



Deliverable D5.1

Innovation action plan



DOCUMENT INFORMATION

PROJECT	
PROJECT ACRONYM	SoBigData-PlusPlus
PROJECT TITLE	SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics
STARTING DATE	01/01/2020 (48 months)
ENDING DATE	31/12/2023
PROJECT WEBSITE	http://www.sobigdata.eu
TOPIC	INFRAIA-01-2018-2019 Integrating Activities for Advanced Communities
GRANT AGREEMENT N.	871042

DELIVERABLE INFORMATION	
WORK PACKAGE	WP5 - NA4 - Accelerating innovation
WORK PACKAGE LEADER	SSSA
WORK PACKAGE PARTICIPANTS	CNR, USFD, UNIPI, UT, SNS, ETHZ, UNIROMA1, CSD, STACC
DELIVERABLE NUMBER	D5.1
DELIVERABLE TITLE	Innovation action plan
AUTHOR(S)	Nicola Del Sarto (SSSA), Alberto Di Minin (SSSA), Alessandro Marchesin (SSSA), Andrea Piccaluga (SSSA)
CONTRIBUTOR(S)	Paolo Ferragina (UNIPI)
EDITOR(S)	Beatrice Rapisarda (CNR), Valerio Grossi (CNR)
REVIEWER(S)	Paolo Ferragina (UNIPI), Rajesh Sharma (UT), Chiara Boldrini (CNR)
CONTRACTUAL DELIVERY DATE	31/03/2020
ACTUAL DELIVERY DATE	12/06/2020
VERSION	1.2
TYPE	Report
DISSEMINATION LEVEL	Public
TOTAL N. PAGES	27
KEYWORDS	Knowledge Management, Network Analysis, Demographic analysis, Survey

EXECUTIVE SUMMARY

This deliverable provides a holistic view of the SoBigData++ consortium taking into account three different dimensions represented by demographic, knowledge management and network analysis. Through a survey sent to the partners participating to the consortium we aimed at codifying contacts and interactions that take place between and within nodes of the network and create a strong community, in order to help actors to share knowledge and information collected in previous activities. The survey is reported in appendix and has been submitted to the partners during the month of march.

This report will work as preliminary snapshot of the state of the art of the consortium and will help the partners to understand which interactions need to be strengthened. Moreover the three analysed aspects highlight that the community of partners is extremely heterogeneous, with two distinct competencies groups (data science and social sciences), that the different activities of the partners have an impact on the knowledge processed (tacit and non-tacit) and that the SoBigData++ network is way dispersed, meaning that there are no clear dominant nodes or ties.

DISCLAIMER

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871042.

SoBigData++ strives to deliver a distributed, Pan-European, multi-disciplinary research infrastructure for big social data analytics, coupled with the consolidation of a cross-disciplinary European research community, aimed at using social mining and big data to understand the complexity of our contemporary, globally-interconnected society. SoBigData++ is set to advance on such ambitious tasks thanks to SoBigData, the predecessor project that started this construction in 2015. Becoming an advanced community, SoBigData++ will strengthen its tools and services to empower researchers and innovators through a platform for the design and execution of large-scale social mining experiments.

This document contains information on SoBigData++ core activities, findings and outcomes and it may also contain contributions from distinguished experts who contribute as SoBigData++ Board members. Any reference to content in this document should clearly indicate the authors, source, organisation and publication date.

The content of this publication is the sole responsibility of the SoBigData++ Consortium and its experts, and it cannot be considered to reflect the views of the European Commission. The authors of this document have taken any available measure in order for its content to be accurate, consistent and lawful. However, neither the project consortium as a whole nor the individual partners that implicitly or explicitly participated the creation and publication of this document hold any sort of responsibility that might occur as a result of using its content.

Copyright © The SoBigData++ Consortium 2020. See <http://www.sobigdata.eu/> for details on the copyright holders.

For more information on the project, its partners and contributors please see <http://project.sobigdata.eu/>. You are permitted to copy and distribute verbatim copies of this document containing this copyright notice, but modifying this document is not allowed. You are permitted to copy this document in whole or in part into other documents if you attach the following reference to the copied elements: "Copyright © The SoBigData++ Consortium 2020."

The information contained in this document represents the views of the SoBigData++ Consortium as of the date they are published. The SoBigData++ Consortium does not guarantee that any information contained herein is error-free, or up to date. THE SoBigData++ CONSORTIUM MAKES NO WARRANTIES, EXPRESS, IMPLIED, OR STATUTORY, BY PUBLISHING THIS DOCUMENT.

GLOSSARY

EU	European Union
EC	European Commission
H2020	Horizon 2020 EU Framework Programme for Research and Innovation
IT	Information Technology
DIKW	Data, Information, Knowledge, Wisdom
TAM	Technology Acceptance Model

TABLE OF CONTENTS

1. Relevance to SoBigData++	7
1.1. Purpose of this document.....	7
1.2. Relevance to project objectives	7
1.3. Relation to other work packages	7
1.4. Structure of the document	7
2. SoBigData++ Demographic	9
2.1. The SoBigData++ Consortium: Human and physical assets deployment.....	9
2.2. The SoBigData++ consortium: Partner’s area expertise	11
3. Knowledge management.....	12
3.1. Knowledge management theory.....	12
3.2. Empirical Analysis	14
4. Network Analysis	17
4.1. Network Theory	17
4.2. Empirical Analysis	18
5. Conclusions.....	24
Appendix A. Survey used for the study	25

1. Relevance to SoBigData++

The Work Package 5 of the SoBigData++ project is centred on innovation activities. It unfolds along the whole 48 months duration of the project and deals with the complexity of building the community and favouring knowledge and information exchange within the actors of the consortium, and between the consortium and the external environment.

Moreover, innovation activities will strengthen the opportunities for cooperation through a privileged communication path for events, projects, hackathon and boot camps, relying on successful cooperation with industrial and institutional stakeholders.

Within this work-package, three innovation reports will be produced.

1.1. Purpose of this document

The present deliverable aims at mapping the SoBigData++ consortium providing a holistic view on how knowledge is exchanged within partners participating to the project. This information will serve as basis for further expansion of the SoBigData++ network.

1.2. Relevance to project objectives

The present deliverable will help to consolidate and boost future expansion of the SoBigData++ network. Currently a growing number of companies, associations, institutions and universities are collaborating with the SoBigData ++ project, and one of the objectives of the project is to expand such a network. Through this activity we aim at codifying contacts that take place between and within nodes of the network creating a strong community, in order to help actors to share knowledge, information, tools and software thanks to the participation in the consortium. Furthermore, we wish to enlarge such a community, trying to involve other companies, organizations and universities.

1.3. Relation to other work packages

The present document aims at mapping the SoBigData++ community and the knowledge exchange within the nodes of the network contributing at the community building of the project. The present deliverable will serve as a basis and as a tool for the implementation of the other work packages related to the community building, named “Critical Data Literacy”, “Ethics and Legal Framework”, “Dissemination”, “Impact and Sustainability” and “Training”.

1.4. Structure of the document

The document is structured as follow. In the next Section 2, we present the demographic characteristics of the consortium, in particular we provide some information about the composition of the consortium and about the characteristics of each single partner. In Section 3, we analyse the knowledge flow happening within the consortium. In order to do that, we introduce a theoretical section in which we summarise the

basic principles of knowledge management theory and then we provide an empirical analysis on the basis of a survey we sent to the partners. In Section 4 we explain network theory and then we provide empirical analysis of the network of the SoBigData++ consortium. In Section 5 we report the conclusions of our analysis.

2. SoBigData++ Demographic

2.1. The SoBigData++ Consortium: Human and physical assets deployment

The first aspect that we analysed is the composition of the consortium. In particular our interest is to understand the commitment in terms of human resources and physical assets that each of the 30 consortium members have allocated to the consortium. In European projects each organization participate to the project with a research unit, which is a part of the whole organization.

Figure 2.1 reports the percentage of human resources that each organization participating to the SoBigData++ project have deployed to the consortium. The results highlight that most of the partners declare a participation to the consortium with 10% or less of the human resources of the organization they belong to. Only five partners report to deploy more than 20% of the human resources of the organization. For sure it should be considered the fact that the size of the different participants varies, so the relative efforts may not be comparable.

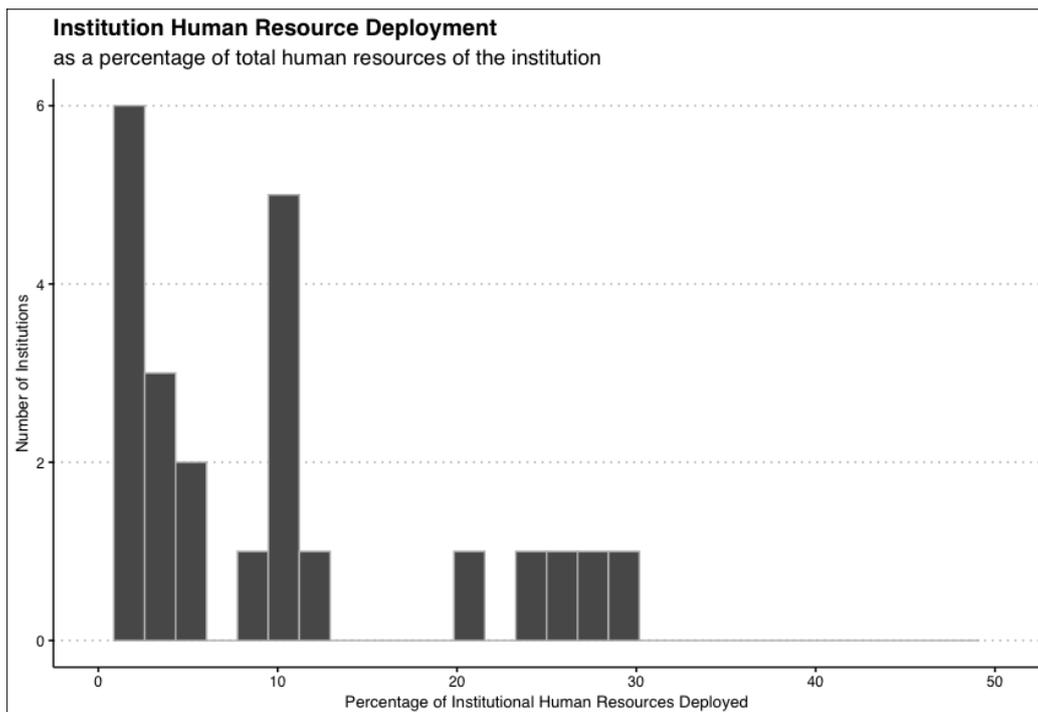


Figure 2. 1 Percentage of human resources deployed to the consortium

Figure 2.2 reports the percentage of assets deployed to the consortium by each partner. Unlike previous results, Figure 2.2 highlights that a higher number of partners declared to deploy to the consortium a higher percentage of assets. According to the results nine partners deploy to the consortium more than 10% of their assets and only 2 deploy to the consortium a percentage close to the 0. The combination of both the figures point out that the effort required to the project partners is higher for what concern the “hard” part represented by assets of organizations. For sure it should be considered the fact that the size of the different participants varies, so the relative efforts may not be comparable.

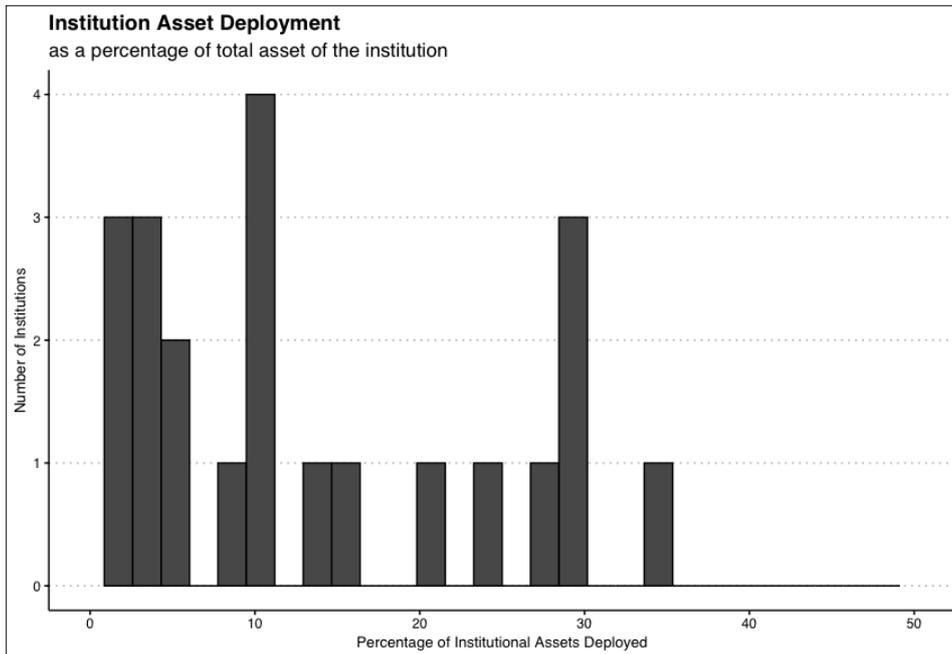


Figure 2. 2 Percentage of assets deployed to the consortium by each partner

Figure 2.3 reports the combination of the two previous results with the information regarding the main activity of the institution, namely Consulting, Research and Teaching, and Production. According to the graph three of the four consultancy companies participating to the project deploy more physical assets than human resources while the fourth allocated the same amount (30%). For what concerns the research and teaching organizations, there is a lot of heterogeneity in terms of percentage of contribution to the project, despite the results may be influenced by the “size” of the involved Institutions in terms of both total personnel and assets.

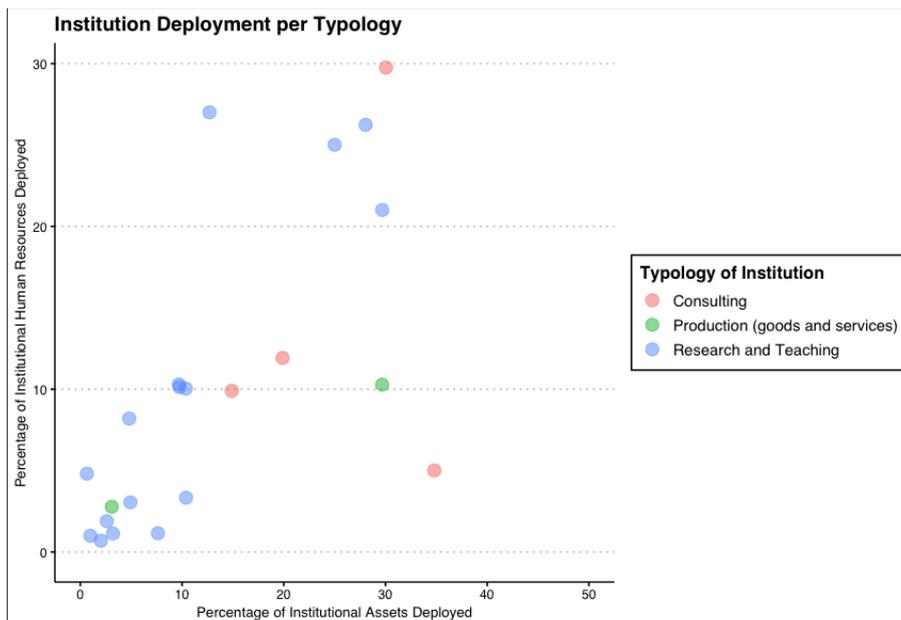


Figure 2. 3 Combination of the two previous results

2.2. The SoBigData++ consortium: Partner's area expertise

In the previous section we analysed the effort that each partner makes available in terms of human resources and physical assets. Here we analysed the main expertise of each partner participating to the project grouped into three main areas, Consulting, Research and Teaching and Production. Consulting firms seem the one more prone to invest resources in the project.

Moreover, we asked to each partner to provide its main area of expertise, divided into computer science, data science, design, ecology, education, IT, multidisciplinary, social sciences.

As expected, the main expertise of the partners of the SoBigData++ project is represented by data science. Partners with the expertise in data science cover all the typology of institutions, and this happens only for this area of expertise. The second main expertise is represented by social sciences and all the institutions with this expertise are research and teaching institutions.

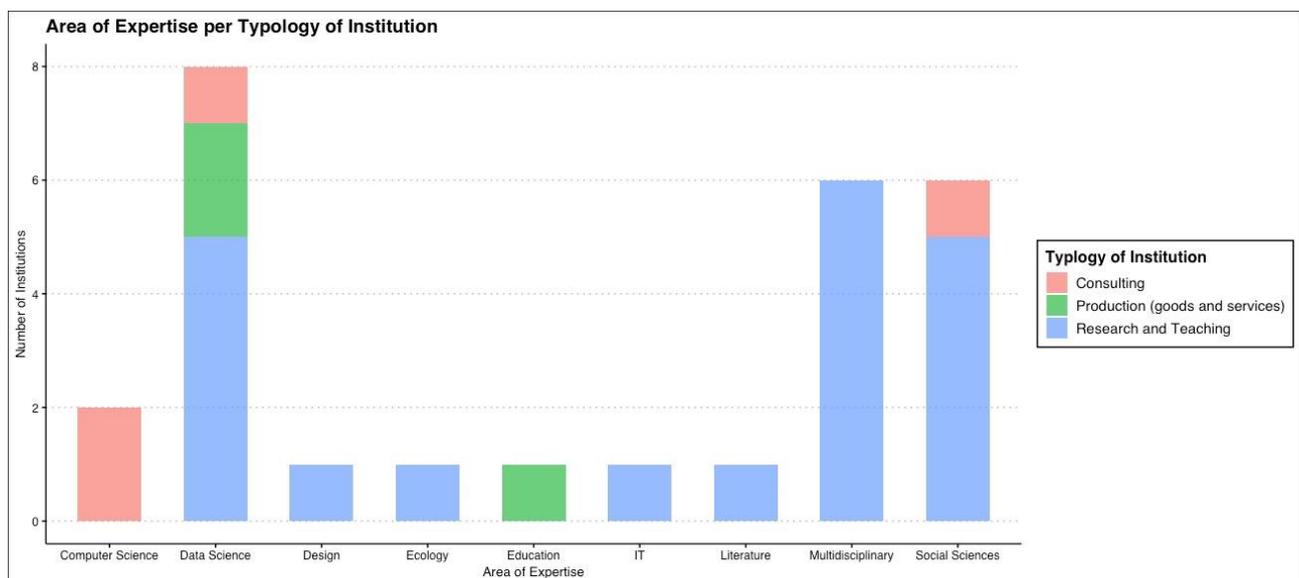


Figure 2. 4 Area of expertise per Typology of Institution

The result of this analysis highlights that the strong heterogeneity among partners belonging to the project may help its knowledge generation and diffusion. For this reason, a strong connection and interaction among partners is desirable and needs to be favoured with specific activities. In particular, the Work Package referring to the community building area may play a pivotal role in expanding the collaboration among partners creating a strong community and helping the success of the project.

3. Knowledge management

3.1. Knowledge management theory

In the last decades, the world has assisted to a progressive dematerialization of the assets of the organizations around the globe: in 1975 the percentage of the contribution of intangible assets to S&P 500 companies' market value was less than 20%. (S&P is a stock market index that measures the stock performance of 500 large companies listed on stock exchanges in the United States). 40 years later, this value reached 80%, so it quadrupled.

Connected to that, it has been possible to monitor an increase in academic research and business implementation on Knowledge Management. First of all, how do we define "knowledge management"? A good definition could be the process of creating, sharing, using and managing the knowledge and information of an organization. It refers to a multidisciplinary approach to achieving organizational objectives by making the best use of knowledge. It would be natural, at this point, to dive deeper, and wonder what is knowledge. In everyday language, we use knowledge all the time. Sometimes we mean know-how, while other times we are talking about wisdom. On many occasions, we even use it to refer to information. Part of the difficulty of defining knowledge arises from its relationship to two other concepts, namely data and information. These two terms are often regarded as lower denominations of knowledge, but the exact relationship varies greatly from one example to another.

Starting from the bottom, data can be defined as *facts and figures which relay something specific, but which are not organized in any way*. Data becomes information once it is organized, structured, categorized, calculate or condensed: providing a schema to a particular fact help convey the first layer of context. Once data is organized in information, the acts of contextualization, synthesis and comparison let information turn into viable knowledge. On top of that, the mastery of chunk of knowledge can be called *wisdom*.

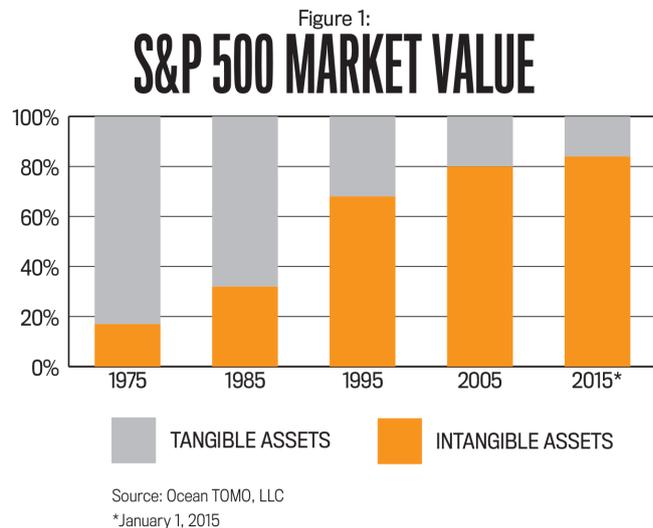


Figure 3. 2 Stock market index

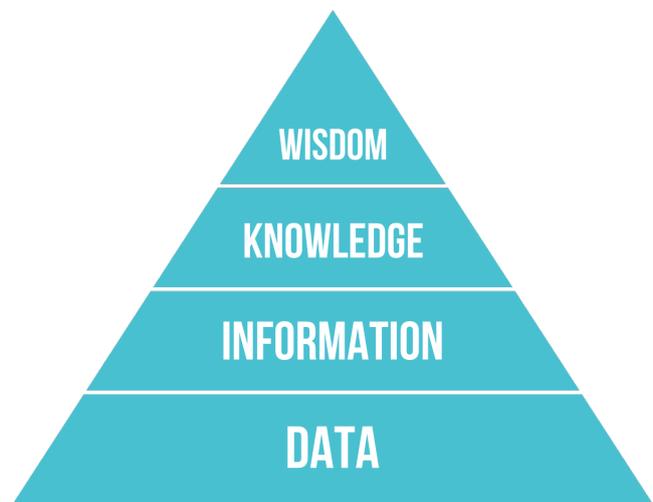


Figure 3. 1 The DIKW Pyramid

Here is an example that may clarify the concepts:

Data	Red, 192.234.235.245.678
Information	South facing traffic light on the corner of Pitt and George Streets has turned red
Knowledge	The traffic light I am driving toward has turned red
Wisdom	I better stop the car

Table 3. 1 Example of application of the Knowledge Management Theory

The next step of this logic would be to understand, how knowledge in an organization develops. According to Nonaka Dynamic Theory of Organizational Knowledge Creation (1994), organizational knowledge is created through a continuous dialogue between tacit and explicit knowledge. It is important to mind that tacit knowledge (as opposed to formal, codified or explicit knowledge) is the kind of knowledge that is difficult to transfer to another person by means of writing it down or verbalizing it. Doing a basic example, explaining to you all the details of the way you can ride a bike will not make you able to ride one without practice.

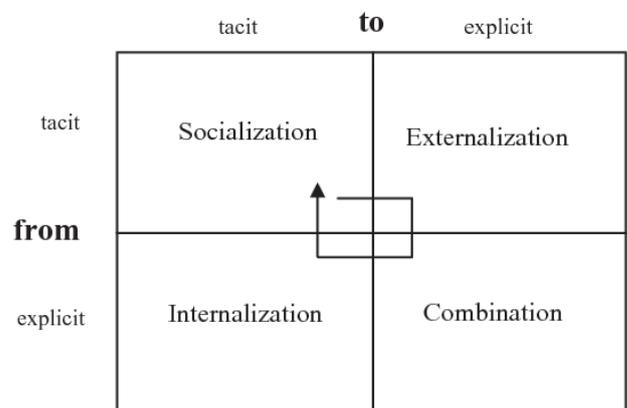


Figure 3. 3 The four patterns of interaction

The nature of this dialogue is examined and four patterns of interaction involving tacit and explicit knowledge are identified, according to the model 4 key practices enables knowledge to grow:

1. Externalization: the act of codifying and documenting a process, a procedure.
2. Combination: the act of using different pieces of knowledge together.
3. Internalization: the act of implementing new knowledge and discoveries into everyday routines.
4. Socialization: the act of sharing with others new concepts through internal dialogues and discussions.

As shown in the picture above, those different practices are not meant to happen separately but are a self-stimulating cycle that expands.

In order to complete the picture of knowledge creation inside organizations, it is worthwhile to mention the role of individuality and human behaviors: according to the technology acceptance model (TAM), the implementation of technology is the output of the behavioral intention to use it. But behavioral intention use is strongly influenced by the perceived usefulness,

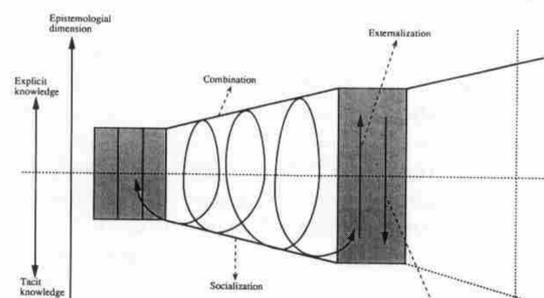


Figure 3. 4 The technology acceptance model (TAM)

ease of use and attitude toward using the technology, which are factors related to the decision maker’s background, personality, state, etcetera.

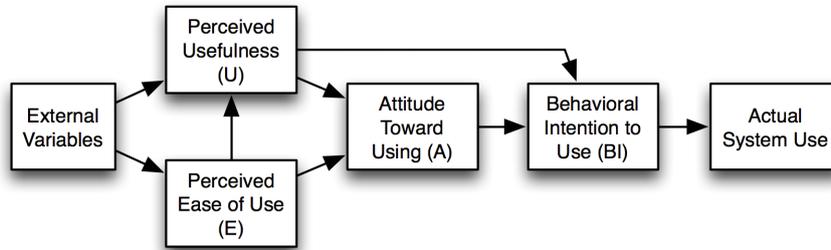


Figure 3. 5 The technology acceptance model (TAM)

3.2. Empirical Analysis

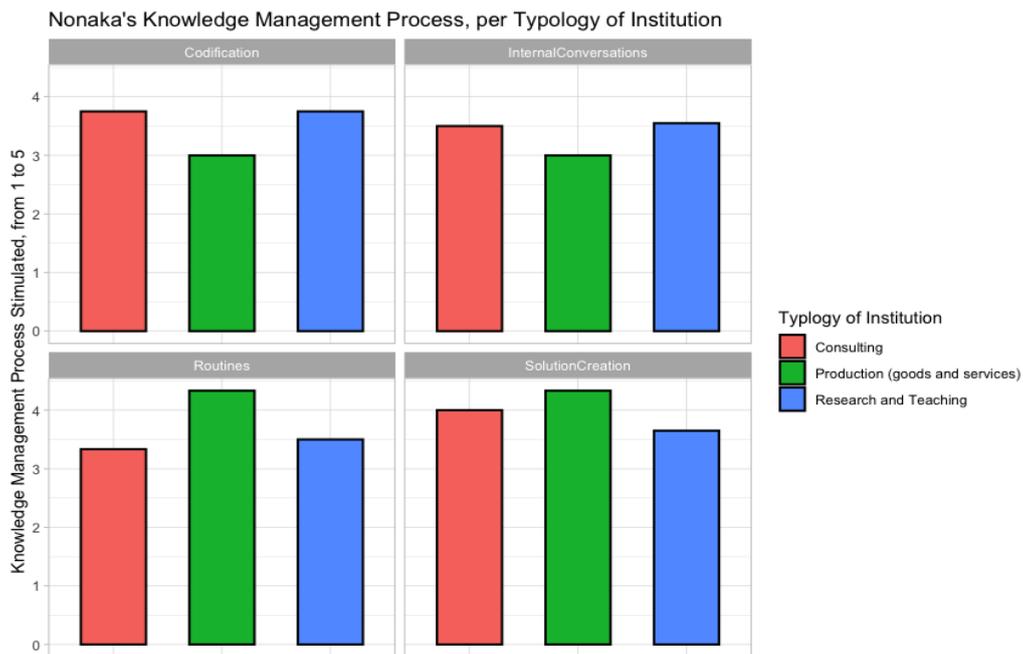


Figure 3. 6 Nonaka’s Knowledge Management Process per Typology of Institution

We asked the different institutions whether they were actually having the knowledge management processes that Nonaka identifies in his theory, and, as reported in figure 5, the answer seems to be positive: for all the 4 processes (codification, internal conversation, routines and solution creation), and all the 3 categories of institutions, the value does not go below the medium level (3). Solution creation is the most common knowledge creation process, while the least common seems to be internal conversation. Productor of goods and services are more likely to create routines and solutions, consulting seems more likely to produce solutions as well, while research and teaching institutions seem more focused on the codification of the new

knowledge created, this difference may be due to the purpose of the projects, which is technology transfer. Instead of pure research.

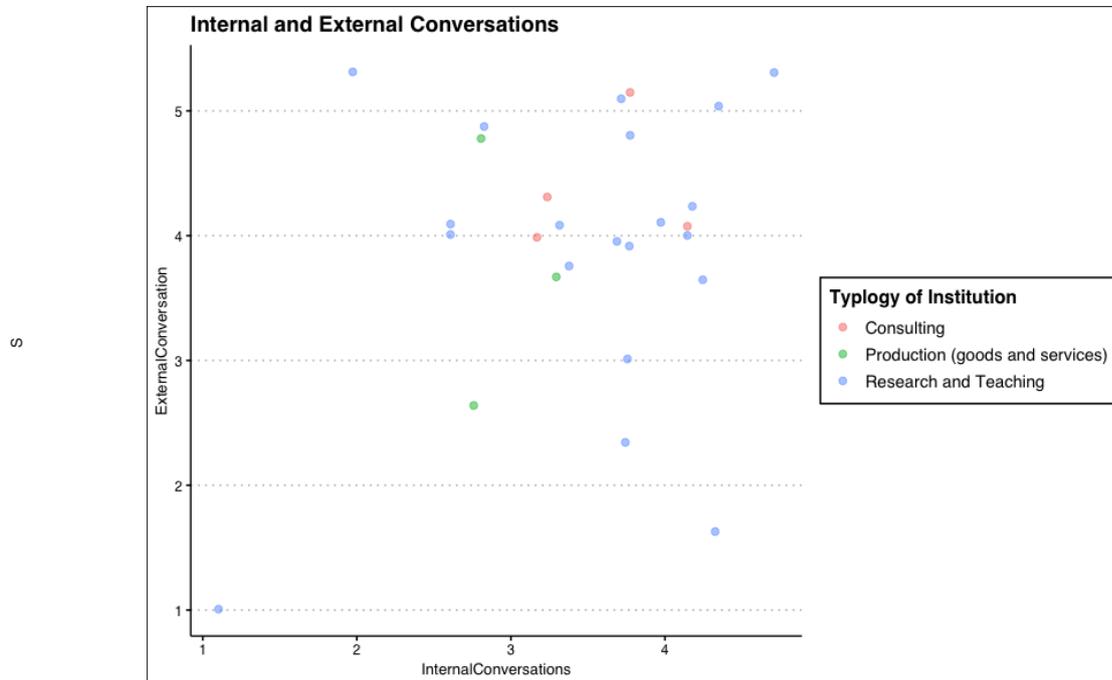


Figure 3. 7 Relationship between internal and external conversations

We also tried to understand whether there is any relationship between internal and external conversations created, as, at first sight, they seem to be correlated. With a deeper analysis we found out that by modeling a linear regression, the R squared of the model would be around 9%, meaning that the level of external conversation is not a consequence of the level of the internal conversation, the two variables do not vary together. In other words, it could be suggested that consulting institutions tend to rely more on external conversations (which is something logical considering their working procedures).

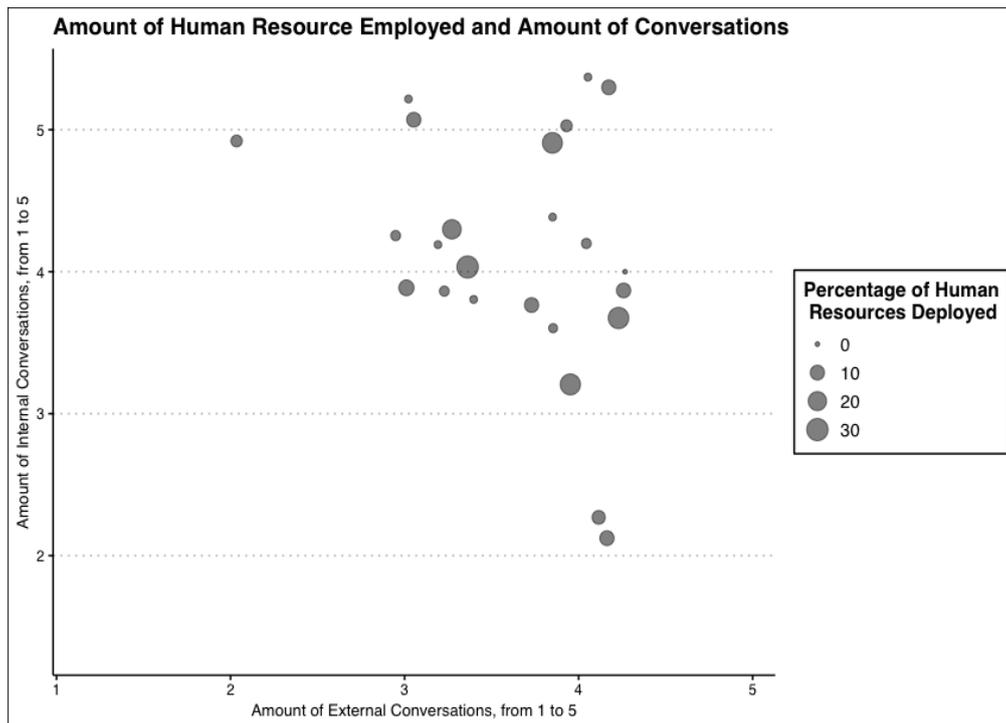


Figure 3. 8 Amount of human resources employed and amount of conversations

Does a higher deployment of human resources mean a higher amount of human interactions, that we measure in conversations? In general, this is not: for any level of human resources deployment, the institutions keep many conversations with their network peers, but it should be highlighted the fact that bigger bubble tend to lay in the bisector of the graph, so the hypothesis should not be completely false.

4. Network Analysis

4.1. Network Theory

Network theory is another important key to working in modern economies. You can identify different types of networks:

- *Information* networks are connections of information objects, like the network of citations between academic papers, world wide web, semantic networks, et cetera.
- *Biological* networks represent observable biological systems, like food web, protein interaction networks.
- *Technological* networks are designed typically for distribution of commodity and services and can be subdivided into *infrastructure* networks (internet, power grid, transportation networks) and *temporary* networks (ad hoc communication networks, sensor networks, autonomous vehicles).

Through network, it is possible to map a wide range of aspects of life, like 15th Century Florentine Marriages, friendships in sports clubs, viral marketing strategies, et cetera.

Different characteristics can be highlighted in the description and representation of a network:

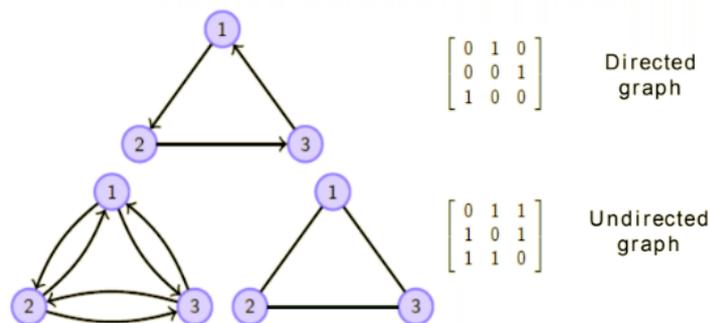


Figure 4. 1 Representation of a network

- It can be *directed*: the links that connect the different nodes have a direction, for instance when communication has a sender and a receiver. Otherwise we call it “undirected”.
- It can be represented through a *matrix*.
- It can be *weighted* so that the value of the different links (edges) has a value.
- In any given graph it is possible to identify:
 - a *walk*, which is a sequence of links,
 - a *path*, which is a walk that does not contain the same node twice,
 - a *cycle*, which is a path with a final link to the initial node
 - a *geodesic*, which is the shortest path (i.e. with minimum number of links) between two nodes.
- *Neighborhood* is the set of nodes linked to a specific node.
- The *degree* is the number of links that each node has.

- The *diameter* is defined as the maximum distance between any two nodes.
- The *centrality* measures how often a given node is in the path that connects two different nodes, and thus the importance of a specific node.
- *Betweenness centrality* is a measure of centrality in a graph based on shortest paths. For every pair of vertices in a connected graph, there exists at least one shortest path between the vertices such that either the number of edges that the path passes through (for unweighted graphs) or the sum of the weights of the edges (for weighted graphs) is minimized. The definition is not always valid: in case the network is not connected enough and oriented, it may not be possible to find the shortest path. The betweenness centrality for each vertex is the number of these shortest paths that pass through the vertex. Betweenness centrality was devised as a general measure of centrality.
- Another proxy for the centrality of a node is the so-called *degree centrality*, it is the number of links incident upon a node, and it can be seen also as the probability of the node to get in contact with anything flowing through the network.
- *Closeness centrality* (or closeness) of a node is a measure of centrality in a network, calculated as the reciprocal of the sum of the length of the shortest paths between the node and all other nodes in the graph.

4.2. Empirical Analysis

The previous introduction to network theory was useful to give some context to the empirical analysis run on SoBigData consortium. Every participant was asked to rate the intensity of their collaboration with each other participant on a scale from 1 to 5, and also to rate the level of collaboration with Individual WP Leaders, Project Coordinators, Project Managers and Individual Task Leaders. The results were used to model a network that has been analyzed starting from the theory mentioned before.

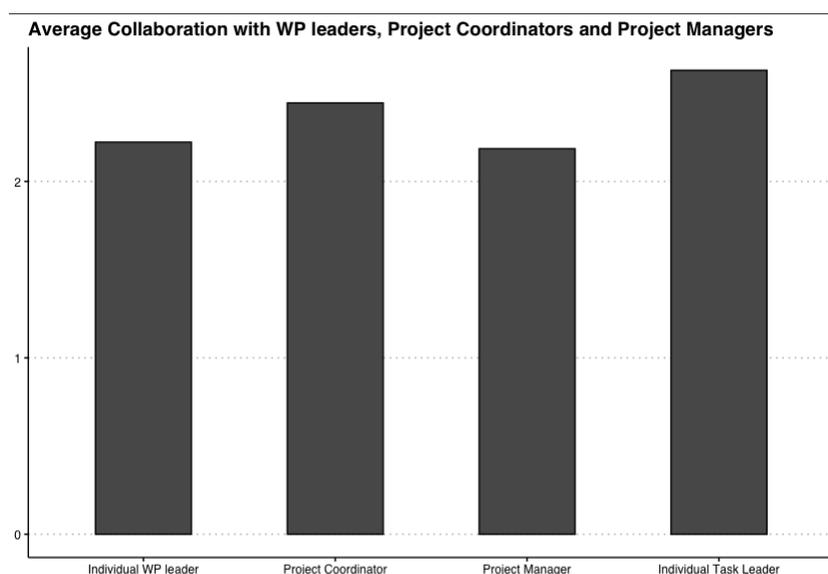


Figure 4. 2 Average collaboration with WP Leaders, Project Coordinators and Project Managers

We wondered what is the level of collaboration with “Individual WP Leaders”, “Project Coordinators”, “Project Managers” and “Individual Task Leaders”. In order to that, the average of the intensity of collaboration of each participant with those 4 actors has been taken. Checking the different results, we found out that individual task leaders seem to be the most involved by the institutions, while the least involved seems to be the project manager. This suggests that the network is decentralized.

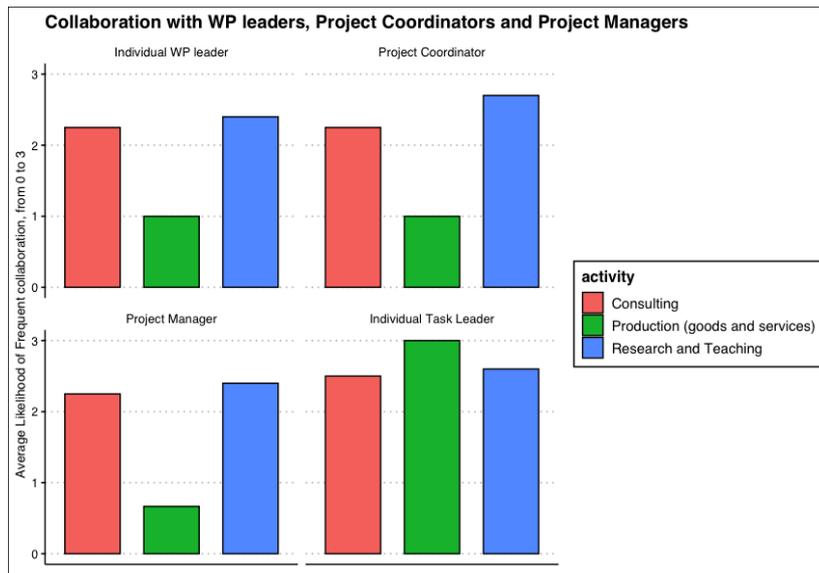


Figure 4. 3 Collaboration with WP Leaders, Project Coordinators and Project Managers

Breaking down the previous graph, it is possible to see that Production institutions are very connected with individual task leaders and very low connected with the other management entities, while Consulting and Research and Teaching institution show a way lower variation among the two.

SoBigData++ Network, only relevant ties

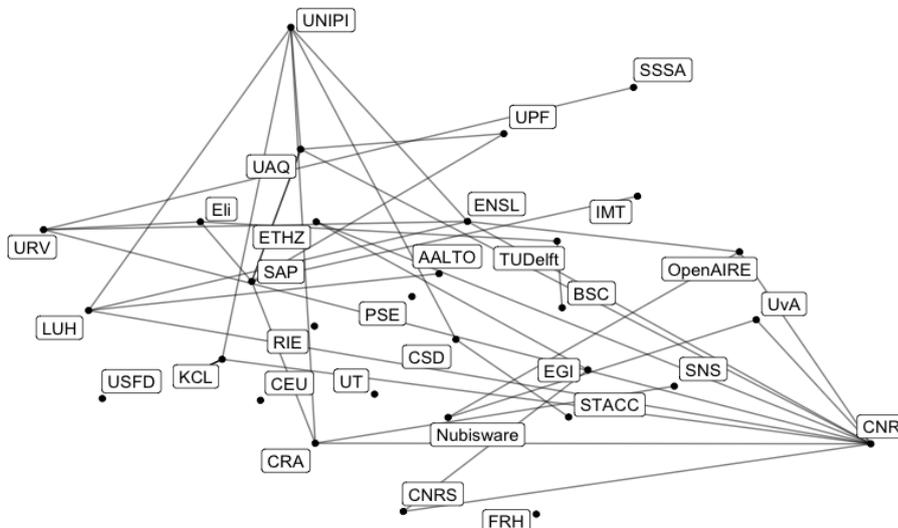


Figure 4. 4 Relevant ties of the SoBigData++ Network

In this first elaboration, the only “heavy ties” are shown. Every institution had to rate their link with every other institution, and this is the result: a low number of links is actually considered relevant. In particular, the institutions with multiple strong ties are the CNR and University of Pisa (UNIPi).

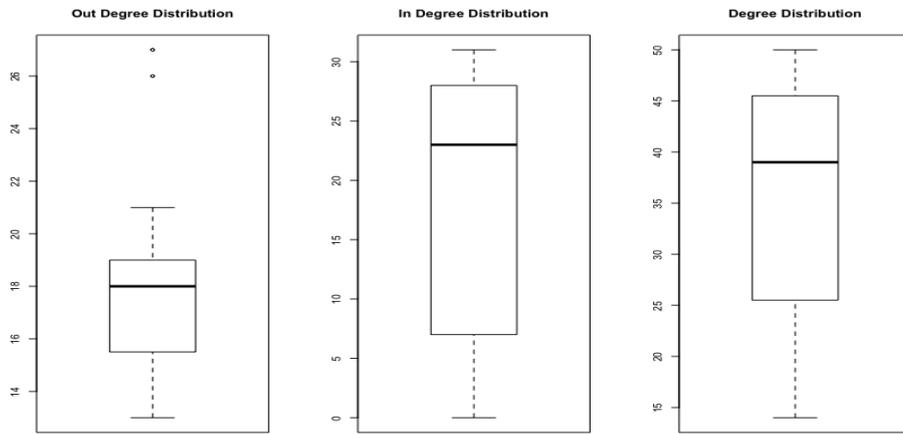


Figure 4. 5 Number of links that start from and end into a particular node

Out Degree and In Degree counts the number of links that start from and end into a particular node, respectively (see Figure 11). The average value of out degree seems to be lower than the average value of in degree, and the distribution looks also steeper and shorter, meaning that each institution is expecting a lower level of the collaboration than the reality with the other institutions.

SoBigData++ Network, by edge betweenness

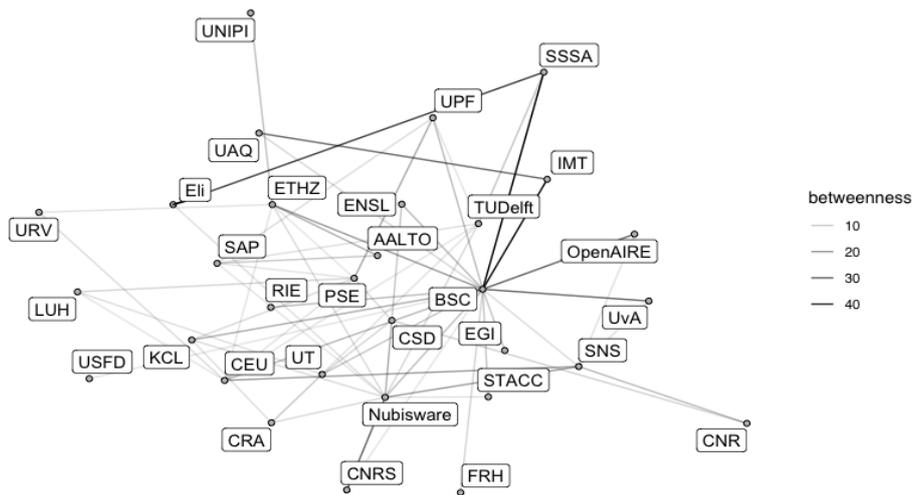


Figure 4. 6 SoBigData++ Network by edge betweenness

Highlighting the betweenness of each link which tells, as mentioned before, degree to which nodes stand between each other, we can notice that there are few key links, in particular the ones connecting Eli, Scuola

Sant’Anna, TuDelft, IMT, OpenAIRE e UvA (see Figure 12). These may be due to the fact that these institutions have few relevant links, meaning that betweenness is spread among fewer links increasing his value.

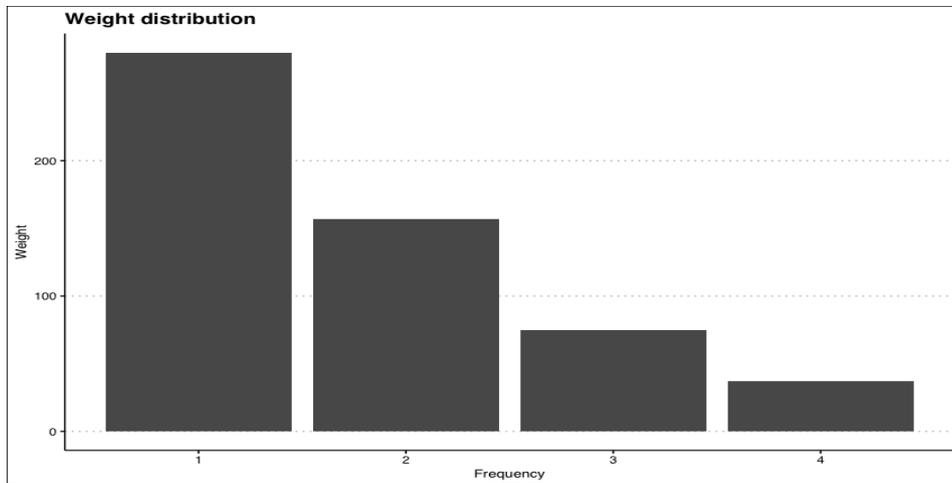


Figure 4. 7 Weight distribution

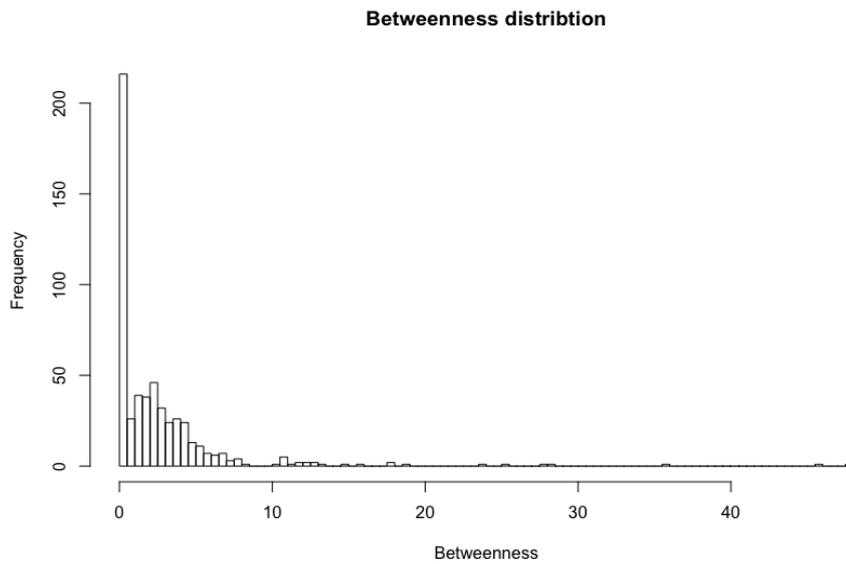


Figure 4. 8 Betweenness distribution

Betweenness is right skewed: the majority of the links has a betweenness that is very close to zero, while very few overcome the value 10.

SoBigData++ Network, by node closeness and degrees

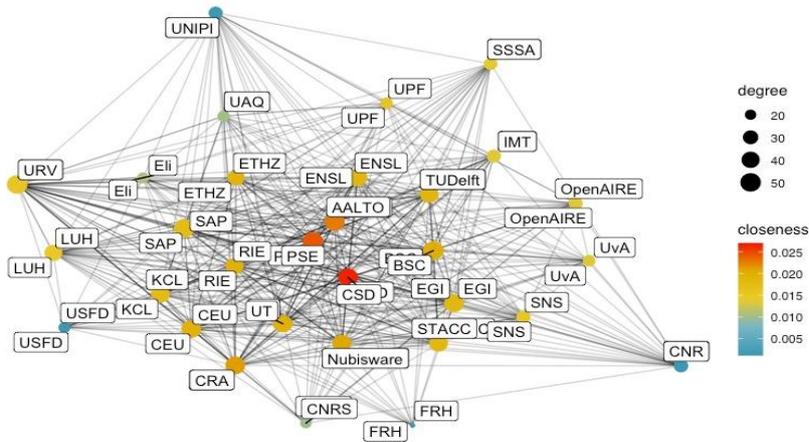


Figure 4. 9 SoBigData++ Network by node closeness and degrees

Closeness is a proxy of the importance and refers to nodes. As mentioned before, it specifies how much a node is central compared to the others. The most important nodes here appear to be PSE, AALTO, and CSD (see Figure 15). The result can be surprising, having in mind that other entities formally play bigger roles, put in place more resources, an instance of that is surely CNR, but this apparent discrepancy could be linked to two facts: closeness value is not only a consequence of the node itself but also of the neighborhood around the node. In second place, bear in mind that the network is not built on budgets and WP task, but on the expectations of the other participant regarding their future collaborations, subjectivity may play a role.

Are degree centrality and closeness centrality correlated? The graph suggests that yes, with few exceptions. Let's go deeper and create a linear model to verify whether central nodes have a higher degree or not.

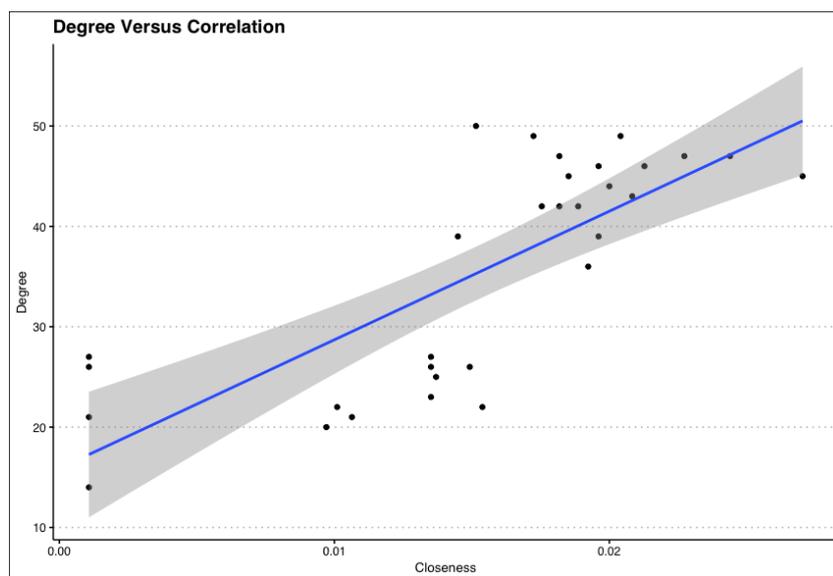


Figure 4. 10 Degrees versus correlation

Closeness seems to be a predictor of the degree of each node, for some extent: the linear model is able to explain in a significant way the correlation between the two variables, the R squared is almost 60%.

```

Call:
lm(formula = b ~ a)

Residuals:
Min 1Q Median 3Q Max
-13.598 -6.516 1.940 4.768 14.700

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.893 3.247 4.895 3.40e-05 ***
a 1280.831 194.983 6.569 3.39e-07 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.278 on 29 degrees of freedom
Multiple R-squared: 0.5981, Adjusted R-squared: 0.5842
F-statistic: 43.15 on 1 and 29 DF, p-value: 3.388e-07
    
```

Are there sub-clusters? Have the institution any tendency to form sub-communities? According to the fast-greedy algorithm, which is one typology of the greedy algorithms used to find out hierarchies in clusters, there are 4 communities that shares many more links among them than others, they are color coded in the following graph. Even if in this two-dimensional view they are not looking close, a 3D model would for sure be able to make the clusters visible.

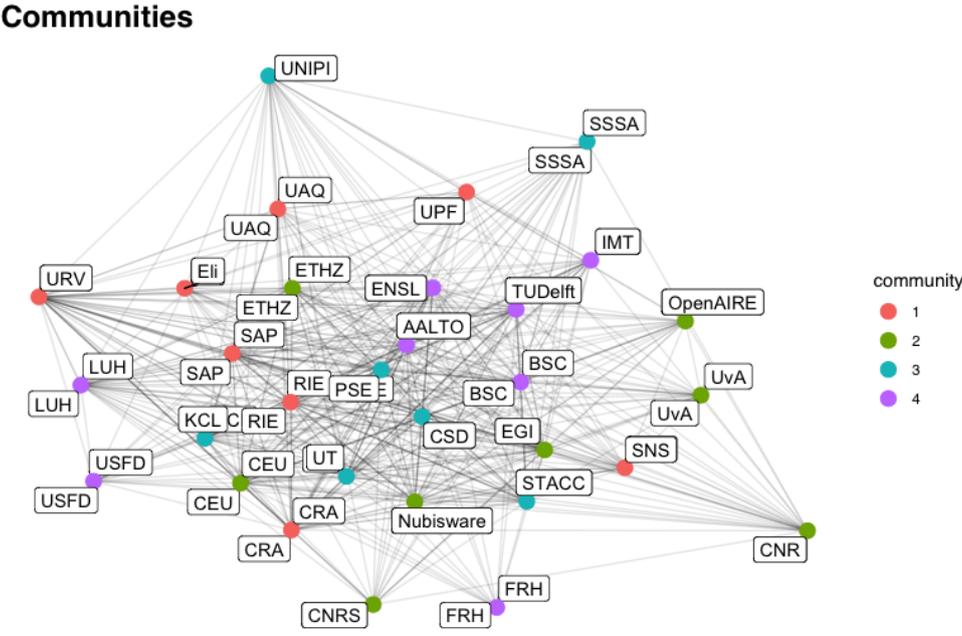


Figure 4. 11 Communities

5. Conclusions

The goal of this report was to focus on the community working on the SoBigData++ project by describing the roles and the complex interactions among the partners, mapping the network and taking a snapshot on how the different partners perceived their interactions and knowledge exchange with other partners.

The community of partners is extremely heterogeneous, with two distinct competencies groups (data science and social sciences) and a prevalent area of activity represented by Research and Teaching.

Knowledge management analysis highlighted what knowledge management processes are deployed within the project, showing that the different activities of partners have an impact on the knowledge processed. In particular our analysis highlighted that producer of goods and services are more likely to create routines and solutions, consulting seems more likely to produce solutions as well, while research and teaching institutions seem more focused on the codification of the new knowledge created as an indirect consequence of the development of SoBigData platform, although the results may be influenced by response biases. In addition, we found that within the SoBigData++ consortium there is a correlation between information exchange within the project and also between the project and the external environment. Such a result is very important as it highlighted that the consortium has a strong relation with external world, and it may be argued that the dissemination of the result can be effective in the future.

Finally, the network analysis highlighted the structure of the SoBigData++ consortium: the network is way dispersed, there are no clear dominating nodes or ties, the ties are averagely weak, except for few of them. It is possible to highlight 4 sub-communities that share more connections between each other than with the rest of the consortium. It is going to be extremely interesting to see how the network will evolve throughout time, whether some ties will strengthen or weaken, and whether the relative importance of the nodes will vary.

Our analysis acts as a snapshot of the SoBigData++ consortium. The preliminary nature of this report is useful to understand the evolution of the consortium during the project. In particular it can be argued that in the future the structure of the network and of the collaboration among partners will change.

Appendix A. Survey used for the study

Survey used for the study

Demographic

In this section, please try to represent (to the best of your knowledge) the view of your organization

1. Please select your institution from the list below:

2. What percentage of your institution's **human resources** are deployed to SoBigData++ Project?

0% 50% 100%

A horizontal slider control with a circular knob at the 0% position. The scale is marked at 0%, 50%, and 100%. To the right of the slider is a checkbox.

3. What percentage of your institution's **assets** are deployed to SoBigData++ Project?

0% 50% 100%

A horizontal slider control with a circular knob at the 0% position. The scale is marked at 0%, 50%, and 100%. To the right of the slider is a checkbox.

4. Please describe your institution's main activity

Consulting

Research and Teaching

Production (goods and services)

5. Please describe your institution's main area of expertise

Knowledge Management

Knowledge management (KM) is the process of creating, sharing, using and managing the knowledge and information of an organization. In this section, we are interested in understanding how knowledge flows within the SoBigData Consortium.

6. How much do you agree with the following sentences:

	1	2	3	4	5
SoBigData++ stimulates conversations within your organization	<input type="radio"/>				
SoBigData++ stimulates conversations with external organizations	<input type="radio"/>				
SoBigData++ generates new knowledge in a written format	<input type="radio"/>				
SoBigData++ integrate existing information into new solutions	<input type="radio"/>				
SoBigData++ develops new routines and best practices	<input type="radio"/>				

7. Has your institution ever participated to a project such as SoBigData++? If yes, describe it briefly.

8. Do you think the **participation** to SoBigData++ project is beneficial to your institution?

9. Do you think your institution is going to benefit from the **final outputs** of SoBigData++ project? If yes, how?

Network Effects

In this section we try to map the network of SoBigData consortium.

10. From 1 (null) to 5 (very high), how do you evaluate the intensity of your collaboration with:

	1	2	3	4	5
SoBigData++ Project Coordinators	<input type="radio"/>				
SoBigData++ Project Manager	<input type="radio"/>				
Individual WP Leaders	<input type="radio"/>				
Individual Task Leaders	<input type="radio"/>				

11. From 1 (null) to 5 (very high), how do you evaluate the intensity of your collaboration with:

	1	2	3	4	5
CNR – Consiglio Nazionale delle Ricerche Italy	<input type="radio"/>				
USFD - The University Of Sheffield United Kingdom	<input type="radio"/>				
UNIP1 – Università di Pisa Italy	<input type="radio"/>				
FRH- Fraunhofer-Gesellschaft Germany	<input type="radio"/>				
UT – University of Tartu Estonia	<input type="radio"/>				
IMT – Scuole Alte Studi Lucca Italy	<input type="radio"/>				
LUH - Leibniz Universität Hannover Germany	<input type="radio"/>				
KCL - King's College London United Kingdom	<input type="radio"/>				
SNS - Scuola Normale Superiore Italy	<input type="radio"/>				
AALTO - Aalto University Finland	<input type="radio"/>				
ETHZ - Eidgenössische Technische Hochschule Zürich Switzerland	<input type="radio"/>				
TU/e - Technische Universiteit Delft Netherlands	<input type="radio"/>				