Application of Advanced Analytical Techniques in Support of Socio-economic Impact Assessment of Innovation Funding Programmes

Katja Berčič
Jožef Stefan Institute
Jamova cesta 39
1000 Ljubljana
Slovenia
katja.bercic@ijs.si

Gabriella Cattaneo
IDC Italia
Viale Monza 14
20127 Milano
Italy
gcattaneo@idc.com

Mario Karlovčec
Jožef Stefan Institute
Jamova cesta 39
1000 Ljubljana
Slovenia
mario.karlovcec@ijs.si

Flavio Fuart
Jožef Stefan Institute
Jamova cesta 39
1000 Ljubljana
Slovenia
flavio.fuart@ijs.si

Gaber Cerle
Jožef Stefan Institute
Jamova cesta 39
1000 Ljubljana
Slovenia
gaber.cerle@ijs.si

ABSTRACT
This paper gives an introduction to the European Programme for Internet Innovation and the needs for advanced analytical techniques in assessment of the socio-economic impact of funded projects. A technical overview and architecture of developed IT tools follows. Broadly, two set of tools were explored. (1) An on-line assessment environment that provides programme managers and companies with feedback about their potential socio-economic impact. For this set of tools the emphasis is on visualisation techniques, therefore a detailed rationale and scientific background for those is provided with usage examples. (2) Statistical modules for identification of funding approaches that, potentially, provide best results ("best practices") have been developed and applied to available data. Through one use-case we have shown that the proposed IT system is useful in measuring the impact of innovation funding programmes, detecting good management practices and giving support to funded projects. The whole system has been published on a public repository with an open source license, giving opportunity to re-use, customize and integrate the reporting system into other impact assessment environments.

Categories and Subject Descriptors
H.4.0 [**Information Systems Applications**]: General—*reporting, statistical analysis, visualisation, networks*

General Terms
Algorithms

Keywords
impact assessment, network visualisation, statistical methods

1. INTRODUCTION
The Future Internet Public-Private Partnership (FI-PPP) is the European programme for Internet innovation. Phase Three of the FI-PPP funding is targeting more that 1000 entrepreneurs, start-ups or SMEs in an attempt to multiply the uptake and impact of the technologies developed in previous phases[1]. Under that framework the Future Internet Impact Assurance (FI-MPACT)[2] Support Action was funded in order to collect and assess the qualitative and quantitative evidence of the potential socio-economic impact of the programme by measuring and projecting market sector economic potential, stakeholder take-up and technological impact of Phase III SME Accelerator projects to 2020. Accelerator projects aim at investing in the strongest start-ups ("subgrantees") across Europe and create an ecosystem where entrepreneurs and business incubators meet.

Based on Key Performance Indicators (KPI) elaborated in the Impact Assessment Guidebook [3], a set of **Impact Assessment** tools were developed to collect empirical data from subgrantees and other interested initiatives. In this context, developed IT tools provide an automated assessment system, driven by a set of KPIs, allowing Accelerator projects, other start-ups and entrepreneurs to measure the potential impact compared to industry standards and the global community of FI-PPP projects. By benchmarking their progress in relation to different business processes, they can identify areas where improvements are needed and measure progress.

Furthermore, a set of scripts for statistical data processing were combined in **Accelerator Benchmarking** tools

---

[1]https://www.fi-ppp.eu/
[2]http://cordis.europa.eu/project/rcn/191426_en.html

to support the identification of accelerators best practices. This quantitative analysis was used to support, guide and strengthen the expert judgement of qualitative indicators of best performing accelerators.

In this paper we present the developed IT tools [2], their usage in practice, some results and possible use of these tools in future.

2. IMPACT ASSESSMENT TOOL

The Impact Assessment tool[3] was focused around funded subgrantees to facilitate mapping of this portfolio, to contribute to the overall impact assessment of the FI-PPP Phase 3 and assist in forecasting the potential impact of this intervention up to 2020. The Self-Assessment tool is open to all interested parties and respondents can undertake the survey at different stages to measure their progress.

The Impact/Self-Assessment tool provides a start-up sanity check, by calculating KPIs for Innovation, Market Focus, Feasibility and Market Needs. Instant feedback is provided by benchmarking respondent's scores with average scores of his/her peers, or a group of most successful peers.

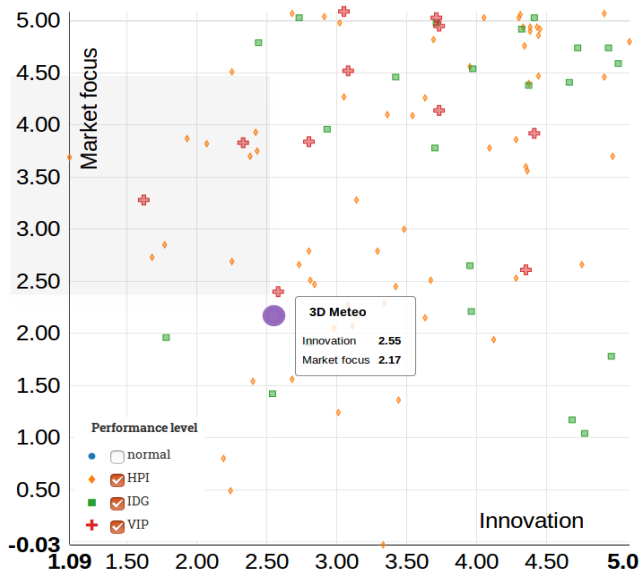## 2.1 Key Performance Indicators Benchmarking

Figure 1: Performance graph based on market focus and innovation variables for a selected project represented with a large violet circle. Different shapes correspond to performance levels: rhombus for Hight Performance Initiatives, squares and crosses for a small selection promising initiatives that received further funding for dissemination activities.

This dynamic benchmarking against other respondents gives entrepreneurs and their mentors a tool to monitor progress and identify areas where additional support is required. The

tool is implemented as a interactive scatter plot that positions projects in two-dimensional space defined with performance indicators. This enables comparison of projects with different performance levels against different pairs of indicators. Performance indicators supported by the tool are: innovation, market focus, feasibility, market needs and mattermark growth. Figure 1 gives an example of performance graph where the selected project is compared with other high performance projects, based on market focus and innovation variables.

2.2 Multi-attribute Based Similarity Tool

In order to provide both the sub-grantees and accelerators better insight in project's performance, a tool that shows the selected project's position in relation to other projects was created. In this tool, similarities among projects are determined using a wide range of attributes that describe projects. This gives users a chance to form their own judgement about "good" or "bad" similarities. The tool is implemented using several well-known data and network analysis methods, and visualized as a network graph. The visualization gives sub grantees and their mentors/reviewers insight into how they compete with other projects and ideas, identify possible similarities, find opportunity windows or search for possible partners. The tool is implemented as an open API using QMiner [5] data analytics platform for processing large-scale real-time streams containing structured and unstructured data. The approach used for implementing multi-attribute based similarity tool consists of following steps: (1) Importing data and feature extraction; (2) Computing the main similarity graph; (3) Generating custom graph for selected project; and (4) Visualization.

In the first step, project data is imported into the system and features are extracted from the data. Prior to importing, data is transformed into JSON file format. The file contains a configuration part that determines which attributes of the data will be used as features in further analysis. This makes the system flexible for feature-set changes. The feature-set consists of numerical, categorical and textual data obtained from questionnaires. Feature extractor for textual data applies English stop-words removal in the tokenization phase and normalizes the word frequencies using TF-IDF weighting.

The main similarity graph shows project similarities based on multiple attributes in form of a two dimensional network graph. This graph is the basis for constructing other custom graphs where a particular selected project is in the focus. The graph is constructed in two steps. First, the multidimensional vector representation of projects obtained in the feature extraction phase is transformed into two dimensional representation by using Multidimensional scaling (MDS) [7] method. A similar approach of dimensionality reduction was used in [4] for visualization of a text document corpus. In the second step of graph construction, the two dimensional projections of projects are connected using Delaunay triangulation [9].

A custom graph is such a graph where one project is in the main focus. This can be a project selected from the list of already imported projects. It can also be generated for a new project that is in the preparation phase, to see

Figure 2: Multi-attribute based similarity graph for an arbitrary project. The selected project is represented with a black dot in centre of the figure.

where is it positioned in relation to other projects. A custom graph is generated by first extracting features from the selected project using the same feature extractors that were used for computing the main similarity graph. In second step, distance between each of the imported project and the selected project is computed (either cosine or euclidean distance can be used). Based on user selected threshold, an edge is created between more similar projects, while less similar projects are excluded from the graph. Finally, projects selected for the custom graph are connected based on the main similarity graph.

The final step of the similarity tool is graph visualization. The graph is visualized using SigmaJS[4] library and a plug-in that implements ForceAtlas2 [6] algorithm. The lay-outing algorithm mimics the physical attraction and repulsion forces, placing well connected and similar project closer on the two dimensional plane. An example of a similarity graph for a project is given in Figure 2.

3. ACCELERATOR BENCHMARKING/BEST PRACTICES REPORTING TOOLS

Accelerator Benchmarking/Best Practices Reporting Tools are written in R [8] programming language and software environment for statistical computing and graphics. In FI-IMPACT, several R modules[5] have been developed to automatically produce statistical reports about Accelerator Benchmarking and Best Practices based on FI-IMPACT databases.

The databases are imported and merged into an R data frame. The tool then produces correlation heatmaps, practice scores with accompanying mean difference t-test results and FI PPP phase projects connectivity information. Most of the results are produced in final-form CSV files. The tool also outputs several sets of diagrams (boxplots for all cat-

---

egorical indicators, CDF diagrams for practice scores and score histograms). All the diagrams are produced in two versions - with all sub-grantees and per-accelerator. Results are grouped into three main categories, as described below. The diagram in Figure 3 shows the structure of the tool.

The *degree of connectivity* tables are a result of analysis of FI-PPP project partnerships. They contain counts of different types of connections. Every pair of projects can have more than one connection, where each such connection between two projects in the three FI-PPP phases represents a partner that appeared in both.

Correlation analysis is performed on all numerical accelerator properties and all numerical subgrantee performance indicators by correlating each possible pair. The results include a heatmap of correlations, a table of correlations and a corresponding table with sizes of samples on which the correlations were computed.

Best practices identification produces notched box plots for categorical accelerator properties, CDF plots for binary accelerator properties, correlation outputs for numerical accelerator properties (heatmap and tables) and histograms for performance indicators. These can then provide insight into which practices coincided with better subgrantee performance.

We define a good practice as an activity performed by one or more of the Accelerators' consortia according to their acceleration plans, which based on objective evidence, is shown to have contributed to the good performance of subgrantees. Here, good performance of the subgrantees is primarily their market success (measured in terms of positive dynamics of revenue growth and customer growth); their ability to convince potential investors and collect additional funding ("traction"); if they are not yet on the market, their market readiness (measured by FI-IMPACT's KPIs scores).

The full report about the usage of developed tools, combined with qualitative data based on face-to-face interviews with Accelerators coordinators and interpretation of results is given in [1]. Some insights were drawn from the results computed by the benchmarking tool, while some of the conclusions formulated with other methods were confirmed by it [1]. For example the geographical scope of the partnerships did not seem to make a major difference, while the presence of professional accelerators had a strong positive influence on most subgrantee performance indicators.

There were several results computed in the benchmarking tool that did not provide any meaningful insights, but did confirm that the data exhibits some expected properties. For example, unsurprisingly, the benchmarking tool showed a strong positive correlation between investment per subgrantee and KPIs measuring feasibility and innovation.

4. CONCLUSIONS

We have shown how a set of IT tools and computational methods can be used to assess the socio-economic impact of a big investment programme, such as FI-PPP European programme for Internet innovation. On one hand, the implementation followed a strict methodology used for bench-
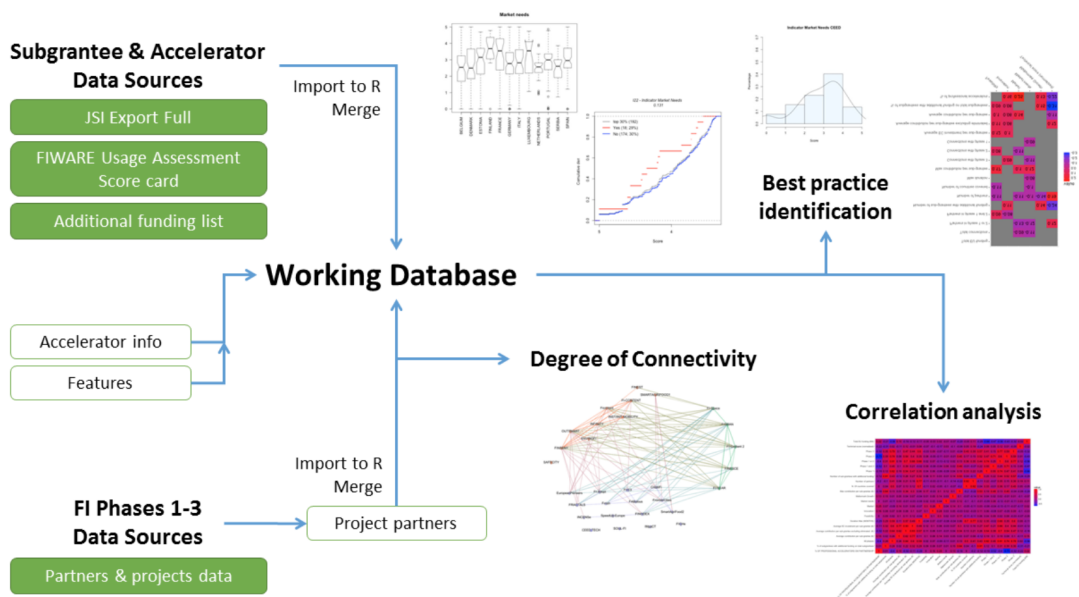
Figure 3: Benchmarking tool inputs and outputs

marking through questionnaires and calculation of respective KPIs, on the other hand, more exploratory techniques were used for statistical analysis and visualisation.

While it is true that many of the results of the statistical analysis implemented in the Accelerator benchmarking tools did not show statistically significant new results, it is also true that the results displayed most of the expected correlations. The statistical analysis explains little of the variations in performance of the subgrantees. In our opinion, the main weakness of the statistical correlation analysis was the very limited available data on actual market performance. However, weak signals, combined with with the results of the qualitative interviews, point to the positive role of professional accelerators within consortia, and positive impacts of practices such as workshops, matchmaking and providing gateways to further funding.

Impact Assessment and Similarity Graph benchmarking tools have been deployed and made available to over 700 subgrantees (projects), their mentors an coordinators. The tool was deemed very useful for long-term monitoring of sub-grantees performance. In this respect, the FI-IMPACT project consortium is drawing plans to assure long-term sustainability of the deployed IT systems.

5. ACKNOWLEDGEMENTS

6. REFERENCES

[1] FI-IMPACT Project. Analysis of accelerators' good practices annex 8.5 to deliverable d2.4 update of impact assesment and forecast, 2016. Annex 8.5 of D2.4 Available from http://www.fi-impact.eu/page/docdownload/477/ [accessed 12 September 2016].

[2] FI-IMPACT Project. Fi-impact online assessment environment, 2016. Deliverable D4.3 Available from http://www.fi-impact.eu/page/docdownload/479/ [accessed 12 September 2016].

[3] FI-IMPACT Project. Impact assessment guidebook, v2, 2016. Deliverable D2.1.v2 Available from http://www.fi-impact.eu/page/docdownload/474/ [accessed 12 September 2016].

[4] B. Fortuna, M. Grobelnik, and D. Mladenic. Visualization of text document corpus. *Informatica (Slovenia)*, 29(4):497–504, 2005.

[5] B. Fortuna, J. Rupnik, J. Brank, C. Fortuna, V. Jovanoski, M. Karlovcec, B. Kazic, K. Kenda, G. Leban, D. Mladenic, A. Muhic, B. Novak, E. Novak, J. Novljan, M. Papler, L. Rei, B. Sovdat, and L. Stopar. Qminer: Data analytics platform for processing streams of structured and unstructured data. In *Software Engineering for Machine Learning Workshop, Neural Information Processing Systems*, 2014.

[6] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PLoS ONE*, 9(6):1–12, 06 2014.

[7] J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.

[8] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013.

[9] P. Su and R. L. S. Drysdale. A comparison of sequential delaunay triangulation algorithms. In *Proceedings of the Eleventh Annual Symposium on Computational Geometry*, SCG '95, pages 61–70, New York, NY, USA, 1995. ACM.