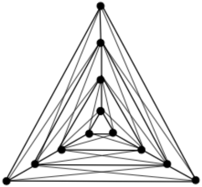
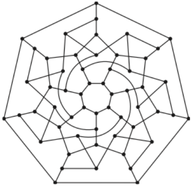
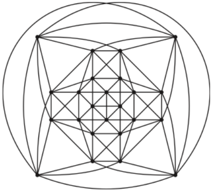
Force-Directed Layout of Node-Link Diagrams

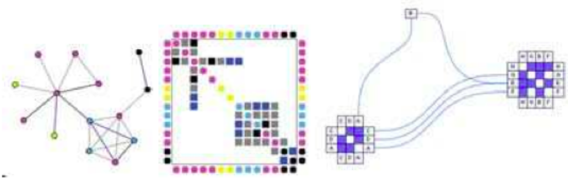
Stephen Kobourov
University of Arizona

Big Graph Drawing: Metrics and Methods
Shonan Seminar, January 12, 2015

Lay Out What?

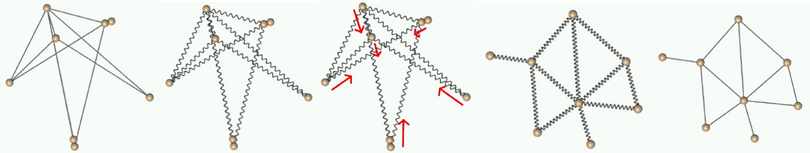
- Plain vanilla graphs
  - nodes or vertices
  - edges or links
  - vertices are unweighted and unlabeled
  - edges are unweighted, unlabeled and undirected
- Node-link diagram visualization
  - force-directed algorithms
  - spring embedders
  - energy-based layouts
  - multi-dimensional scaling (MDS)



(a) Node-link diagram (b) Adjacency matrix diagram (c) Combination

Force-Directed Methods

- Given a graph $G = (V, E)$
  - place vertices as points in $\mathcal{R}^d$
  - route edges in $\mathcal{R}^d$
- Force directed methods define an energy function on layouts
  - based on attractive/repulsive forces (Fruchterman-Reingold)
  - based on graph distances (Kamada-Kawai)
- Energy model
  - iterative improvement
  - minimal energy $\Rightarrow$ good layout



Force-Directed Methods, cont.

- Fruchterman-Reingold: balance of attraction and repulsion

$$F(v) = F_r(v) + F_a(v)$$

$$F_r(v) = \sum_{\forall u \in V} \frac{\kappa^2}{\|pos[u] - pos[v]\|^2}(pos[u] - pos[v])$$

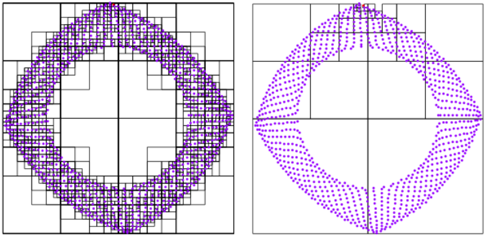
$$F_a(v) = \sum_{u \in Adj(v)} \frac{\|pos[u] - pos[v]\|^2}{\kappa^2}(pos[u] - pos[v])$$

- $\kappa = \sqrt{A_{frame}/|E|}$, ideal edge length
- Kamada-Kawai: match Euclidean distance to graph distance

$$F(v) = \sum_{u \in V} \left(\frac{\|pos[u] - pos[v]\|^2}{(\kappa \times dist_G(u, v))^2} - 1 \right) (pos[u] - pos[v])$$

Faster Force Calculations

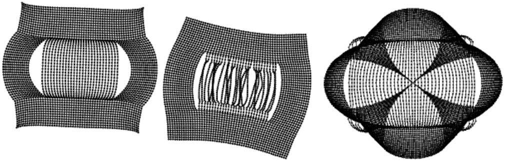
- The $O(n^2)$ repulsive calculation can be approximated
- Barnes-Hut n-body simulation can be done in $O(n \log n)$ time
- Quigley and Eadges (FADE) 2000
- Hachul and Jünger (FM3) 2004
- Hu (sfdp) 2008



Multi-Level Methods

First consider an abstraction, disregarding some of the graph's fine details. This abstraction is then drawn, yielding a "rough" layout in which only the general structure is revealed. Then the details are added and the layout is corrected.

- Hadany and Harel 2000: edge contractions
- Harel and Koren 2000: k-centers
- Walshaw 2000: maximal independent set
- Gajer et al. 2000 (GRIP): vertex filtrations, high dimensional embedding

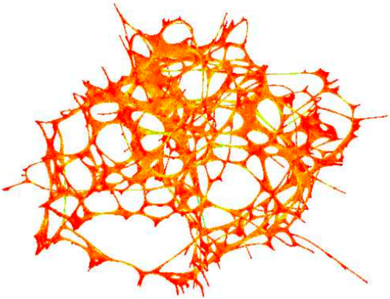


(a) (b) (c)

Fig. 9. (a,b) 55x55 (3025-vertex) sparse grid: (a) beautification considered larger than normal neighborhoods, of radius 35 (b) beautification considered usual neighborhoods of radius 7; (c) 80x80 (6400-vertex) sparse grid with each two opposite corners connected

AT&T: GraphViz and sfdp

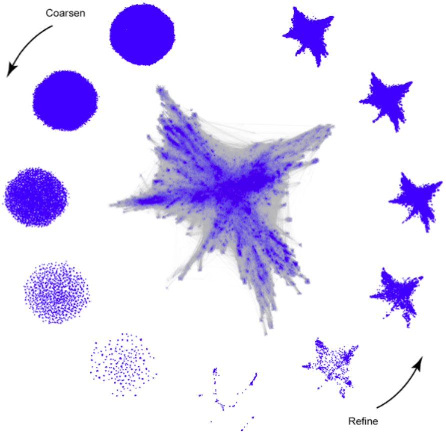
- Multi-scale: edge contractions or maximal independent set
- Barnes-Hut *n*-body simulation with quad-trees



boneS10. $|V| = 914898, |E| = 27276762$

Sandia: VxOrd and OpenOrd

- scalable version of Fruchterman-Reingold with simulated annealing
- recently multi-level and parallelized



Stress Model

The stress model assumes that there are springs connecting all pairs of vertices of the graph, with the ideal spring length equal to the predefined edge length. The energy of this spring system is

$$E = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{2} k_{i,j} (|p_i - p_j| - l_{i,j})^2,$$

- Kamada-Kawai is an example
- neato in GraphViz
- optimize in force-directed fashion
- optimize using stress majorization

MDS Strain Model

Classical Multidimensional Scaling (MDS) tries to fit the inner product of positions, instead of the distance between points.

- LandmarkMDS places subset of nodes by classical MDS; other node positions based on distances to already placed nodes
- PivotMDS uses Singular Value Decomposition
- neato in GraphViz

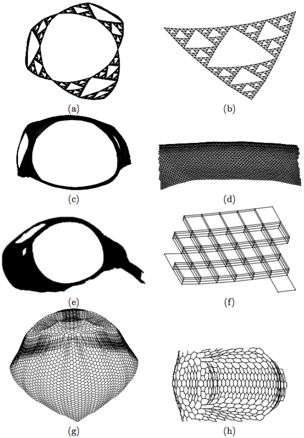
(a) Pivot MDS (b) sparse stress minim. (c) original

Fig. 5. Drawings for a large graph representing the street network in Germany (4 044 153 nodes, 9 564 235 edges, diameter 1 059)

High Dimensional Embedding

The high-dimensional embedding (HDE) algorithm computes vertex positions in k-dimensional space and projects them to a suitably chosen 2D or 3D space.

- k-centers are chosen as in LandmarkMDS
- the graph distances from each vertex to the k-centers form a k-dimensional coordinate system
- find projection to 2D space that minimizes correlations
- very fast $O(n)$ time



Comparative Analysis: Brandenburg et al. 1996

- Algorithms: KK, FR, DH, GEM, TU
- Measures: running time, ratio of longest to shortest edge, number of crossings
- Dataset: 59 graphs from the papers describing the algorithms + K_4 to K_{24}
- Graph sizes: under 100 vertices
- Conclusions: similar results, no clear winner

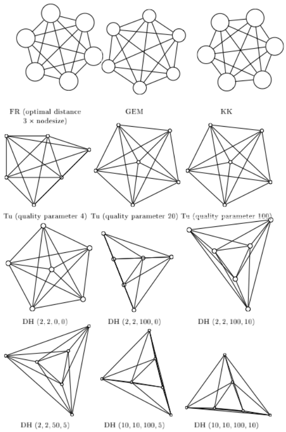
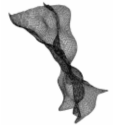


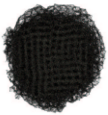
Fig. 5. Several drawings of the K_5

Comparative Analysis: Hachul and Jünger 2005

- Algorithms: FR*, KK*, GRIP, HDE, FM3, ACE
- Measures: running time, aesthetics of layouts
- Dataset: 30 graphs from earlier papers and synthetic ones
- Graph sizes: 930-143,437 vertices and 970-184,532 edges
- Conclusions:
  - multi-scale algorithms work well
  - algebraic methods work fast
  - FM3 slower but better



(c) GRIP

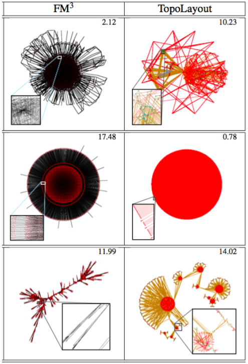


(f) GVA

(i) GVA

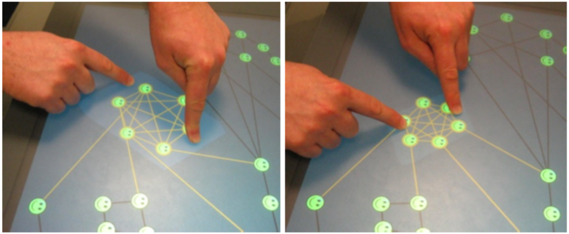
Comparative Analysis: Archambault et al. 2007

- Algorithms: GRIP, HDE, FM3, ACE, TopoLayout
- Measures: running time, structure of layouts
- Dataset: 11 graphs (from earlier papers + tough real world graphs)
- Graph sizes: 1,104-77,251 vertices and 3,232-191,659 edges
- Conclusions:
  - ACE, HDE: only for mesh-like graphs
  - FM3 and TopoLayout are better

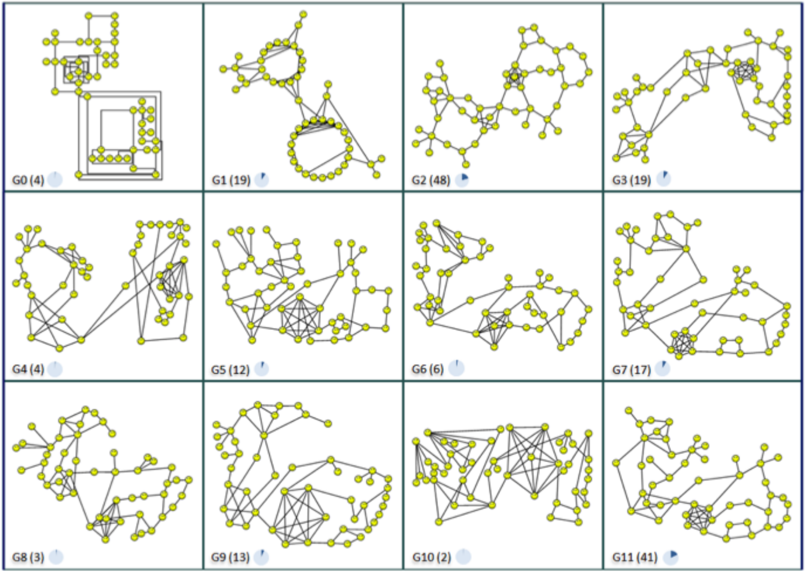


Comparative Analysis: Dwyer et al. 2009

- User-generated layouts
- Automatic layout (orthogonal, circular, force-directed)
- 32/194 users: draw a graph with 50 nodes and 74/77 links
- Tasks: find a K_6, P_4, a cut node, and leaf nodes
- Measures: task accuracy and time, stress, edge crossings
- Conclusions
    - user-generated better than orthogonal/circular aut. layout
    - the most popular layout overall: aut. force-directed layout

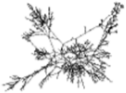


Comparative Analysis: Dwyer et al. 2009



Comparative Analysis: Brandes and Pich 2009

- Algorithms: FM3, HDE, GRIP, MDS (strain, stress)
- Dataset: 10 graphs with 500-3,000 vertices and 700-15,000 edges
- Measures: stress
- Conclusions
  - graph-theoretic distance models work well
  - the very fast MDS methods can hide structure
  - fast MDS as initial layout improves the results

Comparative Analysis: Brandes and Pich 2009

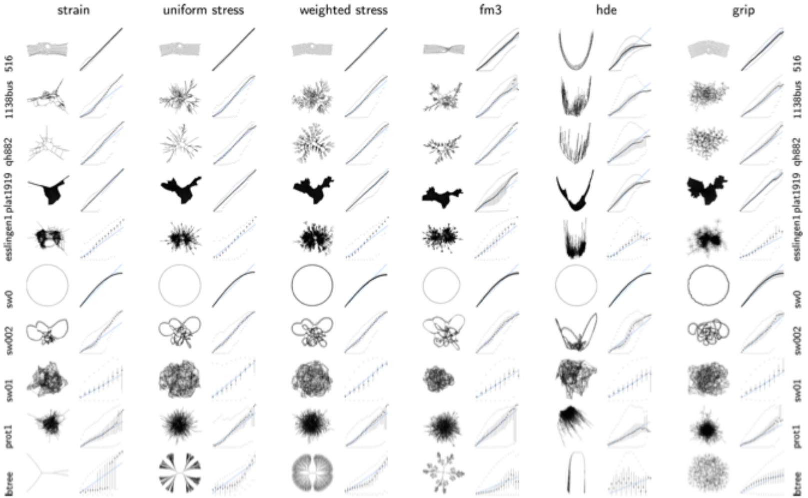
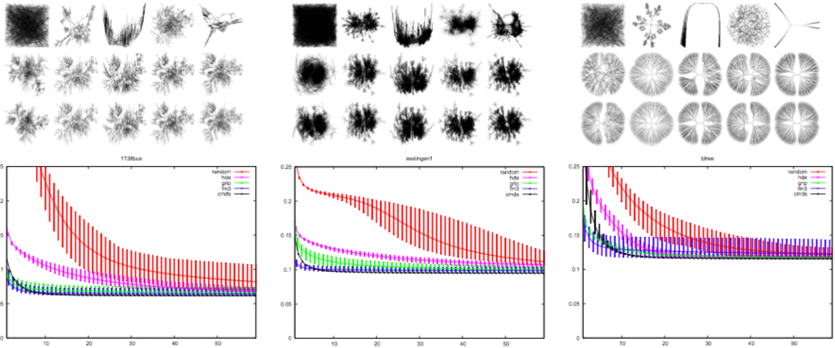


Fig. 2. Drawings the test graphs, and quartile plots of d_{ij} (abscissa) vs. $\|x_i - x_j\|$. Large dots indicate the median, small dots minimum and maximum, and black lines the range of the two middle quartiles (25–75 per cent). The thin blue line with slope 1 is a visual aid.

Comparative Analysis: Brandes and Pich 2009



(a) 1138bus (b) esslingen1 (c) btree

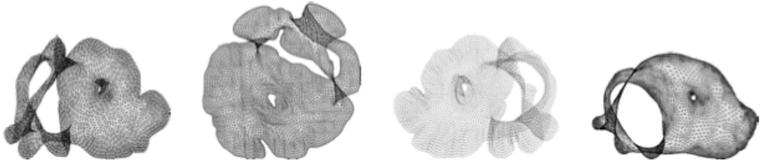
Fig. 3. Upper row: The majorization process with different initializations random, fm3, hde, grip, cmds after $0, 30, 60$ iterations. Lower row: Number of iterations vs. stress. The bars indicate the range of values, the dots the median value, in 25 runs.

Comparative Analysis: Bartel et al. 2010

- Algorithms: multi-scale methods, different coarsening/refining
- Layout: EAD, FR, KK, GEM, NMM, FME, DH
- Dataset: 43 graphs, 34-16,000 vertices and 78-48,000 edges
- Measures: running time, edge lengths, crossings
- Conclusions
  - no clear winner
  - EAD, NMM, FME consistently good
  - KK is best for edge lengths and structure, but too costly
  - some coarsening steps consistently good
  - graph-theoretic distance models work well
  - the very fast MDS methods can hide structure
  - fast MDS as initial layout improves the results

(a) KK (b) NMM (c) FME

Comparative Analysis: Bartel et al. 2010

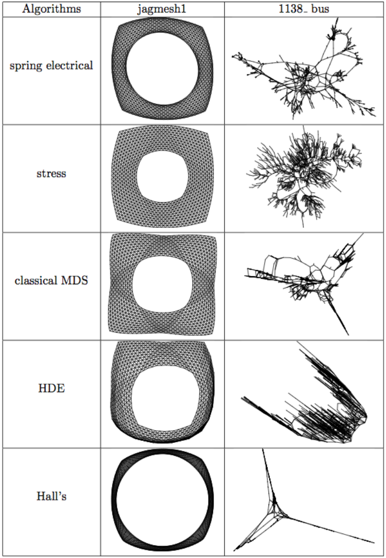


(a) BP,MM,KKFR (b) MP,SM,FME (c) KKS (d) EADS

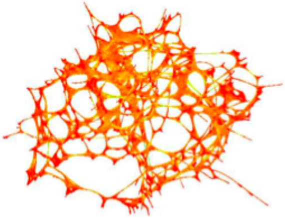
Fig. 2: Layout of the planar graph *3elt* computed with slowest(BP,MM,KKFR, 15 levels, 91s) and fastest overall combination (MP,SM,FME, 5 levels, 2,27s) compared to results of fastest (EADS, 2,77s) and slowest (KKS, 158s) single level methods .

Survey: Hu 2011

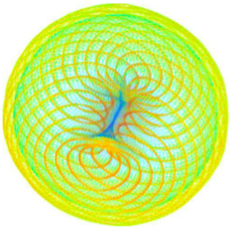
- Algorithms: FR, HDE, MDS (strain, stress)
- Dataset: 2,272 graphs (UF sparse matrices) with up to 27M nodes
- Measures: stress and appearance
- Conclusions: HDE, PivotMDS, LandmarkMDS have trouble with sparse graphs



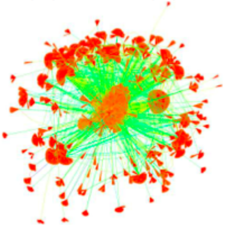
Survey: Hu 2011



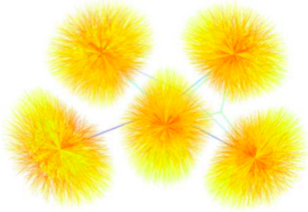
boneS10. $|V| = 914898, |E| = 27276762$



cvxbqp1. $|V| = 40000, |E| = 120000$.



connectus. $|V| = 392366, |E| = 1124842$

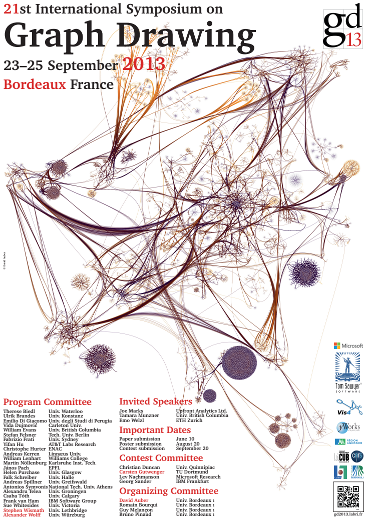


aircraft. $|V| = 11271, |E| = 20267$.

Figure 9: Drawing of some graphs from the University of Florida Sparse Matrix Collection [10]. More drawings can be found at [33].

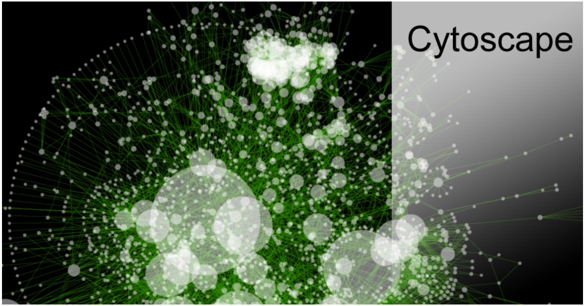
GD Libraries

- GraphViz
- OGDF
- MSAGL
- VTK
- Prefuse
- Tulip
- …



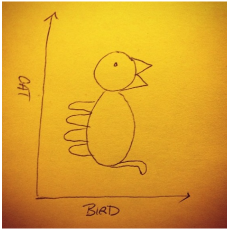
Drawing Software

- TouchGraph
- NodeXL
- Pajek
- Cytoscape
- TopFish
- Gephi
- yEd
- ...



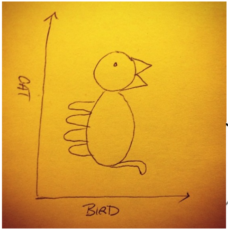
When is drawing good?

- "you will know it when you see it"



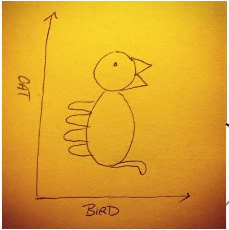
When is drawing good?

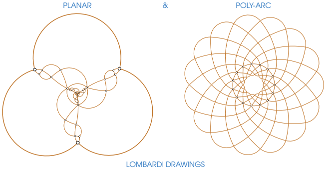
- “you will know it when you see it”
- quantitative measures (e.g., stress, precision/recall)



When is drawing good?

- "you will know it when you see it"
- quantitative measures (e.g., stress, precision/recall)
- qualitative measures (e.g., "aesthetically pleasing")





You will know it when you see it

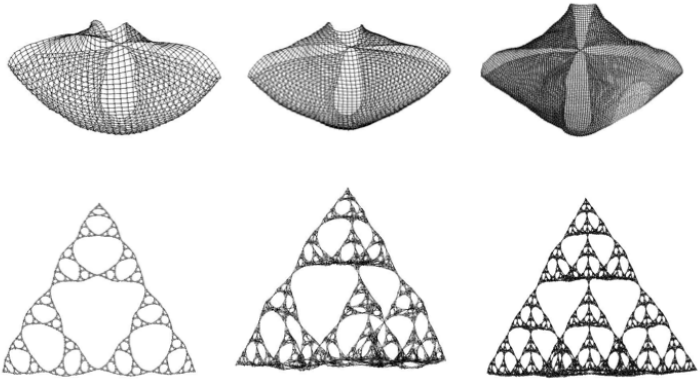
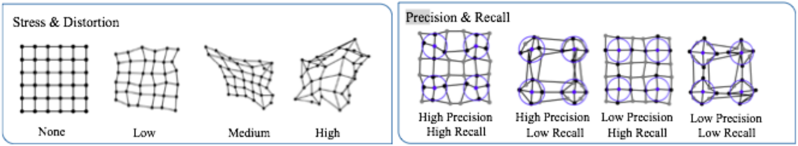


Figure 5.2 Drawings from GRIP. First row: knotted meshes of 1600, 2500, and 10000 vertices. Second row: Sierpinski graphs of order 7 (1,095 vertices), order 6 (2,050 vertices), 3D Sierpinski of order 7 (8,194 vertices) [GK02].

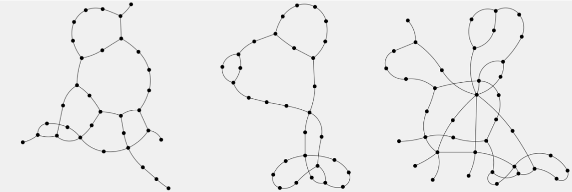
Quantitative measures

- Stress: the classic evaluation function used for MDS
- Precision/Recall measure:
  - false positives (low similarity but small Euclidean distance)
  - false negatives (high similarity but large Euclidean distance)
- Distortion: distances b/n pairs of nodes compared to desired distances



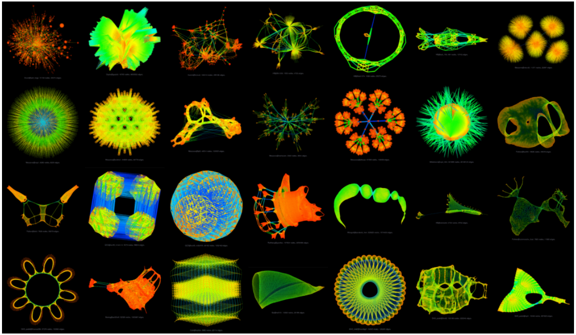
Qualitative measures

- Legibility/readability of the visualization: can one find certain nodes, edges, paths, clusters?
- Utility: which visualizations provide actionable insights and/or prompt more meaningful questions, or are otherwise useful for human decision-making?
- Engagement and enjoyment: beyond "do you like this drawing?"
- Memorability: how well is the shown data remembered?



Datasets

- Rome library: about 11K graphs with 10-100 vertices; 8K are non-planar
- Stanford library (2004): contains MS IM network from 2006 with 240 million nodes and 1.3 billion edges
- Florida Sparse Matrix library: 2.5K graphs with up to 118M nodes and 2B edges



References I

D. Archambault, T. Munzner, and D. Auber.
Topolayout: Multilevel graph layout by topological features.
Visualization and Computer Graphics, IEEE Transactions on, 13(2):305–317, March 2007.

D. Auber, D. Archambault, R. Bourqui, A. Lambert, M. Mathiaut, P. Mary, M. Delest, J. Dubois, and G. Mélançon.
The Tulip 3 framework: A scalable software library for information visualization applications based on relational data.
2012.

G. Bartel, C. Gutwenger, K. Klein, and P. Mutzel.
An experimental evaluation of multilevel layout methods.
In *Proceedings of the 18th International Conference on Graph Drawing*, GD'10, pages 80–91, Berlin, Heidelberg, 2011. Springer-Verlag.

F. Brandenburg, M. Himsolt, and C. Rohrer.
An experimental comparison of force-directed and randomized graph drawing algorithms.
In *Graph Drawing*, pages 76–87, 1996.

U. Brandes and C. Pich.
Eigensolver methods for progressive multidimensional scaling of large data.
In *Proceedings 14th Symposium on Graph Drawing (GD)*, pages 42–53, 2006.

U. Brandes and C. Pich.
An experimental study on distance-based graph drawing.
In *Graph Drawing*, pages 218–229. Springer, 2009.

R. Chernobelskiy, K. I. Cunningham, M. T. Goodrich, S. G. Kobourov, and L. Trott.
Force-directed Lombardi-style graph drawing.
In *Graph Drawing*, pages 320–331, 2011.

References II

M. Chimani, C. Gutwenger, M. Jünger, G. W. Klau, K. Klein, and P. Mutzel.
The open graph drawing framework (OGDF).
Handbook of Graph Drawing and Visualization, pages 543–569, 2011.

R. Davidson and D. Harel.
Drawing graphs nicely using simulated annealing.
ACM Transactions on Graphics, 15(4):301–331, Oct. 1996.

W. De Nooy, A. Mrvar, and V. Batagelj.
Exploratory social network analysis with Pajek, volume 27.
Cambridge University Press, 2011.

T. Dwyer, Y. Koren, and K. Marriott.
Constrained graph layout by stress majorization and gradient projection.
Discrete Mathematics, 309(7):1895–1908, 2009.

T. Dwyer, B. Lee, D. Fisher, K. I. Quinn, P. Isenberg, G. G. Robertson, and C. North.
A comparison of user-generated and automatic graph layouts.
IEEE Trans. Vis. Comput. Graph., 15(6):961–968, 2009.

T. Dwyer, K. Marriott, and M. Wybrow.
Topology preserving constrained graph layout.
In *Graph Drawing*, pages 230–241, 2008.

P. Eades.
A heuristic for graph drawing.
Congressus Numerantium, 42:149–160, 1984.

J. Ellson, E. R. Gansner, E. Koutsofios, S. C. North, and G. Woodhull.
Graphviz - open source graph drawing tools.
In *Graph Drawing*, pages 483–484, 2001.

References III

A. Frick, A. Ludwig, and H. Mehldau.
A fast adaptive layout algorithm for undirected graphs.
In R. Tamassia and I. G. Tollis, editors, *Proceedings of the 2nd Symposium on Graph Drawing (GD)*, volume 894 of *Lecture Notes in Computer Science*, pages 388–403. Springer-Verlag, 1995.

Y. Frishman and A. Tal.
Dynamic drawing of clustered graphs.
pages 191–198.

T. Fruchterman and E. Reingold.
Graph drawing by force-directed placement.
Softw. – Pract. Exp., 21(11):1129–1164, 1991.

P. Gajer, M. T. Goodrich, and S. G. Kobourov.
A fast multi-dimensional algorithm for drawing large graphs.
Computational Geometry: Theory and Applications, 29(1):3–18, 2004.

P. Gajer and S. G. Kobourov.
GRIP: Graph dRawing with Intelligent Placement.
Journal of Graph Algorithms and Applications, 6(3):203–224, 2002.

E. Gansner, Y. Koren, and S. North.
Graph drawing by stress minimization.
In *Proceedings 12th Symposium on Graph Drawing (GD)*, pages 239–250, 2004.

S. Hachul and M. Jünger.
Drawing large graphs woth a potential-field-based multilevel algorithm.
In *Proceedings of the 12th Symposium on Graph Drawing (GD)*, volume 3383 of *Lecture Notes in Computer Science*, pages 285–295. Springer-Verlag, 2004.

References IV

S. Hachul and M. Jünger.
An experimental comparison of fast algorithms for drawing general large graphs.
In *Proceedings of the 13th International Conference on Graph Drawing*, GD'05, pages 235–250, Berlin, Heidelberg, 2006. Springer-Verlag.

R. Hadany and D. Harel.
A multi-scale algorithm for drawing graphs nicely.
Discrete Applied Mathematics, 113(1):3–21, 2001.

D. Hansen, B. Shneiderman, and M. A. Smith.
Analyzing social media networks with NodeXL: Insights from a connected world.
Morgan Kaufmann, 2010.

D. Harel and Y. Koren.
A fast multi-scale method for drawing large graphs.
Journal of graph algorithms and applications, 6:179–202, 2002.

D. Harel and Y. Koren.
Graph drawing by high-dimensional embedding.
In *Graph Drawing*, pages 207–219. Springer, 2002.

D. Harel and M. Sardas.
Randomized graph drawing with heavy-duty preprocessing.
J. Visual Lang. Comput., 6(3):233–253, 1995.

D. Harel and M. Sardas.
An algorithm for straight-line drawing of planar graphs.
Algorithmica, 20(2):119–135, 1998.

References V

J. Heer, S. K. Card, and J. A. Landay.
Prefuse: a toolkit for interactive information visualization.
In *Proc. SIGCHI conference on Human factors in computing systems*, pages 421–430. ACM, 2005.

Y. Hu.
Efficient and high quality force-directed graph drawing.
The Mathematica Journal, 10:37–71, 2005.

T. Kamada and S. Kawai.
An algorithm for drawing general undirected graphs.
Inform. Process. Lett., 31:7–15, 1989.

S. G. Kobourov.
Force-directed drawing algorithms.
Handbook of Graph Drawing and Visualization, pages 383–408, 2013.

Y. Koren, L. Carmel, and D. Harel.
Ace: A fast multiscale eigenvectors computation for drawing huge graphs.
In *Information Visualization, 2002. INFOVIS 2002. IEEE Symposium on*, pages 137–144. IEEE, 2002.

J. Kruskal and J. Seery.
Designing network diagrams.
In *Proceedings 1st General Conference on Social Graphics*, pages 22–50, 1980.

J. B. Kruskal.
Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis.
Psychometrika, 29:1–27, 1964.

M. J. M. Bastian, S. Heymann.
Gephi: an open source software for exploring and manipulating networks.
International AAAI Conference on Weblogs and Social Media, 2009.

References VI

K.-L. Ma and C. W. Muelder.
Large-scale graph visualization and analytics.
Computer, 46(7):39–46, 2013.

S. Martin, W. M. Brown, R. Klavans, and K. W. Boyack.
Openord: an open-source toolbox for large graph layout.
volume 7868, pages 786806–786806–11, 2011.

A. Noack.
Energy models for graph clustering.
J. Graph Algorithms Appl., 11(2):453–480, 2007.

A. Nocaj, M. Ortmann, and U. Brandes.
Untangling hairballs.
In *Graph Drawing*, pages 101–112. Springer, 2014.

A. Quigley and P. Eades.
FADE: graph drawing, clustering, and visual abstraction.
In *Proceedings of the 8th Symposium on Graph Drawing (GD)*, volume 1984 of *Lecture Notes in Computer Science*, pages 197–210. Springer-Verlag, 2000.

P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker.
Cytoscape: a software environment for integrated models of biomolecular interaction networks.
Genome research, 13(11):2498–2504, 2003.

D. Tunkelang.
JIGGLE: Java interactive general graph layout environment.
In *Graph Drawing*, volume 1547 of *Lecture Notes in Computer Science*, pages 413–422. Springer-Verlag, 1998.

References VII

T. Von Landesberger, A. Kuijper, T. Schreck, J. Kohlhammer, J. J. van Wijk, J.-D. Fekete, and D. W. Fellner.
Visual analysis of large graphs: State-of-the-art and future research challenges.
In *Computer graphics forum*, volume 30, pages 1719–1749. Wiley Online Library, 2011.

C. Walshaw.
A multilevel algorithm for force-directed graph drawing.
Journal of Graph Algorithms and Applications, 7(3):253–285, 2003.

R. Wiese, M. Eiglsperger, and M. Kaufmann.
yFiles: Visualization and automatic layout of graphs.
In *Graph Drawing*, pages 453–454, 2001.

Questions

- Quantitative measures beyond stress and strain

Questions

- Quantitative measures beyond stress and strain
- Qualitative measures (aesthetics) beyond edge crossings

Questions

- Quantitative measures beyond stress and strain
- Qualitative measures (aesthetics) beyond edge crossings
- Validated datasets and benchmarks

Questions

- Quantitative measures beyond stress and strain
- Qualitative measures (aesthetics) beyond edge crossings
- Validated datasets and benchmarks
- Real-world tasks for user studies

Questions

- Quantitative measures beyond stress and strain
- Qualitative measures (aesthetics) beyond edge crossings
- Validated datasets and benchmarks
- Real-world tasks for user studies
- Engagement, enjoyment, memorability

Questions

- Quantitative measures beyond stress and strain
- Qualitative measures (aesthetics) beyond edge crossings
- Validated datasets and benchmarks
- Real-world tasks for user studies
- Engagement, enjoyment, memorability
- Intuitive interactive visualization

Questions

- Quantitative measures beyond stress and strain
- Qualitative measures (aesthetics) beyond edge crossings
- Validated datasets and benchmarks
- Real-world tasks for user studies
- Engagement, enjoyment, memorability
- Intuitive interactive visualization
- What did I miss?