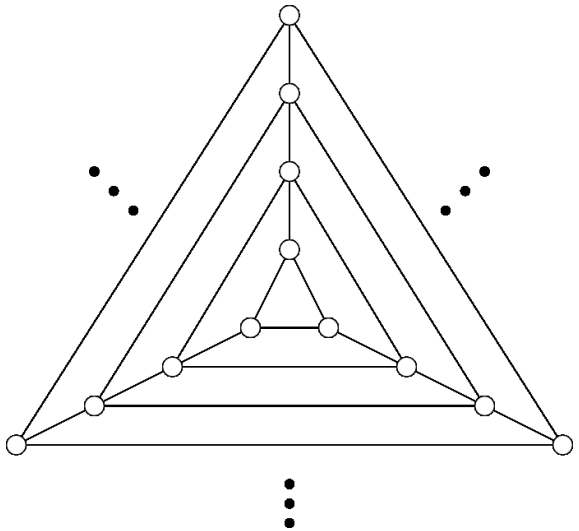

THE UNIVERSITY OF ARIZONA®

# Are Crossings Important for Drawing Large Graphs?

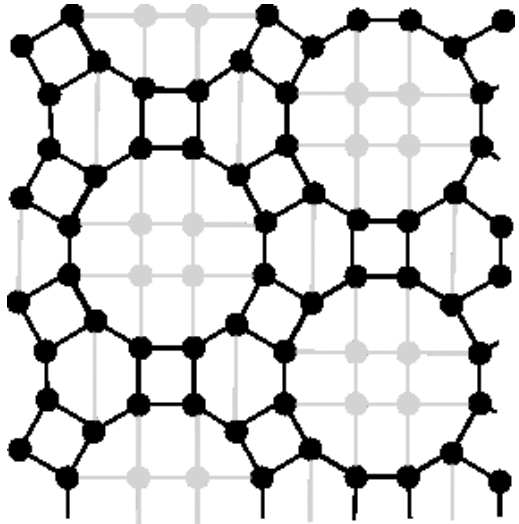
Sergey Pupyrev
University of Arizona

Joint work with Bahador Saket and Stephen Kobourov

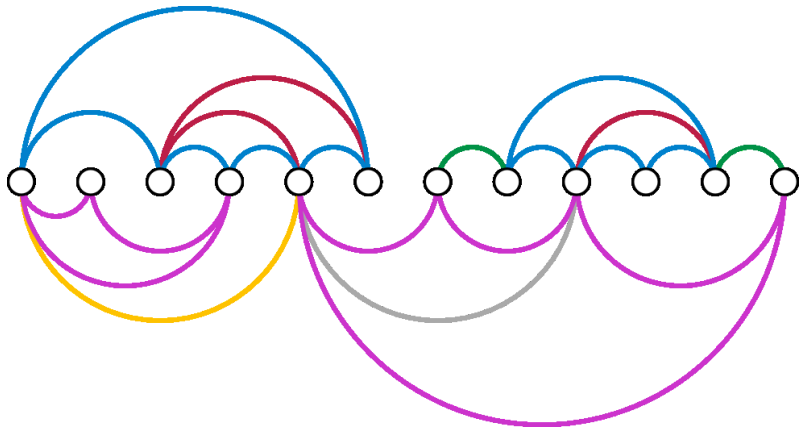
# Graph Drawing *in theory*



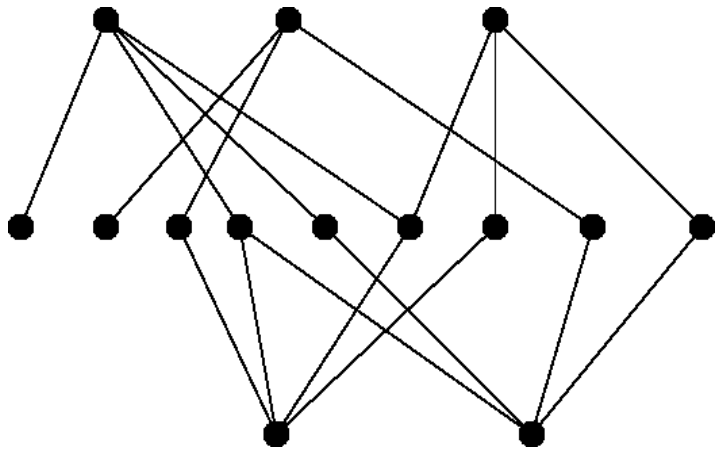
Alam et al. GD'14



Kleist Rahman GD'14

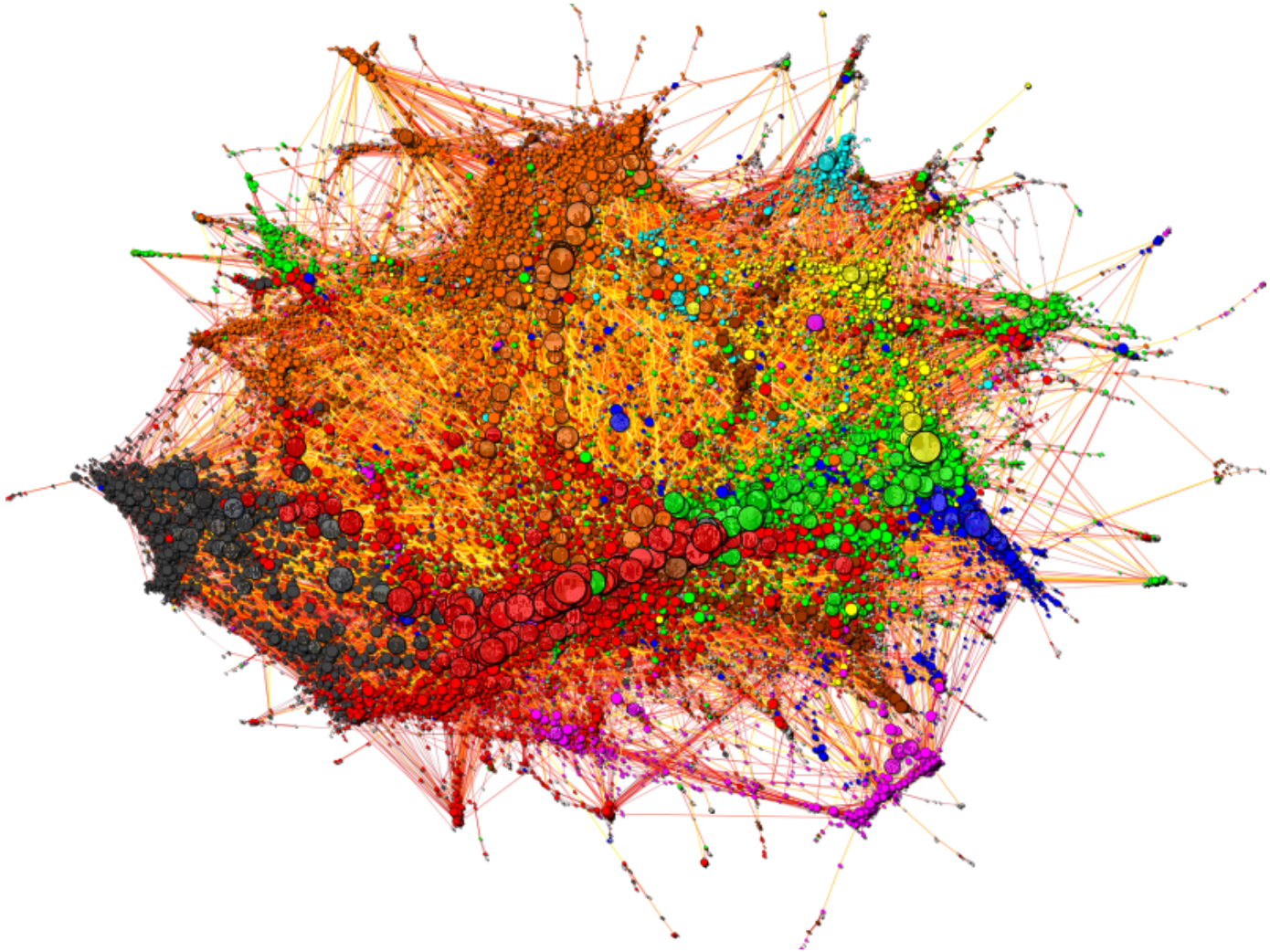


Bannister Eppstein GD'14



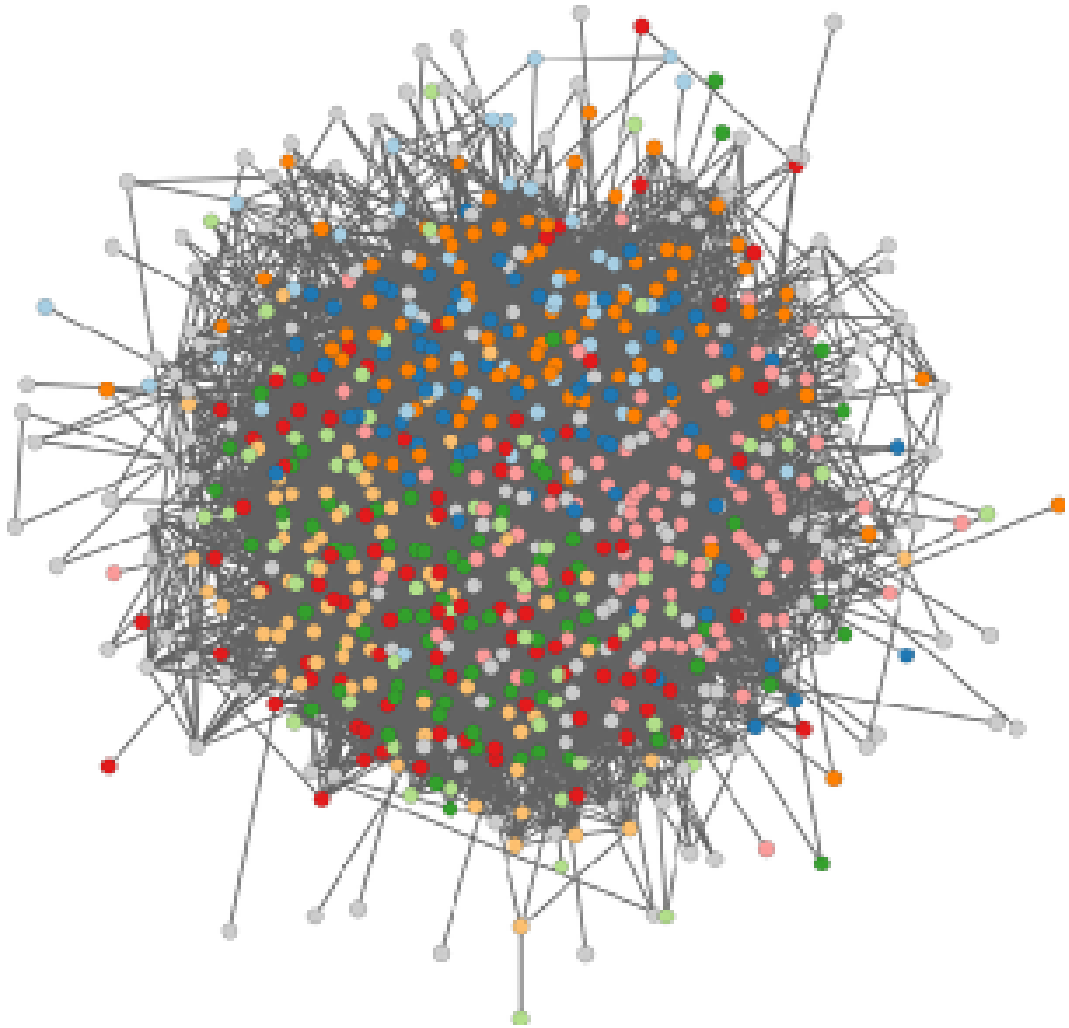
Binucci et al. GD'14

# Graph Drawing *in practice*



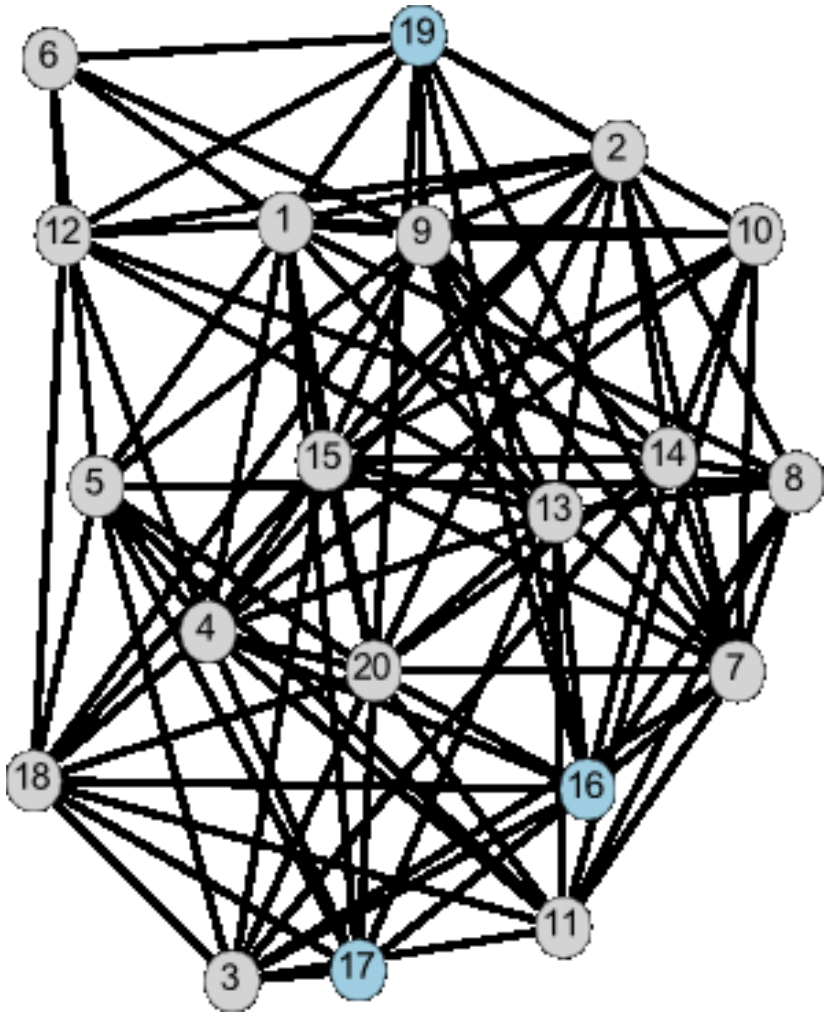
Nepusz 2009

# Graph Drawing *in practice*



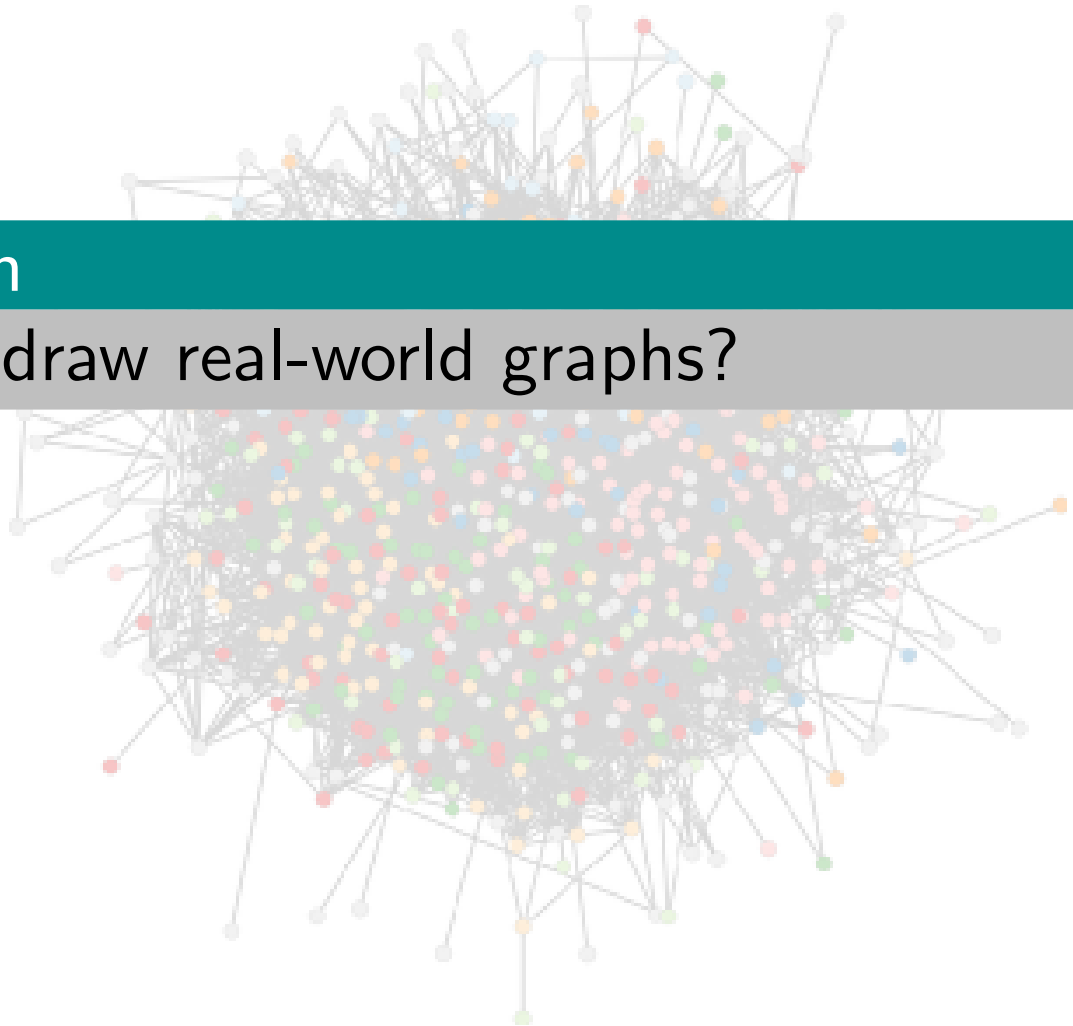
Nocaj Ortmann Brandes GD'14

# Graph Drawing *in practice*




Hu Shi GD'14

# Graph Drawing *in practice*



**Question**

How to draw real-world graphs?

# Aesthetics

- number of edge crossings
- number of edge bends
- angular resolution
- crossing angles
- uniform vertex distribution
- symmetry

# Prior experiments

*"minimizing edge crossings is an important aid to human understanding"*

Purchase Cohen James, GD'96

*"there is strong evidence to support minimising (edge) crosses"*

Purchase, GD'97

*"the most important factors are continuity and edge crossings"*

Ware Purchase Colpoys McGill, IV'02

*"edge crossings and conventions pose significant effects on user preference and task performance"*

Huang Hong Eades, GD'05

*"the number of edge crossings is relatively more important than the size of crossing angles"*

Huang Huang, AI'14

# Prior experiments

*"minimizing edge crossings is an important aid to human understanding"*

Purchase Cohen James, GD'96

**Observation**

Minimizing edge crossings remains the most cited and the most commonly used aesthetic!

*"the most important factors are continuity and edge crossings"*

Ware Purchase Colpoys McGill, IV'02

*"edge crossings and conventions pose significant effects on user preference and task performance"*

Huang Hong Eades, GD'05

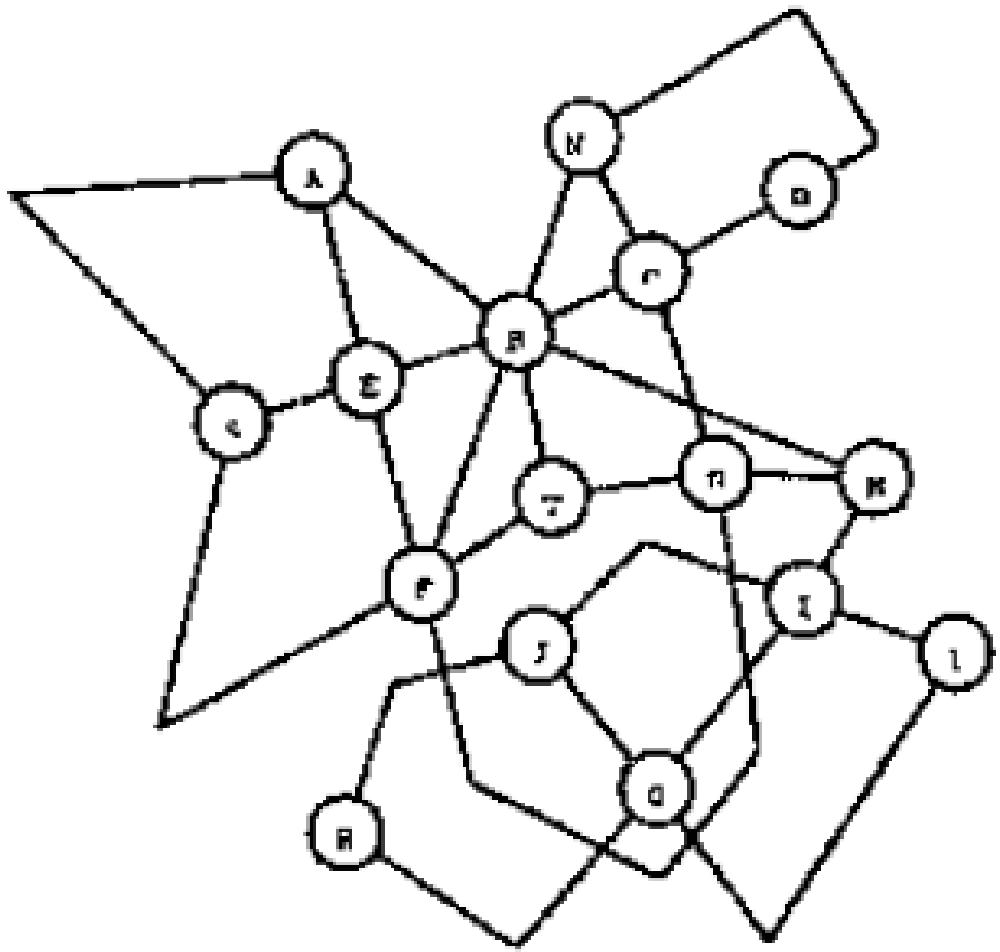
*"the number of edge crossings is relatively more important than the size of crossing angles"*

Huang Huang, AI'14

# Prior experiments

- Purchase Cohen James, GD'96

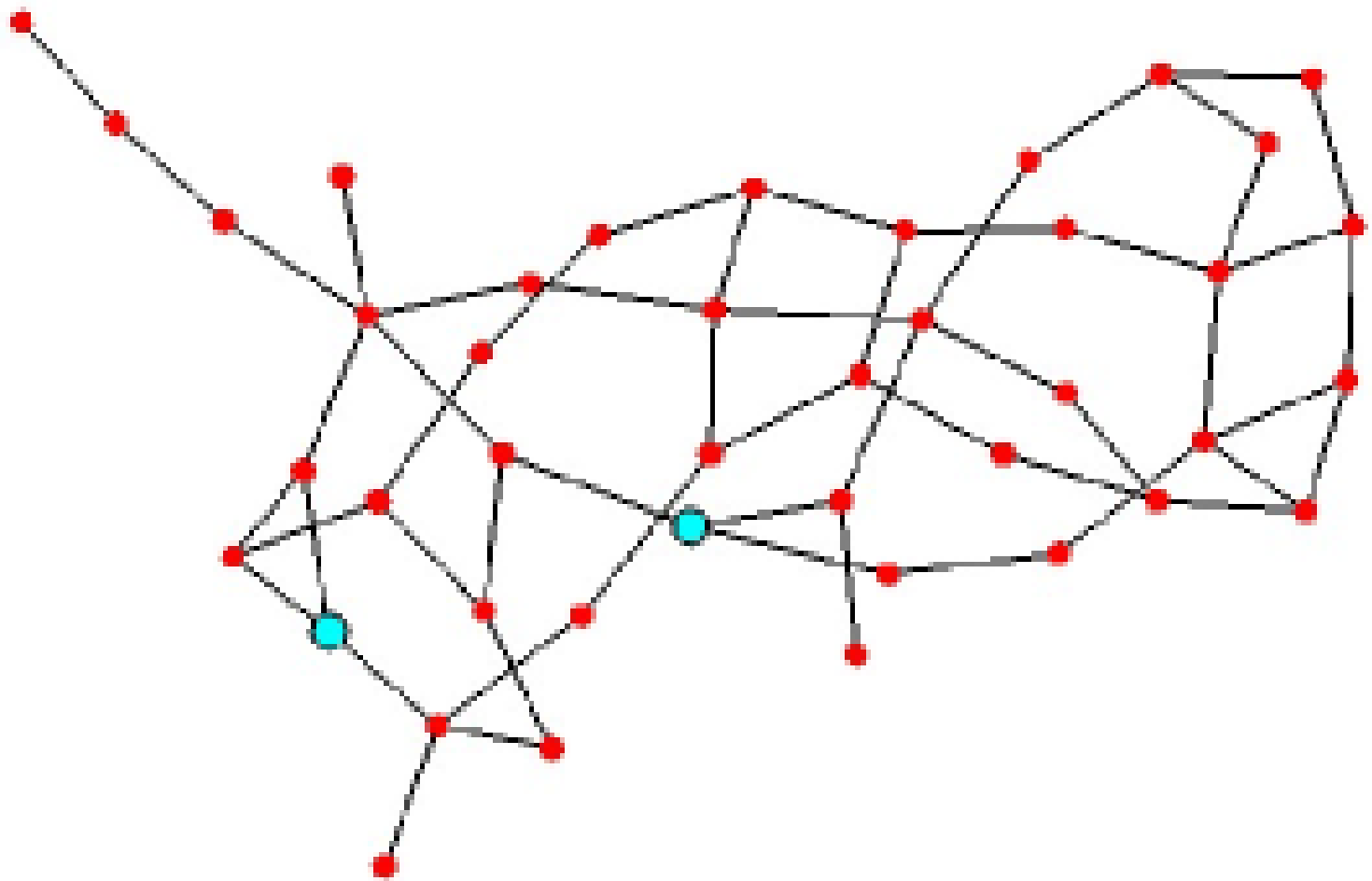
  16 vertices, 18 – 28 edges



# Prior experiments

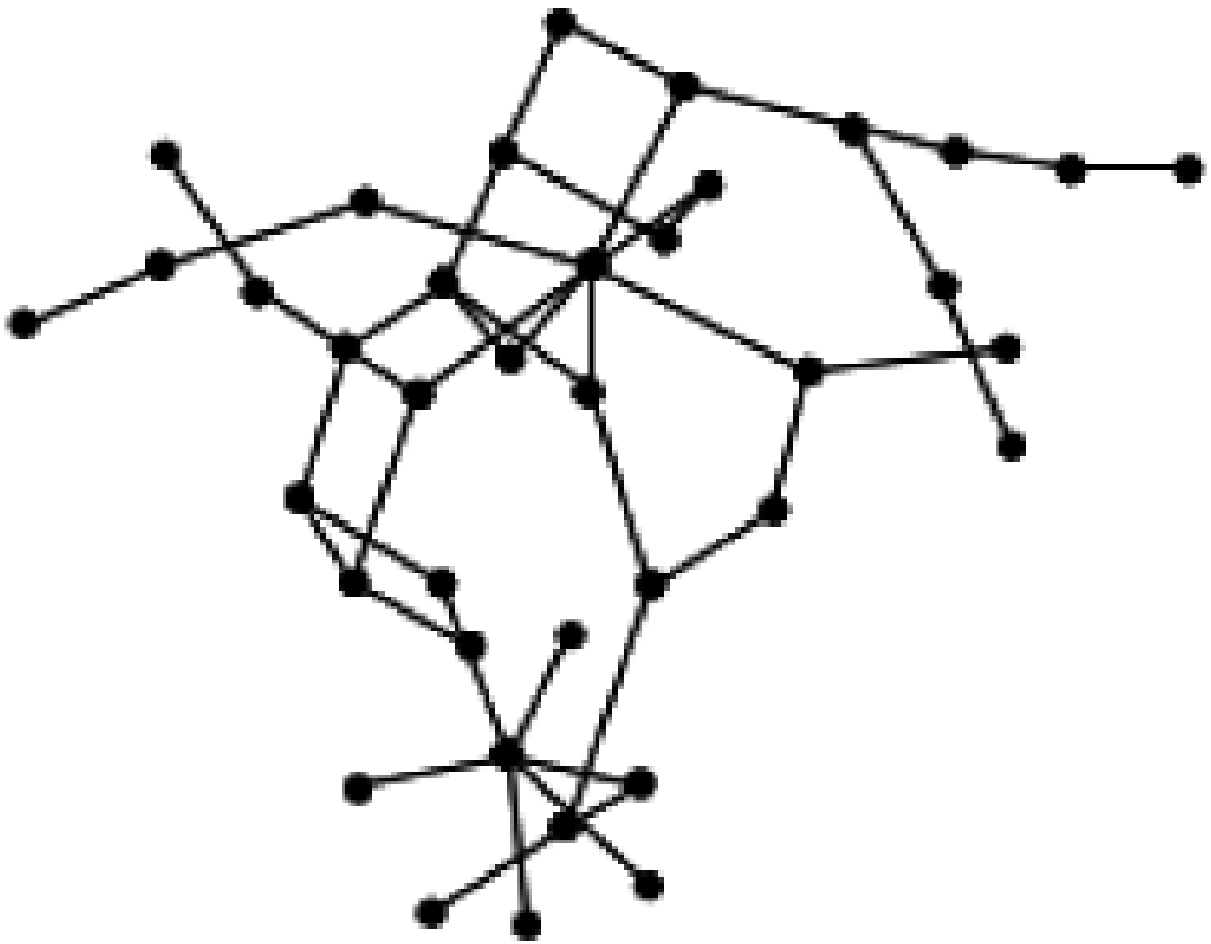
- Ware Purchase Colpoys McGill, IV'02

  42 vertices, $\approx 50 - 60$ edges



# Prior experiments

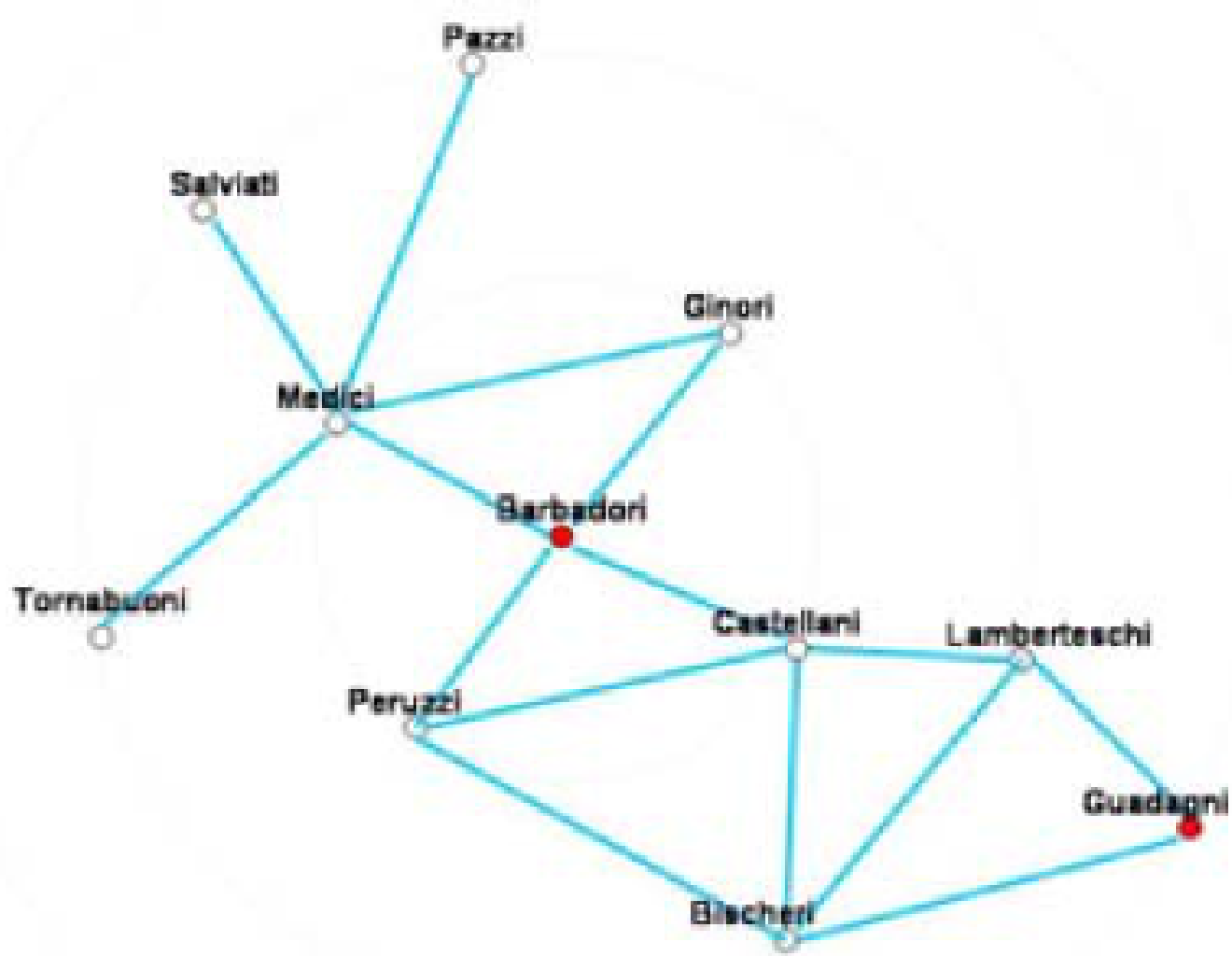
- Huang Huang, AI'14 Huang Eades Hong, VLC'14
  10 – 40 vertices



# Prior experiments

- Huang Eades, APVIS'05 Körner, ACP'11

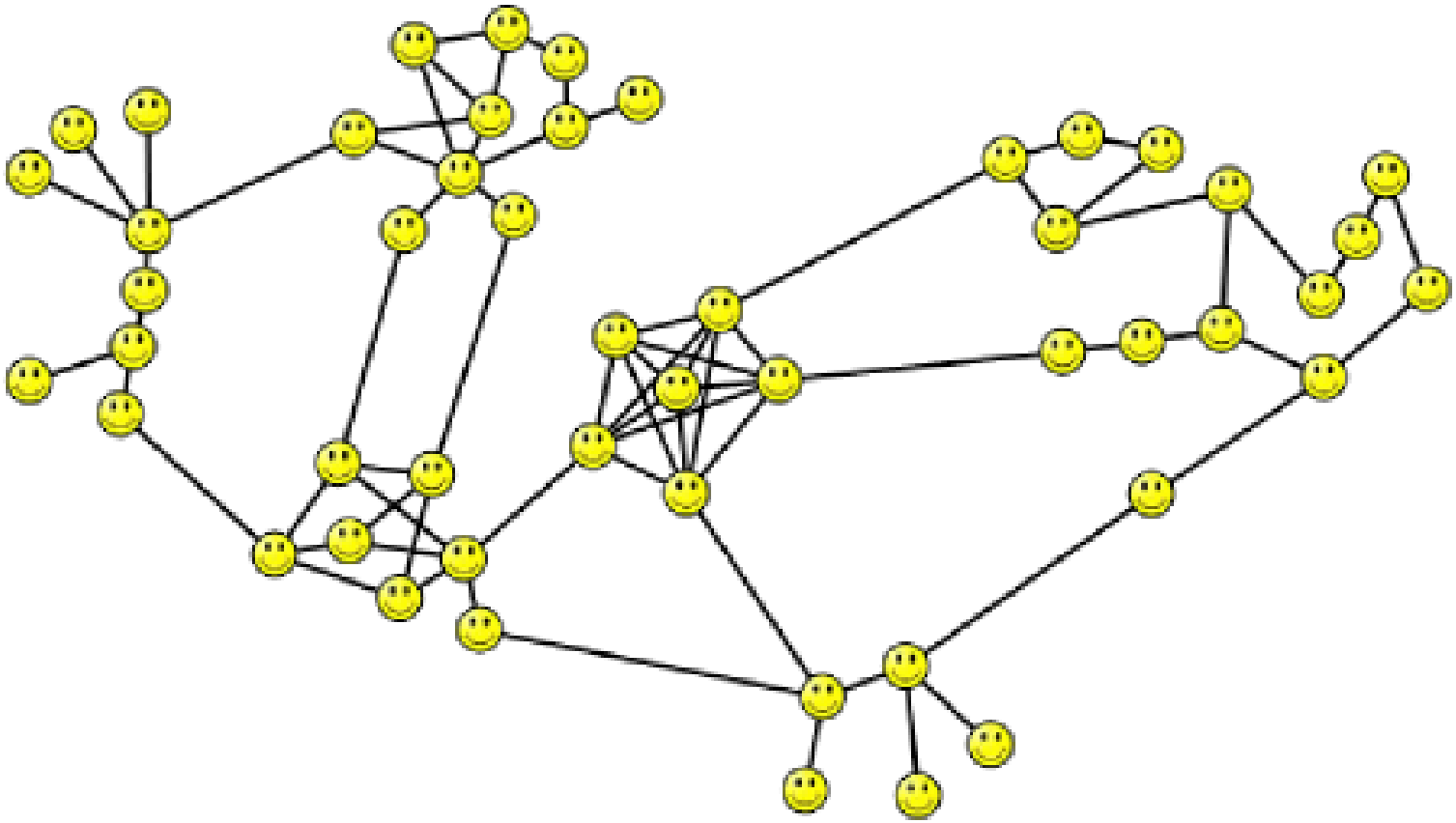
  $9 - 14$ vertices



# Prior experiments

- Dwyer Lee Fisher Quinn Isenberg Robertson North, TVCG'09

  50 vertices, 75 edges



# Prior experiments

- Dwyer Lee Fisher Quinn Isenberg Robertson North, TVCG'09
  - 50 vertices, 75 edges

**Observation 2**

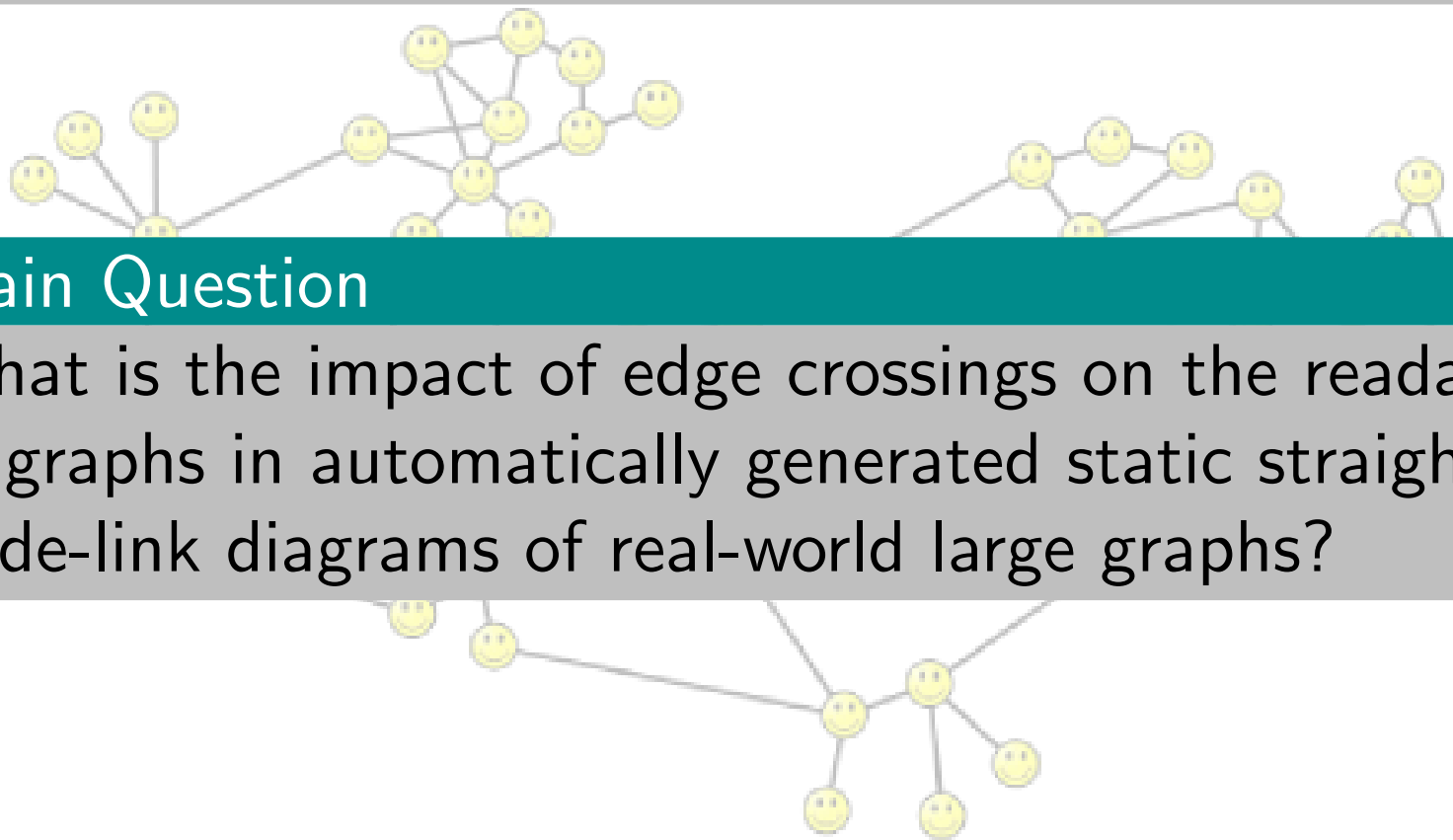
Real-world graphs tend to be large, dense, and non-planar

# Prior experiments

- Dwyer Lee Fisher Quinn Isenberg Robertson North, TVCG'09
  - 50 vertices, 75 edges

**Observation 2**

Real-world graphs tend to be large, dense, and non-planar



**Main Question**

What is the impact of edge crossings on the readability of graphs in automatically generated static straight-line node-link diagrams of real-world large graphs?

# Experiment

- Dataset
- Visualization
- Tasks
- Participants and Apparatus
- Procedure

# Dataset

| graph | $\|V\|$ | $\|E\|$ | density |
|---|---|---|---|
| GD | 506 | 1380 | 2.73 |
| Recipes | 381 | 2171 | 5.70 |

# Dataset

| graph | $\|V\|$ | $\|E\|$ | density |
|---|---|---|---|
| GD | 506 | 1380 | 2.73 |
| Recipes | 381 | 2171 | 5.70 |

- The co-authorship graph for the Int. Symp. on Graph Drawing, 1994-2007. The vertices represent the authors and an edge is between two vertices if the authors published a paper together
- Recipes contain 381 unique cooking ingredients extracted from 56,498 cooking recipes. Edges are created based on co-occurrence of the ingredients in the recipes

Ahn et al., NPG'11

# Dataset

| graph | $\|V\|$ | $\|E\|$ | density |
|---|---|---|---|
| GD | 506 | 1380 | 2.73 |
| Recipes | 381 | 2171 | 5.70 |

- The co-authorship graph for the Int. Symp. on Graph Drawing, 1994-2007. The vertices represent the authors and an edge is between two vertices if the authors published a paper together
- Recipes contain 381 unique cooking ingredients extracted from 56,498 cooking recipes. Edges are created based on co-occurrence of the ingredients in the recipes

Ahn et al., NPG'11

- randomly sampled subgraphs with 40 (small) and 120 (large) vertices, and densities 1.5 (sparse) and 2.5 (dense)

# Visualization

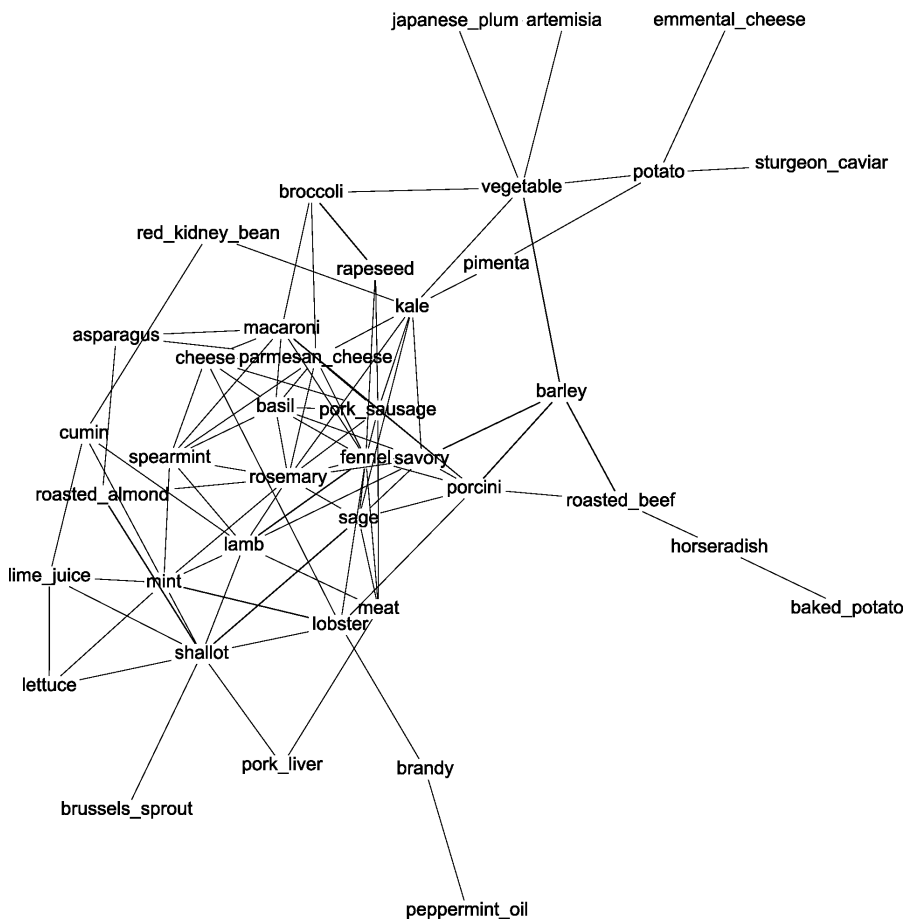
- FDP (force-directed) and NEATO (multidimensional scaling) tools from GRAPHVIZ

# Visualization

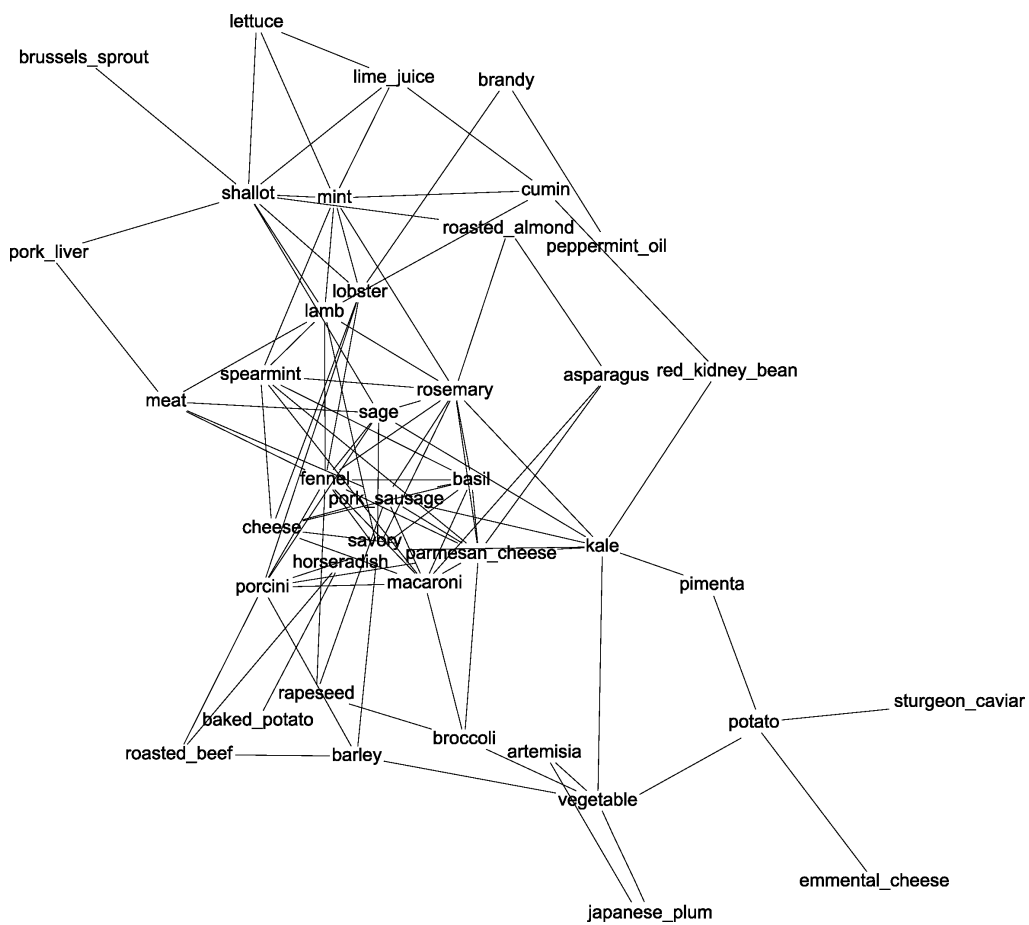
- FDP (force-directed) and NEATO (multidimensional scaling) tools from GRAPHVIZ
- run the algorithms $10,000$ times, varying the initial positions; it gives the drawing with low ($X$) and high ($\approx 2X$) number of crossings

# Visualization

- FDP (force-directed) and NEATO (multidimensional scaling) tools from GRAPHVIZ
- run the algorithms 10,000 times, varying the initial positions; it gives the drawing with low ($X$) and high ($\approx 2X$) number of crossings




139 crossings




259 crossings

# Tasks

**Task 1**: How many edges are in a shortest path between two given nodes?
(connectivity)

**Task 2**: What is the node with the highest degree?
(accessibility)

**Task 3**: What nodes are all adjacent to the given node?
(adjacency)

**Task 4**: Which of the following nodes are adjacent to both given nodes?
(common connections)

- cover a spectrum of the task taxomony for graph visualization by Lee et al., BELIV'06
- standard and commonly encountered in other user evaluations (based on 10+ user studies)

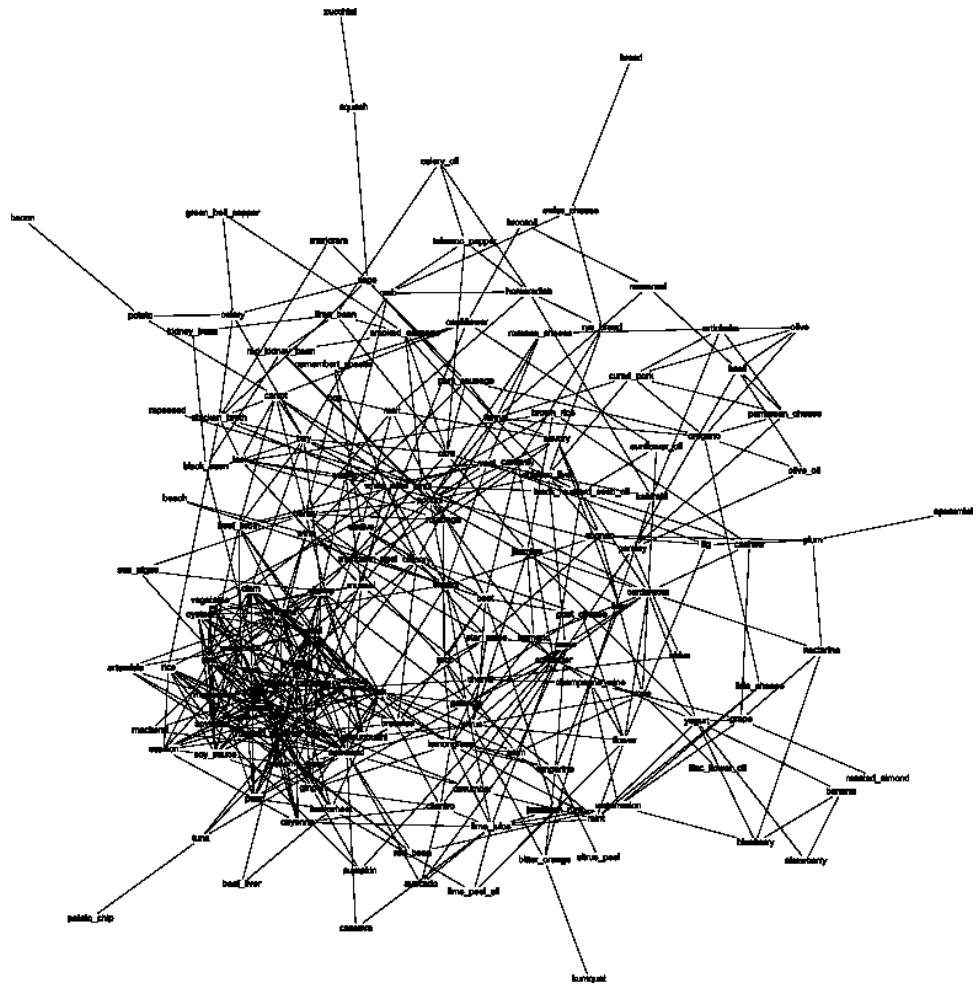
# Procedure: preliminary experiment

- What is a *large* and *dense* graph?

# Procedure: preliminary experiment

- What is a *large* and *dense* graph?

– 150 vertices, 525 edges (density 3.5)

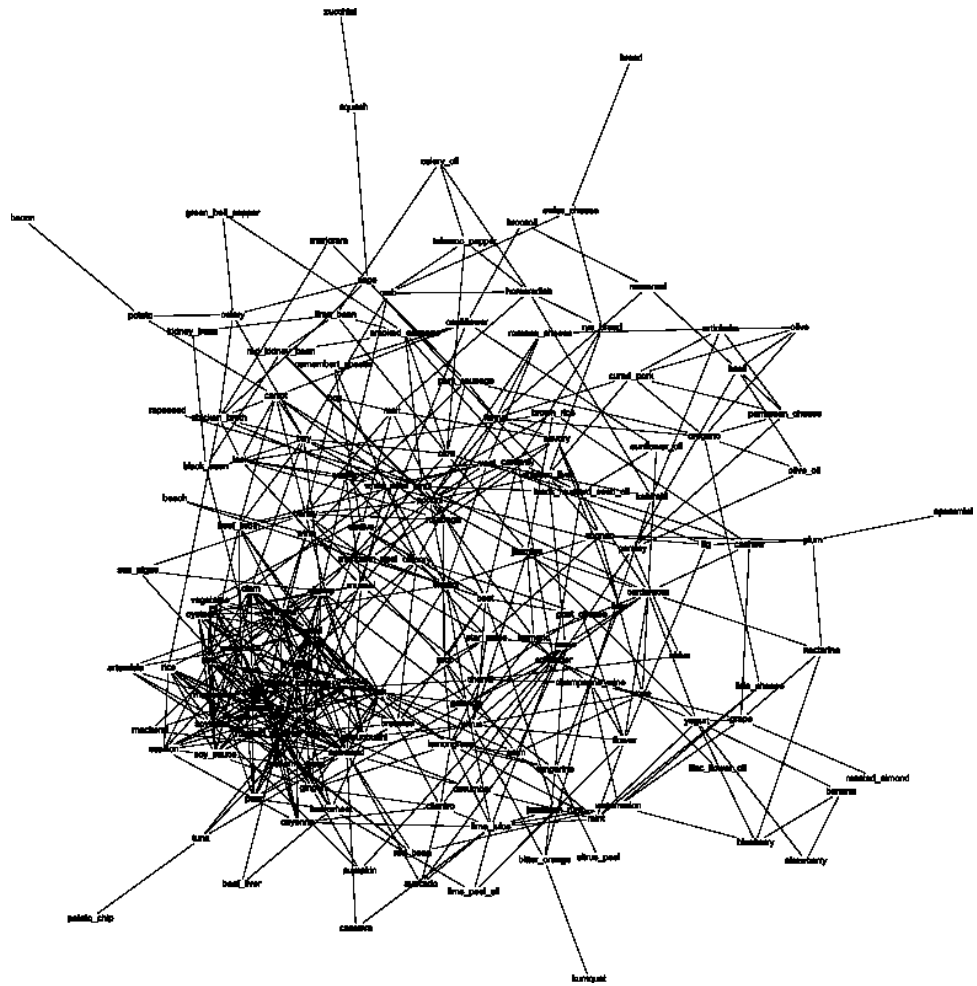


# Procedure: preliminary experiment

- What is a *large* and *dense* graph?

– 150 vertices, 525 edges (density 3.5)

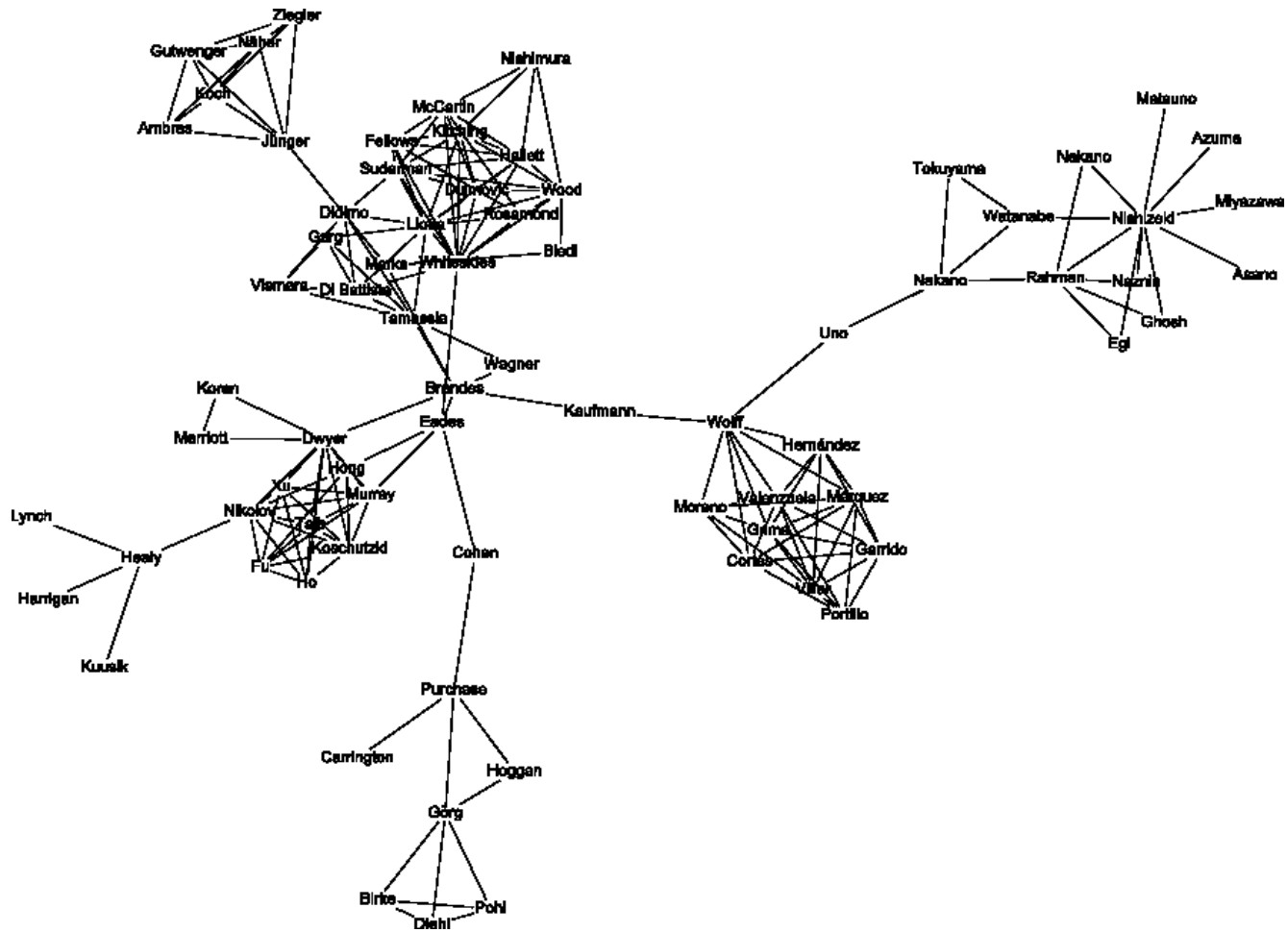
– $\approx$ 180 seconds per Task, $\approx$ 40% accuracy



# Procedure: preliminary experiment

- What is a *large* and *dense* graph?

– 100 vertices, 150 edges (density 1.5)

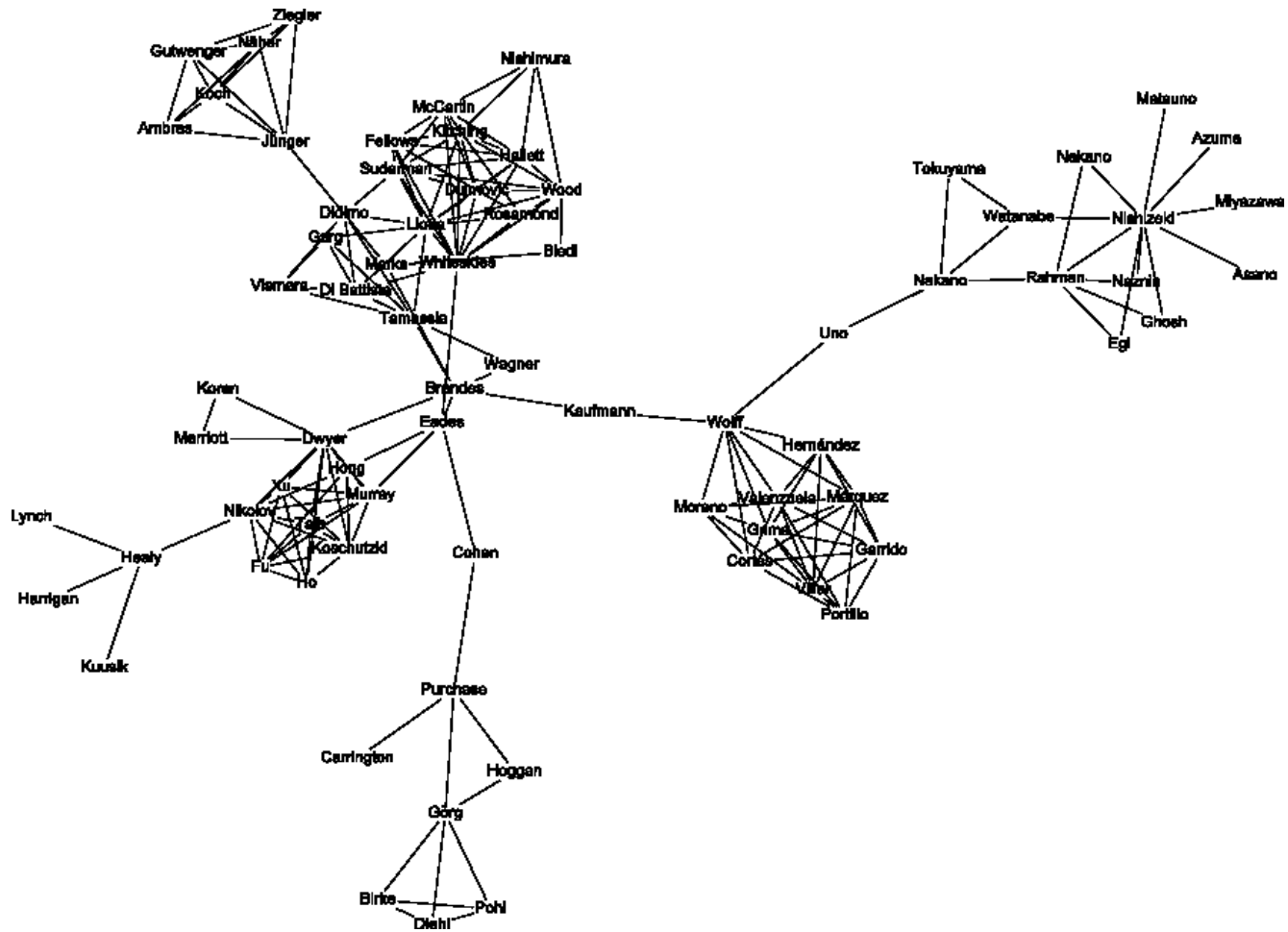


# Procedure: preliminary experiment

- What is a *large* and *dense* graph?

– 100 vertices, 150 edges (density 1.5)

– $\approx$ 60 seconds per Task, $\approx$ 80% accuracy



# Procedure: preliminary experiment

- What is a *large* and *dense* graph?

large $\equiv$ 120 vertices
small $\equiv$ 40 vertices

dense $\equiv$ 3.5 (average 7 neighbors)
sparse $\equiv$ 1.5 (average 3 neighbors)

# Procedure: main experiment

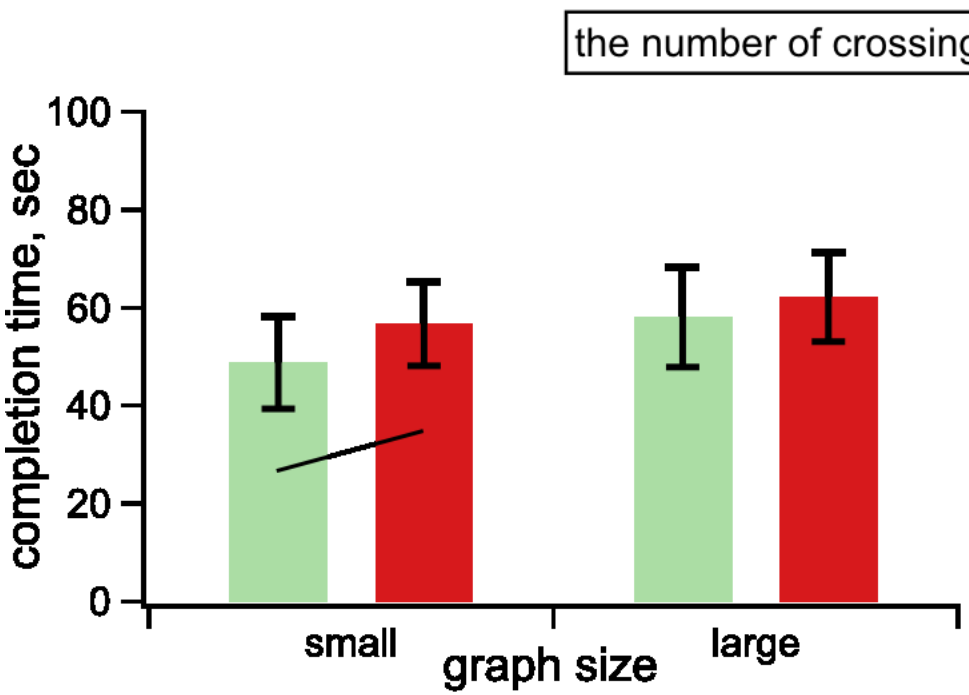
- 64 questions (2 sizes $\times$ 2 number of crossings $\times$ 2 densities $\times$ 2 datasets $\times$ 4 tasks), 26 participants
- online tool with basic interaction (zoom, pan), multiple-choice questions
- record accuracy and completion time

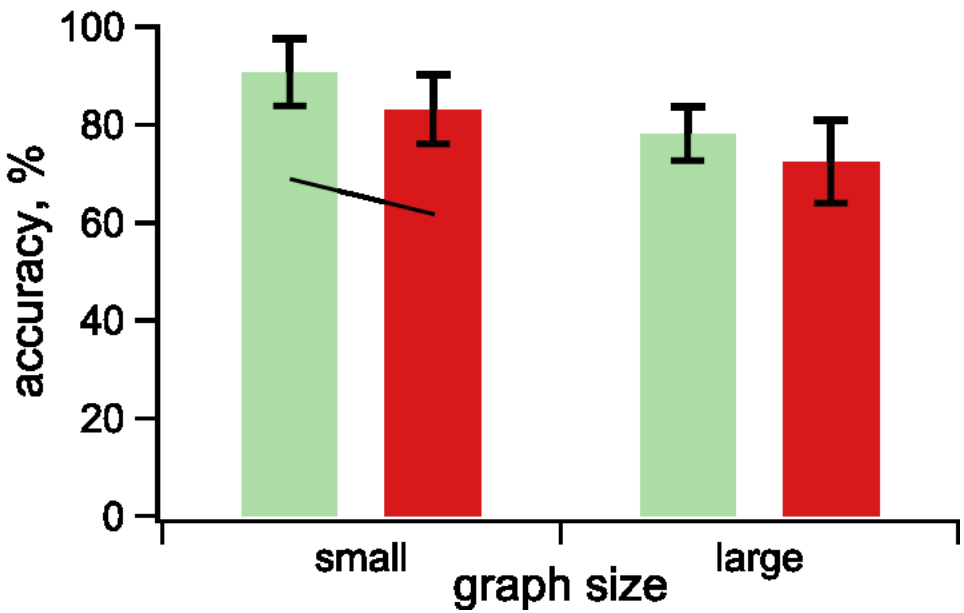
# Hypothesis & Results

**H1** Increasing the number of crossings negatively impacts accuracy and performance time and that impact is *significant* for small graphs but *not significant* for large graphs

# Hypothesis & Results

**H1** Increasing the number of crossings negatively impacts accuracy and performance time and that impact is *significant* for small graphs but *not significant* for large graphs

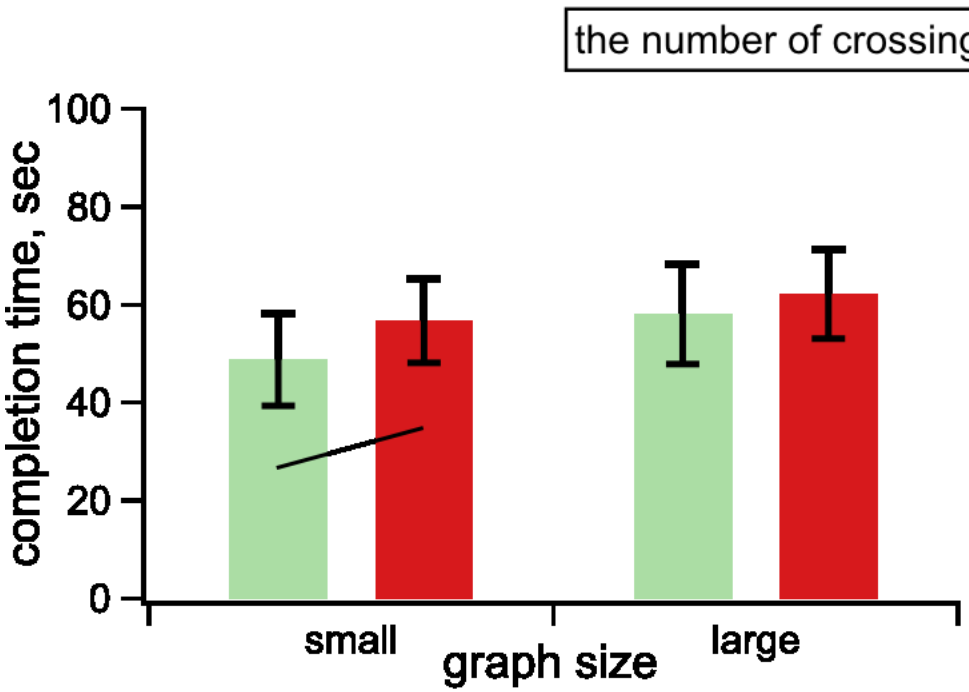


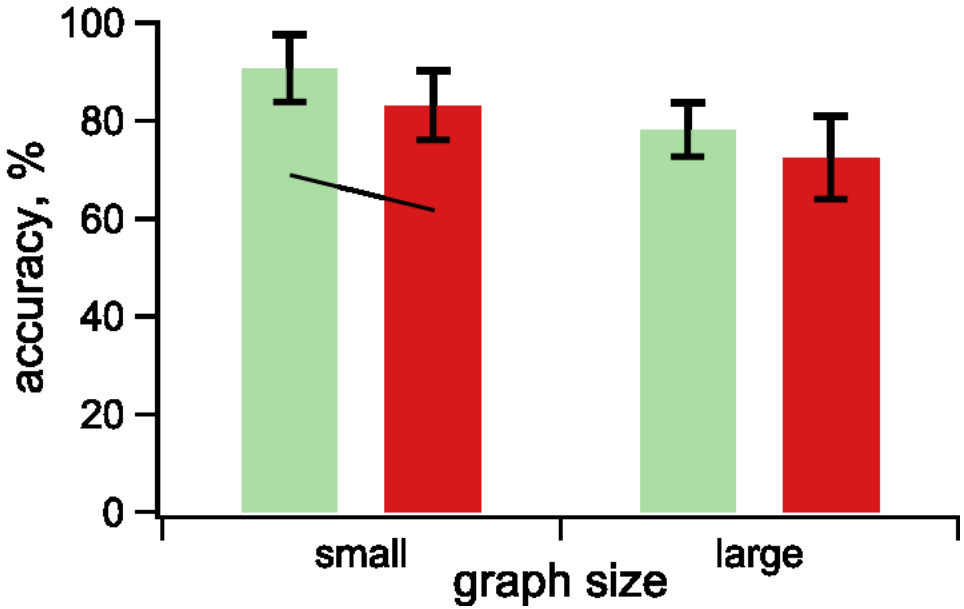



# Hypothesis & Results

**H1** Increasing the number of crossings negatively impacts accuracy and performance time and that impact is *significant* for small graphs but *not significant* for large graphs

Confirmed!



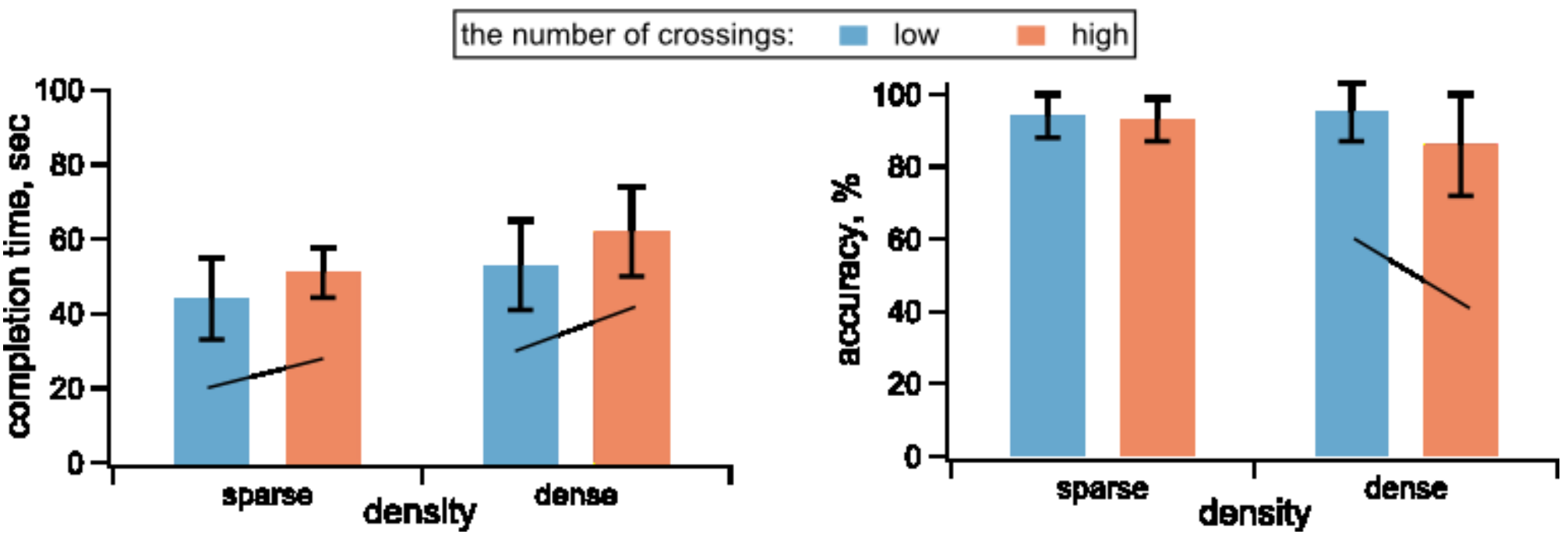



# Hypothesis & Results

**H2** The negative impact of increasing the number of crossings on accuracy and completion time is *significant* for both small sparse and small dense graphs

# Hypothesis & Results

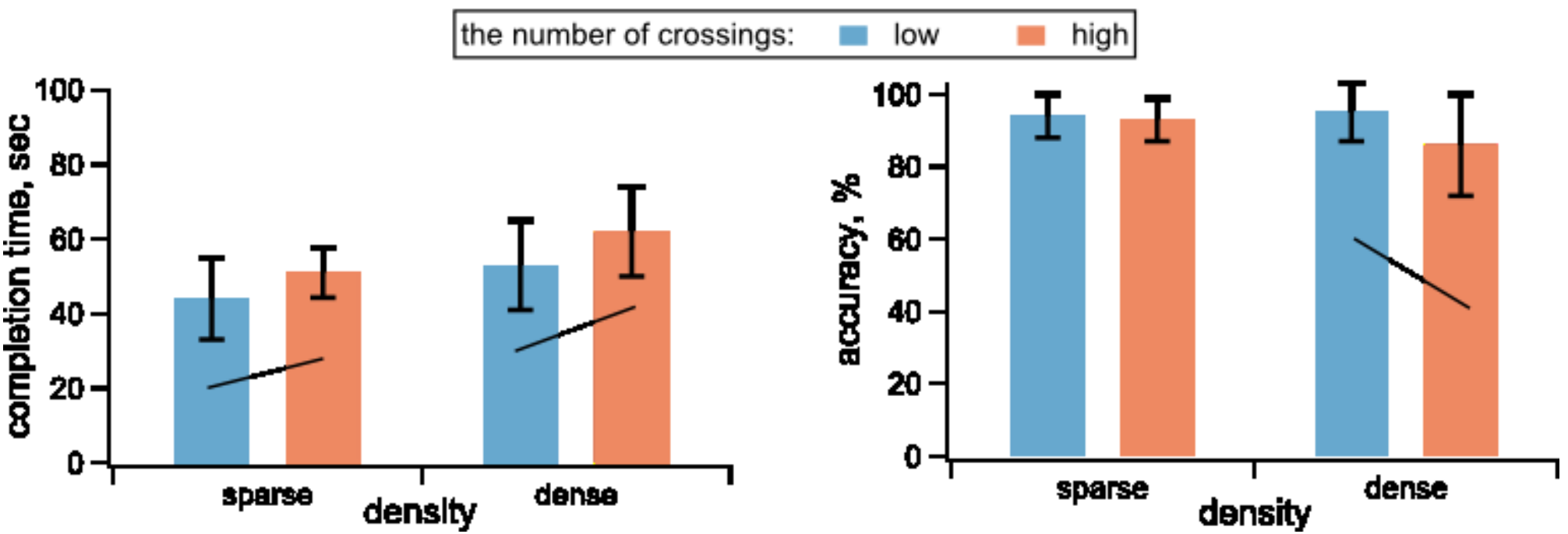
**H2** The negative impact of increasing the number of crossings on accuracy and completion time is *significant* for both small sparse and small dense graphs



# Hypothesis & Results

**H2** The negative impact of increasing the number of crossings on accuracy and completion time is *significant* for both small sparse and small dense graphs

Partially confirmed



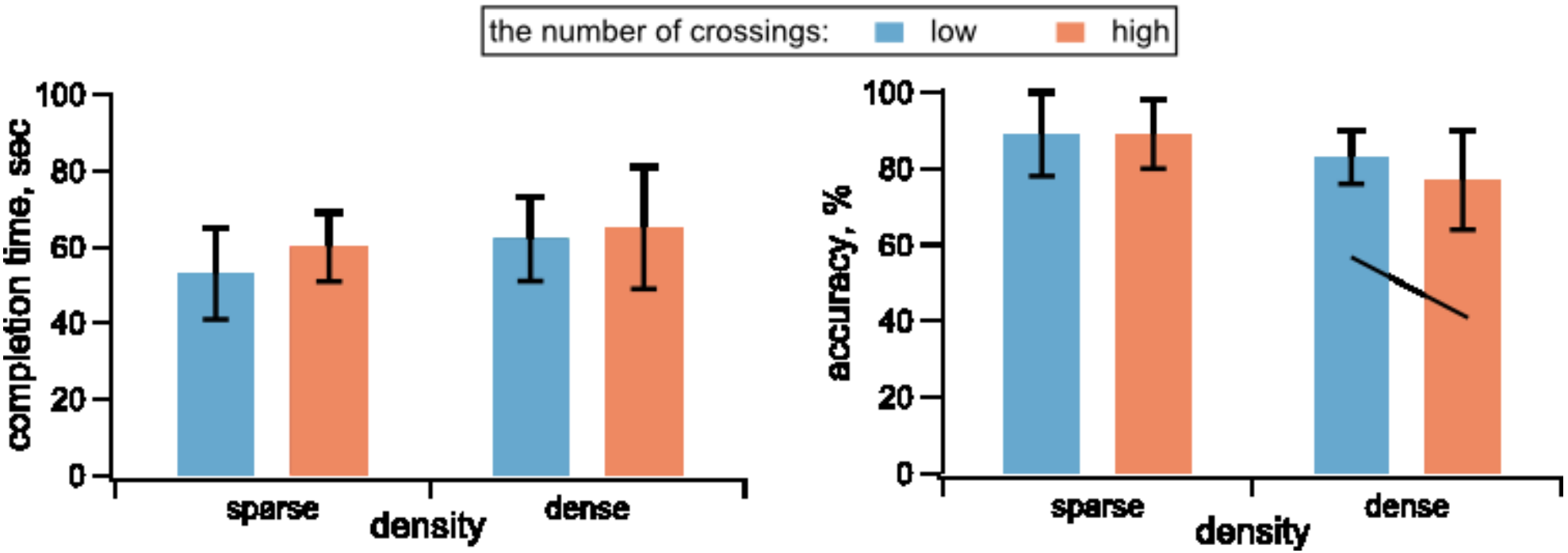
# Hypothesis & Results

**H3** The negative impact of increasing the number of crossings on accuracy and completion time is *not significant* for both large sparse and large dense graphs

# Hypothesis & Results

**H3** The negative impact of increasing the number of crossings on accuracy and completion time is *not significant* for both large sparse and large dense graphs

Partially confirmed



# So, how to draw large graphs?

# So, how to draw large graphs?

Many existing algorithms try to optimize "visual energy" of a layout known as *stress*

# So, how to draw large graphs?

Many existing algorithms try to optimize "visual energy" of a layout known as *stress*

Def.: **Stress** is the variance of edge lengths in the drawing.
For a graph $G = (V, E)$ with $p_v$ being the position of vertex $v \in V$, stress is defined as

$$\sum_{u,v \in V} \frac{1}{d_{uv}^2} (||p_u - p_v|| - d_{uv})^2,$$

where $d_{uv}$ is the ideal distance between vertices $u$ and $v$.

*Lower values of stress correspond to a better layout*

Kamada Kawai, IPL'89
Eades, CN'84

# Stress vs Other Aesthetic Criteria

Question: Does minimizing stress also (possibly indirectly) optimize some of the standard aesthetic criteria?

# Stress vs Other Aesthetic Criteria

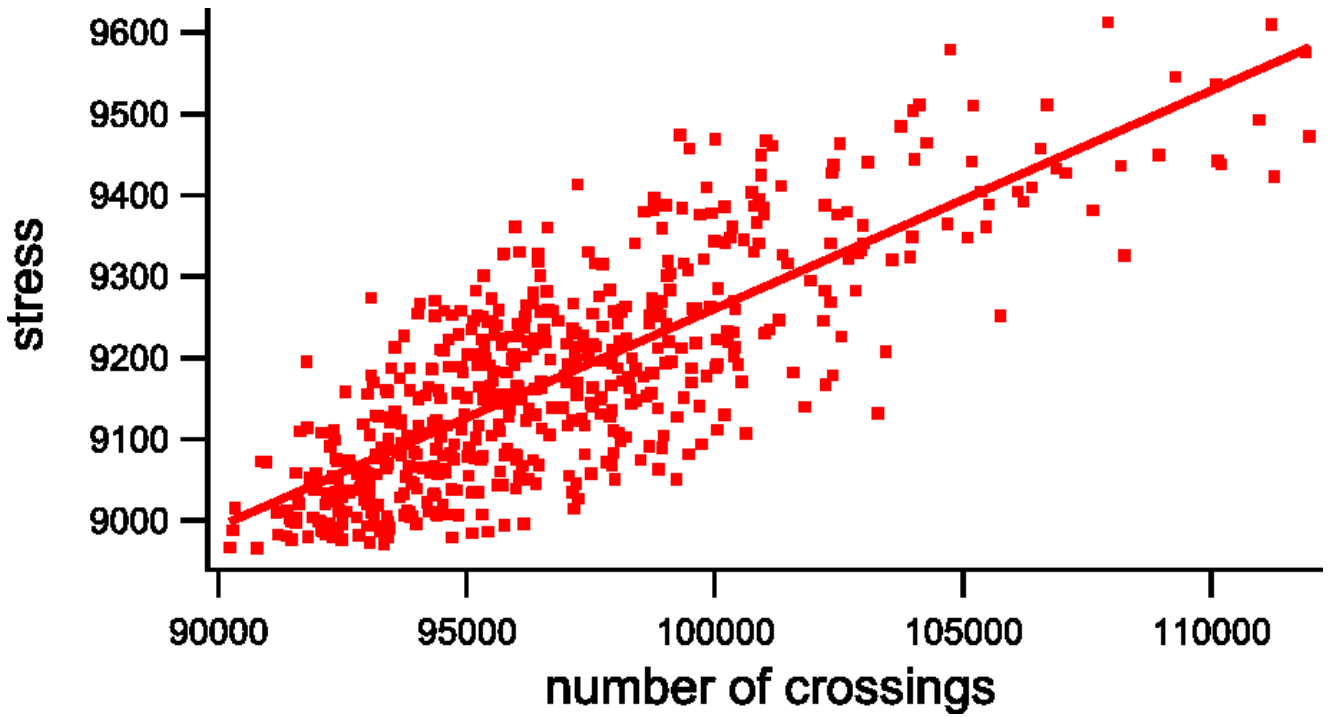
Question: Does minimizing stress also (possibly indirectly) optimize some of the standard aesthetic criteria?

Methodology: Qualitatively analyze layouts produced by force-directed algorithms, with respect to **stress**, **number of crossings**, and **crossing angles**

# Stress vs Other Aesthetic Criteria

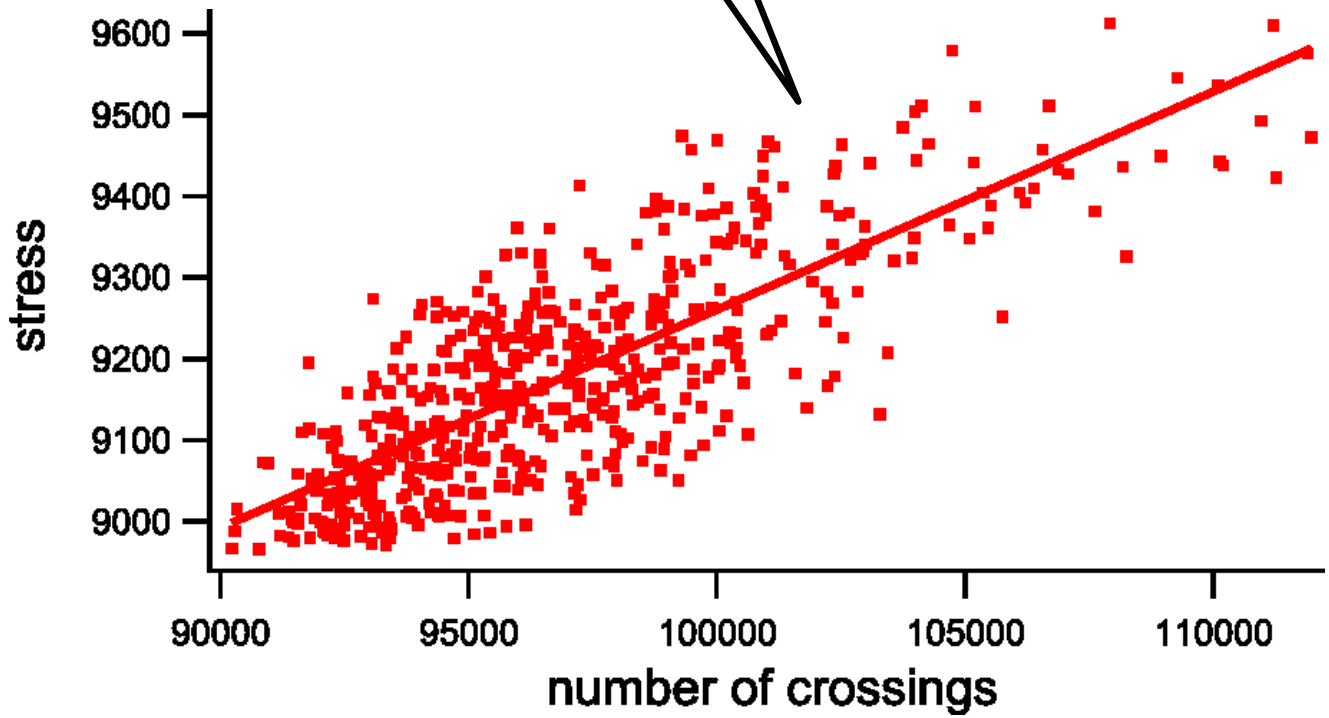
Question: Does minimizing stress also (possibly indirectly) optimize some of the standard aesthetic criteria?

Methodology: Qualitatively analyze layouts produced by force-directed algorithms, with respect to **stress**, **number of crossings**, and **crossing angles**

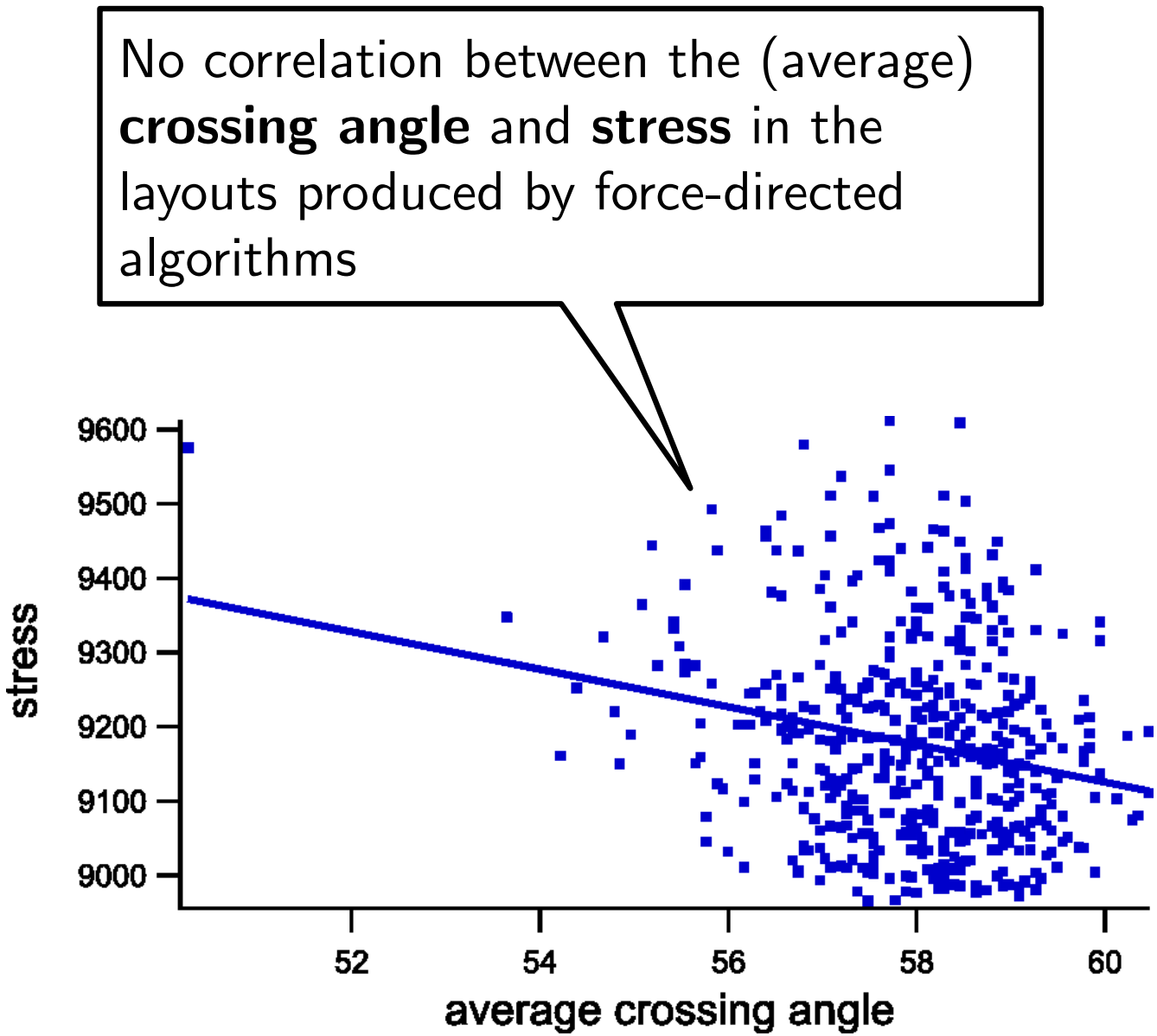


# Stress vs Other Aesthetic Criteria

There is a moderate correlation between the **number of crossings** and **stress** in the layouts produced by force-directed algorithms



# Stress vs Other Aesthetic Criteria



# Conclusions

- minimizing the number of edge crossings in large graphs does not have as significant an impact as in small graphs

# Conclusions

- minimizing the number of edge crossings in large graphs does not have as significant an impact as in small graphs
- traditional energy-based methods might already result in some reduction in crossings

# Conclusions

- minimizing the number of edge crossings in large graphs does not have as significant an impact as in small graphs
- traditional energy-based methods might already result in some reduction in crossings
- our results should be interpreted in the context of the specified graphs, sizes, densities, and tasks

see http://sites.google.com/site/gdpaper2014

# What is next?

- include graphs with more than 120 vertices and density greater than 2.5
- interactive visualizations, non straight-line, non-static drawings
- which other aesthetic criteria are important for large graphs?

# What is next?

- include graphs with more than 120 vertices and density greater than 2.5
- interactive visualizations, non straight-line, non-static drawings
- which other aesthetic criteria are important for large graphs?
- how to draw large graphs?

# What is next?

- include graphs with more than 120 vertices and density greater than 2.5
- interactive visualizations, non straight-line, non-static drawings
- which other aesthetic criteria are important for large graphs?
- how to draw large graphs?

Thank you!