EuroRV3: EuroVis Workshop on Reproducibility, Verification, and Validation in Visualization (2017)
K. Lawonn, N. Smit, and D. Cunningham (Editors)

Where'd it go? How geographic and force-directed layouts affect network task performance Supplemental Information

Scott A. Hale[1,2], Graham McNeill[1], and Jonathan Bright[1]

[1]Oxford Internet Institute, University of Oxford, UK
[2]Alan Turing Institute, UK

1. Pseudo-geographic layout

For our pseudo-geographic layout, we employed a simple approach using multidimensional scaling (MDS) to avoid overly dense clusters while maintaining relative geographic positions. Though typically used for data visualization and dimensionality reduction, MDS is related to force-directed algorithms [Nor04, VK06] and its ability to generate network layouts was demonstrated by Kruskal and Seery almost forty years ago [KS80]. More recently, Venna and Kaski [VK06] have used their neighborhood-based "Local MDS" algorithm for visualizing gene interaction networks.

Given a matrix of pairwise distances, MDS produces a set of points whose pairwise Euclidean distances are as close as possible to those in the original matrix. In our case, we computed the Euclidean distance matrix between nodes, added a number $\lambda > 0$ to every entry (except those on the diagonal) and then computed new coordinates for the nodes using MDS. By adding λ to each entry, the relative increase in distance was greater for pairs of points that were close together and hence, clusters were "stretched out."

There are many variants of MDS. We used the well-known Sammon mapping [Sam69], which focuses on the distances between nearby points. Figure 1b in the main paper demonstrates the approach for the nodes of the network used in our experiments. In this case, using $\lambda = 150$ km spreads out clusters whilst retaining significant geographic information. Our main focus is on comparing geographic and force-directed layouts, but this simple pseudo-geographic layout, which we expect will perform better than a pure geographic layout, provides a stronger, more realistic comparison to the force-directed layout.

2. Experiment detail

Prior to starting each task, each participant was given a concise introduction to the requisite concepts and terminology. For the degree and distance tasks, the participant was then asked three pretest questions about a simple artificial network, followed by three questions about a geographic network representing popular rail commuting routes between UK local authorities. (London was represented by a single node within this network even though it is in fact composed of multiple local authorities.) For the node search task, the participant was asked to find three local authorities on a map as well as three nodes (local authorities) within the commuting network.

All participants completed all three tasks. There were six questions in each section (three pretest and three main questions), and hence eighteen questions in total. We randomized the sequence of the tasks (node search, degree, and path length) to avoid order effects. We had participants first answer three questions on the pretest networks and then perform the same task answering another three questions on the main, commuting network in order to avoid extra mental effort from task-switching. Each participant was assigned to a single layout type at random (i.e., force-directed, geographical, or pseudo-geographical), and we used the same layout for the participant throughout all main sections. All layouts were computed in advance of the experiment; thus, although the Fruchterman-Reingold force-directed layout algorithm is non-deterministic all subjects in the force-directed condition saw the same realization of the layout (Figure 2).

Selecting target nodes for the tasks is not an easy process. We hand-picked nodes and tried to make all questions have a similar, medium level of difficulty. For example, within the node search task, we used well-known cities so that participants were more likely to be able to use their knowledge of UK geography.

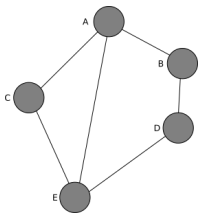
3. Full list of questions

- Node search
  - Pretest: This section will ask you to locate three UK cities on a map of the country. (Figure 3)
  - Main: This set of questions will ask you to locate three UK cities on a network diagram. Each city is represented as a node (circle) within the network.
  - Participants were given three cities randomly from the following list. There was no overlap between the cities in the pretest and main tasks.
    - Birmingham

  - Bristol
  - Cardiff
  - Edinburgh
  - Glasgow
  - Leeds
  - Leicester
  - Manchester
  - Nottingham

- Path length
  - Overall instructions:
    This set of questions will ask you to determine the network distance between two nodes (circles). Distance is measured as the number of edges (lines) between two nodes.
    Network distance is measured as the number of edges (lines) between two nodes. Two nodes connected directly are said to have a distance of 1. If it is only possible to go from one node to another through a third node, then this is distance 2 (A->B->C). If it is only possible to go from one node to another through two additional nodes, this is distance 3 (A->B->C->D). *(An image was included to demonstrate distance.)*
  - Prompt and possible answers for each individual question:
    What is the distance between the two orange nodes?
    1, 2, 3, 4, Other
  - Pretest: Participants were given three graphs generated according to the Watts-Strogatz network model with 15 nodes, each starting with degree 3, and a rewiring probability of 0.1. Selected nodes had a path length of 1, 2, or 3.
  - Main: Participants were given three of the following pairs at random (correct answers in parentheses):
    - Manchester, London (length 1)
    - Birmingham, Leeds (length 2)
    - York, Leicester (length 3)
    - Edinburgh, London (other/disconnected)

- Degree
  - Overall instructions:
    In this section you will be shown network diagrams and asked to compare the "degree" of two nodes (circles).
    Degree refers to the number of edges (lines) connected to each node (circle). **A node with more edges (lines) has a higher degree.** As an example, we can say that node A (degree 3) has a higher degree than node B (degree 2). Nodes A and E have the same degree (3).



  - Prompt and possible answers for each individual question:
    Look at the green and orange nodes (circles) in the network: Which node has a higher degree (more connections)?
    Orange, Green, I don't know
  - Pretest: Participants were given three graphs generated according to the Watts-Strogatz network model with 15 nodes, each starting with degree 3, and a rewiring probability of 0.1.
  - Main: Participants were given three of the following pairs at random (correct answers in parentheses):
    - Manchester, London (London/Green)
    - Birmingham, Leeds (Birmingham/Orange)
    - London, Glasgow (London/Orange)
    - Cardiff, Leicester (Cardiff/Orange)
    - Edinburgh, Glasgow (Glasgow/Green)

The experiment was browser based. All code for the experiment was written in HTML5/CSS/JavaScript by the authors, and existing libraries, such as sigma.js for network layout, were used where possible. All the code is open-source and available freely for replication or to adapt for future experiments (`http://www.github.com/oii-nexus/qa/`). Figure 1 shows screenshots of the degree task fully zoomed out. Participants were able to pan and zoom all visualizations. In addition, the name of a node (local authority name) was shown when participants moused over a node within the main sections of the experiment as shown in Figure 2.

References

[KS80] KRUSKAL J. B., SEERY J. B.: Designing network diagrams. In *Proc. First General Conf. on Social Graphics* (1980), pp. 22–50. 1

[Nor04] NORTH S. C.: Drawing graphs with neato. *NEATO User Manual 11* (2004). 1

[Sam69] SAMMON J. W.: A nonlinear mapping for data structure analysis. *IEEE Transactions on computers 18*, 5 (1969), 401–409. 1

[VK06] VENNA J., KASKI S.: Visualizing gene interaction graphs with local multidimensional scaling. In *ESANN* (2006), vol. 6, pp. 557–562. 1

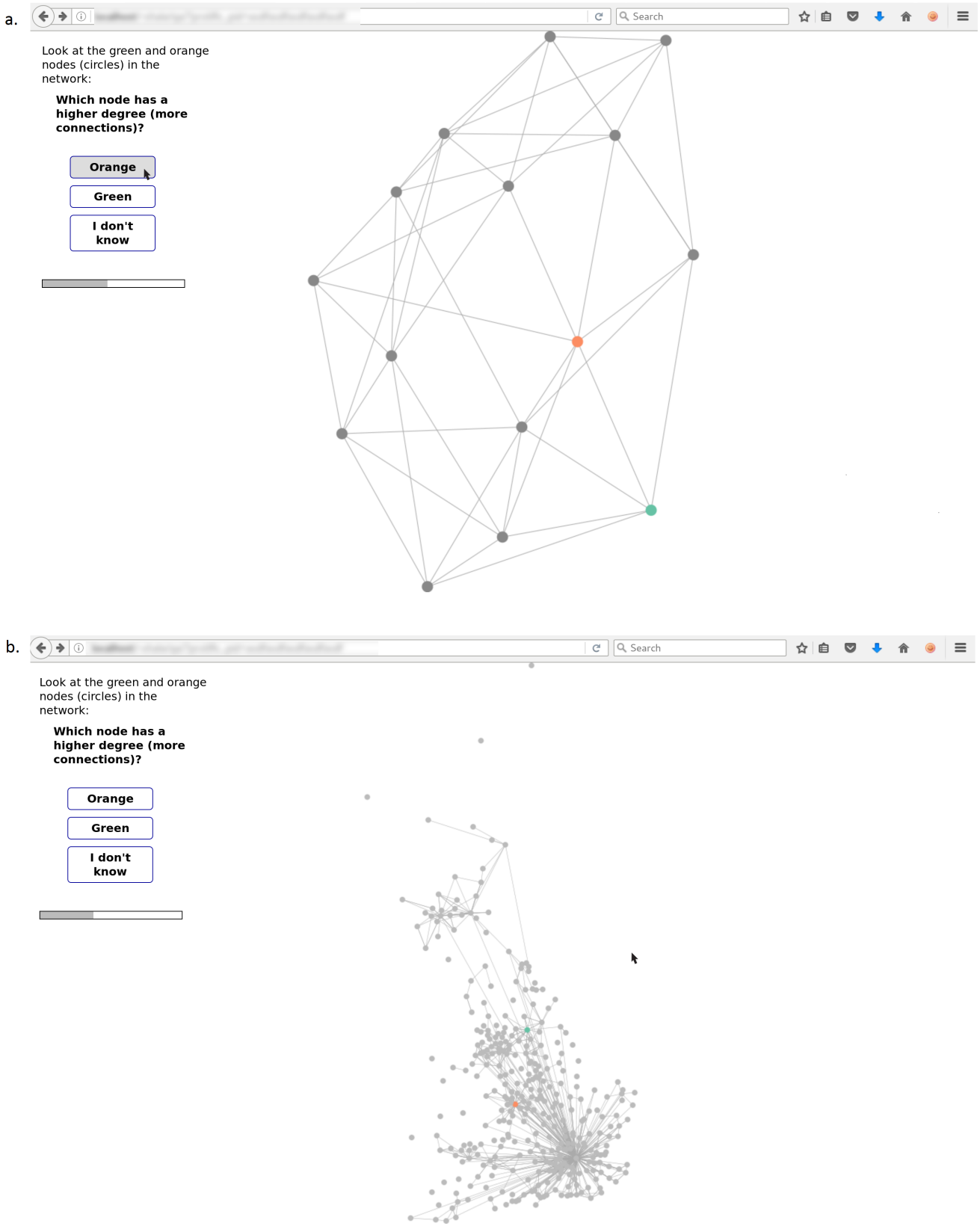


Figure 1: *Screenshots of questions from the degree task for (a) pretest and (b) main sections in the geographic layout condition.*

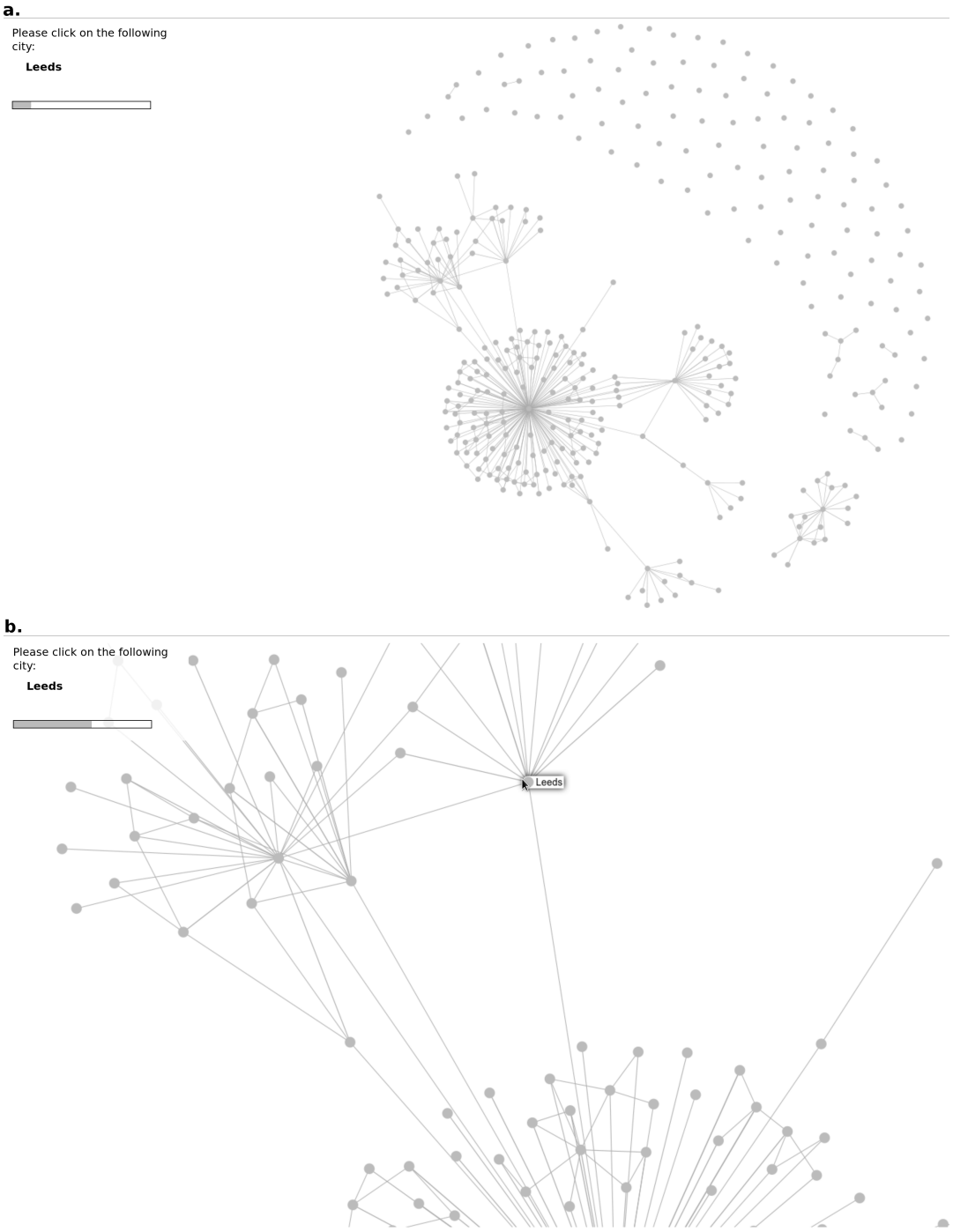


Figure 2: *Screenshots of questions from the node search task in the force-directed layout. (a) Starting interface and (b) after the subject has zoomed and panned to find the target node.*

Figure 3: *A pretest question for the node search task. The map could be zoomed and panned, and the name of a local authority was shown when participants moused over it.*