3.1 Publishable summary

The second year of the GROWTHCOM project has witnessed significant advances in all WPs. The consortium focused on different yet connected topics, ranging from growth and industrialization of countries' economies to technological forecasting and scientific competitiveness, with the final goal to obtain a consistent and systematic view of the global productive ecosystem. We also considered social e-commerce and other on-line systems, which represent a growing sector of modern economies and are attracting an increasing attention from the media, by developing new tools to measure reputation and make recommendations. All these activities have been fuelled by extensive data collection and sharing throughout the consortium. Last but not least, we paid special attention to the dissemination of our results, by means (i) of the organization of a successful summer school, (ii) fruitful interactions with stakeholders and policymakers, (iii) scientific publications, (iv) talks and seminars in conferences, workshops and in scientific institutions. A central role in the dissemination is played by the Growthcom Project web-site www.growthcom.eu, the Facebook and Twitter accounts. In this informatics interfaces we have continuously informed the followers and the public at large with the main news, events and scientific results of our Consortium.

For the final year of the project we plan to have many activities to converge to shared models and results. A crucial infrastructure to deal with this phase will be the shared data platform, developed in WP4, which will allow all the teams to work on the same models and datasets as well as disseminate the consortium's results. The teams in charge of WP1 and WP2 have started to cooperate to compare the works about the Economic Complexity metrics and the trophic structure of countries economies. The same teams have also started to explore the relationships between the Moore's and Wright's laws and the structure of the products network. The important advances described in WP1 and WP3 about link prediction in bipartite networks will be a synergism to the study of the dynamics of the various bipartite networks of interest to the project's activities.

Here we describe the efforts of the various teams in the second year of the project, and in the following section we provide a short summary of the achieved results.

Work Performed

New applications in terms of predictive instruments and methodologies, based on the framework of economic complexity, were considered. Starting from a quantitative, non-monetary metrics for country competitiveness and product complexity, we aimed at providing reliable forecasts on economic systems at the scales of national economies, industrial sectors, and single-company level. Due to the strongly non-linear iterative procedures that define our metrics, the quality of data is of crucial importance. Data collected and validated were obtained from the COMTRADE world trade network dataset and from the e-commerce data at regional level.

Together with these novel datasets, new models, analytical and numerical techniques were developed. In particular, we improved the predictive scheme of economic complexity for assessing the growth potential of countries, and used it to better portray poverty traps and middle-income traps. Besides, we defined and built a hierarchical directed product space, that allows on one hand for strategic planning of countries development, and on the other to make predictions on it. We faced the latter issue also by more complicated techniques, such as the definition of an assist matrix of causal relations between products, and the use of machine learning algorithms.

Much work was devoted to extending the economic complexity approach to bipartite networks other than country-product. We focused on networks of scientific production of countries, and of firms and patents, and, when needed, proposed generalized versions of the algorithm. Importantly, we tackled the fundamental problem of the robustness of the economic complexity metrics with respect to input data, considering both the shape of real matrices and the presence of noisy data.

We tackled the issue of the predictability of technological progress and the evolution of technological ecosystems. Clearly, there are some exploitable patterns in time series of technological progress, and modeling the economy as an evolving complex system of interacting sectors and technologies provides a better understanding of economic dynamics rather than static, aggregated models and allows to make testable predictions.

The work based on these results delivered a number of promising insights. Forecasting based on explanatory variables such as cumulative production seems useful, and natural experiments allow us to evaluate endogeneity biases. These robust relationships can be used in theoretical exercises. We found that optimal investment in learning curve technologies is not necessarily specialized under risk averse preferences. These milestones constitute, in our view, a good background to tackle the challenging research problems ahead, such as forecasting the joint evolution of economic and technological systems and designing the optimal public intervention in this context.
We went further in analysing the "Metrics for reputation and quality in economic systems", where progress was achieved along two main lines, i.e., the role of time in growing complex systems and the reputation in multi-level economic systems. In the former, we have continued our research of growing information networks and we noticed that when the model parameters for the ageing of activity and relevance are varied, PageRank — a seminal node reputation metric for directed networks — can develop various sorts of time biases (e.g., either the recent or the old nodes are preferred). Such biases are detrimental to the algorithm's performance and we show that PageRank is outperformed by the simple in-degree metric in a variety of parameter settings. Therefore, new algorithms that do not suffer from time bias were developed that could be relevant in a broad range of areas where PageRank has established its applications. Another research line focused directly on reputation systems for multi-level (economic) networks. To this end, we studied metrics to measure the influence of users on the behavior of their friends. The multi-level network is composed of a mono-partite social network and a bipartite user-item network (here item is a generic term which can refer to specific objects or other entities). We demonstrated that the currently existing metrics of friends' influence are biased by the presence of highly popular items and users in the data, and as a result can lead to an illusion of friends influence where there is none. In turn, we developed three metrics that do not suffer from the popularity bias and evaluated them on real and artificial datasets. The datasets generated by an artificial model have proven successful in reproducing various patterns observed in real data and can become instrumental in the future research of social influence and its role in system evolution.

Results
We extended the economic complexity metrics to the bipartite network of countries and scientific areas, where links indicate the impact of the scientific articles produced by countries in the various scientific sectors. Such extension is straightforward, and is justified by the triangular shape (or high nestedness) of the network's adjacency matrix. We find that scientific leading countries do not specialise in a few scientific domains, rather, they diversify as much as possible their research system. On the other side, less developed countries are competitive only in scientific domains where also many other countries are present. Diversification thus represents the key element that correlates with scientific competitiveness, as for industrial production. Thus, in its original formulation, the economic complexity algorithm allows in this case to characterize the scientific fitness of each country (i.e., the competitiveness of its research system) and the complexity of each scientific area.
Moreover, we have conceived two generalizations of the economic complexity algorithm for bipartite networks of firms and patents. The first one simply considers the possibility of having a weighting parameter in the standard mathematical expression of one of the two coupled equation. By varying this parameter one can, for instance, increase the linearity of the algorithm, or make high complexity technologies weight more in the fitness computation. This parameter is also important with respect to the convergence of the algorithm. The second generalization instead takes

into account explicitly the distance among technologies. In this way, we can reward coherent companies, i.e., those companies that choose to focus in a well-defined, strategically selected basket of technologies. We are applying this new algorithm to the bipartite network composed by companies and technologies (patent data), with promising preliminary results.
We obtained forecasts of technological progress based on Moore's law, to demonstrate the extent of predictability and have a strong benchmark for future work. This is a genuine milestone in the sense that all future work, by us or other researchers, now faces a clear benchmark showing what is possible using the simplest method. From a statistical point of view, it provides a null hypothesis or null model that other methods shall be able to overcome to show their superiority. We used this benchmark to check whether doing forecasts using cumulative production results in better predictions. Moore's law in- sample analysis has also been very useful when applied to the World War II data, which contains many missing values and is easier to interpret if we make certain assumptions that we believe are valid, based on Moore's law results. Moreover, forecasting using cumulative production has provided the necessary inputs to the work on portfolio optimization, in particular the functional form of the learning curve and basic values of the parameters. The first results on portfolio optimization are promising and will be exploited in the near future.
One of the drawbacks of our previously devised user surprisal metric, which measures a user's ability to be among the first ones to collect ("discover") items that eventually become very popular, is that it requires two parameters to be computed. While there is some simple general intuition behind the choice (while it is generally preferred to set them as low as possible, a too low choice can lead to little statistics and unreliable results), a founded approach was lacking. We have recently made important steps to understand how to choose the parameters that enter its computation (robustness of the user ranking by surprisal is the main criterion).
The aforementioned user surprisal metric remains an important component of the future reputation model for economic networks. We currently explore the use of this metric in a number of settings (at the moment, our main focus is prediction of item popularity from the computer science point of view and bibliometrics). Since social influence, in particular the ability of an individual to influence the others, is an important component of socio-economic systems, newly developed metrics for social influence in a multi-level network contribute importantly to the final reputation system. One of the newly suggested social influence metrics, spreading concentration, highlights the items that are not particularly popular, yet they spread strongly in communities of users. This metric has the potential to evolve into a specific reputation metric that reflects the topological aspects of an item's spreading patterns rather than the item's popularity. Finally, the continually developed and improved model of the growth of information networks remains a crucial test-bench for evaluation and development of various relevant algorithms. Our results on the presence of time bias in some algorithms and very recent first attempts at removing this bias (manuscript in preparation) establish solid algorithmic foundations for the work on reputation in economic networks.
On the dataset side, we have collected many kinds of data, including customer-product relation, user-product interaction, shop-keyword, shop-keyword- time, shop-category, which now can fully support the research of economy complexity, enterprise evolution and the bipartite network analysis. Besides the Alibaba commercial data (which were dealt with during the first year), we consolidated the world trade data (COMTRADE) and the social-event diffusion data at a micro-level. Based on all the distributed data, we built a unified data platform to consolidate those data-sources with very restrict privacy rules. In this platform, we have consolidated all the data (Alibaba, COMTRADE, BACI, etc.).

Finally a strong effort has been developed in the direction (i) of dissemination and communication of the scientific results reached during the project research activity, and (ii) of the interaction with policy-makers, political economical and financial institutions and other stakeholders. About the first point (i) (see report on WP6) the researchers of the Consortium have realized during this 2nd year of activity around 40 publications on scientific journals, and given around 70 talks and seminars in public conferences, workshops and scientific institutions. Moreover a big effort has been displayed in the organization of an extremely successful International Summer School on Economic Complex Systems in Lipari (Italy) on September 2015 (http://lipari.cs.unict.it/LipariSchool/SECS/). It has seen the participation of around 30 people with 20 among students and young researcher, while both invited speakers from important international scientific institutions and Growthcom project investigators have given classes. Finally, a central role in the dissemination and communication of results, news and events has been played by the project webpage (www.growthcom.eu), the Facebook and Twitter Growthcom account (https://www.facebook.com/fp7growthcom/?ref=hl - https://twitter.com/fp7_growthcom).

About the second point (ii) the main investigators of the Consortium have built, during the two years of project activity, interactions and collaborations with the following important public and private institutions: **INET (Institute for New Economic Thinking), Stiglitz Task Force on Industrialization, UK government, Royal Dutch Shell (NL), Azimut Group Investment Fund, Alibaba group (Hangzhou, China), Italian Ministry of Foreign Affairs (MAE), APG Dutch Wealth Fund, GIC Singapore fund, Blackrock fund, Boston Consulting Group (BCG), Templeton Foundation.**