



Dipartimento di Impresa e Management
Cattedra di Sistemi digitali

BIG DATA E DINAMICHE COMPETITIVE

RELATORE

PROF. Mauro Marè

CANDIDATO

Lorenzo Crosta

MATR. 690681

CORRELATORE

PROF. Gianluca Faella

ANNO ACCADEMICO 2017-2018

INDICE

CAPITOLO PRIMO: BIG DATA E SISTEMI DIGITALI

Introduzione

- 1.1 Definizione
- 1.2 Caratteristiche
 - 1.2.1 Focus: Dati strutturati e non strutturati
 - 1.2.2 Segue: Caratteristiche
- 1.3 Sfide dei Big Data
- 1.4 Catena del valore
 - 1.4.1 Generazione
 - 1.4.2 Acquisizione
 - 1.4.3 Raccolta
 - 1.4.4 Analisi
- 1.5 Come le imprese estraggono valore dai dati

CAPITOLO SECONDO: APPLICAZIONI ED IMPLICAZIONI DEI BIG DATA NEI MERCATI MULTILATERALI

Introduzione

- 2.1 Big data nello scenario competitivo delle multi sided platforms
- 2.2 Il Machine Learning nelle multi-sided platforms
 - 2.2.1 Focus: il Machine learning e lo statistical learning
- 2.3 Recommender System
 - 2.3.1 *Collaborative filtering*
 - 2.3.2 *Content-based filtering*
 - 2.3.3 *Hybrid recommender system*
 - Case study: Netflix*
- 2.4 *On-line advertising*
 - Case study: Case study: Google's Business model*

CAPITOLO TERZO: BIG DATA E PRATICHE ABUSIVE

Introduzione

- 3.1 Big data, potere di mercato e pratiche abusive
- 3.2 Big Data & Essential Facilities Doctrine (EFD)
- 3.3 Big Data & Price Discrimination
- 3.4 Big Data e condizioni contrattuali illecite
 - Case study: The Facebook investigations by the Bundescartellamt*
- 3.5 Algorithmic collusion

CAPITOLO QUARTO: BIG DATA ED OPERAZIONI DI CONCENTRAZIONE

Introduzione

- 4.1 *High Tech Concentrations*
- 4.2 *Big Data & M&A*
- 4.3 *Facebook/WhatsApp concentration case*

CAPITOLO PRIMO
BIG DATA E SISTEMI DIGITALI

Introduzione

La raccolta di dati ha da sempre costituito attività d'impresa. Essa prescinde da uno scambio di informazioni da parte dell'offerta, ovvero l'input da cui vengono trasmessi i dati, uomo o macchina, e l'impresa stessa, la quale, solo recentemente ha considerato essi come asset strategico. La raccolta dati, infatti, è da sempre esistita e tante sono state le applicazioni che ne venivano fatte. Basti pensare al vecchio registratore di cassa meccanico, il quale registrava ogni transazione che passava per la cassa. Questo, tra l'altro, ha permesso agli imprenditori di far stare alla cassa anche persone estranee alla famiglia e di assumere nuovi lavoratori, una grande innovazione nel mondo del lavoro. Ora i registratori di cassa moderni registrano una grande quantità di dati, si sa che cosa si vende, in che giorno, a quale ora, con quale moneta, se si usa la carta fedeltà, a che ora le donne vanno a fare la spesa rispetto agli uomini e via dicendo.

I Big Data, come li conosciamo oggi, sono il frutto della grande innovazione che ha contraddistinto l'economia contemporanea, Internet. È proprio grazie a questo strumento, infatti, che le imprese hanno guardato ai dati come un vero e proprio asset ed hanno iniziato a sviluppare i più efficaci meccanismi di raccolta ed analisi. Con l'avvento della *Internet Economy*¹, diversi business sono nati che hanno come core asset i dati e svariate sono le origini che questi ultimi hanno. Essi provengono da utenti o macchine connesse ad Internet e giungono a soggetti terzi, imprese o macchine, per prendere delle decisioni. Essi sono infatti il contrario dell'intuito/sensazioni.

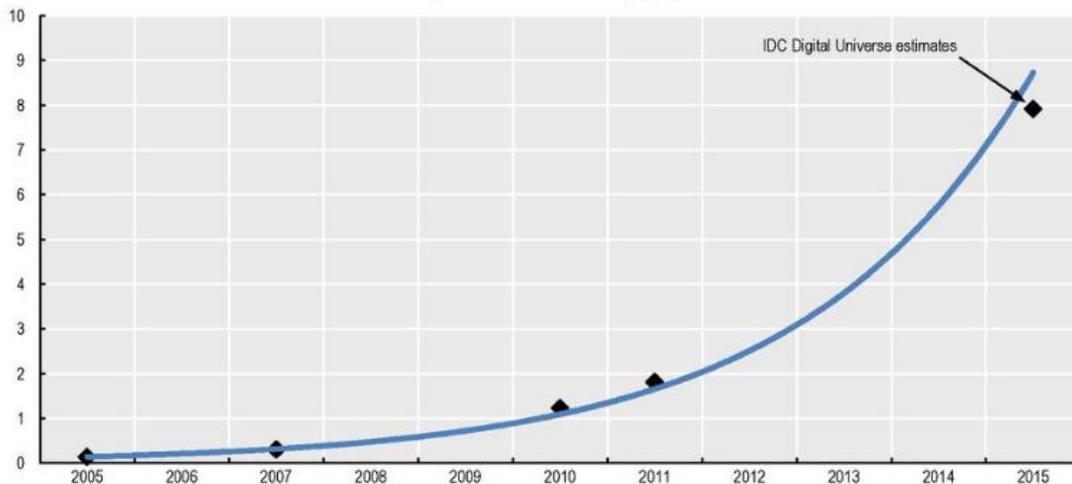
La digitalizzazione di un'attività ha portato ad un effetto moltiplicativo dei dati. Oggi giorno, viviamo in un mondo sempre più interconnesso che genera grosse quantità di informazioni, iniziando dall'uso intensivo che si fa dei social network, motori di ricerca, e-mail e passando infine allo scambio di dati tra macchine (machine to machine). Secondo informazioni tratte da Intel, il 90% cumulativo di dati che ci sono oggi è stato realizzato negli ultimi 3 anni, e la crescita continua. Ogni giorno, vengono generati, raccolti ed elaborati più di 2,5 exabytes (1 exabyte = 1 bilione di gigabytes), con un effetto cumulativo di 8 zettabytes nel 2015 (1 zettabyte = 1 trilione di gigabytes)². Questo aumento così repentino causa il problema di come raccogliere e gestire i dati, ricorrendo ad infrastrutture hardware o software con grandi capacità di calcolo.

¹ Digital economy refers to an economy that is based on digital computing technologies, although we increasingly perceive this as conducting business through markets based on the internet and the World Wide Web. The digital economy is also sometimes called the Internet Economy, New Economy, or Web Economy.

² Fonte: MICHELE AINIS - Economia digitale e Big Data - Roma, 25 ottobre 2017

Figure 1.1. Estimated worldwide data storage

In zettabytes (ZB, trillions of gigabytes)



Source: Based on the IDC (2012) Digital Universe research project.

Come si evince dalla figura, negli ultimi anni c'è stato un incremento esponenziale sia della raccolta dati, ma soprattutto della data analytics. Dal 2009 c'è stato un aumento di dati di 44 volte.

Ma, andando più nello specifico, sono nati sia meccanismi che hanno permesso la raccolta e l'elaborazione di una grossa mole di dati che ha consentito la gestione di tale complessità, sia imprese che traggono profitto grazie alla detenzione (coperta da brevetti, *copyrights*, ma soprattutto dai costi fissi di realizzo dell'infrastruttura di raccolta che costituiscono una barriera all'ingresso per molte imprese), ed all'analisi dei dati.

Realtà leader nella *data storage* e *analytics*, come Apple, Google, Amazon, Facebook, Microsoft, GE, Baidu, Alibaba Group, e Tencent sono solo alcune delle principali aziende affermatesi a livello mondiale. Queste imprese si sono differenziate grazie alla disponibilità di enormi fonti di dati, *data scientist* altamente qualificati, nonché importanti investimenti in infrastrutture. Stesso discorso può estendersi anche ad altre realtà come Uber, Lyft, Didi Chuxing, Palantir, Flipkart, Airbnb, DJI, Snapchat, Pinterest, BlaBlaCar, Ola, Snapdeal e Spotify, aziende i cui modelli di business sono strettamente legati all'ottenimento di dati ed alla loro analisi.

Il valore che i dati generano, grazie alle tecniche di raccolta ed analisi che discuteremo nel proseguo di questo capitolo, ha rivoluzionato e completamente ribaltato il tradizionale rapporto tra consumatori e produttori. In passato, le aziende vendevano i propri prodotti ai rispettivi clienti in cambio di denaro e dati di trascurabile valore. Oggi, le transazioni e in generale ogni interazione con il consumatore, genera preziose informazioni. Questa è la data driven industry.

1.1 Definizione

*“Big data is a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications”*³

Partendo da questa definizione, il termine “Big” sta ad indicare un ammontare di dati tali per cui c’è bisogno di un ingente investimento di capitale sia per la loro raccolta o acquisizione e sia per la loro analisi. I Big data si riferiscono ad un dataset la cui dimensione è troppo elevata per la capacità di calcolo dei software tradizionali per la raccolta e l’analisi dei dati, ovvero il loro volume eccede la capacità di calcolo dei processori dei database convenzionali.

*“Big data is too big, too fast, or too hard for existing tools to process”*⁴

I Big data si riferiscono alla nascita di un *database management* e un *analytical approach* sviluppati per raccogliere, manipolare, estrapolare conoscenza ed infine utilizzare quella conoscenza.

Gli investimenti in Big data includono anche quelli in risorse umane, i cosiddetti *data scientists* con un notevole impatto nella creazione di lavoro, oltre a quelli in soluzioni tecnologiche altamente avanzate, come le piattaforme di database management (ad esempio Hadoop, IBM/Netezza), strumenti di analisi e visualizzazione (come Revolution R), i cosiddetti *text processing*, e le soluzioni di *streaming* in tempo reale.

Il concetto di *database machine* come strumento specializzato nella raccolta e analisi dei dati nacque negli anni Settanta. Con l’aumento del volume di dati, la capacità di raccolta e processuale di un singolo *mainframe computer system* divenne obsoleta e si è passati a sistemi più complessi.

I Big data sono diventati un importante fattore produttivo, un *intangible asset* che richiede capitale umano altamente specializzato, ed attraversano un ciclo produttivo che va dalla loro raccolta e termina con l’elaborazione di informazioni indispensabili per le decisioni strategiche della maggior parte (se non della totalità) delle imprese.

Essi inoltre, come verrà analizzato nei prossimi capitoli, rappresentano una chiave competitiva in differenti dimensioni, verso clienti, fornitori, nuovi entranti e prodotti sostituiti.

1.2 Caratteristiche

Secondo la definizione di Beyer e Laney in “The Importance of Big Data” (2012) ci sono 3 V che caratterizzano i big data: volume, velocità e varietà. Oltre a questi c’è anche un altro aspetto importante che si lega maggiormente al fattore tempo, la variabilità.

³ Big Data Storage and Challenges - M.H.Padgavankar, Dr.S.R.Gupta (2014)

⁴ Big Data Analysis and Storage – Khalid Adam, Mohammed Adam, JASni Mohamed Zain, Mazlina Abdul (2015)



Figura 1.2⁵

Volume: si riferisce all'ammontare di dati. Più dati vuol dire più capacità di immagazzinamento ed elaborazione. Attualmente i più potenti server mondiali raccolgono oltre i 20 zettabytes e sono stimati arrivare a 45 nel 2020.⁶

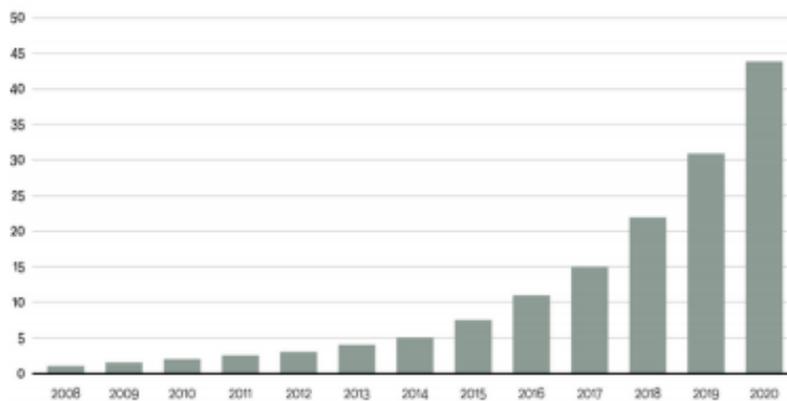


Figura 1.3⁷

Velocità: si riferisce sia alla velocità di generazione di dati, sia alla velocità di immagazzinamento, elaborazione e di estrapolazione di informazioni dall'ammontare dai dati (indicatore di efficienza della *data storage* e *data analytics*).⁸

⁵ Big Data Analysis and Storage – Khalid Adam, Mohammed Adam, JAsni Mohamed Zain, Mazlina Abdul (2015)

⁶ Figure 1.3 indicates that the volume of data stored in the world would be more than 40 zettabytes (1021) by 2020.

⁷ Big Data Analysis and Storage – Khalid Adam, Mohammed Adam, JAsni Mohamed Zain, Mazlina Abdul (2015)

⁸ According to Svetlana Sicular from Gartner, velocity is the most misunderstood big data characteristic (Sicular 2013).



Figura 1.4⁹

Varietà: si riferisce all'esistenza di varie categorie alle quali i big data appartengono. Esso infatti rappresenta un elemento essenziale da conoscere nella loro analisi. I big data contengono testi, audio, immagini, video e molti altri dati strutturati e non, disponibili sia in formato analogico che digitale. Da un aspetto analitico la varietà è la sfida più difficile da affrontare. Molti ricercatori affermano che dominare la varietà e la variabilità è la chiave per il successo della *big data analytics*.

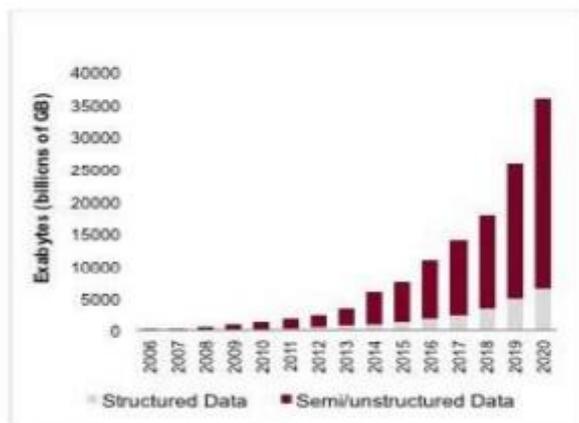


Figura 1.5¹⁰

Variabilità: esso si riferisce alla rapidità con cui un dataset diventa obsoleto nel tempo. Col passare del tempo, che può essere anche molto breve, infatti, un dato può non risultare più compatibile con l'informazione ricercata in un preciso istante. Così come la varietà, esso consiste nell'elemento più difficile da affrontare.¹¹

⁹ Big Data Storage and Challenges - M.H.Padgavankar, Dr.S.R.Gupta (2014)

¹⁰ Big Data Analysis and Storage – Khalid Adam, Mohammed Adam, JASni Mohamed Zain, Mazlina Abdul (2015)

¹¹ One of the big data vendors, IBM has coined additional V for the big data characteristics, which is veracity. By veracity, they address the inherent trustworthiness of the data. As big data will be used e.g. for decision making, it is important to make sure that the data can be trusted. Some researchers mentioned 'viability' and 'value' as the fourth and the fifth characteristics leaving 'veracity' out (Biehn 2013).

1.2.1 Focus: Dati strutturati e non strutturati

I dati, come già affermato più volte, sono complessi da gestire, specialmente quando essi provengono da più risorse (si ricordi la variabilità precedentemente enunciata). Questi dati devono essere collegati tra loro e correlati ad una specifica informazione.

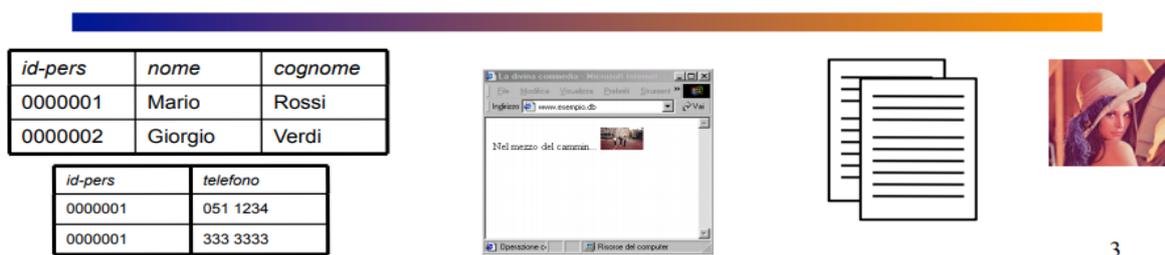
I Big data possono essere definiti come strutturati e non strutturati. Esiste poi una via di mezzo che comprende il 5-10% della totalità dei dati, i dati semi-strutturati.

La struttura dei dati è uno schema con il quale i dati possono essere immagazzinati ed organizzati, in modo tale da renderne semplice l'archiviazione e l'utilizzo.

Figura 1.6¹²

STRUTTURATI

NON STRUTTURATI



I dati strutturati e non strutturati non denotano alcun conflitto tra i due; gli utilizzatori di tali dati scelgono gli uni o gli altri non in base alla loro *data structure*, ma in base all'applicazione che utilizzano: *relational database* per quelli strutturati e diverse altre applicazioni per quelli non strutturati.

Tuttavia, c'è una notevole differenza tra la facilità di analisi dei dati strutturati e la complessità di quelli non strutturati. L'analisi dei dati strutturati, infatti, è un processo tecnologico già maturo; mentre l'analisi dei dati non strutturati rappresenta un'industria nascente con elevati investimenti in R&D, con tecnologie tutt'altro che mature. La sfida per le imprese consiste nell'investire in tecnologie efficienti per l'analisi dei dati non strutturati e aggregare quest'ultimi a quelli strutturati per ottenere informazioni rilevanti di business.

I dati strutturati sono quelli per cui sono facilmente intuibili le loro caratteristiche e possono essere sintetizzate da tabelle che ne permettono una facile lettura e confronto con altri dati. Grazie a questa caratteristica essi sono molto più facili da analizzare rispetto ai dati non strutturati ed hanno, tipicamente, la forma di testo scritto. Essi dunque costituiscono informazioni ad alto grado di organizzazione.

¹² Dati semi-strutturati e non strutturati – Montesi (Unibo)

Figura1.7¹³

<i>id-pers</i>	<i>nome</i>	<i>cognome</i>	<i>id-pers</i>	<i>telefono</i>
0000001	Mario	Rossi	0000001	051 1234
0000002	Giorgio	Verdi	0000001	333 3333

Facendo un banale esempio, raccogliendo notizie su vari attentati di aerei da diversi quotidiani di cronaca, è possibile individuare diversi elementi, raggruppati in cluster, che compongono il fatto.

Intestazione	Attentato 1	Attentato 2	Attentato 3	Attentato 4
Chi	Aereo	Nave	Pullman turistico	Aereo
Cosa	Dirottamento	Sequestro	Sequestro	Dirottamento
Quando	17/02/2014	10/12/2013	10/01/2014	01/11/2013
Partenza	Addis Abeba	Istanbul	Gerusalemme	Londra
Destinazione	Roma	Marsiglia	Hebron	New York
Passeggeri	200	1000	56	250
Feriti	0	5	1	10
Vittime	0	0	1	0
Fonte	Corriere della Sera	Giornale2	Giornale3	Giornale4

Figura 1.8¹⁴

Questo è un esempio di come è possibile strutturare i dati scomponendo le notizie nei loro elementi che la compongono, ovvero i dati.

I dati strutturati, come verrà analizzato più avanti, sono raccolti nei cosiddetti *Relational database management systems* (RDMSs), hanno tipicamente la forma di testo scritto e sono i più semplici da analizzare. Esempi di dati strutturati sono: i numeri di telefono, transazioni finanziarie, vendite a volume ed in valore ecc.

I dati non strutturati o non relazionali, sono essenzialmente l'opposto. Essi hanno al loro interno una struttura, ma è difficile da sintetizzare con schemi o modelli. Dunque non è presente uno schema, come nel caso di oggetti multimediali o file di solo testo narrativo.¹⁵



¹³ Dati semi-strutturati e non strutturati – Montesi (Unibo)

¹⁴ Differenze tra dati strutturati, semi strutturati e non strutturati (Bucap.it)

¹⁵ La disciplina che studia come manipolare questi dati è l'*Information Retrieval*.

Essi, dunque, possono essere costituiti da un testo o da un qualsiasi altro contenuto multimediale, generati da uomo o da macchina. Come vedremo più avanti, essi vengono immagazzinati ed analizzati da database non relazionali come NoSQL.

I tipici dati non strutturati generati dall'uomo includono: file testuali, e-mail, social media, siti web (come Youtube o Instagram), Mobile data, comunicazioni (telefonate, messaggi), Media (MP3, foto digitali, file audio o video) e le cosiddette Business applications (come quelli generati grazie ad Office).

I tipici dati non strutturati generati dalle macchine includono: immagini satellitari, dati scientifici, videosorveglianza digitale e dati provenienti dai sensori.

	Structured Data	Unstructured Data
Characteristics	<ul style="list-style-type: none"> • Pre-defined data models • Usually text only • Easy to search 	<ul style="list-style-type: none"> • No pre-defined data model • May be text, images, sound, video or other formats • Difficult to search
Resides in	<ul style="list-style-type: none"> • Relational databases • Data warehouses 	<ul style="list-style-type: none"> • Applications • NoSQL databases • Data warehouses • Data lakes
Generated by	Humans or machines	Humans or machines
Typical applications	<ul style="list-style-type: none"> • Airline reservation systems • Inventory control • CRM systems • ERP systems 	<ul style="list-style-type: none"> • Word processing • Presentation software • Email clients • Tools for viewing or editing media
Examples	<ul style="list-style-type: none"> • Dates • Phone numbers • Social security numbers • Credit card numbers • Customer names • Addresses • Product names and numbers • Transaction information 	<ul style="list-style-type: none"> • Text files • Reports • Email messages • Audio files • Video files • Images • Surveillance imagery

Figura 1.9¹⁶

I dati semi-strutturati sono dati con struttura parziale. Essi presentano caratteristiche sia dei dati strutturati che di quelli non strutturati. L'e-mail è un tipico esempio di informazione contenente dati strutturati e non; infatti sono presenti sia le connotazioni anagrafiche, sia i contenuti testuali o multimediali.

Il formato principale per la rappresentazione dei dati semi-strutturati è XML.¹⁷ Esso può essere utilizzato sia per rappresentare dati strutturati, ad esempio allo scopo di scambiarli tra diverse applicazioni, che per rappresentare dati semi-strutturati, sfruttandone la flessibilità e la possibilità di indicare sia i dati che lo schema. Nel primo caso, i dati possono risiedere all'origine in un sistema relazionale, per poi essere convertiti in XML. Nel secondo caso, il modello relazionale non risulta essere particolarmente adatto alla gestione di questi dati. XML nasce per scambiare dati tra applicazioni e per rappresentare dati comprensibili anche da esseri umani. Il modello relazionale utilizzato per i dati strutturati, presenta dei limiti che vedremo con un esempio.

Consideriamo un libro e la sua rappresentazione relazionale

¹⁶ Dati semi-strutturati e non strutturati – Montesi (Unibo)

¹⁷ In computing, Extensible Markup Language (XML) is a markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable.



Autore	Titolo	Nascita	Testo
Dante Alighieri	La Divina Commedia	1265	Canto I Nel mezzo del...

Figura 1.10

```

<html>
<head><title>La divina commedia</title></head>
<body>
<h1>La Divina Commedia</h1>
<h2>Canto I</h2>
Nel mezzo del cammin di nostra vita<br/>
...

...
</body>
</html>

```

Il campo testo contiene migliaia di caratteri senza alcuna struttura. Per aggiungere altre informazioni occorre estrapolarle dal testo e modificare la struttura della tabella.

La struttura dei dati semi-strutturati è irregolare o parziale e lo schema è costituito a posteriori (*data guide*), è molto ampio ed evolve rapidamente.

Dunque, riassumendo, la struttura è irregolare o parziale. Lo schema è costruito a posteriori (*data guide*). Lo schema è molto ampio. Lo schema evolve rapidamente. Le differenze tra schema e dati non sono significative. Lo schema viene modificato. Lo schema viene comunicato insieme ai dati. Lo schema non impone vincoli inappellabili. Le interrogazioni riguardano anche lo schema. Nel caso di XML, sono rilevanti l'ordine e l'annidamento reciproci dei dati. Essendo ordinati, i dati sono rappresentati tramite liste.

Relazionali	Semi-strutturati
Netta divisione tra schema e dati	Schema parziale e con caratteristiche simili ai dati
Basati sul concetto di insieme	Basati sul concetto di lista
Non ordinati	Ordinati
Non annidati	Annidati

Figura 1.11¹⁸

Relazionali	Non-strutturati
Netta divisione tra schema e dati	Nessuno schema
Linguaggio di interrogazione	Linguaggio di ricerca
Modello booleano (correttezza & completezza)	Modello basato su Ranking
Aggiornamenti ed interrogazioni parziali	Aggiornamenti totali

¹⁸ Dati semi-strutturati e non strutturati – Montesi (Unibo)

1.2.2 Segue: Caratteristiche

Un aspetto microeconomico importante è la struttura dei costi per la creazione dei big data. Essi implicano, infatti, zero costi marginali di produzione e distribuzione. In altre parole, una volta acquisite sia le competenze tecniche di analisi, sia le infrastrutture di raccolta ed elaborazione, produrre e distribuire dati digitali non apporta costi incrementali (i costi marginali sono praticamente nulli).

A differenza dell'olio, o di qualsiasi altra commodity, i Big Data non consistono in un grande ammontare dello stesso prodotto poiché, con eccezione dei duplicati, i dati digitali sono diversi ciascuno dall'altro. Essi, attraverso codici binari, rappresentano varie sfaccettature del mondo. L'olio ha caratteristiche specifiche, è nero e vischioso. I big data non hanno le stesse proprietà e, ancor di più, non sono costanti nel tempo. Inoltre, la stessa informazione può essere fornita da diversi dati e gli stessi dati possono fornire diverse informazioni. Questa caratteristica è nota come ubiquità dei dati (Paragrafo 3.2). Ad esempio, una ricerca fatta sulla disponibilità a pagare di un *target* di individui per un bene/servizio, fatta sia attraverso i dati provenienti dalle carte di credito, sia dagli acquisti *on-line*, può condurre allo stesso risultato. Così come, processando gli stessi dati sugli acquisti *on-line*, si possono ottenere informazioni sia sugli orari tipici in cui vengono effettuati acquisti sulle piattaforme di e-commerce, sia a quale età gli individui sono maggiormente soliti effettuare acquisti in rete. In sostanza, i dati digitali su cui un'impresa tenta di estrapolare informazioni, spesso ammettono sostituti. Si pensi ancora alle numerose imprese produttrici di applicazioni per gli smartphones che collezionano dati sulla geolocalizzazione dei proprietari dei device.

I big data sono inoltre non rivali e non escludibili nel consumo. In altre parole, una volta divulgati, il consumo di dati di un individuo o entità, non preclude la possibilità di consumo di altri; e nessun consumatore di dati digitali può impedire ad altri di utilizzare gli stessi dati per fini decisionali. Le due entità, infatti, non competono sull'approvvigionamento di dati, bensì sulle decisioni strategiche che ne derivano, che sono frutto di un'analisi ed una raccolta fatta a priori. Quest'ultime due fasi, la raccolta e l'analisi, pongono un dubbio che apre un dibattito molto acceso sul fronte antitrust. Ovvero, ci si chiede se l'accesso ai dati digitali possa essere escluso da chi ne detiene la proprietà. Ci si interrogherà (nel Capitolo 3) sulla possibilità di uso esclusivo dei dati di chi detiene le infrastrutture di raccolta ed elaborazione. Vedremo infatti, che ci sono casi in cui solo un'impresa ha accesso esclusivo a determinati dati. Tuttavia, In questo paragrafo ci si è voluti concentrare esclusivamente sulle caratteristiche peculiare dei dati, rimandando alle implicazioni concorrenziali ai prossimi capitoli.

1.3 Sfide dei Big Data

Le caratteristiche citate nel paragrafo precedente, ovvero le 4 V, l'ubiquità, la strutturabilità, la non rivalità e non escludibilità unite alla struttura dei costi dei dati, apportano delle sfide complesse da affrontare per chi ha a che fare con l'acquisizione, immagazzinamento, gestione ed analisi dei big data. Sviluppare applicazioni, hardware o software, per la raccolta e l'analisi dei big data comportano diverse sfide da affrontare.¹⁹ Qui ne verranno indicate alcune:

- Rappresentazione: molti dataset hanno vari livelli di eterogeneità per tipo, struttura, significati semantici, organizzazione, granularità e accessibilità. La data representation mira a fornire significatività ai dati, sia per l'analisi attraverso i computer, sia per l'interpretazione degli utenti. Ciononostante, un'impropria rappresentazione dei dati potrebbe ridurre il valore dei dati originali e ostacolare l'analisi effettiva dei dati.
- Compressione dati ed eliminazione dati ridondanti: tipicamente, c'è un alto grado di ridondanza dei dati. Ad esempio, molti dati generati dai sensori sono fortemente ridondanti; essi vengono prima filtrati e poi compressi secondo il loro ordine di grandezza.
- Gestione del ciclo di vita dei dati: Il valore dei dati è funzione della "giovinezza" dei dati. Il valore della ricerca analitica fatta su un dataset, pertanto, dipende dal processo decisionale di quali dati scartare e quali invece registrare.
- Gestione analitica: il processo analitico consiste nel processare una massa eterogenea di dati in un tempo limitato, a causa della variabilità dei dati. I tradizionali RDBMSs hanno performance deficitarie in scalabilità ed espandibilità, sono quindi poco efficienti nel processare grossi volumi di dati. I database non relazionali (per i dati non strutturati) stanno iniziando ad essere più performanti e sono candidati a diventare il mainstream nell'analisi dei big data. Essi tuttavia hanno ancora qualche problema in alcune applicazioni. Molte imprese hanno scelto un database che ha un'architettura mista tra il RDBMSs e i database non relazionali (ad esempio Facebook e Taobao).
- Confidenzialità: Bisogna avere confidenza con i dati. Molte imprese non hanno al loro interno le *skills* necessarie per processare grandi ammontari di dati, per questo esternalizzano questa attività a professionisti che hanno a disposizione il tool per la raccolta e l'analisi dei big data.

¹⁹ The sharply increasing data deluge in the big data era brings about huge challenges on data acquisition, storage, management and analysis. Traditional data management and analysis systems are based on the relational database management system (RDBMS). However, such RDBMSs only apply to structured data, other than semi-structured or unstructured data. In addition, RDBMSs are increasingly utilizing more and more expensive hardware. It is apparently that the traditional RDBMSs could not handle the huge volume and heterogeneity of big data. The research community has proposed some solutions from different perspectives. Big Data Storage and Challenges - M.H.Padgavankar, Dr.S.R.Gupta (2014).

- Riservatezza: l'analisi di dati, soprattutto quelli sensibili, deve essere supportata da adeguate misure di sicurezza per prevenirne l'integrità.
- Espandibilità e scalabilità: il sistema di raccolta e analisi dei dati deve processare dati presenti e futuri. L'algoritmo analitico deve essere in grado di processare dataset più complessi ed in espansione.
- Consumo energetico: le fasi di raccolta, elaborazione e trasmissione, con la crescita del volume dei big data richiedono sempre più un notevole consumo di energia. Tuttavia anche i meccanismi di risparmio energetico stanno crescendo al fine di assicurare espandibilità e scalabilità
- Cooperazione: l'analisi dei big data è una ricerca interdisciplinaria, che richiede esperti in differenti campi per "esaurire" tutto il potenziale di un dataset. Diversi ingegneri o scienziati devono infatti poter accedere ai big data.²⁰

1.4 Catena del valore

La *supply chain*²¹ si compone di quattro fasi: generazione, acquisizione, raccolta ed analisi. Le prime due rappresentano la macro fase dell'approvvigionamento del dato grezzo, la raccolta la fase della conservazione e l'analisi quella della produzione in cui vi è l'apporto di nuovo valore, grazie all'elaborazione dei dati grezzi e all'informazione che vi si ricava.

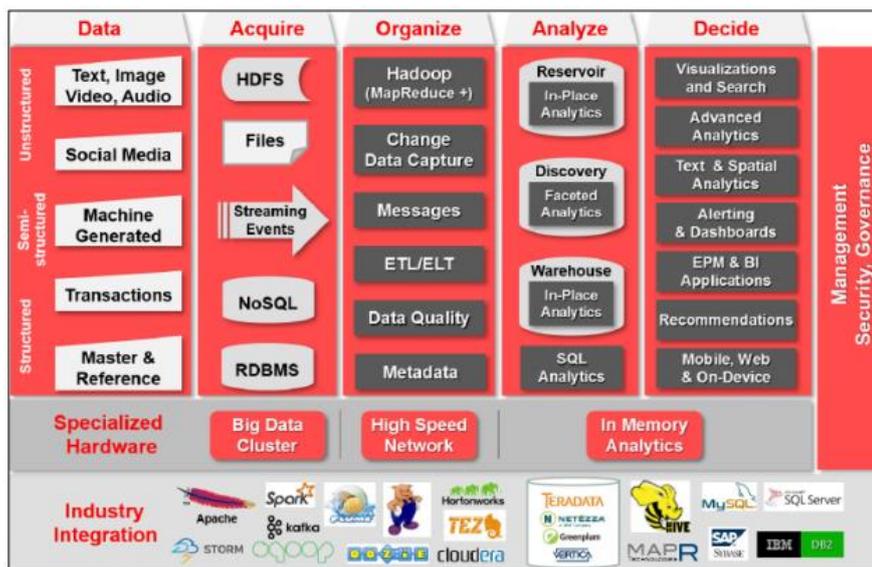


Figura1.12 Fonte: Oracle, Information Management and Big Data, Reference Architecture Overview

²⁰ "A comprehensive big data network architecture must be established to help scientists and engineers in various fields access different kinds of data and fully utilize their expertise, so as to cooperate to complete the analytical objectives." Big Data Storage and Challenges - M.H.Padgavankar, Dr.S.R.Gupta (2014).

²¹ "A supply chain is the collection of processes and resources required to make and deliver a product to the final customer" - University of Pittsburgh, Department of Industrial Engineering.

1.4.1 Generazione

Questa è la prima fase della catena del valore e consiste nella produzione del dato grezzo. Come detto, negli ultimi decenni, con l'avvento di internet prima e con quello dell'*internet of things* dopo, l'ammontare di dati è cresciuto esponenzialmente. Una grossa mole di ricerche in internet, nei motori di ricerca piuttosto che nei blog, di condivisioni sui social network, di scambi di messaggi, e-mail, segnali dai più svariati sensori vengono generati ogni giorno. Questi dati sono strettamente connessi alla vita di tutti i giorni dell'uomo o macchina in questione e sarebbero inutili se presi singolarmente ma, attraverso la registrazione e conservazione di un grosso volume di dati, utili informazioni ne vengono estrapolate (si pensi, ad esempio, agli *hobbies* o agli orientamenti religiosi, politici, sessuali che vengono esplicitati nella propria home page di Facebook; oppure agli spostamenti geografici di tanti automobilisti che permettono a Google di avere informazioni dettagliate sul traffico; oppure ancora alle ricerche che vengono fatte su Amazon che permettono a quest'ultimo di identificare i gusti di una persona e consigliare di conseguenza libri da leggere; è persino possibile prevedere stati d'animo e comportamenti umani...). La tipologia di dati raccolti è estremamente differenziata e le informazioni che ne vengono dedotte sono decisamente utili nelle strategie delle imprese.

1.4.2 Acquisizione

La seconda fase è cruciale, essa consiste nel collezionare, trasmettere ed effettuare una cosiddetta *data pre-processing phase*. Una volta formatosi il dataset di riferimento, bisogna utilizzare un efficiente meccanismo di trasmissione per inviarlo al *tool* di *storage management system* che supporta la fase analitica che si vuole affrontare. Il dataset da trasmettere può contenere molti dati ridondanti o inutili, che necessariamente riducono

lo spazio per la registrazione, ostacolando il processo successivo di analisi. Ad esempio, nei dataset registrati dai sensori per il monitoraggio ambientale, c'è molta ridondanza dei dati. La tecnologia che viene utilizzata in questo caso per l'eliminazione dei dati superflui è la cosiddetta *data compression*. Perciò la fase di *data pre-processing*, in cui c'è l'eliminazione delle inefficienze, è indispensabile per assicurare un'efficiente raccolta ed elaborazione di dati.

1.4.3 Raccolta

La *big data storage* si riferisce all'archiviazione e alla gestione della raccolta (elaborazione) di dati e si focalizza al raggiungimento e della disponibilità all'accesso e dell'attendibilità dei dati.²²

²² "We will review important issues including massive storage systems, distributed storage systems, and big data storage mechanisms." - Big Data Storage and Challenges - M.H.Padgavankar, Dr.S.R.Gupta (2014).

Grazie ad internet, si è sviluppato il *cloud computing*²³ che ha permesso la registrazione automatizzata di dati verso risorse preesistenti, i server. L'obiettivo delle infrastrutture di storage service è quello di fornire lo spazio necessario alla raccolta così da ridurre il costo unitario di archiviazione di ciascun dato e, d'altro canto, essere un'interfaccia funzionale alla fase di analisi. Esse hanno risolto il problema di come collezionare e registrare grandi ammontari di dati, integrare dati eterogenei e come processare dataset di larga scala per permetterne l'analisi.

*"Data collection, data storage, data process, data analysis, data application will be the basic task of the enterprise performance in wisdom economy era. Judgement and decision based on data will become the skills and means of enterprises for development."*²⁴

Come accennato nei primi paragrafi, i *Relational database management systems* (RDBMSs)²⁵ sono i sistemi di registrazione tradizionali, elaborati per memorizzare i dati strutturati. I grandi server processati attraverso il RDBMSs riuscivano a garantire performance elevate nel gestire la variabilità orizzontale dei dati strutturati e accessibilità per le *Big data applications*; tuttavia, con l'aumento del volume di dati e dell'eterogeneità degli stessi in maniera più che proporzionale all'aumento della capacità di immagazzinamento dei RDBMSs, ci si è spostati verso altre soluzioni. Questo a causa dell'enorme volume di dati non strutturati che hanno iniziato ad essere generati. La sfida di riuscire a processare grossi ammontari di dati semi-strutturati e non strutturati, estremamente eterogenei tra loro, ovvero risolvere il problema della *computational scalability*, spetta alle piattaforme di *data analysis*, come Map Reduce. Essa è una piattaforma detta *schema-free* e *index free*; caratteristiche che hanno permesso di dotare tali sistemi di flessibilità e permettere di lavorare con i dati non e semi strutturati. Tuttavia queste piattaforme funzionano solo una volta registrati i dati si servono quindi dei cosiddetti NoSQL²⁶ e NewSQL servers, dove SQL sta per *structured query language* e NoSQL per *"not only SQL"*, che sono strutture complesse di network fra server, per farla breve. Le SQL sono diventate un obiettivo cruciale per tali sistemi.

²³Cloud computing is usually defined as a type of computing that relies on sharing pooling computing resources rather than having local servers or personal devices to handle applications. Divyakant et al. (2011).

²⁴ Big Data Analysis and Storage – Khalid Adam, Mohammed Adam, JASni Mohamed Zain, Mazlina Abdul (2015)

²⁵ Relational database management systems (RDBMSs) are traditional storage systems designed for structured data and accessed by means of SQL. RDBMSs are facing challenges in handling Big Data and providing horizontal scalability, availability and performance required by Big Data applications. In contrast to relational databases, MapReduce provides computational scalability, but it relies on data storage in a distributed file system such as Google File System (GFS) or Hadoop Distributed File System (HDFS).NoSQL and NewSQL data stores have emerged as alternatives to Big Data storage. Big Data Analysis and Storage – Khalid Adam, Mohammed Adam, JASni Mohamed Zain, Mazlina Abdul (2015)

²⁶ NoSQL systems typically adopt the MapReduce paradigm and push processing to the nodes where data is located to efficiently scale read operations. Divyakant et al. (2011)

Le principali caratteristiche di tali sistemi sono la gestione mediante schemi flessibili e l'eliminazione delle inefficienze come la ridondanza di dati, che causano un sovraccollamento di dati nei server duplicandone i costi di raccolta.

Le principali piattaforme esistenti sono Google File System (GFS) e Hadoop Distributed File System (HDFS). Hadoop è un sistema *open source* fondato da Apache Software Foundation. I principali contributori alla realizzazione di tale progetto sono stati Yahoo, Facebook, Citrix, Google, Microsoft, IBM, HP, Cloudera ed altri.²⁷

Queste piattaforme fanno parte della realizzazione di un design di una tecnologia di *cloud computing*, che comprende diversi device fisici e di supporto alla raccolta di dati. Le risorse fisiche si trovano al centro del nodo principale (*master node*) dell'architettura, il quale è il responsabile per il monitoraggio e la gestione delle operazioni giornaliere che si svolgono nei nodi periferici e del normale stato di tali nodi. I dati si trovano registrati in quest'ultimi nodi che riescono ad equilibrare la distribuzione dei dati fra i vari nodi. Il problema della registrazione di dati eterogenei dei RDMSs è risolto frazionando e suddividendo per categorie il totale ammontare di dati.

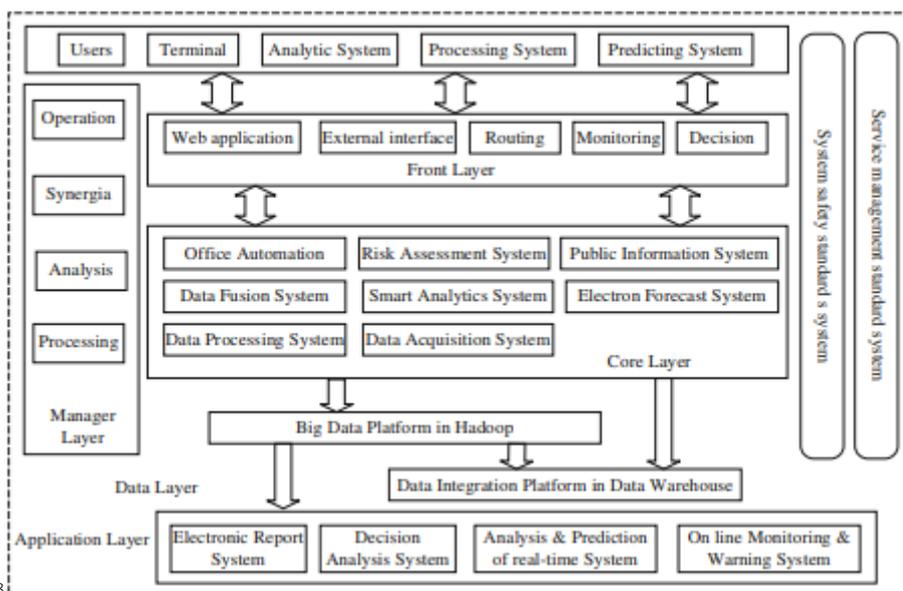


Figura 1.13²⁸

Il sistema di registrazione parte con la richiesta inviata dal client, il cui messaggio passa per il *master node* server attraverso il network con altri server. I dati vengono poi automaticamente dislocati verso i nodi dove ne vengono implementate le operazioni. Questo è il cosiddetto metodo *Partition distributed storage*. Con l'utilizzo degli strumenti di visualizzazione e di analisi i dati arrivano all'utilizzatore finale attraverso il *terminal display*. Questo può collegarsi all'Internet client terminal e accedere al sistema.

²⁷ Storage Architectures for Big Data in the Cloud – Sam Finenberg, HP Storage CT Office (Maggio 2013)

²⁸ Real World Big Data Architecture Splunk, Hadoop, RDMSs – Raanan Dagan (2015)

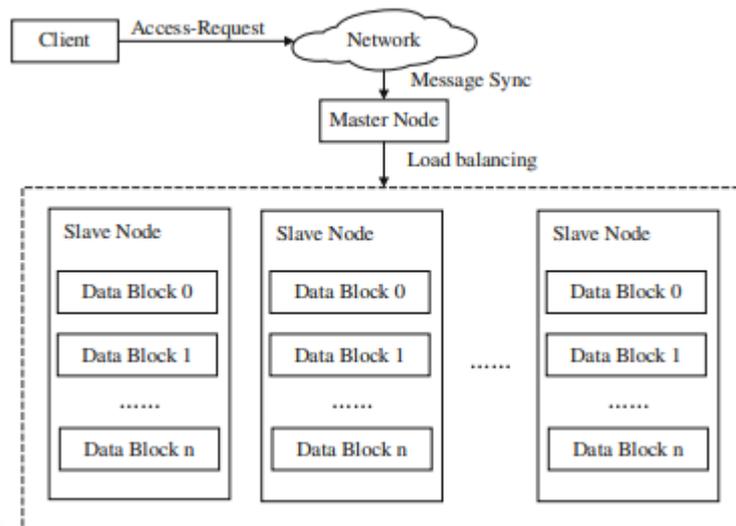


Figura 1.14²⁹

L'assenza di indici comparabili frena le performance di tali sistemi ma questi riescono ad integrarsi con le piattaforme di *Storage Analysis*, fra tutte Map Reduce, la quale riesce ad accedere agli indici dei dati, migliorando di gran lunga la performance della *query*. Tuttavia rimangono ancora diverse sfide da affrontare, ovvero la mancanza di un *query language* standardizzato per le SQL, una limitata ottimizzazione del modus operandi di Map Reduce e della sua integrazione con le RDBMSs e, per ultimo ma decisamente più importante, la gestione in tempo reale (senza quindi passare per strutture di raccolta intermedie) di dati eterogenei.

1.4.4 Analisi

L'analisi è la fase finale della catena del valore dei big data. Essa viene svolta col metodo analitico tradizionale se si tratta di dati strutturati, mentre è conseguita con l'ausilio di architetture analitiche e di software utilizzati allo scopo dell'estrazione e dell'analisi dei big data non strutturati. Lo scopo è quello di estrapolare valore dai dati, sotto forma di informazioni, al fine di implementare decisioni strategiche. Diversi dataset hanno diversi potenziali di valore. La cosiddetta *Data mining* si riferisce alla tecnica di estrazione di informazioni utili dai databases. Per questo motivo c'è bisogno di una fase di *pre-processing* oltre che di un metodo analitico per estrapolare valore dai dati. Questo processo di ricerca ed estrazione di valore dai dati, si compone delle fasi di pulizia, integrazione, selezione, trasformazione, estrazione, valutazione e rappresentazione. I requisiti e le sfide di tale fase analitica sono: trattare big data eterogenei tra loro, avere algoritmi con capacità di calcolo sufficiente a garantire efficienza e scalabilità ed estrapolare informazioni diverse dallo stesso dataset e/o stesse informazioni da dataset simili tra loro. Di seguito verranno espone le diverse tipologie di analisi, oltre ad una breve descrizione del funzionamento di Map Reduce.

²⁹ Real World Big Data Architecture Splunk, Hadoop, RDBMSs – Raanan Dagan (2015)

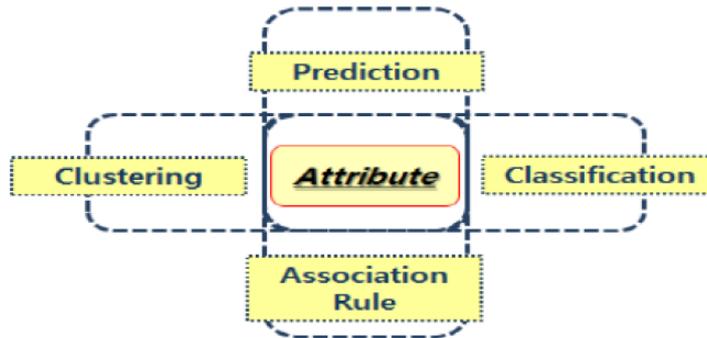


Figura 1.15³⁰

L'analisi tradizionale Consiste nell'utilizzo di metodi statistici per analizzare grossi volumi di dati. L'obiettivo è quello di estrarre dal calderone dei database dati raffinati e utili per i fini strategici che devono essere implementati; ovvero massimizzare il valore dei dati. Questo tipo di analisi viene effettuata tipicamente per i dati strutturati e, a seconda della tipologia dei dati e dei fini a cui l'analisi è volta, ci sono varie tipologie di analisi statistiche da implementare:

- Cluster Analysis: è un metodo statistico per il raggruppamento di dati e per la riclassificazione degli stessi a seconda di determinate caratteristiche che accomuna i dati omogenei. Viene utilizzata quindi per differenziare i dati eterogenei in cluster di dati omogenei tra loro a seconda di determinate caratteristiche che li accomuna. Quindi si avrà elevata omogeneità all'interno dei cluster ed elevata eterogeneità tra cluster diversi.
- Factor Analysis: consiste nel sintetizzare le relazioni tra i dati e diverse variabili attraverso pochi fattori che riescono a rivelare la maggior parte delle informazioni sui big data.
- Correlation Analysis: Serve a verificare se c'è correlazione tra i fenomeni osservati e di che tipo di correlazione si tratta. Dunque è possibile avere relazioni di dipendenza unilaterale, dipendenza reciproca (e quindi correlazione) oppure relazioni e correlazioni inesatte e comunque non giustificate da nessuna variabile causale.
- Regression Analysis: è uno strumento matematico per manifestare la correlazione tra una variabile e molte altre variabili. Essa si basa su una moltitudine di esperimenti su dati osservabili. La retta di regressione, tuttavia, non spiega anche le relazioni che ci sono per casualità, ovvero quelle ingiustificate.

A seconda della tempestività ricercata nell'analisi dei big data, quest'ultima si distingue in *real-time analysis* e *off-line analysis*

³⁰ Big Data Storage and Challenges - M.H.Padgavankar, Dr.S.R.Gupta (2014)

- Real-time analysis: questa tipologia di analisi è usata soprattutto dalle piattaforme di e-commerce e da quelle finanziarie. Poiché i dati mutano costantemente è necessaria un'analisi rapida di quest'ultimi e bisogna fornire report analitici in tempi più brevi possibili. Le architetture esistenti che permettono una real time *analysis* sono i tradizionali RDBMSs che processano parallelamente diversi cluster di dati e le cosiddette *memory-based computing platforms* (come HANA di SAP e Greenplum di EMC)
- Offline Analysis: è un'analisi propria delle architetture che non sono in grado di processare i dati in tempo reale. La maggior parte delle piattaforme di *offline analysis* utilizzano Hadoop come *storage tool* al fine di convertire i dati col minor costo possibile e migliorare quindi l'efficienza del processo di acquisizione dei dati. Esempi di *tools* di questo tipo sono la piattaforma open source Scribe di Facebook, quella di LinkedIn Kafka, Time tunnel di Taobao e Chukwa di Hadoop. Questi strumenti riescono ad acquisire e a trasmettere centinaia di MB al secondo e a soddisfare la domanda.

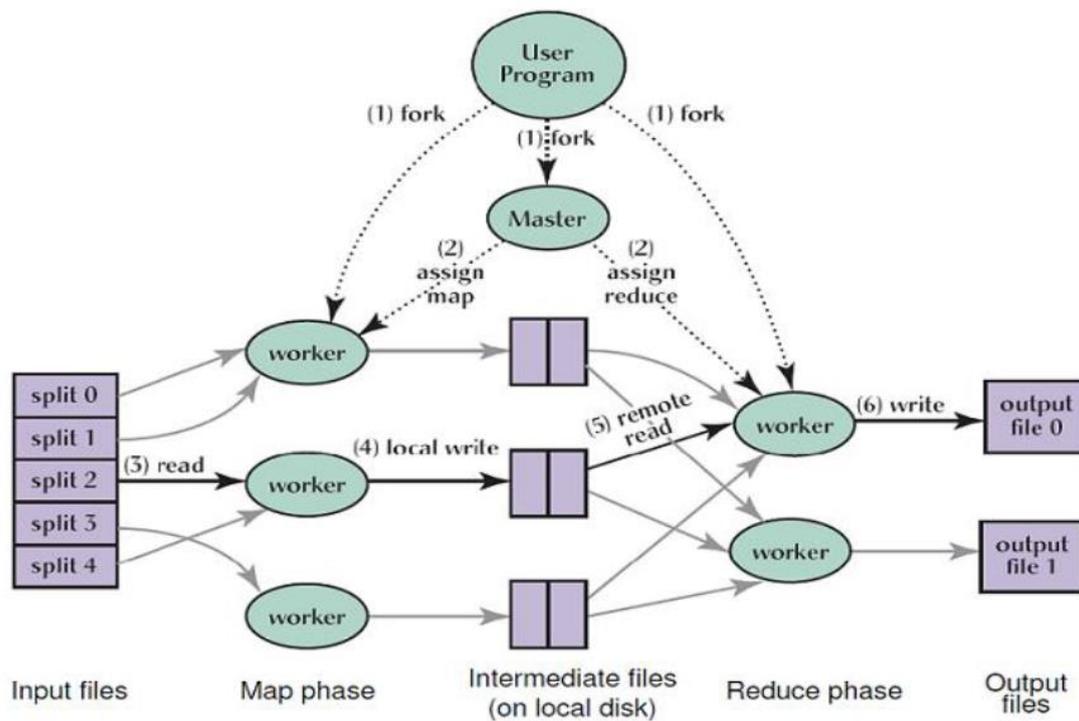
Ci sono poi diversi livelli di analisi dei dati:

- Memory-level analysis: è il caso in cui il totale ammontare dei dati è inferiore allo spazio di memoria disponibile per quello specifico cluster di dataset. Questo può arrivare a contenere addirittura decine di TB. La Memory-level *analysis* è estremamente idonea alla *real time analysis*. Un esempio di architettura di questo tipo è costituito da Mongo DB. Con lo sviluppo del SSD (Solid state driven) la velocità e la capacità dei *memory-level analysis* è migliorata in modo significativo.
- BI analysis: è il caso in cui il volume di dataset è superiore allo spazio di memoria disponibile ma può essere importato dalla cosiddetta BI *analysis environment*. Questi sistemi riescono dunque a superare lo spazio di archiviazione delle *memory-level analysis*.

MapReduce è stato introdotto da Google con l'obiettivo di processare e analizzare grandi dataset su hardware. Esso fornisce un linguaggio di programmazione che si integra perfettamente con il sistema popolare open source Hadoop³¹. MapReduce riesce a raggruppare i dati già processati in gruppi, ad assegnare un valore chiave a ciascun gruppo e a mappare tali gruppi. Esso inoltre riesce a processare i dati in un tempo relativamente breve e risolve le sfide di tolleranza all'errore e di disponibilità di dati dei sistemi tradizionali. Inoltre esso ha spianato la strada per lo sviluppo di applicazioni per la gestione dei dati sempre più affidabili ed efficienti.

³¹ Hadoop comes with its default distributed file system which is Hadoop distributed file system (HDFS). It stores file in blocks of 64 MB. It can store files of varying size from 100MB to GB, TB. Hadoop architecture contains the Name node, data nodes, secondary name node, Task tracker and job tracker. Name node maintained the Metadata information about the block stored in the Hadoop distributed file system. Files are stored in blocks in a distributed manner. The Secondary name node does the work of maintaining the validity of the Name Node and updating the Name Node Information time to time. Data node actually stores the data. Amrit Pal et al. (2014)

Figura1.16



32

1.5 Come le imprese estraggono valore dai dati

Molte imprese colgono appieno il potenziale di un dataset di riferimento, altre meno. Questo dipende dall'efficacia e dall'efficienza del processo di *data storage* e analytics. Dopo aver visto le tecnologie utilizzate per effettuare le diverse tipologie di analisi che si possono fare sui dati, passiamo a vedere come le imprese impiegano i dati raccolti ed analizzati per accrescere e migliorare le performance dei propri processi produttivi.³³

I dati possono essere impiegati in vario modo come ad esempio:

- Ottimizzazione dei costi e dei ricavi: le potenziali applicazioni in quest'ambito sono numerose. Sul lato dei costi, i dati possono essere impiegati per analisi predittive ai fini di opere di manutenzione, efficientamento gestionale, nell'attività di approvvigionamento, *supply chain* e pianificazione logistica. Sul lato dei ricavi, le intuizioni derivanti dai dati possono essere utilizzate per entrare in nuovi mercati, per attività di segmentazione dei consumatori, per migliorare le caratteristiche dei prodotti offerti e per rendere più efficaci i canali di distribuzione. I dati derivati dalle macchine e dai processi, in particolare dai sensori IoT e dal comportamento dei clienti sono i più utili per le operazioni di ottimizzazione. Viceversa, i dati generati internamente, possono essere utili attività di efficientamento.

³² Storage Architectures for Big Data in the Cloud – Sam Finemberg, HP Storage CT Office (Maggio 2013)

³³ Big Data Storage – Martin Strohbach, Jorg Daubert, Herman Ravkin, Mario Lischka

- Marketing e pubblicità: queste attività in genere si basano su transazioni effettuate dall'utente e sui dati comportamentali aggregati derivanti da più fonti, come profili social, informazioni demografiche, cronologia di navigazione on-line e acquisti precedenti. In quest'ottica possono essere analizzati grandi volumi di dati, fondamentale è l'attività di pulizia e scrematura dei dati ai fini della creazione dell'intuizione di valore. In particolare, imprese pubblicitarie altamente tecnologiche e social media con grandi quantità di dati comportamentali di consumo sono solo alcuni dei player cosiddetto fastestgrowing in questo settore.
- Market intelligence: molti fornitori tradizionali di servizi di informazione rientrano in questa categoria. Poche aziende sono in grado di generare macro dati autonomamente, la maggior parte dei fornitori in questo settore aggregano dati da fonti esterne. Poiché i dati conferiscono un chiaro valore per i clienti e non è facile replicarlo, può essere venduto direttamente e, in questo caso, il suo valore è più facilmente determinabile.
- Market-making: le imprese market-making, dalle app ride-sharing ai siti di incontri, giocano un ruolo cruciale nel far incontrare le esigenze di acquirenti e venditori. Queste imprese spesso creano piattaforme per raccogliere le informazioni necessarie per consentire una corrispondenza quanto più efficiente ed efficace. In alcuni casi, i dati di segnalazione pura rappresentano tutto ciò che conta. Ma in altri casi, le preferenze, i dati reputazionali (a garanzia dell'autenticità e della qualità dei partecipanti), i dati sulle transazioni e quelli comportamentali sono determinanti. In questo contesto, economie di scala ed effetti di rete svolgono un ruolo fondamentale.
- Training data for artificial intelligence: machine learning e deep learning necessitano di enormi quantità di training data. Alcuni vengono generati attraverso simulazioni ripetute, in parte sono generati nella sfera pubblica (come ad esempio i dati catastali e quelli climatici), in parte sono aggregati da una varietà di fonti (quali immagini e video o dati comportamentali). Le imprese che con la loro piattaforma producono enormi quantità di dati rilevanti possono godere di un vantaggio importante, in quanto le loro offerte avranno più tempo per imparare e generare ulteriori dati, alimentando il circolo virtuoso. Però, proprio perché c'è una grande varietà di possibilità di utilizzo per i diversi utenti, valorizzare i dati in questo contesto può risultare particolarmente impegnativo.

Questi ecosistemi possono poi sovrapporsi. In alcuni casi, lo stesso set di dati può avere molteplici applicazioni, ciascuna con un diverso valore.

CAPITOLO SECONDO

APPLICAZIONI ED IMPLICAZIONI DEI BIG DATA NEI MERCATI MULTILATERALI

Introduzione

Dopo aver visto le modalità di raccolta ed analisi dei big data, si passerà ora ad esaminare il valore aggiunto che questi apportano al processo produttivo dell'impresa. I dati infatti, una volta raccolti, possono sia essere venduti a soggetti terzi, sia impiegati nei propri processi produttivi, al fine di migliorare l'efficienza interna, intesa come ROI dei *sunk investments* in R&D oppure di quelli per l'acquisto delle infrastrutture di raccolta ed analisi dei dati. Le modalità con cui le imprese ottengono miglioramenti, in termini di ROI, dipendono dalla natura delle imprese stesse. Si può dire, con nessun'ombra di dubbio, che la raccolta e l'analisi dei big data ha portato alla nascita di vere e proprie imprese che utilizzano i dati come *core asset* per i propri processi produttivi, entriamo nella *data driven industry*. Tutti i settori hanno visto crescere in maniera esponenziale la *data storage e analysis*³⁴ e ci sono settori, come quelli creati da internet, ma anche quello bancario e finanziario, il *manufacturing*, la pubblica amministrazione, il trasporto che sono stati rivoluzionati dalla *data driven economy*.³⁵ I processi produttivi delle aziende operanti in tali settori hanno iniziato ad essere guidati dai dati, che hanno iniziato ad orientare le scelte strategiche delle imprese in modo del tutto artificiale grazie agli algoritmi di *machine learning*³⁶. Essi hanno permesso alle imprese di realizzare output in maniera più efficiente e di raggiungere obiettivi soprattutto di natura tecnologica in maniera efficace.

Lo sviluppo emergente dell'innovazione *data-driven* sta portando alla creazione di beni, servizi, strategie di marketing e decisioni aziendali realizzate e pianificate sulla base dei dati a disposizione delle imprese in tutti i settori industriali. Infatti, secondo anche quanto sostenuto nel 2015 dall'OCSE, la richiesta di dati deriva principalmente dalle attività produttive a valle, per le quali il dato rappresenta una risorsa non banale. La stessa OCSE stima una crescita della produttività delle imprese, correlata alla data-driven innovation del 5-10% e una riduzione dei costi amministrativi, per gli enti pubblici, del 15%-20% grazie ad una maggiore efficienza, un maggiore gettito fiscale (proveniente, ad esempio, dall'erogazione di servizi personalizzati) ed un minor rischio di frodi o errori. Secondo il Parlamento Europeo, inoltre, la data economy conta per l'1,9% del PIL europeo (1,6% di quello italiano), incidenza che si stima possa crescere al 3,3% nel 2020 (2,7% per l'Italia).³⁷

³⁴ Cfr: Capitolo 1.

³⁵ Cfr: L'ECONOMIA DEI DATI, TENDENZE DI MERCATO E PROSPETTIVE DI POLICY (It Media Consulting), Gennaio 2018.

³⁶ Si veda *focus: ML*.

³⁷ Cfr: Report OECD, DATA DRIVEN INNOVATION, BIG DATA FOR GROWTH AND WELL-BEING, 6 Ottobre 2015.

In questo capitolo si discuterà delle tipologie di applicazioni della *data storage e analysis* nelle *multi sided platforms*, andando a vedere dapprima in che modo questi impattano sugli aspetti microeconomici di tali piattaforme digitali (2.1), successivamente si analizzerà il loro modo di funzionare con il paradigma del *machine learning* (focus: ML), e le modalità di fare *revenues* che usano le *multisided platforms* (2.2). In particolare algoritmi di raccomandazione e pubblicità on-line (2.3 e 2.4). Infatti studieremo due meccanismi con cui tali imprese digitali monetizzano i propri investimenti nelle strutture di raccolta ed analisi dei dati, il *Recommendation System* e l'*Online Advertising*.

Internet è una vetrina di dati ed ogni dato proviene da qualsiasi ricerca che noi facciamo sul web. Ogni volta che utilizziamo Internet c'è qualcuno che sta collezionando le nostre informazioni. Spontaneamente, tracciamo le nostre preferenze, le nostre attitudini ed abitudini, le nostre specificità. Un mercato di dati esiste eccome. La catena del valore si completa con la vendita delle informazioni che si ricavano dai dati a terzi o a sé stessi in termine l'impiego nei processi produttivi.

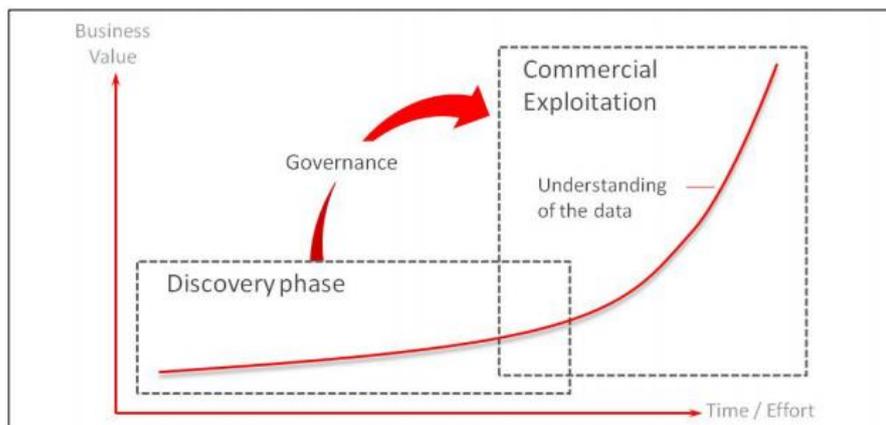


Figura 2.1 Fonte: Oracle, *An Enterprise Architect's Guide to Big Data, Reference Architecture Overview*

Come si nota dal grafico, le imprese sono passate da una fase di *Discovery* in cui i dati hanno iniziato ad entrare in maniera decisa nelle decisioni strategiche, queste hanno così iniziato ad investire nella *data storage e analysis*, apportando così un potente *core asset* nei propri processi produttivi e riuscendo a farsi guidare nelle proprie decisioni dai dati, in modo sempre più automatizzato.

2.1 Big data nello scenario competitivo delle *multi sided platforms*

Il nuovo paradigma tecnologico del ML che si è manifestato negli ultimi anni, ha portato alla nascita di molteplici settori in cui le imprese competono e di molteplici mercati estremamente differenziati tra loro. La competizione ora si sposta forte su chi detiene quei *core asset* descritti nei precedenti paragrafi, con l'impiego di *sunk investments* in R&D, al fine di trovare la cosiddetta *breakthrough innovation*, ovvero un'innovazione di

prodotto o di processo *disruptive*³⁸, che può completamente cambiare la struttura di un'industria. La competizione così descritta porta alla cosiddetta situazione del *winner takes all*³⁹ in cui chi riesce ad ottenere la tecnologia vincente, cambia lo scenario del mercato e si costituisce come impresa dominante, per un certo intervallo di tempo più o meno ridotto a causa della dinamicità dell'industria *high tech*. I nuovi operatori del mercato, riconfigurati dal nuovo paradigma, sono le *multi sided platform*.

La peculiarità di molti di tali mercati, è quella di offrire un servizio agli utenti molto spesso gratuito. Le piattaforme digitali, infatti, raccolgono i dati degli utenti, impiegano tali dati con lo scopo di migliorare il servizio offerto, personalizzandolo a seconda dei gusti dei consumatori finali e riuscendo a monetizzarlo efficacemente grazie ai dati raccolti, i quali aumentano l'attrattività della piattaforma verso i clienti e permettono quindi di porre ai consumatori finali un prezzo irrisorio oppure gratuito. La collezione e l'analisi di dati permette quindi di migliorare i servizi digitali in diversi modi. I principali operatori in tale ambito sono: i motori di ricerca (Google, Bing, Yahoo!), le piattaforme di e-commerce (Amazon, Ebay), Social Network (Facebook, Instagram, Tinder) e tutti i *player* che utilizzano il paradigma del *machine learning* e dell'intelligenza artificiale. I motori di ricerca, ad esempio, raccogliendo ed analizzando i dati provenienti dalle ricerche o *queries* effettuate dagli utenti (ovvero i cosiddetti "*click-and-query data*"), sono in grado di costruire indici di rilevanza e rifletterli nelle pagine del motore di ricerca. Infatti, generalmente, i risultati di una ricerca fatta in precedenza su una simile *query*, sono un segnale rilevante per le ricerche sulla stessa *query* che verranno fatte in futuro. Ovvero, ogni qual volta un utente clicca su un sito, il motore di ricerca lo traccia e conseguentemente apporta tali *click-and-query data* al proprio indice, al fine di aumentare l'efficacia del motore di ricerca che riuscirà a soddisfare, in modo sempre più efficiente, le richieste degli utenti on-line. Ecco un esempio di come i *click-and-query data* costituiscono un input al fine di migliorare i servizi digitali, in questo caso, di ricerca.

Le caratteristiche dei settori digitali "trainati" dalla *data-driven economy* sono molteplici; innanzitutto essi sono costituiti da imprese *multi sided*, ovvero imprese che fronteggiano più Domande ed Offerte. Ad esempio, Facebook offre i propri contenuti gratuitamente agli utenti finali e slot pubblicitari ad inseritori pubblicitari e funge anche da piattaforma per giochi ed app e da motore di ricerca.

Esse sono intermediari di beni e servizi che circolano tramite la loro piattaforma che fornisce tali contenuti secondo i risultati forniti dai propri algoritmi intelligenti applicati e dal *data management*.

³⁸ Disruptive innovations: innovazioni "radicali", che creano nuovi mercati favorendo la crescita di nuovi bisogni della clientela. Tale risultato può essere raggiunto introducendo sul mercato nuove tecnologie, nuovi modelli di business, ovvero utilizzando nuove tecnologie in maniera innovativa. Le "disruptive innovation" garantiscono la competitività nel lungo periodo perché introducono un insieme di funzionalità completamente nuove che ridefiniscono le regole del gioco di un settore.

³⁹ Some online economy markets seem to tend to the creation of significant market power positions (winner takes all or few winners take all), due to the combined effect of scale economies and network effects. Gianluca Faella (2018)

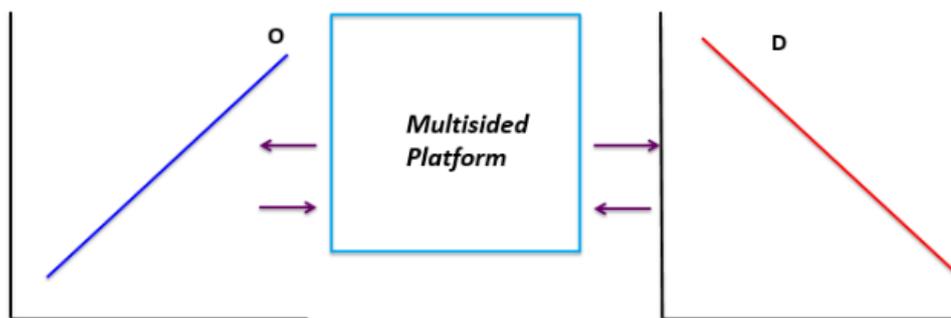


Figura 2.2

Inoltre i mercati *High tech* sono popolati da imprese che hanno nel loro attivo elevati costi fissi (*sunk cost*⁴⁰) di *data storage* e *analysis* da ammortizzare (R&D). Queste aziende hanno costi marginali nel raccogliere ed analizzare i dataset quasi nulli ed il meccanismo di recupero di tali costi fissi è costituito dal porre un prezzo uguale a zero per alcuni clienti, i consumatori finali ed un prezzo positivo se considerata una media fatta da tutti i prezzi praticati a ciascun cliente per ogni *side* della piattaforma. Possono esserci più domande e più offerte a seconda del tipo di piattaforma utilizzato e beni/servizi offerti più differenziati e customizzati con diverse metodologie di *revenues* (*fees, pay for impression, pay per click, pay for visit ecc..*) applicate di conseguenza. Anche la *price discrimination* può essere un'ottima strategia di recupero dei *sunk investments*.

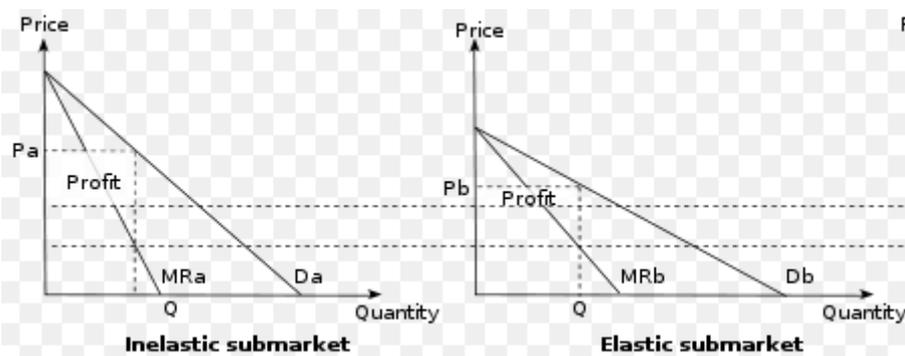


Figura 2.3

Nelle *multi sided platforms* vi sono infatti consumatori più e meno sensibili a variazioni di prezzo e questo è causato dalle diverse aspettative che ripongono sulla piattaforma. Così a due ipotetici consumatori finali vengono offerti due prezzi diversi in virtù della loro diversa curva di domanda caratterizzata ciascuna da una differente elasticità (pendenza). Questa pratica potrebbe essere notevolmente rafforzata dall'*analytics* e dagli algoritmi di *machine learning* cui tali piattaforme dispongono. Esse potrebbero infatti stimare la *willingness to pay* dei consumatori, raccogliendo i dati tracciati da quest'ultimi nei loro acquisti online o customizzando per regione geografica (Paragrafo 3.3).

⁴⁰ In economia, si definiscono *sunk costs* quei costi che sono già stati messi in atto dalle imprese a titolo di costi irrecuperabili. Sono quindi costi che non generano direttamente ricavi e rappresentano immobilizzazioni che non possono essere rivendute e che partecipano solo in maniera indiretta al processo produttivo dell'impresa.

Il percorso di recupero dei costi fissi viene accelerato dal cosiddetto *Network Effect*⁴¹, ovvero l'aumento del numero di *users* e di dati che essi apportano aumentano il valore del dataset (e con esso la complessità della *analytics*) e del prodotto (bene o servizio) finale. Gli utenti crescono linearmente, il valore della piattaforma lo fa esponenzialmente⁴². Il *Network Effect* veniva anche sfruttato dalle vecchie tecnologie come la rete telefonica oppure dai fax (se tutti hanno un fax me lo compro pure io!); ma si pensi anche alla rete di comunicazione via e-mail.

Nelle piattaforme digitali con più *side* si sviluppa il cosiddetto *Indirect Network Effect*.⁴³ Esso è un fenomeno che vede al centro la piattaforma, come intermediario che permette che, con l'aumentare del numero di utenti e di dati di un lato della piattaforma, cresca il valore della piattaforma per mezzo della crescita della percezione che hanno i clienti dell'altro lato della piattaforma, in termini di maggiori aspettative sulle performance e di *willingness to pay*. Ad esempio se aumentano gli *users* di Facebook, aumenterà la domanda di slot pubblicitari degli inseritori pubblicitari e così la loro disponibilità a pagare per un servizio maggiormente personalizzabile; di conseguenza Facebook sarà costretto (ed è ben contento di farlo) ad assecondare tale domanda fornendo più slot pubblicitari ed investendo su un *analytics* che permetta di personalizzare sempre più i contenuti digitali ed ad accrescerne quindi il valore del servizio offerto agli inseritori pubblicitari. Il *Direct e Indirect Network Effect* possono costituire una barriera all'ingresso per molte imprese che vogliono entrare nel settore, andando così a creare posizioni dominanti e situazioni di monopolio. Lo studio di tale variabile ci insegna che c'è una *critical mass* che le imprese devono fronteggiare per fare partire l'effetto rete con un aumento continuo della domanda e più in generale dell'utilizzo (espresso in tempo passato sulla piattaforma) che si fa della piattaforma.

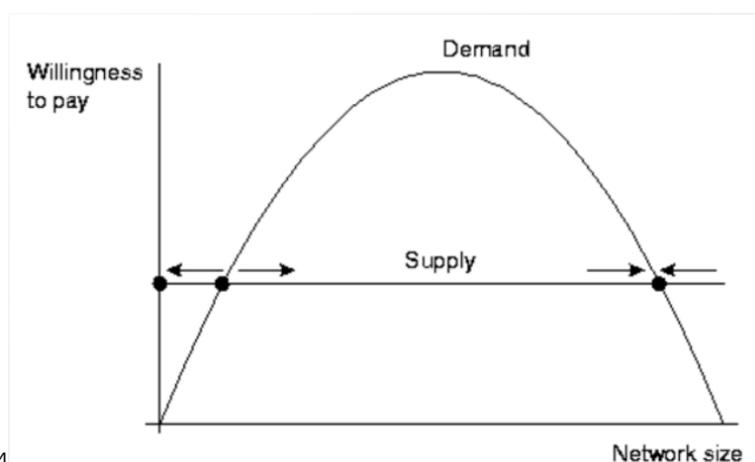


Figura 2.4

⁴¹ The value of a network increases along with the number of users. Gianluca Faella (2018)

⁴² According to Metcalfe's law, if the value of each user in a network is proportional to the number of other users, then the total value of the network is proportional to the square of the total number of users. Metcalfe's law (1993)

⁴³ The value of a platform increases along with the number of applications/services on the platform. Examples are: operating systems, video recorders and game consoles. Gianluca Faella (2018)

Come si evince dalla figura, sono possibili due equilibri uno *small market equilibrium* da cui inizia ad esserci forte aspettativa sul successo del prodotto e la *willingness to pay* dei consumatori aumenta e con esso il valore della piattaforma. Questo permette alla *network size* di crescere in maniera esponenziale fino a raggiungere un picco dove il prodotto è talmente penetrato nei mercati (ed i costi per le architetture di *data storage* e *analysis* ammortizzate) che il prezzo inizierà a scendere per permettere alla *size* di continuare a penetrare su scale internazionali fino ad arrivare al secondo punto di equilibrio, il *large market equilibrium*. Molte imprese utilizzano un *penetration price* molto basso ed in alcuni casi quasi nullo per arrivare a raggiungere lo *small market equilibrium* e poi hanno la possibilità di alzare il prezzo fino a raggiungere un picco in cui un aumento infinitesimalmente incrementale di prezzo fa crollare tutta la *size*. Questo è ovviamente un modello ideale ma viene confermato dalla realtà, infatti imprese come Whatsapp⁴⁴ hanno utilizzato un *penetration price* pari a zero all'inizio del loro ciclo di vita, dopodiché per qualche anno, hanno imposto un basso prezzo. Whatsapp in Italia, dopo il primo anno in cui fu gratuito, impose una *fee* di 0,79 centesimi su Google play e IOS.

Il *direct Network Effect* porta con sé più dati, che inducono le imprese a incrementare la capacità di calcolo dei propri algoritmi, che di conseguenza offrono contenuti personalizzati verso il cliente consumatore finale o *user*, e servizi proliferati o customizzati in base ai cluster di consumatori che si intende raggiungere, verso gli altri clienti che utilizzano la piattaforma. L'algoritmo intelligente impiegato nella fase di analisi, infatti, ha come obiettivo quello di raggiungere gli *users* con un bene/servizio personalizzato, ad esempio la pubblicità o un contenuto correlato ai gusti del consumatore, espressi dalle ricerche effettuate nel passato.

Un'altra caratteristica delle *multi sided platforms* è il cosiddetto *Learning Effect*⁴⁵. Imparare ad utilizzare un prodotto (una piattaforma) ed abituarsi alle caratteristiche di tale prodotto può indurre il consumatore a non voler abbandonare quel prodotto ed esserne così intrappolato (effetto *Lock in*). Il comportamento di un consumatore infatti dipende dalle scelte che egli fa nel passato ed in molti casi è difficile passare ad altri prodotti anche se più efficienti.

Questo è il caso di Whatsapp e Telegram in cui, sebbene Telegram offra più funzionalità di Whatsapp ed anche una maggior efficienza nella trasmissione dei messaggi (data dalla maggiore istantaneità nella ricezione di testi e altri contenuti multimediali), i consumatori hanno scelto Whatsapp probabilmente per un suo *core business* focalizzato solo sui messaggi istantanei (le chiamate, videochiamate e le *stories* sono funzionalità aggiunte

⁴⁴ WhatsApp Messenger is a proprietary, cross-platform, encrypted instant messaging client for smartphones. It uses the Internet to send text messages, documents, PDF files, GIF images, video, user location and audio messages to other users using standard cellular mobile numbers.

⁴⁵ Consumers are path-dependent. Behavioral economics has already highlighted path dependency in all markets. In high-tech markets, non professional users are reluctant to switch to another product. Investment to familiarize with a product is a sunk cost. Examples are: QWERTY keyboard, PC operating systems, etc.

solo di recente), mentre Telegram offriva più funzionalità, come i gruppi ed i *bot* che mettevano in contatto diversi *users*.⁴⁶ Questo è un esempio di *Path Dependence* in cui anche un prodotto (probabilmente) inferiore è riuscito a vincere la concorrenza e a prendersi tutto o quasi il mercato (*Winner takes all*), aumentando così il vantaggio competitivo per la *industry standard owner*. Si ha un tipo di competizione dinamica che porta al susseguirsi di situazioni di quasi monopolio per un lasso di tempo più o meno lungo a seconda dello sviluppo tecnologico in quel settore di riferimento, si ha *competition for the market*. Il valore del prodotto sarà percepito come standard dal consumatore ed aumenta così il suo valore attribuitogli (e la *willingness to pay*).

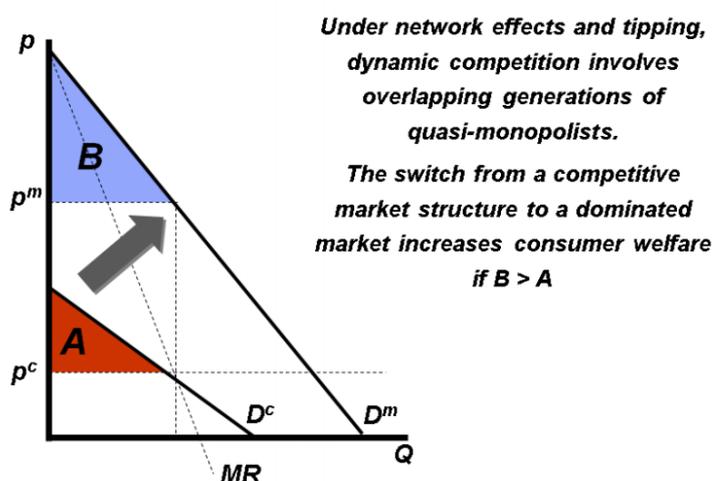


Figura 2.5

Potrebbe così aumentare il benessere del consumatore (*consumer welfare*)⁴⁷ anche se c'è un impatto negativo sulla competizione ed un aumento dei prezzi. Infatti, come si evince dalla figura, se $B > A$ la maggiore disponibilità a pagare o il maggior valore attribuito al prodotto dal consumatore ha un impatto sul benessere del consumatore positivo maggiore dell'impatto negativo dell'aumento del prezzo causato dalla situazione di (quasi) monopolio. I dati e gli algoritmi che li governano portano all'offerta di soluzioni più personalizzate ma questo non è detto che intrappoli gli utenti all'uso di quel prodotto. Tuttavia, se l'analisi si riconduce al tempo speso dagli *users* nell'utilizzo di una piattaforma, potrebbe darsi che esso sia inficiato dalla proliferazione degli utenti tramite ad esempio algoritmi di raccomandazione o pubblicità personalizzata che, potrebbe essere preferita dal consumatore finale rispetto ad una pubblicità generica.

Tuttavia l'effetto *lock in* viene frenato dal *multihoming*, ovvero dall'utilizzo di più piattaforme che fanno o non fanno parte dello stesso mercato rilevante. In fondo, in tali mercati la competizione si riduce ad essere semplicemente un click da parte dello *user*. *“Competition is one click away”*. Questo fenomeno è spiegato dalla natura di tali mercati che nascono *on-line*, essi infatti sono estremamente differenziati ed è persino difficile distinguere un mercato rilevante da un altro più o meno esteso.

⁴⁶ A COMPARATIVE STUDY BETWEEN TELEGRAM AND WHATSAPP IN RESPECT OF LIBRARY SERVICES, International Journal of Library & Information Science (IJLIS), Aprile 2018.

⁴⁷ Consumer welfare refers to the individual benefits derived from the consumption of goods and services.

Si pensi alle piattaforme di Social Network e di quanto sia frammentato tale mercato. Si possono trovare social a scopo di intrattenimento, come Facebook, Instagram, Youtube, Google+, Twitter, Tumblr, altri a scopo professionale, come LinkedIn e Xing, oppure ancora App di comunicazione, come WhatsApp, Viber, Telegram, oppure addirittura le attività sociali e culturali o a scopo ricreativo, come associazioni, circoli e club. Ora quali di questi sono da includere nello stesso mercato rilevante. È chiaro che ci sono diversi *overlaps* tra gli scopi e gli usi che vengono fatti di tali piattaforme. Definire un mercato rilevante è utile soprattutto per capire se un operatore è dominante nel suo settore di riferimento e per calcolarne dunque la *market share*. Come vedremo nell'ultimo capitolo, sebbene la maggior parte di queste piattaforme a più lati offrano al mercato un servizio gratuito con l'intento di ottenere dagli users dati, hanno un prezzo medio che potrebbe essere quasi monopolistico e, in molti casi, al fine di ottenere una mole di dati maggiore o, più semplicemente, col fine di eliminare un forte competitor dal mercato che causa una riduzione del tempo speso nell'utilizzo della piattaforma da parte del consumatore e quindi un minore apporto di dati, effettuano operazioni di fusione o acquisizione aumentando notevolmente la *market share* e quindi la loro posizione di dominanza nel mercato. Questo ovviamente a discapito dei clienti consumatori che vedranno un *overall price* aumentato ed una differenziazione del prodotto costante od in diminuzione. L'impatto sulla *data storage* e *analytics* causato da tali operazioni di concentrazione, ovviamente è notevole; e notevole è pertanto l'impatto sui contenuti che il consumatore finale andrà a vedere, in termini di maggiore personalizzazione e correlazione.

Le piattaforme, oltre che le architetture di *data storage e analysis*, fanno parte dei cosiddetti *system goods*, ovvero sono software complessi dove regna la complementarietà e l'interdipendenza tra le varie parti che compongono l'architettura o piattaforma⁴⁸. Ci sono infatti software che dipendono dalle componenti che li compongono e tali componenti hanno bisogno del software per funzionare. Si possono distinguere così architetture chiuse, semi aperte e aperte o *open source*.

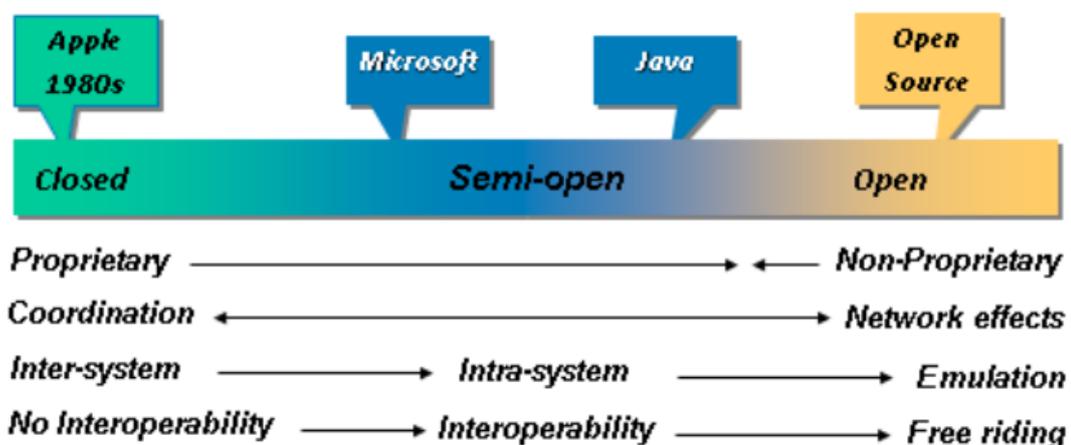


Figura 2.6

⁴⁸ Complex goods made of complementary and interdependent products. In high-technology industries, it is common to see products that are useless unless they are combined into a system with other products. Some examples are: hardware without software, Operating systems without applications, DVD players without content. Gianluca Faella (2018)

Come abbiamo visto il sistema di raccolta più utilizzato, Hadoop⁴⁹ è un'architettura di tipo *open source*, il che significa che essa funge da interfaccia a molti, quasi alla totalità dei sistemi di *data analytics* che esistono. Un esempio di architettura a circuito chiuso è IOS, il sistema operativo di Apple per device mobili (iPhone), compatibile solo con gli standard di Apple e con le sue interfacce di riferimento. Gli *iphones*, infatti, possiedono Safari come browser predefinito, Apple pay come mezzo di pagamento, alcune app predefinite ed App Store, la piattaforma che distribuisce app compatibili con IOS e di proprietà di Apple. Tale piattaforma è a sua volta un esempio di *semi-open architecture*⁵⁰ in cui vi è una concorrenza *intrasystem* ovvero, in questo caso, tra App che devono sviluppare le interfacce idonee per funzionare su IOS ed essere così messe in "vetrina" su Appstore. La disponibilità di app sulla propria piattaforma di distribuzione di app porta anche alla cosiddetta competizione *intersystem*, ovvero vi sarà competizione per sviluppare una piattaforma di distribuzione che sia in grado di interfacciarsi con il sistema operativo e che contenga un numero maggiore di app differenziate tra loro che fanno, o non fanno parte dello stesso mercato rilevante. Android e IOS infatti sono due sistemi che al loro interno sono completamente diversi, uno con versioni sia *open source* sia *semi open* (Android), l'altro con architettura chiusa. Entrambi competono per il sistema operativo, ma anche per le piattaforme di distribuzione di app, motori di ricerca (Safari e Chrome) sia in chiave di competizione tra sistemi (*intersystem*), ovvero come scelta del consumatore finale, sia in chiave di competizione all'interno di un sistema (*intrasystem*).⁵¹

2.2 Il Machine Learning nelle multi-sided platforms

Il proseguo del capitolo è incentrato sulle tecniche di *machine learning* e quindi sui diversi algoritmi intelligenti che ne derivano e sulle applicazioni che, tali sistemi dotati di grande capacità di elaborazione per la gestione di grossi *datasets* in maniera del tutto automatizzata, ne vengono fatte per i processi produttivi delle *multisided platforms*. Analizzeremo dapprima a grandi linee il paradigma del *machine learning* e di come esso abbia rivoluzionato moltissimi settori dell'economia tradizionale, oltre che ad aver creato nuovi mercati attraverso le piattaforme digitali; successivamente si passerà all'analisi degli usi specifici che tali piattaforme ne fanno nei propri processi produttivi, come il *recommendation system* nella condivisione di contenuti multimediali di vario tipo (news, telefilm, post, video, immagini, beni e servizi correlati, pubblicità ecc).

⁴⁹ Cfr: Capitolo 1, Paragrafo 1.4.3

⁵⁰ All architectures have advantages and disadvantages. There is no proof that open systems are more efficient than closed systems. Gianluca Faella (2018)

⁵¹ Depending on the industry conditions, competition may emerge "within systems" or "between systems". There is a trade-off between exploiting network effects (non-proprietary architectures) and quality assurance (closed architectures). System good vendors will try to exploit indirect network effects. Gianluca Faella (2018)

2.2.1 Focus: il Machine learning e lo statistical learning

Nell'ambito del modello previsionale, si collocano la tecnologia di *machine learning* (ML) e di *statistical learning* (SL). Il ML e lo SL utilizzano un approccio di tipo induttivo per formare una rappresentazione del mondo sulla base dei dati che vengono esaminati. Le finalità delle due applicazioni sono le medesime: descrivere trend e andamenti, capire il presente, prevedere il futuro; viceversa, il fulcro delle rispettive attività invece cambia: per il ML l'algoritmo e il suo output sono centrali, per lo SL lo sono i modelli statistici e la quantificazione dell'incertezza. Inoltre, il ML e lo SL sono in grado di ottimizzare e migliorare la rappresentazione della realtà all'aggiungersi di nuovi dati. In questo senso, l'algoritmo "impara" dai nuovi dati inseriti e migliora nel tempo; stessa cosa si può dire per i modelli statistici che diventano più raffinati ed efficaci.⁵²

Dunque, il requisito fondamentale per l'apprendimento automatico è rappresentato da una grande mole di dati, necessari per guidare l'algoritmo. Il ML è uno strumento che esamina grandi quantità di dati in cerca di modelli, quindi genera il codice che consente di riconoscere quegli stessi modelli in nuovi dati. Le applicazioni possono utilizzare questo codice così generato per effettuare migliori previsioni. In altre parole, l'apprendimento automatico può aiutare a creare applicazioni più intelligenti. Il funzionamento del ML si basa sostanzialmente su regole statistiche e trova applicazione in diversi ambiti, fungendo da soluzione ad una vasta gamma di problematiche, quali ad esempio la creazione di un software in grado di determinare, con elevata precisione: se una transazione con carta di credito è da considerarsi fraudolenta; la previsione circa le future entrate di un'azienda; il rischio sul possibile passaggio di un dato cliente ad altro operatore; l'esatto momento in cui un motore a reazione necessita di un intervento di manutenzione; consigli sui film per i clienti; effettuare qualsiasi altra analisi su una serie di dati storici disponibili.

Proprio in virtù del fatto che il ML aiuta a predire il futuro, spesso è incluso nella categoria più ampia di analisi predittiva. Tutto ciò che serve a tal fine sono i dati, software di apprendimento per imparare da questi stessi dati, e i tecnici che sanno come utilizzare i software impiegati.

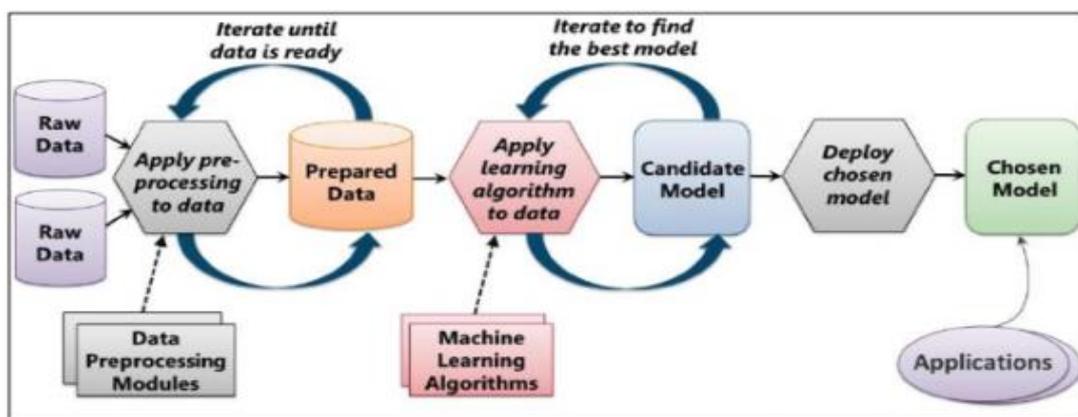


Figura 2.7 Fonte: *Introducing azure machine learning*

⁵² Cfr: "Machine Learning and Statistics: What's the Connection?" Chris Williams, Agosto 2006

Come si evince dalla Figura 1.22, il processo del ML inizia con i dati, di conseguenza maggiore è il numero di dati di cui si dispone, migliori saranno i risultati ottenibili. Proprio in ragione del sempre crescente aumento dei dati a disposizione delle imprese e in generale del mercato, le tecniche di ML stanno diventando sempre più utilizzate. Molti aspetti del processo di *machine learning* non sono affatto semplici. Ad esempio, come individuare quali dati grezzi usare; come elaborare i dati grezzi al fine di creare i *prepared data*⁵³; quali combinazioni tra *prepared data* e algoritmi di apprendimento automatico dovrebbe essere utilizzata per creare il modello migliore. Domande di questo genere vengono generalmente risolte da specialisti di settore, i *data scientist*, ovvero esperti con una vasta gamma di competenze nel campo degli algoritmi di apprendimento automatico complessi. Tali professionisti sono ad esempio in grado di individuare quali, tra gli algoritmi presenti, siano capaci di lavorare più efficientemente in situazioni diverse; o ancora si occupano anche della scrittura di codici essendo particolarmente competenti nel campo dello sviluppo di software. Proprio i *data scientist*, anche per le tecniche di ML così come precedentemente visto per quelle di *data analytics*, distinguono tra tecniche di ML supervisionato e non supervisionato. Le prime, come visto, richiedono necessariamente una conoscenza preventiva di un dato risultato e forniscono potenti strumenti per la previsione e la classificazione delle informazioni. Spesso, tuttavia, non conosciamo l'esito "finale" di un evento. Per esempio, in alcuni casi di frode, potremmo non sapere se una transazione è fraudolenta prima che sia trascorso un certo tempo, anche piuttosto lungo, dopo il verificarsi dell'evento. In questo caso, piuttosto che tentare di prevedere quali transazioni sono da considerarsi fraudolente, potremmo voler usare il ML per individuare quelle operazioni che risultano quanto meno insolite, in modo da tenerne conto per eventuali ulteriori indagini. Le tecniche di ML supervisionato più utilizzate sono:

- Generalized linear models (GLM): una forma avanzata di regressione lineare che supporta diverse distribuzioni di probabilità e funzioni di collegamento, consentendo all'analista di modellare in modo più efficace i dati;
- Decision Trees: un metodo di apprendimento supervisionato che memorizza un insieme di regole atte a dividere una popolazione in segmenti sempre più piccoli e che risultano essere omogenei rispetto alla variabile di destinazione;
- Random forests: un classificatore d'insieme composto da molti alberi di decisione e che dà in uscita la classe che corrisponde all'uscita delle classi degli alberi presi individualmente. Il processo di classificazione si compone in generale di due sottoprocessi in sequenza: la costruzione del modello a partire da dati già classificati (training) e la classificazione vera e propria, in cui il modello precedentemente creato viene utilizzato per associare a nuovi dati un'etichetta di classe,⁵⁴

⁵³ Data preparation is the act of preparing (or pre-processing) raw data or disparate data sources into refined information assets that can be used effectively for various business purposes, such as analysis. Big Data storage and challenges - M.H.Padgavankar, Dr.S.R.Gupta (2014)

⁵⁴ "Appunti di Machine Learning", Maurizio Di Matteo, 30 Maggio 2014

- Gradient boosting machine (GBM): un metodo che produce un modello di previsione attraverso la formazione di una sequenza di alberi di decisione, dove gli alberi successivi correggono gli errori di previsione di alberi precedenti;
- Deep learning: un approccio che si basa su diversi livelli di rappresentazione, corrispondenti a gerarchie di caratteristiche di fattori o concetti, dove i concetti di alto livello sono definiti sulla base di quelli di basso livello. Questo ramo di ML utilizza le reti neurali profonde con più livelli nascosti. Due dei più comuni tipi di reti neurali profonde sono quella convoluzionale e quella ricorsiva. Reti neurali convoluzionali sono spesso utilizzate per il riconoscimento di immagini, elaborando un insieme di elementi, ad esempio, per effettuare un collegamento tra un naso, un volto, e, infine, un oggetto/animale vero e proprio. Questa capacità di riconoscere le immagini è un'importante applicazione per lo sviluppo di veicoli autonomi, che hanno bisogno di riconoscere istantaneamente il loro ambiente circostante. Al contrario, le reti neurali ricorsive vengono utilizzate quando l'insieme delle sequenze e il contesto di riferimento sono importanti, come nel caso del riconoscimento vocale o nell'elaborazione del linguaggio naturale. Il *deep learning* è l'esempio più chiaro della confluenza tra grandi moli di dati, potenza di elaborazione, e algoritmi sempre più sofisticati. In realtà, le reti neurali sono state sviluppate decenni fa, ma di certo non disponevano delle enormi quantità di dati e della potenza di elaborazione necessaria per raggiungere le loro piene capacità. Ora che questi ostacoli sono stati superati, i data scientist stanno rapidamente sviluppando importantissimi progressi nelle tecniche di *deep learning*.⁵⁵

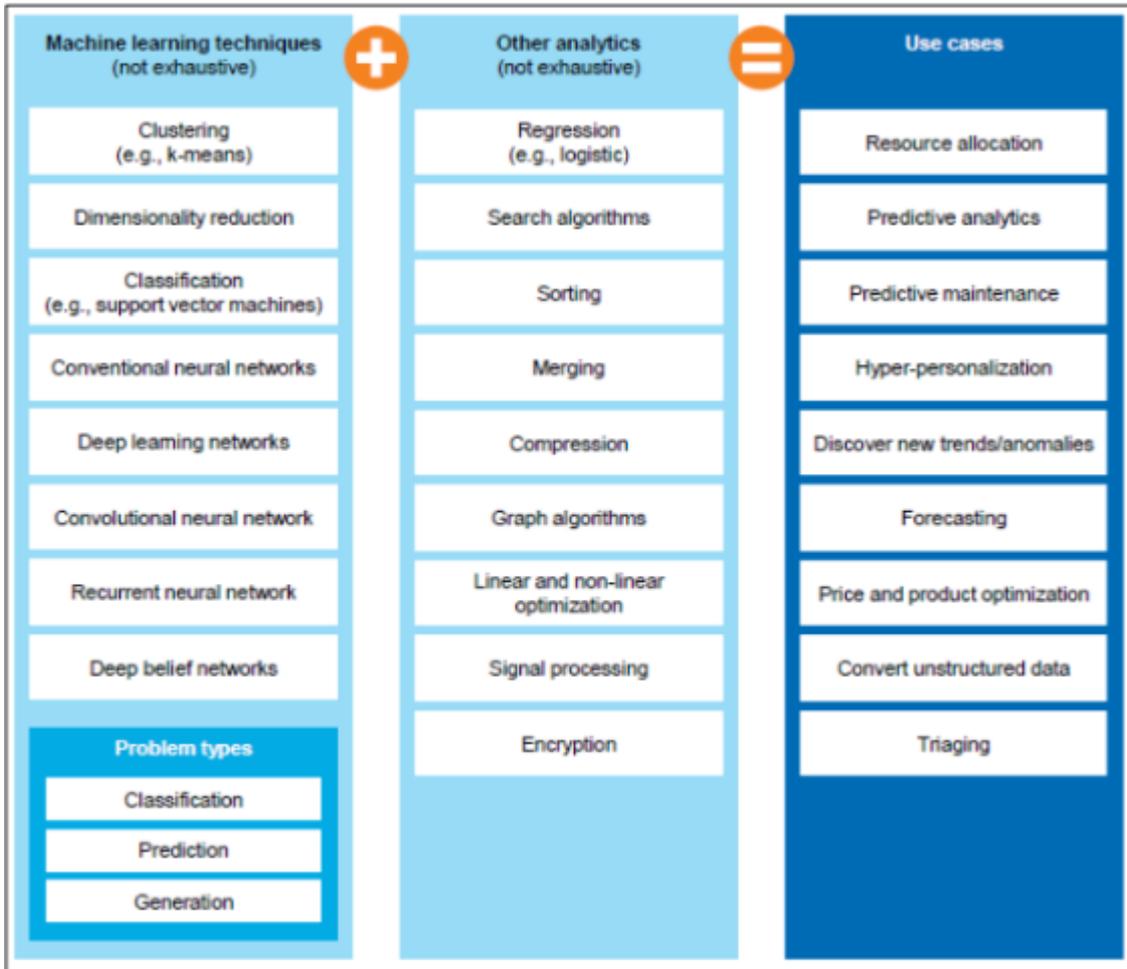
Le tecniche chiave per il ML non supervisionato sono invece:

- Clustering: una tecnica atta a raggruppare oggetti che risultano simili tra loro su molti parametri in segmenti, o cluster. La segmentazione della clientela è un tradizionale esempio di raggruppamento. Esistono diversi algoritmi di clustering differenti, il più diffuso è il cosiddetto "k-means"⁵⁶;
- Anomaly detection: il processo di identificazione degli eventi imprevedibili o dei possibili risultati. In campi come la sicurezza e la frode, non è possibile indagare in modo esaustivo ogni transazione, è necessario segnalare sistematicamente le operazioni più insolite. Si segnala che la tecnica di deep learning, nominata in precedenza, può anche essere utilizzato per il rilevamento di anomalie;
- Dimension reduction: il processo di riduzione del numero di variabili considerate. Considerando che le organizzazioni tendono a raccogliere sempre più dati, il numero di possibili informazioni disponibili per formulare previsioni si espande rapidamente. È sufficiente identificare quali dati garantiscono informazioni di valore per un determinato problema.

⁵⁵ "A brief introduction to Deep Learning", Yangyan Li (2017)

⁵⁶ k-means clustering aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

Figura 2.8



Fonte: McKinsey Global Institute analysis

Le capacità del ML sono le più adatte per la risoluzione di tre grandi ordini di problemi, come illustrato nello schema sottostante:

- classificazione;
- predizione/stima;
- generazione.

Problemi di classificazione riguardano l'osservare la realtà circostante, tra cui l'identificazione di oggetti in immagini e video, e il riconoscimento del testo e dell'audio. La classificazione riguarda anche le associazioni che si trovano nei dati o le segmentazioni in cluster, attività particolarmente utile per segmentare la clientela. Il ML può anche essere usato per predire la probabilità di eventuali risultati. Infine, può essere usato per produrre i contenuti, dall'inserire i dati mancanti a generare il fotogramma successivo all'interno di una sequenza video.

Focus: Le applicazioni del Machine learning e lo statistical learning nei vari settori dell'economia

Il ML e lo SL possono trovare applicazione in un'ampia varietà di ambiti per la risoluzione delle problematiche più disparate.

Nello schema seguente, è proposto uno studio di McKinsey in cui sono stati tracciati i primi 120 casi d'uso del ML nell'ambito di 12 diversi settori. Sull'asse y sono stati riportati il volume di dati disponibili, mentre sull'asse x è stato indicato il potenziale impatto del ML sulle attività interessate sulla base di indagini condotte da oltre 600 esperti del settore. La dimensione della bolla riflette la diversità delle fonti di dati disponibili.

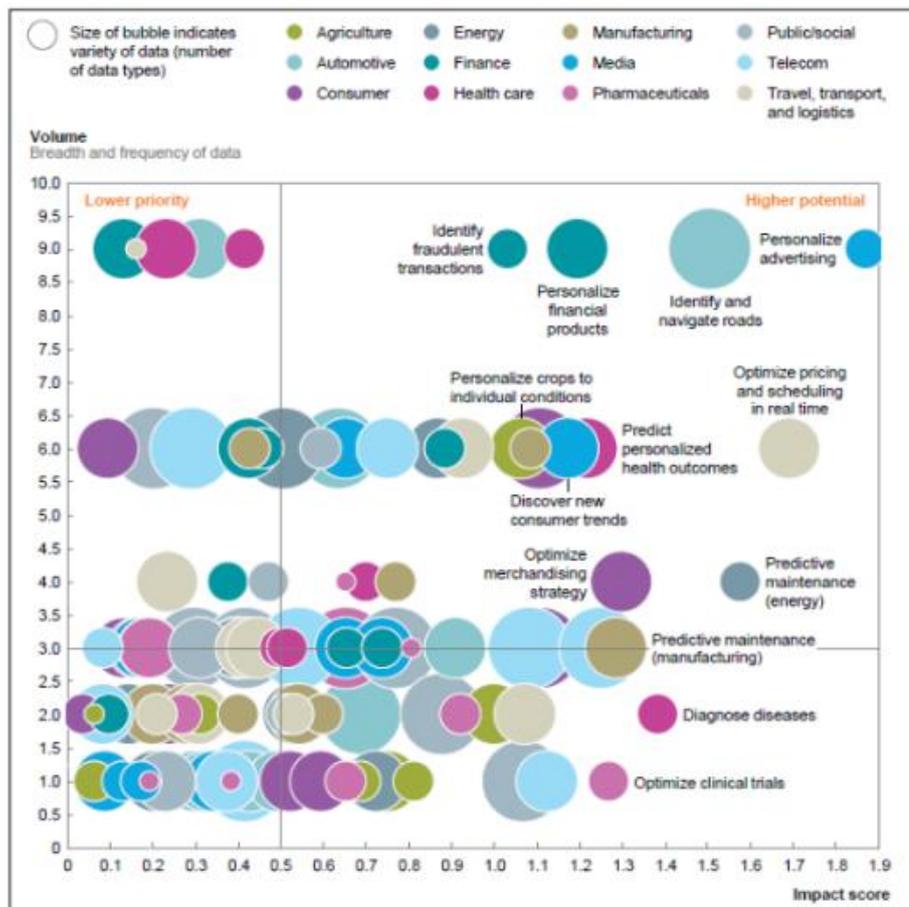


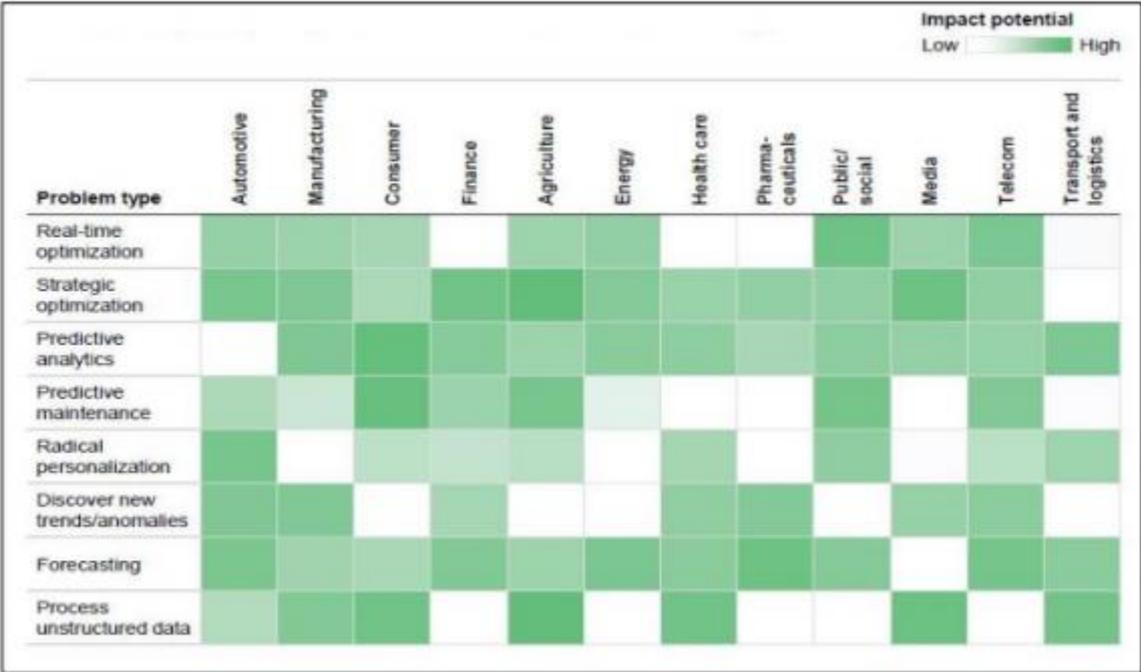
Figura 2.9 Fonte: McKinsey Global Institute analysis

Diversi saranno gli usi specifici del settore che uniscono la ricchezza dei dati con un'opportunità, più grandi saranno le bolle nel quadrante in alto a destra del grafico. Queste rappresentano le aree dove le organizzazioni dovrebbero dare priorità all'uso del ML e prepararsi ad una trasformazione senza precedenti. Alcuni dei casi d'uso in cui si registrano le più alte opportunità includono la pubblicità personalizzata; i veicoli autonomi; l'ottimizzazione dei prezzi, il *routing*, e la pianificazione sulla base di dati in tempo reale sulla mobilità e la logistica; predire risultati personalizzati sulla salute; e l'ottimizzazione delle strategie di merchandising nella vendita al dettaglio.

I casi d'uso nel quadrante in alto a destra si dividono in quattro categorie principali. In primo luogo è evidente che la personalizzazione dei prodotti e dei servizi per i clienti in settori quali prodotti confezionati, finanza e assicurazioni, sanità e media, rappresentano un'opportunità che la maggior parte delle aziende non ha ancora sfruttato appieno.

La seconda categoria è quella dell'analisi predittiva. Ci si riferisce ad esempi come il *triaging* di chiamate al servizio clienti; la segmentazione della clientela sulla base di modelli di rischio, di abbandono e di acquisto; l'individuazione delle frodi e di anomalie bancaria, ovvero di sicurezza informatica; la possibilità di diagnosticare malattie da scansioni, biopsie e altri dati. La terza categoria è rappresentata dall'ottimizzazione strategica, che comprende usi come il merchandising e le procedure di ottimizzazione nella vendita al dettaglio, la programmazione e l'assegnazione di lavoratori in prima linea, e l'ottimizzazione di team e altre risorse attraverso aree geografiche e relazionali. La quarta ed ultima categoria è l'ottimizzazione delle operazioni e della logistica in tempo reale, che include impianti e macchinari automatici atti a ridurre gli errori e migliorare l'efficienza, ottimizzando la gestione della catena di fornitura.

Figura 2.10



Fonte: McKinsey Global Institute

Sempre nell'ambito dello stesso studio, McKinsey ha analizzato e misurato l'impatto del ML all'interno dei diversi settori industriali, come mostrato nella mappa riportata in figura. Ai fini di una più approfondita analisi e di una maggiore comprensione degli usi applicativi del ML, si riportano di seguito una serie di esempi applicativi analizzati più nel dettaglio:

- Prevenzione delle frodi: un esempio di ciò viene fornito dalle piattaforme dei servizi di pagamento. Con più di 150 milioni di portafogli digitali attivi, 200 miliardi di dollari in pagamenti annuali, PayPal è leader nel settore dei pagamenti online. Considerando queste cifre, anche i bassi tassi di frode possono avere una certa consistenza.

Soprattutto agli inizi, l'azienda stava perdendo circa 10 milioni di dollari al mese per il numero di frodi effettuate. Per affrontare il problema, PayPal ha costituito un top team di ricercatori, che hanno utilizzato le tecniche di ML esistenti al fine di per costruire modelli in grado di identificare i pagamenti fraudolenti in tempo reale;

- Targeting digital display: gli algoritmi di ML permettono di fornire pubblicità altamente personalizzata.⁵⁷ Una società tecnologicamente molto avanzata, Dstillery, utilizza il ML per supportare aziende come Verizon e Williams-Sonoma ad indirizzare in tempo reale la display advertising sulle piattaforme. Utilizzando i dati raccolti sulla cronologia di navigazione di un individuo, pagine visitate, clic e acquisti, Dstillery elabora migliaia di previsioni al secondo, gestisce centinaia di campagne alla volta, sovra performando in modo significativo le operazioni di marketing svolte invece da un singolo individuo e registrando un impatto particolarmente efficiente per ciascun dollaro speso;
- Recommending content: esistono infatti algoritmi di ML che permettono alla piattaforma di fornire al consumatore contenuti correlati ai suoi interessi e a quelli di utenti considerati simili.⁵⁸ Per i clienti di Comcast X1, servizio di TV interattiva, la società offre consigli personalizzati in tempo reale sui contenuti, basandosi sulle scelte di visioni precedenti effettuate da ciascun utente. Comcast utilizza tecniche di ML per sviluppare un profilo per ciascun utente in relazione alle rispettive preferenze, raggruppando poi in cluster gli utenti aventi preferenze comuni. Per ciascun cluster, la società tiene traccia e visualizza i contenuti più popolari in tempo reale, in modo da consentire agli utenti di vedere quali contenuti sono attualmente dei trend. Il risultato netto equivale a migliori raccomandazioni, utilizzi più elevate e clienti maggiormente soddisfatti;
- Migliori prospettive di targeting: I marketers⁵⁹ utilizzano modelli di "propensione all'acquisto" come strumenti per determinare le migliori prospettive di vendita, di marketing e i migliori prodotti da offrire;
- Ottimizzazione dei media: NBC Universal archivia centinaia di terabyte di file multimediali per la distribuzione di TV via cavo internazionale. Una efficiente gestione di questa risorsa on-line è necessaria per garantire la distribuzione ai clienti internazionali. L'azienda utilizza il ML per prevedere la domanda futura di ciascun servizio televisivo basato su una combinazione di parametri. Sulla base di queste previsioni, l'azienda archivia quei media per i quali è prevista una bassa domanda. Le previsioni effettuate con l'utilizzo del ML sono certamente più efficaci di regole arbitrarie basate su singoli parametri, come l'età del file. Di conseguenza, NBC Universal riduce i costi complessivi di archiviazione pur mantenendo un alto livello di soddisfazione del cliente.

⁵⁷ Cfr: Paragrafo 2.4

⁵⁸ Cfr: Paragrafo 2.3

⁵⁹ People whose duties include the identification of the goods and services desired by a set of consumers, as well as the marketing of those goods and services on behalf of a company. See also advertising. Big Data Analysis and Storage – Khalid Adam, Mohammed Adam, JASni Mohamed Zain, Mazlina Abdul (2015)

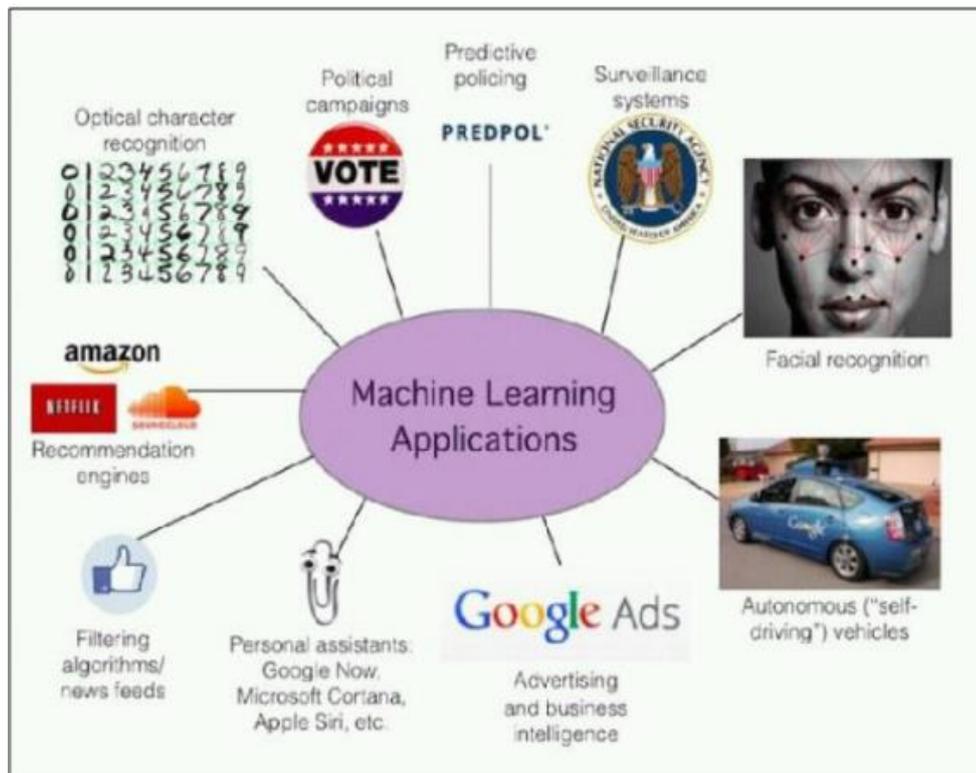


Figura 2.11 Fonte: Elaborazione ITMedia Consulting su dati vari

2.3 Recommender System

Un esempio significativo di applicazione delle tecnologie e degli algoritmi di analisi big data per l'estrazione di valore per il business è rappresentato dai *Recommender Systems*, spesso indicati anche con il termine *Recommendation Engines*. Si tratta di un sistema di filtraggio delle informazioni che mira a prevedere il livello di preferenza che un utente darebbe a un dato oggetto o elemento. Diventati sempre più popolari in tempi recenti ed in contesti di marketing personalizzato e contestuale, tali sistemi abbracciano una vasta area di settori e attività economiche: musica, news, libri, film, servizi finanziari, assicurazioni, ristoranti, persone, ecc.⁶⁰

Utilizzati dalle grandi piattaforme digitali per incrementare l'*engagement* e il tempo speso online sulla piattaforma dagli utenti, aumentare le vendite e minimizzare il tasso di abbandono e quindi la percentuale di *no users*. Questi sistemi costituiscono una valida alternativa agli algoritmi di ricerca dato che aiutano gli utenti a scoprire oggetti o contenuti che difficilmente potrebbero trovare da soli.

La logica elaborativa alla base della produzione delle *recommendation* si basa essenzialmente su due approcci di analisi e filtraggio dei dati: *collaborative filtering* e *content-based filtering*.

⁶⁰ "Perhaps the most important use of recommendation systems is at on-line retailers. We have noted how Amazon or similar on-line vendors strive to present each returning user with some suggestions of products that they might like to buy. These suggestions are not random, but are based on the purchasing decisions made by similar customers or on other techniques we shall discuss in this chapter."

I sistemi *content-based* sono più semplici da realizzare ma producono *recommendation* meno interessanti mentre i sistemi basati sull'approccio *collaborative filtering* consentono di ottenere risultati molto accurati al prezzo di un livello elevato di complessità sia in termini di algoritmi che di volume di dati richiesti. Inoltre, nei casi più avanzati questi approcci sono utilizzati in mondo congiunto dando luogo ai cosiddetti *Hybrid Recommender System*.

2.3.1 Collaborative filtering

Tale approccio è spesso riferito *model-based* in quanto basato sulla definizione di un modello descrittivo dell'utente sulla base del suo comportamento passato (ad esempio oggetti comprati o selezionati, *rating* dato a questi oggetti, o disponibilità a pagare per gli stessi ecc.) e del comportamento passato di tanti altri utenti. A tal fine è necessario raccogliere ed analizzare una grande quantità di dati sulle attività e le preferenze dell'utente. Le *recommendation*⁶¹ sono prodotte tenendo conto del modello e delle preferenze di numerosi altri utenti considerati "simili". L'assunzione di fondo è che gli utenti che hanno mostrato un comportamento e preferenze simili in passato continueranno a mostrarlo in futuro, preferendo quindi oggetti simili a quelli preferiti in passato.

La definizione del modello che sintetizzi il processo d'acquisto o di selezione di contenuti digitali dell'utente sulla base del suo comportamento si avvale della raccolta di dati raccolti in modo implicito ed esplicito. Esempi di raccolta dati in modo esplicito sono:

- chiedere all'utente di valutare un oggetto sulla base di una scala predefinita di valori;
- chiedere all'utente di effettuare una ricerca;
- chiedere all'utente di classificare un elenco di oggetti/elementi in ordine di preferenza;
- presentare all'utente due oggetti a chiedergli di scegliere il migliore tra i due;
- chiedere all'utente di creare una lista di oggetti preferiti.

La modalità di raccolta dati in maniera implicita appare più interessante e si basa sulle seguenti operazioni:

- osservare e memorizzare gli oggetti che l'utente visualizza sull'online store;
- memorizzare gli oggetti comprati online dall'utente;
- analizzare il numero di volte che l'utente visualizza gli oggetti;
- analizzare l'attività dell'utente sui social network per scoprirne le preferenze.⁶²

⁶¹ "Recommender systems apply knowledge discovery techniques to the problem of making personalized recommendations for information, products or services during a live interaction. These systems, especially the k-nearest neighbor collaborative filtering based ones, are achieving widespread success on the Web. The tremendous growth in the amount of available information and the number of visitors to Web sites in recent years poses some key challenges for recommender systems. These are: producing high quality recommendations, performing many recommendations per second for millions of users and items and achieving high coverage in the face of data sparsity." The Role of "Big Data" in Online Platform Competition – Andres V. Lerner (2014)

⁶² Item-based Collaborative Filtering Recommendation Algorithms, Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl (2017)

Uno degli esempi più noti di *collaborative filtering e recommender system* è sicuramente dato dagli algoritmi utilizzati da Amazon basati proprio sull'assunzione che l'utente che ha comprato un dato prodotto molto probabilmente comprerà un altro prodotto già scelto da utenti "simili". Anche le piattaforme Social, quali Facebook e LinkedIn, applicano algoritmi basati su tale approccio per analizzare la rete di collegamenti di un dato utente al fine di suggerirgli nuovi amici, gruppi ecc. Un altro esempio è rappresentato da Spotify che crea una "stazione radio" di canzoni suggerite analizzando regolarmente la musica che un utente ascolta per poi compararla con quella ascoltata da utenti dal comportamento simile. Vengono così suggerite canzoni mai ascoltate dall'utente cercando tra quelle più ascoltate da utenti con interessi simili. L'idea di fondo è che se due canzoni hanno un *rating* simile sono probabilmente canzoni simili o correlate allo stesso artista o genere musicale e se una grande quantità di *users* ascolta le tracce x, y, e z è altamente probabile che esse siano in qualche modo simili.

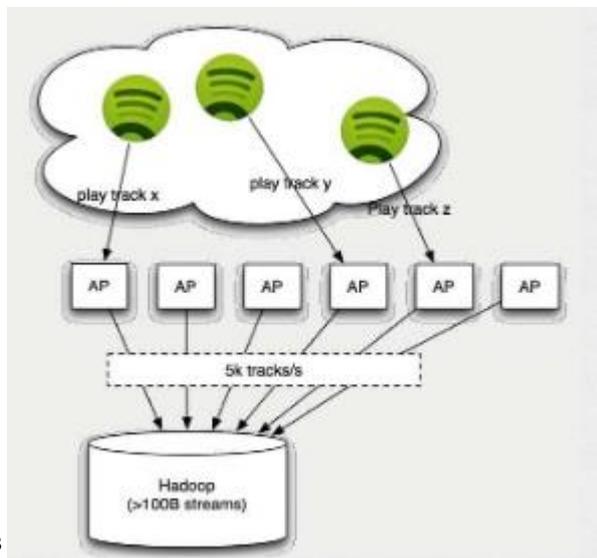


Figura2.12⁶³

$$\left. \begin{matrix} (u_1, i_1, t_1) \\ (u_2, i_2, t_2) \\ \dots \\ (u_n, i_n, t_n) \end{matrix} \right\} 100B \text{ log entries}$$

$$M = \left(\begin{matrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \vdots & \vdots & \dots & \vdots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{matrix} \right) \left. \vphantom{\begin{matrix} c_{11} \\ c_{21} \\ \vdots \\ c_{m1} \end{matrix}} \right\} 10^7 \text{ users}$$

10^7 items

I principali elementi di attenzione per la massima efficacia dell'approccio suddetto sono:

- Accesso ai Big Data: questi sistemi richiedono come prerequisito una grande quantità di dati sull'utente al fine di fornire *recommendatio* accurate;
- Scalabilità: i sistemi suddetti operano frequentemente in situazioni caratterizzati dalla presenza di diversi milioni di prodotti e utenti per cui è necessario un elevato livello di potenza di calcolo e di risorse computazionali per calcolare le *recommendation*;
- Variabilità: dato la vastità di oggetti venduti su una tipica piattaforma di e-commerce, è molto probabile che il rating dato dagli utenti sia relativo ad un insieme limitato di oggetti polarizzando in qualche modo le *recommendation*.

⁶³ Storage for Big Data and Analytics Challenges – Infinidat (2016)

Un modo semplice per descrivere tale approccio è dato dalle seguenti espressioni: “Utenti che hanno scelto questo prodotto hanno scelto anche questi prodotti”; “Utenti simili a te hanno scelto anche questi prodotti”.

2.3.2 *Content-based filtering*

Questa tecnica si basa sull'utilizzo di una descrizione degli oggetti che tiene conto di uno o più di attributi al fine di suggerire oggetti con caratteristiche simili a quelli preferiti in passato dall'utente.⁶⁴ Per ciascun oggetto viene creato un profilo dato da un insieme di attributi e funzionalità. Sulla base del comportamento passato dell'utente viene assegnato a ciascun attributo un peso. Ogni oggetto all'interno del sistema viene così classificato tramite un vettore pesato di attributi. I pesi riflettono l'importanza di ogni attributo o caratteristica per l'utente e possono essere calcolati utilizzando diverse tecniche a partire da una semplice media pesata delle preferenze degli utenti ad approcci più sofisticati basati sul *machine learning* come reti neurali, *cluster analysis*, *decision tree* finalizzati a determinare la probabilità di preferenza di un oggetto per un dato utente. I feedback assegnati direttamente dall'utente, solitamente in termini di “like” o “dislike”, sono usati per assegnare un peso maggiore o minore all'importanza di certi attributi.

Un tipico esempio di *content-based recommender system* è rappresentato da Pandora Radio. Utilizzando gli attributi descrittivi di una canzone o artista (tratti dagli oltre 400 attributi definiti dal progetto “Music Genome”), il sistema viene alimentato con un insieme iniziale di attributi preferiti dall'utente per determinare le canzoni da trasmettere. Le *recommendations* sono man mano raffinate avvalendosi del feedback dato dall'utente privilegiando le canzoni o gli artisti che presentano gli stessi attributi di quelli oggetto di “like” da parte dell'utente e viceversa.

Una delle principali problematiche di questo approccio risiede nella capacità del sistema di applicare in modo corretto ed efficace le preferenze dell'utente su un dato contenuto ad altre tipologie di oggetti o contenuti. Quando il sistema è limitato a formulare *recommendations* sullo stesso tipo di oggetti o contenuti di quelli già utilizzati dall'utente, il valore complessivo del sistema è inferiore al caso in cui è possibile suggerire altri tipi di contenuto o oggetto. Ad esempio, fornire *recommendation* in termini di articoli di news sulla base dei dati raccolti navigando su siti di news è sicuramente utile ma si otterrebbe un valore maggiore se il sistema offrisse *recommendation* in termini di altre tipologie di contenuto come musica, video, discussioni, ecc.⁶⁵

Diversamente dall'approccio *collaborative filtering*, non è necessario disporre inizialmente di una grande quantità di dati.

⁶⁴ “Content-based recommendation systems analyze item descriptions to identify items that are of particular interest to the user.” Content-Based Recommendation Systems, Michael J. Pazzani and Daniel Billsus

⁶⁵ Cfr: “Recommendation system, an introduction”, Dietmar Jannach, Dortmund (DE), 2014

Tuttavia tale approccio risulta abbastanza limitato nel perimetro di applicazione in quanto capace di fornire solo *recommendations* simili al seme iniziale di attributi, caratteristiche e funzionalità. La seguente espressione sintetizza in modo semplice la logica di base: “Se ti è piaciuto questo prodotto, potrebbero piacerti anche questi altri prodotti”. La tabella seguente offre un quadro sinottico dei due metodi descritti.

Figura 2.13

Approccio	Patrimonio informativo richiesto (volume e varietà di dati)	Perimetro di applicazione	Accuratezza recommendation	Livello di complessità
Collaborative filtering	Elevato (problema cold start e sparsità)	Ampio	Alta	Alto
Content-based filtering	Ridotto	Ridotto	Bassa	Basso

Fonte: ITMedia consulting

2.3.3 Hybrid recommender system

I sistemi ibridi combinano i due approcci suddetti per produrre risultati migliori in taluni casi. Tale approccio può essere implementato in diversi modi:

- calcolando separatamente le *recommendation* utilizzando i due metodi precedenti e poi combinando i risultati;
- aggiungendo elementi dell'approccio *content-based* all'approccio collaborative *filtering* (e viceversa);
- unificando i due metodi in un unico modello.

Diversi studi hanno mostrato che le performance dei sistemi ibridi sono superiori in termini di accuratezza delle *recommendation* rispetto a quelle determinate dall'applicazione di un solo modello. Inoltre, i sistemi ibridi consentono di mitigare alcune problematiche tipiche come il *cold start* o la sparsità. La principale criticità nell'implementazione dei sistemi ibridi è data dal loro elevato livello di complessità che richiede adeguate risorse in termini di investimenti per integrazione nelle attività di business e di ricerca e sviluppo.

Sulla base delle suddette considerazioni, emerge un quadro in cui è ragionevole concludere che le piene potenzialità di questi sistemi possano essere sfruttate da soggetti in possesso di:

- rilevante patrimonio informativo soprattutto in termini di volume di dati e in termini di varietà;
- cospicua e diversificata *user base* per contenere la problematica della sparsità;
- adeguate risorse di calcolo;
- adeguate risorse per sostenere gli investimenti necessari sia a livello tecnologico (data center, connettività, tecnologie *cloud*, etc.) che a livello integrazione, ricerca e sviluppo.

Case study: Netflix

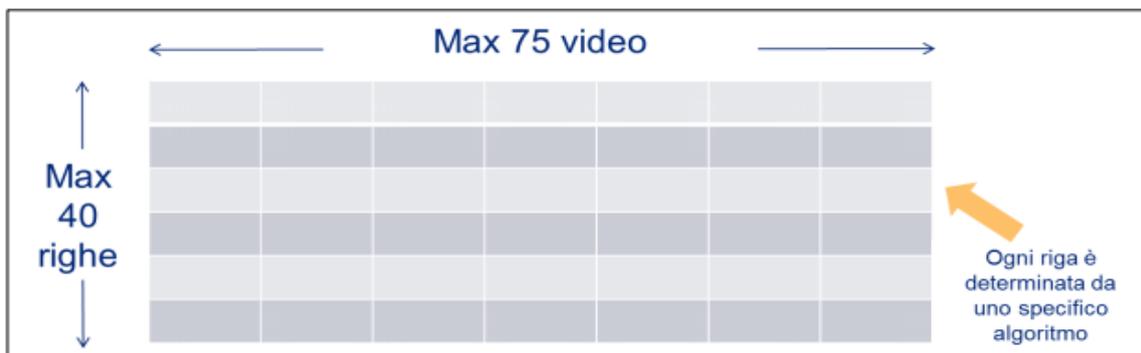
Netflix⁶⁶ rappresenta uno dei più importanti casi di applicazioni di tali tecnologie e in particolare dei sistemi ibridi dato che determina le *recomendation* sia comparando le abitudini di ricerca e visione tra utenti simili (come da approccio *collaborative filtering*) sia suggerendo film con caratteristiche simili a quelle che l'utente ha valutato positivamente (in base all'approccio *content-based filtering*). Quale principale attore nello sviluppo e nella diffusione della Internet TV, Netflix si è adoperata fortemente nell'innovazione del proprio recommender system strutturando e ottimizzando progressivamente tutta una serie di sofisticati algoritmi basati su tecniche di *statistical learning* e *machine learning* (in termini sia di approcci supervisionati come classificazione e regressione, sia di approcci non supervisionati come *clustering*). Un grande impulso alla ricerca in questo ambito è stato dato dalla competizione, denominata Netflix Prize, indetta dalla società nel 2006 mettendo in palio un premio di 1.000.000 \$ al team capace di produrre *recomendation* più accurate del 10% di quelle prodotte dal proprio sistema su un data set di oltre 100 milioni di voti di film. Il premio è stato assegnato il 21 settembre 2009 al team BellKor's Pragmatic Chaos; inoltre, numerosi altri team hanno sviluppato e proposto altri algoritmi che sono stati applicati successivamente in altri settori o mercati. Era in programma una seconda competizione che è stata annullata per via di un accertamento legale da parte della Federal Trade Commission. Come accennato precedentemente, il *recommender system* di Netflix si basa sulla combinazione di diversi algoritmi impiegati per rispondere adeguatamente a diverse esigenze contribuendo alla creazione della *Netflix experience*. Ricerche di mercato e sui consumatori hanno dimostrato che l'utente medio di Netflix perde interesse nei confronti di un contenuto dopo un intervallo di attesa per la scelta che va dai 60 ai 90 secondi, solitamente avendo dovuto scegliere tra 10 o 20 film su uno o due schermi. In questa situazione è auspicabile che l'utente trovi qualcosa di suo gradimento altrimenti il rischio che abbandoni il servizio cresce in modo sostanziale. In questa situazione il *recommender system* deve produrre suggerimenti tali che sia altamente probabile, o meglio ancora certo, che l'utente trovi un contenuto di interesse. La natura stessa dello streaming dei contenuti abilita la raccolta di una grande quantità di dati che descrive esattamente cosa è visto da ogni utente e come viene visto (su che dispositivo, a che ora del giorno, in quale giorno della settimana, con che intensità, ecc.); le informazioni raccolte riguardano anche la posizione all'interno del sito in cui il contenuto è stato scelto e anche gli altri titoli suggeriti ma non selezionati dall'utente.⁶⁷

⁶⁶ "Netflix is a company that handles a big collection of television programs and movies, by streaming it at any time via online (computers or TV). This firm is profitable because the users do a monthly payment to get access to the platform. However, the clients can cancel their subscriptions at any time." Recommendation System for Netflix, Vrije Universiteit Amsterdam, 29 Gennaio 2018

⁶⁷ Cfr: The Netflix Recommender System: Algorithms, Business Value, and Innovation, Carlos A. Gomez-Uribe and Neil Hunt

Il *recommender system* di Netflix capitalizza adeguatamente questo grande patrimonio di dati soprattutto nella costruzione dell'homepage personalizzata per utente. A prescindere dal dispositivo utilizzato, l'homepage è la parte di Netflix che contiene il maggior numero di numero di *recommendation* dando luogo a oltre i due terzi dei minuti visti dall'utente. Il layout della pagina ha una struttura a matrice in cui ogni istanza presentata è una *recommendation* ogni riga contiene video dal "tema" analogo. Ogni riga presenta un'etichetta relativa al tema con l'obiettivo di rendere questo trasparente e più intuitivo per gli utenti. Di solito vi sono circa 40 righe nell'homepage e fino a 75 video per ciascuna riga (questi numeri possono variare sulla base delle caratteristiche hardware del dispositivo utilizzato con l'obiettivo di assicurare la migliore *user experience* possibile). I video di ciascuna riga sono generalmente determinati da uno specifico algoritmo.

Figura 2.14



Fonte: ITMedia consulting

Gli algoritmi utilizzati da Netflix sono di seguito descritti.

- *Personalized Video Ranker (PVR)*: Questo algoritmo è utilizzato per le righe relative a specifici generi. L'algoritmo ordina in modo personalizzato l'intero catalogo di video (o un suo sottoinsieme selezionato per genere) per ogni utente. Il risultato è dato da un elenco di video ordinati per un dato genere (e ciò spiega perché una riga relativa ad uno stesso genere può mostrare risultati completamente differenti per diversi utenti). I risultati possono essere migliorati combinando l'ordinamento personalizzato sulla base del comportamento dell'utente con una piccola componente legata alla popolarità dei film (utilizzata per determinare le *recommendation* della riga "Popular").
- *Top-N video Ranker*: Questo algoritmo produce le *recommendation* per la colonna denominata "Top Picks". Il suo obiettivo consiste nel trovare un set limitato tra le migliori *recommendation* personalizzate all'interno dell'intero catalogo scegliendo tramite l'utilizzo di specifiche metriche solo i video reputati "migliori" tra quelli calcolati con l'algoritmo PVR.⁶⁸

⁶⁸ "There are two types of trends that this ranker identifies nicely: (1) those that repeat every several months (e.g., yearly) yet have a short-term effect when they occur, such as the uptick of romantic video watching during Valentine's Day in North America, and (2) one-off, short-term events, for example, a big hurricane with an impending arrival to some densely populated area, being covered by many media outlets, driving increased short-term interest in documentaries and movies about hurricanes and other natural disasters." The Netflix Recommender System: Algorithms, Business Value, and Innovation, Carlos A. Gomez-Uribe and Neil Hunt

Figura 2.15



Fonte: Netflix, *The Netflix recommender system: algorithms, business value, and innovation*

- *Trending Now*: Questo algoritmo, che determina la riga dal corrispondente nome, punta a sfruttare la rilevanza temporale di alcuni trend che a volte si hanno in intervalli limitati, da pochi minuti ad alcuni giorni, coniugandola con la personalizzazione. Tale algoritmo classifica adeguatamente due tipologie di trend: quelli che si ripetono periodicamente in base a tradizioni, festività, ecc (ad esempio come nel caso di San Valentino); quelli relativi a situazioni o eventi momentanei o improvvisi di varia natura.
- *Continue watching*: Tale algoritmo è utilizzato per contemplare le nuove abitudini di consumo dei contenuti, sia in modo seriale che in modo non episodico tramite la visione di piccoli spezzoni. L'ordinamento dei contenuti visti di recente si basa sulla stima di cosa l'utente intenda continuare a vedere, rivedere o non vedere più. La stima è effettuata tenendo conto di diversi parametri: il punto di abbandono (metà programma, inizio, fine), il dispositivo utilizzato, il tempo intercorso dall'ultima visione, ecc.⁶⁹
- *Video similarity*: Questo algoritmo è utilizzato per determinare le *recommendation* della riga "Because you watched (BYW)". Si tratta di un algoritmo non personalizzato che calcola un elenco di video "gemelli" tra tutti quelli presenti nel catalogo. Si hanno così numerose righe di video "gemelli". La scelta della riga di video "gemelli" da presentare all'utente si basa, invece, su elementi di personalizzazione relativi a molteplici aspetti a partire dai video già visti.⁷⁰
- *Page generation Date*: Netflix utilizza un algoritmo di *Page Generation* che utilizza l'output di tutti algoritmi descritti precedentemente. Per ogni utente tipo vi sono infatti migliaia di righe potenzialmente utilizzabili per l'homepage. Questo algoritmo matematico interamente personalizzato seleziona e ordina in modo ottimizzato le righe più rilevanti per l'utente. Non basandosi su template prefissati, l'algoritmo consente di costruire in modo flessibile homepage senza una certa tipologia di riga (ad esempio BYW o Popular) per un dato utente, homepage con metà righe dello stesso tipo per un altro utente, ecc.

⁶⁹ "Most of our rankers sort unviewed titles on which we have only inferred information."

⁷⁰ "Even though the sims ranking is not personalized, the choice of which BYW rows make it onto a homepage is personalized, and the subset of BYW videos recommended in a given BYW row benefits from personalization, depending on what subsets of the similar videos we estimate that the member would enjoy (or has already watched)." *The Netflix Recommender System: Algorithms, Business Value, and Innovation*

- *Evidence selection*: Si tratta di altri algoritmi orientati a determinare che tipologia di informazioni far vedere nell'area "evidence" posizionata nella parte superiore di ogni contenuto selezionato. Ad esempio, questi algoritmi decidono se mostrare che il film ha vinto un oscar o che il film è simile ad uno visto di recente dall'utente. Altre informazioni riguardano quale immagine mostrare, tra alcune disponibili, per supportare al meglio ogni *recomendation*.
- *Search algorithms*: Tutti gli algoritmi suddetti costituiscono il *recommender system* di Netflix e influenzano l'80% di tutte le ore di contenuto mostrato sulla piattaforma. Il restante 20% deriva dalle ricerche effettuate dagli utenti. Per ottimizzare anche questo aspetto della *user experience*, si utilizzano specifici algoritmi basati su numerosi dati tra quelli relativi ai contenuti visti dall'utente e alle ricerche effettuate, che di fatto trattano la ricerca come un caso particolare di *recomendation*. Come indicato in figura, ad esempio effettuando una ricerca con la parola "fren" un algoritmo restituisce contenuti che contengono tale parola nel titolo ("Frenemies" in figura); un altro algoritmo tenta di prevedere contenuti relativi ad un concetto desumibile dalla chiave di ricerca ("French movies" in questo caso).⁷¹



Figura 2.16 Fonte: Netflix, *The Netflix recommender system: Algorithms, business value, and innovation*

Lo sviluppo e il miglioramento continuo del *recommender system* di Netflix hanno comportato e continuano ad apportare interessanti benefici in termini di valore di business.

Tale sistema è infatti essenziale nell'affrontare in modo efficace il cosiddetto momento della verità ovvero quando un utente inizia una sessione e riesce a trovare qualcosa di interessante in pochi secondi risultando soddisfatto e appagato. Ciò previene efficacemente un'eventuale cancellazione dell'abbonamento.

Inoltre, tramite la possibilità di fornire *recomendation* mirate per ogni singolo utente, si ha la possibilità di sfruttare pienamente l'intero catalogo di contenuti anche in relazione a quei film di nicchia che quasi sicuramente sarebbero un insuccesso sulla TV broadcast a

⁷¹ "The search experience is built around several algorithms. One algorithm attempts to find the videos that match a given query, for example, to retrieve Frenemies for the partial query "fren." Another algorithm predicts interest in a concept given a partial query, for example, identifying the concept French Movies for the query "fren." A third algorithm finds video recommendations for a given concept, for example, to populate the videos recommended under the concept French Movies. Our search algorithms combine play data, search data, and metadata to arrive at the results and recommendations that we offer." The Netflix Recommender System: Algorithms, Business Value, and Innovation

causa di un'audience troppo bassa tale da trainare adeguati ricavi pubblicitari che giustificano l'occupazione dello slot temporale utilizzato. Il *recommender system* permette quindi di distribuire il consumo di contenuti su un ventaglio molto ampio di titoli come ben testimoniato dall'ECS (*Effective Catalog Size*) relativa alla dimensione effettiva del catalogo. Questa metrica descrive quanto diffusa è l'attività di consumo sull'intero catalogo. Se la maggior parte del consumo di contenuti deriva da un singolo video tale valore sarà molto vicino a 1; se invece tutti i video del catalogo generano la stessa quantità di visualizzazioni allora il valore dell'ECS sarà vicino a quello del numero di titoli del catalogo. La curva di colore nero in figura mostra l'andamento dell'ECS in relazione al numero dei video più popolari nel catalogo, spaziando, sull'asse delle ascisse, da 1 a N (con N pari al numero di questi video del catalogo). La curva di colore rosso mostra l'andamento dell'ECS per i primi k video determinati con l'algoritmo PVR. Si osserva che l'effetto della personalizzazione apportata dall'algoritmo PVR consente di ottenere un miglior sfruttamento dei set di video rispettivamente considerati con un picco pari a 4 volte il valore del caso senza il citato algoritmo.

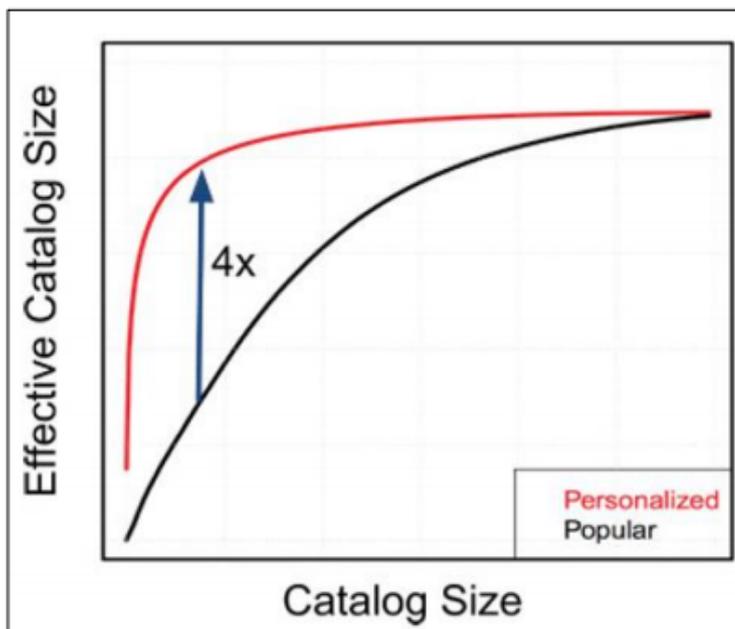


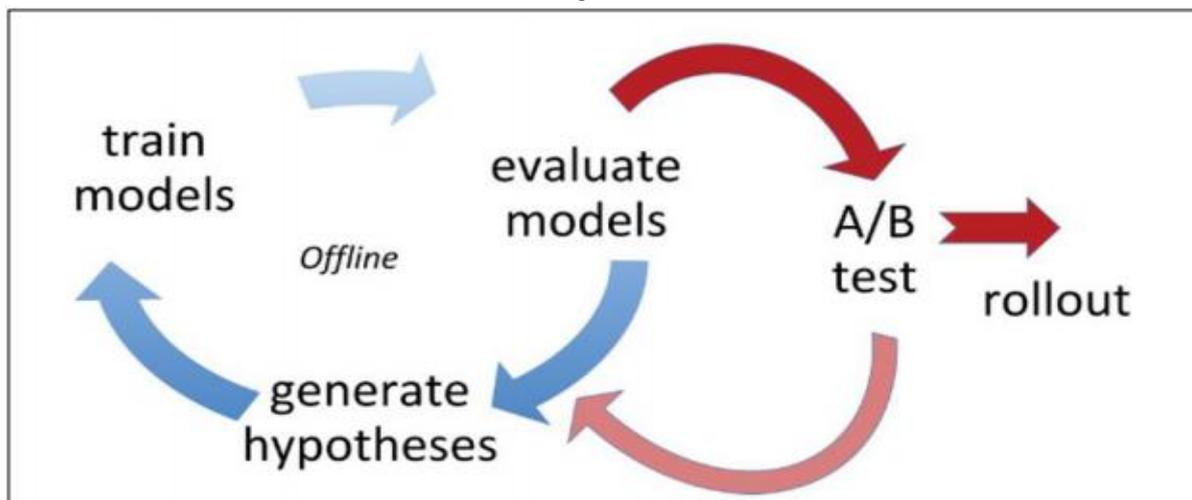
Figura 2.17 ⁷²Fonte: Netflix, *The Netflix recommender system: Algorithms, business value, and innovation*

Grazie alla massimizzazione del catalogo e alla personalizzazione abilitata dalle *recommendation*, Netflix è stata capace nel corso degli anni di abbassare il *churn rate* mensile di diversi punti percentuali giungendo ad un valore attuale molto basso ovvero a singola cifra bassa, dovuto per lo più a problematiche legate al sistema di pagamento

⁷² "The black line is the effective catalog size (ECS) plotted as a function of the number of most popular videos considered in the catalog, ranging from 1 through N (the number of videos in the catalog) on the x-axis. The red line is the effective catalog size for the first k PVR-ranked videos for each member. At a PVR rank corresponding to the median rank across all plays, the ECS in red is roughly 4 times that in black. The values in the x and y axis are not shown for competitive reasons. For more details, see Appendix A. (Right) The take-rate from the first k ranks, as a function of the video popularity rank in black, and as a function of the PVR rank in red. The y-values were normalized through division by a constant so that the maximum value shown equalled 1." The Netflix Recommender System: Algorithms, Business Value, and Innovation

piuttosto che alla scelta esplicita degli utenti di cancellare l'abbonamento. Ciò comporta un aumento complessivo del valore nel tempo degli abbonati e riduce il numero di nuovi abbonati che è necessario acquisire per compensare quelli che abbandonano il servizio. Tutto ciò produce, a detta della società, risparmi pari a circa 1 miliardi di dollari l'anno. L'evoluzione del *recommender system* di Netflix si basa su sforzi costanti di innovazione volti sia a ottimizzare gli algoritmi attualmente impiegati che a svilupparne di nuovi. In particolare, il processo di ricerca e sviluppo è di tipo iterativo-incrementale ed è svolto secondo lo schema riportato in figura. Nella fase off-line, il nuovo algoritmo viene istruito utilizzando specifici *train model* come da paradigma *Machine Learning* basati sui dati storici solitamente impiegati per produrre *recommendation* accurate. Non appena si iniziano ad intravedere risultati incoraggianti in queste sperimentazioni off-line si prepara un test A/B per il nuovo algoritmo. Il test A/B è sistematicamente utilizzato per valutare in modo casuale e controllato la prestazione ottenuta da ogni algoritmo in relazione all'engagement di medio termine degli utenti rispetto al tasso di cancellazione di abbonamenti. Se il risultato del test è positivo, il nuovo algoritmo entra a far parte, in aggiunta o eventualmente in sostituzione di qualche altro algoritmo, dell'insieme di algoritmi utilizzati dal recommender system. Se il risultato del test è negativo o simile a quello ottenuto da altri algoritmi impiegati si hanno due possibilità: abbandonare le attività di ricerca e sviluppo su questo algoritmo o al contrario intensificarle per provare a ottenere in seguito un buon risultato nel test A/B.

Figura 2.18



Fonte: Netflix, *The Netflix recommender system: Algorithms, business value, and innovation*

73

⁷³ "We iterate quickly to prototype an algorithm through offline experimentation by analyzing historical data to quantify how well a new algorithm can predict previous positive member engagement, such as plays. The key underlying assumption, which is not always true, is that members would have engaged with our product in exactly the same way, for example, playing the same videos, had the new algorithm been used to generate recommendations. Once we see encouraging-enough results in offline experiments, we build an A/B test to use the new algorithm to generate recommendations for members. If the A/B test succeeds, we change our product to use that new algorithm by default. If the A/B test is flat or negative, we either abandon the research direction or go back to the offline experimentation world to try to make the new algorithm even better for a possible future A/B test." *The Netflix Recommender System: Algorithms, Business Value, and Innovation.*

Attualmente le attività di ricerca e sviluppo sono orientate a migliorare e ottimizzare le seguenti problematiche: creazione dell'homepage; scelta delle migliori *evidence* per supportare adeguatamente le *recommendation*; coldstart per i nuovi membri sui quali si hanno pochi dati storici limitando di fatto l'accuratezza delle *recommendation* (l'approccio attuale consiste nel proporre a un nuovo utente durante la fase di registrazione una *suvery* durante la quale gli si chiede di selezionare dei video da un insieme creato automaticamente e utilizzato come input per tutti gli algoritmi); creazione di *recommendation* "globali" per supportare l'espansione internazionale di Netflix suggerendo il contenuto migliore in ogni paese in cui l'azienda opera tenendo conto dei vincoli di lingue e sottotitoli disponibili (dato che sarebbe troppo laborioso e costoso tradurre o produrre sottotitoli in tutte le lingue è molto utile sapere le lingue parlate o al limite comprese da ogni utente analizzando le informazioni dei contenuti visionati e delle modalità di visione: con sottotitoli, in lingua originale, ecc.).

2.4 On-line advertising

La pubblicità on-line è definita come ogni forma di pubblicità che arriva al consumatore per mezzo di internet, a prescindere dal *device* di riferimento (pc, tablet, smartphone).

Ci sono vari tipi di pubblicità on-line: *banner advertising*, *video advertising*, *search engine advertising* e *display advertising*. Il primo banner pubblicitario apparve nel 1994, ad opera di AT&T nel sito di hotel "hotWired", pagando per una visibilità di 3 mesi oltre 30000 USD.⁷⁴ Questo momento coincise con la nascita di questo nuovo mercato della pubblicità on-line. Oggi la pubblicità on-line costituisce un settore in grandissima espansione ed attualmente rappresenta una ricchezza di oltre 200 bilioni di dollari. Questo è stato reso possibile grazie al forte incremento della popolazione con accesso ad Internet, che è passato dai 16 milioni del 1995 ai 3,9 bilioni del 2017. Tale settore è composto da numerosi *players* e, a causa dell'elevata differenziazione al suo interno, è costituito da poche piattaforme aventi un'elevata QDM e diversi altri operatori più piccoli. Quindi, quello della pubblicità on-line è un settore con un elevato grado di concentrazione, che vede in Google e Facebook i maggiori *players* attivi, rispettivamente, nei sottosettori della *search advertising* e *non search advertising*.

Nel 1996 DoubleClick iniziò a monitorare il successo dei banner pubblicitari in tempo reale, grazie alla possibilità di misurare il numero di click che quel banner riceveva dagli utenti. Da quel momento nacque così il primo business model delle pubblicità on-line, ovvero il cosiddetto "*cost per impression model*" (CPI) in cui il proprietario del banner pagava una *fee* al sito che ospitava il banner, ogni qual volta che un utente aveva accesso alla pagina e guardava la pubblicità.

⁷⁴"Online advertising", Bundescartellamt, Febbraio 2018

Share of Data Collection Across the Web

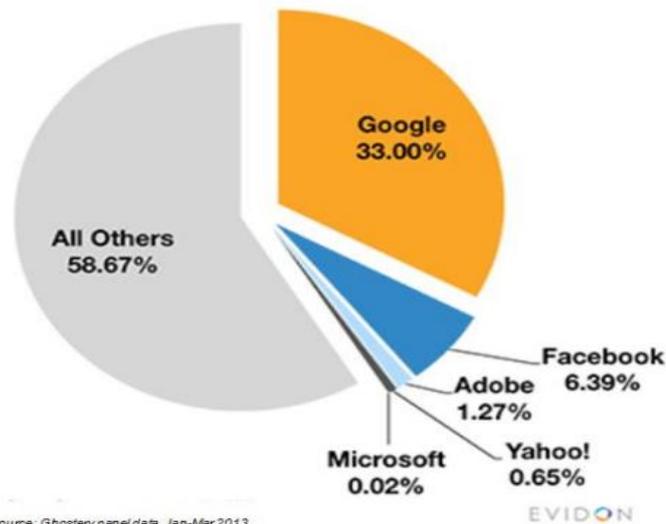


Figura 2.18

Nel 1997 nacque la *pop up advertising* ad opera dell'agenzia Tripod.com; ovvero una pubblicità inserita dentro una finestra (pop-up) che si distingueva in maniera più evidente dal sito che ospitava tale pubblicità. Questa tipologia pubblicitaria ebbe un grosso impatto, fino a che, nei primi anni del 2000, nacquero i cosiddetti *pop-up blockers* che portarono alla sua decadenza.

Il motore di ricerca dominante negli anni 90, GoTo.com, sviluppò, alla fine di quella decade, una nuova tipologia di business model: *pay per click model* (PPC). Questa consisteva nel far pagare una fee all'advertiser ogni qual volta un utente del motore di ricerca cliccava sulla pubblicità. Inoltre questo meccanismo, a monte, prevedeva una forma di asta nella quale, maggiore era l'offerta dell'advertiser, ovvero la sua disponibilità a pagare per click, maggiore visibilità acquistava la pubblicità nella pagina del motore di ricerca. Il modello PPC, in concomitanza con quello dell'asta sono quelli maggiormente utilizzati ancora oggi.

Col passare degli anni e col crescere del volume dei big data raccolti ed analizzati, è stato possibile targettizzare la pubblicità, ovvero far arrivare messaggi pubblicitari distinti per tipologia di cluster di utenti. Ciascun individuo, dunque, in base alle ricerche che effettuava e ai dati che conseguentemente forniva, vedeva pubblicità sempre più allineate ai suoi gusti.

Google, con il suo programma *Google AdWords*⁷⁵, rappresentò il primo esempio di targettizzazione pubblicitaria. Il programma, infatti, permetteva agli inseritori pubblicitari di inserire la loro pubblicità ogni qual volta veniva effettuata una ricerca che comprendesse delle *keywords* che, a priori, venivano vendute da Google all'advertiser. Con la nascita di *Google Analytics*, il motore di ricerca forniva una mole di informazioni sul traffico in internet sempre maggiore e sempre più dettagliate. Questo rafforzò ancor di più la capacità di Google di fornire pubblicità sempre più *target* tizzata, non che il suo potere di mercato.

⁷⁵ Cfr: *Case study: Google's Business models*

Nel 2006 Facebook, che già si stava affermando come social network dominante a discapito di Netscape, pose sul suo sito la prima forma di *display* (o social network) *advertising*, avente la forma di un piccolo banner o link sponsorizzante. Questa tipologia di advertising consiste in immagini, video, testi, grafiche altamente *target* izzabili a seconda degli utenti che accedono al sito.

Nel 2007 Google creò *AdSense for Mobile*, che permetteva agli advertisers di piazzare la pubblicità on-line anche su smartphone e tablet. Così come iniziò a crescere in maniera esponenziale l'utilizzo di tali devices e l'accesso ad internet di tali dispositivi (e tale trend è stimato essere in continua crescita), anche la cosiddetta *mobile advertising* iniziò ad assumere sempre maggiore rilevanza nel settore della pubblicità online.

La pubblicità online, dunque, si differenzia sia per tipologia di mezzo di comunicazione (immagine, testo, audio, video), sia per "luogo" dove viene collocata (website, pop up, video). Le tipologie di pubblicità maggiormente utilizzate sono la *search advertising* e la *display advertising*, senza tralasciare l'importanza del *mobile online advertising* che ha visto crescere notevolmente il suo utilizzo negli ultimi anni.

La *Search advertising* consiste nel collocare pubblicità accanto alla pagina di ricerca che è stata aperta dall'utente a seconda dell'input, o keyword, che è stata inserita nella barra di ricerca del motore di ricerca utilizzato. La pubblicità oltre a potersi collocare accanto, per esempio alla destra dei risultati di ricerca, può situarsi anche in cima ai risultati costituendo la cosiddetta *organic search results*. AdWords⁷⁶ è il prodotto offerto da Google, Bing Ads quello di Microsoft, i principali player nel settore dei motori di ricerca. Nella *search advertising* gli utenti hanno già indicato un interesse concreto quando fanno una *search query* (ovvero una domanda al motore di ricerca). Così facendo, è possibile porre pubblicità su misura, ovvero targettizzata, a determinati cluster di utenti di internet. La *search advertising* viene utilizzata per attrarre gli utenti, sollecitandoli ad acquistare determinati prodotti. In tale tipologia pubblicitaria, gli spazi pubblicitari vengono venduti attraverso il meccanismo dell'asta e il motore di ricerca ci guadagna prendendo percentuali rilevanti sul totale PPC (*pay per click method*) pagato dall'*advertiser*. Il *Search engine optimization* (SEO) è correlato alla *Search engine advertising*.⁷⁷ Gli *Advertisers*, infatti, utilizzano il SEO al fine di dare maggiore visibilità al loro sito nella lista dei risultati organici del motore di ricerca. Questo metodo permette ai proprietari dei siti di posizionarsi in cima ai risultati organici grazie all'attrattività e quindi all'efficacia di tali siti.

⁷⁶ Cfr: *Case study: Google's Business models*

⁷⁷ "Search engine optimisation (SEO) is related to search engine advertising. Advertisers use SEO to get their own website as high on the list of organic results as possible. It covers all those methods which improve a site's positioning in a search engine's unpaid organic results lists. 23 These include optimising the content of individual subpages of a website, incorporating the most appropriate keywords and other metadata, or keeping the volume of data to be transferred when a page is viewed to a minimum." "*Online advertising*", Bundescartellamt, Febbraio 2018.

La *Display advertising* o *banner advertising* è costituita, come detto, da contenuti multimediali, quali immagini, video, testi ecc. Questi vengono situati nei siti, specialmente nei social network, in cui c'è un notevole traffico di utenti. Poiché in passato la *Display advertising* aveva poche opportunità di essere targettizzabile, veniva utilizzata principalmente al fine di avere una cognizione e consapevolezza maggiore di un determinato brand. Ora, grazie ai *Big Data* si è reso possibile tracciare quali brand sono di maggior gradimento per tipologia di consumatore e di conseguenza inserirli nelle *home page* degli utilizzatori dei social networks. È possibile targettizzare in tre modi: per profilo utente, ovvero fare una campagna pubblicitaria orientata a un *target*; andare su un tema e non sulla personalizzazione dell'utente (ad esempio vendo ombrelli e vado su *Meteo.it* perché le persone lo consultano quando c'è rischio di pioggia; *Remarketing* (ad esempio quando si prenota un albergo su Booking su tutti i siti per la prossima settimana appare booking), dipende dalle visite precedenti, è una strategia tipica di acquisti complessi con molta fase di ricerca. Le pubblicità vengono consegnate attraverso i cosiddetti *ad servers*, che registra anche il numero di visualizzazioni e di click che gli utenti effettuano; oppure, attraverso le *data management platforms* (DMPs), che sono emerse solo di recente, che permette agli *advertisers* o alle agenzie pubblicitarie di acquistare spazi pubblicitari insieme ai dati con una procedura simile al *bundle*. In sostanza, il DMPs serve a collezionare, confrontare e gestire larghi volumi di dati che permettono la targettizzazione degli spazi pubblicitari

La *Mobile advertising*, essendo disegnata su misura per dispositivi piccoli, come smartphone e tablet, ha caratteristiche diverse e si lega alle varie app disponibili sui vari distributori ed utilizza dati peculiari dei dispositivi mobili, come la geolocalizzazione. Essa è anche l'unica fonte di *revenue* per moltissime app che forniscono un servizio gratuito e che quindi sono *two sides platforms*, mettono infatti in contatto inserzionisti pubblicitari da un lato e gli utenti che scaricano l'app dalle varie piattaforme di distribuzione.

Case study: Google's Business model

Google può essere vista come una *multi-sided platform* attiva in più mercati rilevanti. Infatti sono svariate le attività che detiene Google nel web. Essa fornisce sia servizi gratuiti, come il motore di ricerca, il browser Google Chrome, Youtube, sia servizi a pagamento come gli slot pubblicitari.⁷⁸

È infatti la *online advertising* la fonte maggiore di *revenue* di Google, che opera in tutti i sottosettori fornendo le più disparate tipologie di pubblicità e mettendo in contatto inserzionisti, siti web e utenti di tutto il mondo.

⁷⁸"L'impatto di Google sull'economia", Deloitte (2014)

Tale piattaforma, infatti, si interfaccia da un lato direttamente con gli inserzionisti pubblicitari che si situano sul motore di ricerca e nella *search engine organic result*⁷⁹, dall'altro con i siti web, anch'essi facenti parte della *search engine organic result* e, indirettamente ed attraverso quest'ultimi, con altri inserzionisti pubblicitari che si collocano in tali siti web. Ovviamente, come detto, Google offre anche molteplici servizi gratuiti agli utenti consumatori finali. Possiamo dunque dire che i lati della piattaforma di Google sono sostanzialmente costituiti da *transaction* e *non transaction markets*, ovvero è composta da mercati in cui vi è scambio di beni (dati vs servizi gratuiti) senza corrispettivo monetario e mercati in cui si scambiano servizi digitali che hanno un loro prezzo (pubblicità ed altro, come l'accesso a taluni servizi digitali).

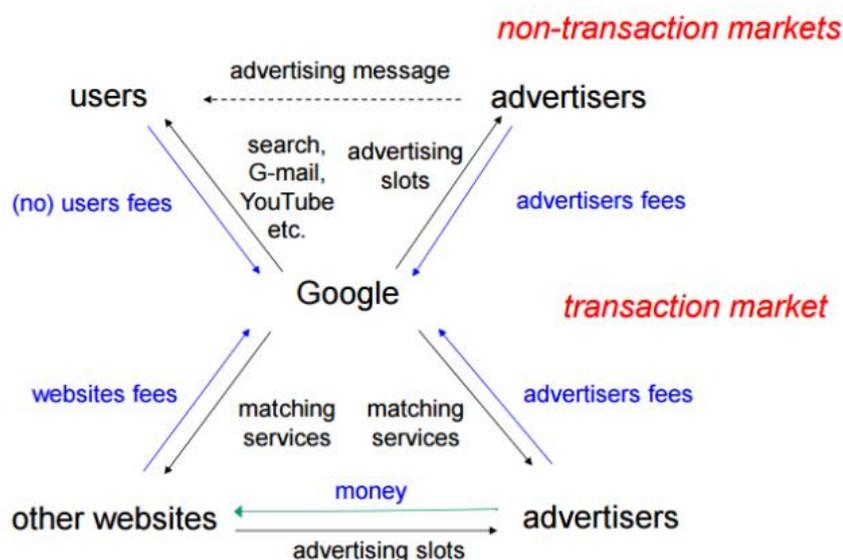


Figura 2.19

Source: Filistrucchi, Geradin, Damme, Affeldt (2013)

Google, dunque, intesa come *online platform*, fa della pubblicità il suo *core business*. Circa l'88% del fatturato di Google è fatto attraverso l'*online advertising*. Negli ultimi anni, tuttavia si è registrata una forte crescita della parte *cloud* (circa del 40%), diventando così una nuova grande fonte di *revenue*.

Passiamo ora ad elencare i vari *business models* che vedono nella raccolta dei dati il *core asset* al fine di fornire pubblicità sempre più efficienti, in termini di visibilità e customizzazione, ai clienti e ai consumatori finali.

Motore di ricerca: è l'asset centrale di Google, si compone di tre processi: crawling, indicizzazione e algoritmo.⁸⁰

⁷⁹ Organic search is a method for entering one or several search terms as a single string of text into a search engine. Organic search results, appear as paginated lists, are based on relevance to the search terms; and exclude advertisements. Whereas, non-organic search results do not filter out pay per click advertising.

⁸⁰ "Personalized *targeting* type reports may include contextual *targeting* when user data, such as cookie ID, isn't available. If you've selected "Non-personalized ads" in your EU user consent settings, you might still see ads under the "Personalized" *targeting* type even though user data isn't being used." "L'impatto di Google sull'economia", Deloitte (2014).

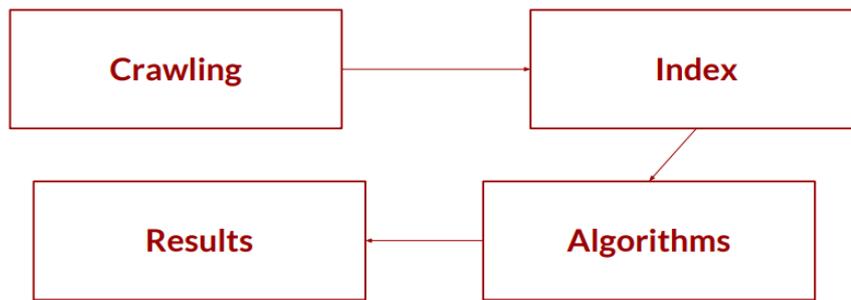


Figura 2.20

Nel primo processo, il *crawling*, vi sono programmi, detti *crawl*, che in tempo reale scansionano e immagazzinano tutto ciò che è situato sulla rete internet.⁸¹ Essi mappano e tracciano tutto, chiunque crea un contenuto web, i *crawlers* lo tracciano, prima che avvenga per la ricerca. chiunque crea un sito o fa una ricerca, viene tracciato dai *crawlers*. Google cerca dentro i crawl che ha creato prima, questo è il motivo tale per cui Google è così veloce. I crawl immagazzinano i dati e creano l'indice. Per l'appunto, i siti che non vogliono essere tracciati devono inserire un codice da inserire nel sito per escludere la visibilità dai *crawlers*. Google non è il primo motore di ricerca, ce ne sono infatti altri, anche meno recenti di Google, che si basavano sulla ricerca di una o più parole. La differenza sostanziale con gli altri *players* del settore è che per Google non tutte le pagine sono autorevoli, ma sono valutate a seconda di quante altre pagine lo linkano (ad esempio, nelle referenze accademiche un autore è più autorevole quanto più è citato da altri autori). Google ha "crawalizzato" tutta la rete e costruisce algoritmi, attraverso i *crawlers* che tracciano e dettano i dati all'indice che, a sua volta, invia gli input all'algoritmo. I sistemi di *ranking* di Google ordinano centinaia di miliardi di pagine web nell'indice di ricerca per fornire risultati utili e pertinenti in una frazione di secondo. Google non monetizza i risultati della ricerca. Non è infatti possibile pagare Google per arrivare primi nelle ricerche (ovvero nella *search engine organic result*), ma la piattaforma automaticamente fornisce risultati quanto più possibile vicini agli interessi dei clienti consumatori finali. Il modello di business di Google, che si è negli anni rivelato vincente e lo ha consacrato come motore di ricerca dominante (ha infatti una QDM del 90% nei Paesi Europei)⁸², è infatti quello di non monetizzare con il motore di ricerca, ma attraverso la pubblicità. Tale servizio, infatti, viene monetizzato attraverso AdWords, il programma di Google che permette agli inseritori pubblicitari di porre i propri annunci sui risultati del motore di ricerca acquistandone le parole chiave (o *queries*), attraverso il meccanismo dell'asta⁸³.

⁸¹ Prima di fare una ricerca, i web crawler raccolgono informazioni da centinaia di miliardi di pagine web e le organizzano nell'indice di ricerca.

⁸² Gli altri motori di ricerca maggiormente utilizzati sono: Bing di Microsoft, Yandex, DuckDuckGo, Boardreader, Dogpile, Creative Commons Search, Giphy, Quora, Vimeo, WolframAlpha, StartPage, Ask.com, SlideShare.

⁸³ Quando un utente effettua una ricerca, il sistema Google Ads trova tutti gli annunci le cui parole chiave corrispondono a tale ricerca. Tra questi annunci, il sistema ignora quelli che non sono idonei, ad esempio gli annunci indirizzati a un altro paese o non approvati in base a una violazione delle norme. Tra gli annunci rimanenti, possono essere pubblicati solo quelli con un ranking dell'annuncio sufficientemente alto. Il ranking dell'annuncio è una combinazione tra offerta, qualità dell'annuncio, soglie del ranking dell'annuncio, contesto della ricerca dell'utente e impatto previsto delle estensioni e degli altri formati dell'annuncio. Diego Ciulli (2018).

AdWords consente infatti alle aziende di sponsorizzare i link dei loro siti internet, nelle pagine del motore di ricerca. Le aziende scrivono brevi inserzioni e fanno delle offerte per le relative parole chiave, attraverso un'asta on-line. Il vincitore è determinato dall'offerta e da un punteggio di qualità dell'annuncio. Gli inserzionisti pagano Google sulla base del costo-per-click (CPC), ovvero i pagamenti sono dovuti solo se qualcuno clicca sull'annuncio sponsorizzato.

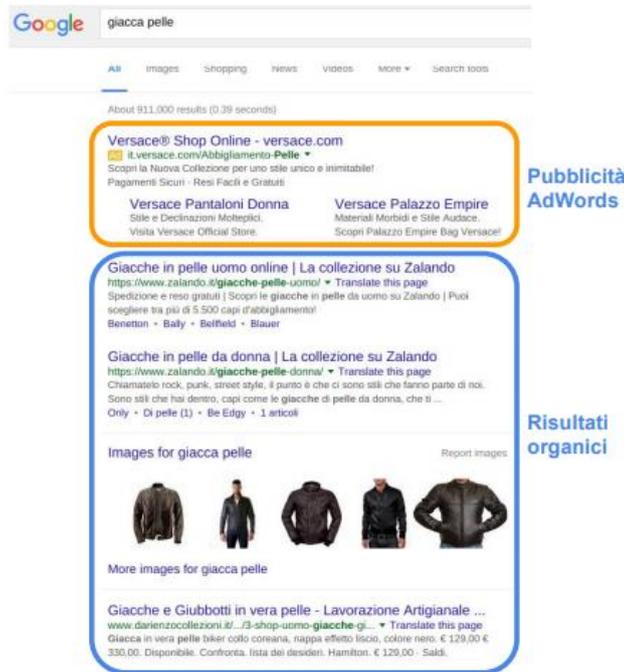


Figura 2.21.

Compito di Google in tal senso è quello di assicurare la cosiddetta *search neutrality*, ovvero garantire che gli algoritmi utilizzati siano neutrali, nel senso che il *ranking* dei risultati, che viene fuori da una ricerca, deve basarsi su metriche oggettive di rilevanza e qualità.

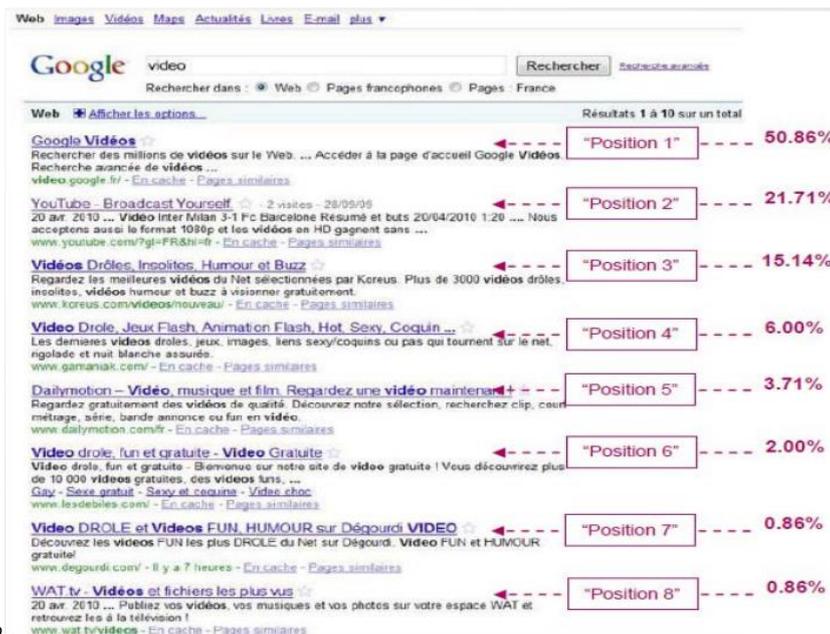


Figura 2.22

84

⁸⁴ "L'impatto di Google sull'economia", Deloitte (2014).

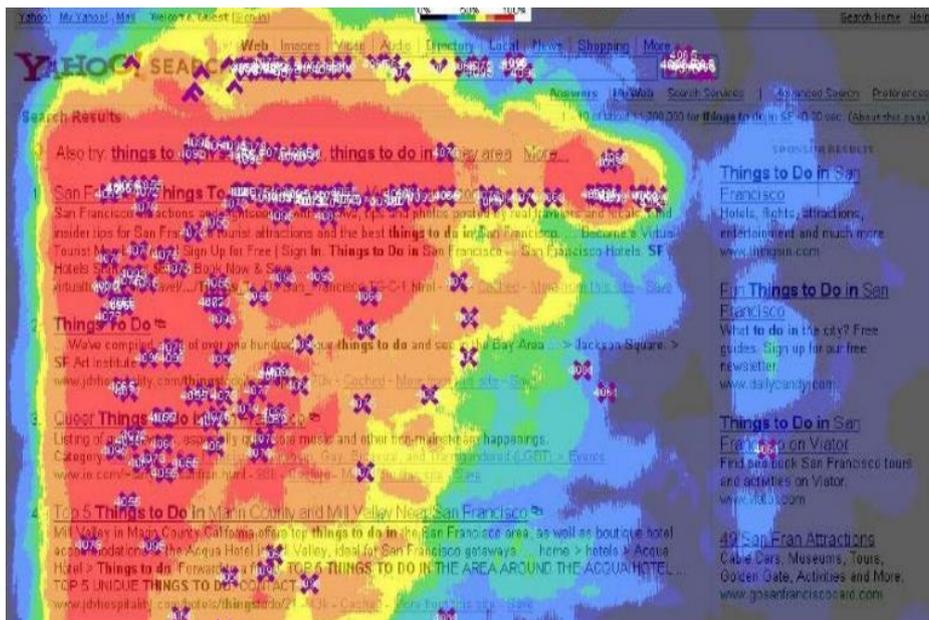


Figura 2.23

Le due figure mostrano l'importanza del ranking. Infatti, circa il 90% dei click vengono fatti per i primi tre risultati che mostra il motore di ricerca, relativamente ad una certa *query*. Il secondo grafico, in particolare, mostra quali sono i punti della *search engine organic result* dove maggiormente si focalizza l'occhio umano. Google quindi inserisce le ricerche più autorevoli nelle prime posizioni dei risultati di una ricerca e via via quelle meno autorevoli, sulla base del numero di click che hanno ricevuto in passato e, tramite AdWords ed il suo meccanismo ad asta, inserisce gli annunci accanto ai risultati della ricerca, dando maggiore rilevanza a coloro i quali hanno pagato di più. Esistono delle tecniche per arrivare primi nel motore di ricerca. Sono nate così delle guide per gli sviluppatori che seguono la logica di far leggere bene il proprio sito web dai *crawlers* in modo da finire tra le prime posizioni di una ricerca. Esistono le cosiddette agenzie SEO, che aiutano i siti a finire in cima alla *search engine organic result* e agenzie web che organizzano e vendono campagne pubblicitarie (Le 10 più grandi aziende SEO in Italia generano da sole oltre 22 milioni di \$ di ricavi annui).⁸⁵ Infine, il meccanismo dell'asta consente alle PMI di gareggiare con aziende più grandi per pubblicizzare i loro prodotti al medesimo pubblico di riferimento.

AdSense: è uno strumento tecnologico che permette a Google di mettere la pubblicità sui siti degli utenti. Infatti, I creatori di contenuti e i proprietari di siti web possono monetizzare i loro contenuti vendendo spazi liberi agli inserzionisti.⁸⁶

⁸⁵ Tra queste, ricordiamo: BizUp, Regex Media, Simpliza, Secretkey, Optimized Group, Netidea Webranking, Ad Sphera Group, Sicomunica, Studio Cappello, Linkomm

⁸⁶ "AdSense permette agli editori di siti web di monetizzare i loro contenuti ospitando pubblicità nei loro siti. AdSense visualizza sui siti web qualsiasi tipo di pubblicità online, inclusi testi, immagini e video. AdSense permette inoltre ai proprietari dei siti web di includere una barra di Google Search nel loro sito e pubblicare annunci insieme ai risultati." Deloitte (2014)



Figura 2.24

AdSense mette in contatto i proprietari di siti e gli inserzionisti attraverso un sistema di compravendita automatizzata (*programmatic advertising*)⁸⁷. La pubblicità, in questo caso, è molto più pervasiva e personalizzata, nel senso che è molto più legata all'individuo che sta effettuando la ricerca. I *banner* sono infatti personalizzati in base alle ricerche effettuate dall'utente in passato. Nel precedente grafico, appare un *banner* pubblicitario di Renault sul sito di Meteo.it. Questo è un *banner* personalizzato in base alle ricerche dell'individuo ed è quindi una forma di pubblicità pervasiva. I proprietari dei siti web che ospitano annunci attraverso AdSense ricevono il 68% dei ricavi generati. La monetizzazione consente ai proprietari di siti e ai *publisher* di offrire gratuitamente ai lettori i loro contenuti. Nell'esempio precedente, Renault paga a Google, che a sua volta paga a Meteo.it. I siti web possono anche porre *banner* pubblicitari senza l'ausilio di AdSense, tuttavia le vendite dirette richiedono trattative commerciali con costi fissi elevati e inaccessibili per le PMI ed inoltre AdSense consente al sito web di verificare il rendimento degli annunci pubblicitari.

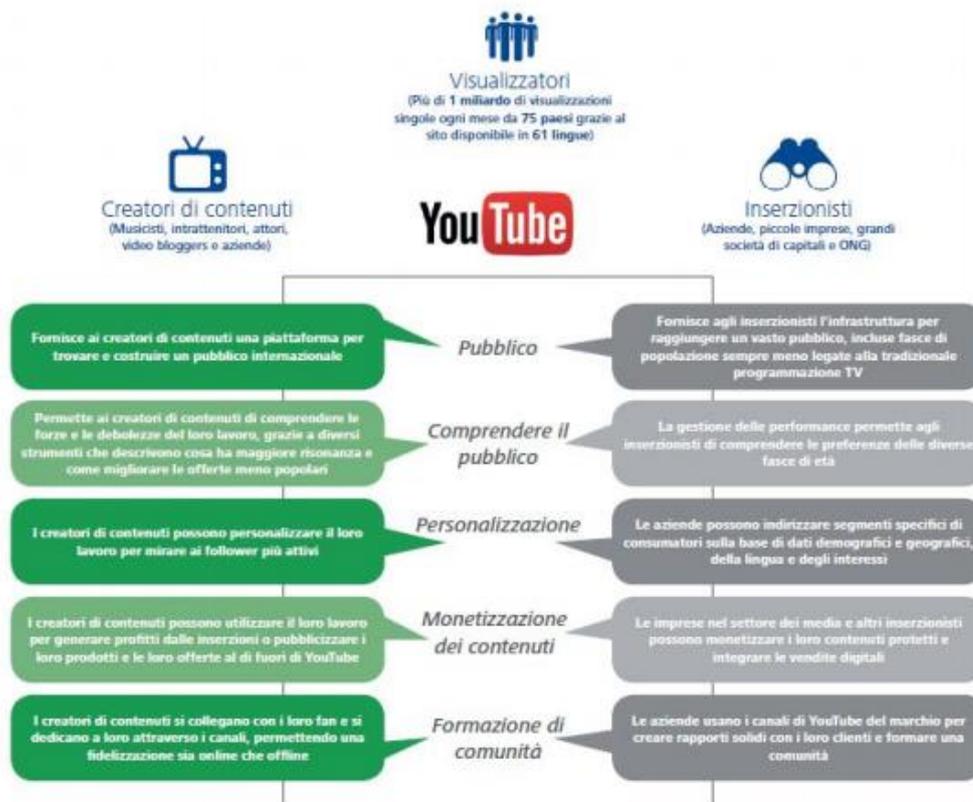
Youtube: I creatori di video possono caricare gratuitamente i propri video su YouTube e scegliere di monetizzare i loro contenuti ospitando pubblicità sul proprio canale.

Tale tipo di pubblicità è costituita da *banners* molto visibili e orientati al *branding*, oppure da video posti all'inizio del contenuto che si intende vedere. Anche in questo caso il modello di *revenues* utilizzato è il CPC (*cost-per-click*) oppure, nel caso dei video, il pagamento avviene solo se il video viene visto per intero senza essere "skippato" dall'utente.

⁸⁷ "Programmatic" ad buying typically refers to the use of software to purchase digital advertising, as opposed to the traditional process that involves RFPs, human negotiations and manual insertion orders. It's using machines to buy ads, basically. "L'impatto di Google sull'economia", Deloitte (2014).



I soggetti che mettono i video, in cambio ottengono la piattaforma gratuita e una parte dei soldi della pubblicità (stesso meccanismo di AdSense). YouTube divide con i creatori dei contenuti i ricavi provenienti dalla pubblicità, nella misura di circa il 55%. Solo i video dei partner ospitano pubblicità. Infine, i diritti d'autore sono molto importanti. La tecnologia Content ID permette ai legittimi proprietari di identificare i loro contenuti nei video caricati dagli utenti YouTube e decidere le azioni da intraprendere (es. silenziare l'audio, bloccare la visualizzazione di un video, monetizzare il video sovrapponendovi della pubblicità). Ad esempio, se sono una famosa *rock band*, posso decidere o di fare ascoltare la mia musica solo nella mia pagina, oppure posso recuperare i soldi mettendo la mia pubblicità su altri video.



Cloud: G-Suite è un servizio basato sul *cloud*, che comprende posta, calendari, messaggistica, archiviazione ed elaborazione *cloud*, fogli di calcolo, presentazioni e siti collaborativi basati sul web. Esso quindi è composto da due filiere: Gmail, Gdox, calendar che sono forniti gratuitamente, oppure a pagamento, se si richiede una versione più potente. Per sviluppare una piattaforma *Cloud*, c'è bisogno di una grande infrastruttura di raccolta ed analisi dati. E poiché i costi per sviluppare tali architetture sono molto ingenti, molte imprese affittano spazi *cloud* a grandi *provider*, come Google e Amazon utilizzando le loro enormi capacità di calcolo.⁸⁸

Maps: Google Maps è un servizio accessibile dal relativo sito web che consente la ricerca e la visualizzazione di carte geografiche di buona parte della Terra. Oltre a questo è possibile ricercare servizi in particolari luoghi, tra cui ristoranti, monumenti e negozi. Inoltre si può visualizzare un possibile percorso stradale tra due punti, oltre a foto satellitari di molte zone con diversi gradi di dettaglio (per le zone che sono state coperte dal servizio si riescono a distinguere in molti casi le case, i giardini, le strade e così via). Le foto sono statiche (non in tempo reale), e una buona parte di loro risalgono alla fine degli anni novanta. Oltre a queste funzioni, Google Maps offre anche una ricerca di attività commerciali sulle stesse mappe. Google Maps è basato su una variante della Proiezione di Mercatore. Se la Terra fosse perfettamente sferica, la proiezione sarebbe la stessa di quella di Mercatore. Google Maps usa le formule della Proiezione sferica di Mercatore, ma le coordinate di Google Maps sono basate sui dati del sistema geodetico mondiale WGS84. La differenza tra una sfera ed un ellissoide WGS84 è la causa di una non precisa proiezione, percepibile in modo particolare su rappresentazioni a grande scala.

Infine, oltre ai modelli di business appena citati, sono presenti numerosi altri prodotti offerti da Google che vedono nei dati la maggiore fonte di input che permette la realizzazione dei suddetti servizi. Essi sono Google Chrome (il *browser*), Google Play (la piattaforma di distribuzione di app), Google Translate, Google Earth e le varie applicazioni nel campo dell'intelligenza artificiale (ad esempio Google Glass⁸⁹ e Google Car⁹⁰).

⁸⁸ Google Suite (formerly Google Apps for Work and Google Apps for Your Domain) is a brand of cloud computing, productivity and collaboration tools, software and products developed by Google, first launched on August 28, 2006 as "Google Apps for Your Domain". Google Suite comprises Gmail, Hangouts, Calendar, and Google+ for communication; Drive for storage; Docs, Sheets, Slides, Forms, and Sites for collaboration; and, depending on the plan, an Admin panel and Vault for managing users and the services. It also includes the digital interactive whiteboard Jamboard and the app development platform App Maker. Google site

⁸⁹ Google Glass is a brand of smart glasses – an optical head-mounted display designed in the shape of a pair of eyeglasses. It was developed by X (previously Google X) with the mission of producing a ubiquitous computer. Google Glass displayed information in a smartphone-like, hands-free format. Wearers communicated with the Internet via natural language voice commands. Google site

⁹⁰ The Google self-driving car project is now Waymo. Waymo began as the Google self-driving car project in 2009. Today, we're an independent self-driving technology company with a mission to make it safe and easy for everyone to get around—without the need for anyone in the driver's seat. Google site

CAPITOLO TERZO

BIG DATA E PRATICHE ABUSIVE

Introduzione

Come visto nel capitolo 1 le imprese lavorano al fine di ottenere i Big Data. Infatti essi, alla pari col capitale ed il lavoro sono input necessari alle piattaforme digitali per permanere nel tempo, grazie alla creazione di beni e servizi sempre più vicini a gusti dei consumatori che evolvono rapidamente. La sfida delle imprese sta nel governare la variabilità, oltre che la varietà dei dati⁹¹.

“firms that collect as many digital data as possible achieve a better understanding of the real world that gives them a competitive advantage over rivals that do not have access to the same big data”⁹²

Come detto, l'utilità di un dato non dipende dal dato in sé, ma dalle risorse materiali ed intellettuali di cui dispone la piattaforma per sviluppare una *data analytics* e una *data storage* efficienti.

Essendo un fattore produttivo, i dati hanno un mercato. Le imprese decidono se gestire internamente o esternalizzare le attività di raccolta ed analisi e diventano così proprietari di dati, decidendo se utilizzarli in tutto, o in parte destinarli al mercato. Nella realtà dei fatti vi sono solo degli specialisti, informatici ed ingegneri che si occupano della raccolta e/o analisi di dati da destinare alla vendita. Le grandi imprese digitali impiegano i dati raccolti ad analizzarli nei propri processi produttivi, sviluppando soluzioni tecnologicamente avanzate e costruendo un business trainato dalla *data driven economy* da un lato e dall'evolversi dei gusti, attitudini ed abitudini dei consumatori finali dall'altro. Gli enormi data set che vengono raccolti ogni giorno possono costituire una barriera all'ingresso per le nuove entranti nel settore, dato il loro elevato ammontare e la richiesta di macchinari, software e personale umano (e finanziario) sempre più all'avanguardia.

Le grandi imprese come Facebook, Microsoft, Yahoo!, Amazon, Uber, Google competono sui dati, detengono enormi architetture di raccolta ed analisi, nonché l'uso esclusivo di tali dati. Queste imprese, teoricamente, non hanno bisogno di acquistare dati, semmai possono decidere se venderli. Tuttavia, ogni impresa ha bisogno dei suoi dati e si dota delle migliori apparecchiature e *skills* per generare valore da questi.

Ma quanto vale un dato?

Esso, come detto, passa dai consumatori alla piattaforma, volontariamente ed ora anche consapevolmente (ricordiamo la normativa sulla *privacy* GDPR in vigore dal 25 Maggio 2018) senza corrispettivo in denaro, a fronte però di un servizio digitale sempre più efficace ed efficiente, che segue il trend del mercato ed i gusti del consumatore. Questo è un tipico esempio di *non transaction market*, ovvero un mercato dove circolano beni e servizi che vengono venduti gratuitamente, nel nostro caso sul web.

⁹¹ Cfr: Capitolo 1, Paragrafo 1.2

⁹² The Role of “Big Data” in Online Platform *Competition* – Andres V. Lerner (26 Agosto 2014)

Tuttavia, il dato un suo valore intrinseco lo ha eccome! Abbiamo visto (nel Capitolo 2) come grazie ai dati si creano nuovi mercati e nuove strategie competitive su cui le imprese digitali fanno leva per sfruttare appieno la tecnologia e le *skills* di cui dispongono. Bene, il *recommendation system* e l'*online advertising* sono le due fonti di *revenues* principali per la tipologia di aziende che abbiamo introdotto un capitolo fa.

Grazie ai dati che il consumatore gli fornisce, queste imprese sfruttano le conoscenze sugli utenti, che acquisiscono dai dati e le impiegano nei propri processi produttivi, che a loro volta genereranno altri dati con valore sempre maggiore e permettono la loro monetizzazione, come detto, o vendendo pubblicità customizzata, o sviluppando algoritmi sempre più efficienti che aumentano o bloccano l'uso che i consumatori fanno della piattaforma. Una sorta di effetto *lock in* per i consumatori, che però, consapevoli e vogliosi di utilizzare la piattaforma gratuita, sono ben lieti di accettare. In fondo "*Competition is one click away*", anche se le scelte di una mole spropositata di consumatori digitali hanno pesanti effetti sull'economia moderna, ormai invasa dal digitale. Dunque dobbiamo essere consapevoli che quel servizio sul web che a noi viene reso gratuitamente, in realtà viene scambiato alla pari con i dati che noi stiamo fornendo (e col tempo che noi impieghiamo sulla piattaforma), ma siamo certi che questo sia il suo reale valore finanziario? Un dato ha un suo costo, costituito da una componente infinitesimale di costi fissi e variabili per la sua raccolta ed elaborazione, ma se qualcuno volesse acquistare un dato si può dire che il suo valore finanziario o di mercato equivale alla sommatoria dei flussi di cassa infinitesimali che si suppone di fare tramite quel dato, in termini di vendita di contenuti pubblicitari o, in maniera ancor più indiretta, tramite l'aumento del tempo speso dagli utenti sulla piattaforma, sempre più ad hoc per loro. Infatti, con l'aumento del numero di utenti e delle loro interazioni con la piattaforma, ne aumentano conseguentemente i dati che la piattaforma immagazzina. Dunque aumentano il volume, la varietà e la variabilità dei dati, che causano una maggiore complessità da gestire. Come detto, la sfida principale delle *digital companies* risiede nella gestione di tale complessità. Con una gestione dei dati efficace ed efficiente, infatti, la piattaforma inizia ad avere sempre maggiore appetibilità verso ogni tipologia di *stakeholder*, andandone ad aumentare in maniera esponenziale il valore e portando così alla costituzione di *dominant online platforms*.

Dunque un dato, o meglio un dataset costituito da una certa varietà, volume e variabilità, ipoteticamente ha un prezzo di mercato, sta alle imprese decidere dove indirizzarlo, se nei propri processi produttivi o destinarlo alla vendita.

In questa sede non si discuterà sull'equità che c'è tra scambio di dati e accesso a piattaforme gratuite e su come viene allocato il differenziale; si vuole mettere alla luce quelli che sono gli impatti nel settore ed in particolare il rapporto competitivo tra le imprese nel settore, tra nuovi entranti e incumbent, all'interno del sistema o tra sistemi (*competition intrasystem* o *intersystem*), in situazioni di competizione dinamica, data dalla rapidità del ciclo di vita di tali settori e dalla ricerca di innovazioni *disruptive* e che porta, molto spesso, alla creazione di monopoli o quasi monopoli (*winner takes all* e *competition for the market*).

Verranno dunque analizzati i più svariati settori digitali in chiave competitiva, partendo dall'idea di base che la *data storage* e la *data analytics*, oltre ad essere le *core activities* per generare valore attraverso i dati, costituiscono barriere all'ingresso per molte imprese che necessitano di determinati dati, con determinate caratteristiche, da inserire nei propri processi produttivi, ma che non dispongono delle risorse finanziarie e intellettuali per intraprendere un'efficace attività di raccolta, analisi ed elaborazione di grossi volumi di dati. Quindi, come già accennato più volte, i settori così descritti sono costituiti da poche grosse imprese incumbent, altre imprese che entrano ed escono rapidamente dal settore, con un tasso di *turn over* molto elevato e altre piccole imprese che non riescono ad entrare. Tutto ciò porta al rafforzamento di posizioni dominanti e a condizioni competitive che possono ostacolare il libero funzionamento del mercato.

Ci si interrogherà dunque se questa struttura dell'industria porta le imprese, con un elevato potere di mercato, ad adottare pratiche abusive creando così situazioni di concorrenza sleale. Vedremo dapprima brevemente come i big data si legano al concetto di abuso di posizione dominante e pratiche abusive (Paragrafo 3.1), poi andremo a stabilire se i dati sono da considerarsi una risorsa critica e quindi essenziale per lo sviluppo dell'industria e per la libera concorrenza, nell'ambito della *Essential facilities doctrine* (Paragrafo 3.2), si vedrà la funzionalità dei dati nell'ambito della *discrimination price* e *algorithmic collusion* (Paragrafi 3.3 e 3.5) e le clausole contrattuali anticompetitive sullo scambio dei dati (Paragrafo 3.4). Queste pratiche rientrano tutte negli Artt 101 e 102 del TFEU e sono potenzialmente condannabili per il regolamento Europeo e Statunitense.

3.1 Big Data, potere di mercato e pratiche abusive

Le autorità garanti della concorrenza sono molteplici nel mondo ed ognuna interviene nel proprio territorio giurisdizionale, anche nel campo digitale. L'obiettivo generale è quello di assicurare che la competizione, nei vari settori dell'economia non venga ostacolata o manipolata in modo da avere conseguenze negative di riflesso verso tutta la società. L'assunzione di fondo è che una forte competizione promuove innovazione, produttività e benessere sociale per i consumatori, attraverso bassi prezzi, migliori prodotti e maggiore varietà di beni/servizi. Mercati aperti e competitivi incentivano infatti lo sviluppo economico. Permangono tuttavia dubbi sulla relazione tra la realizzazione di una *competition policy* efficace e lo sviluppo economico. Infatti, alcuni studi empirici hanno confermato tale relazione, altri, invece, affermano che l'intervento Antitrust può danneggiare la crescita della produttività (Young e Shugart, 2007) e del tasso di innovazione.⁹³ Infatti, ci sono spesso casi in cui le autorità di *competition policy* devono affrontare dei trade-off tra gli obiettivi di benessere sociale che si intendono perseguire. La *competition law* è lo strumento di cui si servono le autorità competenti per promuovere la competizione e il benessere sociale. Essa si compone di tre pratiche principali: contratti restrittivi (Sezione 1 dello Sherman Act; Articolo 101 TFEU; Articoli 2 e 4 della Legge No.

⁹³ Fonte: Chapter 4. *Competition Policy* - Young e Shugart (2007)

287/1990); condotte unilaterali anticompetitive (Sezione 2 dello Sherman Act) e abuso di posizione dominante (Articolo 102 TFEU; Articolo 3 della Legge No. 287/1990); concentrazioni (Sezione 7 del Clayton Act; Regulation No. 139/2004; Articolo 6 della legge 287/1990).⁹⁴

Vengono pertanto perseguiti obiettivi di natura economica e non. Infatti le autorità, a seconda del tipo di settore e delle inefficienze che sono state messe in risalto, perseguono obiettivi di efficienza allocativa, efficienza produttiva, efficienza dinamica, benessere del consumatore (in termini di surplus), e surplus totale (*total welfare*). L'efficienza allocativa si ha quando le risorse a disposizione della società (fattori produttivi, beni, servizi) vengono impiegate in modo da soddisfare i bisogni della società (imprese e consumatori). Questo ottimo paretiano si ottiene, teoricamente, quando il prezzo equivale al costo marginale, in una condizione di concorrenza perfetta; ovvero l'impresa produce ciò che il cliente vuole (ne soddisfa tutta la curva di domanda), al prezzo per cui è disposto a pagare. L'efficienza produttiva viene raggiunta quando l'output è massimizzato, usando la combinazione di risorse a disposizione in maniera efficiente, evitando sprechi e tempi morti nella produzione. L'efficienza dinamica si ha quando c'è una rapida diffusione di nuove tecnologie, nuovi beni e nuovi servizi da destinare alla società e che ne incrementano il benessere. *Consumer welfare* e *total welfare*, infine, sono due concetti microeconomici e si riferiscono alla massimizzazione della ricchezza disponibile di consumatori e produttori e alla minimizzazione delle inefficienze allocative (*deadweight loss*)⁹⁵. Sono possibili molteplici conflitti tra i diversi concetti di efficienza appena descritti e, a seconda del caso pratico e dell'autorità competente, prevarranno gli uni su gli altri.

Tali *trade-off* emergono soprattutto nei settori digitali, caratterizzati da un ciclo di vita molto rapido e da numerose *breakthrough innovations*⁹⁶ da un lato e creazione di monopoli o quasi monopoli dall'altro. Ci sono diverse scuole di pensiero su quali efficienze far prevalere, ma l'efficienza dinamica è sicuramente cruciale per la crescita economica dei settori digitali. L'innovazione, d'altro canto, è fondamentale per aumentare il benessere sociale (consumatore e produttore) ed è di dubbia interpretazione il rapporto che esiste fra innovazione e struttura del settore. Ci sono scuole di pensiero (Schumpeter) che prediligono un rapporto di tipo diretto, ovvero, maggiormente concentrata è la struttura del settore, più alto sarà il tasso di innovazione (e gli sforzi in Ricerca & Sviluppo), focalizzando l'attenzione sulla maggiore disponibilità di risorse critiche di una grande

⁹⁴ Lo Sherman Act disciplina le condotte anticompetitive secondo la giurisdizione USA, gli articoli 101 e 102 del TFEU quelle secondo il regolamento Comunitario; la legge no. 287/1990 disciplina le intese, l'abuso di posizione dominante e le concentrazioni secondo la normativa Italiana; la Regulation No. 139/2004 disciplina le concentrazioni secondo la normativa Europea e il Clayton Act disciplina pratiche anticompetitive come la *price discrimination*, *unfair pricing* e pratiche scorrette secondo la normativa USA.

⁹⁵ Deadweight loss is the fall in total surplus that results from a market distortion, such as a tax. In economics, a deadweight loss (also known as excess burden or allocative inefficiency) is a loss of economic efficiency that can occur when equilibrium for a good or service is not achieved or is not achievable.

⁹⁶ Cfr: innovazione *disruptive*. È un'innovazione che cambia radicalmente lo scenario competitivo e la struttura di un'industria. Non c'è perfetta correlazione tra investimenti in R&D e innovazione *disruptive*; quindi sia imprese grandi che imprese piccole possono trovare l'innovazione vincente, cambiando le dinamiche competitive.

impresa rispetto ad una piccola. Altre scuole di pensiero (Arrow) prediligono un rapporto di tipo inverso, in cui una struttura di settore poco concentrata e quindi fortemente competitiva, porta an tasso innovativo più alto, grazie alla maggiore propensione allo sviluppo di nuove innovazioni, di prodotto e di processo, al fine di vincere la competizione. Tuttavia, l'esperienza empirica ci insegna che sia situazioni di monopolio che di concorrenza hanno portato alla nascita di numerose innovazioni; altri sostengono che è l'oligopolio la struttura dove vi è un maggior incentivo ad innovare. In sostanza non ci sono conclusioni teoriche sul tipo di relazione che esiste tra innovazione e potere di mercato; secondo Scherer, infine, tale relazione può essere sintetizzata da una curva a forma di "u rovesciata" (*inverted U-shaped pattern*), dove ci sarà bassa innovazione ad alti livelli di competizione o a situazioni di monopolio e elevata innovazione in situazioni intermedie di oligopolio.⁹⁷

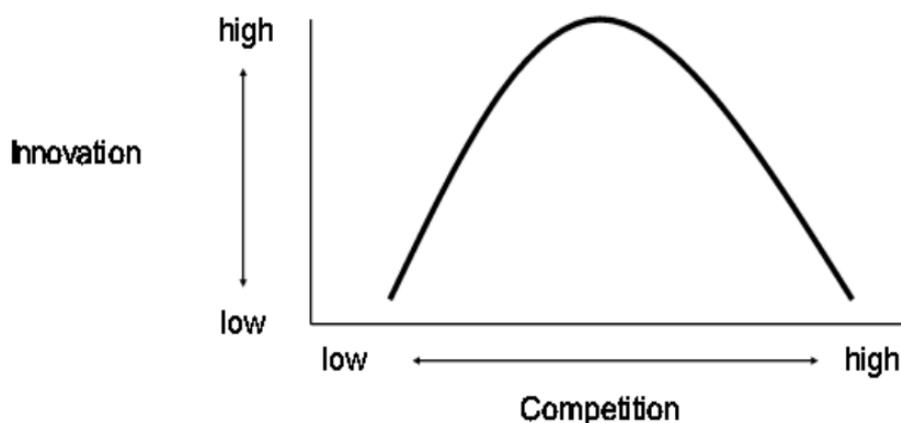


Figura 3.1

Nell'industria *high tech*, dunque, salvaguardare tale tipo di rapporto è fondamentale per assicurare un forte tasso di innovazione nel tempo. Infatti, così come i beni/servizi digitali si rinnovano di continuo, portando cambiamenti dei business model delle imprese esistenti ed alla creazione di nuove tecnologie nel campo digitale, alcune imprese resistono e si rinnovano, altre escono dal mercato e vengono rimpiazzate da nuove entranti che hanno un'innovazione in mano che a sua volta le porterà ad essere grandi incumbent nel futuro o ad uscire immediatamente dal settore se hanno una tecnologia o un'idea di business non allineata ai trend del mercato.

Ogni settore, tuttavia, ha un suo ciclo di vita più o meno lungo, basti pensare a quello dei motori di ricerca, in cui si sono alternati player come Google, Altavista⁹⁸, Bing o quello dei social network, dove c'è stato il passaggio di consegne da MySpace a Facebook.

⁹⁷ L.M. Cabral, *Economia Industriale*, Carocci editore

⁹⁸ AltaVista was a Web search engine established in 1995. It became one of the most-used early search engines, but lost ground to Google and was purchased by Yahoo! in 2003, which retained the brand, but based all AltaVista searches on its own search engine. On July 8, 2013, the service was shut down by Yahoo! and since then, the domain has redirected to Yahoo!'s own search site. Wikipedia

Alcuni *player* (vedi Google) sono riusciti a resistere agli “attacchi” dei competitor e ad innovarsi di continuo, inventando nuovi modelli di business e scoprendo nuovi settori digitali; altri, invece, non sono stati lungimiranti nel capire l’evoluzione del settore continuando ad investire nella tecnologia obsoleta e o sono usciti dal mercato (Altavista, MySpace), o rimangono con quote di mercato piccolissime, collocandosi come imprese di nicchia. Nella realtà dei fatti è evidente di come si siano create posizioni sempre più dominanti, con imprese che hanno diversificato le proprie attività, specializzandosi in uno o più settori, e diventandone quasi monopolisti. Questi sono i casi di Amazon, nell’*e-commerce* e nel *cloud-computing*, Netflix, nello *streaming on demand*, Spotify, nello *streaming* musicale, Google, nei motori di ricerca e non solo, Facebook, assieme ad Instagram e WhatsApp nei social network e nelle app di messaggistica ecc. Molte altre entità, invece, hanno inventato modelli di business di difficile analisi e hanno portato alla creazione dei più disparati mercati rilevanti e sottosectori estremamente frammentari (è il caso ad esempio di Snapchat, il cui business model è di difficile comprensione)⁹⁹. Esistono infine tanti altri player più piccoli, ma a loro è destinata solo una piccola quota del mercato e, molto spesso, quando questi si rivelano vincenti ed iniziano il loro processo di crescita, vengono acquistati da entità più grandi, ma questo lo si vedrà nel Capitolo 4. In questo capitolo ci si vuole interrogare se l’accesso ai big data e la conseguente posizione dominante assunta nel settore, porti le imprese ad adottare atteggiamenti di abuso di tale posizione dominante o di condotte anticompetitive. Ma, andando più nello specifico, si vuol capire se il possesso e lo sfruttamento tramite *analytics* di grossi ammontari di dati, faciliti la creazione di posizioni dominanti con le conseguenti pratiche anticompetitive che ne derivano (Paragrafi 3.2 3.3 3.4), oppure facilitino la creazione di accordi e pratiche concordate che possano restringere e falsare il gioco della concorrenza all’interno del mercato (Paragrafo 3.5) secondo quelle che sono le disposizioni contenute nell’Art 101 e 102 del TFEU¹⁰⁰. Dunque, il prerequisito per infrangere l’Art 102 del TFEU è quello di detenere una posizione dominante; ed abbiamo visto come il possesso dei Big Data rafforzi tale possibilità, per via del *Network effect (direct e indirect)* e delle barriere all’ingresso, rappresentate dalle architetture di raccolta, analisi ed elaborazione di grossi volumi di dati. Tuttavia, ciò che è vietato dalla *Competition law*, non è la mera detenzione di una posizione dominante all’interno del settore, ma l’abuso di tale posizione. L’Art 102 del TFEU e l’Articolo 3 della Legge 287/90 fanno un elenco¹⁰¹ di pratiche che di per sé possono essere abusive.

⁹⁹ Snapchat è persino quotato in borsa, a New York (NYSE)

¹⁰⁰ TFEU: *Treaty on the Functioning of the European Union*

¹⁰¹ Pratiche abusive: imporre direttamente o indirettamente prezzi d’acquisto, di vendita od altre condizioni di transazione non eque; limitare la produzione, gli sbocchi o lo sviluppo tecnico, a danno dei consumatori; applicare nei rapporti commerciali con gli altri contraenti condizioni dissimili per prestazioni equivalenti, determinando così per questi ultimi uno svantaggio per la concorrenza; subordinare la conclusione di contratti all’accettazione da parte degli altri contraenti di prestazioni supplementari, che, per loro natura o secondo gli usi commerciali, non abbiano alcun nesso con l’oggetto dei contratti stessi. Art 102 TFEU

Le imprese imputate possono difendersi dalle accuse di abuso di posizione dominante affermando di promuovere l'innovazione e di fornire prodotti sempre più all'avanguardia, favorendo l'efficienza dinamica, allocativa e produttiva, nonché il *total welfare*.

Il concetto economico di potere di mercato (PDM) riflette l'abilità dell'impresa di fissare prezzi superiori al costo marginale; infatti solo così vi è la creazione di extraprofitto per l'imprenditore. Ma un'impresa è dominante quando c'è un elevato grado di PDM, misurato tramite l'indice di Lerner ($L = (P - MC) / P$)¹⁰². Tuttavia, nello scenario competitivo che stiamo studiando, le *multisided platforms* hanno costi marginali quasi trascurabili e pongono diversi prezzi, ciascuno per ogni cliente della *platform*, così che si usa una media dei prezzi. Anche un prezzo leggermente positivo, dunque, può rilevarsi eccessivo in relazione alla struttura dei costi dell'entità. Infatti, sfruttando le economie di scala e il *network effect*, l'impresa digitale recupera i costi fissi facendo notevoli extraprofitto. Dunque l'indice di Lerner non è molto utile nella definizione di *market power* per le *digital companies*, si fa allora riferimento a metriche più standardizzate, come la *market share* o quota di mercato (QDM)

*"position of economic strength enjoyed by an undertaking which enables it to prevent effective competition being maintained on the relevant market by giving it the power to behave to an appreciable extent independently of its competitors, customers and ultimately of its consumers"*¹⁰³

Dunque, per avere una posizione dominante, non costituisce elemento necessario operare in condizioni di assenza di competizione (situazione di monopolio), ma l'entità deve poter comportarsi in maniera del tutto indipendente dalle strategie dei competitor, è ovvero leader del settore. Infatti, anche imprese che non sono in condizioni di monopolio, ed hanno quindi una QDM non esageratamente elevata (anche il 40% può essere considerata una soglia soddisfacente) possono agire in maniera indipendente. Importante è anche vedere quanto la quota di *market share* sia stabile nel tempo, poiché abbiamo visto che nell'industria *high tech* alto è il grado di dinamicità e di creazione di innovazioni *disruptive* che cambiano gli scenari competitivi.

I dati, come già detto, rappresentano barriere all'ingresso per la maggior parte di nuove entranti nel settore. Essi infatti rafforzano sia il *direct* che l'*indirect Network Effect* e permettono di sfruttare le economie di scala e di ammortizzare i costi fissi. Un'altra barriera all'ingresso può essere rappresentata dall'effetto *lock-in*, che intrappola gli utenti all'uso di una piattaforma, continuando a fornirgli dati e tempo.

Infatti i dati portano a soluzioni più personalizzate per i consumatori ed è probabile che questi spendano maggiore tempo su una piattaforma con contenuti proliferati, o con pubblicità ad hoc o che faccia vedere contenuti correlati ai loro gusti e preferenze.

¹⁰² L'indice di Lerner è uguale al margine di profitto per ogni unità venduta.

¹⁰³ European Court of Justice (ECJ), *United Brands*, 1978.

Inoltre c'è il discorso degli *switching costs*¹⁰⁴ per gli users che, nel cambiare piattaforma, devono trasferire dati ed informazioni personali. Tuttavia questo non è un fattore competitivo rilevante, poiché, come detto, nel digitale, "*competition is one click away*". Nel paragrafo successivo si vorrà capire se il controllo esclusivo di un determinato dataset, con un suo volume, varietà e variabilità, possa escludere all'uso di tale dataset altri *players* del settore. Altro discorso è la definizione di un mercato rilevante, poiché, per determinare se un'impresa è dominante oppure no, bisogna capire dove si colloca e quali sono i *competitors* che offrono beni e servizi sostituti. Per molte imprese digitali, definire un mercato rilevante non è cosa semplice. Riprendendo l'esempio di Snapchat, dove non è ben chiara la tipologia di business model, è veramente difficile trovarne una collocazione nel macrosettore dei social network. Ma, prendendo un esempio più semplice, Facebook, per definire se è o meno dominante è da capire prima in relazione a quale mercato o sottosettore ci si riferisce. Infatti quello dei Social Network è un macrosettore estremamente frammentato al suo interno e ci sono numerosi altri *player* oltre a Facebook. C'è probabilmente stato un periodo in cui Facebook era dominante, ma questo perché il settore era più uniforme ed ammetteva meno sostituti. Successivamente sono cresciuti gli utenti di social, ne sono cresciuti i contenuti, i post, le foto, le interazioni; e così sono nate numerose altre app (Twitter, Instagram, Snapchat, LinkedIn, Google+, Tumblr) ognuna con uno scopo sociale leggermente diverso dall'altra. Sono poi le scelte ed i trend che seguono i consumatori a fare la differenza, nonché la tipologia di consumatori, la loro collocazione geografica, usi, costumi, abitudini ed accessibilità alla rete Internet. Quello dei Social Network può essere un esempio guida di come evolvono i settori digitali. Infatti, se dapprima c'è la nascita del settore per mezzo di innovazioni *disruptive*, è il caso di MySpace per tale settore, poi c'è la sfida per la creazione di un design dominante, o meglio di un prodotto *user friendly*, che rispecchi le esigenze dei consumatori, successivamente chi crea un design dominante diventa monopolista, o quasi; ed infine, quando l'innovazione viene divulgata e vengono costruite le basi di dati per sviluppare soluzioni proliferate verso gli utenti, nascono nuovi *players*, ciascuno con un'idea di business più o meno distinta dall'altro, che iniziano a guadagnare QDM, a discapito di quello che era stato leader dominante fino a poco fa. A seconda del tipo di settore che si analizza, Social Network, streaming di film, streaming musicale, e-commerce, motori di ricerca e tutto il mondo delle App, si avrà un ciclo di vita più o meno lungo, settori dominati da poche grandi o tante piccole imprese, leader, follower e strutture industriali estremamente differenziate, per tipologie di mercati che servono, al loro interno. Tornando al concetto di dominanza, detenere una posizione dominante nel settore non è di per sé illegale, ma sono illegali alcune pratiche, che apparentemente sembrerebbero legali, se condotte da imprese dominanti. In sostanza, le imprese dominanti hanno una sorta di responsabilità di non ridurre o addirittura eliminare il grado di competizione esistente nel mercato.

¹⁰⁴ Switching costs are the costs that a consumer incurs as a result of changing brands, suppliers or products.

*"conduct that, through methods different from normal competition "on the merits", prevents the maintenance or development of the degree of competition still existing in the market"*¹⁰⁵

Una condotta, infine, può essere considerata anticompetitiva anche senza intento anticompetitivo dell'impresa in questione.

L'intervento delle autorità Antitrust deve essere fatto, tenendo in considerazione che, trattandosi di settori con un'evoluzione estremamente dinamica, molte decisioni vanno prese con cautela, cercando di assecondare di volta in volta il rapporto tra innovazione e competizione e decidendo quale obiettivo economico far prevalere. Quindi, sebbene sia vero che i settori digitali, in cui vi è lo scambio di dati e servizi tra utenti e piattaforme, siano costituiti dalla presenza di grandi imprese dominanti, è pur vero che questa situazione di dominanza non è affatto garantita nel futuro ed è giustificata dalle elevate barriere all'ingresso, alimentate dal *Network Effect* e dallo sfruttamento di economie di scala. Tornando all'esempio dei social network, non è chiaro quale sarà il futuro di Facebook, in molti paesi esso sta perdendo quote di mercato e, soprattutto ore di tempo passate dagli *users* a discapito di altre app, come Instagram che stanno vedendo incrementare la propria *market share*. In questo senso, gioca un ruolo fondamentale la tipologia di autorità competente e la regione geografica dove l'entità opera. Infatti, se dal lato Europeo vi è una maggiore attenzione focalizzata sulla massimizzazione dell'efficienza allocativa e del consumer welfare, dal lato USA vi è maggiore concentrazione verso l'efficienza dinamica e il *total welfare*. Uno dei dibattiti più controversi, ad esempio, è quello sul sistema degli *intellectual property rights* (IPRs) e se esso debba prevalere o no a discapito di un intervento Antitrust. Gli IPR (costituiti da *Patents, Copyrights, Trade secrets, Trademarks*) costituiscono una forma di protezione dell'informazione privata. Quest'ultima, infatti, essendo un bene pubblico (non rivale e non escludibile), se condivisa e divulgata, necessita di protezione, poiché, avendo costi di duplicazione e distribuzione praticamente nulli, può causare una riduzione dell'incentivo ad innovare (tutti penserebbero che sarebbe più facile e meno costoso copiare idee altrui piuttosto che investire e crearne di nuove). Quindi gli IPRs servono a garantire un certo grado di incentivo all'innovazione al fine di assecondare l'efficienza dinamica e l'evoluzione tecnologica. D'altro canto, gli IPRs causano aumento del PDM, con la conseguenza che i consumatori pagano di più quantità che potrebbero essere vendute per meno. Ciò dunque causa una riduzione del surplus del consumatore.

¹⁰⁵ ECJ, Hoffman La Roche, 1979; Michelin, 1985

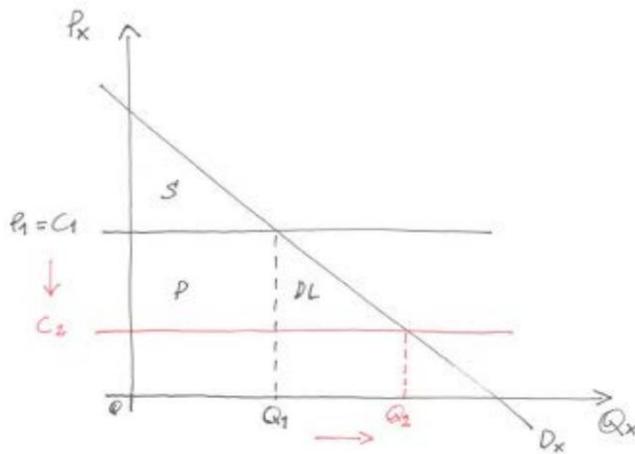


Figura 3.2

Dal grafico si evince che, grazie all'innovazione, la curva dei costi marginali si riduce verso C_2 . In condizioni di concorrenza perfetta, l'impresa produrrebbe la quantità Q_2 aumentando così il surplus del consumatore che passa da una situazione iniziale (S) ad una situazione ($S+P+DL$). Tuttavia il sistema degli IPRs permette all'impresa di vendere sempre una quantità Q_1 ad un prezzo P_1 , ma ottenendo extraprofiti (l'area P). quindi il consumatore paga di più di quello che dovrebbe, il surplus non aumenta (resta costante) ed in più c'è la *deadweight loss* (DL). Quindi, se l'obiettivo delle autorità è quello di garantire efficienza dinamica, si dovrebbe assecondare il sistema degli IPRs, se è quello di garantire l'efficienza allocativa si dovrebbe vietare, poiché il produttore vende una quantità minore a un prezzo più alto di quello che potrebbe e dovrebbe praticare, ed infine, se l'obiettivo è preservare il surplus del consumatore si rimane indifferenti nella scelta, poiché esso non varia. Colui che ci guadagna da tale sistema è sicuramente il produttore, che passa da una situazione senza extraprofiti, ad una situazione con profitto pari all'area P , infatti il surplus totale aumenta di P rispetto alla situazione di partenza (DL rappresenta l'inefficienza allocativa). L'obiettivo del sistema degli IPRs è dunque quello di massimizzare il differenziale tra benefici e costi di protezione e, a seconda della tipologia di informazione protetta, si hanno una diversa durata della protezione e un diverso scopo, inteso come quanto simile può essere un'innovazione senza infrangere gli IPRs di un altro. Più sarà elevata la durata e più sarà ampio lo scopo di copertura degli IPRs, più le imprese avranno incentivo ad investire ed innovare.

Gli IPRs, dunque, sono uno strumento che potrebbe portare a situazioni di monopolio e l'utilizzo esclusivo dei dati, protetti dagli IPRs, rafforza tale posizione. Quindi, chiunque può raccogliere i propri dati, ma nessuno può utilizzare quelli degli altri infrangendo il sistema degli IPRs. Quindi non c'è divulgazione di dati poiché, come vedremo nel Paragrafo successivo, ogni impresa ha bisogno dei suoi dati, non di quelli degli altri.

Dunque queste situazioni di monopolio, come detto, rischiano di portare a situazioni di abuso di posizione dominante. Soffermandoci sugli Artt 101 e 102 del TFEU, per pratiche abusive si intendono le seguenti:

- imporre direttamente o indirettamente prezzi d'acquisto, di vendita od altre condizioni di transazione non eque;
- limitare la produzione, gli sbocchi o lo sviluppo tecnico, a danno dei consumatori;
- applicare nei rapporti commerciali con gli altri contraenti condizioni dissimili per prestazioni equivalenti, determinando così per questi ultimi uno svantaggio per la concorrenza;
- subordinare la conclusione di contratti all'accettazione da parte degli altri contraenti di prestazioni supplementari, che, per loro natura o secondo gli usi commerciali, non abbiano alcun nesso con l'oggetto dei contratti stessi.

Quindi si evince che le pratiche abusive che mettono in atto tali imprese dominanti, riguardano sia il prezzo e le altre condizioni di vendita imposti in modo del tutto non equo (*exploitative abuses*), sia la manipolazione e l'alterazione della struttura dell'industria a discapito della concorrenza (*exclusionary abuses*)¹⁰⁶.

Per quanto riguarda le prime pratiche abusive (*exploitative abuses*), esse sono inerenti alla fissazione di prezzi eccessivamente alti in relazione al valore del servizio fornito, prezzi differenziati a seconda dei clienti serviti, oppure all'imposizione di clausole anticompetitive all'interno di contratti commerciali. Per quanto concerne la prima pratica (*excessive price*) c'è da dire che è difficile capire quando un prezzo è equo oppure no, bisognerebbe infatti conoscere perfettamente la struttura dei costi dell'impresa e relazionare il prezzo col costo marginale o col costo medio. Abbiamo tuttavia visto come la struttura dei costi delle *digital companies* sia diversa dal tradizionale, con elevati costi fissi e costi marginali pressoché nulli. Oppure si può comparare il prezzo praticato col valore economico del bene/servizio reso anche in relazione ad altri beni/servizi comparabili. Tuttavia sono gli stessi IPRs che, come visto, incentivano l'impresa a porre prezzi elevati e a vendere minori quantità e ne giustificano la causa (garantire e proteggere l'innovazione). Quindi, tornando all'esempio precedente, per verificare se Facebook pone condizioni di vendita non eque, o si fa un'analisi dei costi di Facebook, paragonando il costo medio o marginale alla media dei prezzi pattuiti da Facebook ai vari clienti di ogni side che fronteggiano (ricordando che $P=0$ nel lato degli utenti consumatori finali), oppure si vanno a vedere gli altri player del settore (operanti nel settore della pubblicità online) e si compara con i prezzi pattuiti da tali entità. Tuttavia si torna sempre al concetto di definizione di mercato rilevante, ed abbiamo già discusso di quanto ne sia difficile l'identificazione per molti settori digitali.

Per le autorità antitrust Statunitensi, la pratica dell'*excessive price* è anticompetitiva solo se vi è un evidente danneggiamento al cliente o consumatore finale.

¹⁰⁶ Per *exploitative abuse* si intende l'abuso di PDM, ad esempio ponendo prezzi troppo elevati in relazione al bene/servizio offerto, oppure non garantendo l'accesso ad una risorsa essenziale. Per *exclusionary abuse* si intende l'intento di escludere uno o più *player* dal settore, da parte di una o più entità, a seconda se si tratti di pratiche unilaterali o pratiche concordate.

Infatti, per la disciplina concorrenziale USA, il mero possesso di un elevato PDM giustificato da una posizione monopolistica nel mercato e conseguentemente la fissazione di prezzi “troppo” elevati, non è da considerarsi illecito, ma anzi, è un importante elemento per il libero funzionamento del mercato. Come detto, infatti, sono gli extraprofitti da monopolio che attraggono nuove potenziali entranti e incentivano gli investimenti in R&D per effettuare nuove innovazioni.

Secondo le autorità antitrust UE, invece, un prezzo non equo costituisce un abuso di posizione dominante e rientra pertanto nelle pratiche elencate dall’Art 102. Il prezzo deve essere equo in relazione al valore economico del servizio fornito, ma abbiamo visto come questo può essere difficile da definire per le *multisided platforms*. Infatti, secondo la normativa comunitaria antitrust, l’efficienza allocativa e il surplus del consumatore sono gli obiettivi verso cui giungere, piuttosto che l’efficienza dinamica, prevalente nella normativa USA.

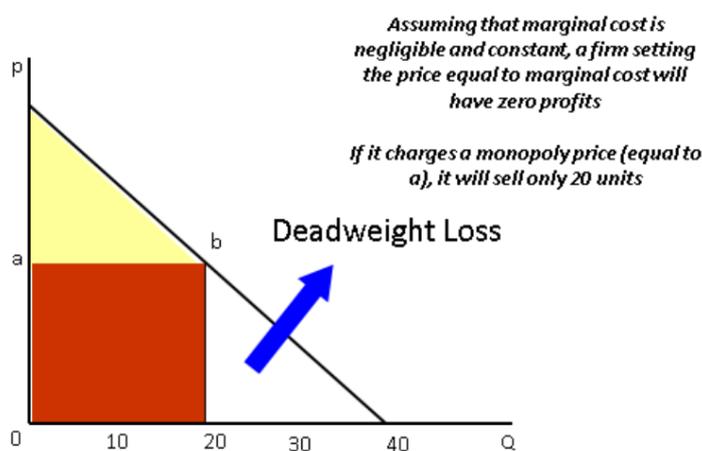


Figura 3.3¹⁰⁷

Vedremo, nel Paragrafo 3.3, come un modo per massimizzare il ROI della raccolta ed analisi dei dati, sia quello di utilizzarli col fine di stimare la *willingness to pay* di ogni *cluster* di consumatore e conseguentemente porre per ciascun *cluster* un prezzo diverso. Si studierà dunque se la *price discrimination* è una strategia notevolmente condizionata ed aggravata dal possesso dei Big Data.

Le altre tre pratiche anticompetitive enunciate dall’Art 102 del TFEU sono dette di *exclusionary abuses*. Infatti esse si riferiscono a quel “dovere” che hanno le imprese dominanti di mantenere inalterata la struttura competitiva del mercato, senza ostacolare i *competitors* attuali e futuri. Non devono in sostanza ridurre (volontariamente e non) il grado di competizione che c’è nel settore attraverso pratiche concordate (Paragrafo 3.5). *“The challenge for an antitrust court lies in stating a general rule for distinguishing between exclusionary acts, which reduce social welfare, and competitive acts, which increase it” (D.C. Circuit, US v. Microsoft, 253 F.3d 34, 2001)*

¹⁰⁷ L.M. Cabral, *Economia Industriale*, Carocci editore

Dunque c'è una differenza sostanziale tra pratiche competitive lecite e pratiche competitive illecite e, se l'impresa in questione ha nel mercato una posizione dominante, tale scrematura si fa più sottile, poiché, come abbiamo visto, il possesso di un PDM elevato costituisce il presupposto per salvaguardare la struttura competitiva da parte dell'impresa dominante. Tuttavia questo sarebbe un modello ideale, nella realtà dei fatti, caratterizzata dalla presenza di conflitti di interesse e giochi di potere, le *big digital companies* adottano atteggiamenti molto spesso anticompetitivi, falsando la competizione nei settori attraverso condotte unilaterali (Art 102 TFEU, Paragrafi 3.2, 3.3, 3.4) o attraverso pratiche concordate con altri player dell'industria (Art 101 TFEU, Paragrafo 3.5).

Sul concetto di condotta unilaterale scorretta (*exclusionary* e *exploitative conduct*), c'è un ampio dibattito. Nel 2009, la Commissione Europea ha identificato due requisiti per verificare se una condotta unilaterale è scorretta oppure no. C'è bisogno sia di un evidente impedimento, per i *competitors* attuali e/o potenziali, ad esercitare il proprio modello di business e a collocare i propri prodotti sul mercato; sia di un evidente danno ai consumatori, in una logica di riduzione del *consumer welfare* e di extraprofiti eccessivamente elevati per le imprese. Bisogna inoltre tenere ben presente alcuni parametri:

- Posizione nel mercato dell'impresa dominante;
- Presenza o meno di barriere all'ingresso e Network Effect;
- Posizione dei competitors;
- *Buyer power* (inteso come importanza strategica dei clienti per l'impresa);
- Evidenza di un evidente impedimento per i *competitors* attuali e/o potenziali;
- Evidenza di una *exclusionary strategy*;
- Estensione della condotta (in termini di mercati e durata).

Passiamo ora ad elencare le pratiche principali di condotte anticompetitive, sia di *exploitative* che di *exclusionary abuses*, che vengono inficcate dalla raccolta e dal possesso di dati ed analytics.

Nel prossimo Paragrafo (3.2) si discuterà se la mera detenzione di dati possa causare il cosiddetto "*Refuse to deal*", ovvero il rifiuto da parte dell'impresa avente una risorsa critica a garantire l'accesso a tale risorsa alle altre imprese, includendo dunque il concetto di *essential facility* e riprendendo quanto visto prima sugli IPRs. Nel paragrafo 3.3 ci sarà un dibattito sul rapporto tra Big Data e *Price discrimination* e di come tale pratica viene adottata al fine di massimizzare il profitto del produttore a discapito del benessere del consumatore. Poi si discuterà se lo scambio di dati e le inefficienze della normativa sulla *privacy* possano indurre all'imposizione di condizioni di scambio non eque per i consumatori (Paragrafo 3.4). infine, si discuterà se le moderne tecnologie di *machine learning* e *deep learning*, rafforzando ed automatizzando le analisi di mercato, possano portare ad equilibri collusivi di medio lungo periodo (Paragrafo 3.5)

3.2 Big Data & Essential Facilities Doctrine (EFD)

Come abbiamo già detto nel Capitolo 2, la raccolta e l'analisi dei Big Data portano le imprese a fornire ai propri consumatori migliori prodotti o servizi e prezzi bassissimi, nella maggior parte dei casi gratuiti. Nel paragrafo precedente abbiamo visto che, tramite il *Network effect* e le economie di scala, le imprese che detengono ed analizzano grossi ammontari di dati possono diventare dominanti. Infatti più utenti apportano più dati alla piattaforma, i quali permettono di ammortizzarne i costi della raccolta e dell'analisi ed inoltre inducono le imprese a migliorare la qualità dei propri servizi offerti e, conseguentemente, ad attrarre più consumatori. Questo è un circolo virtuoso per l'azienda, il cui valore, in questo modo, cresce esponenzialmente.

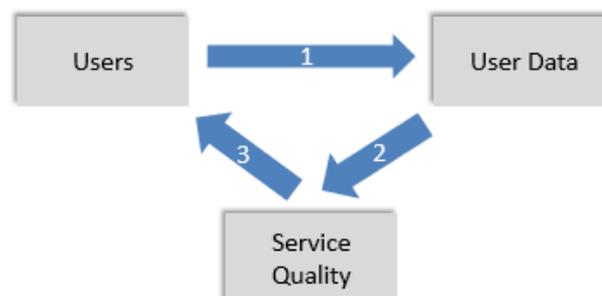


Figura 3.4¹⁰⁸

Secondo questo meccanismo, dunque, le piccole imprese e i nuovi entranti non possono effettivamente competere con tali grandi entità. Infatti esse non dispongono di una elevata base di dati di partenza, dovendo così ricorrere al mercato per acquisire le informazioni che gli servono. Tuttavia nella *data driven economy*, una volta ottenute le infrastrutture di raccolta ed analisi adeguate, quel circolo virtuoso di cui sopra, può essere innescato da ogni impresa, qualunque siano la dimensione e l'età. Quello che è cruciale, infatti, è la dotazione di architetture di raccolta e *data analyst* specializzati, che riescano a sviluppare soluzioni all'avanguardia e a soddisfare le necessità dei consumatori. I dati sono ovunque, ma non tutti hanno la stessa capacità nel tracciarli e conseguentemente elaborare informazioni. Non esiste infatti un unico, o pochi proprietari dei dati, sebbene esistono proprietari di informazioni e *know how* aziendali.

Dunque la domanda che ci si pone in questo paragrafo è la seguente: sono i dati un *Essential Facilities*?

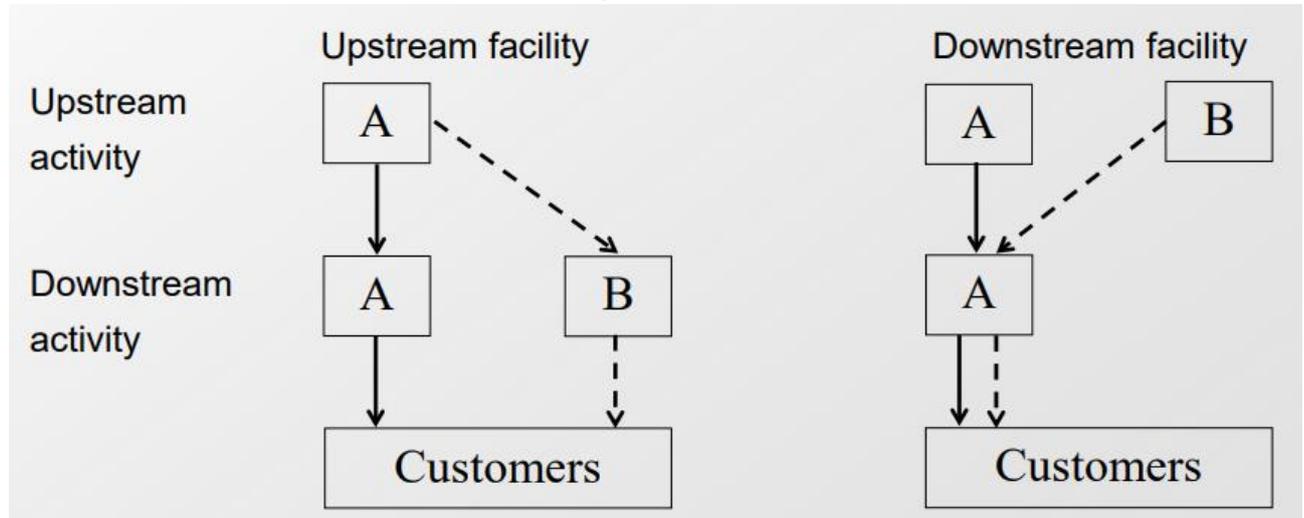
Per prima cosa si torni a quanto detto nel paragrafo precedente circa gli IPRs e l'*exclusionary abuse*. Dicevamo infatti che una delle pratiche unilaterali anticompetitive, condotte dalle imprese dominanti, era quella di negare l'accesso ai *competitor* ad una risorsa essenziale per lo sviluppo di un business in un dato settore.

Alcune imprese integrate verticalmente detengono *facilities*. Queste possono essere risorse di svariato tipo (infrastrutture, *networks*, brevetti, IPRs), appartengono all'*incumbent* e sono indispensabili per fornire beni e servizi ai consumatori finali.

¹⁰⁸ Big Data as a Misleading Facility - Giuseppe Colangelo e Mariateresa Maggiolino (2017)

Duplicare un'essential facilities è impossibile o troppo costoso (si pensi alle reti di telecomunicazione, oppure alle Ferrovie di Stato).

Figura 3.5



Nel grafico si evince che l'essential facility si può collocare a monte o valle della catena produttiva e, nel nostro caso, i dati rappresentano una risorsa che sta a monte. Infatti viene prima la raccolta di una grossa mole di dati, poi le imprese ci ricavano informazioni attraverso le analytics, creano migliori beni e servizi e vendono tali beni e servizi. Dunque tale filiera produttiva ha come punto di partenza la raccolta e l'analisi dei dati, con tutta la loro catena del valore (Capitolo 1).

L'Essential Facilities Doctrine (EFD), ha effetti contrastanti a seconda delle efficienze ricercate dalle autorità Antitrust. Essa, così come il sistema degli IPRs, ha effetti positivi sull'efficienza dinamica, infatti gli extraprofiti da monopolio attirano nuovi investitori che effettuano nuove invenzioni; effetti negativi sull'efficienza allocativa, infatti il risultato del rifiuto di garantire l'accesso ad una facilities essenziale è una minore competizione ed un conseguente aumento del prezzo di mercato, nonché una minore scelta di beni/servizi per il consumatore; essa infine ha effetti neutrali sul Consumer welfare, ma positivi per il surplus dell'impresa. Inoltre, se l'essential facility's owner è integrato verticalmente con un'attività commerciale (e così sono le maggiori digital companies), può sfruttare le economie di scopo¹⁰⁹ e beneficiare così dell'integrazione verticale, aumentando l'efficienza produttiva.

L'EFD, così come il sistema degli IPRs, mette in crisi le autorità Antitrust che tentano di trovare un bilanciamento tra protezione dei diritti di proprietà (nel nostro caso i dati e le informazioni che si ricavano da essi) e il bisogno di salvaguardare la competizione. È difficile cioè bilanciare efficienza allocativa e dinamica.

¹⁰⁹ Si realizzano economie di scopo (adattamento dell'ingl. *economies of scope*, cioè, con traduzione libera, *economie tratte dall'ampliamento del campo di attività*) quando, date le dimensioni degli impianti e, entro certi limiti, anche le quantità delle risorse umane e degli altri input, il costo totale della produzione congiunta o della produzione connessa di due o più beni all'interno di un'impresa è minore della somma dei costi totali sostenuti producendo separatamente, in imprese diverse, gli stessi beni. Economia Industriale (Cabra).

L'European Court of Justice (ECJ) ha stabilito che il rifiuto di garantire l'accesso ad una EF è abusivo se valgono tre condizioni:

- La risorsa è indispensabile poiché non vi sono alternative ed inoltre non è duplicabile;
- Il rifiuto all'accesso elimina ogni forma di competizione nel mercato, che rimane "chiuso" alle imprese che richiedono accesso all'EF;
- Assenza di giustificazione oggettiva.

In USA, invece, la Corte Suprema non ha mai riconosciuto una EFD e, per salvaguardare l'incentivo ad innovare, ritiene il possesso di un elevato PDM un elemento indispensabile per investire in R&D¹¹⁰ e sviluppare innovazioni vincenti. Tuttavia, se il possesso di un potere monopolistico viene accompagnato dalla stipula di contratti volontari con scopo anticompetitivo, il possesso di un EF e la non accessibilità alla stessa è considerata una pratica illecita. In sostanza, secondo la dottrina Statunitense, l'obbligo di garantire l'accesso ad una EF è inapplicabile se: il rifiuto non ha intenzionalità anticompetitiva; è giustificabile ed induce il monopolista a produrre beni o servizi che non si trovano sul mercato.

Ma, tornando alla domanda formulata in precedenza, ci si chiede se i dati sono una EF e se il rifiuto di far accedere i *competitors* a dataset di proprietà dell'impresa dominate possa costituire una pratica abusiva.

Come già accennato nel Capitolo 1, i Big Data non consistono in un grande ammontare dello stesso prodotto, poiché, con eccezione dei duplicati, essi sono diversi gli uni con gli altri. Uno stesso dataset, infatti, può condurre a diverse informazioni, a seconda di quali siano gli interessi dell'impresa proprietaria dei dati e diversi dataset, al contrario, possono condurre alla stessa informazione ricercata. I dati, quindi, spesso ammettono sostituti. Questa caratteristica, come detto, è nota come ubiquità dei dati. Infatti, se Facebook analizzasse i suoi dati con lo scopo di ricercare interessi politici degli utenti, Google potrebbe fare lo stesso considerando i dati provenienti dal motore di ricerca; oppure, se una biblioteca può raccogliere i dati e di conseguenza consigliare i libri da leggere ad un appassionato lettore che frequenta assiduamente la biblioteca, lo stesso può fare Amazon con il suo algoritmo di raccomandazione; oppure ancora, è possibile fare una ricerca sulla disponibilità a pagare di un individuo sia analizzando i dati provenienti dalla carta di credito, sia quelli provenienti dagli acquisti on line.

Tuttavia, diverse imprese possono aver bisogno di differenti informazioni provenienti dagli stessi dati. Riprendendo l'ultimo esempio fatto in precedenza, infatti, un'impresa può essere interessata alla tipologia di acquisti fatti da un individuo, un'altra a quale età gli individui utilizzano di più piattaforme *e-commerce*, un'altra ancora se utilizzano piattaforme di *e-commerce* più le femmine o i maschi. In conclusione, l'esclusività sui dati non implica esclusività sulle specifiche informazioni ricercate ed ottenute tramite l'elaborazione di dati.

¹¹⁰ Ricerca e Sviluppo.

Quindi i Big Data di cui un'impresa necessita, li può avere già un'altra impresa o sono messi a disposizione dagli utenti e quindi tracciabili sul web. Le imprese competono sui Big Data, o meglio, competono sulla qualità delle informazioni che riescono ad estrapolare dai dati, tenendo sempre in considerazione il fatto che ogni impresa ha una determinata mole di dati da tracciare ed esaminare.

*"Digital data are ubiquitous but not always accessible to anybody"*¹¹¹

Ci sono dunque barriere all'accesso ai dati. Tuttavia, una volta che le imprese si dotano delle risorse tecnologiche, umane ed intellettuali adeguate, possono tracciare tutto ciò che avviene all'interno della piattaforma e, come nel caso di Google, anche al di fuori, in altri siti web o piattaforme che richiedono spontaneamente di essere tracciati dal motore di ricerca. In sostanza, quindi, non c'è esclusività nel collezionare i dati digitali degli utenti. Non ci sono infatti accordi o contratti di esclusività tra piattaforma ed utenti. Questi ultimi infatti non sono intrappolati all'uso della piattaforma ed anzi, sono propensi al *multihoming*, fornendo dati a più *player* operanti sul web, in mercati rilevanti diversi e non. Gli utenti, a loro volta, beneficiano del *multihoming*, poiché riescono ad avere servizi digitali specializzati e standardizzati, con caratteristiche, funzioni e qualità di servizio diverse l'una dall'altra. Infatti, così come estremamente differenziati sono i gusti e le specificità che hanno gli utenti del web, ci sono moltissime piattaforme che assecondano tali attitudini. Ci sono infatti sia piattaforme che forniscono lo stesso servizio digitale (Google, Yahoo!, Bing, Ask!), sia piattaforme che soddisfano bisogni diversi (ad esempio Amazon e Google). Si ha infatti, come detto nel secondo capitolo, sia *intrasystem competition*, che *intersystem competition*. Questo perché gli utenti tendono ad utilizzare svariatissime piattaforme e il *multihoming* costituisce un fenomeno che avviene sia all'interno dello stesso mercato rilevante, che tra più mercati rilevanti. Secondo Nielsen, uno dei maggiori *player* di report statistici sul web, il 59% degli utenti di Google ha utilizzato anche Bing, Yahoo! o Ask! nell'arco del periodo di un mese, ed in totale, il 72.6% di utenti utilizza più di un motore di ricerca. Secondo la ricerca, inoltre, il 50% degli users, almeno una volta, ha utilizzato più di un motore di ricerca per soddisfare una *query*. PhoCusWright, un'impresa che fa ricerche sui viaggi, nel 2010 chiese agli utenti quanti siti web visitassero per pianificare un viaggio. I risultati furono che il 59% degli users utilizzava tre o più siti web, il rimanente 41% due o più. Il *multihoming*, dunque, si riflette sull'elevato grado di differenziazione dei servizi digitali offerti da ciascuna piattaforma che cattura dati, diversi o uguali, da stessi soggetti, al fine di effettuare operazioni, come *cluster analysis*, *regression analysis*, *factor analysis*, *correlation analysis*,¹¹² che permettono di monetizzare la raccolta dei dati attraverso la vendita di spazi per pubblicità online targettizzata o con

¹¹¹ The Role of "Big Data" in Online Platform Competition – Andres V. Lerner (26 Agosto 2014)

¹¹² Cfr: Capitolo 1, Paragrafo 1.4.4

la costruzione di algoritmi di raccomandazione (Capitolo 2), resi possibili attraverso l'elaborazione di un elevato ammontare di dati che permettono alla piattaforma di ricavare svariate informazioni.

*"No one firm controls all, most, or even a significant share of user data"*¹¹³

I Big Data si trovano ovunque sul web. Ogni click, ogni visualizzazione, ogni volontaria iscrizione ad un sito, costituisce un dato per una o più piattaforme digitali. Queste ultime hanno quindi accesso ad un elevato ammontare di dati forniti spontaneamente e volontariamente dagli utenti. Un esempio emblematico è costituito da Bing, il motore di ricerca di proprietà di Microsoft. Esso ha accesso ad un enorme volume di Big Data provenienti da svariati mezzi. Per prima cosa dal motore di ricerca (possiede infatti un numero di consumatori pari alla metà di quelli di Google, vedasi il grafico seguente) che fornisce *click-and-query data*; poi da altri servizi come Bing.com, Bing Toolbar e Internet Explorer (browser data). Così Bing ha accesso ad un'elevatissima mole di dati, incluse le ricerche effettuate su Google, Facebook e Twitter tramite Explorer, ed utilizza tali informazioni nel suo *search algorithm*.

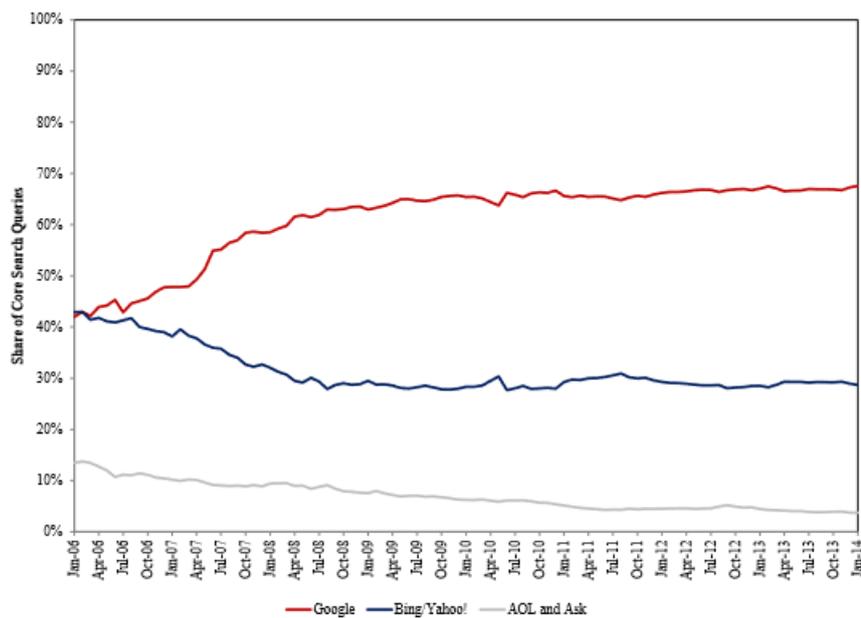


Figura 3.6¹¹⁴

Quindi, ci sono svariati mezzi da cui le imprese possono attingere per avere dati, ma non tutte le imprese dispongono di tali mezzi che, quindi, costituiscono barriere all'ingresso per molti *player*. Una piattaforma, infine, non può impedire ad altre di raccogliere i dati di suo interesse e che ha già prontamente raccolto.

I Big data sono infatti non rivali nel consumo, nel senso che la raccolta e l'analisi da parte di un'impresa non pregiudica la raccolta e l'analisi di un'altra.

¹¹³ Fonte: Big Data as a Misleading Facility - Giuseppe Colangelo e Mariateresa Maggiolino (2017)

¹¹⁴ L'impatto di Google sull'economia", Deloitte (2014)

Ci sono barriere all'ingresso, nel senso che tali imprese debbono dotarsi di risorse tecnologicamente avanzate ed investimenti in R&D per intraprendere un'attività di creazione di valore per mezzo dei dati. Inoltre, il *multihoming* permette a diverse piattaforme digitali di raccogliere dati sullo stesso utente. Quest'ultimo, infatti, a seconda dell'attività che effettua sul web, fornisce nome, cognome, indirizzo IP, localizzazione geografica e molto altro a svariate piattaforme. Ogni piattaforma poi riesce, chi più, chi meno, ad estrapolare valore da quei dati fornitigli spontaneamente dall'utente. Quindi, non è l'impresa con maggiori dati ad avere un vantaggio competitivo, ma l'impresa che riesce, per ogni dataset acquisito, ad estrarre maggior valore, inteso come informazione da inserire nei propri processi produttivi al fine di personalizzare il servizio verso i clienti e monetizzarlo efficacemente.

In conclusione, quindi, non è possibile applicare l'EFD ai Big Data per la loro non escludibilità, data dalla loro natura e dall'ubiquità, la non rivalità e il *multihoming*.¹¹⁵ Ora, per essere più chiari facciamo un esempio. Un'impresa A è una piattaforma digitale che ha una posizione dominante nel mercato dei *social networking services*, mentre l'impresa B vende giocattoli. Supponiamo ora che B chieda ad A i suoi dati e che quest'ultima si rifiutasse senza fornire una giustificazione oggettivamente valida. Ora, l'impresa B si appella all'EFD per ottenere l'accesso ai Big Data di A e l'autorità competente si chiede se effettivamente i dati di A sono indispensabili al business di B. Inoltre si domanda se è impossibile per B creare un suo mezzo di raccolta dati da utilizzare ai fini dei suoi scopi commerciali (ad esempio una piattaforma che raccolga recensioni di madri e figli, oppure dei sensori da apporre sui giocattoli). Tali mezzi di raccolta costituiscono nient'altro che barriere all'ingresso nel mercato dei dati. È economicamente impossibile sviluppare tali mezzi di raccolta dei dati? Se la risposta è negativa, l'EFD non è applicabile.

Ora, assumiamo che tali risorse di raccolta dati non sono duplicabili e che quindi solo A ha la possibilità di raccogliere dati sul consumo di giocattoli. B si appella nuovamente all'EFD, ma anche sta volta non riscontra successo poiché A non è attiva nel mercato dei giocattoli e quindi non utilizza i dati peculiari dei consumatori di giocattoli per i suoi fini commerciali. Nessuna piattaforma digitale infatti opera nei possibili mercati dove estrae dati ed informazioni. Ora, per il proseguo di questo esempio, supponiamo che la piattaforma A produca anche giocattoli. B attacca A dicendo che senza i suoi dati, B non può continuare a produrre giocattoli. Però questo appare come un controsenso, infatti la produzione può continuare anche senza l'ottenimento di dati digitali. Infatti il vantaggio di A risiede nell'estrarre informazioni sui trend di mercato, circa i gusti di bambini e madri che acquistano giocattoli. Tuttavia, è vero che B può avere tali informazioni solo accedendo ai dati di A, così ci si pone il dubbio se è possibile ottenere informazioni circa gusti ed attitudini di madri e bambini nell'acquisto di giocattoli attraverso altri mezzi di raccolta dati, come da motori di ricerca, blog, altri social network ecc. Anche qui, se la risposta è affermativa l'EFD non può essere applicata, poiché non c'è un'unica risorsa essenziale.

¹¹⁵ Exclusivity does not imply either essentiality or monopoly power - Maggiolino (2014)

Tuttavia, se la risposta è negativa, ovvero il prezzo da pagare per ottenere dati da social network, blog e motori di ricerca è insormontabile, si può applicare un EFD, ma questo è uno scenario che nella realtà trova difficile riscontro.

I Big Data, dunque, costituiscono una risorsa essenziale, ad esempio per sviluppare un motore di ricerca, o una piattaforma che raccomanda contenuti digitali, ma, al tempo stesso, tutte le imprese li possono ottenere, o, ancor meglio, tutte quelle imprese che superano le barriere all'ingresso, dotandosi di un efficace meccanismo di raccolta ed analisi. Essi sono inoltre non rivali e non escludibili e, grazie al multihoming, è possibile giungere alle informazioni ricercate da più svariati mezzi. Per tutti questi motivi dunque i Big Data costituiscono una *misleading facilities*¹¹⁶ e, pertanto, una disciplina basata sui fondamenti della EFD non può essere applicata.

3.3 Big Data & Price discrimination

Analizzando dati personali e non, è possibile avere una conoscenza maggiore circa la curva di domanda che l'impresa fronteggia e, tramite gli algoritmi, fare analisi predittive.¹¹⁷

Questo può facilitare pratiche di *price discrimination* tra differenti clienti che acquistano beni o servizi da una medesima *dominant company*. Ovvero è possibile, avendo a disposizione sufficienti dati, dedurre come i *buyers* si comportano in risposta a diversi prezzi, grazie ad analisi predittive. Tutto ciò si ottiene attraverso un'esplorazione accurata della curva di domanda. Ad esempio, Amazon, ha effettuato test, variando di continuo il prezzo di uno stesso DVD. Ha così raccolto numerosi dati aumentando e diminuendo il prezzo del DVD in intervalli di tempo molto brevi. Ad esempio, se un consumatore cercava il DVD alle 10 del mattino, pagava un prezzo minore rispetto ad un consumatore che accedeva alla search query alle 10.15. il seguente grafico mostra le variazioni di prezzo che sono accorse in due anni per un particolare giocattolo per bambini. Tali oscillazioni di prezzo, pertanto, hanno l'effetto di fornire informazioni circa la curva di domanda dei consumatori per quel singolo bene.

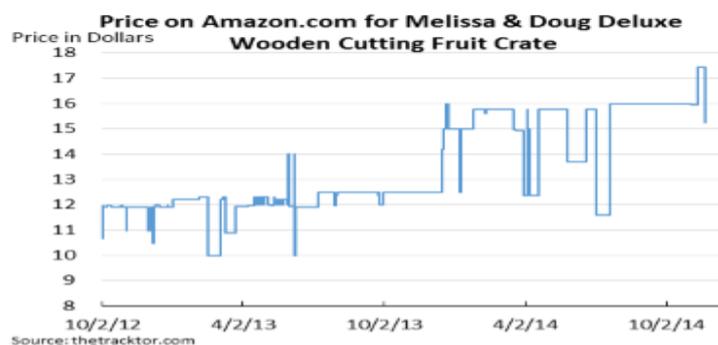


Figura 3.6 ¹¹⁸

¹¹⁶ *Misleading* nel senso di ingannevole. Giuseppe Colangelo e Mariateresa Maggolino (2017)

¹¹⁷ In the marketing context, big data refers to the ability to gather large volumes of data, often from multiple sources, and use it to produce new kinds of observations, measurements and predictions about individual customers. Big Data and Differential Pricing – Obama White House report (Febbraio 2015).

¹¹⁸ Big Data and Differential Pricing – Obama White House report (Febbraio 2015)

La differenziazione dei prezzi, o ciò che gli economisti chiamano *price discrimination*, è la pratica consistente nel porre ai clienti prezzi diversi per uno stesso prodotto. Questa pratica esiste in molteplici settori, anche *off-line*, ad esempio il cinema o il teatro che fanno sconti ai bambini o agli studenti universitari; oppure le compagnie aeree che, generalmente, pongono prezzi di vendita più alti se ci si trova in prossimità della data di partenza; oppure ancora gli sconti per chi possiede la “carta fedeltà” di un supermercato. Nei precedenti esempi, l’obiettivo dell’impresa che applica la strategia di differenziazione dei prezzi, è quello di aumentare i prezzi per coloro i quali sono disponibili a pagare di più, senza perdere l’altro gruppo di clienti più sensibili a variazioni di prezzo. Quindi i prezzi vengono fissati in base alla domanda (elasticità), o alla *willingness to pay*, dei clienti, e non in base alla struttura dei costi dell’impresa.

Ci sono tre tipi (o tre gradi) di *price discrimination*:

- Discriminazione di primo grado o *Personalized Pricing*: avviene quando il venditore pone un prezzo diverso per ciascun *buyer*. Il risultato è la fissazione di un *individual negotiated price* per ogni transazione (ad esempio i prezzi praticati da un *car dealer*).
- Discriminazione di secondo grado o *Quantity Discounts*: avviene quando il prezzo unitario diminuisce all’aumentare della quantità venduta ad un singolo *buyer* (ad esempio gli sconti 3X2 ai supermercati).
- Discriminazione di terzo grado: accade quando i venditori pongono prezzi diversi a differenti categorie di cittadini. Rileva infatti la dimensione demografica (ad esempio le esenzioni per i cittadini con basso reddito, per gli anziani, o per i bambini, oppure ancora per i disabili).

La *price discrimination* non è di per sé una pratica abusiva, infatti, se studiata a livello microeconomico, essa può comportare una maggiore efficienza allocativa, un maggiore surplus totale ed un effetto negativo sul surplus del consumatore, rispetto ad una semplice situazione di monopolio. Il monopolista, infatti, nel caso di strategia di differenziazione di prezzo, si appropria di tutto il surplus dei consumatori, ponendo per ciascuno un prezzo di vendita pari alla *willingness to pay* ed offrendo una quantità pari a quella che si avrebbe in un mercato perfettamente concorrenziale.

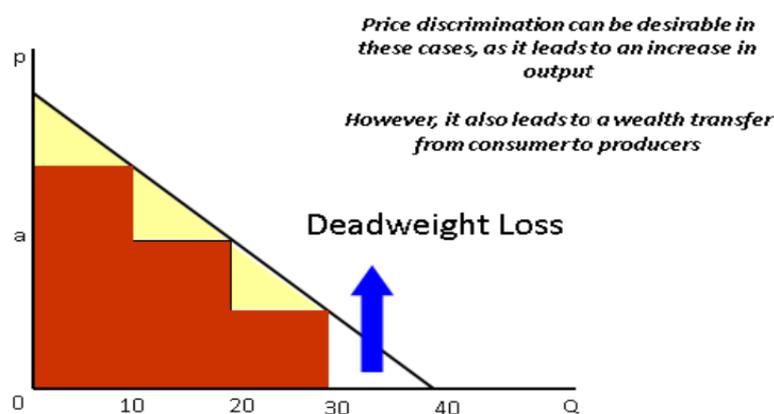


Figura 3.7¹¹⁹

¹¹⁹ EU *Competition law* Enforcement vis-à-vis Exploitative Conducts in the Data Economy Exploring the Terra Incognita – Marco Botta e Klaus Wiedemann.

Rispetto ad una classica situazione di monopolio, qui la *deadweight loss* è minimizzata, infatti l'efficienza allocativa è quasi pari a quella che si avrebbe in situazione di concorrenza perfetta. D'altro canto i consumatori ne potrebbero venire danneggiati, oltre che discriminati. Tuttavia, i consumatori che rappresentano la parte della curva di domanda più rigida, ovvero sono maggiormente sensibili a variazioni di prezzo, se prima non acquistavano il bene/servizio poiché aveva un prezzo di monopolio, ora lo potranno acquistare poiché avrà un prezzo pari alla loro disponibilità a pagare per quel bene.

Nel passato, le imprese studiavano il comportamento dei consumatori facendo esperimenti, via questionario, ad un campione di persone più o meno esteso. La *Big Data storage* e *analytics*, attraverso la raccolta e l'elaborazione di dati personali e non, provenienti da motori di ricerca, cookies¹²⁰, social networks, hanno reso possibile effettuare uno *shift* dalla discriminazione di terzo grado basata su categorie demografiche di clienti alla discriminazione di primo grado o strategia di *personalized pricing*. Infatti i dati permettono di creare profili diversi per ogni consumatore, riuscendo così a stimare cosa sono intenzionati ad acquistare ed a quale prezzo. Gli algoritmi permettono infatti di costruire profili personali basati sulle preferenze di ciascun individuo. È possibile così avere una visione sui comportamenti di acquisto di *cluster* di consumatori e ciò permette alle imprese di elaborare conseguenti strategie di marketing. La piattaforma, tramite gli algoritmi, può analizzare le caratteristiche e la personalità di ogni consumatore e può stimare i loro futuri comportamenti. Utilizzano poi tali conoscenze per offrire i propri prodotti con speciali sconti o determinate condizioni di acquisto (come l'esenzione dal pagamento dei costi di spedizione). La *price discrimination* è così una naturale conseguenza della raccolta di un elevato ammontare di dati da parte delle piattaforme digitali e sta diventando una pratica comune dei settori digitali, quindi essa è una strategia maggiormente efficace nella *data economy* piuttosto che nella "*real*" *economy*.

Nella realtà dei fatti, nonostante le piattaforme hanno una capacità di calcolo elevatissima e riescono a processare numerosi dati strutturati e non, è difficile stimare con esattezza la *willingness to pay* di un consumatore. Ossia è difficile stimare il cosiddetto *reservation price*, ovvero il prezzo massimo al quale il consumatore è disponibile ad acquistare un prodotto o, in altre parole, il prezzo che massimizza il profitto dell'impresa.

Nella *data economy*, la *price discrimination* assume più la connotazione di "*behavioural discrimination*", che rappresenta un mix dei tre gradi di discriminazione descritti sopra. Un esempio eclatante di ciò è data dai motori di ricerca, che possono differenziare la lista dei risultati mostrati nella *search result page* a diverse categorie di consumatori.

¹²⁰ A cookie is a small text file that a web site can place on a user's computer. Each time a user loads a particular web site, the cookie is sent to that site. This allows web sites to "remember" certain information, such as what pages a user has already visited, or whether they are currently logged in to the site. Wikipedia

Questa pratica prende il nome di “*steering*”¹²¹. Essa assomiglia ad una discriminazione di terzo grado poiché *clusterizza* i consumatori in base alle loro attitudini e/o appartenenza demografica. Ad esempio, Google può assegnare un *search ranking* più elevato a prodotti meno cari per consumatori più parsimoniosi; oppure Dell Computer può offrire, attraverso il proprio sito web, lo stesso *laptop* con prezzi diversi a differenti tipi di *buyers* che ne fanno un diverso uso (individuale, accademico, professionale, governativo). Infine, dopo aver cercato un prodotto su Amazon ed Ebay senza aver concluso l’acquisto, la piattaforma potrebbe contattare il potenziale cliente rioffrendo il prodotto, precedentemente visionato, con uno sconto. Conseguentemente i consumatori più pazienti hanno un trattamento favorevole rispetto a consumatori con un’attitudine all’acquisto più impulsiva.

Questi esempi mostrano come le *online platforms* hanno diversi *tools* per discriminare i propri clienti a seconda dei loro comportamenti ed attitudini d’acquisto. Accanto ai dati personali, come età, sesso, educazione, sono essenziali ulteriori informazioni, come gli acquisti *on-line* passati, la geo-localizzazione, la lista dei siti web visionati, le *search queries* effettuate, per permettere alle imprese di effettuare una efficace *behavioural discrimination*.

Alcuni casi empirici hanno confermato che la *behavioural discrimination* è una pratica molto utilizzata nei mercati online, ed in particolare lo “*steering*”. Infatti, se si vanno ad analizzare le stesse *queries* su uno stesso prodotto Amazon, provenienti da differenti paesi in Europa, Asia e USA, si nota che la piattaforma fornisce diversi prodotti ai diversi consumatori collocati in posizioni geografiche diverse, sebbene la *query* fosse la stessa. In particolare l’algoritmo di Amazon è in grado di analizzare sia la condizione economica della regione di provenienza, sia quella del consumatore (grazie ad esempio alla lista di pagine precedentemente ricercate e prodotti precedentemente acquistati). Così, consumatori situati in regioni a basso reddito e con ridotto potere d’acquisto, avranno una lista di prodotti che mette in risalto quelli più economici, al contrario, consumatori più “ricchi” e collocati in paesi ad alto reddito, avranno in risalto prodotti più cari.

Questi casi empirici mostrano come la *behavioural discrimination* è una strategia molto utilizzata nei settori digitali. È tuttavia difficile provare che le *digital companies* adottano sistematicamente tale strategia e che non ci siano oggettive giustificazioni a tale comportamento. È infatti difficile per i consumatori capire se un prezzo individuale è discriminatorio, poiché, nonostante sia in linea col suo *reservation price*, un consumatore non può sapere quali sono i parametri presi in considerazione dagli algoritmi per fornire i risultati delle sue ricerche con i relativi prezzi dei prodotti. Quindi un prezzo di mercato nella *data economy* appare meno trasparente rispetto ad un prezzo di mercato, per uno stesso prodotto, nella *real economy*.

¹²¹ Steering is the practice of showing different products to customers in different demographic groups. In the online environment, steering occurs when a web site alters its search results based on information about a potential customer. Like third-degree *price discrimination*, steering uses information about potential buyers, but not at the individual level. Big Data and Differential Pricing – Obama White House report (Febbraio 2015).

Gli economisti sono generalmente scettici su un intervento delle autorità antitrust contro l'attuazione di una *price discrimination* o *behavioural discrimination*. Secondo buona parte della dottrina, infatti, nel lungo periodo il mercato potrebbe aggiustarsi da solo e far scomparire l'*exploitative conduct*. In altre parole, se il prezzo di un bene/servizio viene percepito troppo alto o discriminatorio, i consumatori smetteranno di acquistare quel bene/servizio e passeranno ad altri fornitori. La pratica, infatti, risulta anticompetitiva se condotta da imprese dominanti che affrontano una concorrenza pressoché nulla, ed operano in un mercato con elevate barriere all'ingresso. Inoltre, la *price discrimination* può avere effetti pro-competitivi. In particolare, gli algoritmi di una piattaforma potrebbero monitorare i prezzi offerti dai *competitors* ed attrarre i consumatori offrendo conseguentemente prezzi più bassi. In questo modo la *price discrimination* potrebbe rafforzare la competizione nel mercato. Essa inoltre rende accessibile un prodotto a consumatori più "poveri", facilitando la distribuzione del benessere tra differenti categorie di consumatori.

Tuttavia, tali effetti pro-competitivi si hanno solo a determinate condizioni. Per prima cosa, la *price discrimination* aumenta il grado di competizione nel mercato se e solo se tutte le piattaforme hanno accesso alle medesime informazioni circa i loro potenziali consumatori, ovvero se ci si trova in condizioni di perfetta simmetria informativa. Nella realtà, le *online platforms* hanno accesso a differenti categorie di *personal data*, quindi ci troviamo davanti ad un mercato caratterizzato da asimmetria informativa e non esiste un'offerta ad hoc (che eguaglia il *reservation price*) per ogni consumatore. Inoltre, poiché l'obiettivo della *price discrimination* è quello di catturare il più possibile il surplus dei vari consumatori, c'è da fare un'accurata analisi per verificare l'impatto generale di tale strategia sia sui consumatori "ricchi" che su quelli "poveri". Infatti, sebbene la *price discrimination* potrebbe facilitare la redistribuzione del benessere tra differenti categorie di consumatori, essa tende ad incrementare il surplus del produttore a discapito del surplus totale di tutti i consumatori.

Ma, allora, quando la *price discrimination* costituisce una pratica abusiva?

Un'impresa dominante infrange l'Art 102 del TFEU quando applica condizioni di vendita dissimili a transazioni equivalenti con diversi *trading partners*. Questo dunque causa uno svantaggio competitivo per molti *buyers*, clienti dell'impresa in questione. L'Art 102 evidenzia che la pratica della *price discrimination* non è abusiva di per sé, ovvero c'è bisogno che questa sia fatta con intento anticompetitivo, necessita cioè dei requisiti dell'*equivalent transaction* e del *competitive disadvantage*. Infine, la Corte di Giustizia Europea (EUCJ) riconosce la possibilità di *objective justification* per giustificarsi ed esentarsi dalla pratica.

Il concetto di "*equivalent transaction*" fu utilizzato ed interpretato dalla EUCJ contro la *trading company* nominata United Brand. Essa, infatti, vendeva banane della stessa marca (Chiquita) a più Paesi e poneva prezzi diversi a seconda dell'elasticità della domanda di ciascun Paese.

In particolare, poneva un prezzo molto elevato in Irlanda rispetto agli altri paesi che United Brand riforniva. La disparità di prezzo violava l'Art 102 poiché United Brands importava banane di simile qualità, con la stessa origine (America Latina e Centrale), nelle stesse navi, con lo stesso Brand e le stesse condizioni di vendita. Perciò l'impresa aveva una struttura di costi, simile per ciascuna banana venduta in ciascuna Nazione, che non giustificava l'adozione di una strategia di *price discrimination*.

Anche il concetto di "*competitive disadvantage*" fu utilizzato per la prima volta in un caso pratico dall'EUCJ. Infatti la compagnia British Airways vendeva biglietti aerei a differenti agenzie di viaggio, ponendo loro prezzi discriminatori in base alla loro disponibilità a pagare. Questo dunque causava un deterioramento nelle posizioni competitive dei clienti discriminati e quindi uno svantaggio competitivo verso le altre agenzie "non discriminate".

Infine, per l'Art 102 non è necessario provare che ci sia una relazione causale tra *price discrimination* e efficienza dinamica, cosa invece molto importante per il regolamento USA. Inoltre la EUCJ non accetta oggettive giustificazioni se vi è discriminazione tra consumatori situati in differenti Stati dell'UE.

I Big Data hanno notevolmente abbassato il costo per ricavare informazioni rilevanti sui clienti e consumatori finali. Attraverso svariati mezzi, infatti, le piattaforme riescono ad estrapolare informazioni circa le caratteristiche dei consumatori. Ad esempio, molti siti web utilizzano l'*Internet Protocol* (IP) al fine di analizzare la localizzazione degli utenti. Oppure, ancora, quando un utente visita un sito web, il proprietario del sito può situare un file chiamato *cookie* così da tracciare tutte le interazioni dell'utente col sito web. Il processo di elaborazione di informazioni attraverso la raccolta e l'analisi dei Big Data rende quindi più semplice identificare nuovi segmenti di clienti e targettizzare gli stessi e i "vecchi" segmenti attraverso piani di marketing e politiche di prezzo ad hoc per ciascun segmento. Tuttavia, stimare la *willingness to pay* di ciascun *cluster* di consumatore è una sfida difficile anche per piattaforme con una grande quantità di dati ed elevata capacità di *computing*. Un altro ostacolo alla realizzazione di una strategia di *personalized pricing* è il grado di competizione che c'è nel settore. È infatti difficile aumentare i prezzi fino ad eguagliare la disponibilità a pagare di un *cluster* di consumatori, poiché questi potrebbero facilmente spostarsi verso fornitori più economici che vendono prodotti sostituti. Per ultimo, è difficile implementare il *personalized pricing*, poiché un attento consumatore potrebbe accorgersi di tale strategia e conseguentemente vedere tale pratica come illecita, andando ad incidere in modo negativo sulla reputazione dell'azienda in questione. Nella realtà dei fatti le *big digital companies* che posseggono un efficace *data storage* e *analytics*, possono implementare tale strategia. Tuttavia, questa può portare sia costi che benefici all'impresa.

Il beneficio principale è che, se il venditore ha un elevato PDM, tale strategia gli permette di espandere la propria QDM, offrendo al mercato una maggiore quantità e raggiungendo quindi una mole di consumatori maggiore con un elevato differenziale di disponibilità a

pagare. Un'università infatti può decidere di mettere una quota di retta annuale fissa e (presumibilmente) elevata, oppure porla in base al reddito familiare dello studente (ISEE). Con la prima strategia l'effetto è un minor numero di studenti iscritti e (presumibilmente) minori profitti. Con una strategia di *personalized pricing*, invece, il numero di studenti iscritti (ed il profitto) aumenterebbe.

D'altro canto, uno dei possibili costi causati dall'implementazione di tale strategia, può essere la riduzione del tasso di innovazione causato dalla focalizzazione sul cosiddetto *versioning*¹²², ovvero un'eccessiva differenziazione di prodotto, con conseguente differenziazione di prezzo (prodotto premium vs prodotto low cost) che causa una minore ottimizzazione delle risorse (se non si hanno sufficienti economie di scopo da sfruttare) e quindi una produzione meno efficiente.

Ad esempio, per un produttore di smartphones può essere meno costoso produrre un singolo design di prodotto piuttosto che un'ampia gamma differenziata di *devices* per ogni *cluster* di cliente individuato. Il *versioning*, infatti è una strategia efficace prevalentemente per gli *information goods*, come libri, film o software, poiché i costi di riproduzione sono tipicamente piccoli comparati col prezzo di vendita. Grazie alla raccolta ed analisi dei Big Data, per le imprese è più efficace ed efficiente intraprendere una strategia di *differential pricing* piuttosto che una strategia di *versioning*.

Un altro aspetto da considerare è che l'effetto sui consumatori, positivo o negativo, dipende da come e dove il *differential pricing* viene utilizzato. In un mercato competitivo con prezzi trasparenti, i benefici superano i costi. Ad esempio, poiché ci sono prezzi molto differenti nella vendita di biglietti aerei, Internet rende facile per molti viaggiatori comparare i prezzi di diverse compagnie aeree e selezionare il migliore per ciascuna tipologia di viaggio. Inoltre il *differential pricing* può aumentare la competizione verso i consumatori maggiormente sensibili ad aumenti di prezzo, che si rivolgono ai fornitori più economici. Inoltre, i consumatori possono intraprendere una strategia di arbitraggio come risposta ad una strategia di *differential pricing* adottata dalle imprese. Ad esempio, se un venditore offre uno stesso prodotto in due città con due prezzi diversi, un astuto consumatore, potrebbe acquistarlo nella città dove il prezzo è basso e rivenderlo ad un prezzo maggiore nell'altra città.

¹²² Firms often produce multiple versions of a product to encourage consumers to self-select into groups that pay different prices, even when it would be more cost efficient to sell a single design. For example, at one time IBM sold two versions of its LaserPrinter Series E, where the only difference between them was a chip that made the low-priced version print more slowly (Deneckere and McAfee 1996). While intentionally disabling some features of a product to facilitate price-discrimination seems perverse, it can nevertheless increase welfare for both firms and consumers if it allows the seller to reach a larger number of customers who would not otherwise be served. EU *Competition law Enforcement vis-à-vis Exploitative Conducts in the Data Economy Exploring the Terra Incognita* – Marco Botta e Klaus Wiedemann.

3.4 Big Data e condizioni contrattuali illecite

Nei precedenti paragrafi abbiamo studiato alcune delle pratiche abusive condotte da imprese aventi nel mercato una posizione dominante. Tale PDM è dato dalla mancanza di prodotti sostituti e dagli IPRs che proteggono l'impresa e le permettono, conseguentemente, di aumentare il proprio potere contrattuale. Le *big digital companies*, infatti, fronteggiano più domande ed offrono più servizi. Quindi stipolano contratti con ogni side della piattaforma.

Qui si discuterà su come tali entità dominanti, in virtù del loro elevato potere contrattuale, adottano pratiche scorrette nella stipula dei contratti con i loro clienti, in particolare con i consumatori finali. L'analisi non è focalizzata pertanto ad enfatizzare il rapporto di contraccambio che c'è tra users o consumatori finali e *platform*, poiché, come già detto, esso prescinde da una transazione senza scambio di denaro, in cui il valore del diritto di processare i dati personali e non degli utenti equivale al valore del servizio gratuito offerto dalla piattaforma, in un rapporto che, apparentemente, sembra di tipo *win to win*.¹²³

Figura 3.8



Tuttavia non è facile stabilire se effettivamente i due beni/servizi scambiati abbiano lo stesso valore e se i consumatori siano consapevoli di fornire preziosi dati ogni qual volta utilizzano la piattaforma e non solo.

Nella realtà dei fatti, viene fatto un utilizzo molto intenso delle piattaforme digitali che, in termini di ore spese, rappresentano uno dei passatempi preferiti dagli utenti. Di conseguenza esse apportano un notevole valore aggiunto e difficilmente gli utenti ne rinuncerebbero (ad esempio, se Instagram mettesse una piccola *fee* per il suo acquisto, è da supporre che non tutti gli utenti rinuncerebbero a scaricarlo, passando magari ad altre piattaforme).

Ma tornando al nostro rapporto di contraccambio tra utenti e piattaforma, in questo paragrafo ci si pone la domanda di come i dati vengono processati, chi detiene il diritto di accesso e di quanto sia trasparente tale contratto di scambio. Gli utenti, infatti, danno il loro consenso a processare i dati personali in cambio della possibilità di utilizzare la piattaforma gratuitamente, ma, molto spesso, forniscono tale consenso in maniera impulsiva e senza sapere la reale destinazione dei propri dati forniti. Essi infatti, senza saperlo, stanno pagando il servizio reso dalla piattaforma, con l'accesso a "pezzi" di *privacy*.

¹²³ EU *Competition law Enforcement vis-à-vis Exploitative Conducts in the Data Economy Exploring the Terra Incognita* – Marco Botta e Klaus Wiedemann, Paragrafo 4.3

Gli utenti di internet, dunque, dichiarano una chiara preferenza per un certo livello di protezione alla *privacy* mentre utilizzano certi servizi *on-line*, mentre il mercato, solitamente, non fornisce più di un'opzione di *privacy*, ovvero, da soddisfare tutte le preferenze di *privacy* di ciascun cliente. A differenti livelli di *privacy* e quindi di dati personali forniti, dovrebbe infatti corrispondere un differente prezzo per l'accesso alla piattaforma. Ossia, più dati fornisco alla piattaforma, meno pago, se li fornisco tutti indistintamente, il servizio mi viene reso gratuitamente. Se non voglio divulgare tutte le mie informazioni, dovrò corrispondere una *fee* alla piattaforma pari al valore dei dati che non sono stati resi. Tuttavia, nella *data economy*, il servizio viene reso, molto spesso, indistintamente gratuito, così che ogni utente che vuole utilizzare una piattaforma digitale si trova nella posizione di dover decidere se fornire ogni tipo di dato accettando i termini e le condizioni alla *privacy* o rinunciare totalmente al servizio della piattaforma. Non ci sono dunque *cluster* di utenti diversi in base al valore che essi riservano alla *privacy protection*.

Tuttavia, ci sono stati casi in cui, nei contratti tra utenti e piattaforma, le condizioni di scambio e i diritti di accesso ai dati non erano ben chiari agli utenti che, dando il loro consenso al trattamento dei propri dati personali alla piattaforma, perdevano il controllo sugli stessi e permettevano ad altre piattaforme o siti web di avere accesso a tali dati e, di conseguenza, proliferare maggiormente il proprio servizio verso l'utente.

Tali pratiche possono rientrare in quelle enunciate dall'Art 102 del TFEU, dove per pratica abusiva si intende anche "altre condizioni di transazioni non eque". Tuttavia l'Art 102 non fornisce una chiara definizione di quando una clausola contrattuale possa definirsi illecita e la domanda chiave che ci si pone in questo paragrafo è: sotto quali circostanze i termini e le condizioni imposte ai consumatori rappresentano un abuso di posizione dominante secondo l'Art 102 del TFEU?

Ci si riferirà, dunque, agli utenti o consumatori finali, in quanto il rapporto di scambio tra questi e la piattaforma è il più difficile da tracciare.¹²⁴ Come detto, essi rinunciano alla loro *privacy* in cambio di un servizio gratuito. Gli altri clienti della *platform* ottengono altri servizi (pubblicità, dati, visibilità ecc.) ad un certo prezzo che, come visto, può o non essere equo, in relazione alla teoria degli *excessive prices* e della *price discrimination* (Paragrafi 3.1 e 3.3), in una logica di *exploitative abuse*.

Le condizioni di scambio possono essere illecite e quindi rappresentare una pratica di *exploitative abuse*, in quanto, in virtù dell'importanza della piattaforma per i consumatori e, conseguentemente, in virtù dell'elevato PDM, la piattaforma impone ai consumatori finali condizioni di scambio dei dati che possono danneggiare quest'ultimi, andando ad intaccare la loro *privacy*.

Il principio generale dell'EU *data protection law* è che è necessario il consenso del *data controller*, ovvero dell'utente consumatore finale, affinché una piattaforma possa avere il diritto di processare i suoi dati personali.

¹²⁴ "Instead, they "trade in" a bit of their privacy in the form of personal data". EU *Competition law* Enforcement.

Secondo l'Art 6 del GDPR, infatti, è il consenso il requisito necessario per intraprendere un'attività di *data storage, analysis e processing*, e, tale consenso, deve essere fornito dagli utenti in maniera del tutto libera ed informata, nel senso che la piattaforma deve mettere in risalto gli utilizzi che fa dei dati personali degli utenti, se questi vengono trasmessi a terze parti e quali sono le implicazioni a livello di personalizzazione del servizio. Infatti, secondo l'Art 4 del GDPR quattro devono essere i requisiti affinché il consenso sia lecito. Esso deve essere fornito liberamente, specificamente per ogni utente, in modo informato e non ambiguo. Ovvero, la piattaforma deve dotare l'utente di tutte le informazioni circa gli usi dei dati e lo deve fare in modo chiaro evitando situazioni ambigue. L'Art 7 del GDPR specifica le condizioni per un valido consenso. Ovvero esso deve essere fornito dal *data controller* e quest'ultimo deve essere libero e consapevole di fornire il proprio consenso alla piattaforma. Inoltre l'Art 9 parla di esplicito consenso, nel senso che il consumatore deve volutamente fornire i dati alla piattaforma per gli usi che quest'ultima, in maniera del tutto esplicita, ne fa e, conseguentemente, la piattaforma deve informare in maniera chiara il consumatore della destinazione dei dati fornitigli.

L'approccio del regolamento GDPR è in linea con la tradizionale letteratura sulla *privacy*, che assume che un'adeguata protezione della *privacy* si ottiene garantendo il controllo dei dati personali al soggetto che emette tali dati. Ovvero l'utente deve avere piena consapevolezza e controllo su cosa viene fatto dalla piattaforma attraverso i suoi dati personali a seguito del consenso fornito da quest'ultimo in modo libero e volontario. Egli deve poter sapere con certezza chi ha accesso ai suoi dati, come questi vengono processati e quali sono le implicazioni di business che ne derivano (ad esempio personalizzazione dei contenuti e *recommendation system*).

Tuttavia, nella maggior parte dei casi, gli atteggiamenti dei consumatori non sono allineati a quanto dice il GDPR¹²⁵ e, spesso, forniscono il loro consenso in maniera superficiale senza avere piena consapevolezza sugli usi che la piattaforma fa dei propri dati. Le piattaforme, d'altro canto, sono consapevoli di ciò ed adottano atteggiamenti che sfavoriscono una chiara e consapevole fornitura di dati, andando pertanto a commettere pratiche illecite, sfavorendo ed occultando alcune delle destinazioni che i dati hanno.

Quello che è appena stato descritto rappresenta il cosiddetto "*privacy paradox*". Questa espressione si riferisce al fenomeno secondo cui la maggior parte dei consumatori rivendica l'importanza della propria *privacy* ed il bisogno di una *data protection*, ma, in realtà, essi non agiscono secondo tali preferenze e desideri. Gli utenti di internet, infatti, spesso divulgano liberamente i propri dati personali, semplicemente accettando i termini e le condizioni alla *privacy*, senza leggerle attentamente e senza capire gli usi e le destinazioni che i dati hanno.

Tuttavia, l'atteggiamento degli utenti dipende dal contesto e dal tipo di informazioni che essi forniscono ed inoltre diversi utenti hanno diverse attitudini verso la *data protection*. Quindi, le preferenze in merito alla *privacy* degli utenti di internet sono eterogenee.

¹²⁵ General Data Protection Regulation.

Tutto ciò, molto spesso, è dovuto a due problemi: la maggiorparte delle *privacy* policies sul web sono eccessivamente lunghe che quasi nessun utente impiega tutto il tempo necessario per leggerle tutte. Allo stesso tempo, il linguaggio utilizzato dalle piattaforme è, molto spesso, di difficile comprensione per molti utenti.

Verrà ora presentato un caso inerente ai termini contrattuali scorretti e dunque pratiche abusive messe in atto da Facebook nel mercato e sanzionate dall’Autorità garante della concorrenza tedesca: il *Bundescartellamt*.

Case study: The Facebook investigations by the Bundescartellamt

Un caso molto emblematico che merita di essere citato nell’ambito dell’exploitative abuse causata dall’imposizione di clausole contrattuali illecite è l’investigazione dell’autorità garante della concorrenza tedesca, il *Bundescartellamt*¹²⁶, contro Facebook. Questo è un caso emblematico che, seppur ancora rimasto irrisolto, riunisce tematiche che comprendono sia aspetti di *data protection* che di *competition law* e rappresenta il primo caso di *competition* in cui vengono applicate norme relative alla *privacy* per la sua risoluzione.

A marzo 2016 il *Bundescartellamt*, ha iniziato formalmente ad investigare le procedure di Facebook. L’autorità affermava di non essere certa circa la consapevolezza da parte degli utenti sulla raccolta ed analisi di alcuni dati che Facebook processava. Così, la violazione della *data protection law*, diventa cruciale per la decisione dell’autorità in merito all’abusività della condotta di Facebook, secondo le disposizioni dell’Art 102 del TFEU.

Il *Bundescartellamt*, nel suo *background paper*¹²⁷, affermava che, nelle situazioni in cui l’accesso ai dati personali degli utenti rappresenta un fattore chiave per il posizionamento nel mercato di un’impresa, non solo le autorità competenti in tema di *data protection*, ma anche le *competition authorities* hanno voce in capitolo e devono pertanto intervenire. A Maggio 2016, infatti, l’autorità, assieme alla *Autorité de la Concurrence Francese*, affermava che, anche se la *data protection* e la *competition law* perseguono diversi obiettivi, bisogna tener conto che il modo di analizzare e processare elevati ammontari di dati personali, possa incidere sul livello di competizione di un’industria.

Il 19 Dicembre 2017 così, il *Bundescartellamt* affermava nella sua valutazione preliminare, che Facebook avesse abusato della sua posizione dominante.

“Facebook is abusing [its] dominant position by making the use of its social network conditional on its being allowed to limitlessly amass every kind of data generated by using third-party websites and merge it with the user’s Facebook account. These third-party sites include firstly services owned by Facebook such as WhatsApp or Instagram, and secondly websites and apps of other operators with embedded Facebook APIs.”

¹²⁶ The Bundeskartellamt is an independent *competition* authority whose task is to protect *competition* in Germany. The protection of *competition* is a key regulatory policy objective in a market economy.

¹²⁷ Fonte: Preliminary assessment in Facebook proceeding: Facebook’s collection and use of data from third-party sources is abusive – Bundeskartellamt (19 Dicembre 2017)

Quindi si evince che, nella sua pubblicazione, il *Bundescartellamt*, fa riferimento sia ad una mancanza di trasparenza verso gli utenti circa le modalità con cui i dati di questi ultimi venivano processati, sia alla forzatura degli utenti sul far loro trasmettere indistintamente tutti i dati personali pena l'impossibilità di utilizzare la piattaforma. Infatti, le preferenze degli utenti in tema di *privacy* non vengono discriminate da Facebook, che quindi forza i consumatori a trasmettere ogni tipo di dato personale proveniente da più fonti, interne ed esterne alla piattaforma. Così Facebook stava abusando della sua posizione dominante, subordinando l'accesso alla piattaforma al trasferimento di ogni tipo di dato personale dagli utenti a Facebook e da altri siti web a Facebook, unendo tali dati e personalizzando così sempre più il servizio verso gli utenti.

Il *Bundescartellamt*, infatti, fa una distinzione tra collezione e uso di dati all'interno del Social Network (*on Facebook*) e quelli provenienti da siti web terzi (*off Facebook*). Solo gli ultimi sono quelli oggetto di investigazione dell'autorità tedesca e si riferiscono a quei siti web e app che condividono le APIs di Facebook, che permettono la condivisione dei dati. Questo non è solo il caso delle altre piattaforme di proprietà di Facebook (WhatsApp e Instagram), ma anche per milioni di altri siti web che, solo apparentemente, non hanno alcun collegamento con Facebook. Tali siti web, infatti, trasferiscono dati personali a Facebook attraverso ad esempio l'utilizzo del "like bottom", dello "share bottom", oppure ancora del "log-in bottom". essi permettono a Facebook di tracciare ogni interazione tra l'utente e il sito web. Secondo l'autorità questi termini davano la possibilità a Facebook di raccogliere dati dagli utenti che navigano su altri siti, anche se tali utenti disabilitavano il cosiddetto "web tracking" attraverso l'inibizione dei *cookies* e anche se non interagivano direttamente con la piattaforma, ad esempio schiacciando il pulsante "mi piace" o "condividi". Le API¹²⁸ condivise con Facebook da siti web terzi erano infatti il presupposto sufficiente al trasferimento dei dati verso il Social Network, che ne analizzava i contenuti. Tali dati poi, come detto, vengono uniti agli altri raccolti direttamente dalla piattaforma o indirettamente attraverso le app di proprietà della piattaforma (WhatsApp e Instagram) per creare profili di utenti sempre più dettagliati e conseguenti alle loro attività sul web. Facebook viene considerata dal *Bundescartellamt* come impresa dominante nel mercato tedesco. Secondo l'autorità, infatti, i consumatori sono intrappolati all'uso della piattaforma e non passano ad un altro Social Network con un identico scopo. Così gli utenti non hanno altra possibilità che accettare i termini e le condizioni imposte da Facebook per utilizzare la piattaforma. Il risultato è che gli utenti non hanno effettivamente acconsentito a tale condivisione di dati tra Facebook e siti web terzi e questo è in contrasto con la *data protection law* che protegge la libertà di decisione dei consumatori riguardo i propri dati personali, che non possono quindi essere forzati a condividere con altre entità.

¹²⁸ An application program interface (API) is a set of routines, protocols, and tools for building software applications. Basically, an API specifies how software components should interact. Additionally, APIs are used when programming graphical user interface (GUI) components. Wikipedia

I termini e le condizioni alla *privacy* sono dunque il centro dell'investigazione e la chiave della valutazione competitiva. È infatti la violazione della *data protection law* il fattore decisivo per la decisione se la condotta Facebook è o no abusiva per l'Art 102 del TFEU.

L'autorità infatti affermava che:

- Gli utenti non avevano liberamente acconsentito ad una tale condivisione ed elaborazione di dati, come richiesto dall'Art 7 del GDPR;
- Gli utenti non avevano ricevuto sufficienti informazioni sugli utilizzi che venivano fatti dei loro dati, in violazione dei requisiti di trasparenza enunciati dall'Art 5,12 e 14 del GDPR;
- Le ragionevoli aspettative degli utenti circa l'elaborazione dei dati che loro stessi fornivano alla piattaforma non erano in linea con gli usi e le destinazioni che la piattaforma faceva dei dati, in violazione dell'Art 6 del GDPR;

I requisiti dell'Art 21 del GDPR sulla profilazione dei dati degli utenti non venivano dunque rispettate.

La violazione della *data protection law*, quindi, in concomitanza con la forzatura verso i consumatori ad accettare di fornire a Facebook dati anche provenienti da terze parti, ha condotto ad una preliminare decisione del *Bundescartellamt* contro Facebook. L'autorità, pertanto, considerava illeciti ed abusivi i termini e le condizioni alla *privacy* imposte dalla piattaforma ai suoi utenti. Ovvero, quindi, tali termini rappresentavano una forma di *exploitative abuse*. L'autorità infatti spiegava nel *Preliminary Assessment* che, in virtù della legge tedesca in merito alla concorrenza, se i termini imposti da Facebook rappresentavano una manifestazione di dominanza ed elevato PDM, tali termini erano da considerarsi *exploitative* e quindi abusivi secondo i principi della legge civile tedesca. In particolare, Facebook è così dominante che è praticamente in grado di dettare i termini contrattuali autonomamente, eliminando l'autonomia contrattuale dell'altra parte.

Su queste basi, dunque, l'Autorità affermava che Facebook, in virtù del suo PDM e della sua posizione dominante nel settore dei social network, ignorava i diritti degli utenti in merito alla *privacy* e protetti dalla *data privacy law*. Così la piattaforma commetteva un *exploitative abuse* della sua dominanza.

Quindi, riassumendo, Facebook è dominante nel mercato tedesco, forza i consumatori ad accettare i propri termini e condizioni contrattuali e fissa tali termini e condizioni alla *privacy* in modo poco trasparente. Tutto ciò viene visto come pratica anticompetitiva dall'Autorità tedesca, in virtù dello sfruttamento della posizione dominante di Facebook, che induce i consumatori ad accettare lo scambio di dati, ponendo clausole contrattuali illecite, che favoriscono tale scambio.

Il *Bundescartellamt*, a differenza di altre autorità, come quella Italiana, non ha il potere di adottare accorgimenti amministrativi e di sanzionare una pratica commerciale illecita sotto la *German consumer law*.

Tuttavia, la scelta di policy seguita dal *Bundescartellamt* comprende sia la *consumer, competition* e *data protection law* poiché condividono un ampio numero di legami, anche se perseguono differenti obiettivi, diversi scopi di applicazione e differenti mezzi di risoluzione. Quindi, sebbene in Europa vi sia una chiara differenza tra *competition law* e *data protection law*, andando a stabilire di volta in volta quale applicare, quando si tratta di imprese dominanti nei loro settori, questa posizione può cambiare, andando a risolvere casi di *abusive practices*, attraverso *policies* di *privacy law*.

3.5 Algorithmic collusion

L'intelligenza artificiale (IA), con le tecnologie del *machine learning* e del *deep learning*¹²⁹ sta avendo un impatto notevole sull'economia globale e sulla società in generale, modificando, in particolare, il modo di competere delle imprese.

Gli algoritmi sono strumenti che imparano dai dati e dall'esperienza. Essi vengono sviluppati per svolgere funzioni ripetitive e costituite da calcoli complessi ed elaborazione di dati che potrebbero essere troppo costosi se condotti da un essere umano. Gli algoritmi sono molto pervasivi nella nostra società, in quanto tracciano, predicono e influenzano i comportamenti degli individui. Essi apportano benefici specialmente in termini di automazione, efficienza e qualità, sia dal lato imprese, sia dal lato clienti/consumatori, poiché gli algoritmi riescono a facilitare e supportare il processo di *decision making* di un individuo.¹³⁰

Questo paragrafo descrive il modo in cui gli algoritmi hanno cambiato lo scenario competitivo delle industrie digitali, offrendo l'opportunità alle imprese di raggiungere accordi collusivi in modi del tutto nuovi che non richiedono il raggiungimento di un accordo in modo tradizionale (scritto o verbale) e non necessitano, quindi, di interazioni umane per la loro riuscita. Le teorie economiche suggeriscono che c'è un considerevole rischio che gli algoritmi¹³¹, aumentando la trasparenza nel mercato e permettendo di scambiare beni con alta frequenza, aumentino il rischio di collusione in settori caratterizzati tradizionalmente da un alto grado di competizione. Questo paragrafo mostra come gli algoritmi possono facilitare la *tacit collusion*¹³², fornendo alle imprese un potente mezzo per comunicare, implementare strategie comuni, monitorare tali strategie e punire eventuali defezioni, intese come deviazioni dall'equilibrio collusivo trovato dall'algoritmo.

¹²⁹ Cfr: Capitolo 2, Paragrafo 2.2

¹³⁰ Algorithms and Collusion, Competition policy in the digital age – OECD (2017)

¹³¹ An algorithm is an unambiguous, precise, list of simple operations applied mechanically and systematically to a set of tokens or objects (e.g., configurations of chess pieces, numbers, cake ingredients, etc.). The initial state of the tokens is the input; the final state is the output. Wilson and Keil (1999).

¹³² Collusione raggiunta senza alcun tipo di accordo esplicito (scritto, verbale)

Non ci sono dubbi circa la potenzialità e l'efficienza degli algoritmi nell'estrarre valore dall'elevato ammontare di dati messi a disposizione. Tuttavia, essi rappresentano anche un nuovo strumento che ha rivoluzionato il modo di operare e di interagire delle imprese e possono causare il rischio che alcuni *players* utilizzino le capacità computazionali di cui dispongono per soddisfare interessi privati congiunti, a discapito di obiettivi sociali.

Rientriamo pertanto nell'ambito di applicazione dell'*exclusionary abuse* dell'Art 101 del TFEU, secondo cui "Sono incompatibili con il mercato comune e vietati tutti gli accordi tra imprese, tutte le decisioni di associazioni di imprese e tutte le pratiche concordate che possano pregiudicare il commercio tra Stati membri e che abbiano per oggetto e per effetto di impedire, restringere o falsare il gioco della concorrenza all'interno del mercato comune." In questo paragrafo si studierà il modo di operare degli algoritmi per facilitare la realizzazione di pratiche concordate al fine di ottenere un equilibrio collusivo.

In letteratura, col termine "collusione" ci si riferisce ad ogni forma di coordinazione o accordo tra *competitors* con l'obiettivo di aumentare i profitti, rispetto ad una situazione di non cooperazione, risultante in un aumento della *deadweight loss*. In altre parole, la collusione è la strategia che porta alla massimizzazione dei profitti congiunti di due imprese, a discapito del surplus dei consumatori. Così, la collusione permette a più imprese di esercitare un PDM che non avevano prima di raggiungere l'equilibrio cooperativo. Tutto ciò risulta in un aumento dei prezzi (cartelli) o di altri parametri che incidono negativamente sulla competizione e riducono il consumer e il total welfare. C'è, pertanto, un trasferimento di ricchezza dai consumatori ai membri del cartello ed un aumento della *deadweight loss*. Da un punto di vista sociale, inoltre, i costi per formare e rinforzare un cartello rappresentano una perdita di benessere, tuttavia, gli algoritmi hanno permesso un notevole abbattimento di tali costi, denominati *rent seeking*. Un altro possibile aspetto negativo causato dalla formazione di cartelli è che, le imprese partecipanti potrebbero vedersi ridurre il loro incentivo ad innovare, a causa della stabilità nel tempo dei profitti risultanti dalla realizzazione dell'equilibrio collusivo, con una perdita in efficienza dinamica.

Ci sono differenti modi per colludere. L'*explicit collusion* si riferisce alla realizzazione di condotte anticompetitive congiunte realizzate mediante l'esecuzione di contratti vincolanti oppure di contratti informali non necessariamente vincolanti, in forma scritta o orale. La maniera più diretta per le imprese di raggiungere un accordo collusivo esplicito è quella di interagire direttamente ed accordarsi su un livello ottimale di prezzo o di output. L'*explicit collusion* è una pratica vietata sia dall'Art 101 TFEU, sia dal regolamento antitrust Statunitense (Sezione 1 dello Sherman Act).

La *tacit collusion*, al contrario, si riferisce a forme di coordinazione anticompetitiva, che vengono raggiunte senza il bisogno di un esplicito contratto, ma che i competitor accettano di mantenere riconoscendo ed assecondando l'interdipendenza con il business di un'altra entità.

In un contesto di *tacit collusion*, il risultato anticompetitivo viene raggiunto automaticamente tra imprese che adottano la loro strategia di massimizzazione dei profitti indipendentemente dai propri competitors. Questo, tipicamente, accade in mercati caratterizzati da elevata trasparenza e con pochi *players*, in cui le imprese beneficiano di un PDM collettivo raggiunto senza nessuna comunicazione esplicita. È questo il caso degli algoritmi che, automaticamente, giungono a risultati, in termini di prezzo, output, market share, ottimizzanti sia per l'impresa che per i competitor, evitando i mezzi di competizione tradizionali e raggiungendo un elevato grado di soddisfacimento, in termini di profitti, per le imprese operanti nel settore.

Tale distinzione tra explicit e *tacit collusion* mostra come, sotto determinate condizioni nel mercato (ad esempio: elevata trasparenza, pochi *players*, prodotti omogenei), la cooperazione sia il normale risultato economico del comportamento razionale di più imprese nel settore, per stabilire strategie di prezzo, output, o più in generale con qualsiasi strumento competitivo a disposizione delle imprese (Ad esempio limitare la pubblicità o restringere gli orari di apertura dei negozi per i settori *brick and mortar*¹³³).

Molti studiosi hanno provato a stimare gli impatti economici dei cartelli e quantificarne il danno è molto difficile. Secondo l'OECD i cartelli in media causano un aumento dei prezzi di circa 15-20%, causando danni ai consumatori per un ammontare di quasi 2 bilioni di dollari all'anno. Altri studi più recenti hanno individuato una media di sovrapprezzo del 17-21% e del 20-25% per i cartelli domestici e del 34% per quelli internazionali.

Figura 3.9

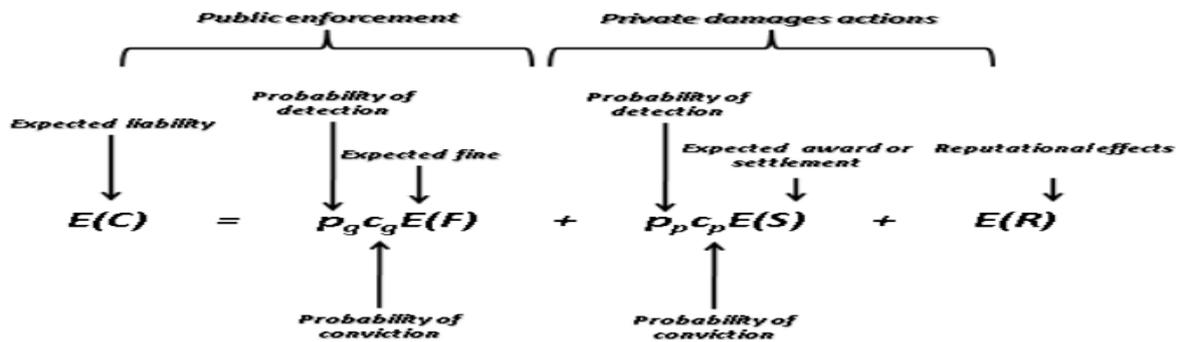
Reference	Number of Cartels	Average Overcharge	
		Mean	Median
<i>Percent</i>			
1. Cohen and Scheffman (1989)	5-7	7.7-10.8	7.8-14.0
2. Werden (2003)	13	21	18
3. Posner (2001)	12	49	38
4. Levenstein and Suslow (2002)	22	43	44.5
5. Griffin (1989), private cartels	38	46	44
6. OECD (2003), excluding peaks	12	15.75	12.75
Total, simple average	102-104	30.7	28.1
Total, weighted average	102-104	36.7	34.6

In Europa, stime ottimistiche hanno rilevato un danno di circa 16.8 bilioni di euro ed un trasferimento di ricchezza di 11.2 bilioni da consumatori a imprese e 5.6 bilioni di *deadweight loss*. secondo stime pessimistiche, invece, il danno ammonterebbe a 261.2 bilioni di euro, di cui 174.14 bilioni trasferiti da consumatori a imprese e 87.07 di *deadweight loss*.

¹³³ brick and mortar refers to a physical presence of an organisation or business in a building or other structure. The term brick-and-mortar business is often used to refer to a company that possesses or leases retail shops, factory production facilities, or warehouses for its operations. Wikipedia

Un'impresa non si muoverà da un equilibrio collusivo fino a quando i benefici supereranno i costi; ovvero, i costi intesi come multa attesa da parte dell'autorità garante della concorrenza e reazione dei competitor ad un'eventuale deviazione dall'equilibrio collusivo, senza tralasciare gli effetti sulla reputazione dell'azienda, non devono superare i benefici, intesi come profitto, dell'impresa partecipante al cartello. In sostanza, un'impresa continuerà a cooperare con i suoi *competitors* fino a quando vale tale relazione:

Figura 3.10



Non è semplice, tuttavia, per le imprese raggiungere un equilibrio collusivo. È infatti difficile trovare un prezzo che soddisfi tutti i membri di un cartello ed inoltre, ogni impresa ha un incentivo a deviare unilateralmente da un equilibrio collusivo se ha la possibilità di aumentare maggiormente le proprie vendite o i prezzi e quindi i propri profitti e, d'altra parte, reggere l'urto della reazione dei competitor. Gli algoritmi facilitano la realizzazione di una collusione tacita, riuscendo a trovare l'ottimo paretiano, inteso come quantità da offrire e prezzo, ad ogni situazione strutturale del mercato e a reagire in tempo reale ad ogni possibile cambiamento di strategia da parte dei competitor.

L'algoritmo quindi riesce a trovare sempre la strategia migliore di prezzo o di quantità dell'impresa in relazione ad altre strategie adottate dagli altri *players* dell'industria¹³⁴, in modo del tutto automatico e senza il bisogno di giungere ad accordi, formali ed informali, con i competitor, ma, semplicemente, calcolando di volta in volta quale strategia si addice maggiormente alla massimizzazione dei profitti per l'impresa e, di riflesso, per gli altri membri del cartello.

Figura 3.11

		Firm B	
		C	NC
Firm A	C	10, 10	5, 15
	NC	15, 5	7, 7

¹³⁴ Calcola automaticamente l'ottimo paretiano, nella prospettiva in cui ogni impresa facente parte del cartello, calcoli a sua volta l'ottimo paretiano.

In una situazione di giochi ripetuti, dunque, la collusione è possibile da sostenere. Ogni impresa o algoritmo deve fronteggiare un *trade-off* e scegliere, di volta in volta, tra benefici presenti e perdite future. Quindi, se il valore scontato delle perdite future eccede i benefici presenti, la collusione è la migliore strategia da perseguire. Per far funzionare il meccanismo collusivo, sono dunque necessari un'elevata abilità nello scovare possibili deviazioni dall'equilibrio cooperativo e un meccanismo di punizione credibile che può ridurre drasticamente il profitto dell'impresa deviante, a causa di guerre di prezzi o altre pratiche competitive.

Indipendentemente dai mezzi usati dalle imprese, è di comune preoccupazione che gli algoritmi possano facilitare il raggiungimento di un equilibrio collusivo e che questo, quindi, possa essere una strategia comune per molti settori digitali. Qui si discuterà su come l'uso degli algoritmi e la crescente disponibilità di dati strutturati e non stiano cambiando le condizioni del mercato, rendendo maggiormente possibili e convenienti pratiche concordate e strategie cooperative che possano falsare ed ostacolare il normale funzionamento dell'industria e le sue dinamiche competitive.

Le caratteristiche strutturali del settore svolgono un ruolo chiave in tal senso. Più un settore è concentrato, maggiori saranno le probabilità di giungere ad un equilibrio collusivo. Il numero di imprese e le barriere all'ingresso sono le due maggiori caratteristiche strutturali che aumentano il rischio di collusione. Un numero elevato di imprese, infatti, non solo rende maggiormente difficile entrare in un equilibrio collusivo, ma riduce anche gli incentivi a colludere, in quanto ciascun player otterrebbe una porzione di extraprofitti da collusione sempre più bassa. D'altro canto, in assenza di barriere all'ingresso, le nuove entranti sarebbero attratte dagli extraprofitti di collusione, le incumbent ad entrare nel cartello, ed aumenterebbe l'incentivo a deviare dall'equilibrio collusivo, in quanto ogni deviazione sarebbe giustificata da un differenziale di profitto troppo elevato che, anche se per un brevissimo lasso di tempo, viene beneficiato dall'impresa.

Inoltre, un elevato numero di clienti rende più difficile giungere alla cooperazione; maggiormente le imprese sono simmetriche, in capacità, tecnologia, market share, organizzazione, struttura dei costi, più facilmente giungono ad un equilibrio cooperativo; più inelastica è la domanda che fronteggia un cartello, maggiore è il prezzo che il cartello riesce a fissare e più conveniente sarà la collusione; infine è di comune accordo che l'elevato grado di differenziazione del prodotto rende la coordinazione più difficile, tuttavia è difficile stabilire se due imprese che vendono prodotti omogenei hanno un alto incentivo a colludere.

Gli algoritmi hanno un notevole impatto sulle caratteristiche della struttura di un'industria. In particolare, molti dei settori digitali in cui gli algoritmi vengono utilizzati per fissare i cosiddetti prezzi dinamici, segmentare i consumatori o migliorare la qualità del proprio prodotto offerto; hanno al loro interno un elevato numero di imprese di grande dimensione, questo è il caso dell'industria dei motori di ricerca, piattaforme di e-commerce, siti web di prenotazione di aerei, treni ed altri mezzi di trasporto e social networks. Molti di questi settori sono caratterizzati, come già detto, anche dalla presenza di elevate barriere all'ingresso, date dall'economie di scala e dal network effect¹³⁵, che permettono alle imprese di crescere, raccogliere ammontari crescenti di dati e sviluppare algoritmi più efficienti ed efficaci. Tuttavia, gli algoritmi costituiscono essi stessi barriere all'ingresso, come conseguenza alla capacità di *data analysis* di un'impresa digitale.

Inoltre, gli algoritmi possono essere utilizzati per identificare ogni possibile minaccia proveniente dal mercato, dando la possibilità agli incumbent di reagire prontamente in modo aggressivo a qualunque strategia di ingresso nel mercato, oppure ad acquisire preventivamente i potenziali competitors. D'altro canto, una volta "superate" le barriere all'ingresso, l'efficienza e l'efficacia di tali algoritmi, possono indurre a giungere ad utili informazioni di mercato per i potenziali entranti che possono far abbassare il costo per l'entrata nel settore.

Infine, un peculiare aspetto degli algoritmi è che rendono il numero di *competitors* nel mercato un fattore meno rilevante per la collusione. Come detto, nei mercati tradizionali, la collusione è più sostenibile in settori con meno *competitors* (mercati oligopolisti), è così più facile giungere ad accordi collusivi, monitorare deviazioni ed implementare efficaci meccanismi di punizione. Gli algoritmi permettono meccanismi di coordinazione, monitoraggio e punizione in maniera del tutto automatica anche in settori poco concentrati, grazie all'abilità e alla rapidità nel collezionare e processare elevati ammontari di dati. In altre parole, il basso numero di imprese è un fattore importante, ma non rappresenta una condizione necessaria per la creazione di un *algorithmic collusion*.

Altre due importanti caratteristiche che rendono più semplice per gli algoritmi giungere ad un equilibrio collusivo, sono il grado di trasparenza nel mercato e la frequenza di interazioni tra imprese. Mentre la trasparenza permette alle imprese di monitorare in maniera efficace le azioni degli altri membri del cartello e individuare atteggiamenti devianti, la frequenza delle interazioni permette di aggiornare continuamente e in tempo reale le condizioni di vendita in base ai trend della domanda. Un esempio di ciò è dato dai cosiddetti *dynamic prices*, i quali permettono a clienti e fornitori di osservare in tempo reale i rapidi cambiamenti di prezzo in diversi settori, ed agire di conseguenza.

¹³⁵ Cfr: Capitolo 2, Paragrafo 2.1

I *pricing algorithms* sono codici computazionali col fine di fissare automaticamente i prezzi che massimizzano il profitto dell'impresa, date certe condizioni. Essi danno la possibilità alle imprese di reagire tempestivamente con un'altra strategia ottimizzante, al mutare di alcune condizioni di mercato. Sono i casi delle piattaforme che vendono biglietti aerei, camere di albergo, biglietti sportivi, trasporti in generale, elettricità e industrie *retail*. Tali algoritmi sono particolarmente utili ad implementare una strategia di prezzo che mutua di continuo nel tempo, il *dynamic pricing* o ad offrire a diversi *cluster* di consumatori prezzi differenti in base alla loro *willingness to pay*, *price discrimination* (Paragrafo 3.3).

Le autorità garanti della concorrenza Tedesca e Francese affermano che l'aumento della trasparenza del mercato e della frequenza delle interazioni, causano un aumento del rischio di collusione:

“Even though market transparency as a facilitating factor for collusion has been debated for several decades now, it gains new relevance due to technical developments such as sophisticated computer algorithms. For example, by processing all available information and thus monitoring and analysing or anticipating their competitors’ responses to current and future prices, competitors may easier be able to find a sustainable supra-competitive price equilibrium which they can agree on.” Autorité de la Concurrence and Bundeskartellamt (2016)

Un'altra importante caratteristica strutturale è data dal grado di innovazione. La dinamicità dell'industria, infatti, riduce il valore attuale degli accordi collusivi, così come la scarsa abilità delle imprese ad innovarsi. In questo campo, gli algoritmi rappresentano un importante mezzo di innovazione, permettendo alle imprese di sviluppare *business models* non tradizionali ed estrarre valore dai dati, col fine di migliorare la qualità e la customizzazione dei propri prodotti. In settori in cui gli algoritmi rappresentano una fonte di vantaggio competitivo, questo è ad esempio il caso dei motori di ricerca, app di navigazione, piattaforme di *matching*, le imprese potrebbero fronteggiare un'elevata pressione competitiva per sviluppare il *best-performing algorithm*.

Sintetizzando in uno schema gli impatti degli algoritmi sulle caratteristiche appena citate, è difficile stabilire se gli algoritmi incrementano o no, in generale, il rischio di collusione.

Figura 3.12¹³⁶

Relevant factors for collusion		Impact of algorithms on the likelihood of collusion
Structural characteristics	Number of firms	±
	Barriers to entry	±
	Market transparency	+
	Frequency of interaction	+
Demand variables	Demand growth	0
	Demand fluctuations	0
Supply variables	Innovation	-
	Cost asymmetry	-

Note: + positive impact; - negative impact; 0 neutral impact; ± ambiguous impact.

¹³⁶ Algorithms and Collusion, Competition policy in the digital age – OECD (2017)

Nonostante gli apparenti effetti ambigui, gli algoritmi hanno sostanzialmente cambiato le caratteristiche strutturali che affettano maggiormente il grado di competizione di un'industria digitale: la trasparenza di mercato e la frequenza delle interazioni.

Tuttavia, la presenza di algoritmi aumenta il rischio di collusione e questo deve essere un fattore di allerta per le autorità antitrust che hanno come obiettivo quello di ridurre il rischio di collusione tacita. Il più grande vantaggio che apportano algoritmi alle strategie concordate delle imprese digitali è che essi riescono a monitorare i comportamenti delle imprese in tempo reale senza il bisogno dell'intervento umano, permettendo così alle imprese di sostituire la *explicit collusion*, con una coordinazione tacita ed automatizzata. Quest'ultima funzione, il monitoraggio, viene svolta dai cosiddetti *monitoring algorithms*. Essi possono processare informazioni sulle decisioni di business dei competitors, su possibili atteggiamenti devianti dall'equilibrio cooperativo e programmare eventuali reazioni immediate. Ad esempio, è possibile un *pricing algorithms* che, automaticamente, reagisce a deviazioni dai prezzi pattuiti dal cartello, assecondando così una possibile guerra dei prezzi, o *trigger strategy*.¹³⁷

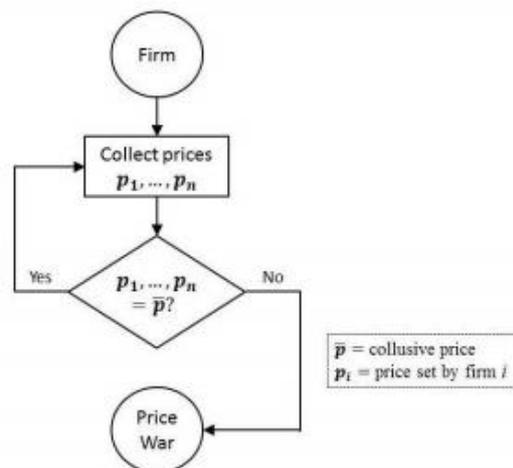


Figura 3.13¹³⁸

Poiché tali algoritmi sono molto veloci ad individuare e sanzionare condotte devianti, le imprese non hanno alcun incentivo a deviare dall'equilibrio collusivo.

¹³⁷ Regardless of the means used by companies, there is a concern that algorithms make collusive outcomes easier to sustain and more likely to be observed in digital markets. This section discusses how the use of algorithms and the increasing availability of business data online are changing market conditions and how that may potentially enhance the likelihood of collusive outcomes. Section 5 will then discuss in further detail what this means for competition enforcement and the challenges algorithms pose under the legal framework that most jurisdictions use to assess competitors' co-operation. Economists have identified the most relevant factors that may increase the likelihood of collusion in a given market.¹² These factors can be divided into structural characteristics, demand-side characteristics and supply-side characteristics. This section discusses how computer algorithms are changing in particular the structural and supply characteristics, so as to make digital markets and some traditional markets more prone to collusion. Algorithms and Collusion, Competition policy in the digital age – OECD (2017)

¹³⁸ Algorithms and Collusion, Competition policy in the digital age – OECD (2017)

Un'altra difficile *task* che riescono a risolvere gli algoritmi, come detto, è quella di riuscire a monitorare i continui cambiamenti strutturali di mercato e, conseguentemente, adottare strategie di pricing concordate con gli altri membri del cartello. Così tali algoritmi, detti *parallel algorithms*, riescono ad automatizzare le decisioni delle imprese partecipanti al cartello e reagire simultaneamente ponendo di volta in volta prezzi massimizzanti ogni qual volta c'è un cambiamento strutturale. È questo l'esempio dei *dynamic prices* citati precedentemente.

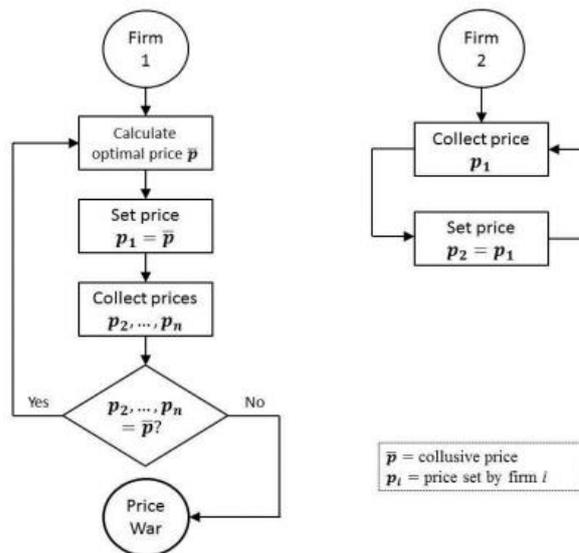


Figura 3.14¹³⁹ Note. Firm 1 is the leader and firm 2 is the follower.

Inoltre, col fine di evitare comunicazioni esplicite, gli algoritmi possono segnalare e rivelare l'intento collusivo agli altri *players* attraverso annunci sul prezzo. Ovvero le imprese possono sviluppare i cosiddetti *signalling algorithms* col fine di comunicare alle altre imprese le proprie strategie di prezzo. Un'impresa può pertanto segnalare ad un'altra che sta aumentando il prezzo e l'altra può, quindi, assecondare tale aumento di prezzo, praticando un prezzo più elevato e massimizzando così i profitti congiunti delle due imprese. Quindi tali algoritmi servono per stabilire e negoziare i termini della collusione prima di intraprendere una strategia di *pricing co-ordination*.¹⁴⁰

¹³⁹ Algorithms and Collusion, Competition policy in the digital age – OECD (2017)

¹⁴⁰ During the early 1990's, the US DOJ investigated tariff fixing activities in the airline industry, where the cartel members were able to implicitly coordinate tariffs using a third party centre and sophisticated signalling mechanisms. The case is described in detail in Borenstein (1999).

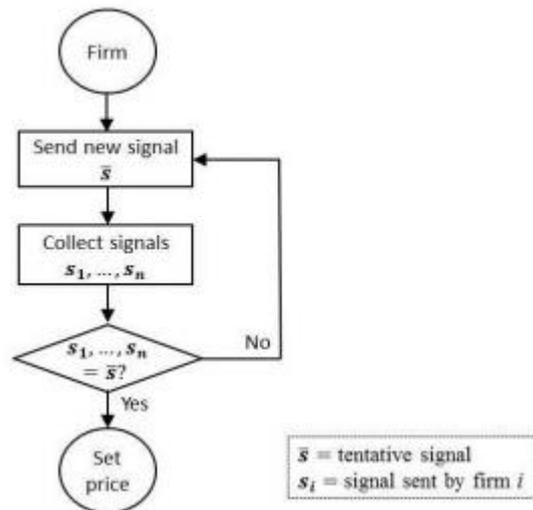


Figura 3.15

Attraverso l'uso del *machine learning* e del *deep learning*, infine, gli algoritmi integrati a tali meccanismi sviluppano una notevole capacità predittiva, imparando costantemente dall'esperienza e riadattandosi alle azioni degli altri player dell'industria. Tutto questo è svolto in maniera automatica, senza il bisogno, quindi, di un intervento umano.

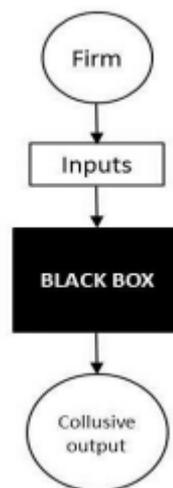


Figura 3.16

Figura 3.17¹⁴¹

Role in implementing collusion	
Monitoring algorithms	Collect and process information from competitors and eventually punish deviations.
Parallel algorithms	Coordinate parallel behaviour, for instance by programming prices to follow a leader, sharing pricing algorithms; or using the same third party algorithm.
Signalling algorithms	Disclose and disseminate information in order to announce an intention to collude and negotiate the common policy.
Self-learning algorithms	Maximise profits while recognising mutual interdependency and readapting behaviour to the actions of other market players.

¹⁴¹ Algorithms and Collusion, Competition policy in the digital age – OECD (2017)

Gli algoritmi hanno così cambiato le caratteristiche dei mercati digitali, aumentandone la trasparenza, la rapidità nel prendere decisioni strategiche e l'abilità delle imprese nel reagire tempestivamente alle azioni dei competitors. Sotto queste circostanze, gli algoritmi possono rendere interdipendenti le azioni dei vari *players* dell'industria senza il bisogno di un'esplicita comunicazione o interazione, aumentando il rischio di collusione tacita e portando così ad un aumento del livello generale dei prezzi. Gli algoritmi possono pertanto distorcere la normale competizione dei "meriti", creando incentivi e meccanismi che facilitano la collusione, dove con i metodi tradizionali non si riusciva a raggiungere. Tuttavia, come visto, gli algoritmi di ML portano alcuni vantaggi competitivi, associati con l'automazione, riduzione dei costi, aumento della qualità e migliore allocazione delle risorse. Al momento, non c'è ancora un'evidenza empirica sugli effetti degli algoritmi sui livelli attuali di prezzo e sul grado di competizione nel mercato. È importante, inoltre, fare una distinzione tra algoritmi che facilitano condotte già regolate dai *framework* vigenti e condotte che, invece, aumentano il rischio di atteggiamenti non regolati da alcuna norma antitrust, la *tacit collusion*. Infatti, gli algoritmi possono direttamente facilitare un equilibrio non competitivo e, essendo uno strumento che elimina il bisogno di un'esplicita comunicazione o interazione tra competitors, sostenere situazioni di collusione tacita.

In generale, secondo l'Art 101 del TFEU, un prerequisito per un intervento antitrust contro accordi e pratiche anticompetitive è l'esistenza di un'evidente forma di coordinazione tra *competitors* che questa abbia un impatto negativo sulla competizione e causa un danneggiamento alla posizione dei consumatori. Come abbiamo visto, l'attenzione delle autorità dovrebbe focalizzarsi su quei mercati che hanno caratteristiche peculiari che causano l'attuazione di accordi collusivi, come l'elevata trasparenza, l'alta frequenza di interazioni e la possibilità di prevedere i comportamenti degli altri *players*.

È possibile adottare un approccio *ex-ante*, consistente nello stabilire un intervento mirato ad ostacolare fusioni ed acquisizioni in settori dove c'è un'elevata attività di mercato guidata dagli algoritmi. Questo approccio potrebbe far sì che si riduca il rischio di una futura coordinazione delle attività di più imprese, sia in settori molto concentrati, che in altri meno.

Tuttavia, è ancora poco chiaro se il sistema regolatorio possa prevenire la formazione di algoritmi di ML che possano raggiungere un equilibrio collusivo. È difficile, dunque, dimostrare il danno verso il consumatore e verso la struttura competitiva del settore. Così ancora nessuna soluzione è stata proposta dalla letteratura antitrust per fermare tali condotte. Inoltre, non ci sono casi o investigazioni in atto che possono fornire un adeguato nesso comparativo e supportare l'evidenza che la collusione tacita danneggia i consumatori e i competitors, attuali e potenziali, che si collocano al di fuori del cartello.

CAPITOLO QUARTO

BIG DATA ED OPERAZIONI DI CONCENTRAZIONE

Introduzione

La rivoluzione digitale sta colpendo tutti i settori in economia. Ogni impresa ha un potenziale inespresso a causa della mancanza di investimenti nelle tecnologie digitali. Come visto nel secondo Capitolo, ci sono settori, come quello finanziario o l'*health care* che sono stati completamente rivoluzionati dal digitale e dalle tecnologie di ML e DL e altri, come quello agricolo, che ancora non ne hanno sfruttato appieno le potenzialità. Il dotarsi delle infrastrutture e delle *skills* per la raccolta e l'analisi dei dati, come abbiamo più volte discusso, rappresenta la barriera all'ingresso per eccellenza per le imprese che vogliono far entrare il ML e il DL nei propri processi produttivi. Così, un crescente numero di imprese, sceglie di acquisire tali competenze piuttosto che svilupparle al loro interno¹⁴². Si stanno così incrementando operazioni di M&A con lo scopo di acquisire Big Data ed implementare le informazioni raccolte, tramite la loro elaborazione nei processi produttivi delle imprese. Le piattaforme digitali, a loro volta, giocano un ruolo fondamentale in tale mercato di M&A. Esse, infatti, avendo accesso ad un grande quantitativo di dati, come vedremo, sono i protagonisti sia di operazioni di concentrazione che hanno come scopo quello di incrementare l'efficacia e l'efficienza della fase di *analysis e processing*, sia di operazioni di concentrazione che permettono a più piattaforme di mettere insieme i propri dati, al fine di avere più informazioni di maggior rilevanza. L'acquisizione di imprese *high tech* sta diventando, dunque, uno strumento importante per stimolare l'innovazione, automatizzando processi decisionali e procedure informative e fornendo ai consumatori prodotti, servizi ed esperienze sempre più personalizzati.

Dopo la crisi finanziaria dei primi anni Duemila, causata dalla crisi delle dot.com¹⁴³, c'è stata una notevole ripresa del mercato finanziario delle imprese *high tech*. Dal 2012 il volume di affari del mercato M&A ha avuto una notevole ripresa e sono aumentati così gli investimenti in imprese tecnologicamente avanzate sia tramite canale diretto, sia tramite la costituzione di fondi di private equity, interessati alla rapida crescita delle imprese digitali. Il volume di affari dell'industria digitale ammontava così, nel 2016, a 700 bilioni di dollari, rappresentando circa il 30% del valore di tutto il mercato M&A (che valeva circa 2.5 trilioni di dollari).

¹⁴² Cfr: BCG M&A Report (2017)

¹⁴³ Cfr: documentary about Dot-Com Bubble, What Happened (2001). Available on Reddit

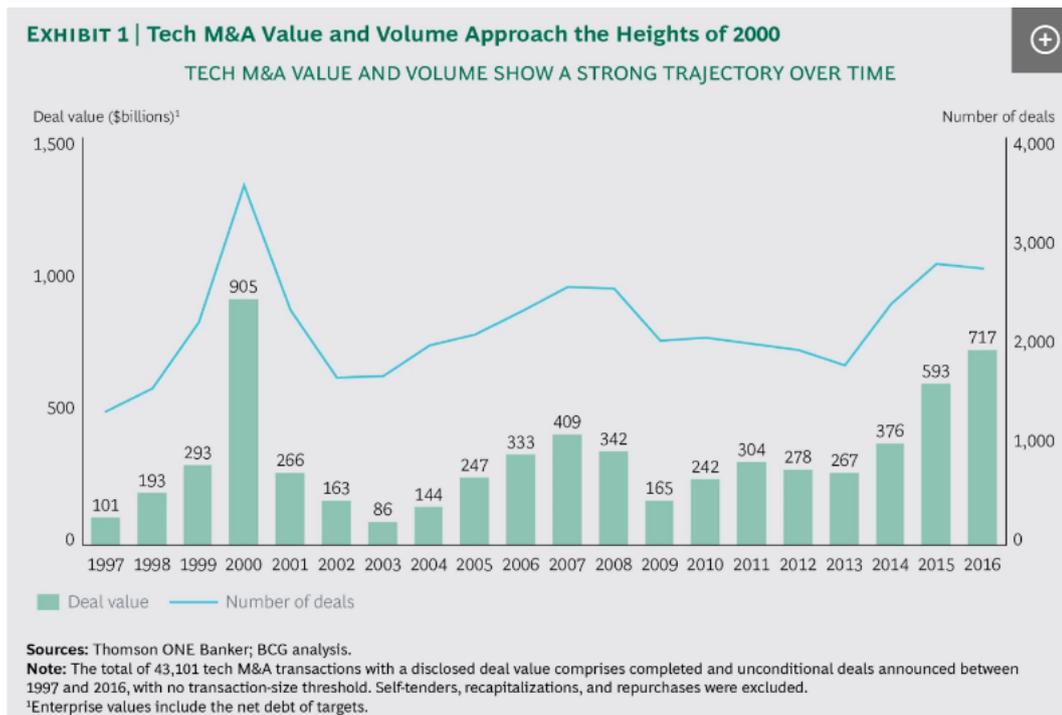


Figura 4.1

Approssimativamente, il 70% di tutte le trattative di M&A che avevano come *target* una *digital company*, coinvolgeva *buyers* al di fuori del mercato digitale. Oggi le imprese digitali sono diverse da quelle dell'epoca delle dot.com dei primi anni Duemila. Ora le imprese *high tech* hanno un business model più o meno definito e, soprattutto, fanno *revenues* in modo chiaro ed esplicito, cosa che invece i primi siti web non facevano, oppure non monetizzavano in maniera efficace la propria attività. Quindi è chiaro che i protagonisti delle operazioni di M&A, col passare degli anni, sono diventate le cosiddette "*technology targets*"¹⁴⁴, ovvero imprese digitalmente avanzate che adottano in maniera intensa il ML e/o il DL nei propri processi produttivi. Secondo uno studio condotto dalla BCG (Boston Consulting Group), l'età media delle *target companies* era di sei anni nell'epoca delle dot.com, mentre nel 2016 era di 14 anni; i ricavi in media dell'impresa *target* erano di 83 milioni di dollari, nel 2016 di 206 milioni; infine, solo il 53% delle imprese *target* aveva un EBIT positivo nel 2000, in comparazione al 65% del 2016¹⁴⁵. Questo potrebbe essere il risultato di una maggiore solidità del business model della *target* e di un migliore sfruttamento delle tecnologie digitali che attirano le attenzioni di potenziali acquirenti che vogliono digitalizzare le proprie attività.

¹⁴⁴ According to BCG "Our goal was to develop a working definition of a "technology target" that goes beyond broad Standard Industrial Classification-based definitions to include companies that have some form of technology as an essential attribute or part of their business model." Thomson ONE Banker; BCG analysis

¹⁴⁵ Cfr: <https://www.bcg.com/publications/2017/corporate-development-finance-technology-digital-2017-m-and-a-report-resurgent-high-tech-marketplace.aspx>

EXHIBIT 2 | Nine Tech Trends Shape the M&A Market

Trend	Number of deals in 2016	Average deal size, 2013–2016	CAGR 2013–2016 ¹ (%)	Landmark deals in 2016	
				Acquirer	Target
Business intelligence and big data and analytics	190	\$376 million	27	Thoma Bravo	Qlik
Cloud and software as a service	160	\$375 million	29	Oracle	NetSuite
Data center, infrastructure, and security	125	\$241 million	12	Apollo Global Management	Rackspace
Fintech	150	\$96 million	6	Société Générale	TagPay
Gaming	30	\$201 million	49	Vivendi	Gameloft
Health care IT	12	\$585 million	59	IBM Watson Health	Truven Health Analytics
Industry 4.0	616	\$349 million	18	Siemens	Mentor Graphics ²
Mobile tech and software applications	246	\$535 million	6	Microsoft	LinkedIn
Smart connectivity and mobility	133	\$376 million	23	General Motors	Cruise Automation
Total	1,662	\$348 million	17		

Sources: Thomson ONE Banker; BCG analysis.
Note: The tech M&A transactions comprise of completed and unconditional deals announced in 2016, with no transaction-size threshold. Self-tenders, recapitalizations, and repurchases were excluded. Only deals with a disclosed deal value were considered in this analysis.
¹The CAGR calculation was based on the number of disclosed deals announced between 2013 and 2016.
²The deal was announced in 2016 and completed in 2017.

Figura 4.2

Il valore delle operazioni di M&A che hanno coinvolto imprese *high tech* è guidato da poche transazioni da bilioni e bilioni di dollari. Nel 2016, ad esempio, 91 operazioni, ciascuna con un valore pari a circa 1 bilione di dollari, ha generato approssimativamente l'80% del valore totale del mercato M&A che coinvolgeva imprese *high tech*. Le due più grandi trattative, ovvero l'acquisizione di ARM Holdings da parte di SoftBank Group, con un valore pari a 31.6 bilioni di dollari e l'acquisizione di LinkedIn da parte di Microsoft, con un valore pari a 26.2 bilioni, costituiscono, invece, l'8% del mercato.¹⁴⁶ Allo stesso tempo, c'è anche un mercato molto attivo di piccole acquisizioni. Più dell'80% del valore del mercato che coinvolge l'acquisizione di imprese *high tech*, infatti, è costituito da transazioni con valore di meno di 100 milioni di dollari. Le operazioni attraverso cui maggiormente si verificano operazioni di M&A che coinvolgono imprese *high tech* sono il *private equity* (PE), il *venture capital* (VC). Nel 2016, circa un terzo di tutte le trattative svolte da fondi di PE o VC riguardava *high tech targets*. La maggiore operazione di concentrazione, tramite *private equity*, è stata l'acquisizione da parte del fondo Apollo Global Management della piattaforma di *cloud computing* Rackspace per 4.3 bilioni di dollari, successivamente l'acquisizione da parte di Thoma Bravo di Qlik Technologies, un'impresa di software e analytics, per 3 bilioni e, per finire, l'acquisizione da parte del fondo Svedese di *private equity* EQT di Press Ganey Holdings, un'impresa di health care software, per 2.4 bilioni. Le transazioni svolte tramite *private equity*, hanno come *target* imprese che sviluppano soluzioni software, piuttosto che hardware, concentrando le proprie attenzioni sulle *mobile technologies*, sviluppo di app, social network, big data e servizi *cloud*¹⁴⁷.

¹⁴⁶ According to BCG "In the large-cap marketplace, there is a trend toward rising overall valuation levels as competition increases for large-scale, must-have assets and as more traditional buyers, backed by substantial M&A bankrolls, seek technological innovations."

¹⁴⁷ See "Cracking the Code in Private Equity Software Deals", BCG Focus, May 2017

Nel settore dei servizi finanziari, circa un quinto delle operazioni ha coinvolto imprese *high tech*. Tra le maggiori transazioni c'è l'acquisizione da parte di LendingTree di Iron Horse Holding, una piattaforma di carte di credito, per 130 milioni di dollari e l'acquisizione di Chayora Holdings, una piattaforma di data storage, da parte di Standard Chartered Bank. Nei settori di manufacturing, il 14% delle operazioni di M&A ha coinvolto imprese *high tech*, soprattutto tra imprese operative nel campo dell'automotive, in cui tale statistica sale ad un quarto. Ad esempio, General Motors ha pagato 1 bilione di dollari per l'acquisizione di Cruise Automation ed ha investito 500 milioni in Lyft, una piattaforma di trasporti on-demand.

Le *high tech targets* dei giorni nostri, sono soprattutto quelle operative nella filiera produttiva dei Big Data (Capitolo 1, Paragrafo 1.4), infatti, le attività di *data storage* e *analysis* stanno incrementando la loro importanza in tutti i settori. Così, molti incumbent, si dotano delle migliori infrastrutture e *skills* per rimanere nel mercato ed offrire sempre migliori prodotti e le nuove entranti, al contrario, una volta superate le barriere all'ingresso per la detenzione e l'analisi dei big data, possono entrare in settori con elevati tassi di crescita e diventare appetibili per le grandi imprese operanti sia nei settori digitali, sia in quelli tradizionali. I settori in cui tali operazioni di M&A sono maggiormente presenti e coinvolgono imprese operanti nell'ambito della raccolta, analisi ed elaborazione di grossi ammontari di dati sono: l'*Industry 4.0*, ovvero il manufacturing automatizzato e l'internet of things (con una crescita annua stimata di oltre il 20%), il *cloud computing*, che fornisce soluzioni basate sulla tecnologia *cloud*¹⁴⁸, il mobile tech e il software application¹⁴⁹.

4.1 High tech Concentrations

Le trattative di M&A danno alla luce operazioni di concentrazione, nel senso che cambiano, più o meno radicalmente, la struttura di un settore. Un settore, infatti, si dice maggiormente concentrato quando ha al suo interno poche imprese con elevata market share, ha ovvero un indice di concentrazione o di omogeneità molto elevato¹⁵⁰.

Quindi, a seguito di operazioni di concentrazione, c'è una riduzione del numero di imprese che operano in modo del tutto indipendente nel mercato.

È infatti il controllo, inteso come capacità di direzionare le strategie aziendali, che viene trasferito da un'entità all'altra.

¹⁴⁸ Si ricordi, in tale settore, l'acquisizione di NetSuite da parte di Oracle per 9.3 bilioni di dollari e l'acquisizione da parte del fondo di *private equity* Vista Equity Partners di GovDelivery per 153 milioni di dollari.

¹⁴⁹ According to BCG, "Among the almost 250 acquisitions in this category is Microsoft's purchase of LinkedIn, which pushed the average deal value to \$535 million. In fact, most 2016 deals were significantly smaller, such as Converse's acquisition of Acision UK Limited for \$136 million. (Acision, now known as Mavenir, was a provider of seamless mobile messaging solutions to service providers and telecom operators)."

¹⁵⁰ In economia, si utilizzano come metriche statistiche per la misurazione del grado di concentrazione di un'industria, l'indice CR4, che compara i ricavi delle 4 maggiori aziende nel settore, con i ricavi totali e l'indice di Herfindahl-Hirschman che viene ottenuto sommando le quote di mercato delle imprese del settore al quadrato.

Questi benefici che il controllo fornisce¹⁵¹ vengono monetizzati dall'impresa *target* a seguito a operazioni di M&A. Un'impresa, infatti, pagherà di più per un'altra impresa, quanto maggiormente riuscirà ad estrapolare benefici dal controllo della stessa e, d'altro canto, l'impresa *target* riuscirà a monetizzare più efficacemente la sua vendita, quanto più alti sono i benefici privati del controllo riconosciuti dall'impresa acquirente. Da un punto di vista finanziario, dunque, il controllo costituisce l'elemento che giustifica il surplus di valore dell'impresa *target*, rispetto ad altri parametri presi in considerazione (ad esempio QDM, o capitalizzazione di mercato) e i suoi benefici rappresentano i flussi di cassa che riescono a fare l'impresa *target* e l'impresa acquirente una volta fuse. Starà poi alla bravura delle due imprese, nonché alla banca d'investimento che cura la trattativa, associare un valore finanziario al possesso integrato del controllo della *target* e dei benefici, intesi come sinergie, che possono scaturire dall'operazione.

Le operazioni di concentrazione sono, dunque, molto frequenti, soprattutto nell'industria digitale.¹⁵² Esse sono quindi trainate dalla tecnologia, soprattutto quella inerente al ML e al DL, che giustifica le sinergie ricercate ed il valore finanziario dell'impresa *target*. Tali sinergie sono costituite soprattutto da economie di scala o di scopo e dal controllo esclusivo di un input o canale di distribuzione, a seconda di dove ci si colloca nella filiera produttiva. Tuttavia, tali operazioni di concentrazione, possono comportare alcuni rischi competitivi. Ovvero, a seguito di un'operazione di M&A, c'è la riduzione di pressioni competitive, che avvengono o all'interno della stessa filiera produttiva o tra segmenti più o meno differenziati; e questo può portare ad un aumento dei prezzi, riduzione delle quantità, peggiore qualità dei prodotti e minore pressione ad effettuare investimenti in R&D. Inoltre, un altro rischio è associato alla maggiore possibilità per le imprese di intraprendere strategie coordinate e pratiche concordate al fine di restringere la competizione (Art 101 TFEU), ovvero intraprendere operazioni di explicit o tacit collusion. C'è da dire, inoltre, che le operazioni di concentrazione possono portare, oltre ai suddetti rischi competitivi, anche ad efficienze, intese come sinergie trovate dalle due imprese che, si traducono in una riduzione dei costi o in un aumento dei ricavi, con possibili effetti positivi anche sui consumatori finali. Per questo motivo, le operazioni di M&A sono maggiormente tollerate rispetto alla formazione di un cartello.

I Big data, come li abbiamo studiati noi, ovvero che attraversano una catena del valore, che parte dalla loro raccolta e termina con l'estrapolazione di importanti informazioni di business, passando per la fase di analisi ed elaborazione tramite l'ausilio di algoritmi, se sono compatibili e vengono messi insieme a quelli di un'altra impresa, possono costituire una preziosa economia di scopo.

¹⁵¹ Detti benefici privati del controllo.

¹⁵² Secondo una ricerca condotta da Accenture le operazioni di M&A, in Italia, sono condotte per il 42% col fine di espandersi in nuovi mercati geografici, il 41% col fine di acquisire nuove competenze digitali, il 40% per espandersi in nuovi settori e il 38% per acquisire le tecnologie più all'avanguardia. <https://www.accenture.com/it-it/insight-tech-led-mergers-acquisitions>.

Due imprese, infatti possono raccogliere dati simili ed analizzarli mettendo in comune infrastrutture e personale umano qualificato, riducendone pertanto i costi associati e permettendo una gestione della raccolta ed una fase di analisi condotte in maniera più efficiente. In sostanza, più dati messi a disposizione, permettono all'impresa fusa di aumentare la capacità di calcolo e a ridurre il costo unitario di raccolta ed analisi associato, grazie allo sfruttamento di economie di scala e al perfezionamento delle tecniche di raccolta ed analisi, che permettono di effettuare una compressione dei dati efficace e ad eliminare i dati ridondanti che scaturiscono dalla fusione di diversi dataset simili tra loro. Tornando alle operazioni di concentrazione, esse sono disciplinate dalla Section 7 of the US Clayton Act¹⁵³; dal Regolamento Europeo 139/2004 e dall' Articolo 5 della Legge 287/90¹⁵⁴. Esse sono discipline che regolano un tipo di controllo svolto ex-ante ed è quello che rileva maggiormente, poiché è molto difficile tornare indietro alla "vecchia" struttura del settore prima della concentrazione. Con tale tipo di controllo si fa un'analisi predittiva sui possibili effetti sul mercato di tali operazioni.

La concentrazione può avvenire in quattro diversi modi: con la fusione di due o più imprese, con l'acquisizione della maggioranza delle azioni aventi diritto di voto in assemblea e quindi costituenti diritto di controllo, con l'acquisizione di un ramo d'azienda o asset strategico e con la costituzione di una joint venture tra due o più imprese. Come appare da tale elenco, è la nozione di controllo che rileva se un'operazione di questo tipo può considerarsi una concentrazione o un semplice investimento o passaggio della proprietà azionaria. Il controllo, infatti, così come definito dall'Art 3 della Merger Regulation, è la possibilità di esercitare un'influenza decisiva sulle decisioni strategiche di un'altra impresa. Non è necessario dimostrare che tale influenza sia esercitata regolarmente da una società su un'altra, ma basta solo che sia reale ed effettiva.

Il controllo può essere esercitato solo o in concomitanza con altre imprese, può essere positivo o negativo, de jure o de facto. Il cosiddetto sole control viene esercitato da una sola impresa su un'altra, riuscendone ad indirizzare le strategie e ad esercitare una decisiva influenza. Questo tipo di situazione si ha quando un'impresa acquista il 100% delle azioni, o la maggioranza dei diritti di voto o acquistando un business o una linea di business.

¹⁵³ Cfr Federal Trade Commission, the Antitrust Laws <https://www.ftc.gov/tips-advice/competition-guidance/guide-antitrust-laws/antitrust-laws>

¹⁵⁴ Secondo l'Art 5 della legge 287/90, L'operazione di concentrazione si realizza: quando due o più imprese procedono a fusione; quando uno o più soggetti in posizione di controllo di almeno un'impresa ovvero una o più imprese acquisiscono direttamente od indirettamente, sia mediante acquisto di azioni o di elementi del patrimonio, sia mediante contratto o qualsiasi altro mezzo, il controllo dell'insieme o di parti di una o più imprese; quando due o più imprese procedono, attraverso la costituzione di una nuova società, alla costituzione di un'impresa comune.

Nel joint control, invece, due o più imprese esercitano congiuntamente un'influenza decisiva su un'altra entità e questo avviene ad esempio, quando l'acquirente acquista il 50% delle azioni di controllo, lasciando il rimanente 50% al venditore oppure con clausole statutarie, come azioni con voto plurimo o quorum rafforzati in assemblea straordinaria, del venditore che gli permettono di mantenere il controllo della società, pur senza detenere la maggioranza delle azioni in assemblea.

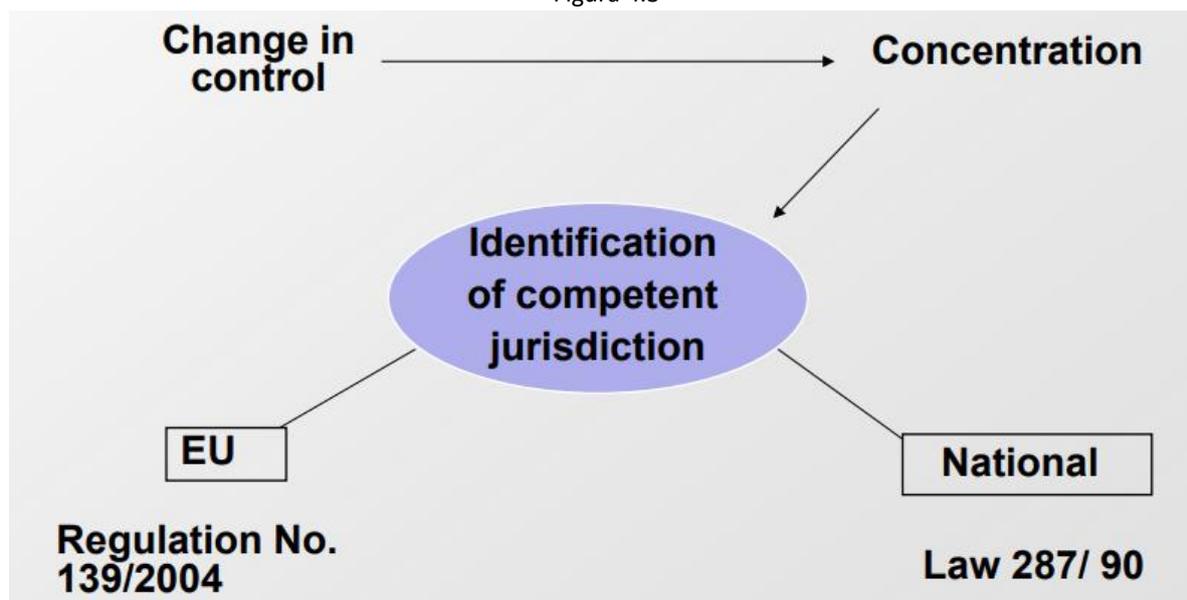
Il controllo si definisce positivo se un'impresa è in grado, individualmente, di prendere decisioni strategiche riguardanti l'altra impresa. Si definisce invece negativo, se l'impresa acquirente non è in grado di assumere decisioni per conto dell'impresa *target*, ma può bloccare le decisioni di quest'ultima, se non vengono ritenute idonee. Un controllo, infine, si definisce *de jure*, quando viene acquisito in seguito a basi legali, come appunto l'acquisto della maggioranza dei diritti di voto o di una linea di business, si definisce *de facto*, quando, nonostante l'impresa acquirente non abbia la maggioranza dei diritti di voto in assemblea e non ci sono clausole statutarie che ne rafforzano il potere di controllo, questo viene raggiunto per mezzo di altre basi, come la possibile comunanza di interessi tra acquirente ed azionista di maggioranza della *target*, patti parasociali, dipendenza strategica della *target* dall'impresa acquirente ecc.

Una Joint Venture (JV) costituisce una concentrazione se c'è la nascita di una nuova entità per mezzo dei conferimenti di altre due o più entità che mantengono, pertanto, il controllo a livello proprietario sulla nuova impresa, che, a sua volta, invece, agisce sul mercato in maniera indipendente ed autonoma ed è specializzata nel suo settore di appartenenza. In genere tali imprese sono tecnologicamente molto avanzate e permettono, pertanto, alle imprese proprietarie di effettuare investimenti congiunti in R&D, mettendo insieme *know how* e capitali. Le JV che non danno vita alla creazione di una nuova entità, danno il via all'esecuzione di contratti di coordinazione e cooperazione che, se stipulati con intento anti-competitivo, possono rientrare nella disciplina delle collusioni, costituendo una pratica abusiva secondo l'Art 101 del TFEU.

Una transazione riguardante intangible assets, come marchio, brevetti o copyrights, può anche considerarsi come una forma di concentrazione se essi rappresentano un'attività economica ed il trasferimento degli IPRs costituisce il passaggio di consegna dell'attività che scaturisce da tali immobilizzazioni immateriali.

Le concentrazioni sono soggette all'obbligo di *prior notification* alla Commissione Europea o all'ICA¹⁵⁵, a seconda del numero e tipologia di imprese coinvolte nell'operazione. Interverrà, pertanto, a seconda dei casi, la normativa comunitaria, dettata dall'Art 1 del regolamento 139/2004¹⁵⁶ se la prior notification va riportata alla Commissione, oppure la normativa nazionale, in Italia con l'Art 16 della Legge 287/1990, se va riportata all'AGCM¹⁵⁷.

Figura 4.3



¹⁵⁵ Italian Competition Authority, AGCM

¹⁵⁶ Una concentrazione è di dimensione comunitaria quando: a) il fatturato totale realizzato a livello mondiale dall'insieme delle imprese interessate è superiore a 5 miliardi di EUR e b) il fatturato totale realizzato individualmente nella Comunità da almeno due delle imprese interessate è superiore a 250 milioni di EUR; salvo che ciascuna delle imprese interessate realizzi oltre i due terzi del suo fatturato totale nella Comunità all'interno di un solo e medesimo Stato membro. 3. Una concentrazione che non supera le soglie stabilite al paragrafo 2 è tuttavia di dimensione comunitaria quando: a) il fatturato totale realizzato a livello mondiale dall'insieme delle imprese interessate è superiore a 2,5 miliardi di EUR; b) in ciascuno di almeno tre Stati membri, il fatturato totale realizzato dall'insieme delle imprese interessate è superiore a 100 milioni di EUR; c) in ciascuno di almeno tre degli Stati membri di cui alla lettera b), il fatturato totale realizzato individualmente da almeno due delle imprese interessate è superiore a 25 milioni di EUR e d) il fatturato totale realizzato individualmente nella Comunità da almeno due delle imprese interessate è superiore a 100 milioni di EUR; salvo che ciascuna delle imprese interessate realizzi oltre i due terzi del suo fatturato totale nella Comunità all'interno di un solo e medesimo Stato membro. Gianluca Faella (2018)

¹⁵⁷ "Le operazioni di concentrazione di cui all'articolo 5, devono essere preventivamente comunicate all'Autorità qualora il fatturato totale realizzato a livello nazionale dall'insieme delle imprese interessate sia superiore a quattrocentonovantadue milioni di euro e qualora il fatturato totale realizzato individualmente a livello nazionale da almeno due delle imprese interessate sia superiore a trenta milioni di euro. Tali valori sono incrementati ogni anno di un ammontare equivalente all'aumento dell'indice del deflatore dei prezzi del prodotto interno lordo." Articolo 5 della Legge 287/90.

Secondo la normativa sulle concentrazioni Europea, dettata dal Regolamento No. 4046/1989 ed in Italia dalla Legge No. 287/90 per verificare quando un'operazione di concentrazione possa costituirsi come pratica abusiva, viene fatto un test di dominanza. Ovvero si vuol verificare se, a seguito dell'operazione di M&A, si viene a creare o a rinforzare una posizione dominante all'interno del settore ed il risultato è un evidente restrizione alla competizione, con un danneggiamento di riflesso verso i consumatori. Il concetto di dominanza qui utilizzato, non è differente da quello introdotto dall'Art 102 del TFEU, ovvero quello di agire in modo totalmente indipendente dai competitors, clienti e consumatori. Tale dominanza può, inoltre, essere individuale, con la creazione o il rafforzamento del PDM di un'entità o collettiva, risultante da una variazione della struttura del settore, che può avvantaggiare anche altri players estranei all'operazione di M&A. Secondo la legge antitrust Statunitense, il Clayton Act, viene adottato un diverso standard nella definizione di condotta abusiva a seguito di un'operazione di concentrazione, detto Substantive Test. Secondo tale approccio, l'autorità deve verificare se c'è una sostanziale riduzione della concorrenza a seguito dell'operazione. Essa parte dal presupposto che ci si trova in mercati oligopolistici in cui nessuna impresa è dominante e non ci sono i presupposti strutturali per ingaggiare operazioni di tacit collusion.

Il Regolamento Europeo No. 139/2004 ha introdotto, nell'Art 2, paragrafi 2 e 3¹⁵⁸ un nuovo test, più simile allo standard USA. In tale test l'operazione di concentrazione non deve significativamente impedire l'effettiva competizione basata sui meriti delle imprese¹⁵⁹. La concentrazione può impedire l'effettiva competizione basata sui meriti in tre modi: creando o rinforzando una posizione dominante di una singola impresa (*individual dominance*); rimuovendo un importante vincolo competitivo, riducendo il grado di competizione nei mercati oligopolisti (*non-collusive oligopolies*) e favorendo intese e pratiche concordate, cambiando la struttura dell'industria e rendendo tali pratiche più raggiungibili alle imprese (*collusive oligopolies*)

Oltre a verificare l'esistenza di effetti anti-competitivi, le autorità legittimate ad esercitare il controllo, oltre a definire il mercato rilevante e geografico, si aiutano con indici strutturali (market shares o QDM, grado di concentrazione) e un numero di fattori addizionali, con l'ausilio anche di tecniche e strumenti econometrici per stimare direttamente gli effetti sui prezzi e sulle quantità, quando si hanno sufficienti dati a disposizione.

¹⁵⁸ "Si ha controllo in presenza di diritti, contratti o altri mezzi che conferiscono, da soli o congiuntamente, e tenuto conto delle circostanze di fatto o di diritto, la possibilità di esercitare un'influenza determinante sull'attività di un'impresa; trattasi in particolare di: a) diritti di proprietà o di godimento sulla totalità o su parti del patrimonio di un'impresa; b) diritti o contratti che conferiscono un'influenza determinante sulla composizione, sulle deliberazioni o sulle decisioni degli organi di un'impresa. Il controllo è acquisito dalla persona o dall'impresa o dal gruppo di persone o di imprese: a) che siano titolari dei diritti o beneficiari dei contratti suddetti; o b) che, pur non essendo titolari di tali diritti o beneficiari di tali contratti, abbiano il potere di esercitare i diritti che ne derivano."

¹⁵⁹ "significantly impede effective competition in the common market or in a substantial part of it, in particular as a result of the creation or strengthening of a dominant position". Gianluca Faella (2018)

In particolare, un'elevata QDM (maggiore o uguale al 50%) costituisce una presunzione di dominanza, ma può essere confutato ad esempio se le imprese hanno possibilità di crescere nel breve termine e se il settore è caratterizzato da elevata dinamicità. Al contrario, una bassa QDM deve essere esaminata alla stregua di altri fattori, come barriere all'ingresso, numero e PDM dei competitors, presenza di vincoli di capacità. Infine, una QDM di meno del 25% presume l'assenza di PDM.

Gli indici di concentrazione che vengono più usati sono il CR4, che somma le QDM dei 4 maggiori players dell'industria e l'indice di Herfindahl-Hirschman (HHI) che somma i quadrati delle QDM di tutti gli operatori del settore. Essi vengono utilizzati soprattutto per le operazioni di concentrazione orizzontali e, in particolare, un HHI < 1000 non fa presumere a problemi di competizione; un HHI compreso da 1000 a 2000 non implica problemi di concorrenza, a meno che la concentrazione coinvolge potenziali o nuovi entranti, una delle parti coinvolte nell'operazione è un importante innovatore, ci sono significanti scambi di azioni tra le parti coinvolte, una delle due parti ha una QDM pre-concentrazione pari o superiore al 50%; infine con un HHI > 2000 possono sorgere seri problemi di competition.

Tuttavia, l'ausilio della QDM e di tali indici strutturali, perde di valore in molti settori *high tech* poiché: l'elevato tasso di innovazione indebolisce il *link* tra struttura di mercato corrente e futura; è necessario considerare anche la possibilità che hanno i nuovi entranti di vincere la competizione con un'innovazione in mano (breakthrough innovation), quando più servizi vengono forniti gratis, è più semplice per i consumatori effettuare il multihomig e passare ad altre piattaforme (competition is one click away)¹⁶⁰; infine avere una piattaforma con più lati, vuol dire avere interazioni competitive con più clienti e tra più clienti della piattaforma. Il caso Microsoft/Skype è emblematico in tal senso, infatti, sia la Commissione che la Corte di Giustizia Europea, hanno considerato le caratteristiche e, in particolare, la dinamicità dell'industria nell'analisi di questo caso. Infatti, sebbene la fusione delle imprese portava ad una QDM congiunta dell'80/90% nel settore della video-comunicazione su Internet, per PC compatibili con Windows, la Corte affermava che questo era un settore caratterizzato da cicli innovativi molto corti, elevata crescita, con la presenza di QDM rilevanti ma che non sono garantite nel futuro e quindi non sono indicative di PDM. Inoltre, poiché il servizio veniva reso gratuito, qualsiasi aumento di prezzo da parte di Microsoft, avrebbe indotto i consumatori a passare ad altre piattaforme di video-comunicazione.¹⁶¹

Nei mercati *high tech*, in sostanza, c'è bisogno di un approccio più dinamico nel definire se un'operazione di concentrazione porti alla creazione di posizioni dominanti nel settore. Bisogna pertanto verificare se: ci sono barriere all'ingresso che ostacolano il passaggio dei consumatori ad altre piattaforme nel caso di un aumento di prezzi o riduzione della qualità dell'impresa risultante dall'operazione; l'impresa fusa controlla un asset che è di per sé in

¹⁶⁰ Cfr Capitolo 2, pagina 6

¹⁶¹ Cfr Case No COMP/M.6281 - MICROSOFT/ SKYPE f

grado di precludere la competizione; e se i competitors posseggono le *skills* e le capacità per sviluppare un'innovazione di successo. Tuttavia, identificare gli attuali e i potenziali innovatori e misurare la loro abilità ad innovare potrebbe risultare estremamente difficile. Gli effetti anti-competitivi, dunque, possono essere di due tipi: effetti non coordinati, o unilaterali, ed effetti coordinati. I primi rappresentano gli effetti più diretti che scaturiscono da operazioni di concentrazione, e si riferiscono alla riduzione del grado di competizione nel settore, che porta una o più imprese ad avere un maggiore PDM. Sia le imprese coinvolte nell'operazione di M&A, sia le altre entità presenti nell'industria possono beneficiare di un rafforzamento del PDM ed applicare, conseguentemente, prezzi più elevati. Oppure le imprese estranee all'operazione, possono beneficiare di un aumento della domanda provocata dall'aumento del prezzo della nuova entità conseguente all'operazione di fusione. È la struttura del settore poi, intesa come numero di imprese e loro QDM, che amplifica o riduce la portata applicativa di tali effetti che scaturiscono da operazioni di M&A. inoltre, anche il grado di sostituibilità gioca un ruolo fondamentale. Infatti, maggiore è il grado di sostituibilità tra i prodotti delle imprese fuse, più alti saranno gli incentivi per queste di aumentare in misura significativa i prezzi. Quindi operazioni di concentrazione che avvengono all'interno dello stesso sotto settore, possono portare ad un maggiore PDM, inteso come possibilità di porre prezzi elevati per le due imprese fuse, senza perdere QDM. Infine, in industrie caratterizzate da elevati *switching costs*, è più difficile per i consumatori cambiare prodotto, rimanendo così intrappolati all'uso dei prodotti dell'impresa fusa, nonostante un incremento dei prezzi. Una concentrazione può, inoltre, cambiare la struttura del settore e le condotte competitive, rendendo la coordinazione delle attività tra diverse imprese una strategia percorribile e stabile nel tempo. Causa, dunque, un aumento del rischio che le imprese adottino strategie di collusione tacita, aumentando congiuntamente i prezzi o riducendo gli output, falsando o restringendo la competizione basata sui meriti. La teoria economica, tuttavia, non afferma l'esistenza di un legame diretto tra concentrazione e pratiche concordate, sono bensì le caratteristiche strutturali dell'industria che possono rendere meno complesso e più stabile l'ambiente economico, favorendo la nascita di accordi collusivi. Secondo la commissione europea, tre sono le condizioni che possono rendere più stabile la coordinazione delle strategie di due o più imprese: un alto grado di trasparenza nel mercato, l'esistenza di mezzi punitivi adeguati e credibili contro possibili deviazioni e l'assenza di forze competitive che provengono dall'esterno del cartello che ne possono ridurre la portata applicativa.

Nei settori *high tech*, nel definire se una concentrazione possa essere dannosa o meno per il grado di competizione esistente nel settore, bisogna analizzare attentamente il tipo di imprese che partecipa all'operazione ed il loro grado di propensione all'innovazione. Nell'*high tech*, infatti, sia imprese grandi che imprese piccole fanno investimenti in R&D e può capitare che le piccole, sebbene non dispongano delle risorse finanziarie delle grandi, riescono a sviluppare importanti innovazioni che inducono le imprese più grandi a fare ingenti esborsi di capitale al fine di acquisirle.

“One of the simplest defences against innovation is to buy up rivals that create innovative products. That’s why, when we look at high-tech mergers, we don’t just look at whether they might raise prices. We also assess whether they could be bad for innovation” (Commissioner Vestager, 18 April 2016)

Quindi lo scenario possibile che può venire a delinearsi è che molte imprese utilizzano le proprie *skills* e capacità per sviluppare una tecnologia da rivendere poi al miglior acquirente. Oppure che le grandi imprese utilizzano le ingenti disponibilità di capitali di cui dispongono per fare acquisizioni tecnologiche, piuttosto che sviluppare innovazioni al loro interno.

La relazione tra concentrazione e innovazione, come visto nel Capitolo 3, è molto labile e poco chiara. Un’elevata concentrazione può permettere alle imprese fuse di catturare maggiori benefici dall’innovazione (che quindi avrà un ROI più elevato) ed avere un incentivo ad innovare maggiore; la concentrazione inoltre può coinvolgere imprese che posseggono asset complementari, così da incrementare l’abilità ad innovare delle imprese fuse; d’altro canto, una competizione meno intensa può ridurre la pressione ad innovare ed inoltre, c’è il rischio che, affrontando poca competizione, le imprese fuse possano sviluppare innovazioni che cannibalizzano i loro stessi profitti. Non esiste, dunque, una presunzione generale che spiega il rapporto tra innovazione e concentrazione nei settori *high tech*.

Secondo la Commissione Europea, ci sono altri fattori che possono aumentare il PDM congiunto di più imprese fuse: il *buyer power*, il tasso di ingresso di nuovi competitors e le efficienze, intese come sinergie trovate dalle imprese coinvolte nell’operazione. Il *buyer power* rappresenta la forza contrattuale dei clienti nelle negoziazioni commerciali dovuta a dimensione, importanza commerciale per il venditore ed abilità di trovare e passare a fornitori alternativi per il cliente. Secondo l’EU Horizontal Merger Guidelines¹⁶², nelle *multi sided platforms*, non è sufficiente che solo un particolare tipo di segmento di clienti non viene inficiato dall’aumento dei prezzi e dall’esplotamento del PDM, bisogna andare a vedere qual è l’effetto per gli altri side della piattaforma ed andare a stimare un effetto generale. Il tasso di ingresso, o entry rate, rappresenta un importante vincolo competitivo per le imprese fuse se l’ingresso è probabile, analizzando l’entità delle barriere all’ingresso, tempestivo¹⁶³ e se è in grado di rimuovere e sconfiggere gli effetti anti-competitivi derivanti dall’operazione di M&A. le efficienze che derivano dalle operazioni di concentrazione sono effetti positivi, come la razionalizzazione delle risorse, economie di scala, economie di scopo, progresso tecnologico, migliori condizioni di acquisto, miglior accesso al capitale e riduzione delle inefficienze.

¹⁶² See <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A52004XC0205%2802%29>

¹⁶³ Normalmente è necessario che l’ingresso avvenga entro 2 anni

In assenza di tali efficienze è difficile giustificare un approccio più “morbido” nei confronti delle operazioni di concentrazione rispetto al trattamento dei cartelli.¹⁶⁴

C'è dunque un bisogno di controbilanciare gli effetti anti-competitivi, determinati *ex-ante* e le possibili efficienze derivanti da tali transazioni di M&A.

Nei primi periodi dalla nascita dell'EU Merger Regulation¹⁶⁵, la Commissione sembrava ignorare tali efficienze e applicava di caso in caso il dominance test affermando che l'intento delle imprese era quello di sfruttare le efficienze per estendere la QDM dell'impresa fusa e di conseguenza diventare un'entità dominante. L'EU Horizontal Merger Guidelines, prendendo spunto dall' US Horizontal Merger Guidelines, ha iniziato a guardare alle efficienze in ottica diversa, soprattutto andando a vedere il rapporto tra queste e propensione all'innovazione. Infatti, se tali sinergie portavano le imprese fuse a sviluppare una più efficiente ed efficace attività di R&D, sviluppando importanti innovazioni di prodotto o di processo gli effetti pro-competitivi potevano superare quelli anti-competitivi risultanti dall'operazione. Le efficienze possono, pertanto, compensare i consumatori degli effetti negativi causati dall'aumento del PDM. Nel grafico seguente, la concentrazione non dovrebbe essere approvata se A (l'aumento del surplus del produttore) è minore di B (la deadweight loss), al contrario, dovrebbe essere approvata solo nel caso in cui i risparmi di costo farebbero sì che vi sia anche un abbassamento di prezzo, ovvero se e solo se P2 fosse minore o uguale di P1.

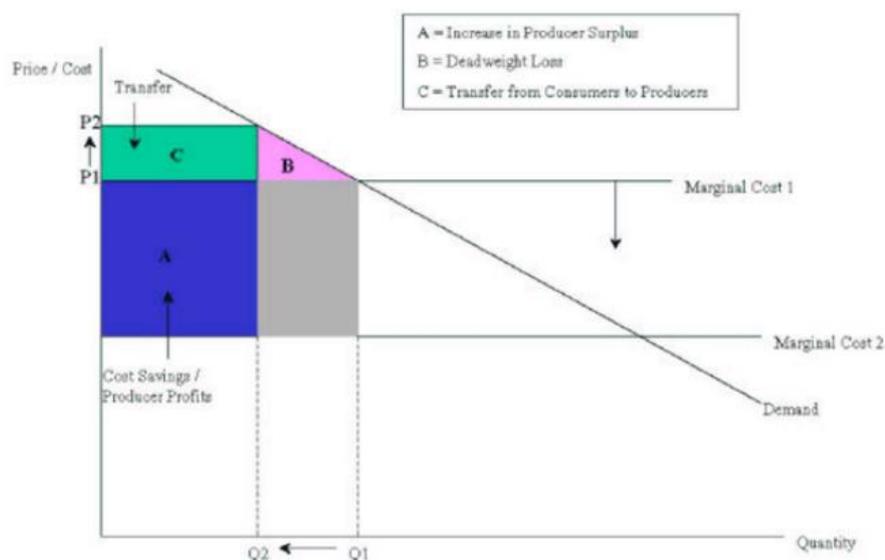


Figura 4.4

Secondo l'EU Horizontal Merger Guidelines, dunque, le efficienze possono controbilanciare gli effetti anti-competitivi se apportano benefici ai consumatori, se sono una diretta conseguenza dell'operazione di M&A e se sono verificabili, ovvero la Commissione deve essere ragionevolmente certa che tali efficienze sono abbastanza consistenti da neutralizzare i potenziali danni derivanti da operazioni di concentrazione.

¹⁶⁴ Cfr Art 101 TFEU

¹⁶⁵ See the paper <https://www.slaughterandmay.com/media/64572/the-eu-merger-regulation.pdf>

Bisogna cioè bilanciare effetti anti-competitivi ed effetti che favoriscono l'innovazione, entrando in un trade off tra inefficienze statiche (alti prezzi e bassi output) di breve periodo ed efficienze dinamiche (innovazioni) di lungo periodo e vedere quali di queste prevalgono, di caso in caso.

4.2 Big Data & M&A

I Big Data, come detto, rappresentano un'economia di scopo che spinge le imprese digitali ad effettuare operazioni di M&A. Le imprese possono, dunque, o mettere insieme i propri dati col fine di ricavare informazioni più dettagliate sui consumatori e per sfruttare le economie di scala, o internalizzare una fase della catena del valore dei Big Data. A seconda della strategia di crescita adottata, avremo una concentrazione orizzontale o conglomerale nel primo caso, una concentrazione verticale nel secondo caso. Sebbene i dati possono rappresentare una sinergia, mettere insieme diversi dataset può risultare complesso e può comportare all'ottenimento di una grossa mole di dati più o meno ridondanti¹⁶⁶. Tuttavia, con un'efficace operazione di pulizia dei dataset, è possibile eliminare le inefficienze in modo più rapido e performante rispetto alla gestione separata di due dataset ed è possibile, dunque, lavorare e concentrarsi su meno dati di quelli che si avrebbero ragionando in modo distinto per ogni dataset. La natura dei diversi dataset, inoltre, ci dirà se l'operazione di concentrazione in questione può definirsi un'integrazione orizzontale o conglomerale, a seconda se la distanza tra i settori o sotto settori sia nitida o molto marcata. Ad esempio, i dataset raccolti da Amazon ed Ebay possono condurre a simili informazioni circa le abitudini di consumo di un cluster di individui che effettua ricerche su entrambe le piattaforme di e-commerce. Molti dati, dunque, sono ridondanti, ma, eliminando le inefficienze, si può giungere ad informazioni più dettagliate, come ad esempio la disponibilità a pagare o verificare quali sono i prodotti che ruotano di più per categoria di merci ricercate. Invece, piattaforme come Ebay e Uber processano dati completamente diversi che, uniti, non permettono di aumentare il valore delle informazioni raccolte, anzi necessitano di differenti infrastrutture e *skills* specializzate per l'uno e per l'altro dataset. Possibili efficienze date da tale tipo di integrazione conglomerale, possono derivare da una razionalizzazione dei fattori produttivi e dalla condivisione di infrastrutture, skills, personale umano e *know how*, che possono permettere alle imprese di crescere e sviluppare nuove soluzioni tecnologiche.

Nel caso di concentrazione orizzontale, dunque, i settori non sono tanto distanti tra loro per tipologia di mercato servito e le imprese coinvolte competono tra loro in termini di QDM, tempo speso dai consumatori sulla piattaforma e modelli di business in generale.

¹⁶⁶ Cfr Capitolo 1, Paragrafo 1.2.2

Un'integrazione orizzontale, in tal caso, oltre a comportare l'eliminazione di un vincolo concorrenziale, dato da un competitor in meno, con conseguente rischio di aumento di PDM e rischio di pratiche collusive ¹⁶⁷, spinge le imprese ad effettuare una gestione della fase di *data storage* più efficiente, eliminando i dati ridondanti, attraverso operazioni di compressione e filtraggio dati e ad aumentare il valore delle informazioni ricercate con l'obiettivo di rendere maggiormente efficace la personalizzazione o clusterizzazione dei clienti o consumatori finali.

ES. CONCENTRAZIONE ORIZZONTALE (Figura 4.5)

Settore: e-commerce

EBAY

AMAZON

Storage
Analysis

Storage & Analysis

Storage
Analysis

Le due imprese, possono, dunque, adottare una strategia di accentrimento delle fasi di storage e analytics, ottimizzando e razionalizzando le risorse a disposizione e ragionando su dataset che includono dati provenienti da entrambi le piattaforme, così da eliminare i dati ridondanti e lavorare con meno dati che permettono di giungere ad informazioni dettagliate sui consumatori, oppure decidere di lasciare la raccolta e l'analisi dei dataset distinta per ogni piattaforma, condividendo però infrastrutture, *skills* e *know how*, così da giungere allo stesso livello di informazioni, con una gestione decentrata dei dataset, che richiede una maggiore complessità da gestire. La scelta su quale strategia adottare, se accentrare o decentrare entrambe le fasi di raccolta e analisi, oppure o l'una o l'altra, dipende da qual è il grado di distanza settoriale tra le imprese, ovvero dipende se le imprese competono o meno nello stesso mercato rilevante. Si avrà infatti integrazione conglomerale se i settori e i rispettivi dataset raccolti sono estremamente differenziati tra loro e l'operazione di concentrazione in tal senso non viene svolta con intento anti competitivo, bensì l'obiettivo appare essere solo la crescita inorganica dell'impresa in questione. In tal caso, infatti, si tenderà a decentrare le fasi di *data storage* e *analysis* e le sinergie che vengono sfruttate sono solo quelle inerenti alla condivisione di risorse tecniche e personali ed avere un *know how* comune ad entrambe le piattaforme. In generale, nel caso di concentrazione conglomerale, non dovrebbero sorgere problemi di competition, tuttavia le piattaforme coinvolte possono produrre beni complementari. Ad

¹⁶⁷ Cfr Paragrafo 3.1

esempio un'impresa che produce sistemi operativi (OS) per pc può differenziarsi entrando nel settore dei browser di Internet con un'acquisizione. In tal caso l'impresa può sfruttare il PDM che possiede nel settore degli OS nell'altro settore, ad esempio può offrire un prodotto ad un prezzo molto basso e coprire le perdite con i profitti nell'altro settore oppure, e questo è più inerente ai settori digitali, adottare operazioni di "tying" e "bundling", senza dare la possibilità al consumatore di acquisire i prodotti separatamente, sfruttando così il PDM di un settore, quello dei sistemi operativi nell'esempio, per espandere la QDM dell'altro settore dei browser. Tale pratica è stata adottata da Microsoft che, per un periodo offriva il proprio OS, Windows, unitamente e integralmente al suo browser Explorer, senza dare la possibilità ai consumatori né di disinstallare il browser predefinito, né di acquisire il sistema operativo senza browser ed, unitamente, vietava ai produttori di PC (OEM¹⁶⁸) di preinstallare altri browsers predefiniti¹⁶⁹.

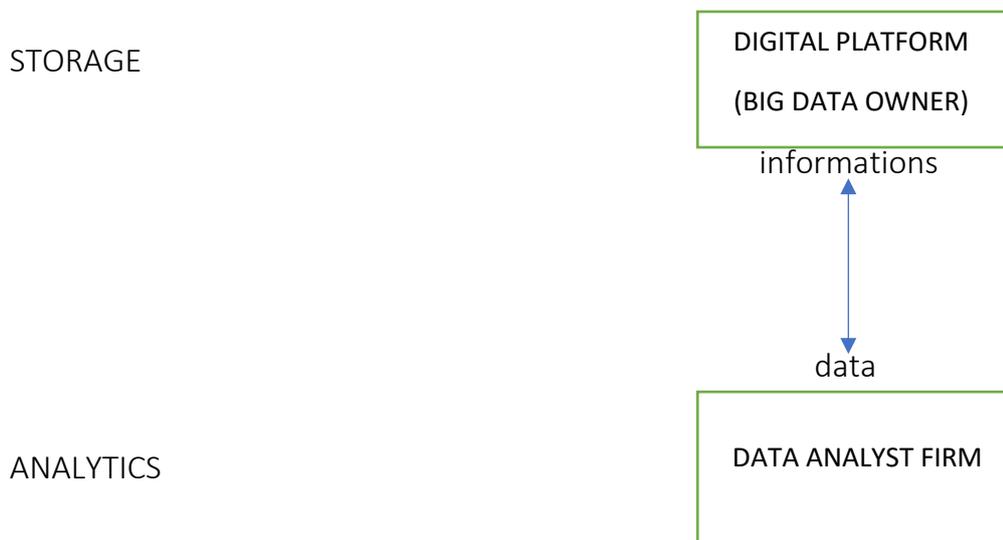
¹⁶⁸ "Original Equipment Manufacturing"

¹⁶⁹ An antitrust investigation started in 1994 had terminated in 1995 with a consent decree, which had prohibited contractual tying, but had explicitly allowed technological bundling. On October 20, 1997, the DOJ alleged that Microsoft violated the 1995 consent decree by bundling IE and Windows. On December 11, 1997, Judge Thomas Penfield Jackson issued a preliminary injunction barring the bundling. On June 23, 1998, the D.C. Circuit Court of Appeals ruled that the 1995 consent decree did not apply to Windows 98, which was supplied with an integrated IE and an IE icon on the desktop, and argued that courts are «ill-equipped to evaluate the benefits of high-tech product design». On May 18, 1998, the DOJ, the Attorneys General of 20 States and the District of Columbia filed a new antitrust suit against Microsoft. In order to overcome the Court of Appeal's interpretation of the 1995 consent decree, the DOJ argued that Microsoft did not simply add functionalities to Windows: Microsoft marginalized Netscape because it represented a potential competitive threat for Windows OS. Microsoft argued, inter alia, that: it was legal to add new functionalities to Windows, as stated also by the 1998 decision of the D.C. Circuit Court of Appeals; it was only competing hard against Netscape, and this competition was welfare-enhancing; it did not have market power in the OS market; competition in the software market was intense and its leading position could be replaced at any time by a competitor or new entrant, its practices enhanced the innovation process rather than limited it, consumers had not been harmed, as they had benefited from low price of the OS, zero price of the browsers, and an acceleration of the innovation process. In the "Findings of Fact" (November 1999) and "Conclusions of Law" (April 2000), the District Court (Judge Thomas Penfield Jackson) held that: the relevant antitrust market was the PC OS market for Intel-compatible computers; Microsoft had a monopoly in that market where it enjoyed a "large and stable market share"; Microsoft's monopoly was protected by the «applications barrier to entry»; Microsoft used its monopoly power in PC OSs to exclude rivals. In particular, it (i) engaged in per se illegal tie-in, (ii) attempted to monopolize the browser market and (iii) illegally maintained a monopoly in OSs; its actions hindered the innovation process and harmed consumers; although various contracts had anticompetitive implications, Microsoft was not guilty of anticompetitive exclusive dealing under Section 1 of the Sherman Act. Microsoft and the DOJ negotiated a settlement under the supervision of Judge Posner, but the negotiations failed because of the opposition of some States. The DOJ then asked for a breakup of Microsoft v Without holding an adequate evidentiary hearing, Judge Jackson ordered the breakup of Microsoft into two "Baby Bills", an OS company and an applications company, together with a number of interim conduct restrictions. Gianluca Faella (2018)

La concentrazione verticale rappresenta l'integrazione tra due imprese che si collocano sulla stessa catena del valore, con attività complementari, ma distinte tra loro. In sostanza le imprese operano su livelli diversi della stessa catena del valore. Tale catena del valore, nel caso nostro, coincide con quella dei Big Data¹⁷⁰, in quanto tali imprese utilizzano i dati come core asset dei propri processi produttivi.

Le imprese, quindi, possono internalizzare le attività di *data storage* o *data analytics*, che precedentemente erano affidate a terzi, passando da forme di partnership commerciali all'inserimento di attività della stessa filiera produttiva all'interno dei propri processi. Un fenomeno molto frequente a cui si assiste ai giorni d'oggi, è l'acquisizione di imprese esperte nel campo dell'analytics che sono in grado di estrapolare importanti informazioni di business, attraverso ad esempio lo sviluppo di potenti algoritmi che permettono di processare elevati ammontari di dati provenienti dalla piattaforma acquirente.

ES. CONCENTRAZIONE VERTICALE (Figura 4.6)



L'integrazione verticale può comportare rischi dal punto di vista competitivo. In particolare, un'impresa può impedire l'accesso al mercato ad un'altra, ad esempio, limitando l'accesso ai competitors ad una materia prima o componente (input foreclosure) o ad un canale distributivo (consumer foreclosure), a seconda se l'impresa si trova a monte o a valle della filiera produttiva, rappresentata in tal caso dalla catena del valore dei Big Data. In questo caso, l'analisi competitiva si svolge ex-ante sulla struttura del settore e sulle condotte delle imprese che possono avvenire post-transazione. In particolare, le autorità competenti vogliono verificare se i competitors hanno accesso a sufficienti fornitori o clienti alternativi e se l'operazione di concentrazione, prontamente notificata, può cambiare gli accordi commerciali tra le parti o facilitare la collusione tra competitors.

¹⁷⁰ Cfr Capitolo 1, Paragrafo 1.4

Un altro strumento utilizzato dalle imprese per esercitare controllo, in questo caso in modo indiretto e congiunto, è costituito dalla creazione di una Joint Venture (JV)¹⁷¹. Una JV è un'impresa comune o società cooperativa, fondata con l'intento di coordinare condotte competitive di due imprese. Essa potrebbe rientrare nell'Art 101 TFEU e rappresentare, pertanto, una pratica concordata anticompetitiva. Tuttavia, nel caso di JV, si ha la creazione di una vera e propria nuova entità, che ha una sua autonomia, con quote partecipative di proprietà di altre imprese. I requisiti per cui una JV implica un'operazione di concentrazione sono che la società frutto della Joint Venture, sia controllata congiuntamente da due o più imprese indipendenti e che l'impresa creata con la Joint Venture operi sul mercato come un ente autonomo, ossia che eserciti stabilmente tutte le funzioni di un'entità economica autonoma. Nel caso dei Big Data, tale strumento viene spesso usato nell'ambito delle operazioni di R&D, affidando alla nuova entità, costituita ad hoc, la gestione di una delle fasi della catena del valore. In tal modo si consente alle imprese madri di beneficiare di importanti vantaggi competitivi ottenuti mediante il raggiungimento di economie di scala o di scopo, la promozione di processi di integrazione verticale od orizzontale e la riduzione dei rischi connessi agli investimenti. Le autorità Antitrust devono valutare caso per caso se i benefici sociali, derivanti da tali forme di cooperazione e coordinazione, in chiave soprattutto tecnologica, superano i costi sociali associati ad un aumento della concentrazione nell'industria. A seconda dei casi, poi, si rientra o no nell'Art 101, altrimenti la fondazione di una JV è lecita e non rappresenta una forma di pratica anticompetitiva.

4.3 Facebook/WhatsApp concentration case

La fusione di WhatsApp con Facebook è stata esaminata dalla Commissione Europea, a seguito della notifica preventiva, inviata da Facebook alla Commissione, in conformità a quanto previsto dall'Art 4 della Merger Regulation¹⁷².

Facebook è una *multi-sided platform* che fornisce i servizi di social networking, consumer communication e photo/video sharing. Essa è inoltre uno dei principali player attivo nella pubblicità on-line e, in particolare, nella *display (o non search) advertising*. I primi tre servizi, che vengono rispettivamente offerti da "Facebook", "Facebook Messenger" e "Instagram", sono gratuiti e vengono monetizzati grazie alla *display advertising* posta da Facebook.

¹⁷¹ Le Joint Ventures si possono classificare a seconda dello scopo per le quali sono poste in essere. In particolare, si distingue tra: a) Joint Ventures strumentali, le quali nascono con la sola finalità di eseguire contratti complessi, di cui un esempio è costituito dalle ATI(7); e b) Joint Ventures operative costituite al fine di esercitare effettivamente un'attività autonoma, ma collaegata a quella delle entità partecipanti.

¹⁷² Le concentrazioni di dimensione comunitaria di cui al presente regolamento sono notificate alla Commissione prima della loro realizzazione e dopo la conclusione dell'accordo, la comunicazione dell'offerta d'acquisto o di scambio o l'acquisizione di una partecipazione di controllo...

Questa, infatti, è molto incisiva poiché presenta un elevato grado di personalizzazione, ottenuto grazie all'ingente ammontare di dati raccolti ed analizzati da Facebook ogni giorno.

WhatsApp è, invece, una piattaforma di consumer communication che permette agli utenti di scambiarsi messaggi istantanei, in maniera efficace e gratuita. Essa, prima di fondersi con Facebook, era disponibile solo come app per smartphones (WhatsApp web è infatti nato il 21 Gennaio del 2015, successivamente all'operazione di fusione dell'entità con Facebook).

L'operazione è costituita dall'acquisizione di WhatsApp ad opera di Facebook, per un ammontare di 19 bilioni di dollari. Il risultato della transazione è, quindi, l'ottenimento di Facebook di un sole control su WhatsApp, costituendo dunque un'operazione di concentrazione ai sensi dell'Art 3 della Merger Regulation.

L'operazione riguarda unicamente il settore del *consumer communication services*, costituito da players che forniscono soluzioni di comunicazione multimediale istantanea, che permette agli utenti di raggiungere amici, familiari e altri contatti in tempo reale.

Storicamente, tale servizio veniva offerto tramite PC. Tuttavia, si è assistito gradualmente ad uno "spostamento" di tale settore sui cosiddetti smart mobile devices, ovvero smartphone e tablet. Oggi, tale settore digitale, presenta notevoli tassi di crescita ed è uno dei primi in tal senso, con un notevole sviluppo di app disponibili per smartphones e tablets. Tali piattaforme di comunicazione istantanea vengono offerte come *stand-alone* app, come ad esempio WhatsApp, Viber, Facebook Messenger e Skype, o come servizio integrato ad una piattaforma di social network, come Facebook per pc, LinkedIn o Direct di Instagram. Tali piattaforme, si differenziano sulla base di vari elementi e funzionalità, come la possibilità di avere chat di gruppo o solo *one-to-one*, le forme multimediali, come testo, messaggi audio, video chat, la possibilità di effettuare chiamate vocali e la condivisione della posizione.

Non tutte le app presentano tutte queste funzionalità. Inoltre, ci sono consumer communication apps disponibili solo in un sistema operativo¹⁷³, come FaceTime o iMessage di Apple ed altre disponibili su diversi sistemi operativi¹⁷⁴. Ad esempio, WhatsApp è disponibile su un'ampia varietà di sistemi operativi su mobile, come IOS, Android, BlackBerry, Windows Phone, Nokia Series, Likewise, Facebook Messenger è attualmente disponibile per Android, IOS, BlackBerry, Windows Phon e Nokia Series. La maggiorparte di tali app, con alcune eccezioni, ad esempio iMessage, è disponibile sia su IOS che su Android, che rappresentano, congiuntamente, l'81% del mercato per i sistemi operativi disponibili per smartphones. Quest'ultimo, dunque, è il mercato di riferimento indicato dalla Commissione, visto che prima della fusione, WhatsApp era disponibile solo su smartphones.

¹⁷³ E vengono pertanto chiamate "*proprietary apps*".

¹⁷⁴ Dette "*cross-platform apps*".

La definizione di mercato rilevante potrebbe abbracciare anche quelli che sono i metodi tradizionali di comunicazione elettronica, come chiamate vocali, SMS, MMS ed e-mails, che possono infatti essere considerati come prodotti sostituti. Tuttavia, ci sono elementi che suggeriscono che tale sostituibilità è imperfetta e che questi siano due servizi distinti. Infatti, sebbene essi soddisfino un bisogno generale comune, i servizi offerti dalle app digitali offrono un'esperienza più ricca al consumatore finale, in termini di funzionalità e contenuti multimediali. Infatti le *consumer communication apps*, forniscono funzionalità aggiuntive come, ad esempio, la possibilità di vedere quando un utente è *on-line*, quando sta scrivendo, oppure quando ha effettuato l'ultimo accesso alla piattaforma. Inoltre, le condizioni di prezzo sono differenti, in quanto le app digitali vengono offerte spesso gratuitamente e senza porre alcun prezzo per ogni messaggio inviato.

Prendendo in considerazione tali caratteristiche strutturali del macrosettore della comunicazione elettronica, la Commissione ha stabilito di restringere il cerchio e l'investigazione al mercato delle *consumer communication apps* disponibili per smartphones.¹⁷⁵ Inoltre, la Commissione affermava che il mercato rilevante, in senso geografico, era quello Europeo, se non addirittura mondiale¹⁷⁶. Infatti, sia WhatsApp che Messenger hanno elevata diffusione in Europa, mentre se l'analisi si espande a tutto il mondo si nota che le QDM sono più frammentate e che gli usi e le abitudini di una popolazione portano a fare scelte distinte da altre. Ad esempio, WhatsApp è molto diffuso in Europa, meno negli USA; LINE e WeChat sono particolarmente popolari in Asia. Tali dinamiche competitive sono il risultato delle preferenze dei consumatori, oltre che dell'ambiente regolatorio e delle condizioni di mercato.

Gli altri tre mercati rilevanti in cui Facebook opera, non sono strettamente correlati al servizio offerto da WhatsApp. Il core service di Facebook è infatti rappresentato dalla piattaforma di social network, che connette più di 1.3 bilioni di utenti di tutto il mondo. Tale settore, che costituisce un recente fenomeno costituito da una rapida crescita che parte dai primi anni del 2000, è costituito dalla nascita di un servizio digitale, che permetteva agli utenti di connettersi, comunicare, condividere e, in generale, comunicare ed esprimere la propria personalità sulla piattaforma. I social network sono, nella stragrande maggioranza, forniti gratuitamente e vengono monetizzati attraverso la pubblicità on-line. Oltre a Facebook, vi sono altre piattaforme che soddisfano, più o meno, gli stessi bisogni, come Google+, LinkedIn, MySpace, Pinterest e InterNations. WhatsApp, al contrario, non è attivo in tale mercato e non è potenzialmente un competitor di Facebook nel settore dei social network. In particolare, WhatsApp è focalizzato sul facilitare una comunicazione semplice e veloce tra utenti, senza la creazione di profili dettagliati contenenti dati come la data di nascita, sesso, lavoro svolto, relazione

¹⁷⁵ "In the light of the above, taking into account that the Parties' combined position would be attenuated in a market including traditional electronic communications services, the Commission will assess the effects of the Transaction in the narrowest relevant product market for consumer communications services, that is the market for consumer communications apps for smartphones." Case No COMP/M.7217.

¹⁷⁶ "In light of the above, the Commission considers that the relevant geographic market for the assessment of the Transaction is at least EEA-wide, if not worldwide." Case No COMP/M.7217.

sentimentale e senza la condivisione di contenuti ed informazioni a tutta la lista di amici dell'utente in questione. Ci sono anche servizi o funzionalità comuni tra social network e consumer communication services, come la possibilità di scambiare contenuti multimediali con altre persone (come messaggi, video, audio e foto). Tuttavia, l'investigazione di mercato rivela un certo numero di differenze strutturali tra i due settori. I social networks, infatti, offrono un'esperienza sociale più ricca rispetto alle *consumer communications apps*, poiché danno la possibilità all'utente di indicare i propri interessi, attività, creare album fotografici ed esprimere le proprie opinioni mediante le interazioni con i contenuti offerti dalla piattaforma e dagli altri utenti. Le funzionalità delle *consumer communications apps*, oggi giorno, sono più limitate e si focalizzano sulla fornitura di un mezzo di comunicazione rapido ed essenziale, piuttosto che sulla creazione di un'esperienza ricca sull'identità digitale di un individuo. Inoltre esse forniscono uno scambio di messaggi che avviene in modo istantaneo ed in tempo reale, mentre, tipicamente, nei social network, i messaggi intesi come commenti su un post, non hanno una risposta immediata ed in tempo reale. Inoltre tali messaggi, nei social network, si rivolgono ad una platea di persone più ampia, data da tutti gli amici digitali dell'utente, mentre nelle consumer communications apps, ci si rivolge, tipicamente, ad un contatto telefonico, oppure ad un gruppo di persone ristretto.¹⁷⁷ Il settore dei social network, infine, si distingue, non a seconda del device o del sistema operativo utilizzato, bensì dall'uso che il consumatore ne intende fare. Esistono infatti social network con scopo di intrattenimento, come Facebook e Google+ ed altri utilizzati a scopo professionale, come LinkedIn e Xing. Tuttavia, anche qui, ci sono sfumature e funzionalità condivise da entrambe le tipologie di piattaforme. In conclusione, secondo la Commissione, consumer communication apps e social network sono due mercati rilevanti distinti e WhatsApp compete, quindi, con Messenger e non con Facebook, inteso solo come piattaforma di social network.

Facebook, come detto, è attivo nel settore della pubblicità on-line, precisamente in quello della *display advertising*, sia per quanto concerne dispositivi fissi, come pc, sia nei dispositivi mobile, come smartphones e tablets.

Facebook, a tal scopo, raccoglie dati riguardanti gli utenti che interagiscono con la piattaforma e li analizza con lo scopo di indirizzare la pubblicità in base alle preferenze dei consumatori, personalizzando i contenuti pubblicitari visibili nella home page degli utenti.

¹⁷⁷ "Moreover, social networks tend to enable communication and information sharing with a wider audience than consumer communications apps, which are more personal and *targeted*. For example, postings on a social network are generally shared with all contacts of a user (unless restricted), while communication on such consumer communications apps as Facebook Messenger and WhatsApp occurs mainly on a one-to-one basis (more than [90-100]% of all messages are one-to-one). Also, even though Facebook Messenger and WhatsApp enable users to create groups, their size is relatively limited (on average in the EEA, [0-5] users for WhatsApp and [0-10] users for Facebook Messenger). By contrast, social networks enable creation of larger groups which can consist of users who do not directly know each other (for example, an event or an interest group)." Case 7217.

WhatsApp, invece, non fornisce contenuti pubblicitari e, pertanto, non raccoglie ed analizza dati personali e non sugli utenti a fini commerciali.

Inoltre i messaggi che gli utenti si inviano su WhatsApp, non sono registrati sui server di WhatsApp, bensì solo sui devices degli utenti o su piattaforme *cloud*.

La Commissione, infine, considerava l'*on-line advertising* un mercato separato dalla *off-line advertising* e, successivamente, frammentava tale settore in search e non search advertising, senza distinzione per tipo di device utilizzato.¹⁷⁸

Focalizzando l'attenzione sul settore dei *consumer communication services*, Messenger possiede circa 250-300 milioni di utenti, di cui 100-200 in Europa e WhatsApp circa 600 milioni, di cui 100-150 in Europa. Ci sono altri players che forniscono la stessa tipologia di servizio ed essi includono sia piattaforme integrate ad altre che forniscono smartphones hardware o sistemi operativi, come iMessage di Apple, BBM di BlackBerry, ChatON di Samsung, Google Hangouts di Google e Skype di Microsoft¹⁷⁹, sia altre piattaforme di messaggistica integrate a piattaforme di social network, come Twitter, Direct di Instagram e sia altre piattaforme stand-alone, come Viber, Telegram, Snapchat, LINE e Wechat. In tale contesto le *consumer communications apps* competono per fornire al consumatore la migliore esperienza comunicativa. D'altro canto, però, i consumatori sono soliti possedere più di un app di comunicazione e tale fenomeno è detto *multihoming*. Tale fenomeno, come detto nei Capitoli precedenti, abbassa, se non annulla gli *switching cost* e frena, dunque, il PDM delle piattaforme dominanti. Tuttavia, le piattaforme competono soprattutto per accaparrarsi maggior tempo dagli utenti e ci sono piattaforme utilizzate moltissimo ed altre molto poco, nonostante siano entrambe presenti sul device mobile dell'utente. Inoltre, anche la dimensione del network gioca un ruolo fondamentale, infatti, piattaforme con una base di clienti più larga, saranno valutate maggiormente dagli utenti rispetto ad altre piattaforme di nicchia. Infine, anche il prezzo dell'app gioca un ruolo fondamentale. Infatti i consumatori di communications app sono molto sensibili a variazioni di prezzo e si aspettano che tale servizio venga offerto gratuitamente.

Ci sono, tuttavia, alcune eccezioni, come ad esempio il penetration price imposto da WhatsApp in Italia, UK, Canada e USA.¹⁸⁰

¹⁷⁸ "The market investigation also supported to a large extent the existence of a further sub-segmentation of the online advertising market between search and non-search advertising. Indeed, the majority of the advertisers who took part in the market investigation considered that search and non-search ads are not substitutable as they serve different purposes (for search ads, mainly generating direct user traffic to the merchant's website, while, for non-search ads, mainly building brand awareness) and, as a result, most advertisers would not be likely to switch from one type to another in the event of a 5-10% price increase. Similarly, the majority of the competitors who took part in the market investigation submitted that search and non-search ads are not substitutable from an advertiser's point of view." Case No COMP/M.7217.

¹⁷⁹ Cfr: Case No COMP/M.6281 - MICROSOFT/ SKYPE.

¹⁸⁰ "For the majority of all other consumer communications apps, monetisation is achieved not through fees but through advertising, stickers and in-app purchases. Facebook Messenger is not currently monetised: it is funded by the monetisation of Facebook's networking platform through advertising." Case No COMP/M.7217.

La market share combinata di Messenger e WhatsApp disponibili per i sistemi operativi di IOS e Android, nel mercato Europeo è di circa 30-40%.¹⁸¹ Tuttavia, in tale settore, la metrica della market share o QDM è poco utile per stimare se un'entità è dominante oppure no. Alcuni affermano, infatti, che la miglior metrica per tale tipo di settore è il minutaggio speso in un mese dall'utente, in quanto, tale misura riesce a stabilire l'importanza dell'app per il consumatore e il suo potenziale valore attraverso la monetizzazione indiretta tramite pubblicità. Tuttavia, la Commissione, sulla base dei risultati dell'investigazione di mercato, non considera questa metrica particolarmente significativa per analizzare il posizionamento di mercato in tale settore e considera, dunque, la QDM come miglior rappresentante del posizionamento delle imprese in tale mercato. Tuttavia, sappiamo come in tali settori caratterizzati da cicli di innovazione brevi e disruptive innovations che possono cambiare da un momento all'altro gli scenari competitivi, un'elevata QDM non è necessariamente indicativa di presenza di un elevato PDM. La Commissione ha inoltre notato che ci sono alcune differenze sostanziali tra Facebook Messenger e WhatsApp, riferite a: modi per identificare l'accesso degli utenti alla piattaforma (numeri di telefono per WhatsApp e Facebook ID per Messenger); contatti (numeri di telefono per WhatsApp e lista di amici per Messenger); *user experience* (più ricca in Messenger rispetto a WhatsApp); *privacy policy* (contrariamente a WhatsApp, Messenger raccoglie dati e li fornisce a Facebook a scopi pubblicitari) e l'intensità di utilizzo delle app (in termini di minuti spesi). Gli unici fattori che, pertanto, rendono Facebook Messenger e WhatsApp strettamente *competitors* sono le funzionalità comunicative (ovvero il core business delle due entità) e la dimensione dei rispettivi networks. Tuttavia, la Commissione ha notato che non esistono funzionalità offerte da Messenger e WhatsApp che non sono anche offerte da altri operatori del settore. Infine, come già ripetuto più volte, il multihoming rende tale competizione più labile, in quanto, Messenger e WhatsApp rappresentano le due principali piattaforme di *communications services* per la maggiorparte degli utenti Europei. Ciò suggerisce che tali piattaforme sono complementari piuttosto che in diretta competizione tra loro.

In conclusione la Commissione non considerava Messenger e WhatsApp come strettamente competitors nel settore delle *communications app*.

la Commissione ha anche stabilito che l'operazione di M&A non causa un aumento degli *switching costs*; infatti nessuna delle due parti ha il controllo su un sistema operativo per smartphone e quindi nessuno di loro può rendere più complicato passare ad altre piattaforme e non sono nemmeno preinstallate su un'ampia base di dispositivi.

¹⁸¹ "WhatsApp: [20-30]%; Facebook Messenger: [10-20]%, followed by Android's messaging platform ([5-10]%), Skype ([5-10]%), Twitter ([5-10]%), Google Hangouts ([5-10]%), iMessage ([5-10]%), Viber ([5-10]%), Snapchat ([0-5]%) and other market players with a share of [0-5]% or less." Case No COMP/M.7217.

Secondo la Commissione, quindi, non ci sono significative barriere che impediscono ai consumatori di spostarsi verso altre app in tale settore¹⁸². Non ci sono inoltre nemmeno elevate barriere all'ingresso, in quanto il settore delle *communications apps* è dinamico ed in forte crescita¹⁸³ ed inoltre, sviluppare e lanciare una nuova app di comunicazione non richiede un significativo ammontare di tempo ed investimento. Il costo principale è infatti rappresentato dai server di raccolta dati che cresce al crescere del network di utenti. Inoltre non ci sono né brevetti, né *know how* né IPR che costituiscono barriere all'ingresso e le tecnologie implementate nelle *communications apps* sono molto standardizzate. Così la Commissione considerava che tale operazione di M&A non aveva l'effetto di aumentare le barriere all'ingresso e di impedire, di fatto, la competizione nel settore.

L'unica barriera all'ingresso e all'espansione riconosciuta dalla Commissione sembra essere il *Network Effect*. Le due piattaforme, infatti, posseggono congiuntamente circa un bilione di utenti. L'esistenza del *Network Effect* non implica necessariamente problemi competitivi in mercati caratterizzati da operazioni di concentrazione.

Tuttavia è possibile che l'entità fusa possa ostacolare la competizione e rendere più difficile l'ingresso o l'espansione nel settore per i *competitors* potenziali ed attuali.

Tuttavia, in settori come quello delle *communications apps*, caratterizzati da elevata dinamicità e bassi *switching costs*, la posizione di mercato rafforzata dal *Network Effect* non è affatto garantita nel futuro.

Inoltre, come già accennato né Facebook, né WhatsApp sono preinstallati su una larga base di *devices*, cosicché non vi è alcun effetto lock-in che intrappola gli utenti all'utilizzo della piattaforma.

Quindi il *Network Effect* esiste nel settore delle *communications apps*, ma la sua presenza non implica, per la Commissione, impedimento e restringimento della competizione. La Commissione affermava che l'operazione di concentrazione tra Facebook e WhatsApp avrebbe portato allo sfruttamento del *Network Effect* con fini anticompetitivi, solo nel caso in cui le due piattaforme avessero unito i propri networks creando un'enorme base di utenti.

Tuttavia questo sarebbe stato possibile solo se i due servizi si integrassero in uno solo, con la creazione di un'unica piattaforma.

Tale integrazione, secondo la Commissione, richiedeva che i profili degli utenti di WhatsApp e quelli di Facebook, diventassero la stessa cosa. Tuttavia questo non è conforme a quelle che sono le prerogative per l'iscrizione alle due piattaforme: numero di telefono per WhatsApp e creazione di un profilo personale per Facebook.

¹⁸² "Facebook estimates that its social networking app is pre-installed in only approximately 10% of the smartphone devices sold in the EEA in 2014, while WhatsApp has only an agreement with Nokia for the pre-installation of its app on some of Nokia's models via a soft-launcher (that is, a pre-installed icon which, when tapped, redirects the user to the app store to download the application)." Case No COMP/M.7217.

Le piattaforme infatti non sono in grado di associare il Facebook ID e il numero di telefono di WhatsApp e viceversa. Quindi tale associazione degli accounts doveva essere fatta manualmente dagli utenti, che, in maniera del tutto volontaria, inserivano il numero di telefono di WhatsApp associandolo al proprio account di Facebook. In conclusione, dunque, la transazione non comporta, secondo la Commissione, problemi strutturali sul livello competitivo dell'industria grazie al multihoming e alla dinamicità della stessa e, soprattutto, non si può parlare di *Network Effect* congiunto, in quanto è impossibile per nessuna delle due piattaforme associare due account dello stesso utente e per l'esistenza di una vasta mole di consumatori che utilizzava già entrambe le piattaforme, senza quindi causare un guadagno netto di utenti per le due entità fuse.¹⁸⁴

Come detto, Facebook, inteso come piattaforma di social network, e WhatsApp non possono essere considerati *competitors*. Tuttavia, molti argomentano che WhatsApp sarebbe potuta diventare una piattaforma di social network, sebbene già non lo fosse. Tuttavia il focus del core business di WhatsApp è la comunicazione istantanea, con un servizio semplice e rapido disponibile, inizialmente, solo su smartphones. Secondo l'investigazione di mercato condotta dalla Commissione, i principali fornitori di *social networking services* sono Facebook, Google+, LinkedIn, Twitter e MySpace. Tali piattaforme, fornivano un'esperienza sociale ricca, tipica caratteristica dei social networks, offrendo ai consumatori la possibilità di creare identità digitali e profili personalizzati che riflettevano la propria personalità e i propri interessi. Essi non sono prodotti sostituiti alle varie piattaforme di *communications services* come WhatsApp, LINE, Snapchat, Viber e Hangouts, bensì sono complementari, nel senso che i consumatori utilizzano gli uni e gli altri. Inoltre, sebbene WhatsApp sia standard e mostri i contenuti in base alle interazioni dell'utente con i propri contatti, Facebook, nella propria *homepage* mostra contenuti dinamici, basati sulle attività di amici, pagine ed altre entità con cui l'utente è connesso. Queste includono post, foto, eventi e la *display advertising*. Tutte le homepages degli utenti, dunque, sono personalizzate e si basano sui loro interessi e sulle loro personalità. WhatsApp, al contrario, non offre né una homepage, né una cronologia di eventi di interesse per l'utente (compleanni, eventi a cui partecipano gli amici ecc.) e la sezione dedicata al profilo dell'utente è molto meno dettagliata rispetto a Facebook, dove vi sono informazioni circa l'educazione, la situazione sentimentale il lavoro ecc.

WhatsApp è perciò una piattaforma più statica e semplice costruita solo con lo scopo di scambiare messaggi multimediali in modo istantaneo.

¹⁸³ "In 2013, the use of messaging and social apps grew by 203%, more than any other type of apps. This growth is expected to continue in the future. New consumer communications apps are continuously offered for download to customers and are expected to be launched also in the future."

¹⁸⁴ "Indeed, on the basis of the estimates provided by the Notifying Party,⁸² in the period between December 2013 and April 2014, between [20-30]% and [50-60]% of WhatsApp users already used Facebook Messenger and between [70-80]% and [80-90]% of WhatsApp users were Facebook users and were therefore already within the reach of Facebook Messenger. Conversely, over the same period 60% to 70% of Facebook Messenger active users already used WhatsApp." Case No COMP/M.7217.

Dunque, secondo la Commissione WhatsApp e Facebook non sono affatto in competizione, poiché appartengono a mercati rilevanti diversi tra loro e l'integrazione, in tal senso, può definirsi conglomerale.¹⁸⁵

I Big Data entrano in gioco nella transazione, considerando il mercato della *on-line advertising*. In tal senso l'operazione di concentrazione può comportare una potenziale *data concentration*, rafforzando, dunque, la posizione di Facebook in tale settore, in particolare in riferimento alla *display advertising*. Poiché solo Facebook è attivo in questo mercato, WhatsApp non raccoglie dati sugli utenti circa la loro età, nome e cognome, genere, gruppo sociale, attività svolte sulla piattaforma, abitudini di consumo e altre caratteristiche di valore per scopi pubblicitari. Inoltre, come detto, WhatsApp non registra nemmeno i contenuti dei messaggi. Così, per tali motivi, l'operazione di concentrazione non porta ad un incremento dell'ammontare di dati disponibili per Facebook ai fini pubblicitari.

Tuttavia, la Commissione ha analizzato due possibili danni potenziali derivanti da un rafforzamento e della posizione di Facebook nel mercato della *on-line advertising*: l'introduzione di pubblicità su WhatsApp e l'utilizzo di WhatsApp come potenziale mezzo di raccolta di dati col fine di incrementare l'efficacia della personalizzazione della pubblicità di Facebook.¹⁸⁶

Secondo tali teorie, dopo la transazione, WhatsApp avrebbe dunque dovuto aggiungere pubblicità on-line sulla propria piattaforma, raccogliendo ed analizzando i dati degli utenti di WhatsApp. Questo avrebbe avuto l'effetto di rinforzare il PDM dell'entità risultante dalla fusione nel settore della pubblicità on-line. Tuttavia, questo non è accaduto e WhatsApp attualmente non vende pubblicità. Infatti WhatsApp avrebbe dovuto cambiare la sua *privacy policy* e ciò avrebbe potuto causare il malcontento di tutti quei consumatori che ripongono un elevato valore sulla *privacy* e sulla sicurezza. Inoltre la concentrazione non ha avuto alcun effetto sugli inseritori pubblicitari, che, in virtù del *multihoming* e dell'elevato grado di differenziazione del mercato della pubblicità on-line, avrebbero potuto passare ad altre piattaforme, come Google+, Youtube, Yahoo!, MSN ecc. a seguito di un aumento dei prezzi dovuto allo sfruttamento del PDM ottenuto dopo l'operazione. Infine, i dati a cui WhatsApp ha accesso, secondo la Commissione, non sono in grado di aumentare l'utilità marginale di Facebook nell'elaborare informazioni di tipo commerciale, non permettono quindi a Facebook di fornire contenuti maggiormente proliferati verso gli utenti.

Infatti WhatsApp raccoglie solo dati come il nome e il numero di telefono associati all'account dell'utente, sebbene ci siano dubbi circa l'utilità del numero di telefono ai fini di incrementare l'efficacia della pubblicità offerta da Facebook.

¹⁸⁵ Cfr Paragrafo 4.2

¹⁸⁶ Altre teorie, invece, suggerivano che Facebook avrebbe potuto dividere WhatsApp in due versioni, una gratuita e monetizzata tramite pubblicità ed un'altra acquisibile con il pagamento di una *fee* senza l'offerta di contenuti pubblicitari.

Tuttavia la Commissione affermava che Facebook non aveva alcuna intenzione di modificare le strategie di WhatsApp in relazione alla raccolta e all'analisi di dati e non c'è, dunque, alcun effetto sulla potenzialità dei dati disponibili per Facebook ai fini dei contenuti pubblicitari personalizzati. Inoltre, come detto, se WhatsApp avesse cambiato *privacy policy*, a seguito della fornitura dei propri dati a Facebook, molti utenti avrebbero cambiato piattaforma passando ad un'altra meno intrusiva. Questo, dunque, ha ridotto l'incentivo di Facebook di raccogliere dati tramite WhatsApp e di trasmetterli alla piattaforma di social network. Inoltre la Commissione affermava che, anche se Facebook utilizzasse WhatsApp col fine di migliorare l'efficacia della sua pubblicità personalizzata, ci continueranno ad essere numerosissimi dati sugli utenti di Internet disponibili per altre piattaforme e, quindi, al di fuori del controllo esclusivo di Facebook. Ci sono, infatti, numerosi altri *player*, oltre Facebook, che raccolgono ed analizzano i dati degli utenti a fini commerciali. Essi includono Google, che è il player con maggiore QDM nel macrosettore della pubblicità on line (vedasi grafico Paragrafo 2.5), Amazon, Apple, Ebay, Microsoft, Yahoo!, Twitter, IAC, LinkedIn, Adobe e Yelp.

In conclusione, basandoci sui risultati dell'investigazione di mercato condotta dalla Commissione Europea, la transazione non suscita seri dubbi sulla sua compatibilità col mercato interno e con il grado di competizione nel settore della *on-line advertising*. Perciò il 03/10/2014 la Commissione dichiarava compatibile l'operazione di acquisizione di WhatsApp da parte di Facebook secondo l'Art 6 della Merger Regulation.¹⁸⁷

¹⁸⁷ 1. La Commissione procede all'esame della notificazione non appena questa le è pervenuta: a) se essa conclude che la concentrazione notificata non rientra nel presente regolamento, lo constata mediante decisione; b) se essa constata che la concentrazione notificata, pur rientrando nel presente regolamento, non suscita seri dubbi per quanto riguarda la sua compatibilità con il mercato comune, decide di non opporvisi e la dichiara compatibile con il mercato comune. Si considera che la decisione che dichiara la concentrazione compatibile riguarda anche le restrizioni direttamente connesse alla realizzazione della concentrazione e ad essa necessarie; c) fatto salvo il paragrafo 2, se la Commissione constata che la concentrazione notificata rientra nel presente regolamento e suscita seri dubbi per quanto riguarda la sua compatibilità con il mercato comune, decide di avviare il procedimento. Fatto salvo l'articolo 9, ciascun procedimento si conclude mediante una decisione conformemente all'articolo 8, paragrafi da 1 a 4, a meno che le imprese interessate non abbiano dimostrato, in maniera soddisfacente per la Commissione, di aver abbandonato la concentrazione. 2. Se la Commissione constata che, a seguito di modifiche apportate dalle imprese interessate, una concentrazione notificata non suscita più seri dubbi ai sensi del paragrafo 1, lettera c), dichiara la concentrazione compatibile con il mercato comune a norma del paragrafo 1, lettera b). La decisione adottata a norma del paragrafo 1, lettera b), può essere subordinata dalla Commissione a condizioni ed oneri destinati a garantire che le imprese interessate adempiano agli impegni assunti nei confronti della Commissione per rendere la concentrazione compatibile con il mercato comune. 3. La Commissione ha facoltà di revocare le decisioni adottate a norma del paragrafo 1, lettera a) o b): a) quando la decisione sia fondata su indicazioni inesatte di cui sia responsabile una delle imprese interessate, o sia stata ottenuta con frode; o b) qualora le imprese partecipanti non osservino uno degli oneri di cui è corredata la decisione. EU Commission.

RIASSUNTO

CAPITOLO PRIMO: BIG DATA E SISTEMI DIGITALI

“Big data is too big, too fast, or too hard for existing tools to process”

Partendo da tale definizione di Khalid Adam, col termine “Big Data” ci si riferisce a dati digitali, misurati in multipli di *bytes*, il cui ammontare è tale per cui c’è bisogno di un ingente investimento di capitale sia per la loro raccolta o acquisizione e sia per la loro analisi. I Big data si riferiscono ad un dataset la cui dimensione è troppo elevata per la capacità di calcolo dei software tradizionali per la raccolta e l’analisi dei dati, ovvero il loro volume eccede la capacità di calcolo dei processori dei database convenzionali. Gli investimenti in Big data includono quelli in risorse umane, i cosiddetti *data scientists*, oltre a quelli in soluzioni tecnologiche altamente avanzate, come le piattaforme di database management (ad esempio Hadoop, IBM/Netezza), strumenti di analisi e visualizzazione (come Revolution R), i cosiddetti *text processing*, e le soluzioni di *streaming* in tempo reale. Secondo la definizione di Beyer e Laney in “*The Importance of Big Data*” (2012) ci sono 3 V che caratterizzano i big data: volume, velocità e varietà. Oltre a questi c’è anche un altro aspetto importante che si lega maggiormente al fattore tempo, la variabilità. Col volume ci si riferisce all’ammontare di dati, più dati vuol dire più capacità di immagazzinamento ed elaborazione. Attualmente i più potenti server mondiali raccolgono oltre i 20 zettabytes e sono stimati arrivare a 45 nel 2020. Con la velocità ci si riferisce sia alla velocità di generazione di dati, sia alla velocità di immagazzinamento, elaborazione e di estrapolazione di informazioni dall’ammontare dai dati (indicatore di efficienza della *data storage* e *data analytics*). Con la varietà ci si riferisce all’esistenza di varie categorie alle quali i big data appartengono. Esso infatti rappresenta un elemento essenziale da conoscere nella loro analisi. I big data contengono testi, audio, immagini, video e molti altri dati strutturati e non, disponibili sia in formato analogico che digitale. Da un aspetto analitico la varietà è la sfida più difficile da affrontare. Molti ricercatori affermano che dominare la varietà e la variabilità è la chiave per il successo della big data analytics. Infine, la variabilità si riferisce alla rapidità con cui un dataset diventa obsoleto nel tempo. Col passare del tempo, che può essere anche molto breve, infatti, un dato può non risultare più compatibile con l’informazione ricercata in un preciso istante. Così come la varietà, esso consiste nell’elemento più difficile da affrontare.

I Big data possono essere definiti come strutturati e non strutturati. Esiste poi una via di mezzo che comprende il 5-10% della totalità dei dati, i dati semi-strutturati. La struttura dei dati è uno schema con il quale i dati possono essere immagazzinati ed organizzati, in modo tale da renderne semplice l’archiviazione e l’utilizzo. I dati strutturati, dunque, sono quelli per cui sono facilmente intuibili le loro caratteristiche e possono essere sintetizzate da tabelle che ne permettono una facile lettura e confronto con altri dati.

Grazie a questa caratteristica essi sono molto più facili da analizzare rispetto ai dati non strutturati ed hanno, tipicamente, la forma di testo scritto.

Essi dunque costituiscono informazioni ad alto grado di organizzazione. Esempi di dati strutturati sono: i numeri di telefono, transazioni finanziarie, vendite a volume ed in valore ecc. I dati non strutturati o non relazionali, sono essenzialmente l'opposto. Essi hanno al loro interno una struttura, ma è difficile da sintetizzare con schemi o modelli. Dunque non è presente uno schema, come nel caso di oggetti multimediali o file di solo testo narrativo. I tipici dati non strutturati generati dall'uomo includono: file testuali, e-mail, social media, siti web (come Youtube o Instagram), *Mobile data*, comunicazioni (telefonate, messaggi), Media (MP3, foto digitali, file audio o video) e le cosiddette Business applications (come quelli generati grazie ad Office). I tipici dati non strutturati generati dalle macchine includono: immagini satellitari, dati scientifici, videosorveglianza digitale e dati provenienti dai sensori. I dati semi-strutturati sono dati con struttura parziale. Essi presentano caratteristiche sia dei dati strutturati che di quelli non strutturati. L'e-mail è un tipico esempio di informazione contenente dati strutturati e non; infatti sono presenti sia le connotazioni anagrafiche, sia i contenuti testuali o multimediali.

La *supply chain* si compone di quattro fasi: generazione, acquisizione, raccolta ed analisi. Le prime due rappresentano la macro fase dell'approvvigionamento del dato grezzo, la raccolta la fase della conservazione e l'analisi quella della produzione in cui vi è l'apporto di nuovo valore, grazie all'elaborazione dei dati grezzi e all'informazione che vi si ricava. Durante la fase di acquisizione, viene fatta la cosiddetta *data pre-processing phase*, in cui, una volta formatosi il dataset di riferimento, bisogna utilizzare un efficiente meccanismo di trasmissione per inviarlo al *tool di storage management system* che supporta la fase analitica che si vuole affrontare. Il dataset da trasmettere può contenere molti dati ridondanti o inutili, che necessariamente riducono lo spazio per la registrazione, ostacolando il processo successivo di analisi. La fase di *data pre-processing*, dunque, consiste nell'eliminazione delle inefficienze ed è, pertanto, indispensabile per assicurare un'efficiente raccolta ed elaborazione di dati.

La *big data storage* si riferisce all'archiviazione e alla gestione della raccolta (elaborazione) di dati e si focalizza al raggiungimento e della disponibilità all'accesso e dell'attendibilità dei dati. Grazie ad internet, si è sviluppato il cloud computing che ha permesso la registrazione automatizzata di dati verso risorse preesistenti, i server. L'obiettivo delle infrastrutture di storage service è quello di fornire lo spazio necessario alla raccolta così da ridurre il costo unitario di archiviazione di ciascun dato e, d'altro canto, essere un'interfaccia funzionale alla fase di analisi. Esse hanno risolto il problema di come collezionare e registrare grandi ammontari di dati, integrare dati eterogenei e come processare dataset di larga scala per permetterne l'analisi.

I Relational database management systems (RDBMSs) sono i sistemi di registrazione tradizionali, elaborati per memorizzare i dati strutturati. I grandi server processati attraverso il RDBMSs riuscivano a garantire performance elevate nel gestire la variabilità orizzontale dei dati strutturati e accessibilità per le Big data applications; tuttavia, con l'aumento del volume di dati e dell'eterogeneità degli stessi in maniera più che proporzionale all'aumento della capacità di immagazzinamento dei RDBMSs, ci si è spostati verso altre soluzioni. Questo a causa dell'enorme volume di dati non strutturati che hanno iniziato ad essere generati. Dunque si è iniziato ad adottare altre soluzioni tecnologiche, come dei cosiddetti NoSQL e NewSQL servers, che hanno permesso la raccolta e l'archiviazione di dati non e semi strutturati.

L'analisi è la fase finale della catena del valore dei big data. Essa viene svolta col metodo analitico tradizionale se si tratta di dati strutturati, mentre è conseguita con l'ausilio di architetture analitiche e di software utilizzati allo scopo dell'estrazione e dell'analisi dei big data non strutturati. Lo scopo è quello di estrapolare valore dai dati, sotto forma di informazioni, al fine di implementare decisioni strategiche. Diversi dataset hanno diversi potenziali di valore. La cosiddetta Data mining si riferisce alla tecnica di estrazione di informazioni utili dai databases.

L'analisi tradizionale Consiste nell'utilizzo di metodi statistici per analizzare grossi volumi di dati. ci sono varie tipologie di analisi statistiche da implementare: Cluster Analysis, Factor Analysis, Correlation Analysis e Regression Analysis. Infine, A seconda della tempestività ricercata nell'analisi dei big data, quest'ultima si distingue in real-time analysis e off-line analysis in base alla capacità delle architetture di processare dati in tempo reale.

In conclusione, i Big Data comportano la nascita di un database management e un analytical approach sviluppati per raccogliere, manipolare, estrapolare conoscenza ed infine utilizzare quella conoscenza.

CAPITOLO 2: APPLICAZIONI ED IMPLICAZIONI DEI BIG DATA NEI MERCATI MULTILATERALI

I principali operatori della *data driven industry*, riconfigurati nel nuovo paradigma del *machine learning*, sono le *multi sided platforms*. Si viene a creare così uno scenario competitivo in cui tali piattaforme impiegano i sunk investments in R&D al fine di trovare la cosiddetta *breakthrough innovation* ovvero un'innovazione di prodotto o di processo *disruptive*, che può completamente cambiare la struttura di un'industria. La competizione così descritta porta alla cosiddetta situazione del *winner takes all*, in cui chi riesce ad ottenere la tecnologia vincente, cambia lo scenario del mercato e si costituisce come impresa dominante, per un certo intervallo di tempo più o meno ridotto a causa della dinamicità delle industrie *high tech*.

La peculiarità di tali piattaforme è quella di fronteggiare più domande e offerte e quindi di porre prezzi differenti per ogni lato della piattaforma. Quindi, nella maggior parte dei casi, ci sarà un gruppo di clienti o consumatori finali che avrà accesso alla piattaforma in maniera del tutto gratuita. Le piattaforme digitali raccolgono i dati degli utenti, impiegano tali dati con lo scopo di migliorare il servizio offerto, personalizzandolo a seconda dei gusti dei consumatori finali e riuscendo a monetizzarlo efficacemente grazie all'elaborazione dei dati raccolti, i quali aumentano l'attrattiva della piattaforma verso i clienti e il tempo speso da quest'ultimi sulla piattaforma.

La struttura dei costi di tali entità, quindi, è costituita da elevati costi fissi, o *sunk costs*, di *data storage* e *analysis*, ammortizzati attraverso il processo di monetizzazione dei dati raccolti, che permettono di porre prezzi positivi verso gli altri lati della piattaforma e da costi marginali pressoché nulli. Il percorso di recupero dei costi fissi viene accelerato dal cosiddetto *Network Effect*, ovvero l'aumento del numero di *users* e di dati che essi apportano aumentano il valore del dataset (e con esso la complessità della *analytics*) e del prodotto (bene o servizio) finale. Gli utenti crescono linearmente, il valore della piattaforma lo fa esponenzialmente.

Nelle piattaforme digitali con più *side* si sviluppa il cosiddetto *Indirect Network Effect*. Esso è un fenomeno che vede al centro la piattaforma, come intermediario che permette che, con l'aumentare del numero di utenti e di dati di un lato della piattaforma, cresca il valore della piattaforma per mezzo della crescita della percezione che hanno i clienti dell'altro lato della piattaforma, in termini di maggiori aspettative sulle performance e di *willingness to pay*.

Il *direct Network Effect* porta con sé più dati, questi inducono le imprese ad incrementare la capacità di calcolo dei propri algoritmi, che di conseguenza offrono contenuti personalizzati verso i consumatori finali o *users*, e servizi più efficaci ed efficienti verso gli altri clienti che utilizzano la piattaforma. L'algoritmo intelligente impiegato nella fase di analisi, infatti, ha come obiettivo quello di raggiungere gli *users* con un bene/servizio personalizzato, ad esempio la pubblicità o un contenuto correlato ai gusti del consumatore, espressi dalle ricerche effettuate nel passato.

Un'altra caratteristica delle *multi sided platforms* è il cosiddetto *Learning Effect*. Imparare ad utilizzare un prodotto (una piattaforma) ed abituarsi alle caratteristiche di tale prodotto può indurre il consumatore a non voler abbandonare quel prodotto ed esserne così intrappolato (effetto *Lock in*). Il comportamento di un consumatore infatti dipende dalle scelte che egli fa nel passato ed in molti casi è difficile passare ad altri prodotti anche se più efficienti.

Tuttavia l'effetto *lock in* viene frenato dal *multihoming*, ovvero dall'utilizzo di più piattaforme che fanno o non fanno parte dello stesso mercato rilevante. In fondo, in tali mercati la competizione si riduce ad essere semplicemente un click da parte dello *user*. ``*Competition is one click away*``. Questo fenomeno è spiegato dalla natura di tali mercati che nascono *on-line*, essi infatti sono estremamente differenziati ed è persino difficile distinguere un mercato rilevante da un altro più o meno esteso.

I dati, in tal caso, rendendo il servizio offerto dalla piattaforma sempre più "su misura" per l'utente, possono intrappolare l'utente all'uso della piattaforma, in quanto ci vorrebbe del tempo affinché l'utente trasmetta i propri dati ad un'altra piattaforma che, conseguentemente, elabora servizi su misura tramite i dati raccolti dall'utente nel tempo. Quindi l'elaborazione dei dati nel tempo può frenare il fenomeno del *multihoming* e portare alla costituzione di imprese dominanti che vengono scelte dagli utenti per il grado di personalizzazione del servizio offerto e per evitare di dover trasmettere gli stessi dati personali ad un'altra piattaforma.

Tornando all'origine dei big data, la loro collezione ed analisi è una conseguenza dell'avvento del nuovo paradigma del *machine learning*. Nell'ambito del modello previsionale, si collocano la tecnologia di machine learning (ML) e di statistical learning (SL). il ML e lo SL utilizzano un approccio di tipo induttivo per formare una rappresentazione del mondo sulla base dei dati che vengono esaminati. Le finalità delle due applicazioni sono le medesime: descrivere trend e andamenti, capire il presente, prevedere il futuro; viceversa, il fulcro delle rispettive attività invece cambia: per il ML l'algoritmo e il suo output sono centrali, per lo SL lo sono i modelli statistici e la quantificazione dell'incertezza. Inoltre, il ML e lo SL sono in grado di ottimizzare e migliorare la rappresentazione della realtà all'aggiungersi di nuovi dati. In questo senso, l'algoritmo "impara" dai nuovi dati inseriti e migliora nel tempo; stessa cosa si può dire per i modelli statistici che diventano più raffinati ed efficaci.

Un esempio significativo di applicazione delle tecnologie e degli algoritmi di analisi big data per l'estrazione di valore per il business è rappresentato dai *Recommender Systems*, spesso indicati anche con il termine *Recommendation Engines*.

Si tratta di un sistema di filtraggio delle informazioni che mira a prevedere il livello di preferenza che un utente darebbe a un dato oggetto o elemento. Utilizzati dalle grandi piattaforme digitali per incrementare l'*engagement* e il tempo speso online sulla piattaforma dagli utenti, aumentare le vendite e minimizzare il tasso di abbandono e quindi la percentuale di no users.

Questi sistemi costituiscono una valida alternativa agli algoritmi di ricerca dato che aiutano gli utenti a scoprire oggetti o contenuti che difficilmente potrebbero trovare da soli.

La logica elaborativa alla base della produzione delle *recommendations* basa essenzialmente su due approcci di analisi e filtraggio dei dati: *collaborative filtering* e *content-based filtering*.

Con l'approccio del *collaborative filtering* si comparano le preferenze dell'utente del passato con le preferenze del passato di altri utenti considerati simili. L'assunzione di fondo è che gli utenti che hanno mostrato un comportamento e preferenze simili in passato continueranno a mostrarlo in futuro, preferendo quindi oggetti simili a quelli preferiti in passato. Uno degli esempi più noti di *collaborative filtering* e *recommender system* è sicuramente dato dagli algoritmi utilizzati da Amazon basati proprio sull'assunzione che l'utente che ha comprato un dato prodotto molto probabilmente comprerà un altro prodotto già scelto da utenti "simili". Anche le piattaforme Social, quali Facebook e LinkedIn, applicano algoritmi basati su tale approccio per analizzare la rete di collegamenti di un dato utente al fine di suggerirgli nuovi amici, gruppi etc. Un altro esempio è rappresentato da Spotify che crea una "stazione radio" di canzoni suggerite analizzando regolarmente la musica che un utente ascolta per poi compararla con quella ascoltata da utenti dal comportamento simile.

Con l'approccio del *content-based filtering* le preferenze dell'utente si misurano in termini di *like* e *dislike*. Per ciascun oggetto viene creato un profilo dato da un insieme di attributi e funzionalità. Sulla base del comportamento passato dell'utente viene assegnato a ciascun attributo un peso. Ogni oggetto all'interno del sistema viene così classificato tramite un vettore pesato di attributi. I pesi riflettono l'importanza di ogni attributo o caratteristica per l'utente e possono essere calcolati utilizzando diverse tecniche a partire da una semplice media pesata delle preferenze degli utenti ad approcci più sofisticati basati sul *machine learning* come reti neurali, *cluster analysis*, *decision tree* finalizzati a determinare la probabilità di preferenza di un oggetto per un dato utente.

Esistono poi dei sistemi ibridi che combinano i due approcci suddetti per produrre risultati migliori in taluni casi. Tale approccio può essere implementato in diversi modi: calcolando separatamente le *recommendation* utilizzando i due metodi precedenti e poi combinando i risultati; aggiungendo elementi dell'approccio *content-based* all'approccio *collaborative filtering* (e viceversa); unificando i due metodi in un unico modello.

Netflix rappresenta uno dei più importanti casi di applicazioni di tali tecnologie e in particolare dei sistemi ibridi dato che determina le *recomendation* sia comparando le abitudini di ricerca e visione tra utenti simili (come da approccio *collaborative filtering*) sia suggerendo film con caratteristiche simili a quelle che l'utente ha valutato positivamente (in base all'approccio *content-based filtering*). Nella *homepage* di Netflix vi sono il maggior numero di *recommendation*, dando luogo a oltre i due terzi dei minuti visti dall'utente.

Esse vengono espletate per mezzo di numerosi algoritmi, che combinano approcci sia di tipo *collaborative* che *content-based filtering*, come il *Personalized Video Ranker*, che ordina i contenuti in base sia al genere che alle preferenze dell'utente espresse nel passato e il Top-N video Ranker che mostra i contenuti preferiti da utenti considerati simili

La pubblicità on-line è definita come ogni forma di pubblicità che arriva al consumatore per mezzo di internet, a prescindere dal device di riferimento (pc, tablet, smartphone).

Ci sono vari tipi di pubblicità on-line: *banner advertising*, *video advertising*, *search engine advertising* e *display advertising*. Tutti questi tipi di pubblicità sono personalizzabili e quindi molto incisive sugli utenti online. I principali *players* in questo macrosettore sono Google e Facebook, che, pertanto, posseggono QDM tali che vengono a costituirsi come imprese dominanti. In particolare, Facebook, tramite anche un'altra piattaforma di sua proprietà Instagram, posiziona *banner* o *link* sponsorizzanti che contengono contenuti di interesse per l'utente. Questo tipo di pubblicità non serve per facilitare il processo d'acquisto, come ad esempio fa la *search advertising*, ma serve ad incrementare il legame tra utente e *brand* e la consapevolezza della forza di tale *brand*. La *search advertising*, di cui Google è il leader indiscusso, invece, consiste nel collocare pubblicità accanto alla pagina di ricerca che è stata aperta dall'utente a seconda dell'input, o keyword, che è stata inserita nella barra di ricerca del motore di ricerca utilizzato. La pubblicità oltre a potersi collocare accanto, per esempio alla destra dei risultati di ricerca, può situarsi anche in cima ai risultati costituendo la cosiddetta *organic search results*. In particolare, Google, attraverso AdSense, AdWords e Youtube riesce a porre pubblicità personalizzata in base alle ricerche, date dalle *queries*, effettuate dall'utente sul web. AdWords crea un rapporto diretto tra insertore e Google, in quanto permette di inserire inserzioni pubblicitarie nella *search engine organic results*, in base a metriche oggettive di rilevanza e qualità. AdSense, invece, consiste nel piazzare la pubblicità all'interno dei siti raggiungibili tramite il motore di ricerca di Google. In questo caso la piattaforma agisce da intermediario e permette di porre pubblicità, sottoforma di *banner* personalizzati in base alle ricerche effettuate nel passato dall'utente. La pubblicità, in questo caso, è molto più pervasiva e personalizzata, nel senso che è molto più legata all'individuo che sta effettuando la ricerca. Tramite Youtube, infine, i creatori di video possono caricare gratuitamente i propri video su YouTube e scegliere di monetizzare i loro contenuti ospitando pubblicità sul proprio canale.

CAPITOLO TERZO: BIG DATA E PRATICHE ABUSIVE

Tema centrale del presente lavoro è l'inserimento del nuovo fenomeno della *data driven economy* nel contesto regolatorio della *competition policy*. Ci si domanda, dunque, se i Big Data con le caratteristiche studiate nel capitolo primo e le applicazioni economiche viste nel capitolo secondo, possano o meno indurre le imprese digitali ad adottare condotte o pratiche che possano ostacolare la normale competizione basata sui meriti.

Tuttavia, l'intervento delle autorità Antitrust deve essere fatto, tenendo in considerazione che, trattandosi di settori con un'evoluzione estremamente dinamica, molte decisioni vanno prese con cautela, cercando di assecondare di volta in volta il rapporto tra innovazione e competizione e decidendo quale obiettivo economico far prevalere.

Infatti, un intervento Antitrust troppo restrittivo rischierebbe di intralciare il tasso innovativo nel settore e quindi c'è bisogno di effettuare un'attenta analisi sull'intenzionalità dell'impresa oggetto di indagine. Bisogna dunque capire se tale entità abbia o meno un intento anticompetitivo oppure se le sue pratiche derivano dallo sfruttamento delle tecnologie che ha a disposizione col fine di incrementare sempre più il tasso di innovazione. Dunque, in questo capitolo, si vuole capire se esiste un rapporto diretto tra detenzione ed analisi di Big Data e pratiche abusive. Quest'ultime possono riguardare sia il prezzo e le altre condizioni di vendita imposti in modo del tutto non equo (*exploitative abuses*), sia la manipolazione e l'alterazione della struttura dell'industria a discapito della concorrenza (*exclusionary abuses*). Per *exploitative abuse*, disciplinata dall'Art 102 TFEU e dalla Sezione 2 dello Sherman Act, dunque, si intende l'abuso di PDM, ad esempio ponendo prezzi troppo elevati in relazione al bene/servizio offerto o condizioni di vendita inique, oppure non garantendo l'accesso ad una risorsa essenziale. Per *exclusionary abuse*, disciplinata dall'Art 101 TFEU e dalla Sezione 1 dello Sherman Act si intende l'intento di escludere uno o più player dal settore, da parte di una o più entità, a seconda se si tratti di pratiche unilaterali o pratiche concordate.

Per quanto riguarda le prime pratiche abusive (*exploitative abuses*), esse sono inerenti alla fissazione di prezzi eccessivamente alti in relazione al valore del servizio fornito, al rifiuto di garantire l'accesso ad una risorsa essenziale, prezzi differenziati a seconda dei clienti serviti, oppure all'imposizione di clausole anticompetitive all'interno di contratti commerciali.

Per quanto concerne la prima pratica (*excessive price*) c'è da dire che è difficile capire quando un prezzo è equo oppure no, bisognerebbe infatti conoscere perfettamente la struttura dei costi dell'impresa e relazionare il prezzo col costo marginale o col costo medio. Abbiamo tuttavia visto come la struttura dei costi delle *digital companies* sia diversa dal tradizionale, con elevati costi fissi e costi marginali pressoché nulli. Oppure si può comparare il prezzo praticato col valore economico del bene/servizio reso anche in relazione ad altri beni/servizi comparabili.

Tuttavia sono gli stessi IPRs che, proteggendo i benefici dell'innovazione, incentivano l'impresa a porre prezzi elevati e a vendere minori quantità, giustificandone la causa.

Per le autorità antitrust Statunitensi, la pratica dell'*excessive price* è anticompetitiva solo se vi è un evidente danneggiamento al cliente o consumatore finale. Infatti, per la disciplina concorrenziale USA, il mero possesso di un elevato PDM giustificato da una posizione monopolistica nel mercato e, conseguentemente, la fissazione di prezzi "troppo" elevati, non è da considerarsi illecito, ma anzi, è un importante elemento per il libero funzionamento del mercato. Come detto, infatti, sono gli extraprofiti da monopolio che attraggono nuove potenziali entranti e incentivano gli investimenti in R&D per effettuare nuove innovazioni.

Secondo le autorità antitrust UE, invece, un prezzo non equo costituisce un abuso di posizione dominante e rientra pertanto nelle pratiche elencate dall'Art 102. Il prezzo deve essere equo in relazione al valore economico del servizio fornito, ma abbiamo visto come questo può essere difficile da definire per le *multisided platforms*. Infatti, secondo la normativa comunitaria antitrust, l'efficienza allocativa e il surplus del consumatore sono gli obiettivi verso cui giungere, piuttosto che l'efficienza dinamica, prevalente nella normativa USA. Quindi, in sostanza, la raccolta e l'elaborazione di grossi ammontari di dati, può portare ad un innalzamento del prezzo medio, inteso come media dei prezzi praticati per ogni singolo lato della piattaforma, ma questo può essere giustificato dall'ottenimento di un servizio sempre più efficace ed efficiente sia verso i consumatori finali, sia verso gli altri clienti della piattaforma, grazie alle attività di raccolta ed analisi intraprese dalla piattaforma.

Un'altra pratica in cui un'impresa sfrutta il proprio PDM con intento anticompetitivo è il cosiddetto *refuse to deal*. In tal caso l'impresa in questione possiede una risorsa critica, indispensabile per lo sviluppo del *business* di un settore, il cui impedimento all'accesso per altri *player* porta a situazioni di monopolio e allo sfruttamento del PDM a discapito dei consumatori.

I Big Data, come detto, costituiscono una risorsa essenziale per le piattaforme *on-line*, essi infatti permettono lo sviluppo di algoritmi automatizzati, che permettono alla piattaforma di aumentare la qualità del proprio servizio e di aumentare la base clienti e il tempo speso da quest'ultimi sulla piattaforma (*engagement*). Tuttavia i Big Data sono "beni digitali" che, doppiamente a parte, sono diversi fra loro.

Uno stesso dataset, infatti, può condurre a diverse informazioni, a seconda di quali siano gli interessi dell'impresa proprietaria dei dati e diversi dataset, al contrario, possono condurre alla stessa informazione ricercata.

I dati, quindi, spesso ammettono sostituti. Questa caratteristica è nota come ubiquità dei dati. L'esclusività sui dati non implica esclusività sulle specifiche informazioni ricercate ed ottenute tramite l'elaborazione di dati. Quindi i Big Data di cui un'impresa necessita, li può avere già un'altra impresa o sono messi a disposizione dagli utenti e quindi tracciabili sul web.

Le imprese competono sui Big Data, o meglio, competono sulla qualità delle informazioni che riescono ad estrapolare dai dati, tenendo sempre in considerazione il fatto che ogni impresa ha una determinata mole di dati da tracciare ed esaminare. Ci sono barriere all'accesso ai dati, tuttavia, una volta che le imprese si dotano delle risorse tecnologiche, umane ed intellettuali adeguate, possono tracciare tutto ciò che avviene all'interno della piattaforma e, come nel caso di Google o Facebook, anche al di fuori, in altri siti web o piattaforme che richiedono spontaneamente di essere tracciate. In sostanza, quindi, non c'è esclusività nel collezionare i dati digitali degli utenti. Non ci sono infatti accordi o contratti di esclusività tra piattaforma ed utenti.

Questi ultimi infatti non sono intrappolati all'uso della piattaforma ed anzi, sono propensi al *multihoming*, fornendo dati a più player operanti sul web, in mercati rilevanti diversi e non.

Gli utenti, a loro volta, beneficiano del *multihoming*, poiché riescono ad avere servizi digitali specializzati e standardizzati, con caratteristiche, funzioni e qualità di servizio diverse l'una dall'altra. Infatti, così come estremamente differenziati sono i gusti e le specificità che hanno gli utenti del web, ci sono moltissime piattaforme che assecondano tali attitudini. Una piattaforma, infine, non può impedire ad altre di raccogliere i dati di suo interesse e che ha già prontamente raccolto. I Big data sono infatti non rivali nel consumo, nel senso che la raccolta e l'analisi da parte di un'impresa non pregiudica la raccolta e l'analisi di un'altra. Ci sono barriere all'ingresso, nel senso che tali imprese debbono dotarsi di risorse tecnologicamente avanzate ed investimenti in R&D per intraprendere un'attività di creazione di valore per mezzo dei dati. Inoltre, il *multihoming* permette a diverse piattaforme digitali di raccogliere dati sullo stesso utente. Ogni piattaforma poi riesce, chi più, chi meno, ad estrapolare valore da quei dati fornitigli spontaneamente dall'utente.

Quindi, non è l'impresa con maggiori dati ad avere un vantaggio competitivo, ma l'impresa che riesce, per ogni dataset acquisito, ad estrarre maggior valore, inteso come informazione da inserire nei propri processi produttivi al fine di personalizzare il servizio verso i clienti e monetizzarlo efficacemente.

In conclusione, quindi, non è possibile applicare l'EFD ai Big Data per la loro non escludibilità, data dalla loro natura e dall'ubiquità, la non rivalità e il *multihoming*. Questo vale in linea generale, poiché, con un'analisi più accurata si evince che esistono *dataset* il cui possesso rappresenta una risorsa essenziale ed escludibile ad altre entità, costituendosi quindi come *facility*.

Un esempio di ciò può essere costituito dai dati sanitari utilizzati nell'industria *health care*, dove, se si esclude il fatto che tale banca dati possa essere sostanzialmente equivalente a quella di un altro operatore (si pensi ad esempio al sistema sanitario di banche dati tedesco, molto simile a quello italiano), l'estensione e la completezza della banca dati del sistema sanitario pubblico sono difficilmente replicabili da operatori privati.

Questi dati potrebbero avere un valore per soggetti terzi, interessati ad utilizzarli, come ad esempio IBM col sistema di intelligenza artificiale Watson.

Analizzando dati personali e non, è possibile avere una conoscenza maggiore circa la curva di domanda che l'impresa fronteggia e, tramite gli algoritmi, fare analisi predittive. Questo può facilitare pratiche di *price discrimination* tra differenti clienti che acquistano beni o servizi da una medesima *dominant company*. Ovvero è possibile, avendo a disposizione sufficienti dati, dedurre come i *buyers* si comportano in risposta a diversi prezzi, grazie ad analisi predittive. La differenziazione dei prezzi, o ciò che gli economisti chiamano *price discrimination*, è la pratica consistente nel porre ai clienti prezzi diversi per uno stesso prodotto. L'obiettivo dell'impresa che applica la strategia di differenziazione dei prezzi, è quello di aumentare i prezzi per coloro i quali sono disponibili a pagare di più, senza perdere l'altro gruppo di clienti più sensibili a variazioni di prezzo.

Quindi i prezzi vengono fissati in base alla domanda (elasticità), o alla *willingness to pay*, dei clienti, e non in base alla struttura dei costi dell'impresa.

La *Big Data storage* e analytics, attraverso la raccolta e l'elaborazione di dati personali e non, provenienti da motori di ricerca, *cookies*, social networks, hanno reso possibile effettuare uno *shift* dalla discriminazione di terzo grado basata su categorie demografiche di clienti alla discriminazione di primo grado o strategia di *personalized pricing*. Infatti i dati permettono di creare profili diversi per ogni consumatore, riuscendo così a stimare cosa sono intenzionati ad acquistare ed a quale prezzo. Gli algoritmi permettono infatti di costruire profili personali basati sulle preferenze di ciascun individuo. È possibile così avere una visione sui comportamenti di acquisto di *cluster* di consumatori e ciò permette alle imprese di elaborare conseguenti strategie di marketing. La piattaforma, tramite gli algoritmi, può analizzare le caratteristiche e la personalità di ogni consumatore e può stimare i loro futuri comportamenti. La *price discrimination* è così una naturale conseguenza della raccolta di un elevato ammontare di dati da parte delle piattaforme digitali e sta diventando una pratica comune dei settori digitali, quindi essa è una strategia maggiormente efficace nella data economy piuttosto che nella "real" economy.

Tra le pratiche di *price discrimination* maggiormente usate dalle *digital companies* c'è lo "steering", che consiste nella manipolazione da parte di un motore di ricerca della *search result page*. Ovvero, con tale pratica, i consumatori vengono *clusterizzati* in base alle loro attitudini e/o appartenenza demografica e questo fenomeno prende il nome di *behavioural discrimination*.

Un'impresa dominante infrange l'Art 102 del TFEU quando applica condizioni di vendita dissimili a transazioni equivalenti con diversi *trading partners*. Questo dunque causa uno svantaggio competitivo per molti buyers, clienti dell'impresa in questione. L'Art 102 evidenzia che la pratica della *price discrimination* non è abusiva di per sé, ovvero c'è bisogno che questa sia fatta con intento anticompetitivo, necessita cioè dei requisiti dell'*equivalent transaction* e del *competitive disadvantage*.

Infine, la Corte di Giustizia Europea (EUCJ) riconosce la possibilità di *objective justification* per giustificarsi ed esentarsi dalla pratica.

Per *equivalent transaction* si intende che l'impresa pone prezzi diversi per beni pressochè identici, mentre per *competitive disadvantage* si intende lo svantaggio competitivo dei clienti che acquistano il bene a prezzi discriminatori.

Nella realtà dei fatti le *big digital companies* che posseggono un efficace *data storage* e *analytics*, possono implementare tale strategia. Tuttavia, questa può portare sia costi che benefici all'impresa. Il beneficio principale è che, se il venditore ha un elevato PDM, tale strategia gli permette di espandere la propria QDM, offrendo al mercato una maggiore quantità e raggiungendo quindi una mole di consumatori maggiore con un elevato differenziale di disponibilità a pagare. Un'università infatti può decidere di mettere una quota di retta annuale fissa e (presumibilmente) elevata, oppure porla in base al reddito familiare dello studente (ISEE). Con la prima strategia l'effetto è un minor numero di studenti iscritti e (presumibilmente) minori profitti. Con una strategia di *personalized pricing*, invece, il numero di studenti iscritti (ed il profitto) aumenterebbe.

D'altro canto, uno dei possibili costi causati dall'implementazione di tale strategia, può essere la riduzione del tasso di innovazione causato dalla focalizzazione sul cosiddetto *versioning*, ovvero un'eccessiva differenziazione di prodotto, con conseguente differenziazione di prezzo (prodotto premium vs prodotto low cost) che causa una minore ottimizzazione delle risorse (se non si hanno sufficienti economie di scopo da sfruttare) e quindi una produzione meno efficiente. Grazie alla raccolta ed analisi dei Big Data, per le imprese è più efficace ed efficiente intraprendere una strategia di *differential pricing* piuttosto che una strategia di *versioning*.

Un altro aspetto da considerare è che l'effetto sui consumatori, positivo o negativo, dipende da come e dove il *differential pricing* viene utilizzato. In un mercato competitivo con prezzi trasparenti, i benefici superano i costi. Ad esempio, poiché ci sono prezzi molto differenti nella vendita di biglietti aerei, Internet rende facile per molti viaggiatori comparare i prezzi di diverse compagnie aeree e selezionare il migliore per ciascuna tipologia di viaggio. Inoltre il *differential pricing* può aumentare la competizione verso i consumatori maggiormente sensibili ad aumenti di prezzo, che si rivolgono ai fornitori più economici.

Quando gli utenti utilizzano una piattaforma digitale, danno esplicitamente il loro consenso a processare i dati personali in cambio della possibilità di utilizzare la piattaforma gratuitamente, ma, molto spesso, forniscono tale consenso in maniera impulsiva e senza sapere la reale destinazione dei propri dati forniti. Essi infatti, senza saperlo, stanno pagando il servizio reso dalla piattaforma, con l'accesso a "pezzi" di *privacy*.

Gli utenti di internet, dunque, dichiarano una chiara preferenza per un certo livello di protezione alla *privacy* mentre utilizzano certi servizi *on-line*, mentre il mercato, solitamente, non fornisce più di un'opzione di *privacy*, ovvero, da soddisfare tutte le preferenze di *privacy* di ciascun cliente.

A differenti livelli di *privacy* e quindi di dati personali forniti, dovrebbe infatti corrispondere un differente prezzo per l'accesso alla piattaforma. Ossia, più dati fornisco alla piattaforma, meno pago, se li fornisco tutti indistintamente, il servizio mi viene reso gratuitamente. Se non voglio divulgare tutte le mie informazioni, dovrò corrispondere una *fee* alla piattaforma pari al valore dei dati che non sono stati resi. Tuttavia, nella *data economy*, il servizio viene reso, molto spesso, indistintamente gratuito, così che ogni utente che vuole utilizzare una piattaforma digitale si trova nella posizione di dover decidere se fornire ogni tipo di dato accettando i termini e le condizioni alla *privacy* o rinunciare totalmente al servizio della piattaforma. Non ci sono dunque *cluster* di utenti diversi in base al valore che essi riservano alla *privacy* protection.

Tuttavia, ci sono stati casi in cui, nei contratti tra utenti e piattaforma, le condizioni di scambio e i diritti di accesso ai dati non erano ben chiari agli utenti che, dando il loro consenso al trattamento dei propri dati personali alla piattaforma, perdevano il controllo sugli stessi e permettevano ad altre piattaforme o siti web di avere accesso a tali dati e, di conseguenza, proliferare maggiormente il proprio servizio verso l'utente.

Le condizioni di scambio possono essere illecite e quindi rappresentare una pratica di *exploitative abuse*, in quanto, in virtù dell'importanza della piattaforma per i consumatori e, conseguentemente, in virtù dell'elevato PDM, la piattaforma impone ai consumatori finali condizioni di scambio dei dati che possono danneggiare quest'ultimi, andando ad intaccare la loro *privacy*.

Il principio generale dell'EU *data protection law* è che è necessario il consenso del *data controller*, ovvero dell'utente consumatore finale, affinché una piattaforma possa avere il diritto di processare i suoi dati personali. Secondo l'Art 6 del GDPR, infatti, è il consenso il requisito necessario per intraprendere un'attività di *data storage, analysis e processing*, e, tale consenso, deve essere fornito dagli utenti in maniera del tutto libera ed informata, nel senso che la piattaforma deve mettere in risalto gli utilizzi che fa dei dati personali degli utenti, se questi vengono trasmessi a terze parti e quali sono le implicazioni a livello di personalizzazione del servizio.

Infatti, secondo l'Art 4 del GDPR quattro devono essere i requisiti affinché il consenso sia lecito. Esso deve essere fornito liberamente, specificamente per ogni utente, in modo informato e non ambiguo.

Ovvero, la piattaforma deve dotare l'utente di tutte le informazioni circa gli usi dei dati e lo deve fare in modo chiaro evitando situazioni ambigue. Tuttavia, nella maggior parte dei casi, gli atteggiamenti dei consumatori non sono allineati a quanto dice il GDPR e, spesso, forniscono il loro consenso in maniera superficiale senza avere piena consapevolezza sugli usi che la piattaforma fa dei propri dati. Le piattaforme, d'altro canto, sono consapevoli di ciò ed adottano atteggiamenti che sfavoriscono una chiara e consapevole fornitura di dati, andando pertanto a commettere pratiche illecite, sfavorendo ed occultando alcune delle destinazioni che i dati hanno.

Quello che è appena stato descritto rappresenta il cosiddetto “*privacy paradox*”. Questa espressione si riferisce al fenomeno secondo cui la maggior parte dei consumatori rivendica l’importanza della propria *privacy* ed il bisogno di una *data protection*, ma, in realtà, essi non agiscono secondo tali preferenze e desideri. Tutto ciò, molto spesso, è dovuto a due problemi: la maggiorparte delle *privacy policies* sul web sono eccessivamente lunghe che quasi nessun utente impiega tutto il tempo necessario per leggerle tutte. Allo stesso tempo, il linguaggio utilizzato dalle piattaforme è, molto spesso, di difficile comprensione per molti utenti.

Un caso molto emblematico che merita di essere citato nell’ambito dell’exploitative abuse causata dall’imposizione di clausole contrattuali illecite è l’investigazione dell’autorità garante della concorrenza tedesca, il *Bundescartellamt*, contro Facebook. Questo è un caso emblematico che, seppur ancora rimasto irrisolto, riunisce tematiche che comprendono sia aspetti di *data protection* che di *competition law* e rappresenta il primo caso di *competition* in cui vengono applicate norme relative alla *privacy* per la sua risoluzione.

Il *Bundescartellamt*, nel suo *background paper* fa riferimento sia ad una mancanza di trasparenza verso gli utenti circa le modalità con cui i dati di questi ultimi venivano processati, sia alla forzatura degli utenti sul far loro trasmettere indistintamente tutti i dati personali pena l’impossibilità di utilizzare la piattaforma. Infatti le preferenze degli utenti in tema di *privacy* non vengono discriminate da Facebook, che quindi forza i consumatori a trasmettere ogni tipo di dato personale proveniente da più fonti, interne ed esterne alla piattaforma. Così Facebook stava abusando della sua posizione dominante, subordinando l’accesso alla piattaforma al trasferimento di ogni tipo di dato personale dagli utenti a Facebook e da altri siti web a Facebook, unendo tali dati e personalizzando così sempre più il servizio verso gli utenti. Il *Bundescartellamt*, infatti, fa una distinzione tra collezione e uso di dati all’interno del Social Network (*on Facebook*) e quelli provenienti da siti web terzi (*off Facebook*).

Solo gli ultimi sono quelli oggetto di investigazione dell’autorità tedesca e si riferiscono a quei siti web e app che condividono le APIs di Facebook, che permettono la condivisione dei dati. Facebook viene considerata dal *Bundescartellamt* come impresa dominante nel mercato tedesco.

Secondo l’autorità, infatti, i consumatori sono intrappolati all’uso della piattaforma e non passano ad un altro Social Network con un identico scopo.

Così gli utenti non hanno altra possibilità che accettare i termini e le condizioni imposte da Facebook per utilizzare la piattaforma. Il risultato è che gli utenti non hanno effettivamente acconsentito a tale condivisione di dati tra Facebook e siti web terzi e questo è in contrasto con la *data protection law* che protegge la libertà di decisione dei consumatori riguardo i propri dati personali, che non possono quindi essere forzati a condividere con altre entità.

La violazione della *data protection law*, quindi, in concomitanza con la forzatura verso i consumatori ad accettare di fornire a Facebook dati anche provenienti da terze parti, ha condotto ad una preliminare decisione del *Bundescartellamt* contro Facebook. L'autorità, pertanto, considerava illeciti ed abusivi i termini e le condizioni alla *privacy* imposte dalla piattaforma ai suoi utenti. Ovvero, quindi, tali termini rappresentavano una forma di *exploitative abuse*. L'autorità infatti spiegava nel *Preliminary Assessment* che, in virtù della legge tedesca in merito alla concorrenza, se i termini imposti da Facebook rappresentavano una manifestazione di dominanza ed elevato PDM, tali termini erano da considerarsi *exploitative* e quindi abusivi secondo i principi della legge civile tedesca. In particolare, Facebook è così dominante che è praticamente in grado di dettare i termini contrattuali autonomamente, eliminando l'autonomia contrattuale dell'altra parte. Su queste basi, dunque, l'Autorità affermava che Facebook, in virtù del suo PDM e della sua posizione dominante nel settore dei social network, ignorava i diritti degli utenti in merito alla *privacy* e protetti dalla *data privacy law*. Così la piattaforma commetteva un *exploitative abuse* della sua dominanza.

L'uso degli algoritmi e la crescente disponibilità di dati strutturati e non stanno cambiando le condizioni del mercato, rendendo maggiormente possibili e convenienti pratiche concordate e strategie cooperative che possano falsare ed ostacolare il normale funzionamento dell'industria e le sue dinamiche competitive.

Questo paragrafo descrive il modo in cui gli algoritmi hanno cambiato lo scenario competitivo delle industrie digitali, offrendo l'opportunità alle imprese di raggiungere accordi collusivi in modi del tutto nuovi che non richiedono il raggiungimento di un accordo in modo tradizionale (scritto o verbale) e non necessitano, quindi, di interazioni umane per la loro riuscita. Le teorie economiche suggeriscono che c'è un considerevole rischio che gli algoritmi, aumentando la trasparenza nel mercato e permettendo di scambiare beni con alta frequenza, aumentino il rischio di collusione in settori caratterizzati tradizionalmente da un alto grado di competizione. Gli algoritmi, dunque, possono facilitare la *tacit collusion*, fornendo alle imprese un potente mezzo per comunicare, implementare strategie comuni, monitorare tali strategie e punire eventuali defezioni, intese come deviazioni dall'equilibrio collusivo trovato dall'algoritmo.

In un contesto di *tacit collusion*, il risultato anticompetitivo viene raggiunto automaticamente tra imprese che adottano la loro strategia di massimizzazione dei profitti indipendentemente dai propri competitors.

Questo, tipicamente, accade in mercati caratterizzati da elevata trasparenza e con pochi *players*, in cui le imprese beneficiano di un PDM collettivo raggiunto senza nessuna comunicazione esplicita.

È questo il caso degli algoritmi che, automaticamente, giungono a risultati, in termini di prezzo, output, market share, ottimizzanti sia per l'impresa che per i competitor, evitando i mezzi di competizione tradizionali e raggiungendo un elevato grado di soddisfacimento, in termini di profitti, per le imprese operanti nel settore.

Gli algoritmi facilitano la realizzazione di una collusione tacita, riuscendo a trovare l'ottimo pareiano, inteso come quantità da offrire e prezzo, ad ogni situazione strutturale del mercato e a reagire in tempo reale ad ogni possibile cambiamento di strategia da parte dei competitors. L'algoritmo quindi riesce a trovare sempre la strategia migliore di prezzo o di quantità dell'impresa in relazione ad altre strategie adottate dagli altri *players* dell'industria, in modo del tutto automatico e senza il bisogno di giungere ad accordi, formali ed informali, con i competitors, ma, semplicemente, calcolando di volta in volta quale strategia si addice maggiormente alla massimizzazione dei profitti per l'impresa e, di riflesso, per gli altri membri del cartello.

Gli algoritmi permettono meccanismi di coordinazione, monitoraggio e punizione in maniera del tutto automatica anche in settori poco concentrati, grazie all'abilità e alla rapidità nel collezionare e processare elevati ammontari di dati. In altre parole, il basso numero di imprese è un fattore importante, ma non rappresenta una condizione necessaria per la creazione di un *algorithmic collusion*.

Altre due importanti caratteristiche che rendono più semplice per gli algoritmi giungere ad un equilibrio collusivo, sono il grado di trasparenza nel mercato e la frequenza di interazioni tra imprese. Mentre la trasparenza permette alle imprese di monitorare in maniera efficace le azioni degli altri membri del cartello e individuare atteggiamenti devianti, la frequenza delle interazioni permette di aggiornare continuamente e in tempo reale le condizioni di vendita in base ai trend della domanda. Un esempio di ciò è dato dai cosiddetti *dynamic prices*, i quali permettono a clienti e fornitori di osservare in tempo reale i rapidi cambiamenti di prezzo in diversi settori, ed agire di conseguenza. I *pricing algorithms* sono codici computazionali col fine di fissare automaticamente i prezzi che massimizzano il profitto dell'impresa, date certe condizioni. Essi danno la possibilità alle imprese di reagire tempestivamente con un'altra strategia ottimizzante, al mutare di alcune condizioni di mercato. Sono i casi delle piattaforme che vendono biglietti aerei, camere di albergo, biglietti sportivi, trasporti in generale, elettricità e industrie *retail*. Tali algoritmi sono particolarmente utili ad implementare una strategia di prezzo che mutua di continuo nel tempo, il *dynamic pricing* o ad offrire a diversi *cluster* di consumatori prezzi differenti in base alla loro *willingness to pay*.

Un'altra importante caratteristica strutturale è data dal grado di innovazione. La dinamicità dell'industria, infatti, riduce il valore attuale degli accordi collusivi, così come la scarsa abilità delle imprese ad innovarsi.

In questo campo, gli algoritmi rappresentano un importante mezzo di innovazione, permettendo alle imprese di sviluppare *business models* non tradizionali ed estrarre valore dai dati, col fine di migliorare la qualità e la customizzazione dei propri prodotti. In settori in cui gli algoritmi rappresentano una fonte di vantaggio competitivo, questo è ad esempio il caso dei motori di ricerca, app di navigazione, piattaforme di *matching*, le imprese potrebbero fronteggiare un'elevata pressione competitiva per sviluppare il *best-performing algorithm*.

Nonostante gli apparenti effetti ambigui, gli algoritmi hanno sostanzialmente cambiato le caratteristiche strutturali che affettano maggiormente il grado di competizione di un'industria digitale: la trasparenza di mercato e la frequenza delle interazioni.

Il più grande vantaggio che apportano algoritmi alle strategie concordate delle imprese digitali è che essi riescono a monitorare i comportamenti delle imprese in tempo reale senza il bisogno dell'intervento umano, permettendo così alle imprese di sostituire la *explicit collusion*, con una coordinazione tacita ed automatizzata.

Quest'ultima funzione, il monitoraggio, viene svolta dai cosiddetti *monitoring algorithms*. Essi possono processare informazioni sulle decisioni di business dei competitors, su possibili atteggiamenti devianti dall'equilibrio cooperativo e programmare eventuali reazioni immediate. Ad esempio, è possibile un *pricing algorithms* che, automaticamente, reagisce a deviazioni dai prezzi pattuiti dal cartello, assecondando così una possibile guerra dei prezzi, o *trigger strategy*.

Un'altra difficile *task* che riescono a risolvere gli algoritmi, come detto, è quella di riuscire a monitorare i continui cambiamenti strutturali di mercato e, conseguentemente, adottare strategie di pricing concordate con gli altri membri del cartello. Così tali algoritmi, detti *parallel algorithms*, riescono ad automatizzare le decisioni delle imprese partecipanti al cartello e reagire simultaneamente ponendo di volta in volta prezzi massimizzanti ogni qual volta c'è un cambiamento strutturale. È questo l'esempio dei *dynamic prices* citati precedentemente.

Inoltre, col fine di evitare comunicazioni esplicite, gli algoritmi possono segnalare e rivelare l'intento collusivo agli altri *players* attraverso annunci sul prezzo. Ovvero le imprese possono sviluppare i cosiddetti *signalling algorithms* col fine di comunicare alle altre imprese le proprie strategie di prezzo. Un'impresa può pertanto segnalare ad un'altra che sta aumentando il prezzo e l'altra può, quindi, assecondare tale aumento di prezzo, praticando un prezzo più elevato e massimizzando così i profitti congiunti delle due imprese. Quindi tali algoritmi servono per stabilire e negoziare i termini della collusione prima di intraprendere una strategia di *pricing co-ordination*.

Attraverso l'uso del *machine learning* e del *deep learning*, infine, gli algoritmi integrati a tali meccanismi sviluppano una notevole capacità predittiva, imparando costantemente dall'esperienza e riadattandosi alle azioni degli altri player dell'industria. Tutto questo è svolto in maniera automatica, senza il bisogno, quindi, di un intervento umano.

Gli algoritmi hanno così cambiato le caratteristiche dei mercati digitali, aumentandone la trasparenza, la rapidità nel prendere decisioni strategiche e l'abilità delle imprese nel reagire tempestivamente alle azioni dei competitors. Sotto queste circostanze, gli algoritmi possono rendere interdipendenti le azioni dei vari *players* dell'industria senza il bisogno di un'esplicita comunicazione o interazione, aumentando il rischio di collusione tacita e portando così ad un aumento del livello generale dei prezzi. Gli algoritmi possono pertanto distorcere la normale competizione dei "meriti", creando incentivi e meccanismi che facilitano la collusione, dove con i metodi tradizionali non si riusciva a raggiungere.

Tuttavia, è ancora poco chiaro se il sistema regolatorio possa prevenire la formazione di algoritmi di ML che possano raggiungere un equilibrio collusivo. È difficile, dunque, dimostrare il danno verso il consumatore e verso la struttura competitiva del settore. Così ancora nessuna soluzione è stata proposta dalla letteratura antitrust per fermare tali condotte. Inoltre, non ci sono casi o investigazioni in atto che possono fornire un adeguato nesso comparativo e supportare l'evidenza che la collusione tacita danneggia i consumatori e i competitors, attuali e potenziali, che si collocano al di fuori del cartello.

CAPITOLO QUARTO: BIG DATA ED OPERAZIONI DI CONCENTRAZIONE

Il dotarsi delle infrastrutture e delle *skills* per la raccolta e l'analisi dei dati, come abbiamo più volte discusso, rappresenta la barriera all'ingresso per eccellenza per le imprese che vogliono far entrare il ML e il DL nei propri processi produttivi.

Così, un crescente numero di imprese, sceglie di acquisire tali competenze piuttosto che svilupparle al loro interno. Si stanno così incrementando operazioni di M&A con lo scopo di acquisire Big Data ed implementare le informazioni raccolte, tramite la loro elaborazione nei processi produttivi delle imprese.

I protagonisti delle operazioni di M&A, col passare degli anni, sono diventate le cosiddette "*technology targets*", ovvero imprese digitalmente avanzate che adottano in maniera intensa il ML e/o il DL nei propri processi produttivi.

I settori in cui tali operazioni di M&A sono maggiormente presenti e coinvolgono imprese operanti nell'ambito della raccolta, analisi ed elaborazione di grossi ammontari di dati sono: l'*Industry 4.0*, ovvero il manufacturing automatizzato e l'*internet of things* (con una crescita annua stimata di oltre il 20%), il *cloud computing*, che fornisce soluzioni basate sulla tecnologia *cloud*, il mobile tech e il software application.

Le operazioni di concentrazione sono, dunque, molto frequenti, soprattutto nell'industria digitale. Esse sono quindi trainate dalla tecnologia, soprattutto quella inerente al ML e al DL, che giustifica le sinergie ricercate ed il valore finanziario dell'impresa *target*. Tali sinergie sono costituite soprattutto da economie di scala o di scopo e dal controllo esclusivo di un input o canale di distribuzione, a seconda di dove ci si colloca nella filiera produttiva. Tuttavia, tali operazioni di concentrazione, possono comportare alcuni rischi competitivi. Ovvero, a seguito di un'operazione di M&A, c'è la riduzione di pressioni competitive, che avvengono o all'interno della stessa filiera produttiva o tra segmenti più o meno differenziati; e questo può portare ad un aumento dei prezzi, riduzione delle quantità, peggiore qualità dei prodotti e minore pressione ad effettuare investimenti in R&D. C'è da dire, inoltre, che le operazioni di concentrazione possono portare, oltre ai suddetti rischi competitivi, anche ad efficienze, intese come sinergie trovate dalle due imprese che, si traducono in una riduzione dei costi o in un aumento dei ricavi, con possibili effetti positivi anche sui consumatori finali. Per questo motivo, le operazioni di M&A sono maggiormente tollerate rispetto alla formazione di un cartello.

I Big data, come li abbiamo studiati noi, ovvero che attraversano una catena del valore, che parte dalla loro raccolta e termina con l'extrapolazione di importanti informazioni di business, passando per la fase di analisi ed elaborazione tramite l'ausilio di algoritmi, se sono compatibili e vengono messi insieme a quelli di un'altra impresa, possono costituire una preziosa economia di scopo.

Due imprese, infatti possono raccogliere dati simili ed analizzarli mettendo in comune infrastrutture e personale umano qualificato, riducendone pertanto i costi associati e permettendo una gestione della raccolta ed una fase di analisi condotte in maniera più efficiente.

In sostanza, più dati messi a disposizione, permettono all'impresa fusa di aumentare la capacità di calcolo e a ridurre il costo unitario di raccolta ed analisi associato, grazie allo sfruttamento di economie di scala e al perfezionamento delle tecniche di raccolta ed analisi, che permettono di effettuare una compressione dei dati efficace e ad eliminare i dati ridondanti che scaturiscono dalla fusione di diversi dataset simili tra loro.

Le operazioni di concentrazione sono disciplinate dalla Section 7 of the US Clayton Act; dal Regolamento Europeo 139/2004 e dall'Articolo 5 della Legge 287/90. Esse sono discipline che regolano un tipo di controllo svolto ex-ante ed è quello che rileva maggiormente, poiché è molto difficile tornare indietro alla "vecchia" struttura del settore prima della concentrazione. Con tale tipo di controllo si fa un'analisi predittiva sui possibili effetti sul mercato di tali operazioni.

La concentrazione può avvenire in quattro diversi modi: con la fusione di due o più imprese, con l'acquisizione della maggioranza delle azioni aventi diritto di voto in assemblea e quindi costituenti diritto di controllo, con l'acquisizione di un ramo d'azienda o asset strategico e con la costituzione di una joint venture tra due o più imprese.

Il Regolamento Europeo No. 139/2004 ha introdotto, nell'Art 2 un nuovo test, più simile allo standard USA. In tale test l'operazione di concentrazione non deve significativamente impedire l'effettiva competizione basata sui meriti delle imprese. La concentrazione può impedire l'effettiva competizione basata sui meriti in tre modi: creando o rinforzando una posizione dominante di una singola impresa (*individual dominance*); rimuovendo un importante vincolo competitivo, riducendo il grado di competizione nei mercati oligopolisti (*non-collusive oligopolies*) e favorendo intese e pratiche concordate, cambiando la struttura dell'industria e rendendo tali pratiche più raggiungibili alle imprese (*collusive oligopolies*)

I Big Data, come detto, rappresentano un'economia di scopo che spinge le imprese digitali ad effettuare operazioni di M&A. le imprese possono, dunque, o mettere insieme i propri dati col fine di ricavare informazioni più dettagliate sui consumatori e per sfruttare le economie di scala, o internalizzare una fase della catena del valore dei Big Data. A seconda della strategia di crescita adottata, avremo una concentrazione orizzontale o conglomerale nel primo caso, una concentrazione verticale nel secondo caso.

Sebbene i dati possono rappresentare una sinergia, mettere insieme diversi dataset può risultare complesso e può comportare all'ottenimento di una grossa mole di dati più o meno ridondanti. Tuttavia, con un'efficace operazione di pulizia dei dataset, è possibile eliminare le inefficienze in modo più rapido e performante rispetto alla gestione separata di due dataset ed è possibile, dunque, lavorare e concentrarsi su meno dati di quelli che si avrebbero ragionando in modo distinto per ogni dataset. La natura dei diversi dataset, inoltre, ci dirà se l'operazione di concentrazione in questione può definirsi un'integrazione orizzontale o conglomerale, a seconda se la distanza tra i settori o sotto settori sia nitida o molto marcata.

Le imprese possono, dunque, o mettere insieme i propri dati col fine di ricavare informazioni più dettagliate sui consumatori e per sfruttare le economie di scala, o internalizzare una fase della catena del valore dei Big Data.

A seconda della strategia di crescita adottata, avremo una concentrazione orizzontale o conglomerale nel primo caso, una concentrazione verticale nel secondo caso.

Sebbene i dati possono rappresentare una sinergia, mettere insieme diversi dataset può risultare complesso e può comportare all'ottenimento di una grossa mole di dati più o meno ridondanti. Tuttavia, con un'efficace operazione di pulizia dei dataset, è possibile eliminare le inefficienze in modo più rapido e performante rispetto alla gestione separata di due dataset ed è possibile, dunque, lavorare e concentrarsi su meno dati di quelli che si avrebbero ragionando in modo distinto per ogni dataset. La natura dei diversi dataset, inoltre, ci dirà se l'operazione di concentrazione in questione può definirsi un'integrazione orizzontale o conglomerale, a seconda se la distanza tra i settori o sotto settori sia nitida o molto marcata. Ad esempio, i dataset raccolti da Amazon ed Ebay possono condurre a simili informazioni circa le abitudini di consumo di un cluster di individui che effettua ricerche su entrambe le piattaforme di e-commerce. Molti dati, dunque, sono ridondanti, ma, eliminando le inefficienze, si può giungere ad informazioni più dettagliate, come ad esempio la disponibilità a pagare o verificare quali sono i prodotti che ruotano di più per categoria di merci ricercate. Invece, piattaforme come Ebay e Uber processano dati completamente diversi che, uniti, non permettono di aumentare il valore delle informazioni raccolte, anzi necessitano di differenti infrastrutture e *skills* specializzate per l'uno e per l'altro dataset. Possibili efficienze date da tale tipo di integrazione conglomerale, possono derivare da una razionalizzazione dei fattori produttivi e dalla condivisione di infrastrutture, skills, personale umano e *know how*, che possono permettere alle imprese di crescere e sviluppare nuove soluzioni tecnologiche.

Nel caso di concentrazione orizzontale, dunque, i settori non sono tanto distanti tra loro per tipologia di mercato servito e le imprese coinvolte competono tra loro in termini di QDM, tempo speso dai consumatori sulla piattaforma e modelli di business in generale.

Un'integrazione orizzontale, in tal caso, oltre a comportare l'eliminazione di un vincolo concorrenziale, dato da un competitor in meno, con conseguente rischio di aumento di PDM e rischio di pratiche collusive, spinge le imprese ad effettuare una gestione della fase di *data storage* più efficiente, eliminando i dati ridondanti, attraverso operazioni di compressione e filtraggio dati e ad aumentare il valore delle informazioni ricercate con l'obiettivo di rendere maggiormente efficace la personalizzazione o clusterizzazione dei clienti o consumatori finali.

Le due imprese, possono, dunque, o adottare una strategia di accentrimento delle fasi di storage e analytics, ottimizzando e razionalizzando le risorse a disposizione e ragionando su dataset che includono dati provenienti da entrambi le piattaforme, così da eliminare i dati ridondanti e lavorare con meno dati che permettono di giungere ad informazioni dettagliate sui consumatori, oppure decidere di lasciare la raccolta e l'analisi dei dataset distinta per ogni piattaforma, condividendo però infrastrutture, *skills* e *know how*, così da giungere allo stesso livello di informazioni, con una gestione decentrata dei dataset, che richiede una maggiore complessità da gestire. La scelta su quale strategia adottare, se accentrare o decentrare entrambe le fasi di raccolta e analisi, oppure o l'una o l'altra, dipende da qual è il grado di distanza settoriale tra le imprese, ovvero dipende se le imprese competono o meno nello stesso mercato rilevante.

Si avrà infatti integrazione conglomerale se i settori e i rispettivi dataset raccolti sono estremamente differenziati tra loro e l'operazione di concentrazione in tal senso non viene svolta con intento anti competitivo, bensì l'obiettivo appare essere solo la crescita inorganica dell'impresa in questione. In tal caso, infatti, si tenderà a decentrare le fasi di *data storage* e *analysis* e le sinergie che vengono sfruttate sono solo quelle inerenti alla condivisione di risorse tecniche e personali ed avere un *know how* comune ad entrambe le piattaforme.

In generale, nel caso di concentrazione conglomerale, non dovrebbero sorgere problemi di competition, tuttavia le piattaforme coinvolte possono produrre beni complementari. Ad esempio un'impresa che produce sistemi operativi (OS) per pc può differenziarsi entrando nel settore dei browser di Internet con un'acquisizione. In tal caso l'impresa può sfruttare il PDM che possiede nel settore degli OS nell'altro settore, ad esempio può offrire un prodotto ad un prezzo molto basso e coprire le perdite con i profitti nell'altro settore oppure, e questo è più inerente ai settori digitali, adottare operazioni di "tying" e "bundling", senza dare la possibilità al consumatore di acquisire i prodotti separatamente, sfruttando così il PDM di un settore, quello dei sistemi operativi nell'esempio, per espandere la QDM dell'altro settore dei browser.

La concentrazione verticale rappresenta l'integrazione tra due imprese che si collocano sulla stessa catena del valore, con attività complementari, ma distinte tra loro. In sostanza le imprese operano su livelli diversi della stessa catena del valore. Tale catena del valore, nel caso nostro, coincide con quella dei Big Data, in quanto tali imprese utilizzano i dati come core asset dei propri processi produttivi.

Le imprese, quindi, possono internalizzare le attività di *data storage* o *data analytics*, che precedentemente erano affidate a terzi, passando da forme di partnership commerciali all'inserimento di attività della stessa filiera produttiva all'interno dei propri processi. Un fenomeno molto frequente a cui si assiste ai giorni d'oggi, è l'acquisizione di imprese esperte nel campo dell'*analytics* che sono in grado di estrapolare importanti informazioni di business, attraverso ad esempio lo sviluppo di potenti algoritmi che permettono di processare elevati ammontari di dati provenienti dalla piattaforma acquirente.

L'integrazione verticale può comportare rischi dal punto di vista competitivo. In particolare, un'impresa può impedire l'accesso al mercato ad un'altra, ad esempio, limitando l'accesso ai competitors ad una materia prima o componente (*input foreclosure*) o ad un canale distributivo (*consumer foreclosure*), a seconda se l'impresa si trova a monte o a valle della filiera produttiva, rappresentata in tal caso dalla catena del valore dei Big Data. In questo caso, l'analisi competitiva si svolge *ex-ante* sulla struttura del settore e sulle condotte delle imprese che possono avvenire post-transazione. In particolare, le autorità competenti vogliono verificare se i *competitors* hanno accesso a sufficienti fornitori o clienti alternativi e se l'operazione di concentrazione, prontamente notificata, può cambiare gli accordi commerciali tra le parti o facilitare la collusione tra *competitors*. I requisiti per cui una JV implica un'operazione di concentrazione sono che la società frutto della Joint Venture, sia controllata congiuntamente da due o più imprese indipendenti e che l'impresa creata con la Joint Venture operi sul mercato come un ente autonomo, ossia che eserciti stabilmente tutte le funzioni di un'entità economica autonoma. Nel caso dei Big Data, tale strumento viene spesso usato nell'ambito delle operazioni di R&D, affidando alla nuova entità, costituita ad hoc, la gestione di una delle fasi della catena del valore. In tal modo si consente alle imprese madri di beneficiare di importanti vantaggi competitivi ottenuti mediante il raggiungimento di economie di scala o di scopo, la promozione di processi di integrazione verticale od orizzontale e la riduzione dei rischi connessi agli investimenti. Le autorità Antitrust devono valutare caso per caso se i benefici sociali, derivanti da tali forme di cooperazione e coordinazione, in chiave soprattutto tecnologica, superano i costi sociali associati ad un aumento della concentrazione nell'industria. A seconda dei casi, poi, si rientra o no nell'Art 101, altrimenti la fondazione di una JV è lecita e non rappresenta una forma di pratica anticompetitiva.

La fusione di WhatsApp con Facebook è stata esaminata dalla Commissione Europea, a seguito della notifica preventiva, inviata da Facebook alla Commissione, in conformità a quanto previsto dall'Art 4 della Merger Regulation.

Facebook è una *multi-sided platform* che fornisce i servizi di social networking, consumer communication e photo/video sharing. Essa è inoltre uno dei principali player attivo nella pubblicità on-line e, in particolare, nella display (o non search) advertising.

WhatsApp è, invece, una piattaforma di *consumer communication service* che permette agli utenti di scambiarsi messaggi istantanei, in maniera efficace e gratuita.

Essa, prima di fondersi con Facebook, era disponibile solo come app per smartphones (WhatsApp web è infatti nato il 21 Gennaio del 2015, successivamente all'operazione di fusione dell'entità con Facebook).

L'operazione è costituita dall'acquisizione di WhatsApp ad opera di Facebook, per un ammontare di 19 bilioni di dollari. Il risultato della transazione è, quindi, l'ottenimento di Facebook di un *sole control* su WhatsApp, costituendo dunque un'operazione di concentrazione ai sensi dell'Art 3 della Merger Regulation.

L'operazione riguarda unicamente il settore del consumer communication services, costituito da players che forniscono soluzioni di comunicazione multimediale istantanea, che permette agli utenti di raggiungere amici, familiari e altri contatti in tempo reale.

Storicamente, tale servizio veniva offerto tramite PC. Tuttavia, si è assistito gradualmente ad uno "spostamento" di tale settore sui cosiddetti smart mobile devices, ovvero smartphone e tablet.

Oggi, tale settore digitale, presenta notevoli tassi di crescita ed è uno dei primi in tal senso, con un notevole sviluppo di app disponibili per smartphones e tablets.

Tali piattaforme di comunicazione istantanea vengono offerte come stand-alone app, come ad esempio WhatsApp, Viber, Facebook Messenger e Skype, o come servizio integrato ad una piattaforma di social network, come Facebook per pc, LinkedIn o Direct di Instagram. Tali piattaforme, si differenziano sulla base di vari elementi e funzionalità, come la possibilità di avere chat di gruppo o solo one-to-one, le forme multimediali, come testo, messaggi audio, video chat, la possibilità di effettuare chiamate vocali e la condivisione della posizione.

Prendendo in considerazione tali caratteristiche strutturali del macrosettore della comunicazione elettronica, la Commissione ha stabilito di restringere il cerchio e l'investigazione al mercato delle *consumer communication apps* disponibili per smartphones. Inoltre, la Commissione affermava che il mercato rilevante, in senso geografico, era quello Europeo, se non addirittura mondiale. Infatti, sia WhatsApp che Messenger hanno elevata diffusione in Europa, mentre se l'analisi si espande a tutto il mondo si nota che le QDM sono più frammentate e che gli usi e le abitudini di una popolazione portano a fare scelte distinte da altre. Tali dinamiche competitive sono il risultato delle preferenze dei consumatori, oltre che dell'ambiente regolatorio e delle condizioni di mercato.

I social networks offrono un'esperienza sociale più ricca rispetto alle consumer communications apps, poiché danno la possibilità all'utente di indicare i propri interessi, attività, creare album fotografici ed esprimere le proprie opinioni mediante le interazioni con i contenuti offerti dalla piattaforma e dagli altri utenti. Le funzionalità delle *consumer communications apps*, oggi giorno, sono più limitate e si focalizzano sulla fornitura di un mezzo di comunicazione rapido ed essenziale, piuttosto che sulla creazione di un'esperienza ricca sull'identità digitale di un individuo.

Inoltre esse forniscono uno scambio di messaggi che avviene in modo istantaneo ed in tempo reale, mentre, tipicamente, nei social network, i messaggi intesi come commenti su un post, non hanno una risposta immediata ed in tempo reale. Inoltre tali messaggi, nei social network, si rivolgono ad una platea di persone più ampia, data da tutti gli amici digitali dell'utente, mentre nelle consumer communications apps, ci si rivolge, tipicamente, ad un contatto telefonico, oppure ad un gruppo di persone ristretto.

Il settore dei social network, infine, si distingue, non a seconda del device o del sistema operativo utilizzato, bensì dall'uso che il consumatore ne intende fare.

Esistono infatti social network con scopo di intrattenimento, come Facebook e Google+ ed altri utilizzati a scopo professionale, come LinkedIn e Xing. Tuttavia, anche qui, ci sono sfumature e funzionalità condivise da entrambe le tipologie di piattaforme. In conclusione, secondo la Commissione, consumer communication apps e social network sono due mercati rilevanti distinti e WhatsApp compete, quindi, con Messenger e non con Facebook, inteso solo come piattaforma di social network.

Focalizzando l'attenzione sul settore dei *consumer communication services*, Messenger possiede circa 250-300 milioni di utenti, di cui 100-200 in Europa e WhatsApp circa 600 milioni, di cui 100-150 in Europa. Ci sono altri players che forniscono la stessa tipologia di servizio ed essi includono sia piattaforme integrate ad altre che forniscono smartphones hardware o sistemi operativi, come iMessage di Apple, BBM di BlackBerry, ChatON di Samsung, Google Hangouts di Google e Skype di Microsoft, sia altre piattaforme di messaggistica integrate a piattaforme di social network, come Twitter, Direct di Instagram e sia altre piattaforme stand-alone, come Viber, Telegram, Snapchat, LINE e Wechat. In tale contesto le *consumer communications apps* competono per fornire al consumatore la migliore esperienza comunicativa. D'altro canto, però, i consumatori sono soliti possedere più di un app di comunicazione e tale fenomeno è detto *multihoming*. Tale fenomeno, come detto nei Capitoli precedenti, abbassa, se non annulla gli *switching cost* e frena, dunque, il PDM delle piattaforme dominanti. Tuttavia, le piattaforme competono soprattutto per accaparrarsi maggior tempo dagli utenti e ci sono piattaforme utilizzate moltissimo ed altre molto poco, nonostante siano entrambe presenti sul device mobile dell'utente. Inoltre, anche la dimensione del network gioca un ruolo fondamentale, infatti, piattaforme con una base di clienti più larga, saranno valutate maggiormente dagli utenti rispetto ad altre piattaforme di nicchia. Infine, anche il prezzo dell'app gioca un ruolo fondamentale.

Infatti i consumatori di communications app sono molto sensibili a variazioni di prezzo e si aspettano che tale servizio venga offerto gratuitamente.

La market share combinata di Messenger e WhatsApp disponibili per i sistemi operativi di IOS e Android, nel mercato Europeo è di circa 30-40%. Tuttavia, in tale settore, la metrica della market share o QDM è poco utile per stimare se un'entità è dominante oppure no.

Alcuni affermano, infatti, che la miglior metrica per tale tipo di settore è il minutaggio speso in un mese dall'utente, in quanto, tale misura riesce a stabilire l'importanza dell'app per il consumatore e il suo potenziale valore attraverso la monetizzazione indiretta tramite pubblicità. Tuttavia, la Commissione, sulla base dei risultati dell'investigazione di mercato, non considera questa metrica particolarmente significativa per analizzare il posizionamento di mercato in tale settore e considera, dunque, la QDM come miglior rappresentante del posizionamento delle imprese in tale mercato.

Tuttavia, sappiamo come in tali settori caratterizzati da cicli di innovazione brevi e *disruptive innovations* che possono cambiare da un momento all'altro gli scenari competitivi, un'elevata QDM non è necessariamente indicativa di presenza di un elevato PDM.

La Commissione ha inoltre notato che ci sono alcune differenze sostanziali tra Facebook Messenger e WhatsApp, riferite a: modi per identificare l'accesso degli utenti alla piattaforma (numeri di telefono per WhatsApp e Facebook ID per Messenger); contatti (numeri di telefono per WhatsApp e lista di amici per Messenger); *user experience* (più ricca in Messenger rispetto a WhatsApp); *privacy policy* (contrariamente a WhatsApp, Messenger raccoglie dati e li fornisce a Facebook a scopi pubblicitari) e l'intensità di utilizzo delle app (in termini di minuti spesi). Gli unici fattori che, pertanto, rendono Facebook Messenger e WhatsApp strettamente *competitors* sono le funzionalità comunicative (ovvero il core business delle due entità) e la dimensione dei rispettivi networks. Tuttavia, la Commissione ha notato che non esistono funzionalità offerte da Messenger e WhatsApp che non sono anche offerte da altri operatori del settore. Infine, come già ripetuto più volte, il *multihoming* rende tale competizione più labile, in quanto, Messenger e WhatsApp rappresentano le due principali piattaforme di communications services per la maggiorparte degli utenti Europei. Ciò suggerisce che tali piattaforme sono complementari piuttosto che in diretta competizione tra loro.

In conclusione la Commissione non considerava Messenger e WhatsApp come strettamente competitors nel settore delle *communications app*. la Commissione ha anche stabilito che l'operazione di M&A non causa un aumento degli *switching costs*; infatti nessuna delle due parti ha il controllo su un sistema operativo per smartphone e quindi nessuno di loro può rendere più complicato passare ad altre piattaforme e non sono nemmeno preinstallate su un'ampia base di dispositivi.

Secondo la Commissione, quindi, non ci sono significative barriere che impediscono ai consumatori di spostarsi verso altre app in tale settore. Non ci sono inoltre nemmeno elevate barriere all'ingresso, in quanto il settore delle communications apps è dinamico ed in forte crescita ed inoltre, sviluppare e lanciare una nuova app di comunicazione non richiede un significativo ammontare di tempo ed investimento.

Il costo principale è infatti rappresentato dai server di raccolta dati che cresce al crescere del network di utenti. Inoltre non ci sono né brevetti, né *know how* né IPR che costituiscono barriere all'ingresso e le tecnologie implementate nelle communications apps sono molto standardizzate. Così la Commissione considerava che tale operazione di M&A non aveva l'effetto di aumentare le barriere all'ingresso e di impedire, di fatto, la competizione nel settore.

L'unica barriera all'ingresso e all'espansione riconosciuta dalla Commissione sembra essere il *Network Effect*. Le due piattaforme, infatti, posseggono congiuntamente circa un bilione di utenti. L'esistenza del Network Effect non implica necessariamente problemi competitivi in mercati caratterizzati da operazioni di concentrazione. Tuttavia è possibile che l'entità fusa possa ostacolare la competizione e rendere più difficile l'ingresso o l'espansione nel settore per i *competitors* potenziali ed attuali.

Tuttavia, in settori come quello delle *communications apps*, caratterizzati da elevata dinamicità e bassi *switching costs*, la posizione di mercato rafforzata dal *Network Effect* non è affatto garantita nel futuro. Inoltre, come già accennato né Facebook, né WhatsApp sono preinstallati su una larga base di devices, cosicché non vi è alcun effetto lock-in che intrappola gli utenti all'utilizzo della piattaforma. Quindi il *Network Effect* esiste nel settore delle *communications apps*, ma la sua presenza non implica, per la Commissione, impedimento e restringimento della competizione. La Commissione affermava che l'operazione di concentrazione tra Facebook e WhatsApp avrebbe portato allo sfruttamento del Network Effect con fini anticometitivi, solo nel caso in cui le due piattaforme avessero unito i propri networks creando un'enorme base di utenti. Tuttavia questo sarebbe stato possibile solo se i due servizi si integrassero in uno solo, con la creazione di un'unica piattaforma. Tale integrazione, secondo la Commissione, richiedeva che i profili degli utenti di WhatsApp e quelli di Facebook, diventassero la stessa cosa. Tuttavia questo non è conforme a quelle che sono le prerogative per l'iscrizione alle due piattaforme: numero di telefono per WhatsApp e creazione di un profilo personale per Facebook. Le piattaforme infatti non sono in grado di associare il Facebook ID e il numero di telefono di WhatsApp e viceversa. Quindi tale associazione degli accounts doveva essere fatta manualmente dagli utenti, che, in maniera del tutto volontaria, inserivano il numero di telefono di WhatsApp associandolo al proprio account di Facebook. In conclusione, dunque, la transazione non comporta, secondo la Commissione, problemi strutturali sul livello competitivo dell'industria grazie al *multihoming* e alla dinamicità della stessa e, soprattutto, non si può parlare di *Network Effect* congiunto, in quanto è impossibile per nessuna delle due piattaforme associare due *account* dello stesso utente e per l'esistenza di una vasta mole di consumatori che utilizzava già entrambe le piattaforme, senza quindi causare un guadagno netto di utenti per le due entità fuse.

I Big Data entrano in gioco nella transazione, considerando il mercato della *on-line advertising*. In tal senso l'operazione di concentrazione può comportare una potenziale data concentration, rafforzando, dunque, la posizione di Facebook in tale settore, in particolare in riferimento alla *display advertising*.

Poiché solo Facebook è attivo in questo mercato, WhatsApp non raccoglie dati sugli utenti circa la loro età, nome e cognome, genere, gruppo sociale, attività svolte sulla piattaforma, abitudini di consumo e altre caratteristiche di valore per scopi pubblicitari. Inoltre, come detto, WhatsApp non registra nemmeno i contenuti dei messaggi. Così, per tali motivi, l'operazione di concentrazione non porta ad un incremento dell'ammontare di dati disponibili per Facebook ai fini pubblicitari.

Tuttavia, la Commissione ha analizzato due possibili danni potenziali derivanti da un rafforzamento e della posizione di Facebook nel mercato della *on-line advertising*: l'introduzione di pubblicità su WhatsApp e l'utilizzo di WhatsApp come potenziale mezzo di raccolta di dati col fine di incrementare l'efficacia della personalizzazione della pubblicità di Facebook.

Secondo tali teorie, dopo la transazione, WhatsApp avrebbe dunque dovuto aggiungere pubblicità on-line sulla propria piattaforma, raccogliendo ed analizzando i dati degli utenti di WhatsApp. Questo avrebbe avuto l'effetto di rinforzare il PDM dell'entità risultante dalla fusione nel settore della pubblicità on-line. Tuttavia, questo non è accaduto e WhatsApp attualmente non vende pubblicità. Infatti WhatsApp avrebbe dovuto cambiare la sua *privacy policy* e ciò avrebbe potuto causare il malcontento di tutti quei consumatori che ripongono un elevato valore sulla *privacy* e sulla sicurezza. Inoltre la concentrazione non ha avuto alcun effetto sugli inseritori pubblicitari, che, in virtù del *multihoming* e dell'elevato grado di differenziazione del mercato della pubblicità *on-line*, avrebbero potuto passare ad altre piattaforme.

Infine, i dati a cui WhatsApp ha accesso, secondo la Commissione, non sono in grado di aumentare l'utilità marginale di Facebook nell'elaborare informazioni di tipo commerciale, non permettono quindi a Facebook di fornire contenuti maggiormente proliferati verso gli utenti. Infatti WhatsApp raccoglie solo dati come il nome e il numero di telefono associati all'account dell'utente, sebbene ci siano dubbi circa l'utilità del numero di telefono ai fini di incrementare l'efficacia della pubblicità offerta da Facebook. Tuttavia la Commissione affermava che Facebook non aveva alcuna intenzione di modificare le strategie di WhatsApp in relazione alla raccolta e all'analisi di dati e non c'è, dunque, alcun effetto sulla potenzialità dei dati disponibili per Facebook ai fini dei contenuti pubblicitari personalizzati. Inoltre, come detto, se WhatsApp avesse cambiato *privacy policy*, a seguito della fornitura dei propri dati a Facebook, molti utenti avrebbero cambiato piattaforma passando ad un'altra meno intrusiva. Questo, dunque, ha ridotto l'incentivo di Facebook di raccogliere dati tramite WhatsApp e di trasmetterli alla piattaforma di social network. Inoltre la Commissione affermava che, anche se Facebook utilizzasse WhatsApp col fine di migliorare l'efficacia della sua pubblicità personalizzata, ci continueranno ad essere numerosissimi dati sugli utenti di Internet disponibili per altre piattaforme e, quindi, al di fuori del controllo esclusivo di Facebook. Ci sono, infatti, numerosi altri *player*, oltre Facebook, che raccolgono ed analizzano i dati degli utenti a fini commerciali. Essi includono Google, che è il player con maggiore QDM nel macrosettore della pubblicità on line, Amazon, Apple, Ebay, Microsoft, Yahoo!, Twitter, IAC, LinkedIn, Adobe e Yelp.

In conclusione, basandoci sui risultati dell'investigazione di mercato condotta dalla Commissione Europea, la transazione non suscita seri dubbi sulla sua compatibilità col mercato interno e con il grado di competizione nel settore della *on-line advertising*. Perciò il 03/10/2014 la Commissione dichiarava compatibile l'operazione di acquisizione di WhatsApp da parte di Facebook secondo l'Art 6 della Merger Regulation.

Bibliografia

Access Barriers to Big Data – Michal S. Gal, Roma, 9 Novembre 2016

Algorithms and Collusion, Competition policy in the digital age – OECD (2017)

Atti del convegno “Piattaforme digitali: Concorrenza, Fisco, Innovazione” – Diego Ciulli

Background information on the Facebook proceeding – Bundeskartellamt, 19 Dicembre 2017

Big Data Analysis and Storage – Khalid Adam Ismail Hammad, Mohammed Adam Ibrahim Fakhaldien, Jasni Mohamed Zain, Mazlina Abdul Majid, Orlando, 10 Settembre 2015

Big Data and Differential Pricing – Obama White House, Febbraio 2015

Big Data as a Misleading Facility – Giuseppe Colangelo e Mariateresa Maggiolino (2017)

Big Data e condotte anticompetitive – Mariateresa Maggiolino, Milano, 18 Gennaio 2018

Big Data e Concorrenza, quale rapporto? (seminario formazione giornalisti) – Gabriella Muscolo, 25 Ottobre 2015

Big Data e economia digitale: il valore dei dati e le nuove sfide per l’antitrust Andrea Pezzoli, Roma, 18 Gennaio 2018

Big Data Storage (Chapter 7) – Martin Strohbach, Jorg Daubert, Herman Ravkin, Mario Lischka

Big Data Storage Architecture Design in Cloud Computing – Xuebin Chen, Shi Wang, Yanyan Dong, Xu Wang, Tangshan, Hebei, China (2016)

Big Data Storage and Challenges – M.H.Padgavankar, Dr.S.R.Gupta, Maharashtra, India (2014)

Big Data tra Privacy e Antitrust – Giuseppe Colangelo

Case No COMP/M.4731 Google/DoubleClick – Regulation (EC) No 139/2004 MERGER PROCEDURE, 11 Marzo 2008

Case No COMP/M.7217 – Facebook/WhatsApp – Regulation (EC) No 139/2004 MERGER PROCEDURE, 03 Ottobre 2014

Corporate tax and the digital economy – HM Trasury, Marzo 2018

Cracking the Code in Private Equity Software Deals”, BCG Focus, May 2017

Data in the EU: Commission steps up efforts to increase availability and boost healthcare data sharing – EU Commission, Brussels, 25 Aprile 2018

Economia digitale e Big Data - MICHELE AINIS, Roma, 25 ottobre 2017

EU Competition Law Enforcement vis-à-vis Exploitative Conducts in the Data Economy Exploring the Terra Incognita – Marco Botta and Klaus Wiedemann

Federal Trade Commission, the Antitrust Laws

L’economia dei dati – tendenze di mercato e prospettive di policy – It Media Consulting, Roma, Gennaio 2018

L’impatto di Google sull’economia”, Deloitte (2014)

L.M. Cabral, Economia Industriale, Carocci editore

Online advertising – Bundeskartellamt, Febbraio 2018

Recommendation system, an introduction, Dietmar Jannach, Dortmund (DE), 2014

Storage for Big Data and Analytics Challenges – Infinidat (2016)

Storage Architectures for Big Data in the Cloud – Sam Fineberg, Maggio 2013

The Netflix Recommender System: Algorithms, Business Value, and Innovation, Carlos A. Gomez-Uribe and Neil Hunt

The Role of “Big Data” in Online Platform Competition – Andres V. Lerner, 26 Agosto 2014

Thomson ONE Banker; BCG analysis