Description of the experimental protocol

Introduction

The experimental protocol is a procedure or a set of procedures for carrying out an experiment. Thus, it includes the description of the different conditions and the course of the experiment.

The experiment consists in participating in a research project by completing a series of tasks and then answering a questionnaire regarding the usability of the project on the one hand, and the user preferences on the other hand.

1. Goal

Graphs are most often used to represent complex systems such as social networks and biological networks. Thus, the visualization of a graph can help to better understand the structure of the data. For this purpose, there are many methods of graph visualization. Using different methods to visualize a graph can provide very different visual appearances. Thus, our study consists, on the one hand, in analyzing the usability of 3D to explore and analyze data represented as graphs, and on the other hand, in identifying the best representation to visualize graphs.

2. Studied factors

2D radial representations are projected onto three differents 3D surfaces:

- ✓ The half-sphere
- ✓ The cone
- ✓ The torus portion

Thus, for 2D and each of the 3D surfaces, each participant will have to complete a certain number of tasks related to graphs. We therefore propose the following three tasks:

- **Task 1** : Find the node that has the greatest degree among the most central node's neighbors.
- **Task 2** : Find a least central node that has at least 2 neighbors.
- **Task 3** : Find a node of degree at least 3 and that has the highest clustering coefficient except 100%

3. Experimental Material

The experiment is done using a WebGL version of our prototype. So, each participant could perform the experiment remotely on his own laptop with the following minimum characteristics

- RAM : 1Go
- Screen size: 14 inch

4. Implementation

a. Graphs to use

For this research project, we need 6 different graphs that have equivalent topological characteristics. For this, we will use the Stochastic Block Model (SBM) graph generation algorithm present in the *igraph* network analysis library. These 6 graphs are respectively used for the 2D representations (uniform, central emphasis and peripheral emphasis) and the three 3D projection surfaces (half-sphere, cone and torus portion). We then built 24 configurations with the various representation surfaces so that each surface and graph is performed at least once as first, using the concept of the Latin square. A square Latin is an *N*x*N* array filled with *N* different symbols in such a way that each symbol occurs exactly once in each row and exactly once in each column. But our configurations follow a distribution order between the 2D and the 3D surfaces. Indeed, if a configuration starts with the 2D surfaces (and ends with the 3D surfaces) and the first surface is the one that emphasizes the center, then the first 3D surface will be the torus portion, since it is the one that highlights the center. So, our configurations are built so that some participants start with the 2D surfaces and end with the 3D surfaces. Others start with the 3D surfaces and end with the 2D surfaces.

b. Course of the experiment

Each participant had approximately one hour and a half to complete three tasks on each surface (2D and 3D). The purpose of our evaluation is to show that the 3D proposals could be better to explore and to analyze graphs whatever the interest (the central or peripheral nodes, the dense areas), compared to the 2D representations, since they optimized the spatial distribution of nodes and we improved the edges drawing by

projecting them onto the surfaces. So, there could be less time in exploration, less clicks and more accurate responses to different tasks, because the perception of the nodes connectivity is improved. Moreover, we want to analyze the usability of the 3D for exploring and analyzing graphs. On the other hand, we want to identify the best representation that could be used to visualize graphs.

At the end of the experiment, each participant completes questionnaires related to the usability of the system and the user experience. Since our experiment is done remotely, we organized a video conference for each participant in order to supervise the experiment's process. The experiment consists of a training phase and an evaluation phase. Before starting the training phase, each participant is instructed about the experiment procedure, its environment, navigation and interaction techniques. For example, when the mouse hovers a node, a tooltip shows its clustering coefficient value and its degree. He is also given the essential notions about graphs in order to ensure that he has the useful knowledge for the experiment. In the training phase, the participant is asked to perform the above tasks on a small graph (the karate club's graph) and on each surface. Once familiar with the system, he moves on to the evaluation phase, but with generated graphs. If the participant is ready to start the training or the evaluation, he clicks on a start button to see the first task to complete and the next task is automatically displayed after validating the previous task's response.

5. Measured variables

a. Variables

During the experiment, we measure these following variables:

- The execution time of the task
- The number of clicks
- The efficiency score

b. User performance evaluation

Ci-dessous les formules permettant d'évaluer les performances de l'utilisateur pour chacune des tâches de l'expérimentation :

- Task 1:

$$score_i = 100 * (deg_i / deg_{ideal}) \; if \; d(ctr, i) = 1, \; otherwise \; score_i = 0$$

where deg_i is the degree of the selected $node_i$. deg_{ideal} is the greatest degree among the central node's neighbors and $d(ctr,i)$ is the shortest distance between the central node and $node_i$. Thus, $node_i$ must be directly connected to the central node, i.e. $d(ctr,i)$ must be equal to 1.

- Task 2:

$$score_i = 100 * (1 - c_i)/(1 - c_{ideal}) \; if \; c_{ideal} \neq 1, \; otherwise \; score_i = 0$$

where c_i and c_{ideal} are respectively the centrality value of the $node_i$ and that of the ideal node. Furthermore, the score is 0 if the degree of the selected node is less than 2. Indeed, it is easy to check that the degree of the selected node is at least 2. Thus, the score is 0 if the condition is not met. Otherwise, the score varies from 0 at the center to 1 for a node of degree at least 2 and the most on the periphery.

- Task 3: score_i = 100 * ccf_i

$$score_i = 100 * (ccf_i - ccf_{worst})/d \; if \; d > 0, \; otherwise \; score_i = 0$$

where $d = ccf_{ideal} - ccf_{worst}$. ccf_i, ccf_{worst} and ccf_{ideal} are respectively the clustering coefficient of the $node_i$, the worst clustering coefficient and the highest clustering coefficient except 100%. So, the score is 0 if the degree of the selected node is less than 3 or if the clustering coefficient of the selected node is 100%.

6. Statistical processing of results

In addition to the measured variables, the responses on the usability of the 3D radial visualization and those related to user preferences will be analyzed. The results of this analysis will allow us to validate or not the usability of our approach. Moreover, we could identify the most appropriate representation or combination of representations to visualize a graph according to its size and topological structure.