ICCGI 2020

The Fifteenth International Multi-Conference on Computing in the Global Information Technology

ISBN: 978-1-61208-793-1

October 18 – 22, 2020

ICCGI 2020 Editors

Sandra Sendra Compte, Universitat Politecnica de Valencia, Spain
Mary Luz Mouronte-López, Universidad Francisco de Vitoriam Spain

ICCGI 2020

Foreword

The Fifteenth International Multi-Conference on Computing in the Global Information Technology (ICCGI 2020), held between October 18–22, 2020, continued a series of international events covering a large spectrum of topics related to global knowledge concerning computation, technologies, mechanisms, cognitive patterns, thinking, communications, user-centric approaches, nanotechnologies, and advanced networking and systems. The conference topics focus on challenging aspects in the next generation of information technology and communications related to the computing paradigms (mobile computing, database computing, GRID computing, multi-agent computing, autonomic computing, evolutionary computation) and communication and networking and telecommunications technologies (mobility, networking, bio-technologies, autonomous systems, image processing, Internet and web technologies), towards secure, self-defendable, autonomous, privacy-safe, and context-aware scalable systems.

This conference intended to expose the scientists to the latest developments covering a variety of complementary topics, aiming to enhance one's understanding of the overall picture of computing in the global information technology.

The integration and adoption of IPv6, also known as the Next Generation of the Internet Protocol, is happening throughout the World at this very moment. To maintain global competitiveness, governments are mandating, encouraging or actively supporting the adoption of IPv6 to prepare their respective economies for the future communication infrastructures. Business organizations are increasingly mindful of the IPv4 address space depletion and see within IPv6 a way to solve pressing technical problems while IPv6 technology continues to evolve beyond IPv4 capabilities. Communications equipment manufacturers and applications developers are actively integrating IPv6 in their products based on market demands.

IPv6 continues to represent a fertile area of technology innovation and investigation. IPv6 is opening the way to new successful research projects. Leading edge Internet Service Providers are guiding the way to a new kind of Internet where any-to-any reachability is not a vivid dream but a notion of reality in production IPv6 networks that have been commercially deployed. National Research and Educational Networks together with internationally known hardware vendors, Service Providers and commercial enterprises have generated a great amount of expertise in designing, deploying and operating IPv6 networks and services. This knowledge can be leveraged to accelerate the deployment of the protocol worldwide.

We take here the opportunity to warmly thank all the members of the ICCGI 2020 Technical Program Committee, as well as the numerous reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to ICCGI 2020. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the ICCGI 2020 organizing

committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that ICCGI 2020 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the area of computing in the global information technology.

ICCGI 2020 Chairs

ICCGI 2020 Steering Committee

Constantin Paleologu, University Politehnica of Bucharest, Romania
Peter Kieseberg, St. Pölten University of Applied Sciences, Austria
Yasushi Kambayashi, Nippon Institute of Technology, Japan

ICCGI 2020 Publicity Chair
Jose M. Jimenez, Universitat Politecnica de Valencia, Spain
Jose Luis García, Universitat Politecnica de Valencia, Spain

ICCGI 2020 Industry/Research Advisory Committee

Zeeshan Pervez, University of the West of Scotland, UK
Ulrich Reimer, University of Applied Sciences St. Gallen, Switzerland
Juho Mäkiö, Hochschule Emden / Leer, Germany

ICCGI 2020

Committee

ICCGI 2020 Publicity Chair

Jose M. Jimenez, Universitat Politecnica de Valencia, Spain
Jose Luis García, Universitat Politecnica de Valencia, Spain

ICCGI 2020 Industry/Research Advisory Committee

Zeeshan Pervez, University of the West of Scotland, UK
Ulrich Reimer, University of Applied Sciences St. Gallen, Switzerland
Juho Mäkiö, Hochschule Emden / Leer, Germany

ICCGI 2020 Technical Program Committee

António Abreu, ISEL/IPL, Portugal
Safa'a AbuJarour, University of Potsdam, Germany
Abdelouhab Aitouche, EEA - French School of High Studies in Engineering, France
Mohammad Yahya Al-Shamri, King Khalid University, Saudi Arabia
Fernando Almeida, INESC TEC & University of Porto, Portugal
Nancy Arana Daniel, University of Guadalajara, Mexico
Jocelyn Aubert, Luxembourg Institute of Science and Technology (LIST), Luxembourg
Michaela Baumann, University of Bayreuth, Germany
Robert Bestak, Czech Technical University in Prague, Czech Republic
Dorota Bielinska-Waz, Medical University of Gdansk, Poland
Fernando Bobillo, University of Zaragoza, Spain
Zorica M. Bogdanovic, University of Belgrade, Serbia
Eugen Borcoci, University "Politehnica" of Bucharest (UPB), Romania
Jean-Louis Boulanger, CERTIFER, France
Christos Bouras, University of Patras, Greece
Juan Carlos Burguillo-Rial, Universidade de Vigo, Spain
DeJiu Chen, KTH Royal Institute of Technology, Sweden
Albert M. K. Cheng, University of Houston, USA
Rebeca Cortazar, University of Deusto, Spain
Pietro Cunha Dolci, Santa Cruz do Sul University, Brazil
Beata Czarnacka-Chrobot, Warsaw School of Economics, Poland
António Dourado, University of Coimbra, Portugal
Ahmed El Oualkadi, Abdelmalek Essaadi University, Morocco
Sylvia Encheva, Western Norway University of Applied Sciences, Norway
Manuel Filipe Santos, University of Minho, Portugal
Dirk Frosch-Wilke, University of Applied Sciences Kiel, Germany
Félix J. García, University of Murcia, Spain
Apostolos Gkamas, University Ecclesiastical Academy of Vella of Ioannina, Greece

Table of Contents

Implementation of Cloud Computing in Albania and Related Security Concerns

Ergest Alite
Faculty of Information Technology
Polytechnic University of Tirana
Tirana, Albania
e-mail: ergest.alite@albtelecom.al

Olimpjon Shurdi
Faculty of Information Technology
Polytechnic University of Tirana
Tirana, Albania
e-mail: oshurdi@fti.edu.al

***Abstract*—Cloud Computing Technology is one of the Information and Communications Technologies (ICTs) which was given the most attention to during this recent decade. Since the trend of Cloud Computing Technology is already well-oriented in the World, we will see how this trend is affecting Albania and how this technology has been adopted in the local market. To identify this, we have conducted a research method based on two theories: the diffusion of innovation (DOI) model, and the technology organization environment (TOE) framework. This paper will present three Cloud Technologies' cases implemented in Albania by public and private companies, as well the technical challenges and economic opportunities these implementations caused to the respective organizations. In the end, we will briefly present the results of the research, impacted factors, and the approach that should be followed by each player in order to have such technology more present in the Albanian market.**

Keywords- Cloud Computing; Security Concern; Regulation Normatives; Technological Challenges; Diffusion of Innovations (DOI); Technology Organization Environment (TOE).

I. INTRODUCTION

Cloud Computing technology is revolutionizing Information Technology (IT) services nowadays. Almost all major IT hardware and software companies are focusing on delivering these services. Following Amazon's $ 2 billion investment in IBM infrastructure-based services in March 2006 [15], all other global companies have followed this strategy by offering various services, such as Infrastructure as a Service (IaaS), Platform as a Service (PaaS), Storage as a Service (StaaS) or Software as a Service (SaaS) [17]. As per such development, we can mention the services provided by Microsoft with Azure, or Google with G Suite and Oracle with Oracle Cloud [16]. All application development companies are trying, through customer requests, to switch all their services from their offices to the cloud platforms of their respective companies. Microsoft is performing this through pushing its customers to get Cloud Services during respective Enterprise Agreement renewal [19]. SAP has already announced that it will provide support for on-premise platforms up to 2025, and after this time it will interupt support services for customers who still have on-primise platforms. Consequently, all customers running SAP applications will have to migrate their on-premise services to the cloud by 2025 [18].

The focus of this paper is primarily to identify the adoption, and Implementation of Cloud Computing in Albania. In section 2, it is provided a general overview of Cloud Computing trend in the world, and in Albania. Section 3 is depicting a general overview of Cloud Computing implementations which happened in Albania in both public and private sectors. Section 4 explains the research method, based on the DOI model, and on the TOE theory framework. In section 4, the survey results, and the main reasons of low adaptability, and implementations in Albania have been provided. In the end, conclusions and future works have been depicted accordingly.

II. CLOUD COMPUTING TRENDS AND APPROACHES

Considering the above mentioned important and fast technological and commercial developments, all existing IT services will be shifted towards a centralized hardware and application development for all clients. Although the strategy from global companies and technology leaders is clear, most clients and local companies are not yet aware, and do not own a strategy on how to approach accordingly. Most national companies are attempting to primarily pursue the first two approaches; the 1st approach, being based on establishing a partnership with global companies, and become resellers of their services; and the 2nd approach, being based on making own investments and operating independently for similar services. Both ways have their pros and cons. Basically, becoming a reseller of a global company has its benefits because you do not have to invest in infrastructure, maintenance, human resources, or in the field of products and marketing, for the reason that all these items are granted from the provider. On the other hand, a disadvantage of this approach is the fact that the client is always the provider's client. In order to illustrate this case we can mention one example from 3G Ireland. This company became Microsoft Cloud Service Provider (CSP) and started selling Office 365 products. After a period of 1 year, it was identified that the revenue trend generated by this service was quite as low as not even covering the investment performed for marketing campaigns, or other product operations expenditures. In this regard, it was much evident that such service was not profitable enough for the company. As a result, the company's board of directors made the decision of terminating the service, and no longer marketing it. However, the existing 10,000 customers were

 ISBN: 978-1-61208-793-1

not affected by this decision at all, as they were secured uninterrupted service, with the only understandable difference that the service provider was no longer 3G Ireland, but Microsoft. Another major disadvantage of this method are the small profit margins, which make such service in small markets to result as non-profitable, just like 3G Ireland did. In the second approach, companies decided to invest themselves for providing such services, which substantially has its own benefits raleted to the fact that: service is local, customers remain with the company, and there exists higher trust from customers, and faster access to services. The disadvantages of this method though are related to the fact that investment costs are very high, leading to the need of more resources, and higher operational costs (maintenance, and support staff). Furthermore, when it comes to small market countries, this kind of approach is explicitly not profitable because products will have expensive prices compared to global providers due to high TCO.

However, despite the overall above picture, many telecommunication companies such as Deutche Telekom, Telefonica, Telecom Italia, etc., have extensively implemented their cloud services platforms in many countries around the world. Moreover, upon UK's exit from the EU, such companies have accomplished to make USA-based companies like Microsoft, Google and Amazon host their Cloud infrastructure on their premises for services they provide in the European Union market. Most of Google, Microsoft or Amazon platforms are now hosted at Telefonica' Data Centers or Deutche Telekom' Data Centers all across Europe.

A. World trends

Cloud Computing is winning a huge significance nowadays by influencing almost all technological developments in the field of information technology. All countries are experiencing quite a significant and explosive situation where international companies established in the USA, such as Amazon, Google, Microsoft, etc., strongly operate in the markets around the world with their Cloud Computing services. Worth mentioning is the fact that after the substantial investment realized by Amazon back in March 2006, the current $100 billion Cloud Computing market is controlled by Amazon at 33% and they are generating $33 Billions of revenue a year from Cloud Computing services. Additionaly, this market is followed by Microsoft Azure with 18% and Google Cloud with 8%. We can clearly distinguish the created advantage from such companies in global market share. In the top-8 list there are as well two Chinese companies such as Alibaba Cloud and Tencent Cloud, where both own 7% of the widespread market. The Chinese case is clearly related to the potential of the domestic market which is reflecting in such positioning on respective global market shares [1].

B. Situation in Albania

It is worth noticing that in Albania Cloud Computing technology is not yet finding widespread support and application among users and providers of information and telecommunication services. This situation is characterized by several factors that directly influence the development of such technology:

- Firstly, the Albanian market is considered small and therefore demand is extremely trivial.
- Secondly, required investments (CAPEX) for Cloud Computing technology are relatively substantial and do not justify the low market demand.
- Thirdly, there is no legal regulatory framework in place to protect and stimulate the development of such technology within the Albanian territory.

Despite the above mentioned, in Albania some improvement is needed to be done when it comes to Internet Services. At first, the Albanian Government has to call for the development of broadband internet access from all licensed operators throughout the country. This is a key component for the provisioning of Cloud Computing services. Although the number of Fixed Telephony lines has almost doubled from 2015 to the 4th quarter of 2019, this is still not enough, as many houses, and rural areas have no internet access at all [2].

Significant progress has been carried out by Mobile Telephony operators with the implementation of 4G Services, but even such development lacks full coverage of the entire territory. Such services were only deployed in high density residential areas, mainly in the big cities. Regarding 5G Technology, licenses have already been acquired by all operators, but its implementation has not started yet. There is no information regarding any implementation start-up, nor any timeline on when all the country's territory will be covered by the 5G technology. Furthermore, noteworthy is the fact that the absence of broadband fixed-line telephony services in rural and remote areas has made it impossible to offer private Wi-Fi services. To support our claim we can mention the fact that in one of the most important tourist attraction located in the northern part of country, the Albanian Alps, called Tamara, the mobile telephony signal coverage, and the 3G were enabled as late as the end of year 2016 [3]. Although the fiber optic infrastructure is being established altogether, for this area, fixed broadband telephony services are not yet present. To sum up, without spread developmet of Internet Access and Broadband Services we cannot pretend to have high demand, and need for Cloud services in Albania.

Although the indicators of providing Cloud Computing services are not at their optimum in the current domestic market, there is an expectation of market reaction and interest for the upcoming years. This will come as a result of the Albanian businesses' own demand for such services (example Hotels, Agricultures Companies, etc.). This spirit will be further intensified by the current brain drain situation (the migration of IT specialists to European Union countries), and this lack in Human Resources will drive businesses to choose such services as a feasible, and effective option. However, globalization and pricing policies set by big players are negatively affecting developing countries, like Albania and are not favorizing respective demand coming from these countries. Price differentiation for these countries is imperative in order to boost service

 ISBN: 978-1-61208-793-1

request as well as to extend usage of Cloud Computing technology throughout the country.

In a time when in the region, and around the world it is largely articulated that Cloud Computing is the technology of the future, at a summit organized in Zurich, Switzerland by HP company in October 23rd, 2012, it was claimed that the income generated from Cloud Computing for Swisscom (the largest IT company in Switzerland) was 5-6% of the total revenue, and the forecast for the following three years was that this figure would go up to 7-9%. So, this figure is viewed as very small, and the investment return for this technology is very low in comparison with other technologies such as NGN New Generation Networks (Voice-Internet-TV), with a rough estimation of 20% [14].

III. Cloud Computing Implementations in Albania

Currently some ongoing investments are being made in Albania. These investments are related to both public and private sectors. In the following sections we will depict general information for every case and some high level information respectively:

A. *Public Sector – Services provided by National Agency for Information Society*

In 2008, the Albanian Government started the implementation of Cloud Computing technologies for its purposes. This implementation was designed, and followed by the National Agency for Information Society (NAIS) [4]. The project implementation was carried out in partnership with the Infosoft Systems Company; the hardware systems devices used were those produced by HP Company, and the softwares used the applications platforms of Microsoft Company [6]. The Cloud Computing model implemented in this project was that of Private Cloud. The main objectives of this project consolidated the government IT resources into a single agency, reduced costs and improved the IT services of state agencies and ministries.

Although initially it was thought of a little structure and minimal use, it has already turned into a big and very important platform, which holds more than 100 different applications, such as public websites, intranets, government existing applications, etc. [7]. This implementation was enabled in order to centralize, and standardize the infrastructure used before in all state institutions of the country. In this regard, all ministries and state agencies receive better, and more qualitative service, and moreover it makes them only focused on services to citizens, not to the infrastructure, and its maintenance.

B. *Private Sector – Services provided by ALBtelecom*

Regarding the private sector, it is worth mentioning the most significant implementations to have attempted to join the global trend, and provide Cloud services locally in Albania. In April 2013, for the first time in the Albanian market, the implementation of a real public cloud infrastructure was launched. This implementation was carried out by ALBtelecom, and it was aimed for the entire range of businesses in Albania; starting from SOHOs, and ending with large businesses. This infrastructure was providing guarantee, quality, and faster delivery time of the given service through two data centers, one in Tirana, and the other in Elbasan, (as stated by Mr. Vahap Yeroglu, CIO of ALBtelecom) [12]. According to the publications [9], [13], this platform offered hosting of all kinds of applications, to all customers, ranging from websites to specific applications, such as those for the sales management, human resources management, Customer Relations Management (CRM) and finance management. Through this infrastructure ALBtelecom is able to offer private cloud services as well, customize the types of services for specific customers. It is reported that infrastructure can hold up to 30,000 clients [10]. The offering of this technology makes ALBtelecom the first private company in Albania to give this service recipients the opportunity to keep the data service within the Albanian territory.

C. *Private Sector – Services provided by Vodafone Albania*

On April 19th, 2013, Vodafone launched its new service, called Vodafone Cloud, in the Albanian market. This service enabled all prepaid and postpaid subscribers saving their personal data such as photos, videos, contacts and various files containing important information for these customers. This type of service could be accessed from different devices, such as computers, or smartphones, with various operating systems installed. The initiative undertaken by Vodafone became an ongoing strategy of Vodafone Group, offering the cloud Services to all its subscribers in all subsidiaries of the company. It is worth mentioning the fact that although Vodafone Albania has implemented some initiatives to provide the Cloud technology, they have been group-borrowed efforts from other countries rather than direct investments made in Albania. In this context, at group level, Vodafone has signed a contract with Microsoft for the provision of cloud services. As a result, Vodafone becomes Cloud Services Provider of Microsoft. Moreover, recently Vodafone has as well signed an agreement with Amazon to be Amazon Web Services Provider, at group level [8].

IV. Theoretical Background

The most well-known theories about technology adoption at company level are the DOI model, and the TOE framework [20]. Although there are studies where some authors have conducted respective research in one of such methods, or both of them, we will consider the second option to be more effective, and accurate [21], [27].

A. *Diffusion of innovation (DOI)*

DOI is a theory that examines how, and in which way new technologies, especially IT ones are adopted, and accepted among different communities and cultures [22]. This theory suggests that it is principally based on features of the technology, and on the perception of users regarding the system. Roger analyses the spreading of innovation, and has suggested some characteristics which influence the adoption of innovation, such as: relative advantage, complexity, and compatibility [24].

Relative advantage is the degree to which an innovation can bring benefits to an organization. Compatibility refers to the degree to which an innovation is consistent with existing business processes, practices, and value systems. Complexity considers the degree to which an innovation is difficult to use [25].

B. *Technology, organization, and environment (TOE) framework*

The process by which a company or organization adopts and implements technological innovations is influenced by the technological context, the organizational context, and the environmental context [23].

The technology context refers to characteristics of the technologies which are available for possible adoption by the organization, and the current state of technology in the organization. The organizational context consists of the organizational structure, the presence of innovation – enabling processes such as informal communication, and strategic behavior of top management, as well the size, and slack resources of the organization. The environmental context mixes nearby market elements, such as competitive pressure and regulatory support [26].

V. RESEARCH METHOD

We will base our research method by combining the DOI model, and the TOE framework, once they complement each other, which seems to be a better solution to this analysis, since it combines different contexts, as previously mentioned [29].

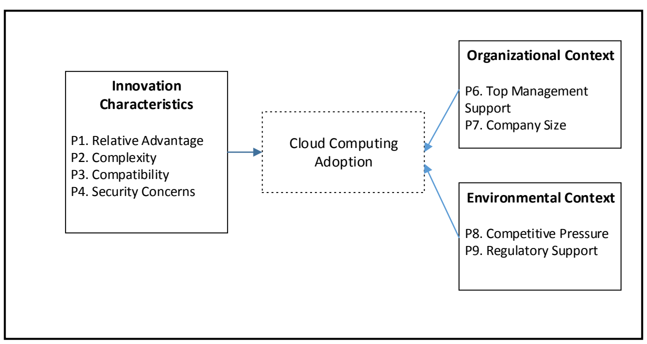

Figure 1. The research method.

The research method Fig. 1 includes the context of innovation defined in the DOI theory, and the two contexts presented on the TOE framework. The technology context was not included in our study due to the fact that such domain is quite complete, and easily deployable.

A. *Inovations Characteristics*

Considering the adoption, and implementation of Cloud Computing, we believe there are 4 variables in the context of the characteristics of innovation: relative advantage, complexity, compatibility, and security concerns [30].

- Relative advantage

Rogers [22] has defined relative advantage as the degree to which an innovation is perceived as being better than the idea it supersedes. In our analysis we will propose: P1. Relative advantage will positively influence Cloud Computing implementation.

- Complexity

Rogers has defined complexity as the degree to which an innovation is perceived to be relatively difficult to understand and use. In the same spirit, we will propose: P2. Complexity will negatively influence Cloud Computing implementation.

- Compatibility

Rogers has defined compatibility as the degree to which innovation fits with the potential adopter's existing values, previous practices, and current needs. In this regard, we will propose: P3. The high compatibility will positively influence Cloud Computing implementation.

- Security concern

With shared computing resources, security is a critical issue in Cloud Computing. Moving to cloud, a new security layer will convince companies or not. In this regard we will propose: P4. Security concerns will positively influence Cloud Computing implementation. [28]

B. *Organizational Context*

The organizational context is defined in terms of resources available to support the adoption of the innovation [31].

- Top Management support

Top management plays an important role because Cloud Computing implementation may involve integration of resources, and reengineering of processes. In this regard, we will propose: P5. Top management support will positively influence Cloud Computing adoption.

- Company size

A determinant factor in Cloud Computing implementation is company size. It is understandable that big companies have more advantages than smaller ones, considering the big number of resources, from human and technological aspects. In this regard, we will propose: P6. Firm size will positively influence Cloud Computing adoption.

C. *Organizational Context*

Environmental context includes the company's competitors, and the regulatory environment.

- Competitive pressure

Competitive pressure has long been recognized in the innovation diffusion literature as an important driver on the technology diffusion. It refers to the level of pressure felt by the firm from competitors within the industry [20], [32]. With the implementation of Cloud Computing, companies greatly benefit from a better understanding of market visibility, greater operation efficiency, and more accurate data collection. In this

 ISBN: 978-1-61208-793-1

regard, we will propose: P7. Competitive pressure will positively influence Cloud Computing implementation.

- Regulatory support

This refers to the support given by the authority in order to encourage the increase of Information Systems (IS) innovations in companies. Governments could facilitate the adoption of Cloud Computing by creating rules to protect businesses in the use of this system. In this regard, we will propose: P8. Regulatory support will positively influence Cloud Computing implementation [30].

VI. SURVEY RESULTS, SECURITY AND PRIVACY CONCERNS AND MISSING REGULATORY SUPPORT

In order to verify this research method, a survey was taken, among 250 IT community members all around Albania. 55% of the people were from the private sector, 30% from public sector, and the rest, i.e. 15% freelancers, and employees of non-profit organizations.

The main result from this survey was the fact that Cloud computing is not usually adopted, and implemented in Albania. One of the main reasons according to outcomes was the fact that there exists a big concern regarding security, and privacy. It consist of the fact that users think their data may be used for unknown purposes, without them being aware, without prior consent, or approval. Due to this concern, it is worth noting that the security and privacy of users' data on cloud platforms is regulated differently in different countries. Currently, in Europe the law on the protection of personal data, called General Data Protection Regulation (GDPR) applies to all US companies operating in Europe. By this law, all these companies are obliged to store, process and transmit the data of European users within the bonderies of the European Union. Consequently, all companies have been forced to build and operate their own computing infrastructures in the European Union countries for the services provided to European citizens.

In Europe, the general structure related to data protection and freedom of movement was established by Regulation (EU) 2016/679, (General Data Protection Regulation) [6]. National displacements realized by each Member State, assume considering national law as a guiding criterion. The European Commission, and the European Institutions have taken decisions, and carried out informative campaigns through explanatory, and orientation documents. The content of these documents (as in the case of the European Network and Information Security Agency (ENISA), [15] which speaks for the basic nature of the legal structure) have been adapted by all global companies operating in Europe.

In Albania, the law does not predict exactly what obligations (processing, transfer, and security) should be applied to the data located in the cloud. Currently there is only the law on personal data protection [9] which does not contain, nor treat the above mentioned features for Cloud Computing. Albanian legislation on personal data protection does not make it mandatory to process, store, and transmit data of Albanian users within the territory of Albania, but is more in line with the European laws on personal data protection, allowing Albanian users to use the services of international providers operating under the terms of the European Union. In this way, the economic activity of the Albanian operators as well as the revenues that the state institutions can benefit from providing, and receiving these services within the territory of the country are directly affected. In this regard, Albanian legislation has a huge missing regulatory framework which doesn't help adoption, and implementation of Cloud Computing inside the country's territory.

VII. CONCLUSIONS

Although Cloud computing can be seen as the new phenomenon which is set to revolutionize the way we use the Internet, there is much to be cautious about. There are many new technologies emerging at a rapid rate, each of which embracing technological innovations, with the potential of making humans lives easier. However one must be very careful to understand the limitations, and the security risks posed in utilizing these technologies. Cloud computing is no exception. Almost all large information technology companies such as Microsoft, Google, Amazon, Oracle, IBM, etc. provide opportunities of cooperation with local companies in order to use local cloud platforms of these companies for meeting legal obligations in the respective countries. With their support, local companies realize spending cuts, taking subsidies for infrastructure, and increasing the base number of customers consequently generating more revenue.

We have shown here that Albania has already hit the road towards the Cloud Computing era. Although implementations are few in number, they are very important from the technological and economic viewpoints.

In the Albanian legislation, currently there exists an open space for legal obligations to be met by companies for storing, processing, and securing of data that are residing in cloud computing platforms.

The lack of national laws for personal data protection has made the provision, and development of cloud computing services even less demanding, and with a shift towards Europe's service, mainly in Germany. The government's approach damages both corporations, and consumers. It creates an unfavorable climate for the provision of cloud computing services in the country by not providing legal obligations, and national security. Being in such conditions, Albanian customers prefer to get Cloud Computing services outside of Albania, for the sole reason that foreign companies are not interested in the Albanian data. But on the other hand, they do forget something crucial: data are the future gold, and should be taken care of as such.

This study was intended to explain the process of technology adoption, and to identify factors that affect the adoption of Cloud Computing in Albania. The model was designed based on the DOI and the TOE theory framework. The theoretical research carried out resulted in a set of contexts that may influence the adoption of Cloud Computing: characteristic innovations, organizational, and

 ISBN: 978-1-61208-793-1

environmental context. The study is a contribution in defining the model of research, and development in the dimensions that constitute it, as well it is a resource for all companies, and researchers that may use the conclusions of this study to expand their knowledge in this field, eventually developing other externalities for such emerging technology in Albania.

REFERENCES

[1] F. Richter "Amazon Leads $100 Billion Cloud Market", Statista, Feb 11, 2020. [Online]. Available from: https://www.statista.com/chart/18819/worldwide-market-share-of-leading-cloud-infrastructure-service-providers/, [retrieved: 04, 2020].

[2] Electronic and Postal Communications Authority (AKEP) "Statistical Indicators of Electronic Communications Market", Q4, 2019, Published on 11.02.2020. [Online]. Available from: https://akep.al/wp-content/uploads/2019/01/Raport-T4_2019.pdf, [retrieved: 05, 2020].

[3] M. Cakmak, ALBtelecom, and Eagle Mobile CTO "Work on fiber optic coverage, telephone signal and 3G services in the Kelmend area begins" ALBtelecom Mag, pp. 2-4, Jul 2016.

[4] E. Alite and J. Imami, ISTI 2014 "Technical Challenges and Economic Opportunities of Cloud Computing Implementations in Albania", Tirana, pp. 2-5, Jun 2014.

[5] E. Hasa "Albania consolidates government IT with HP and Microsoft cloud solution" *HP Case Study – Government of Albania*, pp. 1-3, January 2013.

[6] E. Hasa "Albania Standardizes Government IT with Cloud Services, Lifts Productivity 70 Percent" *Microsoft Case Study – Government of Albania*, pp. 2-5, February 2012.

[7] Albtelecom Press Release "Interview with Deputy General Director of Albtelecom and Eagle Mobile, Mr. Vahap Yeroglu", Tirana, Albania, pp. 1-2, January 20, 2014.

[8] A. Mucobega "ALBtelecom Unique Services E-Cloud", Tirana, Albania, pp. 1-2, April 8, 2013.

[9] Eagle Magazine "Albtelecom's Cloud Computing", Tirana, Albania, pp. 1-4, April, 2013.

[10] Telecompaper.com "Albtelecom brings cloud computing services to Albania", Tirana, Albania, April 10, 2013.

[11] Official Document, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). [Online]. Available from: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A32016R0679, [retrieved: 05, 2020].

[12] E. Alite and O. Shurdi, ISTI 2015 "Implementation of Cloud Computing platforms in Albania. Data Security and financial impact for storing them within national boundaries", Tirana, pp. 2-5, June, 2015.

[13] Act, No. 9887, date 10.03.2008, changed with Act No. 48/2012 "For protection of personal data", [Online]. Available from: http://www.inovacioni.gov.al/files/pages_files/ligji_nr_9887_date_10_03_2008_i_ndryshuar_me_nr__48__2012__per_mbrojtjen___.pdf, [retrieved: 05, 2020].

[14] J. Joseph, "PSTN services migration to IMS are SPs finally reaching the tipping point for large scale migrations?," *2010 14th International Telecommunications Network Strategy and Planning Symposium (NETWORKS)*, Warsaw, 2010, pp. 3-6, doi: 10.1109/NETWKS.2010.5624927.

[15] M. Siegel, F. Gibbons "Amazon enters the Cloud Computing business" - Standford University School of Engineering, pp. 8-9, May 2008.

[16] E. Jones "Cloud Market Share – a Look at the Cloud Ecosystem in 2020" [Online]. Available from: https://kinsta.com/blog/cloud-market-share/, [retrieved: 05, 2020].

[17] N. Drake and B. Turner "Best cloud computing services of 2020: for Digital Transformation" [Online]. Available from: https://www.techradar.com/best/best-cloud-computing-services, [retrieved: 05, 2020].

[18] D. Howlett "SAP announces critical changes in maintenance policy post 2025 - is it enough?" [Online]. Available from: https://diginomica.com/sap-announces-critical-changes-maintenance-policy-post-2025-it-enough, [retrieved: 05, 2020].

[19] Metrix Data 360 "Negotiating a Microsoft Enterprise Agreement: What to Expect" [Online]. Available from: https://metrixdata360.com/negotiate-microsoft-enterprise-agreement/, [retrieved: 05, 2020].

[20] T. Oliveira and M. F. Martins "Literature Review of Information Technology Adoption Models at Firm Level". The Electronic Journal Information Systems Evaluation, 14**,** pp.110-121, 2011.

[21] T.R. Leinbach "Global E-Commerce: Impacts of National Environment and Policy" edited by Kenneth L. Kraemer, Jason Dedrick, Nigel P. Melville, and Kevin Zhu. Cambridge: Cambridge University Press. *The Information Society,* 24**,** pp. 123-125, 2008.

[22] E.M. Rogers "Diffusion of Innovations, 5th Edition", Free Press, 2003.

[23] L. G. Tornatzky and M. Fleischer "*The Processes of Technological Innovation",* Massachusetts, Lexington Books, 1990.

[24] K. Zhu, S. Dong, S.X. Xu and K.L. Kraemer "Innovation diffusion in global contexts: determinants of post-adoption digital transformation of European companies", European *Journal of Information Systems,* 15**,** pp. 601-616, 2006.

[25] S.K. Lippert and C. Govindarajulu "Technological, Organizational, and Environmental Antecedents to Web Services Adoption" *Communications of the IIMA,* 6, pp. 146-158. 2006.

[26] K. Zhu, S. Xu, and K.L. Kraemer "Internet Technology Diffusion at the Firm Level: Empirical Evidence from Asia-Pacific Region, Europe, and North America" *In:* International Conference on Information Systems: International Symposium on Asia-Pacific, Washington DC, 2004.

[27] H.F. Lin and S.M. Lin "Determinants of e-business diffusion: A test of the technology diffusion perspective" *Technovation,* 28**,** pp.135-145, 2008.

[28] R. Schneiderman "For Cloud Computing, the Sky Is the Limit", *Signal Processing Magazine, IEEE,* 28, pp. 15-18, 2011.

[29] M. Espadanal and T. Oliveira "Cloud Computing Adoption by firms", MCIS 2012 Proceedings. 30, pp. 3, 2012.

[30] M. Espadanal and T. Oliveira "Cloud Computing Adoption by firms", MCIS 2012 Proceedings. 30, pp. 4-6, 2012.

[31] C. Low, Y. Chen and M. Wu 'Understanding the determinants of cloud computing adoption" *Industrial Management & Data Systems,* 11, pp. 1006-1023, 2011.

S. C. Misra and A. Mondal "Identification of a company's suitability for the adoption of cloud computing and modelling its corresponding Return on Investment" *Mathematical and Computer Modelling,* 53, pp. 504-521, 2011.

Comparison of Embedding Objectives for Next Generation Networks

1st Stelios Prekas
Dept. of Informatics & Computer Engineering
University of West Attika
Egaleo, Greece
Email: sprekas@uniwa.gr

2nd Panagiotis Karkazis
Dept. of Informatics & Computer Engineering
University of West Attika
Egaleo, Greece
Email: p.karkazis@uniwa.gr

3rd Panagiotis Trakadas
General Department
National & Kapodistrian University of Athens
Evia, Greece
Email: ptrakadas@uoa.gr

***Abstract*—Virtualization is considered a key factor for the commercial success of the future Internet and next generation networking in general. But, the plethora of modern day specialized applications migrating to an outsourced hosting environment, introduces new challenges into the process of optimizing the mapping of virtual resources over physical ones. In this paper, we categorize the variety of the current distinctive mapping objectives and performance metrics and we thoroughly compare the most common proposed embedding solutions with different optimization strategies under the same simulation circumstances, evaluating their suitability for serving network applications judging by their experimental results.**

***Keywords*—Virtualization; Next generation networks; Virtual network embedding; Embedding objectives; Performance metrics.**

I. Introduction

Network virtualization stands out as the catalyst technology for the future Internet and a major milestone to its architectural transition [1]–[3]. By its nature, network virtualization enables the co-existence of multiple intangible networks over a shared physical infrastructure. However, network virtualization regardless of having been established as the cornerstone of the modern Internet architecture, it faces a crucial decision of whether transitioning to new and innovative technological concepts or remaining limited by legacy architectures. Nevertheless, the majority of business and mainstream commercial electronic applications already face a transition of their own, dealing with the migration to outsourced virtual hosting infrastructures. Such online cloud based services are known as *Infrastructure as a Service* (IaaS).

IaaS is the delivery of hardware, either servers, storages or networks, and associated software, like operating systems or file systems, as a service [4]. It is the evolution of traditional hosting that does not require any long term commitment regarding in-house investments in data center equipment and other infrastructures and allows corporate administrators or even end users, to provision computational and network resources transparently and on-demand. The architectural philosophy of IaaS also led to the redesign of that market's business model. The role of the Internet Service Provider (ISP) is divided into two new ones; the Infrastructure Provider (InP), who deploys and maintains the physical equipment and the Service Provider (SP) [4], who has the responsibility to provide end-to-end services and applications to the end users. Thus, the appropriate conditions were created for new concepts to emerge like the Next Generation Networks (NGNs), that remodel the known and established architectures by separating and decoupling the underlying technologies and infrastructures from the services that run on top.

An NGN is a packet-based networking environment able to provide digital services including telecommunications, across multiple broadband, Quality of Service (QoS) enabled transport technologies [5]. Inside an NGN, all service-related functions are independent from underlying transport-related technologies. They offer unrestricted access to different SPs and support generalized mobility, which allows consistent and ubiquitous provision of services to end users [6].

The aforementioned massive virtual convergence of specialized applications has also added new levels of complexity to the already complex virtualization concept of Virtual Network Embedding (VNE). VNE, also being independent from the underlying networking technologies, is a fully compatible concept with NGNs, such as the 5G ecosystem and in particular its management and orchestration (MANO) platform. The introduction of these formulations will further allow to combine such technologies with the Software-defined Networking (SDN) and the Network Functions Virtualization (NFV) concepts [6], [7]. Studies regarding VNE, research novel strategies whose main goal is to solve the problem of mapping Virtual Networks (VNs) over substrate resources in an optimal way. This is in fact the gist of the modern virtualization philosophy. Through dynamic allocation of virtual resources onto existing hardware, the benefits gained by all the counterparts from the physical underlay, can be maximized. Even so, optimality is the key to the success of VNE. Arbitrary satisfaction to the demands of incoming service requests, may only lead to poor management of substrate networks and computational assets, thus wasting resources, capital, energy and time.

A substantial number of VNE algorithms have already been researched and developed [8], providing targeted solutions for each arising particular need. Nevertheless, the increasing demand in specialized NGN oriented applications, constantly springs new problem formulations resulting in new embedding objectives [7], [9]. A multitude of customized guaranteed services to the end user can only be made feasible by solutions

 ISBN: 978-1-61208-793-1

that incorporate combinatorial network features, like self-configuration and organization, with regard to different objectives, ranging from guaranteeing the QoS, ensuring economical profit, providing survivability, controlling energy efficiency or establishing network security [10]–[14].

In this paper, we document the variety of distinguished VNE objectives and performance metrics, we thoroughly compare the most common proposed VNE solutions with different optimization strategies under the same simulation circumstances and we investigate their suitability for serving NG applications judging by their experimental results. The remainder of this paper is organized as follows. In Section II, we present the state of the art and the current distinctive VNE objectives. Performance evaluation of the most common VNE algorithms under the same simulation environment are presented in Section III. Finally, Section IV concludes the paper and presents future work.

II. State of the Art

The study of VNE deals with the process of optimally binding demanded resources of the incoming Virtual Network Requests (VNRs) with nodes and links of the Substrate Network (SN). The VNE problem can be formulated mathematically. What is more, by reduction from other computer science optimization problems, like the multicommodity flow or k-multiway separator problems, VNE can be classified as an NP-hard decision problem itself [15].

A. VNE Problem formulation

Several papers have provided similar formulations for VNE [8], [12], [15]–[17]. Taking all into account, the VNE life-cycle can be broken down into the following main phases.

1) ***SN and VNR modeling:*** The common formulation directives typically suggest that the substrate network is denoted by an undirected graph $SN = (N_s, L_s)$ where N_s represents the set of substrate nodes and L_s the set of substrate links. Let $VN^i = (N^i, L^i)$ be a set of $i = 1, ..., n$ VNRs, where N^i and L^i represent the sets of virtual nodes and virtual links of the VN^i, respectively.

2) ***Setting capacities of element parameters:*** In addition, let $\widehat{R} = \prod_{j=1}^{m} R_j$, where $\widehat{R}$ is a vector space of node and link resource vectors, over resource sets $R1, ..., Rm$.
Let $rc : N_s \cup L_s \rightarrow \widehat{R}$ be a function that sets the capacity thresholds of the elements, either nodes or links, of the substrate network. Let $rd : N^i \cup L^i \rightarrow \widehat{R}$ be a function that sets the demand thresholds of the elements, either nodes or links, of all the VNs of the VNRs.

3) ***Establishing networking constraints:*** Depending on the specifications of each applicable scenario, there are some restrictions to be considered before the launch of the mapping process. The most obvious case is that the candidate substrate resources have to be adequate in order to support the demand requirements of the virtual resources being mapped. Exceptional cases may emerge, where redundancy is required, thus even more substrate resources may have to be reserved. Moreover, there are algorithms that are developed with embedded support for the distance constraint. The authors in [16] suggest that the function M_N :

$$\begin{matrix} M_N : N^i \rightarrow N_S \\ \forall (n^V, m^V) \in N^i \end{matrix} \ni \begin{cases} M_N(n^V) \epsilon N_S \\ M_N(m^V) = M_N(n^V) \iff m^V = n^V \end{cases} \tag{1}$$

which implements the mappings of virtual nodes upon substrate ones, should comply with the following constraint:

$$dis\left(loc(n^V), loc(M_N(n^V))\right) \leq D^V \tag{2}$$

where $dis(j, k)$ measures the distance between the location of two nodes j and k and D^V is the non-negative value expressing how far virtual node n^V of location $loc(n^V)$, can be mapped regarding the location $loc(M_N(n^V))$ of the candidate substrate node calculated by $M_N(n^V)$.
Additionally, the concept of the Hidden Hop (HH) property, first introduced in [18] and [19], establishes the portional consumption of the computational resources of those substrate nodes, which forward network packets within an embedded virtual path that aggregates more than one physical link. The importance of this realistic feature is highlighted in our study (see Section III-C *Evaluation results*).

4) ***Launching the desirable mapping strategies:*** A substrate resource is partitioned to host several virtual resources. A single substrate node can host several virtual nodes. In some cases, substrate resources can also be combined to create new virtual resources. This is the case for a virtual link which spans several physical ones, to form a network path in the SN. In this case, a virtual link between two virtual nodes n^V and m^V is mapped to a path in the SN that connects the substrate hosts of n^V and m^V. Each substrate link may then be part of several virtual links. As such, the mapping of virtual links to substrate paths describes a $N : M$ relationship [8].

Furthermore, there is the case where specialized algorithms can be utilized in order to map a single virtual link over several substrate paths, with flexible splitting ratios depending on the solution's constraints and objectives. This is called the path-splitting (PS) technique [12] and provides resource redundancy and demand distribution.

From the studied literature we can derive that the VNE problem can be divided in two interdependent sub-problems. The mapping of the demands of the virtual nodes over the available resources of the physical nodes, and the embedding of the demands of the virtual links interconnecting those virtual nodes, onto the residual network resources of the physical paths.

Taking the above into account, we can adopt the following summarized functions, suggested in [8], as the ones comprising the embedding process.

$$f_i : N^i \rightarrow N_S \tag{3}$$

$$g_i : L^i \rightarrow SP \subseteq SN \tag{4}$$

The function from (3) is used for the mapping of the virtual nodes belonging to N^i, over the physical ones belonging to N_s. And the function from (4) is used for the mapping of the virtual links belonging to L^i, onto a substrate path SP,

 ISBN: 978-1-61208-793-1

consisted of physical links belonging to L_s. Together, they form an embedding for each VN^i.

In any case, substrate resources should be spent economically, therefore the mapping procedure has to be optimized. Throughout the VNE oriented literature, every study adopts different objective strategies regarding the individual perspective of the research. Many pursuit the maximization of the revenue metric. The revenue of a VNR can be defined as the weighted sum of its demands. The total revenue is the sum of the revenue of all successfully mapped VNRs. For instance, the authors in [16] and [12] define the revenue of a VNR as follows:

$$\underset{G^V \epsilon VN^i}{\mathbb{R}(G^V)} = \sum_{e^V \epsilon L^i} b(e^V) + \sum_{n^V \epsilon N^i} c(n^V) \quad (5)$$

where $b(e^V)$ and $c(n^V)$ are the bandwidth and CPU requirements for virtual link e^V and virtual node n^V, respectively.

Whereas, in [17] it is stated that the main aim is to decrease the number of congested substrate links, since the authors argue that the substrate nodes do not affect the rejection probability of a VNR. The aforementioned study sets two main objectives; at first, the authors denote the set of α-congested substrate links by defining ζ as follows:

$$\zeta = \{e^s \epsilon L^s : B_{e^s} \leq (1-\alpha) B_{e^s}^{max}\} \quad (6)$$

where $B_{e^s}^{max}$ is the bandwidth capacity of physical link e^s and $0 \leq \alpha \leq 1$. The objective is to reconfigure the accepted VNRs (VN_{AR}^i) within the SN, in the aim of minimising the number of α-congested links.

$$minimize_{VN_{AR}^i}(|\zeta|) \quad (7)$$

Secondly, they formulate the minimization of the reconfiguration cost,

$$minimize_{AR}(\alpha\phi_n + b\phi_e) \quad (8)$$

where $0 \leq \alpha \leq 1$ and $0 \leq b \leq 1$ are the weights of the migrated virtual nodes (ϕ_n) and virtual links (ϕ_e) respectively.

*5) **Outcome evaluation:*** Depending on each individual embedding objective, the mapping strategies can be evaluated by examining the values of their performance metrics (see Section III-B *Performance Metrics*) and comparing them with the results of other VNE solutions, bearing identical or similar mapping objectives. Their evaluation can be achieved by the means of dedicated software tools, that can simulate the networking environment and stage the desired circumstances. In Table I, we present the most distinctive embedding objectives, distinguished among the related literature, as well as the performance metrics that the solutions proposed and tried to optimize. The final column suggests a possible matching between the documented solutions popular network application categories.

B. *Variations of embedding objectives*

While some approaches generically seek to optimize the common performance objectives, such as acceptance ratio, rejection rate, revenue or embedding cost, others are specially focused on specific aspects of the VNE problem. These distinctive objectives that have been researched throughout the existing literature, are summarized as follows.

*1) **Limit energy consumption:*** Energy consumption has always been a key profit factor for the InPs. Moreover, the rapid pacing of business digitization has led to an immense increase of the energy footprint of the IaaS industry. Although the energy expenses alone are a major concern of the InPs, excess IaaS energy consumption has an environmental impact as well. Nevertheless, the energy consumed by the mapped VNs and their respective applications can be managed and limited to an accepted minimum by specialized VNE algorithms, which can grant the mapping process energy awareness by tidying up the mapped virtual resources and enabling the hibernation of the substrate idle ones [10].

*2) **Ensure security robustness:*** Data integrity and transactional privacy are the two most important ingredients of communication trustworthiness. Additionally, regarding virtualization environments, where multiple VNs are layered over the same physical infrastructure. Throughout related literature,

TABLE I. DISTINCTIVE OBJECTIVES AND METRICS.

References	Algorithms	Performance Objectives	Distinctive Metrics	Candidate Applications
I. Fajjari, N. Aitsaadi, G. Pujolle and H. Zimmermann (2011) [17]	Greedy VN Reconfiguration	-Minimize rejection rate -Reduce cost of reconfiguration	-Rejection rate -Reconfiguration cost	Ad hoc and sensor networks
N. M. K. Chowdhury, M. R. Rahman and R. Boutaba (2009) [16]	(MIP), D-ViNE, R-ViNE	-Increase AR & revenue -Decrease cost	-AR, Revenue & Cost -Node/link utilization	Generic network applications
M. Yu, Y. Yi, J. Rexford and M. Chiang (2008) [12]	Modularized algorithm ("greedy" node mapping and Unsplittable Flow)	-Optimize link utilization -Minimize cost -Maximize revenue	- Cost & Revenue -Splitting Ratio	Fault-sensitive applications
Y. Zhu and M. H. Ammar (2009) [20]	VNA-I & VNA-II	-Minimize both the maximum node stress and maximum link stress	-Node/link stress	Generic network applications
J. Lischka and H. Karl (2009) [21]	VNM algorithm (based on SID)	-Efficient use of the underlying resources	-Revenue-to-Cost Ratio	Generic network applications
L. Gong, Y. Wen, Z. Zhu and T. Lee (2014) [14]	GRC-VNE (heuristic)	-Maximize revenue of the InP -Minimize the cost of the InP	-Revenue -Globall resource capacity	IaaS business software
Botero et al (2012) [10]	VNE-EA (MIP)	-Minimize the inactive substrate links and nodes after a mapping	-Inactive links and nodes -Incurred cost in the SN -The percentage of accepted VNRs	Energy consuming Cloud Data Centers
S. Liu, Z. Cai, H. Xu and M. Xu (2014) [11]	Heuristic algorithm	-Maximize Acceptance ratio -Maximize the long-term Revenue to Cost Ratio	-Acceptance ratio -Long-term revenue -Long-term Revenue to Cost Ratio	Online banking
X. Zhang, C. Phillips and X. Chen (2011) [13]	Integer Linear Program	-Find effective backup paths during the mapping process	-Average mapping delay -Cost of resilience	Multi-layer networks

early attempts have been made to introduce security aware VNE algorithms which involve both resource and security restrictions, by taking into account the specific security requirements of each incoming VNR [11].

*3) **Achieve VN survivability:*** The survivability of a network application can be achieved by means of increasing the resilience of the underlying network, either partially or fully. This means that fallback resources should be calculated during the VNR analysis. These backup resources (nodes, links or both) are bound to the embedded VN until its lifecycle is fulfilled. Fault-sensitive applications require total transparency whenever transitions between the primary and the backup resources take place and vice versa, thus granting a recovery from failure seamless to the application users [12], [13].

*4) **QoS driven embeddings:*** There are situations where VNRs are submitted with QoS-specific requirements. This fact can effectively differentiate the approach a SP should adopt in order to process each VNR. For instance, the given constraints and metrics regarding a VN that provides IP telephony services are weighted and evaluated quite differently in contrast to the ones related to a VN that provides peer-to-peer services [12], [13].

*5) **Maximize economical profit:*** We should not forget that VNE is a real life problem that incorporates the request of a commercial service that involves the leasing of networking and computational resources of a specific or several Infrastructure Providers. One way or another, all the entities involved throughout the commercial chain, spend money for the realization of the mappings of VNs over substrate resources. So, a general approach is to pursuit an objective of maximizing the metric of revenue [14]. This objective is directly proportional to the maximization of the acceptance ratio and, at the same time, the minimization of the embedding cost.

III. Performance evaluation

In this section, we first describe the simulation environment and the methods implemented for comparison reasons. Then, we present the adopted evaluation metrics and the noteworthy evaluation results.

A. Simulation environment

For the realization of our experiments we used the ALEVIN simulation tool [22]. ALEVIN is a Java based framework that focuses on modularity and efficient handling of arbitrary parameters of substrate resources and VNR demand, as well as on supporting the integration of new and existing algorithms and evaluation metrics [23]. A set of algorithms with high impact due to the popularity of their publication among the existing literature, has already been implemented [18]. Each computational experiment has been performed on a x86 architecture computer with an Intel quad core 2,67GHz processor and 8GB of RAM. Although the scale of the undermentioned simulated experiments did not produce the same volume of results as in a real life environment, large scale simulations requiring powerful testbeds, are a goal of future work.

1) Substrate network: By using ALEVIN's scenario generator, we created a physical network, modeled as an undirected graph. The size and scale of the substrate network and the capacities of its resources remain the same for all the simulation scenarios. A meshed topology of 50 nodes has been automatically generated, with a CPU resource capacity of 100 processing units (CPU cycles) for each physical node, and a bandwidth resource capacity of 100 link utilization units (BW units) for every substrate link.

2) Virtual network requests: The current architecture of the ALEVIN framework provides the ability to simulate VNRs in an off-line manner exclusively. Thus, two main simulation approaches have been adopted, regarding the VNRs to be mapped. The first deals with a large number of small scaled VNs and the second with a small number of VNs with greater scale, both with a progressive increase in the demanded CPU and Bandwidth resources. The exact specifications of all the simulation scenarios are presented in Table II.

TABLE II. VALUES OF PHYSICAL AND VIRTUAL PARAMETERS.

	Physical Network			Virtual Network			
Scenario ID	Number of nodes	CPU cycles	BW units	Number of VNRs	nodes per VNR	CPU cycles	BW units
sid1	50	100	100	50	3	15	15
sid2	50	100	100	50	3	30	30
sid3	50	100	100	50	3	30	50
sid4	50	100	100	20	10	15	15
sid5	50	100	100	20	10	30	30
sid6	50	100	100	20	10	50	50

3) Comparison method: In our evaluation, we compared four embedding solutions referenced in [18] and [23], that combine different node mapping and link mapping algorithms. The notations that we used to refer to these different strategies are presented in Table III.

TABLE III. COMPARED EMBEDDING STRATEGIES.

Notation	Embedding Solution Description [Reference]
GreedykSP	Greedy Available and K Shortest Paths [20]
GreedySplit	Greedy Available and Path Splitting [12]
CNLMsplit	Coordinated Node and Link Mapping with Path Splitting [16]
CNLMkSP	Coordinated Node and Link Mapping with k Shortest Paths [16]

B. Performance Metrics

In our experiments we used several performance metrics for evaluation purposes. We measured the average acceptance ratio (AR), the cost-revenue relationship (COSTREV) and the running time (RT), for VNRs of different proportions and with different progressive resource demands, as described in the previous sub-section. Here follows a brief description of the significance of the aforementioned metrics.

*1) **AR:*** describes the number of VNRs that could be completely embedded by the embedding algorithm, divided by the total number of VNRs. A simple conclusion could be that the higher the AR, the more efficient the embedding strategy. But, this can be proven as a rather unidimensional approach.

*2) **COSTREV:*** measures the proportion of cost spent in the substrate network, taking into account the revenue that has been mapped. However, when a set of VNRs has been included in the mapping process, probably the algorithms may not be able to map all of them. In this case, when the percentage of mapped revenue is not too high, the COSTREV metric is not a good indicator of the network mapping quality. That's why this metric can be modified, multiplying it by the percentage of mapped revenue.

*3) **RT:*** the time spent by each algorithm to complete its trials of mapping the entirety of all the incoming VNRs.

C. Evaluation results

In Figures 1 to 4, we present the most important of the produced results of our experiments. In the following paragraphs we analyze and discuss distinctive conclusions of the examined solutions, based on these results.

*1) **Coordinated node and link mapping (CNLM):*** All the studied cases that use the CNLM strategy performed comparatively well regarding the overall number of successfully mapped VNs. Moreover, as shown in Figure 2, the CNLM's embedding efficiency deteriorates in a slower rate than the other mapping strategies, as the VNRs' size and numbers become greater.

On the other hand, the solutions that were using CNLM performed these successful mappings with disproportionally higher cost in RT, as is profound in Figure 1, ranking them the worst for applications where promptly embeddings are crucial. This strategy seems best fitted for applications with a long term life-cycle, expecting high quality results no matter the RT cost and having no need for VN reconfiguration or mapping migration.

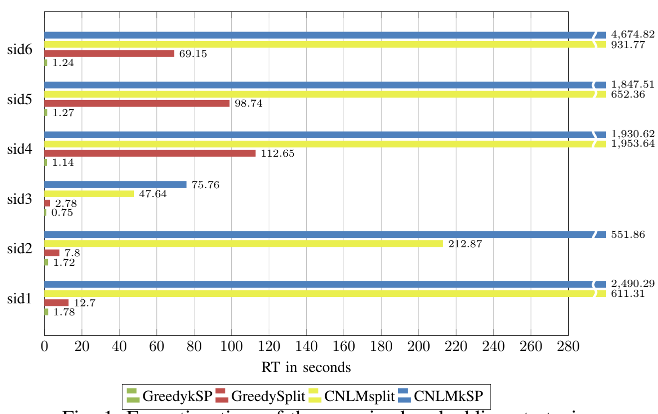

Fig. 1. Execution time of the examined embedding strategies.

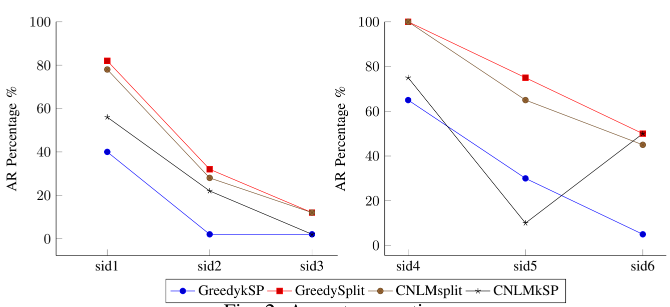

Fig. 2. Acceptance ratio.

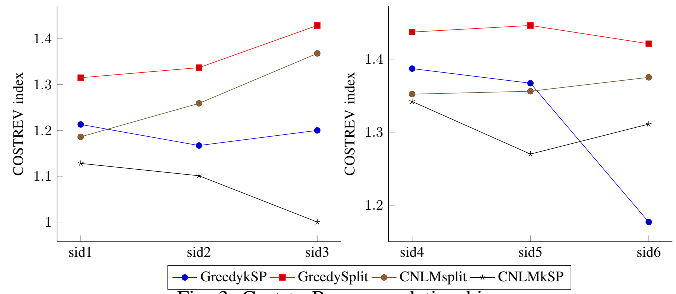

Fig. 3. Cost to Revenue relationship.

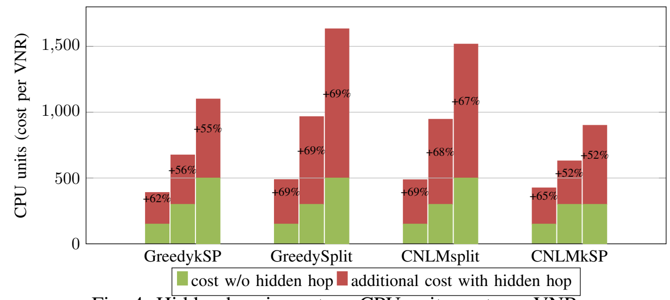

Fig. 4. Hidden hop impact on CPU units cost per VNR.

*2) **Link mapping with path splitting:*** Algorithms using the PS mechanism outperformed all the other studied solution in the field of most mapped VNRs, with a balanced RT. This resulted in fewer rejected VNRs and the algorithms obtained satisfactory AR percentages, as shown in Figure 2. However, due to the use of the PS technique, the cost of the embeddings in every scenario is slightly higher compared to the other solutions with similar ARs, as depicted by the COSTREV index in Figure 3. Moreover, the embedding cost increases even more when the HH property is activated. As depicted in the stacked bar chart of Figure 4, it is easily discerned that while the HH property is taken into account, the solutions that adopt the PS strategy are burdened with a higher cost in substrate CPU units per embedded VNR, almost by 10%, in contrast to the other studied algorithms.

Additionally, the fact that a number of virtual links are mapped over multiple physical paths, enforces the utilization of extra physical computational and networking resources, thus making the technique suitable of applications demanding redundancy and robustness. On the other hand, such utilization of physical resources results in spending increased amounts of operational energy, classifying the solution potentially as the least green among the examined strategies.

*3) **K shortest path link mapping:*** While the RT of the algorithm utilizing the GreedykSP is fast in every scenario as depicted in Figure 1, algorithms adopting the k shortest path link mapping technique perform poorly overall (Figure 2), especially when a large number of VNRs are involved, regardless of being small in size. But in spite of this weakness, the algorithms involving the K shortest path link mapping procedure, produces cheaper embeddings resource-wise, as Figure 3 indicates. Consequently, such approaches should be avoided for fault-sensitive applications and seem more qualified for stateless, express request serving grids like ad-hoc or sensor networks.

IV. Conclusion and future work

Our research studied the most common proposed VNE solutions under progressively increasing demands and within different optimization objective scopes. The goal was to compare novel strategies to the VNE problem, each one distinctive for its optimization approach, and derive the suitability of each solution for any of the key aspects of the VNE problem.

By analyzing the simulation outcomes, it was comprehended that each strategy could yield per case satisfactory results provided that the specifications of the scenario examined tend to be compatible with its optimization objective. However, by further analyzing the produced results of the trials presented in the previous section, it is discerned that although the values of the performance metrics can be reviewed and contemplated about, different network applications will always have varied requirements. All algorithms produce results and values for the same set of metrics and under the conditions of the same parameters at any given scenario, regarding the physical and virtual topologies as well as their respective resources. Some algorithms may excel in certain domains, where others may seem to produce poorer results. Yet still, depending on the prism of another embedding objective, the failure can become the prominent choice. Nevertheless, each proposed VNE solution is developed with predefined objective functions, thus obeying to certain preset constraints and rules. In order to conduct more comprehensive and accurate simulations, all examined solutions should be capable of functioning with other objective functions and not adhere to a certain set of constraints.

Mechanisms and techniques presented by their authors as capable of providing an optimal solution to the VNE problem, lack the architectural structure and ingenuity that could produce optimal solutions, tailored to address a variety of modern day scenarios. This need becomes more timely now that the NGN ecosystem may rely its success entirely on adaptive virtualization architectures. This novel per use case scalability feature, namely a VNE solution matrix, obsoletes the established practice of repeatedly modifying each algorithm entirely in order to meet the optimization objective specifications of a particular network application. Instead, it can provide a variety of metric specific optimal solutions, by even combining a set of metrics into a modular class of service VNE algorithmic toolbox, thus achieving the coveted goal of making available customized end-to-end services to the end users, in an truly optimal manner. This innovative concept will play a pivotal role in our future work.

V. Acknowledgment

This work is partially supported by University of West Attika.

References

[1] K. Tutschku, T. Zinner, A. Nakao, and P. Tran-Gia, “Network virtualization: Implementation steps towards the future internet,” Electronic Communications of the EASST, vol. 17, 2009, p. 14. [Online]. Available: http://dx.doi.org/10.14279/tuj.eceasst.17.216

[2] P. Barham et al., “Xen and the art of virtualization,” ACM SIGOPS operating systems review, vol. 37, no. 5, 2003, pp. 164–177.

[3] F. Alvarez et al., “An edge-to-cloud virtualized multimedia service platform for 5g networks,” IEEE Transactions on Broadcasting, vol. 65, no. 2, 2019, pp. 369–380.

[4] S. Bhardwaj, L. Jain, and S. Jain, “Cloud computing: A study of infrastructure as a service (iaas),” International Journal of engineering and information Technology, vol. 2, no. 1, 2010, pp. 60–63.

[5] I. T. Union, Next Generation Networks – Enhancements to NGN. ITU-T, Sep. 2016, chapter 1, pp. 2, language = english,.

[6] P. Trakadas et al., “Comparison of management and orchestration solutions for the 5g era,” Journal of Sensor and Actuator Networks, vol. 9, no. 1, 2020, p. 4.

[7] A. Pol et al., “Advanced nfv features applied to multimedia real-time communications use case,” in 2019 IEEE 2nd 5G World Forum (5GWF). IEEE, 2019, pp. 323–328.

[8] A. Fischer, J. F. Botero, M. T. Beck, H. De Meer, and X. Hesselbach, “Virtual network embedding: A survey,” IEEE Communications Surveys & Tutorials, vol. 15, no. 4, 2013, pp. 1888–1906.

[9] Y. Shekhawat et al., “Orchestrating live immersive media services over cloud native edge infrastructure,” in 2019 IEEE 2nd 5G World Forum (5GWF). IEEE, 2019, pp. 316–322.

[10] J. F. Botero et al., “Energy efficient virtual network embedding,” IEEE Communications Letters, vol. 16, no. 5, 2012, pp. 756–759.

[11] S. Liu, Z. Cai, H. Xu, and M. Xu, “Security-aware virtual network embedding,” in 2014 IEEE International Conference on Communications (ICC). IEEE, 2014, pp. 834–840.

[12] M. Yu, Y. Yi, J. Rexford, and M. Chiang, “Rethinking virtual network embedding: substrate support for path splitting and migration,” ACM SIGCOMM Computer Communication Review, vol. 38, no. 2, 2008, pp. 17–29.

[13] X. Zhang, C. Phillips, and X. Chen, “An overlay mapping model for achieving enhanced qos and resilience performance,” in 2011 3rd International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). IEEE, 2011, pp. 1–7.

[14] L. Gong, Y. Wen, Z. Zhu, and T. Lee, “Toward profit-seeking virtual network embedding algorithm via global resource capacity,” in IEEE INFOCOM 2014-IEEE Conference on Computer Communications. IEEE, 2014, pp. 1–9.

[15] E. Amaldi, S. Coniglio, A. M. Koster, and M. Tieves, “On the computational complexity of the virtual network embedding problem,” Electronic Notes in Discrete Mathematics, vol. 52, 2016, pp. 213–220.

[16] N. M. K. Chowdhury, M. R. Rahman, and R. Boutaba, “Virtual network embedding with coordinated node and link mapping,” in IEEE INFOCOM 2009. IEEE, 2009, pp. 783–791.

[17] I. Fajjari, N. Aitsaadi, G. Pujolle, and H. Zimmermann, “Vnr algorithm: A greedy approach for virtual networks reconfigurations,” in 2011 IEEE Global Telecommunications Conference-GLOBECOM 2011. IEEE, 2011, pp. 1–6.

[18] J. F. Botero et al., “Flexible vne algorithms analysis using alevin,” in Euroview 2011, 2011, in Proc. of the 11th Würzburg Workshop on IP: Joint ITG, ITC, and EuroNF Workshop ”Visions of Future Generation Networks”.

[19] J. F. Botero, X. Hesselbach, A. Fischer, and H. De Meer, “Optimal mapping of virtual networks with hidden hops,” Telecommunication Systems, vol. 51, no. 4, 2012, pp. 273–282.

[20] Y. Zhu and M. H. Ammar, “Algorithms for assigning substrate network resources to virtual network components.” in INFOCOM, vol. 1200, no. 2006. Citeseer, 2006, pp. 1–12.

[21] J. Lischka and H. Karl, “A virtual network mapping algorithm based on subgraph isomorphism detection,” in Proceedings of the 1st ACM workshop on Virtualized infrastructure systems and architectures, 2009, pp. 81–88.

[22] Alevin2 - tool for the evaluation of algorithms for embedding virtual networks. [Online]. Available: https://sourceforge.net/p/alevin/wiki/home/ [retrieved: May, 2020]

[23] A. Fischer et al., “Alevin-a framework to develop, compare, and analyze virtual network embedding algorithms,” in Open-Access-Journal Electronic Communications of the EASST, 2011, pp. 1–12.

Quasi-physical Model of Knowledge

In the search for a unified basis of integration of technology and society

Maxim Polyakov
Managing Partner
Noosphere Ventures Inc.
Menlo Park, USA
email: mvpolyakov@i.ua

Igor Khanin
Professor
National University of Water
Management and Nature Resources Use
email: khanin.ig@gmail.com

Nikolai Bormatenko
Scientist
Noosphere Ventures Inc.
Dnipropetrovsk, Ukraine
email: bormatenko@segater.com

***Abstract* — The domain of research is the process of understanding the field of information phenomena, including its practical level, in solving problems such as fragmentation of data, which is due to their diversity and variability. The solution to such problems requires immersion in the ontology (essence) of information phenomena and a relevant cognitive model. The problem of creating such a model is interesting, in particular, in that reducing the fragmentation of data will make data, including Big Data, and then programs and organizations, more unified in form, more diverse in content, in general - more useful. To solve the problem, the following conditions had to mature: an empirical base, formed by digitizing information practices; principles of convergence of heterogeneous fundamental ideas; reconstruction of understanding how cognition occurs in various fields of phenomena. The presence of conditions made it possible to develop a Quasi-Physical Model of Cognition (QPMC), where the object of cognition is the result of the convergence of the physiosphere, biosphere and infosphere (Vernadsky's noosphere); the method of cognition is the embodiment of forms of consciousness in artifacts; the ontology of the bodies of knowledge and innovation is signs constructions (programs, databases and knowledge, organizations. The QPMC model is addressed to those who in the field of information phenomena solve practical and theoretical problems of information and communication technology, cognition, semiotics, economics, knowledge management, which, unlike, for example, "digitization" or modeling, require reflection as a better understanding of the innovator's own actions.**

Keywords- infosphere; noosphere; model of knowledge; conscious phenomena; quasi-physical effects; innovation development; sign ontology; data ontology; organization ontology; noosphere ontology; convergence of knowledge.

I. INTRODUCTION

Albert Einstein formulated the principle that a problem cannot be solved within the framework of the belief system (one might say the "model of knowledge") in which it arose. Literally: "We can't solve problems by using the same kind of thinking we used when we created them."

In the 20th century, scientific knowledge faced a series of crises in physics. Overcoming them was akin to a "drama of ideas". There was a need to form new models of cognition that correspond to the new realities brought forth by special and general theories of relativity, atomic and nuclear physics and quantum theory. This forced researchers to turn to philosophy and ask questions of knowledge. In addition to Einstein were Niels Bohr, who received physical and philosophical education, Max Born, Werner Heisenberg, Louis de Broglie, Erwin Schrödinger, Peter Kapitsa and others.

These were new areas, but they belonged to the well-known world of physical phenomena. Solutions were found, but many did not agree with them at the time and still disagree today. Kirilyuk [12] evaluated the status of sciences where mature cognition of the physiosphere has faced many of the problems specific to the sphere of informational phenomena (infosphere), which is yet at the infancy of the path of cognition. According to Kirilyuk, the desire for a commercial result at the expense of understanding the essence of the phenomena studied (fundamental science) leads to the accumulation of unresolved problems. Nobel Laureate Geim [8] also acknowledges the insufficient productivity in basic research: "Finally, in my dream, humans realize social media can make some people very rich but cannot save the planet. The latter requires new fundamental discoveries." The Manifesto of the Slow-Science movement [38] supports this: "Don't get us wrong—we do say *yes* to the accelerated science of the early 21st century… However, we maintain that this cannot be all. Science needs time to think."

Kirilyuk cites other factors that also slow down the processes of cognition. Among them is what can be interpreted as a violation of the Hegelian rule of ascent from the abstract to the concrete. In our interpretation, it consists in applying mathematical and other high-level abstractions directly to empirical material. Cognition skips requisite steps of fundamental and applied theories. Ascent is replaced with risky leaps from deep-lying mathematical abstractions to concrete empirical material on the surface.

This article differs from previous articles developing the same topic by adapting the idea of sign bodies in combination with the idea of cognition as the embodiment of forms of consciousness. Due to this, the categories of noospheric thinking formed an integral system of categories embodied in the architecture of the QPMC, models of the noosphere, Paradigm Innovative Development (PIDev) and Vertical Integration of Knowledge (VIK). As a result, the architecture of QPMC and the graphic representation of each model included in the QPMC, are a combination of basic ideas through abduction and induction methods Thus, the framework of the mental structure of the article is formed by its drawings: the architecture of the QPMC model, the structure of the object of cognition, phylogenesis and ontogenesis of knowledge, ontology of the sign. It is convenient to start from these drawings. The presence of synonyms for some terms in the article is a temporary phenomenon. For example, the term “quasi-physical” is borrowed from philosophy. In QPMC it is specifically programs, databases and knowledge, projects and organizations, as sign (hyperphysical) constructions. Synonymy preserves the continuity between the carry over idea and its concretization. This is necessary for understanding and the subsequent development of QPMC. The establishment of rigid terminology would mean that the development of knowledge in the corresponding direction has been completed.

Article is dedicated to cognitive model. Section 2 describes the problem of infosphere cognition and background information of QPMC. Next Section gives a description of fundamental principles of QPMC in role of method and agenda of cognition of infosphere. Section 4 presents explanatory power and prognostic ability of QPMC. Finally, Section 5 contains salient points of QPMC and real contribution of implementing of this model.

II. PROBLEM STATEMENT

Vernadsky [33] classified "noosphere" as a completely different sphere of phenomena. Today it requires scientific knowledge. The concept is open to question. Most often, the noosphere is associated with information phenomena, that is, the infosphere. Vernadsky’s creative researcher Gruzman [10] specified that Vernadsky’s philosophical and meta-scientific studies should be considered in the context of his works on the theory and history of science.

Due to the lack of an empirical base, Vernadsky was unable to formulate his views on the noosphere into a constructive model of cognition. He used the concepts of matter and energy. He did not have a sufficient base of empirical knowledge of information, more precisely, of signs. At the same time, however, Vernadsky managed to put the whole world into one word (noosphere) as an object of knowledge. LaRouche [6] is one who managed to appreciate the significance of this step.

The work of philosopher physicists at the level of reflection in terms of the cognitive process was continued by Kuhn [14]. He focused on the phylogenesis of knowledge and the social aspects of cognition. His “Structure of Scientific Revolutions,” in particular the concept of paradigm, precipitated a lively discussion. Popper [25] was particularly interested in drawing a demarcation line between scientific and non-scientific knowledge. But the primary source of scientific knowledge is not scientific, but rather empirical knowledge. Losev [15] wrote: “What is science? A systematic representation of knowledge gained from experience (through the medium of external feelings), i.e. exposition and explanation of empirical phenomena.” Popper should have taken Kuhn's [14] position in order to think more broadly in terms of logical transitions from pre-scientific knowledge to paradigmatic (meta-scientific) and then to scientific. Perhaps it was psychologically difficult for an ambitious person such as Popper. As a result, attention has been concentrated on important, but limited questions of the truth and social role of paradigms (scientific achievements), rather than the structures of scientific revolutions (model of the genesis of knowledge).

Most likely Mamardashvili was under the influence of these discussions when he wrote the book “The Arrow of Knowledge”. Mamardashvili [18] formulated the principle of “destruction-reconstruction of understanding”, which in a more categorical form corresponds to a paradigm shift, that is, to a change in the cognitive model during the scientific revolution.

A detailed picture of cognition was given by Losev [16], using the concept of a name (which is actually a sign and information as a message). It can be considered as the most complete picture of the world as a noospheric object of cognition, in which there is substance, energy, and signs (information) in their relationship and development. It is difficult to work within this picture - it should be illustrated in relation to the current moment and the actual tasks of cognition.

The use of computers in economics and society has exposed and exacerbated the problems of the development of the infosphere, whose main features determine the structure of the noosphere. Under these conditions, Einstein's cognitive principle gains particular relevance. Today, the infosphere uses inherited and not always conscious models of cognition. It can be said that today’s knowledge of the infosphere is conveyed through inherited and poorly formulated models of knowledge. These interpretations then establish the nomenclature of scientific specialties, the standards for teaching scientific disciplines, as well as the methods of conducting scientific work and writing articles and books, etc.

Inherited models reflect the specificity of cognition in the post-paradigmatic phase of the development of natural sciences or the pre-paradigmatic phase in which humanitarian or practical knowledge is found. The question arises: is it possible to bring together the spheres of phenomena which are studied by the natural and humanitarian sciences? These are generally considered irreducible to a common basis. To solve this problem, one can apply Marx’s theories on transformed forms of consciousness. These thoughts were further developed by Mamardashvili. He introduced the concept of “quasi-physical effect of non-physical (conscious) phenomena” into philosophical discourse [18].

 ISBN: 978-1-61208-793-1

Today, knowledge of the infosphere is approaching the paradigm phase and it needs a model that covers all phases of the development of cognition, including pre-paradigmatic (empirical-heuristic), paradigmatic (scientific revolution) and post-paradigmatic (scientific). Such a model is necessary, especially in order to solve the problem of the ontology of information (more precisely, the sign). This problem has a thousand-year history. Examples of attempts to address this in the context of modern cognition may be found in the efforts of the scientific axiom of Stamper [31] and the monograph Tanaka-Ishii [32].

Knowledge of the ontology (essence) of data is necessary for the development of the applied paradigms of computer programs, data, business organizations, and economics. Knowledge of these ontologies aims to optimize and unify the structure of programs, data and business organizations. These results will serve as a tool for optimizing the information (more precisely, sign) infrastructure of the economy, from the enterprise level and higher, by reducing its rigidity (resistance to changes) and fragmentation, both of which are caused by the presence of a complexity barrier due to semantic diversity and data variability [24].

Moreover, according to Martens [20], the essence (ontology) of the economy can be defined as an information (or rather, a sign) machine for the production of materialized knowledge. Saussure, Hayek, and Coase [27] also noted the similarities of economics and sign systems, including language. Therefore, from both a practical and theoretical perspective, optimization of the sign structure is the primary task of the economy.

For a long time, continuous attempts have been made to create models of practical work in the infosphere, including elements of cognition. Some of these are called methodologies. Examples include Object Oriented Analysis and Design (OOAD) [3]. The subject of this methodology is software systems. Furthermore, these are: Theory and Practice of Business Processes (TPBP) [28]; Enterprise and System Architecture (ESA) [35]; and Semantic Technologies (ST) [1], as well as other tools .

Each is typically based on some significant and more or less adequate and principled position. For OOAD, this is “a program as an object, not an algorithm”, TPBP is “a program is a tracing-paper from business processes”, ESA is “overcoming the disintegration between technologies (information) and business,” ST - “technologies must take into account the semantics of data.” These tools have not become cognitive models that ensure the transition of the infosphere from pre-scientific to scientific knowledge. Obviously, such a model needs a comprehensive and well-integrated basis of the empirical, philosophical and scientific.

It would not be an exaggeration to mention cognitology or cognitive sciences (of knowledge). However, the attention of these disciplines is mainly the manipulation of knowledge-bearing data, or the processes of formalizing knowledge expressed in natural languages [9].

In fact, there are many models designed specifically for the knowledge of humanitarian phenomena or in claiming universality. These include the tektology of Bogdanov [2], the cybernetic approach of Wiener [34], the praxeology of Slutsky [30] and Kotarbinsky [13], among others. These and similar works can be considered as experiments in the search for a real model of cognition of non-physical phenomena. They have fulfilled their tasks and today they are potential subjects of an instructive history of knowledge and sources of positive and negative experience.

The “philosophy of information” by Floridi [7] focusses directly on the application of philosophy to explain the phenomena of information. Its default object, practically without restrictions, is any phenomenon to which the word “information” can be attributed. Such phenomena form the empirical basis of Floridi’s philosophy of information. However, according to Losev’s definition (“What is philosophy? — A worldview compiled by synthesizing scientific information”), philosophy is not directly connected to empirical foundations, but rather to the achievements of science. Moreover, science is “A systematic exposition of knowledge gained from experience (through the medium of external feelings) [15].”

Thus, the development of science relies on achievements of empirical knowledge and the development of the philosophy of scientific achievements. More precisely, the dependence is stepwise and mutual.

The activities of the research community of The International Federation for Systems Research (IFSR) [37] are directed at forming the foundations of the information sciences, as in the case of the cognitive model. Its members working on the Fuschl Conversations project state: “We want to build a general theory that conceptualizes reality as a field containing meaningful human social interactions, as well as technology and nature [4].”

The goal is ambitious, but is it achievable? Indeed, the history and current state of cognition show that cognition of the whole is effected in parts, due to the limited capabilities of a person. In this case, the knowledge of each part is divided into stages. Periods of continuous evolutionary development of cognition are replaced by discrete transitions between areas and stages of cognition. This means that a single object (synchrony) and the process (diachrony) of cognition require a rational decomposition that corresponds to the current values of development. It is doubtful that a “theory of everything” could be created, but it is possible and necessary to build a model of cognition, according to which a large whole is divided into parts in a certain way, and the processes of their cognition are divided into phases. At the same time, the procedures for knowing each part in each phase should be regulated with universal terms.

The article may seem abstract, and therefore complex. Indeed, one can engage in the digitization of simple information practices without realizing how this is done. But in order to learn how information processes, characterized by a higher diversity, work, it’s necessary to understand how cognition of these processes works. In this case, their mutual abstraction is a tool to simplify the problem.

The subject of the article (model of knowledge) is self-sufficient. However, it is part of a broader plan and aims to achieve practical goals. Based on the QPMC, a solution to the problem of the ontology (essence) of the sign is proposed [23]. Based on the ontology of the sign, applied ontologies of programs and data are developed [24]. They are used as a basis for the development of data infrastructure, architecture of programs and organizations. The QPMC model relates to the theory and practice of cognition. As a practice, it is needed when it becomes impossible to cognize and create without reflecting. So far, there are enough analogies, conjectures and associations for creativity, its practicality is not obvious. QPMC is necessary for solving problems such as a barrier of semantic diversity (fragmentation) of data and the tasks of developing a data infrastructure, architecture of programs and organizations resulting from them.

It should be understood that this article is devoted to the self-sufficient topic of knowledge of the sphere of information phenomena. At the same time, it describes the problems of this sphere, for the solution of which the QPMC is intended, and provides links to articles where this is discussed in more detail. Section 4 of the article provides schematic examples of explanatory, prognostic, and productive strengths of the QPMC model. In subsequent works, each area of application of the QPMC is supposed to be considered from the perspective of the final economic result.

III. THE ARCHITECTURE OF THE QUASI-PHYSICAL MODEL OF COGNITION

In order to apply Einstein's above-mentioned cognitive principle to the infosphere, the Quasi-Physical (Noospheric) Model of Cognition (QPMC) is proposed, as shown in Figure 1. Vernadsky's noospheric thinking is embodied and transformed in this model, taking into account the characteristics of current times.

The main sections of the model depicted in Figure 1 are as follows: Consciousness, Cognitive-Creative Activity (CCA) of consciousness, the sphere of phenomena, the totality of CCA effects of consciousness, the model of knowledge ontogenesis (Vertical Integration and the Parabola of Knowledge), the model of knowledge phylogenesis (Paradigm Innovative Development), the ontology of phenomena, and transformation of the sphere of phenomena. The definitive structure of noospheric thinking consists of these and other QPMC concepts.

The scientific activity of consciousness occurs within the framework of actual being, which grows out of potential being (meon, meonal environment) as a result of the cognitive and creative practical activity of consciousness. CCA of consciousness triggers phenomena and forms pre-scientific knowledge about them. They are fixed in consciousness, specifically in natural language. These descriptions can be varied, integrated, differentiated, formalized by digitization, etc. Problem-oriented lexicons can be derived from natural language, and can be called ontologies. These operations, as a connection by association and fact rather than by essence and meaning, can be attributed to modeling. However, the scientific knowledge modeled after natural sciences is different.

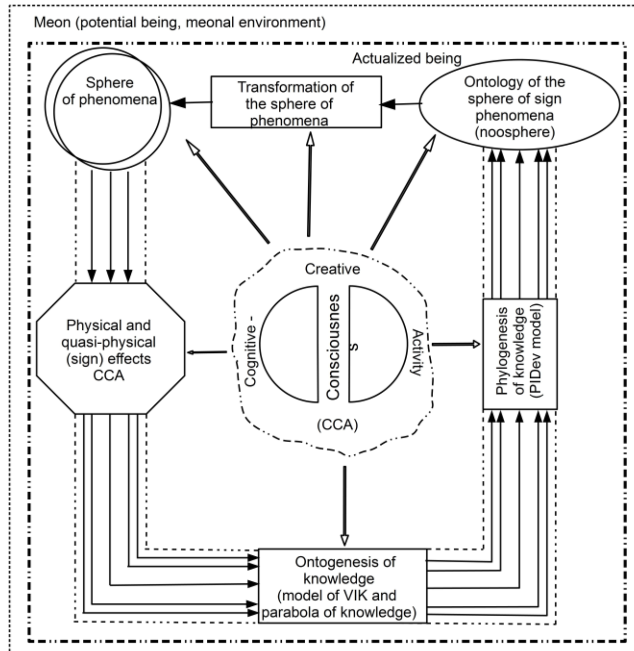

Figure 1. Architecture of quasi-physical Model of Cognition.

Pre-scientific knowledge corresponds to their embodiment in the effects caused by CCA. These can be natural or artificial objects (artifacts). First of all, they should be distinguished by their nature (ontology), which is established by way of acts of scientific knowledge (Figure 1). The cycles of knowledge can be repeated many times. In this case, each block can be connected to any previous block through feedback loops.

Thus, the consciousness arising from life continues the practical (not yet theoretical) cognitive-creative development of a potential being. The result of this activity is the formation of an empirical knowledge base in the form of a targeted combination of physical and quasi-physical effects.

On this basis, it is impossible to deny the existence of the empirical roots of mathematical knowledge, even though some researchers may at times have difficulty trying to establish them. There can be no science for science. The deepest abstractions should be the quintessence of infinitely diverse empirical phenomena.

Scientific knowledge begins with immersion from the concrete (many effects, that is, the results of activity) into the abstract, and continues through the ascent from the abstract to the concrete. This provides the opportunity to optimize existing and create fundamentally new and more effective (optimal) artifacts. This is the principle of ontogenesis, that is, the Vertical Integration of Knowledge (VIK). The ontogenesis of knowledge involves individual phenomena and their structures, forming a hierarchy up to the sphere of phenomena. The phylogenesis of the sphere of phenomena and its structures depends on the breadth of coverage and completeness of the ontogenesis of knowledge of the phenomena. One might say that phylogenesis is a change in the degree of paradigmization of these structures.

With the accumulation and ordering of practical knowledge, the studied and transformed sphere of phenomena passes from the empirical-heuristic (pre-paradigmatic) to the paradigmatic, and then to the scientific (post-paradigmatic) phase of development. The ontology of sign phenomena is the result of these processes in the infosphere. It enables subjects of innovative activity to think and act based on an understanding of the essence of phenomena, that is, essentially. Today, natural sciences are close to this. QPMC implies the convergence of the natural sciences and the humanities. Therefore, the construction of the ontology of the subject area is a fundamental requirement of this model.

The fundamental principles of QPMC are the well-known laws of philosophical logic. Mamardashvili [24] formulated them in a generalized form as fundamental philosophical abstractions. There are Plato's embodiment of cognition, Descartes's cogito and Marx's criterion of truth (practice). Similar principles, upon which the world is built, are contained in the principle of sufficient reason as amended by Losev [17]. These are also Peirce's [21] categories of "Firstness," "Secondness," and "Thirdness," and particularly the relationship between them.

The classical principles in QPMC are supplemented by ideas about transformed forms of consciousness and the quasi-physical effects of non-physical phenomena [19]. These give QPMC its quasi-physical nature. The connection between the fundamental traditions of cognition and its development and the convergence of natural science and humanitarian knowledge is a fundamental feature of QPMC against the background of such teachings as Shchedrovitsky's [29] "methodology". This methodology at its core breaks with the traditions of scientific knowledge. The foundation for it is thinking and activity [29]. But it is not Peirce's "Firstness", that is, this is not the ontology as "quality" [21].

A. Quasi-physical model of the object of knowledge

The objects of knowledge for QPMC are phenomena (manifestations of unknown entities yet to be determined) as well as the structures in which they consist, up to the world level which contains all of the phenomena. World structure is divided into spheres, each of which consists of phenomena of one nature (one quality, one ontology). Their incarnations (effects) are physical bodies and living organisms. Since ancient times, signs have been called the fundamental essence. Peirce [21] gave this thought a definitive form. His "firstness" (simply "quality") can be compared with *ratio fiendi*, that is, the need to become, or the "physical" necessity. "Secondness" (relations) is the need for mathematics, that is, relations in space and time (*ratio essendi*), especially the need for action, or activity of consciousness (*ratio agendi*). Peirce's "thirdness" means a universal connection, including consciousness. One can compare the *ratio cognoscendi* of the law of foundation with thirdness. According to Peirce [21], thirdness is accomplished by signs, which are studied by semiotics. Peirce provides a concise and accurate answer to the question "What function does signs perform?" Firstness and secondness can be interpreted as the signified, with thirdness as the signifier part of signs. We can then say that the objective world, that is, the world as a result of scientific knowledge, is constructed of signs.

Vernadsky paid little attention to the words "sign" or "information". The idea of the world phenomena interconnection is expressed by the word "noosphere", i.e., sphere of mind and thought. It would be naive to assume that his point of view maintained confidence in human intelligence. Most likely, he implied that thoughts, and subsequently the signs they produced, were, are, and always will be elements of the world order.

At the same time, Vernadsky was a researcher of the geosphere and biosphere, and has achieved a great deal in his field. Exploring and acting, he represented the world as a whole. His thinking was global and historical. "Social history," he argued, "is a continuation of natural history [33]." From Vernadsky's point of view, it follows that the biosphere contains the geosphere, but is not reduced to it. This is not accidental, considering they each consist of different entities. Similarly, the noosphere must also contain the biosphere and not be reducible to it. It should be a sphere of sign, or informational phenomena (if we understand information as messages).

The idea of the noosphere is encoded in Peirce's frame of reference, and the idea of the Peirce's sign likewise in Vernadsky's system. Their belief systems can serve as answers to each other. The biosphere is the whole world as it was yesterday in the scope of its scientific knowledge, but the noosphere is the further development of the world today. The main features of the described model are presented in graphical form in Figure 2.

In Figure 2, the vertical structure of the noosphere is superimposed on the vertical of integrations and a parabola of knowledge, which are discussed below. The full circle corresponds to the noosphere. It covers a segment that corresponds to the biosphere. In turn, it contains a segment symbolizing the physiosphere (an extension of the concept of the geosphere). Concentric rings and parabolas of knowledge show that a unified model of cognition is applied to the knowledge of various spheres. It is characterized by levels of abstraction and the asynchronous nature of the development of scientific knowledge of various fields. The most developed is the physiosphere, followed by the biosphere. The scientific knowledge of the noosphere is just beginning.

It should be noted that phenomena differ from effects (objects) in their uncertainty, as do objects from phenomena in their more complex structure. We can say that objects are phenomena that have found their embodiment and application.

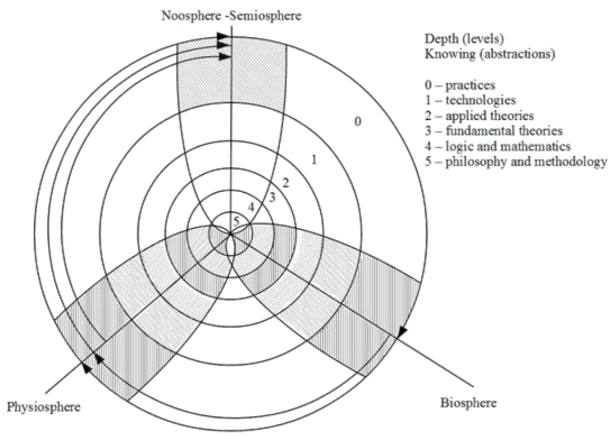

Figure 2. Quasi-physical model of the object of cognition.

By comparing the views of Peirce and Vernadsky, we can draw a conclusion about the fundamental role that signs play in the noosphere and in the world. The pansemiotism of Peirce and Vernadsky's belief in the power of knowledge is also justified by the hidden potential within information technologies so far. Therefore, it is important to know the ontology of signs. For this it is necessary, first of all, to separate non-physical phenomena from physical ones and to find out how they differ in principle from each other.

The practical basis for addressing non-physical phenomena is information. In natural language, the word information is synonymous with message words. However, since Claude Shannon applied it in the sense of one of the measures of the recipient's attitude to the message, in science it has been used as a homonym. Therefore, in order to avoid ambiguity, particularly in fundamental matters, non-physical phenomena can be called signs or sign bodies, since messages (information) actually consist of signs.

In order to not limit knowledge to philosophizing, only strongly formalized sign formations ought to be included in the empirical base of integral knowledge. Let these be called sign constructions. Today, computer programs and the data they process are maximally formalized.

Thus, the physical effect is a fragment of the meonal environment (potential being). Consciousness perceives it as a whole, consisting of physical components interconnected by physical connections. Physical effects can be natural or artificial (artifacts) objects. Natural objects are also products of at least cognitive activity, although cognitive and creative activities are difficult to separate. An example of creative activity alone can be the replication of patterns.

Consciousness in the process and as a result of cognition transforms being. Biological phenomena and their corresponding effects possess all the properties of physical phenomena and effects. Therefore, by skipping the biological effects, one can immediately switch to the quasi-physical effects.

A quasi-physical effect is a fragment of actual being. Consciousness only perceives it as a whole, consisting of two physical or quasi-physical components, which, unlike physical effects, are connected in the mind by a collaborating relation. Between physical and quasi-physical effects (objects) there is no insurmountable gap. For example, in the case of a computer program, by connecting parts (by loading program text and data into a computer system), an object created as a quasi-physical thing is transformed into an autonomous physical thing. This is a computer that operates under the control of the signals within the text of the program, printed on a machine-readable medium. Of course, in order to take liberties to think and speak this way, it is necessary to postulate the existence of temporal objects (things-processes), which are characterized by structure and states.

Thus, a quasi-physical object consists of physical objects that fall into two parts, interconnected by means of consciousness with a collaborating relation. In certain phases of its existence, a quasi-physical object can become a physical object.

B. *Model of Paradigm Innovative Development (PIDev)*

The PIDev model (Figure 3) resembles Kuhn's structure of the scientific revolutions, but with fundamentally different characteristics.

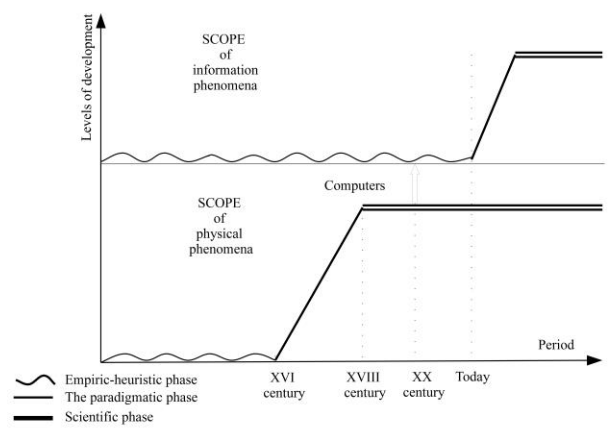

Figure 3. Model of Paradigm Innovative Development (PIDev).

First, PIDev is an integral model. Its objects of knowledge are macrostructures. These are the spheres of phenomena or their components. The PIDev model allows us to differentiate and compare them according to the degree of maturity of knowledge. In addition, the PIDev model combines several phases of cognition within the sphere of phenomena. As an integral model, PIDev could be called a model of the phylogenesis of knowledge.

Figure 3 compares the PIDev models, one relating to the physiosphere and the other to the infosphere. The wavy line indicates the empirical-heuristic (pre-paradigmatic) phase, the oblique does so to the paradigmatic, and the double line represents the scientific (post-paradigmatic) phases.

The diagram shows that computers emerged in the physiosphere in the scientific phase of its development. They are applied to information practices (effects of sign phenomena) which are related to the empirical-heuristic phase of the development of the infosphere. The knowledge of the infosphere falls behind that of the physiosphere. Therefore, IT is a physical data processing technology applied to information practices. To solve the problem of the disintegration of IT and business, it is necessary to bridge the gap between the development of knowledge of physical data processing technologies and information practices.

Secondly, the fundamental features of cognition identified by Kuhn were based on the experiences of matured natural sciences in the scientific phase of development. In contrast, the PIDev model begins with the formation of objects of knowledge and the science corresponding to it. In this case, the knowledge of such a sphere of phenomena emerges in the form of innovative development, and this depends more on innovative business than on official science. Business forms the empirical base needed for scientific knowledge of the infosphere. Business may not, however, be aware that this clears the path to knowledge. If the integration of knowledge and business is done purposefully, it can serve as the archetype of innovative development, termed the “knowledge economy”.

Thus, thirdly, the PIDev model is an integration of knowledge and management (business). The cooperative development of management and knowledge is innovative development. In the empirical-heuristic phase, its main driving force is enterprise personnel, from workers to top managers. Vernadsky [33], who studied the history of knowledge in Europe, drew attention to the important role of the masses in preparing the natural-science revolution.

Fourthly, the main task of the pre-paradigmatic phase of innovative development is to formalize existing practices. In doing so, they use methods based on experience (analogies) or guesses (heuristics). An example is the digitalization of information practices through data processing technologies. Although IT is called information, it still exists without proper philosophical and scientific justification in terms of information, more precisely, signs. Moreover, attempts to form it are constantly repeated. However, we are not able to solve this problem within the new conditions of the development of the infosphere using existing humanitarian models of cognition, whether explicit or hidden, or with models borrowed from the natural sciences. For the scientific revolution of the infosphere to be possible, an act of “destruction - reconstruction of understanding” must occur [18]. Only then can a cognitive model relevant to the problems of developing the infosphere be formed.

The model mentioned above may consist of known parts, while at the same time give an unexpected overall picture of what is happening. Cognitive structures such as Marx’s transformed forms of consciousness, quasi-physical effects of non-physical phenomena [19], ascent from the abstract to the concrete [11], Firstness, Secondness and Thirdness [21], information, sign, knowledge, innovation, etc., can be viewed in a new light, pose questions and suggest answers, in accordance with graphic representation of QPMC, PIDev, VIK, ontology of sign and corresponding references. Such a model can become an effective tool for strengthening innovation. A striking example is evidenced in Mendeleev’s periodic table of the elements.

C. Vertical Integration and Parabola of Knowledge

Phylogenesis, or macroscopic development of knowledge regarding the sphere of phenomena, occurs as a result of the accumulation of “micromutations” in the bodies of the infosphere, including programs, data and knowledge bases, enterprises. These are individual innovative acts. They can even use deeply abstract philosophical or scientific innovations, but at the same time they must result in a concrete practical, including commercial, result. The logic behind the development of an empirical knowledge base is crucial.

The process of formation of innovations, which can be called the “ontogenesis of knowledge”, is presented in Figure 4 by the model of Vertical Integration and the Parabola of Knowledge (VIK).

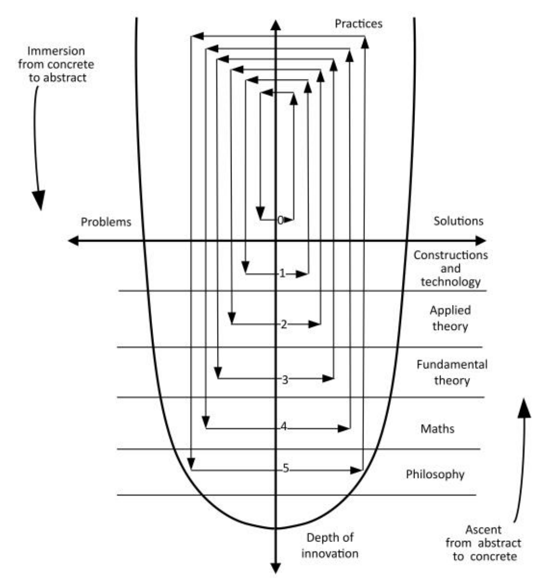

Figure 4. Model of Ontogenesis of Knowledge (VIK).

The zone of the upper half of the figure is reserved for specific (practical, materialized) knowledge. In this case,

Copyright (c) IARIA, 2020. ISBN: 978-1-61208-793-1

specific processes are considered to be temporal things. This upper half corresponds to the zero level of abstraction.

The lower zone is reserved for abstract knowledge, which does not refer to individual things but to their sets, ignoring the uniqueness of each elements within the set. The zone is divided into five levels of abstraction: structures and technologies; applied theories; fundamental theories; mathematics; and philosophy, including methodology.

The zone of the left half is occupied by problems, and on the right are solutions. The right branch of the parabola symbolizes the ascent from the abstract to the concrete [11]. Moreover, the shape of the parabola (the left branch) suggests the need to supplement the Hegelian figure of knowledge with a symmetrical figure, which is an immersion from concrete to abstract.

The inline rectangles in the figure are innovative cycles. If innovative changes only affect practical knowledge, then innovation is characterized by a zero level. The digitalization of information practices is at such a level. It is based on physical data processing technologies, the formation and development of which is within the responsibility of the physiosphere, and not the infosphere.

Figure 4 shows inline innovation cycles of different depths. In these cycles analysis (immersion from the concrete into the abstract) alternates with synthesis (the ascent from the abstract to the concrete). The Ranganathan model [26], called the spiral of knowledge, reduces to a similar alternation. The concurrence reinforces the fundamental importance of this pattern. Additionally, both the VIK graphic model and Hegel's “ascent” depict the concrete above the abstract. This coincidence, which also manifests at the figurative level, is not accidental. The empirical base of knowledge is inexhaustible and is constantly expanding. At the same time, levels of abstraction are limited to one center.

As illustrated in Figure 4, the VIK model expands the interpretation of the term “innovation”. According to the model, innovations are changes that can occur at any of the six levels of abstraction shown in the figure. The depth of innovation is determined by the maximum level of abstraction. Innovations that influence philosophy, methodology, mathematics and/or fundamental theory are paradigmatic. VIK is not so much a classification of knowledge as it is a unit of knowledge with meaning, fullness and completeness. It is not pertinent to divide knowledge into scientific, educational and professional subjects until the vertical of knowledge is formed in accordance with the levels of abstraction.

Scientific revolutions can occur in fields both old and emerging. As development accelerates, the time intervals between them ought to be reduced. Prior to a scientific revolution, problems are constantly arising that need to be addressed. Old sciences struggle with them, creating new scientific subjects. We are then to wait until some “invisible college” finds a radical solution to address the problem. In this case, the inevitable question is the formation of organizational structures to carry out the vertical integration of knowledge. The prototypes of such structures may be research networks where appropriate vertically oriented associations can be created.

The VIK model clearly demonstrates the usefulness of QPMC. It is an alternative to elemental empirical, heuristic and associative cognition. The VIK model shows that abstractions are the result of simplification of practical knowledge and, in turn, are used to systematize them.

IV. The Innovative Potential of Noospheric Thinking

Considered by Polyakov et al. [24], Marx's thesis on practice as the main criterion of truth in a series of fundamental philosophical abstractions takes the last place (in order, but not in meaning). Consequently, noospheric thinking (such as QPMC), like any model or theory, should be evaluated in terms of its innovative potential.

The explanatory power and prognostic ability of models and theories are closely associated. The explanatory power of the cognitive model allows us to understand what is happening with the knowledge of today’s sphere of phenomena. Predictive ability helps anticipate what events may occur in the development process.

As a communal (common) intellectual capital, the cognition model should have theoretical and practical productivity and serve as an environment for the formation of fundamental and applied theories, as well as for the development and application of structures and technologies that can be of a physical, combined or sign nature.

A. Examples of explanatory power of noospheric thinking

Noospheric thinking uses well-known abstract statements and formulations of practical problems. The use of many of them is not limited to the infosphere. In this case, QPMC clarifies their meaning.

1) Plato's allegory of the cave.

Plato’s Cave is often used as an argument in defense of idealism. Mamardashvili [24] saw in it the formulation of the problem of the fundamental philosophical abstraction of the embodiment of the understood. From the QPMC standpoint, this allegory is a paradigm of knowledge that has not yet lost its relevance. It is linked to the zero-innovation cycle as part of the VIK model and parabola of knowledge.

2) Hegel's ascent from the abstract to the concrete.

In the VIK model Hegel’s imaginative vision and logic are explained and developed. If concreteness is an ascent, then abstraction is an immersion in the essence of things, and not a separation from them. Therefore, in the VIK model, abstraction is also a movement toward the foundation, an immersion in depth. Indeed, it is a method of immersing from the concrete into the abstract. This method, based on the empirical basis, forms paradigms, systematizing the results of empirical-heuristic knowledge. Thus, in the VIK model, Hegel’s imaginative vision and logic are explained

 ISBN: 978-1-61208-793-1

and developed. Indeed, abstraction is an immersion in the essence of things, and not a separation from them. Then concretization can only be an ascent.

3) Unnatural modeling is a simplified version of the Hegelian ascent from the abstract to the concrete

The VIK model and the parabola of knowledge enables us to see a simplified version of the Hegelian ascent from the abstract to the concrete in non-natural modeling methods, particularly in the mathematical modeling of information phenomena. The simplification consists of the absence of transitional steps between abstractions and practice

The use of QPMC as a tool to move from modeling to cognition based on understanding is considered by Polyakov et al. [22].

4) IT status today

The logical-conceptual apparatus of QPMC can be used to determine the status of IT today. It follows that IT can be called technology if it is data processing technology. It is also possible to apply the term “information” to IT, if we mean information practices. Thus, IT today is a physical data processing technology applicable to informational (noospheric) practices.

5) What is a program?

A program is a well formalized sign effect of a non-physical programming phenomenon. To understand the program, one must understand the sign. To understand the sign, one need to understand the program. It is therefore no wonder that after hundreds of monographs and dissertations on the topic of “what is a program”, this problem does not disappear from the registers of scientific papers [36]. It should be noted that to understand in this case means to find a suitable abstraction - the key to each of the many diverse objects. Such an abstraction should make the diverse and complex array of objects uniform and simple.

6) What is data?

Within the structure of the sign paradigm from the QPMC model, data is defined as the designating part of sign construction [23]. This, above all, allows to eliminate the perception of data as local phenomena, inalienable from consciousness.

7) What is an organization?

Organization, in particular, business can be determined through the program. Indeed, a software application for economic purposes is a model of a fragment of an economic organization. All agreed upon programs for all fragments of the organization would form a complete model of the organization. The data structure reflects the architecture of the organization. The data processing algorithms correspond to the organization management function implemented by data users. A similar idea about the similarity of programs and organizations was expressed by Brödner [5]. The Quasi-physical Model of Cognition opens the possibility for its application.

8) What is the economy?

Today’s answer to this question is ambiguous. Martens worked with many emerging economies in the world. To better understand how they will respond to external influences (assistance, loans or investments), he represented them through the abstraction of a knowledge-producing information machine. He considered such an abstraction to appear productive [20].

Indeed, the economy is controlled through information, and all that it accomplishes is knowledge embodied in things or processes. Institutional science in this case can be considered as outsourcing. QPMC actually implements this idea in a detailed and in-depth format. It offers a convergence of cognitive and economic activities.

B. Examples of prognostic ability QPMC

Based on QPMC (noospheric thinking), we can make the following predictions:

1) A decrease in the intensity of innovative ideas in the physiosphere may occur, which may require the search for new areas of phenomena suitable for intensive innovative development. One such area is the infosphere. Opinions that its potential has been exhausted do not correspond to facts;

2) The stocks of semantically simple, rarely changing information practices suitable for digitization by existing data processing technologies are near exhaustion;

3) There is both a need and an opportunity to create an ontological theory of signs on the empirical basis of programs and data for economics and business;

4) In the near future, it is possible to develop (“upgrade”) existing IT to a state of truly informational, that is to say, sign technology;

5) Creating methods to increase the flexibility of data structures and to bring them closer to the status of an infrastructure resource in the near future is likely to happen; and

6) As the quasi-physical approach develops, modeling as a tool for the innovative development of the infosphere will give way to inventions based on ontological theories.

C. Noospheric thinking (QPMC) as the intellectual capital of a knowledge corporation.

When there are scientific revolutions, business organizations that claim to be the “knowledge corporation” cannot tarry. They accumulate empirical knowledge, thus it must be systematized. To accomplish this, corporations must have relevant cognitive models capable of solving the current theoretical and practical problems in the development of the infosphere.

D. Examples of possible theoretical productivity QPMC:

1) Formation of the paradigm of ontology of the sign and the theory of sign construction on the empirical basis of computer programs, databases and business organizations;

2) Formation of the paradigm and theory of computer programs based on the paradigm of ontology of the sign;

3) Formation of the paradigm and data theory based on the paradigm of ontology of the sign;

4) Formation of a paradigm (architecture) and the theory of economic organization based on the paradigm of ontology of the sign; and

5) Formation based on the theory of knowledge economy as a semiotic machine that produces knowledge in sign and reified form.

E. Examples of QPMC practical productivity should be:

1) The development of flexible, unified data infrastructures that reduce fragmentation of a single data field, from the enterprise level to the global economy scale;

2) The development of information and software tools to support intellectual activities in terms of imparting meaning to textual works;

3) The optimization and integration of computer program architectures; and

4) The optimization and integration of business organizations architectures.

V. CONCLUSION

QPMC is based on the concepts that form a hierarchical structure. The peak of this hierarchical structure is the architecture of the QPMC, followed by: the structure of the noosphere; PIDev and VIK models. Further, VIK, for example, is based on the division of objects of knowledge according to the levels of abstraction and integration of levels within the framework of the Hegel's concept of ascent from the abstract to the concrete, which is supplemented by immersion from the concrete to the abstract.

The real contributions of QPMC are:

- to establish a connection between many ideas from literary sources. (The generality of these literary sources is not obvious);
- to establish a connection between these ideas with the empirical base of the infosphere.

The implementing of QPMC consist in controlling the transition of infosphere cognition from the pre-paradigmatic to the post-paradigmatic phase of development.

The main result of the formation of QPMC is the creation of tools for the theoretical and practical solution of the problems of the infosphere; for example, the architecture of programs and organizations and the usefulness of ICT depends on the problem of fragmentation of data. For this, with the help of QPMC a paradigm of ontology of a sign has been developed, the concept of data and programs have been defined. See section IV of this article as well [22] [24].

The first iteration of the formation of QPMC, which can be considered as an independent topic, is completed, but in this case it is considered only as one of the stages in understanding the sphere of information phenomena.

From the point of view of cognitive activity, QPMC represents a fairly complete core (paradigm) of the theory of knowledge, connecting the empirical base of information phenomena and a number of basic ideas. The sources of its further development will be the results of practical application (replenishment of the empirical base) and the deepening of ties with basic ideas.

REFERENCES

[1] T. Berners-Lee, J. Hendler and O. Lassila, "The Semantic Web," Scientific American, vol. 5, pp. 34–43, 2001.

[2] A. A. Bogdanov, Tectology: Universal Organizational Science. Moscow: Economics, 1989.

[3] G. Booch, Object-Oriented Analysis and Design with Applications. Redwood City, CA: Addison Wesley Longman Publishing Co., Inc., 2004.

[4] S. Brier, The Necessity of Trans-scientific Frameworks for Doing Interdisciplinary Research. [Online]. Available from: https://www.researchgate.net/publication/237477137_Fuschl_Conversations_Foundations_of_Information_Science_Towards_a_New_Foundation_of_Information_Cognitive_and_Communication-Science 2020.06.12

[5] P. Brödner, "Software is orgware – a semiotic perspective on computer artifacts," Proceedings of the International Conference on User-driven IT Design and Quality Assurance, vol. 5, pp. 1–8, 2005.

[6] Current Strategic Studies: The Vernadsky Strategy by Lyndon H. LaRouche, Jr. [Online]. Available from: https://larouchepub.com/eiw/public/2001/eirv28n18-20010504/eirv28n18-20010504.pdf 2020.06.12

[7] L. Floridi, The Blackwell Guide to the Philosophy of Computing and Information. Oxford: Wiley-Blackwell, 2003.

[8] A. Geim, Be Afraid, Very Afraid, of the Tech Crisis. [Online]. Available from: https://www.scienzainrete.it/files/Andre_Geim__Be_afraid%2C_be_very_afraid_of_the_world%27s_tech_crisis__Financial_Times.pdf 2020.06.11

[9] G. Gorman, The Uniqueness of Knowledge Management – or the Emperor's New Clothes? [Online]. Available from: https://www.emeraldgrouppublishing.com/librarians/info/viewpoints/unique.htm 2020.06.10

[10] H. Gruzman, Dogged Genius of Russia. [Online]. Available from: http://lit.lib.ru/g/gruzman_g/genii.shtml 2020.06.12

[11] G. W. F. Hegel, Science of Logic. Vol. 1. Moscow: Musl', 1970.

[12] A. P. Kirilyuk, "The Last Scientific Revolution", in Against the Tide: A Critical Review by Scientists of How Physics & Astronomy Get Done, M. Lopez Corredoira and C. Castro Perelman, Eds. Boca Raton: Universal Publishers, pp. 179–217, 2008.

[13] T. Kotarbinsky. A Treatise on Good Work. Moscow: Economics Publ., 1975.

[14] T. Kuhn, The Structure of Scientific Revolutions. Chicago: University of Chicago Press, 2012.

[15] A. F. Losev, Higher Synthesis. Unknown Losev. Moscow: CheRo, 2005.

[16] A. F. Losev, The Philosophy of Name. Moscow: Pravda, 1990.

[17] A. F. Losev, Music as a Subject of Logic. Moscow: author's edition, 1927.

[18] M. Mamardashvili, The Arrow of Cognition: Sketch of Naturally-Historical Gnosiology. Moscow: School "The languages of Russian culture," 1997.

ISBN: 978-1-61208-793-1

[19] M. Mamardashvili, The Experience of Physical Metaphysics. Moscow: Progress-Traditsiya, 2008.

[20] B. Martens, The Cognitive Mechanics of Economic Development and Institutional Change. New York: Routledge, 2004.

[21] W. Nöth, “Charles Sanders Pierce,” in Criticism and Semiotics, vol. 3-4, I. Silantyev and V. Shatin, Eds. Novosibirsk: Institute of Philology SO RAN, pp. 5–32, 2001.

[22] M. Polyakov, I. Khanin, and N. Bormatenko, “From modeling to technologization of conscious phenomena based on understanding,” 9th International Conference on Information and Social Science, 2019, pp. 468-477, doi:10.18848/1832-3669/CGP/v14i03/27-45.

[23] M. Polyakov, I. Khanin, and N. Bormatenko, “Quasi-physical approach to solving the semiotics practical productivity paradox,” 4th Annual International Conference on Philosophy: Yesterday, Today & Tomorrow (PYTT 2016), Singapore, November 28–29, 2016, pp.1-10, doi:10.5176/2382-5677_PYTT16.3

[24] M. Polyakov, I. Khanin, N. Bormatenko, and S. Kosenchuk, “Ontology of sign: a key to information and technological advancement of the knowledge society,” The International Journal of Technology, Knowledge, and Society, vol. 14, 2018, pp. 27–45, doi:10.18848/1832-3669/CGP/v14i03/27-45.

[25] K. Popper, “Science: conjectures and refutations, philosophy of science: a personal report,” in British Philosophy in the Mid-Century, C. A. Mace, Eds. London: George Allen and Unwin, pp. 155-191, 1957.

[26] S. Ranganathan, “Library science and scientific method,” in Annals of Library Science, vol.4. India: NISCAIR-CSIR, pp. 19–32, 1957.

[27] E. Saburov, “Economy as drama of ideas,” in Social Sciences and Contemporary World, vol. 3, S. Valentey, Eds. Moscow: Nauka, pp. 83–89, 2010.

[28] A. Scheer, ARIS: Business Process Frameworks. Moscow: Vest-Meta Technologia, 1999.

[29] G. P. Shchedrovitsky, Sign and Activity. Moscow: Vostochnaja literature, 2005.

[30] E. Slutsky, “Etude to the problem of building formal and praxeological foundations of the economy,” Economic thought. Reader. Kiev: Knowledge, pp. 498–509, 2007.

[31] R. Stamper, Exploring the Semantics of Communication Acts [Online]. Available from: http://www.vits.org/konferenser/lap2005/Paper%2014-LAP.pdf. 2020.06.09

[32] K. Tanaka-Ishii, Semiotics of Programming. New York: Cambridge University Press, 2010.

[33] V. Vernadsky, Works on the History of Science. 2nd edition. Moscow: Science, 1988.

[34] N. Wiener, Cybernetics: Or Control and Communication in the Animal and the Machine. 2nd revised ed. Paris: Hermann & Cie, Camb. Mass. (MIT Press), 1961.

[35] J. A. Zachman, “A framework for information systems architecture,” IBM Systems Journal, vol. 26, pp. 276–292, 1987.

[36] https://anr.fr/Project-ANR-17-CE38-0003 2020.06.11

[37] http://ifsr.org/ 2020.06.11

[38] http://www.slow-science.org/ 2020.06.12

Advances of Generative Adversarial Networks: A Survey

Dirk Hölscher*, Christoph Reich*, Martin Knahl*, Frank Gut* and Nathan Clarke†
*Institute for Data Science, Cloud Computing and IT Security, Furtwangen University, 78120 Furtwangen, Germany
Email: {dirk.hoelscher, christoph.reich, martin.knahl, frank.gut}@hs-furtwangen.de
† Centre for Security, Communications and Network Research, Plymouth University, Plymouth, U.K.
Email: n.clarke@plymouth.ac.uk

***Abstract*—Generative Adversarial Networks (GANs) are part of the deep generative model family and able to generate synthetic samples based on the underlying distribution of real-world data. With expanding interest new discoveries and recent advances are hard to follow. Recent advancements to stabilize training, will help GANs to open up new domains using adjusted architectures and loss functions. Various findings show, that GANS can be used to generate not only images, but is also useful for text and audio creation. This paper, presents an overview of different GAN architectures, giving summaries of the underlying fundamentals of each presented GAN. Furthermore, this paper presents look into four application domains and lists additional domains. Additionally, this paper summaries datasets and metrics used to evaluate GANs and present recent scientific advancements.**

Keywords–generative adversarial networks; machine learning; deep learning.

I. Introduction

Generative Adversarial Networks (GANs) have revolutionized deep generative models used to learn data distributions with unsupervised learning, to generate synthesized samples for various domains coming from a number of sources resembling the true data distribution. Introduced by Goodfellow et al. [1] in 2014 it has been the focus of many researchers to apply the concept to new domains, introduce new loss functions and architectures, and new approaches to stabilize the training process. Applicability of GANs is growing rapidly and reaches new domains while achieving better results by applying adapted architectures and joining domain typical methods. The shared goal of all GANs is to reach a Nash Equilibrium [2], meaning that neither the discriminator nor the generator can be further improved. The idea of GANs originates from the game theory, a classical two-player zero-sum game [3], with only one winner. Nowadays, GANs are not only used to create synthesized images [4]-[5], but they can also create text [6], perform image and video translations [7]-[8], video summaries [9], copy objects into another image [10], help to reconstruct archaeological findings [11] or create images from text descriptions [12]. Each of these GANs is based on a different architecture with varying loss functions. All these variations and specialized architectures combined with missing unified evaluation methods and datasets [13] make it difficult to compare and evaluate the performance of GANs. Therefore, it is important to find available and commonly used metrics and datasets used in the scientific community to be able to compare the different approaches and underlying architectures regarding their performance. This survey gives an overview for researchers and summarizes the current state-of-the-art of GANs based on six research questions about architectures, domains, evaluation (metrics and datasets), prevailing problems and advances in research.

The paper is organized as follows: Section II introduces the methodology used for this survey and gives a short overview about publications. In Section III, common and novel GAN architectures are listed and summarized. Afterwards, Section IV gives a short overview and describes research done in four domains, where different styles of GANs are deployed. Section V summarizes datasets used for evaluation by the previous described GAN architectures and lists GAN evaluation methods. Furthermore, a database search was conducted to find the most prominent evaluation methods. Section VI takes a look at occurring problems while training GANs and shows advances in research. SectionVII concludes this paper.

II. Methodology and Overview

The following section will detail the methodology used to create the survey and shows the process how contributions were searched and selected. Furthermore, this section gives an overview about publications distribution in recent years.

A. Research Questions

The Research Questions (RQ) regarding GANs are as follows:

RQ1: What GAN architectures exist?
RQ2: In what domains are GANs utilized?
RQ3: What datasets are used to evaluate GAN performance?
RQ4: What metrics are deployed to validate GANs?
RQ5: What challenges exist when working with GANs?
RQ6: Advances in research?

All subsequent Sections are structured to answer the above mentioned RQs.

B. Database Search

The search was conducted using the following four databases:

- IEEE Xplore
- ACM Digital Library
- NIPS Proceedings
- arXiv (used to find additional publications in the first three mentioned databases)

Search terms were focused on clear easy structures. A general search without using any specialized search string yields a result of more than 430,000 results for IEEE Xplore and ACM alone. By searching titles, full text and abstracts about the full term and synonym achieved the best results. Searching was done with the following search term: *"Publication Title": "generative adversarial network*" OR "gan" AND "Abstract": "generative adversarial network*"*. With gaining popularity of GAN, more publications starting to appear as shown in Figure 1.

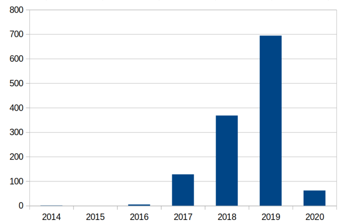

Figure 1. Annualised Number of publications

A big chunk of publications were published in conference proceedings (89%) while only 11% were journals. There is a continuous rise in publications following 2014 with 2015 being an outlier for the searched databases (arXiv excluded, due to limited filtering options and inclusion of only peer-reviewed publications). At the time of this survey 2020 had 62 publications, excluding early access titles. It can be assumed that publications will continue to increase in the coming years.

III. Generative Adversarial Network Architectures

With the introduction of GANs by Goodfellow et al. [1] in 2014, various new adapted architectures addressing a variety of problem domains, were introduced in the following years. This section will list and describe some of the most prominent and concept wise interesting GAN architectures and is related to RQ1. The different architectural patterns are sorted by their date of publication to indicate the growth and progress in recent years. Figure 2 shows a generic GAN architecture with all components, as well as the input for Generator(G) and Discriminator(D). G takes a random noise vector as input and generates a fake sample forwarded to D as input. D know the real data distribution, and classifies the input either as being real or fake and backpropagates the error.

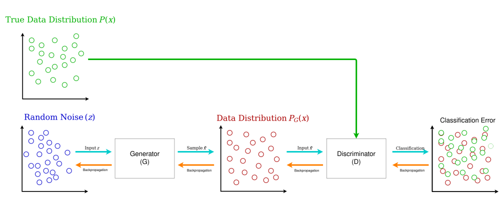

Figure 2. Generic GAN Architecture

A. Generative Adversarial Nets

The first GAN introduced by Goodfellow et al. in 2014 [1], introduced the concept of GAN's as a two-player minimax game where two neural networks G and D compete against each other. The generator takes random noise as input and tries to generate an output consistent with the original data. The discriminator tries to classify if the input came from the original data or from the generator. This can be images for example. The goal of D is to maximize the probability to distinguish samples that came from the original data and sample generated by G. At the same time G is trained to minimize the probability of getting caught by D, meaning that generated samples forwarded to D are classified as part of the original data. The equation of the minimax game is given with the following function V(G,D) also known as a vanilla GAN:

$$\begin{aligned}\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] \\ + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))]\end{aligned} \tag{1}$$

Where x are the data real instances, $E_{x\sim p_{data}(x)}$ is the expected value over all x $D(x)$ is D's probability estimation that an instance from x is real, z is the noise input vector, $G(z)$ is the G's generated output using z, $E_{z\sim p_z(z)}$ is the expected value over all generated fake samples and $D(G(z))$ is the probability that a fake instance is classified as real.

B. Conditional Generative Adversarial Nets

The cGAN by Mirza and Osindero [4] is a extension of the vanilla GAN, where both D and G are conditioned using additional information, for example class labels. This can be achieved by adding another input layer for D and G that contains the conditioned information. The input noise from the vanilla GAN is combined with additional information and combined to form a joint hidden representation (e.g., a multilayer perceptron (MLP) hidden layer). The discriminator takes the original data as well as the conditioned information as input for a discriminative function. This minimax game can be represented as the following equation given by the authors:

$$\begin{aligned}\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x|y)] \\ + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z|y)))]\end{aligned} \tag{2}$$

The equation of the minimax game is the same as (1) with the additional of the condition y to the discriminative function. Figure 3 illustrates the structure of a cGAN with the conditioned input. where the blue circles are the random noise z and green is the conditioned information y fed into the layers of the network. Both z and y are G's input, where y acts an restriction for instance creation. D utilizes y to determine if a sample is real or fake and knows the real data instances as well as the condition y.

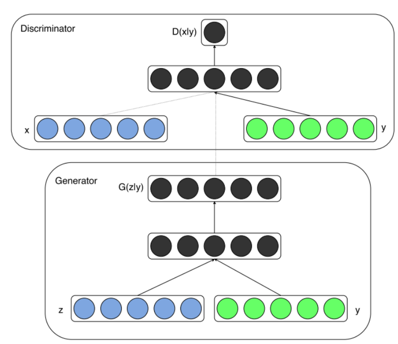

Figure 3. cGAN Architecture [4]

C. InfoGAN

InfoGAN proposed by Chen and Deng [14] in 2016 splits the input random noise vector into two vectors. This is done to make the GAN learn meaningful representations. The first vector contains incompressible noise(z), while the second vector (latent code) targets the data's semantic features(c). G takes both vectors as input and is represented as *G(z,c)*. To prevent trivial codes, where the generator ignores c to satisfy:$P_G(x|c) = P_G(x)$, meaning G is able to ignore the

 ISBN: 978-1-61208-793-1

latent vector to generate samples. Therefore, an information-theoretic regularization is introduced stating that between *c* and *G(z,c)* the amount of mutual information must be high. Based on information theory, mutual information is the measurement of learned knowledge between two random variables *Y* and *X*. This can be expressed as the difference between two entropies (X and Y) with their mutual information (I):

$$I(X;Y) = H(X) - H(X|Y) = H(Y) - H(Y|X) \quad (3)$$

This ensures that the information in c is not lost in G. This leads to an information-regularized minimax game with the following equation:

$$\min_G \max_D V_I(D,G,Q) = V(D,G) - \lambda I(c; G(z,c)) \quad (4)$$

The equation is a modified version of (1) with λ added as a mere factor, to ensure the use of latent code in the network. A mutual information term can be hard to maximize but is achievable by utilizing a lower bound auxiliary distribution *Q(c|x)* known as Variational Information Maximization [15]. The auxiliary distribution Q is implemented as a neural network. D and Q share every convolutional layer with one fully connected one, which is responsible for the parameters of the conditional distribution *Q(c|x)*. If c is of categorical nature, softmax is used to represent Q, while continuous c is dependent on the true posterior.

D. Wasserstein GAN

The Wasserstein GAN (WGAN) by Arjovsky, Chintala and Bottou [16] is an extension of the vanilla GAN with benefits of improved stability and a loss function correlating with G's convergence. WGAN updates D more often for each training iteration than it updates G. WGAN utilizes an approximation of the Earth-Mover (EM) distance or Wasserstein metric [17] to establish above mentioned benefits when compared with vanilla GANs. The EM distance as shown in (5) is the optimal cost of transporting an amount of mass from *x* to *y* to transform $\mathbb{P}_r$ into $\mathbb{P}_g$. Meaning the infimum or cheapest cost for any transport with γ being all calculated plans and $W(\mathbb{P}_r, \mathbb{P}_g)$ being the probability distribution of x and y.

$$W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\gamma \in \Pi(\mathbb{P}_r, \mathbb{P}_g)} \mathbb{E}_{(x,y)\sim\gamma}[\|x - y\|] \quad (5)$$

Therefore, a function f solving the maximization problem of (6) must be found.

$$W(\mathbb{P}_r, \mathbb{P}_\theta) = \sup_{\|f\|_L \leq 1} \mathbb{E}_{x\sim\mathbb{P}_r}[f(x)] - \mathbb{E}_{x\sim\mathbb{P}_\theta}[f(x)] \quad (6)$$

Sup is the least upper bound and symbolized by f is 1-Lipschitz function [18]. To do this, a parameterized neural network with compact space (W) and weights w is trained with backpropagation. This means all functions of w will be K-Lipschitz. This is achieved by fitting the weights into a fixed value range after each gradient update. The Wasserstein distance can be calculated with the found and learned 1-Lipschitz function.

E. Deep Convolutional Generative Adversarial Network

In 2015 Radford, Metz and Chintala [19] introduced DC-GANs. One way to use DCGANs is by learning representations of unlabeled images to use them in supervised learning. Such a feat can be achieved by using parts of D and G for feature extraction in supervised tasks. DCGANs combine the idea of connecting GANs and Convolutional Neural Networks (CNN) by incorporating architectural changes of CNNs. The first change made was the use of all convolutional networks without deterministic spatial functions and allowing the learning of spatial downsampling by using strided convolutions. G profits from spatial downsampling and additionally learns its own discriminator. The second is the elimination of fully connected layers. Instead, DCGANs utilize global average pooling as an trade-off between training stability and model convergence. G has one connected layer which is the input of the noise vector shaped into a 4-dimensional tensor used to build the convolution stack. D's last layer is flattened and fed through a single sigmoid output. The next change pertains batch normalization, helping to prevent some occurring training problems in deeper models due to poor initialization. Furthermore, the addition of batch normalization helped deep generators to learn and prevented collapse on a single point. Every layer is fitted with batch normalization expect the generator's output and the discriminator's input layer. All layers of G use ReLU (Rectified linear unit - a linear function for values greater than zero and non-linear for negative values) as activation function except the output layer which uses Tanh (hyperbolic tangent - is a nonlinear function between $-1, 1$, averaging around zero). For D leaky rectified is used for activation. Figure 4 shows the generator's architecture which takes a 100x1 noise vector (z) and runs it through the generator to map it into a 64x64x3 output $G(Z)$. The input noise is reshaped and expanded into a 4-dimensional tensor to form a vector of size 1024x4x4.

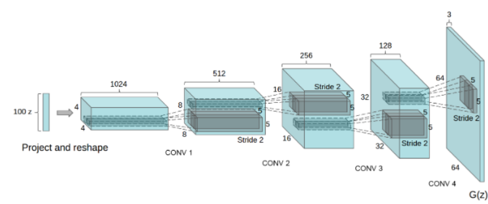

Figure 4. DCGAN Architecture [19]

An extension to DCGAN is called Regularized DCGAN (RDCGAN) by Mehralian and Karasfi [20]. Representation learning is as good as DCGANs ability but prevents mode collapse (see VI-A) with the help of an encoder (E) with the benefit of i) stabilize training and ii) provide more features. Based on DCGAN, RDCGAN is an unsupervised method that can learn good representations of images and by training through adversarial learning the discriminator and encoder can be reused for image classification tasks. DCGAN's structure persists with an added encoder to learn image encodings and feed them to the generator for input reproduction. To show the efficiency of RDCGAN's classification abilities, a classifier is situated at the top of the network. As with vanilla GAN D is trained to maximize $D(x)$ to identify images from the original data, while G is minimizing $D(G(z))$ for getting caught with fake images. While training G, D is frozen and the the encoder is only frozen while G is training to create fake images. The goal of this approach is to generate fake images with noise to fool D and reconstruct encoded images from the input. Therefore, G has two loss functions to enable G not only to generate images but to have the ability to reconstruct images. With l_{con}a reconstruction loss function the distance between real and generated image from E can be minimized and is

 ISBN: 978-1-61208-793-1

depicted in the following equation:

$$\mathrm{l}_{con} = \mathbb{E}_{x \sim P_{data}(x)} \|x - G(E(x))\|_2 \tag{7}$$

To enforce G to reconstruct images which will pass through Ds tests the encoder is defined as follows:

$$\mathrm{l}_{encoder} = \mathbb{E}_{x \sim P_{data}(x)} \left(\log(D(G(x))) + \varepsilon\right) + \mathrm{I}_{con}) \tag{8}$$

By minimizing E, it can extract more useful features for G to reconstruct the image and D will classify them as real images with a higher chance. For G to fool D, G has the following object function trying to minimize log(D(G(z))):

$$\mathrm{l}_{generator} = \mathbb{E}_{z \sim P_z(z)} \left(\log(D(G(z))) + \varepsilon\right) + \mathrm{I}_{encoder}) \tag{9}$$

F. Image-to-Image Translation

In [7], Zhu, Park, Isola and Efros, present an unpaired Image-to-Image translation architecture called CycleGAN. Paired Image-to-Image translation ($\{x_i, y_i\}_{i=1}^N$) means for each x_i there is an corresponding y_i and vice versa. Unpaired translations differ, because there is no information given which x_i matches y_i. Therefore, unpaired approaches have a source $\{x_i\}_{i=1}^N \in X$ and specify a target set for the translation $\{y_i\}_{i=1}^M \in Y$. The idea of cycle consistency is that $G : X \rightarrow Y$ and $F : Y \rightarrow X$ are consistent meaning that the translation of an image from the source X to the target Y and back will yield the starting image from source X again. Forward cycle consistency ($x \rightarrow G(x) \rightarrow F(G(x)) \approx x$) is the ability to translate each image from X back to its original state. The same is true for each image from Y, called backward cycle consistency ($y \rightarrow F(y) \rightarrow G(F(y)) \approx y$). This results in the following equation:

$$\begin{aligned} \mathcal{L}_{cyc}(G, F) = \mathbb{E}_{x \sim p_{data}(x)} [\|F(G(x)) - x\|_1] \\ + \mathbb{E}_{y \sim p_{data}(y)} [\|G(F(y)) - y\|_1] \end{aligned} \tag{10}$$

Two discriminators(D_X and D_Y) are used to differentiate images $\{x\}$ and translated images $\{F(y)\}$ for D_X and $\{y\}$, $\{G(x)\}$ for D_Y. The adversarial loss function for D_Y is:

$$\begin{aligned} \mathcal{L}_{\mathrm{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)} [\log D_Y(y)] \\ + \mathbb{E}_{x \sim p_{data}(x)} [\log(1 - D_Y(G(x))] \end{aligned} \tag{11}$$

The complete equation for the specified problem is as follows:

$$G^*, F^* = \arg \min_{G,F} \max_{D_x, D_Y} \mathcal{L}(G, F, D_X, D_Y) \tag{12}$$

CycleGAN architecture is based on the neural style transfer in [21]. The network structure contains two stride-2 and two $\frac{1}{2}$-strides convolutions, as well as residual blocks. D is using PatchGANs with a patch size 70x70 to classify if these patches are real or were generated by G. To stabilize the training the negative log likelihood in (12) is replaced by least square loss and results in the following equation:

$$\begin{aligned} \mathcal{L}_{\mathrm{LSGAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{data}(y)} \left[(D_Y(y) - 1)^2\right] \\ + \mathbb{E}_{x \sim p_{data}(x)} \left[D_Y(G(x))^2\right] \end{aligned} \tag{13}$$

For the occurring problem of model oscillation both discriminators are updated with a history of previous generated images instead of using the newest generated images.

In contrast, the Image-to-Image translation architecture presented by Isola et al. in [22], is a paired approach. In this case, G must not only fool D but must also achieve near ground truth output. Therefore, L1 distance is chosen:

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z} [\|y - G(x, z)\|_1] \tag{14}$$

This leads to the objective's equation of:

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G) \tag{15}$$

The architecture is based on [19] (see Section III-E)). To prevent some bottlenecks G is U-Net [23] shaped with connections between each layer i and $n - i$ with n being the overall quantity of layers. All channels of i are connected with the channels of $n - i$. Their used D is called PatchGAN also used in the above CycleGAN and tries to classify a patch of size NxN in an image of being real of fake. D runs this patch across the whole image averaging the results of all patches to generate the output of D.

G. StyleGAN

StyleGAN by Karras, Laine and Aila [5] is an enhancement from their previous published ProGAN [24]. StyleGAN proposes a new generator by including style transfer ideas. G starts learn from a continuous input and at each convolution layer, adjusts the style based on latent code (see Section III-C). In combination with noise the network can distinguish and separate attributes and random variations. Instead of starting from the input layer, StyleGAN's architecture starts from a learned constant. The latent code (z) is embedded into a latent space Z. Instead of putting z directly into G's input layer, a non-linear mapping is used to map the input into an intermediate latent space W ($f : \mathcal{Z} \rightarrow \mathcal{W}$)G is controlled by the latent space through adaptive instance normalization at each layer. StyleGANs architecture is shown in Figure 5

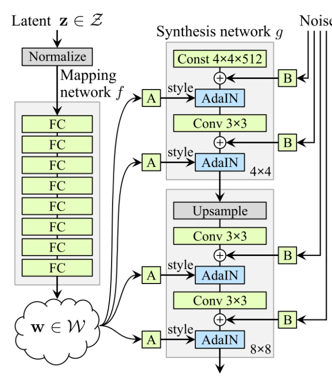

Figure 5. StyleGAN Generator Architecture [5]

and shows a progressively growing G. The network starts with 4x4 images and are trained until stable and then the size is doubled. The network consists of 8 layers while the synthesis network g consists of 18 layers, each resolution has two layers till 1024^2 is reached. The last layer's output is converted into RGB by using a 1x1 convolution. The depicted mapping network takes a random point from latent space and creates a style vector which is transformed and utilized after each convolution. Adaptive instance normalization (AdaIN) standardizes the the output of feature maps to a standard Gaussian and adds the style vector as bias. Additional noise is added for style variations of the generated images. StyleGAN enables image synthesis control with the help of scale-specific modifications for each style. A style is drawn from each learned distribution and based on the collected styles the synthesis network can generate new images. Style mixing is used to use more than one latent space to generate a defined percentage of images during training. While such an image is generated at some random point one latent space is switched for another.

H. StackGAN

StackGAN [25] and their enhancement StackGAN+++ by Zhang et al. [12] is a stacked GAN approach where GANs are stacked in a tree-like shape. StackGAN-v1 is used for text-to-image synthesis. Therefore, a text description is encoded resulting in a text embedding φ_t. This first GAN is generating low resolution images based on φ_t focusing on rough outlines and object colors. Gaussian conditioning variables $\hat{c}_0$ are sampled from a conditioning augmentation to create more training pairs from a Gaussian distribution ($\mathcal{N}\left(\mu_0\left(\varphi_t\right), \Sigma_0\left(\varphi_t\right)\right)$). A random variable z is used to train the generator (G_0) and discriminator (D_0) shown in the following equations:

$$\begin{aligned}\mathcal{L}_{D_0} &= \mathbb{E}_{(I_0,t)\sim p_{data}}\left[\log D_0\left(I_0,\varphi_t\right)\right] \\ &+ \mathbb{E}_{z\sim p_z, t\sim p_{data}}\left[\log\left(1-D_0\left(G_0\left(z,\hat{c}_0\right),\varphi_t\right)\right)\right]\end{aligned} \tag{16}$$

$$\begin{aligned}\mathcal{L}_{G_0} &= \mathbb{E}_{z\sim p_z, t\sim p}\left[-\log D_0\left(G_0\left(z,\hat{c}_0\right),\varphi_t\right)\right] \\ &+ \lambda D_{KL}\left(\mathcal{N}\left(\mu_0\left(\varphi_t\right),\Sigma_0\left(\varphi_t\right)\right)\|\mathcal{N}(0,I)\right)\end{aligned} \tag{17}$$

The goal is to maximize L_{D_0} and minimize L_{G_0}. I_0 is the real image and t the original description form p_{data}, z is the noise vector and λ is used a regulation parameter to balance the terms in (17), set to 1 for the tests conducted in the paper. A pre-trained character level CNN-RNN text encoder is used to map text descriptions to the feature space of images learning a correspondence between images and descriptions. First φ_t is fed into a fully connected layer to generate the necessary parts for the shown Gaussian distribution to sample $\hat{c}_0$. The discriminator compresses the text embedding into N_d dimensions with a fully connected layer and then used to form a $M_d \times M_d \times N_d$ tensor. The image gets downsampled until it reaches $M_d \times M_d$ dimensions and is linked with the text tensor. The result is forwarded to a 1×1 convolutional layer to learn features of the text and image. The decision score is made by single node fully connected layer.

Stage-II GAN is built on top of Stage-I GAN and tries to generate high-resolution images. It is conditioned on images with low resolutions and as well as text embedding to correct previous mistakes made by Stage-I GAN. The low-resolution images of G_0, resulting in $s_0 = G_0\left(z,\hat{c}_0\right)$, and the Gaussian latent variables $\hat{c}$ are used to train D and G of the Stage-II GAN by maximizing D and minimizing G as shown in the following equations:

$$\begin{aligned}\mathcal{L}_{D} &= \mathbb{E}_{(I,t)\sim p_{data}}\left[\log D\left(I,\varphi_t\right)\right] \\ &+ \mathbb{E}_{s_0\sim p_{G_0}, t\sim p_{data}}\left[\log\left(1-D\left(G\left(s_0,\hat{c}\right),\varphi_t\right)\right)\right]\end{aligned} \tag{18}$$

$$\begin{aligned}\mathcal{L}_{G} &= \mathbb{E}_{s_0\sim p_{G_0}, t\sim p_{data}}\left[-\log D\left(G\left(s_0,\hat{c}\right),\varphi_t\right)\right] \\ &+ \lambda D_{KL}\left(\mathcal{N}\left(\mu\left(\varphi_t\right),\Sigma\left(\varphi_t\right)\right)\|\mathcal{N}(0,I)\right)\end{aligned} \tag{19}$$

During this stage the Stage-II GAN is not fed with additional random noise, as the occurrence of randomness should have already happened and been stored in s_0. Gaussian variables from both stages have the same pre-trained encoder and generate the same φ_t. The main difference are different fully connected layers for condition augmentation to generate differing deviations and means. Stage-II GAN is able to learn new information which were previously omitted by Stage-I GAN. G of the Stage-II GAN is constructed as an encoder-decoder network using residual blocks. As with Stage-I generator φ_t is used to generate $\hat{c}$. The results s_0 of the first GAN are is downsampled until it reaches a size of $M_g \times M_g$. All encoded features are fed into residual blocks, able to learn multi-modal representations. In the last step a decoder is used to generate a high-resolution image. The discriminator receives an extra downsampling block to handle the larger image size. For training, D works with positive samples of real images and their textual descriptions and negative samples consisting of real images and mismatched textual descriptions and generated images with their textual description.

Upsampling is done by nearest-neighbor block followed by a stride-1 convolution with 3×3. Each layer except the final layer apply batch normalization and ReLU activation function. Each residual block consists of stride-1 convolution with 3×3, batch normalization and ReLU activation. Two of these blocks are used in Stack-I GAN with a 128×128 resolution and is doubled for 256×256 models. Downsampling layers, with the exception of the first block, have Batch normalization, a LeakyReLU activation and stride-2 convolution with 4×4.

I. Copy-Pasting GAN

Arandjelovic and Zisserman introduced a Copy-Pasting GAN in [10]. The approach is about unsupervised object discovery, where G learns how to discover objects and pasting it into another so that D will be fooled. The main difference with other GAN and object discovery methods is G, which does not generate objects and its solely responsibility is to detect and segment existing objects. Instead of creating a whole new image G combines two images by selecting and copying a section of the source image to the destination image. Therefore, the generator is restricted. A convolutional network takes the source image (I_s) to generate a segmentation mask. A new image is generated by copying and blending (I_s) into the new image (I_d) using the copy mask ($m_\theta(I_s)$) with values of m are in $[0,1]$, θ symbolizing trainable parameters and $\odot$ being the element-wise product, resulting in the following equation:

$$\begin{aligned}I_c &= C\left(I_s, I_d, m_\theta\left(I_s\right)\right) \\ &= m_\theta\left(I_s\right)\odot I_s + \left(1-m_\theta\left(I_s\right)\right)\odot I_d\end{aligned} \tag{20}$$

The same masked is used across all channels. This applies a limit to the copying and pasting operation of G by forcing it to paste the object at the same location without any transformation. G is implemented as a U-Net with a 1-channel output, resulting in ($m_0(I_s)$). D is also a U-Net with an average-pooled encoding in the middle and passed on to fully connected layer for classification.

IV. GAN Application Domains

The field of application for GANs is constantly growing and finds its way into various domains. The following section, will list additional GAN architectures extending the concept of the architectures presented in Section III and list them based on their domain together with a short summary and if applicable the architectural type from Section III. This section extends the findings of RQ1 from the previous section and introduces

 ISBN: 978-1-61208-793-1

exemplary domains to show the variety of GANs. Furthermore, the domains in this section answers RQ2.

A. Autonomous Driving

Autonomous driving is complex domain with various tasks and complex image structures depicting various objects of different types and distances. There are several different use cases for autonomous driving where GANs can help to depict situations that are highly unlikely and difficult to capture (e.g., pedestrian crossing the street relatively close to the vehicle or even getting by it). In [26] Choi, Jeong, Park and Ha, propose an image generation GAN based on DCGAN (see Section III-E) to create new situational images with the help of feature maps. This approach extracts a feature map (e.g., of pedestrians) and places them in another scene. A more specialized approach is presented in [27] placing pedestrians with different poses into preexisting scenes by using an encoder to extract a person from a scene and given to mask estimation network deciding which pixel should be taken and insert into the original image to place pedestrians. In [28], GANs are used for augmentation in the domain of autonomous driving vehicle. First a basic introduction about GANs is given. Afterwards, the paper introduces a CycleGAN (see III-F) experiment is conducted where the condition of front mounted camera is changed from soiled to clean and vice versa. In [29] the authors present DeepRoad a validation framework for autonomous driving systems. The framework can generate new images with various weather conditions based on real-world weather scenes. Using synthetic images the consistency of the system can be tested, and validates images for DNNs with their VGGNet features by measuring the distance between input and training images and is closely related to CycleGAN (see III-F).

B. Archaeology

In [11] Hermoza nad Sipiran introduce ORGAN a 3D reconstruction GAN to restore archaeological objects. ORGAN is based on an encoder-decoder 3D DNN on top of a GAN based on cGANS (Section III-B). With two loss functions (completion loss and Wasserstein loss) the network is trained to predict the missing parts of a damaged object. To compensate differences between archaeological objects, the cGAN is conditioned on variable containing information about the culture, region or the object itself. ORGAN is able to reconstruct objects with nearly 50% missing. Another approach is the 3D reconstruction of objects from single photographs by Kniaz, Remondino and Knyaz [30]. They use a Z-GAN (based on pix2pix, Section III-F) for image-to-voxel translation. The generator's encoder is unchanged while the 2D kernels are changed into 3D convolutional kernels. Like a U-Net, Z-GAN has skip connections (in accordance with U-Net definition) helping to transfer high-frequency components of the input to the 3D shape. The generated reconstructions are reviewed and compared with domain experts and shows the strength of the approach.

C. Video Editing

In [9], Zhang, Zhao, Kampffmeyer and Tan, propose DTR-GAN a dilated temporal relational GAN for video summarization. DTR-GAN allows a frame-level summarization without loosing viable information within the video. Therefore, a DTR module is constructed capturing relations between neighbouring multi-range frames. One DTR layer consists of four DTR units of different size to capture the relations between frames. G tries to learn to identify key frames. A Temporal Encoding Module integrates Bi-LSTM (Long Short Term Memory) layer as well as a three layer DTR network. A confidence score is predicted for key frame capturing confidence. D learns the relations between original video and summary using a three-player loss to ensure that D can recognize and learn the difference between valid summaries and trivial ones. Mocycle-GAN [8] proposed by Chen, Pan, Yao, Tian and Mei, is an unpaired video-to-video translation process comparable with image-to-image translations (Section III-F). Mocycle-GAN transfers videos from a source domain to a target domain with two generators G_X and G_Y for cross domain frame synthesis, two discriminators (D_X and D_Y) to distinguish between real and synthetic frames and a motion translator (M_X) for cross domain motion translation. Real frames are first translated into synthetic frames with G_X and further transformed via inverse mapping of G_Y into reconstructed frames. Furthermore, a FlowNet [31] is used to obtain optical flows for motion representation before and after forward cycles. While training Mocycle-GAN utilizes three temporal constraints for structural appearance and temporal continuity exploration. Adversarial learning using adversarial constraints ensures realistic synthetic frames, frame and motion cycle consistent constraint is used for inverse translations for frames and motions while the third motion translation constraint validates cross domain motions transfers. DCVGAN [32] propose an advanced video generating GAN using optical information as well as 3D geometrical information with depth video. The generator is based on MoCoGAN [33] and splits the latent space into content and motion. A RNN models and generates video dynamics and motion latent vectors, while a CNN is responsible for depth image generation from latent vectors. Colour to depth translation is achieved by using a pix2pix(Section III-F) based architecture. Using U-Net structures the colour image generator is based on an encoder-decoder architecture. Two discriminators one for images, using randomly selected colour and depth images to test if it is an original sample or generated, and one for videos taking the same input as video and evaluating its temporal features. With the addition of depth images DCVGAN can outperform MoCoGAN.

D. Medicine

In [34] the authors propose a GAN for treatment recommendations based on patient-centric literature retrieval. The goal of the approach is to measure the similarity of gene mutations. Therefore, the input is encoded as two-hot vectors with the number of genes used as vector dimensions. G outputs one entire batch per training step while D receives one sample and the batch average. Furthermore, D is parallel trained with an online training scheme receiving only a single sample and a batch average of zero with a training mode indication. A metric is used to rank patient document pairs, by looking for genes mentioned in both records (patient and documents). To gain treatment adequacy the authors separate D and G and condition both on the patient information vector. G is trained as a feed-forward network to predict patient treatments and D is used to distinguish between real and synthetic vectors. Scores are calculated for D and G by taking the conditioned output of D and G and comparing it with the treatment vector by applying cosine similarity. Both scores are fused to obtain the final document score. In [35], Xie, Xu and Li, propose a GAN for CT reconstruction of images taken from limited angles with artefacts. The proposed GAN is based on cGAN (Section III-B) and uses Wasserstein distance (Section III-D) for training stability. The proposed GAN uses perceptual loss as well as adversarial loss as loss functions. The proposed approach shows that it can remove artefacts while retaining the CT's texture details.

E. Additional GAN Domains

The applicability of GANs is not limited to the four previous shown domains. In [36], Wen, Singh and Raj, proposes a framework to reconstruct faces based on their voices by matching identities between generated faces and speakers. Engel, Agrawal, Chen, Gulrajani, Donahue and Roberts [37] generated high quality audio samples using GANs with latent and pitch vectors with global conditioning instead of WaveNet and rapidly sped up the process. The importance of GANs for augmentation are presented and discussed in [38]. In [6] an actor-critic conditional GAN is presented. The GAN can learn surrounding context from text to fill in blanks. Parts of the text are deleted and the model tries to fill the missing pieces to look like the original. ChinaStyle[39] introduces a new dataset set with six categories and 1913 total images about Chinese traditional figure painting. Furthermore, the authors introduce MA-GAN a style transfer GAN for image translation of portraits into Chinese paintings. A more exotic domain can be found in [40]. RamenGAN is a cGAN using an additional discriminator to create ramen with rounder dishes. The other proposed RecipeGAN a WassersteinGAN is used to receive ingredients of a dish via image search.

V. GAN Evaluation Methods

This section, will list and highlight common datasets used to evaluate new approaches and will further categorize all presented GANs from Section III. Furthermore, this section, will show and introduce commonly used metrics to evaluate the performance of GANs. All models are generative ones but differ in their approaches and thus use different datasets for evaluation. Categorizing by datasets is one possible way to filter GANs. Furthermore, this section, will present GAN evaluation measures.

A. Datasets

Table I summarizes datasets used by well known GAN architectures for image generation. The next architectural style are Style-Transfer GANs as summarised in table II which maps the properties from one domain to another. For example, this includes the transfer from day to night or hand drawing to real image.

TABLE II. Dataset overview Style-Transfer GANs

Architecture	Cityscape	CMP Facades	ImageNet
CycleGAN	X	X	X
Pix2Pix	X	X	X

The final Table III summarizes specialized architectures for text-to-image processing and image manipulation.
The variance of used datasets is quite vast and limits model comparison immensely. As shown above based solely on literature, a comparison between some models is possible, but cannot be guaranteed. For this reason, Lucic, Kurach, Michalski, Bousquet and Gelly [13] propose a simple dataset for model evaluation. The above shown and currently used datasets are either too simple or complex to achieve meaningful evaluation results. Therefore, a data manifold is created which allows efficient computation between sample distance to manifold. The problem is transformed into a computational problem, where the precision is higher the closer the samples from model distribution are to the manifold. High recall in this case means that G is able to generate samples relatively close to the manifold. The proposed manifold is of convex polygons. For single-channel grey images ($x \in \mathbb{R}^{d^2}$) the sample distance to the manifold is the squared Euclidean distance from $\hat{x} \in \mathbb{R}^{d^2}$ to C_3, the manifold's closest sample, as shown in the following equation:

$$\min_{x \in \mathcal{C}_3} \ell(x, \hat{x}) = \sum_{i=1}^{d^2} \|x_i - \hat{x}_i\|_2^2 \tag{21}$$

The approximate of a solution is found by gradient descent on the convex polygon's vertices, with each iteration being a valid convex polygon. Using random initial solutions the algorithm is executed 5 times to minimize false-negative samples. The findings of RQ3 are summarized in the tables of this section and give an overview of the different GAN architectures and their datasets utilized for evaluation.

B. GAN Evaluation

This subsection, will highlight the most important metrics and measurements used to evaluate GANs. The results are based on the conducted literature review performed to answer RQ4. The variety of datasets also applies to metrics. There are numerous evaluation metrics with a few stand out measurements. There is a difference between evaluating the generated images produced by a GAN and the model itself. In [41] Borji presents and discusses the pros and cons of evaluation measurements for GANs. The author creates a list with important characteristics an evaluation method should fulfil and tests evaluation methods against these characteristics. Any method used to evaluate GANs should favour high fidelity samples with high diversity which have disentangled latent spaces, well-defined boundaries, are sensitive to transformations and distortions, can withstand human judgment and have low complexity.

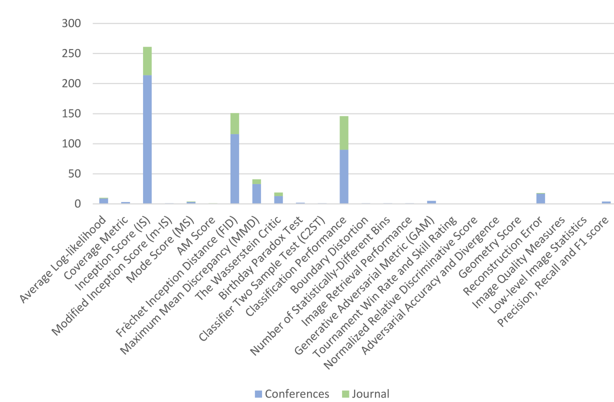

Figure 6. GAN evaluation metrics overview [41] and literature excerpt

24 quantitative measurements and 5 qualitative measurements are listed. The following Figure shows the quantitative metrics collected by [41] and shows a literature excerpt to give an indication of the metrics acceptance within the scientific community. Full text as well as abstract searches were

ISBN: 978-1-61208-793-1

TABLE I. DATASET OVERVIEW OF COMMON GAN ARCHITECTURES

Architecture	MNIST	CIFAR-10	TFD	SVHN	FFHQ	ImageNet	MIR Flickr 25,000	YFCC100M 2	CelebA	LSUN
GAN	X	X	X							
cGAN	X					X	X	X		
InfoGAN	X			X					X	
WassersteinGAN										X
DCGAN	X	X		X		X			X	X
StyleGAN					X					X

TABLE III. DATASET OVERVIEW STACKGAN AND CP-GAN

Architecture	CIFAR-10	CLEVR	Flying Chairs	CUB	MS COCO	Oxford-102	LSUN	ImageNet
StackGAN	X	X	X					X
CP-GAN				X	X	X	X	

conducted for IEEE Xplore and ACM Digital Library with the concatenated search term:*"Full Text Only":"generative adversarial networ*" or "gan" and "metric name" and "Abstract":"generative adversarial networ*"*. Figure 6 shows the most used metrics for GAN evaluation namely Inception Score, Fréchet Inception Distance, Classification Performance and the Maximum Mean Discrepancy, summarized in the next sections.

C. Inception Score (IS)

The Incetion Score [42] applies the Inception Model (a pretrained model) to every generated image to receive $p(y|x)$ (conditional label distribution), ideally having a low entropy, meaning that the generated images have meaningful objects. The marginal $\int p(y|x = G(z))dz$ will have a high entropy if the model generated images of high variety. This results in the following metric:

$$exp(\mathbb{E}_x \mathrm{KL}(p(y|x) \| p(y))) \tag{22}$$

D. Fréchet Inception Distance (FID)

The Fréchet Inception Distance [43] is an improvement of the above described Inception Distance. The coding layer of the Inception Model is used to replace x by generalizing polynomials x (mean and covariance) to obtain vision related features. The coding units follow a multidimensional Gaussian. The distance between the two Gaussian (real-world and synthetic samples) the Fréchet distance (Wasserstein-2 distance) to assess the quality of generated samples given by the following equation:

$$\begin{aligned} d^2((m, C, (m_w, C_w))) = {} & |||m - m_w\|_2^2 \\ & + Tr\left(C + C_w - 2\left(CC_w\right)^{1/2}\right) \end{aligned} \tag{23}$$

E. Classification Performance

Classification performance summarized in [41] is to apply unsupervised representation learning algorithms as feature extractors for labeled datasets and then evaluate their performance.

F. Maximum Mean Discrepancy (MMD)

The Maximum Mean Discrepancy [44][45] is a test to determine if the distribution of p and q are different based on drawn samples from each of them. This is done by finding a smooth function with the properties of being large if the points were drawn from p and small for samples drawn from q. The difference of values from a mean function between two samples is called the MMD. The greater the distance the more likely it is for the taken samples to come from different distributions. The goal of MMD is to answer if $p \neq q$ when p and q are distributions from x and $X := \{x_1, \ldots, x_m\}$, $Y := \{y_1, \ldots, y_n\}$ are distributed from p and q independently and identically.

VI. GAN Challenges

Application domains as well as specialized GAN architectures are flourishing and the potential and advances seem promising to not only achieve results of higher quality with better performance but also to stabilize the training process and prevent non-convergence or mode collapse. This section, will describe two occurring challenges in accordance with RQ5 for GANs as well as recent advances and new insights.

A. Mode Collapse and Non-Convergence

GAN networks are still prone to some problems like mode collapse or non-convergence. Mode collapse [46][47] is lack of diversity in generated samples. The problem lies in the minimax game where G tries to fool D. If G finds a sweet-spot (i.e., concentrate on a single mode), G is more likely to abuse this sweet-spot to produce a more plausible output. D is able to learn the pattern of this sweet-spot and can always classify it to be fake. The next iteration of D can get stuck and G can abuse this by creating the best output for the current D.

Non-convergence [46][48] is D's problem to distinguish between real and fake samples. With G improving, D's performance will get worse. With a near perfect G, D is forced to randomize its output (coin flip) and the feedback becomes negligible. The problem is if G is trained past this point and adjust to given feedback, the quality of generated samples will drop. Depending on the architecture and dataset some GANs are more prone to non-convergence as others.

B. Advances in Research

Improved architectures enabling higher resolutions, stabilizing training preventing mode collapse or non-convergences will further strengthen the applicability of GANs. This subsection will introduce some advances in research compliant with RQ6. In [49] Chavdarova, Gidel, Fleuret and Lacoste-Julien, show that training a GAN with stochastic gradient noise can prevent the convergence of standard game optimization methods. Quality aware GANs are proposed in [50]. The authors use a variation of structural similarity (SSIM) index and a quality aware discriminator gradient penalty function as regularizers for GANs. Learning disentangled representations unsupervised is proposed in [51]. By introducing a similarity constraint the authors show, that their approach is able to distinguish different representations by using conditions. In [52] a super resolution reconstruction method is introduced.

 ISBN: 978-1-61208-793-1

The combination of GANs and wavelet transformation used to train wavelet decomposition coefficients improves the quality of reconstructed images. Adiga, Attia, Chang and Tandon [53] propose two performance metrics mode-collapse divergence (MCD) and Generative Quality Score (GQS) to measure the quality of generated samples, created to capture the impact of mode collapse. Yang, Li, Qi and Lyu [54] introduce a method to predict images synthesized by GANs. The authors identified a difference between the location of facial landmarks due to lacking global constraints. Another approach by McCloskey and Albright [55] show that normalization steps by G leads to a detectable suppression of image characteristics.

VII. Conclusion

The presented survey summarizes the evolution of GANs by providing an overview of conducted research of the last 5 years, showing different architectural designs and loss functions. GANs will continue to evolve and further reshape deep generative models to produce more realistic samples. Image and video augmentation, restoration or 3D modelling are only a few domains in which GANs are already helping to produce new samples. The research contributing to GANs will find approaches to further stabilize training, minimize mode collapse and non-convergence. The varying architectural patterns use different datasets and metrics for evaluation. There is still a need to find a uniform accepted evaluation method consisting of datasets and metric. As technology advances, so will GANs and their possibilities. New generator and discriminator architectures will help to achieve the ambitious goal of generating realistic samples. This will be one of the biggest challenges as shown in Section VI-B as synthesized samples are still distinguishable from real samples. There are various domains (e.g., autonomous driving vehicles) with occurrences of rare high impact events, that are hard to capture (e.g., pedestrian crossing the street in close proximity to the vehicle). Combining the various fields of open challenges will help generating indistinguishable fake samples of such rare events, required for the evaluation of existing machine learning models to ensure they are able to recognize these rare events.

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in Advances in Neural Information Processing Systems 27, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2672–2680, [Online] http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf [retrieved: 05-2020].

[2] M. Brückner and T. Scheffer, “Nash equilibria of static prediction games,” in Advances in Neural Information Processing Systems 22, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, Eds. Curran Associates, Inc., 2009, pp. 171–179, [Online] http://papers.nips.cc/paper/3755-nash-equilibria-of-static-prediction-games.pdf [retrieved: 05-2020].

[3] X. Chen and X. Deng, “Settling the complexity of two-player nash equilibrium,” in 2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS’06), Oct 2006, pp. 261–272.

[4] M. Mirza and S. Osindero, “Conditional Generative Adversarial Nets,” nov 2014, [Online] http://arxiv.org/abs/1411.1784 [retrieved: 05-2020].

[5] T. Karras, S. Laine, and T. Aila, “A style-based generator architecture for generative adversarial networks,” in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2019-June. IEEE Computer Society, jun 2019, pp. 4396–4405.

[6] W. Fedus, I. Goodfellow, A. M. Dai, and G. Brain, “MaskGAN: Better Text Generation via Filling in the,” Tech. Rep., feb 2018.

[7] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks,” in Proceedings of the IEEE International Conference on Computer Vision, vol. 2017-Octob. Institute of Electrical and Electronics Engineers Inc., dec 2017, pp. 2242–2251.

[8] Y. Chen, Y. Pan, T. Yao, X. Tian, and T. Mei, “Mocycle-GAN: Unpaired video-to-video translation,” in MM 2019 - Proceedings of the 27th ACM International Conference on Multimedia. New York, NY, USA: Association for Computing Machinery, Inc, oct 2019, pp. 647–655, [Online] http://dl.acm.org/doi/10.1145/3343031.3350937 [retrieved: 05-2020].

[9] Y. Zhang, X. Zhao, M. Kampffmeyer, and M. Tan, “DTR-GAN: Dilated temporal relational adversarial network for video summarization,” in ACM International Conference Proceeding Series. New York, New York, USA: Association for Computing Machinery, may 2019, pp. 1–6, [Online] http://dl.acm.org/citation.cfm?doid=3321408.3322622 [retrieved: 05-2020].

[10] R. Arandjelović and A. Zisserman, “Object Discovery with a Copy-Pasting GAN,” may 2019, [Online] http://arxiv.org/abs/1905.11369 [retrieved: 05-2020].

[11] R. Hermoza and I. Sipiran, “3D reconstruction of incomplete archaeological objects using a generative adversarial network,” in ACM International Conference Proceeding Series. New York, New York, USA: Association for Computing Machinery, jun 2018, pp. 5–11. [Online]. Available: http://dl.acm.org/citation.cfm?doid=3208159.3208173

[12] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, “StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 8, aug 2019, pp. 1947–1962.

[13] M. Lucic, K. Kurach, M. Michalski, O. Bousquet, and S. Gelly, “Are Gans created equal? A large-scale study,” in Advances in Neural Information Processing Systems, vol. 2018-Decem, 2018, pp. 700–709.

[14] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, “InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets,” in Advances in Neural Information Processing Systems, 2016, pp. 2180–2188, [Online] https://arxiv.org/abs/1606.03657. [retrieved: 05-2020].

[15] S. Mohamed and D. Jimenez Rezende, “Variational information maximisation for intrinsically motivated reinforcement learning,” in Advances in Neural Information Processing Systems 28, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Curran Associates, Inc., 2015, pp. 2125–2133, [Online] http://papers.nips.cc/paper/5668-variational-information-maximisation-for-intrinsically-motivated-reinforcement-learning.pdf [retrieved: 05-2020].

[16] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein GAN,” jan 2017, [Online] http://arxiv.org/abs/1701.07875 [retrieved: 05-2020].

[17] J. Gao and H. Tembine, “Distributionally Robust Games: Wasserstein Metric,” in Proceedings of the International Joint Conference on Neural Networks, vol. 2018-July. Institute of Electrical and Electronics Engineers Inc., oct 2018.

[18] J. Heinonen, “Lectures on Analysis on Metric Spaces.” Springer, New York, NY, 2001, ch. 6 - Lipschitz Functions, pp. 43–48.

[19] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” in 4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings, Y. Bengio and Y. LeCun, Eds., 2016, [Online] http://arxiv.org/abs/1511.06434 [retrieved: 05-2020].

[20] M. Mehralian and B. Karasfi, “RDCGAN: Unsupervised representation learning with regularized deep convolutional generative adversarial networks,” in 2018 9th Conference on Artificial Intelligence and Robotics and 2nd Asia-Pacific International Symposium, AIAR 2018. Institute of Electrical and Electronics Engineers Inc., dec 2018, pp. 31–38.

[21] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in Computer Vision – ECCV 2016, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham: Springer International Publishing, 2016, pp. 694–711.

[22] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5967–5976.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[24] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in International Conference on Learning Representations, 2018, [Online] https://openreview.net/forum?id=Hk99zCeAb [retrieved: 05-2020].

[25] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. Metaxas, "StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks," in Proceedings of the IEEE International Conference on Computer Vision, vol. 2017-Octob. Institute of Electrical and Electronics Engineers Inc., dec 2017, pp. 5908–5916.

[26] S. Y. Choi, H. J. Jeong, K. S. Park, and Y. G. Ha, "Efficient Driving Scene Image Creation Using Deep Neural Network," in 2019 IEEE International Conference on Big Data and Smart Computing, BigComp 2019 - Proceedings. IEEE, feb 2019, pp. 1–4, [Online] https://ieeexplore.ieee.org/document/8679269/ [retrieved: 05-2020].

[27] A. Vobecký, M. Uřičář, D. Hurych, and R. Škoviera, "Advanced Pedestrian Dataset Augmentation for Autonomous Driving," Tech. Rep., 2019.

[28] M. Uřičář, P. Křížek, D. Hurych, I. Sobh, S. Yogamani, and P. Denny, "Yes, we GAN: Applying adversarial techniques for autonomous driving," Electronic Imaging, no. 15, feb 2019, pp. 48–1–48–17, [Online] http://arxiv.org/abs/1902.03442 [retrieved: 05-2020].

[29] M. Zhang, Y. Zhang, L. Zhang, C. Liu, and S. Khurshid, "Deeproad: GaN-based metamorphic testing and input validation framework for autonomous driving systems," in ASE 2018 - Proceedings of the 33rd ACM/IEEE International Conference on Automated Software Engineering. Association for Computing Machinery, Inc, sep 2018, pp. 132–142.

[30] V. V. Kniaz, F. Remondino, and V. A. Knyaz, "Generative Adversarial Networks for Single Photo 3D Reconstruction," in ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 42, no. 2/W9. Copernicus GmbH, jan 2019, pp. 403–408, [Online] https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XLII-2-W9/403/2019/ [retrieved: 05-2020].

[31] A. Dosovitskiy, P. Fischer, E. Ilg, P. Häusser, C. Hazirbas, V. Golkov, P. v. d. Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," in 2015 IEEE International Conference on Computer Vision (ICCV), Dec 2015, pp. 2758–2766.

[32] Y. Nakahira and K. Kawamoto, "DCVGAN: Depth Conditional Video Generation," in Proceedings - International Conference on Image Processing, ICIP, vol. 2019-Septe. IEEE Computer Society, sep 2019, pp. 749–753.

[33] S. Tulyakov, M. Liu, X. Yang, and J. Kautz, "Mocogan: Decomposing motion and content for video generation," in 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 2018, pp. 1526–1535.

[34] L. von Werra, M. Schöngens, E. D. Gamsiz Uzun, and C. Eickhoff, "Generative adversarial networks in precision oncology," in Proceedings of the 2019 ACM SIGIR International Conference on Theory of Information Retrieval, ser. ICTIR '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 145–148, [Online] https://doi.org/10.1145/3341981.3344238 [retrieved: 05-2020].

[35] S. Xie, H. Xu, and H. Li, "Artifact removal using gan network for limited-angle ct reconstruction," in 2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA), Nov 2019, pp. 1–4.

[36] Y. Wen, R. Singh, and B. Raj, "Face Reconstruction from Voice using Generative Adversarial Networks," Tech. Rep., 2019, [Online] https://github.com/cmu-mlsp/reconstructing_faces_from_voices [retrieved: 05-2020].

[37] J. Engel, K. K. Agrawal, S. Chen, I. Gulrajani, C. Donahue, and A. Roberts, "GANSynth: Adversarial neural audio synthesis," in International Conference on Learning Representations, 2019, [Online] https://openreview.net/forum?id=H1xQVn09FX [retrieved: 05-2020].

[38] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," Journal of Big Data, vol. 6, no. 1, dec 2019.

[39] Y. Wang, W. Zhang, and P. Chen, "Chinastyle: A mask-aware generative adversarial network for Chinese traditional image translation," in SIGGRAPH Asia 2019 Technical Briefs, SA 2019. New York, New York, USA: Association for Computing Machinery, Inc, nov, pp. 5–8, [Online] http://dl.acm.org/citation.cfm?doid=3355088.3365148 [retrieved: 05-2020].

[40] Y. Ito, W. Shimoda, and K. Yanai, "Food image generation using a large amount of food images with conditional gan: Ramengan and recipegan." New York, NY, USA: Association for Computing Machinery, 2018, [Online] https://doi.org/10.1145/3230519.3230598 [retrieved: 05-2020].

[41] A. Borji, "Pros and cons of GAN evaluation measures," Computer Vision and Image Understanding, vol. 179, feb 2019, pp. 41–65.

[42] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in Advances in Neural Information Processing Systems, 2016, pp. 2234–2242.

[43] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," Tech. Rep., 2017.

[44] A. Gretton, K. M. Borgwardt, M. Rasch, B. Schölkopf, and A. J. Smola, "A Kernel Method for the Two-Sample-Problem," Tech. Rep., [Online] www.kyb.mpg.de/bs/people/arthur/mmd.htm [retrieved: 05-2020].

[45] I. Tolstikhin, B. K. Sriperumbudur, and B. Schölkopf, "Minimax estimation of maximum mean discrepancy with radial kernels," in Advances in Neural Information Processing Systems, 2016, pp. 1938–1946.

[46] Z. Zhang, M. Li, and J. Yu, "On the convergence and mode collapse of GAN," in SIGGRAPH Asia 2018 Technical Briefs, SA 2018. New York, New York, USA: ACM Press, dec, pp. 1–4, [Online] http://dl.acm.org/citation.cfm?doid=3283254.3283282 [retrieved: 05-2020].

[47] "Common Problems - Generative Adversarial Networks - Google Developers," [Online] https://developers.google.com/machine-learning/gan/problems [retrieved: 05-2020].

[48] "GAN Training - Generative Adversarial Networks - Google Developers," [Online] https://developers.google.com/machine-learning/gan/training [retrieved: 05-2020].

[49] T. Chavdarova, G. Gidel, F. Fleuret, and S. Lacoste-Julien, "Reducing noise in gan training with variance reduced extragradient," in Advances in Neural Information Processing Systems 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 393–403, [Online] http://papers.nips.cc/paper/8331-reducing-noise-in-gan-training-with-variance-reduced-extragradient.pdf [retrieved: 05-2020].

[50] K. Parimala and S. Channappayya, "Quality aware generative adversarial networks," in Advances in Neural Information Processing Systems 32, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Curran Associates, Inc., 2019, pp. 2948–2958, [Online] http://papers.nips.cc/paper/8560-quality-aware-generative-adversarial-networks.pdf [retrieved: 05-2020].

[51] X. Li, L. Chen, L. Wang, P. Wu, and W. Tong, "SCGAN: Disentangled Representation Learning by Adding Similarity Constraint on Generative Adversarial Nets," IEEE Access, vol. 7, 2019, pp. 147 928–147 938.

[52] Z. Huang and C. Jing, "Super-resolution reconstruction method of remote sensing image based on multi-feature fusion," IEEE Access, vol. 8, 2020, pp. 18 764–18 771.

[53] S. Adiga, M. A. Attia, W. T. Chang, and R. Tandon, "ON the tradeoff between mode collapse and sample quality in generative adversarial networks," in 2018 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2018 - Proceedings. Institute of Electrical and Electronics Engineers Inc., feb 2019, pp. 1184–1188.

[54] X. Yang, Y. Li, H. Qi, and S. Lyu, "Exposing GAN-synthesized Faces Using Landmark Locations," in Proceedings of the ACM Workshop on Information Hiding and Multimedia Security - IH&MMSec'19. New York, New York, USA: ACM Press, jul 2019, pp. 113–118, [Online] http://dl.acm.org/citation.cfm?doid=3335203.3335724 [retrieved: 05-2020].

[55] S. McCloskey and M. Albright, "Detecting GAN-Generated Imagery Using Saturation Cues," in Proceedings - International Conference on Image Processing, ICIP, vol. 2019-September. IEEE Computer Society, sep 2019, pp. 4584–4588.

Poster: Studying The Social Networks in Educational Forums

Mary Luz Mouronte-López

Higher Polytechnical School
Universidad Francisco de Vitoria
Madrid, Spain
Email: maryluz.mouronte@ufv.es

***Abstract*—This research aims to carry out a topological study of social networks located in university forums of the Moodle platform. Graphs are built visualizing the structure of the nodes and links and calculating statistical parameters such as: degree, betweenness centrality, clustering coefficient, PageRank, EigenVector centrality and assortativity. The communities' structure was also estimated. This study analyzes how students and faculties work and socialize in the educational environment, which help to know more precisely the level of involvement of each student as well as to improve some learning and methodological aspects. Several subjects and forums are analyzed (theoretical and practical contents).**

Keywords–social networks; graph; topological study

I. Introduction

Virtual Educational Platforms help to increase the motivation of the students in the courses, providing mechanisms to facilitate the interaction between students and teachers. These platforms include applications such as: content management, collaborative communication and monitoring and evaluation tools, administration utilities, as well as other functionalities such as notebook, content searches, etc. Several research exists about tools used for online teaching blogs, discussion boards, wikis and 3D virtual [1][2]. Studies on Learning Management Systems (LMS) have also been carried out [3][4]. There also are investigations that pretend to gain some insight into how teachers and students use the LMS. The interaction between students in online discussion Forums has been studied in detail using Social Network Analysis [5] and other methods [6][7]. This paper analizes the social interactions that happened in Moodle, when this platform was used in the context of a university course. Several topological parameters and the structure of communities were calculated. 14 forums each with an average of 115 students were studied. Three types of forums were considered: news and questions forums, practical exercise forums and theoretical content forums.

II. Analyzing the social networks

The XML file of the Moodle forums was analyzed and processed using software programs implemented in Python. These programs were designed, built and tested, following the typical life cycle of any software component. The interactions in each forum were represented in a graph G = (V; S), where V is the set of nodes corresponding to students and faculties and S is the set of links between them. The package networkx was used in order to analyze the structure of the social network. The following parameters [8][9], and their statistical distributions were calculated:

A. Betweenness centrality

The betweenness centrality of a node n can be defined as:

$$b(n) = \sum_{u \neq n \neq w} \frac{\sigma_{uw}(n)}{\sigma_{uw}} \tag{1}$$

Where
σ_{uw} is the total number of shortest paths from node u to node w
$\sigma_{uw}(n)$ is the number of those paths that pass through n

B. Node clustering coefficient

The clustering coefficient C(n) of a node n can be defined as:

$$C(n) = \frac{2 * t(n)}{d(n) * (d(n) - 1)} \tag{2}$$

Where
t(n) is the number of triangles containing n.
d(n) is the degree of n

C. EigenVector centrality

The EigenVector centrality of a node n can be defined as

$$x_n = \frac{1}{\lambda} \sum_{j=1}^{N} x_j = \frac{1}{\lambda} \sum_{j=1}^{N} Aij * xj \tag{3}$$

Where
A_{ij} is element ij of the Adjacency Matrix, such as A_{ij}= 1 if node i is attached to node j and 0 otherwise. This equivalent to $A * X = \lambda * X$ where λ is the largest EigenValue associated with A and X is its associated EigenVector.

D. PageRank

The PageRank centrality, PR, of a node n can be defined as:

$$PR(n) = (1 - \alpha) + \alpha * \sum_{w \in V : w \to n} \frac{PR(w)}{k(w)} \tag{4}$$

Where α, damping parameter, $\in$ [0,1]
$PR(w)$ is the PageRank of the node w which is linked to n.

E. Assortativity

Assortativity of a network evaluates the probability of connection between pairs of nodes [9].

 ISBN: 978-1-61208-793-1

TABLE I. IN EACH FORUM, AVERAGRE MÍNIMUM DISTANCE BETWEEN NODES <L>, AVERAGE BETWEENESS <B>, AVERAGE PAGERANK <PR> (CONSIDERING α=0.85), AVERAGE EIGENVECTOR CENTRALITY <EV>, AVERAGE DEGREE <K> AND AVERAGE CLUSTERING <C> VALUES.

	F1	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*F8*	*F9*	*F10*	*F11*	*F12*	*F13*	*F14*
<l>	1.22	1.13	1.34	1.01	1.78	1.15	1.85	1.11	1.13	1.18	1.02	1.15	1.85	1.94
<b>	0.007	0.007	0.012	0.009	0.007	0.013	0.008	0.006	0.015	0.013	0.006	0.008	0.010	0.002
<PR>	0.0002	0.0005	0.0090	0.0031	0.0042	0.0096	0.0063	0.0036	0.0107	0.0114	0.0043	0.0072	0.0078	0.0017
<EV>	0.0018	0.0013	0.0017	0.0078	0.0067	0.0100	0.0088	0.0056	0.0013	0.0238	0.0054	0.0086	0.0095	0.0025
<K>	16.10	15.01	13.01	65.0	16.12	8.10	12.13	15.67	25.20	12.30	18.50	20.13	15.25	10.13
<C>	0.912	0.813	0.912	0.812	0.910	0.876	0.950	0.876	0.910	0.923	0.987	0.887	0.988	0.865

TABLE II. IN EACH FORUM, NUMBER OF COMMUNITIES PER TEORETHICAL (T), PRACTICAL EXCERCISES (P) AND NEWS AND QUESTIONS FORUMS.

	F1	*F2*	*F3*	*F4*	*F5*	*F6*	*F7*	*F8*	*F9*	*F10*	*F11*	*F12*	*F13*	*F14*
T	2	-	-	2	-	-	3	3	2	2	-	-	-	-
P	-	5	4	-	5	4	-	-	-	-	6	-	-	-
NQ	-	-	-	-	-	-	-	-	-	-	-	2	2	2

III. RESULTS

Table I shows the average mínimum distance between nodes <l>, the average betweeness <b>, the average PageRank <PR>, the average EigenVector centrality <EV>, the average degree <K> and finally, the average clustering <C> values for each discussion forum. The Figure 1 shows as an example the betweeness distribution in the Forum F1.

We also measure the similarities between vertices by means of Walktrap Algorithm [3] which uses random walks on G to identify communities. This method creates a sequence of partitions (μ_k) $1 \leq k \geq n$, and chooses the best partition of the network, calculating Q_k for each partition and selecting the partition that maximizes this parameter. The modularity Q is defined as the fraction of edges within communities minus the expected value of the same quantity for a random network. The Table II depicts, in each forum, the number of communities per teorethical (T), practical excercises (P) and News and Questions Forums. It can observed that the highest number of communities happened for the practical exercises forums and the lowest number occurred for new and questions forums. The Figure 2 shows as an example the communities in the Forum F5 for the practical forum.

IV. CONCLUSIONS

The research allows to establish a methodology to analyze the interactions between students and faculties in educational forums.

The density and cohesion of the components have been studied. It has also identified the more participatory persons as well as the position that each of them occupies in the network as a whole (power relationships), which has been carried out through the analysis of different types of centrality (Betweenes, PageRank, Degree, EigenVector, Degree). Several groups of persons which are especially cohesive have also been detected. These persons and groups had a decisive influence on the results, particularly in the practical exercises. The forums related to news and general questions as well as those which refer to theorical contents presented less participation and communities. All forums were characterized by a low minimum distance between nodes, which facilitated the propagation of the answers and solutions. The nodes also presented a high average degree and assortativity. The obtained results show that this methodology allows to analize the interacts that happen in Virtual Educational Plaforms, which can help to improve the learning contexts increasing the participation and involment of the students.

This research can be continued by analyzing other topological parameters of the network. A study of the dynamical behaviour (changes in topology and node status over the time) can also be carried out.

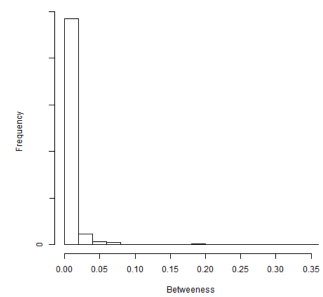

Figure 1. In Forum F1, betweeness distribution

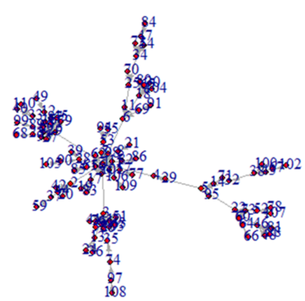

Figure 2. In Forum F5, communities for the practical forum

ACKNOWLEDGMENTS

This work was funded by Telefónica Chair at Francisco de Vitoria University.

REFERENCES

[1] S. Gregory and M. Bannister-Tyrrell, “Digital learner presence and online teaching tools: higher cognitive requirements of online

 ISBN: 978-1-61208-793-1

learners for effective learning," RPTEL, vol. 12, no. 18, 2017, https://doi.org/10.1186/s41039-017-0059-3.

[2] R. Kraleva, V. Kralev, and M. Sabani, "Digital learner presence and online teaching tools: higher cognitive requirements of online learners for effective learning," Engineering and Information Technology, no. 9, 2019, pp. 1190–1198, https://doi.org/10.18517/ijaseit.9.4.9437.

[3] M. Embi, A. Hamat, and A. Sulaiman, "The Use of Learning Management Systems Among Malaysian University Lecturers," International Journal of Learning, no. 18, 2012, pp. 61–70, https://doi.org/10.18848/1447-9494/CGP/v18i04/47554.

[4] D. Weaver, C. Spratt, and S. Nair, "Academic and student use of a learning management system: Implications for quality," Australasian Journal of Educational Technology, no. 24, 2008, https://doi.org/10.14742/ajet.1228.

[5] C. U. Bookman, Ed., Analyzing Social Networks. California, U.S.A. : SAGE Publishing, 2018.

[6] N. Fariza, M. Nor, N. Razak, and J. Aziz, "E-learning: Analysis of online discussion forums in promoting knowledge construction through collaborative learning." WSEAS Transactions on Communications, no. 9, 2010.

[7] K. Downing, T. Lam, T. Kwong, W. Downing, and S. Chan, "Creating interaction in online learning: a case study," ALT-J, vol. 15, no. 3, 2007, pp. 201–215.

[8] M. E. Newman, "he structure and function of complex networks," SIAM Review, vol. 45, 2003, pp. 167–256.

[9] ——, "Assortative Mixing in Networks," Physical review letters, vol. 89, no. 20, 2002, p. 48109–1120.