ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

NETWORK NEUTRALITY IN THE INTERNET AS A TWO SIDED MARKET CONSTITUTED BY CONGESTION SENSITIVE END USERS AND CONTENT PROVIDERS

Ph.D. THESIS

Özgür KAPLANLIOĞLU

Department of Economics

Economics Programme

MARCH 2023

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

NETWORK NEUTRALITY IN THE INTERNET AS A TWO SIDED MARKET CONSTITUTED BY CONGESTION SENSITIVE END USERS AND CONTENT PROVIDERS

Ph.D. THESIS

Özgür KAPLANLIOĞLU
(412152004)

Department of Economics

Economics Programme

Thesis Advisor: Prof. Dr. Bülent GÜLOĞLU
Thesis Co-Advisor: Prof. Dr. Sencer ECER

MARCH 2023

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YOĞUNLUĞA DUYARLI SON KULLANICILAR VE İÇERİK ÜRETİCİLERİNDEN OLUŞAN İNTERNETTE AĞ TARAFSIZLIĞI

DOKTORA TEZİ

Özgür KAPLANLIOĞLU
(412152004)

İktisat Anabilim Dalı

İktisat Programı

Tez Danışmanı: Prof. Dr. Bülent GÜLOĞLU
Eş Danışman: Prof. Dr. Sencer ECER

MART 2023

Özgür KAPLANLIOĞLU, a Ph.D. student of ITU Graduate School student ID 412152004, successfully defended the dissertation entitled "NETWORK NEUTRALITY IN THE INTERNET AS A TWO SIDED MARKET CONSTITUTED BY CONGESTION SENSITIVE END USERS AND CONTENT PROVIDERS", which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Bülent GÜLOĞLU** Istanbul Technical University	....
Co-advisor :	**Prof. Dr. Sencer ECER** Istanbul Technical University	....
Jury Members :	**Prof. Dr. Alp ÜSTÜNDAĞ** Istanbul Technical University	....
	Prof. Dr. Fuat Erdal Anadolu University	....
	Dr. Öğr. Üyesi Murat Güven Sakarya University	....
	Dr. Öğr. Üyesi Mustafa Doğan Istanbul Technical University	....
	Doç. Dr. Ali Coşkun Bogazici University	....

Date of Submission : 14 February 2023
Date of Defense : 1 March 2023

To my wife Ayşe,
without whom this study would have been completed two years earlier.

FOREWORD

This dissertation has been written to fulfill the graduation requirements of the Istanbul Technical University Economics PhD Programme. I was engaged in researching and writing this dissertation from June 2019 to June 2022.

My research question was formulated together with my professor, Sencer ECER. The research was difficult but also quite timely considering all the debate that is still going on in the telecommunication industry. Fortunately, conducting extensive investigation has allowed me to answer the question that we identified. I felt very lucky that, both Sencer ECER and my advisor Bulent GÜLOĞLU were always available and willing to answer my queries.

I would like to thank my supervisors for their excellent guidance and support during this process.

I also benefitted from debating issues with my friends and family. If I ever lost interest, you kept me motivated. My parents and my wife deserve a particular note of thanks: your wise counsel and kind words have, as always, served me well.

I hope you enjoy your reading.

March 2023 — Özgür KAPLANLIOĞLU
(Management Consultant)

TABLE OF CONTENTS

Page

ABBREVIATIONS

CP	**:** Content Provider
EU	**:** European Union
FCC	**:** Federal Communications Commission
ISP	**:** Internet Service Provider
OTT	**:** Over The Top
US	**:** United States

SYMBOLS

p_x	: Unit price of data for End Users
c_x	: Marginal cost of unit data consumption for ISP
p_B	: Unit price of bandwidth under discrimination
c_B	: Marginal cost of installing unit bandwidth for ISP
B	: Total bandwidth installed by the ISP
B_Θ	: Total bandwidth purchased by the discriminated CP
a_θ, b_θ	: Shape parameters of End Users' demand function for CP θ
a_Θ, b_Θ	: Shape parameters of End Users' demand function for the discriminated CP
$a_{\Theta'}, b_{\Theta'}$	: Shape parameters of End Users' demand function for the remaining CPs
a, b	: Shape parameters of End Users' aggregate demand function
d_θ	: Congestion sensitivity of CP θ
d_Θ	: Congestion sensitivity of the discriminated CP
$d_{\Theta'}$	: Aggregate congestion sensitivity of the remaining CPs
d	: Aggregate congestion sensitivity
r_θ	: Advertisement revenue of CP θ on unit data
r_Θ	: Advertisement revenue of the discriminated CP on unit data

LIST OF FIGURES

Page

NETWORK NEUTRALITY IN THE INTERNET AS A TWO SIDED MARKET CONSTITUTED BY CONGESTION SENSITIVE END USERS AND CONTENT PROVIDERS

SUMMARY

In this paper, I 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. I 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. I then study the implications of the violation of network neutrality on total data consumption, congestion, and capacity investment.

Under network neutrality, the decision variables of the ISP are end-user price (p_x) and network bandwidth (B). I found that, in equilibrium, because of the monopolistic nature of the market p_x is greater than the price under competitive equilibrium, and B is lower than its socially optimum value. Thus, under neutrality, the ISP undersupplies both the capacity and the data.

Under discrimination, the ISP is allowed to charge an access fee on unit bandwidth to one of the content providers (the discriminated CP). To reflect a scenario of great practical value, I choose the discriminated CP from one of the big OTTs such as YouTube, Instagram, Facebook, Netflix, etc. In this setting, to access the network, the discriminated CP needs to buy bandwidth from the ISP. However, the bandwidth bought by the discriminated CP is not for exclusive usage of the discriminated CP. It rather acts as an upper bandwidth limit for the discriminated CP. Under discrimination, the decision variables of the ISP are the end-user price (p_x), the network bandwidth (B), and the bandwidth price (p_B).

I found that when allowed the ISP always prefers to deviate from network neutrality by charging a positive price for bandwidth (p_B). Also, the ISP sets p_B just enough to keep the discriminated CP in the market. Comparing the equilibrium outcomes, I show that under discrimination, the ISP charges a lower price to end-users. However, the discrimination also leads to less network bandwidth installed. Both the lower end-user price and the lower network bandwidth contribute to the congestion. Thus, under discrimination, the congestion is higher than under neutrality.

Considering its adverse effects on the network bandwidth and congestion, although the end-user price is lower under discrimination, I recommend that the network neutrality principle should not be abolished.

YOĞUNLUĞA DUYARLI SON KULLANICILAR VE İÇERİK ÜRETİCİLERİNDEN OLUŞAN İNTERNETTE AĞ TARAFSIZLIĞI

ÖZET

Bu çalışmada, Network Neutrality (Ağ Tarafsızlığı) ilkesinin ihlal edilmesi durumunda, son kullanıcılar ve içerik üreticilerinden (CP) oluşan bir veri piyasasında, fiyat ve ağ yoğunluğu gibi denge büyüklüklerinin hangi yönde değişeceğini ve bunun nasıl politika sonuçlarına sebep olabileceğini inceledim.

Ağ tarafsızlığı son yıllarda oldukça tartışmalı bir ağ ilkesi haline gelmiştir. Bu ilke en basit tanımı ile internet servis sağlayıcılarının şu iki kurala katı bir şekilde uymalarını gerektirir. Birinci kural, internet servis sağlayıcılarının (ISP) ağ üzerinde hareket eden veri paketlerinin hangi içerik üreticisinden çıktığına göre ayrımcılığa gitmemeleri gerektiğidir. İkinci kural ise ISP'lerin veri paketlerinin içeriğine göre bir ayrımcılığa gitmemelerini şart koşmaktadır. Dolayısı ile ağ tarafsızlığı ilkesinin getirmiş olduğu bu iki kural ISP'lerin içerik üreticileri üzerinden fiyat ayrımcılığı (price discrimination) gibi iş modellerine gidebilmelerini engellemektedir. Bu ilke, ağ üzerinde özellikle sınırlı sayıda içerik üreticisinin sebep olduğu ağ yoğunluğu (congestion) sorunları nedeni ile son yıllarda ISP'ler tarafından tartışılır ve hatta çoğu kez ihlal edilir hale gelmiştir. Bu yüzden ağ tarafsızlığının hukuki dayanağı da birçok ülkede politika yapıcılarınca tartışılır hale gelmiştir.

Çalışmamda, monopol bir İnternet Servis Sağlayıcı (ISP) tarafından işletilen, son kullanıcı ve içerik üreticilerinden oluşan basit bir iki taraflı piyasa (two-sided market) tasarladım. Bu modelde, ağ tarafsızlığı geçerli iken, ISP son kullanıcıları kullandıkları birim veri üzerinden doğrusal bir şekilde fiyatlamaktadır (p_x). İçerik üreticileri (CP) ise, ISP'nin altyapı yatırımı yaparak sağladığı bant genişliğini (B) herhangi bir fiyat ödemeden, ağ tarafsızlığı ilkesi çerçevesinde kullanmaktadır. ISP ayrıca kendi ağı üzerinden tüketilen birim veri için sabit bir sınırsal maliyete (marginal cost) sahiptir. Literatürde de işaret edildiği gibi ağ tarafsızlığını ihlal etmenin birden fazla yolu bulunmaktadır. Bunlardan en yaygın olan ve literatürde en sıklıkla çalışılmış olan iki yöntem "throttling" yani "boğma" ve "tiered service" yani "katmanlı servis" yöntemleridir. Ben bu çalışma özelinde ağ tarafsızlığının ihlal edilmesini, ISP'nin belirlediği tek bir içerik üreticisine (discriminated CP), sahip olduğu bant genişliğini (B) belirli bir birim fiyat (p_B) üzerinden satması ($3rd$ degree price discrimination) olarak tanımladım. Fakat içerik üreticisi tarafından satın alınan bant genişliği (B_Θ), bu içerik üreticisine özel olarak ayrılmayıp, sadece bir üst hız limiti olarak uygulanmaktadır. Bu tarz hız limitlemeleri günümüzde de birçok operatör tarafından sıklıkla kullanılmakta olduğundan dolayı model için tercih edilmiştir.

Bu noktada, discriminated CP'nin sahip olduğu nitelikler ile ilgili bir parantez açmak gereği bulunmaktadır. Modelim bir $3rd$ degree price discrimination modeli olduğu için ayrımcılığa uğrayan içerik üreticisinin sıradan ve rastgele bir içerik üreticisi olması beklenemez. Bu içerik üreticisinin ağ yoğunluğuna (congestion) son derece hassas ve çalışmanın sonuçlarına etkisi göz önünde bulundurulduğunda üzerinden oldukça fazla veri tüketilen bir içerik üreticisi olması gerekmektedir. Günümüz internet piyasasında bu şartları sağlayan birkaç içerik üreticisi bulunmaktadır. Bunlar YouTube, Instagram, Facebook, Netflix gibi büyük OTT (Over the Top) oyunculardır. Ayrıca literatürde de görüldüğü üzere ISPler ile ağ tarafsızlığı noktasında en çok bu içerik üreticileri hukuksal sorunlar yaşamaktadırlar.

Ağ bant genişliğinin (B) modele etkisi ağ yoğunluğu (congestion) üzerindendir. Modelde, son kullanıcıları ağ yoğunluğuna hassas bir şekilde tanımladım. Daha açık bir ifade ile, son kullanıcılar yoğunluğun yüksek olduğu bir ağda yarar (utility) kaybı yaşamakta ve veri tüketimlerini azaltmaktadırlar. Ağ bant genişliğinin (B) içerik üreticisine etkisi ise reklam gelirleri üzerindendir. Modelde içerik üreticisinin, kendisi üzerinden tüketilen birim veri başına sabit bir reklam geliri olduğunu varsaydım. Bu durumda içerik üreticisi üzerinden tüketilen veri miktarı arttıkça kârı da artacaktır. Bu açıdan düşünüldüğünde herhangi bir içerik üreticisinin bant genişliği satın almak için güçlü bir teşviği bulunmaktadır. Bant genişliği satın almadığı takdirde ağ yoğunluğu sebebi ile veri tüketimi düşecek ve içerik üreticisinin gelirleri azalacaktır.

Ağ tarafsızlığı şartları altında, ISP denge durumunda, son kullanıcıya vereceği birim veri fiyatına (p_x) ve ağ bant genişliğine (B) karar verir. ISP'nin monopol olmasından dolayı her iki denge büyüklüğünün de toplumsal olarak en elverişli olan değerlerden (socially optimal) daha düşük olduğu gözlenmektedir.

Ağ tarafsızlığının ihlal edildiği şartlar altında ise, ISP denge durumunda, son kullanıcıya vereceği birim veri fiyatına (p_x), ağ bant genişliğine (B), ve ağ tarafsızlığı durumundan farklı olarak içerik üreticisine sağlayacağı bant genişliğinin birim fiyatına (p_B) karar verir. Bu durumda ortaya çıkan bazı sonuçlar önemlidir.

Öncelikle, denge durumunda ISP'nin daima, ağ bant genişliği fiyatını (p_B) içerik üreticisini pozitif bir miktarda ağ bant genişliği almasını sağlayacak şekilde, aşırı yüksek bir fiyat olmamak koşulu ile belirlediği ortaya çıkmıştır. Buna göre ağ tarafsızlığından sapmanın piyasadaki içerik çeşitliliğini azaltmadığı ortaya konulmuştur. Bu durumu şu şekilde açıklamak mümkündür. ISP'nin ayrımcılığa uğrayan içerik üreticisini (discriminated CP) piyasa dışında tutmasının iki etkisi vardır. Birincisi, son kullanıcılar bu içerik üreticisine ulaşamadıkları için veri tüketimi düşecektir. İkincisi ise, bu içerik üreticisinin piyasada olmamasından dolayı ağ yoğunluğu (congestion) düşecek ve bu durum veri tüketiminin artmasına sebep olacaktır. Model birbiri ile ters yönde hareket eden bu iki etkiden daima birincisinin galip geleceğini göstermiştir. Bu yüzden kârını en yüksekleyen (profit maximizer) bir ISP için ayrımcılığa uğrayan içerik üreticisini (discriminated CP) makul bir ağ genişliği fiyatı ile piyasada tutmak daha rasyoneldir.

Bir diğer sonuç ise, monopol ISP'nin ağ bant genişliği birim fiyatını daima 0'dan büyük şekilde belirleyeceğidir. Bu durumu şu şekilde açıklayabiliriz. Ağ tarafsızlığından sapıldığı durumda ISP'nin iki tane gelir kaynağı vardır. Bunlardan birincisi son kullanıcılardan sağladığı veri kullanım geliri ikincisi ise ayrımcılığa uğrayan içerik üreticisinden (discriminated CP) elde ettiği bant genişliği kullanım geliridir. ISP bant genişliği fiyatını 0 tuttuğu takdirde ikinci gelir kaynağından mahrum kalacaktır. Bu durum ayrımcılığa uğrayan içerik üreticisinin (discriminated CP) bant genişliğinin tamamına ulaşmasına imkan verdiği için son kullanıcılardan elde edilen veri gelirini arttıracaktır. Fakat ayrımcılığa uğrayan içerik üreticisinin (discriminated CP) ağ genişliği talep fonksiyonunun sürekli olduğu düşünüldüğünde ve mevcut ağ genişliğinin bir üst sınır olduğu düşünüldüğünde ISP veri tüketim gelirlerini azaltmadan ve aynı zamanda ağ genişliği için 0'dan daha büyük bir fiyat belirleyerek önceki durumuna göre daha kazançlı hale gelebilmektedir. Bu nedenle kârını en yüksekleyen (profit maximizer) bir ISP için ağ tarafsızlığı durumu daima ağ tarafsızlığının ihlali durumuna göre daha az kârlıdır ve ISP, ağ tarafsızlığı ilkesi gevşetildiği takdirde onu ihlal etmeyi tercih edecektir.

Çalışmanın sonuçlarından bir diğeri de, iki taraflı piyasaların (two-sided market) belirleyici bir özelliğini doğrular niteliktedir. Bilindiği üzere iki taraflı piyasalarda, platformu işleten oyuncu (bu çalışmanın özelinde bu ISP'dir), piyasanın iki tarafının çapraz elastikiyetlerine (cross-group elasticity) göre, taraflardan birisini tek taraflı durumda gerçekleşecek olan denge fiyatının üzerinde, diğerini ise altında fiyatlamaktadır. Bu çalışmada da, son kullanıcı fiyatının (p_x) yükselmesine sebep olan denge durumlarında, ağ bant genişliği birim fiyatının (p_B) düştüğü, ve tersi durumda da yükseldiği görülmektedir. Bu sonuca bağlı olarak, içerik üreticisinin birim veri başına elde ettiği reklam geliri artarsa, ISP'un ağ bant genişliği fiyatını (p_B) yükselteceği ve son kullanıcı fiyatını (p_x) ise düşüreceği görülmüştür. Bir başka ifade ile, içerik üreticisinin işi daha kârlı hale geldikçe, bu duruma ISP ağ bant genişliği birim fiyatını yükselterek karşılık verecektir. Benzer şekilde ağ bant genişliğini arttırmak için gereken yatırım maliyeti düştükçe, ISP bu duruma ağ genişliği birim fiyatını (p_B) düşürerek ve son kullanıcı fiyatını (p_x) ise yükselterek karşılık verecektir.

Çalışmanın diğer önemli sonuçlarından birisi de ağ tarafsızlığı ve ağ tarafsızlığının ihlal edilmesi durumlarındaki denge büyüklüklerinin karşılaştırılması halinde ortaya çıkmaktadır. Bu karşılaştırmaya göre ISP, ağ tarafsızlığının ihlal edildiği durumda, son kullanıcıya ağ tarafsızlığı durumuna kıyasla daha düşük bir birim veri fiyatı belirlemektedir. Fakat ISP'nin ne kadar ağ bant genişliği yatırımı yapacağına bakıldığında, ağ tarafsızlığı durumuna göre daha az yatırım yaptığı görülmektedir. İlk durum son kullanıcı için yarar arttırıcı gibi görünse de ikinci durum ağ yoğunluğunu arttırıcı yönde etki yapacağı için son kullanıcının yararını düşürücü yönde etki yapar.

Ağ tarafsızlığın ihlal edildiği durumda, ağ yoğunluğunun, ağ tarafsızlığı durumuna göre nasıl değişeceğine bakıldığında ise, ayrıştırılan içerik üreticisi için (discriminated CP) ağ yoğunluğunun daima olumsuz yönde etkilendiğini, geri kalan içerik üreticileri için ise, belirli denge durumlarında, ağ yoğunluğunun olumlu yönde değişebileceği, fakat bu olumlu değişimin her koşulda gerçekleşmeyeceği görülmüştür. Ağın tamamına bakıldığında ise genel ağ yoğunluğunun olumsuz yönde etkilendiği görülmektedir.

Tüm bu sonuçlar ışığında genel kanaatim; alt yapı yatırımına ve ağ yoğunluğuna olan olumsuz etkilerinden ve son kullanıcıya olan yararlarının ikna edici olmamasından dolayı, ağ tarafsızlığı ilkesinden sapılmaması yönündedir.

1. INTRODUCTION

In most of the ISPs globally, the amount of 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 their networks in total. What would it be like if ISPs were able to charge a price from these OTTs to let them continue to use their infrastructure to provide their services? How would consumer welfare be affected if these OTTs accept to pay the price? Such questions are entirely related to the current network neutrality debate in academics and industry.

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 that started in academics and industry on whether network neutrality should be relaxed or abolished or maintained as it is, 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 explicit network neutrality constraints [1]. Following this change, major ISPs and network owners in the US started discriminating on the upstream side of the market, in which some of the cases were taken to court. 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 ISP and the CP [2].

From an economics perspective, network neutrality is a broad term with no simple definition that economists have consensus on. 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 a practice of third-degree price discrimination situation). Likewise, ISP's division of its bandwidth into predetermined partitions for exclusive usage by content providers who contract with ISP for preferential treatment also counts as a clear violation of network neutrality. Such a tiered service consisting of "slow lanes" and "fast lanes" actually depicts a second-degree price discrimination setting and again has an economic interest too [1].

The economics literature on network neutrality is far from being rich. Existing studies analyze different aspects of the above-mentioned two possible modes of violation of network neutrality and how it affects welfare, profits, investment incentives, content variety, and innovations. However, there is still some controversy on this principle. Proponents claim that network neutrality guarantees the openness of the internet. Without network neutrality, large and financially powerful content providers (like Google, Amazon, Netflix, Yahoo, etc.) dominates the internet. This would hinder the innovation on edge and prevents new startups with innovative and revolutionary ideas from coming into existence [3]. On the other hand, opponents argue that with the abolishment 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.

I organized the paper as follows. I summarize related literature in Section 2. In Section 3; I explain how the model specifies End Users, ISP, and CPs in detail. In Section 4 and Section 5, I derive the equilibrium for network neutrality and discrimination cases respectively. How discrimination affects the equilibrium on different aspects such as data volume, congestion, etc. is discussed in Section 6. Finally, concluding remarks are presented in Section 7.

1.1 Literature Review

The literature around network neutrality is still underdeveloped. The followings are some of the notable studies.

Choi et al. [4] consider a broadband network with a monopolistic ISP. They build a model where it is assumed that the internet consists of only two symmetric CPs. So in their modeling scheme, there is a continuum of end-users with a unit mass, two content providers competing for market share, and a monopolistic ISP. Market shares of content providers are determined by the classical Hotelling approach [5]. End users are homogeneous in the sense that each has the same demand for content. Content providers in Choi et al.'s model 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 two profit margins and the difference between the two profit margins 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. show; under discrimination that, the network access fee is unambiguously lower than it is under network neutrality. Although I utilize a linear pricing scheme instead of a constant network access fee for end-users; in my model, I also show that the price for end-users is strictly lower under discrimination. In Choi et al.'s model, ISP's incentives to practice the discriminatory regime depends on its bargaining power and the magnitudes of CPs' profit margins. My model shows that the ISP always prefers to discriminate as long as the discriminated CP's business is profitable. However, my model differs from Choi et al.'s in the sense that no competition is assumed between CP's and ISP's discrimination method is charging a network access fee to a designated CP based on unit bandwidth rather than offering prioritized service. In Choi et al.'s model non-prioritized content provider (CP2) is always worse off under discrimination in terms of profit, whereas, the prioritized content provider (CP1) may be better off under discrimination if relative magnitudes of CPs margins 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. 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 too. My 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 [6] compare network neutrality with discrimination in the fashion of either practicing tiered service or charging an access fee to CPs. Their model differs from Choi et al.'s in several points. Unlike Choi et al., they don't assume constant data traffic. Consumption decision is given endogenously by homogeneous households (end-users). In the model, the more congestion end-users are exposed to, the less content is consumed. The more expensive the content is, the less content is consumed by households also. Besides unlike Choi and Kim, Economides and Hermalin allow CPs to directly sell their content to end-users. This adds a second source of revenue to their revenue stream. In my model, though, I don't allow such trade between end-users and CPs either.

Economides and Hermalin 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 reveal the ISP's incentives to charge content providers. Lastly, they switch to a dynamic setting and try to understand the investment incentives of the monopolistic ISP. The principal result is that 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 is consumed by end-users. Another result that is attached to the previous one is that blocking access to a positive measure of CPs are necessarily reduce 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. In other conditions tiering is always welfare inferior. When it comes to ISP's incentives to charge CPs, it is shown that in the majority of the cases such pricing schemes are welfare inferior. However, it is also shown that ISP's charging CPs (even though it reduces total welfare) will cause smaller network access fees on the end-users side. Economides and Hermalin reveal that if adjustment function is multiplicatively separable in delay time (congestion) and content type, then ISP will always prefer discrimination vis-a-vis network neutrality. To summarize, when bandwidth is taken constant (static setting), there would be an inefficiency caused by preferring a tiering regime (discrimination) vis-a-vis network neutrality. Yet, Economides and Hermalin find that, under dynamic setting (also under certain assumptions regarding the functional form of the adjustment function and incentive-compatibility constraints being satisfied), ISP will always install unambiguously more bandwidth in discrimination vis-a-vis network neutrality which has an enhancing effect on 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 [7] study short-run and long-run effects of discrimination (practicing two service tiers to CPs as priority class and best-effort class) on content variety, welfare, and capacity investment incentives. They found, under discrimination, content variety may or may not be higher than it is under neutrality. They show equilibrium content variety is closely related to how CPs are distributed. If it is 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. On the opposite case (CP distribution is right-skewed), prioritization causes fewer CPs to be active in the equilibrium compared to network neutrality in the short run. Sticking to the assumption that contents are uniformly distributed, the ISP always prefers prioritization against 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 also investigate the investment incentives of ISP in the long run. They reveal that under the assumption of uniformly distributed contents the ISP's investment incentives would be higher. However, for non-uniform distributions of CPs, the result is not clear.

Bourreau et al. [8] consider a similar model to Krämer and Wiewiorra's with some important differences. The major distinction is that, they 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 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. compare the equilibrium under network neutrality and discrimination. In both cases, 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, ISPs decide subscription fees for the end-users and network capacities. Likewise, under discrimination, at the second stage, end users decide which ISP to join and, CPs decide not only which ISP to join but also whether to pay for the prioritization. At 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, 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 dominates. Bourreau et al. show that the capacity effect always dominates and conclude that congestion would be lower under discrimination. 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. 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. 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 it is not profitable. In other words, competition may drag ISPs into a "prisoner's dilemma" type of situation.

2. MODEL

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

2.1 End-users

In this study, I don't model end-users as individuals but rather the collection of all internet users. In this regard, end-users have differentiated aggregate demand for each CP θ with different congestion sensitivities where θ represents an arbitrary CP (CPs are websites and web applications like YouTube, Facebook, Gmail, cnn.com *etc.*). They pay a unit price p_x to the ISP for their data consumption and they reduce their data consumption under congestion proportional to their congestion sensitivity to a given CP. I define data demand for an arbitrary CP θ as following.

$$x_\theta = x_\theta^0 - d_\theta \left(\frac{x^0}{B} \right) \qquad \text{where; } x^0 = \sum_\theta x_\theta^0 \tag{2.1}$$

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. I 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 denote the end-users' congestion sensitivity to CP θ. The higher d_θ is, the more data consumption reduced due to the congestion. In defining the congestion, $\left(\frac{x^0}{B}\right)$, I prefer to use 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 modeled by the parameter d_θ. The congestion causes dis-utility to end-users which makes data demand under zero-congestion to always be greater than nonzero-congestion ($x_\theta \leq x_\theta^0$).

Next I spell out the demand functions under network neutrality and under discrimination separately.

2.1.1 Under network neutrality

2.1.1.1 Data demand for an arbitrary CP (θ)

I consider a linear demand function under zero-congestion which effectively leads to the following demand function for the CP θ under nonzero-congestion under network neutrality regime;

$$x_\theta(p_x, B) = (a_\theta - b_\theta\, p_x) - d_\theta \left[\frac{(a - b\, p_x)}{B} \right] \tag{2.2}$$

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 $x^0 = (a - b\, p_x)$.

I 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 [9]. With this assumption, I particularly have in mind the bandwidth variable that I explicitly model. Having constant elasticity for this variable is reasonable for the average household that I am modeling. To satisfy constant elasticity property, I assume $\frac{a_\theta}{b_{\theta,}}$ is constant for all θs. I can readily show that the elasticity is indeed constant.[1]

Considering user-time is a limited resource and to prevent unbounded demand, which could potentially violate user-time constraints, I also assume that $a < \bar{a}$ where $\bar{a}$ is finite and positive.

Rearranging equation 2.2 and using constant elasticity assumption with respect to the bandwidth I have the following:

[1]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

$$x_\theta(p_x,B) = (a_\theta - b_\theta\, p_x)\left[1 - \frac{d_\theta\left(\frac{b}{b_\theta}\right)}{B}\right] \tag{2.3}$$

As seen in equation 2.3 the congestion (through the bandwidth) has a counter-clock wise rotational effect on the demand (Figure 2.1).

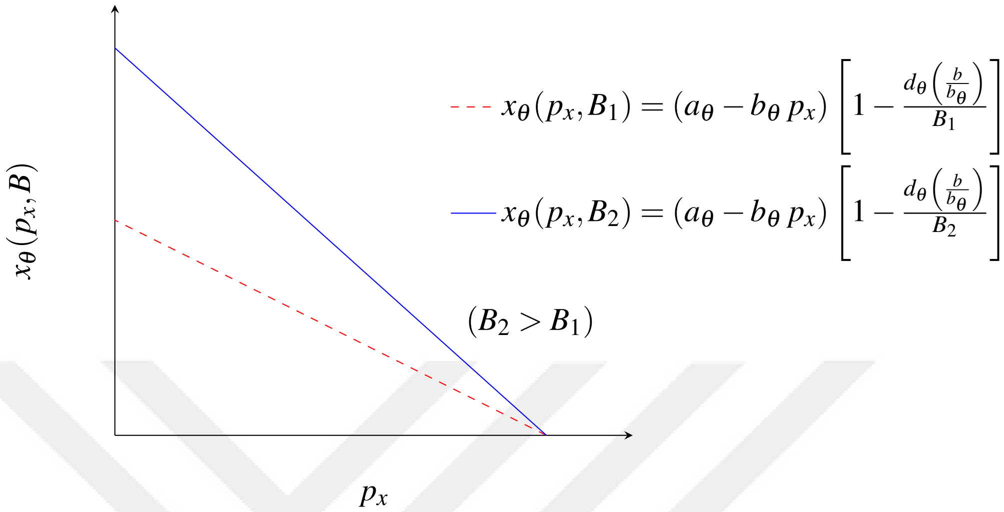


Figure 2.1 : Data demand for an arbitrary CP (θ).

2.1.1.2 Total data demand

Then, the aggregate data demand under network neutrality is given by;

$$\sum_\theta x_\theta(p_x,B) = x(p_x,B) = (a - b\,p_x) - d\frac{(a-b\,p_x)}{B} \tag{2.4}$$

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

2.1.2 Under discrimination

Under discrimination, to keep the model tractable , I assume that a single CP is discriminated by the ISP. I denote the discriminated CP as Θ in the rest of this paper, whereas the collection of all the remaining CPs is denoted as $\Theta^{'}$. 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 which is specified by the ISP. I denote this price as p_B. In other words, the ISP sells bandwidth to the discriminated CP. 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 my setting B_Θ is not for the exclusive usage of the discriminated CP. It acts as a upper bandwidth limit. It means that the discriminated CP is capped by B_Θ, but the remaining CPs can also use B_Θ when the network is overly utilized. In this respect, p_B can be thought as a network access fee rather than a unit price for a commodity (bandwidth).

2.1.2.1 Data demand for the discriminated CP (Θ)

Under discrimination the demand for CP Θ is given by

$$x_\Theta(p_x, B, p_B) = (a_\Theta - b_\Theta\, p_x) - d_\Theta \left[\frac{(a_\Theta - b_\Theta\, p_x) + \frac{B_\Theta}{B}\left(a_{\Theta'} - b_{\Theta'}\, p_x\right)}{B_\Theta} \right] \tag{2.5}$$

where $\left(a_{\Theta'} = a - a_\Theta\right)$, $\left(b_{\Theta'} = b - b_\Theta\right)$, and $\left(d_{\Theta'} = d - d_\Theta\right)$. As can be seen, under discrimination, congestion is defined as

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

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

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

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 $(a_\Theta - b_\Theta\, p_x) + \left(a_{\Theta'} - b_{\Theta'}\, p_x\right) = (a - b\, p_x)$ (Figure 2.2).

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

I summarize above ideas on how congestion term behaves under different relative bandwidth allocations in Figure 2.2. It is insightful to compare CP Θ's demand under discrimination with the demand under neutrality. Rearranging equation 2.5

$$x_\Theta(p_x,B,p_B) = (a_\Theta - b_\Theta\, p_x) - d_\Theta\left[\frac{a-b\,p_x}{B}\right] - d_\Theta\left[\frac{\left(\frac{B}{B_\Theta}-1\right)(a_\Theta - b_\Theta\, p_x)}{B}\right] \quad (2.6)$$

Comparing equation 2.2 with equation 2.6, the first two terms in the CP Θ's demand function are as same as the ones in the demand function under neutrality. The third term reflects the effect of the discrimination. If $B_\Theta < B$,everything else being the same, the demand under discrimination is always less than the demand under neutrality. When $B_\Theta = B$ equation 2.6 converges to equation 2.2.

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

Figure 2.2 : Congestion term for the discriminated CP (Θ) under discrimination.

2.1.2.2 Data demand for the remaining CPs $\left(\Theta^{'}\right)$

Under discrimination, the overall data demand for the remaining CPs, $\Theta^{'}$, is given by

$$x_{\Theta'}(p_x, B, p_B) = \left(a_{\Theta'} - b_{\Theta'}\, p_x\right) - d_{\Theta'} \left[\frac{\left(a_{\Theta'} - b_{\Theta'}\, p_x\right) + \frac{B_\Theta}{B}\left(a_\Theta - b_\Theta\, p_x\right)}{B} \right] \tag{2.7}$$

where the congestion is defined as;

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

Similar to the previous case, as B gets larger relative to B_Θ it becomes as if all bandwidth is exclusively used by $\Theta^{'}$.

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

As B_Θ gets closer to B, it acts as if network neutrality is imposed (Figure 2.3).

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

I can also compare the remaining CPs',Θ', demand under discrimination with the demand under neutrality. Rearranging equation 2.7

$$x_{\Theta'}(p_x, B, p_B) = \left(a_{\Theta'} - b_{\Theta'}\, p_x\right) - d_{\Theta'} \left(\frac{a - b\, p_x}{B} \right) + d_{\Theta'} \left[\frac{\left(1 - \frac{B_\Theta}{B}\right)\left(a_\Theta - b_\Theta\, p_x\right)}{B} \right] \tag{2.8}$$

Comparing equation 2.2 with equation 2.8; first two terms reflect the situation under neutrality. The third term, on the other hand, is positive and reflects the effect of discrimination. If $B_\Theta < B$, everything else being the same, under discrimination the demand for remaining contents $\Theta^{'}$ is always greater than it is under neutrality. The reason is; when $B_\Theta < B$, less congestion is experienced by the remaining CPs, $\Theta^{'}$.

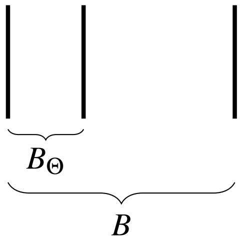
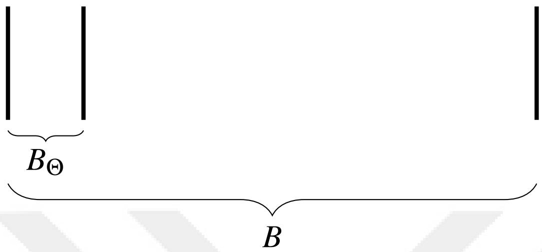
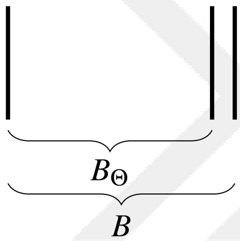
	Bandwidth Allocation	Congestion Term
a)	 (when $B_\Theta < B$)	$= \left[\frac{\left(a_{\Theta'} - b_{\Theta'}\, p_x\right) + \frac{B_\Theta}{B}\left(a_\Theta - b_\Theta\, p_x\right)}{B} \right]$
b)	 (when $B_\Theta \ll B$ or $\frac{B_\Theta}{B} \to 0$)	$\approxeq \left[\frac{\left(a_{\Theta'} - b_{\Theta'}\,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

Figure 2.3 : Congestion term for the remaining CPs ($\Theta^{'}$) under discrimination.

.

2.1.2.3 Total data demand $\left(\Theta + \Theta^{'}\right)$

Combining equation 2.6 and equation 2.8, total data demand under discrimination is given by;

$$x(p_x, B, p_B) = \left(1 - \frac{d}{B}\right)(a - b\,p_x) - \left[\frac{d_\Theta\left(\frac{B}{B_\Theta} - 1\right) - d_{\Theta'}\left(1 - \frac{B_\Theta}{B}\right)}{B}\right](a_\Theta - b_\Theta\,p_x) \tag{2.9}$$

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

Henceforth I assume that $d_\Theta > d_{\Theta'}$. The discriminated CP is one that very sensitive to congestion more than total of the remaining CP market. This assumption gives us an idea about what kind of a CP is considered for discrimination. This could also be considered as the criterion for third-degree price discrimination. This is intuitive in the sense that a CP is willing to pay for bandwidth only if its demand highly sensitive to congestion, *e.g.*, big OTTs like YouTube, Instagram, Facebook and Netflix.

Comparing equation 2.4 with equation 2.9, 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 $\Theta^{'}$. Thus, overall effect is negative.

2.1.2.4 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 equation 2.5

$$x_\Theta(p_x, B, p_B) = (a_\Theta - b_\Theta\,p_x)\left[1 - \frac{d_\Theta\left(\frac{b_\Theta + b_{\Theta'}\left(\frac{B_\Theta}{B}\right)}{b_\Theta}\right)}{B_\Theta}\right] \tag{2.10}$$

Rearranging equation 2.7

$$x_{\Theta'}(p_x, B, p_B) = \left(a_{\Theta'} - b_{\Theta'}\, p_x\right)\left[1 - \frac{d_{\Theta'}\left(\frac{b_{\Theta'} + b_{\Theta}\left(\frac{B_{\Theta}}{B}\right)}{b_{\Theta'}}\right)}{B}\right] \tag{2.11}$$

Demand functions specified above in equation 2.10 and equation 2.11 are counterparts of equation 2.3 under discrimination. It is easy to see that as $B_{\Theta} \rightarrow B$, equation 2.10 and equation 2.11 are both converge to equation 2.3.

2.2 Content Providers

The CPs are different in size and in congestion sensitivity. The CPs' revenue come from advertisements, and they don't directly trade with end users. I 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 don't like congestion because it leads to a decrease the data consumed through them and consequently a decrease in their advertisement revenue. In this manner, CPs have incentive to purchase network bandwidth.

2.2.1 Under network neutrality

2.2.1.1 Profit function of an arbitrary CP (θ)

Under network neutrality, an arbitrary CP θ make profit from advertisement at a unit profit rate r_{θ} which leads to following profit function.

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

Substituting $x_{\theta}(p_x, B)$ with equation 2.2 I have;

$$\Pi_{\theta} = \left[(a_{\theta} - b_{\theta}\, p_x) - d_{\theta}\frac{(a - b\, p_x)}{B}\right] r_{\theta} \tag{2.12}$$

Where r_{θ} is advertisement revenue per unit data.

2.2.2 Under discrimination

2.2.2.1 Discriminated CP's (Θ) bandwidth demand

Under discrimination, the discriminated CP, Θ, has to pay p_B on the basis of unit bandwith and the demand for bandwidth (B_Θ) is determined by solving its profit maximization problem;

By using equation 2.5 define $\widehat{x}(p_x, B, p_B)$

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

Then;

$$\begin{aligned} \max_{\widehat{B}} \quad & \widehat{\Pi} = \widehat{x}(p_x, B, p_B) r_\Theta - \widehat{B}\, p_B \\ \text{s.t.} \quad & \widehat{B} \leq B, \\ & \widehat{B} \geq 0 \end{aligned} \tag{2.13}$$

Where B_Θ solves $\max_{\widehat{B}} \widehat{\Pi}$. I define the profit under optimality as Π_Θ.

$$\begin{aligned} & \frac{\partial \widehat{\Pi}}{\partial \widehat{B}} = \lambda_1 \\ & (\lambda_1 \geq 0 \text{ and } B_\Theta = B) \\ & \frac{\partial \widehat{\Pi}}{\partial \widehat{B}} = -\lambda_2 \\ & (\lambda_2 \geq 0 \text{ and } B_\Theta = 0) \\ & \frac{\partial \widehat{\Pi}}{\partial \widehat{B}} = 0 \\ & (0 < B_\Theta < B) \end{aligned}$$

Then;

$$
\begin{aligned}
&B_\Theta = \sqrt{\frac{d_\Theta}{p_B+\lambda_1}\,(a_\Theta - b_\Theta\, p_x)\, r_\Theta} \\
&(\lambda_1 \geq 0 \text{ and } B_\Theta = B) \\
&B_\Theta = \sqrt{\frac{d_\Theta}{p_B-\lambda_2}\,(a_\Theta - b_\Theta\, p_x)\, r_\Theta} \\
&(\lambda_2 \geq 0 \text{ and } B_\Theta = 0) \\
&B_\Theta = \sqrt{\frac{d_\Theta}{p_B}\,(a_\Theta - b_\Theta\, p_x)\, r_\Theta} \\
&(0 < B_\Theta < B)
\end{aligned} \tag{2.14}
$$

2.2.2.2 Profit function of the discriminated CP (Θ)

Similar to the network neutrality case, under discrimination, CP Θ makes profit on the unit data consumed through it at rate r_Θ, but unlike network neutrality, this time it has to pay p_B to the ISP for the unit bandwidth that it uses, which leads to following profit function. In the below, B_Θ is already chosen to be profit maximizing.

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

Substituting $x_\Theta(p_x, B, p_B)$ with equation 2.5 I have;

$$
\begin{aligned}
\Pi_\Theta = &\left[(a_\Theta - b_\Theta\, p_x) - d_\Theta \left(\frac{a - b\, p_x}{B}\right)\right] r_\Theta \\
&- d_\Theta \left[\frac{\left(\frac{B}{B_\Theta} - 1\right)(a_\Theta - b_\Theta\, p_x)}{B}\right] r_\Theta \\
&- B_\Theta\, p_B
\end{aligned} \tag{2.15}
$$

Where B_Θ is the bandwidth purchased by CP Θ. CPs other than Θ has the same profit function as before;

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

2.3 Internet Service Provider

In this study, the ISP is modeled as a monopoly. It mediates the two sided market, where end users constitutes one side and CPs the other. Under network neutrality constraints, it only charges to end users on the basis of 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 is per unit bandwidth.

2.3.1 Under network neutrality

Under neutrality, the ISP's profit function is

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

Substituting $x(p_x, B)$ with equation 2.4 I have;

$$\Pi_{ISP} = \left(1 - \frac{d}{B}\right)(a - b\,p_x)\,(p_x - c_x) - B\,c_B \tag{2.16}$$

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

2.3.2 Under discrimination

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

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

Substituting $x(p_x, B, p_B)$ with equation 2.9 I have;

$$\begin{aligned}\Pi_{ISP} = &\left(1 - \frac{d}{B}\right)(a - b\,p_x)\,(p_x - c_x)\\ &- \left[\frac{d_\Theta\left(\frac{B}{B_\Theta} - 1\right) - d_{\Theta'}\left(1 - \frac{B_\Theta}{B}\right)}{B}\right](a_\Theta - b_\Theta\,p_x)\,(p_x - c_x)\\ &- B\,c_B\\ &+ B_\Theta\,p_B\end{aligned} \tag{2.17}$$

Where p_B is the price charged to CP with type Θ per unit bandwidth and B_Θ is bandwidth demanded by Θ.

3. EQUILIBRIUM UNDER NEUTRALITY

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

1. The ISP decides 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.

Stage 1

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 (equation 2.16).

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

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

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

for interior solution we have[1]

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

I solve the equilibrium and obtain the following equilibrium values.[2]

[1] The second order conditions are shown to be satisfied. See the appendix

[2] the proof is given in the appendix section

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

$$p_x^* = \frac{1}{2}\left(\frac{a}{b} + c_x\right) \tag{3.1}$$

$$B^* = \sqrt{\frac{d}{c_B}\,(a - b\,p_x^*)\,(p_x^* - c_x)} \tag{3.2}$$

The monopoly price p_x^* is not related to 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. Remember congestion has a rotational effect on the demand. We know that, when demand is linear (i.e is in the form of $k_0 - k_1\,p$) and marginal cost is constant, 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, equilibrium monopoly price does not depend on $\left(1 - \frac{d}{B}\right)$.

If we consider capacity as a quality property and since the pricing decision does not depend on capacity when congestion has only rotational effect on the data demand, using the seminal result by Spence [10] we can conclude that 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 if we look at regulator's perspective (in which I assume that the regulator is welfare maximizer) we see that the ISP under-supplies the capacity since welfare maximizing capacity is given by $\sqrt{\frac{d}{c_B}\,(a - b\,p_x^*)\left(\frac{1}{2}\left(\frac{a}{b} - p_x^*\right) + p_x^* - c_x\right)} < B^*$.

I 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 bandwidth as the aggregate congestion sensitivity, d, increases. One may define $(a - b\,p_x^*)\,(p_x^* - c_x)$ as maximum potential profit of the ISP for given a 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.

4. EQUILIBRIUM UNDER DISCRIMINATION

I then study the equilibrium under discrimination. Under discrimination, apart from the neutrality case, there is an extra decision point where the discriminated CP chooses how much bandwidth to purchase. The equilibrium is the subgame perfect Nash equilibrium (SPNE) of the following two-stage game.

1. The monopolistic ISP decides total bandwidth to install, B^{**}, unit price of data for end-users, p_x^{**} and 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 end-users' decision is trivial. They just respond based on their aggregate data demand as described in equation 2.9. The discriminated CP, on the other hand, decides based on its profit maximization problem as defined in equation 2.13.

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

The solution to this problem gives us the discriminated CP's demand for bandwidth (B_Θ) as given at section (2.2.2).

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

Stage 1

At the first stage, the ISP chooses optimal values for p_x, B and p_B by solving its profit maximization problem taking end-users' and the discriminated CP's best responses which come from the Stage 2 into consideration using the profit function defined in equation 2.17.

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

In the above profit function $x(p_x, B, p_B)$ is the end-users' best response (i.e. their data demand) and B_Θ is the discriminated CP's best response (i.e. its bandwidth demand) which come from the Stage 2" given p_x, B and p_B. The first order conditions of the problem stated above is the following.

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

for interior solution I have

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

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

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

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

I first investigate the possibility of the first outcome in the next proposition.

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

Another way to express this proposition is that the ISP does not set such an excessively high p_B that the discriminated CP does not buy any bandwidth at all. 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. In the other hand, it 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 decreasing effect of the discriminated CP's exclusion offsets the increasing effect of the lower congestion and consequently the exclusion of the discriminated CP reduces the total data demand which would not be profitable for the ISP. Thus, the ISP keeps the discriminated CP in the market.[1] The condition $\left[1-\frac{d_\Theta\left(\frac{b}{b_\Theta}\right)}{B}-\frac{d_{\Theta'}}{B}\right]>0$ realistically holds true for reasonable congestion ranges.

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

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

Under discrimination, the ISP has two revenue sources. First, it makes revenue from the data consumed by end-users, ($x(p_x,B,p_B)\,p_x$). Second, it makes 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. However, since the buy-able bandwidth is capped by B and as a result the 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.[2]

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

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

$$\begin{aligned}p_x^{**}=&\frac{1}{2}\left(\frac{a}{b}+c_x\right)\\&-\frac{1}{2}\left(\frac{d_\Theta\,b_\Theta\,r_\Theta}{b\,(B^{**}-d)-b_\Theta\left[d_\Theta\left(\frac{B^{**}}{B_\Theta^{**}}-1\right)-d_{\Theta'}\left(1-\frac{B_\Theta^{**}}{B^{**}}\right)\right]}\right)\frac{B^{**}}{B_\Theta^{**}}\end{aligned}\tag{4.1}$$

$$B^{**}=\sqrt{\frac{d}{c_B}\,(a-b\,p_x^{**})\,(p_x^{**}-c_x)-\frac{\left[d_\Theta-d_{\Theta'}\left(2\frac{B_\Theta^{**}}{B^{**}}-1\right)\right]}{c_B}\,(a_\Theta-b_\Theta\,p_x^{**})\,(p_x^{**}-c_x)}\tag{4.2}$$

$$p_B^{**}=\frac{d_\Theta}{(B_\Theta^{**})^2}\,(a_\Theta-b_\Theta\,p_x^{**})\,r_\Theta\tag{4.3}$$

[1] the proof is given in the appendix section
[2] the proof is given in the appendix section
[3] the proof is given in the appendix section

Recall that, under network neutrality, the equilibrium price p_x^* is not related to congestion sensitivity d and equilibrium bandwidth B^*. Unlike network neutrality, under discrimination, the equilibrium price p_x^{**} is a function of 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 the profit margin of the discriminated CP. This comes from 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 [11]. 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 discriminated CP decides to purchase is given by following implicit expression,[4]

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

Proposition 5 reflects the equilibrium condition for the interior solution where it simply says that everything else being the same greater profit margin on the end-users 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 [11]. Because, a greater profit margin on the end-users side means a higher price for end-users, and a greater portion of the bandwidth for the discriminated CP then means a lower unit bandwidth price for the discriminated CP.

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. I assume that $\frac{d_\Theta}{d_{\Theta'}}$ is strictly greater than 1. $\left(1-\frac{r_\Theta}{p_x^{**}-c_x}\right)$ must be less than 1 and consequently p_x^{**} must be greater than (c_x+r_Θ). By Proposition 4, I show $p<\frac{1}{2}\left(\frac{a}{b}+c_x\right)$ for interior solution. 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. Advertisement rates on unit data is usually far less than operator profit on unit data

[4]the proof is given in the appendix section

consumption. In this respect, it is quite reasonable to assume that the interior solution exists.

In Proposition 4, I 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 marginal increase in d_Θ on equilibrium outcome. Because, at the equilibrium p_x^{**}, B^{**} and B_Θ^{**} are function of d_Θ itself. Thus, d_Θ cannot be increased freely without changing equilibrium values p_x^{**}, B^{**} and B_Θ^{**}. Fortunately, it is both possible and interesting to study the effect of marginal changes in cost of unit bandwidth investment (c_B) and advertisement rate (r_Θ) on the equilibrium outcome (assuming that the discriminated CP has zero marginal cost per unit data consumed through it; 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 bandwidth price (p_B) and an increase in data price (p_x). The lower price for bandwidth (p_B) makes the discriminated CP purchase even more bandwidth (B_Θ). The growth in 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 Proposition 4 and Proposition 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, data price (p_x) decreases to maintain the demand for data. High bandwidth price (p_B) makes the discriminated CP purchase a lesser 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 incentive 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 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 to get profit. 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 distortion on how much bandwidth is supplied. In other words, under network neutrality, the monopoly short-supplies the bandwidth. Under discrimination, there is an additional distortion caused by two-sidedness: Unconstrained monopoly supplies the bandwidth even less proportional to the difference between congestion sensitivity of the discriminated CP's service and the aggregate congestion sensitivity of the remaining CPs.

5. NEUTRALITY VS DISCRIMINATION

In this section, I compare the equilibrium under network neutrality and discrimination in different aspects.

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

Proposition 6 restates the result given at Proposition 5 while comparing the equilibrium price under network neutrality and the equilibrium price under discrimination. Remember, Proposition 5 is all about how price is allocated between the end-users and the discriminated CP. It does not say anything about how the prices look compared to neutrality. One result of the Proposition 5 is that, as the share of bandwidth purchased by the discriminated CP increases, end-user price increases too. As shown at 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^{**} dictates 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 total bandwidth purchased by the discriminated CP. This is the mechanism I reveal at Proposition 5. On top of this, Proposition 6 demonstrates that, regardless of how price is allocated between end-users and the charged CP, the end-user price (i.e., the unit price of the data) is always lower under discrimination than under neutrality.[1] Depending on the allocation, it gets closer to neutrality price or moves away.

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

[1] the proof is given in the appendix section

As mentioned earlier, total bandwidth (B) is the main instrument for the ISP to set prices to the two sides of the market. It can make total bandwidth scarcer (under-supplies it) and bandwidth price (p_B) goes up, and similarly, it can make it abundant and end-user price (p_x) goes 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 charged CP (d_Θ), and towards other CPs $(d_{\Theta'})$ and advertisement rate (r_Θ). Proposition 7 demonstrates that, under discrimination, total equilibrium bandwidth is always less than total equilibrium bandwidth under neutrality.[2] As stated in the following proposition, the private incentives of the unconstrained monopolistic ISP lead to more congested service for the charged 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, it is not guaranteed to have a better experience with the remaining CPs in terms of congestion. 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.[3]

[2]the proof is given in the appendix section

[3]the proof is given in the appendix section

6. CONCLUSIONS

In this paper, I 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), I 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, I depict a two-sided market with third-degree price discrimination.

In my 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). I 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. The first result is that the ISP always prefers to discriminate. Putting 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.

I 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 the results in combination, my 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

[1] **Economides, N.** (2015). Economic Features of the Internet and Network Neutrality, Working Papers 15-01, NET Institute, https://ideas.repec.org/p/net/wpaper/1501.html.

[2] **Greenstein, S.**, **Peitz, M. and Valletti, T.** (2016). Net Neutrality: A Fast Lane to Understanding the Trade-Offs, *Journal of Economic Perspectives*, *30*(2), 127–50, https://www.aeaweb.org/articles?id=10.1257/jep.30.2.127.

[3] **Economides, N.** (2015). Testimony on Network Neutrality to US Congress, Working Papers 15-14, New York University, Leonard N. Stern School of Business, Department of Economics, https://ideas.repec.org/p/ste/nystbu/15-14.html.

[4] **Pil Choi, J. and Kim, B.C.** (2010). Net neutrality and investment incentives, *The RAND Journal of Economics*, *41*(3), 446–471, https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1756-2171.2010.00107.x, https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1756-2171.2010.00107.x.

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

[6] **Economides, N. and Hermalin, B.E.** (2012). The economics of network neutrality, *The RAND Journal of Economics*, *43*(4), 602–629, http://www.jstor.org/stable/41723347.

[7] **Krämer, J. and Wiewiorra, L.** (2012). Network Neutrality and Congestion Sensitive Content Providers: Implications for Content Variety, Broadband Investment, and Regulation, *Information Systems Research*, *23*(4), 1303–1321, http://www.jstor.org/stable/42004258.

[8] **Bourreau, M.**, **Kourandi, F. and Valletti, T.** (2015). Net Neutrality with Competing Internet Platforms, *The Journal of Industrial Economics*, *63*(1), 30–73, https://onlinelibrary.wiley.com/doi/abs/10.1111/joie.12068, https://onlinelibrary.wiley.com/doi/pdf/10.1111/joie.12068.

[9] **Graves, P. and Sexton, R.** (2006). Demand and Supply Curves: Rotations versus Shifts, *Atlantic Economic Journal*, *34*(3), 361–364, https://ideas.repec.org/a/kap/atlecj/v34y2006i3p361-364.html.

[10] **Spence, A.M.** (1975). Monopoly, Quality, and Regulation, *The Bell Journal of Economics*, *6*(2), 417–429, http://0-www.jstor.org.divit.library.itu.edu.tr/stable/3003237.

[11] **Rochet, J.C. and Tirole, J.** (2003). Platform Competition in Two-Sided Markets, *Journal of the European Economic Association*, *1*(4), 990–1029, https://ideas.repec.org/a/tpr/jeurec/v1y2003i4p990-1029.html.

APPENDICES

APPENDIX A.1 : Proofs of the propositions

APPENDIX A.1

Proofs

Proposition 1

Proof.

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

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^*;

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

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.

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

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

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

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

Proof. 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 equation 2.4. Then, in this case, the profits of the ISP are equal to

$$x(p_x^{**},B^{**},p_B^{**})\,(p_x^{**}-c_x)-B^{**}\,c_B \tag{A.1.1}$$

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

$$x(p_x^{**},B^{**},p_B)\,(p_x^{**}-c_x)-B^{**}\,c_B+B_\Theta\,p_B \tag{A.1.2}$$

If $B_\Theta^{**}=0$, then we must have the former expression equation A.1.1 to be greater than or equal to the latter expression equation A.1.2. In other words,

$$\begin{aligned} &x(p_x^{**},B^{**},p_B^{**})\,(p_x^{**}-c_x)-B^{**}\,c_B \geq \\ &x(p_x^{**},B^{**},p_B)\,(p_x^{**}-c_x)-B^{**}\,c_B+B_\Theta\,p_B \qquad \forall p_B \end{aligned} \tag{A.1.3}$$

Now we want to show that equation A.1.3 is impossible.

Recall equation 2.9;

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

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

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

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

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

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

In the above inequality, by continuity of the demand functions, the second expression on the left hand side can made arbitrarily small by setting a positive B_Θ^{**} so that the inequality holds. At that B_Θ^{**} the inequality stated in equation A.1.3 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.

$\square$

Proposition 3

Proof. Remember in equation 2.14 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 equation 2.13, it is seen that first constraint is binding and 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 following should hold;

$$\frac{\partial B_\Theta(p_x^{**},B^{**},p_B^{**})}{\partial p_B}=0 \tag{A.1.4}$$

The ISP's profit is;

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

Optimality implies that;

$$x(p_x^{**}, B^{**}, p_B^{**} = 0)\,(p_x^{**} - c_x) - B^{**}\,c_B \geq \\ x(p_x^{**}, B^{**}, p_B)\,(p_x^{**} - c_x) - B^{**}\,c_B + B_\Theta\,p_B \quad \forall p_B \tag{A.1.5}$$

From equation A.1.4 we deduce that, if we increase p_B^{**} infinitesimally, the charged CP's demand for bandwidth Θ remains the same as before. Define $p_B = p_B^{**} + \varepsilon = \varepsilon$. This price will lead to folllowing profit;

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

Because B_Θ^{**} is the same at both case, $x(p_x^{**}, B^{**}, p_B^{**} = 0) = x(p_x^{**}, B^{**}, p_B^{**} = \varepsilon)$. Since we have an extra positive term $(B_\Theta\,p_B)$ on the right hand side of equation A.1.5, this contradicts with the optimality condition stated in equation A.1.5 and consequently $p_B = \varepsilon$ is a profitable deviation. Thus, at the equilibrium $p_B^{**} > 0$ should hold.

□

Proposition 4

Proof. First I need to solve the charged CP's, Θ, maximization problem equation 2.13 to get the expression for charged CP's bandwidth demand B_Θ;

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

From equation 2.14 the solution is;

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

$$\text{where } (\lambda_1 \geq 0 \textit{ and } B_\Theta = B), (\lambda_2 \geq 0 \textit{ and } B_\Theta = 0)$$

In Proposition 2, I showed that $B_\Theta^{**} > 0$. Thus I concluded that the second constraint is not binding and $\lambda_2 = 0$. Proposition 3 implies that as long as the first constraint 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;

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

must hold. Solving for interior solution for the ISP;

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

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

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

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 noncross derivatives of the demand function x() to 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;

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

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

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

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

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

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;

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

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

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

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;

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

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

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

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

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

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

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

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;

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

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

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

$\square$

Proposition 5

Proof. At the equilibrium;

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

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

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

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

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

Then;

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

Remember 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;

$$p_x^* = \frac{1}{2}\left(\frac{a}{b} + c_x\right) \tag{A.1.6}$$

From Proposition 4 we have;

$$\begin{aligned} p_x^{**} =& \frac{1}{2}\left(\frac{a}{b} + c_x\right) \\ & - \frac{1}{2}\left(\frac{d_\Theta\, b_\Theta\, r_\Theta}{b\,(B^{**} - d) - b_\Theta\left[d_\Theta\left(\frac{B^{**}}{B_\Theta^{**}} - 1\right) - d_{\Theta'}\left(1 - \frac{B_\Theta^{**}}{B^{**}}\right)\right]}\right)\frac{B^{**}}{B_\Theta^{**}} \end{aligned} \tag{A.1.7}$$

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

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

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 equilibirum $x(p_x^{**}, B^{**}, p_B^{**}) > 0$.
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 equation A.1.7 is positive. Then $p_x^{**} < p_x^*$.

$\square$

Proposition 7

Proof. From Proposition 1 we have;

$$B^* = \sqrt{\frac{d}{c_B}(a - b\,p_x^*)(p_x^* - c_x)} \tag{A.1.8}$$

From Proposition 4 we have;

$$B^{**} = \sqrt{\frac{d}{c_B}(a - b\,p_x^{**})(p_x^{**} - c_x) - \frac{\left[d_\Theta - d_{\Theta'}\left(2\frac{B_\Theta^{**}}{B^{**}} - 1\right)\right]}{c_B}(a_\Theta - b_\Theta\,p_x^{**})(p_x^{**} - c_x)} \tag{A.1.9}$$

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 equation A.1.9 is positive.

Thus; $B^{**} < B^*$ holds.

$\square$

Proposition 8

Proof. Under neutrality, congestion is the same for both Θ and $\Theta^{'}$, given as;

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

Under discrimination, from equation 2.6 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 equation 2.8 congestion for $\Theta^{'}$ 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 $\Theta^{'}$ may be better or worse off depending on the equilibrium quantities $p_x^{**}, p_x^{*}, B^{**}$, and B^{*}.

From equation 2.9 overall congestion $\Theta+\Theta^{'}$ 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 $\left(\Theta+\Theta^{'}\right)$ is worse under discrimination. $\square$

CURRICULUM VITAE

Name Surname : Özgür KAPLANLIOĞLU

EDUCATION

- **B.Sc.** : 2010, Middle East Technical University, Faculty of Engineering, Department of Civil Engineering
- **M.Sc.** : 2015, Bogazici University, The Institute for Graduate Studies in Sciences and Engineering, Department of Financial Engineering

PROFESSIONAL EXPERIENCE:

- 2019 - present Engagement Manager at McKinsey & Company
- 2017 - 2019 Big Data & AI Manager at Vodafone