从小数据到大数据：
可视化研究的选题经验探讨

汪云海

What is visualization research?

Jessica Hullman

Mar 13, 2019 · 13 min read · Listen

The purpose of visualization is insight, not pictures: An interview with visualization pioneer Ben Shneiderman

Few people in visualization research have had careers as long and as impactful as Ben Shneiderman. We caught up with Ben over email in between his travels to get his take on visualization research, what's worked in his career, and his advice for practitioners and researchers. Enjoy!

Multiple Views: *One of the main purposes of this blog is to explain to people what visualization research is to practiti[...]ssibly, laypeople. How would you answer the question "what is visualization research"?*

https://medium.com/multiple-views-visualization-research-explained/the-purpose-of-visualization-is-insight-not-pictures-an-interview-with-visualization-pioneer-ben-beb15b2d8e9b

The visualization research goal

Ben S:

Ben S: First let me define information visualization and its goals, then I can describe visualization research.

Information visualization is a powerful interactive strategy for exploring data, especially when combined with statistical methods. Analysts in every field can use interactive information visualization tools for:

- more effective detection of faulty data, missing data, unusual distributions, and anomalies
- deeper and more thorough data analyses that produce profounder insights, and
- richer understandings that enable researchers to ask bolder questions.

Like a telescope or microscope that increases your perceptual abilities, information visualization amplifies your cognitive abilities to understand complex processes so as to support better decisions. In our best moments, information visualization users work on problems that address the grand challenges of our time, such as the UN Sustainable Development Goals.

What is visualization research？

Ben S: Visualization research seeks new visual displays, control panels, features, and workflows that improve the capabilities of users. To accomplish these goals, visualization researchers develop perceptual and cognitive theories that guide design, in concert with developing new tools. Visualization researchers also develop quantitative and qualitative evaluation methods to validate their hypotheses and refine their theories.

https://medium.com/multiple-views-visualization-research-explained/the-purpose-of-visualization-is-insight-not-pictures-an-interview-with-visualization-pioneer-ben-beb15b2d8e9b

My research goal

- Develop new **algorithms**, **tools** and **systems** that enhance people's ability to understand and communicate data


IDEAS Lab

Home People Publications Events Seminars Code & Data

About Us

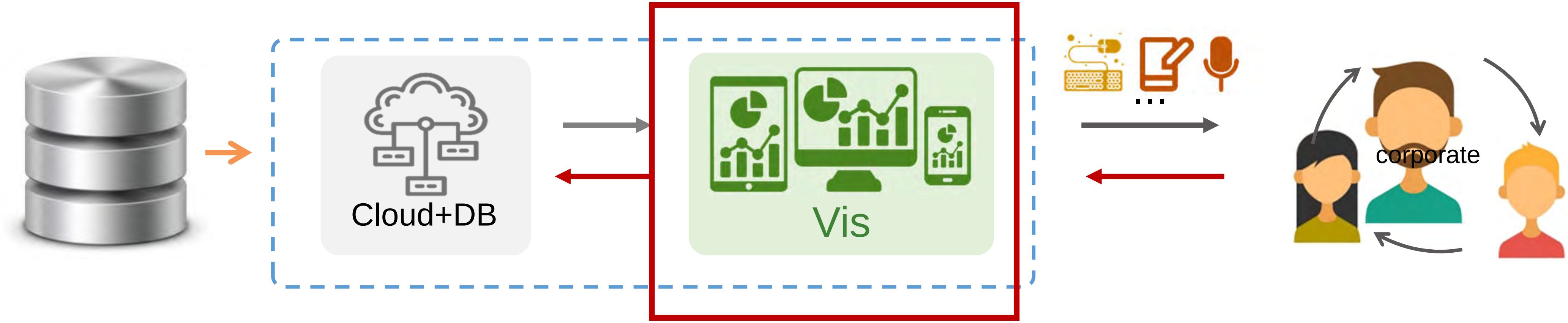
Interactive Data Exploration System (IDEAS) Lab (formerly known as "VisLab") was founded 2016 at Shandong University, Qingdao, China by Prof. Yunhai Wang. Its mission is to enhance people's ability to understand and communicate data through the design of automated visualization and visual analytics systems.

We study the perceptual and mathematical foundations of visualization in order to improve the efficiency of interactive data analysis, while developing interactive systems for data visualization and analysis.

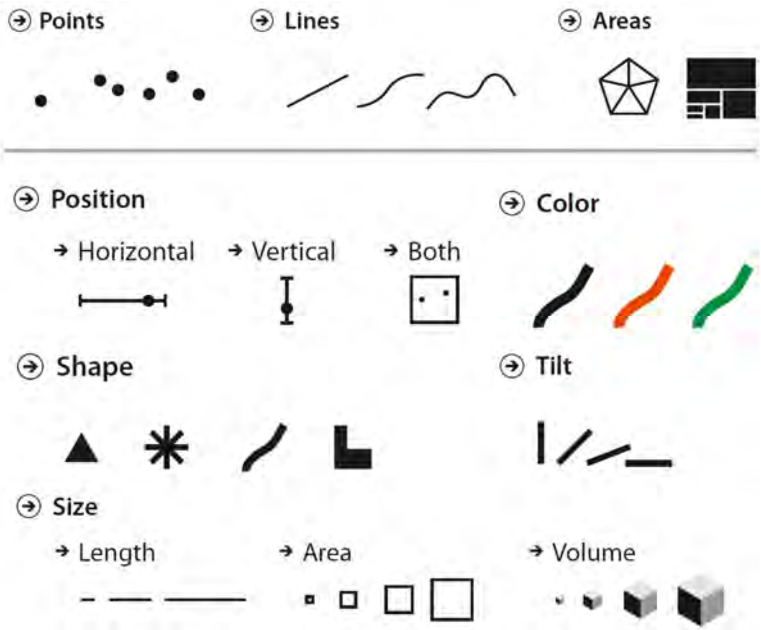
https://ideaslab.wang/

Human Data System

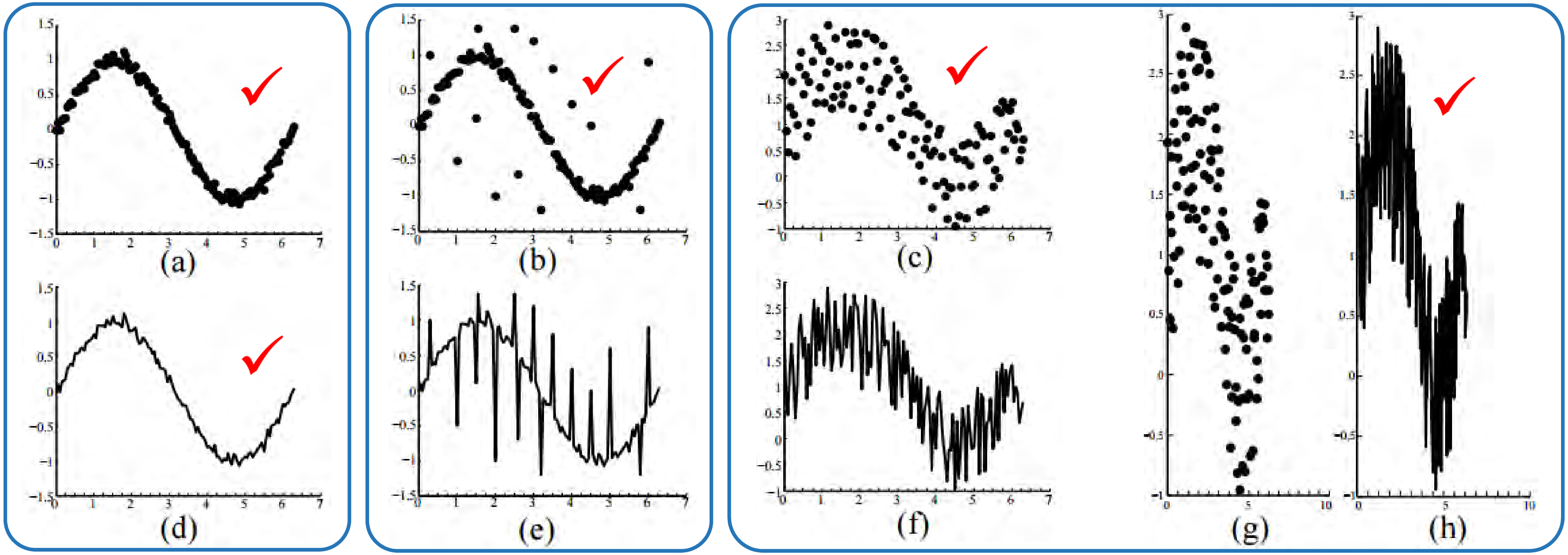
- Develop new **algorithms**, **tools** and **systems** that enhance people's ability to understand and communicate data



Large design space



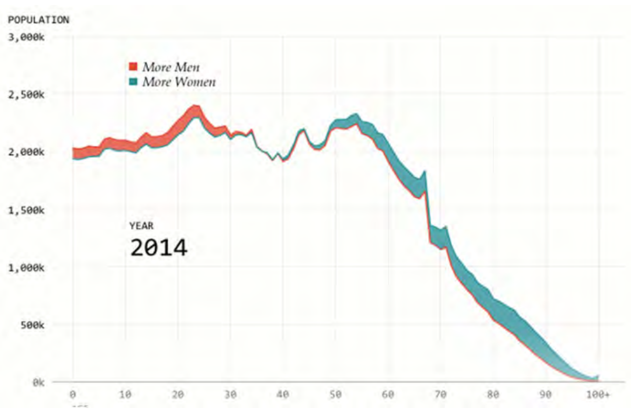
One example:
Selecting right representation for visualizing trends in time series

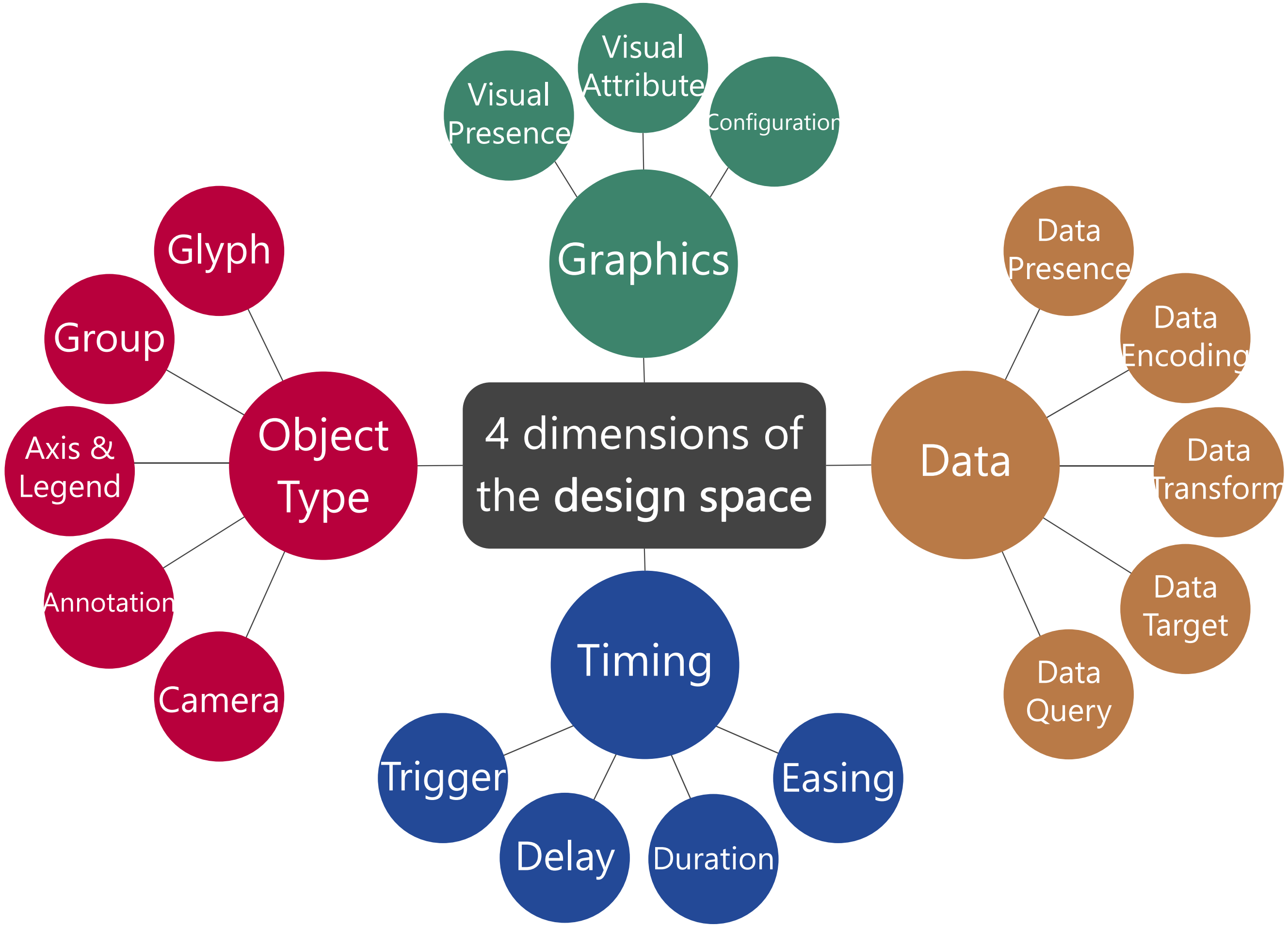


little noise Outlier Heavy noise Aspect ratio

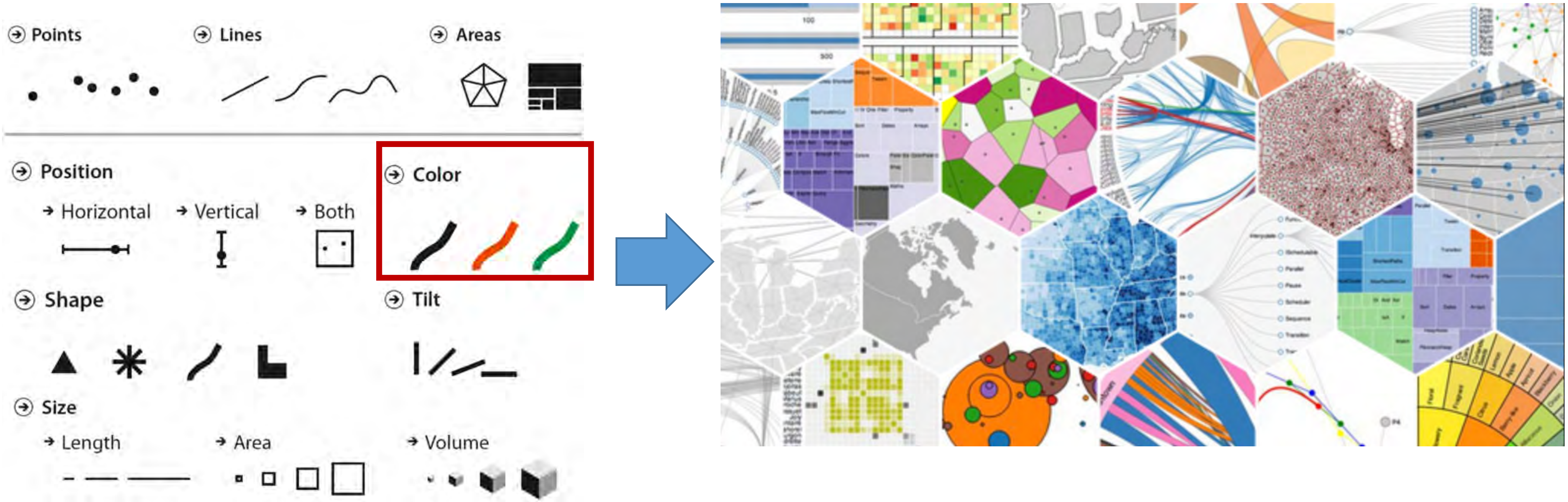
Selecting the right representation for time series visualization
IEEE TVCG 2018 (presented in IEEE InfoVis 2018)

It is even larger for chart animations

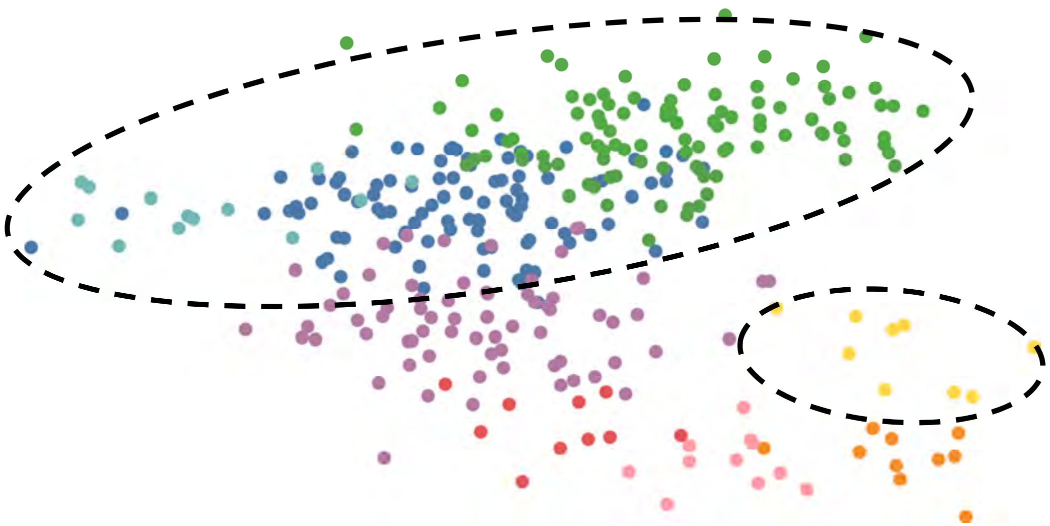




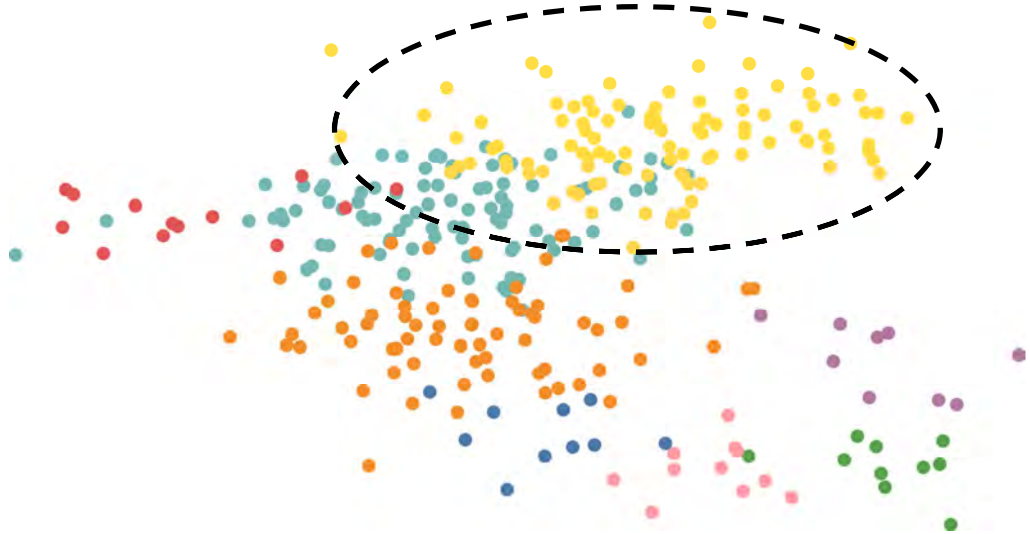
Let’s start from color design



Which one is better?

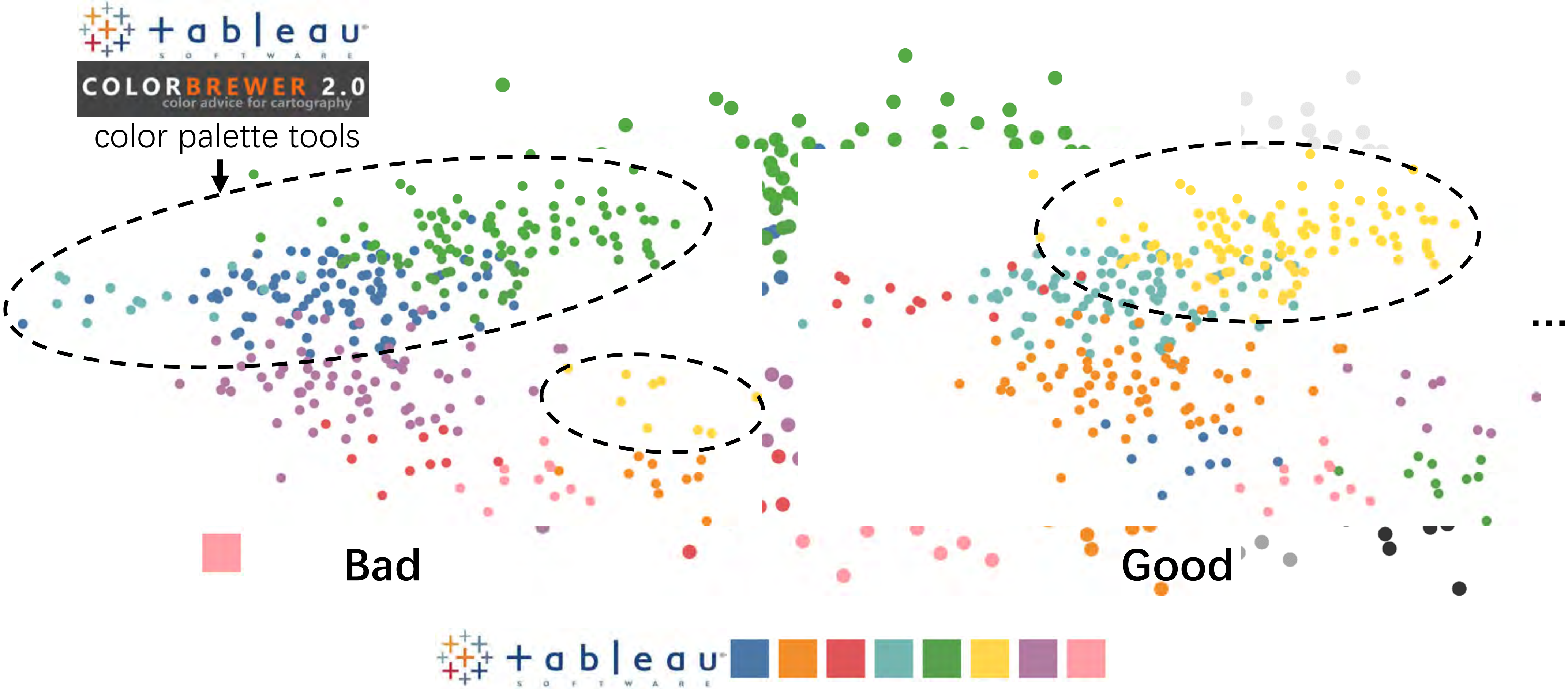


Bad



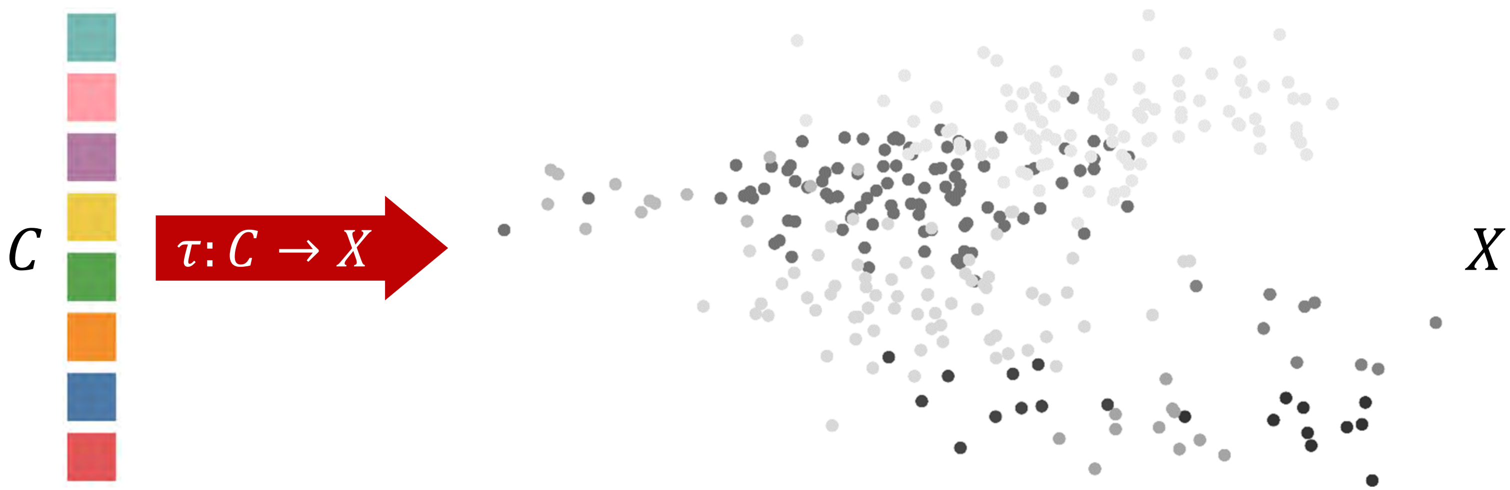
Good

Multiclass scatterplots colorization



Problem formulation

- Given a dataset X of m classes with a color palette C of p colors ($p \geq m$)



- Our goal is to find the best color assignment which can maximizes the visual class separability with respect to human perception.

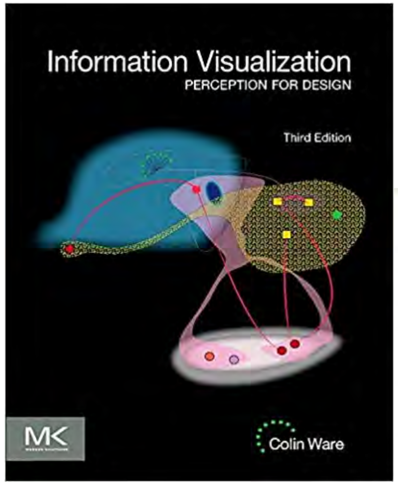
Perceptual factors

For color-coded labeling, human judgments are heavily influenced by two color factors:

- Point color distinctness
- Point color contrast with background

Our separation factors:

- Distinctness
- Contrast with background

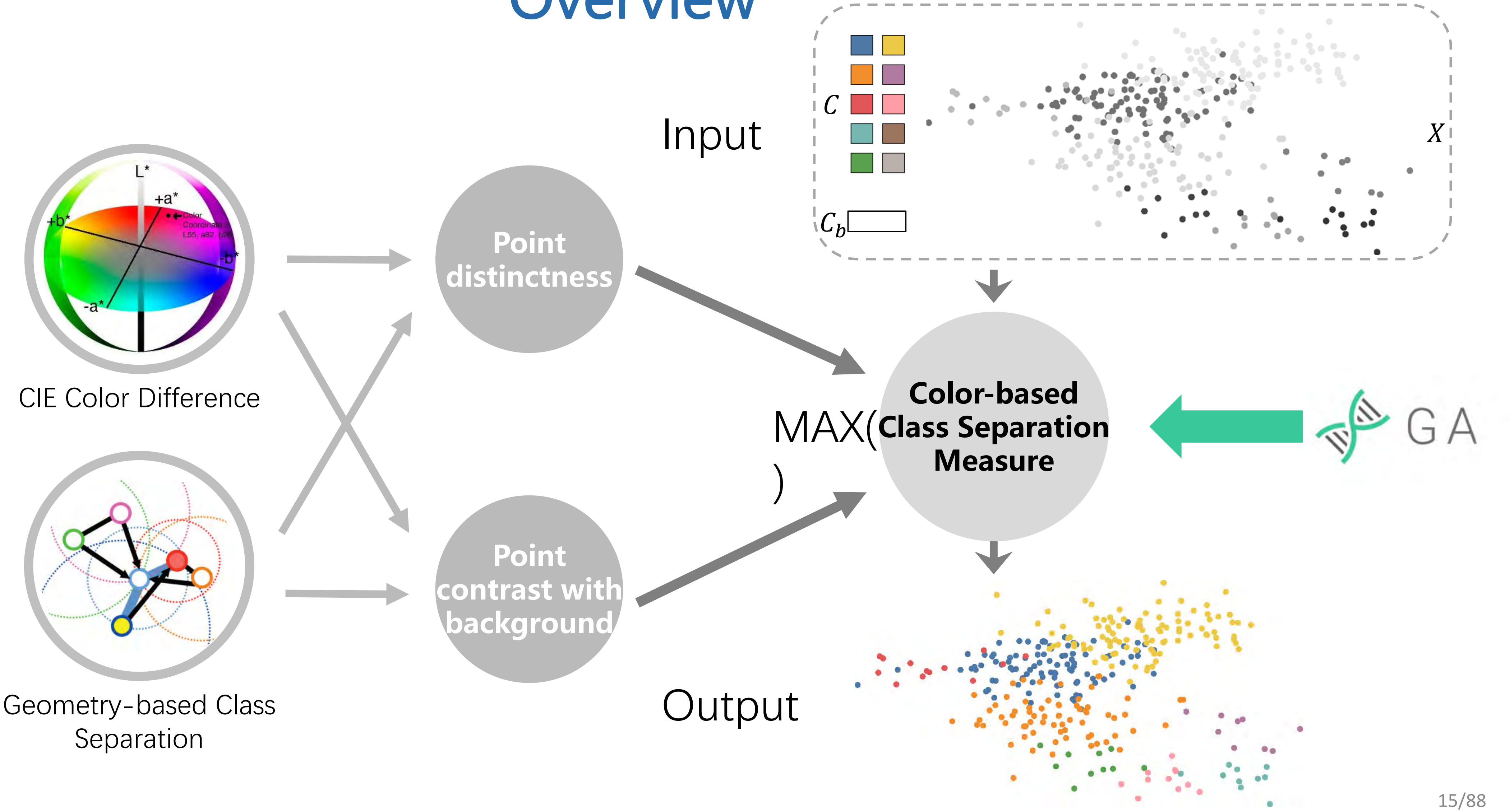


Distinctness. A uniform color space, such as *CIEluv*, can be used to determine the degree of perceived difference between two colors that are placed close together. It might be thought that an algorithm based on *CIEluv* could be used to simply choose a set of colors that are most widely separated, but most color scheme design problems are too complex for this; background colors, symbol sizes, and

Contrast with background. In many displays, color-coded objects can be expected to appear on a variety of backgrounds. Simultaneous contrast with background colors can dramatically alter color appearance, making one color look like another. This is one reason why it is advisable to have only a small set of color codes. A method for reducing contrast effects is to place a thin white or black border around the color-coded object. This device is commonly used with signal lights; for example, train signals are displayed on large black background discs. In addition, we should never display codes using purely chromatic differences with the background. There should be a significant luminance difference in addition to the color difference.

[Ware. 2012, p122-p123]

Overview


Input
C
C_b
X
L*
+a*
+b*
-b*
-a*
CIE Color Difference
Geometry-based Class Separation
Point distinctness
Point contrast with background
MAX(
Color-based Class Separation Measure
)
GA
Output

Palettailor: Discriminable Colorization for Categorical Data

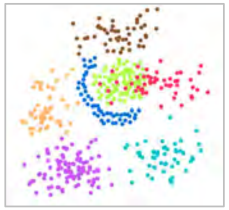
Input Data

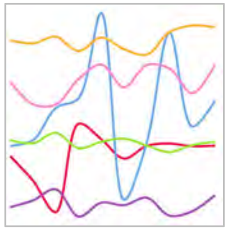
Choosing a palette

Discriminable Automated Colorization
[Wang et al. 2018]

Our Goal

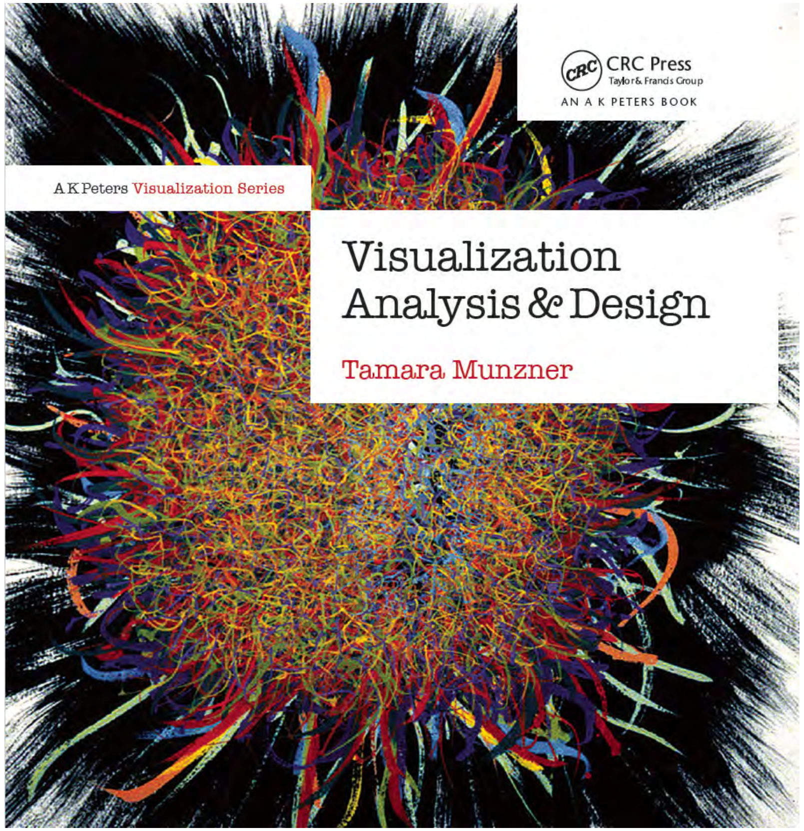
Design a fully-automated approach that creating and assigning color palettes to different visualizations with higher discrimination quality





http://www.palettailor.net

Can we support color highlighting?

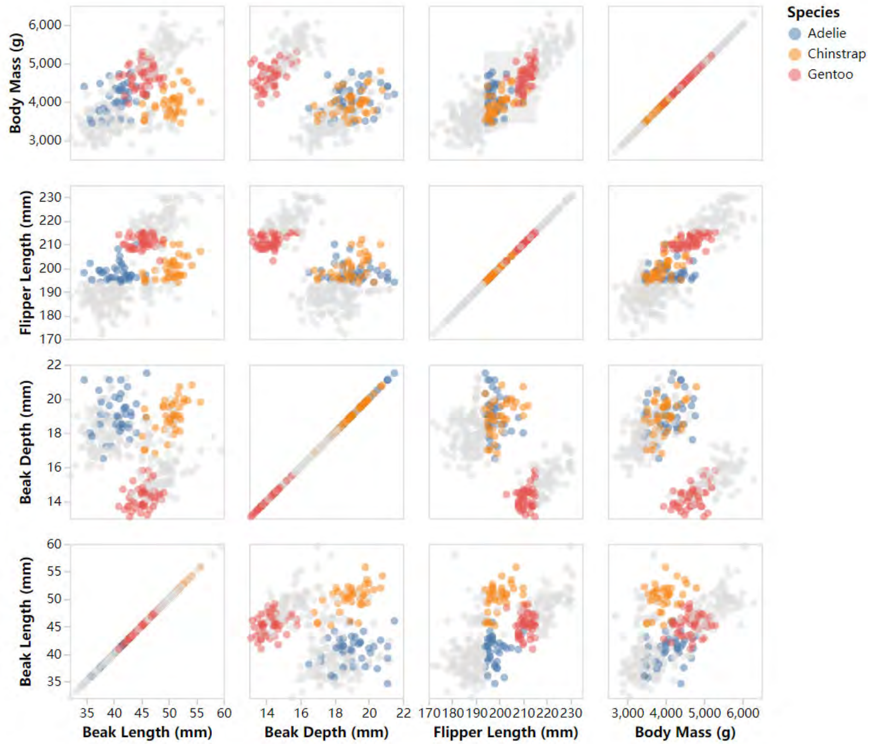


11. Manipulate View

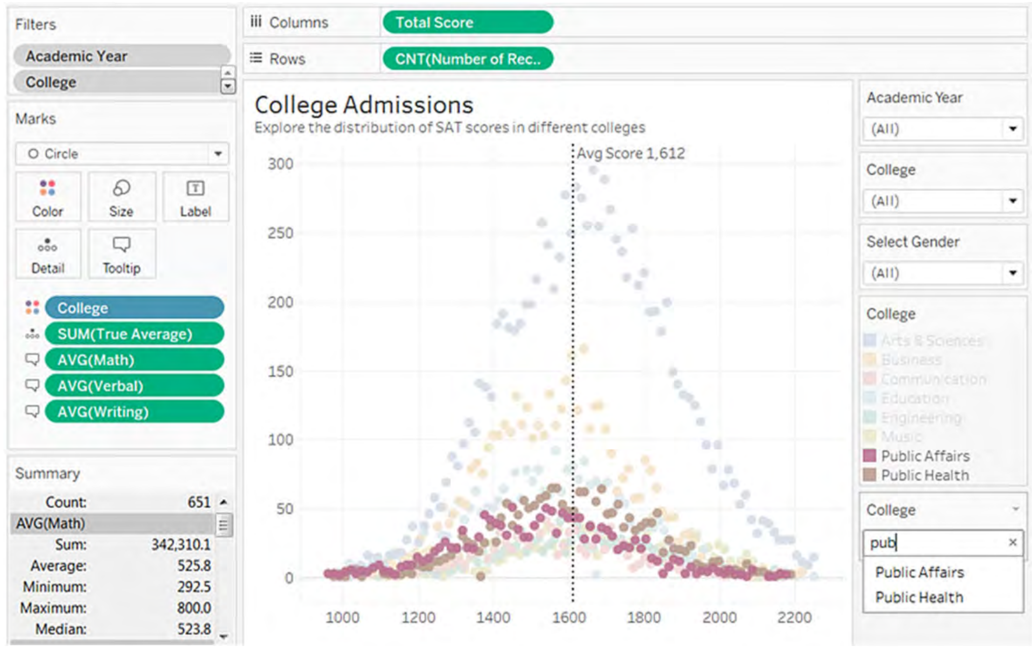
design choice to highlight selected items by changing their color. Another important caveat is that the highlight color should be sufficiently different from the other colors that a visual popout effect is achieved with sufficient hue, luminance, or saturation contrast.

A fundamental limitation of highlighting by color change is that the existing color coding is temporarily hidden. For some abstract tasks, that limitation constitutes a major problem. An alternative design choice that preserves color coding is to highlight with an outline. You could either add an outline mark around the selected object or change the color of an existing outline mark to the highlight color. It's safest to highlight the items individually rather than to assume the selection set is a spatially contiguous group, unless you've built that selection constraint into the vis tool.

Color design for Interactive Visualizations

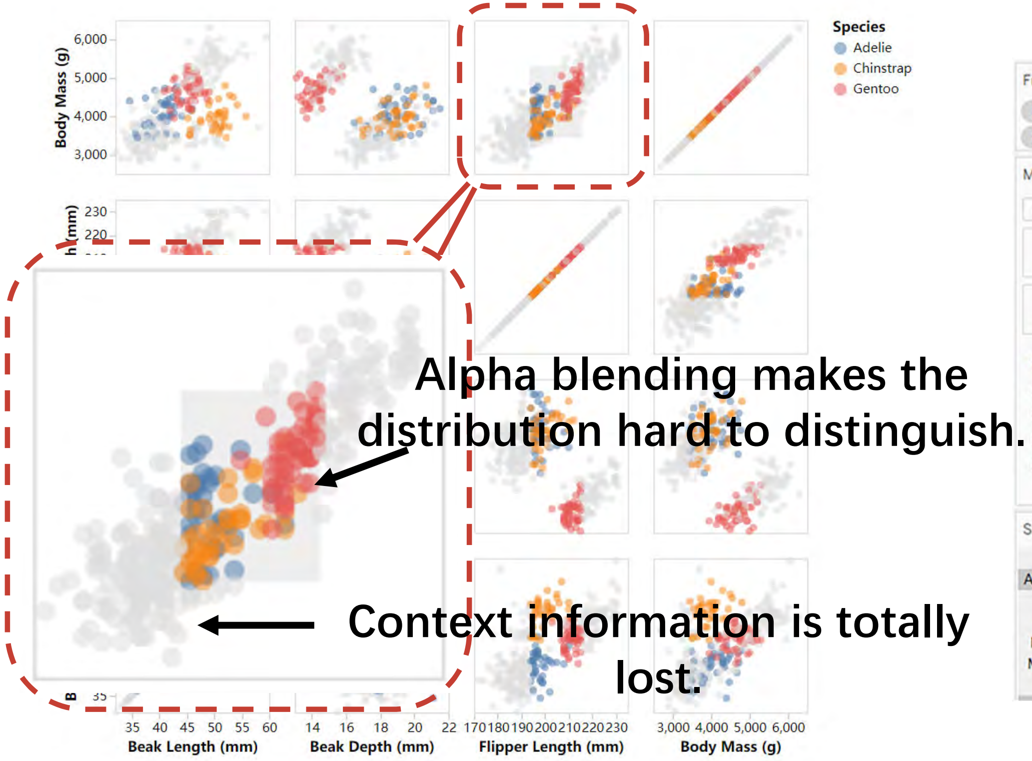


Brushing Scatterplot Plots
Example from Vega

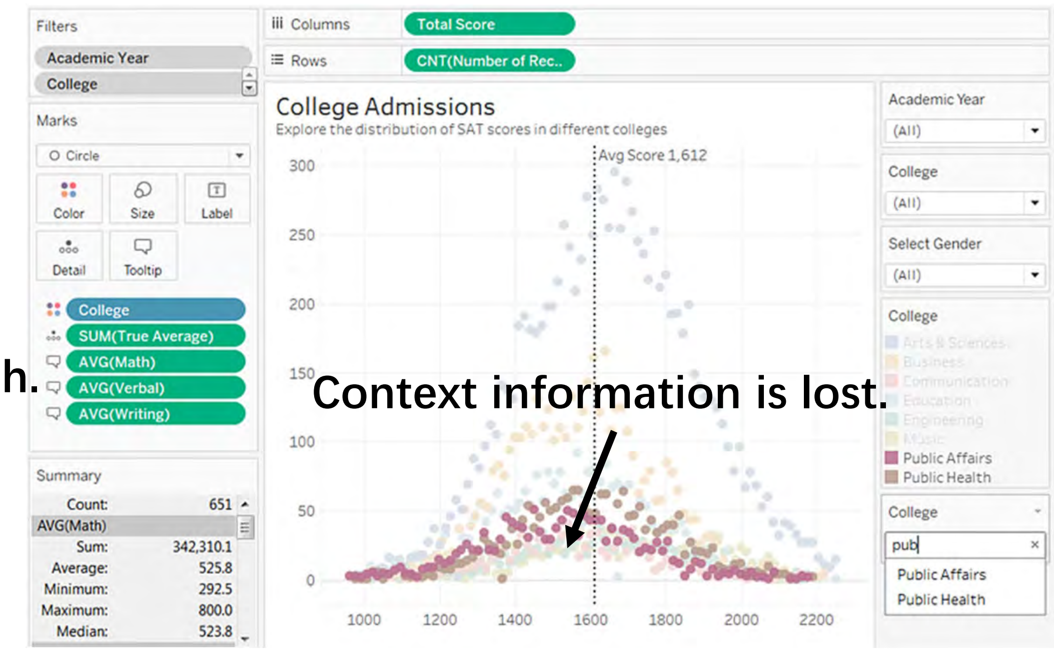


Highlighting Data Points in Context
by Tableau Highlighter

Color design for Interactive Visualizations



Brushing Scatterplot Plots
Example from Vega



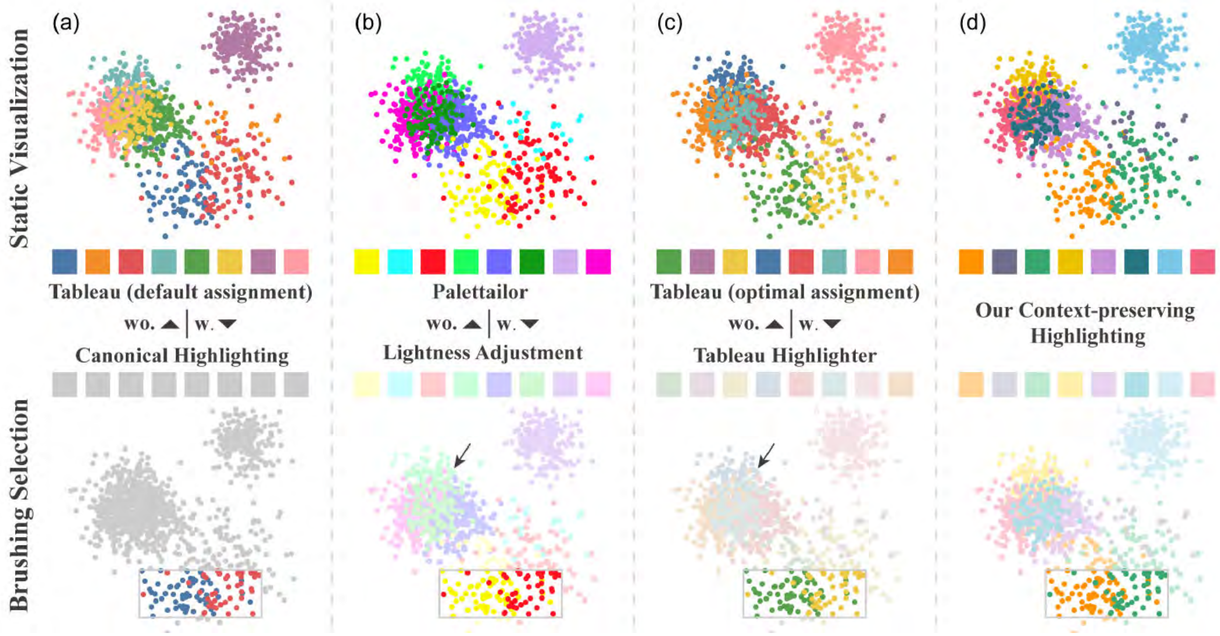
Highlighting Data Points in Context
by Tableau Highlighter

Design requirements

- **DR1:** highlighting the selected data points as much as possible to deliberately attract user attention;
- **DR2:** maximizing the visual discrimination between classes for efficiently exploring the data, for the selected and non-selected classes; and
- **DR3:** maintaining color consistency for data points when they are dynamically highlighted or de-emphasized.

How to solve this issue?

- Dynamically combined two pre-created palettes: one for highlighting and the other for delighting



Submission ID: 6939

This video has voice over

Interactive Context-Preserving Color Highlighting for Multiclass Scatterplots

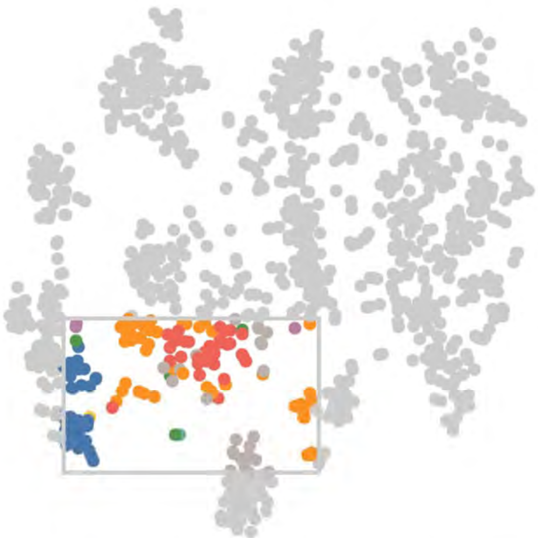


Tableau (default assignment) with Canonical Highlighting

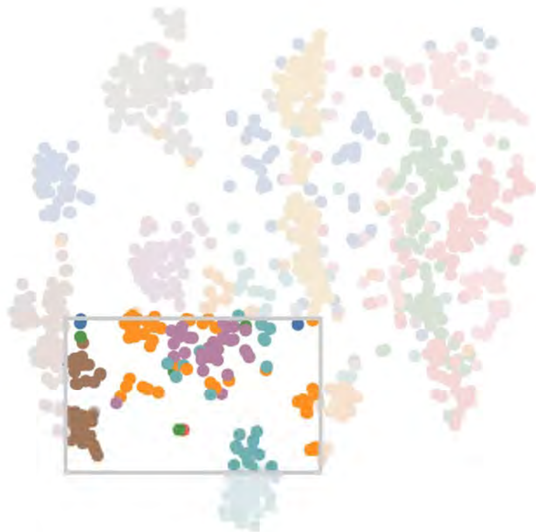
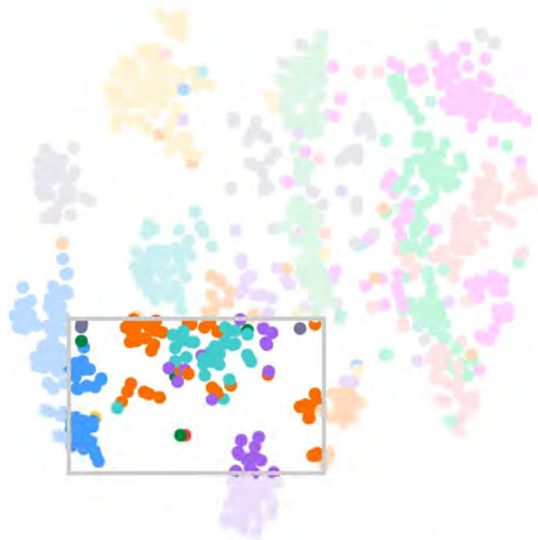
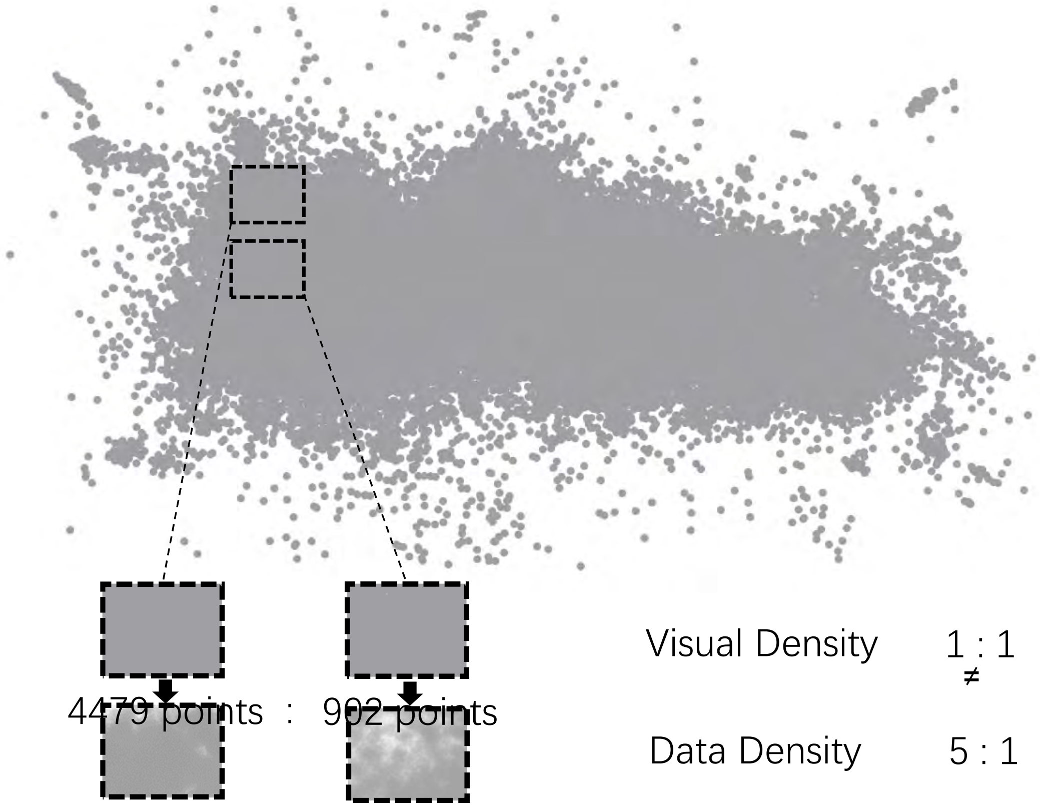


Tableau (optimal assignment) with Tableau Highlighter



Our Context-preserving Highlighting

If the number of data points is too large?

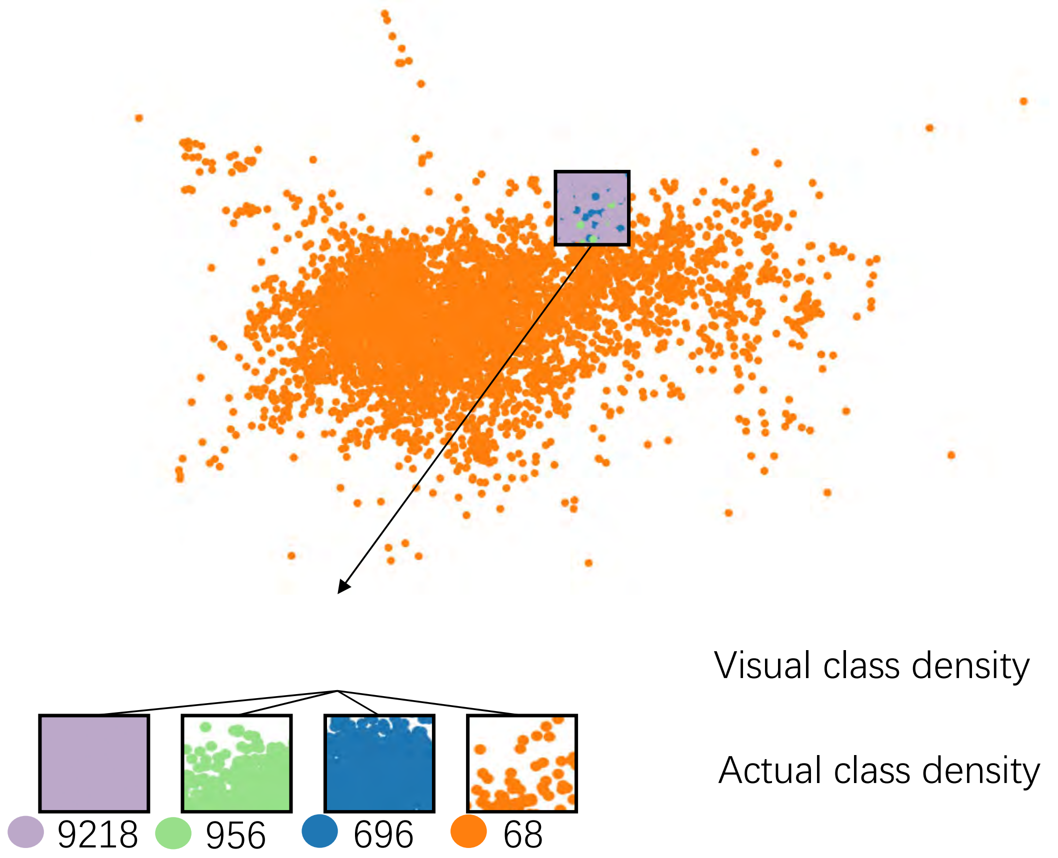


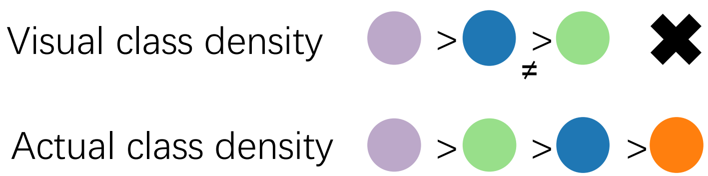
distorts relative data density

Overdraw problem for multi-class data

~100k points ➡ 1600 x 900 pixels

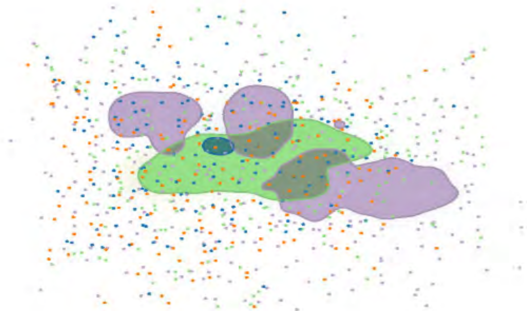
Distorts relative class density





Overdraw reduction for scatterplots

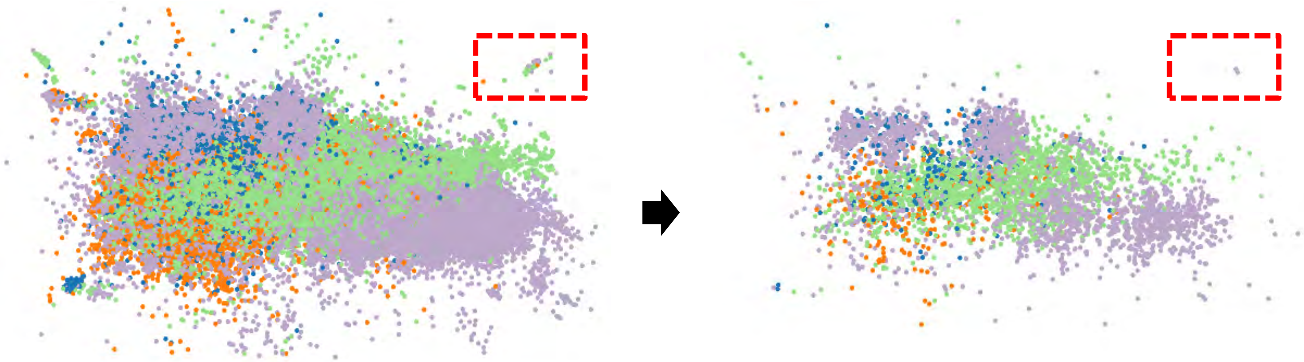
[Micallef et al. 2017]
Appearance optimization



[Mayorga and Gleicher 2013]
Density Estimation

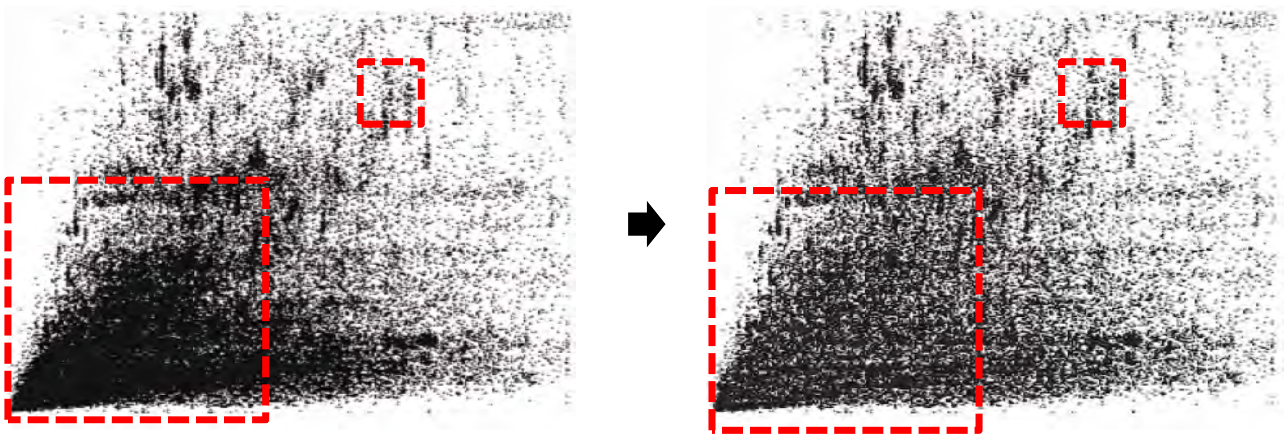
Visual Sampling

Visual sampling



Missing outliers

Random sampling



Without multi-class information

Keeping outliers

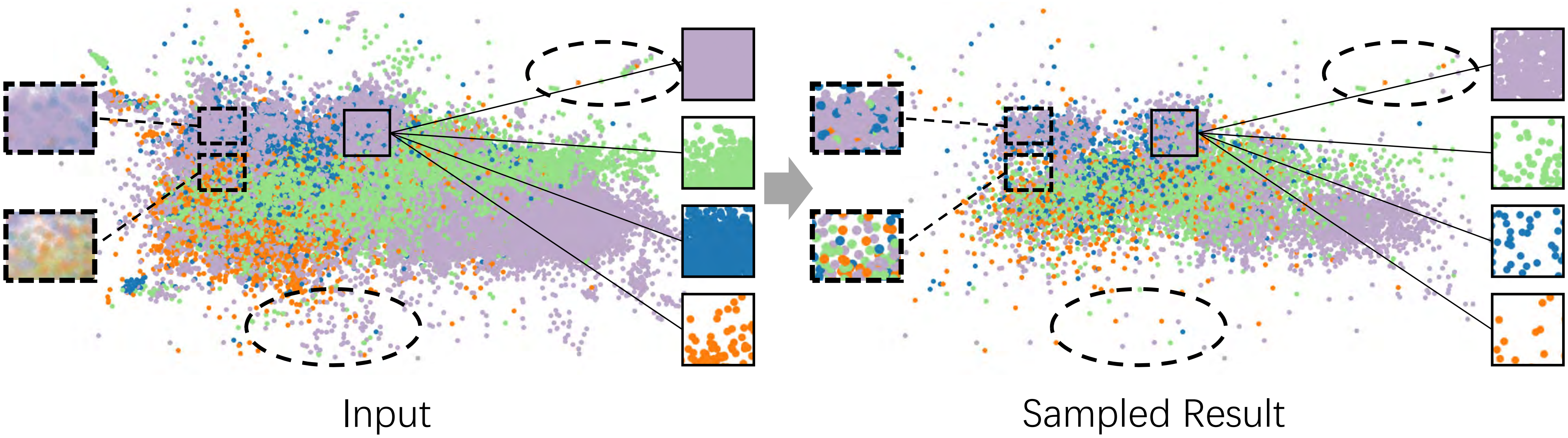
Distorting relative densities

Non-uniform sampling [Bertini and Santucci 2004]

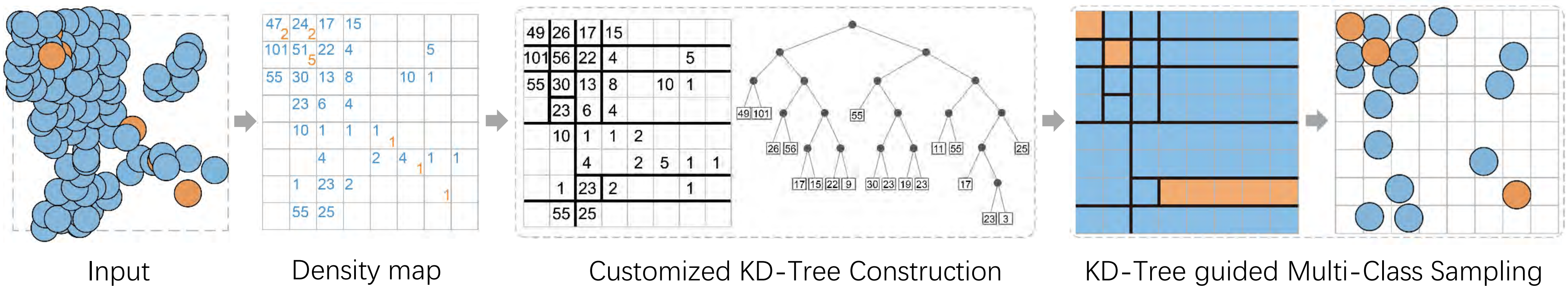
Our goal

Design a perception-driven sampling method which allows to

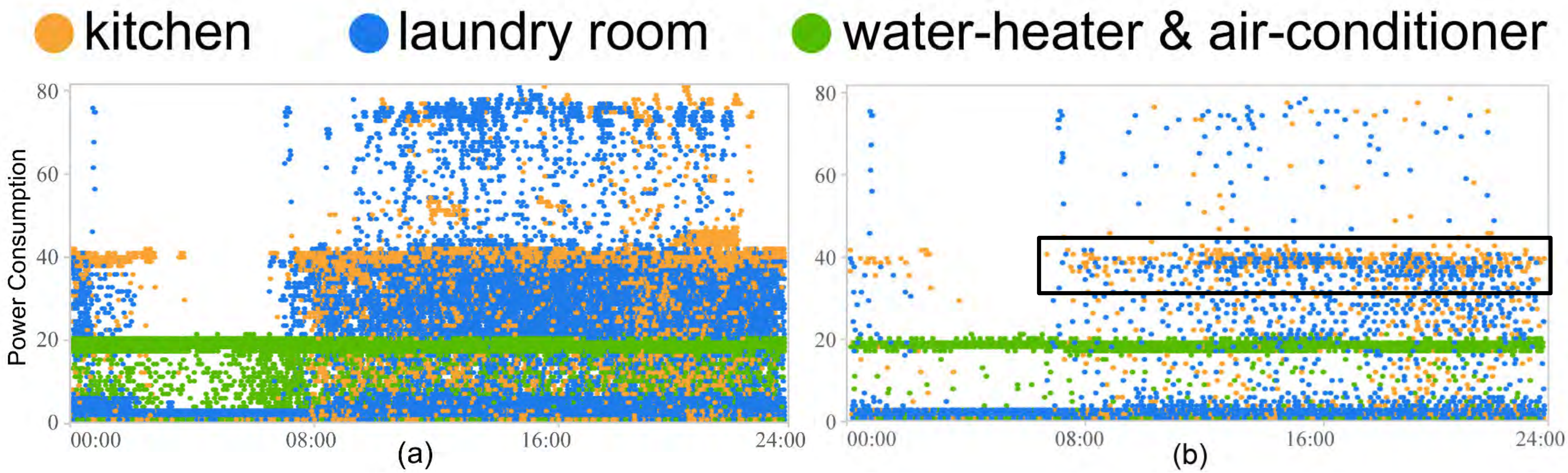
- preserve relative data density while keeping outliers in sparse areas; and
- preserve relative class density while ensuring the visibility of rare classes.



Algorithm overview

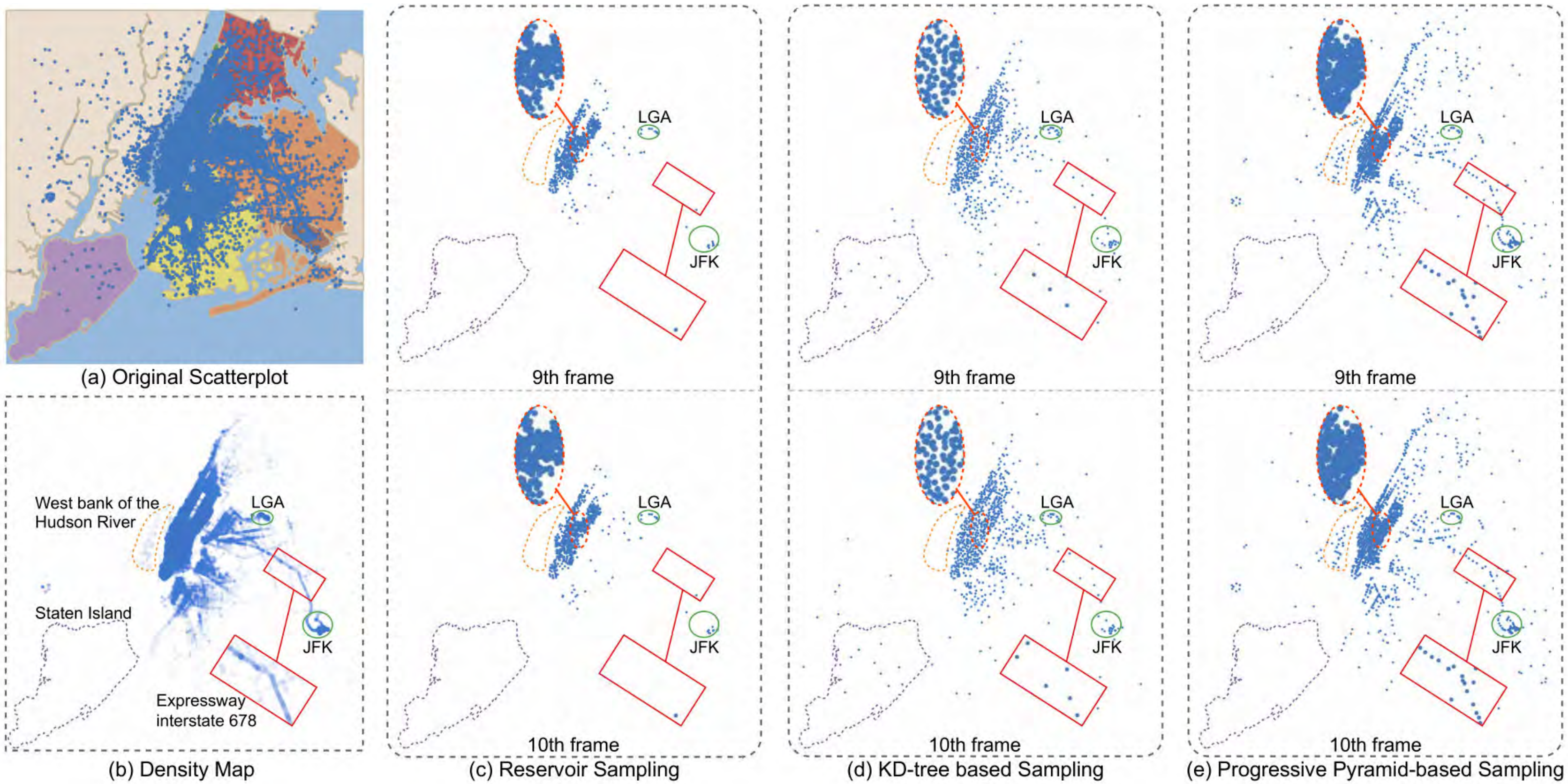


Example- Electric Power Consumption



Sampling: 1,570K → 3K

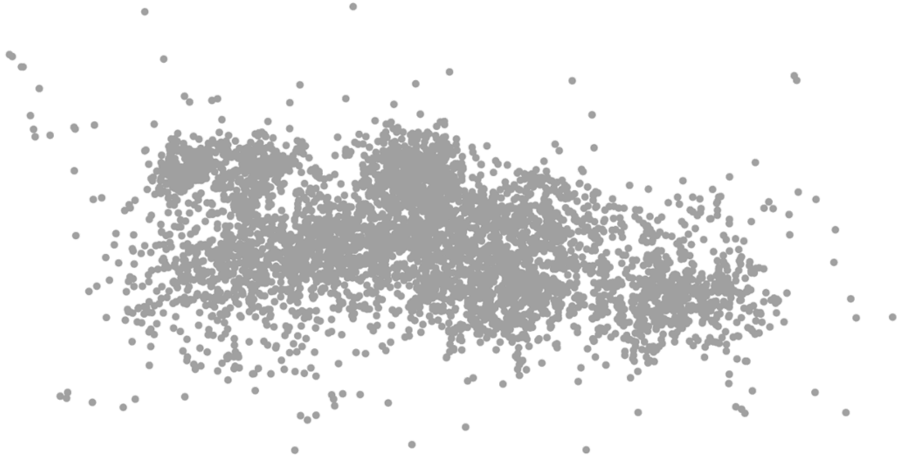
How to sample very large data or streaming data?



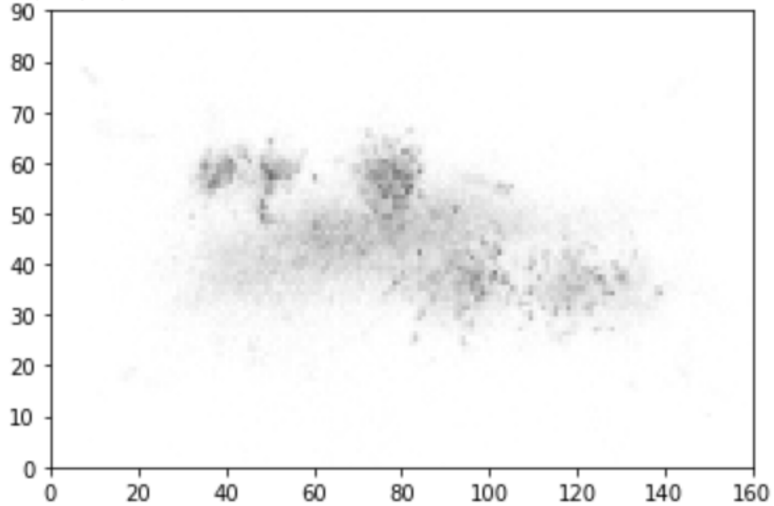
Pyramid-based Scatterplots Sampling for **Progressive** and **Streaming** Data Visualization (IEEE VIS 2021)

A common flaw of existing sampling methods

Scan the **whole** dataset?



Random Sampling



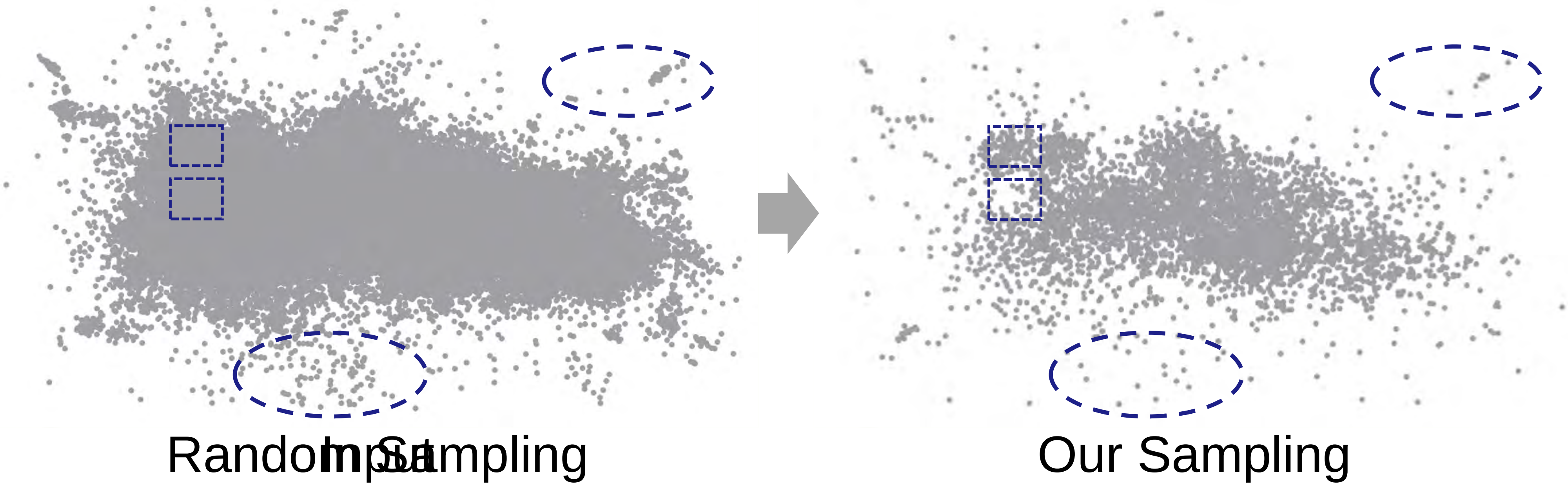
Perceptual Sampling based on density map

- Multi-class blue noise sampling [Chen et al. 2014]
- Multi-view Z-order sampling [Hu et al. 2019]
- KD-tree-based sampling [Chen et al. 2019]
- …

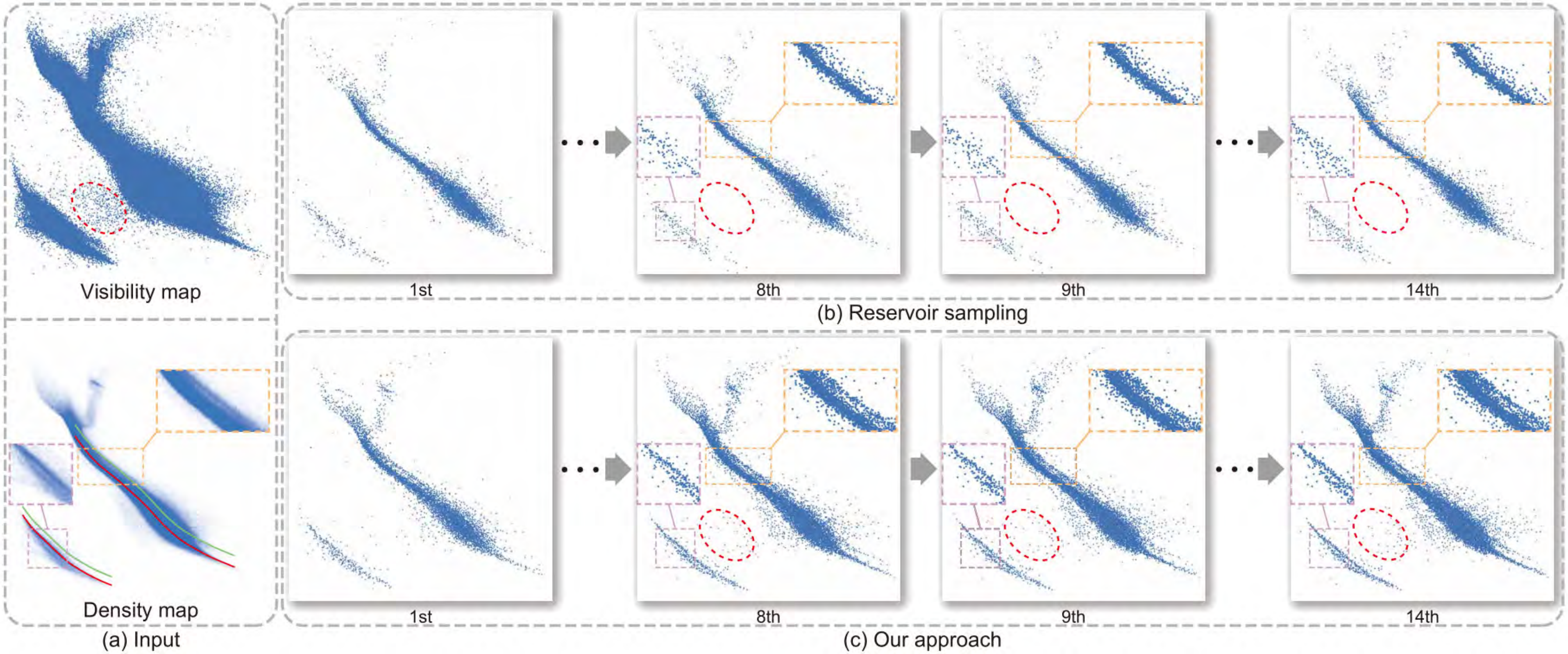
Temporal coherence

Our goals

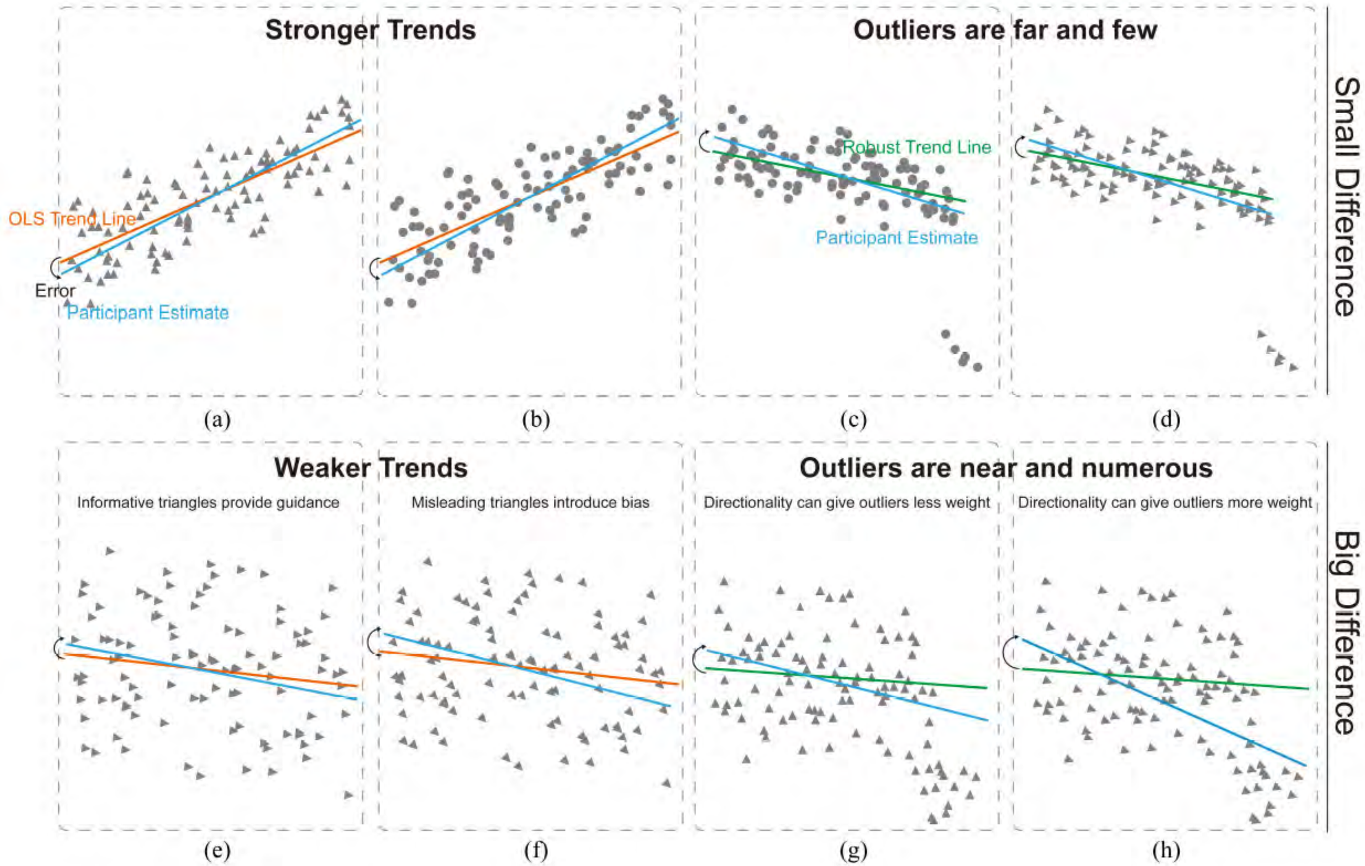
- Design a perceptual sampling method which allows to
  1. preserve relative data density while keeping outliers in sparse areas;
  2. maintain temporal coherence between successive frames; and
  3. run efficient enough to support user interaction (< 1s).



Case: Hertzsprung-Russell diagram



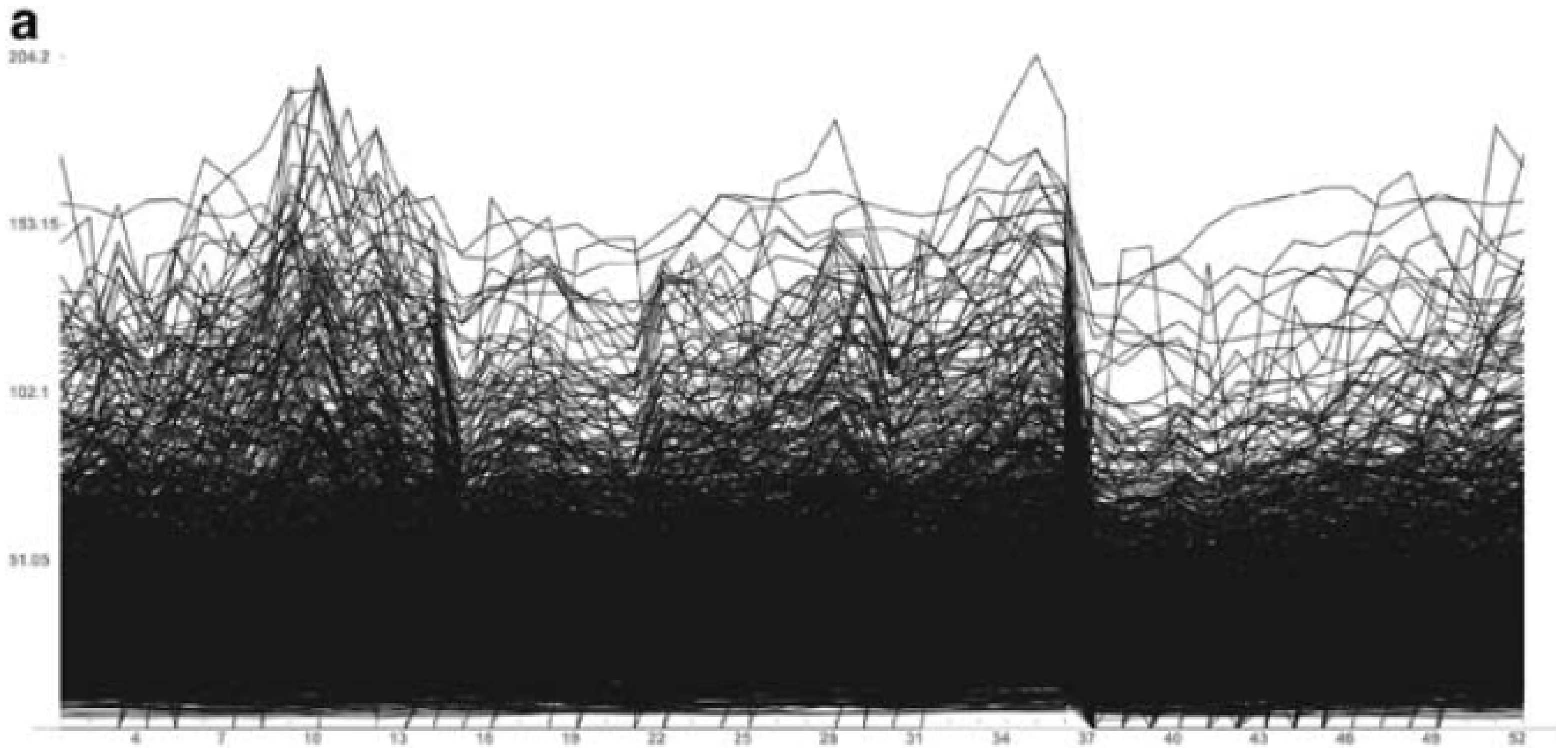
If we focus on correlation, how to design scatterplots?



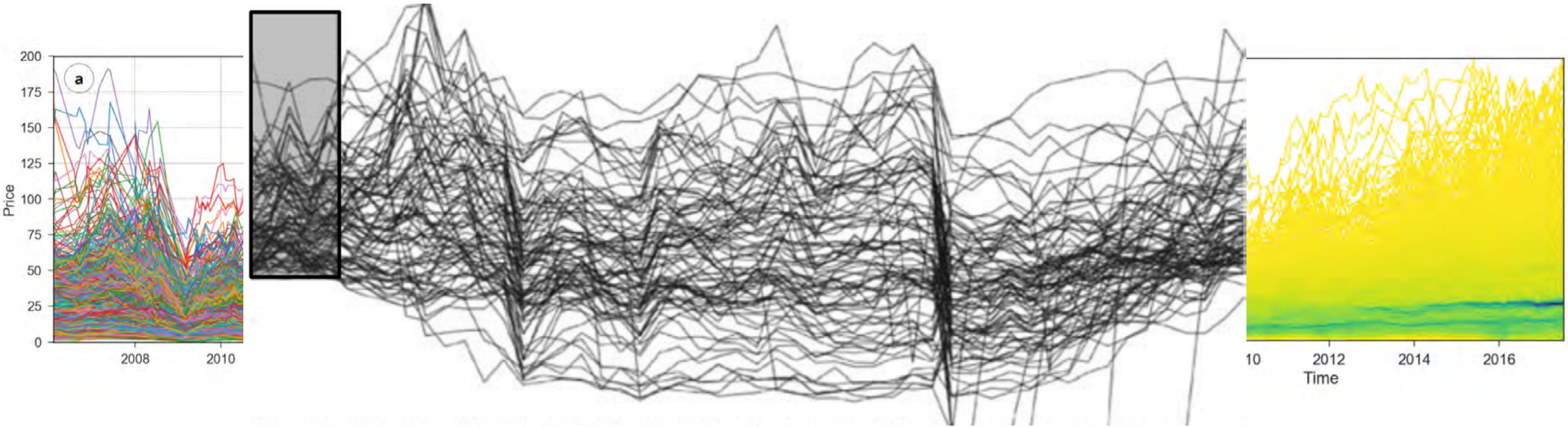
Data-Driven Mark Orientation for Trend Estimation in Scatterplots

ACM CHI 2021

If we have too many lines, can we interact with them?



Time-series Data Visualization



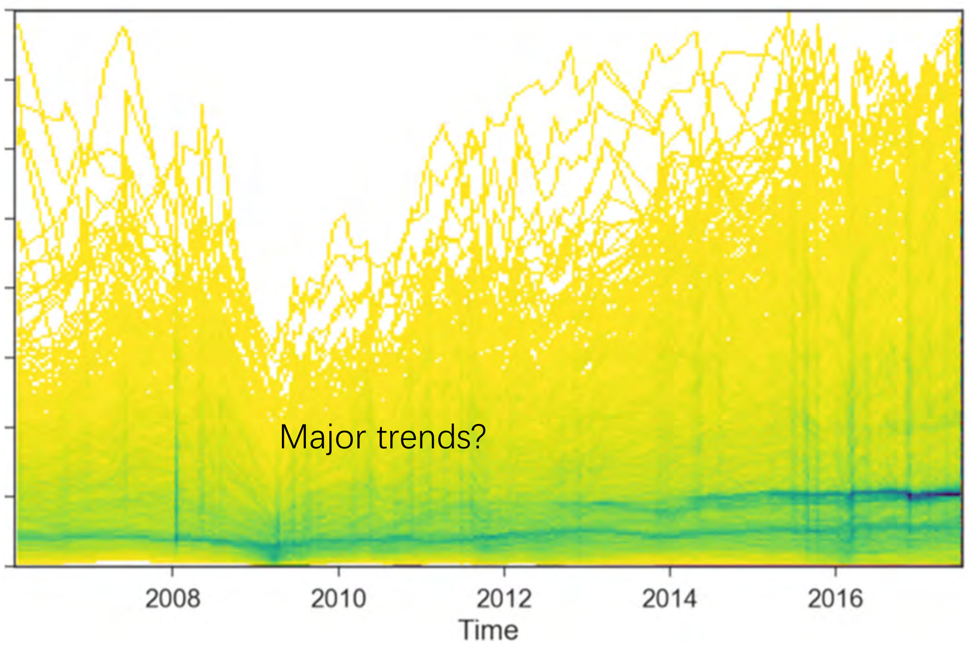
Line Graph

Density Line Chart

[Dominik Moritz and Danyel Fisher, 2018]

[Harry Hochheiser and Ben Shneiderman, 2004]

Time-series Data Visualization

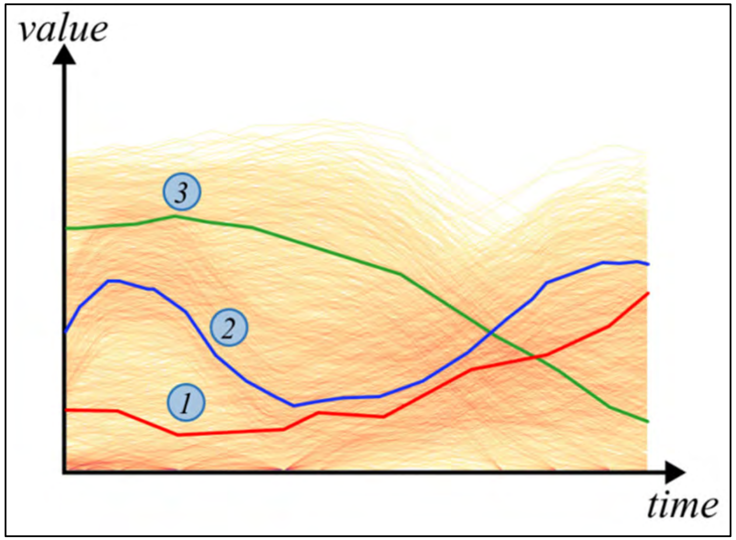


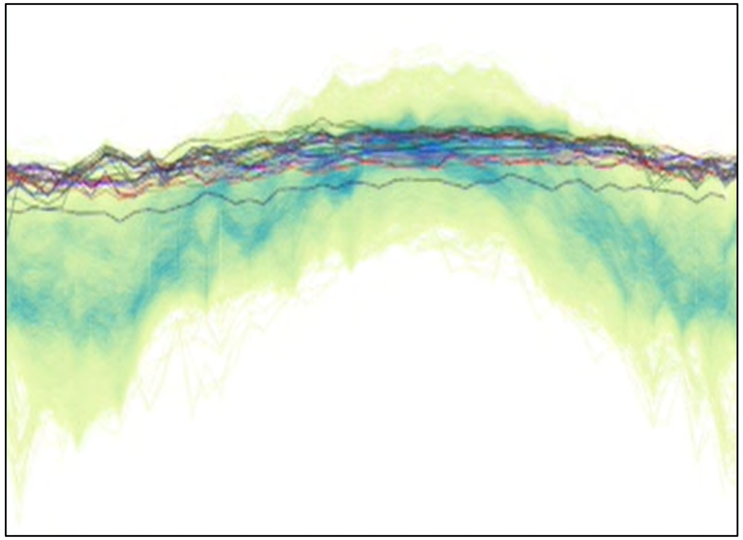
Requirements that cannot satisfy:

- fast rendering
- fast querying
- representative lines

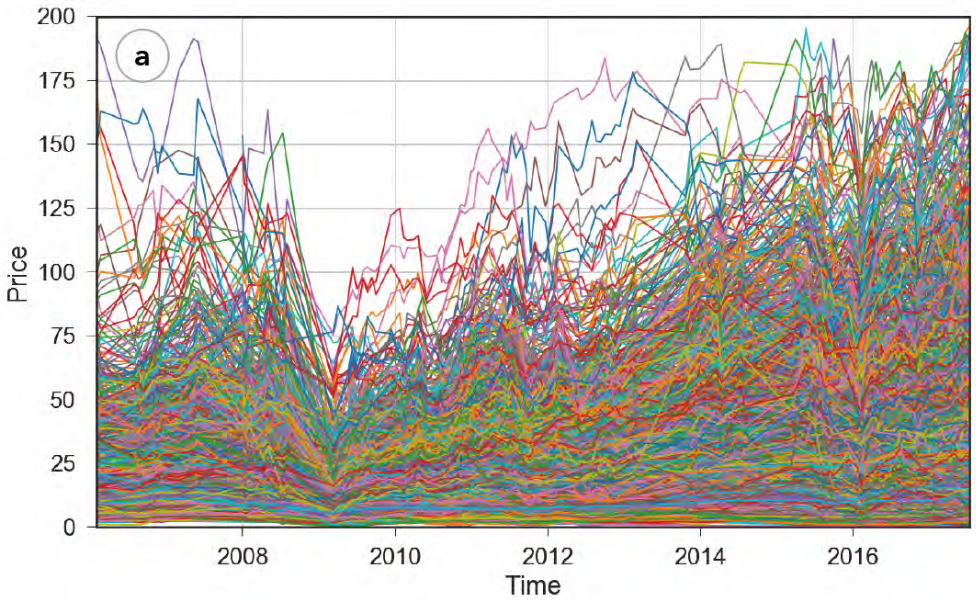
Our goal

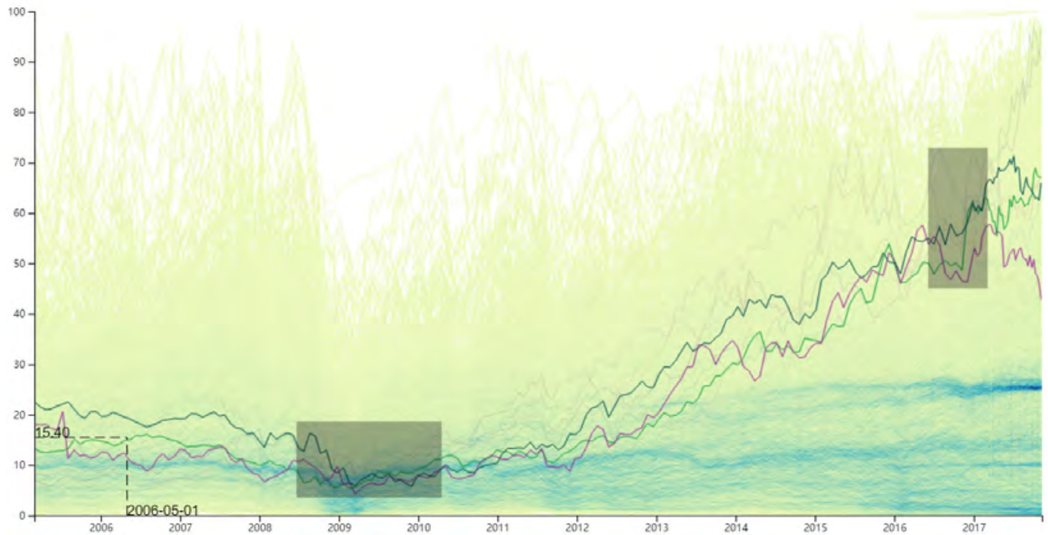
- Design an exploration system enables incremental query of many time series and fast density field computation for selecting representative lines with rich interactions.



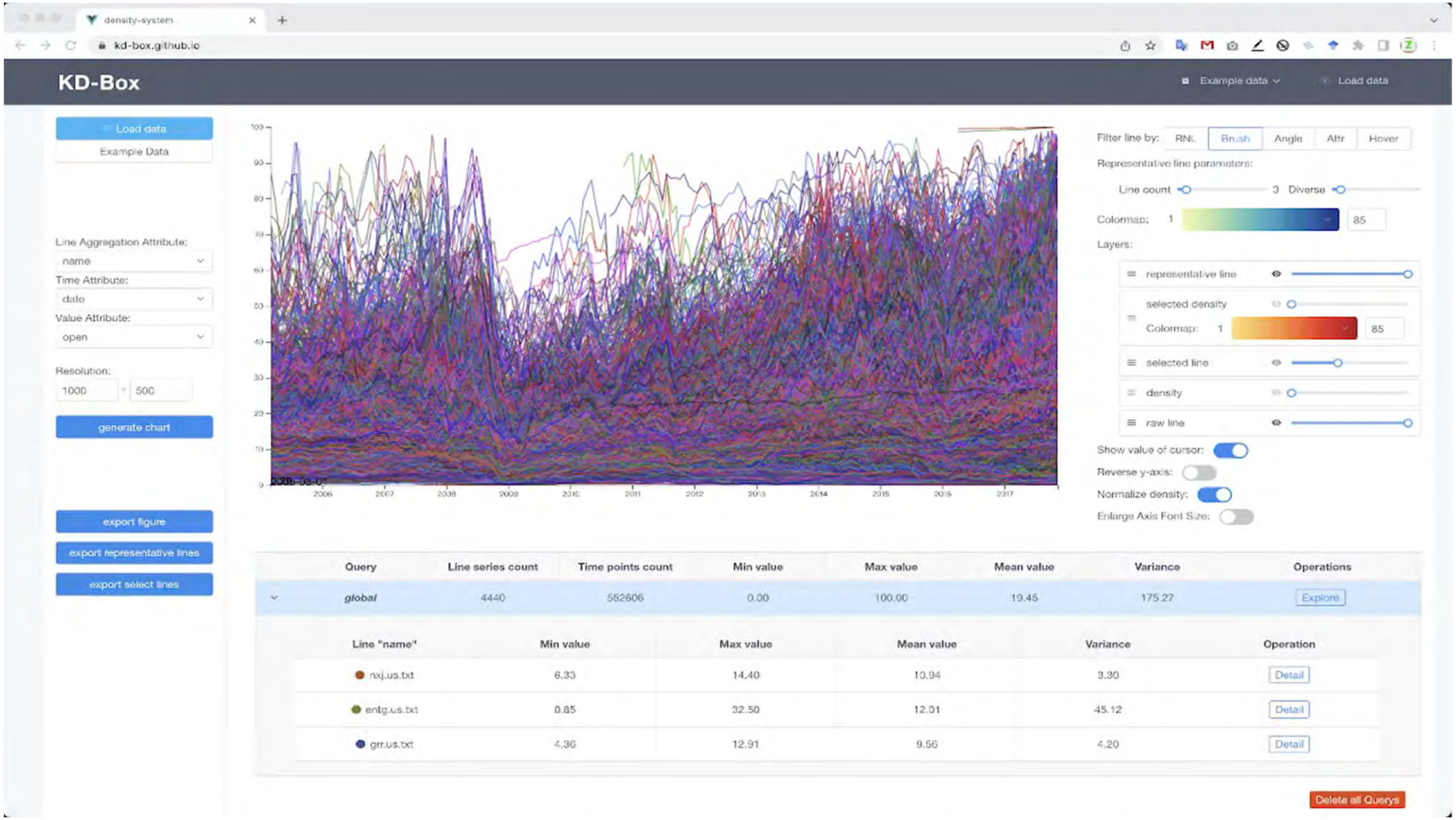


Our goal



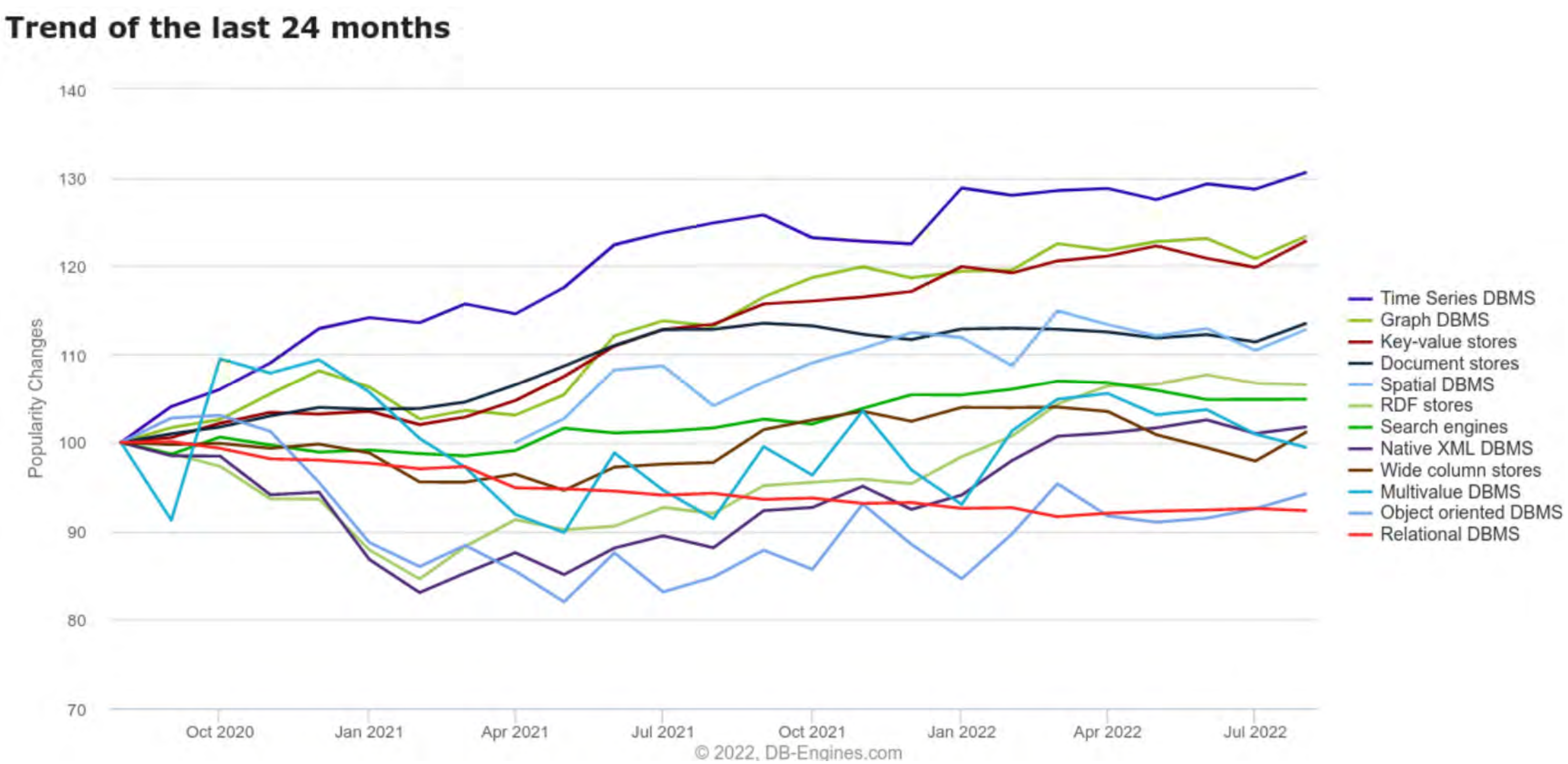


https://kd-box.github.io/



https://kd-box.github.io/

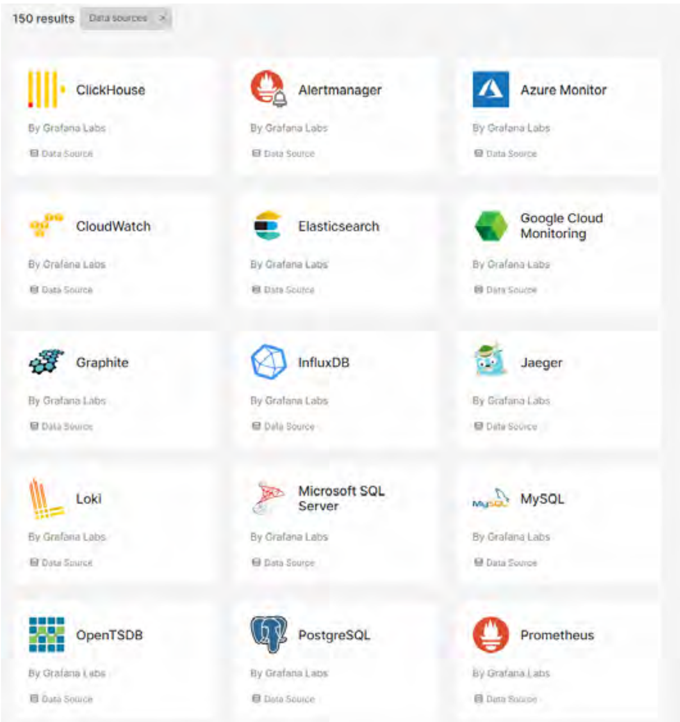
Time series database



Time Series Data Visualization

Grafana

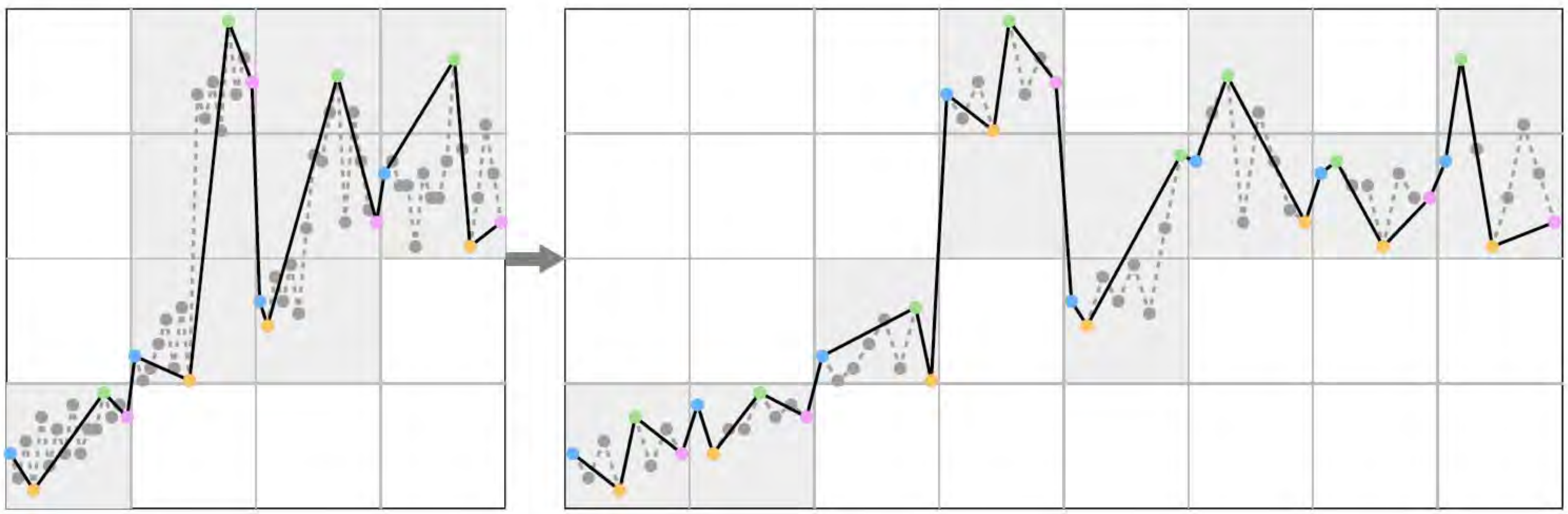
An open-source software application used for analytics and monitoring



What data is in database

1. Database query is slow
2. Network transmission data delay is high
3. Data loss leads to inaccurate visualization

M4 Aggregation

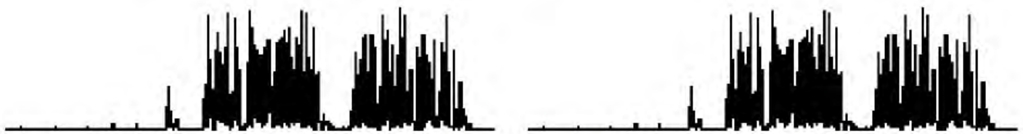


Produces error-free visualizations!

1. Uwe Jugel, Zbigniew Jerzak, Gregor Hackenbroich, Volker Markl: *M4: A Visualization-Oriented Time Series Data Aggregation*. Proc. VLDB Endow. 7(10): 797-808 (2014)

M4 Query

```
SELECT t,v FROM Q JOIN
(SELECT round($w*(t-$t1)/($t2-$t1)) as k, --define key
        min(v) as v_min, max(v) as v_max, --get min,max
        min(t) as t_min, max(t) as t_max  --get 1st,last
        FROM Q GROUP BY k) as QA          --group by k
ON k = round($w*(t-$t1)/($t2-$t1))        --join on k
   AND (v = v_min OR v = v_max OR         --&(min|max|
        t = t_min OR t = t_max)           --  1st|last)
```



b) resulting image == expected image

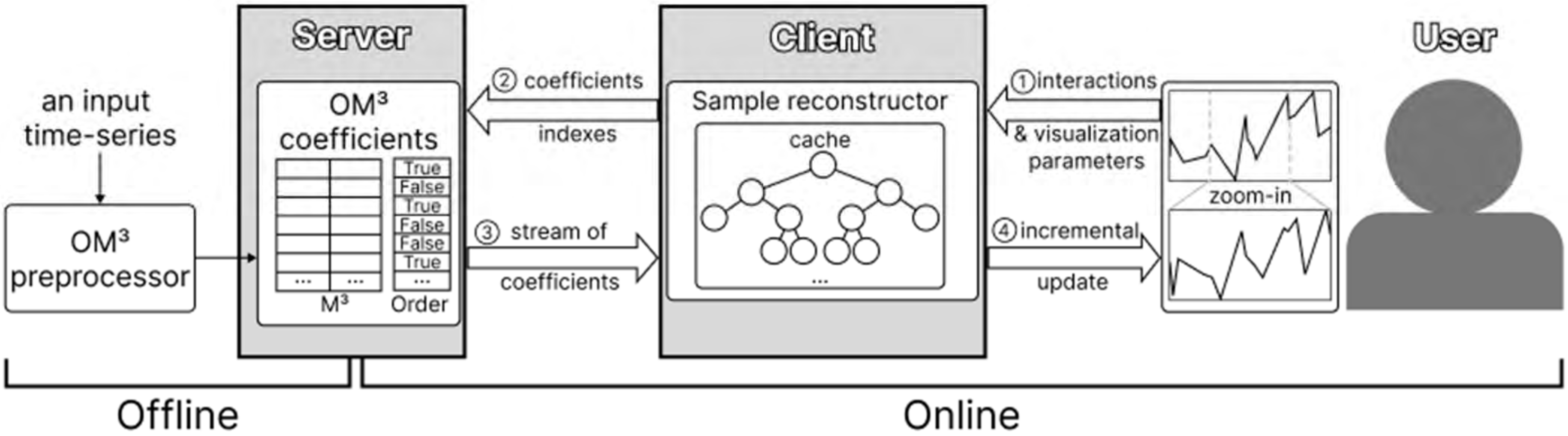
1. Its query execution has **high** interaction latency $O(n+4\omega)$
2. It might result in **data redundancy** because of the joint operation
3. It **cannot reuse previous query results** to support incremental update

Our Desiderata

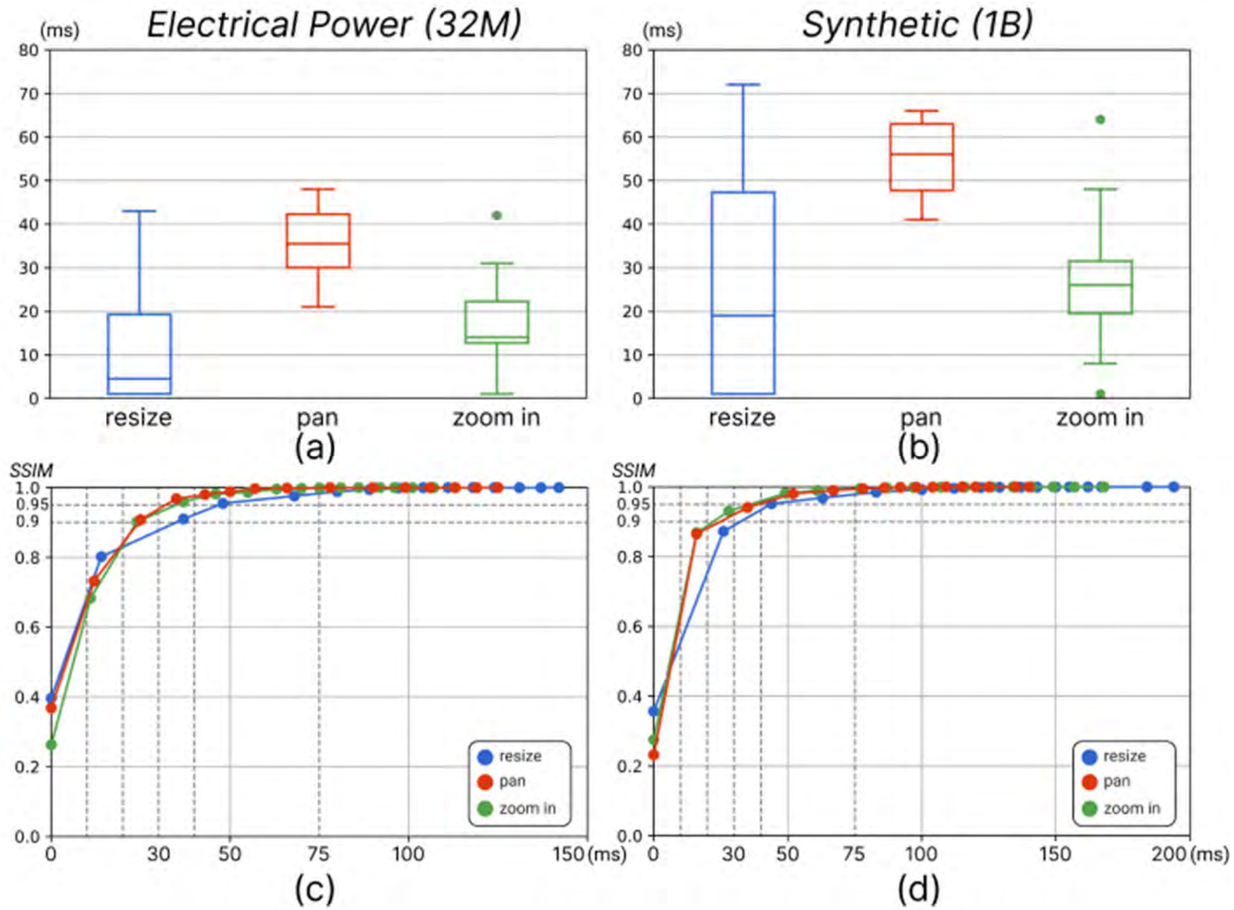
- Ensure **error-free** line visualizations **at any scale**;
- Minimize **query latency and amount of data transfer** to the visualization clients;
- Support **progressive** refinement of intermediate visualization

1. **Yunhai Wang etc.** : **OM^3: An Ordered Multi-level Min-Max Representation for Interactive Progressive Visualization of Time Series***. Accepted by SIGMOD'23*

OM³ based Architecture

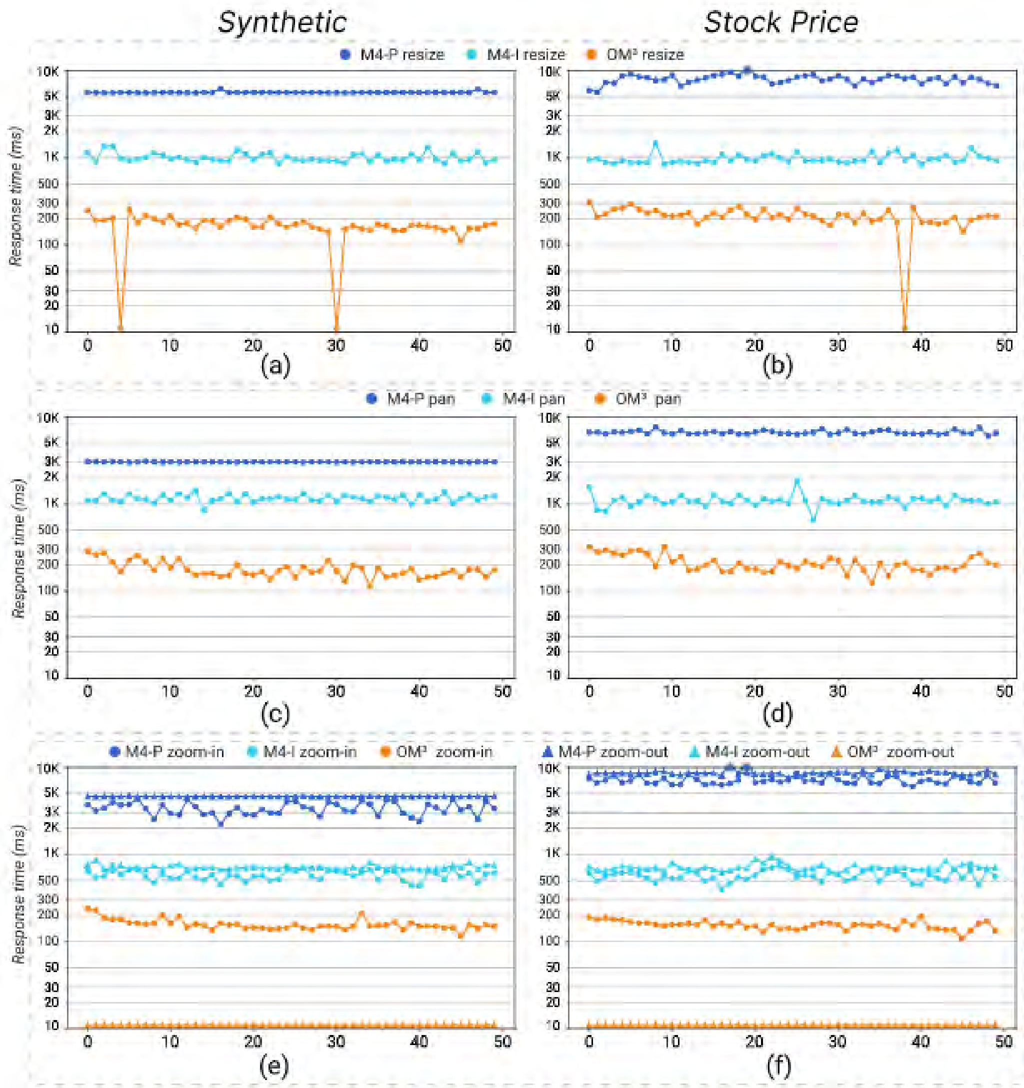


Evaluation: Progressive Visualization

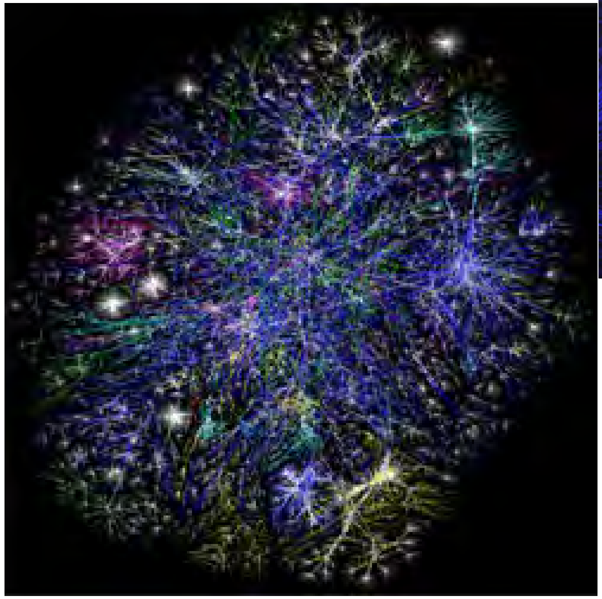


Evaluation: Interactive Visualization

The dataset has **8M** record

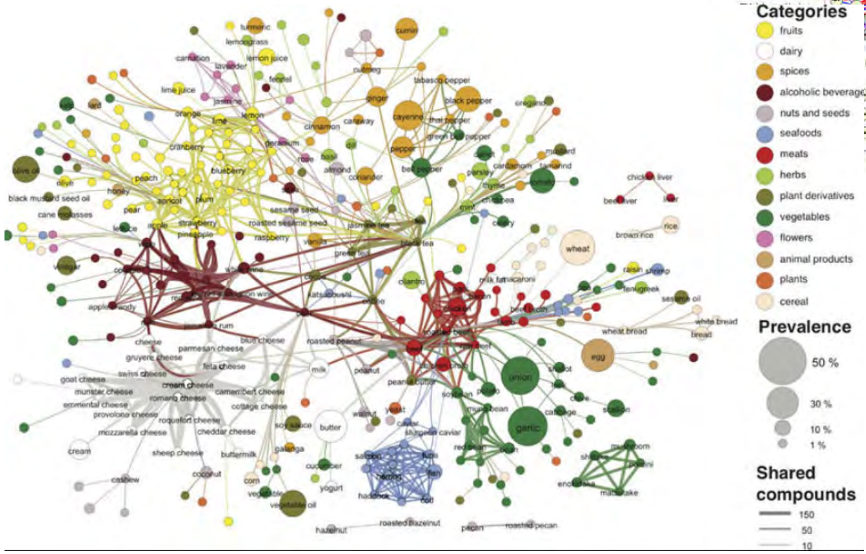


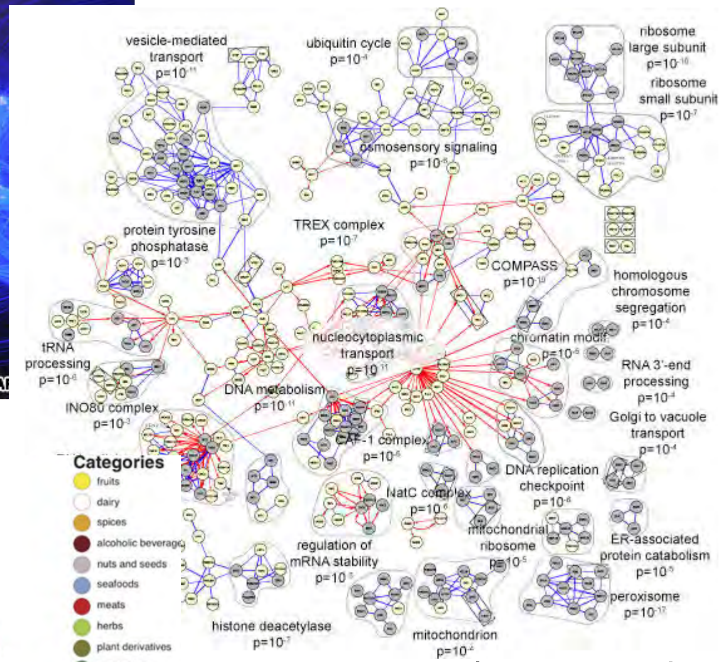
Points + lines: node-link diagrams



The Internet [2005]

Social Networks [2010]





Gene Regulatory Graph

[Decourty 2008]

Flavor network

[Anh 2011]

Challenges in Graph Processing

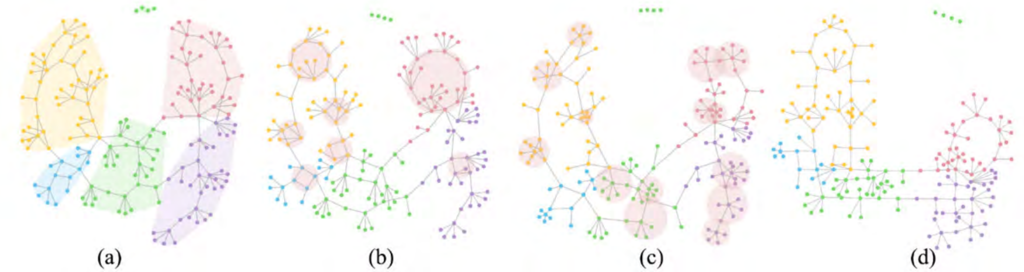
The Ubiquity of Large Graphs and Surprising Challenges of Graph Processing VLDB 2018 Best paper

Siddhartha Sahu, Amine Mhedhbi, Semih Salihoglu, Jimmy Lin, M. Tamer Özsu
David R. Cheriton School of Computer Science
University of Waterloo

Table 15: The graph processing challenges selected by the participants.

Challenge	Total	R	P
Scalability (i.e., software that can process larger graphs)	45	20	25
Visualization	39	17	22
Query Languages / Programming APIs	39	18	21
Faster graph or machine learning algorithms	35	19	16
Usability (i.e., easier to deploy, configure, and use)	25	10	15
Benchmarks	22	12	10
Extract & Transform	20	6	14
More general purpose graph software (e.g., that can process offline, online, and streaming computations)	20	9	11
Graph Cleaning	17	7	10
Debugging & Testing	10	2	8

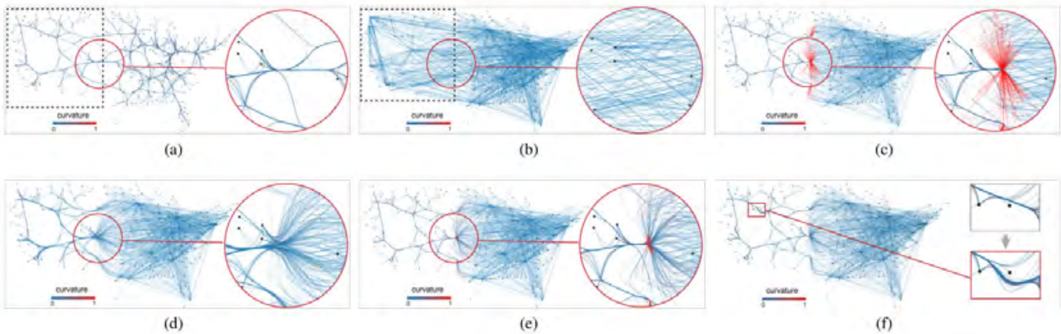
Our constrained Graph visualization before 2022



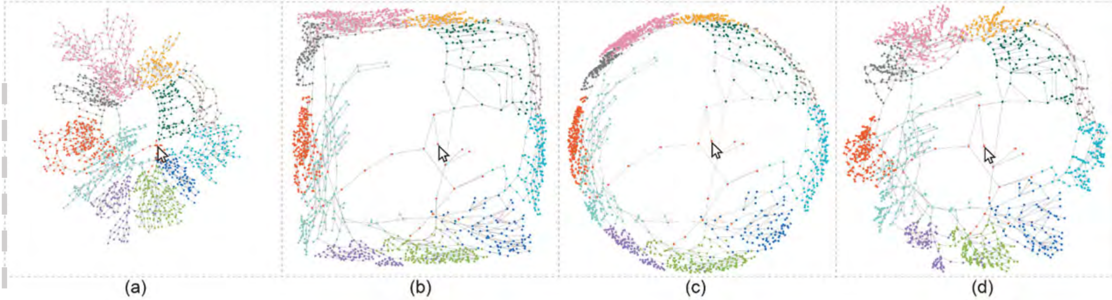
Revisiting Stress Majorization as a Unified Framework for Interactive Constrained Graph Visualization

IEEE Transactions on Visualization and Computer Graphics
(Proc. InfoVis 2017), 2018

Interactive Structure-aware Blending of Diverse Edge Bundling Visualizations

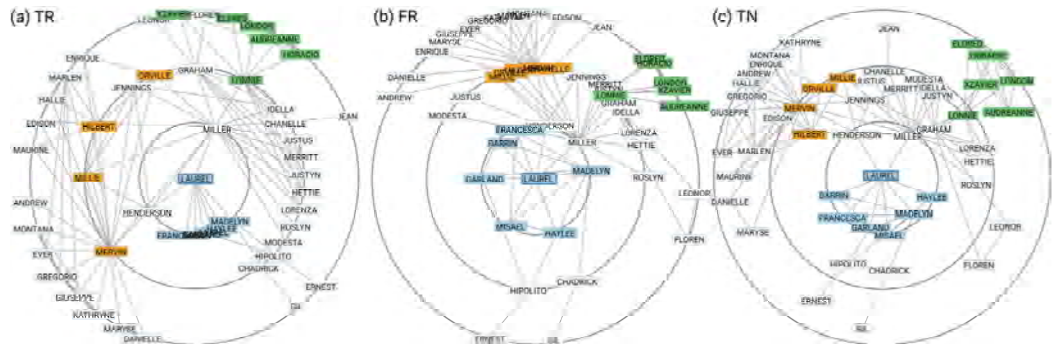


IEEE Transactions on Visualization and Computer Graphics
(Proc. InfoVis 2019), 2020



Structure-aware Fisheye Views for Efficient Large Graph Exploration

IEEE Transactions on Visualization and Computer Graphics
(Proc. InfoVis 2018), 2019

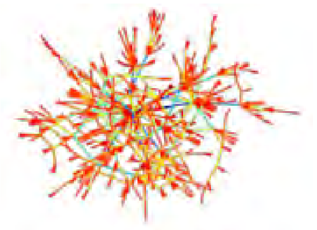


Target Netgrams: An Annulus-constrained Stress Model for Radial Graph Visualization

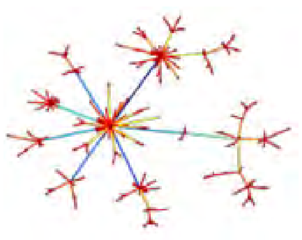
IEEE Transactions on Visualization and Computer Graphics,
2022

Graph Layout methods

Spring-electrical models

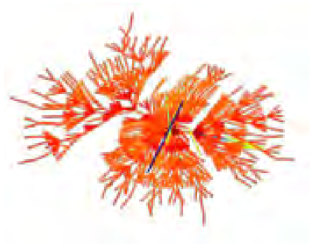


Force directed placement
[T. M. Fruchterman and E. M. Reingold, 1991]

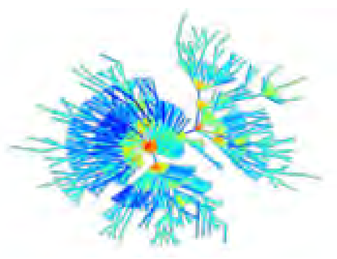


LinLog
[A. Noack, 2003]

Stress models

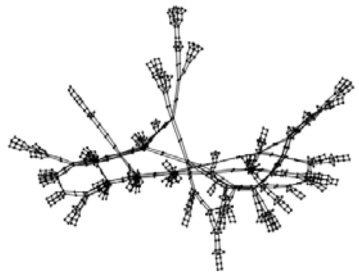


Stress model
[E. R. Gansner and Y. Koren, 2004]

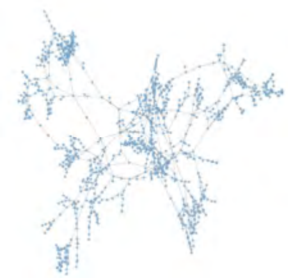


Sparse stress model
[M. Ortmann and U. Brandes. 2016]

Hybrid models

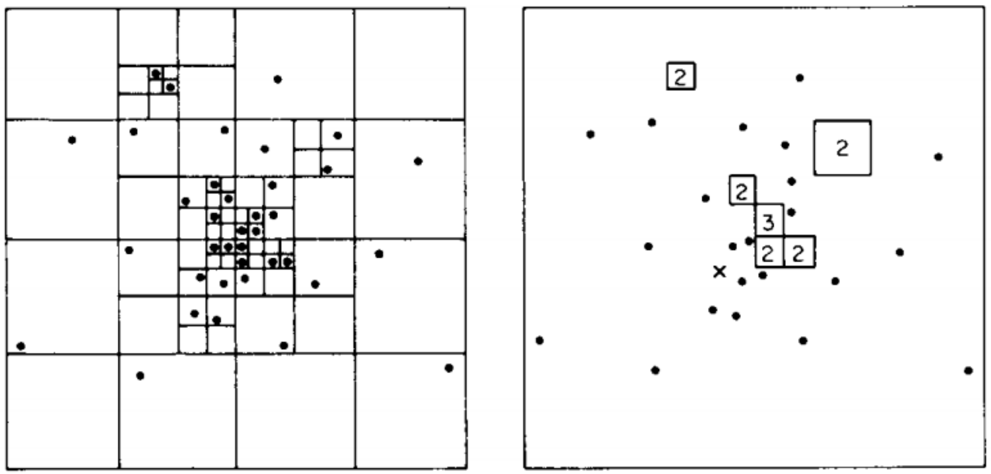


localized stress model
[Y. Hu and Y. Koren, 2009]



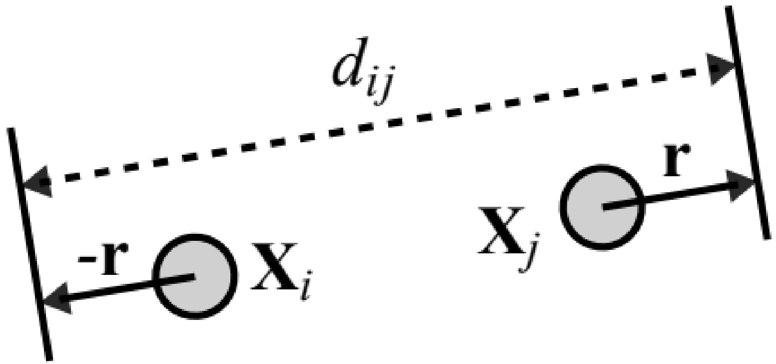
Maxent
[E. R. Gansner and Y. Hu, 2012]

Graph Layout Solvers



Barnes & Hut approximation

[J. Barnes and P. Hut., 1986]



Stochastic Gradient Descent

[J. X. Zheng and S. Pawar, 2019]

Graph Layout Packages

Graphviz

Complex APIs and parameters

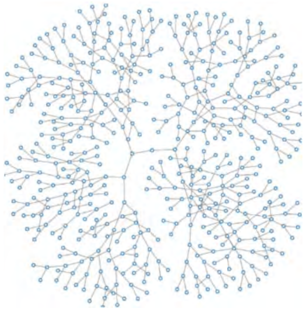
Tulip

Hard to understand

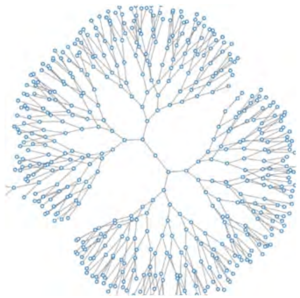
OGDF

D3

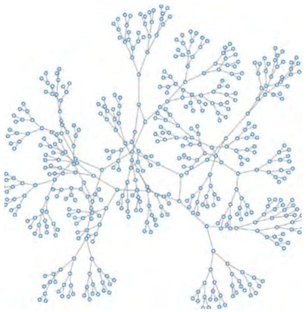
What is the next?



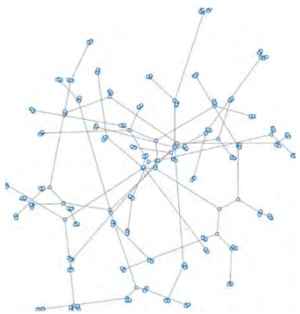
Force directed placement



Stress model



Maxent



LinLog

Why is the differences?

What is the connection?

How to evaluate?

A Unified framework is needed!

Proposed Framework

Force-Directed Placement

$$\forall\{i,j\}\in E,\quad F^a_{i,j}=||\mathbf{x}_i-\mathbf{x}_j||^2*\mathbf{e}_{i,j},$$

$$\forall\{i,j\}\in V^2,\quad F^r_{i,j}=-\frac{1}{||\mathbf{x}_i-\mathbf{x}_j||}*\mathbf{e}_{i,j},$$

$$F_i=\sum_{\{i,j\}\in E}F^a_{i,j}+\sum_{\{i,j\}\in V^2}F^r_{i,j}$$

Stress Model

$$U=\sum_{\{i,j\}\in V^2}\frac{(||\mathbf{x}_i-\mathbf{x}_j||-d_{ij})^2}{d_{ij}^2},$$

$$\frac{\partial U}{\partial x_i}=-\sum_{\{i,j\}\in V^2}\frac{2(||\mathbf{x}_i-\mathbf{x}_j||-d_{ij})}{d_{ij}^2}*\mathbf{e}_{i,j}.$$

Maxent

$$U=\begin{cases}\sum_{(i,j)\in S}\frac{(||\mathbf{x}_i-\mathbf{x}_j||-d_{ij})^2}{d_{ij}^2}+\alpha\sum_{(i,j)\notin S}\frac{sgn(q)}{||\mathbf{x}_i-\mathbf{x}_j||^q}, \text{if } q\neq 0\\ \sum_{(i,j)\in S}\frac{(||\mathbf{x}_i-\mathbf{x}_j||-d_{ij})^2}{d_{ij}^2}+\alpha\sum_{(i,j)\notin S}\ln||\mathbf{x}_i-\mathbf{x}_j||, \text{if } q=0\end{cases}$$

$$\frac{\partial U}{\partial x_i}=-\left(\sum_{(i,j)\in S}\frac{2(||\mathbf{x}_i-\mathbf{x}_j||-d_{ij})}{d_{ij}^2}-\alpha\sum_{(i,j)\notin S}\frac{1}{||\mathbf{x}_i-\mathbf{x}_j||}\right)*\mathbf{e}_{i,j}.$$

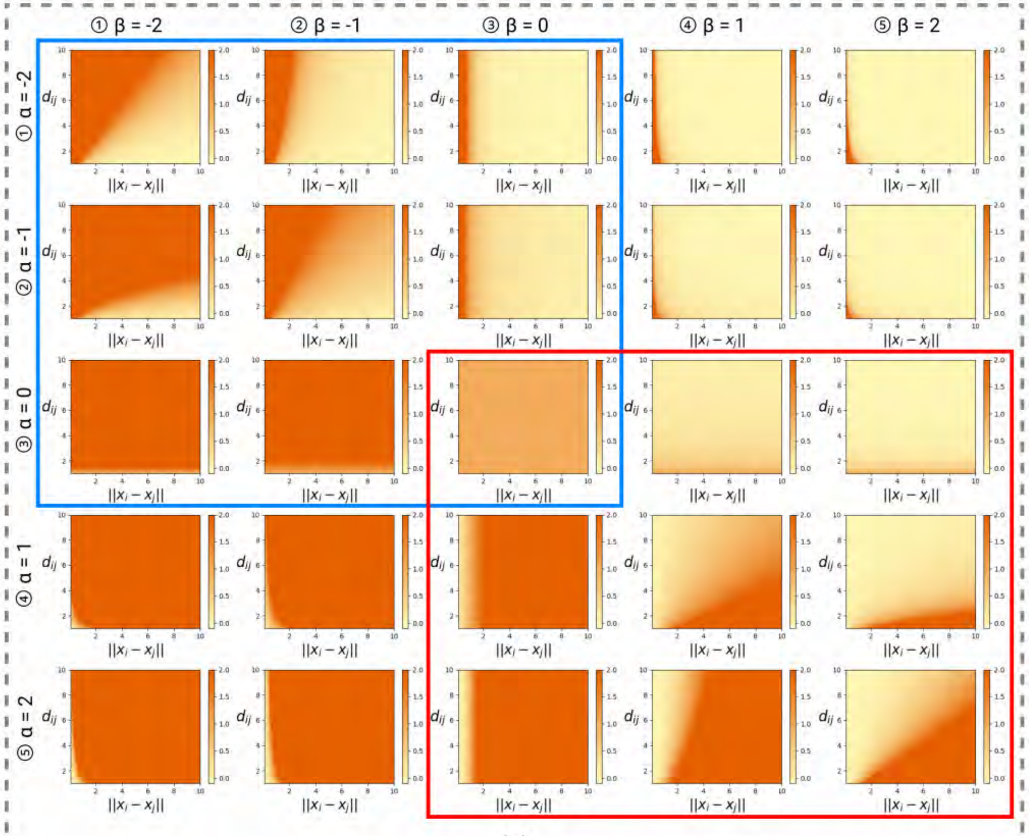
$$F_{i,k}=\sum_{(i,j)\in\Omega_k}\omega_k*\frac{||\mathbf{x}_i-\mathbf{x}_j||^{\alpha_k}}{d_{ij}^{\beta_k}}*\mathbf{e}_{k,ij},\quad F_i=\sum_k^l F_{i,k}$$

Quotient based Force Function

Quotient based force functions and their parameters of different layout methods

Method	Attractive Force	$\{\omega_1, \alpha_1, \beta_1, \Omega_1\}$	Repulsive Forces	$\{\omega_2, \alpha_2, \beta_2, \Omega_2\}$
FDP [10]	$\sum_{(i,j)\in E} \|\|\mathbf{x}_i - \mathbf{x}_j\|\|^2 \mathbf{e}_{ij}$	$\{1,2,0,E\}$	$\sum_{\{i,j\}\in V^2} \frac{-1}{\|\|\mathbf{x}_i - \mathbf{x}_j\|\|} \mathbf{e}_{ij}$	$\{-1,-1,0,V^2\}$
FA2 [22]	$\sum_{(i,j)\in E} \|\|\mathbf{x}_i - \mathbf{x}_j\|\| \mathbf{e}_{ij}$	$\{1,2,0,E\}$	$\sum_{\{i,j\}\in V^2} \frac{k_{fa}}{\|\|\mathbf{x}_i - \mathbf{x}_j\|\|} \mathbf{e}_{ij}$	$\{k_{fa}, -1, 0, V^2\}$
LinLog [31]	$\sum_{(i,j)\in E} 1 * \mathbf{e}_{ij}$	$\{1,1,0,E\}$	$\sum_{\{i,j\}\in V^2} \frac{-1}{\|\|\mathbf{x}_i - \mathbf{x}_j\|\|} \mathbf{e}_{ij}$	$\{-1,-1,0,V^2\}$
SM [14]	$\sum_{\{i,j\}\in V^2} \frac{2\|\|\mathbf{x}_i - \mathbf{x}_j\|\|}{d_{ij}^2} \mathbf{e}_{ij}$	$\{2,1,2,V^2\}$	$\sum_{\{i,j\}\in V^2} \frac{-2}{d_{ij}} \mathbf{e}_{ij}$	$\{-2,0,1,V^2\}$
MARS [27]	$\sum_{(i,j)\in P\times V} \frac{2\|\|\mathbf{x}_i - \mathbf{x}_j\|\|}{d_{ij}} \mathbf{e}_{ij}$	$\{2,1,1,P\times V\}$	$\sum_{(i,j)\in P\times V} -2\mathbf{e}_{ij}$	$\{-2,0,0,P\times V\}$
SSM [32]	$\sum_{(i,j)\in P\times V \cup E} \frac{2\|\|\mathbf{x}_i - \mathbf{x}_j\|\|}{d_{ij}^2} \mathbf{e}_{ij}$	$\{2,1,2,P\times V\cup E\}$	$\sum_{(i,j)\in P\times V \cup E} \frac{-2}{d_{ij}} \mathbf{e}_{ij}$	$\{-2,0,1,P\times V\cup E\}$
Maxent [13]	$\sum_{\{i,j\}\in S} \frac{2\|\|\mathbf{x}_i - \mathbf{x}_j\|\|}{d_{ij}^2} \mathbf{e}_{ij}$	$\{2,1,2,S\}$	$(\sum_{\{i,j\}\in S} \frac{-2}{d_{ij}} + \sum_{\{i,j\}\in V^2} \frac{-\alpha sgn(q)}{\|\|\mathbf{x}_i - \mathbf{x}_j\|\|^q})\mathbf{e}_{ij}$	$\{-2,0,1,S\}$, $\{-\alpha sgn(q), -q, 0, V^2\}$

Parameter Effect



(a)

Relation:

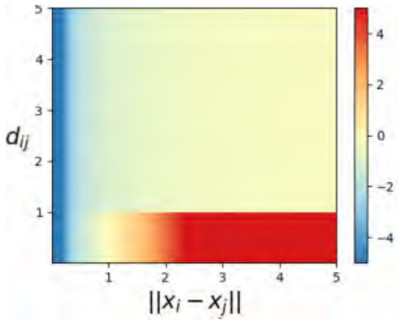
1. If $\alpha\beta$ is zero, the force magnitude is purely determined by $||\mathbf{x}_i - \mathbf{x}_j||$ or d_{ij};
2. If $\alpha\beta$ is smaller than zero , the factors d_{ij} and $||\mathbf{x}_i - \mathbf{x}_j||$ have the same positive or negative effect on the force magnitude; and
3. If $\alpha\beta$ is larger than zero, the factors d_{ij} and $||\mathbf{x}_i - \mathbf{x}_j||$ have the opposite effect on the force magnitude.

Guideline:

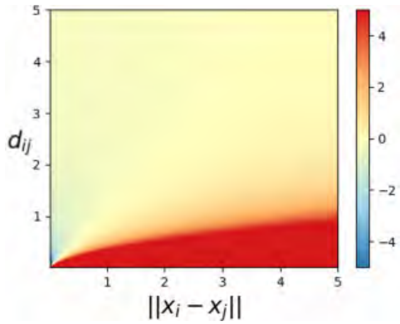
- **G1:** For the attractive force, the exponent parameters are suggested to satisfy: $\alpha \geq 0, \beta \geq 0$; and
- **G2:** For the repulsive force, the exponent parameters are suggested to satisfy: $\alpha \leq 0, \beta \leq 0$.

A Balanced Stress Model (BSM)

The resultant forces for different graph-theoretical distance and Euclidian distance:

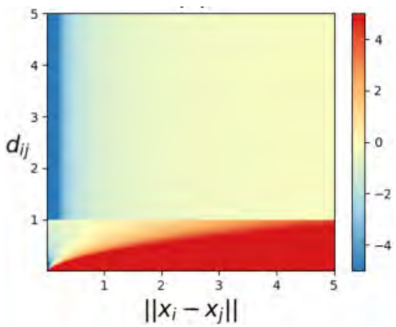


Force-Directed Placement

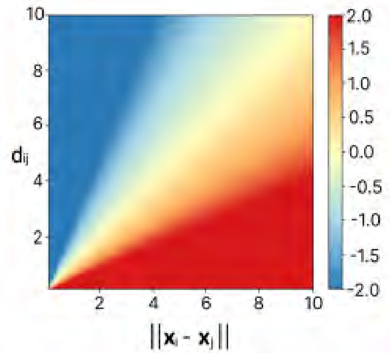


Stress Model

Weak repulsive force, not balance



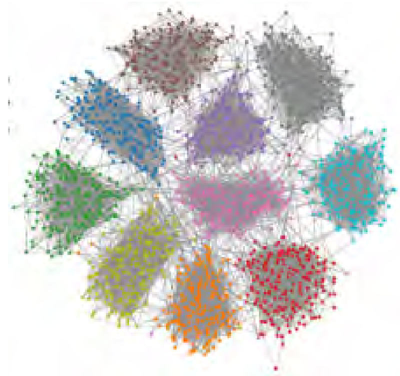
Maxent



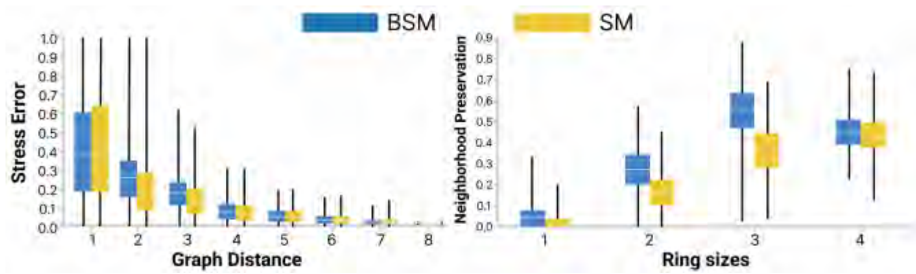
Balanced Stress Model

$$F_i = \sum_{i,j \in V^2} \left(\frac{||\mathbf{x}_i - \mathbf{x}_j||}{d_{ij}} - \frac{d_{ij}}{||\mathbf{x}_i - \mathbf{x}_j||} \right) * \mathbf{e}_{ij},$$

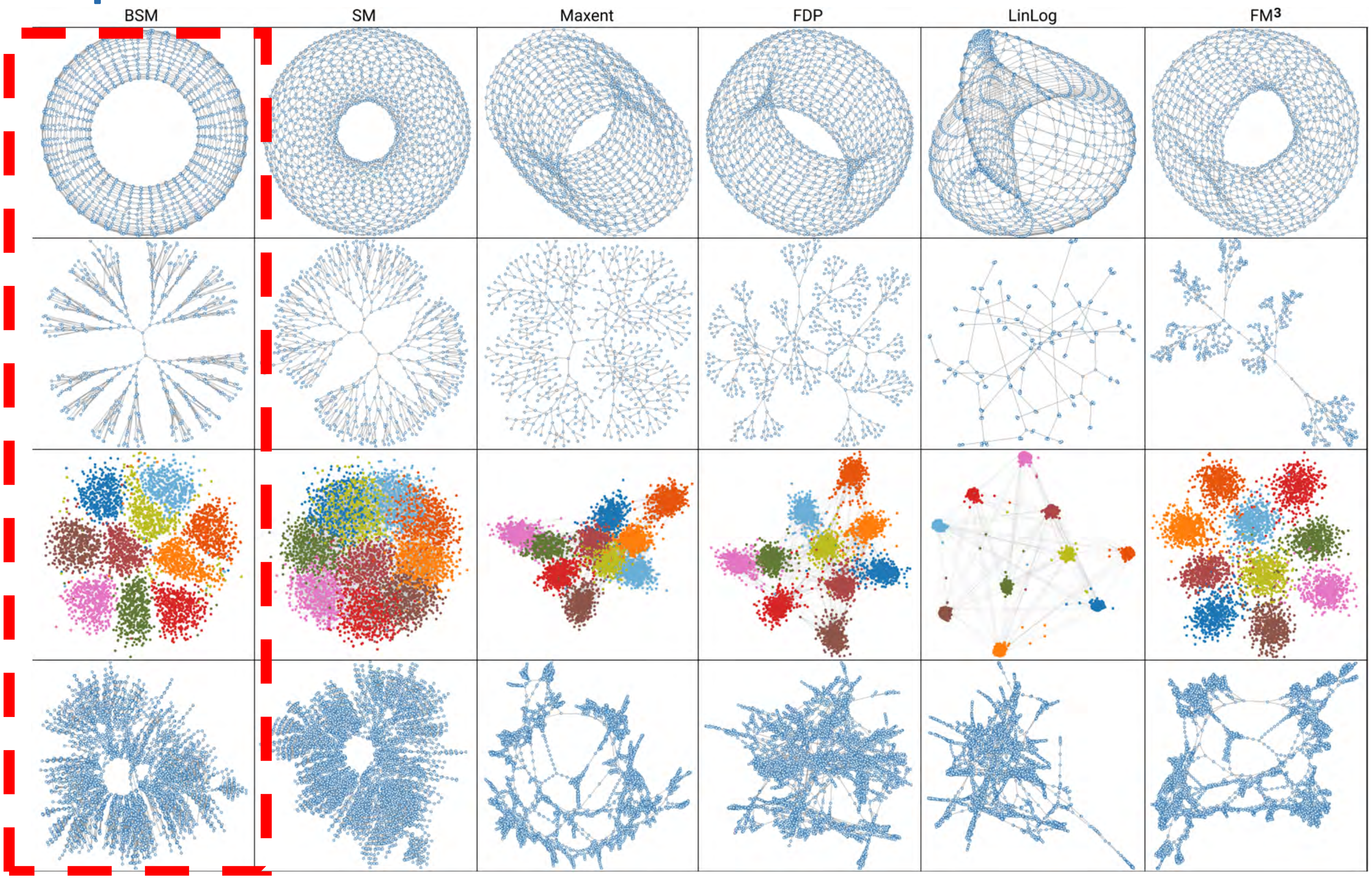
Stress Model



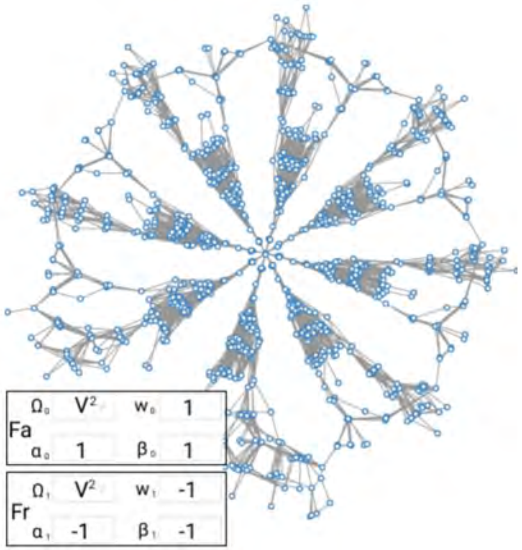
Balanced Stress Model

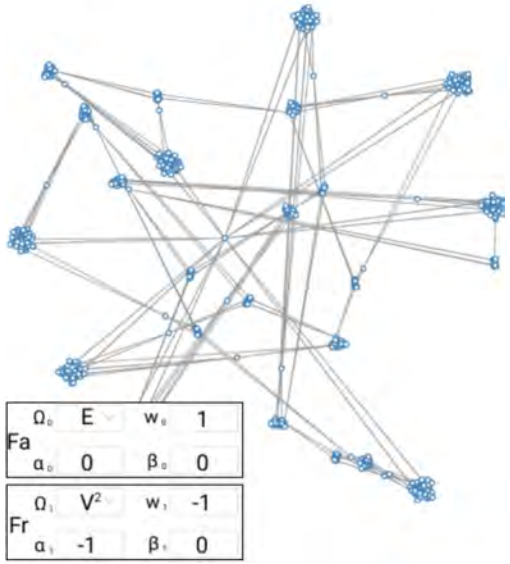


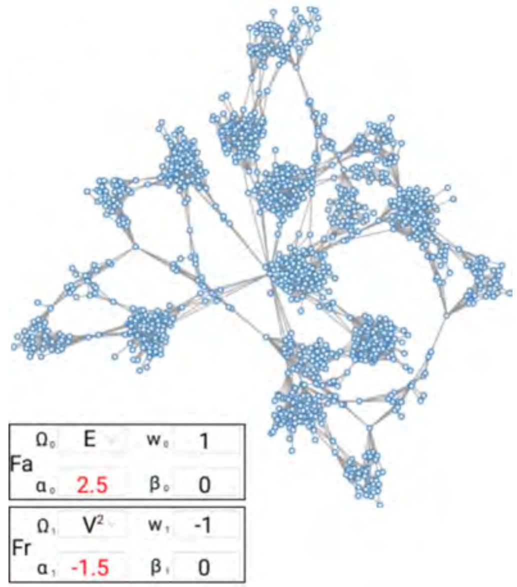
Comparison between Different Methods



Usage Scenario







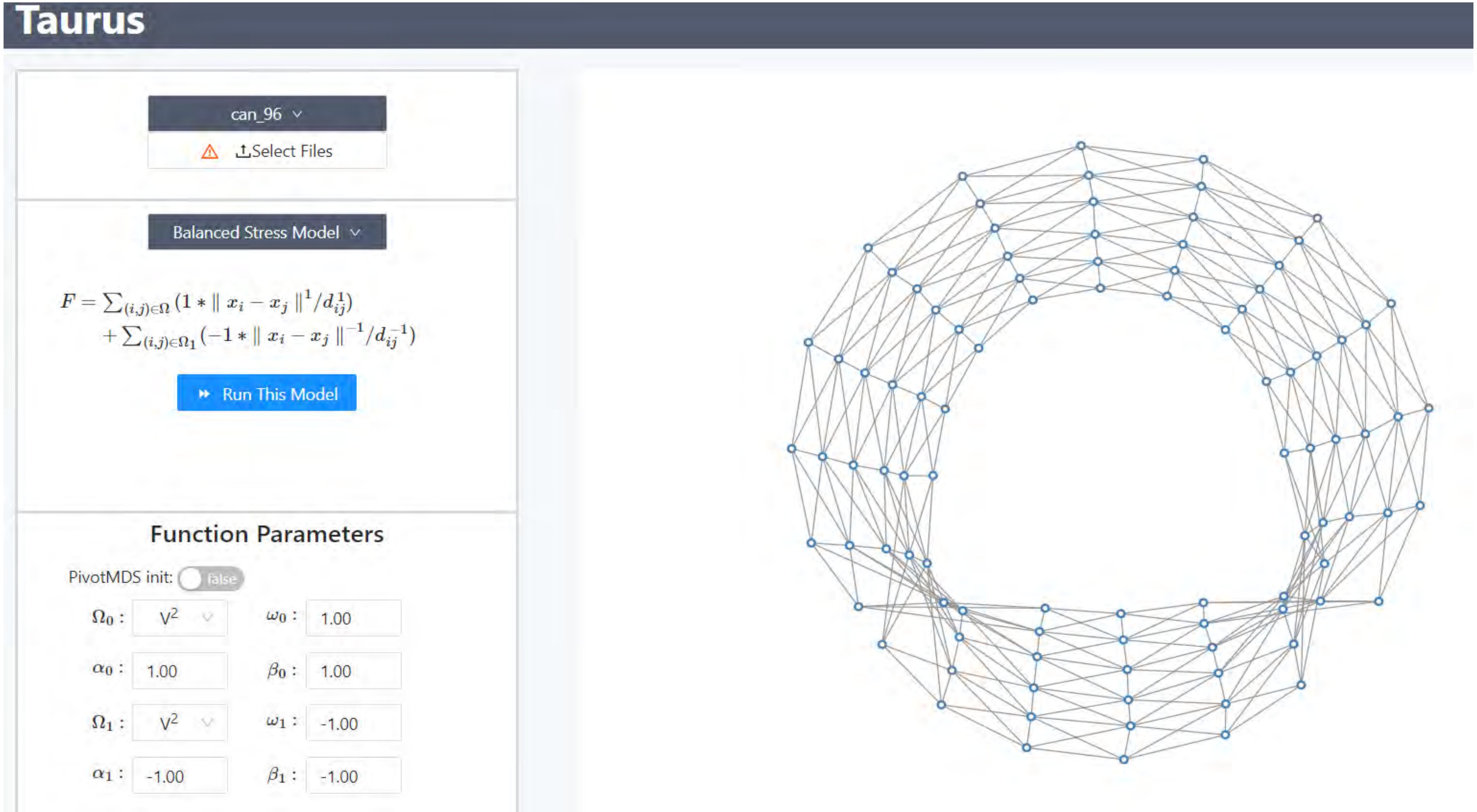
Balanced Stress Model

LinLog

User customized method

Code: https://github.com/Ideas-Laboratory/Taurus

Library & Demo



However, it can only handle small graphs so far

- This model depends on the graph shortest path distance, which is **hard** to compute for large graphs

$$F_{i,k} = \sum_{(i,j)\in\Omega_k} \omega_k * \frac{||\mathbf{x}_i - \mathbf{x}_j||^{\alpha_k}}{d_{ij}^{\beta_k}} * \mathbf{e}_{k,ij}, \quad F_i = \sum_{k}^{l} F_{i,k}$$

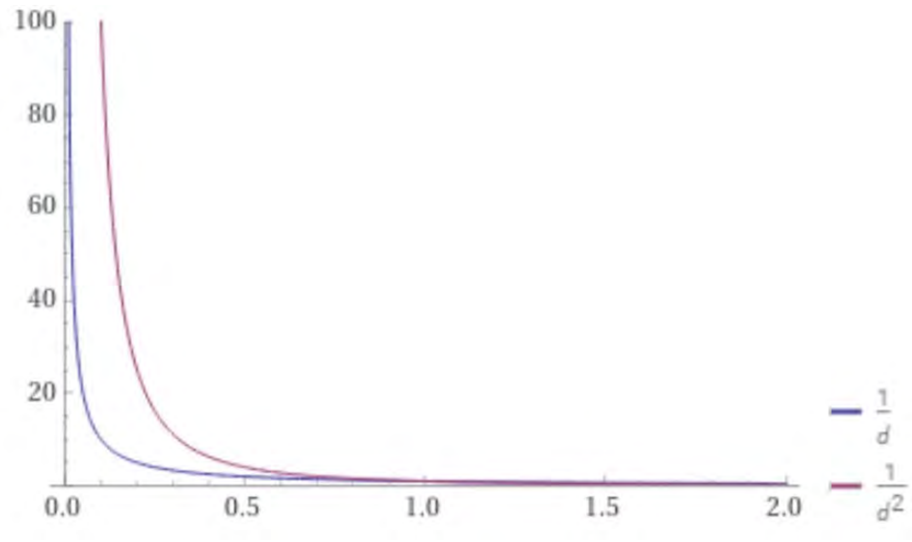
FDP does not have this issue!

P1: Nodes connected by an edge should be drawn close to each other

P2: Nodes should not be drawn too close to each other in general

Repulsive forces $F^r(i,j) = -||\mathbf{x}_i - \mathbf{x}_j||^{-q}, \quad i \neq j,$

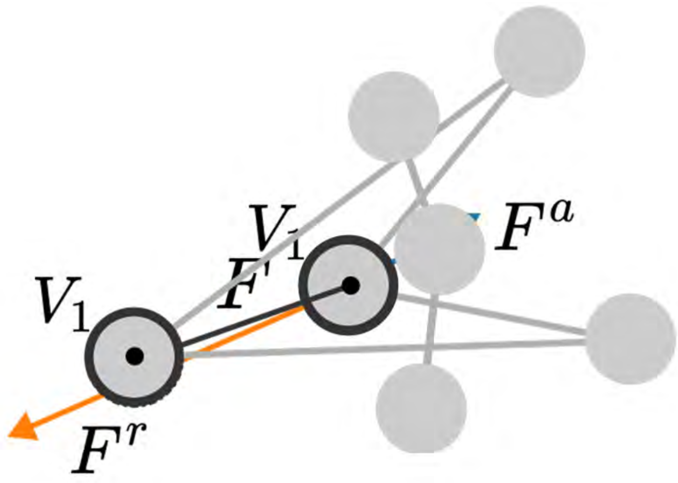
Attractive forces $F^a(i,j) = \alpha||\mathbf{x}_i - \mathbf{x}_j||^{p}, \quad i \leftrightarrow j,$

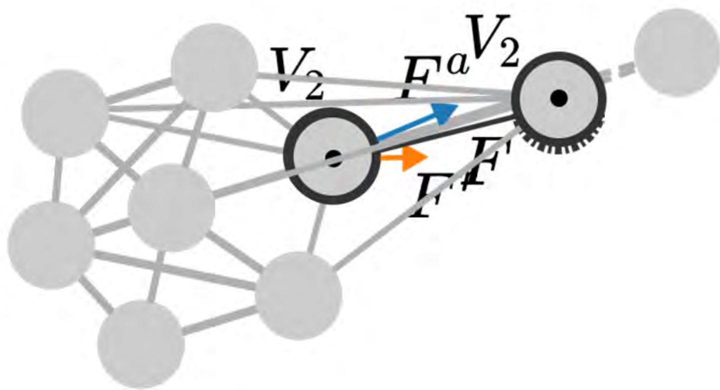


Does this model seem reasonable?

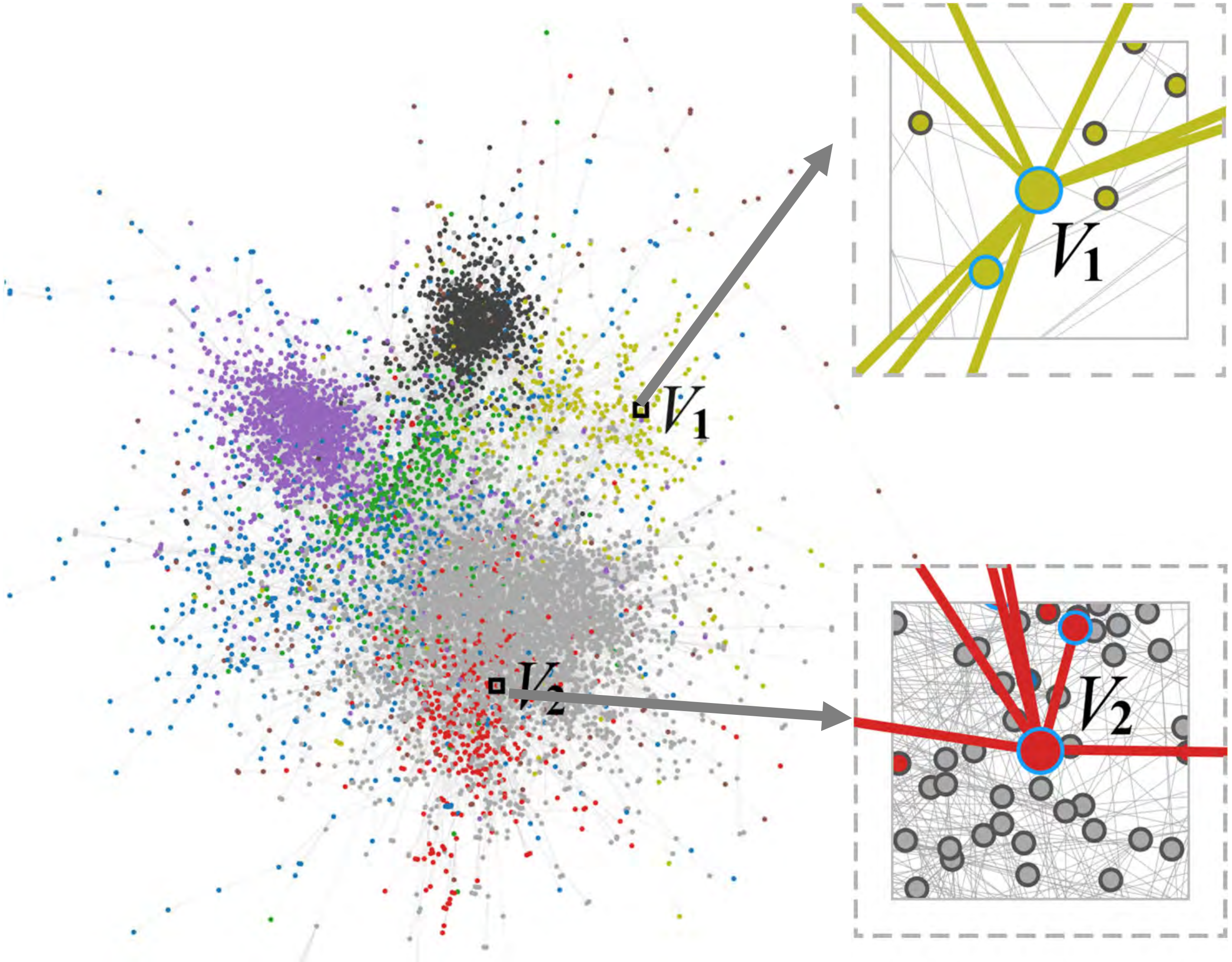
P1: Nodes connected by an edge should be drawn close to each other

P2: Nodes should not be drawn too close to each other in general

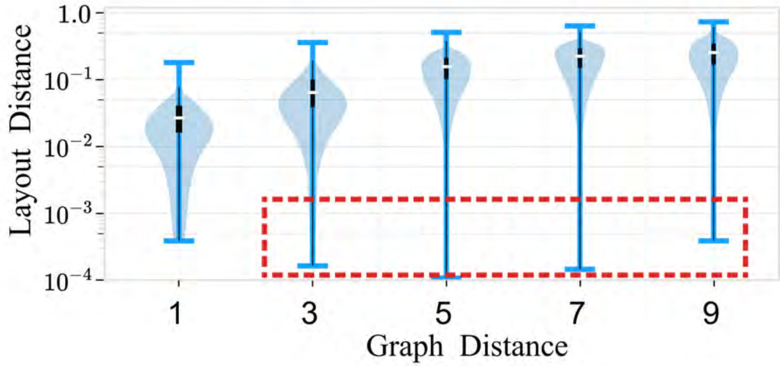




A real example



Nearby nodes are not neighbors



Many far-away nodes are nearby

Revisiting Force-directed placement

P1: Nodes connected by an edge should be drawn close to each other

P2: Nodes should not be drawn too close to each other in general

P3: Nodes connected by an edge should be drawn closer to each other than unconnected nodes

A desired short-range Force!

- 作为排斥力，短程应该弱一些，应该是有界的;

R1. $\exists \varsigma > 0 \text{ s.t. } 0 < f(d) \leq \varsigma, \ \forall d > 0;$

- 作为排斥力，应该是个短程力。无穷远处应该接近零;

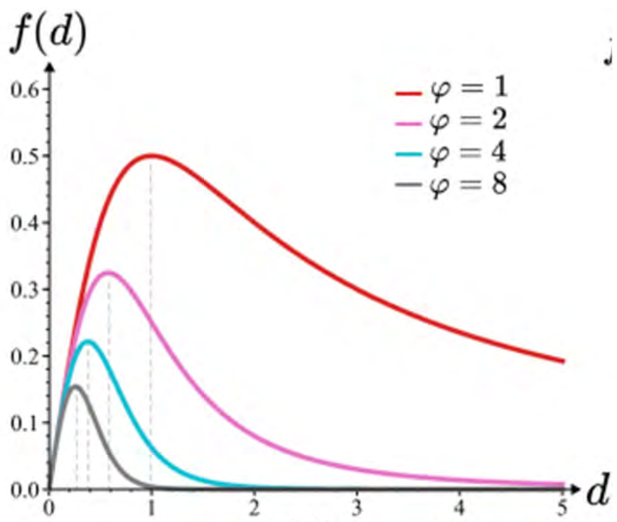
R2. $f(d) \sim d^{-q} \quad \text{as } d \to \infty, \ \text{where } q > 0;$

- 作为吸引力，短程应该接近线性，以便于增加短程吸引力而不改变长程;

R3. $f(d) \sim d \quad \text{as } d \to 0.$

基于t分布的力满足上述要求

$$f(d) = \frac{d}{(1 + d^2)^{\varphi}}$$



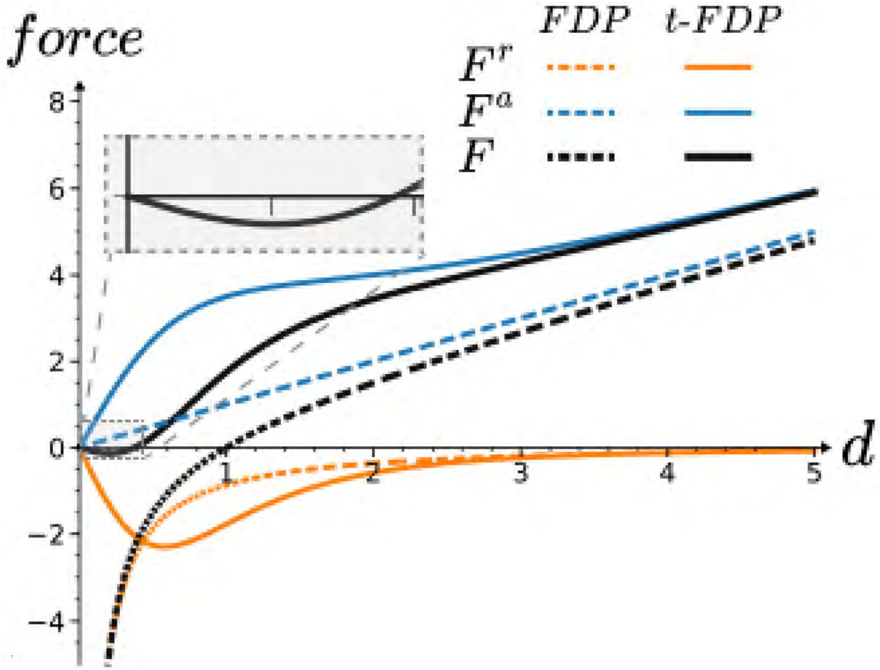
(a)

t-force

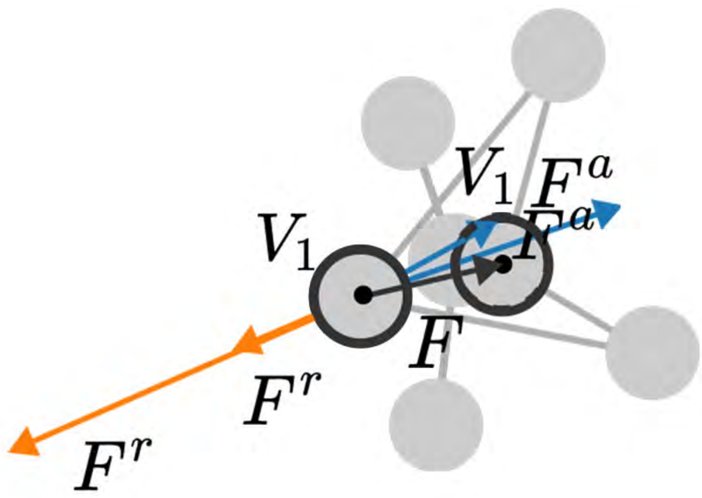
The t-FDP model

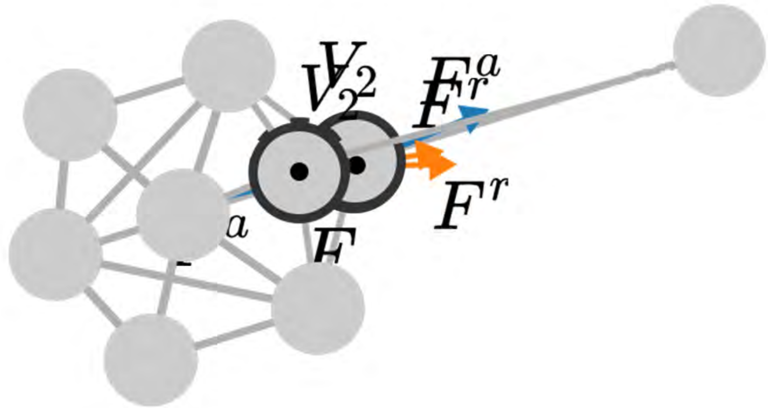
$$F^r(i,j) = -\frac{||\mathbf{x}_i - \mathbf{x}_j||}{(1 + ||\mathbf{x}_i - \mathbf{x}_j||^2)^\gamma} \frac{\mathbf{x}_i - \mathbf{x}_j}{||\mathbf{x}_i - \mathbf{x}_j||}$$

$$F^a(i,j) = \left(||\mathbf{x}_i - \mathbf{x}_j|| + \frac{\beta ||\mathbf{x}_i - \mathbf{x}_j||}{1 + ||\mathbf{x}_i - \mathbf{x}_j||^2} \right) \frac{\mathbf{x}_i - \mathbf{x}_j}{||\mathbf{x}_i - \mathbf{x}_j||}$$

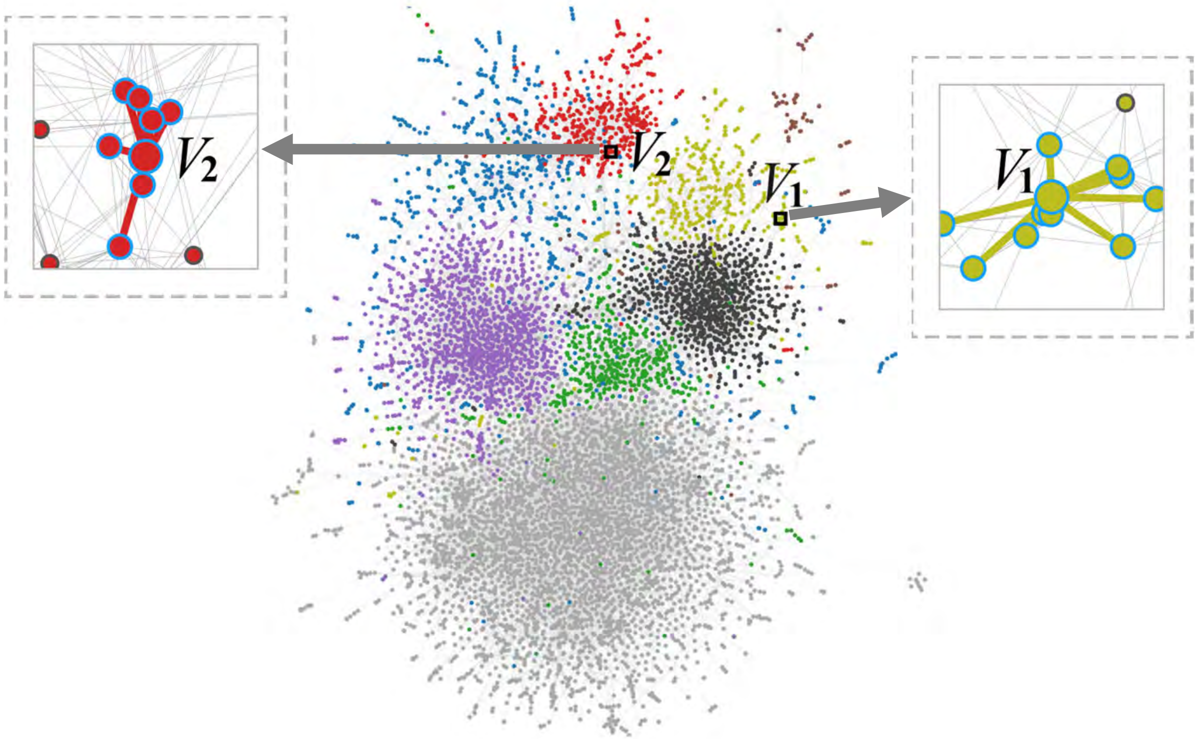


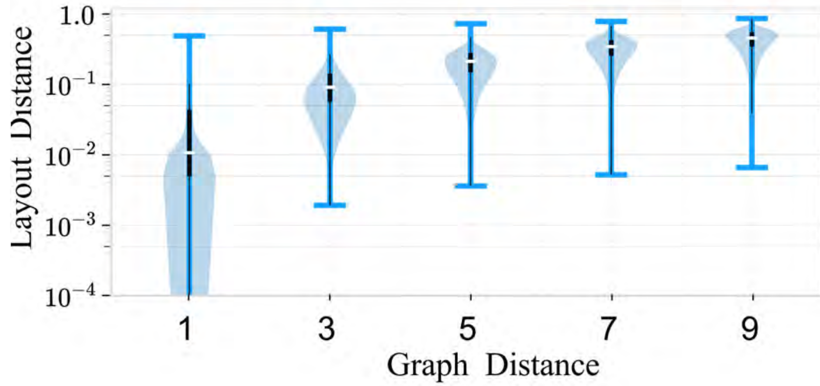
The t-FDP model





A Real Example





Convolution-based Formulation

$$
\begin{aligned}
F^r(i) &= \sum_{j=1, j\neq i}^{n} \frac{||\mathbf{x}_i - \mathbf{x}_j||}{(1+||\mathbf{x}_i - \mathbf{x}_j||^2)^\gamma} \frac{\mathbf{x}_i - \mathbf{x}_j}{||\mathbf{x}_i - \mathbf{x}_j||} \\
&= \sum_{j=1, j\neq i}^{n} \frac{\mathbf{x}_i - \mathbf{x}_j}{(1+||\mathbf{x}_i - \mathbf{x}_j||^2)^\gamma} \\
&= \mathbf{x}_i \sum_{j=1}^{n} \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) - \sum_{j=1}^{n} \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j)\mathbf{x}_j
\end{aligned}
$$

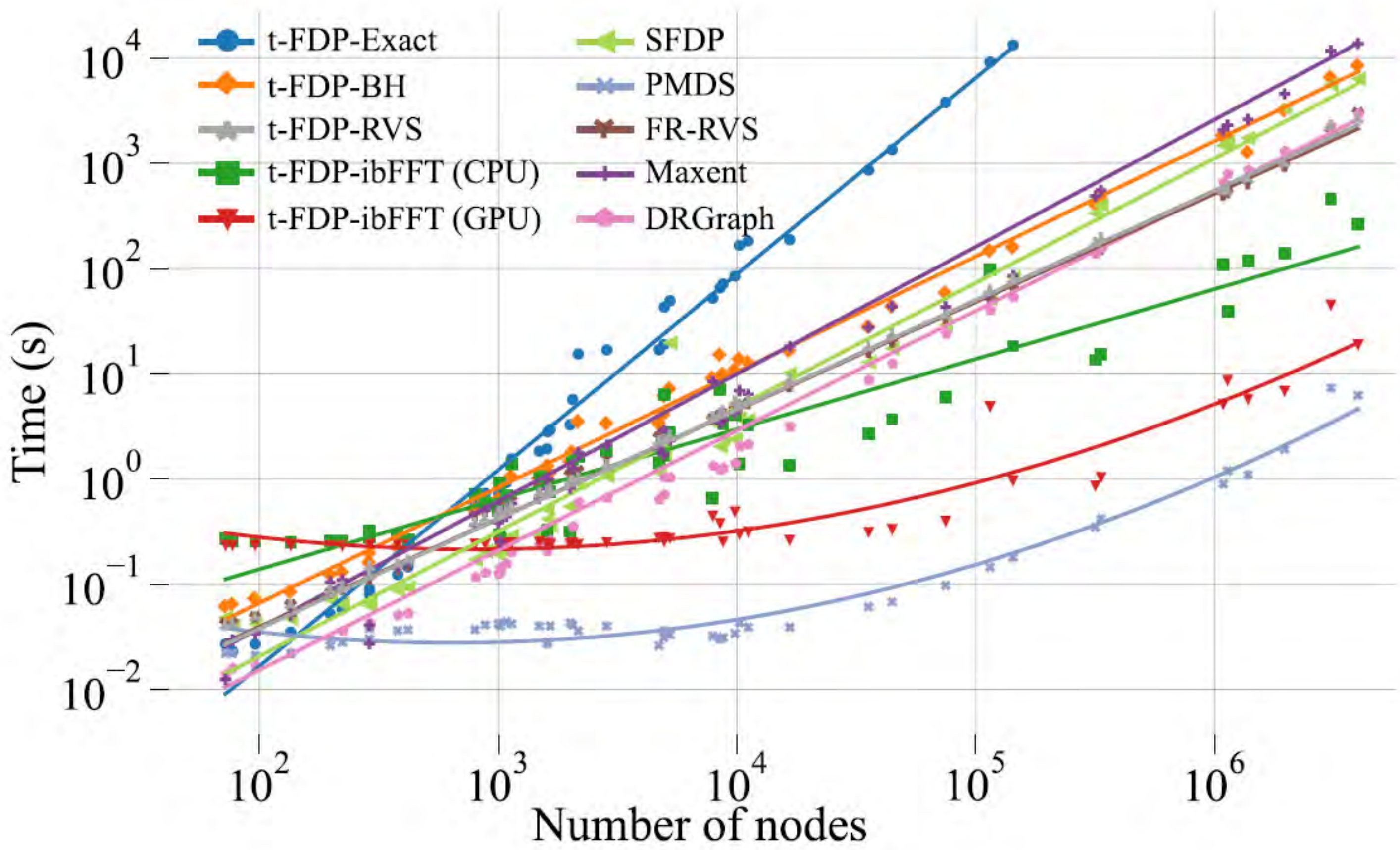
$$
\mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{(1+||\mathbf{x}_i - \mathbf{x}_j||^2)^\gamma}.
$$

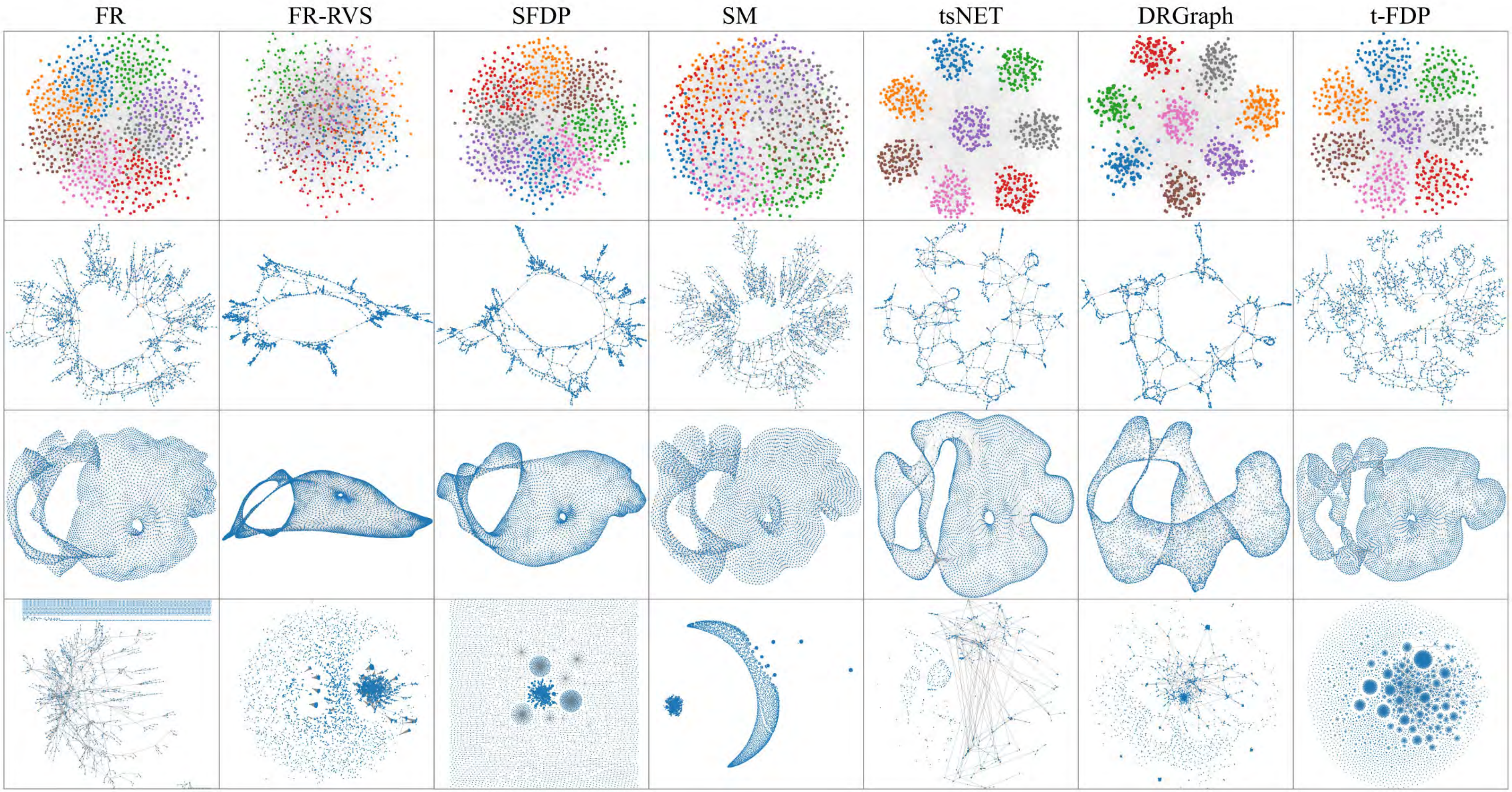
$$
\psi(\mathbf{x}_i) = \sum_{j=1}^{n} \mathbf{K}(\mathbf{x}_i, \mathbf{x}_j) v_j.
$$

FFT-based Acceleration

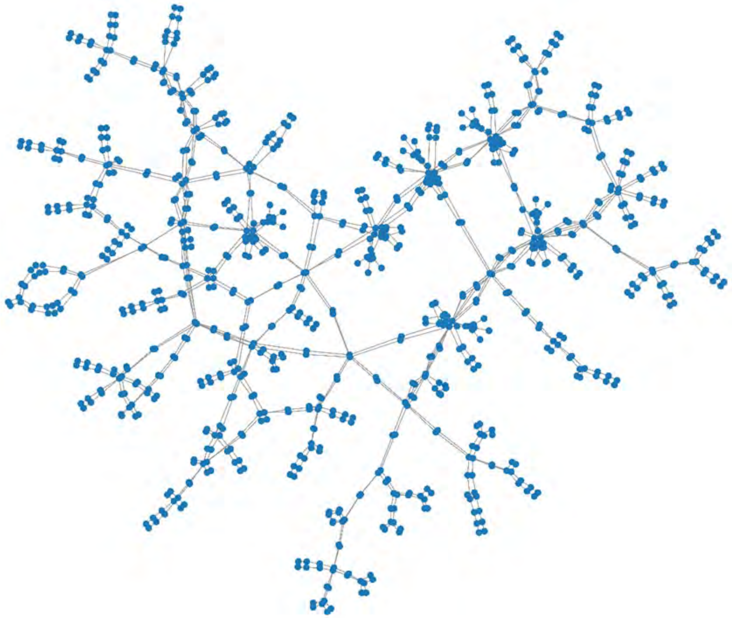
- projecting all data points $\mathbf{x}_i$ onto the grid by using Lagrange polynomials with a time complexity $O(k^2 n)$;
- computing the interaction of the grid nodes, which can be accelerated by FFT with a complexity $O(2k^2 \log k)$; and
- back-projecting the interaction of all grid nodes to the original points with time complexity $O(k^2 n)$.

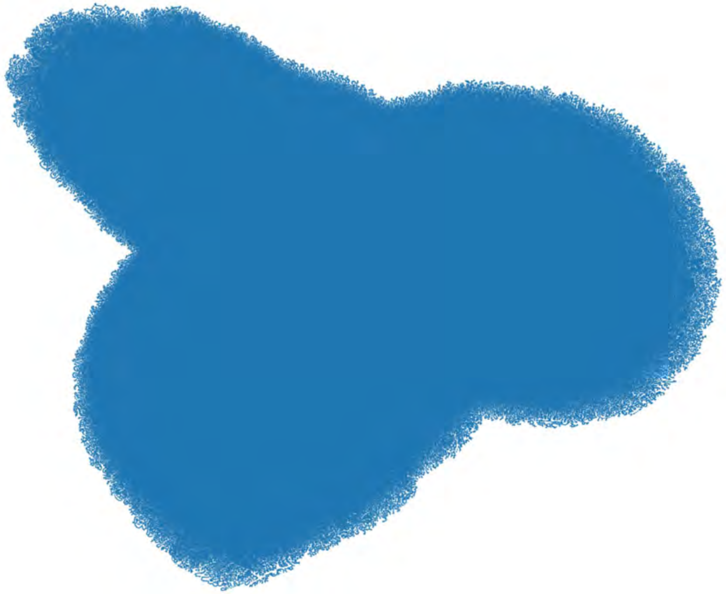
Performance Comparison

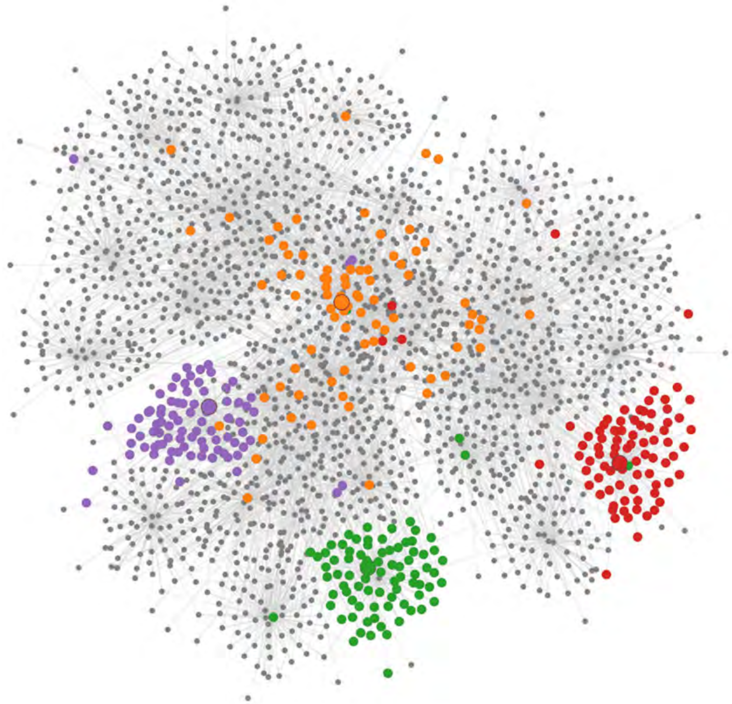



FR
FR-RVS
SFDP
SM
tsNET
DRGraph
t-FDP

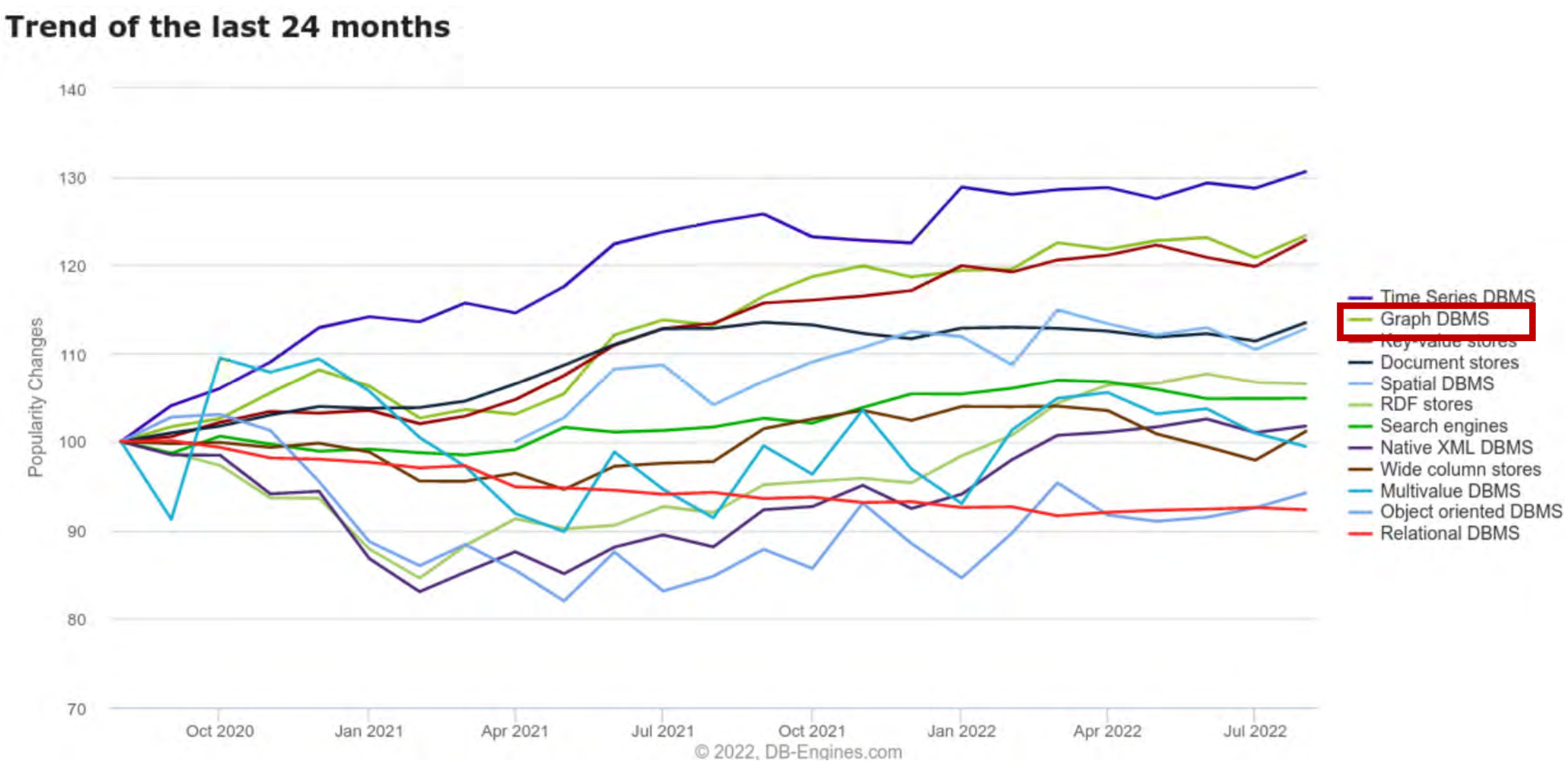
Interactive Exploration



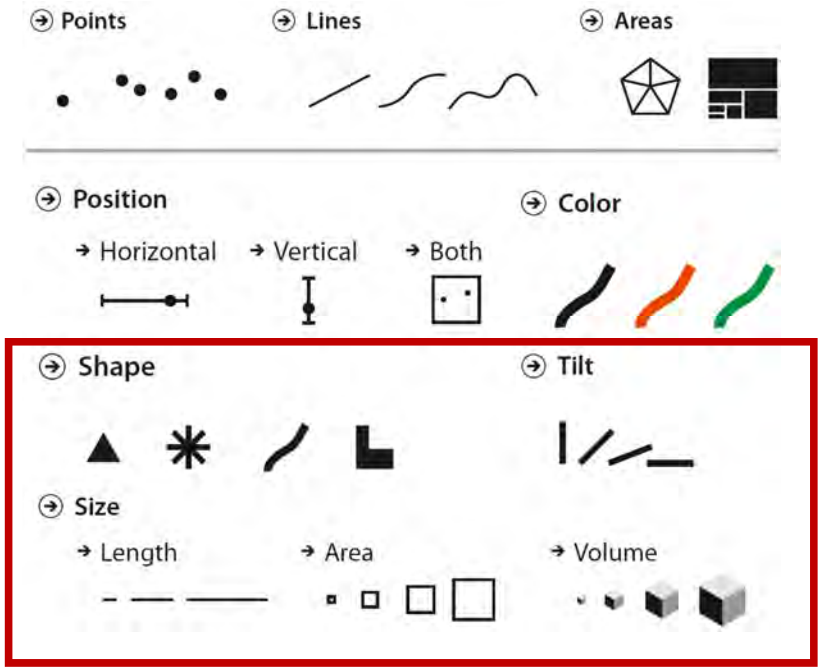


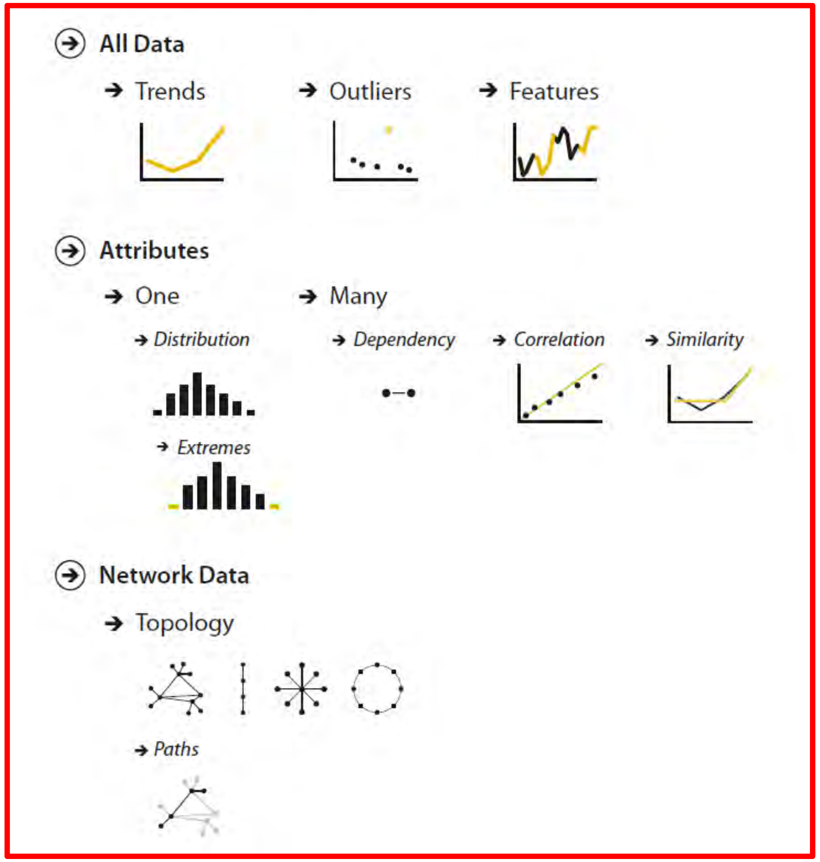


But we have not solve the issue of even large graphs

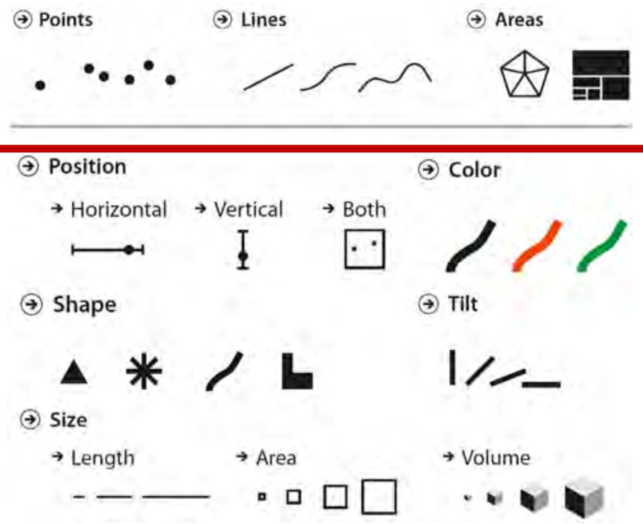


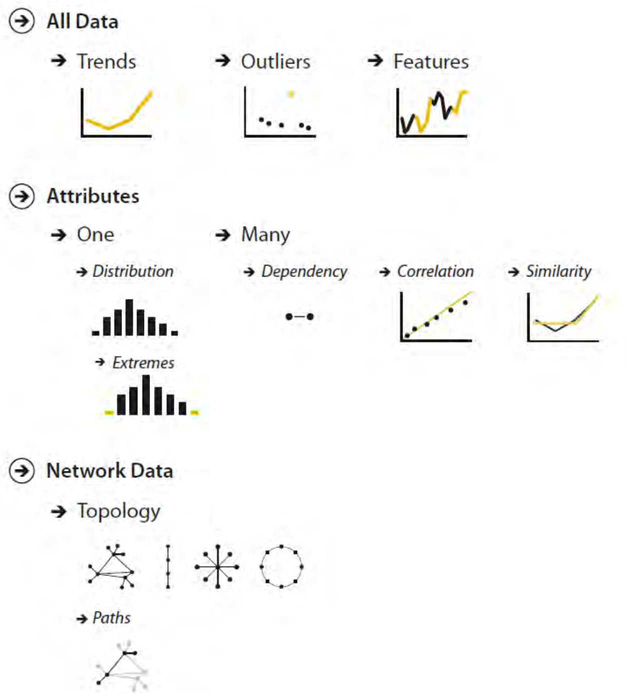
Data Size x Design space x Task Targets





Data Size x Design space x Task Targets x Users





- Library/Toolkit Designer
- Application Programmer
- End-user

Chart Animation Authoring Tools – Programming Tools

Low-level Grammar

D3

- A transition operator
- A collection of interpolation functions

StarDust

- Produce animation with better performance
- Similar API as D3

Pros: Facilitating the creation of highly-customized animations
Cons: Verbose specification

High-level Grammar

gganimate

- High-level grammar
- Only support charts created with ggplot2

Vega-Lite

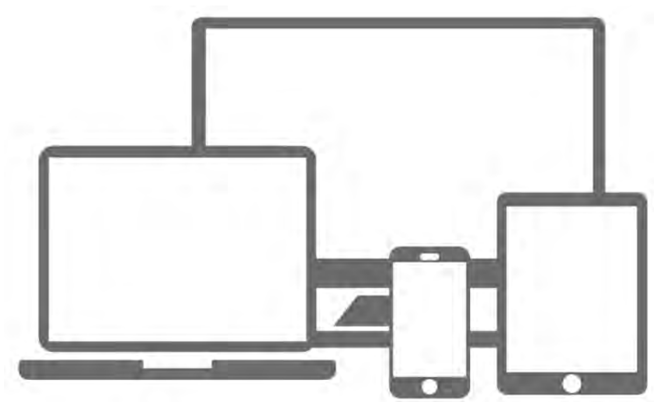
- Concise specifications for visualizations
- Limited support for animations

Pros: Easy to understand
Cons: Steep learning curve

Declarative Language Design for Chart Animations

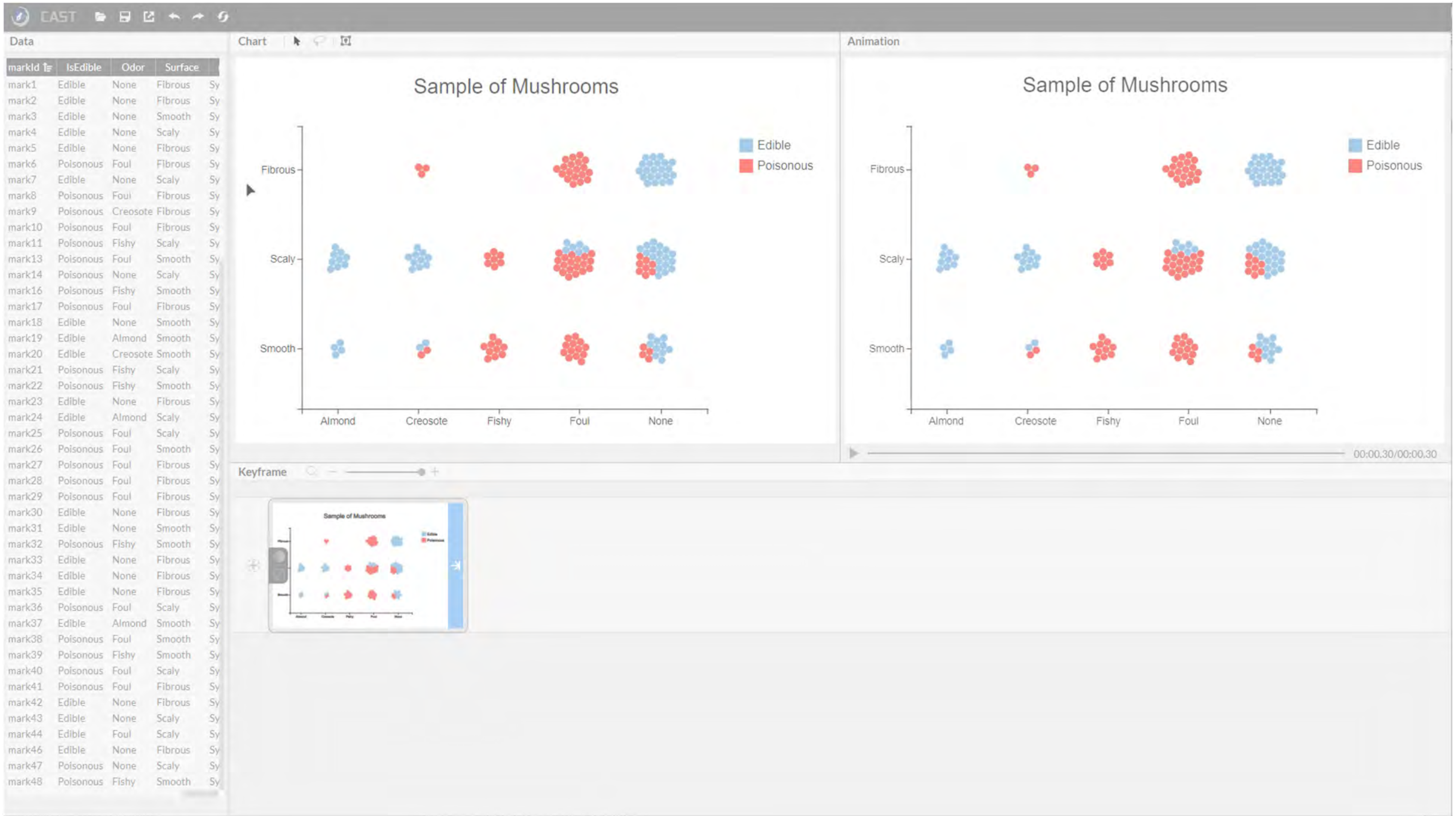
- **Canis** - A **high-level grammar** that enables declarative specifications of data-driven chart animations
- **Canis compiler** that automatically synthesizes Lottie JSON specifications, which support rendering animations across multiple platforms.

```
{
  "charts":
    [{"source": "./donut.dsvg"}],
  "animations": [{
    "selector": ".path",
    "effects": [{
      "type": "wipe bottom",
      "easing": "easeInCubic",
      "duration": 1200
    }]
  }]
}
```



https://chartanimation.github.io/canis/ (homepage)
https://github.com/ChartAnimation/canis-compiler (github)

Visual Programming of Chart Animations



ACM CHI 2021, Best Paper Honorable Mention Award

Project page: https://chartanimation.github.io/

Summary

- Color and Sampling for Point-based visualizations
- Interaction of Density map for line-based visualizations
- Mathematical layout models for node-link diagrams (point + line)

敬请指正

谢谢！

http://www.yunhaiwang.net/

cloudseawang@gmail.com