Dynamics of the network economy: a content analysis of the search engine trends and correlate results using word clusters

Murat Yaşlıoğlu

School of Business,
Istanbul University, Turkey
Email: murat@yaslioglu.com

Abstract: Network economy is a relatively untouched area, strategic approach to the dynamics of this new economy is quite limited. Network economy is about the networks. Thus, it was decided to follow up the information on the internet including almost every kind of documentation. First, a deep relation analysis using trends was conducted to find out the related topics to new economy's dynamics: network effect, network externalities, interoperability, big data, open standards and social media. After the relation analysis, correlates of aforementioned keywords were analysed. Finally, all the clean 'top results' on the web were collected by the help of Linux command line tools into various, large text files. These files were analysed by the help of Nvivo qualitative analysis tool to form clusters. By the broad information available at hand, an extensive discussion on each result is written. It is believed that this new research approach will also guide many future researches on various subjects.

Keywords: network economy; network effect; network externalities; interoperability; big data; open standards; network strategy; methodology; analytics; word clusters; search engines.

Reference to this paper should be made as follows: Yaşlioğlu, M. (2020) 'Dynamics of the network economy: a content analysis of the search engine trends and correlate results using word clusters', *Int. J. Data Analysis Techniques and Strategies*, Vol. 12, No. 1, pp.1–29.

Biographical notes: Murat Yaşlıoğlu is an Associate Professor of Management and Strategy at the Istanbul University, Istanbul Business School, Department of Management and Organization. He had his PhD degree in 2013 and teaches mainly strategic management and research methodology in various state and private universities.

1 Introduction: network economy

The narrow point of view on trade surplus and deficit has failed against the complex nature of digitised economy. While the great recession of 2008 suppressed global trade and financial flow; on the contrary, the flow of data has increased drastically (Blaug, 1997; North, 2016; Tapscott, 2015). This connectivity binds the world tighter than flow and trade of manufactured goods. Since 2004, compared to the flow of manufactured goods and services, the amount of data flow has grown 50 times larger. Moreover, it is

Copyright © 2020 Inderscience Enterprises Ltd.

expected to become ten times larger in the upcoming five years (Bauer and Erixon, 2016). With the help of digital platforms such as Alibaba, Amazon, eBay the costs of doing business is shrinking and eventually small businesses are becoming tiny multinationals. These changes surely bring challenges along with opportunities. Global competition for all is boosting, price pressure is severe and new digital models are emerging every other day (Goldfarb et al., 2015). The network of data is vulnerable to cyber-attacks. More interestingly the data created becomes more chaotic and massive alongside their possible offerings. 21st century's 'globalisation' depends much on intensive intangible flows and participation of emerging economies. Network infrastructure and platforms are becoming almost as important as transportation. Perhaps most importantly, the exchanges of free services are taking the place of monetary transactions (Manyika et al., 2016). The components of the network processes differ in many aspects, these aspects may be accounted for manufacturing too. However, this new way of value creation requires companies to have some distinct capabilities such as; being more agile, leaner and more data driven. Small and medium sized enterprises have the opportunity to adapt themselves more easily compared to bigger corporations.

2 Literature review: an introduction to network economy

Classical economic theory and strategic management theory are based on sort of an action and reaction principle. It is suggested that if a product's or service's demand increases, as a reaction by providers, its supply is also expected to increase. Eventually, as the supply increases, because of abundance of current production, the price of that particular product/service is expected to decrease. In a free market, the equilibrium price is set by the invisible hand (McGuire and Olson, 1996). A greatly abundant product/service is almost free, a scarce one however is more expensive. That is the only reason water is cheaper than petroleum; coal is cheaper than its elder brother, diamond (Pellegrin, 2016; Porter, 2011).

Strategy theory suggests certain ways of gaining competitive advantage; with increasing perceived value of the products/services by differentiating, producing with less costs (cost leadership) or finding an unexploited area of the market and generating a 'niche'. In all cases, with its more recent name, companies should 'innovate'. Either innovate their products/services or the way they produce them. Innovation is done through finding new ways of creating form, place, time or possession utilities. If the utility created is 'new' enough to be scarce, or otherwise different, then value assigned to the whole will increase. Innovation by its definition should increase value. Resource-based view tries to come up with an answer to the question "why some companies have superior profits than others?" and assumes that a firm's resources and capabilities are the primary drivers of competitive advantage and economic performance [Narayanan, (2000), pp.100–134; Lengnick-Hall, (1992), pp.399–406; Leonard-Barton, 1995]. This implies, if a firm's resources are valuable and rare, then it shall create competitive advantage; if they are also hard to imitate and well organised to exploit then this competitive advantage is sustainable. Long thing short, competitive advantage in this perspective is based on rarity, in other words 'scarcity'.

However, what we call network dynamics are quite different. Value creation is not only based on scarcity or rarity. On the contrary, digitalisation lead to a way that rare is not valuable, abundance of digitals creates more value than their scarcity. A smartphone

has no value without billions of users, an operating system will be less valuable if only a few users happen to use it, a game console is useless with its range of games and users, internet of things is no 'things' without their variety and count. For example, a video cassette standard VHS has become more successful compared to its superior Betamax because it was more abundant; mobile networks' value increases with more users connecting. All of aforementioned examples are the results of so-called 'network effect'.

The economic theory behind the network effect is 'network externalities'. Network externalities represent economies of scale on the demand side. It commonly exists in network industries which happen to connect devices or things with equivalents or related ones. This is why a consumer simply wants to register on a larger social network when he/she is to choose between competitors. This is the reason Facebook, Twitter, Instagram are the most valuable and profitable companies worldwide. Katz and Shapiro (1994) have defined two types of externalities:

- *Direct network externalities:* a consumer's utility depends not only on price but also on a network scale in the service. Because sole benefit of the service is less than its social benefit. This is why a phone network with more users is attractive because of easier and cheaper connectivity to other clients.
- *Indirect network externalities:* even though the scale of the network does not directly affect the consumer's choice, the options and opportunities comes thereby are more variable. This is why an operating system, game platform or a hardware base which come with more support is preferred. For instance, network effects are famous for causing lock-in with the most-mentioned instances like Microsoft products and the QWERTY keyboard.

Network effect strongly relies on expectations of the consumers. If consumers expect greater expansion on an interconnection, then they will assign a higher value to it. Eventually, "it will actually take precedence over others in the manner of a self-fulfilling prophecy" (Ida, 2009). Aforementioned externalities are not necessarily opposed to but quite outbound the definition of rarity or scarcity in classical definition. But we will try to put these approaches together later.

The utility and therefore value increase through network externalities or 'scale of economies on the demand side' can be achieved through (Ida, 2004; Katz and Shapiro, 1985):

1 *Value generated by direct involvement:* when a customer subscribes to the network, he/she directly gains advantage because of wide connectivity and information network. For instance, a new subscriber to WhatsApp network will benefit free messaging service with widest range of users. The total value of the product therefore is higher than the product itself without connectivity.

2 *Value generated for current network members:* current members' total value increases with new subscribers. Because current subscribers will enjoy benefits of increased network and information flow.

3 *Value generated for potential members:* every new member to the network will increase the expected value for potential subscribers to the network.

The generation of value within network economy also has some prerequisites. Firstly, whether or not a platform has emerged 'interoperability' is a must among interfaces on

vendor and customer side. Secondly, open rather than closed standards are more preferable to boost the interoperability without lock-in, as a safeguard of the economic system. Interoperability is a characteristic of a product or a system where its interfaces can be integrated to and as well as work with other systems or platforms. Interoperability should apply to current features and/or future developments. Interoperability can be measured by the ability of different domains to exchange information, services, goods to attain shared goals. Hence, it implies systems to work together in a harmony. Interoperability manifests the potential to enlarge the network and therefore increases the appeal to the customers (Honkola et al., 2009). With increasing appeal, as network economy suggests, expected external value of network rises. It is mainly attained by increasing the number of potential connections and by attracting new clients to the network. Even though it has various degrees, interoperability with higher levels can help to reduce uncertainty, lock-in and severe price competitions (Shapiro and Varian, 2013). Companies striving to achieve higher levels of interoperability should cooperate and standardise their products/services. This brings companies to a stage where they should decide whether to cooperate with their competitors.

Open standards are created by various vendors, academics, professionals, e.g., together and these standards are open to public or any others' use. The common most common examples for this are 'open source software', to which public or individual users can also contribute. Various vendors and individuals (even those who were not part of the original group) can benefit the standards' documents to create products that implement the common definitions in the standard (Hippel and Krogh, 2003; Lakhani et al., 2003). And thus, they are '*interoperable by design*', with no specific liability or advantage for any customer for choosing one product over another on the basis of standardised features. Because of open standards and eventually interoperability, companies have to compete on the basis of the quality of their implementation, price/performance ratio, usability of their interface, marketing activities, unconstrained supply and hyper-scaling the possible platforms. Closed standards of communication protocols and/or interfaces under the control of a sole company may give it the opportunity to exploit the network effect up to a level of monopoly. Microsoft was broadly acknowledged by others to keep its monopolistic position through aforementioned means for a very long time (Fitzgerald, 2006). "It is impossible to predict which firm will be able to appropriate the economies of network because a path dependency can occur, depending on the initial condition. Also, since positive feedback is caused by the economies of network, lock-in occurs and excess inertia or momentum matters. Consequently, the complex network economy is very different from the simple world that standard economics, like the general equilibrium theory, has presumed" (Ida, 2012).

3 Research methodology: application of a new approach

Previous section focuses on differences among current strategies and network economy requirements of strategy. It is mostly mentioned in the literature that the network effect is merely possible through "interoperability, exploitation of big data, application of open standards, enlarging the network itself and network externalities." But, as investigated and researched by some scholars, putting them in a formal 'strategy basis' is yet very hard (Shapiro and Varian, 2013; Achrol and Kotler, 1999; De Man, 2004; Applegate et al., 2002; Economides, 1996; Choi and Whinston, 2000; McGee and Sammut-Bonnici,

2002; Murdoch and Detsky, 2013; Meltzer and Glushko, 1998; Bilbao-Osorio et al., 2013; Zhu et al., 2006). More importantly, it is perhaps harder to come up with an answer to the question; "what distinguishes sources of success from failure in network economy where externalities are already present?" There are many industry examples indicating, the companies are aware of the possibilities and opportunities network economy offers; but, the lacking side is a foundation for the theory behind competitive advantage within networks. The broadest and most grounded researches in this area are Shapiro, Katz and Varivan's works on network effects and competition (Katz and Shapiro, 1985, 1994; Shapiro and Varian, 2013).

Firstly, we use analytic tools to investigate the importance of the topic and show the trend on network economy, using keywords' search trends on the world wide web. Key assumption of this part is: "given the importance on related subject, the subject should be processed into the web or network itself." The keywords which are looked on the internet-based search trends are; "network effect, network externalities, interoperability, open standards, big data." These keywords are taken from the literature as key elements of the network economy. Additionally, the keyword 'social media' is also added. Because, the most infamous result of the network effect and therefore, the biggest platform that network economy governs is the 'social media'. Therefore, evaluation of social media is accounted a good case to observe the impacts of *network effect and network economy*.

To reveal the importance assigned by users, worldwide query of the keywords and related terms should have been extracted. For this purpose, a trend analysis is on Google search engine for last 60 months (five years) is conducted (Choi and Varian, 2012; Carneiro and Mylonakis, 2009). The trend analyses results are categorised as:

1 *Interest over time:* numbers represent search interest relative to the highest point on the chart for the given region and time (for this research: worldwide, five years). A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. Likewise, a score of 0 means the term was less than 1% as popular as the peak. Hereby, another important metric is the keyword rank through the entire worldwide web, because interest over time scores degrees the peak and bottom interest for any given keyword. So, to indicate the importance of the keyword or topic an absolute indicator is necessary. Monthly average search volume is presumed to be a good reflector.

2 *Related topics:* "users searching for your term also searched for these terms." These metrics are categorised as 'top and rising'. '*Top*' *metrics* are the most popular terms. Scoring is on a relative scale where a value of 100 is the most commonly searched term, a value of 50 is a term searched half as often, and a value of 0 is a term searched for less than 1% as often as the most popular one. '*Rising*' *metrics* are related terms with the biggest increase in search frequency since the last time period. Results marked '1,000' implies tremendous increase, probably because these terms are new and had few (if any) prior searches.

These results are filtered, aggregated and cut down in relation to the given keywords to accumulate information to be used for third stage of the research which intends to qualitatively analyse and cluster most relevant information on the world wide web (Sharma and Suman, 2013). Results are gathered together with second stage of the research which seeks correlation thru web search results of any given period; Appendix 1.

congregates the abridged and simplified results of this aforementioned, initial stage of the research.

Results for *interest over time* and *related topics* gives out many clues and insights on a term or topic which is being investigated. But, these insights may be missing in perspective. Answer to the following question is crucial; "what might be also trending as the given topic trends?" or in other words "what are the positively (or perhaps negatively) correlated search queries with a given term?" So, as second stage of the research; most correlated internet queries with the researched topics will give us some clues about the applications, practises, related trends and places. So-called second stage and the first stage are complementary to each other; first stage looks for keyword relations while second stage seeks for insights from web content big data. Without direct relation to keyword trends, insight may have been missing.

Web search activity correlations has previously been shown useful for providing estimates of real-world activity in a variety of contexts, with the most common being health and economics. Examples in health include influenza (Ginsberg et al., 2009; Eysenbach, 2006), acute diarrhea, chickenpox, listeria (Pelat et al., 2009; Wilson and Brownstein, 2009). Examples in social sciences include movie box office sales (Goel et al., 2010), computer game sales, music billboard ranking, general retail sales, automotive sales (Choi and Varian, 2009), travel, investor attention (Da et al., 2011), and initial claims for unemployment (Choi and Varian, 2012). *Web search correlation tool* searches across millions of candidate query time series to find the best matches. The objective of search engine correlation is to surface the queries in the database whose spatial or temporal pattern is most highly correlated (R^2) with a target pattern. Method employs a novel approximate nearest neighbour (ANN) algorithm over millions of candidate queries in an online search tree to produce results similar to the batch-based approach (Vanderkam et al., 2013). The technique used for this calculation is a combination of vector quantisation called as 'asymmetric hashing (AH)' and a second order pass exact search (Gersho and Gray, 1992). AH is used to calculate approximate distances among search vectors. It is done through mapping the 'Pearson correlation' distance onto a simpler form (Bagui et al., 2009; Shrivastava and Li, 2014).

Using big data environments to come up with causal research is virtually impossible, the best way to make sense of this amount of information is an attempt to find correlations of keywords with other sources (Vanderkam et al., 2013). *Correlate* is like trends in reverse. With trends (please refer to initial stage of the research), you type in a query and get back a series of its frequency (over time, or in each place). With correlate, you enter a data series (the target) and get back queries whose frequency follows a similar pattern (Mohebbi et al., 2011).

The AH algorithms for Euclidean distance and for Pearson correlation with missing indices both operate in three steps: the construction of a lookup table, the calculation of approximate distances using that table and, finally, a reordering of the top candidates using exact distances. For the distance metrics used for correlate, the standard definition of Pearson correlation between two times series are used.

$$r(u,v)=\frac{\operatorname{cov}(u,v)}{\sigma_u\sigma_v}=\frac{\sum_{i=1}^{n}\left[\left(u_i-\mu(u)\right)\left(v_i-\mu(v)\right)\right]}{\sqrt{\sum_{i=1}^{n}\left[\left(u_i-\mu(u)\right)^2\right]}\sqrt{\sum_{i=1}^{n}\left[\left(v_i-\mu(v)\right)^2\right]}}$$

Then a distance function is calculated with the standard definition of Pearson distance function [$dp(u; v) = 1 - r(u; v)$]. Searching in AH works by calculating approximate distances between the target and database vectors. For AH, Pearson correlation distance is replaced by a simpler function by the help of squared Euclidean distance.

$$u' = \frac{u - \mu(u)}{2N|u - \mu(u)|}$$

$$v' = \frac{v - \mu(v)}{2N|v - \mu(v)|}$$

And then

$$d_p(u, v) = |u' - v'|^2$$

To index a vector v, find the centroid closest to v in each chunk. Concretely, set.

$$h_i(v) = \arg\min\left(\left|\pi_i(v) - c_i^j\right|\right)$$

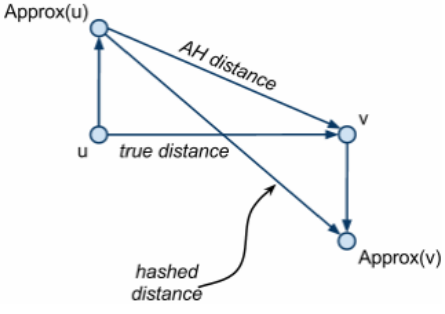
$$h(v) = \left(h_1(v), h_2(v), ..., h_{N/k}(v)\right)$$

For each vector, this results in N/k integers between 1 and 256, one for each chunk. Combine the integers into a N/k byte hash code, $h(v)$. AH then, is used to compute approximate distance between vectors.

$$Approx(v) = \left(c_1^{h_1(v)}, c_2^{h_2(v)}, ..., c_{N/k}^{h_{N/k}(v)}\right)$$

Unlike algorithms involving locality-sensitive hashing, AH only hashes the database vector, thus eliminating one source of error (Liu et al., 2004, 2009; Vanderkam et al., 2013). An illustration of this reduced error can be seen in Figure 1.

Figure 1 Asymmetric hashed distance of vectors (see online version for colours)



Nearest Approx(v) results for any given 'u (entered topic)' then taken into account whilst showing the correlations (Stephens-Davidowitz and Varian, 2014). Results of top approximates are congregated on Appendix 1 along with results of first stage of the research. Both of these stages' results are used to create a 'mind map' for later stage of the analyses.

The later and final stage of the research includes the qualitative analysis and text mining of the content of all queries for the keywords both obtained initially from theory

and from research stages. Every keyword is ranked using SERPs keyword rank checker tool[1] within top two search engines (Google and Yahoo!). SERPs' tool ranks top 200 web results for each keyword. For our research, approximate amount of data processed is about 1,400 (7 keywords * 200 web results for each) website size (around 10,000 pages of raw text data). Websites are dumped to text files *without* hyperlinks. In order to do so, a batch file with Linux command line was created. The command line tools and batch execution script can be found in Appendix 2. Raw data of text-after being cleaned from advertisements, links, unrelated stuff, irrelevant menu items and others is analysed by the help of Nvivo text analytics tool to create a cluster tree.

Web content is deliberately divided into two sections for convenience. First section is only for theory keywords (network economy, network effect, network externalities, interoperability, big data), and second section is for the keyword 'social media[2]'. Finally, at the conclusion of the research; trends, correlate results and logical trees are again analysed together in order to come up with managerial and academic implications (Chauhan and Kaur, 2017).

Figure 2 Research methodology summary

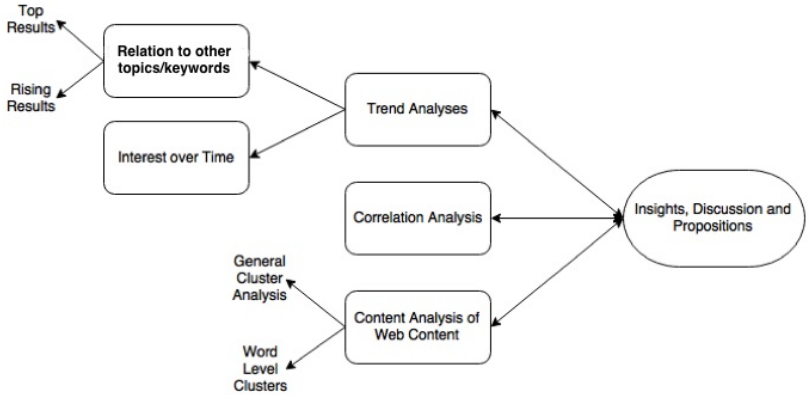


Figure 2 summarises the analyses described in detail as methodology follows three distinct but interconnected steps to come up with concise and comprehensive insights:

1 Keywords' relation is investigated through search engines' top and rising trends results. (Please check Appendix 1 for results)

2 Keywords' correlation with ongoing searches is analysed. (Please check Appendix 1 for results)

3 Keywords' top search engine results are accumulated, downloaded, clarified and cleaned to be used for a content cluster analysis using Nvivo.

4 Results

Results from the 'related search analysis' for 'top and rising results' together with 'correlate analysis' are summarised and congregated in Appendix 1. A node was created using all sources and references downloaded from the web, *node level* cluster analysis[3] of the nodes helps us to categorise the literature. Figure 3 shows the results of this analysis.

Figure 3 Nodes clustered by word similarity using Pearson correlation coefficient (see online version for colours)

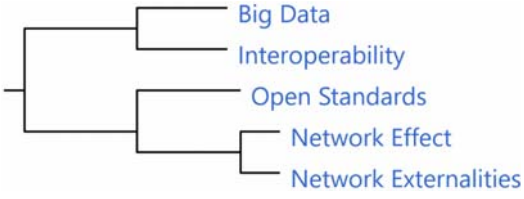


Table 1 Word level cluster analysis

Network economy			
Network effect and externalities	*Open standards*	*Big data*	*Interoperability*
• Firms	✓ Work and implementation	➢ Research	✓ Management
• Markets	✓ Organisation policy	➢ Company	✓ Learning
• Benefits	✓ Open source	➢ Consumer	----------
• Externalities	✓ Proprietary rights	➢ Analytics	o www and .gov
• Cost	✓ Government	➢ Pdf	o Privacy
• Computer	✓ Innovation	➢ Data reports	o Public organisation
----------	----------		o Tech. support
➢ Network	❖ Internet		o Provide needs
➢ Demand	❖ Software property		o Security
➢ Economics	❖ Web development		----------
➢ Competition	❖ Process standards		➢ Information systems
➢ Compatibility	❖ Digital business		➢ Information services
➢ Size and scale			➢ Quality
➢ Social networks			----------
➢ Consumers			✓ Healthcare
➢ Product and price equilibrium			✓ Clinical patient
➢ Value model			✓ Electronic exchange and interoperability

✓ Users			
✓ Standardisation			
✓ Better and economic products			

Note: Incl. sub clusters marked with different type of bullets.

As figure indicates; big data, interoperability, open standards, network effect and network externalities clustered separately. These results lead us to investigate the network economy in four separate contexts:

1 network effect and network externalities

2 open standards

3 big data

4 interoperability

A *word level* cluster analysis[4] resulted with insights deeper than contextual division. Appendix 3 shows the results of the word level analysis. For convenience purposes word level analysis is compacted in Table 1.

Table 1 includes the *Pearson correlation cluster* results from the web content of the keyword analysis. All the results of the analyses are used to come up with discussion and implications.

5 Discussion of the results and implications

Network effect and its 'sine qua non', network externalities are only existent where computer networks are available. Computer networks offer the possibility of linkage through devices. Network economy is mostly about value creation through connection of things, most of the time by the internet (Uzzi, 1996). Network externalities are about size and scale of the network created; and this network can only be sized enough for competitive advantage if the business/value model is supported by compatibility of the platforms those are supposed to operate it. Even though the platforms are different, it is crucial to create better and more economic products/services with highest degree of standardisation across platforms. Thus, another natural result of network effect and externalities is cost advantage, as each user joining the network actually increases value and decreases marginal costs per unit.

Resource-based view suggests that a core competence is crucial for competitive advantage and strategic success; new value strongly depends on the possibilities offered by connectivity. However, resource-based view also suggests that a strategic resource must be rare in order to create competitive advantage. But on the contrary, connectivity is all about the abundance. So, the only possible and remaining way to grasp and sustain competitive advantage is through a perfect and rare business model (Lavie, 2006). This type of model should create some economies of scale on the demand side and actually should benefit a spill-over effect from networking possibilities. If this spill-over from network grows big enough to sustain a competitive advantage, it becomes a huge barrier to entry and perhaps leads to a monopoly (Sjöholm, 1999). This is clearly visible on social networking sites such as Facebook, Instagram or messaging apps such as WhatsApp. Some of the most advanced examples for these business models are Uber, AirBnB and Bitcoin.

Uber taxi service is a classic two-sided marketplace where more cars on the network attract more travellers and vice versa. But, it is not simply the number of cars and travellers that attract each other but the levels of participation of both sides. A higher participation from drivers is useful only if it results in higher availability and consequently, lower waiting time for passengers. Similarly, a higher participation from passengers is useful to drivers only when it means lower downtime and, potentially, the ability to charge higher prices (Cramer and Krueger, 2016).

In Airbnb's case, it enjoys marketplace-like network effects (example: eBay) in that the value of the network increases as both suppliers and consumers join. Every time a new supplier joins, number of options increase for the travelling consumer, making it more likely to attract demand. Every time a travelling consumer searches, it potentially benefits the property owner/manager, making it more likely to attract suppliers. That's a flywheel that, properly spun up, starts to accelerate partially from its own momentum (Murhc, 2014).

Bitcoin on the other hand is by its nature a money unit that is created, transferred and exchanged on the internet. It is a type of digital currency in which encryption techniques are used to regulate the generation of units of currency and verify the transfer of funds, operating independently of a central bank.

Correlation analysis of the network effect and externalities gives us some out of the context clues, each can be investigated more deeply in several researches. The correlation results and insights for network effect are:

- *Common crimes:* possibly, some common crimes are results of internet connectivity and social media where people familiar and unknown to each other can reach each other with less social boundaries and limitations.
- *Media messages:* most of the media messages, besides traditional ways, travels through the networks. The network effect and network externalities foster these messages to spread faster, easier but uncontrolled and unconstrained.
- *Morally right:* social networks and use of information technologies are shaping and being shaped by moral values and ethics. Since social boundaries and social pressure become less effective on people, we feel freer to do whatever we want at the cost of some moral and ethical values. 'Common crimes', which showed up in correlation analysis and mentioned above, is also a side effect of this.
- *Social rules:* this correlation also supports network effects' relation to social rules and norms. It is quite related to common crimes, media messages, moral and ethical values.
- *Direct competition:* network effect is created through the connectivity of the devices as aforementioned, acting on a limited number of platforms. This has a by effect meaning companies are exposed to direct competition, that is a situation in which two or more businesses offer products or services that are essentially the same; as such, the businesses are competing for the same potential market (Rao et al., 2000).
- *Universal language:* this concept refers to a hypothetical or historical language spoken and understood by all or most of the world's population. In some contexts, it refers to a means of communication said to be understood by all living things, beings, and objects alike (Markel, 1996). Being vastly interconnected may require a universal language, beginning with computers to understanding each other on a single coding, extending to humans connected to use a single language to be able to connect. Yet, English is somewhat filling this gap.

The results for '*network externalities*' side of the correlations analysis can be listed and described as:

- *Media effects:* "media influence is the actual force exerted by a media message, resulting in either a change or reinforcement in audience or individual beliefs. Media effects are measurable effects that result from media influence, or a media message. Whether that media message has an effect on any of its audience members is contingent on many factors, including audience demographics and psychological characteristics. These effects can be positive or negative, abrupt or gradual, short-term or long-lasting. Not all effects result in change: some media messages reinforce an existing belief. Researchers examine an audience after media exposure for changes in cognition, belief systems, and attitudes, as well as emotional, physiological and behavioural effects." With economies of scale on the demand side, and also with its result network effect; media messages (also enlisted under correlations of network effect) spread hundred times faster than it was ten years ago (Potter, 2012). Being this infectious, media effect also becomes more striking and efficient; either positively or negatively.
- *Technology effect:* network externalities require technology to be widespread and reachable to occur. So, being too much connected and dependent on technology also have side effects on social, psychological, physiological nature of people. One of the most criticised effects of technology is so-called isolation, depression, lack of social skills, obesity, stress, poor sleeping habits, pain, constant distraction, e.g., (Tenner and Rall, 1997).
- *Sociological terms:* as other results also indicated both for network effect and externalities sociological terms are prone to change and distortion by the networking through means of technology. This area seems to be a very problematic among all. Not to forget, correlation results are evident symptoms of the main search terms. These are the unnoticed side effects of things and concepts which people sometimes do not even notice.
- *Maximum matching:* maximum matching is investigated under the graph theory, a maximum matching in is a matching that contains the largest possible number of edges connecting the devices, computers, people etc. There may be many maximum matchings in a given set of units or devices (Vazirani, 1994; Cormen et al., 2001). One of the significant challenges in assessing the role of externalities is that the underlying networks generating them are often unobservable or difficult to pin down. Connections among devices might be accomplished by means of different communication networks according to connectivity and service requirements. In particular, while attributes such as income, professional qualification, and education are frequently available, other important measurements of social connection – friendship, a shared professional history, etc., are more difficult to obtain. Beyond the scarcity of data, the matching literature lacks a definitive framework to account for externalities, while still enabling empirical evaluations (Baccara, 2012). This insight offers some possibilities for future research; for instance, possible matching of devices to create maximum connectivity, evaluation of the efficiency of matching protocols in terms of welfare, social ties, demand and supply for the identified externalities are some key areas to investigate in relation with maximum matching.

- *Benefits of the internet:* internet is both a platform and an element of network externalities. Network effect nowadays is created through internet and its offerings in the form of social media, internet of things, e.g.

Interoperability is a characteristic of a product or system, whose interfaces are completely understood, to work with other products or systems, present or future, in either implementation or access, without any restrictions. Internet searches on interoperability shows a high interest with 12,000 searches per month and an average score of 80 over 100. Its assigned importance is an evidence and result of its usage in health and emergency sectors. Among all, data and system interoperability seems the most important subjects in this area. 'Data with data: application and operating systems' interoperability' and 'system with system: interoperability of the architecture', finally 'data with system: software and hardware interoperability' are the sub-topics of interoperability. Incident command systems, electronic health records system, federal agency system, unified computing (like Cisco), virtual machine systems (e.g., VMware and parallels), cloud storage and computing are the most intense applications of interoperability among various types of information systems (Malhotra, 2010, Brailer, 2005; Sheth, 1999). Within every each of them there is a vast amount of different types of operating systems, hardware, software computer languages; but yet systems still communicate, work and exchange flawlessly without any compatibility issues. Interoperability necessitates great deal of management and organisation activities. This is because all the systems vary within and among each other. Moreover, persons and organisations using these systems are much different from each other. Then we can say that the management and organisation functions required to interoperate such systems are not only required on intra-user level but also on inter-user level.

The results for '*interoperability*' side of the correlation analysis can be listed and described as:

- *Java interface:* Java is a computer language with few implementation dependencies as possible. It can run on any platform regardless of computer architecture, sandboxed in a virtual machine. It is referred most of the time as 'write once, run anywhere'. Java actually is a contemporary key for interoperability, actually its use for 'internet of things' offers many possibilities. Internet of things implies the interconnection via the internet of computing devices embedded in everyday objects, enabling them to send and receive data (Breg et al., 1998). To be able to do so, highest degree of interoperability is crucial, and java interface offers a blueprint to it. That is why it is not a surprise to see it correlate to interoperability with a coefficient around 0.99.

- *Software development, quality and design:* all of the three keywords correlate over 0.98 to interoperability since we have already discussed the importance of software to software and software to hardware compatibility. It is impossible to achieve interoperability without cross-platform operations.

- Unified modelling language (UML): UML is a standardised modelling language enabling developers to specify, visualise, construct and document artefacts of a software system. Thus, UML makes these artefacts scalable, secure and robust in execution. UML is an important aspect involved in object-oriented software

development and yet inevitable for highest degree of interoperability and therefore internet of things (Booch, 2005).

- *Private IP:* private internet protocol addresses are used for private networks such as local area networks. LANs can be referred as closed networks (if not open to wide area networks obviously). The possible reason for this correlation is the security concerns among interoperable machines, especially future internet of things. Without a private network, all these machines are prone to security risks. This is quite an important and if the system is vulnerable very risky side effect of interoperability is one of the most delicate concerns.

Open standards are standards made available to the general public and are developed (or approved) and maintained via a collaborative and consensus driven process. Open standards facilitate interoperability and data exchange among different products or services and are intended for widespread adoption (OFE Ltd., 2008). Those standards are a result of collaborative action and therefore, inclusion of various parties ensures them not to be dominated by one or few interest groups. Availability is public and easy, that makes them accessible and relatively cheaper to acquire. The end product or service is flexible, not closed or limited by one party, resulting its applicability wider and hence, it is customisable by the end-user. Open standards are widely used by software developers; the product is often referred as open-source model (Corrado, 2005). Open source software can be used as an antonym to proprietary software which are held by a benefactor and almost always of major restrictions of use and change. It has become a technical standard for computers and computer networks, and because of its wide availability most of the de facto standards arose from commonly accepted open standards nowadays (Farrell, 1989; Pace, 2009). Also, some proprietary software which have created standards for common use also evolved into de facto standards with the inclusion of open source software; best examples are Microsoft's DOC and XLS format files, Adobe's PDF and Flash, MP3, HTML. One of the most advanced and integrated model of open standards is OpenStack. OpenStack is an open-source infrastructure as-a-service (IaaS) cloud operating system that controls large pools of compute, storage, and networking resources throughout a datacenter, all managed through a dashboard that gives administrators control while empowering their users to provision resources through a web interface (Sefraoui et al., 2012; Opensource, 2015). OpenStack is managed by the OpenStack foundation, a non-profit that oversees both development and community-building around the project. It is compatible and interoperable with other cloud platform APIs such as IBM and Google. The aim is to provide the computer industry with the opportunity to build a hosting architecture, massively scalable which is completely open source, while overcoming the constraints and the use of proprietary technologies (Sefraoui et al., 2012).

These examples can be extended by non-software applications. Because, open standards are not limited with software even though best and easiest illustrations are in that area. For instance; left driver seat side, seat-belts, QWERTY keyboard are some open standards which have become de facto (Acemoglu and Robinson, 2006). Some examples can also be given as a hybrid model of software and non-software, such as credit card systems (Anil Kumar and Ravi, 2008). Cross-fit, a high-intensity fitness programme incorporating elements from several sports and types of exercise and a branded fitness regimen created by Greg Glassman, is a perfect contemporary example of non-software open standard (Bowles, 2015). Because of its open-source model nature

affiliates can develop their own programming, pricing, and instructional methods. Many athletes and trainers see themselves as part of a contrarian, insurgent movement that questions conventional fitness wisdom. The company says this de-centralised approach shares some common features with open source software projects and allows best practices to emerge from a variety of approaches (Velazquez, 2008).

Another comprehensive and extensive non-software example of open standards model, which is by definition a collaborative and publicly available standard pack, is the US's Common Core State Standards (CCSS) for education. This can be overlooked as an open standard since it is far from being a software or for-profit business model, yet it is by its nature and development is one of the most collaborative standards. The creation of the standards involves parties such as; university professors, governor offices, advisory councils, teachers, school chiefs, researchers, experts, state commissioners from various states. CCSS are accepted by almost all states in USA, yet are free to benefit, but actually have no use for parties other than K-12 schools; however, being a state standard for education does not compromise its open nature (Kendall, 2011; Dalton, 2012).

The results for the correlation analysis for '*open standards*' can be listed and described as:

- *Java data objects:* the java data objects (JDO) specification is part of the java community process. JDO is a java application program interface (API) for transparent persistence. It works with both object and relational databases as well as other types of systems. Sun Microsystems has donated JDO to the open source community, so it is one of the biggest APIs gone open from proprietary, which will facilitate java to integrate into any systems, open and interoperable (Jordan and Russell, 2003).

- *Unix for windows:* most versions of Unix are proprietary and handled mainly by system architecture builders. Yet three late versions of BSD branch of Unix ended up open source. Mainly both Unix and Windows are proprietary, however Unix for Windows points out desperate need of integration and interoperability among these two-major system and PC architectures. There emerged some open source implementations of POSIX systems such as Unix, BSD or Linux to Windows which is mainly preferred by PC users; Cygwin (Meister, 2016).

- *Model driven:* model driven architecture (MDA) is a framework based on the UML and other industry standards for visualising, storing, and exchanging software designs and models. However, unlike UML, MDA promotes the creation of machine-readable, highly abstract models that are developed independently of the implementation technology and stored in standardised repositories. There, they can be accessed repeatedly and automatically transformed by tools into schemas, code skeletons, test harnesses, integration code, and deployment scripts for various platforms (Kleppe et al., 2003). The driving force behind the MDA is the fact that a software system will eventually be deployed to one or more platforms, used separately or together. Platforms are subject to change over time – and they change at different, typically higher, rates than the higher-level models of the system, which in turn tend to grow increasingly independent of the target platforms. This paper provides an introduction to the MDA's response to this conundrum (Mellor et al., 2002).

- *Umbrello (UML):* Umbrello is a free software for Unix like platforms as well as Windows and handles all the standard UML diagram types. It can reverse engineer code written in C++, IDL, Pascal/Delphi, Python, and Java, as well as import XMI files generated by external tools from PHP or Perl code and export to various programming languages (Seifert and Wieland, 2003). Umbrello allows the distribution of model contents by exporting to document or web page [Doc(x) and/or HTML] formats (Toth, 2006). This aids in collaborative development efforts where team members may not have direct access to Umbrello, where programmers have programmed in various computer languages that need interchanging or in cases where model contents should be published on a web site.

'*Big data*', among all constitutes of network dynamics, is the most searched one with a peak score average of 95/100 in recent five years. Numbers show a great interest increase doubling every year since 2012. And on average of past five years it has above 60,000 hits per month, reaching up to 200,000 searches per month in the final six months of this period. Usage of big data is a great concern and competitive capability for contemporary competition. To be able to make sense of big data, analytics is very important. However, big data analytics is greatly different from standard data analysis techniques, because there is not just one simple database. Databases, even the largest ones that one can think of is not large enough to consider as big data. And big data analysis is beyond a mere data mining (Singh et al., 2015). It includes learning through machines, and machines' learning as well, it is neither classified nor categorised. And strikingly, the incomprehensible amount of data will even increase more when internet of things become a common tradition.

Hadoop, or formally called Apache Hadoop is an open-source platform initially supported by Apache Software Foundation. Hadoop provides a simple framework to analyse both structured and unstructured data across distributed clusters of computers (Borthakur, 2007). Hadoop can either run on a single server or may be dispersed to thousands of computers offering local computation to analyse vast amounts of data. HDFS is highly fault-tolerant and can be deployed on low-cost hardware (Zikopoulos and Eaton, 2011). HDFS provides high throughput access to application data and is suitable for applications that have large datasets (Shvachko et al., 2010). So basically, it provides computation across machines, where data fragment is available at, however under one file system which is distributed. This gives the opportunity of computation power, scalability, cost effectiveness, flexibility, resiliency to failure (White, 2012). So, Hadoop makes big data computation more possible than ever. On the other hand, Apache Spark is designed to accelerate analytics on mainly Hadoop (HDFS), Cassandra, Amazon S3 and Openstack Swift; it provides complementary tools that comprehend a machine learning library (MLlib), a graph processing engine and stream processing (Bosagh et al., 2016). These tools are cornerstones of the architecture that makes big data analysis possible, efficient and effective.

Given the importance, it is obvious from trends analysis that the academic certificates, courses and Master's degrees have become majorly popular. Two out of all, hits the trends results and not surprisingly took the lead on either education or application of big data analytics; Bangalore and Hyderabad cities of India. These cities have become hits for big data and Hadoop training and education. Coursera, an open learning and course platform that also offers degrees in certain specialisations, offers videos and documentations about big data specialisation. Not surprisingly, Coursera's top

specialisation is 'big data' and its related subjects such as data science, machine learning, Hadoop and Spark.

Social media has become a hype, its use among public and businesses has grown exponentially and still growing. It is the market of network economy (Tapscott, 1996; Couldry, 2012). It is both a factor and outcome of the network effect. So, with this power, it has become the target of companies, marketers, users or in other words everyone. To benefit it, one has to understand ways of collecting and making sense of the information in it; not only with functionalising but also with strategising. Conventional ways of data and information gathering gets reshaped with the force of social media and now involves gathering intelligence from a wider range of public and sources, using analytics to clarify insights. This is not to suggest a total replacement of traditional ways of intelligence gathering, but social media should emerge as a strong complement. There are two traditional sources of information to support competitive analyses; first is the primary sources information such as managers, competitors, suppliers, department reports, customer data within the company, e.g., and the other is the secondary resources such as published journals, articles, market researches, e.g., (Hox and Boeije, 2005). Social media acts on a different plane, it operates on social spaces with peoples' real-time conversations and behaviours. If you can find vital and accurate information, and right experts and analytics to make use of that information, then there is not much need of traditional databases and published work. For this, the process of mapping sources of information becomes and evolves into mapping people and their conversations; gathering data becomes engaging a mapped network of experts and online community and tracking through on-time basis; synthesising and analysing becomes structuring complex and mining relevant data; and simply communicating that information becomes embedding new ways of thinking into strategic processes simultaneously and preaching appropriate information via micro publications (Harrysson et al., 2012).

Yet trend results indicate a dark side of social media as well. Bullying (cyberbullying) and addiction are among top results of trend analysis. Cyberbullying reflects a venue (other than face to face contact) through which verbal and relational forms can occur. Most of the time bullying cannot be observed on social media directly but it can be followed by the traces that are in the form of complaints and/or 'cry outs' about the bullying episode (Xu et al., 2012). Social media addiction on the other hand is defined as "an individual is addicted when an individual's psychological state, which includes both mental and emotional states, as well as their scholastic, occupational and social interactions, is impaired by the overuse of the medium" (Beard, 2005). The most associated problem with the addiction of social media is depression and loss of self-esteem (Young and Rogers, 1998; Griths et al., 2014), which is a serious problem for every generation or age group.

The results for the correlation analysis for 'social media' are very high (all above $r = 0.95$) and can be listed and described as:

- *Infographic:* a visual representation of information or data, e.g., as a chart or diagram. "A good infographic is worth a thousand words", social media is a fast-consuming service and thus, users may skip long written texts (Smiciklas, 2012). Infographics may be a good use to capture attention and tell what you want to tell in short notice.

- *Campaigns:* work in an organised and active way towards a goal. For instance, "people who campaigned against child labor." Social media offers the best medium for campaigns, it helps messages to go 'viral' and eventually reach as many as possible.
- *Bully:* as also trends results indicated bullying is a rising problem. Trends results puts forward the recognised part of the issue, however correlation analysis also shows it is bigger than it is accomplished and credited for. Both results ache for caution.
- *What does high, can you leave, why is my, why are my, can you call, whose number is, why would my, how to know, quotes for girlfriend:* these correlations are all above r = 0.98, and they are all about personal or intimate information seeking. People mostly use social media for their relationships, the searches about ordinary questions reveal a problem rather than a simple information query. People lacking social and inter-personal ties may become obliged to look solutions to their problems through internet. This eventually can lead to depression. Also, there are a good number of articles focusing on 'identity representation' and social media, in which it is said to create issues (Wilson et al., 2012).
- *Psychopaths:* another warning sign from the correlation analysis (r = 0.97). There are not many articles about psychopathy and social media relationship, but one comprehensive big data research also warns about social media's prediction on psychopathic behaviours (Sumner et al., 2012). Results are not robust enough to come up with a solid conclusion as any other behavioural research but warning is there for the ones to notice; such as our trends and correlation results.

6 Conclusions

The network economy has its own dynamics, and these dynamics require a strategic approach rather than a simple functional level effort. Network effect and externalities are the baseline for a network economy. Table 2 summarises so-called network dynamics of the economy.

To be able to achieve a competitive advantage in such economy, companies or institutions should be aware of distinctions. And rather than informally dealing them as 'issues', they should have certain strategies to achieve a strategic turnaround. As discussed earlier, many of the cases have succeeded with a deliberate effort on business models focusing on networks and their dynamics. It is important to understand the information provided by the community.

Networks are formed of users; companies have little power on the networks because every user has different needs which virtually make it impossible to have 'generic strategies' to rule them all. Strategy should be adaptive, flexible as in the old definition, however that is not enough. Networks' nature is chaotic since it is formed of a social crowd. It is possible for almost identical systems embodied in identical environments to demonstrate far-out contradictive demeanours, even when the foundations are highly simple and totally deterministic. Moreover, assertions are commonplace enough to administer to any system; including organisations, institutions, teams, groups or individuals, and thus they are admissible to a broad variety of social domains. Perhaps

using a metaphor here will be accurate; network economy requires strategies to be like play-dough: soft and colourful enough to be desired and shaped by any party in the network but resilient, durable and impregnable enough to sustain.

Table 2 Comparison of old strategy dynamics vs. network dynamics

	Strategic dynamics	*Network dynamics*
Economy basis	Economies of scale	Network externalities
Prerequisite (cogwheels)	Supply and demand equilibrium	Network effect
Challenges	Productivity, efficiency, effectiveness	Agility, lean operations, big data management
Source of competitive advantage	Scarcity, rareness	Abundance
Competitive strategies	Differentiation, cost leadership, niche	Interoperability, open standards, network effect
Value drivers	Form, place, time, possession utilities	Form, place, time, possession utilities + network utility
Market basis	Free movement of goods/services, market exchange	Free movement of information, social media

Utility is completely dependent on the model; a good product is never enough. There have been more advanced web sites compared to Facebook or Twitter, but their touch on network's necessities was better than any other. There have been many alternatives to Uber, but user-friendly interface and inter-dependency ability have made it top provider. Or finally, AirBnB was the first one to recognise a niche and fulfil it with abundance. Carbon-copying the infrastructure of a software, website, application, operating system or architecture, machinery and interior of a coffee shop, market, company is easy; the problem begins where system deploys on the network. Hence, network utility is the fifth additional utility where form, place, time and possession utility applies. Enablers of this utility are interoperability, open standards, big data and eventually network effect up to certain amounts in every model; *quantum satis*.

References

Acemoglu, D. and Robinson, J.A. (2006) 'De facto political power and institutional persistence', *The American Economic Review*, Vol. 96, No. 2, pp.325–330.

Achrol, R.S. and Kotler, P. (1999) 'Marketing in the network economy', *The Journal of Marketing*, Vol. 63, pp.146–163, Special issue.

Anil Kumar, D. and Ravi, V. (2008) 'Predicting credit card customer churn in banks using data mining', *International Journal of Data Analysis Techniques and Strategies*, Vol. 1, No. 1, pp.4–28.

Applegate, L., Austin, R. and McFarlan, F.W. (2002) *Corporate Information Strategy and Management: the Challenges of Managing in a Network Economy (Paperback Version)*, McGraw-Hill, Inc., Ohio. USA

Baccara, M. (2012) 'A field study on matching with network externalities', *The American Economic Review*, Vol. 102, No. 5, pp.1773–1804.

Bagui, S., Just, J. and Bagui, S.C. (2009) 'Deriving strong association mining rules using a dependency criterion, the lift measure', *International Journal of Data Analysis Techniques and Strategies*, Vol. 1, No. 3, pp.297–312.

Bauer, M. and Erixon, F. (2016) *Competition, Growth and Regulatory Heterogeneity in Europe's Digital Economy*, European Centre for Political Economy (ECIPE) Working Paper, No. 2.

Beard, K.W. (2005) 'Internet addiction: a review of current assessment techniques and potential assessment questions', *Cyber Psychology & Behavior*, Vol. 8, No. 1, pp.7–14.

Bilbao-Osorio, B., Dutta, S. and Lanvin, B. (2013) *The Global Information Technology Report 2013*, pp.1–383, World Economic Forum, Davos, Switzerland.

Blaug, M. (1997) *Economic Theory in Retrospect*, Cambridge University Press, Cambridge, UK.

Booch, G. (2005) *The Unified Modelling Language User Guide*, Pearson Education, India.

Borthakur, D. (2007) *The Hadoop Distributed File System: Architecture and Design*, pp.12–13, Apache Software Foundation Press, Hadoop Project Website [online] https://hadoop.apache.org/docs/r1.2.1/hdfs_design.pdf (accessed 7 May 2018).

Bosagh Zadeh, R., Meng, X., Ulanov, A., Yavuz, B., Pu, L., Venkataraman, S. and Zaharia, M. (2016) 'Matrix computations and optimization in apache spark', in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, pp.31–38.

Bowles, N. (2015) *Exclusive: on the Warpath with CrossFit's Greg Glassman* [online] https://www.maxim.com/maxim-man/crossfit-greg-glassman-exclusive-2015-9 (accessed 7 May 2018).

Brailer, D.J. (2005) 'Interoperability: the key to the future health care system', *Health Affairs*, Vol. 24, No. 1, p.W5.

Breg, F., Diwan, S., Villacis, J.E., Balasubramanian, J., Akman, E. and Gannon, D. (1998) 'Java RMI performance and object model interoperability: experiments with Java/HPC++', *Concurrency – Practice and Experience*, Vol. 10, Nos. 11–13, pp.941–955.

Carneiro, H.A. and Mylonakis, E. (2009) 'Google trends: a web-based tool for real-time surveillance of disease outbreaks', *Clinical Infectious Diseases*, Vol. 49, No. 10, pp.1557–1564.

Chauhan, R. and Kaur, H. (2017) 'A feature-based selection technique for reduction of large scale data', *International Journal of Data Analysis Techniques and Strategies*, Vol. 9, No. 3, pp.207–221.

Choi, H. and Varian, H. (2009) *Predicting Initial Claims for Unemployment Benefits*, pp.1–5, Google Inc., Mountain View, CA, USA.

Choi, H. and Varian, H. (2012) 'Predicting the present with Google trends', *Economic Record*, Vol. 88, No. s1, pp.2–9.

Choi, S.Y. and Whinston, A.B. (2000) 'Benefits and requirements for interoperability in the electronic marketplace', *Technology in Society*, Vol. 22, No. 1, pp.33–44.

Cormen, T.H., Leiserson, C.E., Rivest, R.L. and Stein, C. (2001) *Introduction to Algorithms*, 2nd ed., MIT Press, Cambridge, MA, USA

Corrado, E.M. (2005) 'The importance of open access, open source, and open standards for libraries', *Issues in Science and Technology Librarianship*, Vol. 2005, No. 42, DOI: 10.5062/F42F7KD8 [online] http://www.istl.org/05-spring/article2.html.

Couldry, N. (2012) *Media, Society, World: Social Theory and Digital Media Practice*, pp.108–112, Polity Press, Malden, MA 02148, USA.

Cramer, J. and Krueger, A.B. (2016) 'Disruptive change in the taxi business: the case of Uber', *The American Economic Review*, Vol. 106, No. 5, pp.177–182.

Da, Z., Engelberg, J. and Gao, P. (2011) 'In search of attention', *The Journal of Finance*, Vol. 66, No. 5, pp.1461–1499.

Dalton, B. (2012) 'Multimodal composition and the common core state standards', *The Reading Teacher*, Vol. 66, No. 4, pp.333–339.

De Man, A.P. (2004) *The Network Economy*, pp.102–118, Edward Elgar Pub, Northampton, MA, USA.

Economides, N. (1996) 'Network externalities, complementarities, and invitations to enter', *European Journal of Political Economy*, Vol. 12, No. 2, pp.211–233.

Eysenbach, G. (2006) 'Infodemiology: tracking flu-related searches on the web for syndromic surveillanc', in *AMIA Annual Symposium Proceedings*, American Medical Informatics Association, p.244.

Farrell, J. (1989) 'Standardization and intellectual property', *Jurimetrics*, Vol. 30, No. 1, pp.35–50.

Fitzgerald, B. (2006) 'The transformation of open source software', *Mis Quarterly*, Vol. 30, No. 3, pp.587–598.

Gersho, A. and Gray, R.M. (1992) *Vector Quantization I: Structure and Performance. In Vector Quantization and Signal Compression*, pp.309–343, Springer, USA.

Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S. and Brilliant, L. (2009) 'Detecting influenza epidemics using search engine query data', *Nature*, Vol. 457, No. 7232, pp.1012–1014.

Goel, S., Hofman, J.M., Lahaie, S., Pennock, D.M. and Watts, D.J. (2010) 'Predicting consumer behavior with web search', *Proceedings of the National Academy of Sciences*, Vol. 107, No. 41, pp.17486–17490.

Goldfarb, A., Greenstein, S.M. and Tucker, C.E. (2015) 'Introduction to 'economic analysis of the digital economy'', *Economic Analysis of the Digital Economy*, pp.1–17, University of Chicago Press, Chicago, IL, USA.

Griths, M.D., Kuss, D.J. and Demetrovics, Z. (2014) 'Social networking addiction: an overview of preliminary findings', *Behavioral Addictions: Criteria, Evidence and Treatment*, Chapter 6, pp.119–141, Academic Press, Waltham, MA 02451, USA.

Harrysson, M., Metayer, E. and Sarrazin, H. (2012) 'How 'social intelligence' can guide decisions', *McKinsey Quarterly*, Vol. 4, No. 1, pp.81–89.

Hippel, E.V. and Krogh, G.V. (2003) 'Open source software and the 'private-collective' innovation model: issues for organization science', *Organization Science*, Vol. 14, No. 2, pp.209–223.

Honkola, J., Laine, H., Brown, R. and Oliver, I. (2009) 'Cross-domain interoperability: a case study', in *Smart Spaces and Next Generation Wired/Wireless Networking*, pp.22–31, Springer, Berlin Heidelberg.

Hox, J.J. and Boeije, H.R. (2005) 'Data collection, primary vs. secondary', *Encyclopedia of Social Measurement*, Vol. 1, No. 1, pp.593–599.

Ida, T. (2004) *Bottleneck Monopolies and Network Externalities in Network Industries: Introduction to Network Economics*, Faculty of Economics, Kyoto University, Japan.

Ida, T. (2009) *Broadband Economics: Lessons from Japan*, pp.118–126, Taylor and Francis, Routledge imprint, London, UK.

Ida, T. (2012) *Broadband Economics* [online] http://www.econ.kyoto-u.ac.jp/~ida/2Kyouiku/1Gakubu/2008/BB-E-3.pdf (accessed 15 January 2017).

Jordan, D. and Russell, C. (2003) *Java Data Objects*, O'Reilly Media, Inc., Sebastopol, CA 95472, USA

Katz, M.L. and Shapiro, C. (1985) 'Network externalities, competition, and compatibility', *The American Economic Review*, Vol. 75, No. 3, pp.424–440.

Katz, M.L. and Shapiro, C. (1994) 'Systems competition and network effects', *The Journal of Economic Perspectives*, Vol. 8, No. 2, pp.93–115.

Kendall, J.S. (2011) *Understanding Common Core State Standards*, ASCD, Alexandria, VA 22311-1714, USA.

Kleppe, A.G., Warmer, J.B. and Bast, W. (2003) *MDA Explained: the Model Driven Architecture: Practice and Promise*, Addison-Wesley Professional, Boston, MA 02116, USA.

Lakhani, K.R. and Von Hippel, E. (2003) 'How open source software works: 'free' user-to-user assistance', *Research Policy*, Vol. 32, No. 6, pp.923–943.

Lavie, D. (2006) 'The competitive advantage of interconnected firms: an extension of the resource-based view', *Academy of Management Review*, Vol. 31, No. 3, pp.638–658.

Lengnick-Hall, C.A. (1992) 'Innovation and competitive advantage: what we know and what we need to learn', *İ*, Vol. 18, No. 2, pp.399–429.

Leonard-Barton, D. (1995) *Wellsprings of Knowledge: Building and Sustaining the Sources of Innovation*, University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship, SSRN [online] https://ssrn.com/abstract=1496178 (accessed 7 May 2018).

Liu, T., Moore, A.W., Gray, A.G. and Yang, K. (2004) 'An investigation of practical approximate nearest neighbor algorithms', *Proceedings of the 12th International Conference on Neural Information Processing Systems*, pp.825–832, MIT Press, Cambridge, MA, USA.

Liu, T., Rosenberg, C.J. and Rowley, H.A. (2009) *U.S. Patent No. 7,539,657*, U.S. Patent and Trademark Office, Washington, DC.

Malhotra, R. (2010) 'Using a meta-modelling knowledge management layer – a new approach to designing the enterprise information architecture', *International Journal of Data Analysis Techniques and Strategies*, Vol. 2, No. 4, pp.307–335.

Manyika, J., Lund, S., Bughin, J., Woetzel, J., Stamenov, K. and Dhingra, D. (2016) *Digital Globalization: the New Era of Global Flows*, McKinsey Global Institute, Summit, NJ 07901, USA.

Markel, M.H. (1996) *Technical Communication: Situations and Strategies*, St. Martin's Press, Brooklyn, NY 10030, USA.

McGee, J. and Sammut-Bonnici, T. (2002) 'Network industries in the new economy', *European Business Journal*, Vol. 14, No. 5, pp.116–32.

McGuire, M.C. and Olson, M. (1996) 'The economics of autocracy and majority rule: the invisible hand and the use of force', *Journal of Economic Literature*, Vol. 34, No. 1, pp.72–96.

Meister, B. (2016) *PuTTY/Cygwin Tutorial* [online] http://www.cs.dartmouth.edu/~campbell/cs50/putty-cygwin-tutorial.pdf (accessed 5 April 2017).

Mellor, S.J., Scott, K., Uhl, A. and Weise, D. (2002) 'Model-driven architecture', in *International Conference on Object-Oriented Information Systems*, Springer, Berlin Heidelberg, pp.290–297.

Meltzer, B. and Glushko, R.J. (1998) 'XML and electronic commerce: enabling the network economy', *SIGMOD Record*, Vol. 27, No. 4, pp.21–24.

Mohebbi, M., Vanderkam, D., Kodysh, J., Schonberger, R., Choi, H. and Kumar, S. (2011) *Google Correlate Whitepaper*, Google Inc., Mountain View, CA, USA.

Murdoch, T.B. and Detsky, A.S. (2013) 'The inevitable application of big data to health care', *Jama*, Vol. 309, No. 13, pp.1351–1352.

Murhc, S. (2014) *Does Airbnb have Network Effect?* [online] https://www.quora.com/Does-Airbnb-have-network-effect (accessed 7 March 2017).

Narayanan, V.K. (2000) *Managing Technology and Innovation for Competitive Advantage*, Pearson Ed. India Press, Noida 201 309, Delhi, India.

North, D.C. (2016) 'Institutions and economic theory', *The American Economist*, Vol. 61, No. 1, pp.72–76.

OFE Ltd. (2008) *The Importance of Open Standards in Interoperability*, OFE Onepage Brief No. 1. [online] http://openforumeurope.org/wp-content/uploads/2015/06/OFE-Open-Standards-Onepage-2008.pdf (accessed 7 March 2017).

Opensource (2015) *What is OpenStack?* [online] https://opensource.com/resources/what-is-openstack (accessed 13 January 2017).

Pace, A.K. (2009) 'De facto standards', *Library Technology Reports*, Vol. 38, No. 2, pp.41–42.

Pelat, C., Turbelin, C., Bar-Hen, A., Flahault, A. and Valleron, A.J. (2009) 'More diseases tracked by using Google trends', *Emerging Infectious Diseases*, Vol. 15, No. 8, pp.1327–1328.

Pellegrin, J. (2016) *The Political Economy of Competitiveness in an Enlarged Europe*, Palgrave, New York, NY 10010, USA.

Porter, M.E. (2011) *Competitive Advantage of Nations: Creating and Sustaining Superior Performance*, Simon and Schuster, New York, NY 10020, USA.

Potter, W.J. (2012) *Media Effects*, Sage, Thousand Oaks, CA 91320, USA.

Rao, A.R., Bergen, M.E. and Davis, S. (2000) 'How to fight a price war', *Harvard Business Review*, Vol. 78, No. 2, pp.107–120.

Sefraoui, O., Aissaoui, M. and Eleuldj, M. (2012) 'OpenStack: toward an open-source solution for cloud computing', *International Journal of Computer Applications*, Vol. 55, No. 3, pp.38–42.

Seifert, T. and Wieland, T. (2003) 'Prerequisites for enterprises to get involved in open source software development', *Institute for Computer Science*, Vol. 2, No. 1, p.1.

Shapiro, C. and Varian, H.R. (2013) *Information Rules: a Strategic Guide to the Network Economy*, Harvard Business Press, Boston, MA 02100, USA.

Sharma, S.K. and Suman, U. (2013) 'An efficient semantic clustering of URLs for web page recommendation', *International Journal of Data Analysis Techniques and Strategies*, Vol. 5, No. 4, pp.339–358.

Sheth, A.P. (1999) 'Changing focus on interoperability in information systems: from system, syntax, structure to semantics', in *Interoperating Geographic Information Systems*, pp.5–29, Springer, USA.

Shrivastava, A. and Li, P. (2014) 'Asymmetric LSH (ALSH) for sublinear time maximum inner product search (MIPS)', in *Advances in Neural Information Processing Systems*, pp.2321–2329.

Shvachko, K., Kuang, H., Radia, S. and Chansler, R. (2010) 'The hadoop distributed file system', in *2010 IEEE 26th symposium on Mass Storage Systems and Technologies (MSST)*, IEEE, pp.1–10.

Singh, A., Rana, A. and Ranjan, J. (2015) 'Data mining techniques and its effect in customer relationship management', *International Journal of Data Analysis Techniques and Strategies*, Vol. 7, No. 4, pp.406–427.

Sjöholm, F. (1999) 'Technology gap, competition and spillovers from direct foreign investment: evidence from establishment data', *The Journal of Development Studies*, Vol. 36, No. 1, pp.53–73.

Smiciklas, M. (2012) *The Power of Infographics: Using Pictures to Communicate and Connect with your Audiences*, Que Publishing, Indianapolis, IN 46240, USA.

Stephens-Davidowitz, S. and Varian, H. (2014) *A Hands-on Guide to Google Data*, Tech. Rep.

Sumner, C., Byers, A., Boochever, R. and Park, G.J. (2012) 'Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets', in *2012 11th International Conference on Machine Learning and Applications (ICMLA)*, IEEE, Vol. 2, pp.386–393.

Tapscott, D. (1996) *The Digital Economy: Promise and Peril in the Age of Networked Intelligence*, Vol. 1, McGraw-Hill, New York.

Tapscott, D. (2015) *The Digital Economy*, McGraw-Hill Education, New York, NY 10030, USA.

Tenner, E. and Rall, J.E. (1997) *Why Things Bite Back: Technology and the Revenge of Unintended Consequences*, p.F1, Vintage Books, New York.

Toth, K. (2006) 'Experiences with open source software engineering tools', *IEEE Software*, Vol. 23, No. 6, pp. 44-52.

Uzzi, B. (1996) 'The sources and consequences of embeddedness for the economic performance of organizations: the network effect', *American Sociological Review*, Vol. 61, No. 4, pp.674–698.

Vanderkam, D., Schonberger, R., Rowley, H. and Kumar, S. (2013) *Nearest Neighbor Search in Google Correlate*, Dynamic.Stx8.Com [online] http://www.google.com/trends/correlate/nnsearch.pdf (accessed 14 January 2017).

Vazirani, V.V. (1994) 'A theory of alternating paths and blossoms for proving correctness of the general graph maximum matching algorithm', *Combinatorica*, Vol. 14, No. 1, pp.71–109.

Velazquez, E. (2008) 'Sweatstorm', *Muscle & Fitness*, May, pp.142–154 [online] http://www.crossfittroy.com/files/Sweatstorm.pdf (accessed 8 April 2017).

White, T. (2012). *Hadoop: the Definitive Guide*, O'Reilly Media, Inc., Sebastopol, CA 95472, USA.

Wilson, K. and Brownstein, J.S. (2009) 'Early detection of disease outbreaks using the internet', *Canadian Medical Association Journal*, Vol. 180, No. 8, pp.829–831.

Wilson, R.E., Gosling, S.D. and Graham, L.T. (2012) 'A review of Facebook research in the social sciences', *Perspectives on Psychological Science*, Vol. 7, No. 3, pp.203–220.

Xu, J.M., Jun, K.S., Zhu, X. and Bellmore, A. (2012) 'Learning from bullying traces in social media', in *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Association for Computational Linguistics, pp.656–666.

Young, K.S. and Rogers, R.C. (1998) 'The relationship between depression and internet addiction', *CyberPsychology & Behavior*, Vol. 1, No. 1, pp.25–28.

Zhu, K., Kraemer, K.L., Gurbaxani, V. and Xu, S.X. (2006) 'Migration to open-standard interorganizational systems: network effects, switching costs, and path dependency', *Mis Quarterly*, Vol. 30, pp.515–539, Special issue.

Zikopoulos, P. and Eaton, C. (2011) *Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data*, McGraw-Hill Osborne Media, New York, NY 10010, USA.

Notes

1 https://serps.com/tools/rank-checker/

2 Please refer to methodology introduction paragraph.

3 'Node level' indicate where level of analysis for the research in on 'nodes' which is namely any given topic/term. Here results are given as nodes' clusters.

4 'Word level' indicate where level of analysis for the research is on 'words' rather than 'nodes'. So raw text is clustered independent of the topic/terms which they belong to.

Appendix 1

*Results of relation and correlate for keywords**				
Keywords	*Monthly average search volume*[1]	*Interest over time*	*Related results (top and rising)*	*Correlate results with correlation coefficient*
Network economy	70 searches per month	Slightly declining since 2012 – almost always below 50 score.	• Computer networks • Marketing • Internet • Portable document format (PDF) • Innovation • Business model • Information technology • SAP SE • UCL Institute of Education	-
Network effect	1,300 searches per month	Steady for five years and above the average of 70 score.	• Computer networks • Social networking • Externality • Strategy • Competitive advantage • Business model • Value, product, goods and services • Uber • Resource • Anderssen Horowitz • Cartel • Airbnb • Bitcoin	• 0.7905 common crimes • 0.7867 media messages • 0.7798 morally right • 0.7793 social rules • 0.7773 what is sustainable agriculture • 0.7772 direct competition • 0.7758 the universal language

Notes: - – some keywords had insufficient data to calculate correlations, * – trends and correlation results are refined since some of their relation to the subject is limited, or mis-interpreted with some other contexts (e.g., expectations theory is related to network effect, but if looked deeply it is namely the neural networks), [1]only in English language, [2]social media is taken as a case affected by network effect and basically governed by network economy dynamics.

Appendix 1 (continued)

*Results of relation and correlate for keywords**				
Keywords	*Monthly average search volume*[1]	*Interest over time*	*Related results (top and rising)*	*Correlate results with correlation coefficient*
Network externalities	980 searches per month	Steady interest over five years, high peaks every year to 100 score. Average of 65.	• Network effect • Spillover effect • Decision making • Computer network • Marketing • Technology • Service • Business model • Monopoly • Barriers to entry	• 0.8953 media effects • 0.8946 technology effects • 0.8868 sociological terms • 0.8618 maximum matching • 0.8589 benefits of the internet
Interoperability	12,100 searches per month	Steady interest over five years with a high average of 80 scoring. Yet there are deep points every year with 40.	• Data • System • Incident command system • Software • VMWare • Electronic health record • Communication • Information system • Management • Organisation • Federal emergency system • Cisco unified computing	• 0.9868 Java interface • 0.9854 software development • 0.9845 unified modelling language • 0.9836 web forms • 0.9807 software quality • 0.9803 private ip • 0.9795 software design
Open standards	260 searches per month	Steady interest over five years with an average score of 60.	• Open-source model • Software • Proprietary software • Computer • Technical standard • DeFacto standards • CrossFit • OpenStack • Internet of things • Common core state • Computer network	• 0.9596 Java data • 0.9513 Unix for windows • 0.9379 model driven • 0.9334 Umbrello (unified modelling language)

Notes: - – some keywords had insufficient data to calculate correlations, * – trends and correlation results are refined since some of their relation to the subject is limited, or mis-interpreted with some other contexts (e.g., expectations theory is related to network effect, but if looked deeply it is namely the neural networks), [1]only in English language, [2]social media is taken as a case affected by network effect and basically governed by network economy dynamics.

Appendix 1 (continued)

*Results of relation and correlate for keywords**				
Keywords	*Monthly average search volume*[1]	*Interest over time*	*Related results (top and rising)*	*Correlate results with correlation coefficient*
Big data	60,500 searches per month	Great interest increase doubling every year since 2012. Hit a peak score average of 95 in recent years.	• Analytics • Apache Hadoop • Data • Cloud computing • Technology • Database • Data mining • Machine learning • Internet of things • IBM • Apache Spark • Coursera • Hyderabad • Academic certificate, course • Bangalore • Master's degree	-
Social media[2]	74,000 searches per month	Rising interest through five years from an average score of 60 to 100.	• Social media marketing • Website • Buffer (buffer.com) • Bullying (school, cyber-) • Communication • Social networking • Strategy • Addiction • LinkedIn • Telegram • Snapchat • Google Scholar	• 0.9935 infographic • 0.9932 campaigns • 0.9903 bully • 0.9900 what does …(ty, ttyl, nvm) mean • 0.9871 what does high • 0.9803 can you leave • 0.9791 why is my • 0.9782 why are my • 0.9779 can you call • 0.9778 whose number is • 0.9765 why would my • 0.9765 how to know • 0.9760 quotes for girlfriend • 0.9669 psychopaths

Notes: - – some keywords had insufficient data to calculate correlations, * – trends and correlation results are refined since some of their relation to the subject is limited, or mis-interpreted with some other contexts (e.g., expectations theory is related to network effect, but if looked deeply it is namely the neural networks), [1]only in English language, [2]social media is taken as a case affected by network effect and basically governed by network economy dynamics.

Appendix 2

For Mac

- Install homebrew app to get linux command line tools:
  /usr/bin/ruby -e''$(curl -fsSL https://raw.githubusercontent.com/Homebrew/install /master/install)''
- Install wget command line tool:
  brew install wget
- Use wget to download a website content to a text file (results.txt):
  wget -qO- http://.../ | sed -e 's/<[^>]>//g'>results.txt*
- Use wget to download a website content to a text file stripping the html codes (results.txt):
  wget -qO- http://.../ | sed -e 's≠[^>]>//g;s/^ //g'>results.txt*
- For Linux:
  Download and install "lynx browser and parallel" from the source of preference. (applications)
- Use lynx to dump the website content to text file (results.txt):
  Lynx – dump – nolist https://webpage.com/>results.txt
- To download multiple web pages (listed in sites.txt file) at once create a single bash alias (results.txt):
  while read 1; do echo "##########"; echo $1; echo "##########"; lynx –dump –nolist "$1"; done < sites.txt > results.txt
- Create list of websites, as plain text, into a ".txt" file e.g. sites.txt and use;
  parallel -j 16 -a sites.txt lynx -dump -nolist > all_results.txt

Appendix 3 (see online version for colours)