Imperial College London

INTERIM REPORT

IMPERIAL COLLEGE LONDON

DEPARTMENT OF COMPUTING

Uber Waiting Times and Network Effects

Author:
Remi, Kaan Uzel

Supervisor:
Dr. Yves-Alexandre de Montjoye
Ali Farzanehfar

Second Marker:
Prof. Antonio Filieri

January 23, 2020

Abstract

Hotel companies not owning a single bed, taxi companies not owning a single car, the worlds' most diverse store not owning a single till; in the 21st century we are witnessing a rapid transformation of our way of life, greatly driven by such digital platforms. Ride hailing services in particular have been quite present in our lives, disrupting how we think about transport. As a platform business these services benefit from network effects: their value increases according to their number of users. Due to the existence of network effects, these services are believed to benefit a lot from first mover advantage. However, no one has studied whether there are limits in the network effect for ride hailing services, something crucial to competition watchdogs when deciding if they should let a service like this into the city. In this interim report, we investigate the existing scholarship studying network effects in competitive environments using quantitative techniques. The current literature presents extensive analysis and modelling of various graph based methods, looking at the apparition of "small-world" or "scale-free" phenomena, as well as some structural properties. Although very useful, these models are often very abstract and rarely lend themselves to empirical falsification. Here we report preliminary results on graph based models of platform competition designed in a falsifiable way. We will then validate these models using a diverse range of open data sets on ride hailing services, a major one being the New York TLC.

Acknowledgements

I would like to express my sincerest thanks and gratitude to my supervisor Ali Farzanehfar, for his guidance, support, and for proposing this project.

Contents

Chapter 1

Introduction

Since the late 18^{th} early 19^{th} centuries, the world has been seeing a surge in population of urban cities. This global scale urbanisation quickly turned into a huge economical growth perspective for multiple sectors. Today, with our the ever-increasing degree of digitisation in cities, the market for ride-hailing platforms (RHP) has exploded. In about a decade we have seen dozens of such companies (Uber, Lyft, Kapten, ViaVan, ...) grow to form a multi-billion dollar industry.

Uber-style platforms typically benefit from some form of network externality which results in a rich-get-richer dynamic (Matthew effect) [1], also called preferential attachment [2]. This occurs when an increase in the number of users of a product or service, directly increases its value. For example, additional users riding with a RHP will mean a richer market for drivers which incentivises them to join Uber. With this increase in drivers, waiting times are reduced which in turn attract additional riders to join Uber.

This property of a platform business is referred to as a network effect. Intuition suggests that for businesses like Uber, there could be some inherent limit to how much they can grow. For example, for Uber passengers, the difference between a waiting time of 30s and 1 min is not significant, and so, once Uber reaches a saturation such that waiting times are down to 1 minute, any growth in the number of drivers, will only have a marginal effect on the number of riders.

We propose here an agent based model that can simulate the network effects acting on various such platforms in a city, influenced by its topology. We also investigate the limits of these network effects, as well as the applicability of our model using open-source individual level trip data from multiple cities across the world.

The main objectives of this project is creating an agent based modelling of the local NE in play on ride-hailing platforms, as well as their limits. A key feature we are also basing ourselves on is having as few free-parameters as possible. This will ensure the agent is resilient, extensible, and able to have correct predictions in a variety of settings. The model will be made on the assumption that new users join RHPs according to a system resembling preferential attachment [3]. The model should also be flexible and incorporate the unique topology of each city, for this the population density of its different neighbourhoods will be used to further influence the growth of the RHPs being considered. Our final model should be capable to reproduce observed patterns of market-share between platforms in large cities, where data is most available as well as giving us information on the limits of the NE.

There are a number of challenges in our planned investigation. The main challenge is being able to reproduce observed patterns in data from large cities. The evolution of market shares of RHP such as Black-Cab, Uber, Lyft, etc. are known in a place like London (U.K.). The capacity of our model to reproduce such known evaluations will be a testimony of its success. We will discuss metrics further in the evaluation chapter.

This work will shine a much needed, data-driven light on competition between RHPs. Furthermore, the planned empirical validation of our model opens the door for competition authorities in cities of all sizes to use this research to help guide their decision making regarding the introduction of RHPs in their jurisdictions.

Chapter 2

Background

2.1 What are network effects

A network effect (NE) is an economical or business term that describes a situation where the value of a product or service increases with its number of users [4]. Classical examples of such that have been studied extensively include the citation pattern of scientific papers [5], the World Wide Web (WWW), or even the collaboration graph of movie actors [3]. The term "Network Effect" is used because these observed phenomena result from certain dynamical properties that are intrinsic to some particular types of networks.

These examples naturally exhibit NEs. A scientific paper which has been cited thousands of times is present in a substantial amount of other papers, gaining more exposure and increasing its odds of being cited again. Similarly, a website that has a certain reputation and many frequent users is more likely to be shared, increasing its odds of being linked on another page. Finally, a famous actor will naturally be offered more important roles in upcoming movies and therefore will increase his number of co-stars.

In pop-culture, "The rich get richer" is often the term used to describe such apparently unlimited exponential growth. In the literature, one can often find the terms: "Matthew Effect" [6] from the biblical verse *"For to every one who has will more be given, and he will have abundance..."*, "Cumulative Advantage" [7] or more recently "Preferential Attachment" [3].

2.2 Local vs global network effects

The examples laid out in the previous section are all instances of *global* (or *direct*) NEs. Typically, networks that profit from a direct NE have a single type of node, and the addition of a new node directly increases the value of each user. For example, when the telephone was introduced to the public, each additional phone allowed each owner to call an additional person, increasing their individual value for the service [8].

A typical example of a global NE could be the indexing of a search engine such as Google. As far as we know, Google's indexing grossly works by expanding its graph of known (indexed) websites by crawling through these [9]. What this means is that they go through their known websites in search for new links which they have yet to index. Naturally, having more indexed websites exponentially increases the number of discoverable links, and therefore introducing a global network effect.

Not all networks benefit directly from growth alone: those exceptions are called *local* (or *indirect*) NEs. Platforms that strongly profit from these are ride-sharing or ride-hailing platforms (RHP). If we model Uber as a network, it would be a bipartite network with two different types of nodes. Once the platform has a high number of rider nodes, the business becomes quite lucrative for drivers, increasing the number of driver nodes. This new increase in drivers cuts the waiting time for riders, which in turn increases the number of rider nodes.

Although not as obvious, local network effects are far from being the exception. Take any messaging app such as Facebook Messenger, WhatsApp or Snapchat. The overwhelming majority of their users are only aware of their own contacts, they only interact through the service with people they are directly connected to. In order words, if we were to draw an undirected graph of such a platform, each node would only be aware of its immediate neighbours - essentially creating minuscule local networks [10]. Agents have comparatively low degrees, and are unaware of the overall structure of the graph.

The case of Uber is an interesting one because it also represents a "second" layer of locality: although the increase in drivers is correlated in an increase in users, this is the case only in local geographically areas. RHPs benefit from these local effects, but this has yet to be rigorously studied. In addition to geography, there might also be waiting times that come into effect, which we will study in this paper. Although Uber is a global network, it consists of many smaller local clusters of bipartite graphs, within which we observe indirect NEs.

2.3 The implications for competition

Network effects have always had a considerable impact on financial markets. Historically, economists and competition watchdogs viewed NEs as a significant barrier to entry and protective of strong market positions. Now, more modern economic has recognised the various limits of NEs and negative consequences of platform growth [11]. Old literature suggested that networks exhibiting direct network effects (such as the telephone network) rapidly scaled to a monopoly. The widespread view was that once a particular platform reached a certain scale, it wasn't profitable to build a competing company [12]. As a result, the former would have significant market power, and competition enforcers tended to share this view.

As the understanding of multi-sided platforms advanced, so did the economic literature on both direct and indirect network effects [13]. Rapidly, economists recognised that the existence of NEs didn't necessarily ensure a strong first mover advantage: they could suffer from negative externalises just as much as positive ones. For example, as a search engine gains users it becomes attractive to advertisers, which are beneficial for the business. On the other hand, having more advertisers (or even users) has no effect on user demand [14]. In other cases, this impact might even be negative. Take paper or online newspapers, the increase in number of readers is correlated with the number of potential advertisers, but more advertising diminishes the value for the readers [15]. For social networks, which are an ever-expanding class of NE-prone platform, growth can even invite competition [16]. As the number of profiles gets large, the number of users who are trying to use the platform for disruptive or illegal activities also becomes problematic. This also naturally comes at a higher infrastructural, network and maintenance cost.

The explosion of digital platforms also brings its own set of problem for network effects. With social media, ride-hailing and dating platforms being free to use, there is no consumer "lock-in": the cost of moving to a competing service is often close to none. Additionally, while they may have a preference, users are not exactly limited to using a single platform (e.g. always using Uber instead of Black Cabs). This concept is called "multi-homing" and has been investigated by Rochet & Tirole [13].

All things considered, profiting from a network effect is not the one single secret to a successful business. Platforms on a multi-sided market must capture as many positive externalities, while also mitigating the negative effects that naturally come with growth. Such examples would be Facebook, that all-the-while competing with the (at the time $\approx$ 2004) giants Myspace and Friendster, enforced its strict policy for "real" profiles, requiring a valid college email addresses. It also rigorously enforced its terms of service, banning what it thought was obscene and nude content [17]. On the other hand at Myspace, news reports of minors who lied about their age and child sex predators who preyed on them caused public concern. Advertisers subsequently abandoned the platform and the site floundered [11].

With the development of online platforms benefiting from NE, some may seem to hold a monopoly. This seems to be the case in the mind of most people when they think about Uber, although in cities like London or New York, the competition is considerable. On the other-hand, this competition might not exist in smaller cities such as Clermont Ferrand in France, leading to a real monopoly. Research has yet to determine the reason for this divide, and give a model for predicting when each situation is most likely to occur.

2.4 Graphed based models of network effects

At first, networks of complex topology have been explained using Erdős and Rényi's (ER) random graph theory [18]. Their model starts by generating N vertices, and connects each pair of vertices with a probability p. From this setup, the probability that a randomly chosen vertex has k edges follows a Poisson distribution $P(k) = e^{-\lambda}\frac{\lambda^k}{k!}$, with parameter:

$$\lambda = N\binom{N-1}{k}p^k(1-p)^{N-1-k}$$

Unfortunately, this research dates back to 1960, where data on large networks was nonexistent. Due to this, the theory couldn't be tested on real world data. Today, we have access to virtually unlimited data about all kinds of networks. This allows to empirically validate previously proposed models.

A more recent model proposed by Watts and Strogatz (WS) is the small-world model [19]. In this model, N vertices are aligned to form a $1-D$ lattice, each vertex connected to two nearest other ones. An additional edge is then drawn to any other vertex with probability p. Because this can spontaneously generate "shortcuts" between otherwise far-apart vertices, this process decreases the average distance between each of them, leading to a small-world phenomenon [3] (also commonly known as "6 degrees of separation" [20]).

A common point between both the ER and WS models is that the probability of finding a highly connected node decreases exponentially with k, i.e. vertices with high degree are virtually non-existent. This conflicts with what we can observe in empirical data. Indeed, in (very) large networks, such highly connected vertices actually have a large chance of occurring, following a power-law tail. Barabási and Albert (BA) [3] explain that two generic aspects of real-world networks are missing in these two models.
First, both networks start with N vertices, and no new one are ever added. The models simply attach edges between them using two different methods. In contrast, real-world networks are open, they form by the constant addition of new vertices from their environment. New actors join the industry quite often; the WWW grows exponentially over time with the addition of new websites; and research papers are constantly being published.
Secondly, random networks assume a random and uniform probability for two vertices to be connected. In the real-world we commonly see what BA coin as "Preferential Attachment". Newly published papers are much more likely to cite well known and peer-reviewed research rather than unknown works. This means that the probability of a new paper citing one that already has many citations (a higher degree node) is much higher than citing a paper with few citations (a low degree node). The variety of such existing examples illustrate that the way a new vertex links to existing ones is far from being uniform.

Barabási and Albert base their model on exactly these two features:

- Continuous growth (of the population size N),
- Preferential Attachment (of new nodes to high degree nodes).

Their network starts with a small number of vertices (m_0) and grows by the addition of a single vertex at each time step. New nodes are then connected by $m(\leq m_0)$ edges to m other (different) vertices. Preferential attachment is then implementing by saying that each new node has a probability Π to be linked to node i depending on the degree k_i of that vertex. That is:

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j}$$

This network evolves into a scale-invariant state with the probability that a vertex has degree k follows a power-law distribution with exponent $\gamma_{model} = 2.9 \pm 0.1$.

From comparing these models, we can see that BA's work related best to real-world networks. Thanks to this, we will take inspiration in his work to hopefully bring a new insight on NE profiting businesses and their limits.

2.5 Local network effect models

Current local NE models often focus on precise modelling of a complex network, by setting specific rules for individual clusters within them. Arun Sundararajan (AS) bases his model on the fact that individual vertices have no knowledge at all about the underlying structure of the network they are a part of [10]. The model is graph based and contains N fixed vertices which represent agents. Each agent is associated to a set G_i, the neighbour set of vertex i. This can be seen as the friends, or contacts of the agent. If $j \in G_i$, we can say that j is a neighbour of agent i. This creates an undirected graph of agents connected to their respective "close" acquaintances. A visualisation of such a neighbourhood can be seen in Figure 2.1. In addition, each vertex is initialised with an unchangeable parameter $\theta_i \in [0, 1]$ which influences what AS describes as its payoff value. At each time-step, agents make a binary decision a_i to adopt or not a new abstract network good. The decisions are influenced by their neighbours and modifies their reward, or payoff π_i:

$$\pi_i(a_i, a_{i-1}, G_i, \theta_i) = a_i[u([\sum_{j \in G_i} a_j], \theta_i) - c]$$

where this payoff is dependent on a value function $u(x, \theta_i)$ that quantifies how many neighbours have adopted the good.

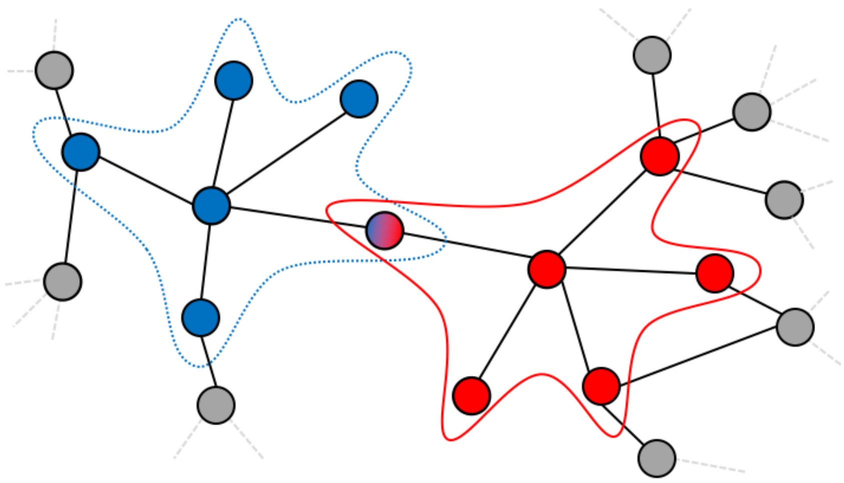


Figure 2.1: Two agents and their neighbourhood clusters G_{blue} and G_{red}

AS's model successfully represents the fact that if your close family, friends or co-workers adopt something, you will to some extent be incentivised to do the same. His goal was to identify the overall adoption trend of the network, and how local adoption can influence the global result, taking a game theory-like approach.

One notable feature that AS describes in the "future work" section of his paper is that agents could adopt one of many incompatible goods, in a dynamical/evolutionary way. In our model we propose just that, where users are able to join one of n RHPs based on a localised measure of the city's topology.

2.6 Lack of empirical validation

Throughout each of the models that we've seen this far, there has been no empirical work on competition and ride hailing platforms. Furthermore, BA verified their model on networks of at most a few hundred-thousand vertices, and the same goes for AS and WS [3, 19, 10]. While this can seem like a substantial size, real world networks now easily go above tens of millions of nodes with a good amount of this data now being collected on astronomical scale.

Obviously, most of this real-world data is not made public, but some have used public datasets of millions of records to build alternative simulations.

R. Tachet et al. have built a "shareability" model of rides and tested their findings on empirical data from metropolises ranging from San Francisco, London and Singapore [2]. Their argument is that our increasingly connected urban areas drastically increase the number of unique trips (made in a personal car) that are similar enough to be merged into one car-pooling, all the while keeping the time delay to a minimum. This would reduce congestion as well as societal, environmental and economical cost in a city.

In their model, two trips are defined to be shareable if they would incur a sharing delay of no more than Δ minutes, relative to a single ride. The authors of the paper suggest a formula for the sharability (S) of rides for any given city:

$$S = 1 - \frac{1}{2L^3}(1 - e^{-L})(1 - (1 + 2L)e^{-2L}) \qquad \text{with} \quad L = \lambda \Delta^3 \frac{v^2(C)}{|\Omega(C)|}$$

with v the average traffic velocity of the city, λ the average rate at which taxi rides are available, and Ω the city's area.

What they were able to demonstrate was that in all the metropolises/megalopolises that they studied, sharability rapidly saturated to near 100% as both the average of trips/h or L grew.

2.7 Contributions

Through this review we've seen how rich the existing literature is in terms of network modelling and theoretical results. Nonetheless, none have yet to study local NEs in the context of ride-hailing platforms using publicly available empirical data. The interesting feature of RHPs is that they naturally exhibit NEs, and it is not known whether or not this growth has a limit.

Chapter 3

Project Plan

In this project we first take inspiration from the literature (particularly Barabási and Albet [3]) to build an agent based simulation of the network effects in play for ride-hailing platforms. We then using existing open-source datasets to extend this model to take into account geographical topological data and to investigate the performance of the model. From this we hope to understand the limits of RHP network effects.

3.1 Simulated Environment

To get initial results, we implemented a simplistic agent based simulation in Python. At each time-step, a new user is generated and added to one of the available ride-sharing platforms. Each platform keeps track of their respective market-shares as a fraction of users they have, over the total users that have been generated. Figure 3.1 below shows the evolution of market shares when a user selects the platform to join uniformly. Note that due to the method of selection, this does not exhibit preferential attachment.

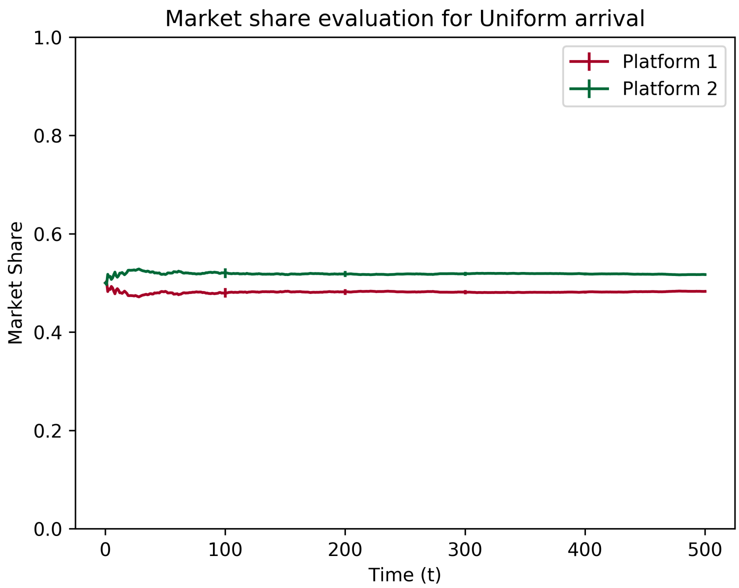


Figure 3.1: Market Share Evolution for 500 time-steps with Uniform selection

Figure 3.2 shows this same evolution, but instead the choice is made proportionally to each platforms' current market-share. Note here that preferential attachment is in play.

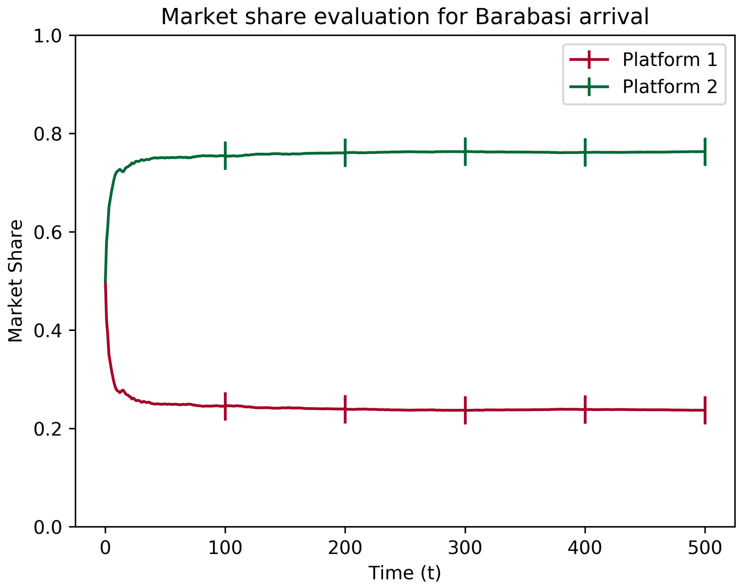


Figure 3.2: Market Share Evolution for 500 time-steps with preferential attachment

After viewing these preliminary results, we wanted to investigate the strength of the first-mover advantage in this very simple model. To do this, we decided to introduce a delay before the market entry of a secondary platform and observe the effect of this delay on the final market shares. Surprisingly the market-shares always lead to a monopoly situation, even with the smallest possible Δ in introduction time. In this model, the newly introduced platform has no chance to gain users. Again this is illustrated in Figure 3.3, where even with a 0.14% (of total simulation time) late arrival condemns the new platform to nearly no part of the market.

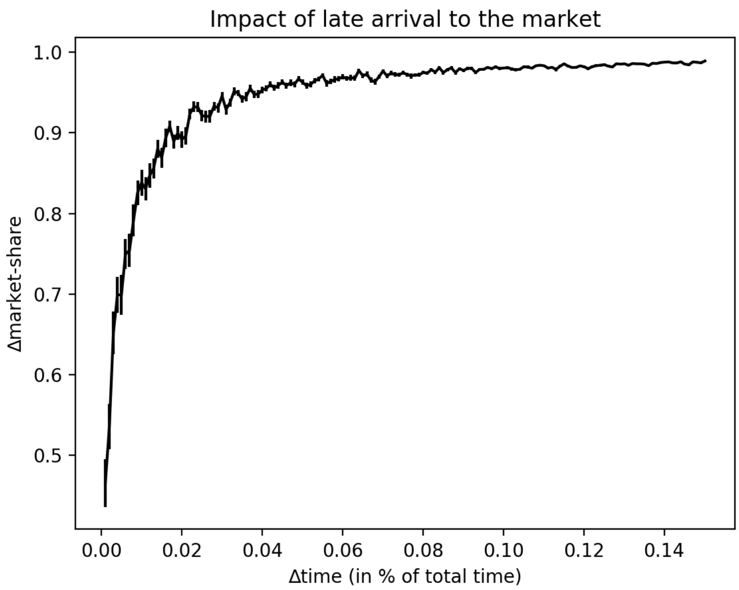
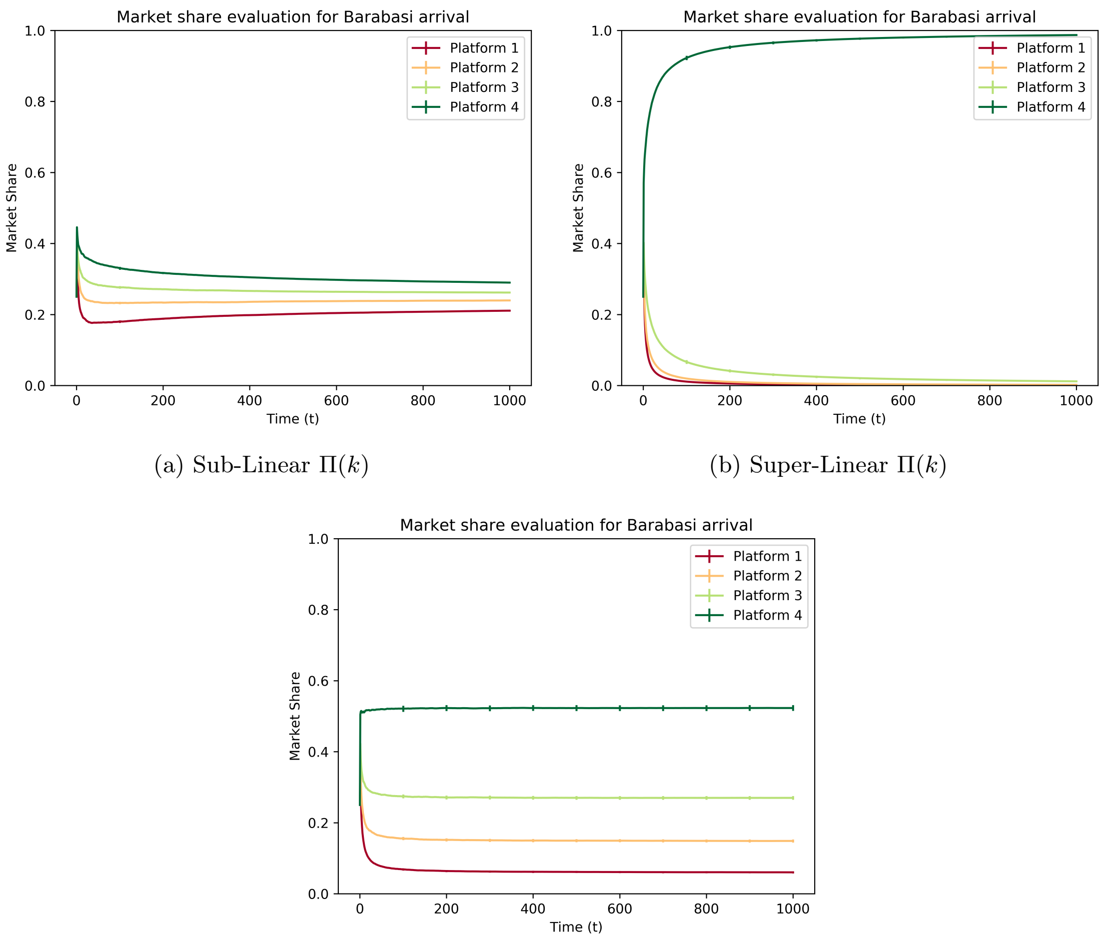


Figure 3.3: Impact of Δ in arrival time (% of total simulation) on market share

The figure demonstrates on the y axis the absolute difference in terms of market-shares between the two considered platforms. The x axis represents how late the secondary platform joined the market, relative to the total simulation time. For example, given this graph we can see that if we ran a simulated environment for 100,000 time-steps (say, seconds), and that the second platform came to market after just $20(= 100,000 * 0.0002)$s, the resulting market shares would be 95% of the market for the first platform, and only 5% for the second one. The divide goes to nearly 100% v.s. 0% after only a few hundred seconds. This means that even in our simple model, there is an extremely strong first-mover advantage.

An extra bit of work was done to reproduce the effects of changing the linearity of Barabasi's model as suggested by himself in a later paper [21]. We are indeed able to reproduce the effects of sub-linearity as well as super-linearity: with $\alpha = \frac{1}{2}$, $\alpha = 1$ and $\alpha = 2$. As we can see in Figure 3.4, preferential attachment completely disappears for the sub- and super-linear models. The former yields the same experimental results as having uniform selection, and the latter leads to a single platform rapidly dominating the market.



(a) Sub-Linear $\Pi(k)$

(b) Super-Linear $\Pi(k)$

(c) Linear $\Pi(k)$

Figure 3.4: Effect of changing the linearity in preferential attachment

3.2 Topological Environment

One key remaining part of the modelling process is to incorporate some fort of geographic topology into the existing network effect (NE). We are currently investigating two independent avenues: the population densities of cities, and the other related to the distance of platform users from each other. These ideas directly relate to the waiting times that users experience, with shorter waiting times resulting in stronger NEs.

The idea is to not only take into consideration the degree of platform nodes when linking a new user node, but to also factor its distance from the average user of each platform.

In essence, the network would be $2-$dimensional and laid out on an $x - y$ plane representing an arbitrary area, usually a city. The topology would represent the population of ride-sharing demand density throughout the city. Platform nodes would carry with them a pair of values, corresponding to the average position of their connected user nodes. At each time-step, a user node would be generated according to the topology, which represents a probabilistic mapping in the grid. An edge is then drawn between new nodes and a platform based on the latter's size and distance to the user.

Before the implementation can begin, data must be found. At the moment a single 1.1 billion records long dataset has been found and will be used to create a model of New York City [22]. Once we've established a local database for these records (>> 300GB) we'll be able to run some analysis on the data and determine the best way to map this to topological information about the city. This being done, we'll be able to start the creation of the above laid-out model.

A third and final avenue would of course be to directly integrate waiting times to our model, although the specifics of this are still to be decided upon.

3.3 Evaluation

The final part of the project will be the evaluation of the developed model. We'll need to use datasets based on various other cities, such as San Francisco, London, Vienna or Singapore (used by R. Tachet et al. in [2]) to generate multiple models. Each model will be assessed using the methods presented in the following chapter. Based on the results, we will then either try to fix our model, finding reasons for which it doesn't work; or look at ways to extend it if it is successful. Such extensions could be taking into consideration the social network of users, their ability to join multiple platforms ("multi-homing" [11]), or their ability to leave a platform.

Chapter 4

Evaluation

The aim of this project being to successfully model the network effects present in urban ride-sharing (ride-hailing) platforms, our main evaluation metric here will be the reproduction of their historical data. We aim to be capable to reproduce the evolution of (e.g.) black cabs market share evolution after the introduction of uber in London, with particular interest in modelling the variations of the graph.

The model could be "trained" using a subset ($\approx 80\%$) of a given dataset, and evaluated on the remaining data. Such training will involve the tuning of any parameter used in the final model, such as the particular method to induce preferential attachment.

As an analytical approach to this evaluation, the mean square error (MSE) has been showed to be a powerful tool for this, mainly in regression problems for machine learning.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

where $\hat{Y}$ is the vector of points from the observed data, and Y the vector of points produced by our model.

We will use this to compare graphs point by point and give an accuracy measure using MSE. Due to the fact that there is no existing model doing this particular modelling, we will not be considering any benchmarks and we will simply report our findings, commenting on why the model works well, or why it isn't able to generalise.

Other metrics such as precision, recall or even f1 score are often used for model evaluation, but mainly in classification problems, and not in a regression problem such as this one. An alternative to the MSE would be determining if our model is constantly over-estimating the market share of a platform instead of underestimating them. This will be made more constant towards the end of the project, as discussed in the planning section.

Chapter 5

Conclusion

There is a real lack of empirical analysis on the implications for competition that Uber-style platforms represent. Network effects govern their booming growth, and regulators currently have no empirically validated tool to guide their decision makings in accepting or rejecting the entry of such platforms in their jurisdictions. Here we aim to offer an agent based simulation that could be tailored to a city's specific topology and shed light on this choice, by understanding the limits of ride-hailing platform network effects.

Bibliography

[1] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.

[2] Remi Tachet, Oleguer Sagarra, Paolo Santi, Giovanni Resta, Michael Szell, SH Strogatz, and Carlo Ratti. Scaling law of urban ride sharing. *Scientific reports*, 7:42868, 2017.

[3] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.

[4] Carl Shapiro, Shapiro Carl, Hal R Varian, et al. *Information rules: a strategic guide to the network economy*. Harvard Business Press, 1998.

[5] Sidney Redner. How popular is your paper? an empirical study of the citation distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 4(2):131–134, 1998.

[6] Robert K Merton. The matthew effect in science: The reward and communication systems of science are considered. *Science*, 159(3810):56–63, 1968.

[7] Derek J De Solla Price. Networks of scientific papers. *Science*, pages 510–515, 1965.

[8] Matthew T Clements. Direct and indirect network effects: are they equivalent? *International Journal of Industrial Organization*, 22(5):633–645, 2004.

[9] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. 1998.

[10] Arun Sundararajan. Local network effects and complex network structure. *The BE Journal of Theoretical Economics*, 7(1), 2007.

[11] Daniel O'Connor. Understanding online platform competition: Common misunderstandings. *Internet Competition and Regulation of Online Platforms (May 2016), Competition Policy International*, 2016.

[12] Jeffrey Rohlfs. A theory of interdependent demand for a communications service. *The Bell journal of economics and management science*, pages 16–37, 1974.

[13] Jean-Charles Rochet and Jean Tirole. Platform competition in two-sided markets. *Journal of the european economic association*, 1(4):990–1029, 2003.

[14] Giacomo Luchetta. Is the google platform a two-sided market? *Journal of Competition Law and Economics*, 10(1):185–207, 2014.

[15] E Glen Weyl. A price theory of multi-sided platforms. *American Economic Review*, 100(4):1642–72, 2010.

[16] Andrei Hagiu and Simon Rothman. Network effects aren't enough. *Harvard business review*, 94(4):64–71, 2016.

[17] INQUISITR. Lindsay lohan banned by facebook. https://www.inquisitr.com/10537/lindsay-lohan-banned-by-facebook/, 2008. [Online; accessed 22-January-2020].

[18] Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[19] Duncan J Watts and Steven H Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440, 1998.

[20] John Guare. *Six degrees of separation: A play.* Vintage, 1990.

[21] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.

[22] NYC Taxi Limousine Commission. TLC trip record data. https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page, 2020. [Online; accessed 21-January-2020].