Quality of Life through Quality of Information
J. Mantas et al. (Eds.)
IOS Press, 2012

doi:10.3233/978-1-61499-101-4-544

Ophiucus: RDF-based Visualization Tool for Health Simulation Models

Andrew SUTCLIFFE[a,1], Anya OKHMATOVSKAIA[b], Arash SHABAN-NEJAD[b], David BUCKERIDGE[b]
[a] *School of Computer science, McGill University, Montreal, Canada*
[b] *Department of Epidemiology and Biostatistics, McGill University, Montreal, Canada*

Abstract. Simulation modeling of population health is becoming increasingly popular for epidemiology research and public health policy-making. However, the acceptability of population health simulation models is inhibited by their complexity and the lack of established standards to describe these models. To address this issue, we propose Ophiuchus – an RDF (Resource Description Framework: http://www.w3.org/RDF/)-based visualization tool for generating interactive 2D diagrams of population health simulation models, which describe these models in an explicit and formal manner. We present the results of a preliminary system assessment and discuss current limitations of the system.

Keywords. Agent-based simulation, ontology, population health, RFD graph, information visualization

Introduction

Simulation modeling of population health is a powerful technique that has been used for over a decade by epidemiologists and public health researchers to study population-level health phenomena emerging from complex interactions of mechanisms at the levels of individuals, populations, and healthcare systems. With the advances in the computing technology and increasing availability of electronic health data, it continues to gain popularity and becomes more influential in informing public health policy. However, the inherent complexity of these models and the lack of standards for describing them present a major problem towards communicating the details of the models between model developers and model users.

To address this issue, we are developing SimPHO, an ontology that provides semantic support for formal encoding of population health simulation models, and an accompanying application, Ophiuchus – a domain-specific automated visualization tool for generating interactive model diagrams. We believe that this tool will help make simulation models more accessible and transparent, and eventually improve the quality of these models by facilitating their review and validation.

Ophiuchus addresses three major problems associated with the presentation of health simulation models. First, the few existing standard protocols for describing the models [1, 2] are insufficient for the purpose of model reproducibility or formal validation due to the ambiguity of textual representation [3], which motivates the use of

[1] Corresponding Author. Andrew Sutcliffe, E-mail: andrew.sutcliffe@mail.mcgill.ca

diagrams. Second, most of the existing formalisms proposed for diagrammatic representation of simulation models [3, 4] are specific to a certain class of models and/or represent only a particular aspect or view of the model, and are therefore incomplete, when used in isolation. A combination of diagrammatic approaches ensuring multifaceted representation is preferable [3]. Finally, contemporary agent-based models of population health are often large, complex and overwhelming for the user, when all information about the model is presented at once, even if presentation is consistent and clear. In the following section we describe how we used the existing body of knowledge on graph visualization to design the system.

1. Methods

To achieve complete, multifaceted, unambiguous representation, we use an RDF ontology SimPHO (ontology for Simulation Modeling of Population Health) to encode information about a simulation model in a standard, formal and expressive way. Ophiuchus refers to SimPHO as the knowledge source and automatically visualizes a subset of encoded entities and relationships, while providing additional information on-demand in an interactive manner. RDF representation allows for automatic conversion to and between the existing graphical formalisms (e.g. UML), and for formal validation of the models at the conceptual level using the known RDF inference models and logical reasoners. In contrast to the existing domain-independent RDF-visualization tools [5, 6], which are not optimal for efficient communication of health simulation models, Ophiuchus takes advantage of the domain knowledge to emphasize essential relationships in health simulation models, such as causality, event sequencing, etc.

To address the problem of graph size and complexity, we employed known techniques for interactive incremental exploration [7]. One specific approach called "semantic zooming" allows displaying more detailed content for part of the graph, which is currently in focus. We group elements of the graph into clusters based on their content, and allow the user to expand and collapse these clusters as necessary. The semantics that forms a basis for clustering is explicitly defined in the underlying SimPHO ontology. For example, grouped together may be all variables representing upstream health determinants.

Ophiuchus is a modular application implemented using the open-source tools and libraries. We used Protégé-OWL API for Java [8] to query the ontology and the C++ libraries by Open Graph Definition Framework (OGDF) [9] for graph layout. The components communicate with each other using XML-based intermediate representations, as explained in the next section. The generated diagrams are deployed through a web-based interface in order to make them readily available to a large user community including population health simulation model developers and users, such as epidemiologists, statisticians, public health researchers and practitioners.

2. Results

2.1. System Architecture

Figure 1 illustrates the functional components of Ophiuchus and their corresponding inputs and outputs. Depending on the desired visual formalism, Ophiuchus first retrieves a subset of the relationships encoded in the SimPHO using Protégé OWL API [8]. The retrieved set of relationships is specific to the type and application of the diagram being generated. For example, the initial prototype of Ophiuchus focuses on generating causal diagrams [10], commonly used by epidemiologists, who might be primarily interested in visualizing causal relationships between the variables (including direct/indirect causation, effect modification). In causal diagrams, nodes represent variables and directed edges (arrows) denote causal relationships. Another graphical representation, event graphs [11], would focus on events, their enablers and triggers.

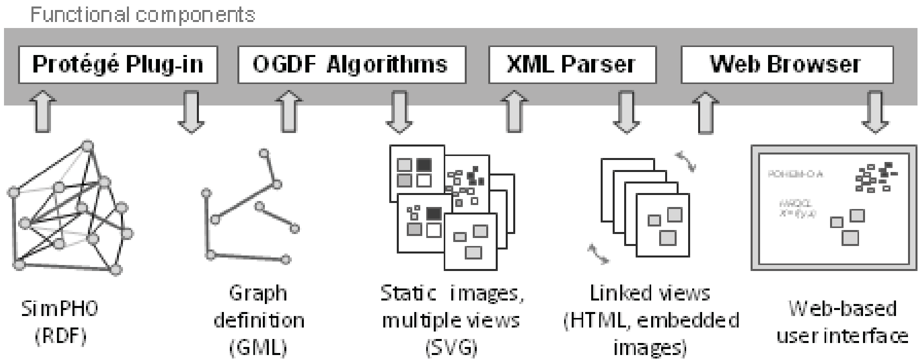

Figure 1. Ophiuchus architecture

The semantic data extracted from the ontology is translated to create a graph definition template, which is then used to perform automatic graph layout and generate 2D images in SVG[2] format by the OGDF algorithm. We have selected orthogonal layout for edge drawing, which was shown to reduce visual confusion on edge crossings [12]. The resulting SVG images are parsed, embedded in HTML and linked together to be published online at the front end.

To achieve the effect of semantic zooming, Ophiuchus has to generate a set of static SVG diagrams corresponding to different views of the same graph with different combinations of collapsed and expanded clusters. Since layout algorithms are complex, all SVG views are generated in advance before the model is first put online.

2.2. User Interface

The front-end of the Ophiuchus is a web page separated into three content sections performing different functions: *Title, Main,* and *Overview* (Figure 2). The *Title* pane contains general navigation information: the name of the simulation model and the name of the node or the cluster the user is currently exploring. The *Main* pane covers most of the screen area, and it is the principle location of interaction between the user and the diagram. Pointing at the node causes a simple *information box* to appear next to the cursor providing the user with additional information on what he is looking at. By clicking on the cluster, the user will navigate to a view where the cluster has been expanded and all of its components are visible. The *Overview* pane on the top right contains a fully expanded causal diagram of the simulation model to provide the user with contextual information, which may easily be lost when focusing/zooming on a fragment of the graph. The overview pane can be expanded and minimized as needed.

[2] Scalable Vector Graphics: http://www.w3.org/Graphics/SVG

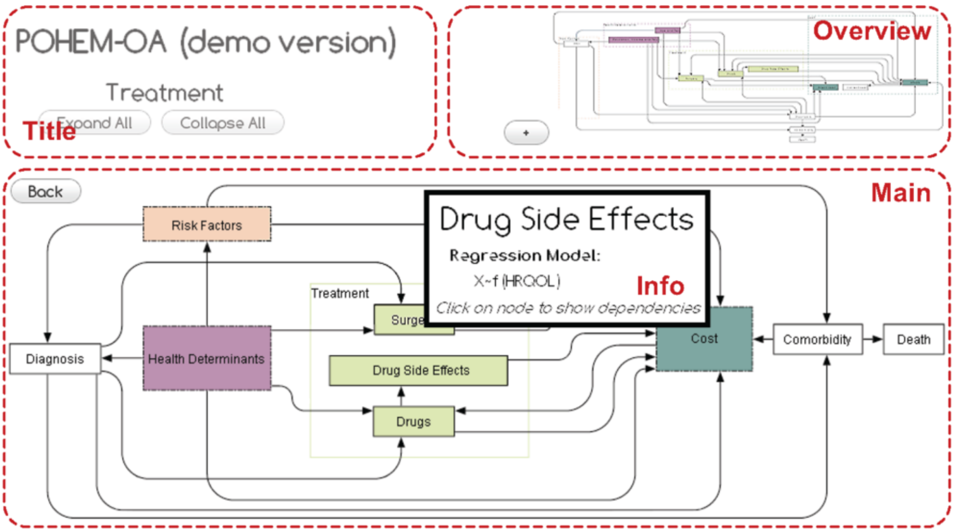

Figure 2. Ophiuchus user interface

2.3. Assessment

To perform a preliminary assessment of the system, we have used POHEM-OA, a thoroughly validated agent-based simulation model of osteoarthritis [13], which we encoded in SimPHO. The interactive diagram of POHEM-OA model was deployed online and made available to a mixed group of twenty users for unguided testing. The user group included researchers and professionals in epidemiology, public health, statistics and simulation modeling. Some of the users were familiar with POHEM-OA model as developers or users, while others had little or no exposure to the model. The users received no formal instruction on navigation or graphical notation.

Eleven users out of twenty provided their feedback. Positive comments were related to the interactive aspects of the diagram, such as being able to expand the clusters and display additional information about elements. Several limitations were pointed out by multiple respondents, which could be grouped into 3 general issues:

- Changing locations of the nodes between views appears confusing to the users.
- Sparseness of the fully expanded graph resulting in a “spaghetti” diagram.
- Distracting nature of the pop-up information boxes. The users would prefer to use buttons or triggers to bring up additional information.

The users also listed desirable additions to the subsequent versions of the system, for example the ability to turn on and off specific types of nodes and links.

3. Discussion

The generally positive feedback we received from the subjects in our informal study is encouraging and suggests that Ophiuchus could become quite helpful for the intended user-group by addressing the existing lack of specialized tools to describe and explain the agent-based models of population health in scientific and professional communication. The apparent demand for a tool like Ophiuchus, however, emphasizes the need to tackle the issues revealed during the assessment.

The most acknowledged problem – node relocation when exploring the graph – stemmed from the lack of control over the behavior of the layout algorithm. The usual heuristics used by graph layout algorithms are centered on reducing the number of edge crossings, lengths of paths and enforcing straight edges [12]. Applying these heuristics independently to different views of the diagram (with clusters expanded vs. collapsed) resulted in variable node locations, which was a source of confusion for most users. To address this issue, the subsequent views of the graph could be generated by a dynamic layout algorithm, which would take care of the changes in the graph (collapsing/expanding clustered nodes), while preserving the parts of the graph that did not change as much as possible. Another option is to employ constrained graph layout techniques [14] to partially preserve node locations between the views, or to alleviate the problem through the use of animation, i.e. relocate the nodes in a visible motion.

In addition to addressing the limitations pointed out by the users, our future work will include adding support for multiple types of diagrams (e.g. event graphs, UML-style diagrams, etc.). So far our evaluation was informal and focused entirely on the user experience of freely navigating the generated graphs. For a formal assessment, usability studies would have to be conducted on a larger sample of target users and to incorporate specific use scenarios and objective measurements. We will also need to evaluate the performance of the system with larger RDF graphs in terms of the processing time, as well as the size and usability of the resulting diagrams.

References

[1] Grimm V, Berger U, DeAngelis DL, Polhill JG, Giske J, Railsback SF. The ODD protocol: A review and first update. Ecological Modelling 2010; 221: 2760-2768
[2] CISNET: Standardized Model Documents, http://cisnet.cancer.gov/profiles/ (last access May 2012)
[3] Onggo S. Methods for Conceptual Model Representation. In: Robinson S, Brooks RJ, Kotiadis K, van der Zee DJ. (eds.) Conceptual Modeling for Discrete-Event Simulation. CRC Press; 2011.
[4] Pooley RJ. Towards a standard for hierarchical process oriented discrete event simulation diagrams part 1: a comparison of existing approaches. Transactions of the Society for Computer Simulation International 1991; 8: 1-20
[5] Katifori A, Halatsis C, Lepouras G, Vassilakis C, Giannopoulou E. Ontology visualization methods - a survey. ACM Computing Surveys 2007; 39: 10-es
[6] Deligiannidis L, Kochut KJ, Sheth AP. RDF data exploration and visualization. Proceedings of the ACM first workshop on CyberInfrastructure: information management in eScience-CIMS '07.39; 2007
[7] Herman I, Melancon G, Marshall MS. Graph visualization and navigation in information visualization: A survey. IEEE Transactions on Visualization and Computer Graphics 2000; 6: 24-43
[8] Stanford Center for Biomedical Informatics Research: Protégé, http://protege.stanford.edu (last access May 2012)
[9] Open Graph Drawing Framework, http://www.ogdf.net/doku.php/start (last access May 2012)
[10] Greenland S, Pearl J, Robins JM. Causal Diagrams for Epidemiologic Research. Epidemiology 1999; 10: 37-48
[11] Buss AH. Modeling with Event Graphs. In: Charnes JM, Morrice DJ, Brunner DT, Swain JJ. (eds.) Winter Simulation Conference. pp. 153-160; 1996.
[12] Ware C, Purchase H, Colpoys L, McGill M. Cognitive measurements of graph aesthetics. Information Visualization 2002; 1: 103-110
[13] Kopec JA, Sayre EC, Flanagan WM, Fines P, Cibere J, Rahman MM, Bansback NJ, Anis AH, Jordan, JM, Sobolev B, Aghajanian J, Kang W, Greidanus NV, Garbuz DS, Hawker GA, Badley EM. Development of a population-based microsimulation model of osteoarthritis in Canada. Ostheoarthritis and Cartilage 2010; 18: 303-311
[14] Dwyer T. Scalable, Versatile and Simple Constrained Graph Layout. Computer Graphics Forum. 2009; 28: 991-998