Network Visualization Literacy

Angela Zoss

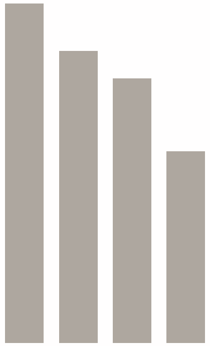
Tuesday, March 27, 2018

Public Dissertation Defense

netvislit.org

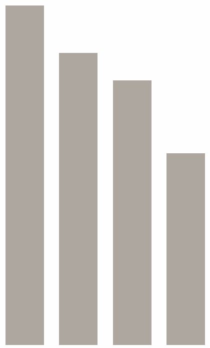
Table of Contents

How well can people read visualizations?

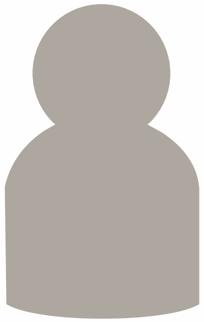


Chart

How well can people read visualizations?

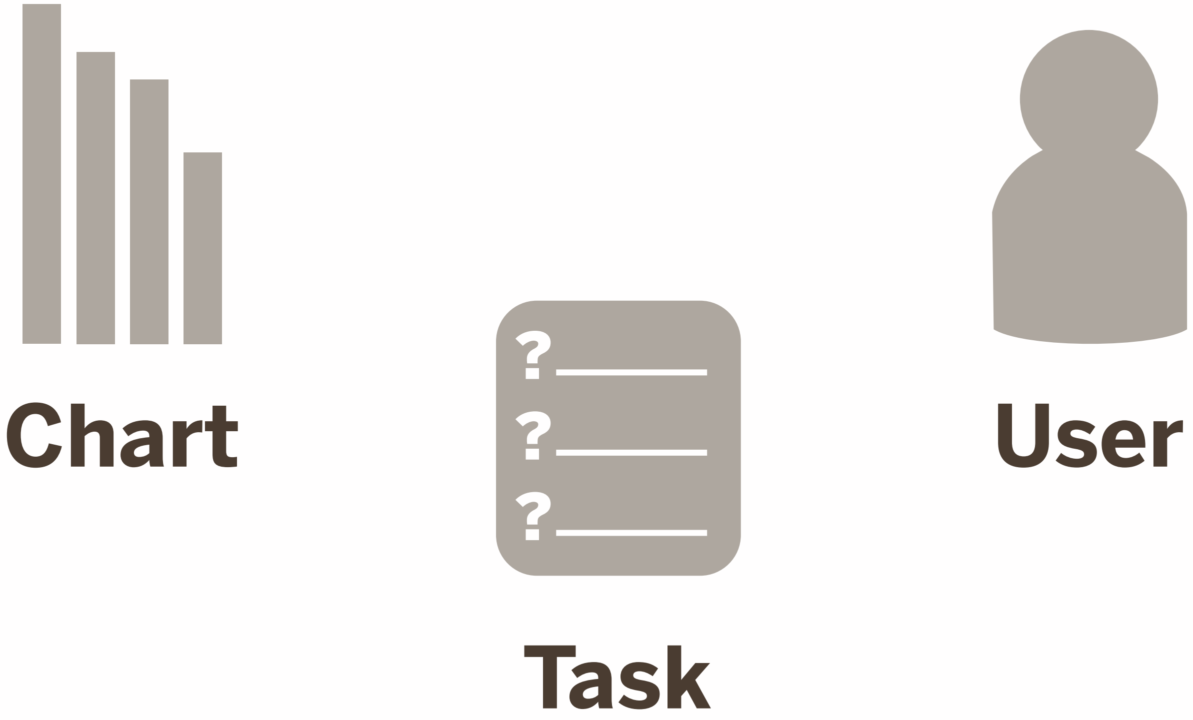


Chart



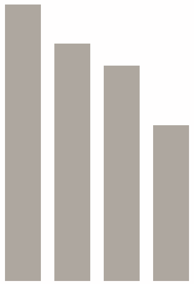
User

How well can people read visualizations?



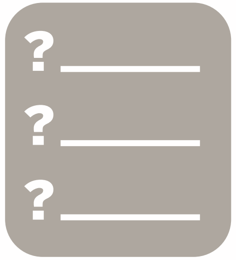
Testing Network Visualization Literacy

Testing Network Visualization Literacy



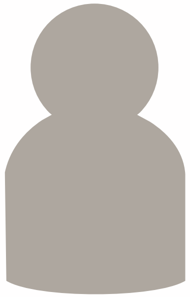
Chart

- **Data**
  - Properties (number of nodes, edges, etc.)
  - Real-world vs. generated
- **Encoding**
  - Algorithms
- **Design**
  - Element size
  - Element color



Task

- **Validity**
  - Ecological
- **Phrasing**
  - Abstract vs. Concrete



User

- **Skills**
  - Prior experience
- **Indiv. Diff.**
- **Envir.**

Study 1	Gathering Tasks
Study 2	Testing Literacy
Study 2A	Graphic Design and Phrasing
Study 2B	Layout and Expertise

Network Visualization Literacy Research Questions

Study 1: Tasks
What do people need to do with network visualizations?

Study 2a: Design
What visualization features might influence usage of netvis?

Study 2b: Training
Does training in network science help?

Study 1: Tasks for Network Visualizations

An Opinion Survey of Network Science Researchers

Research Question

What network analysis **tasks** are appropriate for testing network visualization interpretation across user expertise levels?

Measurable network properties

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

Measurable network properties

But which are **most important**?

Easiest to **estimate** from a visualization?

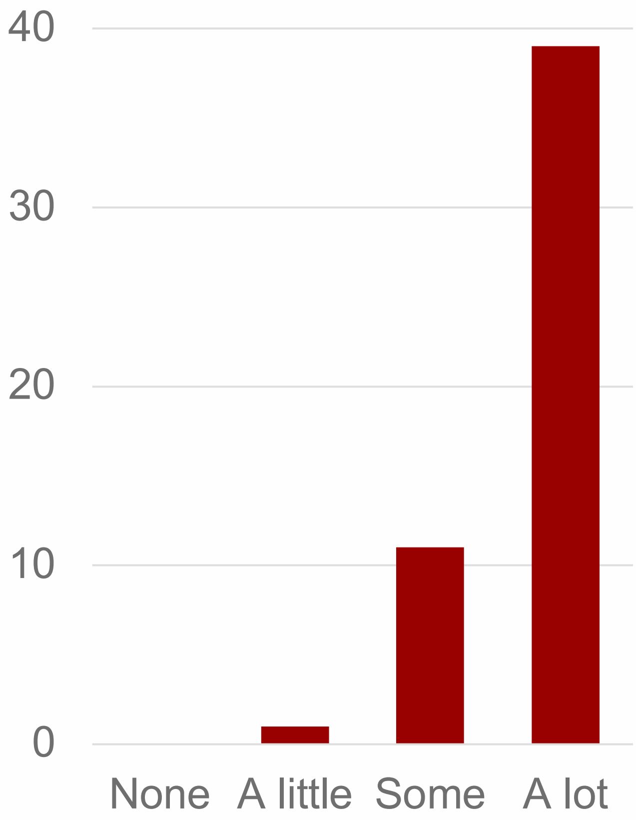
Level	Candidate task
Element (node)	1. Closeness Centrality 2. Eigenvector Centrality 3. Node Betweenness Centrality 4. Node Degree
Element (link)	5. Link Betweenness Centrality 6. Loops
Small groups	7. Component Size 8. Modularity 9. Number of Components
Full network	10. Average Degree 11. Average Path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

Survey Participants

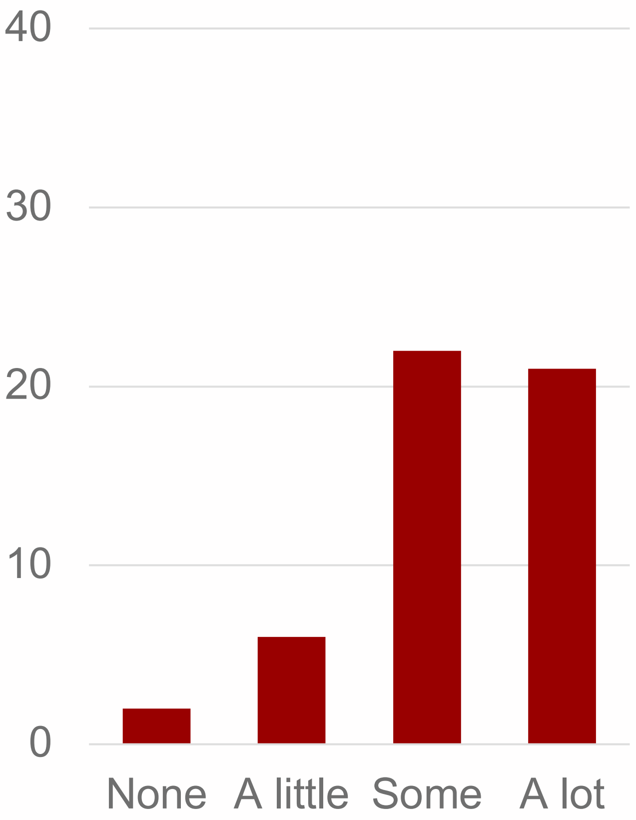
Open invitation to SOCNET listserv (n=51)

Experience as **consumer** and **producer** of network science research?

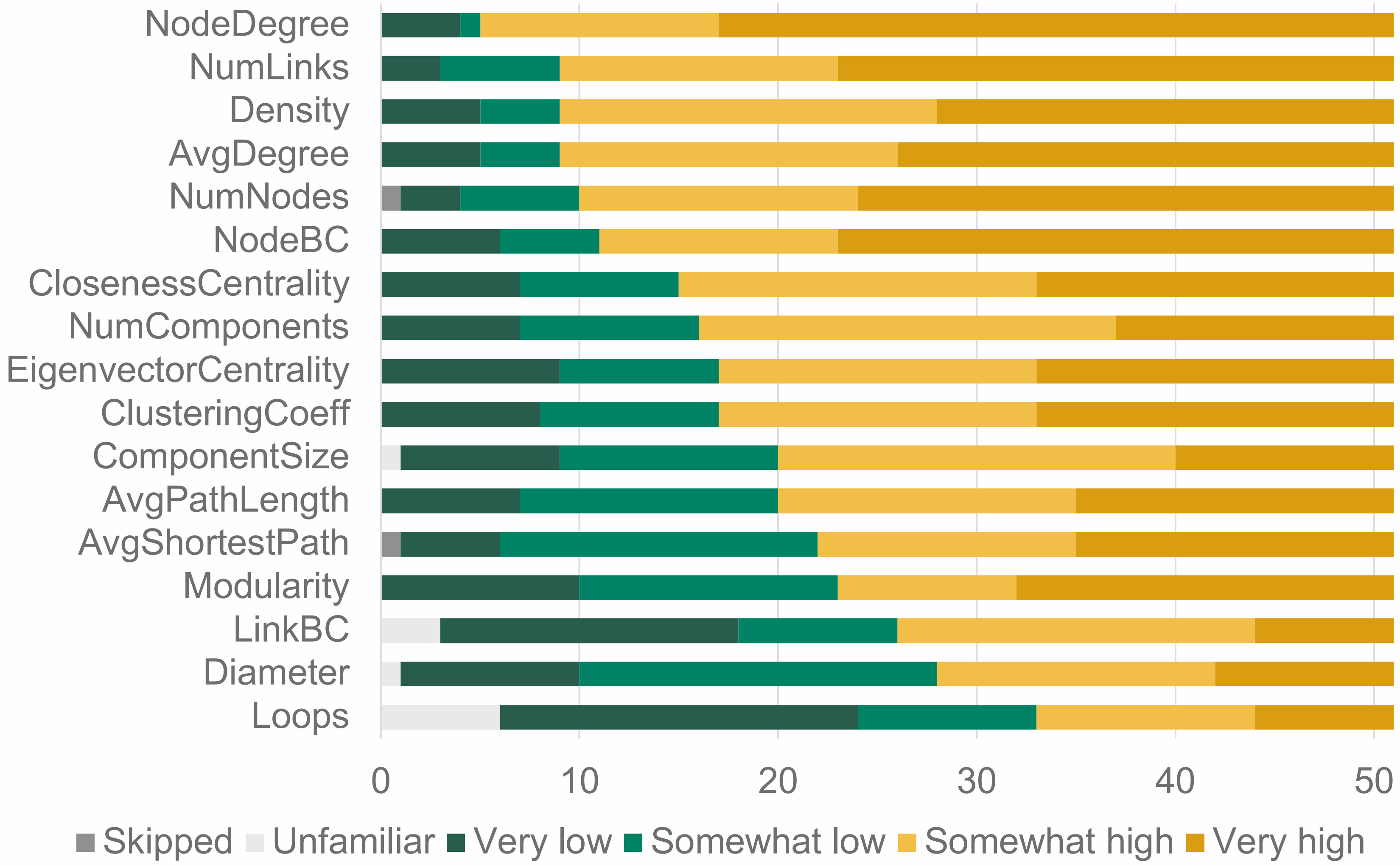
Consumer

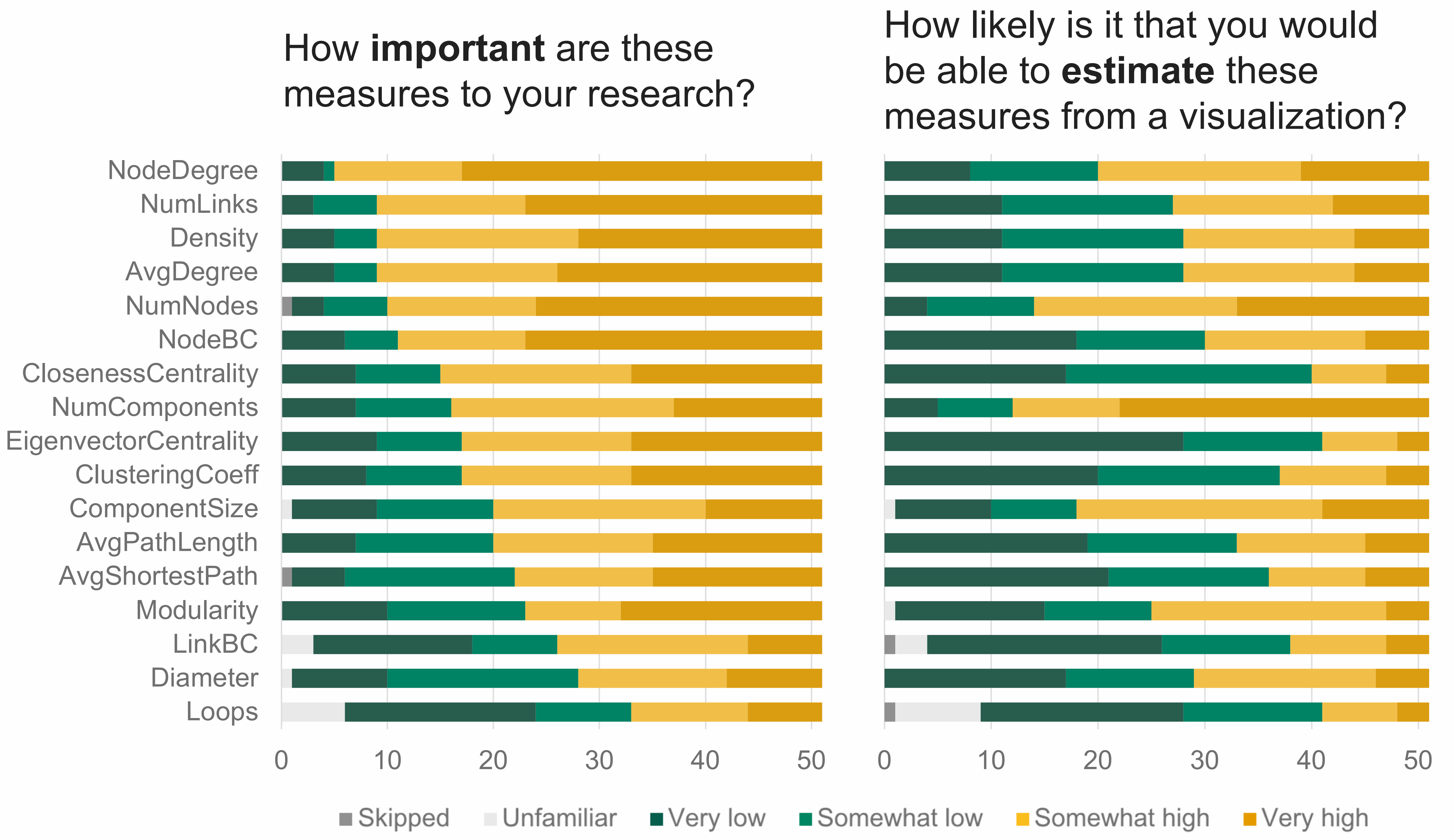


Producer



How **important** are these measures to your research?

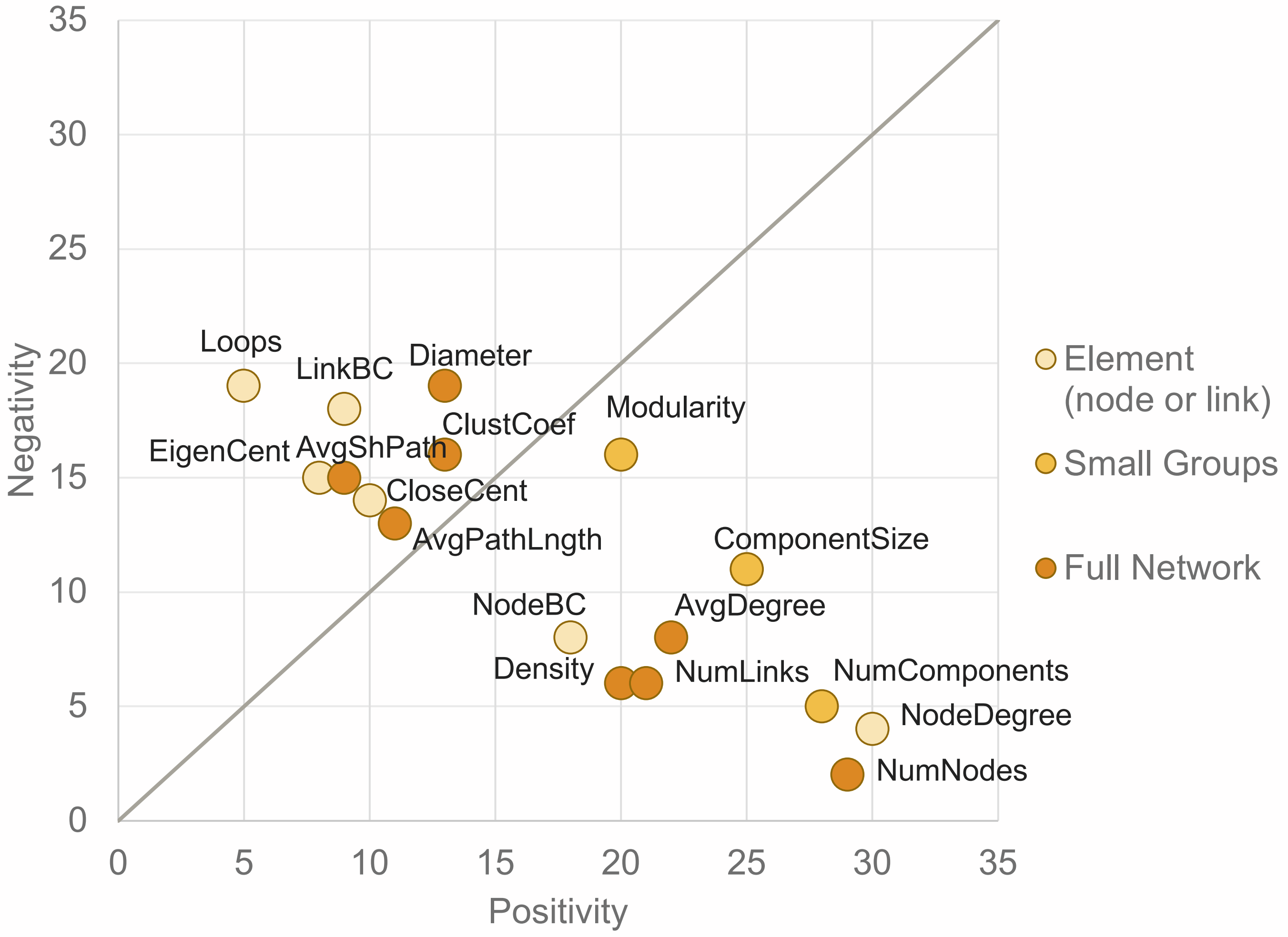



How **important** are these measures to your research?
How likely is it that you would be able to **estimate** these measures from a visualization?
NodeDegree
NumLinks
Density
AvgDegree
NumNodes
NodeBC
ClosenessCentrality
NumComponents
EigenvectorCentrality
ClusteringCoeff
ComponentSize
AvgPathLength
AvgShortestPath
Modularity
LinkBC
Diameter
Loops
0
10
20
30
40
50
Skipped
Unfamiliar
Very low
Somewhat low
Somewhat high
Very high

Combining Importance and Estimability

Positivity
number of people rating measure **somewhat or very high** on both importance and estimability

Negativity
number of people rating measure **somewhat or very low** on both importance and estimability



Final tasks

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

Study 2a: Impact of Graphic Design, Context

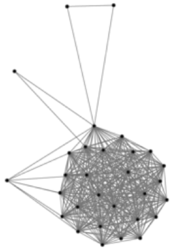
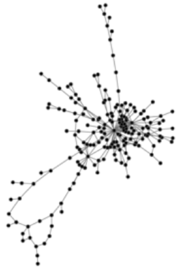
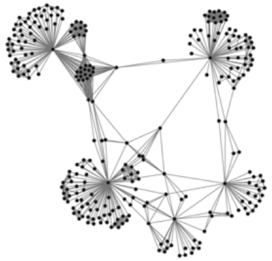
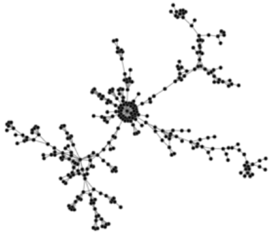
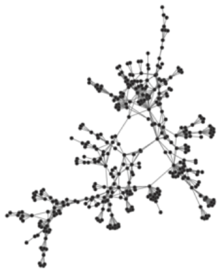
A performance analysis of workers on Amazon's Mechanical Turk

Research Questions

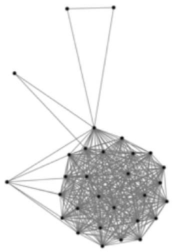
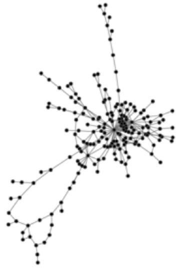
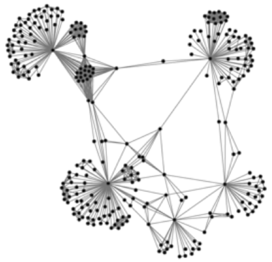
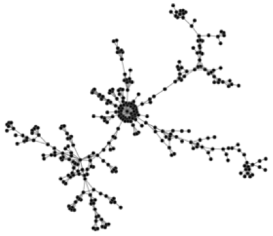
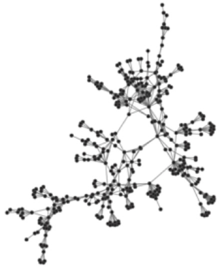
- Which network measures are hardest to assess from a network visualization? Which are easiest?
- How do network properties a (e.g., number of nodes, density) or its context (e.g., concrete vs. abstract question phrasing) affect the ability of users to interpret the visualization?

Network datasets

7 real-world datasets

	0	1	3	5	7	8	9
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

A note on density

	0	1	3	5	7	8	9
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**

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~~

Final tasks

Numerical Response

- Degree of highest degree node
- Number of clusters (and confidence)
- Number of nodes
- Number of links
- Average degree

Click Response

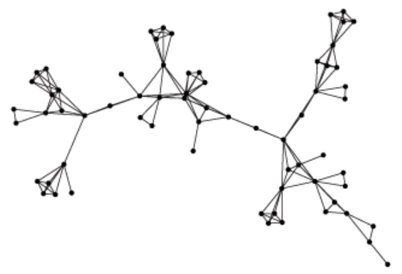
- Highest degree node
- Highest betweenness centrality node

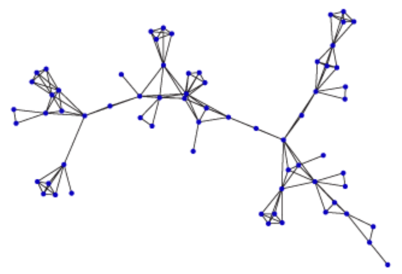
Percentage Response

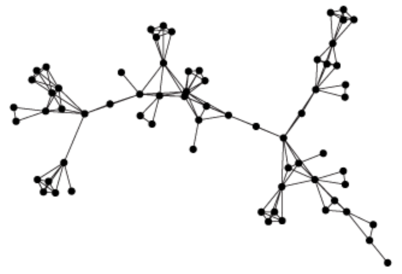
- Size of largest cluster

Graphic Conditions (between subjects)

- **Baseline**
  NLD with GEM layout
- **Concrete phrasing**
  Using “person” and “relationship” rather than “node” and “link”
- **Color**
  change all nodes to a different color
- **Size**
  make all nodes slightly larger







Task phrasing

Degree of Highest Degree Node

Abstract

Find the node with the most links.

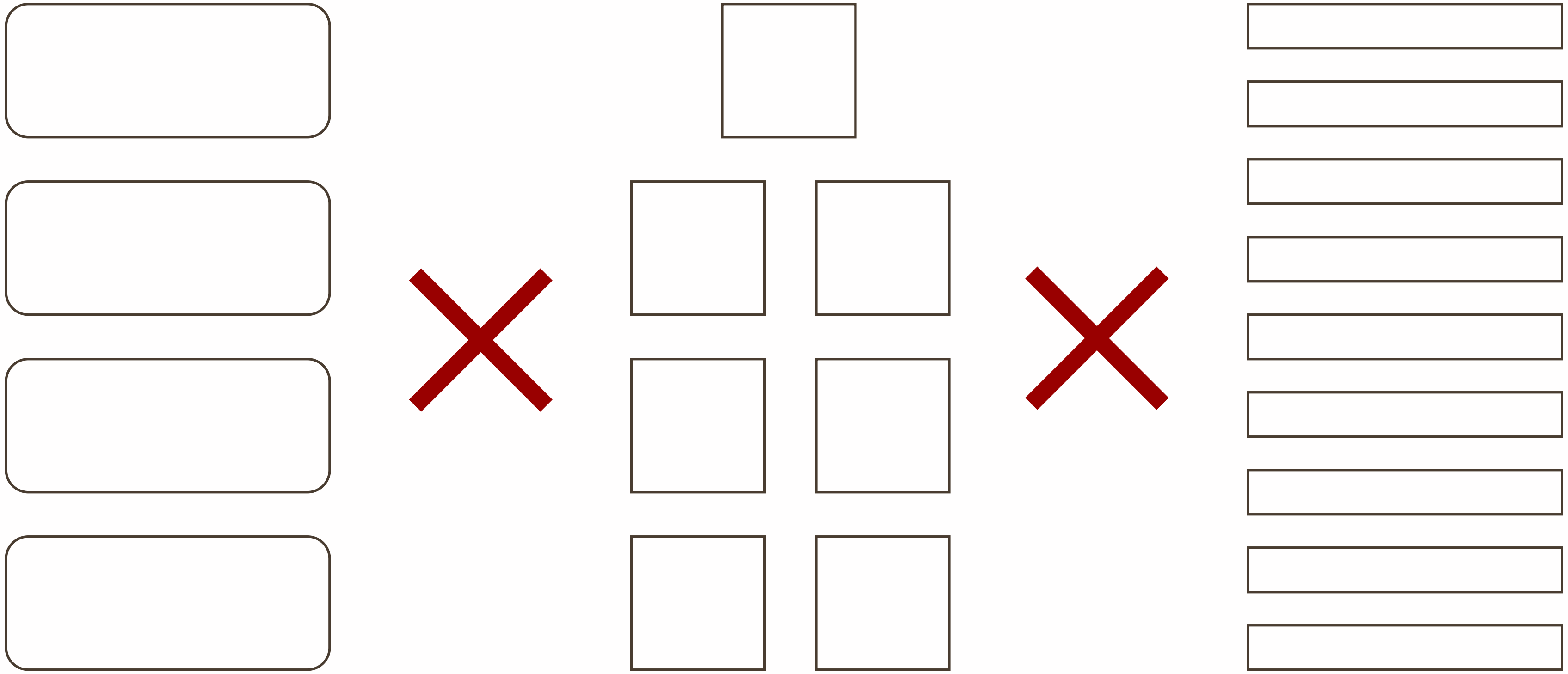
About how many links does it have?

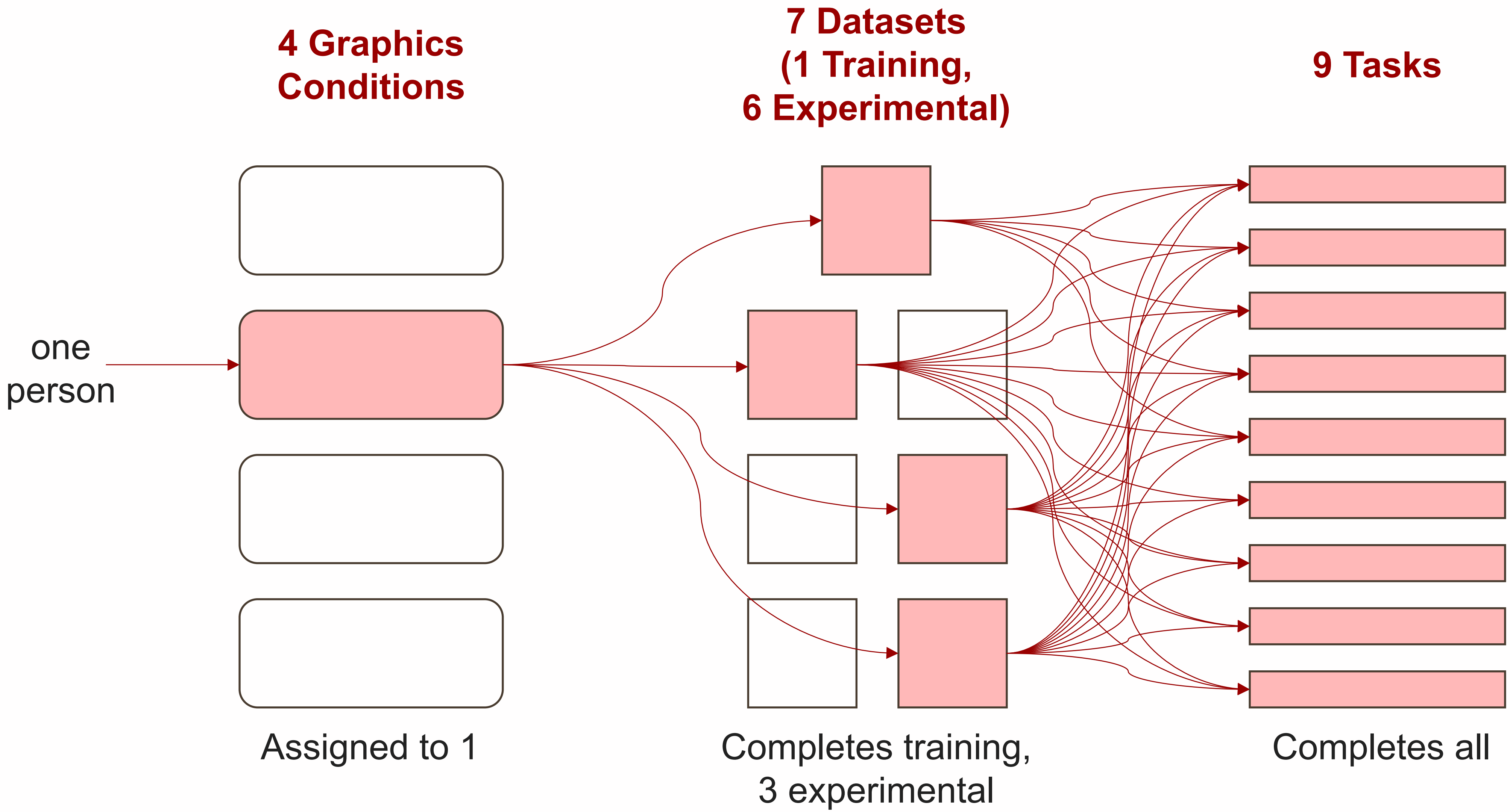
Concrete

Find the most popular person.

About how many friends does he or she have?

4 Graphics Conditions × **7 Datasets (1 Training, 6 Experimental)** × **9 Tasks**




4 Graphics Conditions
7 Datasets (1 Training, 6 Experimental)
9 Tasks
one person
Assigned to 1
Completes training, 3 experimental
Completes all

Training

- Quick introduction to terminology, instructions for study
- Examples of several tasks, with slightly extended instructions
    1. Number of nodes
    2. Number of links
    3. Highest degree node (click task)
    4. Number of clusters
    5. Percentage of nodes in largest cluster
    6. High Betweenness Centrality nodes (click task)
- Correct answers, with explanations

Amazon Mechanical Turk

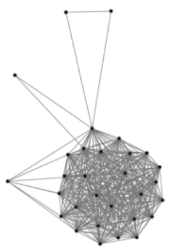
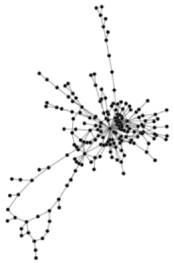
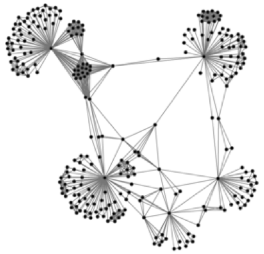
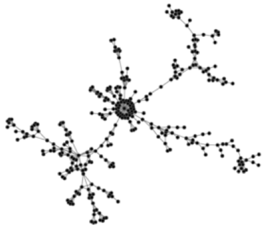
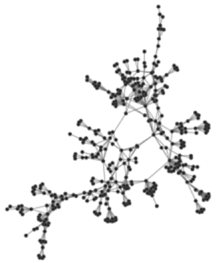
Selection criteria:

- 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

	1	3	5	7	8	9
network						
control	46	49	49	44	50	47
phrasing	54	51	51	52	52	52
color	52	51	52	50	51	54
size	52	55	52	51	49	48

Omitted users (n=55)

- answered at least 20 of the 22 possible non-click training and experimental questions
- were in the lowest 20% of the participants for total duration
- were also in the top 20% of the participants for average error

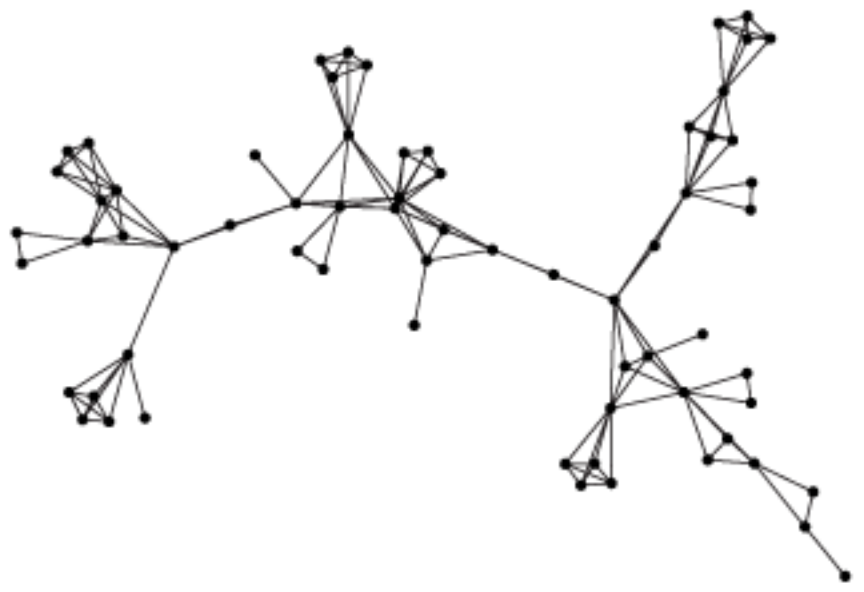
  -or-

- failed to complete a dataset (provided fewer than eight responses)

On Measuring Accuracy

On task accuracy

Numerical Responses

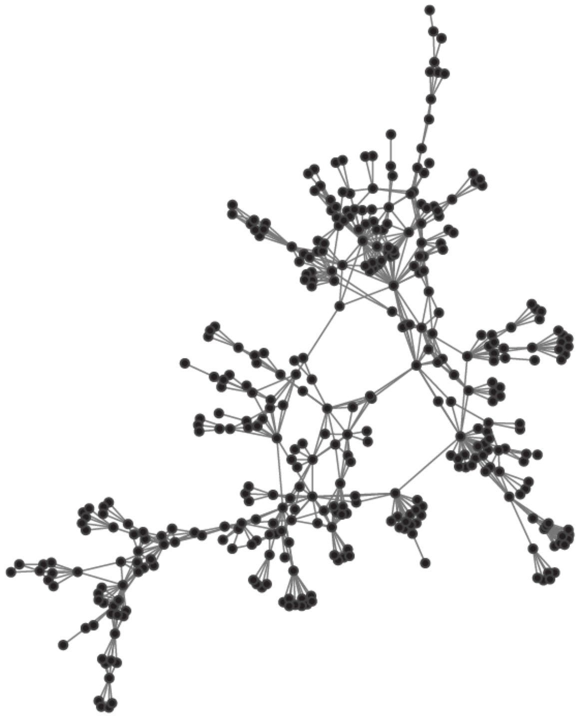


Example: Number of Nodes

Correct Answer:	67
Example Response:	50
Absolute Error:	|67-50| = 17
Error Percentage:	17/67 = 25.4%
Log Absolute Error:	log10(|67-50| + 1) = **1.25**

On task accuracy

Numerical Responses

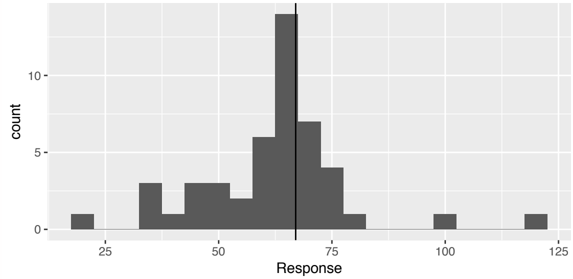


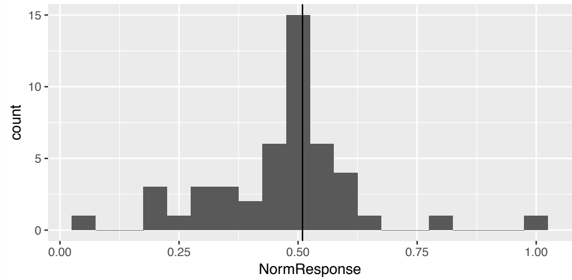
Example: Number of Nodes

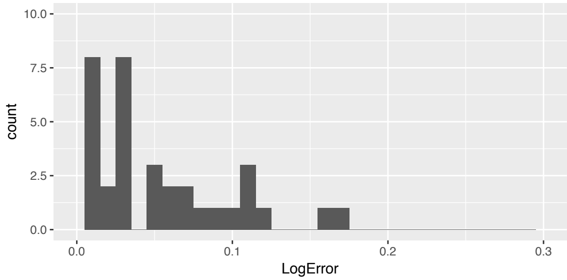
Correct Answer:	379
Example Response:	475
Absolute Error:	|475-379| = 96
Error Percentage:	96/379 = 25.3%
Log Absolute Error:	log10(|475-379| + 1) = **1.97**

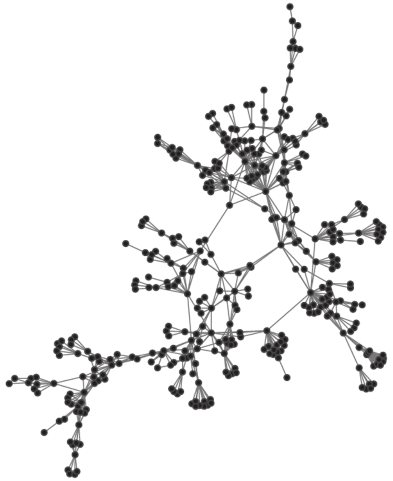
Log Absolute (Normalized) Error

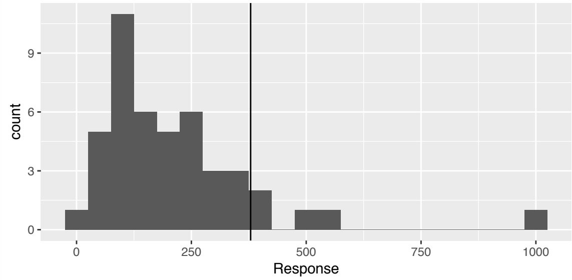
	Responses	Normalized Responses	Log Abs Norm Error

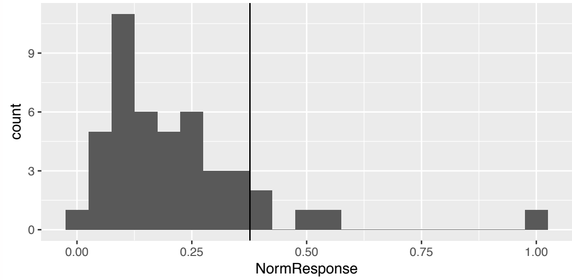


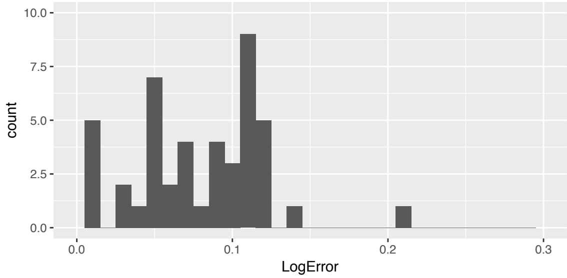












Analyzing non-numeric tasks

Click tasks

Use rank of selected node for the measure being analyzed

(correct answer has rank of 1)

Percentage tasks

Model the response itself, rather than accuracy

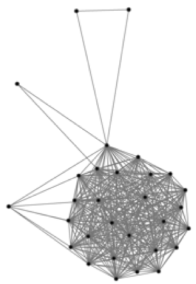
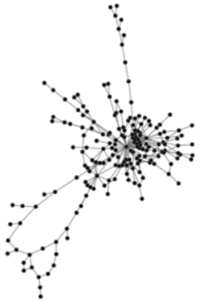
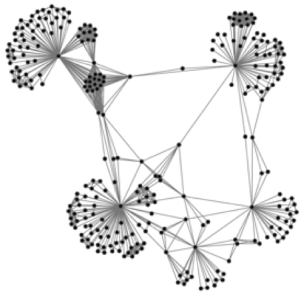
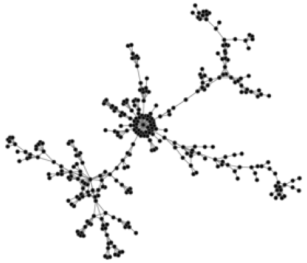
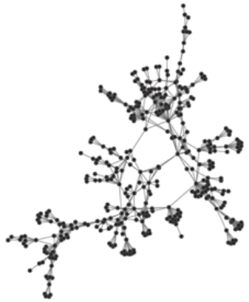
Hypotheses

1. **Varying network properties:**
   Performance on numerical assessments will decline as network size and density increase.
2. **Varying context:**
   Performance on numerical assessments will be higher with the use of concrete (informal) phrasing than with abstract (formal) phrasing.
3. **Varying design:**
   Performance on numerical assessments will be reduced by larger node size and unaffected by using a different (i.e., non-black) dark color.
4. **Varying task:**
   Performance on research tasks that involve clusters will be lower than performance on node- or graph-based tasks.

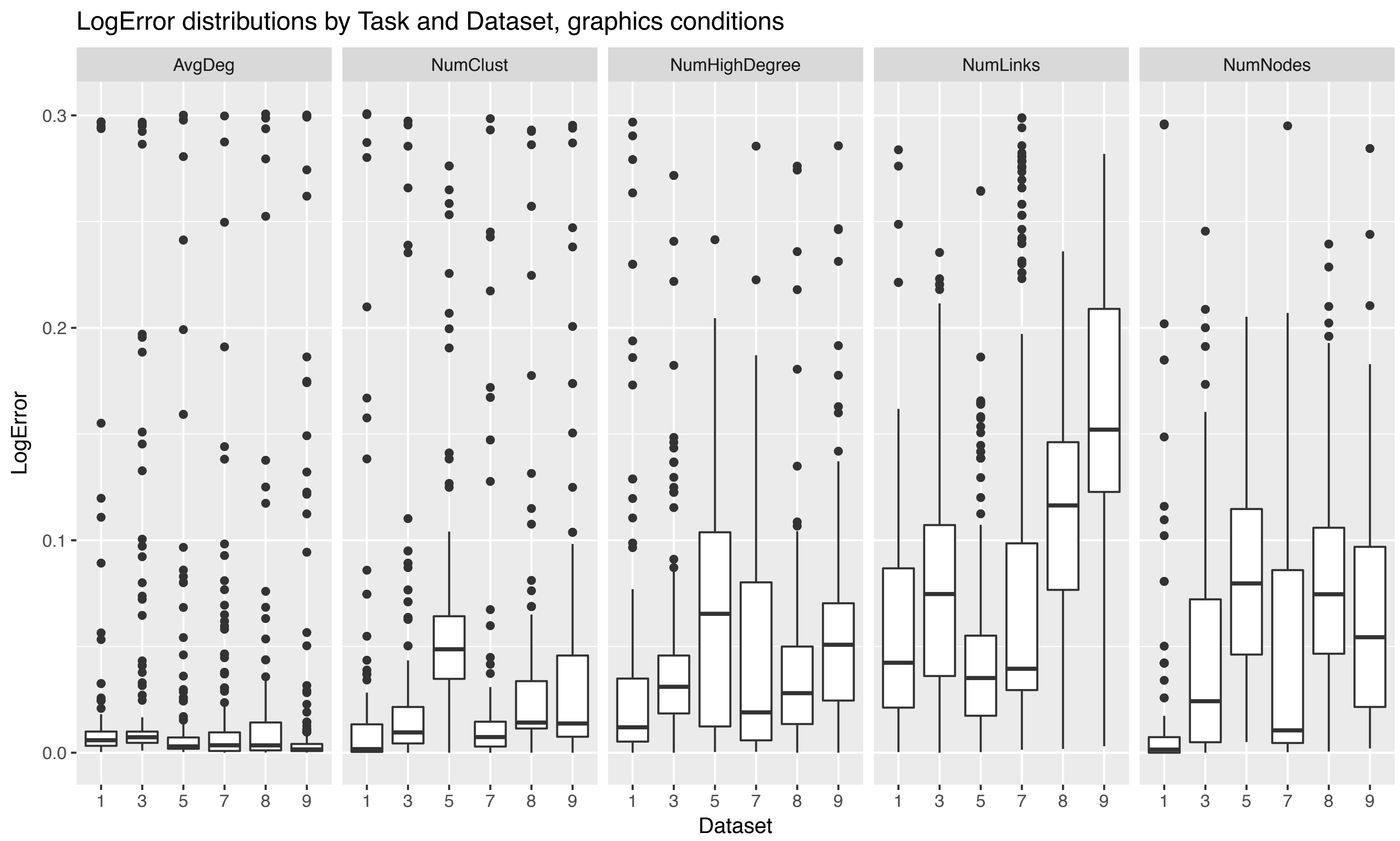
Hypothesis 1:
Larger networks will be harder

Performance will decrease as networks increase in node size, link size, and density.

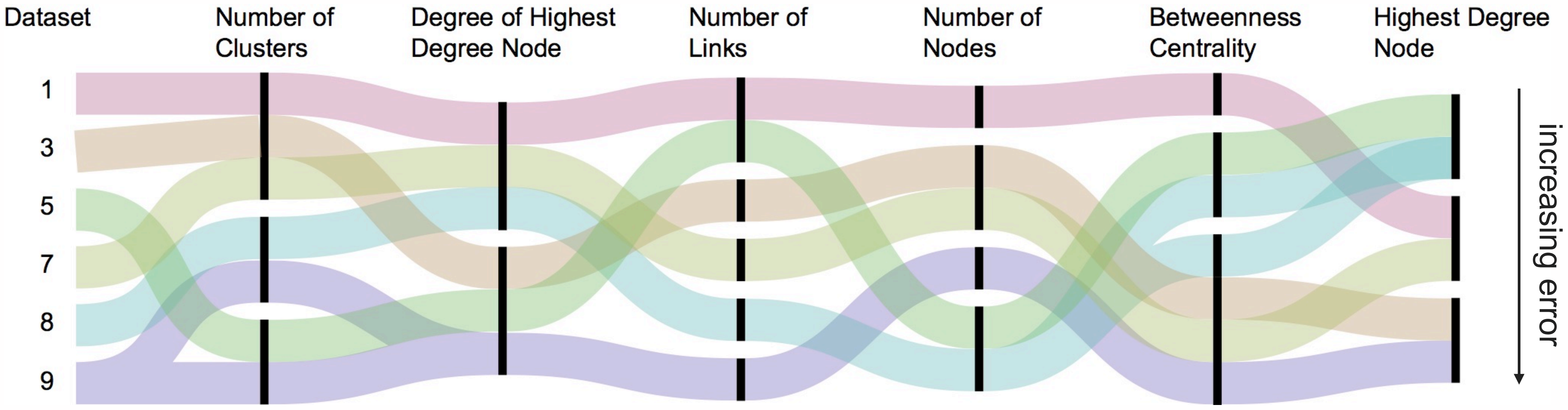
Network datasets

	1	3	5	7	8	9
network						
nodes	30	67	184	270	321	379
edges	337	143	246	932	583	914
density	0.775	0.065	0.015	0.026	0.011	0.013

Task difficulty by dataset



Size matters for extremes (partial support)

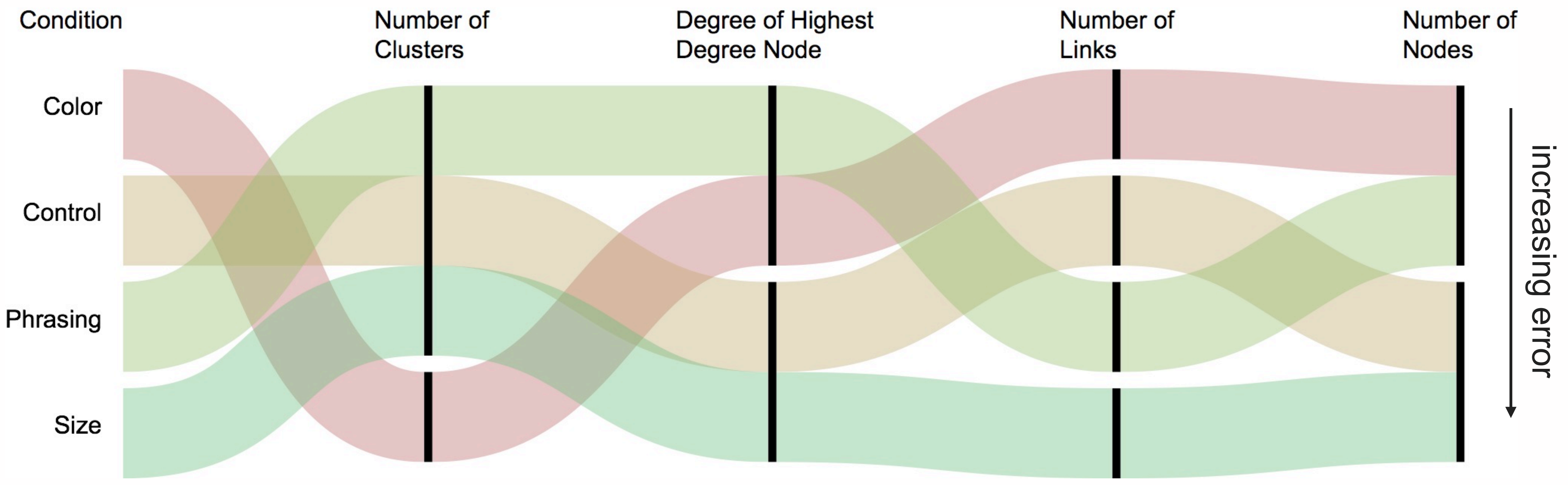


Dataset NS for Average Degree

Hypothesis 2: Informal phrasing will be easier

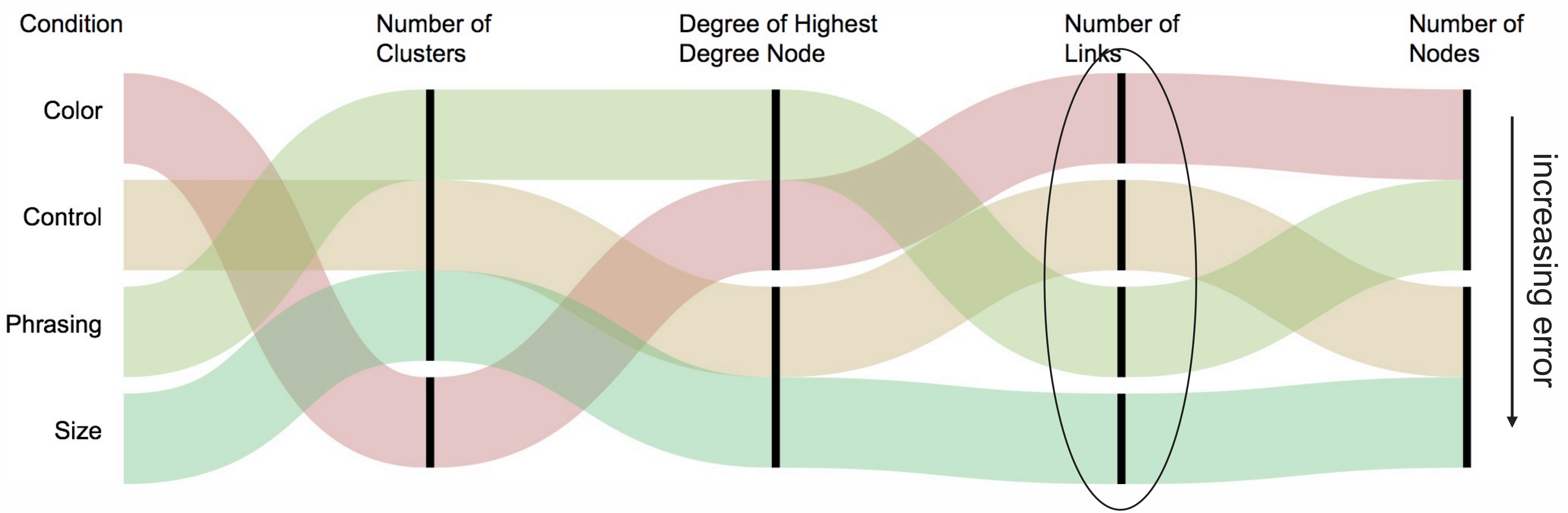
Performance will improve when questions are phrased using informal terminology.

Phrasing does perform best, except for Number of Links task.



Condition NS for Average Degree, Betweenness Centrality, Highest Degree Node

Phrasing does perform best, except for Number of Links task.

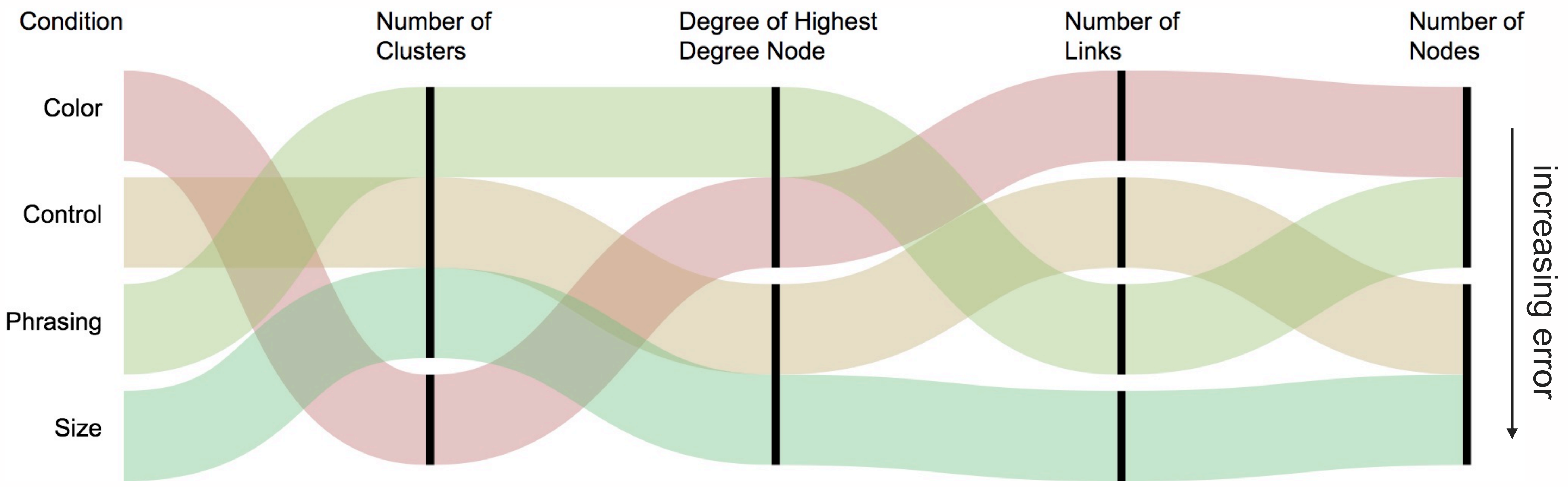


Condition NS for Average Degree, Betweenness Centrality, Highest Degree Node

Hypothesis 3: Changing node color will make no difference

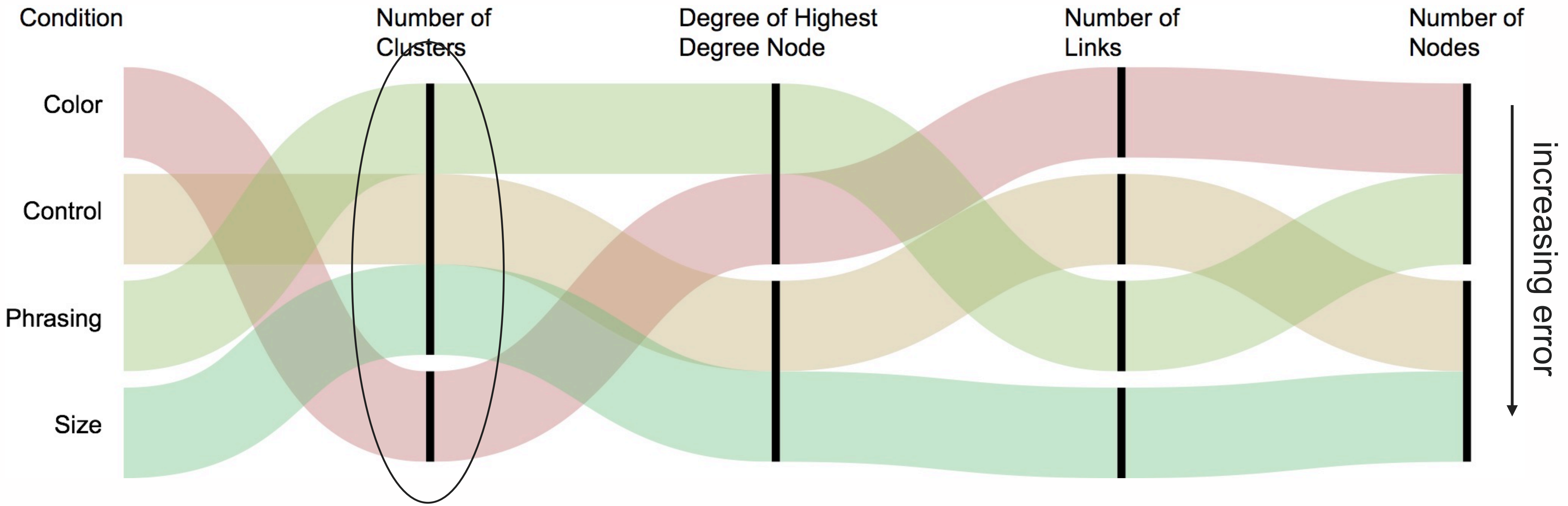
Performance when nodes are changed to an alternate color will be the same as performance on the control condition.

Color almost always better than control.



Condition NS for Average Degree, Betweenness Centrality, Highest Degree Node

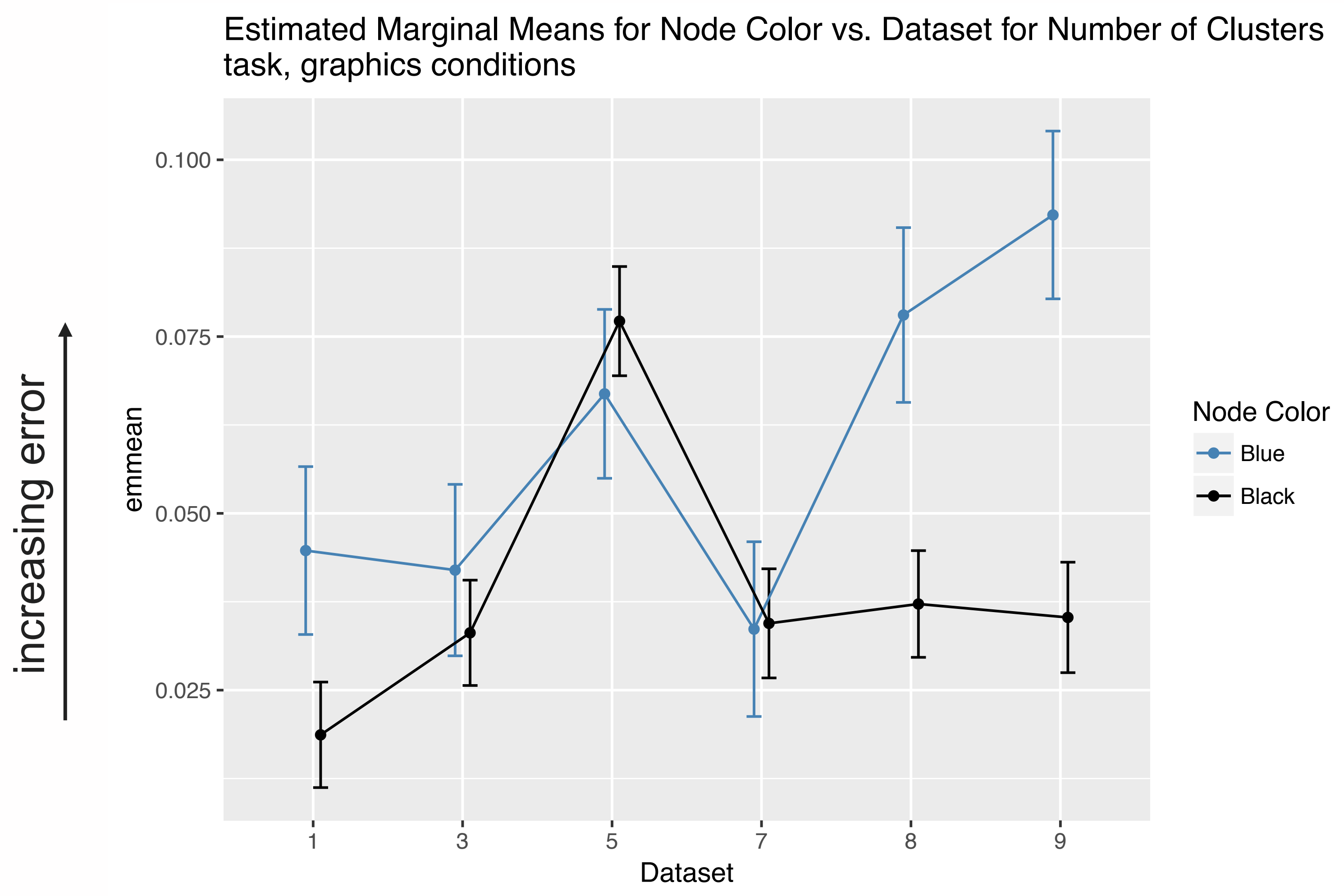
Color almost always better than control.

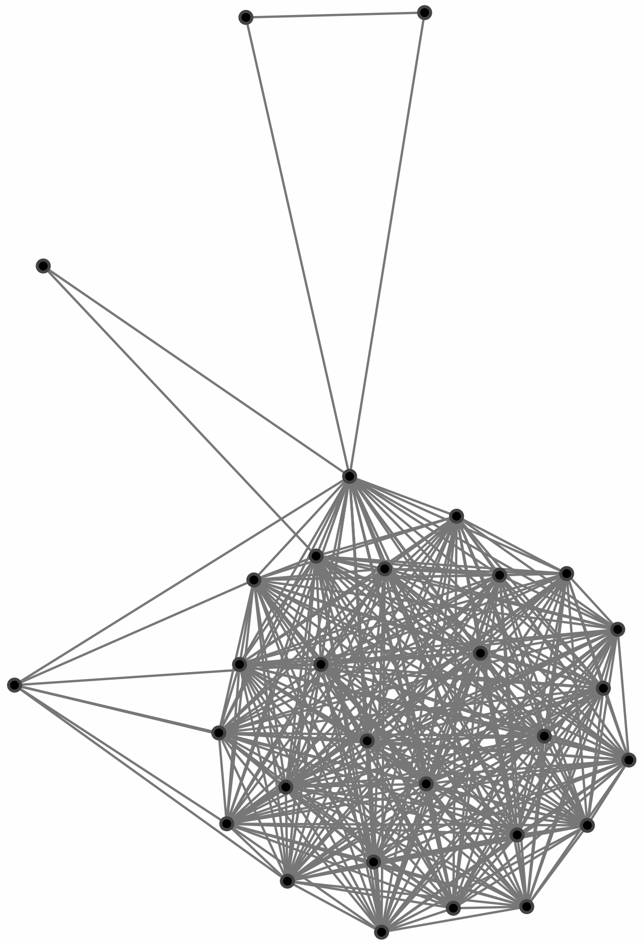


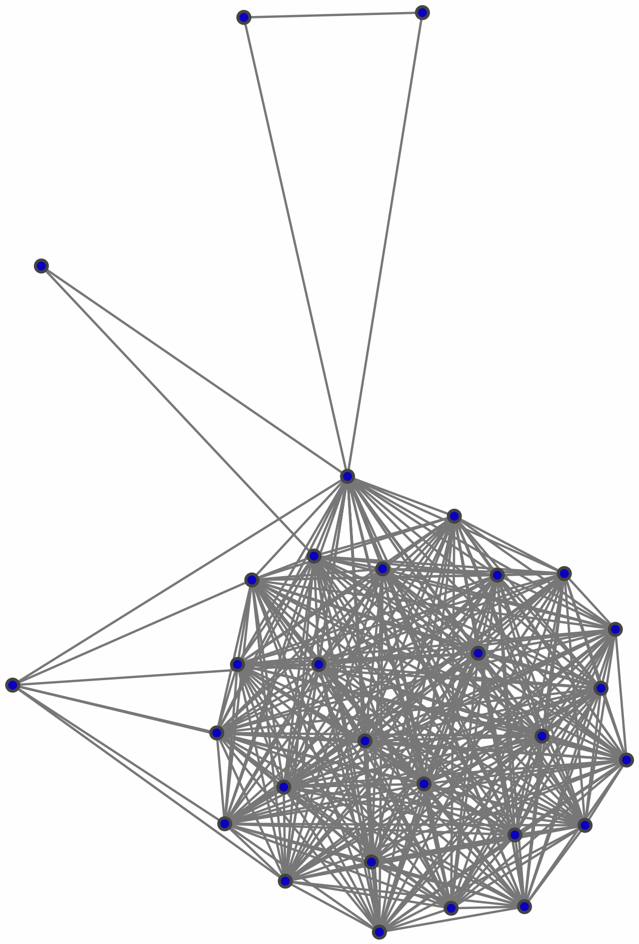
Condition NS for Average Degree, Betweenness Centrality, Highest Degree Node

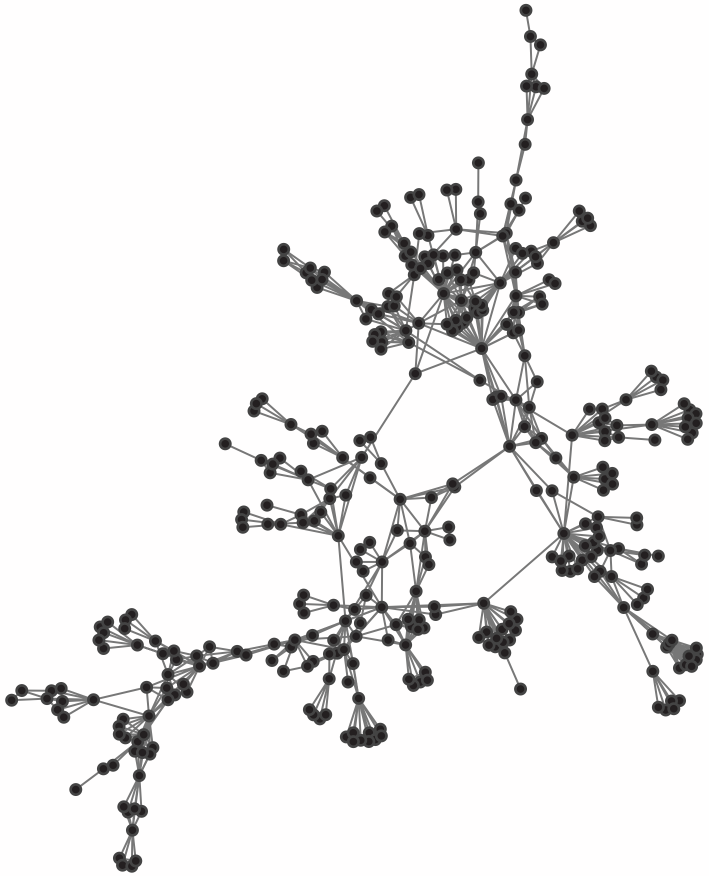
Number of clusters, fixed effects

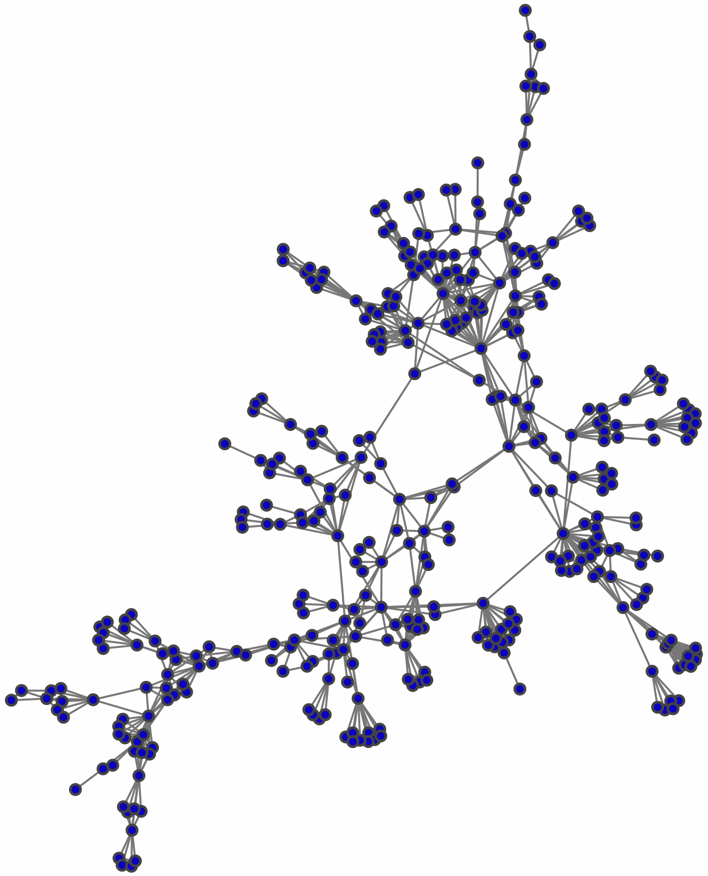
Color of nodes vs. Dataset







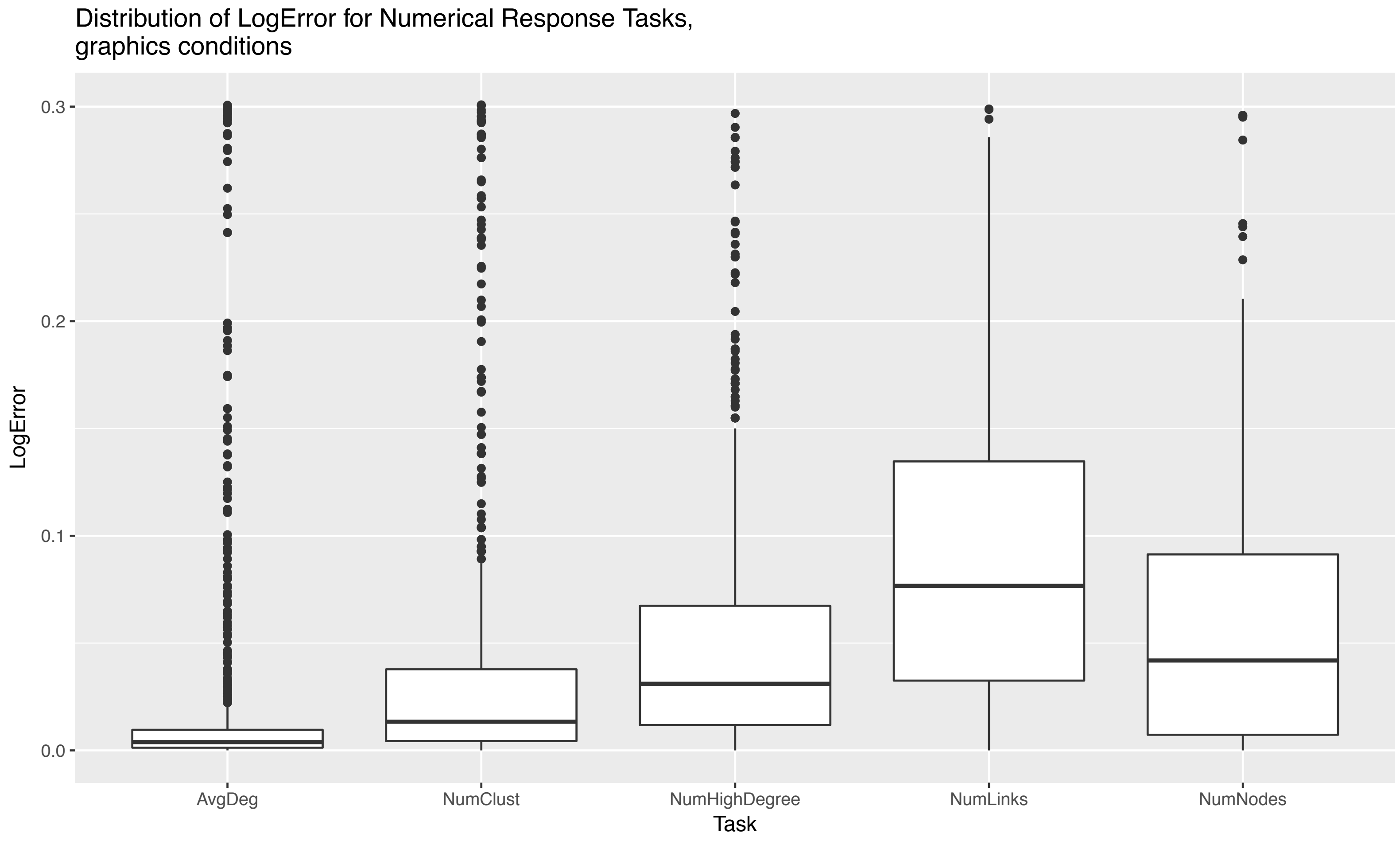




Hypothesis 4:
Tasks involving clusters will be harder

Performance will decline on the number of clusters and size of clusters tasks.

Number of links actually hardest



Results

Hypothesis 1 (varying network properties)	partial support
Hypothesis 2 (varying context)	mostly supported
Hypothesis 3 (varying design)	not supported for color
Hypothesis 4 (varying task)	not supported

Study 2b: Impact of Layout, Expertise

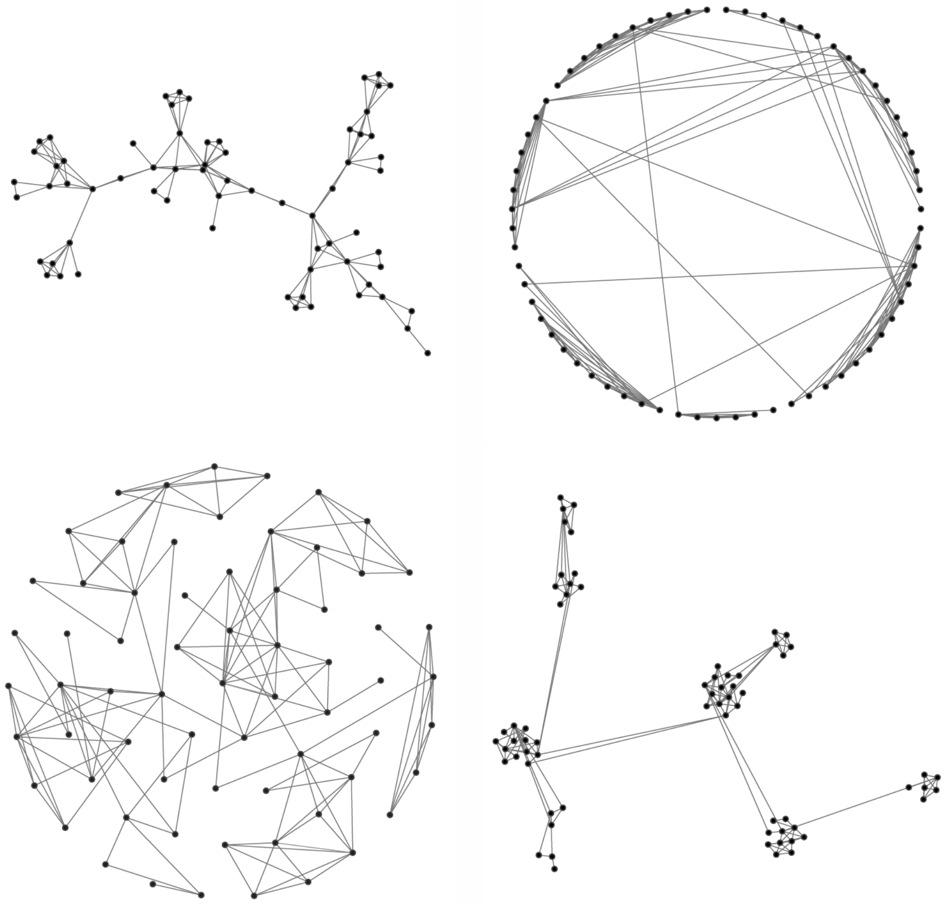
A comparison between Turkers and IUNI Network Science researchers

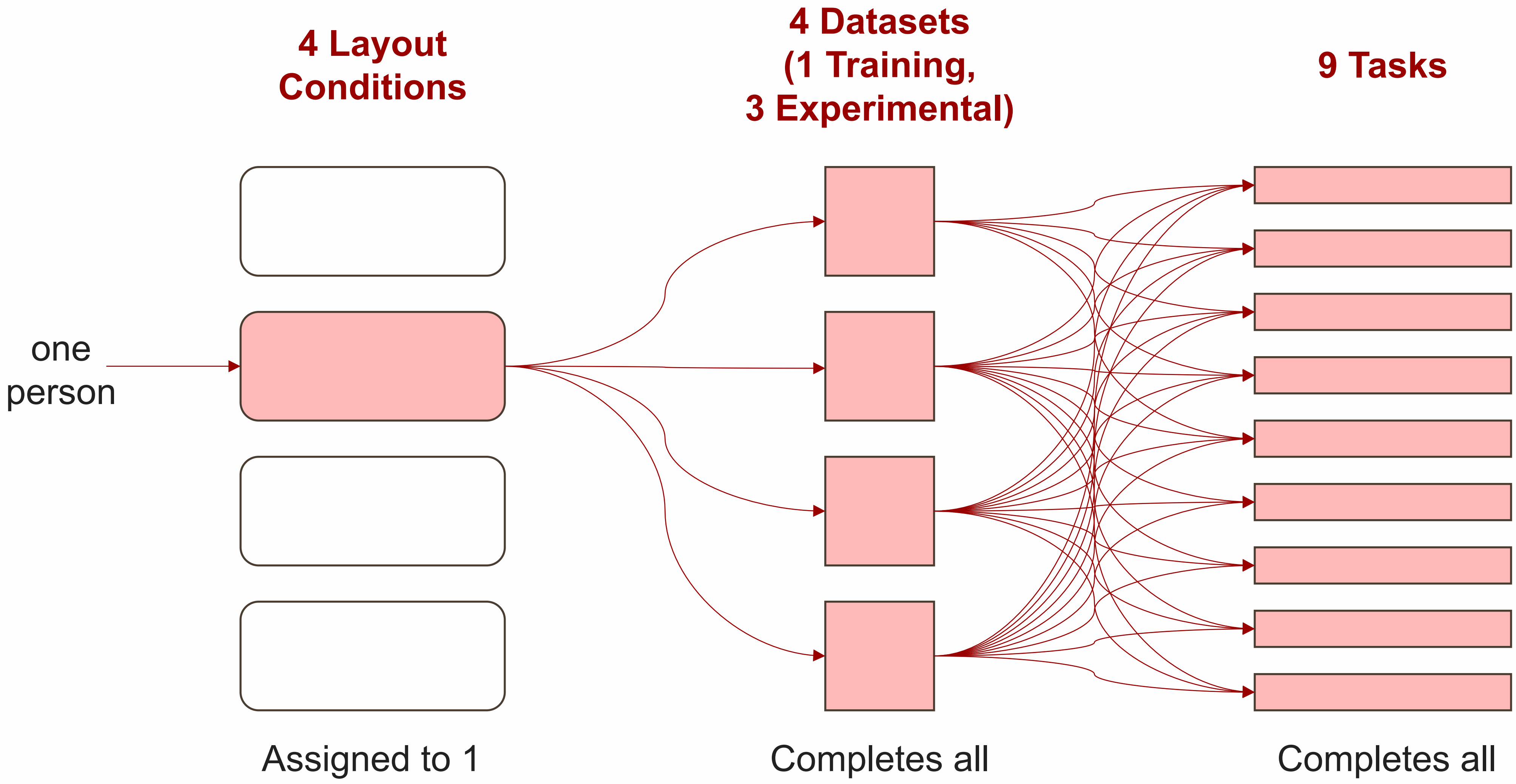
Research Questions

- How differently do network science experts and novices perform when reading network visualizations?
- What kinds of manipulations to network layout affect the overall ability of users to read the visualization? In what ways?
- How do the differing priorities of layout algorithms (e.g., a focus on revealing clusters) affect performance on specific network interpretation tasks (e.g., cluster detection)?

Layout Conditions (between subjects)

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




4 Layout
Conditions
4 Datasets
(1 Training,
3 Experimental)
9 Tasks
one
person
Assigned to 1
Completes all
Completes all

IU Network Science community (n=238)

Selection criteria:

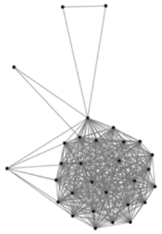
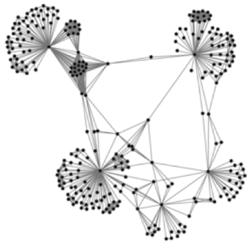
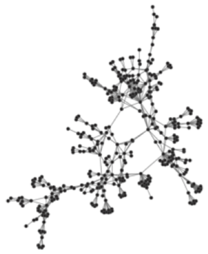
- Affiliated with IUNI, CNS program, or other network science training

Compensation:

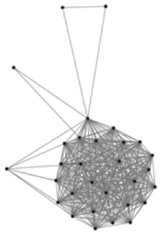
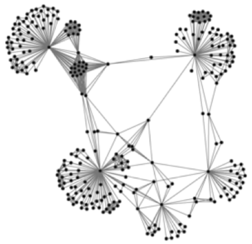
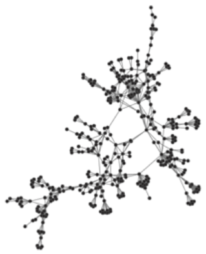
- Pilot: drawing for two $50 Amazon Gift Cards
- Graduate Students: $10 Amazon Gift Cards, pizza
- Faculty/Staff: randomly assigned to two conditions:
  $10 Amazon Gift Card, $10 donation to IU diversity scholarship

Participants

IU NetSci

network			
control	23	23	22
circle	19	17	19
frucht	17	17	16
openord	22	23	23

MTurk

network			
control	46	44	47
circle	46	44	44
frucht	49	51	52
openord	49	48	49

Hypotheses

5. **Varying expertise:**
   Network scientists will perform better than novices on numerical assessments tasks, even when layout changes.
6. **Varying layout:**
   Different layouts will relate to performance improvements on certain tasks:
   a) *Use of the OpenOrd layout, which prioritizes clustering, will relate to better performance on clustering tasks.*
   b) *Use of the Fruchterman-Reingold layout, which prioritizes even node distribution, will relate to better performance on tasks that involve counting nodes, locating nodes, or assessing/comparing node properties.*
   c) *Use of the Circular layout will relate to decreased performance on all tasks.*

H5: Network experts will perform better.

Network science expertise will improve performance, even across layout changes.

H5 not supported.

None of the task performance models retained **network science training** as a significant predictor of accuracy, even when collapsing across condition increases the sample size.

H5 not supported.

None of the task performance models retained **network science training** as a significant predictor of accuracy, even when collapsing across condition increases the sample size.

☺ Users don’t need much training to perform as well as experts!

H5 not supported.

None of the task performance models retained **network science training** as a significant predictor of accuracy, even when collapsing across condition increases the sample size.

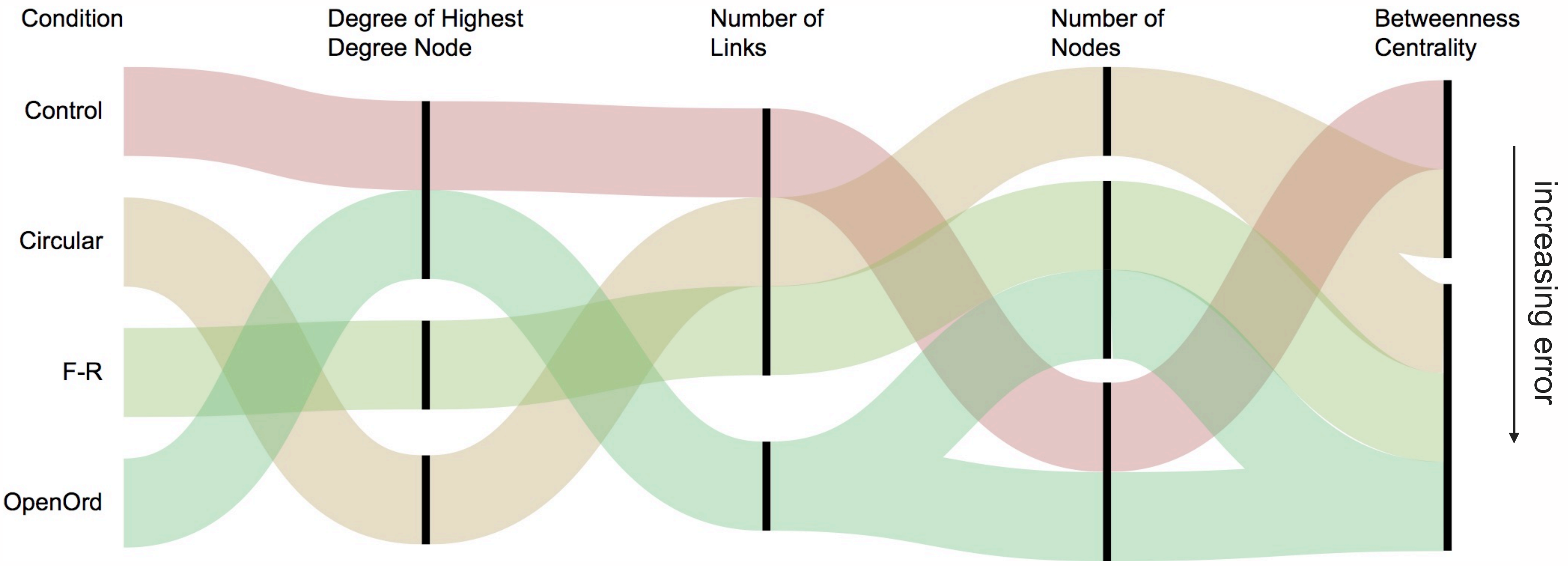
☺ Users don’t need much training to perform as well as experts!

☹ Not even experts get especially high accuracy scores.

H6: Different layouts will improve performance on different tasks.

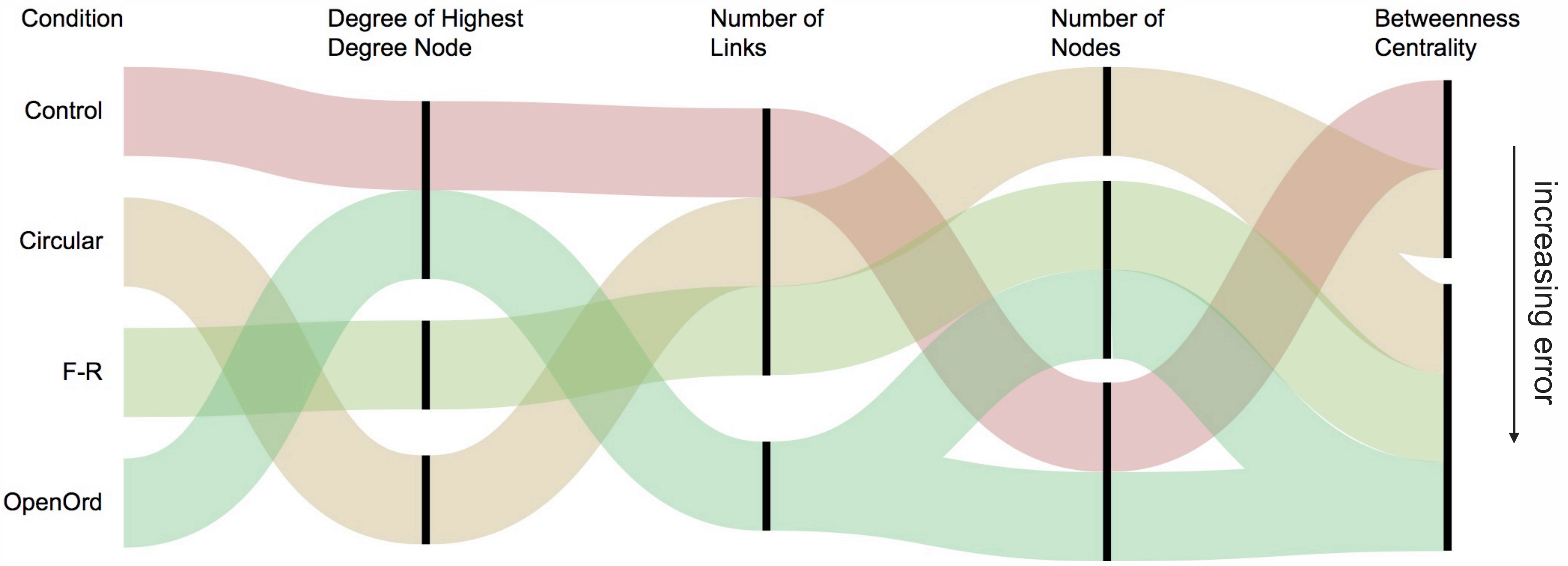
a) OpenOrd will improve performance on clustering tasks.
b) Fruchterman-Reingold will improve performance on node tasks.
c) Circular layout will impair performance on all tasks.

Layout vs. Task



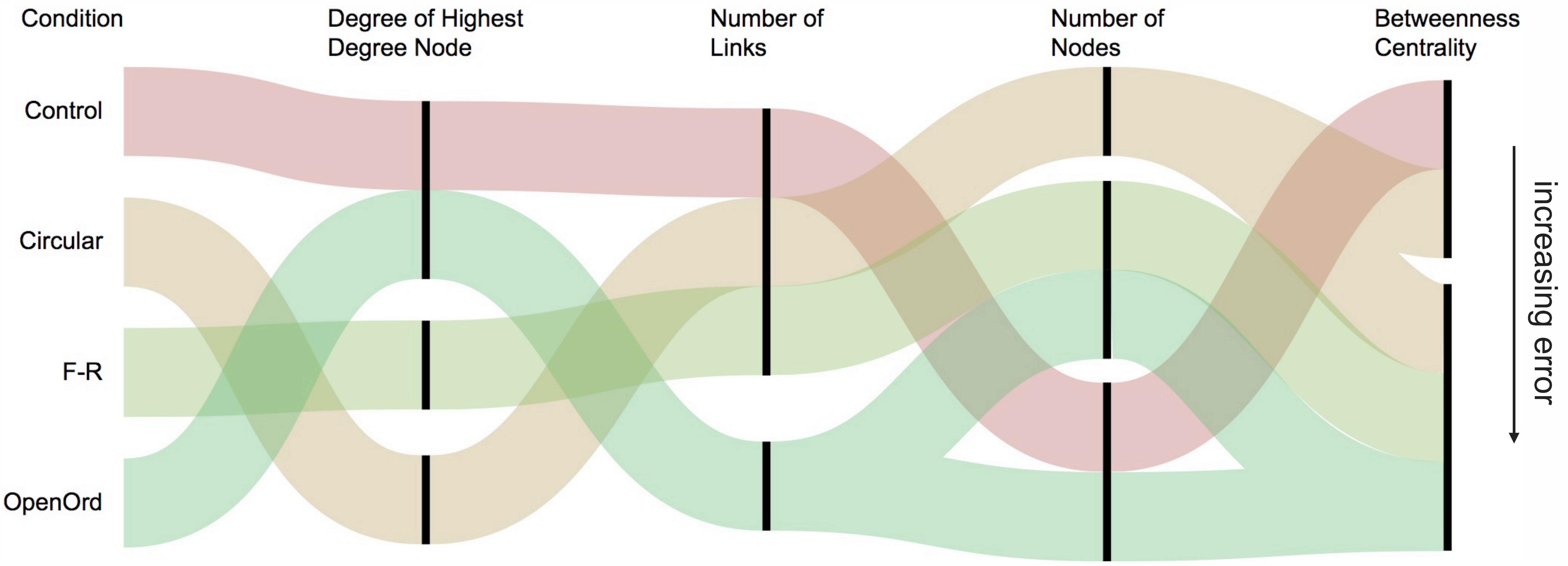
Condition NS for Average Degree, Number of Clusters. Result unknown for Highest Degree Node.

~~OpenOrd~~: Condition NS for cluster tasks



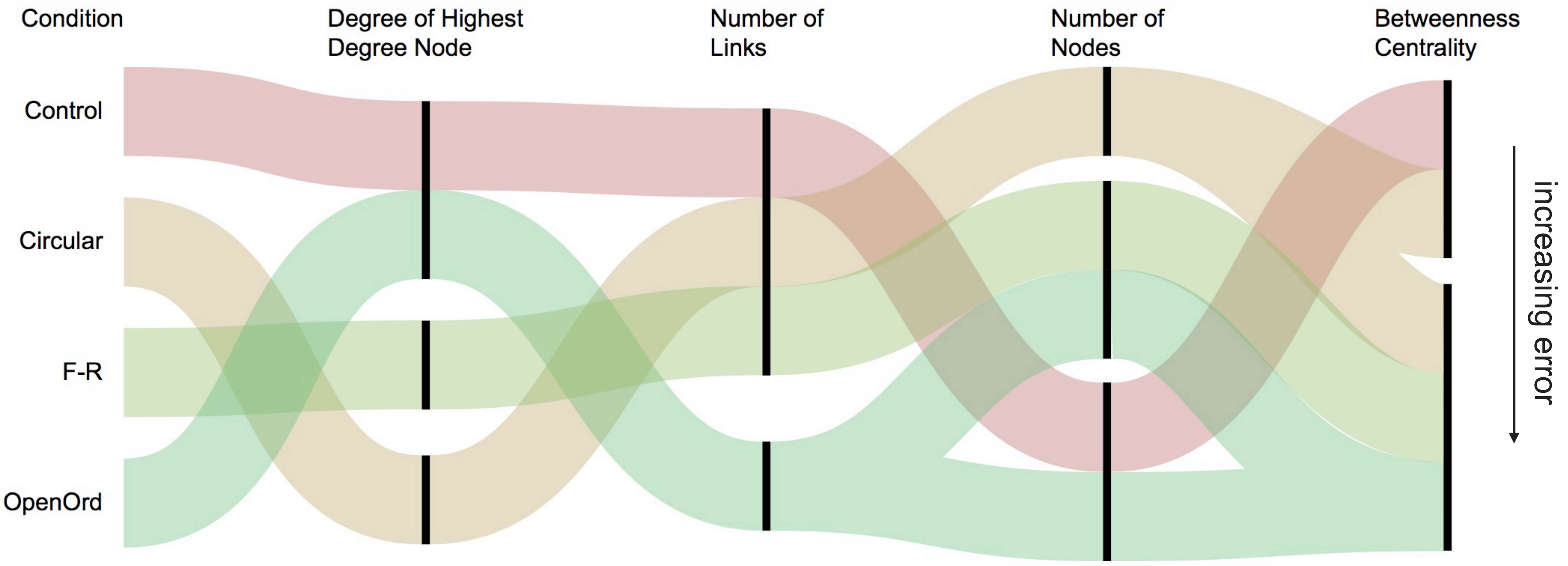
Condition NS for Average Degree, Number of Clusters. Result unknown for Highest Degree Node.

~~FR~~: FR not especially good for node tasks



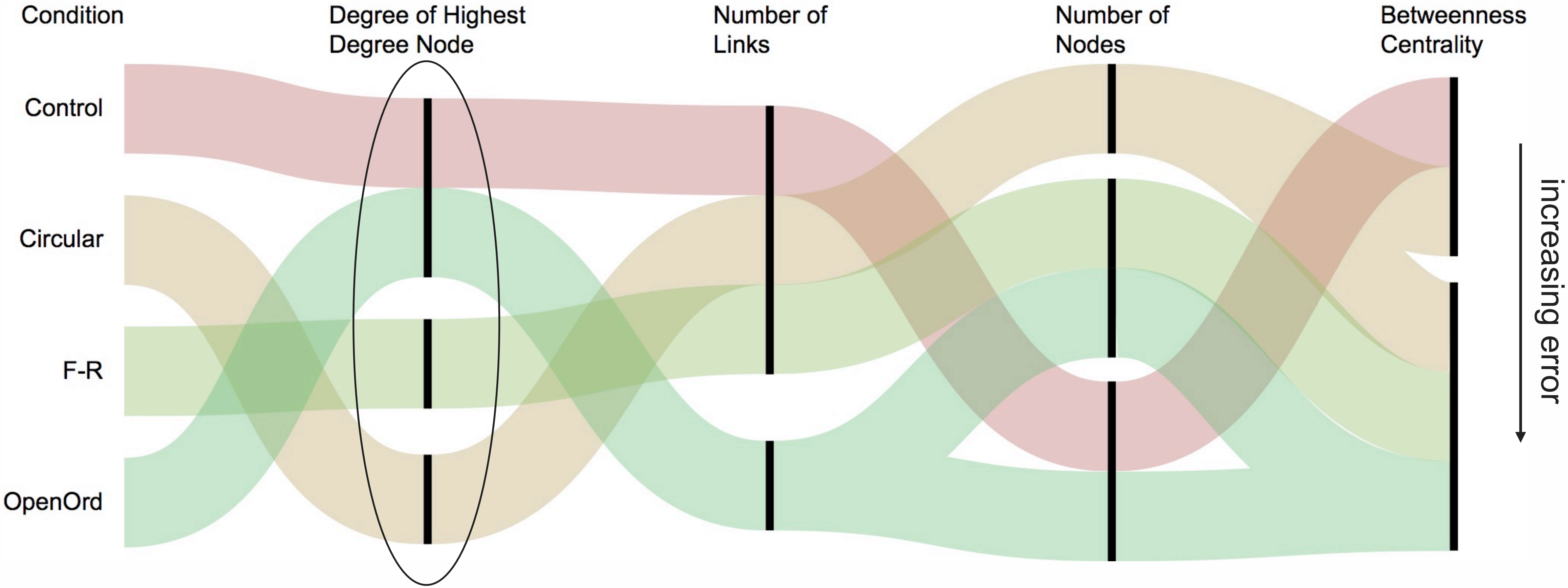
Condition NS for Average Degree, Number of Clusters. Result unknown for Highest Degree Node.

~~Circular~~: Circular is generally fine!



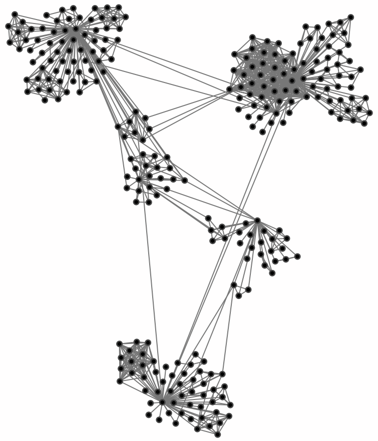
Condition NS for Average Degree, Number of Clusters. Result unknown for Highest Degree Node.

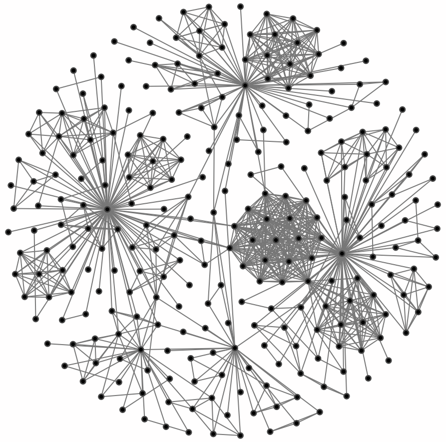
~~Circular~~: Circular is generally fine!



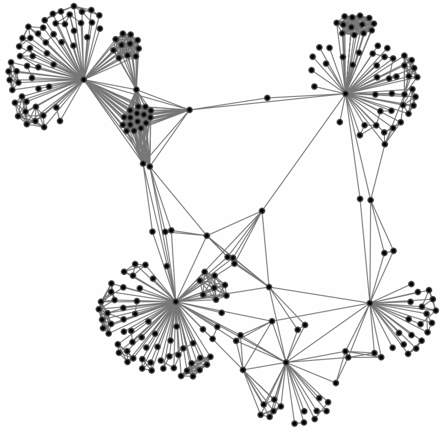
Condition NS for Average Degree, Number of Clusters. Result unknown for Highest Degree Node.

Dataset 7 especially problematic

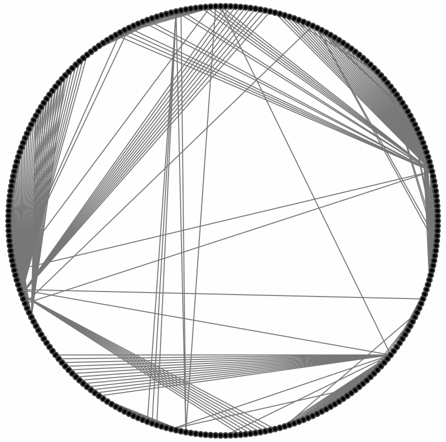




>



>



Results

Hypothesis 5 (varying expertise)	not supported
Hypothesis 6a (OpenOrd)	not supported
Hypothesis 6b (Fruchterman-Reingold)	not supported
Hypothesis 6c (Circular)	not supported

Conclusions

Recommendations

- Including brief training should allow novices to read network visualizations well (enough)
- A brighter node color may indeed improve performance
- Layout algorithm may influence task performance, but the relationship is complex.
  - Force-directed is fine overall
  - Circular is fine for tasks that don't require estimating degree

Challenges

- Selecting and operationalizing task
- Operationalizing error
- Recruiting experts

Possible extensions

- Different tasks
- Different populations
- Data overlays
- Interactivity
- 3D visualizations
- Qualitative data on interpretation strategies

References

- Arnheim, R. (1969). *Visual thinking.* Berkeley: University of California Press.
- Du, F., Plaisant, C., Spring, N., & Shneiderman, B. (2017). Finding similar people to guide life choices: Challenge, design, and evaluation. In S. Fussell & G. Mark (Eds.), *CHI '17: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 5498-5544). New York, NY: ACM.
- Healey, C. G., & Enns, J. T. (2012). Attention and Visual Memory in Visualization and Computer Graphics. *IEEE Transactions on Visualization and Computer Graphics, 18*(7), 1170-1188.
- Sprague, D., & Tory, M. (2012). Exploring how and why people use visualizations in casual contexts: Modeling user goals and regulated motivations. *Information Visualization, 11*(2), 106-123.
- Ware, C. (2012). *Information Visualization: Perception for Design*. (3rd Edition). Morgan Kaufman.
- Zoss, A. M. (2014). Cognitive processes and traits related to graphic comprehension. In M. L. Huang, W. Huang (Eds.), *Innovative Approaches of Data Visualization and Visual Analytics*, pp. 94-110, IGI Global. DOI:10.4018/978-1-4666-4309-3.ch005

Questions?

amzoss@iu.edu

netvislit.org