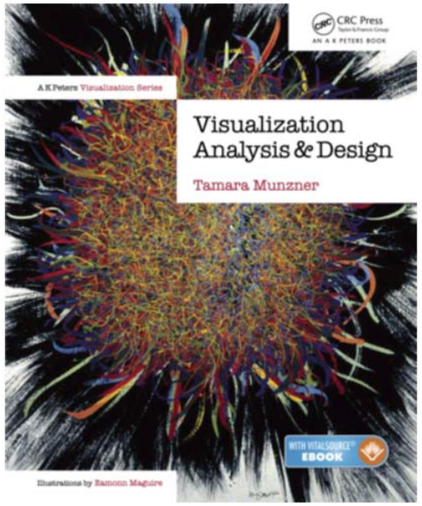
Task-driven Automated Data Visualization

Yunhai Wang

Shandong University


Data

Why are we using Visualization?



Tamara Munzner 2011:

"Computer-based visualization systems provide visual representations of datasets intended to help people carry out some **tasks** more effectively"

Data + Task = Visualization?[1]

Ask which visualization is **best** suited to pursue a given task
on given input data

[1] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. A design space of visualization tasks. Visualization and Computer Graphics, IEEE Transactions on, 19(12):2366–2375, 2013.

Domain Tasks

Infinite numbers of domain tasks!

Table 1. Questions for the analysis of conserved syntenic data, with the scale and relationship addressed by each. The scales are: g, genome; c, chromosome; b, block; and f, feature. The relationships are: p, proximity/location; z, size; o, orientation; and s, similarity.

	question	*scale*				*relationship*			
		g	*c*	*b*	*f*	*p*	*z*	*o*	*s*
1	Which chromosomes share conserved blocks?	X				X			
2	For one chromosome, how many other chromosomes does it share blocks with?	X	X			X			
3	What is the density of coverage and where are the gaps on: chromosomes? blocks?	X	X	X		X			
4	Where are the blocks: on chromosomes? around a specific location on a chromosome?	X	X			X			
5	What are the sizes and locations of other genomic features near a block?		X			X	X		
6	How large are the blocks?		X				X		
7	Do neighboring blocks go to the same chromosome? relative location on a chromosome?	X	X			X			
8	Are the orientations matched or inverted for: block pairs? feature pairs?		X	X				X	
9	Do the orientations match for pairs of: neighboring blocks? features within a block?		X	X				X	
10	Are similarity scores alike: with respect to neighboring blocks? within a block?		X	X					X
11	Are the paired features within a block contiguous?			X		X			
12	How large is a feature relative to other genes within a block?			X			X		
13	What are the sizes, locations, and names of features within a block?			X		X	X		
14	What are the differences between individual nucleotides of feature pairs?				X				X

Genome

Mizbee, Infovis'10

3.1 Domain Tasks

Three main tasks in forecast calibration have been identified through close collaboration with a group of forecasters, observations on their routine work, and detailed discussions about their working flow.

T1 **Generating Initial Forecast** An initial forecast is produced using a post-processing method, such as the analog method. This forecast serves as a basis for the subsequent calibration process.

Weather

T2 **Detecting Regions of Interest (ROI)** The initial forecast derived after the post-processing step could have biases. Thus, forecasters need to detect ROIs where biases exist.

T3 **Applying Detailed Calibrations** Calibrations are applied to the detected ROIs statistically or manually according to professional knowledge of the experts regarding the current weather state.

Visual voting, Scivis 2015

The analysts usually conduct a complex task through a set of low level tasks. These tasks typically focus on the behavior of the seller, such as:

T1 Identifying time periods and/or sales categories of interest.

T2 Identifying transactions with interesting patterns in specific attributes (e.g., payment amount $\geq$ 500) and examining their detailed information.

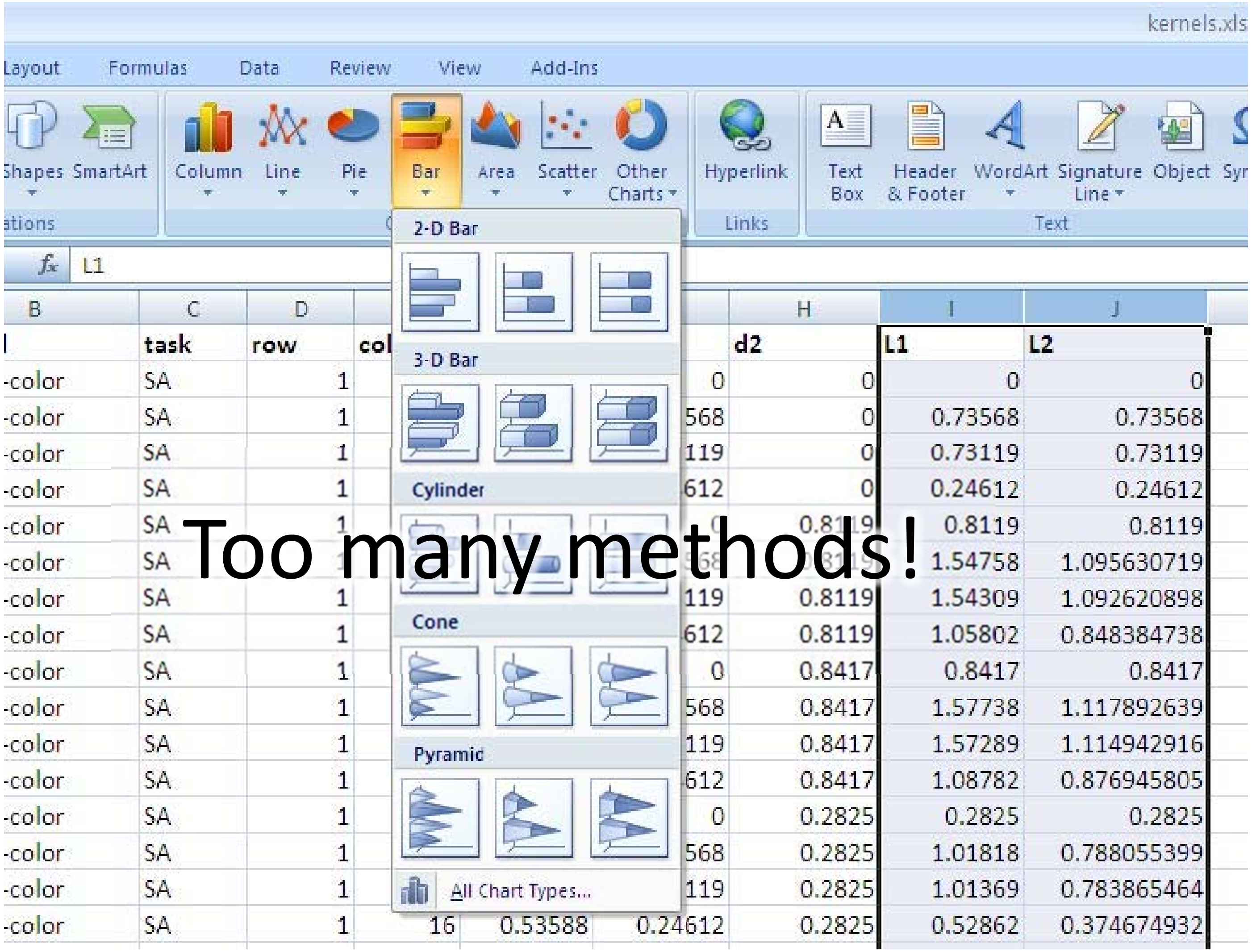
E-transactions

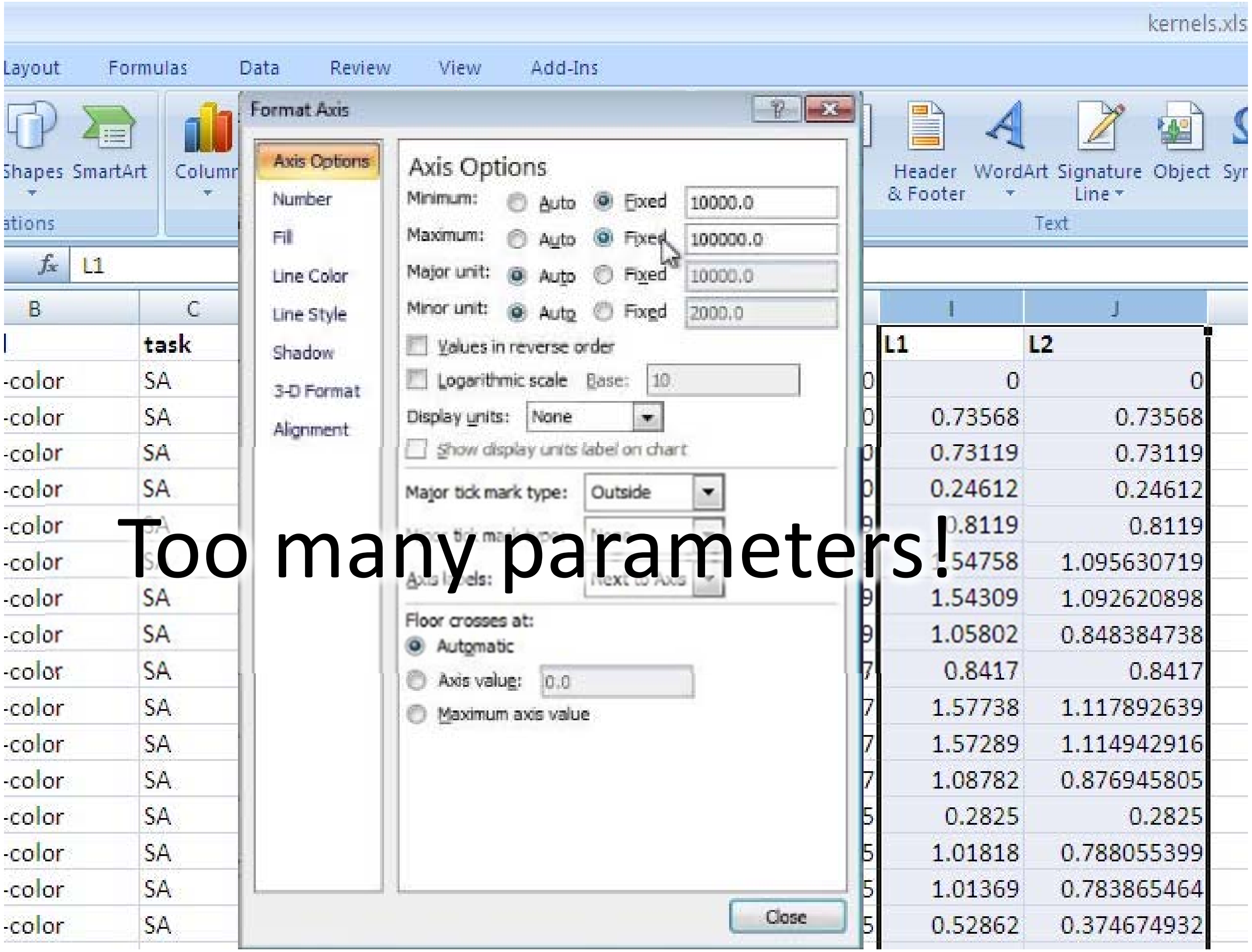
T3 Identifying sellers with interesting transaction patterns, such as a seller making frequent transactions with small payment amounts.

T4 Examining the transaction patterns of a specific seller.

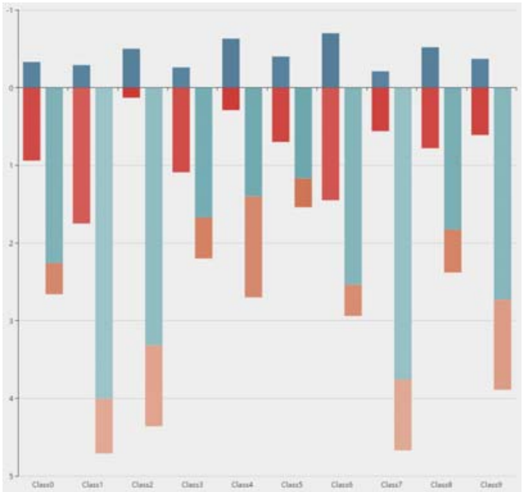
VAST 2014

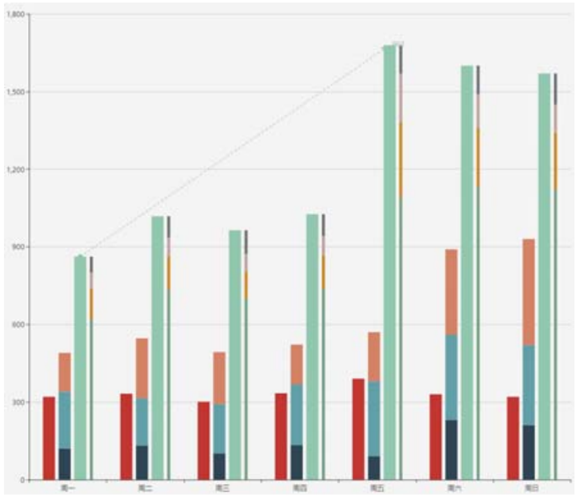
Too many methods!

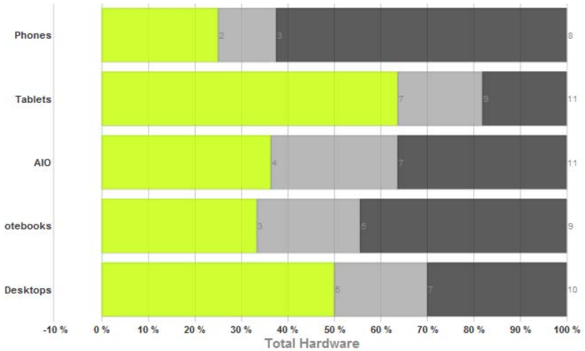


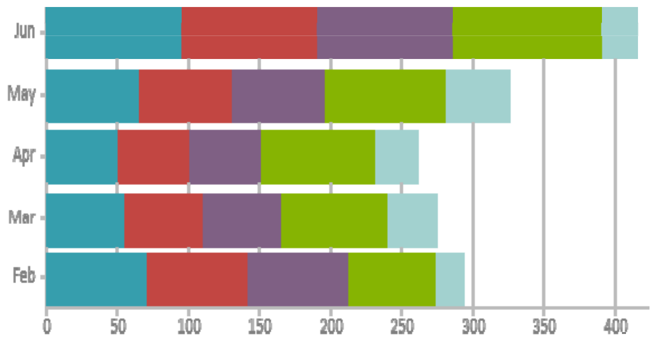


Too many parameters!



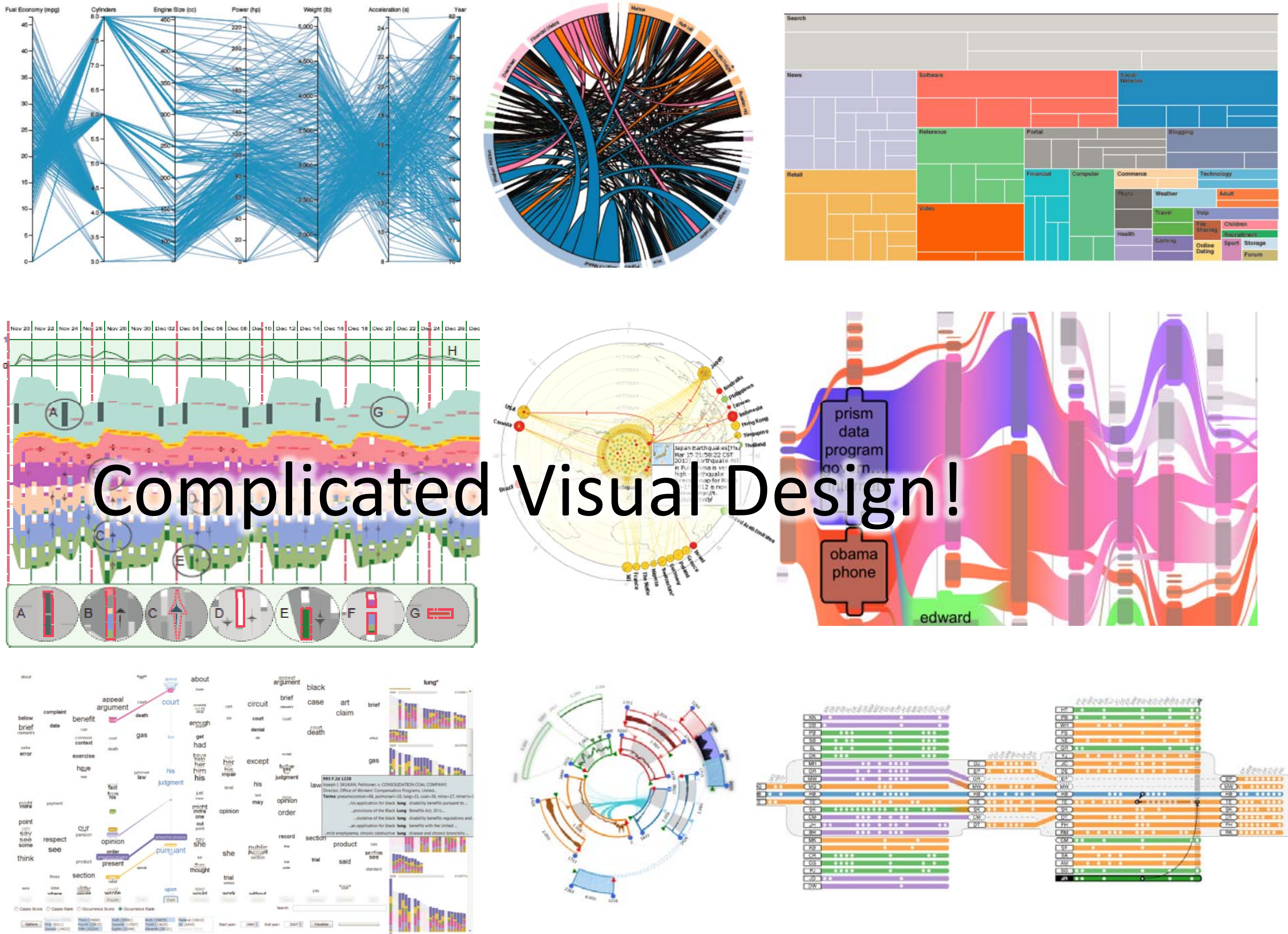


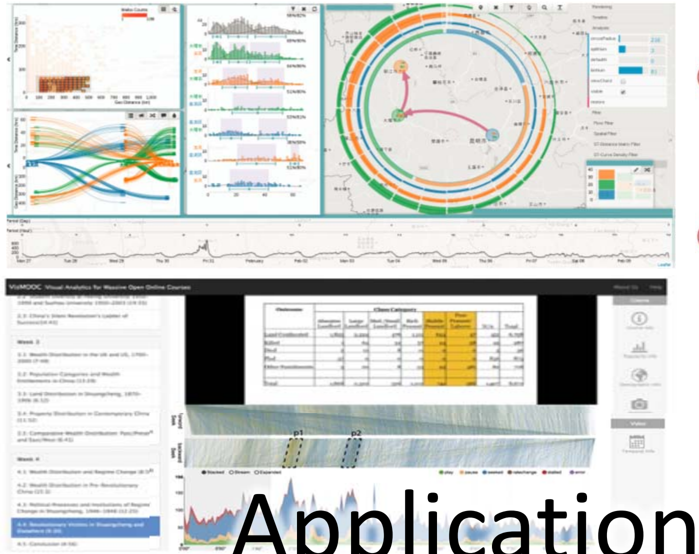


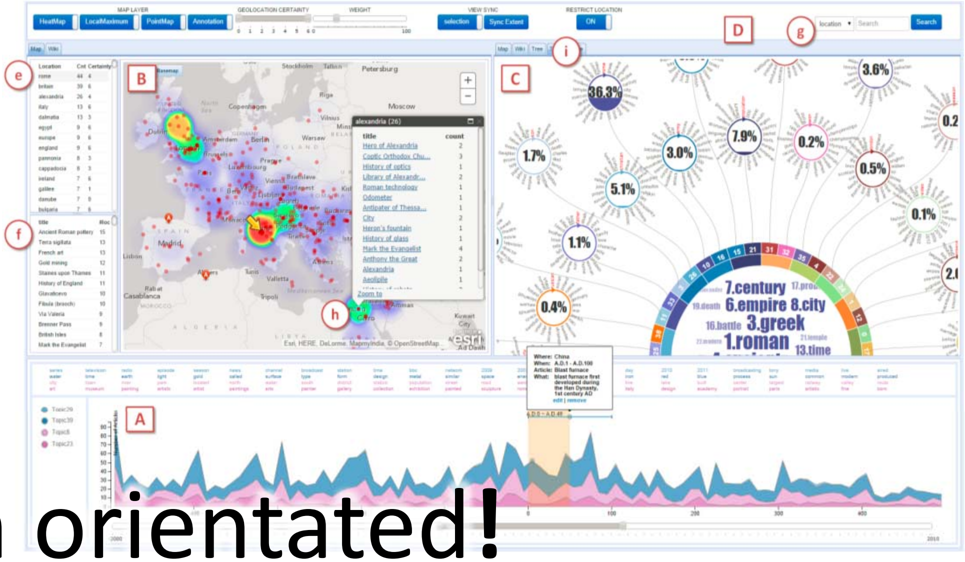


A **tedious** procedure!

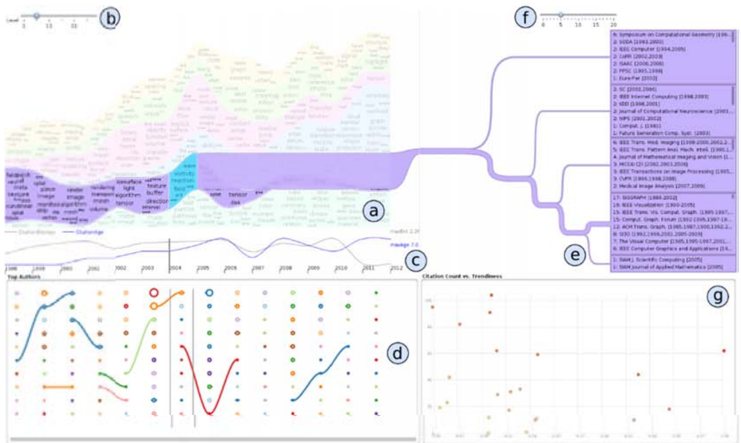
Unfortunately, our community pays little attention on this!

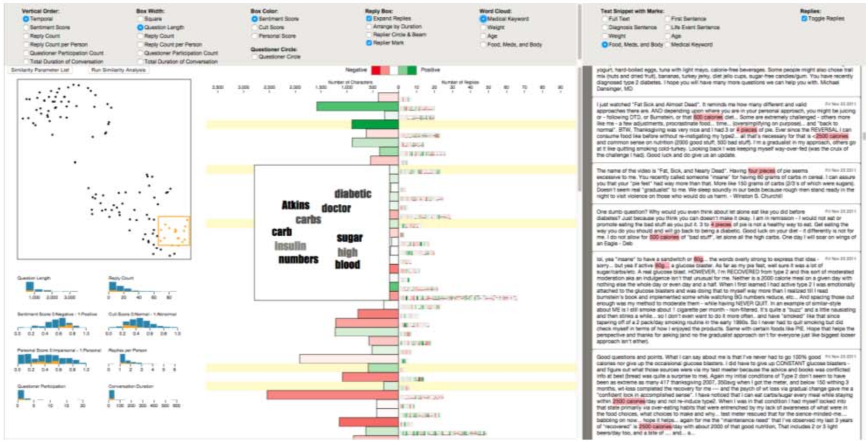

Complicated Visual Design!
prism data program
obama phone
edward





Application orientated!





Gaps!

- It is **not easy** to make sense of unfamiliar visualization[1];

- Most users have **few** experience in visual design but we have **many** different applications.

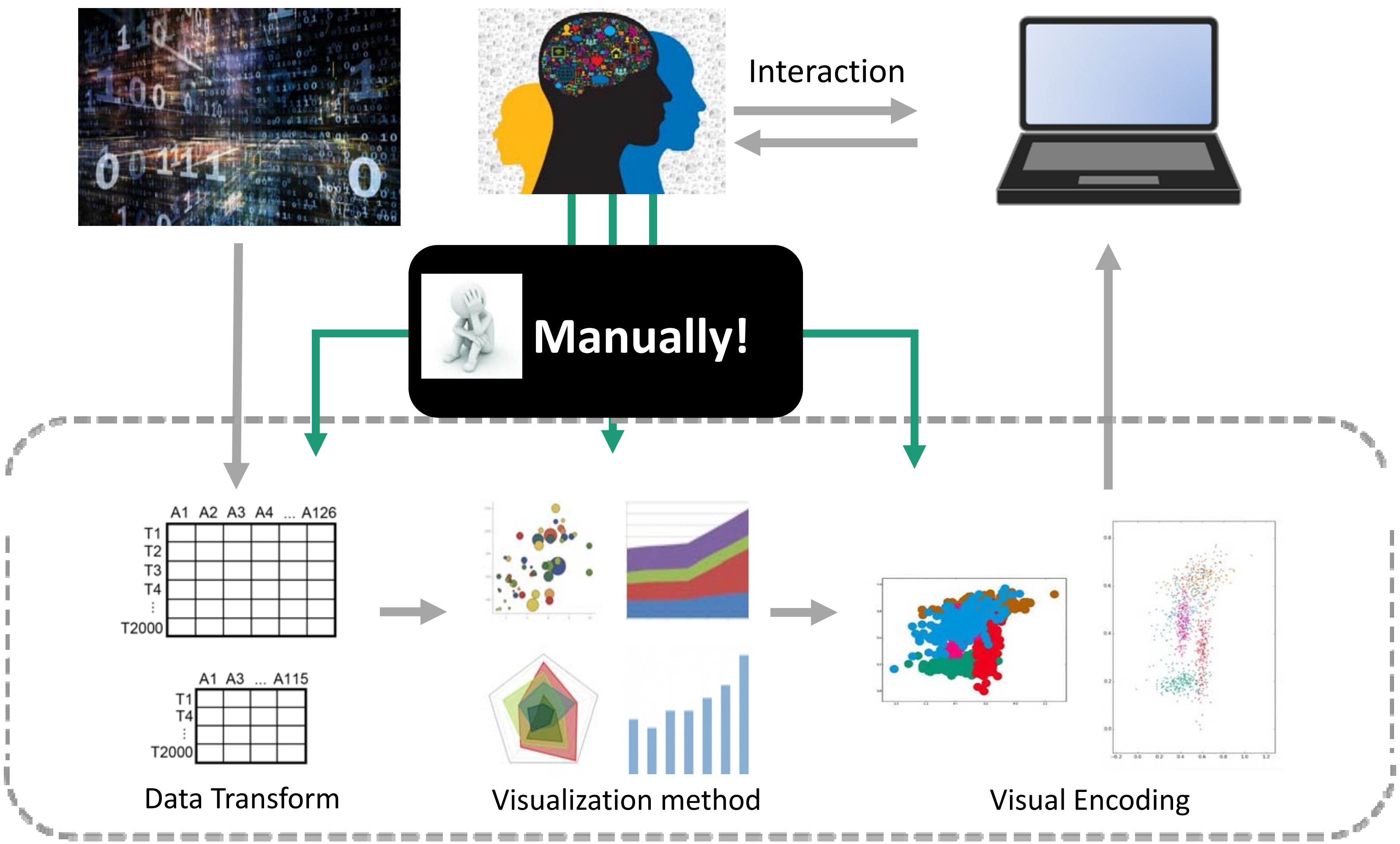
[1] Sukwon Lee, Sung-Hee Kim, Ya-Hsin Hung, Heidi Lam, Youn-Ah Kang, Ji Soo Yi: How do People Make Sense of Unfamiliar Visualizations?: A Grounded Model of Novice's Information Visualization Sensemaking. IEEE Trans. Vis. Comput. Graph. 22(1): 499-508 (2016)

Our solution

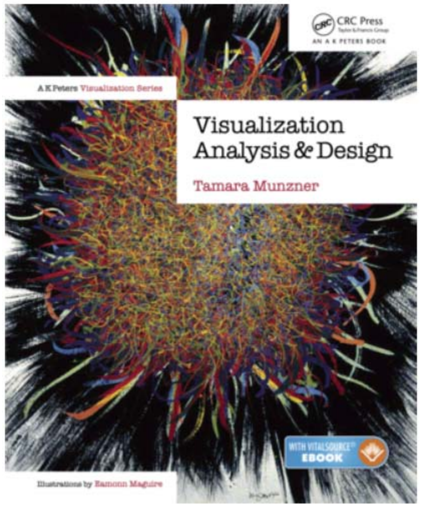
Task-driven Automated

Data Visualization

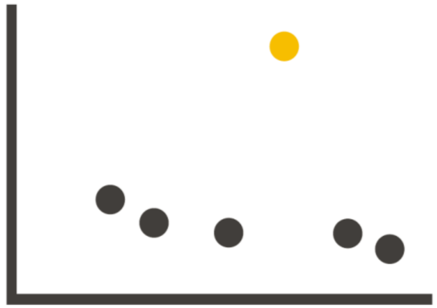
Traditional Visualization Pipeline



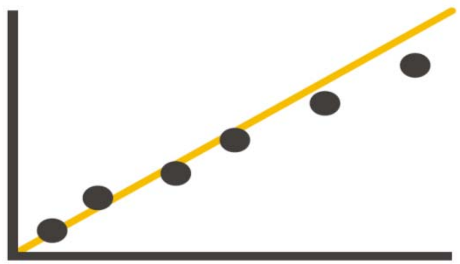
Abstract tasks



Trend

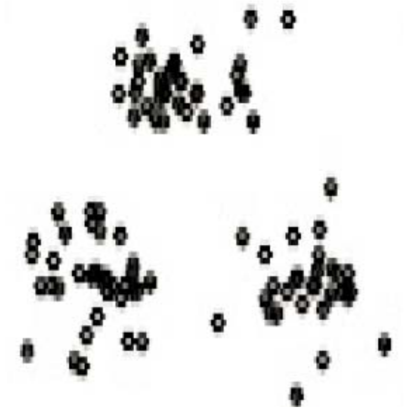


Outlier



Correlation

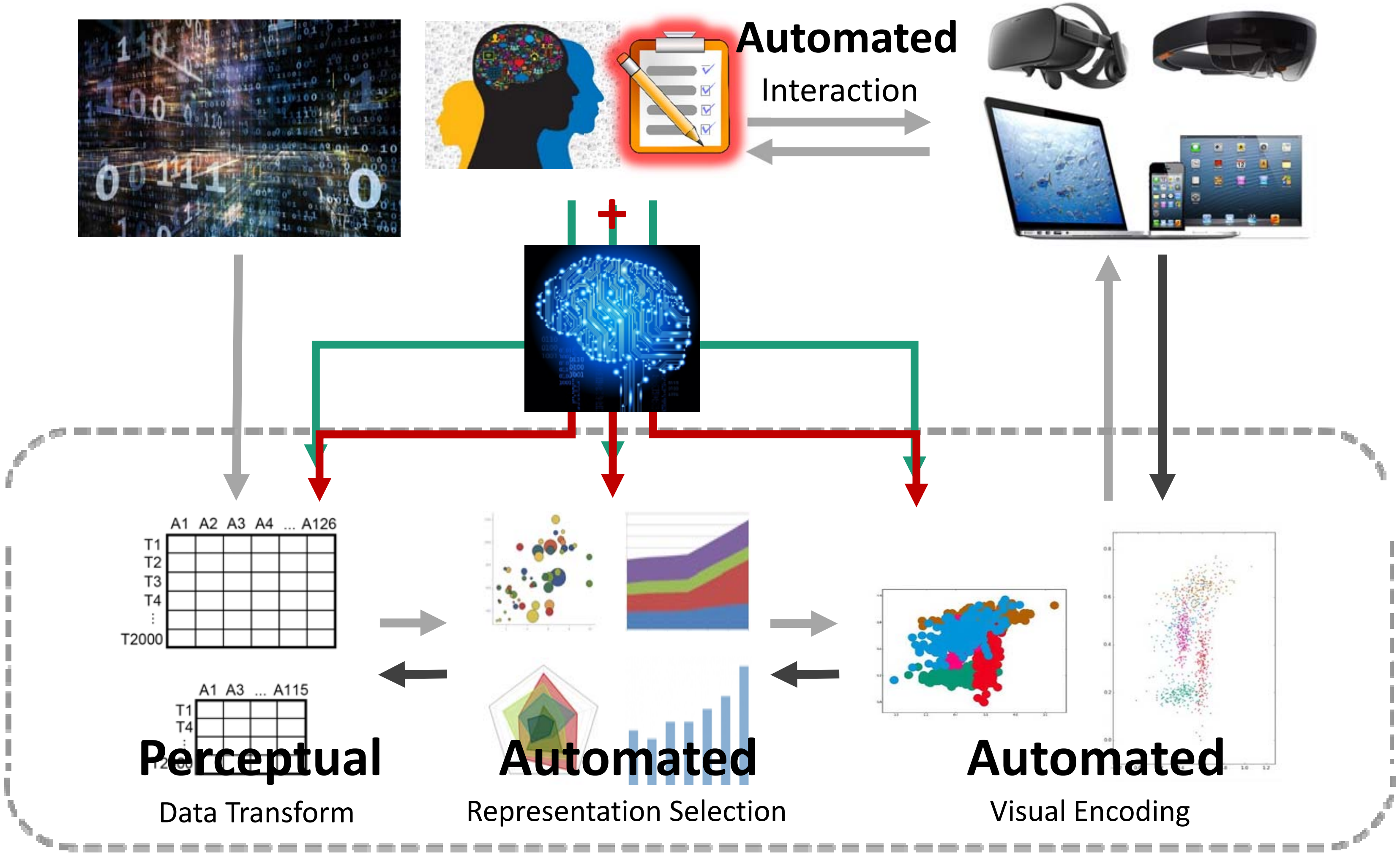
Distribution



Clusters

Similarity

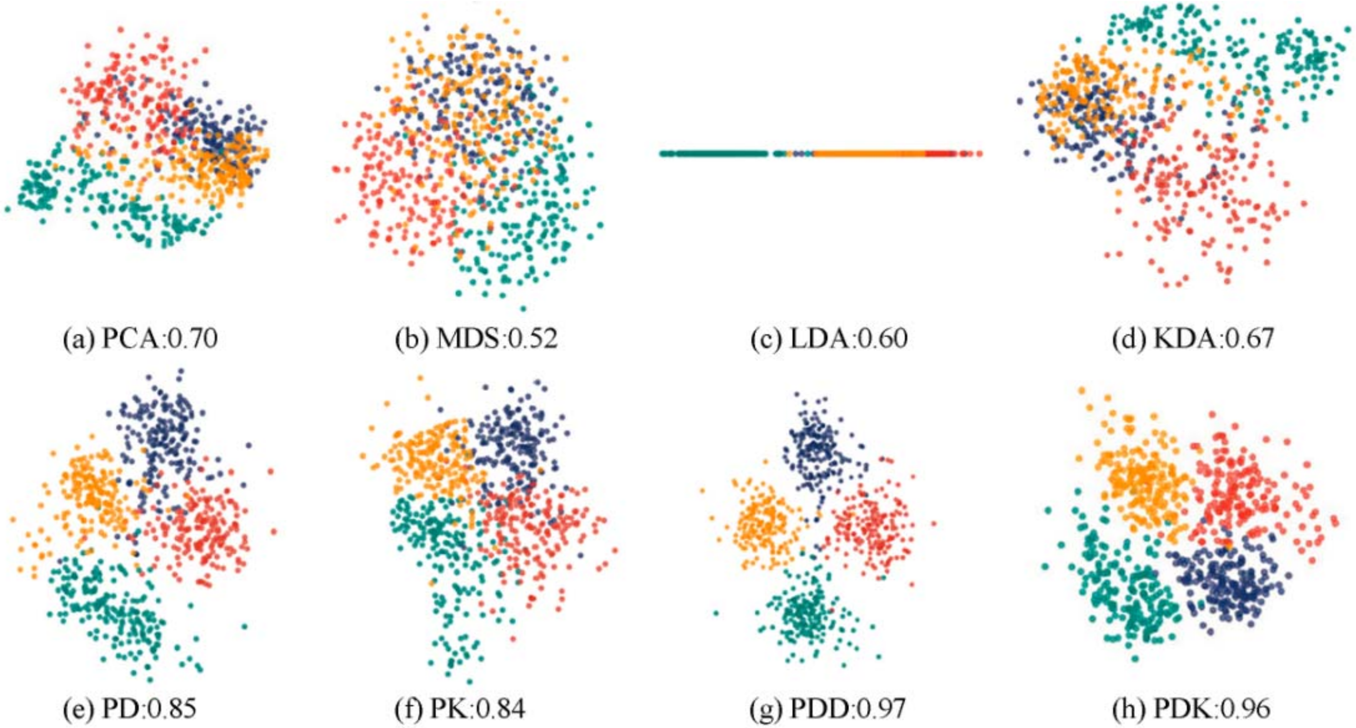
Task-driven Automated Visualization Design



A Perception-Driven Approach to Supervised Dimensionality Reduction for Visualization

Yunhai Wang [1] Kang Feng [1] Xiaowei chu[1] Jian Zhang[2] Chi-Wing Fu[3]
Michael Sedlmair[4] Xiaohui Yu[1] Baoquan Chen[1]
[1]Shandong University [2]CNIC, CAS [3]the Chinese University of Hong Kong [4]University of Vienna, Austria

IEEE Transactions on Visualization and Computer Graphics 2017

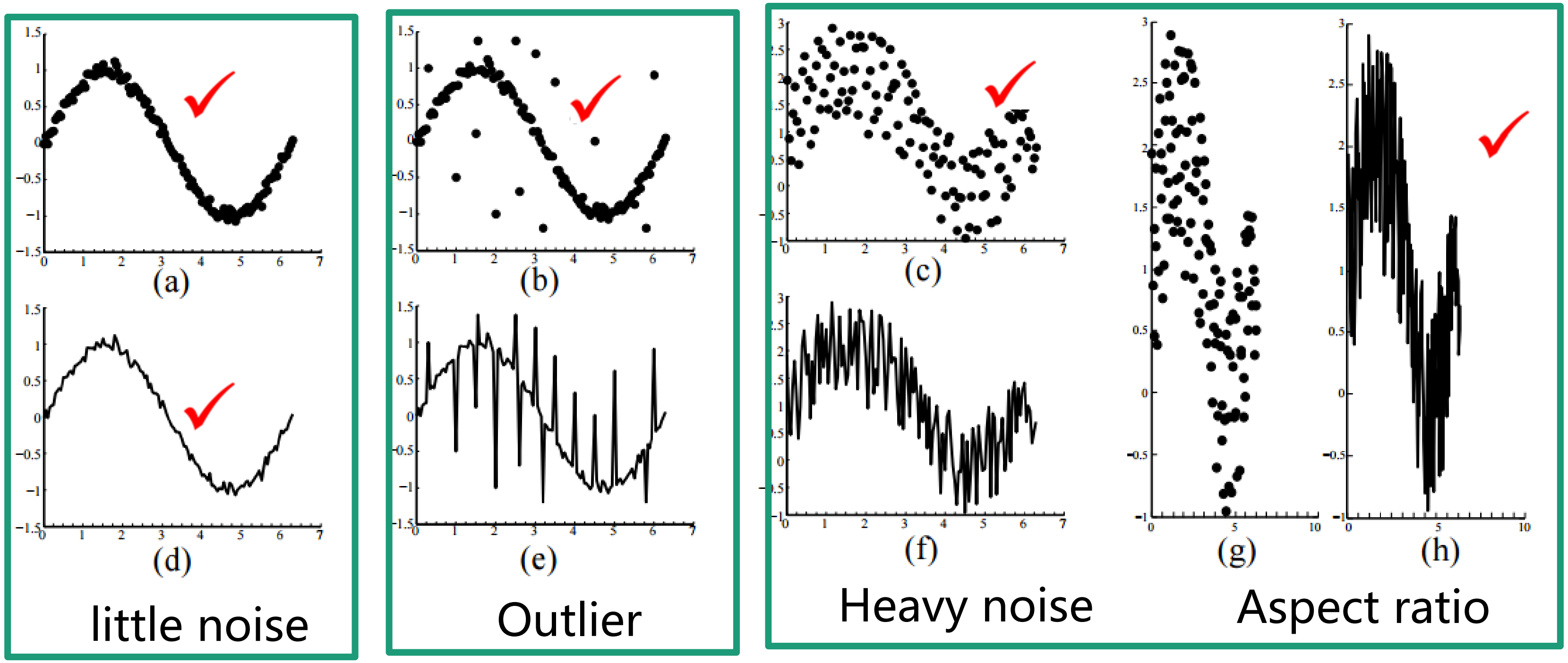


Perceptual Data Transform

Line Graph or Scatter Plot? Automatic Selection of Methods for Visualizing Trends in Time Series

Yunhai Wang [1*] Fubo Han [1*] Lifeng Zhu[2] Oliver Deussen[3] Baoquan Chen[1]
[1]Shandong University [2]Southeast University [3]University Konstanz

IEEE Transactions on Visualization and Computer Graphics 2017



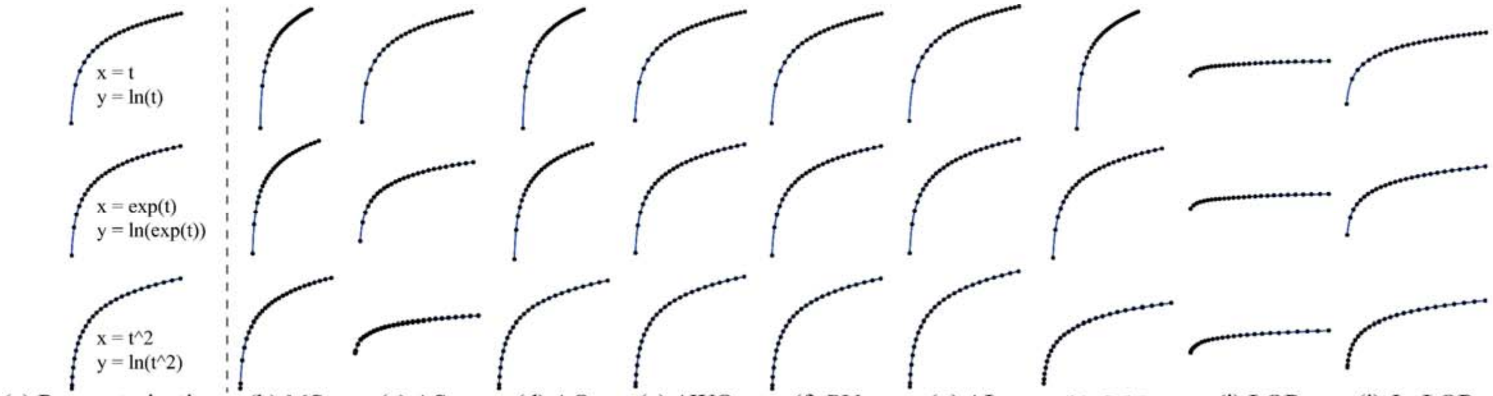
Automated Visualization Selection

Is There a Robust Technique for Selecting Aspect Ratio in Line Charts?

Yunhai Wang [1] Zeyu Wang [1] Fubo Han [1] Lifeng Zhu [2] Jian Zhang [3]
Changhe Tu [1] Oliver Deussen [4] Baoquan Chen [1]
[1]Shandong University [2]Southeast University [3]Computer Network Information Center [4]University of Konstanz

Submitted to IEEE Transactions on Visualization and Computer Graphics

Under Minor Revision

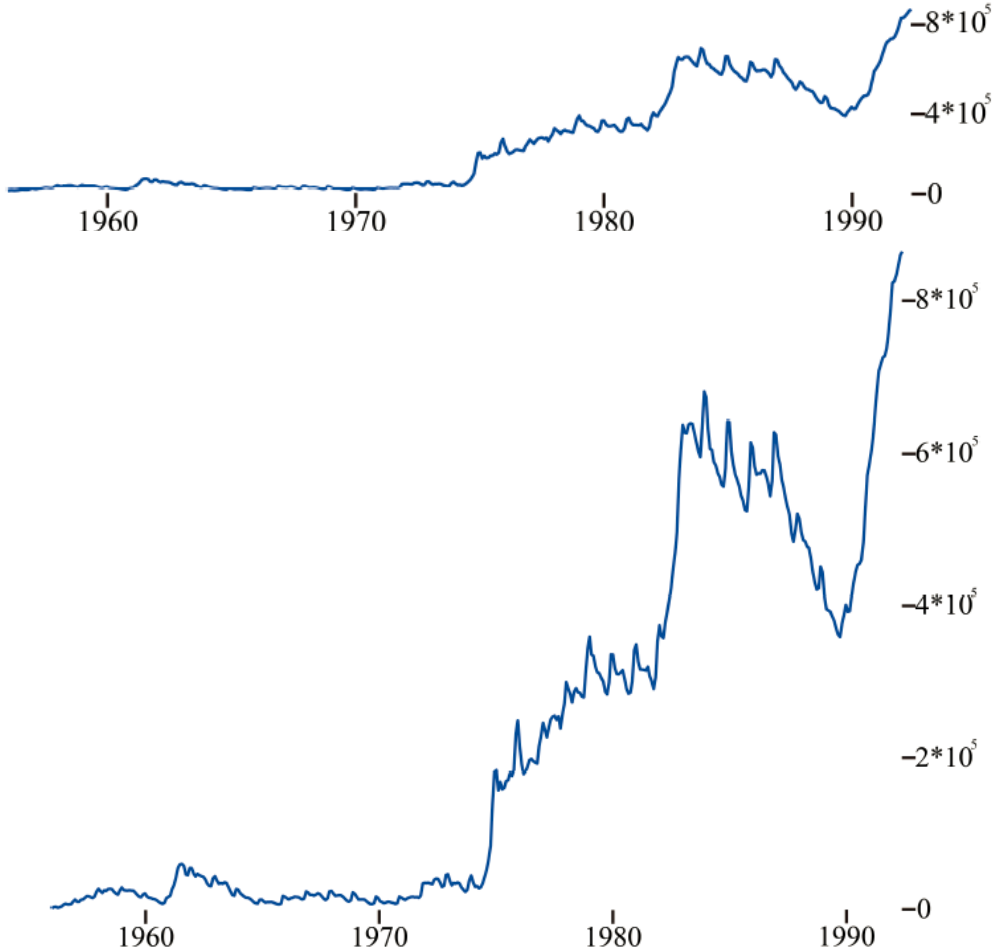


Methods	Means ± std
MS	1.767 ± 0.419
AS	0.685 ± 0.264
AO	1.465 ± 0.434
AWO	0.995 ± 0.001
RV	1 ± 0
AL	1.044 ± 0.004
GOR	0.895 ± 0.206
LOR	0.129 ± 0.068
L_1-LOR	0.515 ± 0.008

(a) Parameterizations (b) MS (c) AS (d) AO (e) AWO (f) RV (g) AL (h) GOR (i) LOR (j) L_1-LOR (k) Aspect ratios

Automated Visual Encoding

Aspect ratio strongly impacts graphical perception

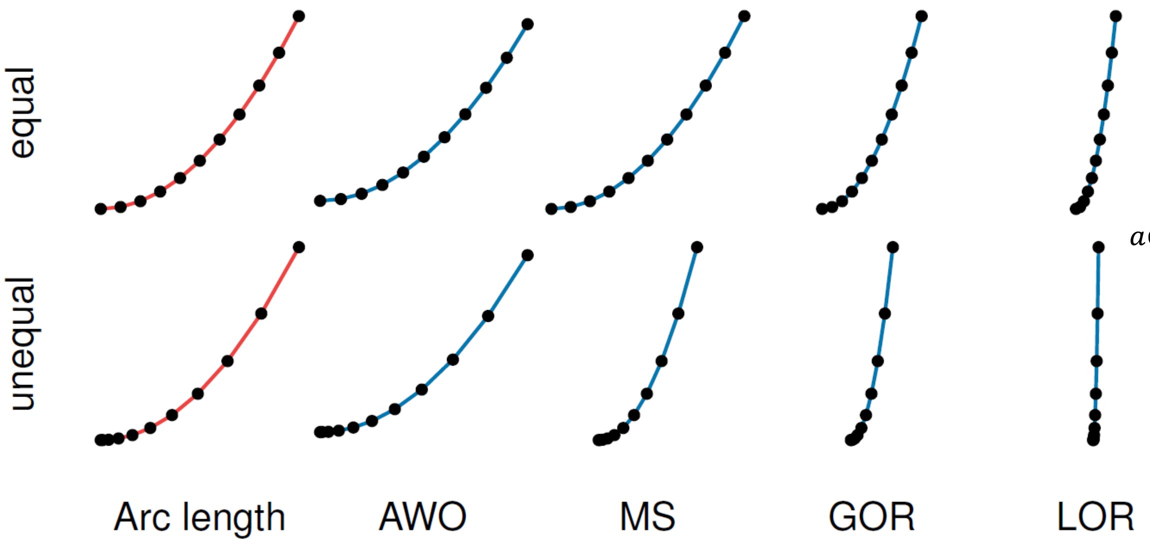


(b) Aspect ratio = 0.752

Many aspect ratio selection methods!

Name of Method	Abbr.	Ref.	Def.	Para. Invar.
Median Absolute **S**lope	MS	[3]	Eq. 3	×
Average Absolute **S**lope	AS	[7]	Eq. 4	×
Average Absolute **O**rientation	AO	[3]	Eq. 5	×
Arc Length **W**eighted Average Absolute **O**rientation	AWO	[4]	Eq. 6	✓
Global **O**rientation **R**esolution	GOR	[7]	Eq. 8	×
Local **O**rientation **R**esolution	LOR	[7]	Eq. 9	×
L_1-norm based **L**ocal **O**rientation **R**esolution	L_1-LOR	Ours	Eq. 17	✓
Resultant **V**ector	RV	[8]	Eq. 10	✓
Arc **L**ength	AL	[9]	Eq. 11	✓

Our Contribution I



$$\min_{a\in(0,\infty)} \sum_{i=1}^{n} \left\| \frac{\Delta x_i}{\sqrt{a}}, \sqrt{a}\Delta y_i \right\|$$

$$= \min_{a\in(0,\infty)} \frac{1}{\sqrt{a}} \sum_{i=1}^{n} \|\Delta x_i, a\Delta y_i\|$$

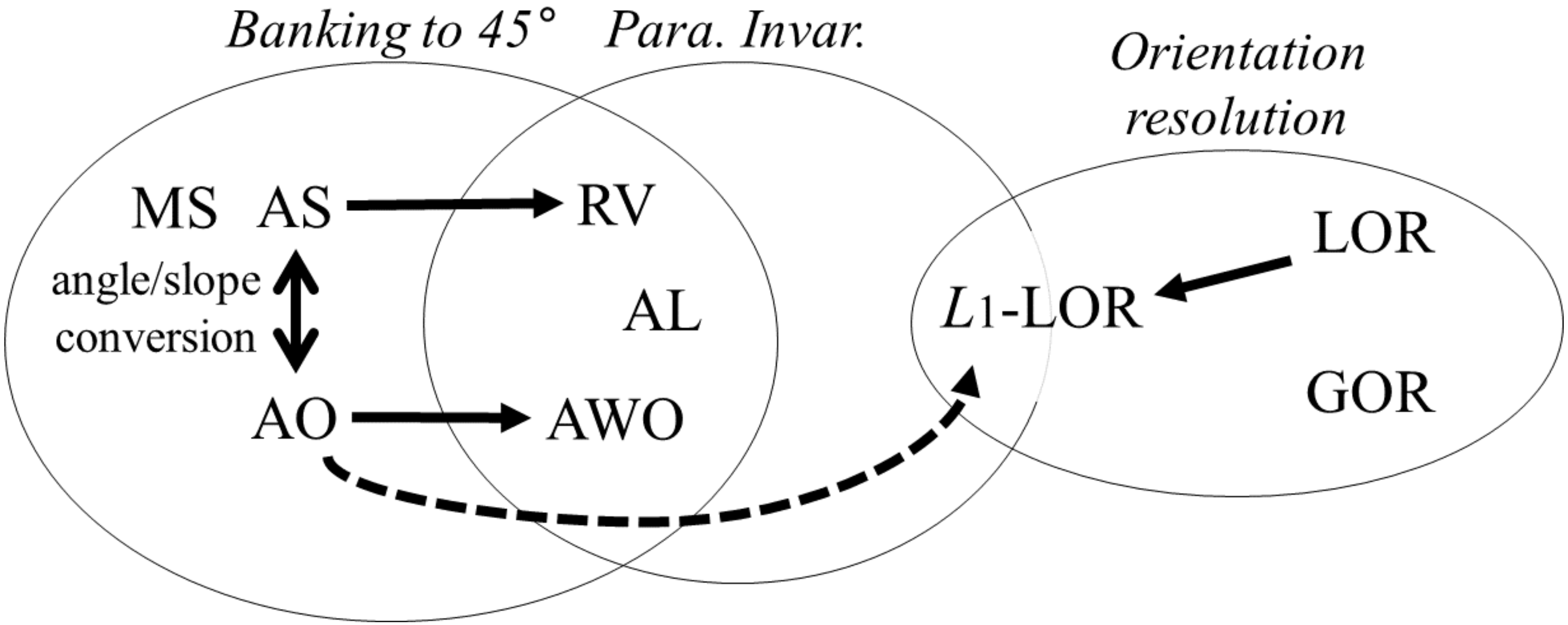
$$= \min \frac{1}{\sqrt{a}} \int_C ds$$

Arc-length (AL)

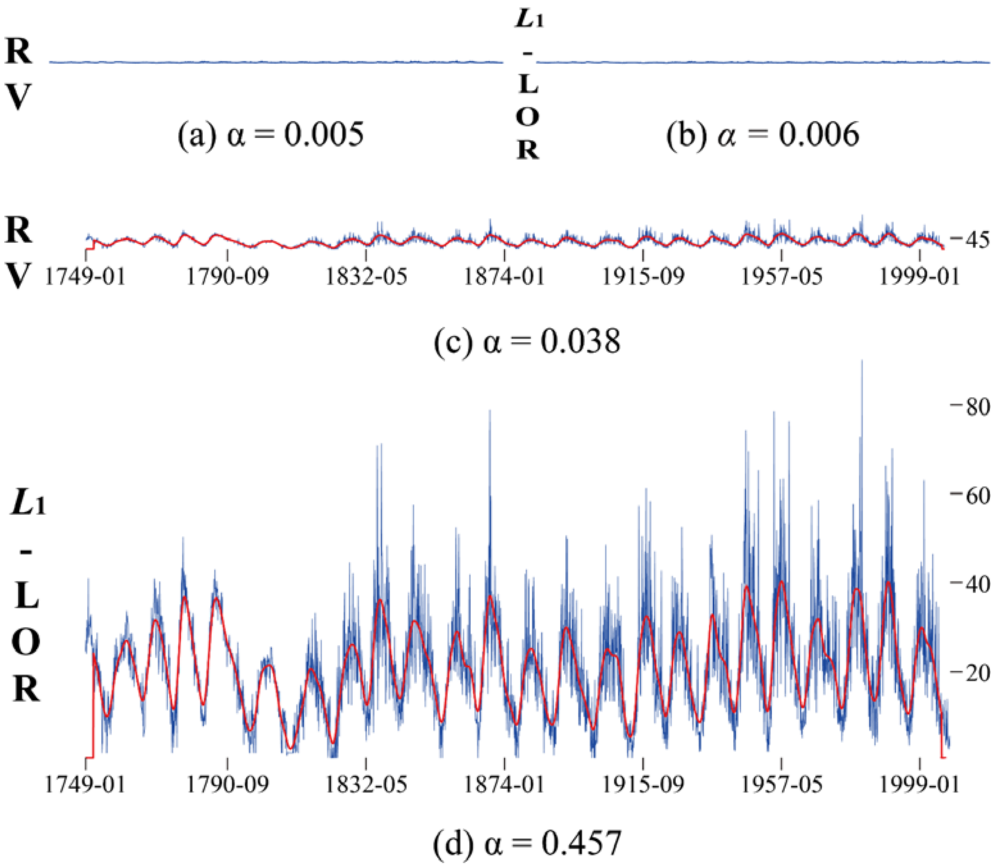
Why some methods are parameterization invariant?

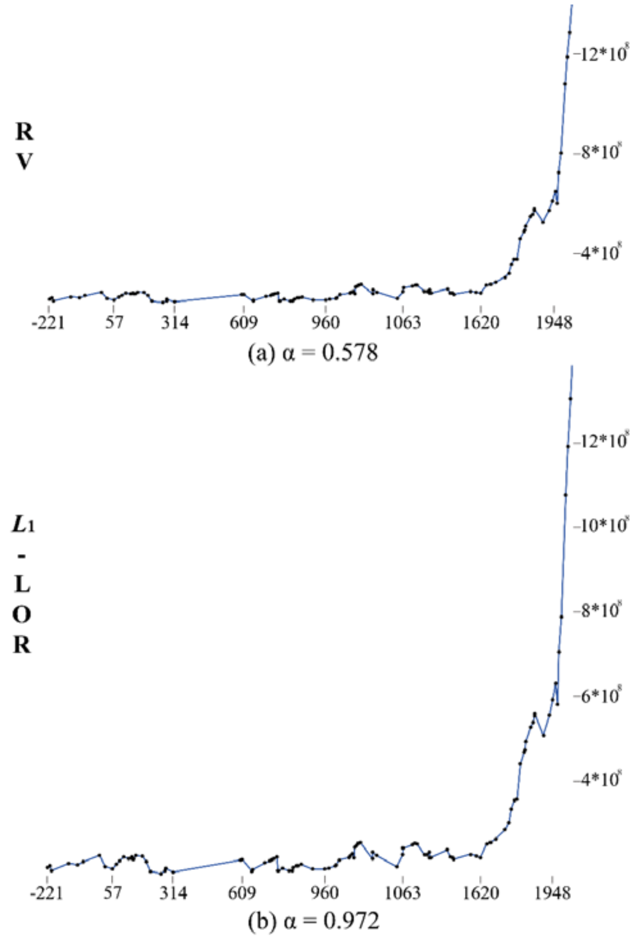
Line integral

Our Contribution II



Our Contribution III: dual-scale banking





EdWordle: Consistency-preserving Word Cloud Editing

VIS 2017

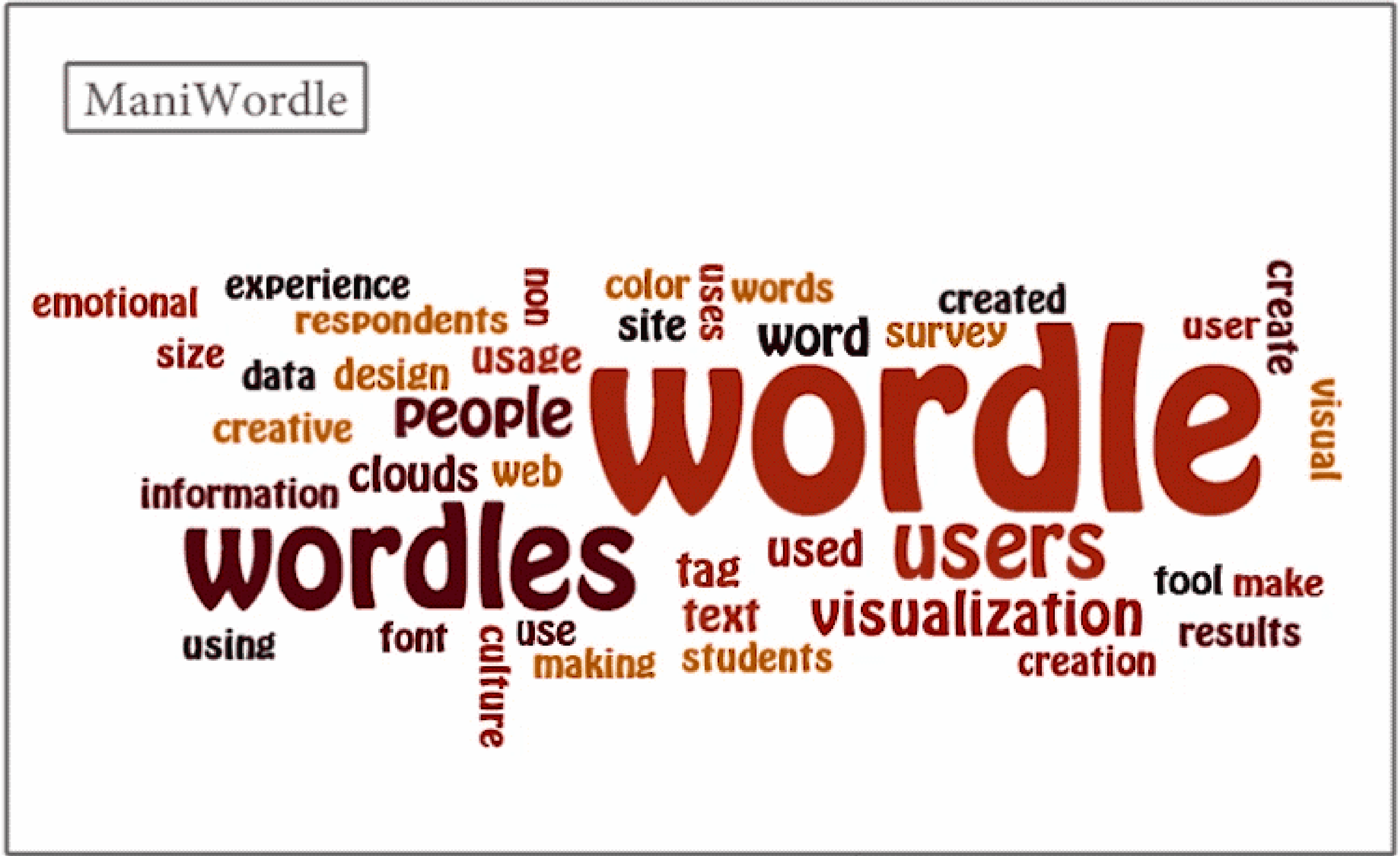
Yunhai Wang, Xiaowei Chu, Chen Bao, Lifeng Zhu
Oliver Deussen, Baoquan Chen and Michael Sedlmair

IEEE Transactions on Visualization and Computer Graphics
(Proc. IEEE InfoVis 2017)

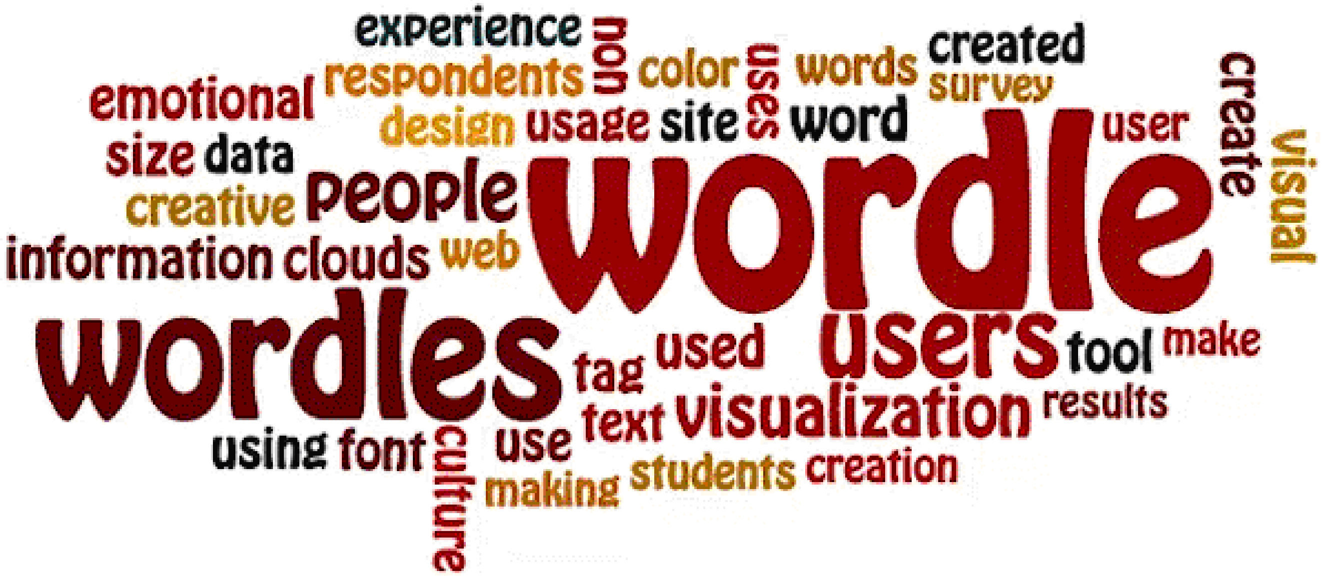
Try EdWordle at www.edwordle.net

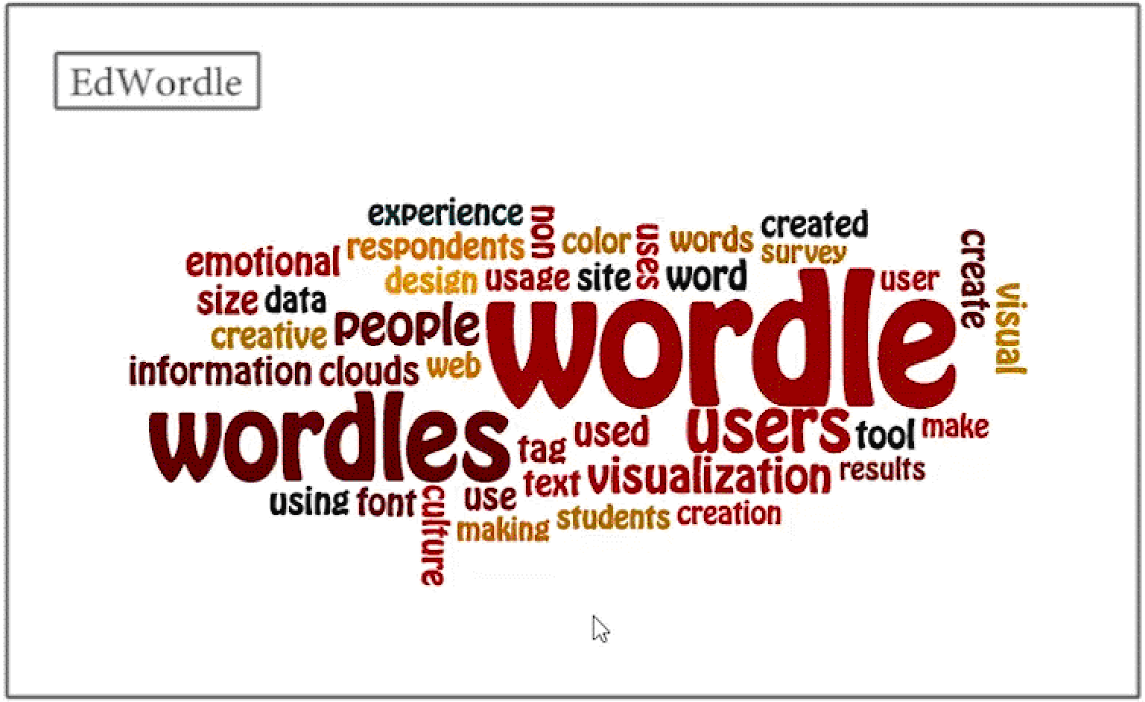
Fig. 1. Result of a case study with a professional writer who sought to visualize a BBC news feed: the left image shows the input Wordle layout; the right image shows the layout that was created using EdWordle. The writer ordered related words into semantically meaningful groups, one group per story. Each group was organized spatially together and color-coded, creating a layout that the user referred to as a *"storytelling cloud"*.

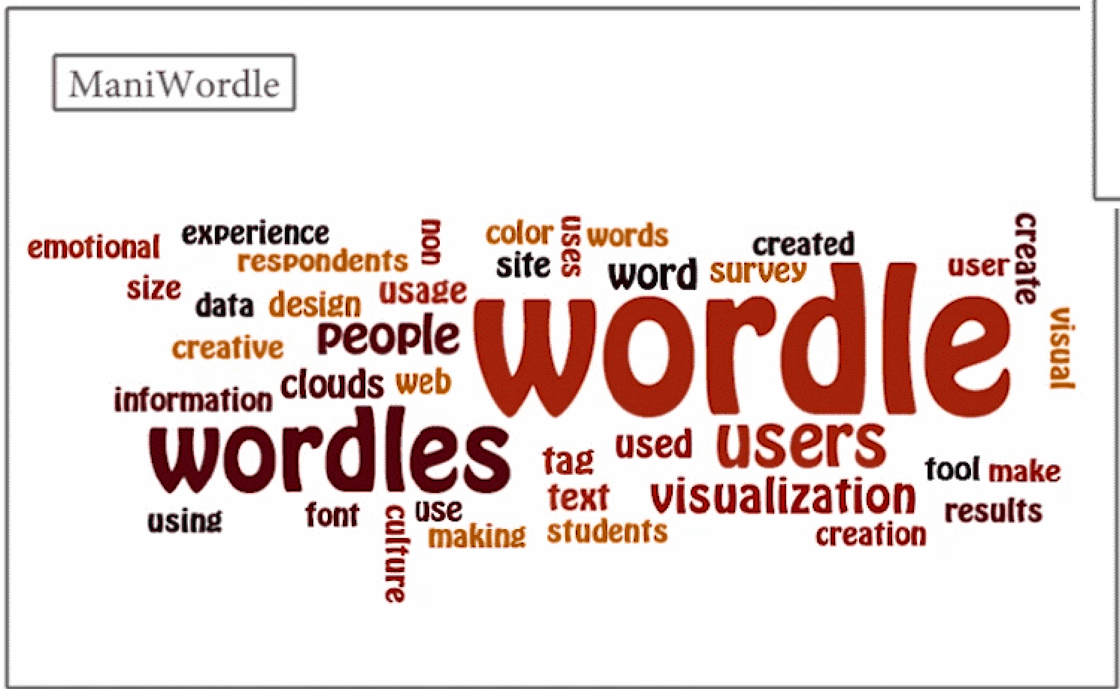
Automated Visual Interaction


ManiWordle
emotional
experience
respondents
non
color
uses
words
created
create
size
data
design
usage
site
word
survey
user
wordle
creative
people
visual
information
clouds
web
wordles
used
users
tag
fool
make
text
visualization
using
font
use
culture
making
students
results
creation

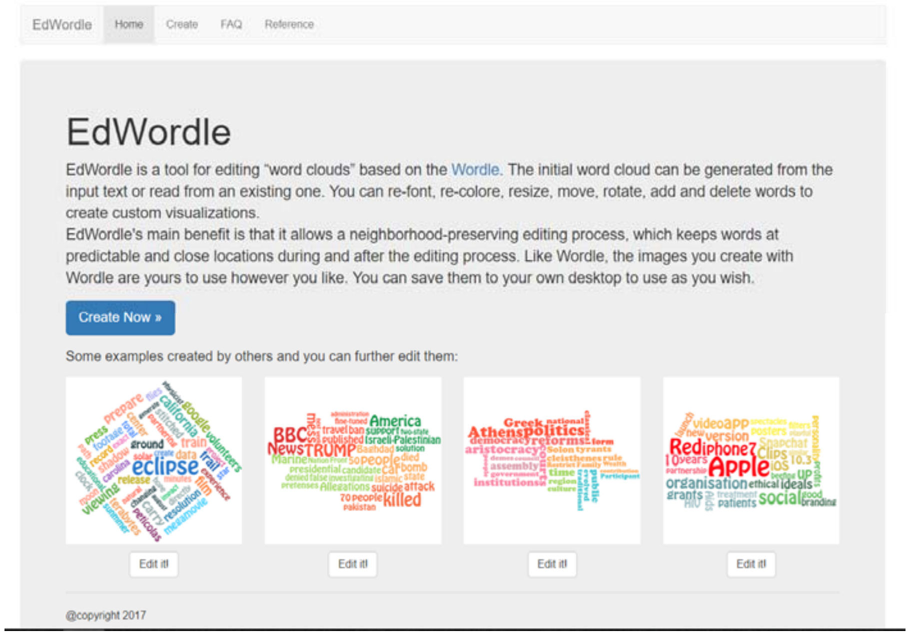
EdWordle

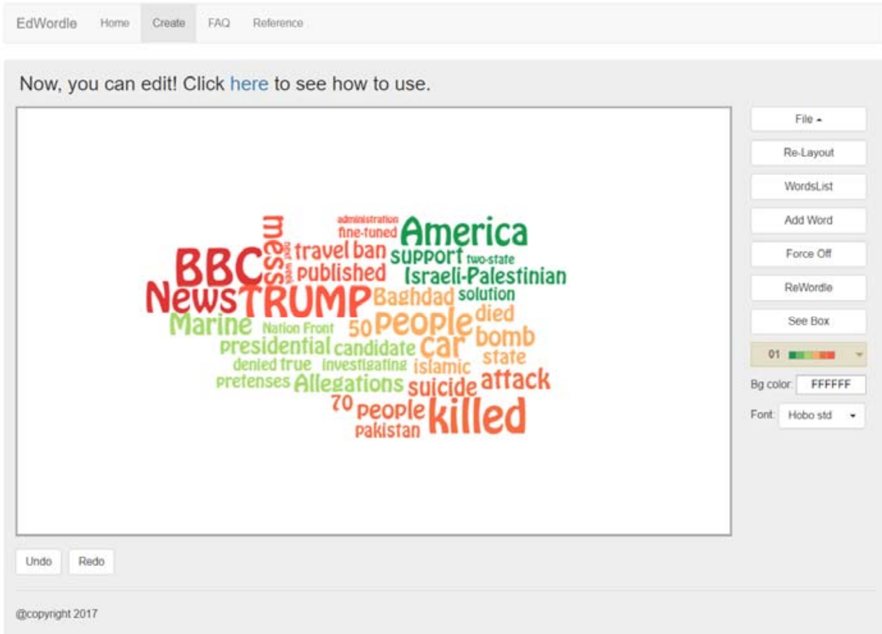

experience
non
color
uses
words
created
survey
respondents
emotional
design
usage
site
word
user
create
visual
size
data
creative
people
wordle
information
clouds
web
wordles
tag
used
users
tool
make
text
visualization
results
using
font
culture
use
making
students
creation


EdWordle
experience non color uses words created
emotional respondents survey
design usage site word user create
size data visual
creative people
information clouds web wordle
wordles users tool make
tag used
text visualization results
using font culture use
making students creation


ManiWordle
emotional experience non color uses words created create
respondents site word survey user
size data design usage
creative people visual
information clouds web wordle
wordles tag used users tool make
text visualization results
using font culture use making students creation

Our System: www.edwordle.net





Consistency-Preserving Wordle Editing !

VIS 2017

John A Guerra Gómez and Krist Wongsuphasawat liked

Robert Kosara @eagereyes · 7h

EdWordle: Consistency-preserving Word Cloud Editing – move words without everything moving around. #IEEEVIS edwordle.net

1 10 7

VIS 2017

Petra Isenberg @dr_pi · 7h

#ieeevis the paper I presented yesterday as an EdWordle - and using fall colors because we like them better right at this time of the year

2 8

VIS 2017

Revisiting Stress Majorization as a Unified Framework for Interactive Constrained Graph Visualization

Yunhai Wang, Yanyan Wang, Yinqi Sun, Lifeng Zhu, Kecheng Lu
Chi-Wing Fu, Michael Sedlmair, Oliver Deussen, and Baoquan Chen

IEEE Transactions on Visualization and Computer Graphics
(Proc. IEEE InfoVis 2017)

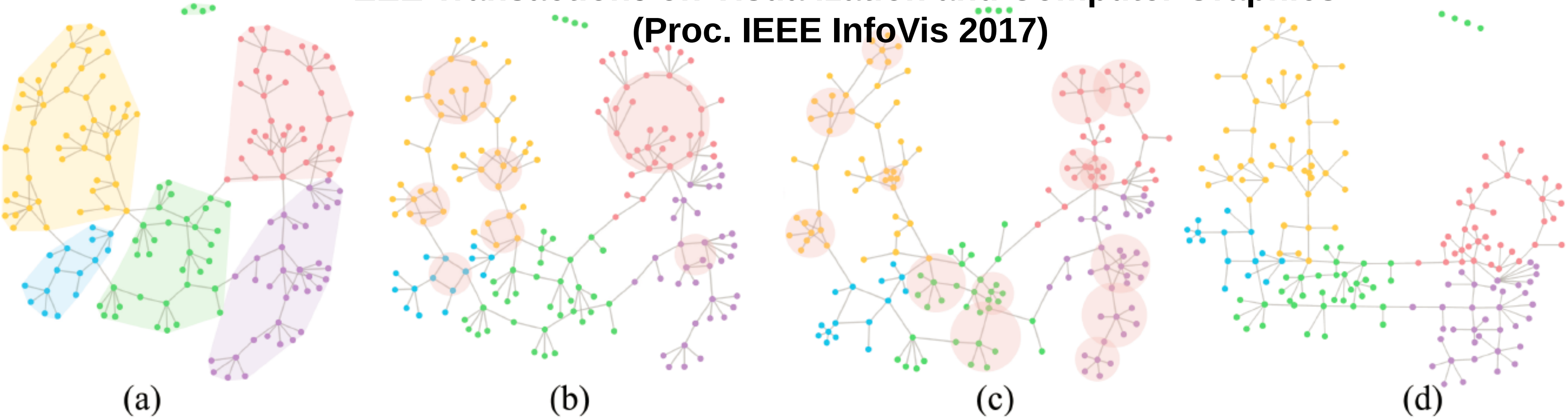


Fig. 1. Our unified framework for constrained graph visualization allows us to create graph layouts with various constraints: a) cluster non-overlap (CN); b) CN + circle constraint (CC); c) CN + star constraint (SC); and d) CN + CC + SC + edge direction constraint.

Automated Visual Interaction

VIS 2017

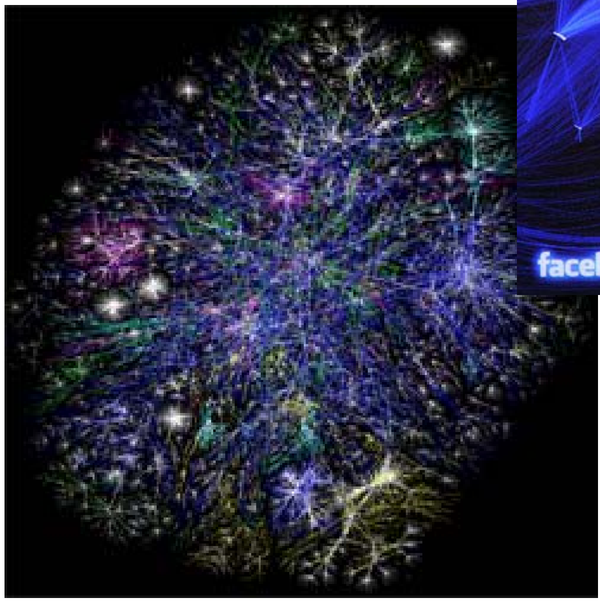
[1] Shandong University [2] Southeast University [3] The Chinese University of HongKong
[4] University of Vienna [5] Konstanz University and VCC SIAT

Revisiting Stress Majorization as a Unified Framework for Interactive Constrained Graph Visualization

Yunhai Wang[1], Yanyan Wang[1], Yinqi Sun[1], Lifeng Zhu[2], Kecheng Lu[1]

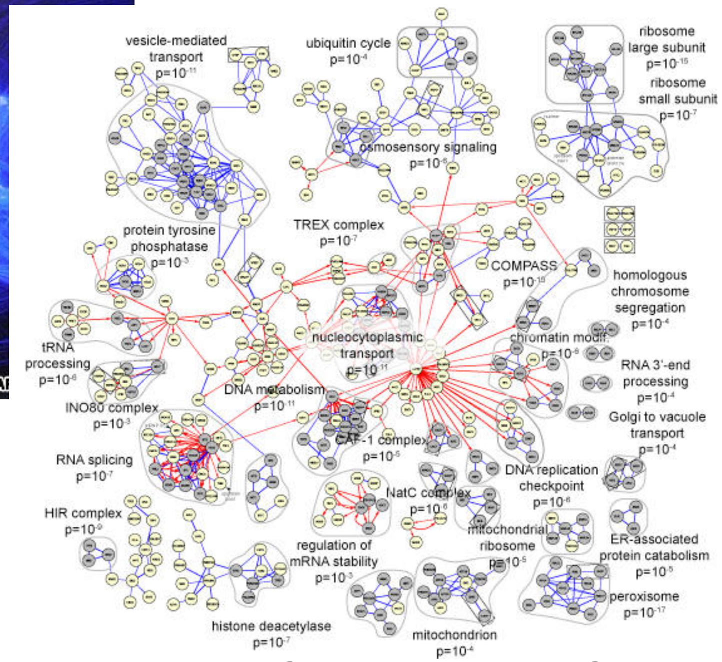
Chi-Wing Fu[3], Michael Sedlmair[4], Oliver Deussen[5], and Baoquan Chen[1]

Graphs are everywhere!



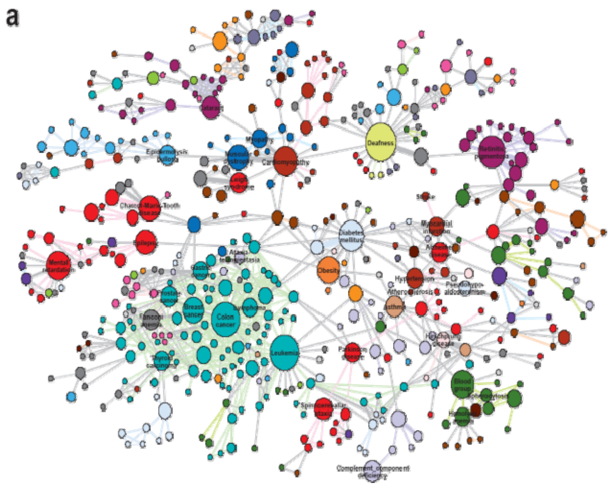
The Internet [2005]

Social Networks [2010]



Gene Regulatory Graph

[Decourty 2008]

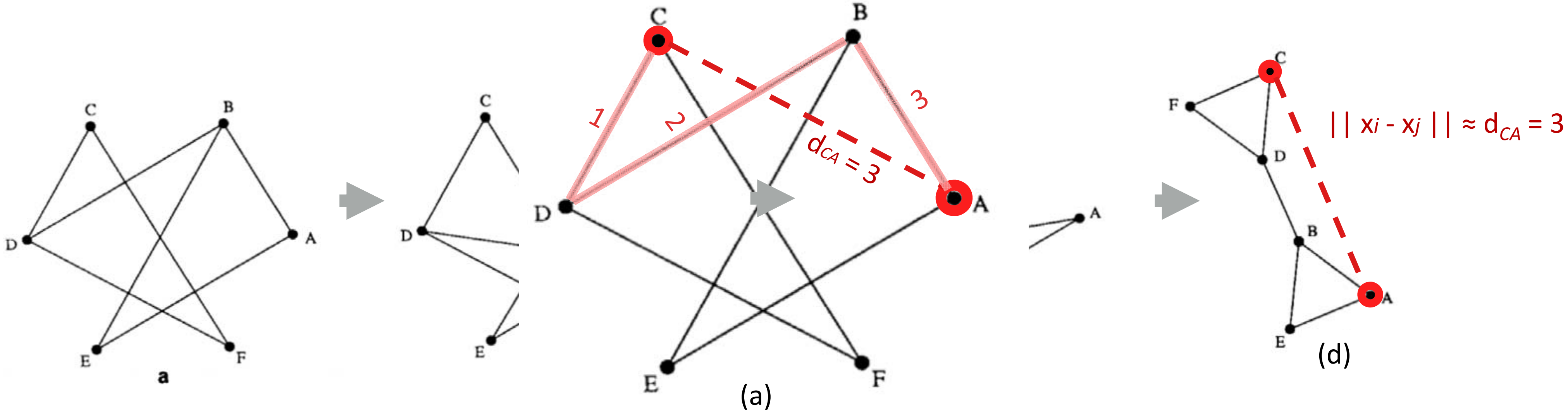


Human Disease Network

[Barabasi 2007]

Node-link Graph Visualization

- Stress model: $$\min \sum_{i,j \in V} w_{ij}(\| x_i - x_j \| - d_{ij})^2$$



The energy minimization process

d_{ij} is the shortest path

[Kamada et al. 1989]

[Kamada et al. 1989]

Task-driven Graph Layout: User constraints

Design various constraints to help explore the structures of interest.

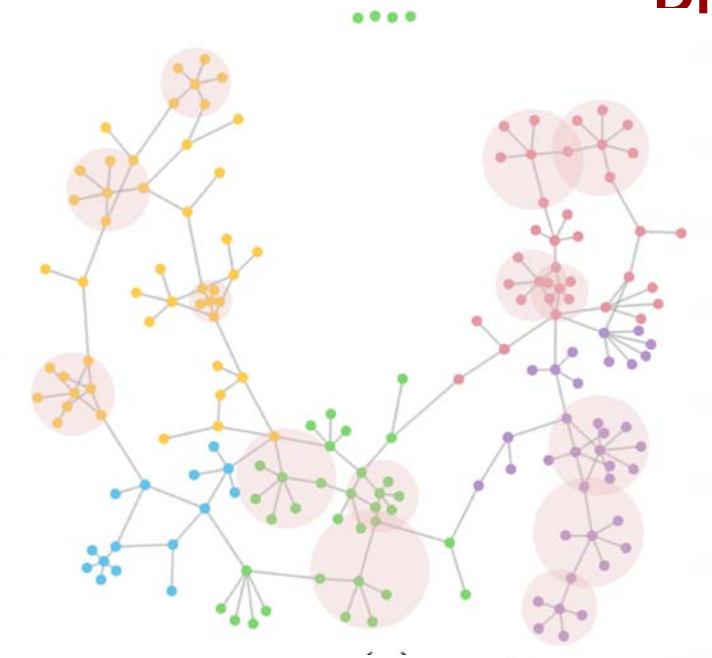
Directed Edges

all above

Aesthetic Criteria:

- Minimizing cluster overlapping
- Maximizing the minimum angle between edges leaving a node (star)
- Circles are salient structures
- Maximizing edge orthogonality
- Minimizing edge crossing
- Maximizing symmetry
- ……

[H. C. Purchase et al. 2002]



Clusters

Circles

Stars

Constrained Stress Model Layout

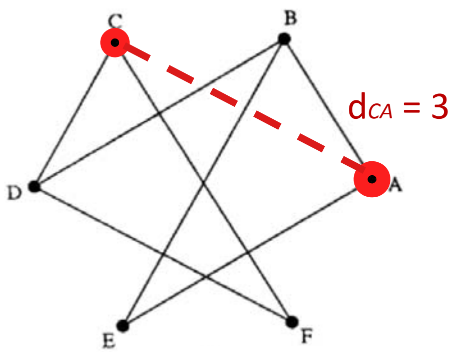
Layout produced from our framework

Core Idea

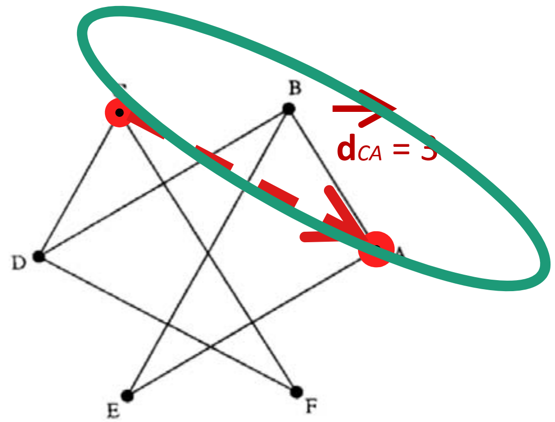
Introducing **edge vector** in the "stress model" (reformulating stress majorization) to provide **inherent (unified)** support to general layout constraints.

» Classical stress model: distance only

» We: **distance** + **angle** (i.e. **vector**)



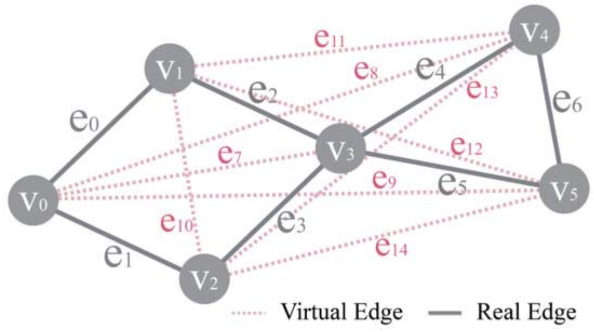
Stress moc **scalar**



Ours: **tor**

Stress Model

- Given a graph G(V,E)



- Stress model:

$$\min S(\mathbf{X}) = \min \sum_{i<j} w_{ij}(\| \mathbf{x}_i - \mathbf{x}_j \| - d_{ij})^2$$

Stress Majorization

$$S(\mathbf{X}) = \sum_{i<j} w_{ij}\, d_{ij}^{\,2} + \sum_{i<j} w_{ij} \,\|\, \mathbf{x}_i - \mathbf{x}_j \,\|^2 - 2 \sum_{i<j} w_{ij} d_{ij} \|\, \mathbf{x}_i - \mathbf{x}_j \,\|$$

$\Downarrow$

$$S(\mathbf{X}) \leq \sum_{i<j} w_{ij}\, d_{ij}^{\,2} + \sum_{i<j} w_{ij} \,\|\, \mathbf{x}_i - \mathbf{x}_j \,\|^2 - 2 \sum_{i<j} w_{ij} d_{ij} \frac{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{z}_i - \mathbf{z}_j)}{\|\, \mathbf{z}_i - \mathbf{z}_j \,\|}$$

Reformulation in Vector Form

$$S(\mathbf{X}) = \sum_{i<j} w_{ij}\, d_{ij}^{\,2} + \sum_{i<j} w_{ij} \| \mathbf{x}_i - \mathbf{x}_j \|^2 - 2 \sum_{i<j} w_{ij} d_{ij} \| \mathbf{x}_i - \mathbf{x}_j \|$$

$$\mathbf{d}_{ij} = d_{ij} \frac{(\mathbf{z}_i - \mathbf{z}_j)}{\| \mathbf{z}_i - \mathbf{z}_j \|}$$

$$S(\mathbf{X}) \leq \sum_{i<j} w_{ij}\, d_{ij}^{\,2} + \sum_{i<j} w_{ij} \| \mathbf{x}_i - \mathbf{x}_j \|^2 - 2 \sum_{i<j} w_{ij} d_{ij} \frac{(\mathbf{x}_i - \mathbf{x}_j)^T (\mathbf{z}_i - \mathbf{z}_j)}{\| \mathbf{z}_i - \mathbf{z}_j \|}$$

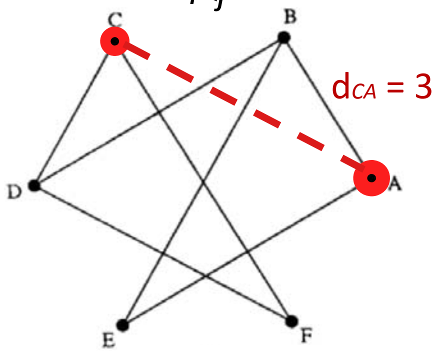
Edge vectors

Reformulation in Vector Form

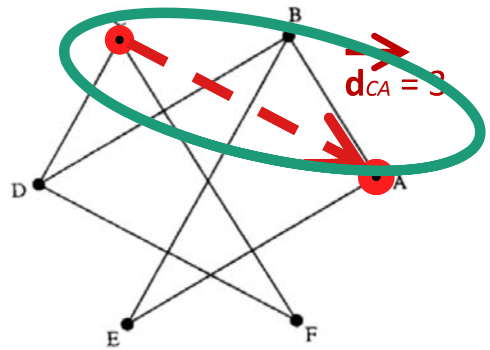
$$S(\mathbf{X}) \leq \sum_{i<j} w_{ij}\, d_{ij}^{\,2} + \sum_{i<j} w_{ij} \| \mathbf{x}_i - \mathbf{x}_j \|^2 - 2\sum_{i<j} w_{ij}\, (\mathbf{x}_i - \mathbf{x}_j)^{\mathrm{T}} \mathbf{d}_{ij} \quad \frac{(\mathbf{x}_i - \mathbf{x}_j)^{\mathrm{T}}(\mathbf{z}_i - \mathbf{z}_j)}{\| \mathbf{z}_i - \mathbf{z}_j \|}$$

$$= \sum_{i<j} w_{ij} \| \mathbf{x}_i - \mathbf{x}_j - \mathbf{d}_{ij} \|^2 \qquad \mathbf{d}_{ij} = d_{ij} \frac{(\mathbf{z}_i - \mathbf{z}_j)}{\| \mathbf{z}_i - \mathbf{z}_j \|}$$

Stress model: $\sum_{i<j} w_{ij} (\| \mathbf{x}_i - \mathbf{x}_j \| - d_{ij})^2$



Stress moc scalar



Ours: tor

Incorporating User Constraints

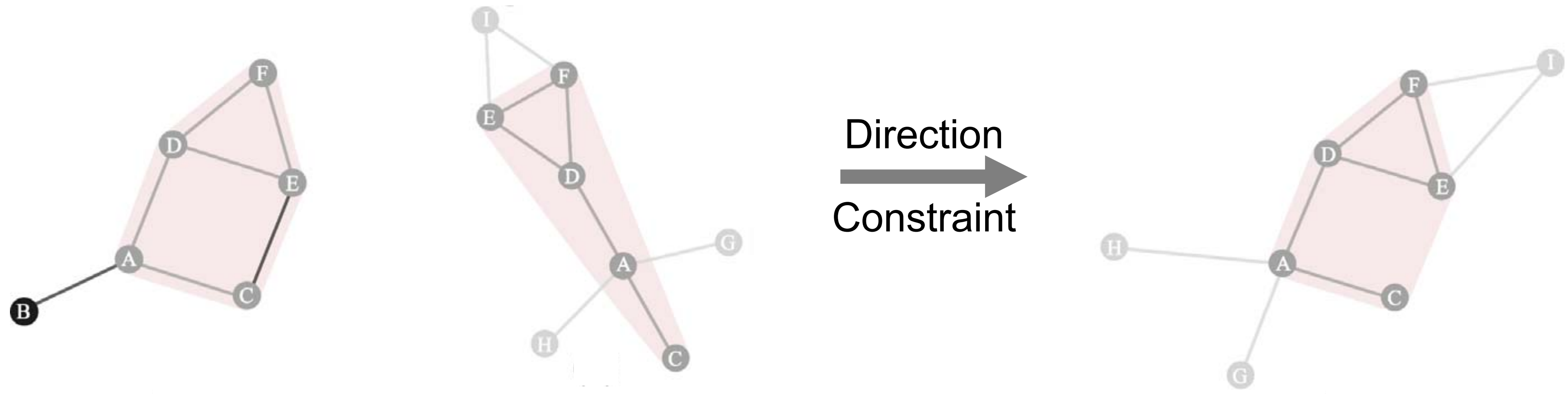
- Define **d'**$_{ij}$ — target edge vectors for constraints
  - $\|\mathbf{d'}_{ij}\|$ **distance** constraints between node pairs
  - $\vec{\mathbf{d'}}_{ij}$ **direction** constraints between node pairs

User Constraints

1. Direct Constraint
2. Metrics-based Constraint
3. Shape-based Constraint

1. Direct Constraint

- Edge **Length** / **Direction** Constraint
  - Temporal Coherent Dynamic Graph Visualization

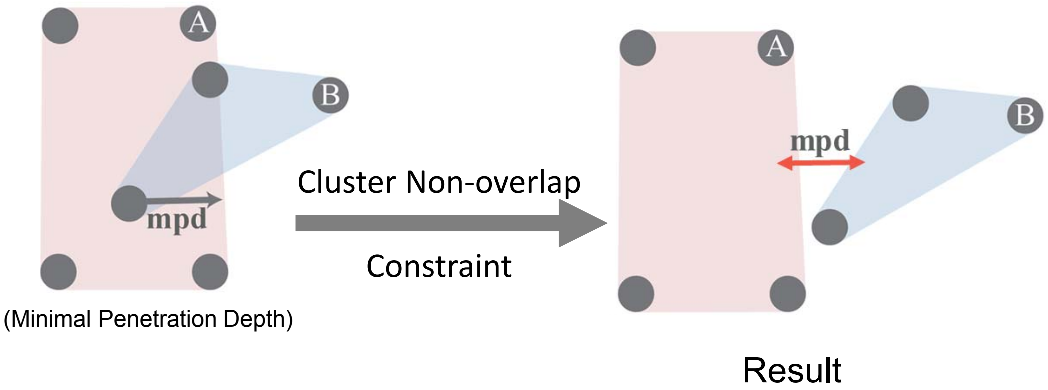


Layout at time t

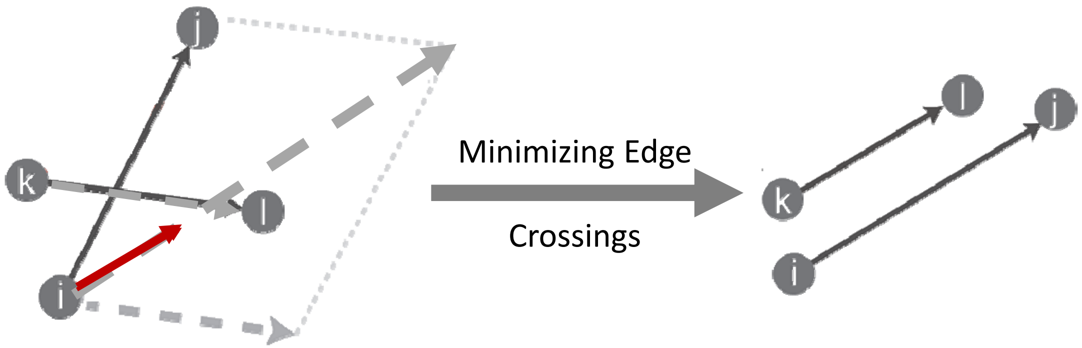
(Unconstrained)
Layout at time t+1

(Constrained)
Layout at time t+1

2. Metrics-based Constraint

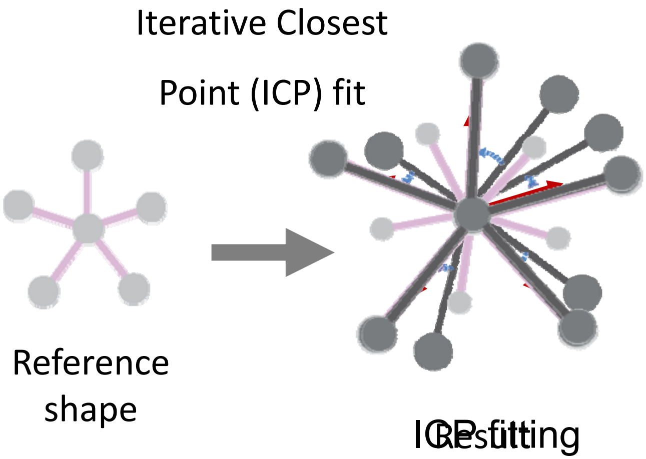


Non-overlap Constraint

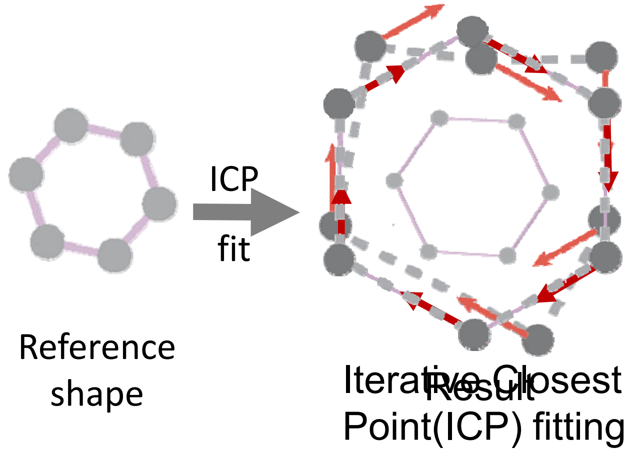


Minimizing Edge Crossings

3. Shape-based Constraint

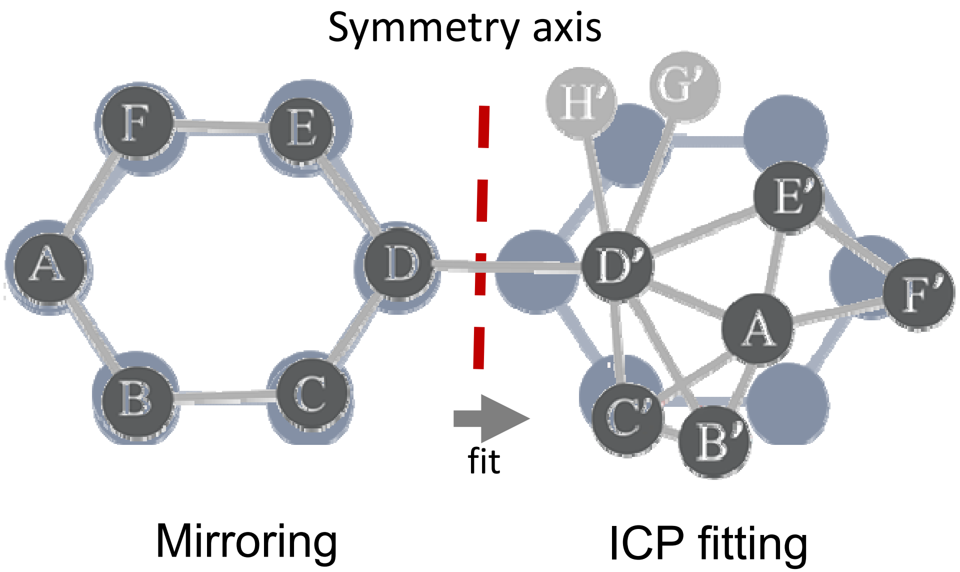


Star Constraint



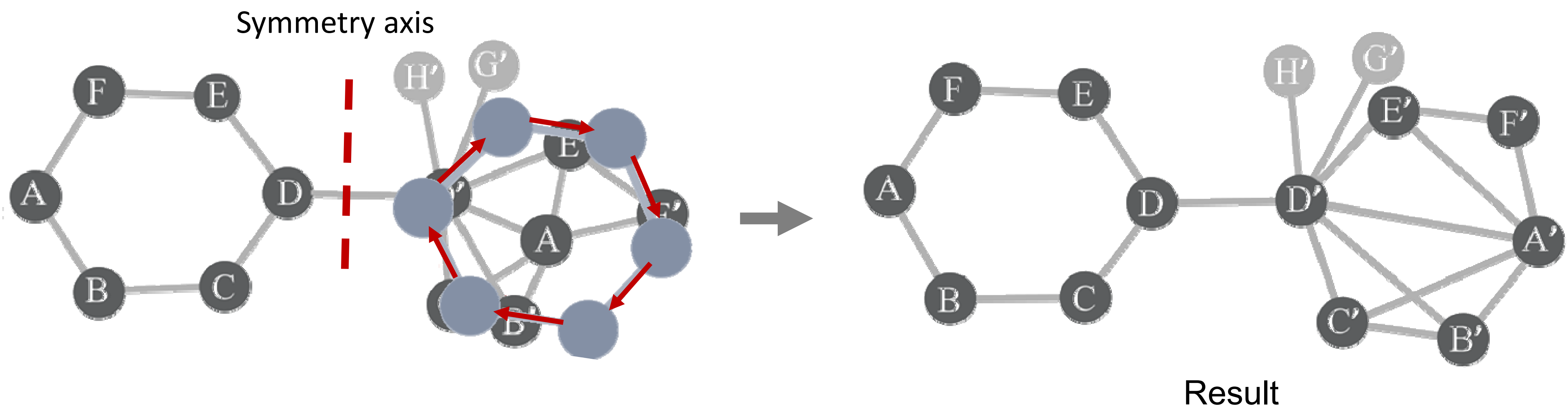
Circle Constraint

3. Shape-based Constraint



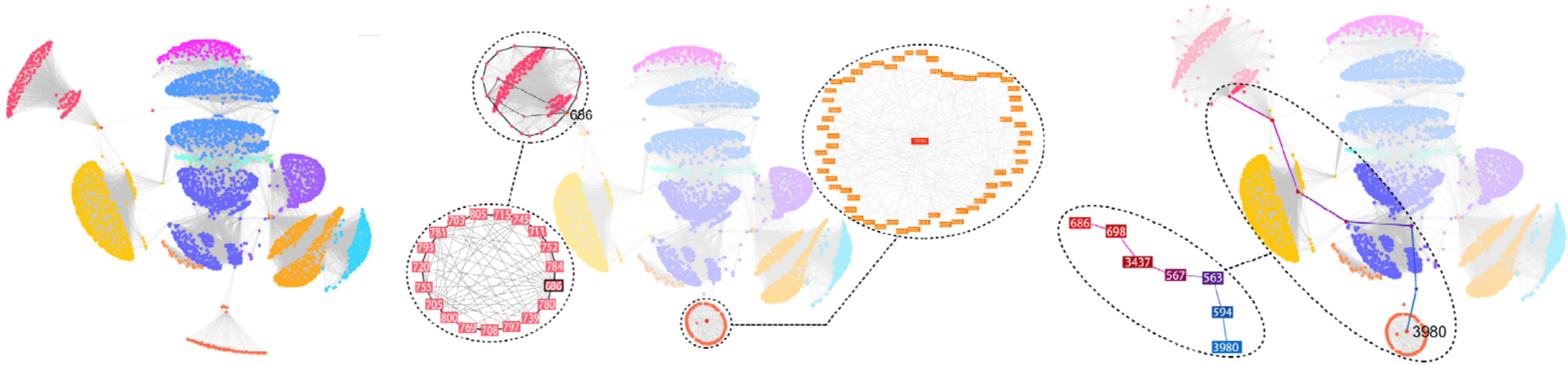
Symmetry Constraint

3. Shape-based Constraint



Symmetry Constraint

Constraint-based Graph Exploration



(a) Cluster non-overlap constraints

+ Circle constraints
star constraints

(b)

+ Path constraints

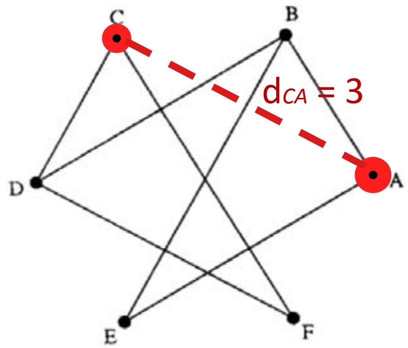
(c)

Facebook4039 Graph

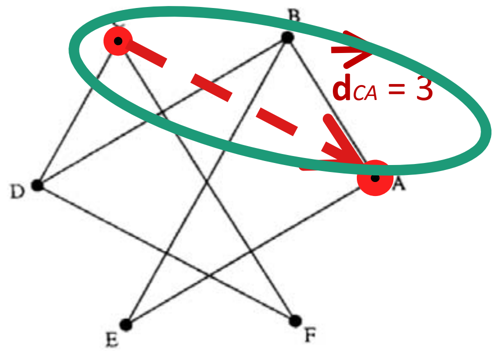
4039 nodes, 88234 edges

More details can be found from:
http://www.yunhaiwang.org/infovis17/vectormds/index.html

Code, system, and data are available at:
https://github.com/Yanyan-Wang/vectorized_stress_majorization

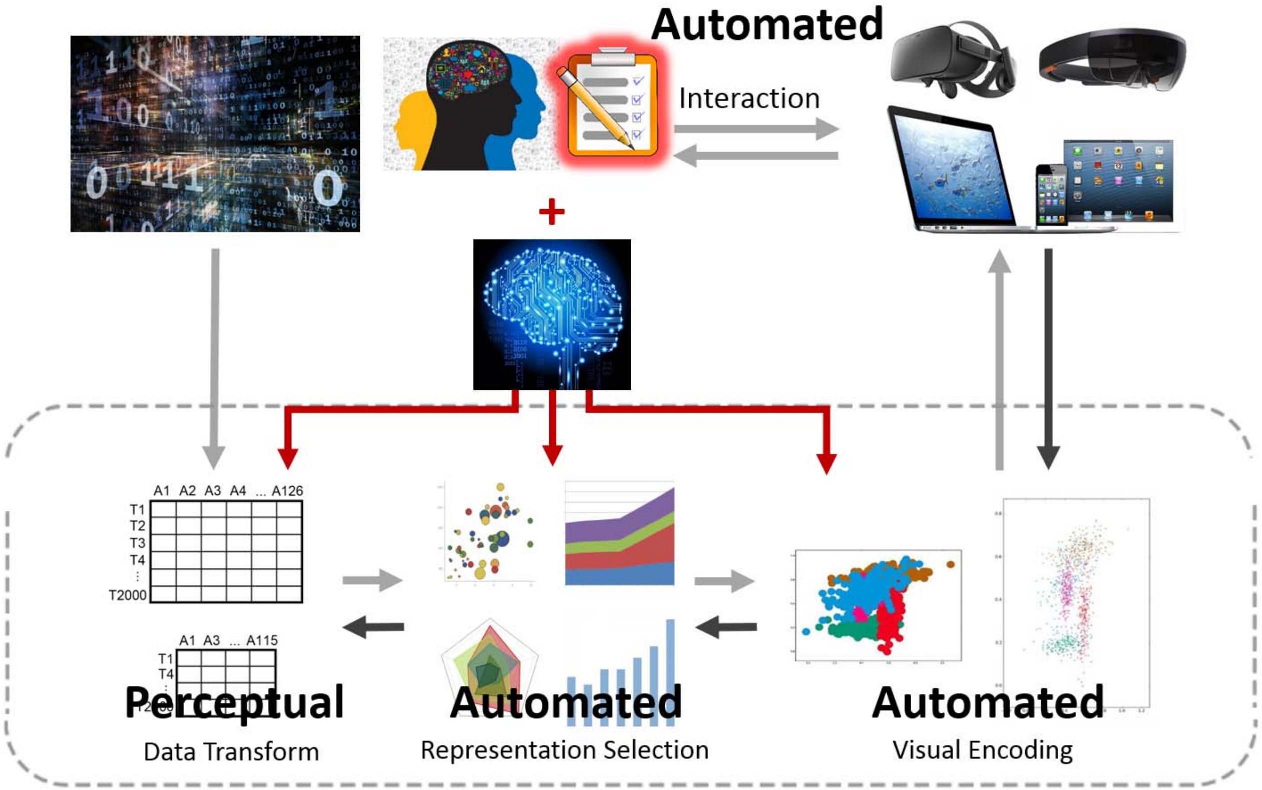


Stress moc **scalar**



Ours: **tor**

Thanks!



http://www. yunhaiwang.org/