OPAALS PROJECT

Contract n° IST-034824

WP10: Sustainable Research Community Building in the Open Knowledge Space

D10.19 – Extension of Google maps visualisation for large scale social networks

Project funded by the European Community under the "Information Society Technology" Programme

Contract Number: IST-034824

Project Acronym: OPAALS

Deliverable N°: D10.19

Due date: May 2010

Delivery Date: July 2010

Short Description: In this code deliverable, means to use EvESim and Wille SNA toolkits co-operatively for geospatial visualisation of large social networks are described.

Author: SUAS (Raimund Eder, Thomas J. Heistracher, Thomas Kurz, Christoph Rücker), TUT (Jukka Huhtamäki, Jakko Salonen), IPTI (Fernando Colugnati)

Partners contributed: SUAS, TUT, IPTI

Made available to: OPAALS Consortium and European Commission

Versioning

Version	Date	Name, organization
0.1	15/07/2010	Initial version
0.2	04/08/2010	Submission for internal review
1.0	31/08/2010	Final submission

Quality check

Internal Reviewers: Marc McLaughlin (WIT), Amir Razavi (UNIS)

Dependencies:

Achievements*	Structured and open import facilities for EvESim, various geospatial and social visualisation options, redesigned simulation framework with agent distribution
Work Packages	WP3, WP10
Partners	SUAS, TUT, IPTI
Domains	Computer Science, Social Science
Targets	
Publications*	See deliverable D10.20
PhD Students*	None
Outstanding features*	Utilization of EvESim and Wille for the visualisation of large-scale social networks. Complete EvESim redesign on the basis of the Eclipse Modelling Framework.
Disciplinary domains of authors*	R. Eder (Information Technologies, Software and Systems Engineering) T. Heistracher (Information Technologies, Software and Systems Engineering, Biophysical Modelling) T. Kurz (Information Technologies, Software and Systems Engineering, Interpersonal Communication) C. Ruecker (Information Technologies, Software and Systems Engineering) J. Huhtamäki (Visualisation, Software and Systems Engineering) J. Salonen (Visualisation, Software and Systems Engineering) F.A.B. Colugnati (Statistics and Social Network Analyst)

The information marked with an asterisk (*) is provided in order to address Recommendation n. 4 from the Year 2 review report

Contents

1 Introduction

Visual *social network analysis* (SNA) is a powerful method enabling the analyst to gain insight into the structure of social networks and to communicate the findings to others [1]. From the beginning of SNA and its precursor sociometry, visualisation has been a key part of the analysis process, cf. Moreno ([2]):

"We have first to visualize [...] A process of charting has been devised by the sociometrists, the sociogram, which is more than merely a method of presentation. It is first of all a method of exploration. It makes possible the exploration of sociometric facts. The proper placement of every individual and of all interrelations of individuals can be shown on a sociogram. It is at present the only available scheme which makes structural analysis of a community possible."

Wasserman and Faust (1994) define *sociogram* as "a picture in which people (or more generally, any social units) are represented as points in two-dimensional space, and relationships among pairs of people are represented by lines linking the corresponding points". While sociograms are usually visualised as (socio)graphs composed of nodes representing the social actors, and edges representing the connections between them, expressive complementing visualisation approaches exist.

One of the most interesting alternative approaches for sociogram visualisation is geospatial social network visualisation in which graph nodes are laid out according to their geographical locations instead of an abstract graph layout algorithm such as Kamada-Kawai [3] or Fruchterman-Reingold [4] commonly used in different SNA tools. Freeman ([1]) points out an early example of overlaying social network data on a geographical map presented by Leonard and Loomis as early as in 1941. Although modern tools including Google Maps and Google Earth introduce means to visualise data on a geographical map, there are no general means to visualise large geospatial (social) networks.

In this deliverable, means to use EvESimulator (EvESim) [5, 6, 7] as a general tool for visualising large geospatial social networks are introduced. While the

descriptions of the visualisation views *per se* are covered in deliverable *D10.20 - Emulation and testing of OPAALS P2P infrastructure by utilisation of EvESim*, the deliverable at hand gives an overview of the data formats and data management processes enabling the geospatial visualisation in a variety of application contexts. More particularly, we describe how the *Wille SNA Toolkit* can be used to collect, aggregate and refine – in a word *preprocess* – SNA data to be used in EvESim and other SNA tools. This approach to represent our work was selected because i) deliverable D10.19 is a code deliverable, and ii) we wanted to encapsulate the comprehensive description of all EvESim-related information in one deliverable so that it can be used as the up-to-date documentation of the framework.

In detail, the present deliverable includes the following contributions:

- Geospatial visualisation views implemented into EvESim, which is available at `http://evesim.org/index.php?id=47`.
- A set of Wille components included in the Wille SNA Toolkit for data preprocessing and transformations between data formats, available at `http://wiki.tut.fi/Wille/SNAToolkit`.
- The definition of the EvESim Network Description format including a schema and examples, which is available at `http://evesim.org/index.php?id=47` in the "Example Files" section.
- Detailed instructions including screenshots to use EvESim are available at `http://evesim.org/index.php?id=117`

This document is structured as follows: In Chapter 2, the means to import social network data including the geographical locations of actors to EvESim are introduced. Wille components enabling the use of EvESim in different usage contexts are described in Chapter 3. Finally, Chapter 4 concludes the work and relates to future research aspects.

2 Interfaces for Network Description

The main interface and common ground for the geographical representation of Digital Ecosystems is the so called *Network Description XML*. This file schema was introduced in *D10.7 – Visualisation Service for P2P Infrastructure and EvESim based on Google Maps* as a joint effort of TUT and SUAS. The XML structure including the according XML schema is still valid and used as the common format for social network data in OPAALS.

An example for such a NetworkDescription can be seen in Listing 2.1. This example is a small part of a real-world network extracted from Guigoh. Network nodes exist within the *nodes*-tag. Each node has a type, which is an arbitrary text to categorize the node. The *uri*-attribute uniquely identifies the node using a standard Uniform Resource Identifier (URI). Additionally to that, each node has a *name* and a *location* which is used for geovisualisation purposes. Within the *connections*-tag, the individual *connection* entries contain the connections to the other nodes within the network. Each connection contains a reference to another node within the network. Besides that, *strength* and *type* can be added as attributes as well.

Many interfacing components and services were built in OPAALS for exporting and importing data in the Network Description XML format. Examples are the Guigoh data export service by IPTI, the Agropedia data export by IITK as well as the interfaces described in the following chapter, implemented by TUT. Whereas SNA formats (e.g. Pajek) do not include geographical data and geographical formats, and geographical formats such as KML do not introduce means to represent P2P-specific data, this XML format tries to cover the specific needs of Digital Ecosystems without claiming completeness. Its schema specifies the way of saving the network data information needed for EvESim visualisation in XML format [7, 8].

```
1 <NetworkDescription xmi:version="2.0"
2   xmlns:xmi="http://www.omg.org/XMI"
3   xmlns="http://www.fh-salzburg.ac.at/NetworkDescription">
4   <nodes>
```

```
5     <node type="GuigohUser"
6       uri="http://www.opaals.org.br/ProfileView.do?id=80">
7       <name>Christoph</name>
8       <location lat="47.800499" lng="13.04441" name="Salzburg, Austria"/>
9       <connections>
10        <connection strength="1"
11          to="http://www.opaals.org.br/ProfileView.do?id=22"
12          type="contact from Guigoh"/>
13        <connection strength="1"
14          to="http://www.opaals.org.br/ProfileView.do?id=32"
15          type="contact from Guigoh"/>
16      </connections>
17    </node>
18    <node>
19      ...
20    </node>
21  </nodes>
22 </NetworkDescription>
```

Listing 2.1: Example of NetworkDescription XML

3 EvESim and Wille SNA Toolkit

In this chapter, we describe the means to use Wille SNA Toolkit and EvESimulator co-operatively to conduct data-driven visual geospatial social network analysis. Two approaches exist. First, Wille SNA Toolkit can be used to collect and preprocess data for EvESimulator. Second, Wille SNA Toolkit enables the use of datasets in EvESimulator-compliant format (see Chapter 2 for the description) with additional network analysis tools including NodeXL, Pajek, Vizster and others.

Figure 3.1 shows an overview of the ecosystem of SNA data sources and tools and the role of Wille SNA Toolkit in it: Adapters are created to collect data from different source systems to be processed with Wille pipelines composed of different Wille components. Components implement various refinement and transforming operations to the data. Finally, data is exported to different analysis and visualisation software for further processing or, alternatively, a visual representation is created directly in a Wille pipeline.

Wille SNA Toolkit is developed with *Wille2*, the latest generation of Wille Visualisation System. Wille Visualisation System is a framework for building visualisation applications. For more information on Wille Visualisation System, please visit Wille homepage at `http://wiki.tut.fi/Wille`.

The objective of Wille SNA Toolkit is to streamline data harvesting and processing procedures related to visualisation-oriented, data-driven social network analysis done with Wille. The toolkit also enables the development of context-sensitive social network analysis and visualisation applications [9]. Examples of possible data sources for the analysis include wikis, social networking services and Web APIs. The emphasis of Wille SNA Toolkit is on supporting visual analyses of social networks but it can also be used to export data e.g. to statistical analysis tools.

Effectively, the co-operation between Wille SNA Toolkit and EvESim is enabled by the two Wille services `sna.evesim2graph` and `sna.graph2evesim`. The following two sections define their respective usage.

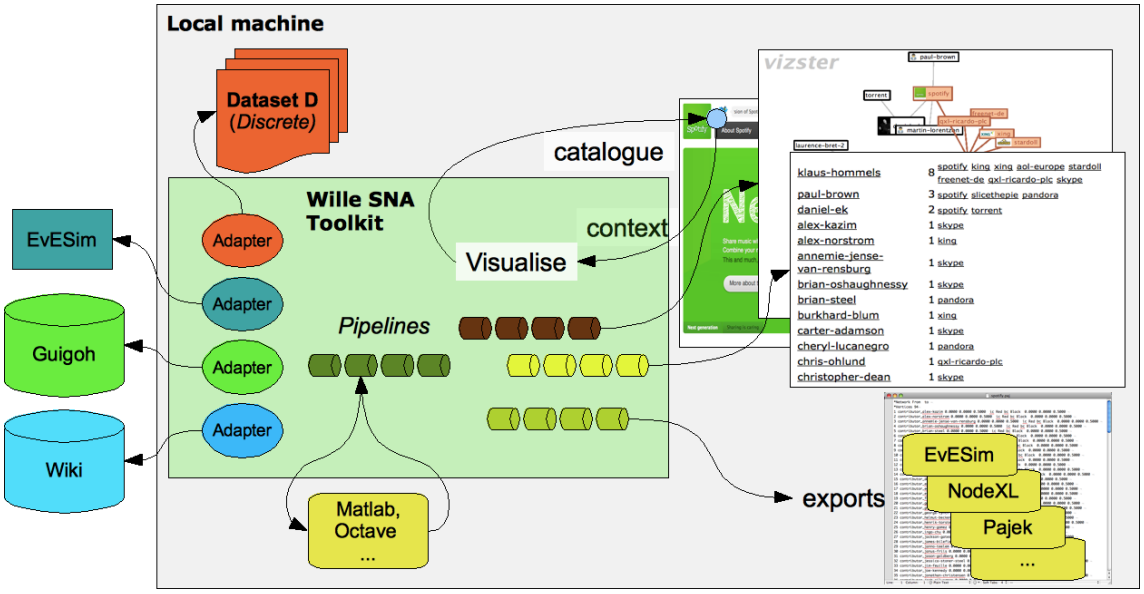

Figure 3.1: Wille SNA Toolkit and an ecosystem of SNA data and tools

3.1 Using EvESim data with SNA tools

Service `sna.evesim2graph` transforms data from EvESim Network Description format to a graph format used internally in Wille SNA Toolkit, *i.e.*, the service acts as one of the adapters in Figure 3.1. During the final phase of OPAALS, a social networking platform Guigoh has been the primary data source for social network data. Thus, an exporter utility for reading data from Guigoh in EvESim Network Description format was developed by Lia Carrari Lopes of IPTI. Further, this functionality has been wrapped as Wille service `reader.guigoh` to be used as a Wille component collecting data for Wille pipelines as shown in Figure 3.1. With this setting, Wille SNA Toolkit can be used to feed Guigoh-originated data to tools including Pajek, NodeXL, Vizster and others in addition to EvESim.

With `sna.evesim2graph`, the following transformation can be conducted. Listing 3.1 represents a data set in EvESim Node Description format. The example is truncated for the sake of brevity. Listing 3.2 shows an excerpt of the dataset after it has been processed with `sna.evesim2graph`.

```
1  <node type="GuigohUser" uri="http://www.opaals.org.br/ProfileView.do?id=80">
2    <name>Christoph</name>
3    <location lat="47.800499" lng="13.04441" name="Salzburg, Austria"/>
4    <connections>
5      <connection strength="1" to="http://www.opaals.org.br/ProfileView.do?id=22"
```

```
          type="contact from Guigoh"/>
      <connection strength="1" to="http://www.opaals.org.br/ProfileView.do?id=32"
          type="contact from Guigoh"/>
    </connections>
  </node>
```

Listing 3.1: EvESim Network Description example

```
      <actor id="1001">
         <name>Christoph</name>
         <type>GuigohUser</type>
         <image>http://matriisi.ee.tut.fi/~huhtis/demo/2010/innovationecosystems/
             static/nopicture.png</image>
         <url>http://www.opaals.org.br/ProfileView.do?id=80</url>
         <location>Salzburg, Austria</location>
         <lat>47.800499</lat>
         <lng>13.04441</lng>
      </actor>
   </actorlist>
   <connectionlist>
      <connection>
         <source>1001</source>
         <target>1003</target>
         <strength>1</strength>
         <type>contact from Guigoh</type>
      </connection>
      <connection>
         <source>1001</source>
         <target>1002</target>
         <strength>1</strength>
         <type>contact from Guigoh</type>
      </connection>
```

Listing 3.2: Dataset in Wille SNA Toolkit graph format

Next, graph data can be processed with any of the graph-processing Wille components including `sna.metrics`, `sna.graph2graphml`, `sna.graph2pajek` and others to create the different pipelines presented in Figure 3.1. More importantly, due to the use of a unified format for graps, new Wille components built to analyse and visualise graphs or export graph immediately support the datasets imported from EvESim.

Processing the example dataset with `sna.graph2pajek`, for example, leads to a representation in Listing 3.3. The data is now ready to be imported to Pajek for further analysis.

```
*Network generated with Wille SNA Toolkit
*Vertices 4
1 contributor_1003 0.0000 0.0000 0.5000  ic Red bc Black  0.0000 0.0000 0.5000
2 contributor_1002 0.0000 0.0000 0.5000  ic Red bc Black  0.0000 0.0000 0.5000
3 contributor_1004 0.0000 0.0000 0.5000  ic Red bc Black  0.0000 0.0000 0.5000
```

```
4 contributor_1001 0.0000 0.0000 0.5000  ic Red bc Black  0.0000 0.0000 0.5000
*Arcs
4 1
4 2
2 1
2 4
2 3
1 4
1 2
1 3
3 1
3 2
```

Listing 3.3: Example dataset in Pajek format

3.2 Feeding SNA data to EvESim

The service `sna.graph2evesim` serialises data in Wille SNA Toolkit graph format to EvESim Network Description documents. In addition, service `sna.graph.geocode` can be used to add latitude and longitude information into graphs that include location information only in textual format. This functionality enables the use of EvESimulator as a social network analysis tool for any data source that Wille SNA Toolkit is built to support. Currently, Wille SNA Toolkit supports social network analysis with data from Guigoh, OPAALS Wiki (`http://wiki.opaals.eu`), and Innovation Ecosystems Network Dataset (`http://www.innovation-ecosystems.org/data-driven/`). As new data-collecting procedures are built into Wille SNA Toolkit, the data collected is ready to be exported to EvESim. A detailed description of the usage of the data sources with Wille SNA Toolkit is provided in [10] and at the Wille SNA Toolkit homepage `http://wiki.tut.fi/Wille/SNAToolkit`.

3.3 On data formats

The rationale behind the usage of a tailored vocabulary for graphs in Wille SNA Toolkit is to keep the format as simple as possible. The simplicity of the format serves as means to streamline the creation of new adapters for enabling SNA work of new sources of data.

The Wille SNA Toolkit graph format follows the logical structure of GraphML (`http://graphml.graphdrawing.org`), a *de facto* graph representation format. The representation in both of the formats is separated into nodes and edges. First, nodes

are described and then edges between the nodes are enumerated. The difference between formats lies in the representation of node and edge data. In Wille SNA Toolkit graphs, the data is represented with simple key-value-pairs whereas in GraphML the data representation is more complicated, thus adding the workload of applying the format.

4 Concluding Remarks

Geospatial visualisation of large social networks introduces a valuable tool for visual social network analyses. It this deliverable, we have described means to use EvESim to visualise the structure of social networks over a geographical map.

To enable the geospatial visualisations, new user interfaces have been implemented into EvESim, which are described elsewhere. Moreover, a data format for representing network data including geographical information has been defined and a mechanism for importing data to EvESim was implemented.

Further, means for the co-operative use of EvESim and Wille SNA Toolkit are provided for developers of new network analysis applications.

While EvESim and Wille SNA Toolkit both use a dedicated format for representing the sociograph data, components to transform data between formats are provided for application developers as a part of the Wille SNA Toolkit.

Future work enabling a more streamlined means to use EvESim and Wille SNA Toolkit include defining processes for automating the communication between the two tools.

Bibliography

[1] Linton C. Freeman. Visualizing social networks. *Journal of Social Structure*, 1(1):1–15, 2000.

[2] Jakob Levy Moreno. *Who Shall Survive?: Foundations of Sociometry, Group Psychotherapy and Sociodrama.* Beacon House Inc., Beacon, NY, USA, 1953.

[3] T. Kamada and S. Kawai. An algorithm for drawing general undirected graphs. *Inf. Process. Lett.*, 31(1):7–15, 1989.

[4] Thomas M. J. Fruchterman and Edward M. Reingold. Graph drawing by force-directed placement. *Softw. Pract. Exper.*, 21(11):1129–1164, 1991.

[5] Salzburg University of Applied Sciences. EvESimulator. `http://www.evesim.org`, Last Accessed: 15/04/2009.

[6] T. Kurz, J. Noguera, R. Eder, and T. Heistracher. Peripheral simulation framework for digital ecosystems. In *1st international OPAALS Conference on Digital Ecosystems*, 2007.

[7] T. Kurz R. Eder C. Adelberger and T. Heistracher. Evaluation of two p2p-approaches for a distributed simulation framework. In *2nd international OPAALS Conference on Digital Ecosystems*, 2008.

[8] C. Adelberger. *Evolutionary Framework Simulation (Diploma Thesis).* Fachhochschule Salzburg GmbH, Salzburg, Austria, 2008.

[9] Jaakko Salonen and Jukka Huhtamäki. Launching Context-Aware visualisations. In *Proceedings of 3rd International OPAALS Conference on Digital Ecosystems, Aracaju, Brazil, 22- 23 March 2010*, Berlin Heidelberg, March 2010. Springer.

[10] O. Nykänen, J. Salonen, and H. Huhtamäki. Wille visualisation toolkit for developers with a concise oks visualisation application catalogue for end-users. OPAALS project Deliverable D10.18.