximizing capacity is given by $\sqrt{\frac{d}{c_B}(a - b\, p_x^*)\left(\frac{1}{2}\left(\frac{a}{b} - p_x^*\right) + p_x^* - c_x\right)} < B^*$.

We also study some interesting comparative statics. The monopoly invests less in bandwidth, B^*, as the unit cost of capacity, c_B, increases, and it invests more in bandwidth as the aggregate congestion sensitivity, d, increases. One may define $(a - b\, p_x^*)(p_x^* - c_x)$ as the maximum potential profit of the ISP for a given price because it reflects the zero-congestion situation. In this respect, the monopoly invests more in bandwidth if the potential profit from eliminating congestion is relatively high.

6. Equilibrium Under Discrimination

We now study the equilibrium under discrimination. The equilibrium is the subgame perfect Nash equilibrium (SPNE) of the following two-stage game.

1. The ISP determines the total bandwidth to install, B^{**}, the unit price of data for end-users, p_x^{**}, and the unit price of bandwidth for the discriminated CP, p_B^{**},.
2. End-users decide how much data to consume, $x(p_x^{**}, B^{**}, p_B^{**})$ and the discriminated CP chooses how much bandwidth to purchase, B_Θ^{**}

Stage 2. As in the neutrality case, the decision of the end-users, who just respond based on their aggregate data demand as described in (9), is trivial. The discriminated CP, on the other hand, decides based on its profit maximization problem as defined in (13), that is,

$$\max_{\hat{B}} \quad \hat{\Pi} = \hat{x}(p_x, B, p_B) r_\Theta - \hat{B}\, p_B$$
$$\text{s.t.} \quad \hat{B} \leq B,$$
$$\hat{B} \geq 0$$

The solution to this problem gives us the discriminated CP's demand for bandwidth (B_Θ)

$$B_\Theta = \sqrt{\frac{d_\Theta}{p_B + \lambda_1}(a_\Theta - b_\Theta\, p_x) r_\Theta}$$
$$(\lambda_1 \geq 0 \text{ and } B_\Theta = B)$$
$$B_\Theta = \sqrt{\frac{d_\Theta}{p_B - \lambda_2}(a_\Theta - b_\Theta\, p_x) r_\Theta}$$
$$(\lambda_2 \geq 0 \text{ and } B_\Theta = 0)$$
$$B_\Theta = \sqrt{\frac{d_\Theta}{p_B}(a_\Theta - b_\Theta\, p_x) r_\Theta}$$
$$(0 < B_\Theta < B)$$

Stage 1. In the first stage, the ISP chooses the optimal values for p_x, B and p_B by solving its profit maximization problem (17). The ISP takes the end-users' and the discriminated CP's best responses, which come from Stage 2.

$$\max_{p_x, B, p_B} \quad \Pi_{ISP} = x(p_x, B, p_B)(p_x - c_x) + B_\Theta\, p_B - B\, c_B$$

The first order conditions of the problem stated above is the following.

$$\frac{\partial \Pi_{ISP}}{\partial p_x} \leq 0 \quad \text{(with equality when } p_x^{**} > 0\text{)}$$
$$\frac{\partial \Pi_{ISP}}{\partial B} \leq 0 \quad \text{(with equality when } B^{**} > 0\text{)}$$
$$\frac{\partial \Pi_{ISP}}{\partial p_B} \leq 0 \quad \text{(with equality when } p_B^{**} > 0\text{)}$$

For interior solution we have,

$$\frac{\partial \Pi_{ISP}}{\partial p_x} = 0 = \quad \frac{\partial x(p_x^{**}, B^{**}, p_B^{**})}{\partial p_x}(p_x^{**} - c_x) + x(p_x^{**}, B^{**}, p_B^{**}) + p_B^{**}\frac{\partial B_\Theta(p_B^{**}, p_x^{**})}{\partial p_x}$$

$$\frac{\partial \Pi_{ISP}}{\partial B} = 0 = \quad \frac{\partial x(p_x^{**}, B^{**}, p_B^{**})}{\partial B}(p_x^{**} - c_x) - c_B$$

$$\frac{\partial \Pi_{ISP}}{\partial p_B} = 0 = \quad \frac{\partial x(p_x^{**}, B^{**}, p_B^{**})}{\partial p_B}(p_x^{**} - c_x) + B_\Theta^{**} + p_B^{**}\frac{\partial B_\Theta(p_B^{**}, p_x^{**})}{\partial p_B}$$

As far as the discriminated CP is concerned, this game has three possible outcomes. First, the ISP could set the unit bandwidth price so high that the discriminated CP does not choose to buy any bandwidth at all, and it is effectively out of the market. Second, the ISP sets the unit price of bandwidth to zero or very close to zero, which effectively will lead to neutrality. And lastly, the ISP sets an intermediate price, and the discriminated CP buys a share of the available bandwidth. We next cover these possible outcomes.

Proposition 2 (Blocking Access Entirely). *Under discrimination, the discriminated CP always buys a positive amount of bandwidth, ($B_\Theta^{**} > 0$) if* $\left[1 - \frac{d_\Theta\left(\frac{b}{b_\Theta}\right)}{B} - \frac{d_{\Theta'}}{B}\right] > 0$ *holds.*

Thus, the ISP sets a price p_B that does not push the discriminated CP out of the market. If the ISP pushes the discriminated CP out of the market by setting p_B too high, the end-users' data demand is affected in two ways. On one hand, since the discriminated CP is not available, the overall data demand tends to decrease. On the other hand, the data demand tends to increase due to the lower congestion caused by the exclusion of the discriminated CP. However, the condition $\left[1 - \frac{d_\Theta\left(\frac{b}{b_\Theta}\right)}{B} - \frac{d_{\Theta'}}{B}\right] > 0$ guarantees that the negative effect of the exclusion dominates the positive effect of the lower congestion. Consequently, the exclusion of the discriminated CP reduces the total data demand, which would not be profitable for the ISP. Thus, in equilibrium, the ISP keeps the discriminated CP in the market.[7]

We next investigate the possibility of the second outcome in the next proposition.

Proposition 3 (Incentives to Deviate from Network Neutrality). *The ISP strictly prefers to deviate from network neutrality ($p_B^{**} > 0$).*

Under discrimination, the ISP has two revenue sources. First, it earns revenue from the data consumed by end-users, $(x(p_x, B, p_B)\, p_x)$. Second, it earns revenue from the bandwidth purchased by the discriminated CP, $(B_\Theta\, p_B)$. If the ISP sticks with network neutrality and set $p_B = 0$, then it would lose the second revenue stream. Since the bandwidth for sale is capped by B and, the

---

[7] See the Appendix for the proof.

discriminated CP's bandwidth demand is discontinuous in p_B, the ISP can still set the $p_B > 0$ and make a profitable deviation. Thus, it prefers to abandon network neutrality.[8]

By Proposition 2 and Proposition 3 we know that we have an interior solution. We solve the equilibrium for interior solution and obtain the following equilibrium values.[9]

Proposition 4 (Equilibrium Quantities Under Discrimination). *Under discrimination, the equilibrium unit data price, p_x^{**}, the bandwidth, B^{**} and the unit bandwidth price, B_Θ^{**}, are as follows.*

$$p_x^{**} = \frac{1}{2}\left(\frac{a}{b} + c_x\right) - \frac{1}{2}\left(\frac{d_\Theta\, b_\Theta\, r_\Theta}{b(B^{**} - d) - b_\Theta\left[d_\Theta\left(\frac{B^{**}}{B_\Theta^{**}} - 1\right) - d_{\Theta'}\left(1 - \frac{B_\Theta^{**}}{B^{**}}\right)\right]}\right)\frac{B^{**}}{B_\Theta^{**}} \tag{20}$$

$$B^{**} = \sqrt{\frac{d}{c_B}(a - b\, p_x^{**})(p_x^{**} - c_x) - \frac{\left[d_\Theta - d_{\Theta'}\left(2\frac{B_\Theta^{**}}{B^{**}} - 1\right)\right]}{c_B}(a_\Theta - b_\Theta\, p_x^{**})(p_x^{**} - c_x)} \tag{21}$$

$$p_B^{**} = \frac{d_\Theta}{(B_\Theta^{**})^2}(a_\Theta - b_\Theta\, p_x^{**})r_\Theta \tag{22}$$

Recall that, under network neutrality, the equilibrium price p_x^* is not related to the congestion sensitivity d and the equilibrium bandwidth B^*. Unlike network neutrality, under discrimination, the equilibrium price p_x^{**} is a function of the congestion sensitivity of the discriminated CP, the aggregate congestion sensitivity of the remaining CPs, the total aggregate congestion sensitivity, the two equilibrium bandwidths, and the profit margin of the discriminated CP. This difference is due to the two-sidedness of the market.

In two-sided markets, the platform allocates the total price between market participants based on cross-group elasticities. This allocation usually leads to a market equilibrium where one side is priced above the marginal cost, whereas the other side is priced below the marginal cost (Rochet and Tirole, 2003). The following proposition, together with Proposition 4, displays the effect of two-sidedness on the equilibrium quantities.

Proposition 5 (Two Sidedness). *Under discrimination, the share of total bandwidth that the discriminated CP decides to purchase is given by following implicit expression,*[10]

---

[8] See the Appendix for the proof.

[9] See the Appendix for the proof.

[10] See the Appendix for the proof.

$$\frac{d_{\Theta}}{d_{\Theta'}}\left(1-\frac{r_{\Theta}}{p_x^{**}-c_x}\right)=\left(\frac{B_{\Theta}^{**}}{B^{**}}\right)^2$$

Proposition 5 reflects the equilibrium condition for the interior solution that, all else the same, greater profit margin on the end-user's side, $(p_x^{**}-c)$, implies a greater share of bandwidth, $(\frac{B_{\Theta}^{**}}{B^{**}})$, is purchased by the discriminated CP. In other words, this condition is another demonstration of the classical two-sided market outcome (Rochet and Tirole, 2003): in two-sided markets one side of the market has a higher profit margin whereas the other side is the loss leader. In our case, when p_x^{**} is high, then $(\frac{B_{\Theta}^{**}}{B^{**}})$ is high and accordingly p_B^{**} is low. Therefore, high end-user price, p_x^{**}, only comes with a low bandwidth price, p_B^{**}, and vice versa.

It is worth mentioning that Proposition 5 only applies for sufficiently low values of r_{Θ}. The right-hand side (RHS) of the equation is bounded by 1. We assume that $\frac{d_{\Theta}}{d_{\Theta'}}$ is strictly greater than 1. Then, $\left(1-\frac{r_{\Theta}}{p_x^{**}-c_x}\right)$ must be less than 1, and consequently p_x^{**} must be greater than (c_x+r_{Θ}). By Proposition 4, we show $p<\frac{1}{2}\left(\frac{a}{b}+c_x\right)$ for interior solution. So, if r_{Θ} is, for instance, greater than $(\frac{1}{2}\left(\frac{a}{b}-c_x\right))$, then the interior solution would not be possible. Indeed, advertisement rates on unit data is usually far less than operator profits on unit data consumption.

In Proposition 4, we derive the implicit expressions for p_x^{**}, B^{**}, p_B^{**} and B_{Θ}^{**}. Therefore, by merely using the Proposition 5's result it is quite difficult to understand the effect of a marginal increase in d_{Θ} on the equilibrium outcome. Fortunately, it is possible to study the effect of marginal changes in the cost of unit bandwidth investment (c_B) and the advertisement rate (r_{Θ}) on the equilibrium outcome (assuming the discriminated CP has zero marginal cost per unit data consumed through it; and hence r_{Θ} is equivalent to profit margin of the discriminated CP as well.).

Using the results presented in Proposition 4 and Proposition 5, it is easy to see that when bandwidth investment becomes less expensive (c_B), the ISP installs more bandwidth (B). This makes bandwidth less scarce, resulting in a decrease in the bandwidth price (p_B) and an increase in the data price (p_x). The lower price for the bandwidth (p_B) makes the discriminated CP purchase even more bandwidth (B_{Θ}). The growth in the data price (p_x) increases the profit margin of the ISP on the end users' side, consequently by Proposition 5, the share of bandwidth purchased by the discriminated CP increases.

Using Propositions 4 and 5, when the advertisement rate (r_{Θ}) rises, the monopolistic ISP charges more for bandwidth (p_B) and installs less bandwidth (B) to make capacity even scarcer. As a result, the data price (p_x) decreases to maintain the demand for the data and hence the revenues. The high bandwidth price (p_B) makes the discriminated CP purchase a lower share of the total installed capacity. This result is interesting because as the discriminated CP's business becomes profitable, the ISP will have much stronger private incentives to make capacity even scarcer.

As mentioned before, as the advertisement rate (r_{Θ}) increases, the data price (p_x) decreases, and bandwidth price (p_B) increases. A greater profit margin on advertisement makes the other side of the market (end-users) more attractive for the discriminated CP. This results in a higher bandwidth price and lowers the data price on the end-user side. How price allocation will be affected is not clear when there is an increase in the congestion sensitivity of the discriminated CP's service (d_Θ). Such an increase makes both sides more attractive for each other. End-users like the discriminated CP's content flow through a wider bandwidth because they are more sensitive to congestion now. Similarly, the discriminated CP is willing to pay more for bandwidth because it prefers to get congestion-sensitive end-users on board. Thus, an increase in the congestion sensitivity of the discriminated CP's service causes an ambiguous effect on the price allocation.

Under network neutrality, we see that the monopoly causes a downward distortion on the supply of bandwidth. Under discrimination, there is an additional distortion caused by two-sidedness: Unconstrained monopoly supplies the bandwidth even lower in proportion to the difference between congestion sensitivity of the discriminated CP's service and the aggregate congestion sensitivity of the remaining CPs.

7. Network Neutrality vs Discrimination

In this section, we compare the equilibrium under network neutrality and discrimination.

Proposition 6. *The ISP charges a lower price to end-users under discrimination than under neutrality* $(p_x^{**} < p_x^{*})$.

Proposition 6 compares the equilibrium price under network neutrality and the equilibrium price under discrimination. Recall that, Proposition 5 is all about how price is allocated between the end-users and the discriminated CP. It does not compare with the case of neutrality. One result of the Proposition 5 is that as the share of bandwidth purchased by the discriminated CP increases, the end-user price increases too. As shown in Proposition 4, the second component in the equilibrium end-user price formula is a factor of $(\frac{B^{**}}{B_{\Theta}^{**}})$. For the discriminated CP to get a larger piece from the overall bandwidth, the ISP lowers the bandwidth price (p_B^{**}) and increases the bandwidth investment (increase B^{**}). Lower p_B^{**} and higher B^{**} dictate higher end-user price. Interestingly, lower price for the end-users can only come up with less bandwidth investment and higher bandwidth price, leading to a smaller share of bandwidth purchased by the discriminated CP. This is the mechanism we reveal at Proposition 5. Moreover, Proposition 6 demonstrates that, regardless of how price is allocated between the end-users and the charged CP, the end-user price is always lower under discrimination than under neutrality.[11]

Proposition 7 (Investment Incentives). *Under discrimination, the ISP invests less in capacity than it does under neutrality* $(B^{**} < B^{*})$.

As mentioned earlier, the total bandwidth (B) is the main instrument for the ISP to set prices for each of the sides of the market. It can make total bandwidth scarcer and bandwidth price (p_B) go up, and similarly, it can make it abundant and end-user price (p_x) go up. The market equilibrium for total bandwidth is determined by factors such as cost of bandwidth investment (c_B), congestion sensitivity of end-users towards the discriminated CP (d_{Θ}) and towards the remaining CPs $(d_{\Theta'})$ and the advertisement rate (r_{Θ}). Proposition 7 demonstrates that under discrimination, total equilibrium bandwidth is always less than the total equilibrium bandwidth under neutrality.[12,13] As stated in the following proposition, the private incentives of the ISP lead to a more congested service for the discriminated CP.

Proposition 8 (Congestion). *Discrimination leads to a more congested service for the discriminated CP,* Θ. *For the remaining CPs, a less congested service is not for certain. Despite the ambiguity in the remaining CPs' congestion, the overall congestion is worse under discrimination.*

Although under discrimination the data price for the end-users is lower, Proposition 8 states that the end-users experience more congestion with the discriminated CP. Moreover, a better

---

[11] See the Appendix for the proof.

[12] See the Appendix for the proof.

[13] The assumption, $d_{\Theta} > d_{\Theta'}$ is sufficient but not necessary for Proposition 7 to hold.

experience is not guaranteed with the remaining CPs in terms of congestion.[14] Whether the overall congestion level would be better off under discrimination is an important question. The assumption ($d_\Theta > d_{\Theta'}$) causes the change in overall congestion to be dominated by the increase in the discriminated CP's congestion.[15]

[14] See the Appendix for the proof.

[15] The assumption, $d_\Theta > d_{\Theta'}$ is sufficient but not necessary for Proposition 8 to hold.

8. Conclusions

In this paper, we study the economic implications of abolishing network neutrality. Although there are many ways and modes of violating network neutrality (prioritization and blocking access are the ones exhaustively studied in the economics literature), we define abolishing network neutrality as setting a positive price for unit bandwidth to a single content provider (the discriminated CP) for network access. In this sense, we depict a two-sided market with third-degree price discrimination.

In our setting, there is a monopoly internet service provider (ISP) mediating a two-sided market between end-users and content providers. ISP's pricing is linear. In other words, it charges end users a price p_x for per unit data consumption. Under neutrality, CPs' access to the ISP's infrastructure is free of charge. Under the discriminatory regime, the ISP charges the discriminated CP a price p_B for per unit bandwidth. For the remaining CPs, the network access is free of charge as before (third-degree price discrimination).

Under network neutrality, because of the market distortion caused by the ISP, the equilibrium end-user price is above the socially optimum price (i.e., the marginal cost). We find that a similar distortion is present for the equilibrium bandwidth. In the absence of competitive pressure, the ISP under-supplies capacity.

Under discrimination (i.e., when the ISP is allowed to charge a single CP for unit bandwidth) due to the two-sidedness of the market, the equilibrium outcomes change. Our first result is that the ISP always prefers to discriminate. Put differently, profit-maximizing conditions always dictate the ISP to set a positive bandwidth price for the discriminated CP. Another result is that the discriminated CP always chooses to purchase a piece of the available bandwidth. In other words, the ISP's pricing strategy is such that it helps the discriminated CP to stay in the market. The discrimination leads to a pricing scheme that would lead the discriminated CP to pay only for a share of the total capacity. Thus, discrimination does not deteriorate content variety, but it causes the throttling of the discriminated CP's content.

We find that under discrimination end-user price is always lower than it is under neutrality. Considering the data demand and the end-user price are being inversely related, one may conclude that abolishing network neutrality increases overall data consumption. However, data demand decreases with congestion. One of the most important results of the paper is that under discrimination due to the private incentives of the ISP, the equilibrium bandwidth would be lower than it is under neutrality. More precisely, the ISP has lower incentives to invest in bandwidth under discrimination. Although a lower end-user price tends to increase the equilibrium data consumption, a lower bandwidth tends to decrease it. Therefore, we can't say anything definitive regarding the overall data consumption. There are certain parameter intervals where the overall data consumption is greater than in network neutrality and the total welfare is higher.

According to the model, under discrimination, end-users experience more congestion on the discriminated CP's content. Moreover, we see that the overall congestion is dominated by the discriminated CP's congestion, and it increases, too.

Considering our results in combination, our study does not support the abolition of network neutrality. Such a policy would lower the end-user price, but it would also hinder infrastructure investments. Moreover, the model shows that congestion would be worse for the discriminated CP and overall congestion would also increase.

References

Bourreau, Marc, Frago Kourandi, and Tommaso Valletti. 2015. "Net Neutrality with Competing Internet Platforms." *The Journal of Industrial Economics* 63 (1): 30–73. https://doi.org/https://doi.org/10.1111/joie.12068.

Economides, Nicholas. 2015a. "Testimony on Network Neutrality to US Congress." Working Papers 15-14. New York University, Leonard N. Stern School of Business, Department of Economics.

Economides, Nicholas. 2015b. "Economic Features of the Internet and Network Neutrality." Working Papers 15-01. NET Institute.

Economides, Nicholas, and Benjamin E. Hermalin. 2012. "The Economics of Network Neutrality." *The RAND Journal of Economics* 43 (4): 602–29. http://www.jstor.org/stable/41723347.

Graves, Philip, and Robert Sexton. 2006. "Demand and Supply Curves: Rotations versus Shifts." *Atlantic Economic Journal* 34 (3): 361–64. https://doi.org/10.1007/s11293-006-9021-2.

Greenstein, Shane, Martin Peitz, and Tommaso Valletti. 2016. "Net Neutrality: A Fast Lane to Understanding the Trade-Offs." *Journal of Economic Perspectives* 30 (2): 127–50. https://doi.org/10.1257/jep.30.2.127.

Hotelling, Harold. 1929. "Stability in Competition." *The Economic Journal* 39 (153): 41–57. http://www.jstor.org/stable/2224214.

Krämer, Jan, and Lukas Wiewiorra. 2012. "Network Neutrality and Congestion Sensitive Content Providers: Implications for Content Variety, Broadband Investment, and Regulation." *Information Systems Research* 23 (4): 1303–21. http://www.jstor.org/stable/42004258.

Pil Choi, Jay, and Byung-Cheol Kim. 2010. "Net Neutrality and Investment Incentives." *The RAND Journal of Economics* 41 (3): 446–71. https://doi.org/https://doi.org/10.1111/j.1756-2171.2010.00107.x.

Rochet, Jean-Charles, and Jean Tirole. 2003. "Platform Competition in Two-Sided Markets." *Journal of the European Economic Association* 1 (4): 990–1029.

Spence, A. Michael. 1975. "Monopoly, Quality, and Regulation." *The Bell Journal of Economics* 6 (2): 417–29. http://0-www.jstor.org.divit.library.itu.edu.tr/stable/3003237.

Appendix

Proofs

Proposition 1

Proof.

$$\frac{\partial \Pi_{ISP}}{\partial p_x} \leq 0 \quad \text{(with equality when } p_x^* > 0\text{)}$$
$$\frac{\partial \Pi_{ISP}}{\partial B} \leq 0 \quad \text{(with equality when } B^* > 0\text{)}$$

Then for interior solution we have,

$$\frac{\partial \Pi_{ISP}}{\partial p_x} = 0 = \frac{\partial x(p_x^*, B^*)}{\partial p_x}(p_x^* - c_x) + x(p_x^*, B^*)$$
$$\frac{\partial \Pi_{ISP}}{\partial B} = 0 = \frac{\partial x(p_x^*, B^*)}{\partial B}(p_x^* - c_x) - c_B$$

Substituting $\frac{\partial x(p_x^*, B^*)}{\partial p_x}$ and $\frac{\partial x(p_x^*, B^*)}{\partial B}$,

$$\frac{\partial x(p_x^*, B^*)}{\partial p_x} = -b\left(1 - \frac{d}{B^*}\right)$$
$$\frac{\partial x(p_x^*, B^*)}{\partial B} = \frac{d}{(B^*)^2}(a - b\, p_x^*)$$

Then solving for p_x^*;

$$\begin{aligned}\frac{\partial \Pi_{ISP}}{\partial p_x} = 0 \quad &= \left(1 - \frac{d}{B^*}\right)[-b(p_x^* - c_x) + (a - b\, p_x^*)] \\ &= [-b(p_x^* - c_x) + (a - b\, p_x^*)] \\ &= [a - 2\, b\, p_x^* + b\, c]\end{aligned}$$
$$p_x^* = \frac{1}{2}\left(\frac{a}{b} + c_x\right)$$

Then solving for B^*;

$$\frac{\partial \Pi_{ISP}}{\partial B} = 0 = \frac{d}{(B^*)^2}(a - b\, p_x^*)(p_x^* - c_x) - c_B$$
$$c_B = \frac{d}{(B^*)^2}(a - b\, p_x^*)(p_x^* - c_x)$$
$$(B^*)^2 = \frac{d}{c_B}(a - b\, p_x^*)(p_x^* - c_x)$$
$$B^* = \sqrt{\frac{d}{c_B}(a - b\, p_x^*)(p_x^* - c_x)}$$

The second order conditions are given below.

$$H_{\Pi_{ISP}(p_x^*,B^*)} = \begin{bmatrix} \frac{\partial^2 x(p_x^*,B^*)}{\partial p_x^2}(p_x^*-c_x)+2\,\frac{\partial x(p_x^*,B^*)}{\partial p_x} & \frac{\partial^2 x(p_x^*,B^*)}{\partial p_x\,\partial B}(p_x^*-c_x)+\frac{\partial x(p_x^*,B^*)}{\partial B} \\ \frac{\partial^2 x(p_x^*,B^*)}{\partial p_x\,\partial B}(p_x^*-c_x)+\frac{\partial x(p_x^*,B^*)}{\partial B} & \frac{\partial^2 x(p_x^*,B^*)}{\partial B^2}(p_x^*-c_x) \end{bmatrix}$$

$$= \begin{bmatrix} -2b\left(1-\frac{d}{B^*}\right) & \frac{d}{{B^*}^2}[a-2\,b\,p_x^*+b\,c] \\ \frac{d}{{B^*}^2}[a-2\,b\,p_x^*+b\,c] & -\frac{2d}{{B^*}^3}(a-b\,p_x^*)(p_x^*-c_x) \end{bmatrix}$$

$$= \begin{bmatrix} -2b\left(1-\frac{d}{B^*}\right) & 0 \\ 0 & -\frac{2d}{{B^*}^3}(a-b\,p_x^*)(p_x^*-c_x) \end{bmatrix}$$

Above hessian is diagonal at optimal. Since each diagonal element is negative the matrix is negative definite, which shows the equilibrium values are strict local maximums. ☐

Proposition 2

<u>*Proof.*</u> Let $B_\Theta^{**}=0$ in equilibrium.

Because the total data demand when $B_\Theta^{**}=0$ is only composed of the other CPs, we have;

$$x(p_x^{**},B^{**},p_B^{**})=\left(1-\frac{d_{\Theta'}}{B^{**}}\right)(a_{\Theta'}-b_{\Theta'}\,p_x^{**})$$

This expression follows from (4). Then, in this case, the profits of the ISP are equal to

$$x(p_x^{**},B^{**},p_B^{**})(p_x^{**}-c_x)-B^{**}\,c_B \qquad (A.1)$$

On the other hand, for any other p_B such that $B_\Theta>0$, the profis of the ISP are,

$$x(p_x^{**},B^{**},p_B)(p_x^{**}-c_x)-B^{**}\,c_B+B_\Theta\,p_B \qquad (A.2)$$

If $B_\Theta^{**}=0$, then we must have the former expression (A.1) to be greater than or equal to the latter expression (A.2). In other words,

$$\begin{aligned} &x(p_x^{**},B^{**},p_B^{**})(p_x^{**}-c_x)-B^{**}\,c_B \geq \\ &x(p_x^{**},B^{**},p_B)(p_x^{**}-c_x)-B^{**}\,c_B+B_\Theta\,p_B \qquad \forall p_B \end{aligned} \qquad (A.3)$$

Now we want to show that (A.3) is impossible.

Recall (9) that,

$$x(p_x^{**},B^{**},p_B)=\left(1-\frac{d}{B^{**}}\right)(a-b\,p_x^{**})-\left[\frac{d_\Theta\left(\frac{B^{**}}{B_\Theta}-1\right)-d_{\Theta'}\left(1-\frac{B_\Theta}{B^{**}}\right)}{B^{**}}\right](a_\Theta-b_\Theta\,p_x^{**})$$

Using $\left[1-\frac{d_{\Theta}\left(\frac{b}{b_{\Theta}}\right)}{B}-\frac{d_{\Theta'}}{B}\right]>0$ we have,

$$\left(1-\frac{d}{B^{**}}\right)(a-b\,p_x^{**})>\left(1-\frac{d_{\Theta'}}{B^{**}}\right)(a_{\Theta'}-b_{\Theta'}\,p_x^{**})$$

We can show that there exists a finite $p_B<\infty$ such that $0<B_{\Theta}\leq B^{**}$ that satisfy following;

$$x(p_x^{**},B^{**},p_B)=\left(1-\frac{d}{B^{**}}\right)(a-b\,p_x^{**})-\left[\frac{d_{\Theta}\left(\frac{B^{**}}{B_{\Theta}}-1\right)-d_{\Theta'}\left(1-\frac{B_{\Theta}}{B^{**}}\right)}{B^{**}}\right](a_{\Theta}-b_{\Theta}\,p_x^{**})$$

$$>x(p_x^{**},B^{**},p_B^{**})=\left(1-\frac{d_{\Theta'}}{B^{**}}\right)(a_{\Theta'}-b_{\Theta'}\,p_x^{**})$$

In the above inequality, by continuity of the demand functions, the second expression on the left-hand side can be made arbitrarily small by setting a positive B_{Θ}^{**} so that the inequality holds. At that B_{Θ}^{**}, the inequality stated in (A.3) does not hold because the right-hand side has an extra positive term, $B_{\Theta}\,p_B$, which violates the inequality. Thus, in equilibrium, $B_{\Theta}^{**}>0$ should hold. ☐

Proposition 3

Proof. Remember in (14) the charged CP's demand for bandwidth is defined as,

$$B_{\Theta}=\sqrt{\frac{d_{\Theta}}{p_B+(\lambda_1-\lambda_2)}(a_{\Theta}-b_{\Theta}\,p_x)r_{\Theta}}$$

Assume, at the equilibrium, $p_B^{**}=0$, consequently $B_{\Theta}^{**}=B^{**}$. Then, from the charged CP's maximization problem (13), it is seen that first constraint is binding and the second is not. Thus, $\lambda_1>0$ and $\lambda_2=0$. Then,

$$B_{\Theta}^{**}=\sqrt{\frac{d_{\Theta}}{\lambda_1}(a_{\Theta}-b_{\Theta}\,p_x^{**})r_{\Theta}}$$

Because constraint is binding $\lambda_1>0$ then the following should hold,

$$\frac{\partial B_{\Theta}(p_x^{**},B^{**},p_B^{**})}{\partial p_B}=0 \qquad (A.4)$$

The ISP's profit is,

$$\Pi_{ISP}=x(p_x^{**},B^{**},p_B^{**}=0)(p_x^{**}-c_x)-B^{**}\,c_B$$

Optimality implies that,

$$x(p_x^{**},B^{**},p_B^{**}=0)(p_x^{**}-c_x)-B^{**}\,c_B\geq$$
$$x(p_x^{**},B^{**},p_B)(p_x^{**}-c_x)-B^{**}\,c_B+B_{\Theta}\,p_B \qquad \forall p_B \qquad (A.5)$$

From (A.4) we deduce that, if we increase p_B^{**} infinitesimally, the charged CP's demand for bandwidth Θ remains the same as before. Define $p_B = p_B^{**} + \epsilon = \epsilon$. This price will lead to following profit,

$$\Pi_{ISP} = x(p_x^{**}, B^{**}, p_B = \epsilon)(p_x^{**} - c_x) - B^{**}\, c_B + B_\Theta^{**}\, p_B$$

Because B_Θ^{**} is the same at both cases, $x(p_x^{**}, B^{**}, p_B^{**} = 0) = x(p_x^{**}, B^{**}, p_B^{**} = \epsilon)$. Since we have an extra positive term $(B_\Theta\, p_B)$ on the right-hand side of (A.5), this contradicts the optimality condition stated in (A.5) and consequently $p_B = \epsilon$ is a profitable deviation. Thus, at the equilibrium $p_B^{**} > 0$ should hold. ☐

Proposition 4

<u>Proof.</u> First, we need to solve the charged CP's, Θ, maximization problem (13) to get the expression for charged CP's bandwidth demand B_Θ;

$$\max_{\hat{B}} \quad \hat{\Pi} = \hat{x}(p_x, B, p_B) r_\Theta - \hat{B}\, p_B$$
$$\text{s.t.} \quad \hat{B} \le B,$$
$$\hat{B} \ge 0$$

From (14) the solution is;

$$B_\Theta = \sqrt{\frac{d_\Theta}{p_B + (\lambda_1 - \lambda_2)} (a_\Theta - b_\Theta\, p_x) r_\Theta}$$

where $(\lambda_1 \ge 0\ and\ B_\Theta = B), (\lambda_2 \ge 0\ and\ B_\Theta = 0)$

In Proposition 2, we show that $B_\Theta^{**} > 0$. Thus, we conclude that the second constraint is not binding and $\lambda_2 = 0$. Proposition 3 implies that if the first constraint is binding, the ISP has an incentive to increase p_B. Thus, at the equilibrium the first constraint should not be binding and $\lambda_1 = 0$ should hold. Thus, at the equilibrium,

$$B_\Theta^{**} = \sqrt{\frac{d_\Theta}{p_B^{**}} (a_\Theta - b_\Theta\, p_x^{**}) r_\Theta}$$

The ISP's maximization problem is,

$$\max_{p_x, B, p_B} \quad \Pi_{ISP} = x(p_x, B, p_B)(p_x - c_x) + B_\Theta\, p_B - B\, c_B$$

In the equilibrium,

$$\frac{\partial \Pi_{ISP}}{\partial p_x} \le 0 \quad \text{(with equality when } p_x^{**} > 0\text{)}$$
$$\frac{\partial \Pi_{ISP}}{\partial B} \le 0 \quad \text{(with equality when } B^{**} > 0\text{)}$$
$$\frac{\partial \Pi_{ISP}}{\partial p_B} \le 0 \quad \text{(with equality when } p_B^{**} > 0\text{)}$$

must hold. Solving for interior solution for the ISP,

$$\frac{\partial \Pi_{ISP}}{\partial p_x} = 0 = \quad \frac{\partial x(p_x^{**}, B^{**}, p_B^{**})}{\partial p_x}(p_x^{**} - c_x) + x(p_x^{**}, B^{**}, p_B^{**}) + p_B^{**}\frac{\partial B_\Theta(p_B^{**}, p_x^{**})}{\partial p_x}$$

$$\frac{\partial \Pi_{ISP}}{\partial B} = 0 = \quad \frac{\partial x(p_x^{**}, B^{**}, p_B^{**})}{\partial B}(p_x^{**} - c_x) - c_B$$

$$\frac{\partial \Pi_{ISP}}{\partial p_B} = 0 = \quad \frac{\partial x(p_x^{**}, B^{**}, p_B^{**})}{\partial p_B}(p_x^{**} - c_x) + B_\Theta^{**} + p_B^{**}\frac{\partial B_\Theta(p_B^{**}, p_x^{**})}{\partial p_B}$$

For the first order conditions stated above to characterize a maximizer, the Hessian matrix of second derivatives should be negative definite. One sufficient condition for this to hold is that the second non-cross derivatives of the demand function $x(p_x^{**}, B^{**}, p_B^{**})$ be negative and high in absolute value. For $\frac{\partial^2 \Pi_{ISP}}{\partial^2 p_x}$ this condition easily holds at sufficiently low p_x values. For $\frac{\partial^2 \Pi_{ISP}}{\partial^2 p_B}$ it is sufficient to have d_Θ not to be too big compared to $d_{\Theta'}$, which makes sense because a single CP wouldn't be dominating the CP market in terms of congestion sensitivity. $\frac{\partial^2 \Pi_{ISP}}{\partial^2 B}$ is shown to be negative already. The intuition for this is that there is an upper limit for the data demand regardless of the bandwidth installed.

Expressions for $\frac{\partial x(p_x,B,p_B)}{\partial p_x}$, $\frac{\partial x(p_x,B,p_B)}{\partial B}$ and $\frac{\partial x(p_x,B,p_B)}{\partial p_B}$ are,

$$\begin{aligned}\frac{\partial x(p_x, B, p_B)}{\partial p_x} = \quad & -b\left(1 - \frac{d}{B}\right) \\ & + \left(\frac{d_\Theta}{B_\Theta^2} - \frac{d_{\Theta'}}{B^2}\right)(a_\Theta - b_\Theta\, p_x)\frac{\partial B_\Theta(p_B, p_x)}{\partial p_x} \\ & + b_\Theta\left[\frac{d_\Theta\left(\frac{B}{B_\Theta} - 1\right) - d_{\Theta'}\left(1 - \frac{B_\Theta}{B}\right)}{B}\right]\end{aligned}$$

$$\frac{\partial x(p_x, B, p_B)}{\partial B} = \quad \frac{d}{B^2}(a - b\, p_x) - \frac{\left[d_\Theta - d_{\Theta'}\left(2\frac{B_\Theta}{B} - 1\right)\right]}{B^2}(a_\Theta - b_\Theta\, p_x)$$

$$\frac{\partial x(p_x, B, p_B)}{\partial p_B} = \quad \left(\frac{d_\Theta}{B_\Theta^2} - \frac{d_{\Theta'}}{B^2}\right)(a_\Theta - b_\Theta\, p_x)\frac{\partial B_\Theta(p_B, p_x)}{\partial p_B}$$

Expressions for $\frac{\partial B_\Theta(p_B,p_x)}{\partial p_x}$, $\frac{\partial B_\Theta(p_B,p_x)}{\partial B}$ and $\frac{\partial B_\Theta(p_B,p_x)}{\partial p_B}$ are;

$$\frac{\partial B_\Theta(p_B,p_x)}{\partial p_x} = -\frac{1}{2}\left(\frac{d_\Theta\, b_\Theta\, r_\Theta}{p_B\, B_\Theta}\right)$$

$$\frac{\partial B_\Theta(p_B,p_x)}{\partial B} = 0$$

$$\frac{\partial B_\Theta(p_B,p_x)}{\partial p_B} = -\frac{B_\Theta}{2\,p_B}$$

Then rewriting $\frac{\partial \Pi_{ISP}}{\partial p_B} = 0$;

$$\begin{aligned}\frac{\partial \Pi_{ISP}}{\partial p_B} = &\quad 0 \\ = &\quad \left(\frac{d_\Theta}{(B_\Theta^{**})^2} - \frac{d_{\Theta'}}{(B^{**})^2}\right)(a_\Theta - b_\Theta\, p_x^{**})(p_x^{**} - c_x)\frac{\partial B_\Theta(p_B^{**},p_x^{**})}{\partial p_B} \\ &\quad +B_\Theta^{**} \\ &\quad +p_B^{**}\frac{\partial B_\Theta(p_B^{**},p_x^{**})}{\partial p_B}\end{aligned}$$

Rearranging the terms;

$$-\frac{\left(B_\Theta^{**} + p_B^{**}\frac{\partial B_\Theta(p_B^{**},p_x^{**})}{\partial p_B}\right)}{\frac{\partial B_\Theta(p_B^{**},p_x^{**})}{\partial p_B}} = \left(\frac{d_\Theta}{(B_\Theta^{**})^2} - \frac{d_{\Theta'}}{(B^{**})^2}\right)(a_\Theta - b_\Theta\, p_x^{**})(p_x^{**} - c_x)$$

Substituting $\frac{\partial B_\Theta(p_B^{**},p_x^{**})}{\partial p_B}$ with $-\frac{B_\Theta^{**}}{2\,p_B^{**}}$ on the LHS of above equation;

$$-\frac{\left(B_\Theta^{**} + p_B^{**}\frac{\partial B_\Theta(p_B^{**},p_x^{**})}{\partial p_B}\right)}{\frac{\partial B_\Theta(p_B^{**},p_x^{**})}{\partial p_B}} = -\frac{\left(B_\Theta^{**} - p_B^{**}\frac{B_\Theta^{**}}{2\,p_B^{**}}\right)}{-\frac{B_\Theta^{**}}{2\,p_B^{**}}} = p_B^{**}$$

Then,

$$p_B^{**} = \left(\frac{d_\Theta}{(B_\Theta^{**})^2} - \frac{d_{\Theta'}}{(B^{**})^2}\right)(a_\Theta - b_\Theta\, p_x^{**})(p_x^{**} - c_x)$$

Recall that from CP's maximization problem we have,

$$p_B^{**} = \frac{d_\Theta}{(B_\Theta^{**})^2}(a_\Theta - b_\Theta\, p_x^{**})r_\Theta$$

Rearranging $p_B^{**} = \left(\frac{d_\Theta}{(B_\Theta^{**})^2} - \frac{d_{\Theta'}}{(B^{**})^2}\right)(a_\Theta - b_\Theta\, p_x^{**})(p_x^{**} - c_x)$ we have,

$$p_B^{**} = \frac{d_\Theta}{(B_\Theta^{**})^2}(a_\Theta - b_\Theta\, p_x^{**}) r_\Theta \left[\frac{p_x^{**} - c_x}{r_\Theta}\left(1 - \frac{d_{\Theta'}}{d_\Theta}\frac{(B_\Theta^{**})^2}{(B^{**})^2}\right)\right]$$

Thus, we have,

$$\left[\frac{p_x^{**} - c_x}{r}\left(1 - \frac{d_{\Theta'}}{d_\Theta}\frac{(B_\Theta^{**})^2}{(B^{**})^2}\right)\right] = 1$$

Rewriting $\frac{\partial \Pi_{ISP}}{\partial p_x} = 0$;

$$\begin{aligned}\frac{\partial \Pi_{ISP}}{\partial p_x} = &\quad 0\\ = &\quad \left(1 - \frac{d}{B^{**}}\right)(a - b\, p_x^{**})(p_x^{**} - c_x)\\ &+ \left(\frac{d_\Theta}{(B_\Theta^{**})^2} - \frac{d_{\Theta'}}{(B^{**})^2}\right)(a_\Theta - b_\Theta\, p_x^{**})(p_x^{**} - c_x)\frac{\partial B_\Theta(p_B^{**}, p_x^{**})}{\partial p_x}\\ &- \left[\frac{d_\Theta\left(\frac{B^{**}}{B_\Theta^{**}} - 1\right) - d_{\Theta'}\left(1 - \frac{B_\Theta^{**}}{B^{**}}\right)}{B^{**}}\right](a_\Theta - b_\Theta\, p_x^{**})(p_x^{**} - c_x)\\ &+ p_B^{**}\frac{\partial B_\Theta(p_B^{**}, p_x^{**})}{\partial p_x}\end{aligned}$$

Substituting $(a - b\, p_x^{**})(p_x^{**} - c_x)$ with $(a - 2b\, p_x^{**} + b\, c_x)$

and $(a_\Theta - b_\Theta\, p_x^{**})(p_x^{**} - c_x)$ with $(a_\Theta - 2\, b_\Theta\, p_x^{**} + b_\Theta\, c_x)$ we have,

$$\begin{aligned}\frac{\partial \Pi_{ISP}}{\partial p_x} = &\quad 0\\ = &\quad \left(1 - \frac{d}{B^{**}}\right)(a - 2b\, p_x^{**} + b\, c_x)\\ &+ \left(\frac{d_\Theta}{(B_\Theta^{**})^2} - \frac{d_{\Theta'}}{(B^{**})^2}\right)(a_\Theta - b_\Theta\, p_x^{**})(p_x^{**} - c_x)\frac{\partial B_\Theta(p_B^{**}, p_x^{**})}{\partial p_x}\\ &- \left[\frac{d_\Theta\left(\frac{B^{**}}{B_\Theta^{**}} - 1\right) - d_{\Theta'}\left(1 - \frac{B_\Theta^{**}}{B^{**}}\right)}{B^{**}}\right](a_\Theta - 2\, b_\Theta\, p_x^{**} + b_\Theta\, c_x)\\ &+ p_B^{**}\frac{\partial B_\Theta(p_B^{**}, p_x^{**})}{\partial p_x}\end{aligned}$$

Recall that $p_B^{**} = \left(\frac{d_\Theta}{\left(B_\Theta^{**}\right)^2} - \frac{d_{\Theta'}}{(B^{**})^2}\right)(a_\Theta - b_\Theta\, p_x^{**})(p_x^{**} - c_x)$ and $\frac{\partial B_\Theta^{**}(p_B^{**}, p_x^{**})}{\partial p_x} = -\frac{1}{2}\left(\frac{d_\Theta\, b_\Theta\, r}{p_B^{**}\, B_\Theta^{**}}\right)$. Then,

$$\frac{\partial \Pi_{ISP}}{\partial p_x} = 0$$

$$= \left(1 - \frac{d}{B^{**}}\right)(a - 2b\, p_x^{**} + b\, c_x)$$

$$- \frac{d_\Theta\, b_\Theta\, r}{B_\Theta^{**}}$$

$$- \left[\frac{d_\Theta \left(\frac{B^{**}}{B_\Theta^{**}} - 1\right) - d_{\Theta'}\left(1 - \frac{B_\Theta^{**}}{B^{**}}\right)}{B^{**}}\right](a_\Theta - 2\, b_\Theta\, p_x^{**} + b_\Theta\, c_x)$$

We assume $\frac{a_\Theta}{b_\Theta} = \frac{a}{b}$. Rearranging,

$$\frac{\partial \Pi_{ISP}}{\partial p_x} = 0$$

$$= \frac{b(B^{**} - d) - b_\Theta \left[d_\Theta \left(\frac{B^{**}}{B_\Theta^{**}} - 1\right) - d_{\Theta'}\left(1 - \frac{B_\Theta^{**}}{B^{**}}\right)\right]}{B^{**}}\left(\frac{a}{b} - 2\, p_x^{**} + c\right)$$

$$- \frac{d_\Theta\, b_\Theta\, r}{B_\Theta^{**}}$$

Solving the implicit expression for p_x^{**},

$$p_x^{**} = \frac{1}{2}\left(\frac{a}{b} + c_x\right) - \frac{1}{2}\left(\frac{d_\Theta\, b_\Theta\, r_\Theta}{b(B^{**} - d) - b_\Theta \left[d_\Theta \left(\frac{B^{**}}{B_\Theta^{**}} - 1\right) - d_{\Theta'}\left(1 - \frac{B_\Theta^{**}}{B^{**}}\right)\right]}\right)\frac{B^{**}}{B_\Theta^{**}}$$

Given $\frac{\partial x(p_x, B, p_B)}{\partial B} = \frac{d}{B^2}(a - b\, p_x) - \frac{\left[d_\Theta - d_{\Theta'}\left(2\frac{B_\Theta}{B} - 1\right)\right]}{B^2}(a_\Theta - b_\Theta\, p_x)$. Then,

$$\frac{\partial \Pi_{ISP}}{\partial B} = 0$$

$$= \frac{d}{(B^{**})^2}(a - b\, p_x^{**})(p_x^{**} - c_x)$$

$$- \frac{\left[d_\Theta - d_{\Theta'}\left(2\frac{B_\Theta^{**}}{B^{**}} - 1\right)\right]}{(B^{**})^2}(a_\Theta - b_\Theta\, p_x^{**})(p_x^{**} - c_x)$$

$$- c_B$$

Rearranging and solving for implicit expression of B^{**},

$$B^{**} = \sqrt{\frac{d}{c_B}(a - b\,p_x^{**})(p_x^{**} - c_x) - \frac{\left[d_\Theta - d_{\Theta'}\left(2\frac{B_\Theta^{**}}{B^{**}} - 1\right)\right]}{c_B}(a_\Theta - b_\Theta\,p_x^{**})(p_x^{**} - c_x)}$$

□

Proposition 5

Proof. At the equilibrium,

$$\frac{\partial \Pi_{ISP}}{\partial p_B} \leq 0 \quad \text{(with equality when } p_B^{**} > 0\text{)}$$

must hold. Solving for interior solution,

$$\frac{\partial \Pi_{ISP}}{\partial p_B} = 0 = \frac{\partial x(p_x^{**}, B^{**}, p_B^{**})}{\partial p_B}(p_x^{**} - c_x) + B_\Theta^{**} + p_B^{**}\frac{\partial B_\Theta(p_B^{**}, p_x^{**})}{\partial p_B}$$

Expression for $\frac{\partial x(p_x, B, p_B)}{\partial p_B}$ is,

$$\frac{\partial x(p_x, B, p_B)}{\partial p_B} = \left(\frac{d_\Theta}{B_\Theta^2} - \frac{d_{\Theta'}}{B^2}\right)(a_\Theta - b_\Theta\,p_x)\frac{\partial B_\Theta(p_B, p_x)}{\partial p_B}$$

Expression for $\frac{\partial B_\Theta(p_B, p_x)}{\partial p_B}$ is,

$$\frac{\partial B_\Theta(p_B, p_x)}{\partial p_B} = -\frac{B_\Theta}{2\,p_B}$$

Then rewriting $\frac{\partial \Pi_{ISP}}{\partial p_B} = 0$;

$$\begin{aligned}\frac{\partial \Pi_{ISP}}{\partial p_B} &= 0 \\ &= \left(\frac{d_\Theta}{(B_\Theta^{**})^2} - \frac{d_{\Theta'}}{(B^{**})^2}\right)(a_\Theta - b_\Theta\,p_x^{**})(p_x^{**} - c_x)\frac{\partial B_\Theta(p_B^{**}, p_x^{**})}{\partial p_B} \\ &\quad + B_\Theta^{**} \\ &\quad + p_B^{**}\frac{\partial B_\Theta(p_B^{**}, p_x^{**})}{\partial p_B}\end{aligned}$$

$$-\frac{\left(B_\Theta^{**} + p_B^{**}\frac{\partial B_\Theta(p_B^{**}, p_x^{**})}{\partial p_B}\right)}{\frac{\partial B_\Theta(p_B^{**}, p_x^{**})}{\partial p_B}} = \left(\frac{d_\Theta}{(B_\Theta^{**})^2} - \frac{d_{\Theta'}}{(B^{**})^2}\right)(a_\Theta - b_\Theta\,p_x^{**})(p_x^{**} - c_x)$$

$$-\frac{\left(B_{\Theta}^{**}+p_{B}^{**}\frac{\partial B_{\Theta}(p_{B}^{**},p_{x}^{**})}{\partial p_{B}}\right)}{\frac{\partial B_{\Theta}(p_{B}^{**},p_{x}^{**})}{\partial p_{B}}}=-\frac{\left(B_{\Theta}^{**}-p_{B}^{**}\frac{B_{\Theta}}{2\,p_{B}}\right)}{-\frac{B_{\Theta}}{2\,p_{B}}}=p_{B}^{**}$$

Then,

$$p_{B}^{**}=\left(\frac{d_{\Theta}}{(B_{\Theta}^{**})^{2}}-\frac{d_{\Theta'}}{(B^{**})^{2}}\right)(a_{\Theta}-b_{\Theta}\,p_{x}^{**})(p_{x}^{**}-c_{x})$$

Recall that from CP's maximization problem we have,

$$p_{B}^{**}=\frac{d_{\Theta}}{(B_{\Theta}^{**})^{2}}(a_{\Theta}-b_{\Theta}\,p_{x}^{**})r_{\Theta}$$

Rearranging $p_{B}^{**}=\left(\frac{d_{\Theta}}{(B_{\Theta}^{**})^{2}}-\frac{d_{\Theta'}}{(B^{**})^{2}}\right)(a_{\Theta}-b_{\Theta}\,p_{x}^{**})(p_{x}^{**}-c_{x})$ we have,

$$p_{B}^{**}=\frac{d_{\Theta}}{(B_{\Theta}^{**})^{2}}(a_{\Theta}-b_{\Theta}\,p_{x}^{**})r_{\Theta}\left[\frac{p_{x}^{**}-c_{x}}{r_{\Theta}}\left(1-\frac{d_{\Theta'}}{d_{\Theta}}\frac{(B_{\Theta}^{**})^{2}}{(B^{**})^{2}}\right)\right]$$

Thus, we have,

$$\left[\frac{p_{x}^{**}-c_{x}}{r}\left(1-\frac{d_{\Theta'}}{d_{\Theta}}\frac{(B_{\Theta}^{**})^{2}}{(B^{**})^{2}}\right)\right]=1$$

□

Proposition 6

Proof. From Proposition 1 we have,

$$p_{x}^{*}=\frac{1}{2}\left(\frac{a}{b}+c_{x}\right) \qquad (A.6)$$

From Proposition 4 we have,

$$p_{x}^{**}=\frac{1}{2}\left(\frac{a}{b}+c_{x}\right)-\frac{1}{2}\left(\frac{d_{\Theta}\,b_{\Theta}\,r_{\Theta}}{b(B^{**}-d)-b_{\Theta}\left[d_{\Theta}\left(\frac{B^{**}}{B_{\Theta}^{**}}-1\right)-d_{\Theta'}\left(1-\frac{B_{\Theta}^{**}}{B^{**}}\right)\right]}\right)\frac{B^{**}}{B_{\Theta}^{**}} \qquad (A.7)$$

By Proposition 2 we know $B_{\Theta}^{**}>0$.

Recall that,

$$x(p_{x},B,p_{B})=\left(1-\frac{d}{B}\right)(a-b\,p_{x})-\left[\frac{d_{\Theta}\left(\frac{B}{B_{\Theta}}-1\right)-d_{\Theta'}\left(1-\frac{B_{\Theta}}{B}\right)}{B}\right](a_{\Theta}-b_{\Theta}\,p_{x})$$

Then,

$$x(p_x, B, p_B) = \frac{b(B-d) - b_\Theta\left[d_\Theta\left(\frac{B}{B_\Theta} - 1\right) - d_{\Theta'}\left(1 - \frac{B_\Theta}{B}\right)\right]}{B}\left(\frac{a}{b} - p_x\right)$$

At the equilibrium $x(p_x^{**}, B^{**}, p_B^{**}) > 0$.

Thus; $b(B^{**} - d) - b_\Theta\left[d_\Theta\left(\frac{B^{**}}{B_\Theta^{**}} - 1\right) - d_{\Theta'}\left(1 - \frac{B_\Theta^{**}}{B^{**}}\right)\right] > 0$ must hold.

Then; 2^{nd} term on the (A.7) is positive. Then $p_x^{**} < p_x^*$. ☐

Proposition 7

Proof. From Proposition 1 we have,

$$B^* = \sqrt{\frac{d}{c_B}(a - b\,p_x^*)(p_x^* - c_x)} \qquad (A.8)$$

From Proposition 4 we have,

$$B^{**} = \sqrt{\frac{d}{c_B}(a - b\,p_x^{**})(p_x^{**} - c_x) - \frac{\left[d_\Theta - d_{\Theta'}\left(2\frac{B_\Theta^{**}}{B^{**}} - 1\right)\right]}{c_B}(a_\Theta - b_\Theta\,p_x^{**})(p_x^{**} - c_x)} \quad (A.9)$$

Let $f(p) = (a - b\,p_x)(p_x - c_x)$. From first order conditions it can be shown that, $p_x^* = \frac{1}{2}\left(\frac{a}{b} + c_x\right)$ maximizes $f(p)$.

Then,

$$\frac{d}{c_B}(a - b\,p_x^{**})(p_x^{**} - c_x) < \frac{d}{c_B}(a - b\,p_x^*)(p_x^* - c_x)$$

We assume $d_\Theta > d_{\Theta'}$. Then, 2^{nd} term on (A.9) is positive.

Thus; $B^{**} < B^*$ holds. ☐

Proposition 8

Proof. Under neutrality, congestion is the same for both Θ and Θ', given as,

$$\frac{a - b\,p_x^*}{B^*}$$

Under discrimination, from (6) congestion for Θ is given by,

$$\frac{a - b\,p_x^{**}}{B^{**}} + \left[\frac{\left(\frac{B^{**}}{B_\Theta^{**}} - 1\right)(a_\Theta - b_\Theta\,p_x^{**})}{B^{**}}\right]$$

By Proposition 6 we have $p_x^{**} < p_x^*$ and by Proposition 7 we have $B^{**} < B^*$. So, we know that $\frac{a-b\,p_x^{**}}{B^{**}} > \frac{a-b\,p_x^*}{B^*}$, then,

$$\frac{a-b\,p_x^{**}}{B^{**}} + \left[\frac{\left(\frac{B^{**}}{B_\Theta^{**}} - 1\right)(a_\Theta - b_\Theta\,p_x^{**})}{B^{**}}\right] > \frac{a-b\,p_x^*}{B^*}$$

Thus, congestion is worse off for Θ under discrimination.

From (8) congestion for Θ' is given by,

$$\frac{a-b\,p_x^{**}}{B^{**}} - \left[\frac{\left(1 - \frac{B_\Theta^{**}}{B^{**}}\right)(a_\Theta - b_\Theta\,p_x^{**})}{B^{**}}\right]$$

Because the second term in the above expression is positive, congestion for Θ' may be better or worse off depending on the equilibrium quantities p_x^{**}, p_x^*, B^{**}, and B^*.

From (9) overall congestion $\left(\Theta + \Theta'\right)$ is given by,

$$\frac{a-b\,p_x^{**}}{B^{**}} + \left[\frac{\left(\left(\frac{d_\Theta}{d}\right)\left(\frac{B^{**}}{B_\Theta^{**}} - 1\right) - \left(\frac{d_{\Theta'}}{d}\right)\left(1 - \frac{B_\Theta^{**}}{B^{**}}\right)\right)(a_\Theta - b_\Theta\,p_x^{**})}{B^{**}}\right]$$

Since we assume $d_\Theta > d_{\Theta'}$ holds, the second term in the above expression is positive. By Proposition 6 we have $p_x^{**} < p_x^*$ and by Proposition 7 we have $B^{**} < B^*$. So, we know that $\frac{a-b\,p_x^{**}}{B^{**}} > \frac{a-b\,p_x^*}{B^*}$, then,

$$\frac{a-b\,p_x^{**}}{B^{**}} + \left[\frac{\left(\left(\frac{d_\Theta}{d}\right)\left(\frac{B^{**}}{B_\Theta^{**}} - 1\right) - \left(\frac{d_{\Theta'}}{d}\right)\left(1 - \frac{B_\Theta^{**}}{B^{**}}\right)\right)(a_\Theta - b_\Theta\,p_x^{**})}{B^{**}}\right] > \frac{a-b\,p_x^*}{B^*}$$

Thus, overall congestion $(\Theta + \Theta')$ is worse under discrimination.

□