Brussels, 30 October 2015

COST 041/15

DECISION

Subject: **Memorandum of Understanding for the implementation of the COST Action "European Cooperation for Statistics of Network Data Science" (COSTNET) CA15109**

The COST Member Countries and/or the COST Cooperating State will find attached the Memorandum of Understanding for the COST Action European Cooperation for Statistics of Network Data Science approved by the Committee of Senior Officials through written procedure on 30 October 2015.

COST Association
Avenue Louise 149 I 1050 Brussels, Belgium
t: +32 (0)2 533 3800 I f: +32 (0)2 533 3890
office@cost.eu I www.cost.eu

MEMORANDUM OF UNDERSTANDING

For the implementation of a COST Action designated as

COST Action CA15109
EUROPEAN COOPERATION FOR STATISTICS OF NETWORK DATA SCIENCE (COSTNET)

The COST Member Countries and/or the COST Cooperating State, accepting the present Memorandum of Understanding (MoU) wish to undertake joint activities of mutual interest and declare their common intention to participate in the COST Action (the Action), referred to above and described in the Technical Annex of this MoU.

The Action will be carried out in accordance with the set of COST Implementation Rules approved by the Committee of Senior Officials (CSO), or any new document amending or replacing them:

a. "Rules for Participation in and Implementation of COST Activities" (COST 132/14);
b. "COST Action Proposal Submission, Evaluation, Selection and Approval" (COST 133/14);
c. "COST Action Management, Monitoring and Final Assessment" (COST 134/14);
d. "COST International Cooperation and Specific Organisations Participation" (COST 135/14).

The main aim and objective of the Action is to critically assess commonalities and opportunities for cross-fertilization of statistical network models in various applications, such as economics, sociology, epidemiology, ecology and biology, with particular attention to scalability in the face of Big Data, while creating a broad and inclusive research community. This will be achieved through the specific objectives detailed in the Technical Annex.

The economic dimension of the activities carried out under the Action has been estimated, on the basis of information available during the planning of the Action, at EUR 56 million in 2015.

The MoU will enter into force once at least five (5) COST Member Countries and/or COST Cooperating State have accepted it, and the corresponding Management Committee Members have been appointed, as described in the CSO Decision COST 134/14.

The COST Action will start from the date of the first Management Committee meeting and shall be implemented for a period of four (4) years, unless an extension is approved by the CSO following the procedure described in the CSO Decision COST 134/14.

COST is supported by the EU Framework Programme Horizon 2020
COST Association, International not-for-profit organisation/Association international sans but lucrative
Register of legal Entities Brussels: 0829090573

COST Association
Avenue Louise 149 I 1050 Brussels, Belgium
t: +32 (0)2 533 3800 I f: +32 (0)2 533 3890
office@cost.eu I www.cost.eu

TECHNICAL ANNEX

OVERVIEW

Summary

A major challenge in many modern economic, epidemiological, ecological and biological questions is to understand the randomness in the network structure of the entities they study: for example, the SARS epidemic showed how preventing epidemics relies on a keen understanding of random interactions in social networks, whereas progress in curing complex diseases is aided by a robust data-driven network approach to biology.

Although analysis of data on networks goes back to at least the 1930s, the importance of statistical network modelling for many areas of substantial science has only been recognised in the past decade. The USA is at the forefront of institutionalizing this field of science through various interdisciplinary projects and networks. Also in Europe there are excellent statistical network scientists, but until now cross-disciplinary collaboration has been slow.

This Action aims to facilitate interaction and collaboration between diverse groups of statistical network modellers, establishing a large and vibrant interconnected and inclusive community of network scientists. The aim of this interdisciplinary Action is two-fold. On the scientific level, the aim is to critically assess commonalities and opportunities for cross-fertilization of statistical network models in various applications, with a particular attention to scalability in the face of Big Data. On a meta-level, the aim is to create a broad community which includes researchers across the whole of Europe and at every stage in their scientific career and to facilitate contact with stakeholders.

Areas of Expertise Relevant for the Action	Keywords
● Mathematics: Statistics ● Computer and Information Sciences: Machine learning algorithms	● data science ● networks ● statistics ● big data ● stochastic modelling

Specific Objectives

To achieve the main objective described in this MoU, the following specific objectives shall be accomplished:

Research Coordination

- to initiate novel transdisciplinary research links within academia and with industrial stakeholders with special attention to Early Career investigators, gender balance, and COST Inclusiveness Target Countries.
- to communicate existing approaches from one statistical network application field to another.
- to develop novel theory and methodology in important fields of network applications.
- to seed future research projects to channel efforts and maximize efficiency.
- to identify pressing research needs in statistical network data analysis, modelling and inference.

Capacity Building

- to aid and stimulate Early Career Investigators in the field of Statistical Network Science.
- to achieve geographical and demographical diversity with special attention to gender balance and COST Inclusiveness Target Countries, throughout the Action.

COST is supported by the EU Framework Programme Horizon 2020

COST Association, International not-for-profit organisation/Association international sans but lucrative Register of legal Entities Brussels: 0829090573

COST Association
Avenue Louise 149 I 1050 Brussels, Belgium
t: +32 (0)2 533 3800 I f: +32 (0)2 533 3890
office@cost.eu I www.cost.eu

● to complement the Action’s physical meetings with innovative online initiatives to stimulate more inclusive and continuous collaborative ties between Action participants.

COST is supported by the EU Framework Programme Horizon 2020
COST Association, International not-for-profit organisation/Association international sans but lucrative
Register of legal Entities Brussels: 0829090573

COST Association
Avenue Louise 149 I 1050 Brussels, Belgium
t: +32 (0)2 533 3800 I f: +32 (0)2 533 3890
office@cost.eu I www.cost.eu

DESCRIPTION OF THE COST ACTION

1. S&T EXCELLENCE

1.1. Challenge

1.1.1. Description of the Challenge (Main Aim)

A major challenge in many modern economic, epidemiological, ecological and biological questions is to understand the randomness in the network structure of the entities they study: for example, the SARS epidemic showed how preventing epidemics relies on a keen understanding of random interactions in social networks, the financial crises in 2008 arose from structural changes in the financial lending network induced by the subprime mortgage collapse, whereas progress in curing complex diseases is aided by a robust data-driven network approach to biology.

This Action aims to facilitate interaction and collaboration between diverse groups of statistical network modellers, establishing a large and vibrant interconnected and inclusive community of network scientists. The aim of this interdisciplinary Action is two-fold. On the scientific level, the aim is to critically assess commonalities and opportunities for cross-fertilization of statistical network models in various applications, with particular attention to scalability in the face of Big Data. On a meta-level, the aim is to create a broad community which includes researchers from across the whole of Europe and at every stage in their scientific career and to facilitate contact with stakeholders.

1.1.2. Relevance and timeliness

Networks are an important modelling paradigm for many current questions, such as (i) infectious disease spreading, (ii) financial risk in the global financial network, (iii) genetic disease and network pharmacology networks, (iv) gossip and other viral events in big social networks and (v) cost of queues in traffic networks. The European economy and society could benefit enormously from new insights in these problems. This Action brings together statistical experts from various network modelling communities to borrow from each other’s expertise to develop these new insights.

Within Europe various groups of scientists are at the global frontier of statistical network science. A number of framework programmes and other national instruments support these talented scientists individually. Despite differences in terminology, statistical network analysis is at the heart of many of these initiatives. Individual countries have recognized the need for collaborative projects, such as the upcoming Statistical Network Analysis programme in the Newton Institute in 2016 and the Dutch National Science Foundation 10-year gravitation programme on networks. However, Europe is significantly lagging behind the USA in acknowledging the need for collaboration across the field. In the US, various places (e.g. Santa Fe Institute, Center for Complex Network Research at Northeastern University, Los Alamos, Kansas, Michigan, Indiana, Notre Dame, The Northwestern Institute on Complex Systems at Northwestern University) have set up network science units, encouraging novel interactions and connecting in a more relevant way to the needs of industry and other partners than the traditional departments. This novel scientific unit of activity is reflected in annual meetings, such as the large IEEE Network Science conferences. In Europe there is no equivalent integrated structure of network research or collaboration.

Recent technology makes it possible to support collaboration of communities which meet in person only a few times per year through online seminars, blogs and user groups. This can result in a vibrant environment that would be easily accessible across Europe and would overcome barriers which are created through geographical location or practical restrictions regarding travel, as these may affect researchers with caring responsibilities.

1.2. Specific Objectives

1.2.1. Research Coordination Objectives

This Action entices scientists to cross national and disciplinary boundaries in order to work on a common set of methodological questions. The objectives are

(i) to initiate novel transdisciplinary research links within academia and with industrial stakeholders, with special attention to Early Career Investigators, gender balance and COST Inclusiveness Target Countries;
(ii) to communicate existing approaches from one statistical network application field to another;
(iii) to develop novel theory and methodology in important fields of network applications;
(iv) to seed future research projects to channel efforts and maximize efficiency;
(v) to identify pressing research needs in statistical network analysis, modelling and inference.

Such collaborations would have practical benefits in the short and medium term in industry, government and academia. Some stakeholders are directly involved as Network Proposers.

1.2.2. Capacity-building Objectives

The Action network aims to build capacity by focussing on various groups of researchers, as follows:

1. To aid and stimulate Early Career Investigators in the field of Statistical Network Science
2. To achieve geographical and demographical diversity with special attention to gender balance and COST Inclusiveness Target Countries, throughout the Action.
3. To complement the Action's physical meetings with innovative online initiatives to stimulate more inclusive and continuous collaborative ties between Action participants.

This capacity building will be carried out by creating transdisciplinary and transnational focus groups whose aim is to work on and secure funding for:

- a) (Big) data in network data science;
- b) Network Modelling;
- c) Network Inference and Prediction.

1.3. Progress beyond the state-of-the-art and Innovation Potential

1.3.1. Description of the state-of-the-art

(Big) data in network data science. Sampling network data can be a hazardous task. For "small" networks often somewhat arbitrary boundaries have to be drawn in order to define a network object that does not have edges to vertices outside the network. For larger networks, sampling network data often means getting an incomplete view of the underlying network. Both approaches have their advantages and drawbacks. Whereas ignoring parts outside the network allows one to apply standard network models, its generalizability may be limited. On the other hand, sampling vertices of the full network would allow proper generalization, but inference in the face of a large number of missing edges is rather challenging.

Just as in the famous poem "The Blind Men and the Elephant", indirect sampling, which research in genomics, sociology, as much as finance relies on, brings an additional challenge to network modelling. In sociology, questionnaires provide only a proxy of the underlying social network; in finance, the reliance on particular financial summaries show a similar partial picture of reality. The integration of disparate data sources (often across vastly different scales) poses additional

COST is supported by the EU Framework Programme Horizon 2020
COST Association
Avenue Louise 149 I 1050 Brussels, Belgium
t: +32 (0)2 533 3800 I f: +32 (0)2 533 3890
office@cost.eu I www.cost.eu

challenges in robust inference, model construction and validation. In large scale homogeneous networks, deterministic flow models have been fairly successful. However, at all scales, fluctuations and indeed bifurcations arising from stochastic effects and from heterogeneities are not completely understood. Moreover, understanding the effects of variability (across different measurements, in time, and across different individuals) on the dynamics and output of the system poses a real challenge.

Network data are complex and getting a feel for the data before starting the modelling process requires visualization. Until recently, most network visualizations relied on static representations of a graph. However, a network topology is rarely represented well in two dimensions. Novel approaches rely on dynamic three dimensional representations, which improve the inspection of a network. However, for very large networks, visualization remains a challenge.

Network modelling. Currently there are a number of network models available, with little exchange between communities. In probability theory, there is an increasing interest in studying the behaviour of complex networks, but the interface with the statistics community is usually very limited, and hence issues such as efficient parameter estimation and model simulation are not yet studied in great detail except in a few models. The oldest area for data-driven network modelling is social network analysis, which goes back to the 1960's. This model uses exogenous as well as endogenous covariates to explain the joint existence of all the edges in the network. While intuitively appealing, there are instabilities in the estimation of parameters, and even today the computational effort in fitting these models is so intensive that only networks up to a few thousand nodes can be examined.

The most basic class of models, Bernoulli random graphs, assumes independence between edges and the same probability for any edge to occur. Under this model, any two friends of an agent are no more likely to be friends themselves than two complete strangers would be. This assumption is often not realistic, and while this type of model can serve as a null model, there is not much interest in it from the point of view of data analysis. From probability theory, the limiting behaviour of dense or moderately dense graphs can be described by so-called graphons. There are now some ideas available for estimating such graphons from data, for example using an analogue of histograms. Still a number of open questions arise; foremost, the statistical interpretation of the graphon limit is not clear. Moreover, many real-world networks are sparse.

In so-called stochastic block models it is assumed that the nodes are of differing types, the edges in the network are independent but their probabilities depend on the type of the nodes in the edge. This model has been shown to fit well to networks such as metabolic networks where each node has approximately the same degree. When there is a substantial variation in the degrees, which is the case in many real networks, the model can be distorted. It has been proposed to include additional parameters in the model which reflect the degrees. In this case the number of parameters increases with the number of nodes, so statistical estimation issues arise, as well as the danger of over-fitting the model so that the predictive power is very low. Stochastic block models should also link in naturally with detecting communities in networks, but interpreting the classes as communities is often not straightforward. While community detection is one of the main problems in network analysis, there currently are not many models available which are easy to interpret in this setting. This lack of models generates a pressing question about assessing significance of communities which are suggested by outputs of algorithms.

Network Inference and Prediction. From the mathematical side, models for temporal evolution of networks and models for correlated networks are available, and there is a growing literature on queuing networks, but the statistical aspects of these models are under-explored. Processes on networks such as traffic and the spread of disease have been studied analytically, but usually not from a statistical viewpoint. In statistics, graphical models have emerged as a good paradigm for

COST is supported by
the EU Framework Programme
Horizon 2020
COST Association
Avenue Louise 149 I 1050 Brussels, Belgium
t: +32 (0)2 533 3800 I f: +32 (0)2 533 3890
office@cost.eu I www.cost.eu

inference in the presence of dependence, and there are obvious links to social network models. Exploiting these links is in its infancy. In addition, methods for making predictions from the inferred network and for visualizing the inferred network and the uncertainty in the estimation and predictions, are under-explored for large networks.

Given the lack of information interchange and the complexity of networks, different applications often lend themselves to different network models, such as duplication-divergence models for protein-protein interaction networks. Parameter estimation in these models often relies on Markov Chain Monte Carlo (MCMC) methods, which are computer-intensive, or on method-of-moment approximations which are not yet fully understood. A useful option for assessing statistical significance is through Monte Carlo tests which are based on simulating from the underlying hypothesised model. However, there is a lack of efficient algorithms for simulating from large complex network models.

Network inference involves fitting mathematical models to observed network data. The type of inference method used – Bayesian, likelihood-based, generalised estimating equations – depends highly on local expertise within the disciplines. Within each of these disciplines, most of the efforts in recent years have been on the development of efficient methods for parameter estimation and model selection, with significant progress being made on the development of sparse inference approaches and efficient MCMC-based schemes. At present the developed methods can still only hold networks of limited complexity, as most inferential schemes do not scale better than quadratic in the number of nodes and linear in the number of samples.

1.3.2. Progress beyond the state-of-the-art

COLLABORATIVE THEME 1. EXPLORING (MASSIVE) NETWORK DATA SETS
Experimental science, evidence-based medicine and effective policies are based on data. Very large network data sets often exceed the limits of standard data processing and statistical techniques and hence sub-sampling approaches are promising. Sampling network data, however, is tricky. The Action aims to develop and share effective data collection strategies.

a. Sampling networks
Many network data objects are large, unwieldy files, that are often stored in distributed form in a distributed database. Standard statistical methods require access to the whole object, typically in working memory, to do inference. For massive network data, this is fundamentally impossible. Novel data storage approaches need to take scalability of the procedures into account at a very basal level, whereby preprocessing network data into smaller collections of sufficient statistics would be an essential step. Deciding on which features of the network are essential is ongoing work; while some features may strongly depend on the network in question, there are some universally used summaries, such as the density, which can help compare different data sets and guide the choice of tools for the network analysis.

b. Sampling and intervention
When faced with potentially dangerous infectious diseases, such as the recent SARS outbreak, the Asian bird flu and the H1N1 epidemic in 2009, a successful vaccination scheme should target potential highly infectious sources, e.g., the highly connected nodes in the social network. However, in open, dynamic societies, information on such nodes is not readily available and efficient realistic protocols have to be based on local network information. The statistical friendship paradox states that in a typical network situation one's friends have on average more friends than oneself. Successful vaccination schemes have used this idea by vaccinating randomly selected contacts of randomly selected people. This counter-intuitive idea is an important example of the strange reality behind networks. It shows that thinking of infectious diseases in terms of a spatial process in

Euclidean space is misleading. This Action will work on a variety of novel sampling strategies in the context of infectious diseases.

COLLABORATIVE THEME 2. NETWORK MODELLING

There are several network models commonly used throughout network science. Cross-comparison and joint development will boost scientific activity. Underlying issues are the efficient estimation of parameters, the identification of important microscopic and mesoscopic structures in networks, and scaleable simulation methods.This Action will combine expertise from the behaviour of network models where agents are interchangeable with models where agents have specific characteristics to address the statistical interpretation of the long-term behaviour as well as to propose a unifying framework for models which includes tools for parameter estimation and for simulations. Special emphasis will be given to scalable models. Network models can be used to guide sampling approaches from networks by ensuring that the sampling scheme captures much of the dependence in the data. They are also used to guide statistical inference through limit theorems and simulation constructions. Specifically, the Action will address the following questions.

a. Models for flows on networks

This Action will address the effect of hub-like or scale-free properties of networks on flows on networks such as traffic flow. Modelling first percolation times will give important insight in how fast traffic spreads though the network. First passage percolation on random and complete graphs with various edge weight distributions has attracted substantial attention.

Infectious disease spreading can be viewed as a diffusion process along the edges of a network formed by the relevant interactions between individuals, as, for example, sexual and proximity contacts. At larger geographical scales, it can also be viewed as a diffusion along the network of traveling individuals connecting different locations where host mix and the transmission occur. Therefore the evolution dynamics of an epidemic and its impact on the population is significantly affected by the network features such as its topology, the distribution of weights and its temporal evolution. The mathematical study of network theory provides a framework to address this complex interplay. Studies of branching tree phenomena and reaction-diffusion processes on networks have been adapted to the modelling of disease spreading. Network statistical characterization and node centrality ranking provide understandings on the impact of network structure on the epidemic dynamics. First-passage percolation is intrinsically linked to susceptible-infected-susceptible epidemic models and hence there will be some cross-fertilisation in the modelling as well as parameter estimation.

b. Graphical models

Probabilistic graphical models combine a rigorous formulation of their mathematical and computational properties with the immediateness of a graphical representation that enables effective communication between scientists from different fields. The study of polypathology in HIV positive patients naturally requires the analysis of the association structure of several variables concerning different pathologies but also background variables related to the use of drugs, gender, sex habits, age at infection, CD4 related measures etc. Graphical models are a natural modelling tool for such complex intercorrelations.

If one examines the interaction structure between the most important international stock market returns, one ought to discriminate between direct and induced correlations between them. Graphical models encode for conditional independence and can be visualised by a graph, where vertices represent the stocks and the dependence among them are visualised by edges. Graphical models applied to financial markets can be used to investigate the information flow between the major international financial markets. Graphical models are linked with social network models, and this link will be exploited in the proposed project to transfer ideas between the two areas, with influence on modelling flow processes on networks.

c. Simulation

The network perspective also underlies sophisticated computational models such as epidemic models. Empirical network data, such as social contacts and mobility patterns, allow for estimation of the parameters in the disease diffusion. By simulating epidemic events in realistic scenarios one can provide projections on the unfolding epidemic and assessments of the efficacy of intervention strategies. In traffic modelling, the aim is often to devise infrastructure network adjustments to increase traffic flow. Also here, there are typically good estimates available for various flow parameters based on current traffic patterns. In both scenarios, the Action will develop tests for intervention strategies in massive parallel simulations of the network process. An intermediate research step are efficient implementation tools of network structures and dynamics.

d. Beyond dyadic networks

Most networks are represented as dyadic relationships, visualised as an edge between pairs of nodes. In contrast, many financial, biological and social mechanisms involve more than two participants, requiring at least three. For example, in 2008 loans collapsed simultaneously. An example of a three-way social processes is gossip, where two people choose a third person, an “object”, to talk about. These mechanisms rely on so-called triadic, three-way, data structures, involving three actors, for instance the ‘sender’ and ‘receiver’ of gossip and the ‘object’ of their gossip. This Action will explore new stochastic models and statistical methods for analysing such triadic or higher-order interactions.

COLLABORATIVE THEME 3. NETWORK INFERENCE & PREDICTION

The type of inference method used – Bayesian, likelihood-based, generalised estimating equations – depends highly on local expertise within the disciplines. This collaborative theme aims to share statistical competences and focus on the most effective method given the particular situation. One of the central motivations of network modelling is the complex task of prediction. Machine learners, balancing the “bias-variance trade-off,” have started to develop robust and general-use prediction algorithms. If these algorithms can be extended to the field of statistical network models, it would serve an important practical use in many applied fields.

a. Likelihood-free methods

Numerous important developments in statistical computing have occurred recently which rely in some way on "likelihood-free" techniques. These methods all overcome the mathematical intractability of a complex stochastic or statistical model by using computer simulation. Approximate Bayesian Computation (ABC) is perhaps the most well-known application of the likelihood-free approach, but it also underlies other methods, including special cases of particle MCMC (pMCMC). Many problems involving stochastic networks are mathematically intractable, yet amenable to computer simulation, and hence natural targets of likelihood-free methods.This part of the theme has strong connections with the simulation aspect of the networks modelling theme.

b. Sparse, structured inference

Most natural, man-made and social networks have some aesthetic quality, which relates to the fact that there are not too many links, that most nodes have very few links and only a few nodes have a lot. This possibly relates to an ontological principle that most real-life networks are sparse and have some sort of small-world property, or to an epistemological principle that our minds are only capable of considering such networks as relevant explanations for reality. Whatever the motivation, this has led to a branch of statistics called sparse inference. The aim is to find the most parsimonious model that still explains the data appropriately. The Action aims to extend such approaches to network model inference.

COST is supported by
the EU Framework Programme
Horizon 2020

COST Association
Avenue Louise 149 I 1050 Brussels, Belgium
t: +32 (0)2 533 3800 I f: +32 (0)2 533 3890
office@cost.eu I www.cost.eu

c. Predicting network flow
Infectious diseases spread among a heterogeneous contact network. The effect of drugs spread through the genomic network. The effect of a traffic accident will spread through a mobility network. Huge amount of data on host behaviour have recently become available. For example, networks, which trace the activities and interactions of hosts, social patterns, transportation fluxes and population movements on a local and global scale, are being measured. An important objective in such network is to predict robustly the effect of some event in the network on the rest of the network.This part of the theme is strongly connected to the aspect of modelling network flow in the networks modelling collaborative theme.

1.3.3. Innovation in tackling the challenge

There are several key innovative aspects in tackling the scientific challenge. We deal with them in a pointwise manner:

- *Massive influx of data:* Network can be massive. Most modellers have so far shied away from such challenges. By combining the efforts of experts in various fields, the Action aims to start dealing with large scale network problems.
- *Uniform approach:* By starting a constructive discussion among network modellers throughout Europe, the Action will make a start with providing a unifying approach to network data, modelling and analysis. This unified approach will involve creating benchmark network data sets, which can help to galvanize the community to provide and improve standards. (GANTT 4a)
- *Improving visibility:* The relative invisibility of statistical network scientists is partly the result of the scatter of the field. The Action will itself be a novel focus point for network expertise, which means that 3rd parties, whether it be industry, government, academia or other parts of society, will be able to find the experts.
- *Integrated toolkit:* The Action will build forth on existing software to develop an integrated toolkit for network analysis. (GANTT 4b)
- *Training and outreach:*Teaching material will be made available at the COST Action website by recording the lectures at the COST Summer School meetings.

For the meta-level aim of creating a broad community, seminars will be held both live and online (if financial resources suffice) with questions allowed, as already successfully implemented, for example, by the Royal Statistical Society. A website will host all past talks but also online tutorials and teaching resources. Through email lists and blogs, discussions within focus groups can easily be facilitated. A mentoring scheme with online contact through skype or Webex will be able to reach researchers across the whole of Europe. To kick-start this effort, a workshop will be organised in a COST Inclusiveness Target Country within the first year of the Action.

1.4. Added value of networking

1.4.1. In relation to the Challenge

Whereas, e.g., Spain and Italy have strong statistical physics traditions, the UK has a strong statistics background. Transnational collaboration within the field of statistical network science, therefore, is almost a prerequisite for successful implementation of the challenges described in this Action. Moreover, there are traditionally strong subject boundaries within Europe, which the individual scientists find difficult to overcome. Statistical network applications come from biology, sociology, economics, engineering and epidemiology, none of which traditionally are even present within the same faculty or university sub-unit.

Statistical network science is an interdisciplinary discipline with socio-economic potential within many fields. The need for educating and training young people in this field will have a high economic pay-off. Also involvement of researchers of COST Inclusiveness Target Countries will have a disproportionally large multiplier effect. Therefore, the Action uses various networking activities to encourage involvement of young researchers and researchers from COST Inclusiveness Target Countries.

1.4.2. In relation to existing efforts at European and/or international level

The field of statistical network science is moving very fast, partially in response to related initiatives, such as Big Data and High-throughput Science, as e.g. in genetics and astronomy. This Action will connect to those challenges by forging strategic links. Through its individual members, the Action already connects to a number of such initiatives and the Action will draw on them to make statistical network science part of these important challenges. Milestones will include participation in other networks of active and recently completed programmes (e.g. 10-ECRP-044 “Social Influence in Dynamic Networks”; ICT-2007.8.0-255987 “FOC, Forecasting Crisis”; ICT-2011.5.6-288501 “CRISIS, Complexity Research Initiative for Systemic Instabilities”; FP7/2007-2013-278433 "Predemics"). Moreover, towards the end of the COST Action, the MC will aim to be involved in formulating new Horizon2020 Programme calls.

The Action will collaborate with initiatives, such as the 2016 Newton Institute programme on Theoretical Foundations for Statistical Network Analysis, to create synergies between existing smaller initiatives and a truly pan-European Action on this important theme.

COST International Partner Country (IPC) interest and involvement in this Action, specifically from important academic leaders in the US, is an important acknowledgment of the Action's potential and stimulant for organizing a strong coherent European presence in the field of statistical network science. While developing cross-European collaborative ties, the Action will strategically involve participants from COST IPCs in various challenges.

2. IMPACT

2.1. Expected Impact

2.1.1. Short-term and long-term scientific, technological, and/or socioeconomic impacts

The Action is expected to attract the interest of industrial stakeholders, in particular those developing statistical network modelling tools to assist decision processes. These analytic tools are used by a broad range of clients offering applications and services in finance, health, agriculture, manufacturing, engineering, communications, energy and forensic science. Software for network analytics allow creating real-world solutions for fraud detection, credit default prediction, operational risk management, medical diagnosis, health monitoring, risk analysis, data mining, troubleshooting, safety assessment, forensic identification and more. Short-term impact of this COST Action is mutual collaboration between this type of stakeholders and Action investigators. Application inspired theoretical development by Action investigators will benefit from further development and improvement of the modelling tools, which are core activities of this type of stakeholders. Medium term impact will be the improvement of complex decision making in a variety of economic contexts. This will have a direct impact on the economy in Europe.

Based on advances in chemical biology and network science, network pharmacology is a distinctive new approach to drug discovery. It involves application of network analysis to determine the set of proteins most critical in any disease, and then chemical biology to identify molecules capable of targeting that set of proteins. By addressing the true complexity of disease and by seeking to harness the ability of drugs to influence many different proteins, network pharmacology differs from conventional drug discovery approaches, which have generally been based on highly specific targeting of a single protein. Network pharmacology has the potential to provide new treatments for complex diseases where conventional approaches have failed to deliver satisfactory therapies. By means of proprietary platforms in network pharmacology, companies can analyse networks of proteins associated with particular diseases. It then identifies drug candidates with optimal impact on these networks. Of particular interest is the use of these platform in the context of cancer and degenerative diseases of the nervous system.

Financial institutions and markets are highly interconnected, but only recently has an expanding literature started to emerge to map these interconnections and to assess their impact on financial risks and returns. Companies dealing with financial network analytics can help financial professionals see a connected picture of risk. Network algorithms reveal interconnected risks and the aim is to create cloud-based solutions for distilling its complexity into visual decision-making signals. Typical clients of this stakeholder include the world's largest central banks, infrastructure providers and leading financial institutions. Continuous development of the network analytic tools is paramount to enable clients to oversee complex financial networks, visualize global market dynamics, identify systemic risks, and much more. The long-term goal of this approach is to avoid repetition of the 2008 financial crash, whereby the problems in the subprime mortgage market in the US spread through and were amplified in the global financial network.

Short-term impact of the Action is training a new generation of scientists in an societal important theme, network analysis, strengthening the cohesion of scientists working on various aspects of statistical network science, and producing new analysis tools for practitioners.

2.2. Measures to Maximise Impact

2.2.1. Plan for involving the most relevant stakeholders

The Action has already attracted a substantial number of stakeholders as part of the proposers of the network. These stakeholders cover a variety of interests spanning the challenges mentioned before, to wit, infectious disease networks, genomic networks, financial networks, social co-evolution networks, semantic networks, complex disease networks and traffic networks. In the first year, outreach activities are planned to attract mutually beneficial stakeholders by visiting other networks and organizing satellite meetings as part of conferences organised by potential stakeholders.

There is a clear expectation on behalf of the stakeholders that development of statistical network science will have a positive impact on the profitability of their activities. On the other hand, these applied demands spark relevant scientific questions. In order to bring stakeholders and academics involved in the Action together, various meetings are planned:

(i) Sandpit meetings, whereby stakeholders present a problem, on which groups of junior and senior academics together will work over two days. (GANTT 2f)

(ii) Speed date meetings, as part of the Action conference, whereby stakeholders and academics talk for 5 minutes in order to get to know each other. (GANTT 2d)

(iii) Short-Term Scientific Missions (STSMs) whereby academics, particularly young researchers, visit the stakeholder for a short on-site collaboration.

2.2.2. Dissemination and/or Exploitation Plan

The Dissemination Coordinator will oversee the dissemination of the Action's results:

- Exploitation: at the first sandpit meeting at the end of the first year, a White Paper will be presented about the needs of s and industry and society in general with respect to statistical network science. This White Paper will form the basis of further collaborations with the stakeholders (GANTT 5a).

- Website: all Action members will have editing access to a Wiki-based website. The WG Leaders have the responsibility to keep their section up to date by encouraging submission of papers and software tools by its members to the Action's website. (GANTT 1a)

- Publications: all academic Action members are expected to publish their results in peer-reviewed scientific journals and in written reports to stakeholders. Members are also encouraged to make pre-print technical reports available on the website (for instance by linking to Arxiv). As a result of the COST Action joint papers between participants are expected. The aim is to get at least one joint paper per participant on average. (GANTT 4c)

- STSM reports: each researcher that makes use of STSM funding will write a STSM report stating what has been achieved during the visit.

- White papers: the MC in collaboration with the WG Leaders will produce three white papers to give an overview of the current and projected activities within the Action. An early Industrial and Societal Needs White Paper and a Final Report will also be part of the overall dissemination strategy of the Action. (GANTT 5b-e)

- Presentations: Action members will present their work to peers, stakeholders and the general audience. A novel way of doing this is through online seminars that will be organized on a regular basis throughout the lifetime of the Action, if funds suffice. (GANTT 1c)

- Outreach activities: outreach to members of COST Inclusiveness Target Countries through online meetings, as well as the first year outreach conference organized in such a country. Moreover, Action members will be involved in related activities, such as conference of various application fields that use network analysis. (GANTT 1c & 2a)

2.3. Potential for Innovation versus Risk Level

2.3.1. Potential for scientific, technological and/or socioeconomic innovation breakthroughs

The Action has identified a number of scientific and technological challenges of various degrees of complexity. The first aim of the Action, namely applying network analysis tools from one field to another, is a low-risk- high-impact strategy, which promises serious gains for many applications. By collaborating closely with collaborators in industry and the applied sciences, the potential for innovation is maximized. Sandpit meetings and industrial speed date meetings will guarantee close communication of the Action participants and the stakeholders.

Beyond the cross-fertilization described above, several serious challenges need to be addressed that are riskier, but that require close collaboration for forging ahead. Especially making analysis strategies scalable for Big Data scenarios requires new ideas. Although the risks of this part of the project are higher, it also has a proportionally large pay-off, when successful.

The aim of the Action is – in collaboration with stakeholders – to pave the way for serious innovation, which may translate in better economical decision making, lower risk of financial crises, novel strategies to deal with infectious disease containment and new drug targets for various diseases. Even partial success in any of these fields will be enormously beneficial for society at large.

The aim of creating a partly online community which includes researchers across the whole of Europe and at every stage in their scientific career could serve as a model for research development across Europe and for research collaborations with considerable geographical spread. This community could also provide an example for supporting researchers with caring responsibilities by providing seminars which do not require any travel. Travel-less seminars will also help reduce the ecological footprint of the Action.

This part of the Action carries the risk of low participation. To counter-act this risk the Action will hold a workshop in a COST Inclusiveness Target Country within the first year of coming into existence. Participation will be monitored closely and, through email campaigns and targeted contact emails, a wide range of researchers will know about the Action within a short period of coming into existence.

3. IMPLEMENTATION

3.1. Description of the Work Plan

3.1.1. Description of Working Groups

There are three Working Groups in this Action. They are responsible for organizing (possibly joint) WG meetings, encouraging interactions in their WG and producing after one and a half year a white paper about the current state and future challenges in their field.

WG1. Exploring (massive) network data sets
This group will deal with sampling data from, of and on networks. How can one determine whether a viral advertising campaign was successful? What is the difference between sampling a gene regulatory network in a transsectional or longitudinal manner and does it matter? How can one sample from highly or poorly connected nodes in an infectious disease network?

WG2. Network Modelling
Percolation models, diffusion models, graphical models, ordinary and stochastic differential equation models and many other models have been proposed to describe networks and their behaviour. The WG will focus on comparing and developing network models for existing and novel applications, such as finance, sociology, epidemiology and biology.

WG3. Network Inference & Prediction
Network inference consists of computationally identifying network model parameters from data and as such it builds forth on the activities of the first two groups. Inference provides explanations and descriptions of phenomena. Network prediction builds forth on inference to deal with diverse questions, such as “What is the best advertising or vaccination strategy?”, “Is this network activity a sign of fraud?” and “Will this drug be effective?”.

Furthermore, depending on the target audience, the WGs are responsible for implementing the following Tasks.

i. Live online seminar series
One limitation of an international network is the reduced face-to-face interactions. New technology makes it possible to organise (live) online research seminars. If sufficient funds are available, the aim is to organise Action-wide internal research seminars throughout the year, whereby members including PhD students, Postdocs and other junior researchers present their research findings in low-key, but international and interdisciplinary setting. (GANTT 1c)

ii. Blogs and Wikis on Action website
Members of the Action are encouraged to store their publications in online, freely-accessible publication archives, such as Arxiv. Also PhD students, Postdocs and other junior researchers can

COST is supported by
the EU Framework Programme
Horizon 2020

COST Association
Avenue Louise 149 I 1050 Brussels, Belgium
t: +32 (0)2 533 3800 I f: +32 (0)2 533 3890
office@cost.eu I www.cost.eu

make use of this medium to “publish” their results and invite feedback. The Action will make a Blogging and/or Wiki tool available on its website to encourage discussion on scientific results. (GANTT 1b)

iii. Summer Schools & online lectures
Early Career Investigators in the Action will have a variety of backgrounds. Some may come from quantitative psychology, others from mathematical finance, statistics or physics, to name a few. In order to deal with this variety of backgrounds, basic skills can be taught by means of online lecture material. The Action's website will contain a variety of lecture materials; some by its own members, but it will also draw on available lecture material online, such as for example a COPASI tutorial on YouTube. Furthermore, the Action will organise Summer Schools, typically as part of an Action conference, with lectures on converging topics in statistical network science. (GANTT 1d & 2c)

iv. ECI meetings
Junior members of the Action will organise Early Career Investigator Meetings, where the aim is to exchange results to one's peers only. (GANTT 2b)

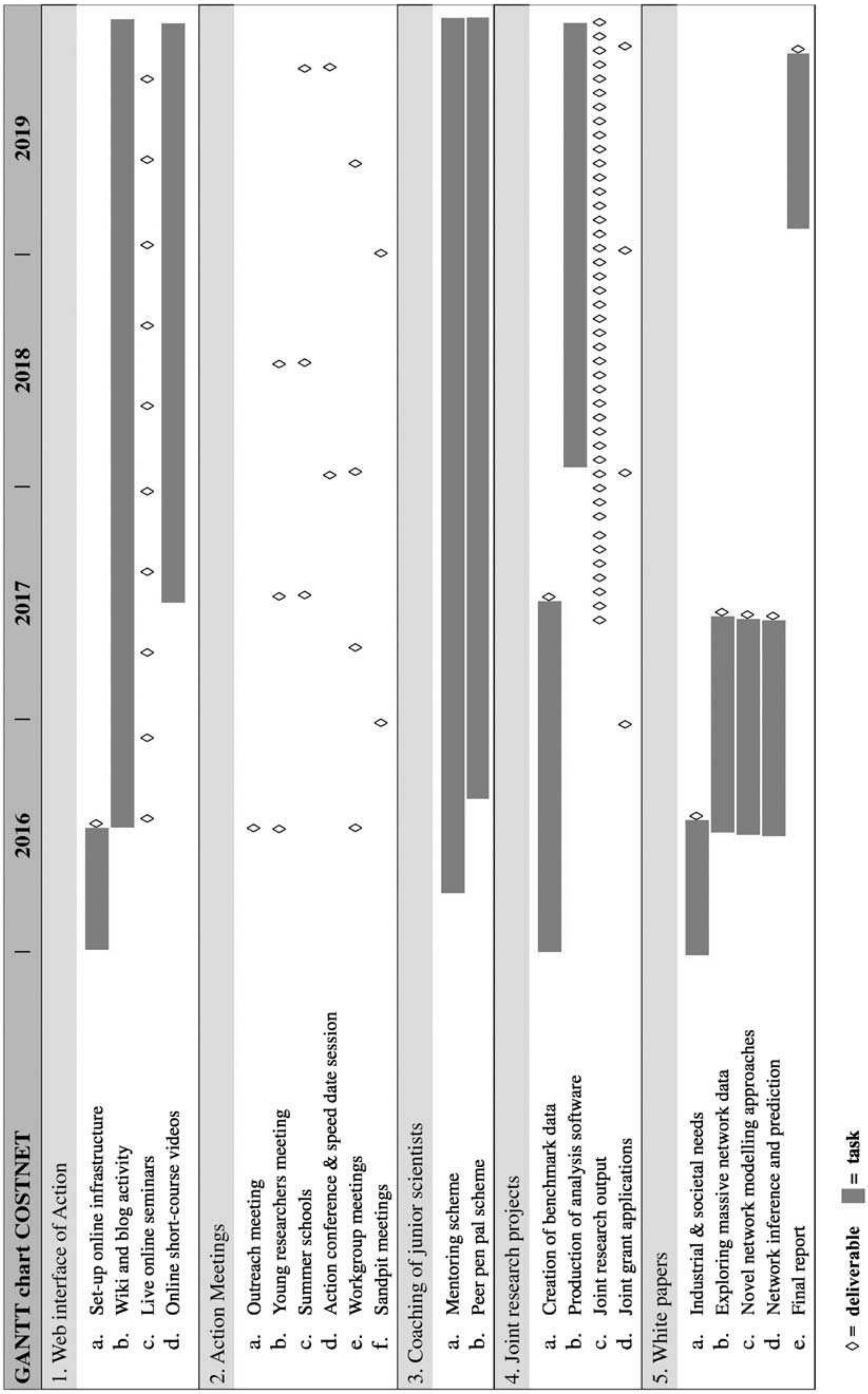
v. Action conference & WG meetings
There will be three Action conferences throughout the duration of the Action. The first one will be a special Outreach meeting, organized in the first year in a COST Inclusiveness Target Country, to encourage researchers from these countries to join the COST Action. At the second meeting the attention will be on the three White Papers produced by the Working Groups, to map a path for European collaboration on important themes in statistical network modelling and inference. At the final conference the Action will reflect on its achievements and the way ahead in a Final Report. (GANTT 2d-e)

vi. Mentoring scheme: joint PhD supervision
This Action will encourage the “adoption” of a colleague's PhD student or Postdoc, where this is deemed valuable for both sides. Transdisciplinary mentoring schemes are felt as being particularly attractive to reinforce the objectives of the Action. Currently, various agreements between Italy, the Netherlands and the UK are in place. Also joint PhD supervision with COST IPC Action member in the US and the Action's stakeholders are considered. (GANTT 3a)

vii. Peer Penpal scheme
Besides “vertical” support and mentoring, the Action will encourage “horizontal” interactions by introducing an online Peer Penpal Scheme. This scheme will link one PhD student or Postdoc to a fellow peer in another institution involved in the Action. The aim of this scheme is to foster a community spirit among Early Career statistical network Investigators and to expose them early on in their career with international contacts. (GANTT 3b)

viii. Platform for academic and industrial collaboration
An important feature of the Action is the involvement of industrial partners. They will be part of the fabric of all collaborations, but there are several specific Tasks to encourage collaboration with academic partners. The sandpit meetings will see teams of PhD students, early career and senior researchers work in teams for a one and a half day on a problem proposed by one of the industrial partners. The aim is to present a novel way at solving a problem, while at the same time (i) adding value to the Stakeholder, (ii) providing training for Early Career Investigators and (iii) forging ties between academia and industry. (GANTT 2f)

3.1.2. GANTT Diagram



COST is supported by
the EU Framework Programme
Horizon 2020

COST Association
Avenue Louise 149 I 1050 Brussels, Belgium
t: +32 (0)2 533 3800 I f: +32 (0)2 533 3890
office@cost.eu I www.cost.eu

3.1.3. Risk and Contingency Plans

The main risks to the collaborative nature of the Action are the following:

(i) Lack of engagement of academic partners and stakeholders.

(ii) Small number of participants from COST Inclusiveness Target Countries, Early Career Investigators and lack of gender balance.

(iii) Little trans-disciplinary collaboration.

The Action has strived very hard to come up with innovative ways to engage partners, to reach out to other parts of Europe and Early Career Investigators, to ensure gender balance and to forge collaborative ties. The Action's point of departure is that one size does not fit all. The Action's management plan foresees a wide range of activities, from organizing young researcher meetings, via an outreach Action meeting in a COST Inclusiveness Target Country to sandpit meetings with industrial stakeholders. In addition, making more innovative use of modern online technologies, such as live online seminars, blogging and wiki functionality, and on demand streaming of lectures, will help the Action to minimize the risk of involvement of the scientific and industrial community. Devolving responsibility to individual Task Coordinators will help to keep the organization at a grass roots level. The partially overlapping nature of various Action instruments will mean that if some Tasks fail to gain traction, they can be substituted by others.

3.2. Management structures and procedures

The organisation of this Action will conform to the COST Rules. The Action will be coordinated by a Management Committee (MC). A Core Group (CG) is elected at the first MC meeting: it will consist of the Action Chair, Vice Chair and Leaders for the Working Groups (WGs). The WG Leaders are researchers who will coordinate the WG tasks and events. They will be responsible for a) the coherence of the scientific work and b) the completion of specific deliverables and milestones. The WG Leaders will report directly to the Management Committee. The CG is supported by three Specific Tasks Coordinators: a STSM Coordinator (administering and monitoring Short-Term Scientific Missions by Action members), a Dissemination Coordinator (responsible for the Action website and monitoring the publication strategy) and an Outreach Coordinator (monitoring the virtual network and outreach and implementation policy). The CG will administer a Gender Mainstreaming plan (monitored by the Vice Chair) and a plan for Early Career Investigators (monitored by the Chair). This COST Action will respect an appropriate gender balance in all its activities and the Management Committee will place this as a standard item on all its MC agendas. The Action will also be committed to considerably involve Early Career Investigators. Functions of Scientific Leaders of WGs and Specific Task Coordinators can be combined.

The CG, supported by Specific Tasks Coordinators, manages the day-to-day operations of the Action, monitors milestones and prepares documents for the annual MC meetings, including a financial plan to adequately share the resources within the Action budget. The Annual MC meetings take place at workshops or conferences organised by the Action. The MC ensures strong interactions between WGs to ensure cross-fertilisation and to stimulate cooperation with related COST Actions and networks. An outreach policy will be developed by the CG in discussion with the MC.

Each WG consists of researchers and stakeholders from various disciplines. It is possible and encouraged that Action members are involved in more than one WG. The MC will invite Action members to the WGs. Each group is chaired by a WG Leader who is responsible for the activities organised by the WG. Each WG organises each year a, possibly joint, S&T Meeting or Workshop to achieve and disseminate the Action's milestones and deliverables.

3.3. Network as a whole

The aim of the Action is to build an inclusive and transdisciplinary collaboration to put statistical network science on a solid footing within Europe. This means that the Action aims for gender equality, an even geographic distribution and a strong presence of Early Career Investigators. The Action will include a wide variety of modellers and analysts with a background in genetics, finance, sociology, neuroscience, logistics and mathematical theory. Its senior scientists are acknowledged experts in their respective fields. Together they create a critical mass to make real headway in solving some of the challenges described in this Action.

The Action ascribes to the importance of gender balance. The Vice Chair of the MC will be responsible for the Gender Mainstreaming Policy of the Action. The Action aims at a 50-50 gender balance in the MC. Gender equality will be monitored throughout the life-time of the Action.

Training, coaching and development of Early Career Investigators is an important objective in this Action. They play an important role in developing statistical network science as a mature scientific field with European prominence. The scale of their involvement in this Action requires innovative tools to serve and engage them as part of the Action. The Action, directly monitored by the Chair, will use a combination of innovative activities, described under 3.1.1., to achieve these goals.

The Action will aim at achieving a good geographic distribution, in particular by involving participants from Inclusiveness Target Countries. To support this aim in the first year a special outreach Action conference is organized in one of these countries to encourage scientists and stakeholders from local institutions and organisations to participate.

COST is supported by
the EU Framework Programme
Horizon 2020

COST Association
Avenue Louise 149 I 1050 Brussels, Belgium
t: +32 (0)2 533 3800 I f: +32 (0)2 533 3890
office@cost.eu I www.cost.eu