Socio-semantic network data visualization

Alexey Drutsa[1,2], Konstantin Yavorskiy[1]

[1] Witology
alexey.drutsa@witology.com, konstantin.yavorskiy@witology.com
http://www.witology.com
[2] Moscow State University, Dep. of mech. & math.
http://www.math.msu.su

Abstract. The paper is devoted to some information visualization problems arising in the course of the development of the software package *WitoAnalytics* that enables to analyze and visualize data resulting from the socio-semantic network of the Witology web-platform. The work on the software is in progress. The article contains a short overview of the first software capabilities to visualize some types of socio-semantic network subgraphs.

Keywords: information visualization, data visualization, socio-semantic network, graph

1 Introduction

The Witology company is engaged in solution of some specific real-world problems by constructing active human community, while developing collective mind of participants. In order to achieve the goal a collaborative software platform has been developed and used in the company. Its essential difference from other similar systems consists in direct involvement of specifically trained facilitators in the community. In connection with a possible large number of community members, there is a need of a visual representation of data on their activity on the platform. The data can be used by facilitators as analytical material allowing them to quickly make the right decisions.

At the present time, there is a large number of software designed for analysis and visualization of social networks data (Social Network Analysis, SNA). They include both wide-section programs to analyze all kinds of graphs such as, for example, UCINet [3], Pajek [4] and Cytoscape [5], and programs for the text analysis, for example, Discourse Network Analyzer [6] and AutoMap [7]. Furthermore the class of SNA programs includes specialized software for the analysis of social networks, for example, NodeXL [8], which allows you to retrieve, analyze and visualize data from networks such as Twitter and Facebook. Since the Witology platform is a socio-semantic network [2], then it requires a special analysis software package, adjusted to analysis and visualization of this type of network. Note that the main focus of the research is the scientific field named *information visualization* [9, 10], rather than technological problems of implementation of various methods.

2 Problem statement

In the paper [2] a general model of socio-semantic network is defined as a triplet $\mathbb{G} = (G, C, A)$, such that

- $G = \{V, E_1, \ldots, E_k; \pi, \delta_1, \ldots, \delta_k\}$ is a social network — weighted oriented multi-graph, where V is a set of network members, $E_1, \ldots, E_k \subset V \times V$ are different relations between the members, $\pi : V \to \Pi$ is a *user profile* function and $\delta_i : E_i \to \Delta_i$ $(i \in \{1, \ldots, k\})$ denotes parameters of corresponding relation;
- $C = \{T, R_1, \ldots, R_m; \theta, \gamma_1, \ldots, \gamma_m\}$ is a content multi-graph, where T is a set of all generated content elements (texts, media, evaluations, tags etc), $R_1, \ldots, R_m \subset T \times T$ are different relations between the content elements, $\theta : T \to \Theta$ denotes a function that corresponds to content element parameters and $\gamma_i : R_i \to \Gamma_i$ $(i \in \{1, \ldots, k\})$ denotes parameters of corresponding relation;
- $A \subset V \times T$ is a authorship relation between the social graph and the content.

For such graph analysis the following task is posed: to develop a series of visualizations for the most significant activity of the participants in the platform that would convey the activity in the most informative manner. For instance it could be user evaluations, text generation and etc. Such visualizations should demonstrate both time slices of the database and data change over time.

3 Results

In order to solve the task a specialized software package (hereinafter referred to *WitoAnalytics*) was developed. As mentioned above, the software developed by paper authors could be regarded as one of many SNA softwares, but adjusted to the analysis and visualization of a particular type of graph — socio-semantic network of the Witology platform. The network presented in the current article has more than 500 members and, but the visualizations contains around 200 major network members. At the moment the package allows you to build multiple WitoAnalytics monocot graph visualization and visualization of a bipartite graph.

3.1 User estimation graph

Consider the following oriented weighted subgraph of socio-semantic network: $G_e = \{V_e, E_e, \delta_e\}$, where $\delta_e : E_e \to [-k, k] \times \mathbb{N}$ is a bidimensional edge weight, the first component corresponding to the average value of vertex estimates (in some range $[-k, k]$) and the second component corresponding to the number of the estimates. Hereinafter the subgraph will be called as *user estimation graph*. Such graph could result from user content estimation data taking into account the author relations for the estimations and for the content which is estimated (like texts, etc).

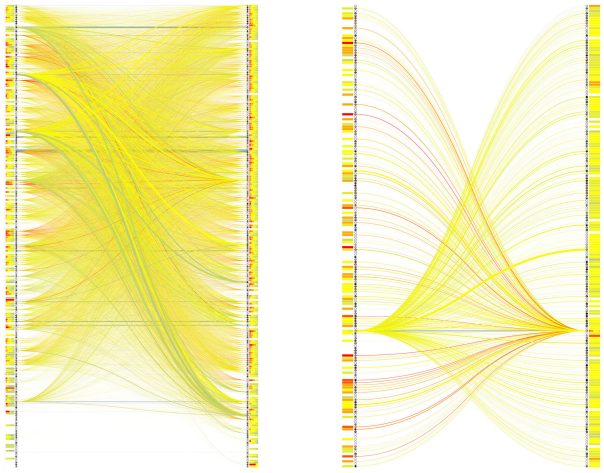

Fig. 1. (A) — The visualization "Elka" (spruce, rus.), (B) — visualization of a local user neighborhood

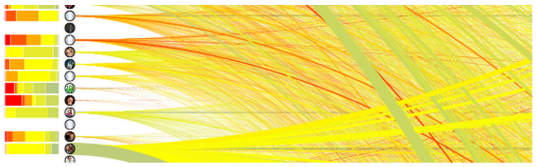

Fig. 2. Scaled-up part of the visualization "Elka"

The two following visualizations of user estimation graph are proposed. The first is a bipartite representation, where each element from V_e is associated with two nodes situated on a plain, their vertical coordinates being equal. In this case the direction of edges coincides with direction of horizontal axis. The visualization is named as "Elka" (spruce, rus.) and its example is presented in figure 1 (A). Here the edge thickness corresponds to the number of estimates between nodes, and the edge color corresponds to the average value, diagonal edges being marked out with special color. Histograms of out-estimate distribution (on the left) and in-estimate distribution (on the right) are displayed near the nodes. A local user neighborhood of the user estimation graph is presented in figure 1 (B), that is only the edges connected with a fixed user are displayed and the nodes without visible edges are removed. Figure 2 contains a scaled-up part of the visualization "Elka" presented in figure 1 (A).

The second variant of user estimation graph visualization is a monocot representation, where each element from V_e is associated with only one node situated on a circumference. In order to distinguish in-edges and out-edges for a node all

the in-edges have same joining angle to the node and all the out-edges have another same joining angle, in-angle and out-angle being not coincided and defining directions, that are symmetric with respect to the radius connected the node. The visualization is named as "Solntse" (sun, rus.) and its example is presented on the figure 3.

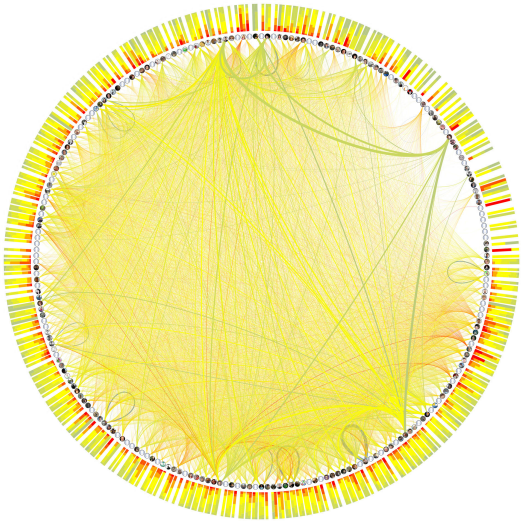

Fig. 3. The visualization "Solntse" (sun, rus.)

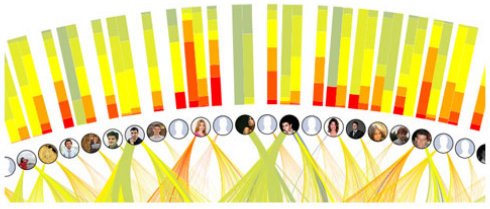

Fig. 4. Scaled-up part of the visualization "Solntse"

The visualization "Elka" allows us to quickly and accurately provide overall picture of estimations between users, and to identify the nature of evaluations of individual users stood out against a background of other users. Thus, for instance, one can see in figure 1, that all users on average have neutrally esti-

mated each other. At the same time, some nodes stand out among them, their estimates are almost completely negative, or, conversely, are positive. Such users, for example, may be taken under special control by facilitators. In addition, such visualization could be used in order to instantly find a negative evaluation conspiracy of a user group against an individual node. This would be expressed in several broad red lines, leading to one of the nodes in the right column, and other its in-edges on average would not have red color.

Unfortunately, the visualization "Elka" cannot identify so-called "mark up" groups, in which an agreement between users on mutual positive estimation exists. Thus, even a group with two members must be a kind of thick green intersecting edges in the visualization, their symmetry check is quite time-consuming process for a large amount of nodes. To solve this problem the visualization "Solntse" can be very suitable, because in this case incoming and outgoing edge ends of a node coincide.

3.2 Idea support graph

Let's consider a restriction of socio-semantic graph $\bar{\mathbb{G}} = (\bar{G}, \bar{C}, A)$, where content $\bar{C}$ contains only one relation $\bar{R}$, which is strict partial order relation on the set $\bar{T}$, and $\bar{G}$ contains also only one relation $\bar{E}$ induced by the ratio of A as follows:

$$v\bar{E}w \Longleftrightarrow \exists t, \tau \in \bar{T} \mid vA\tau \ \wedge \ wAt \ \wedge \ t\bar{R}\tau \ \wedge \ \tau \in \bar{T}',$$

where $\bar{T}'$ — the set of all maximum elements from $\bar{T}$ relatively $\bar{R}$. Then such subgraph $\bar{\mathbb{G}}$ will be called as *idea support graph*. Idea support graph is visualized by WitoAnalytics as follows. The nodes $\bar{V}$ are allocated on an outer concentric circumference, and the nodes $\bar{T}'$ are allocated on an inner concentric circumference. Size of the nodes and their deviation from the line of the circumference corresponds to the number of edges. The visualization is named as "Glaz" (eye, rus.) and its example is presented in figure 5.

3.3 Short review of current WitoAnalytics capabilities

In the current state WitoAnalytics has the following list of capabilities:

- visualizations of user text estimation (5 types, that include both general view of the graph, and individual user view);
- visualizations of user actions like "content creation", "content evaluation", "content commenting" and etc;
- visualization of user group diversity (dendogram visualizations, adjacency matrix visualizations, histograms and densities);
- valued graph clusterization (3 methods, that include random max-clique search algorithm);
- N-gram and word extracting from user content.

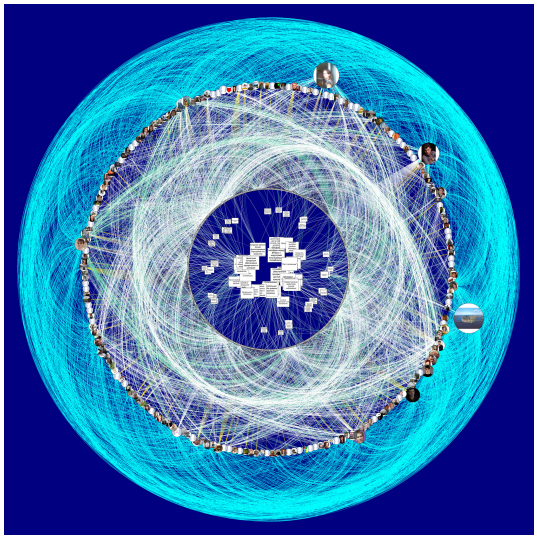

Fig. 5. The visualization "Glaz" (eye, rus.)

4 Prospect

Since Witology is a relatively young company the work on the analysis and visualization of socio-semantic network data of the platform is the unfinished project, in the framework of which one has to solve many analytical problems and problems of visualization known as information visualization problems [10]. They include the following questions:

- what data to visualize, for example, to detect collusion and "mark up" groups of participants for many different subgraphs of the platform;
- how to place nodes and edges;
- which thresholds and for which the parameters of nodes and edges should be set.

References

1. Alexey Drutsa, Konstantin Yavorskiy, *Visualizatsia dannikh sociosemanticheskoy seti*, Dokladi po komputernim naukam i informatsionnim tekhnologiyam, Natsionalniy Otkritiy Universitet "INTUIT", Moscow, 1, 2012, pp.112-118 (in russian).
2. Rostislav Yavorskiy, *Research Challenges of Dynamic Socio-Semantic Networks*, `http://www.witology.com`.
3. Steve Borgatti, Martin Everett and Lin Freeman, UCINET, Analytic Technologies, `http://www.analytictech.com/ucinet/`.
4. Pajek, `http://vlado.fmf.uni-lj.si/pub/networks/pajek/`.
5. Cytoscape, `http://www.cytoscape.org/`.

6. Philip Leifeld, Discourse Network Analyzer, http://www.philipleifeld.de/discourse-network-analyzer/.
7. Auto Map, Casos, http://www.casos.cs.cmu.edu/projects/automap/.
8. NodeXL, CodePlex, http://nodexl.codeplex.com/.
9. Z. V. Apanovich, *Metodi visualizatsii informatsii - naukoemkoe napravlenie IT*, Komp'uternie instrumenti v obrazovanii, No 2, 2010 (in russian).
10. Z. V. Apanovich, *Ot risovania graphov k visualizatsii informatsii*, (preprint) Novosibirsk, 27 p., 2007 (http://www.iis.nsk.su/files/preprints/148.pdf) (in russian).