Network Visualization Literacy

Indiana University Network Science Institute
Open Science Forum
November 1, 2017

Angela Zoss

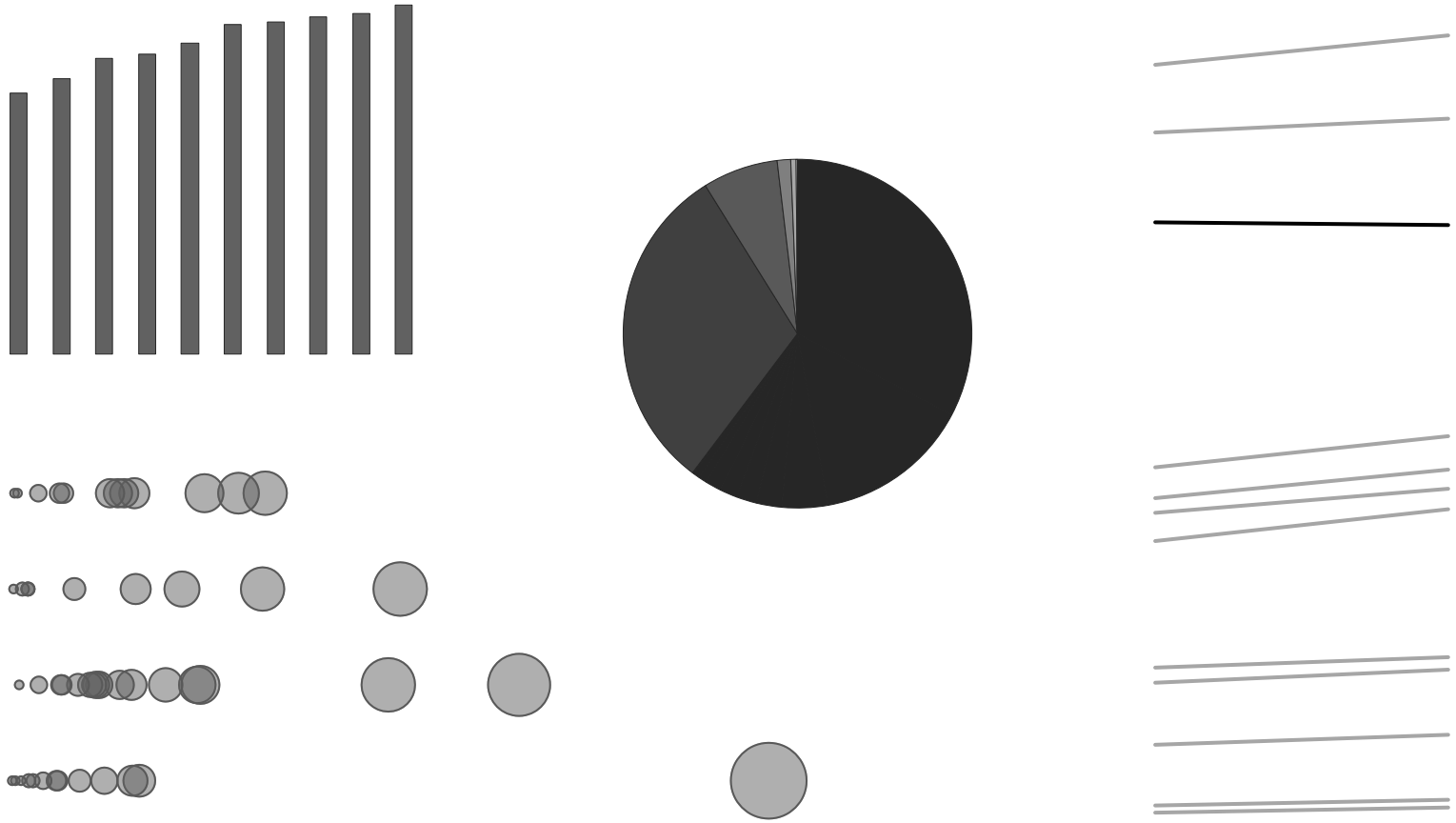
Doctoral Candidate
Department of Information and Library Science

netvislit.org

Outline

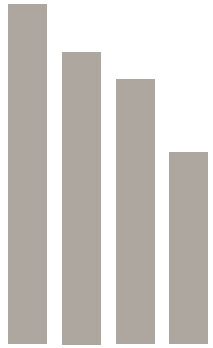
- What is visualization literacy?
- What is network visualization literacy?
- Selecting tasks for NetVisLit
- Results of NetVisLit performance studies

Visualization Literacy



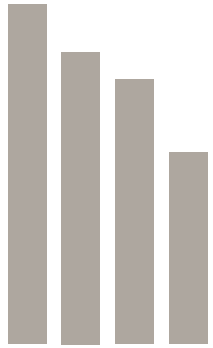
http://guides.library.duke.edu/vis_types

How well can people read visualizations?

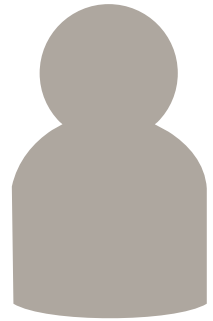


Chart

How well can people read visualizations?

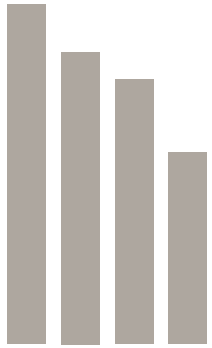


Chart

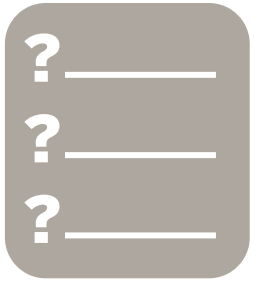


User

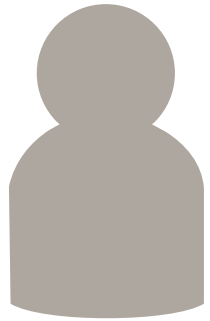
How well can people read visualizations?



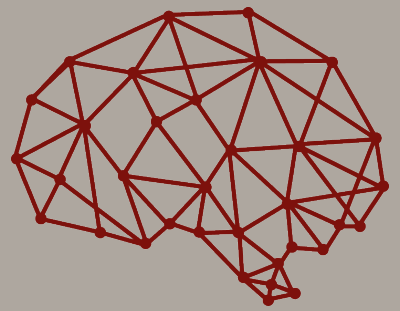
Chart



Task

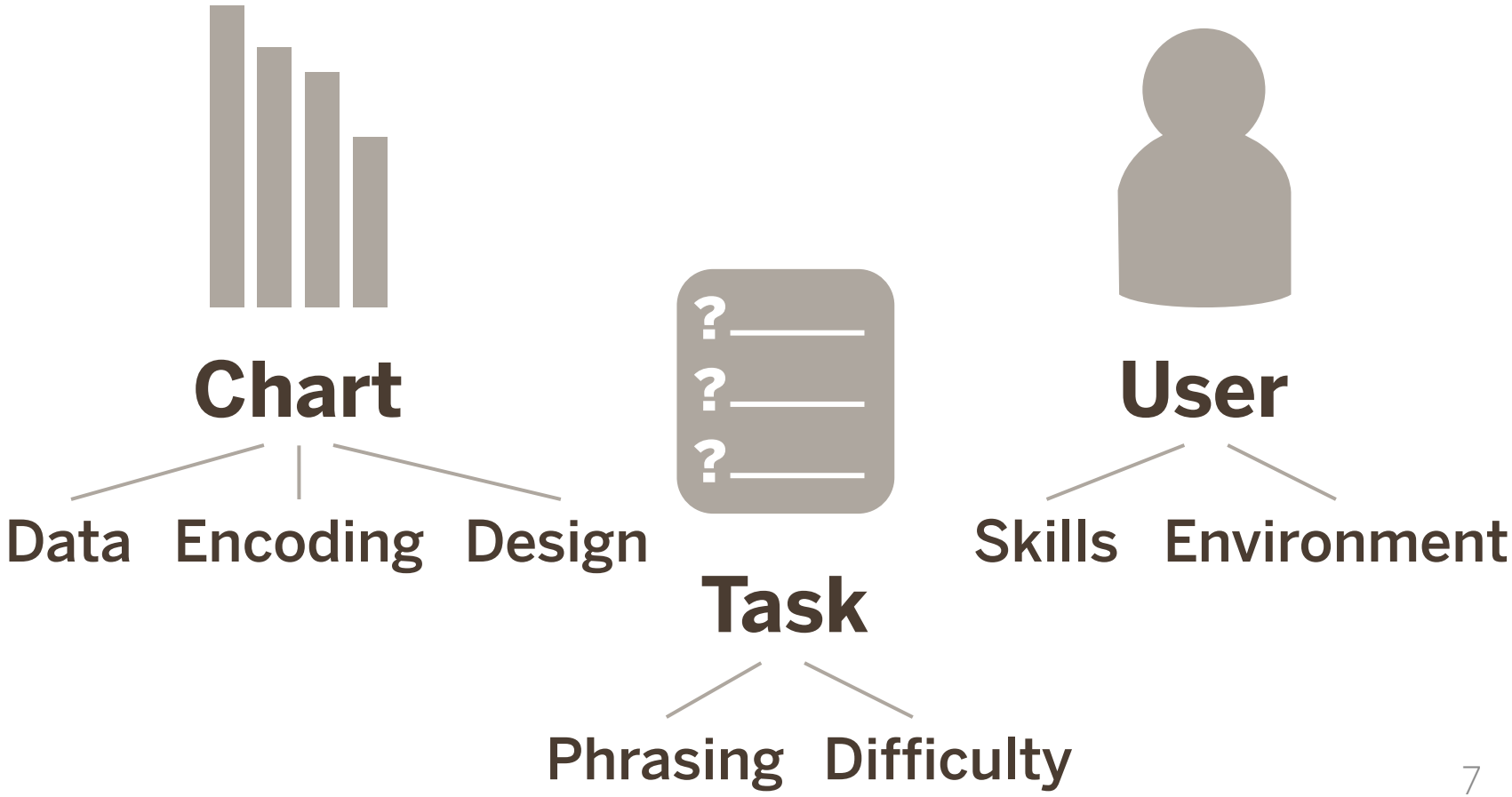


User

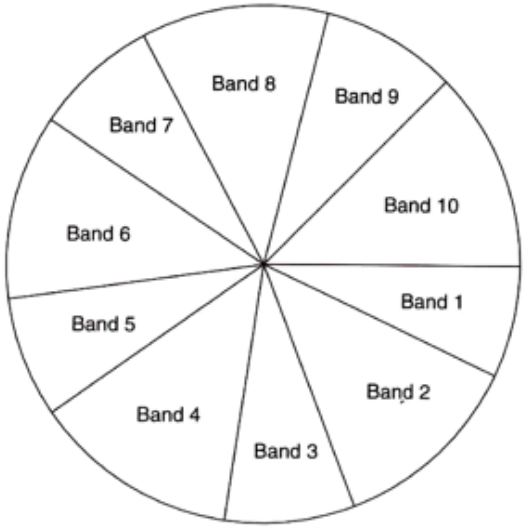


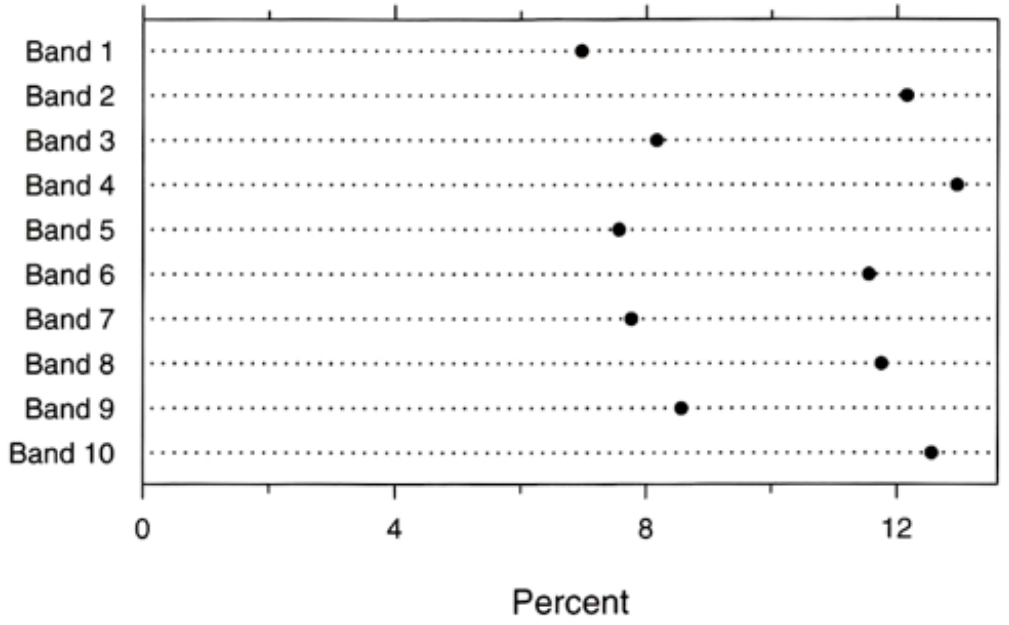
netvislit.org

How well can people read visualizations?



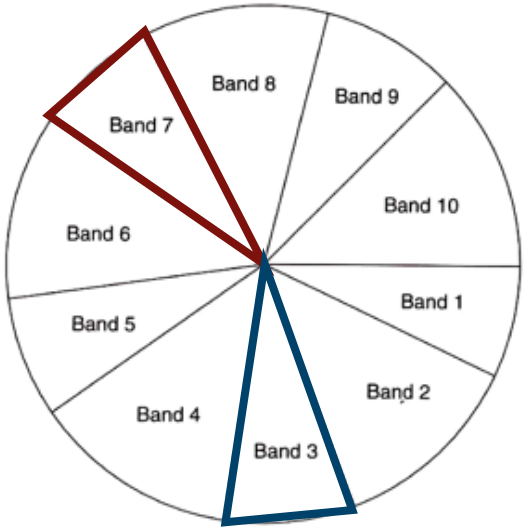
Encoding

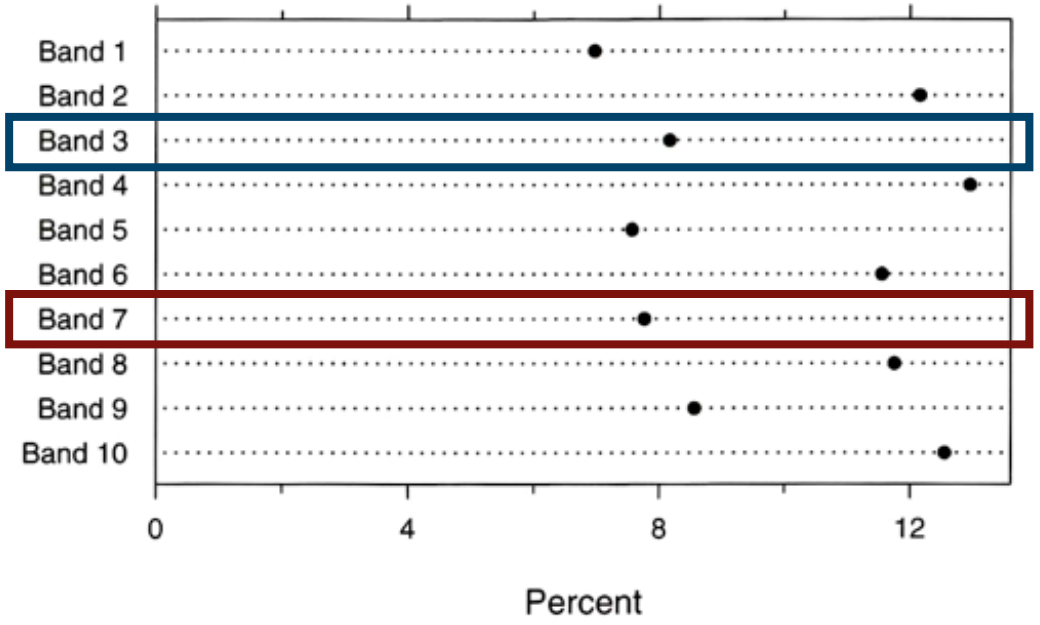




Cleveland (1994)

Encoding



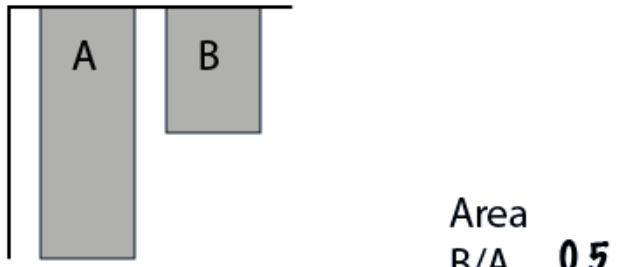


Cleveland (1994)

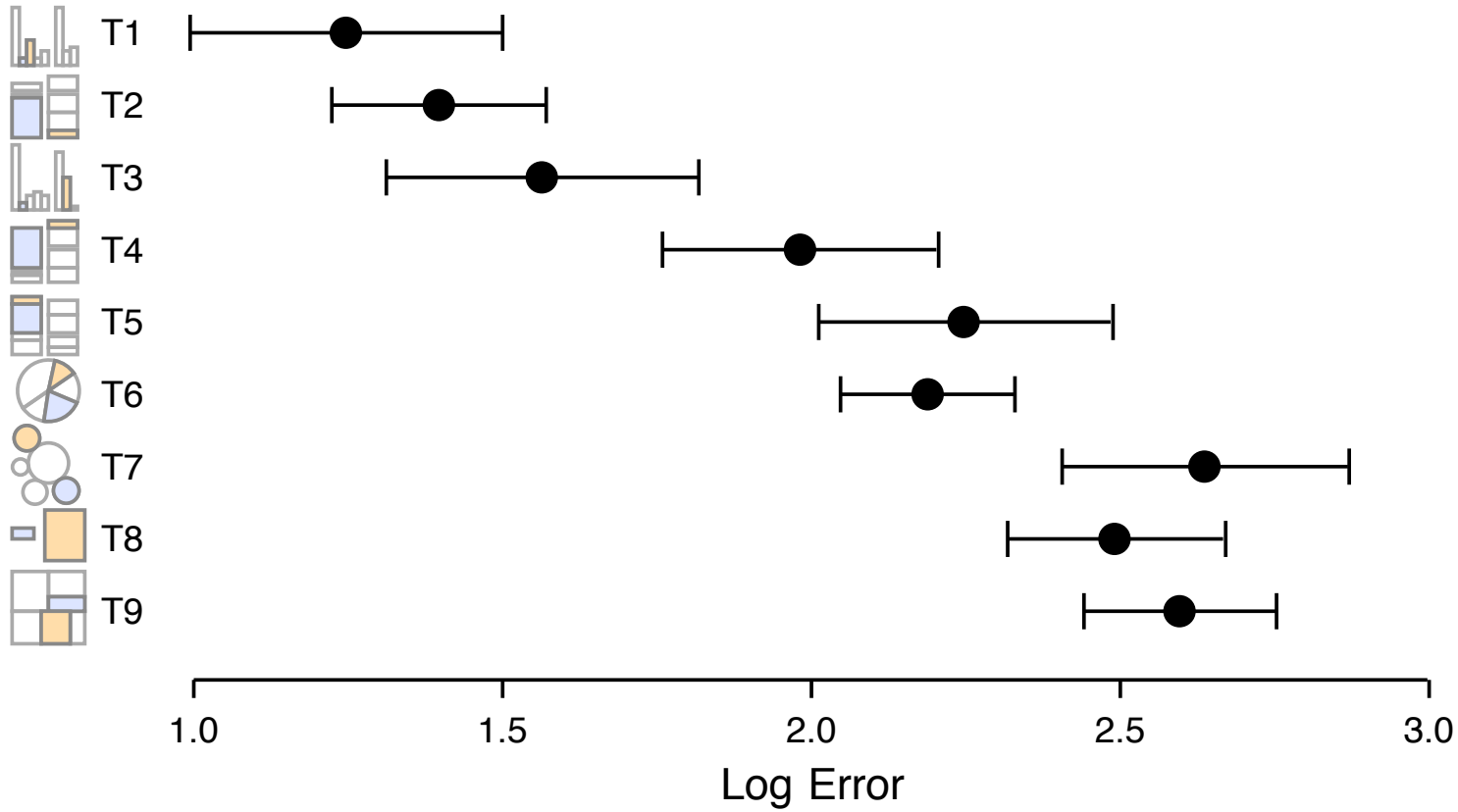
Encoding

Based on a visual inspection, guess what is the area fraction of B/A for each pair?

For example:



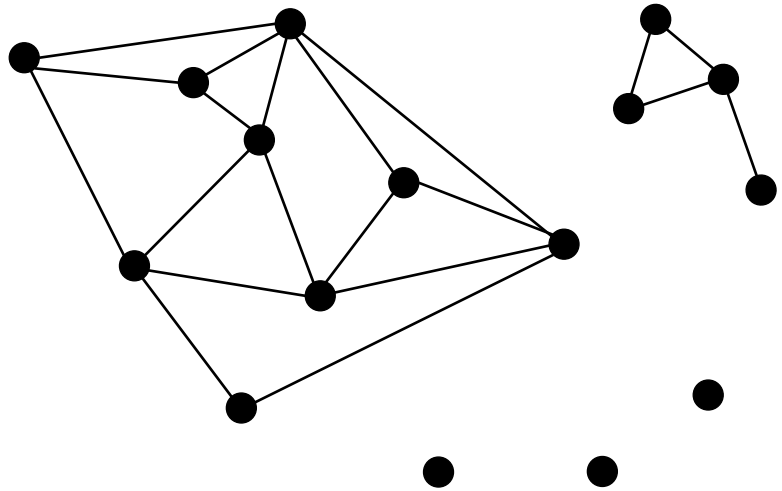
Encoding

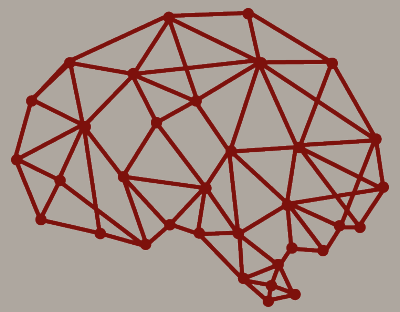


Heer & Bostock (2010)
http://vis.stanford.edu/files/2010-MTurk-CHI.pdf

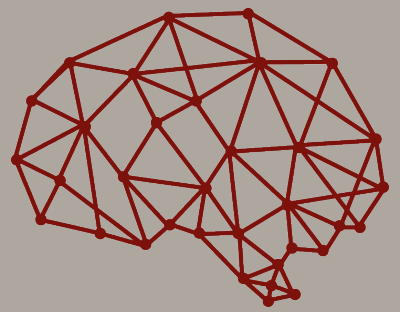
Network Visualizations

	A	B	C	D
A		1	0	1
B	1		1	0
C	0	1		0
D	1	0	0	





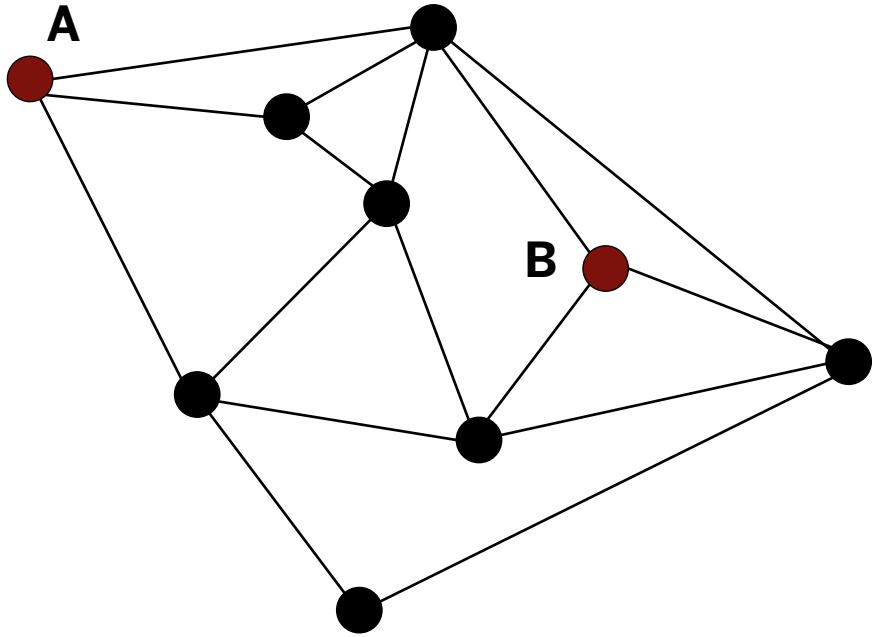
netvislit.org



netvislit.org

Network Visualization Literacy Tasks

Shortest path between A and B?



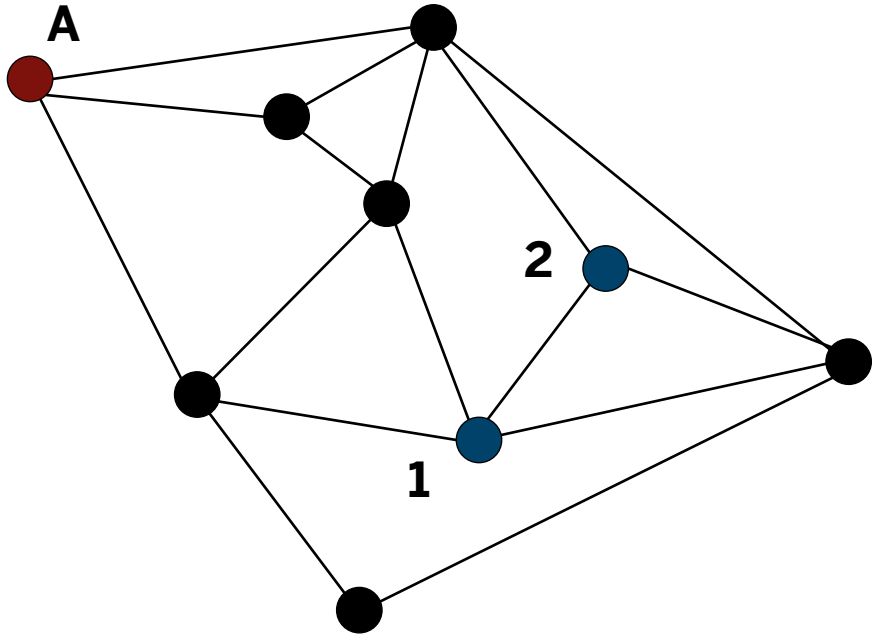
Design variations:

- edge bends
- edge crossings
- layout symmetry
- link exit angle
- orthogonal grid
- path continuity

Purchase, 1997; Purchase, 2000;
Purchase, Carrington, & Allder, 2002;
Purchase, Cohen, & James, 1997;
Ware, Purchase, Colpoys, & McGill, 2002;
Huang, 2013; Huang, 2014;
Huang, Eades, Hong, & Lin, 2013;
Huang & Huang, 2011;
Huang, Huang, & Lin, 2016

Network Visualization Literacy Tasks

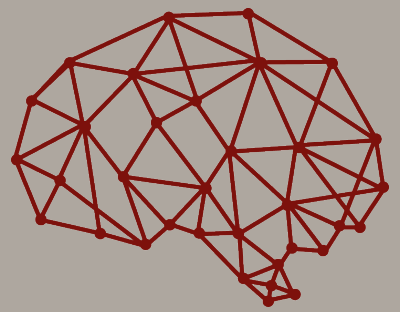
Which is more similar to A?



Design variations:

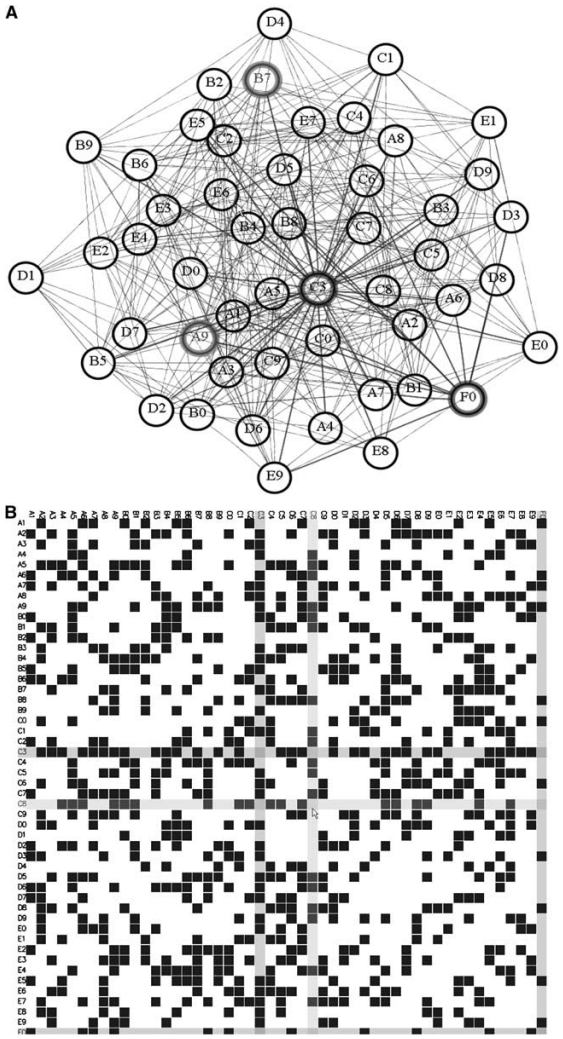
- Euclidean distance
- Measured path length
- # nodes on path
- Design of path (width, darkness, hue)

Fabrikant & Montello, 2008;
Fabrikant et al., 2004;
Fabrikant, Ruocco, Middleton, Montello, & Jörgensen, 2002



netvislit.org

Network Visualization Literacy Tasks



Tasks:

- # nodes
- # edges
- most connected node
- find node by label
- find link by label
- find common neighbor
- find path between node

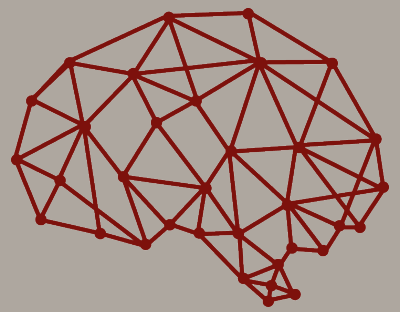
Design variations:

- data sizes (20, 50, 100 nodes)
- data densities (.2, .4, .6)
- layout conditions (node-link, matrix)

Ghoniem, Fekete, & Castagliola, 2005

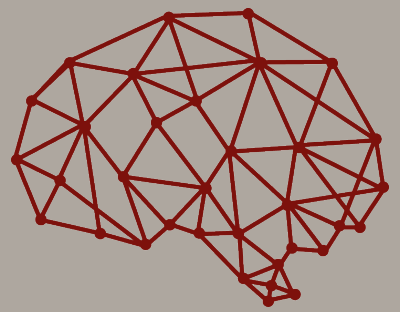
Gaps

- Tasks based on real-world network usage
- Small changes in graphic design
- Layouts vs. tasks
- Differences in user experience



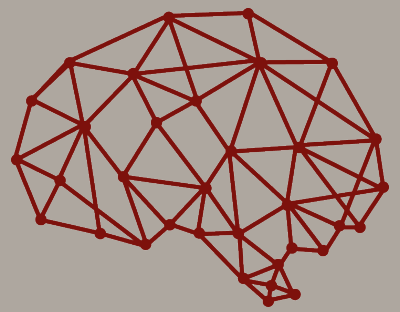
Statement of purpose

The purpose of this research is to explore the tasks for which network visualizations are best suited, taking into account the experience level of the user and the properties of the visualization.



Study A:
Network science task selection

What network measures do network science experts consider most important? Easiest to estimate using a visualization?

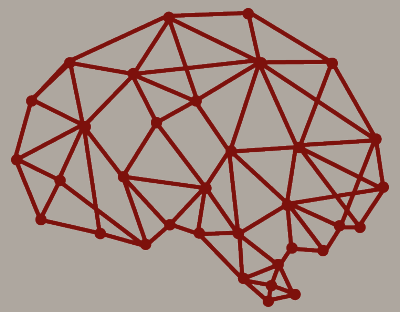


Candidate tasks

Level	Candidate task
Element (node)	1. Closeness Centrality 2. Eigenvector Centrality 3. Node Betweenness Centrality 4. Node Degree
Element (link)	5. Link Betweenness Centrality 6. Loops
Small groups	7. Component Size 8. Modularity 9. Number of Components
Full network	10. Average Degree 11. Average Path Length 12. Average Shortest Path 13. Clustering Coefficient 14. Density 15. Diameter 16. Number of Links 17. Number of Nodes

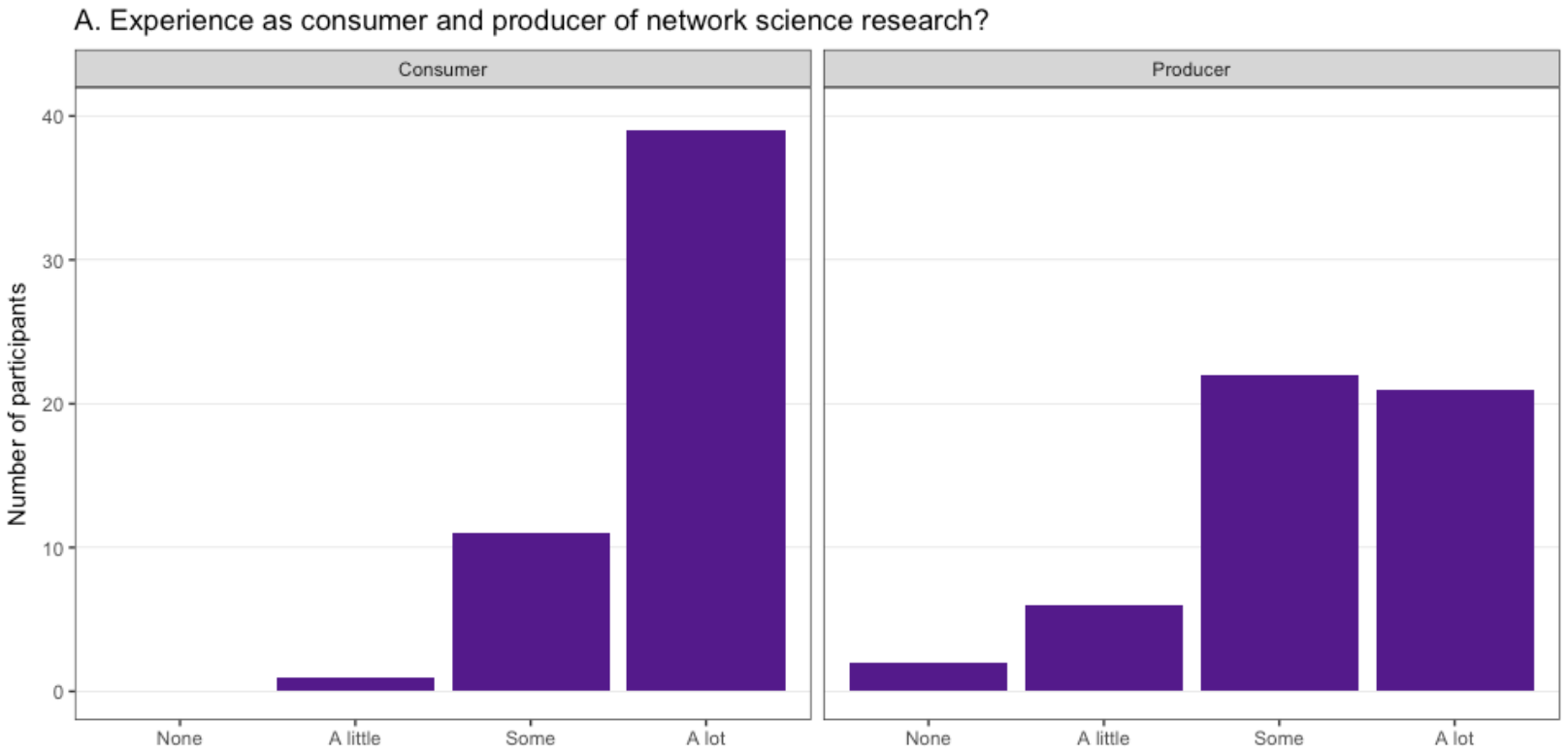
Questions

- How **important** are these measures to your research?
- How likely is it that you would be able to **estimate** these measures from a visualization?



Participants

Open invitation to SOCNET listserv (n=51)

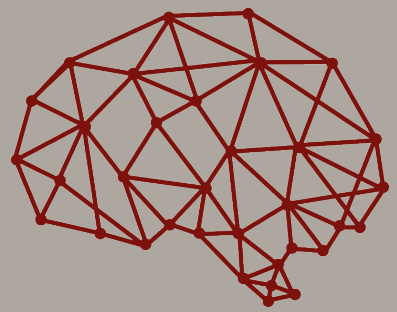


VISUALIZATION LITERACY

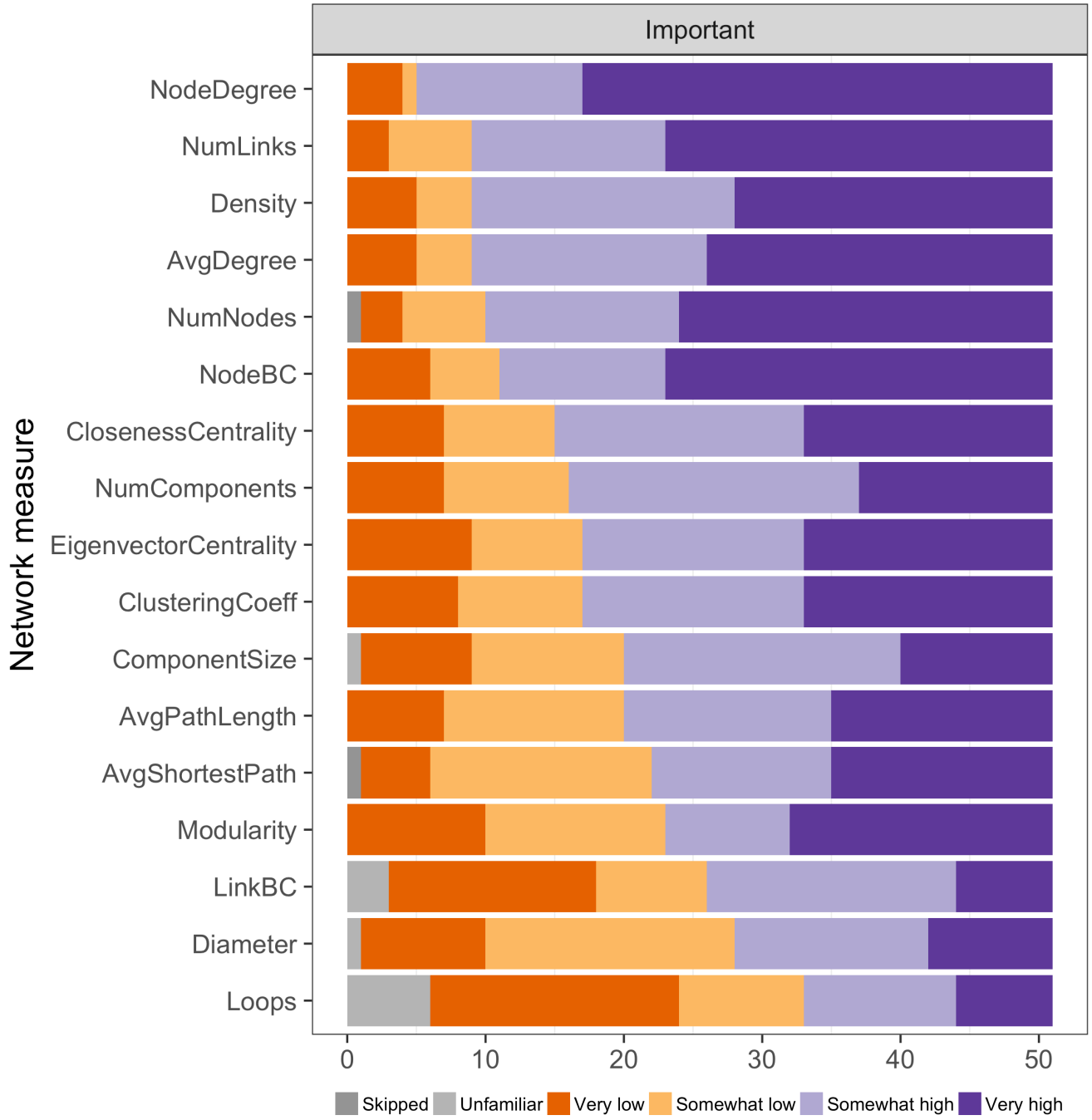
NETVISLIT

TASK SELECTION

TASK PERFORMANCE



netvislit.org

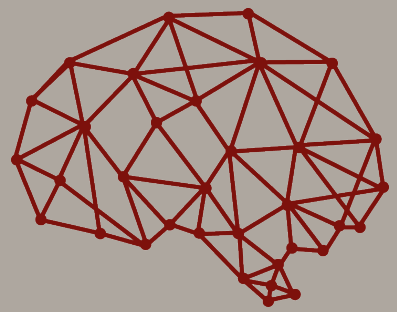


VISUALIZATION LITERACY

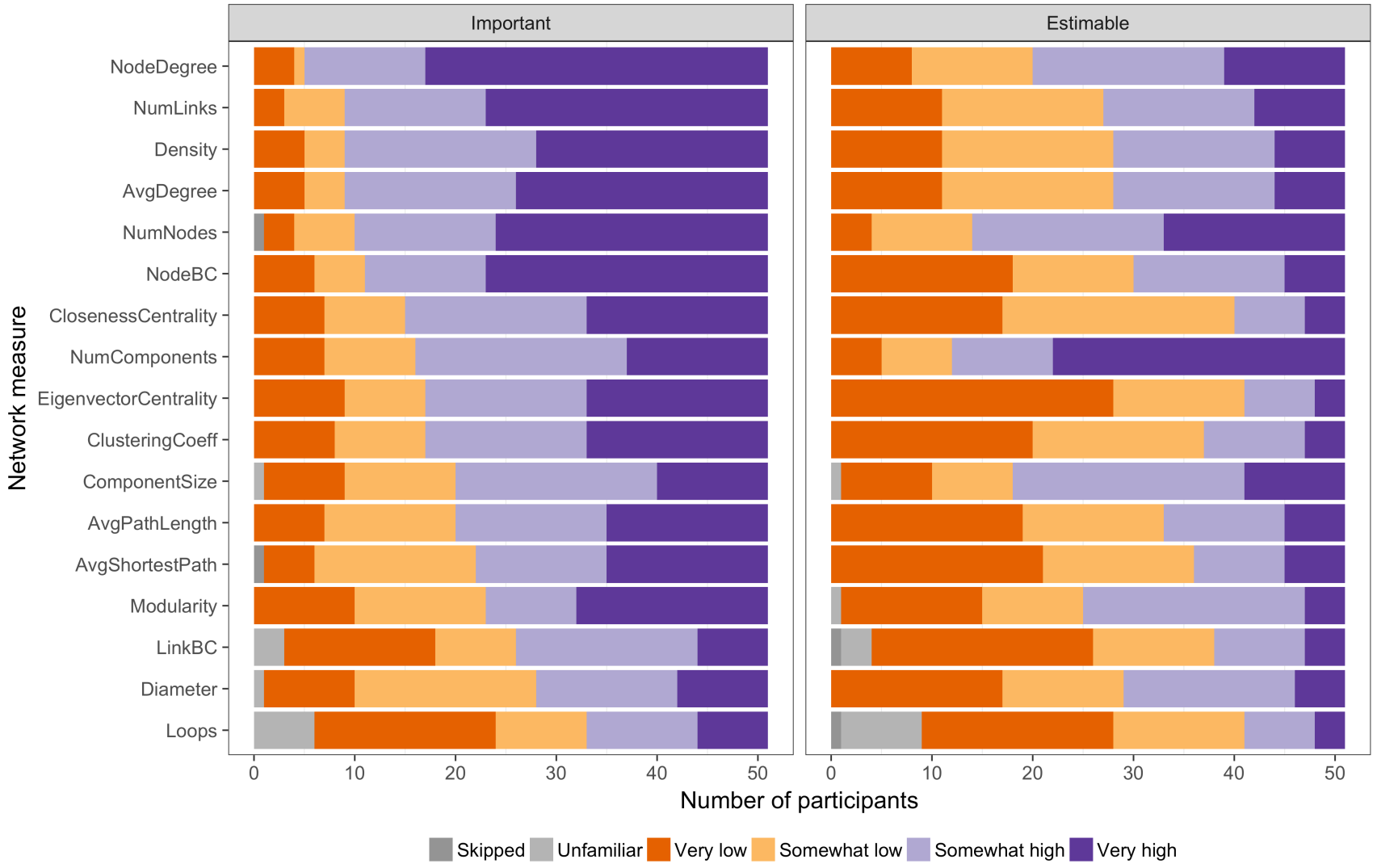
NETVISLIT

TASK SELECTION

TASK PERFORMANCE



netvislit.org

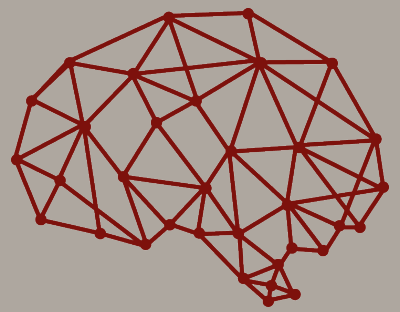


VISUALIZATION LITERACY

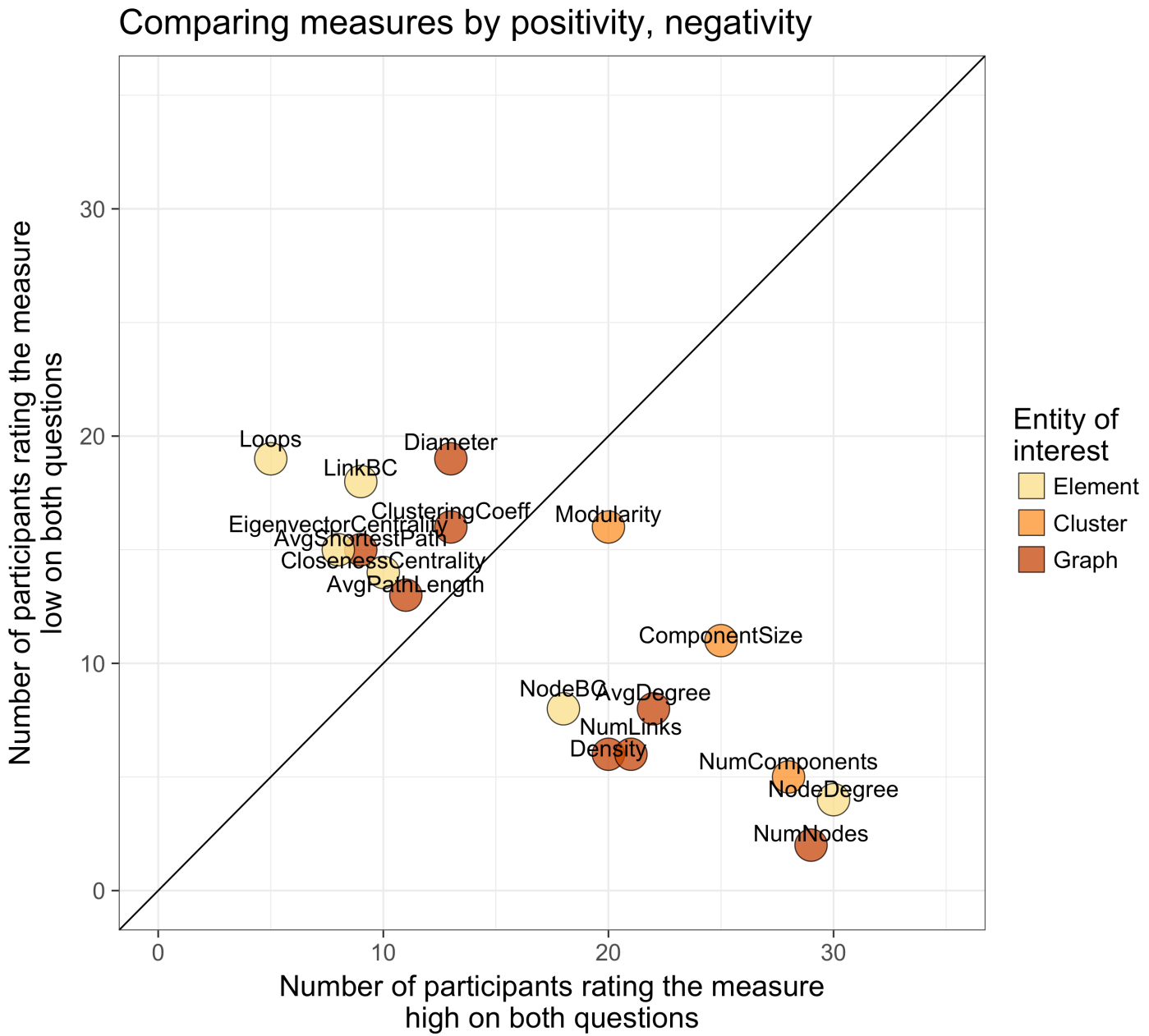
NETVISLIT

TASK SELECTION

TASK PERFORMANCE

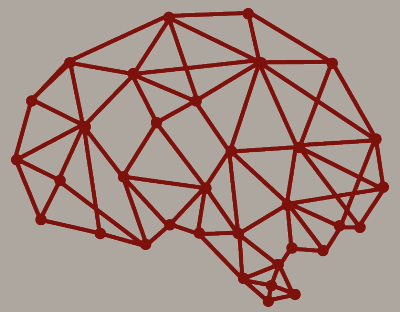


netvislit.org



Final tasks

Level	Task
Element (node)	1. Node Degree 2. Node Betweenness Centrality
Small group	3. Number of Components 4. Component Size
Full network	5. Number of Nodes 6. Number of Links 7. Average Degree 8. Density

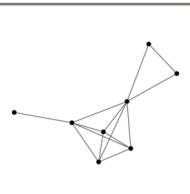
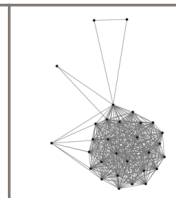
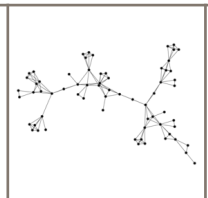
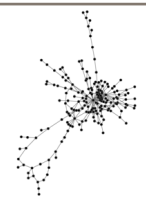
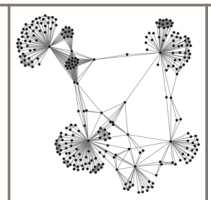
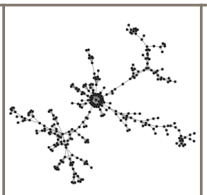
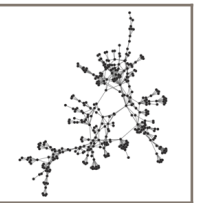


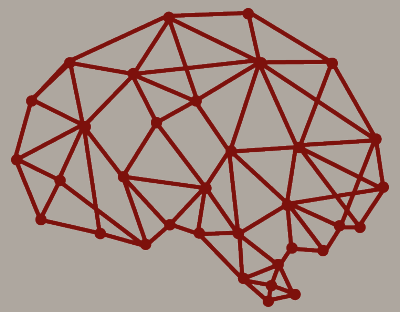
Study B: Visualization task performance

How do a network's **properties** (e.g., number of nodes, density), **design** (e.g., color, size, layout) or **context** (e.g., concrete vs. abstract question phrasing) affect the ability of users to interpret the visualization?

Network datasets

7 real-world datasets

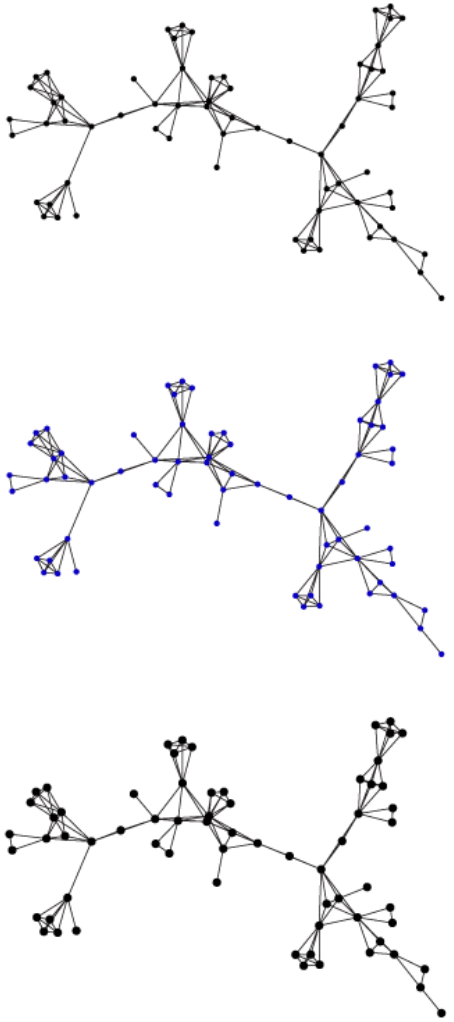
network							
nodes	8	30	67	184	270	321	379
edges	14	337	143	246	932	583	914
density	0.5	0.775	0.065	0.015	0.026	0.011	0.013

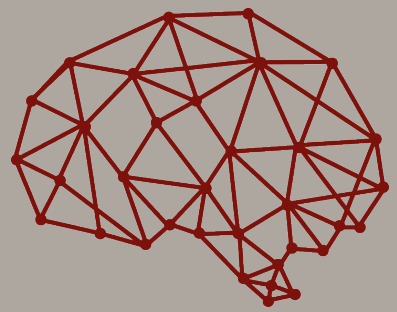


netvislit.org

Graphic Conditions (between subjects)

- Baseline
  NLD with GEM layout
- Concrete phrasing
  Using "person" and "relationship" rather than "node" and "link"
- Color
  add a solid color to the nodes
- Size
  make all nodes slightly larger



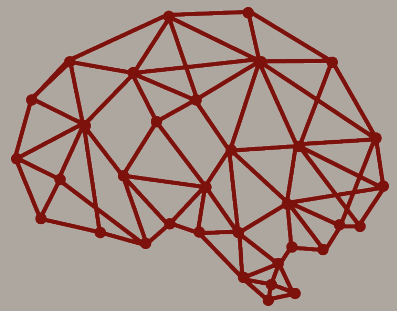


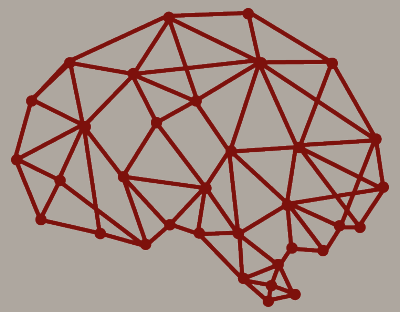
Task phrasing – formal

Measure Name	Question Phrasing (Technical)
Node degree	Find the node with the most links. About how many links does it have? Click on the node with the most links. (Your last click will be the only click recorded.)
Node betweenness centrality	Find any nodes that bridge gaps between clusters, rather than being closely connected to a single cluster. Click on each of those nodes. (If you see a lot of these nodes, please choose at most five that seem to be clear examples.)
Cluster detection	If you were asked to estimate the number of clusters in this network, about how confident would you be in your estimation?
Number of unconnected components	How many clusters do you see in this network? Please type the number below.
Component size distribution	Find the largest cluster in the network, and look at the nodes in that cluster. What percentage (approximately) of the total nodes in the network can be found in the largest cluster?
Number of nodes	About how many total nodes are in this network?
Average degree or degree distribution	About how many links does each node in this network have, on average?
Number of links	About how many total links are in this network?

Task phrasing – informal

Measure Name	Question Phrasing
Node degree	Find the most popular person. About how many friends does he or she have? Click on the person with the most friendship connections. (Your last click will be the only click recorded.)
Node betweenness centrality	Find any people who bridge gaps between friend groups, rather than being closely connected to a single friend group. Click on each of those people. (If you see a lot of these people, please choose at most five who seem to be clear examples.)
Cluster detection	If you were asked to estimate the number of tightly-knit friend groups in this community, about how confident would you be in your estimation?
Number of unconnected components	How many tightly-connected friend groups do you see in this community? Please type the number below.
Component size distribution	Find the largest friend group in the network, and look at the people in that group. What percentage (approximately) of the total people in the community can be found in the largest friend group?
Number of nodes	About how many total people are in this community? Please type the number below. (For larger communities, the number can be an approximation, but please type only numbers into the box.)
Average degree or degree distribution	About how many friendship connections does each person in this community have, on average?
Number of links	About how many total connections are there in this community?

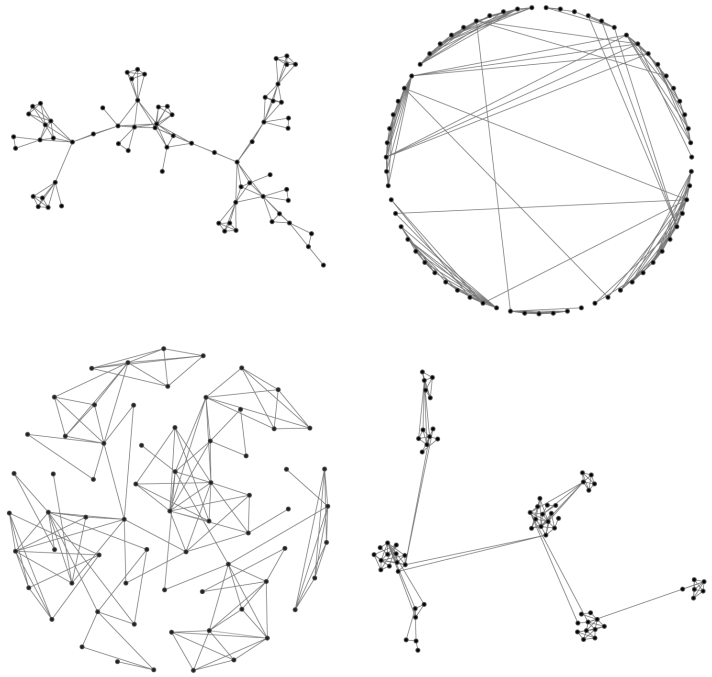




netvislit.org

Layout Conditions (between subjects)

- GEM layout
  force-directed layout
- Circular layout
  nodes positioned by cluster assignment
- Fruchterman-Reingold
  nodes evenly distributed
- OpenOrd
  emphasizes clusters



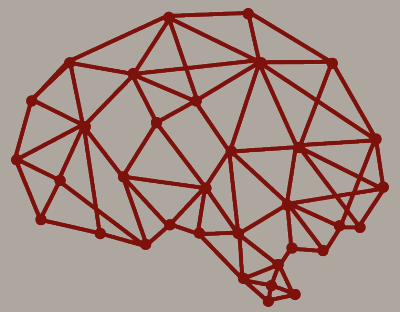
Amazon Mechanical Turk

Selection criteria:

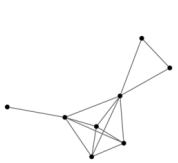
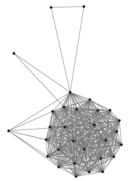
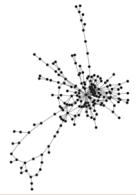
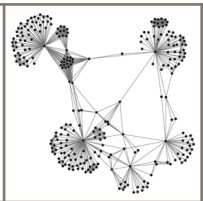
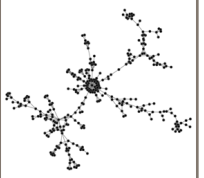
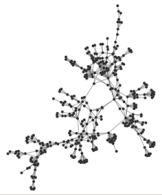
- located in the United States
- approval rate for the worker is at least 95%
- number of approved tasks is at least 100

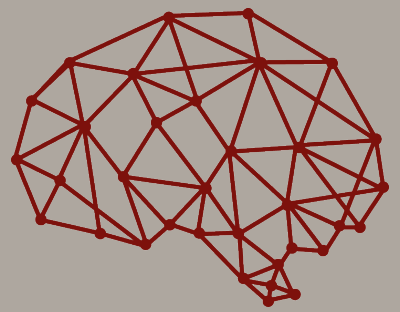
Compensation:

- $3.50 for a 25-30 minute study

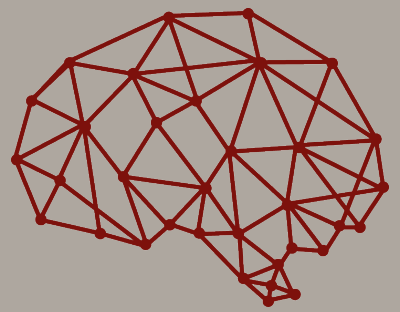


AMT Participants

network							
control	104	50	51	51	50	52	49
phrasing	108	54	52	53	53	53	54
color	109	52	53	54	53	52	55
size	109	54	56	53	55	52	52
circle	102	50	47	48	46	47	47
frucht	105	49	50	51	52	49	53
openord	112	54	54	57	53	54	55



netvislit.org



netvislit.org

IU Network Science community

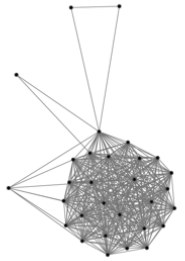
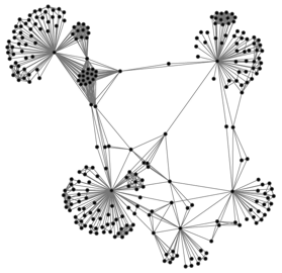
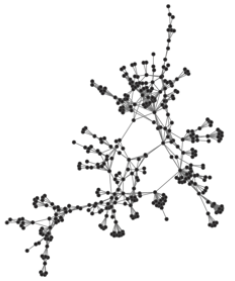
Selection criteria:

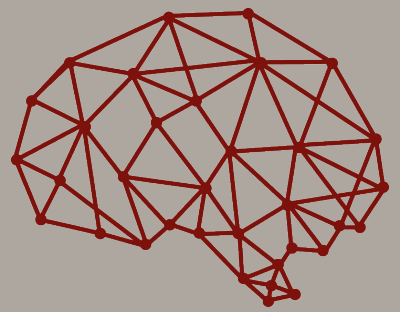
- Affliated with IUNI, CNS program, or other network science training

Compensation:

- Pilot: drawing for two $50 Amazon Gift Cards
- Graduate Students: $10 Amazon Gift Cards, pizza
- Faculty/Staff: randomly assigned to two conditions: $10 gift card, $10 donation

IU NetSci Participants

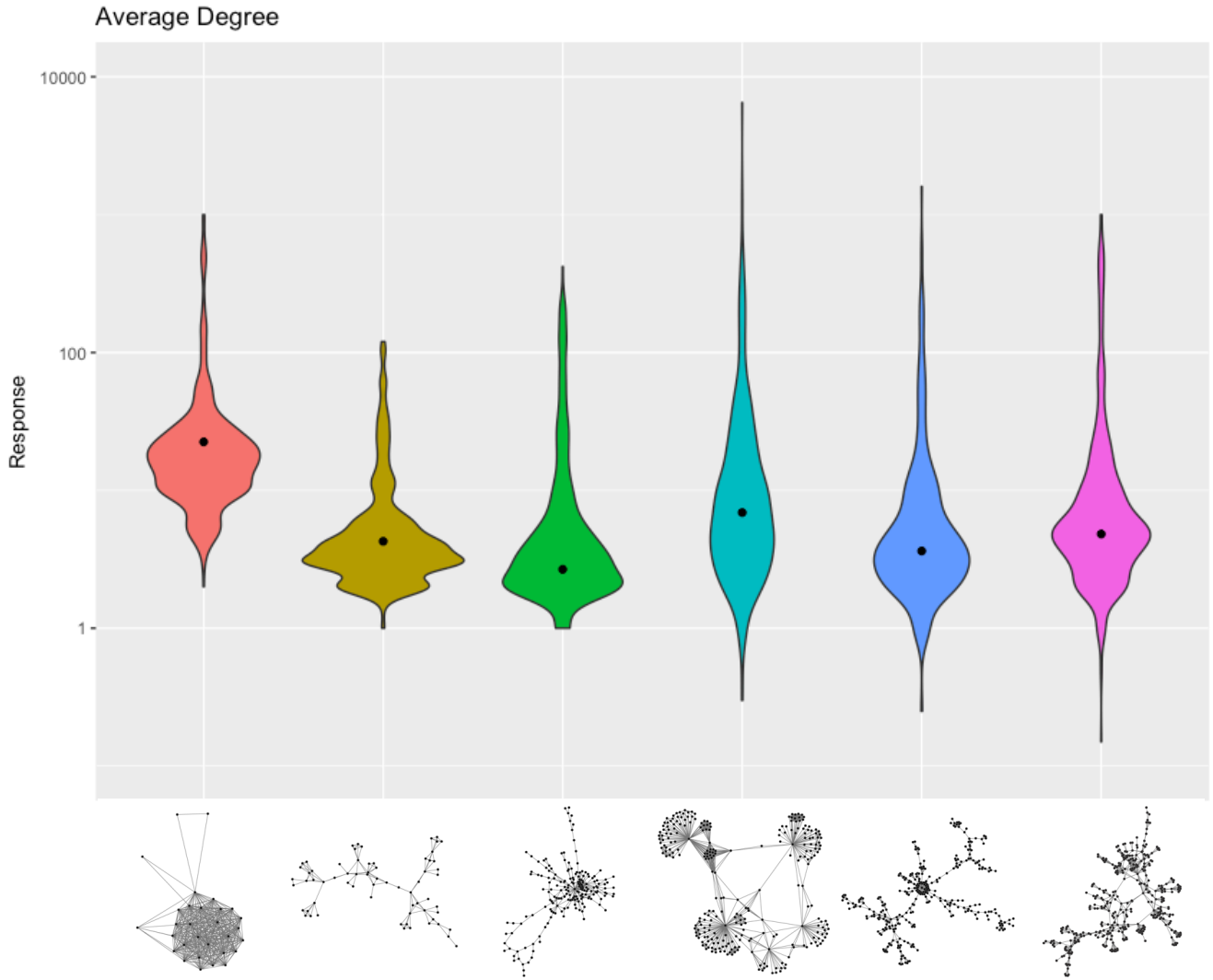
network				
control	23	21	23	22
circle	20	18	17	19
frucht	17	17	17	16
openord	23	22	23	23



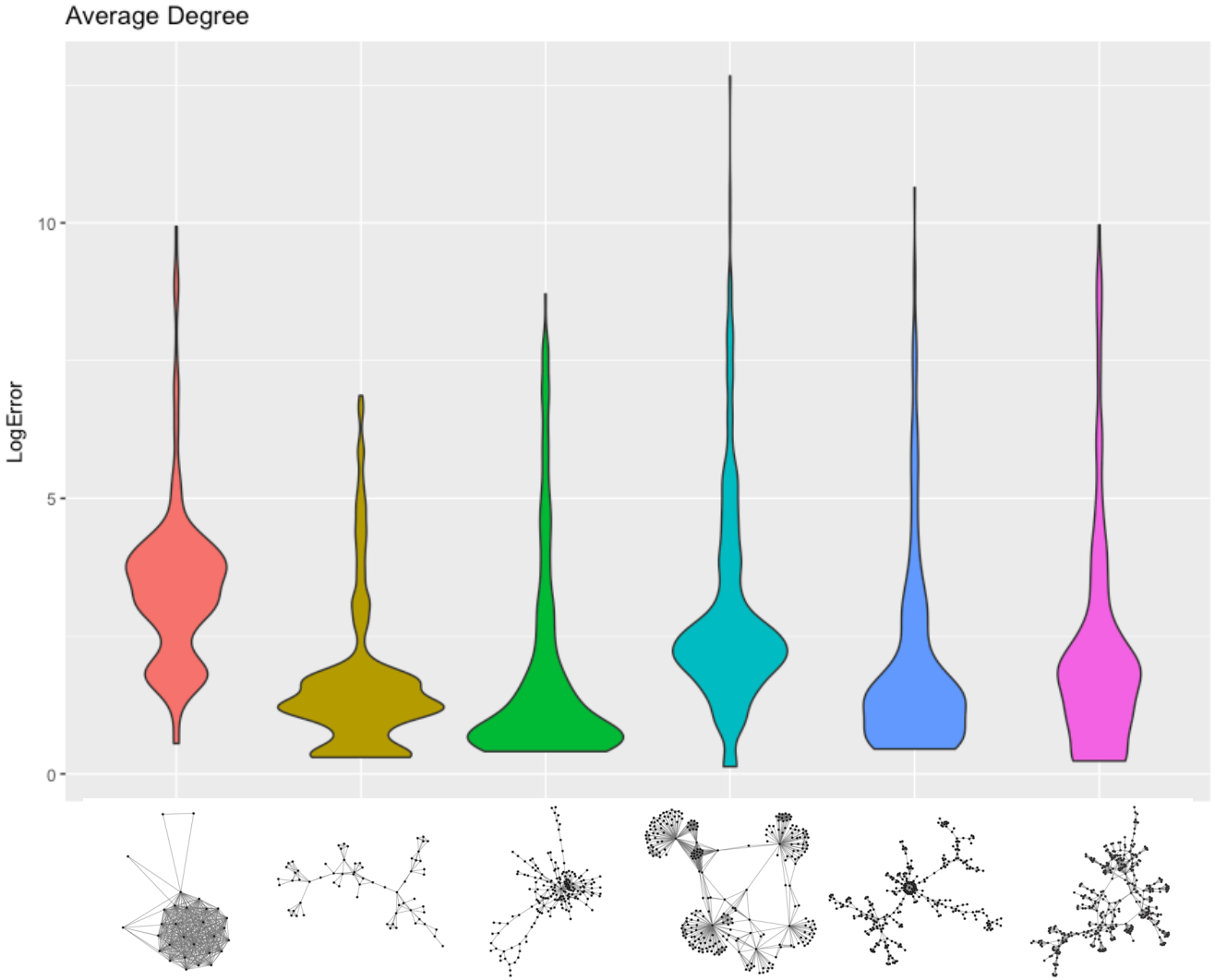
netvislit.org

Results: Tasks

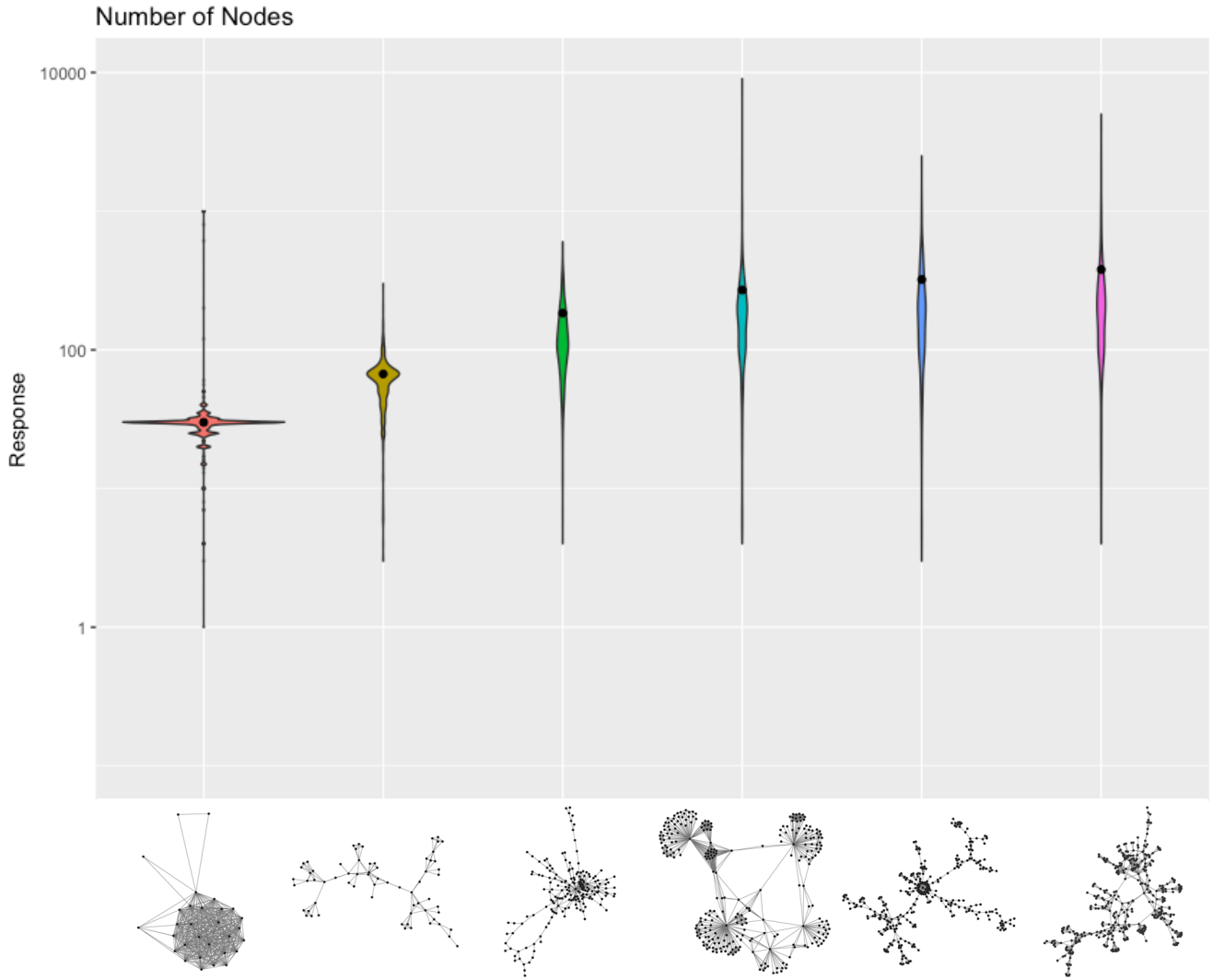
Average Degree (Response)



Average Degree (Error)

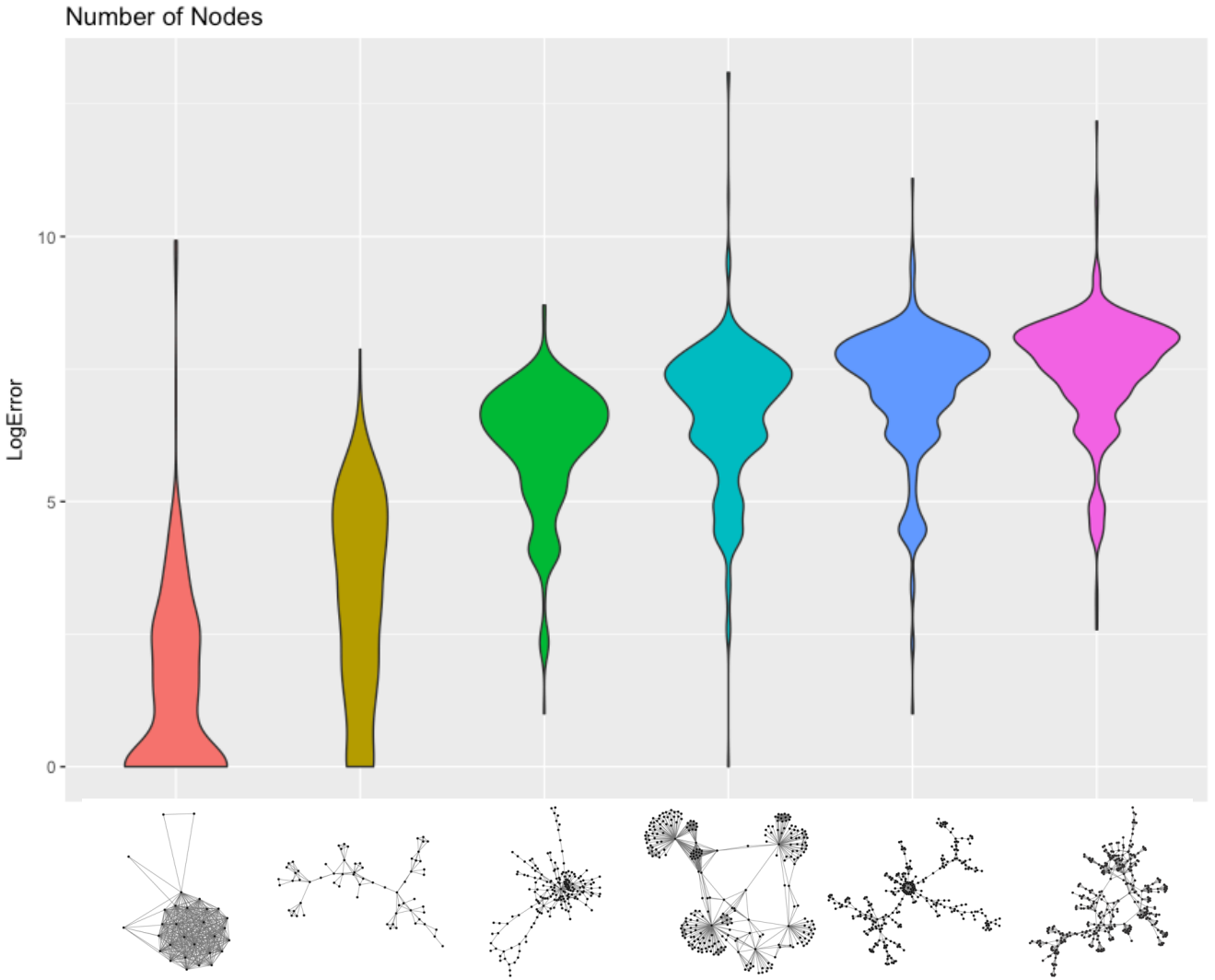


Number of Nodes (Response)

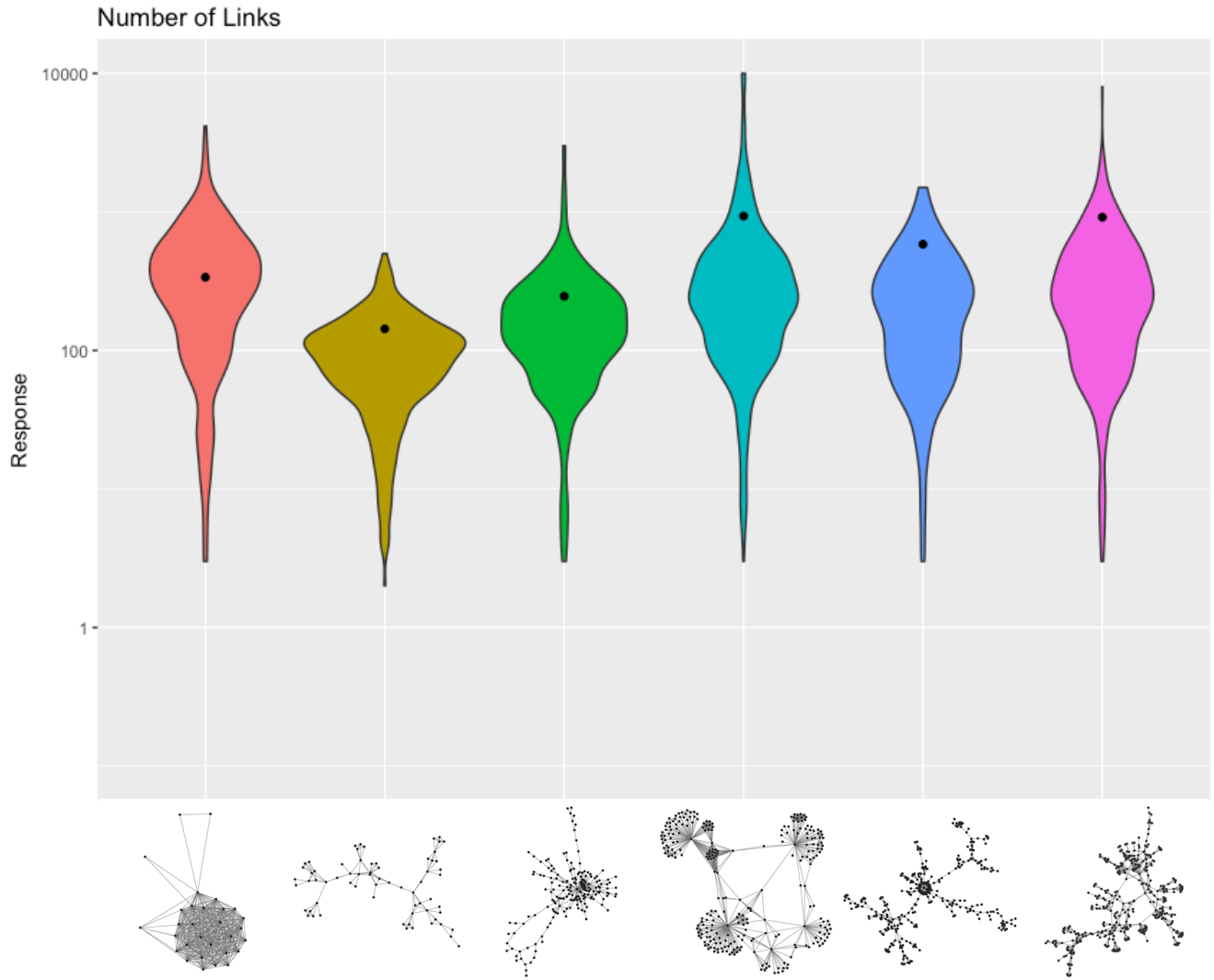


netvislit.org

Number of Nodes (Error)

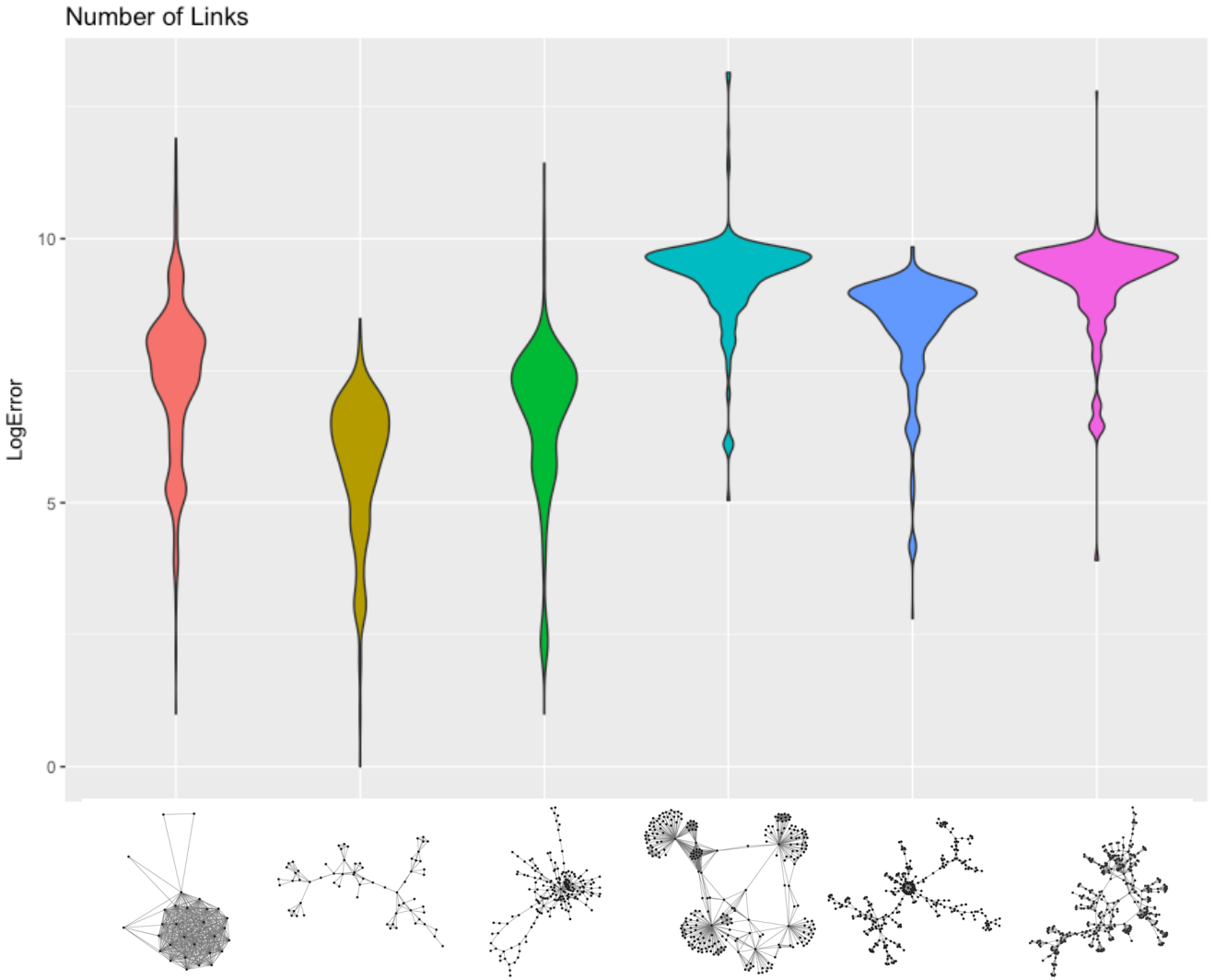


Number of Links (Response)

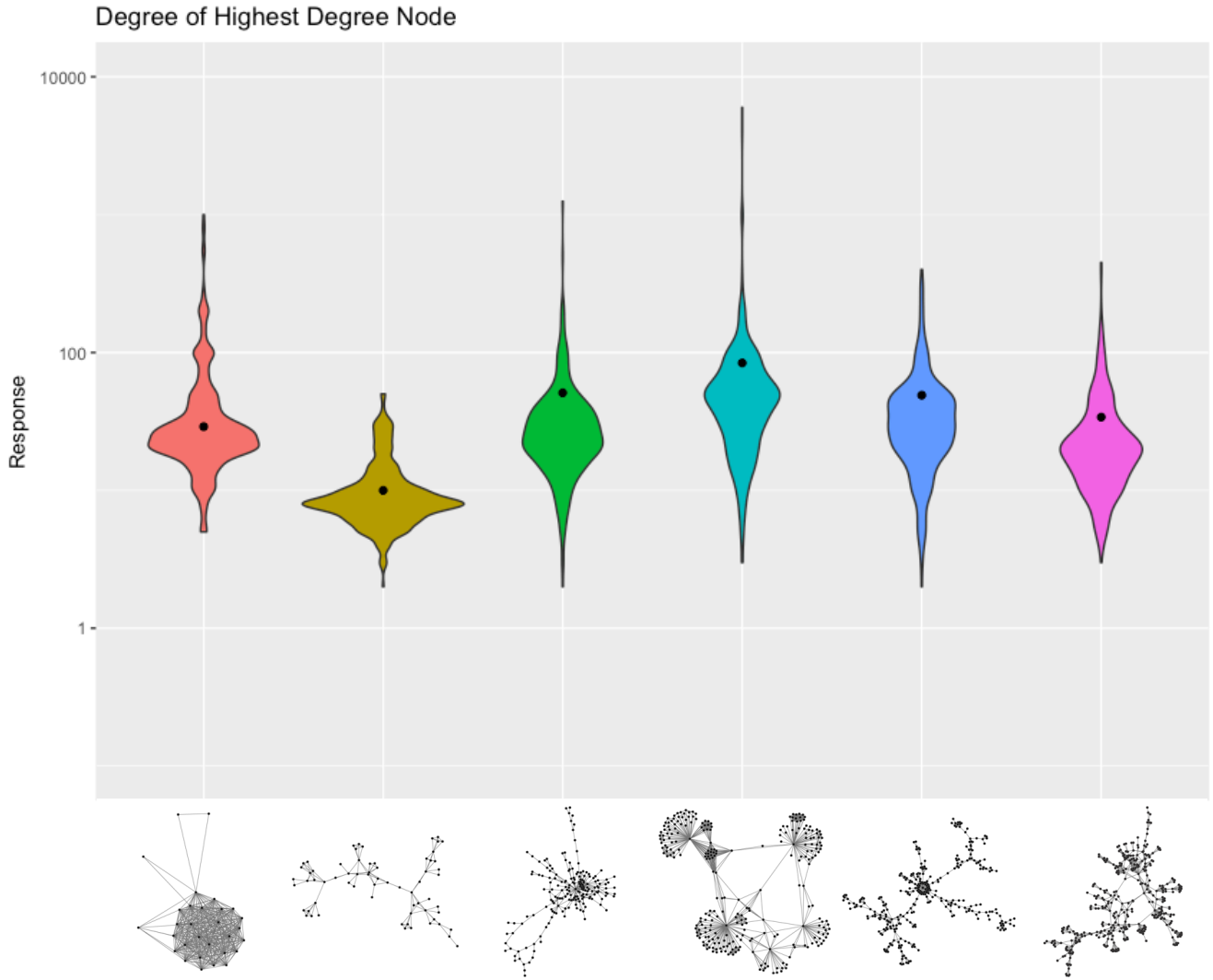


netvislit.org

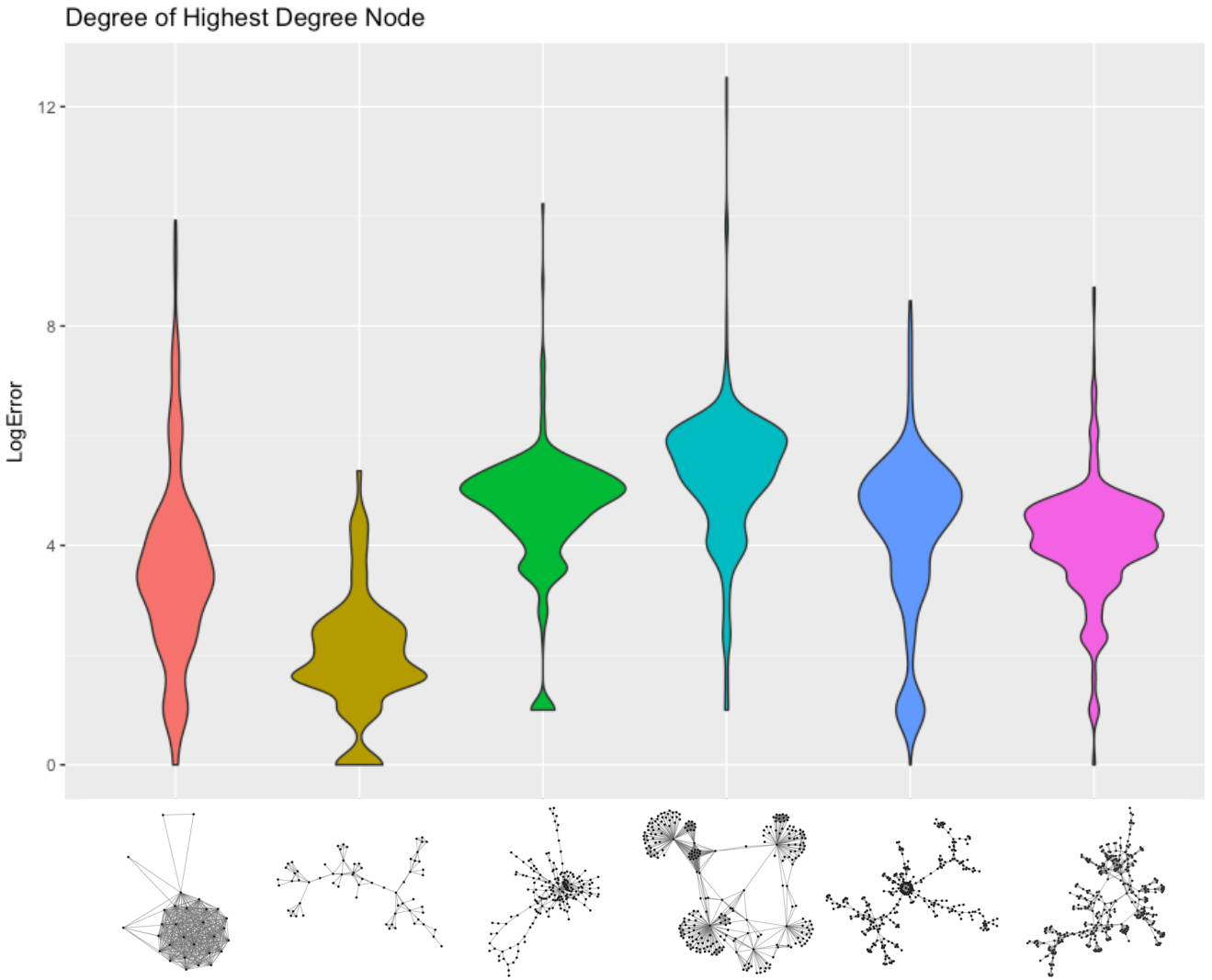
Number of Links (Error)

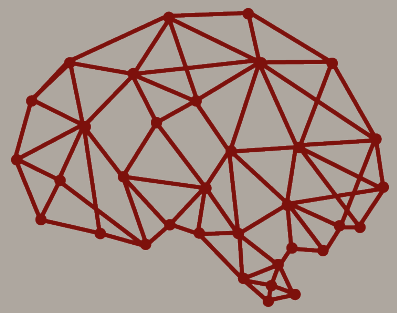


High Degree Node (Response)



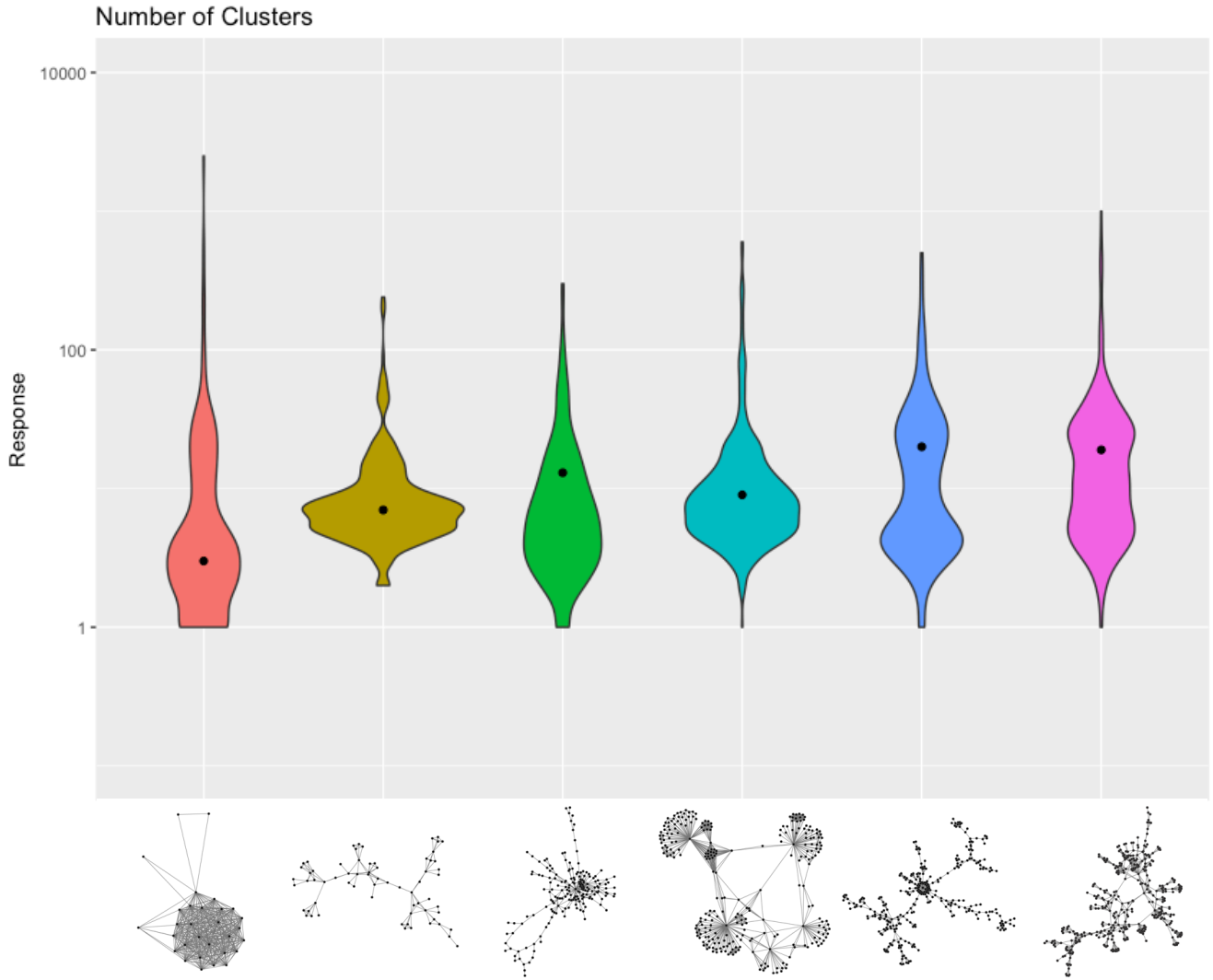
High Degree Node (Error)



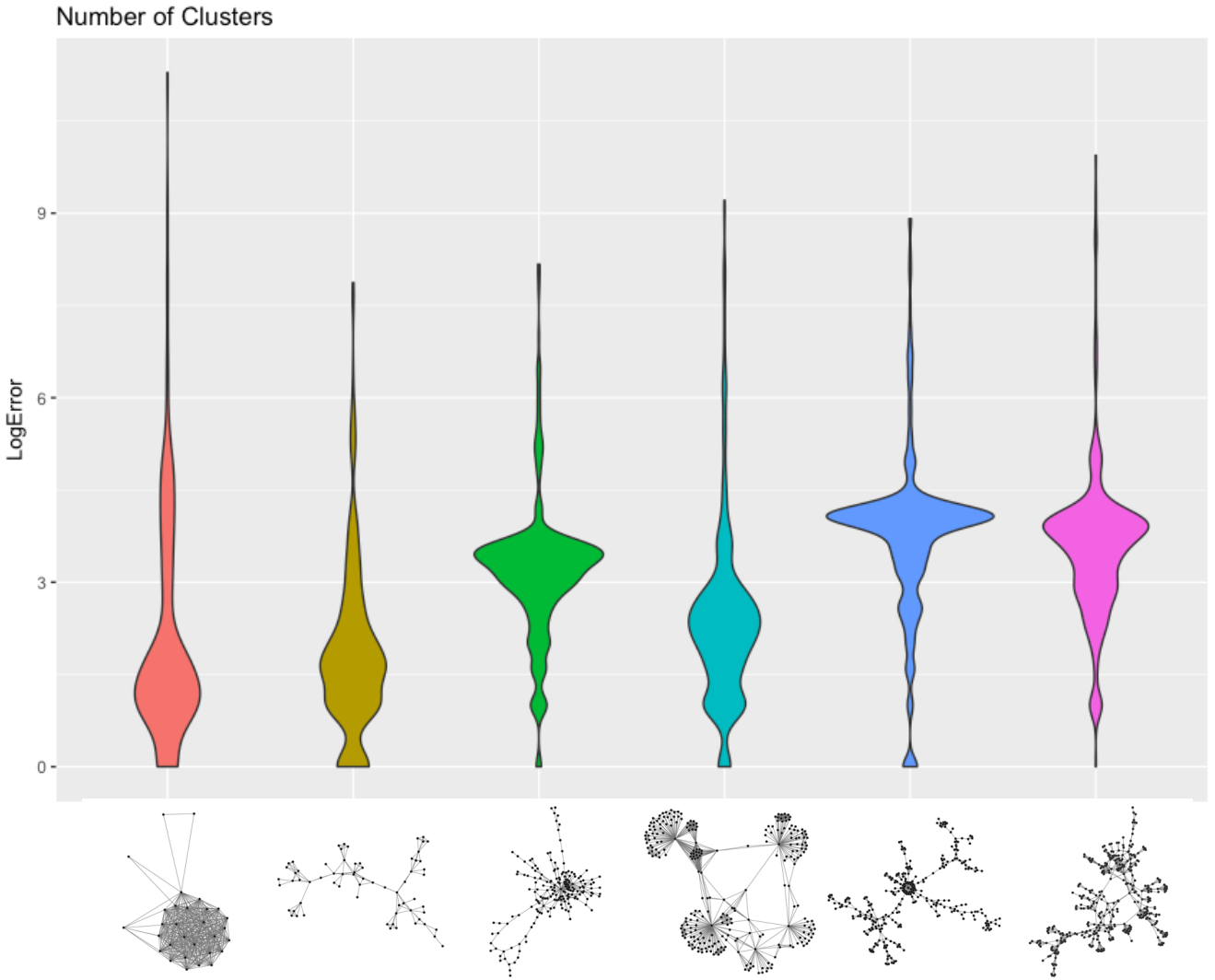


netvislit.org

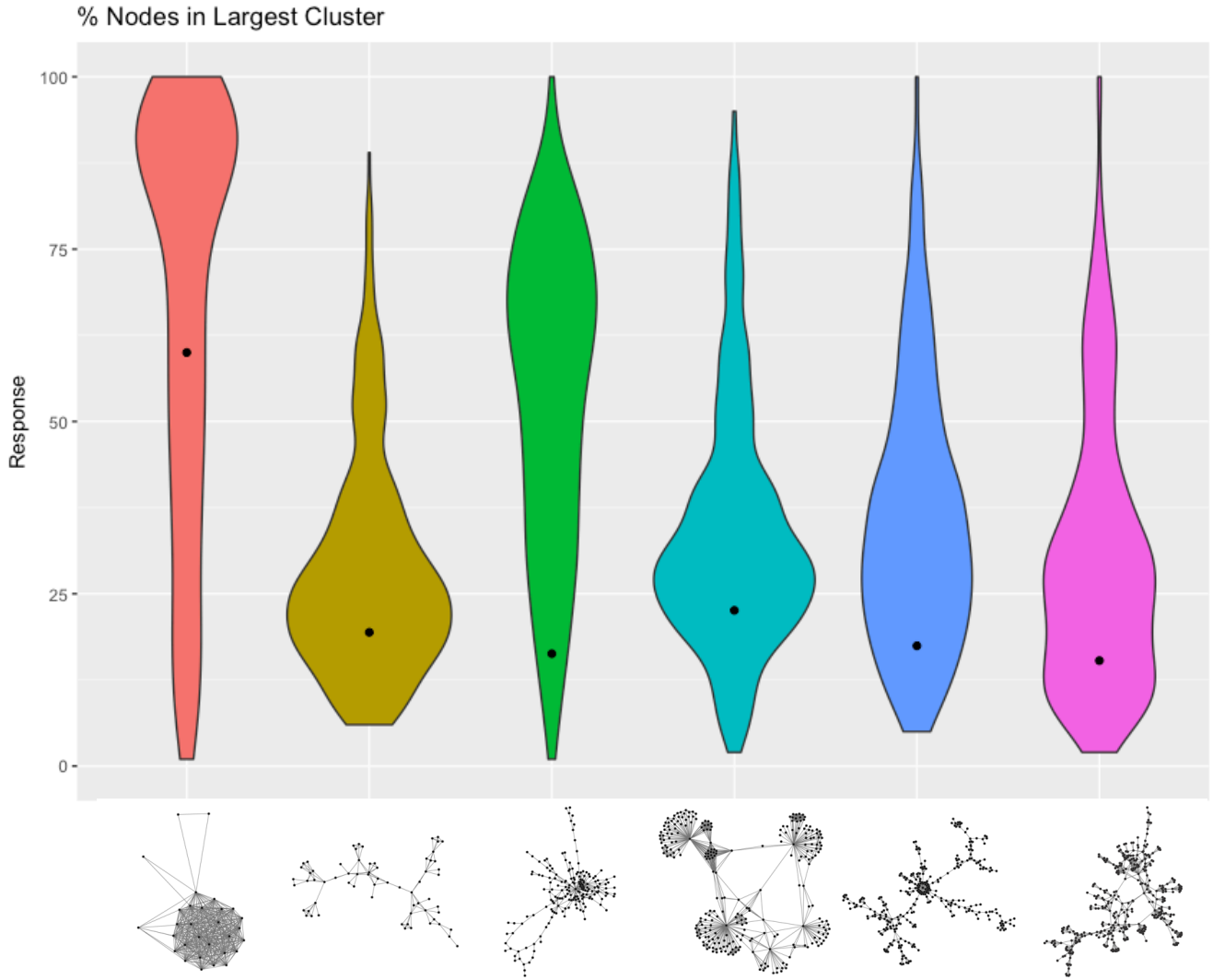
of Clusters (Response)



of Clusters (Error)

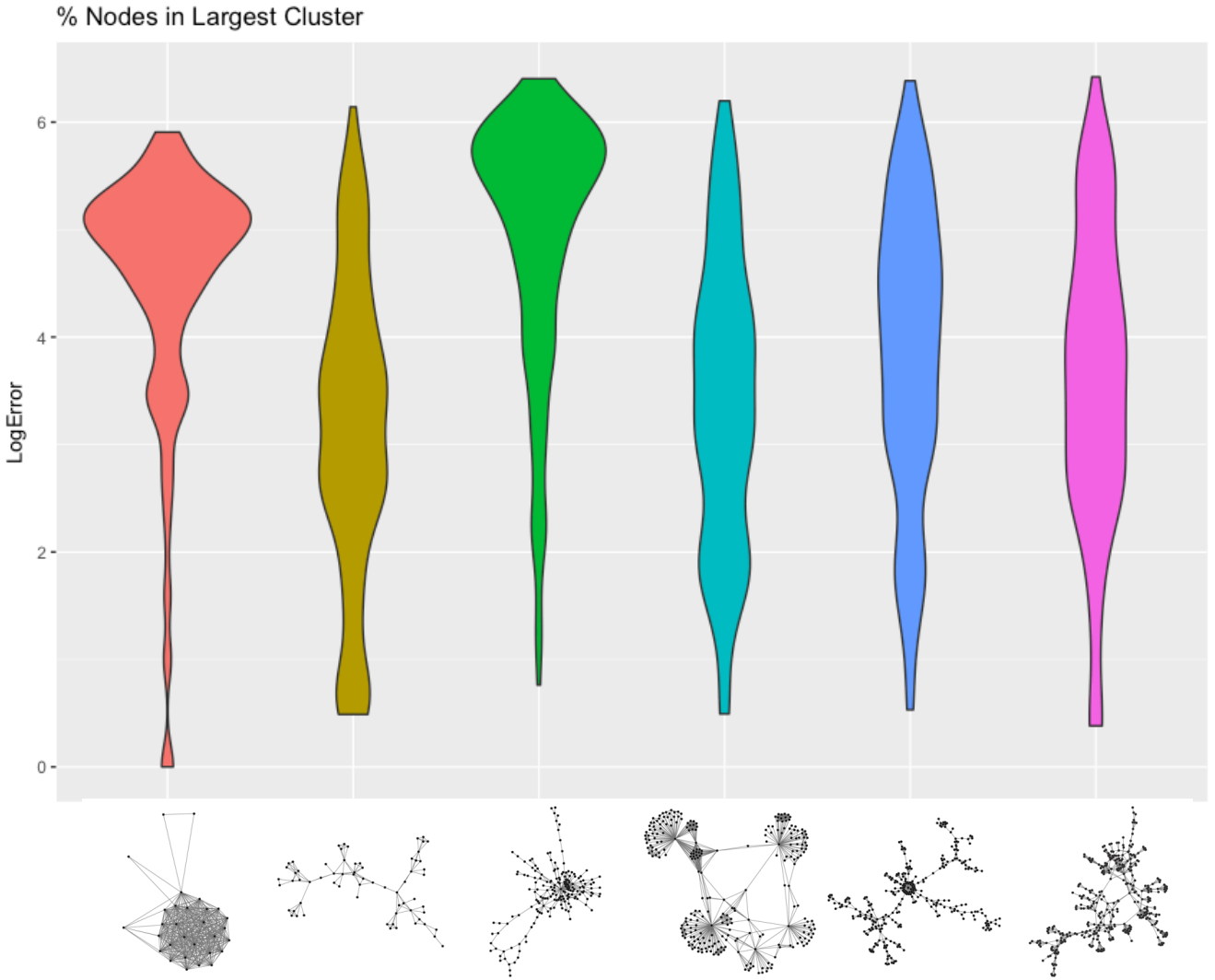


Largest Cluster (Response)



netvislit.org

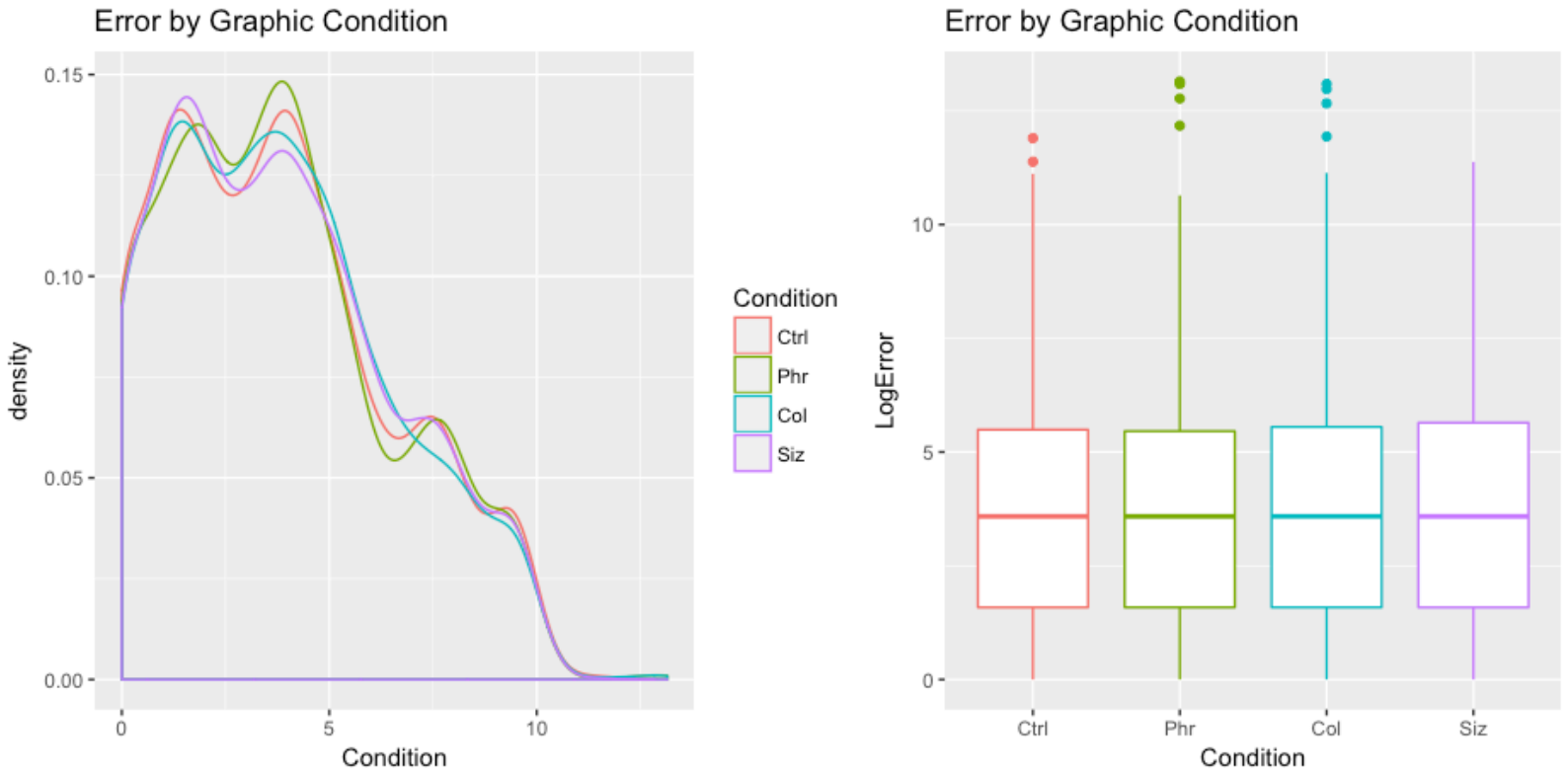
Largest Cluster (Error)



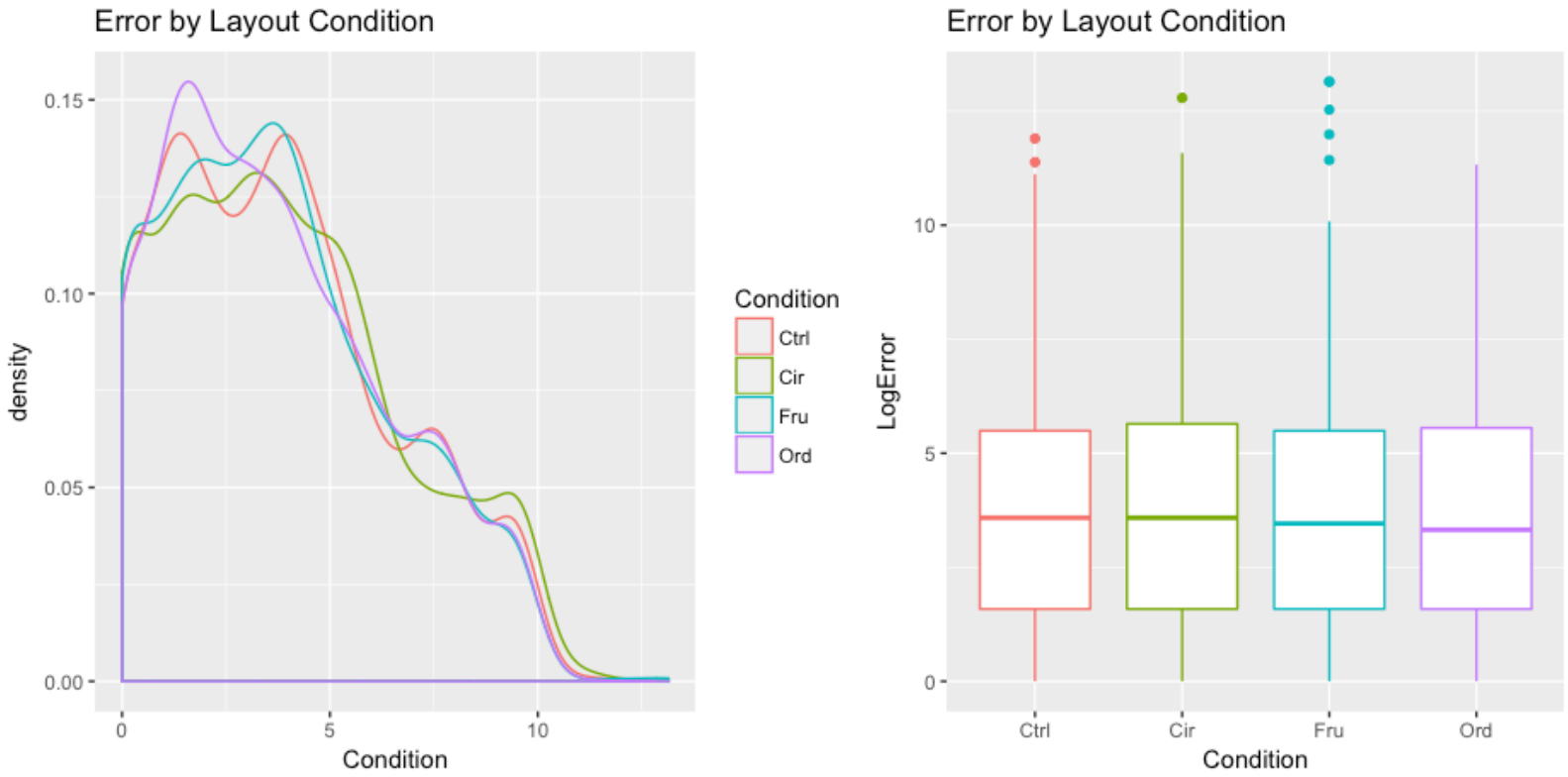
netvislit.org

Results:
Conditions and Expertise

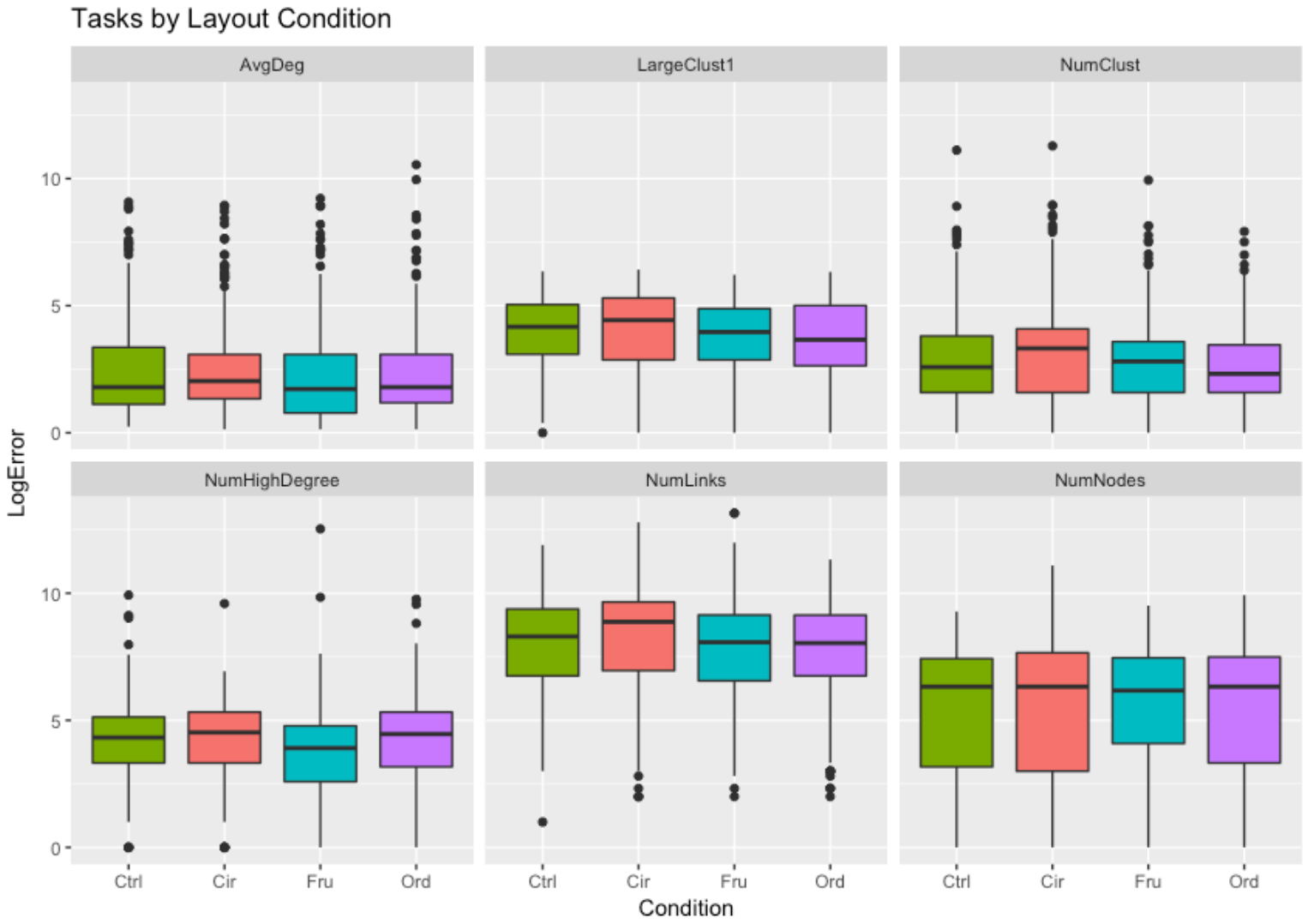
Graphic Conditions (Error)



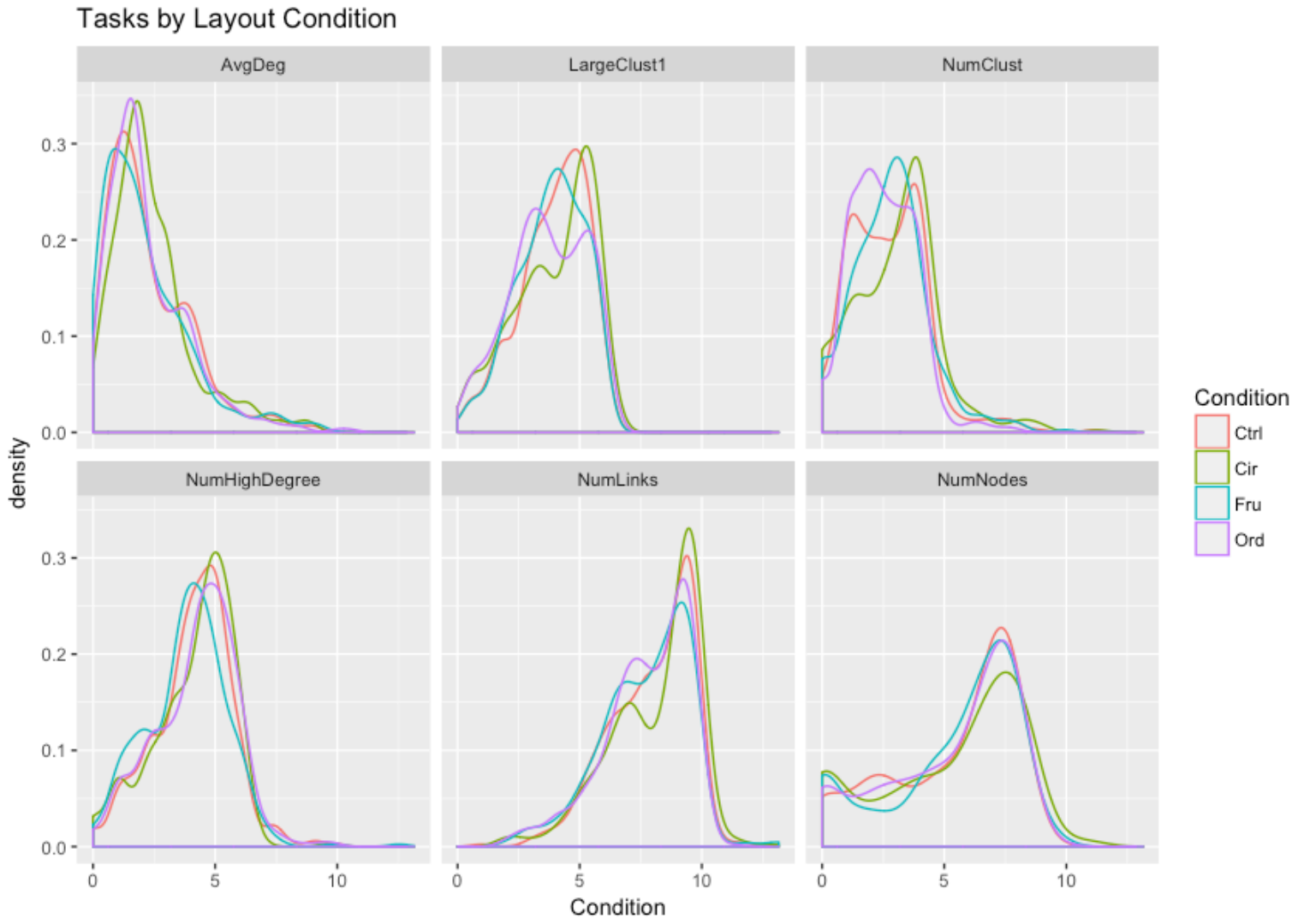
Layout Conditions (Error)



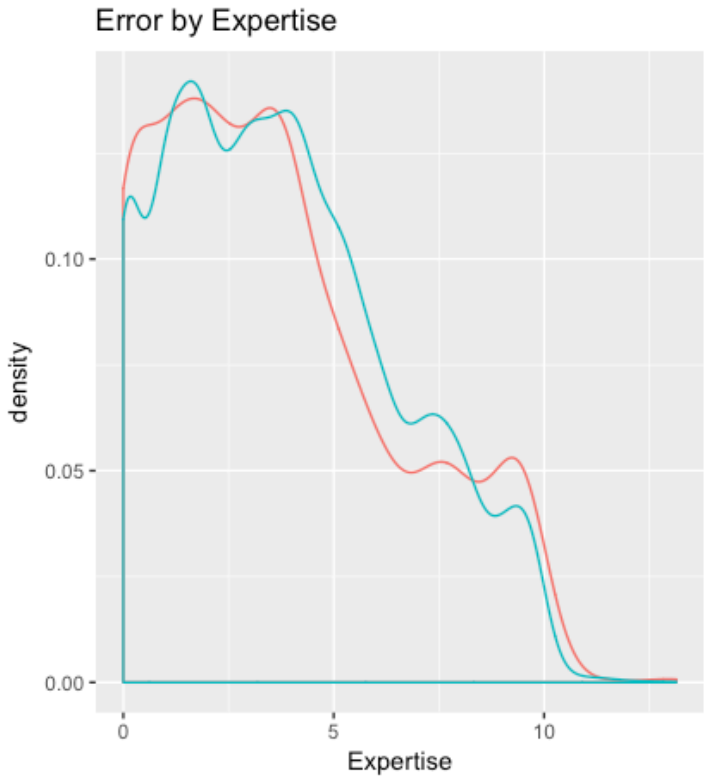
Tasks by Layout

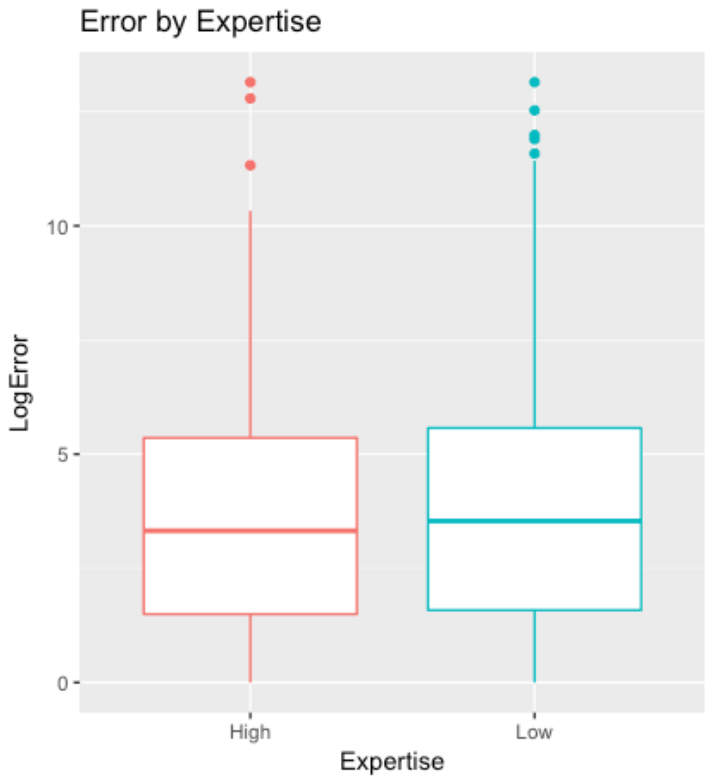


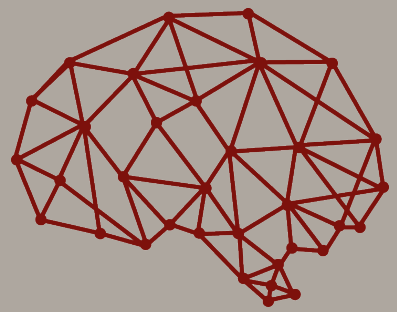
Tasks by Layout



Expertise (Error)



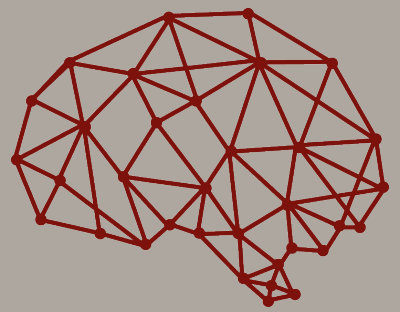




netvislit.org

Future work

- Complete analysis
- Different conditions
  (e.g., variable encoding)
- Qualitative study on interpretation

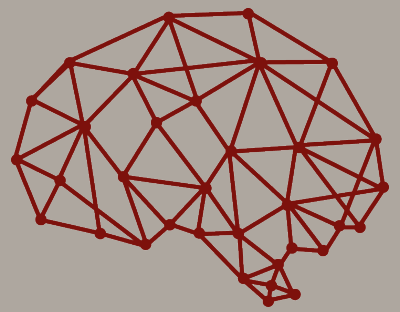


Thank you!

Questions?

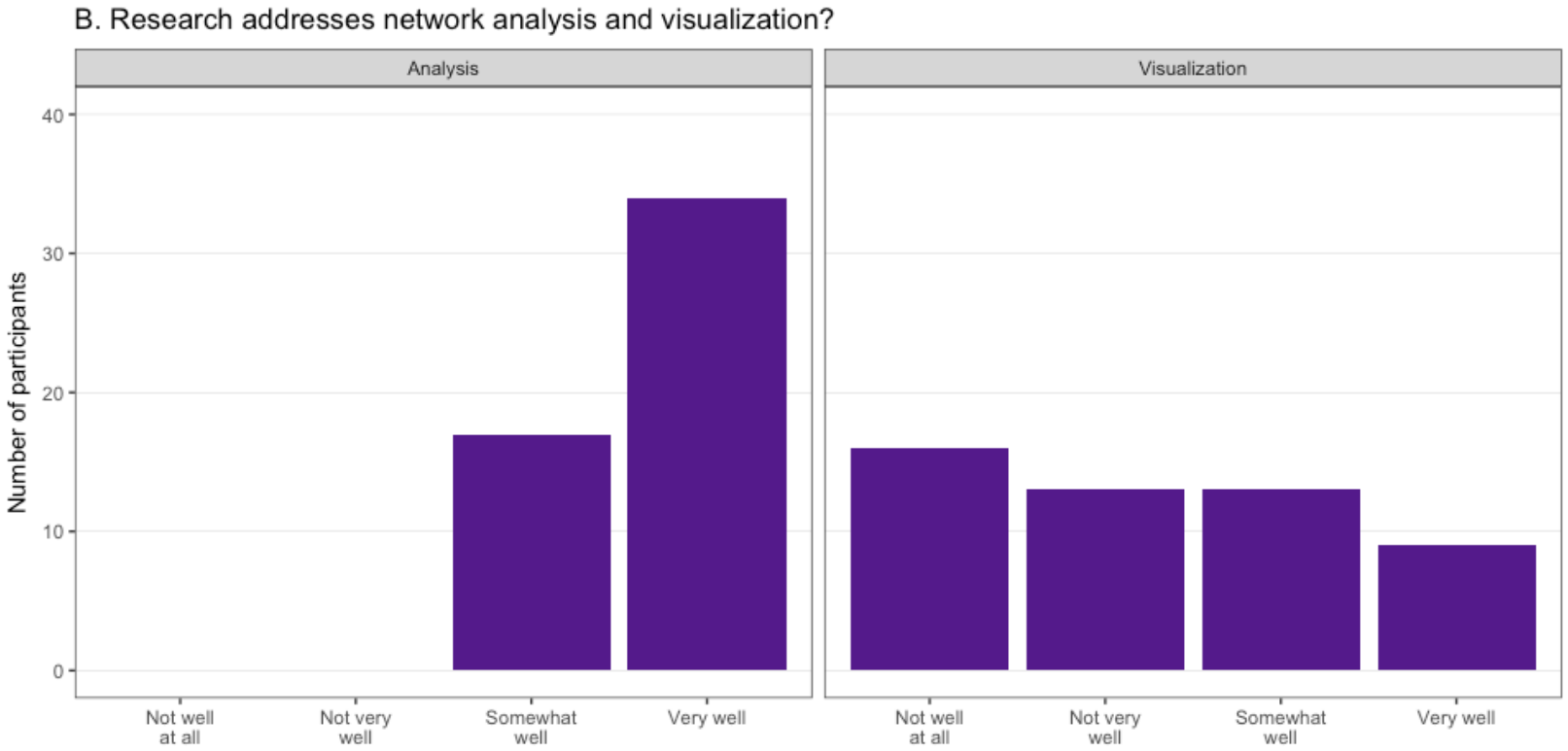
amzoss@indiana.edu

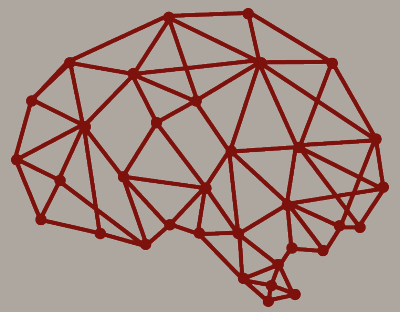
netvislit.org



netvislit.org

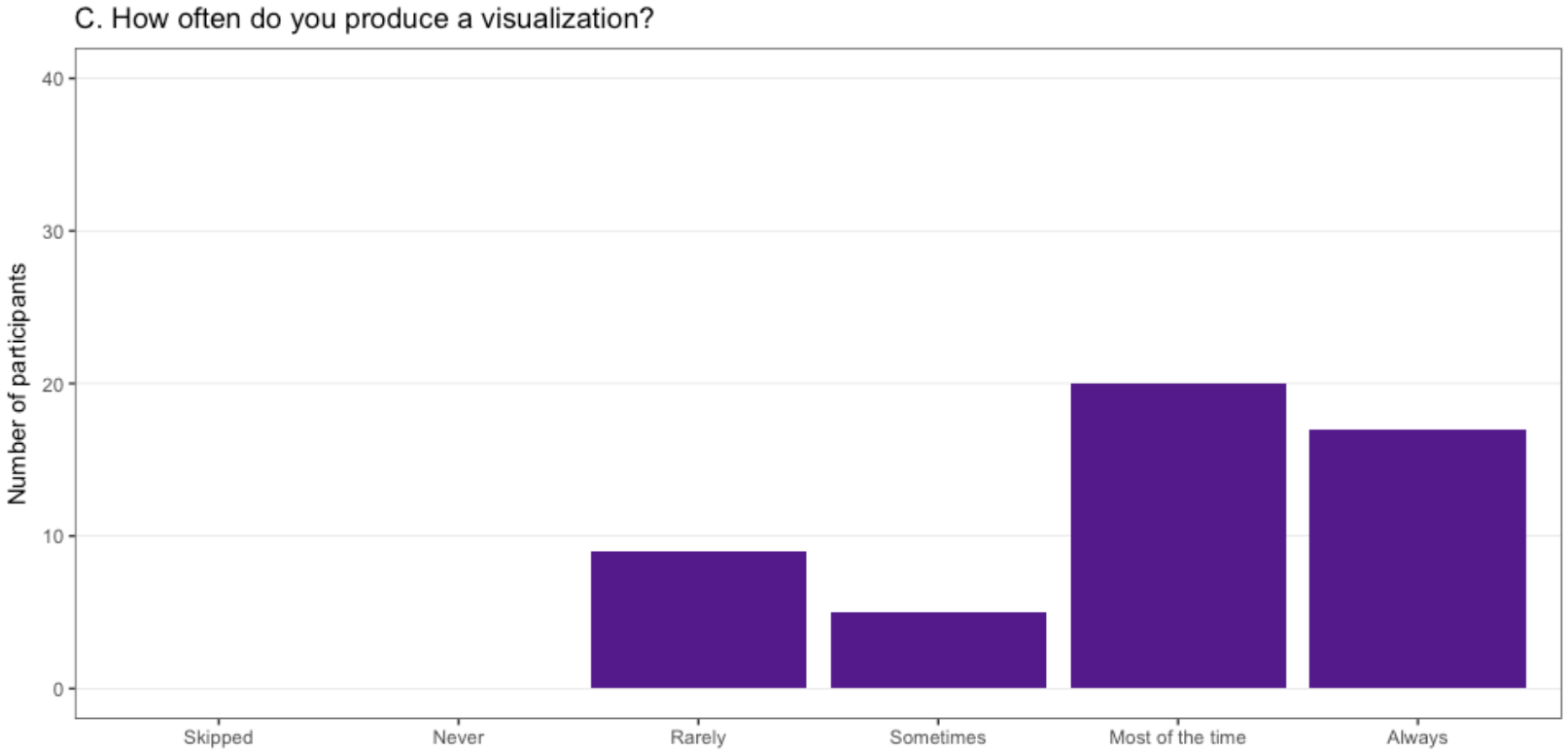
Participants



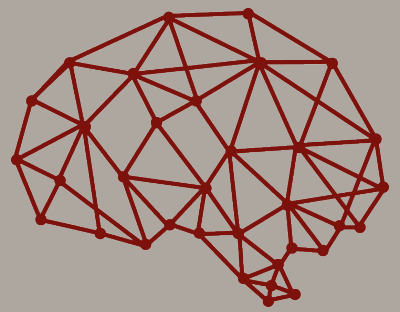


netvislit.org

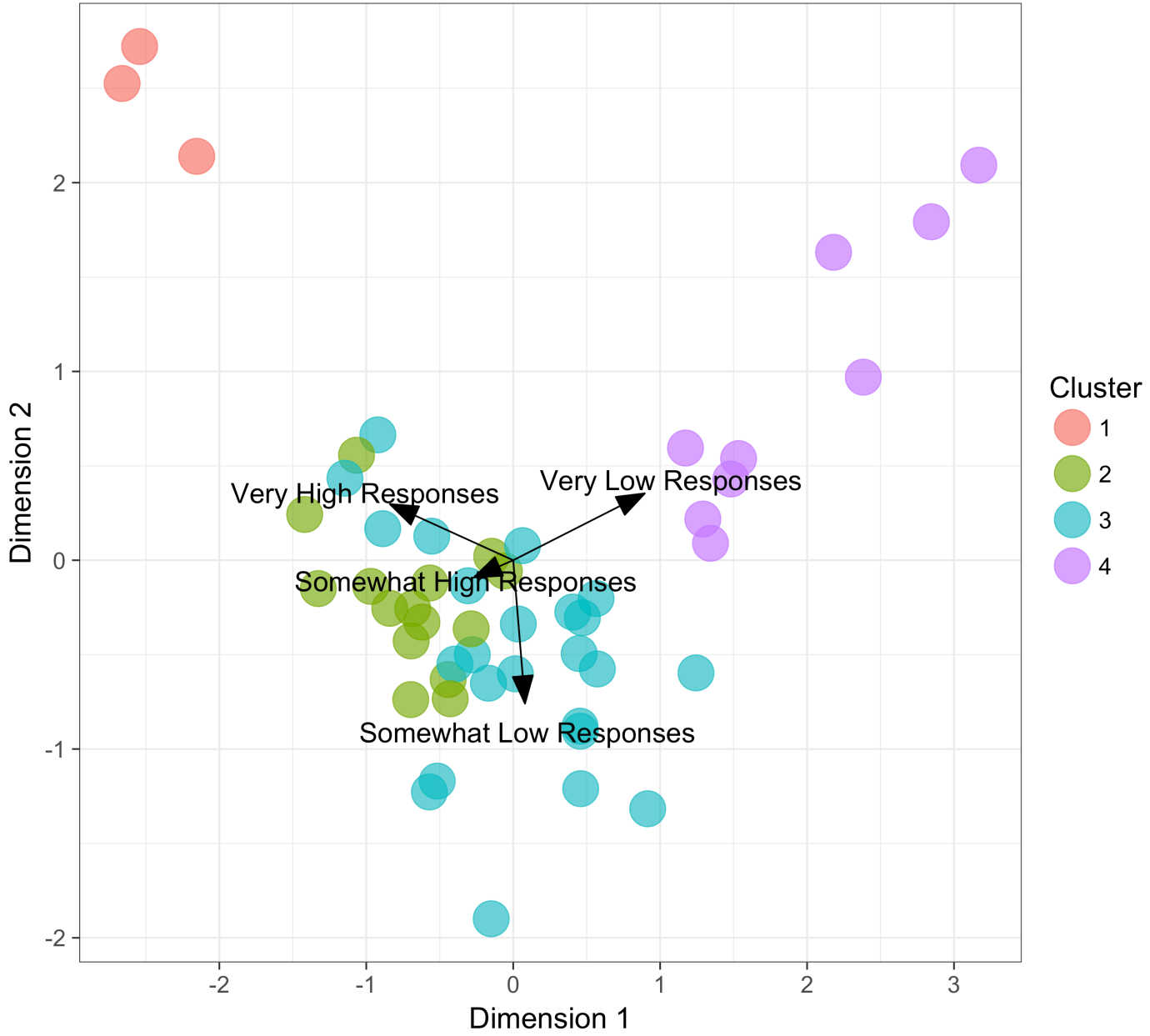
Participants



- Visualization Literacy
- NetVisLit
- Task Selection
- Context and Design
- Layout
- Expertise



netvislit.org



VISUALIZATION LITERACY

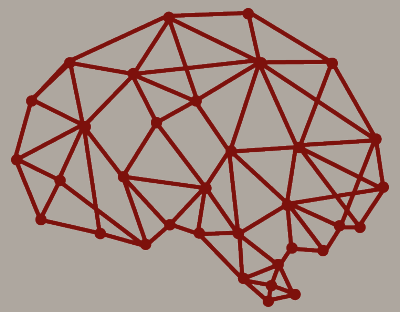
NETVISLIT

TASK SELECTION

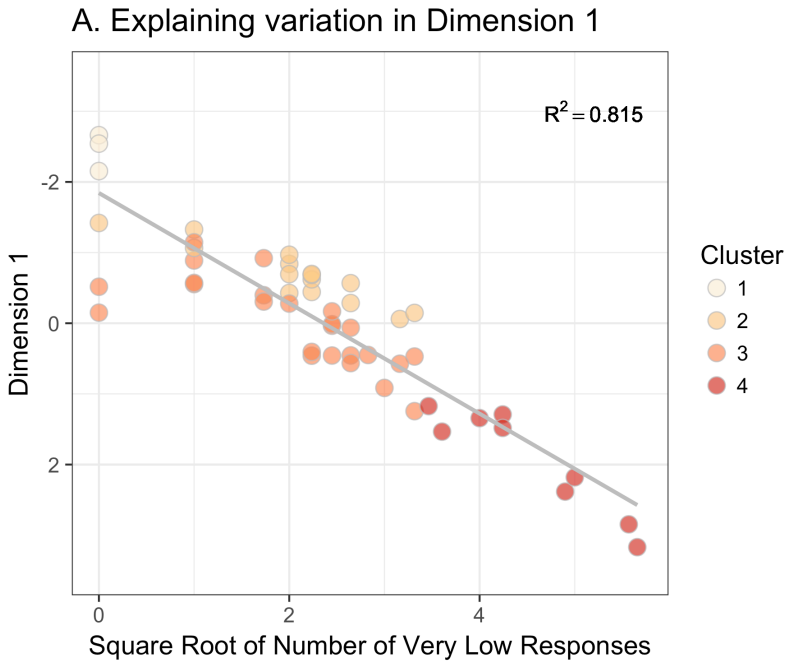
CONTEXT AND DESIGN

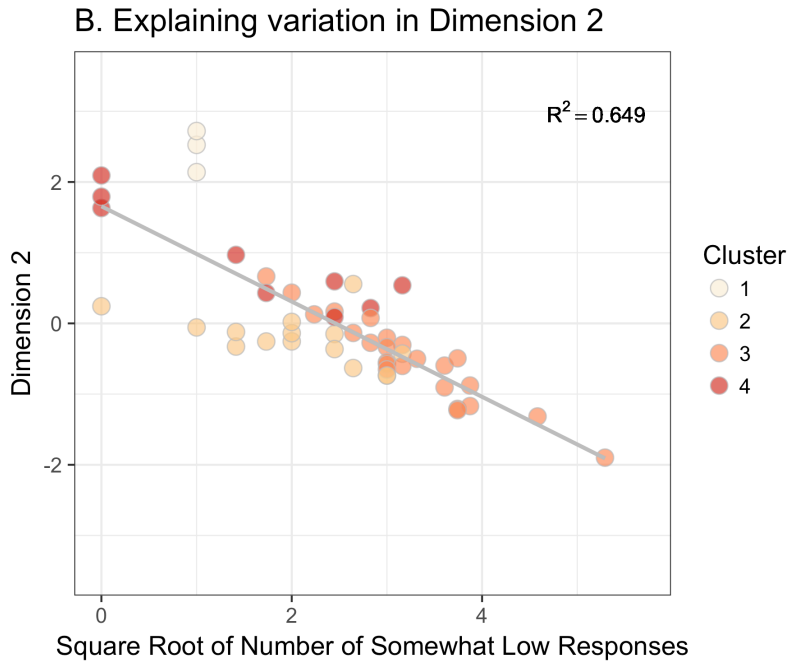
LAYOUT

EXPERTISE



netvislit.org





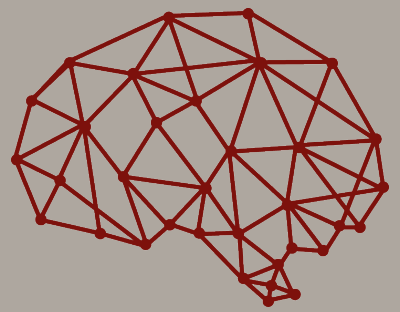


Figure creation: Study B

- Used Sci2 and GUESS for GEM layout
- Used Adobe Illustrator for clean-up:
  - Resize to fit in a 729 pixel x 729 pixel square
  - Change all nodes to a uniform width and height
  - Change all edges to a uniform width
  - Standardize node and link colors

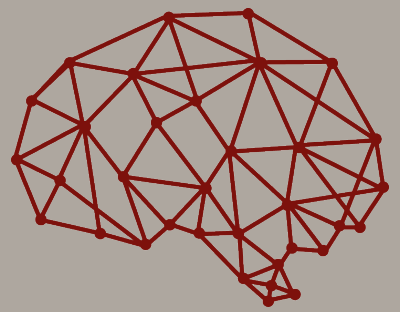


Figure creation: Study C

- Used Gephi for three additional layouts:
  - OpenOrd
  - Fruchterman-Reingold
  - Circular
- Used Adobe Illustrator for clean-up:
  - Resize to fit in a 350 pixel x 350 pixel square
  - Change all nodes to a uniform width and height
  - Change all edges to a uniform width
  - Standardize node and link colors