Connectome Simulation:

Stability Analysis of the *c. elegans* Connectome

The title must be interesting and convincing and fit to our track-record

Proposal for a FWF Stand-Alone Project

by

Andreas HOLZINGER

Holzinger Group, HCI-KDD, Institute for Medical Informatics, Statistics and Documentation, Medical University Graz, Austria

Graz, September, 15, 2016

Current Date

Requested Funding:

357.987 EUR (3 PhD for the duration of 36 months)

Get actual amounts

OEFOS 2012 Discipline: 102 Computer Science - 033 Data Mining

Keywords: connectomes, c. elegans, evolutionary algorithm, neuron simulation, OpenWorm

Confirm?

UPDATE!

Note: This proposal is original and has not been submitted to any other grant authority.

Ethical Declaration: This proposal does not raise any ethical issues.

Remark: This type of proposal is limited to a total number of 26 pages and the project duration is limited to a max. duration of 36 Months.

Abstract:

This project aims to investigate the degree of accuracy in connectome simulations sacrificed by the use of less biologically realistic neuron models as part of that simulation, particularly in the context of simpler models usually requiring less cost in terms of processing power to run the simulation. The first overall goal is to establish a baseline for the scientific community when deciding on the realism required for a particular study to yield feasible results while attempting to conserve processing power, particularly in the context of simulating connectomes of complex organisms. In order to do this, we will implement a simulation framework tailored to our needs, which specifically means a high degree of modularity to support continually repeating the same study with several different neuron models. Moreover, we need to improve the technical foundation to handle larger, more complex connectomes than currently available, in order to be able to use the same framework in future research on the connectomes of more complex organisms. The other great research goal is to test the resilience of already established neural networks with respect to perturbations, that is random (forced) reconfiguration of the internal connectome structure or environment influences on single components (disturbing forces on neurons and synapses). The main question in this area is to define standard connectome behavior and establish metrics of deviation from it, so we can effectively measure at which point the natural behavior of an organism collapses. This project will employ 3 PhD students; since these students are expected to complete their degrees in three years, this effectively limits our time-frame to 36 months.

Check

Contents

1 Scientific Aspects

1.1 Motivation for our Research

The field of Connectomics is very promising in terms of furthering our understanding of nervous systems. While it has already very successfully increased our understanding of biological neural networks, connectomics still has several large challenges to overcome, to the point where over 20 years after the connectome of *c.elegans* has been found as the first connectome of a multi-cellular organism, we are still not able to fully accurately simulate its nervous system, with progress for more complex neural networks lagging even further behind. There are several distinct obstacles that hamper progress within this field:

1. The data collection itself has several technical limitations imposed on it. In order to accurately not only map a nervous system anatomically in its entirety, but also find the appropriate parameters that govern the operation of every single cell and synapse within it, a large variety of different techniques need to be utilized, and some of those may prove impossible to perform on a given organism.

2. As connectomics aims to analyse ever more complex organism, these technical difficulties get exacerbated by ethical considerations, which severely limit the possible techniques that can be used.

3. Finally, all these problems get severely inflated by the sheer amount of data required.

All these problems mean that finding the full connectome of any organism is a very lengthy task, albeit one that produces partial results at a constant rate.
Some techniques have in the past been used to work around this problem. From a neurophysiological point of view, some measurements that could not be conducted on e.g. *c. elegans* have been conducted on closely related species, with the results being extrapolated . More recently, evolutionary algorithms have been used to find missing parameters in order to be able to simulate the connectome accurately. This however will always raise the issue of whether the simulation can actually be used to infer explanations on the connectome underlying them, since it is ultimately possible that the simulation works entirely differently, only accidentally producing similar resulting behaviour.
So far complete datasets on complex organism, while certainly theoretically feasible, appear not achievable within the immediate future (Gjorgjieva et al., 2014) (Mikula, 2016).
There should be something more here...

Add Ref from THAT book!

I do not like this paragraph at all

text

1.2 Scientific Questions, Hypotheses and Goals

The main scientific questions in our project are aimed at *feasibility* and *scientific value* of the use of abstracted simulations in *connectomics*. Having extensive experience in we will evaluate the current state of such simulations and their applicability to future studies on more complex organism than *c. elegans*.

something relevant

It is therefore our goal to explore and justify the following statements in detail. They are the key hypotheses of our proposal, and will be explored though the simulation of a specific connectome, namely that of *c. elegans*:

1. Studies will be performed to prove that it is feasible to work with partial simulations to obtain workable, if preliminary data. Considering the considerable timeframe involved in finding connectomes, this will lead to earlier findings in the future by being able to simulate unfinished connectomes with an relevant degree of accuracy while increasing researchers' awareness of the exact degree of uncertainty inherent in the use of simplified models in this specific use case.

2. Considering the computational requirements for simulating the entire connectome, let alone running this simulation through an evolutionary algorithm, and the overarching goal of getting preliminary data from incomplete connectomes, it will be explored how far the computational model can be abstracted without influencing the results beyond tolerances.

3. The simulations can be used to gain insight into the stability of the *c. elegans* nervous system, which will give useful data as to how optimized biologically evolved networks actually are.

Should we say something about a comparison to the in vivo studies about this?

4. As macro-scale connectomes have been used to gain valuable insights into the inner workings of human brains despite only providing a quite abstracted and simplified, we will show that the same principle can be applied to micro-scale connectomes: That even simplified and incomplete models and simulations based on these can yield valuable insights and should be considered regardless of known or unknown inaccuracies to the actual known details of neuron operation.

5. Since the current connectome for c. elegans is not entirely complete out of necessity, lacking some essential parameters for simulating exact neuron functions, the implementation will find these missing parameters by means of evolutionary algorithms, matching the evolved behaviour as well as possible to the expected behaviour as expressed by *c.*

Needs rephrasing, should we even keep this?

elegans. The project will explore how much of the function of the connectome is predetermined by the physiological structure of it, and how much is governed by the exact parameters of the neurons - parameters which at least to a certain degree have the potential to adapt over time. Pursuant to this the evolution will be repeated using different target behaviour, in order to find the threshold at which the network as given is no longer able to perform the function due to it's differences to the one the structure was intended for.

Scope?

1.3 Scientific Relevance and Innovative Aspects

Connectomics is a promising field of study to further our understanding of nervous systems. However, progress in this field is slow due to the sheer amount of data required and the difficulty in measuring that data. This problem has been worked around by using macro-scale connectomes, which do appear to provide useful data while still presenting a significant level of abstraction from the actual workings of the brain.

Ref in comments, couldn't fetch that paper

Conversely, the field of micro-scale connectomics has focussed on providing detailed data on the exact workings of neurons and their interconnections into a nervous system to the point of best possible match to known parameters from in vivo measurements. While this discipline focusses on accuracy, the amount of data to be measured is quite staggering. Also, as the example of *c. elegans* shows, when the expected data set is completed, it often leads to important distinctions in hitherto disregarded details that now also need to be captured (e.g. (Bargmann, 2012) and (Izquierdo & Beer, 2013)

Given this reliance on a set of data that can take enormous amounts of time to compile, the ability to work with the incomplete datasets would allow further research, basing their studies on such simulations, to get a head start, allowing researchers to get preliminary findings before the dataset is entirely completed.

Another aspect of the same problem is the fact that accurate micro-scale simulations require vast quantities of processing power to complete. While this has proven completely necessary to accurately model the exact details of observed behaviour, simulations that have arguably taken place before our understanding of neurophysiology reached current levels and that were by today's standards hampered by grossly underpowered computer equipment, such as (Kimura & Onami, 2005) were already quite successful in generating useful knowledge.

Bit about how this will be implemented by us

1.4 Importance of the expected results

The main areas of contribution will correspond with our key hypotheses as well as the work package structure that follows; therefore we content ourselves with a very concise list at this point:

- An evaluation of current and past simulation models for neuron function and neural network function in regards to their accuracy and processing simplicity. While research in this area tends to focus on finding the best possible accuracy, our evaluation will look at the trade-off between accuracy and processing power, and how much processing power can be saved by using a simplified model while staying within a given tolerance for the accuracy of the results.
- An evaluation of current connectome simulations concerning the divergence of results with different levels of abstraction and simplification in the details of the simulation will give valuable insight into how simplified data as a foundation for simulation can yield valuable preliminary results.
- Based on this evaluation a tool will be developed with the aim of providing researchers an accessible tool to simulate neural networks without specialist knowledge about the computer science behind it being required.
- This tool will be used to conduct a study concerning the stability of the neural network of *c. elegans*. Various detrimental stimuli will be applied to the network to gain data about which level of interference is required for certain functions to cease.

Is this relevant in "importance of expected results"?

2 Work Plan

2.1 Deliverable Overview TEMP

TO BE DELETED

- Papers:
  1. Evaluation of neuron models and their characteristics concerning the simplification/accuracy trade-off
  2. Evaluation of current state-of-the-art simulations and their applicability
  3. Implementing a modular simulation engine
- Conference contributions:

1. Our Implementation of the simulation engine (once finished)
2. Result comparison: Found connectome (EA) vs previously found, analysis of differences in regards to changes in sim accuracy.
3. Study results: Impact of computationally less accurate models on overall results of simulation.

2.2 PhD-Split TEMP

TO BE DELETED

1. **Wolfgang:** Technical implementation of the simulation and neuron models, EA
2. **Bernd:** Visualization and graph-based resilience studies.
3. **NN:** To be announced upon project start.

The main workload involved in this project will be handled by three PhD students supported by the senior researchers and regular staff from the research group. While the maximum timeframe for a project could be longer, we expect the PhD students to finish their degrees within three years. Thus, the overall duration for this project will also be limited to 36 months.
Our research will be split into 3 distinct work packages (WP), where WP 1 deals with evaluation of the current state-of-the-art for neuron and connectome simulations, as well as implementing our simulation framework based on these findings. WP 2 deals with visualization and editing of connectomes for use in our simulation engine, and WP 3 finally uses our completed framework to conduct stability studies on the connectome of *c. elegans*.

2.3 WP 1: Implementing the Simulation

CopyPaste some stuff here from above? Need the emphasis/padding?

Motivation: In order to study simulations, these simulations must obviously first be implemented. Most simulation tools that already exist in the area of connectomics attempt to replicate every known detail of the neural network, not just as a whole but also on a cellular level. While this has obvious accuracy benefits when attempting to discern the actual operation of neurons and their interconnectivity, it also bears a cost in terms of processing power requirements. If one is interested in investigating a more complex connectome than that of *c. elegans*, one quickly finds the necessary computations exceeding the feasibility threshold. Since this study attempts to research the applicability of simulations that use slightly abstracted neuron models in an effort to reduce this computational complexity, it will be necessary to

implement our own simulations. As considerable effort can be invested in doing so, we will look at existing tools as a starting point, and aiming to adapt one of these as opposed to starting our own tool from scratch. With this in mind, *OpenWorm* (Szigeti et al., 2014) appears to be a very promising tool for us to investigate, due to its modular design and multi-scale nature. *OpenWorm* is a modular simulation engine specifically tailored to *c. elegans*, with the stated goal of the OpenWorm Foundation being simple access to in-depth simulation of the worm not only to researchers, but also to other interested parties such as artists or, in fact, any other interested individual. As such, it attempts to provide the greatest amount of accuracy in simulation while still providing a certain level of ease-of-use. In order to achieve this goal, an *open science* approach has been chosen for the project, with a large number of contributors still advancing the simulation. Since development on *OpenWorm* began, parts of its core modules have already been updated to reflect new neurophysiological findings about *c. elegans*, increasing the complexity of the program while at the same time increasing its accuracy, such as (M. et al., 2013). Due to the open-science approach used by *OpenWorm* and the resulting ease of access to the assets involved, as well as the modular nature of the implementation allowing for modification without prohibitive effort, *OpenWorm* is likely a very suitable candidate for a starting framework which we can adapt to suit our needs. However, due to the complexity of the software involved, this decision is not a trivial one, and some work must still be invested in evaluating the exact advantages and disadvantages of using *OpenWorm* as a foundation for our work.

Selected Related Work: As we are trying to implement not just a single simulation, but rather a modular framework able to cope with a variety of different neuron models, *OpenWorm* was chosen as the prime candidate. The original rationale for the necessity for such a framework is presented in (Szigeti et al., 2014). There the authors also identify several of the key issues with integrative simulations in the field of connectomics, mainly in the use of traditional academic structures, which impede large-scale cooperation and sharing of data, which they overcome by using an open science approach. This gives us the advantage of not only providing the framework to build on, but also providing a platform for publicizing our findings after our studies are concluded.

The framework *OpenWorm* provides has also in the meantime been expanded to include more and newer models to work with (for example (M. et al., 2013)), proving that the goal of the original creators to implement a truly modular tool-kit was indeed successfully met.

The authors in also explicitly state that the possibility of using different neuron models to investigate the relation between biological accuracy and the scope of behaviours reproduced

by the simulation is indeed a secondary ambition of providing this framework ("As stated in the Introduction a secondary ambition of the project is to explore heuristically how the complexity of behaviors reproduced by the models scales with biological realism." Szigeti et al. (2014)). However, to the best of our knowledge, such a study has not yet been conducted.

Other tools to be evaluated are the NEURON simulation program (Hines, 1994) , which, while providing good simulations of single cells and theoretically being able to simulate networks as well, includes no syntax for neural networks, and GENESIS (Bower et al., 2003), which is commonly used by researchers when simulating large-scale neural networks.

Ref only - book chapter. Include as reference?

Finally, in order to support its effort to establish a standard format for connectome information, we consider it important to use *NeuroML* (Gleeson et al., 2010) for our simulation. This format is also already used by *OpenWorm*, but even if we decide against that framework, *neuroML* will be used by this project.

Objectives:

O1: Definition of simulation requirements. As a first step in implementing our simulation or even evaluating existing tools the requirements for our studies need to be defined in order to provide a goal for the evaluation and subsequent adaptation. This in itself is no trivial task. While we attempt to study the effects of simplification and abstraction on the accuracy of these models, the exact nature of these abstractions must be chosen out of a very large selection of possible models, each having their respective benefits and drawbacks. Overall, since the aim is research into the effects of simplification on a model, several models with varying degrees of simplification from the state-of-the-art best fit to neurophysiological data need to be evaluated in regards to their relevance and expected results. We consider it important though that the models chosen for this approach be chosen at this very early stage of the process, allowing the implementation to be tailored to their requirements. This mainly means a large amount of modularity and customizability in the finished simulation tools, in order to permit a large selection of models to be used in our study. However, since modularity and customizability in software usually comes at the cost of an elevated level of effort spent on the actual implementation, there is a certain trade-off involved, where some degree of flexibility of the framework may need to be sacrificed in order to allow for more time for the actual study. For this reason we must decide at the very least on a few types of models to use, as well as specifying their requirements on the simulation. This will be used as a starting point for the requirement specifications used for the next two objectives.

O2: Evaluation of existing tools. Instead implementing an entire new simulation frame-

work, our focus will be on the adaptation of an existing one, namely *OpenWorm*. This framework is very modular, a fact that has been used in the past to update its internal workings to accommodate newly discovered neurophysiological insights. However, simplifying this tool from its original form has not yet been done to the best of our knowledge. Also, the tool's development has been focussed on the connectome of *c. elegans*, with the possibility of also simulating other neural networks so far not explicitly featuring in any documentation. However, it is still possible that *OpenWorm* does include such a possibility with only minor adapation. Thus the tools needs to be evaluated for its usefulness for this project. While *OpenWorm* seems to have the most potential as a base for our implementation, there is of course the possibility that it is in fact unsuitable for our work. In this case, our evaluation will be expanded to include other tools such as . Even if *OpenWorm* proves to be the best candidate, these other tools will be evaluated as well in order to gain an overview over the current landscape of neuron simulation tools. Once the framework we will use in this study has been identified, the next step will be the identification and implementation of additional features required to shape the simulation for use in our study. While *OpenWorm* has been chosen as the prime candidate mainly due to its functionality and modular nature, it also has one additional large benefit: The Open Science approach practiced by the *OpenWorm* team means that documentation and source code are easily accessible, allowing us to do this without extended effort.

Get some other tools

O3: Creating the simulation toolset After deciding on which framework to use and how it must be adapted, the next step is to actually implement the simulation according to our specification. This not only means adapting parts of the selected framework to our purpose (such as replacing the neuron model), but also implementing several alternative models (such as one for neurons affected by a disease) that will subsequently be used in our simulations to emulate faults in the system.

O4: Ease of Use This study, and the toolset it provides, will require a certain amount of effort for setting up for each individual study performed. It is thus our goal to also implement tools that allow this process to be performed through an easy-to-use editor that would not only cut down on the time required for this initialization, but also allow researchers without expertise in programming to use our tools effectively.

Tasks:

T1: Identification and specification of suitable neuron and neural network models for use in the study. While this point could be designated as a simple starting point for the remainder of this project, we are committed to providing a thorough analysis on the

advantages and disadvantages of each possible model, to the point where the output should be in-depth enough to qualify for publication as a paper on its own merit without relying on the remaining parts of this project for substance.

T2: Evaluation of existing tools. We will evaluate *OpenWorm*, as well as several competing simulation tools, particularly their inner workings for compatibility with our goals.

T3: Adaptation of chosen tool and implementation of additional features. Even adapting an existing tool for our purpose, there will be some features we will have to implement ourselves. The main goal is to have the simulation as adaptable as possible, to deal with a lot of possible scenarios during the actual studies.

T4: Implementation of an Evolutionary Algorithm. Finally, based on the adapted simulation, an evolutionary algorithm will be implemented in order to not only find, but also fine-tune the various parameters within the connectome. While this has been done on *c. elegans* and the resulting datasets are available (amongst others as a part of the *OpenWorm* framework as detailed in (Szigeti et al., 2014)), part of our study will necessitate some new fine-tuning, in order to adapt the network to a slightly abstracted operation of the single neurons that compose it. While this can also be based on pre-existing work, some implementation effort needs to be expended here for adaptations. It should be noted that this evolutionary algorithm should be flexible enough to deal with bigger and more complex connectomes in the future, in order to provide a tool that is not only useful for *c. elegans*, but has a large variety of possible applications in future connectomics research. This will explicitly have to include an implementation that deals with the technical problems of simulating and evolving a large connectome, such as the immense amounts of data needing to be handled, even though this is not a problem yet.

Deliverables:

D1: Specification of suitable models and their benefits / drawbacks when used in a simulation, along with their requirements for the simulation. This deliverable is considered to be substantial enough to warrant publication as a paper in a journal such as

insert journal name here. Or conf?

D2: Simulation framework. This will be the base simulation framework to allow highly customized simulations of connectomes based on simplified neuron models. As one of the main tasks of this project, the work should provide sufficient material to support a paper regarding the implementation well before completion. Furthermore, the completion of this task should also be presented at a suitable conference such as in order to inform the scientific community at large of our efforts as well a taking part in a platform where the conclusions of our work can be discussed with other experts in the field directly.

insert conf here

D3: Connectome Editor. This editor

will allow easy use of our framework, even by non-programmers. It will enable researchers to use our tools quickly and easily as a part of iterative studies. **D4: Evolutionary algorithm.** This will allow researchers to use our editor to alter the connectome within higher tolerances than supported by the original connectome parameters, by allowing the connectome to adapt its normal function to the now altered parameters.

2.4 WP 2: Connectome Visualization and Editor

Motivation: Once the simulation is operational, it will be important for researchers to be able to use it effectively, preferably without the need for specialist knowledge in the exact function of the simulation itself. Even if such knowledge is present, it is a large time investment to have to interfere with the simulation's code every time a minor adjustment needs to be made. While this is generally true, it holds particular relevance here, since our study will revolve around repeating a similar experiment in the simulation several times, adjusting parameters and neuron models between runs. While simulating the connectome is a rather computation-extensive process, the resulting patterns can be very complex and difficult to read from tables of numbers alone. It is therefore necessary to implement a visualization tool that allows researchers to *see* the connectome in operation, in order to benefit from humans' ability to detect patterns and changes in them visually.

Our project team has experience from the creation of the graph visualization tool *Graphinius* , which we will put to use for this tool as well. The overall goal is to create a front-end for our simulation, which will not only provide an accessible way of defining the parameters of a simulation, but also serves as a easy way to observe the function of the connectome, when possible in real-time, in a fully interactive client. While our experience with *Graphinius* provides us with knowledge about the creation of graph-based visualization tools, the level of complexity targeted by our framework, we will also evaluate existing tools for connectome visualization and incorporate our findings into this project.

Graphinius Reference!

Selected Related Work: While the exact requirements on the editing part of this work package will be defined as part of *WP 1*, the visualization part will be more interesting from a technical point of view. As mentioned before we will base this tool on the *Graphinius* framework due to our familiarity with the platform and the benefits this gives to development. We will however also be evaluating other existing tools for their relevance to our effort. Specifically we are looking at the *neuroConstruct* and *ConnectomeViewerFramework*.

Ref again

This sounds really awkward.

NeuroConstruct (Gleeson et al., 2007) as a tool is an inspiration of what we want our edit-

ing tool to be. It allows for the definition of cell models, the free construction of a neural network based on these single cells in 3D space, and the automatic generation of script files for simulator packages to carry out simulations. It does however only support NEURON (Hines, 1994) and GENESIS (Bower et al., 2003) as its simulation engines. It also allows the finished simulation to be loaded for visualization, which is certainly a desirable feature for researchers. Finally, it uses the NeuroML format, which is also used by *OpenWorm*, allowing for compatibility.

ConnectomeViewerFramework (Gerhard et al., 2011) on the other hand does not use NeuroML, and while the user interface promises a higher degree of usability than *neuroConstruct*, it also requires the user to have some knowledge of Python for scripting. However, it does have superior support for large-scale connectomes and their visualization. Since one goal of this project is compatibility with more complex connectomes, some of the best practices of large-scale connectome visualization need to be taken into consideration if one does not want to look at uncountable numbers of single neurons.

Phrasing

Objectives:

O1: Connectome Visualization. In order to better steer the simulations, researchers need a tool with which to get immediate feedback on the operation of the simulation. This tool needs to be able to visualize not only the structural layout of the connectome being simulated, but also the current state of the network. There are also two time-scales necessary for this tool to operate. Given our goal of providing tools that will be able to deal with more complex connectomes in the future, we must assume that the complexity of the network being investigated may be such that the simulation will not be able to process it in real-time, necessitating the visualization to run as slow as the simulation. On the other hand, the simulation will produce output which can be viewed using the same tool at a later date, at which point the tool should be able to handle real-time data.

O2: Connectome Editor We will also implement an editor for the base connectome data, so that the network to be simulated can be easily adjusted by researchers without expert knowledge in programming. This editor will allow not only the alteration of parameters within the existing connectome structure, but also the adding and removing of neurons and synapses, as well as the editing of parameters of the simulation itself. Importantly, it will also permit the swapping of the used neuron models for individual neurons, which will permit the simulation to emulate certain conditions where single neurons suffer from some restriction of their function, such as MS. It will also seamlessly integrate with the evolutionary algorithm in order to fine-

integrate with the paragraph further up

Confirm/Find alternate example. Ref!

tune the network to the changed parameters.

O3: Expanded Control through the Visualization With the editor and its methods of altering the base connectome data in place, the visualization tool will be expanded to allow limited control over the network during operation. The expected use case for this would be stability studies, where the connectome is running within parameters as usual when some part of its operation is compromised, as can happen in nature through injury or illness.

Tasks:

Deliverables:

Write this up.

2.5 WP 3: Conducting studies

Motivation: With the simulation in place, studies can be performed to ascertain the level of inaccuracies projected into the results by the use of a simplified simulation. In order to find this, our simulation will be used to perform stability studies on the connectome of *c. elegans*. By repeating the connectome stability studies performed by , we will have an example of an *in vivo* study we can use to match our results against, giving us information about the accuracy of our simulation. This will be used to generalize about the effectivity of using simplified neuron simulations to gain preliminary results about more complex connectomes in the future.

get that reference

This needs to be properly expanded after checking the in vivo studies

2.5.1 Simulation

Once the evolutionary algorithm has concluded and the missing parameters are found, the simulation can commence. For this it is necessary to implement a framework so that researchers can easily repeat the simulation with slightly altered parameters. At this point the software should also definitely be easy to use, so that the research can proceed without usability issues in the repeated simulations unnecessarily extending the time-frame of the project.

Tasks:

T1: Stability studies on the connectome of *c. elegans*. The simulation chosen to test our hypothesis is a study concerning how damage to single neurons in the network affects the function of the network as a whole. Apart from the direct data regarding the stability of the connectome in regards to single point damage events, this study will also be repeated using different simulation models with varying degrees of accuracy. The goal is to ascertain the extent to which the found results will differ given these varied models. This will show how much variations in accuracy of the used model actually influence the end results of the study, shedding light on the usefulness of simulations using simulations which may not be

state-of-the-art in terms of simulation accuracy, but bring benefits in the use of processing power.

T2: Connectome Stability Tool. A tool will be created to allow researched to easily and quickly start simulations and affect them in real time. This tool will be made available on a open-source basis to facilitate easier studies in future.

Licence?

Confirm?

T3: Connectome Alteration Language As a final part of the toolset a command language will be defined in order to allow researches quick and easy alterations to an otherwise fixed simulated network, such as replacing some random neurons with one following a slightly different behaviour in order to simulate certain diseases. Like the other tools this will be defined with modularity and adaptability in mind, so that it will be usable in future simulations of different connectomes. This standardized language will greatly improve the accessibility of our tools.

Tenuous...Nee catchy name though

Deliverables:

D1: Connectome of *c.elegans*. The complete c. elegans connectome with the parameters found by the evolutionary algorithm will be published, to allow for easy verification of the results. As actually matching the found parameters with the actual physiological properties of the neurons would certainly be beyond the scope of this project, it would benefit future research if this data were to be compared by a more neurobiologically minded research group to comment on differences or similarities. One comparison however is possible: The comparison between our dataset and ones previously discovered by evolutionary algorithms, where possible by algorithms using the simulation chosen as the base for our study. This will allow us to find similarities and differences, highlighting the impact the altered simulation model has on the final result. Some results, such as some information paths being vastly different as a result of minor differences in simulation accuracy, will yield new insights into the way neurons interact and how stable and/or adaptable a biological neural network is. This comparison will be submitted for a conference.

Find conference

D2: Analysis of the impact of different accuracies of neuron models on the results of connectome simulations. This will be an in-depth analysis of the results of our overall project. These will be presented at a conference

again, which one

also a paper?

Is this enough?

3 Organizational Aspects

mostly copy of hedatbio

3.1 Work Organization, Supervision and Risk Management

The lead applicant, Andreas Holzinger, will act as project leader and will work directly on this project and allocate a significant amount of time to the HEDATBIO project. He is associate professor for Computer Science and a member of the doctoral school for Computer Science at Graz University of Technology, hence he is in the position to supervise the involved PhD students and to bring in new or additional work-force on demand.He has extensive project management experience and know-how in software development and his diverse scientific background will be a further success factor for this project. Moreover, this project has the full commitment of the Holzinger Group, the applicant's institute and the University, so all technical facilities and organizational support is ensured.

accurate?

The PhDs employed by this project will also be supported by the PostDocs, who will only be hired after the proposal is granted in order to find the best possible international candidates for ensuring the full success of HEDATBIO.

Support from undergrads?

We take care on risks at three levels (scientific risks, management risks, technical risks):

Scientific risks lie in the uncertainty of research. We follow the approach that key elements in managing uncertainty are reflective learning and sense-making as well as a good communication strategy and a well-balanced atmosphere to stimulate thinking and problem solving. We take measures to control progress of work, ensure detailed and clear definition of architecture/interfaces and focus on implementing key features in sane iterations. To further reduce risks we have incorporated an international Scientific Advisory Board and will ensure regular communication/feedback.

accurate?

Managerial risks include lack of resources and/or staff changes forced upon the project by one or more collaborators, therefore the quality of the outcome might decrease as it depends on having access to high-quality resources and staff. We will ensure a good level of communication to talk about any problems that arise and will strive to quickly bring in new scientific staff if necessary. For this we are well prepared as we can allocate additional human resources from permanent staffs, at short notice.

Technical risks may include difficulties in data acquisition. This risk is minimal, since we

are planning to use open source data sets as well as locally available data from both sides. All equipment is at our disposal and this project has full commitment by both Institutes and both Universities.

Rewrite this paragraph

3.2 Strategies for Dissemination of Results

1. Science-to-Science: We will produce internal progress reports. The most promising ones will be extended to peer-reviewed conferences papers, symposia and workshop contributions and the most valuable results will be developed into solid international journal publications. The project team also plans to organize workshops/special sessions at international conferences to disseminate the gained knowledge, and to make algorithms and tools accessible to the international scientific community. The lead applicant has established an HCI-KDD expert network since 2011, which will also serve as an international dissemination platform.

2. Science-to-Business: As our algorithmic libraries will be open source and available online, businesses might find it interesting to utilize or embed our software in their products. While we have no co-operation with existing commercial vendors of KDD software at present, good business cases might open up in the future as more and more algorithms become available on the platform. As ClowdFlows is web based and follows the ideas of 'executable paper', it could for example benefit editors of journals vetting submissions, students trying to learn from real world examples, or engineers collaborating on prototyping new algorithms.

do we have s2b?

3. Science-to-Public: We will inform the public about our research and aim to disseminate our knowledge broadly on a regional level, for example in public exhibits, at public Open-lab days ("FWF Lange Nacht der Forschung" and "ARRS Night of researchers"), in regional newspapers and local German and Slovenian speaking workshops, etc. A project web-site HEDATBIO will be available and provide a showcase.

accurate?

3.3 Economic, Social and Practical Impact

The main goal of this project is on *enabling and supporting future connectome studies by giving researchers the tools to estimate the impact of the trade-off of processing power vs. accuracy of results in simulations*, but in order to achieve this we will have to devise new tools of practical importance. While this knowledge will accelerate future connectome studies, by allowing simulations to save processing power where absolute accuracy is not required, these tools will further reduce the time investment required for such studies, by giving researchers, particularly those without expert knowledge in computer science neuron simulation the ability to define detailed studies out-of-the-box. While this knowledge is mainly of use for the research

community, any acceleration of neurophysiological research and the understanding of the brain it promises could have a significant impact on future medical developments.

3.4 Career Benefits for those Involved

The PhD students involved will have an excellent opportunity to achieve knowledge in an extremely interesting and stimulating setting within a realistic time horizon. There are several promising research directions worth being pursued as a PhD within this project.
All project members have the opportunity to achieve expert know-how in an area that is becoming ever more important in the future, particularly considering full connectome simulations of complex organisms are certainly on the way.

3.5 Infrastructure

entire section copy of hedatbio with SLO taken out

The Institute for Medical Informatics, Statistics and Documentation, Medical University Graz is working on biomedical information systems, with emphasis on making clinical data usable. This includes the development and evaluation of algorithms, software and statistical methods, from data acquisition to data analytics. The Institute offers statistical expertise for biomedical research projects, data management for clinical trials and data extraction and reporting services for medical research. The dedication is to deliver high-quality, methodical contributions and to develop software systems for the support of clinical researchers with a focus on information quality (Holzinger & Simonic, 2011). The Institute maintains a Quality Management System, has long-standing software engineering expertise, and is ISO 9001 (certification number: Q-11627/0) certified in both project management and software development. All necessary equipment is available and this project has the full support of both the Institute and the University.

3.6 National and International Cooperation

entire section copy of hedatbio

The lead applicant is employed at the Medical University Graz, located at Graz University Hospital, where excellent local co-operations to relevant clinicians and biological domain experts are established. There are well-established cooperations with Graz University of Technology, where the project applicant teaches the main required lecture Biomedical Informatics at the Faculty of Computer Science and supervises engineering students. On an international level the Group is well connected to the international HCI-KDD network, which the project

applicant established. To guarantee that HEDATBIO will be successful, we will collaborate with international partners whose use cases connect to ours.

We will involve three international experts as our **scientific advisory board** who will also help in dissemination of our work: Prof. Dr. Ning ZHONG from Japan, and Prof. Dr. Nitesh CHAWALA from the USA, and Prof. Dr. Igor JURISICA from Canada.

3.7 Project Team

Note: Short CVs of team members are attached in separate documents

entire section copy of hedatbio

Assoc.Prof. Dr. Andreas HOLZINGER, PhD is the project applicant and principal investigator, head of the Research Unit HCI-KDD at the Medical University Graz, and currently Visiting Professor for machine learning in health informatics at Vienna University of Technology. His interdisciplinary experience in management of several national and EU Projects (e.g. REACTION Remote Accessibility to Diabetes Management and Therapy in Operational Healthcare Network; EMERGE Emergency Monitoring and Prevention) will be very beneficial in this project. He will support and supervise the PhDs on all theoretical and managerial aspects, providing an ideal research climate for his advanced students.

Bernd Malle, PhD student Bernd has finished his Master studies in Software Development at Graz University of Technology, supervised by Prof. Dr. A.Holzinger. He is interested in graph based Machine Learning and initiated the project 'Graphinius' connecting a graph library to an in-browser code editor & visualization module. The combination of theoretical interests and practical ambition makes Bernd an ideal candidate for this project, with a particularly good match for the connectome resilience studies.

Wolfgang Scherer, PhD student Wolfgang completed a MSc by Research in Electronics from the University of York (York, UK) with his thesis "Training Spiking Neural Networks by Evolutionary Algorithms." Since then he is finishing a second Master degree at Graz University of Technology in Software Development. With his previous work on simulations of neural networks as well as evolutionary algorithms, he is a very good candidate for handling the implementation of the tools in question.

Mach ich noch oder soll ich schon fertig schreiben?

Sonstige? 1 NN PhD?

clarify?

1 x N.N. PhD One more PhD student *will be hired once the project has been granted.* This third PhD will work on:

define workpackages for 3rd phd

4 Financial Aspects

4.1 In-kind contribution of partners

confirm/upda

94,000 EUR (senior staff costs, organization of international workshops), all necessary equipment will be provided; open access costs will be covered; invitations for the international scientific advisory board will be covered;

4.2 Grant applications of partners

In order to successfully carry out this project three full time PhDs over the whole project duration of 36 months is needed; travel costs of 2,000 EUR in the first year, 3,000 EUR in the second year, and 6,000 EUR in the third year, shall be exclusively used for traveling to conferences for the PhDs. This results in 340,940 EUR, to which the obligatory overhead of 5% must be added, which results into the total grant application of **357,987 EUR**.

4.3 Funding per year

table omitted from he-datbio. include here?

	Year 1	**Year 2**	**Year 3**	**Total**
Bernd MALLE	€ 36,660	€ 36,660	€ 36,660	€ 109,980
Wolfgang SCHERER	€ 36,660	€ 36,660	€ 36,660	€ 109,980
PhD N.N.	€ 36,660	€ 36,660	€ 36,660	€ 109,980
Travel Costs	€ 2,000	€ 4,000	€ 8,000	€ 16,000
5% obligatory Overhead				€ 17,047
Requested funding				€ 357,987

Table 1: Requested funding per year.

5 References

Note: Due to the page limit, this list contains only limited related work.

Bargmann, C. I. (2012). Beyond the connectome: How neuromodulators shape neural circuits. *BioEssays*, 34(6), 458–465.

Bower, J. M., Beeman, D., & Hucka, M. (2003). The GENESIS simulation system. *The Handbook of Brain Theory and Neural Networks*, (August 2000), 475–478.

Gerhard, S., Daducci, A., Lemkaddem, A., Meuli, R., Thiran, J.-P., & Hagmann, P. (2011). The connectome viewer toolkit: an open source framework to manage, analyze, and visualize connectomes. *Frontiers in neuroinformatics*, 5(June), 1–15.

Gjorgjieva, J., Biron, D., & Haspel, G. (2014). Neurobiology of caenorhabditis elegans locomotion: Where do we stand? *BioScience*, 64(6), 476–486.

Gleeson, P., Crook, S., Cannon, R. C., Hines, M. L., Billings, G. O., Farinella, M., Morse, T. M., Davison, A. P., Ray, S., Bhalla, U. S., Barnes, S. R., Dimitrova, Y. D., & Silver, R. A. (2010). NeuroML: A language for describing data driven models of neurons and networks with a high degree of biological detail. *PLoS Computational Biology*, 6(6), 1–19.

Gleeson, P., Steuber, V., & Silver, R. A. (2007). neuroConstruct: A Tool for Modeling Networks of Neurons in 3D Space. *Neuron*, 54(2), 219–235.

Hines, M. (1994). *The Neuron Simulation Program*, (pp. 147–163). Springer US: Boston, MA.

Holzinger, A. & Simonic, K.-M. (2011). *Information Quality in e-Health. Lecture Notes in Computer Science LNCS 7058*. Heidelberg, Berlin, New York: Springer.

Izquierdo, E. J. & Beer, R. D. (2013). Connecting a Connectome to Behavior: An Ensemble of Neuroanatomical Models of C. elegans Klinotaxis. *PLoS Computational Biology*, 9(2).

Kimura, A. & Onami, S. (2005). Computer simulations and image processing reveal length-dependent pulling force as the primary mechanism for C. elegans male pronuclear migration. *Developmental Cell*, 8(5), 765–775.

M., V., A., P., & P., G. (2013). Integration of predictive-corrective incompressible SPH and hodgkin-huxley based models in the OpenWorm in silico model of C. elegans. *BMC Neuroscience*, 14(Suppl 1), no pagination.

Mikula, S. (2016). Progress Towards Mammalian Whole-Brain Cellular Connectomics. *Frontiers in Neuroanatomy*, 10(June), 1–7.

Szigeti, B., Gleeson, P., Vella, M., Khayrulin, S., Palyanov, A., Hokanson, J., Currie, M., Cantarelli, M., Idili, G., & Larson, S. (2014). OpenWorm: an open-science approach to modeling Caenorhabditis elegans. *Frontiers in computational neuroscience*, 8(November), 137.