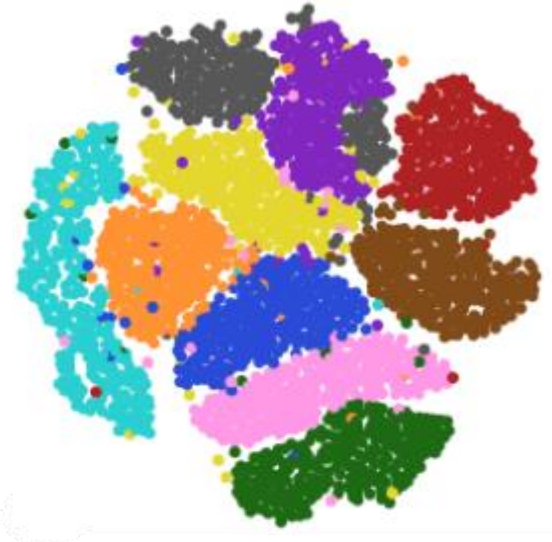
Methods for Reducing Visual Clutter

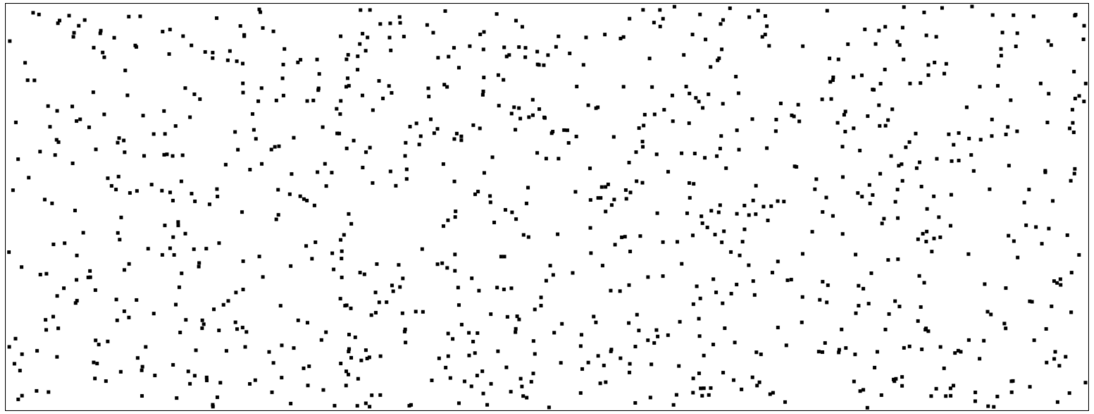
Yu-Shuen Wang, CS, NYCU

Original data

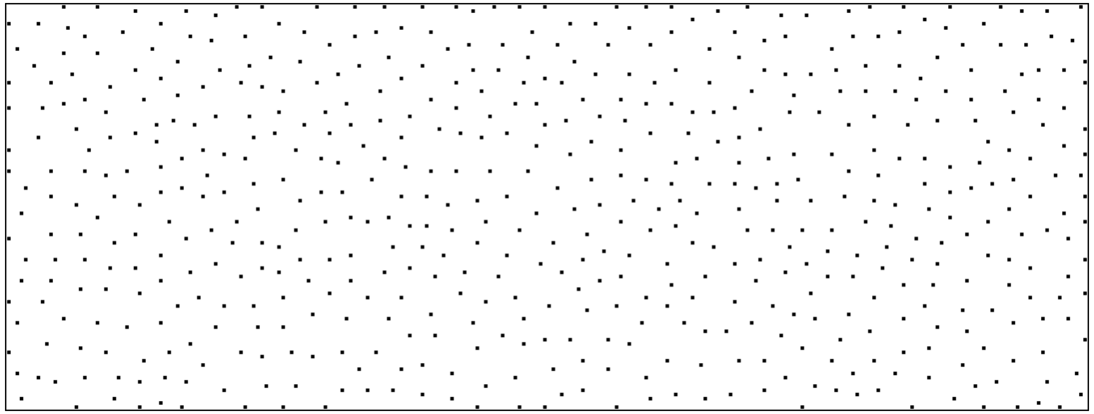
Random sampling



Blue noise sampling



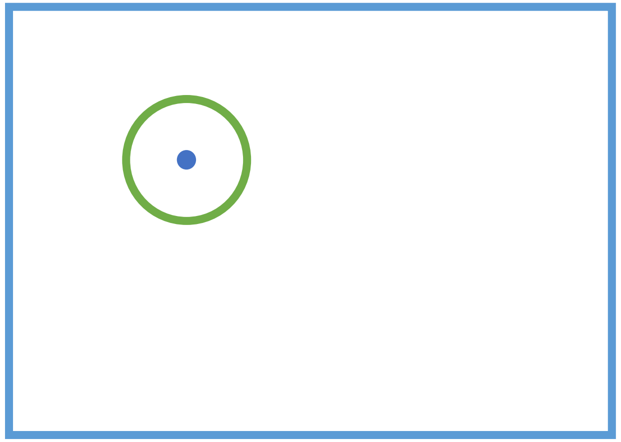
Random noise



Blue noise sampling

Blue noise sampling

- Suppose we are sampling the 2D space $D = [0,1]^2$. When the space is empty, we have a constant pdf equal to one for the entire space: $f_{X,Y}(x, y) = 1$.
- After inserting the first sample, a rejection disk with radius r (the minimum distance between samples) centered on the sample will have zero value in the pdf. The remaining space will now have a pdf equal to $1/(1 - \pi r^2)$.
- For subsequent samples, the analysis is the same, except that some disks may overlap.



Efficient Computation of Blue Noise Point Sets through Importance Sampling

- Break down the 2D pdf $f_{X;Y}(x;y)$ into a 1D conditional pdf and a 1D pdf: $f_{X,Y}(x,y) = f_{Y|X}(y|x)f_X(x)$, where $f_X(x)$ is the marginal density function of x found by integrating $f_{X,Y}(x,y)$ over y:

$$f_X(x) = \int_0^1 f_{X,Y}(x,y)dy$$

- Compute the coordinates (x_i, y_i) of a new sample i in two steps:
  1. random generation of x_i based on $f_X(x)$ and then
  2. selecting y_i randomly according to $f_{Y|X}(y|x_i)$.

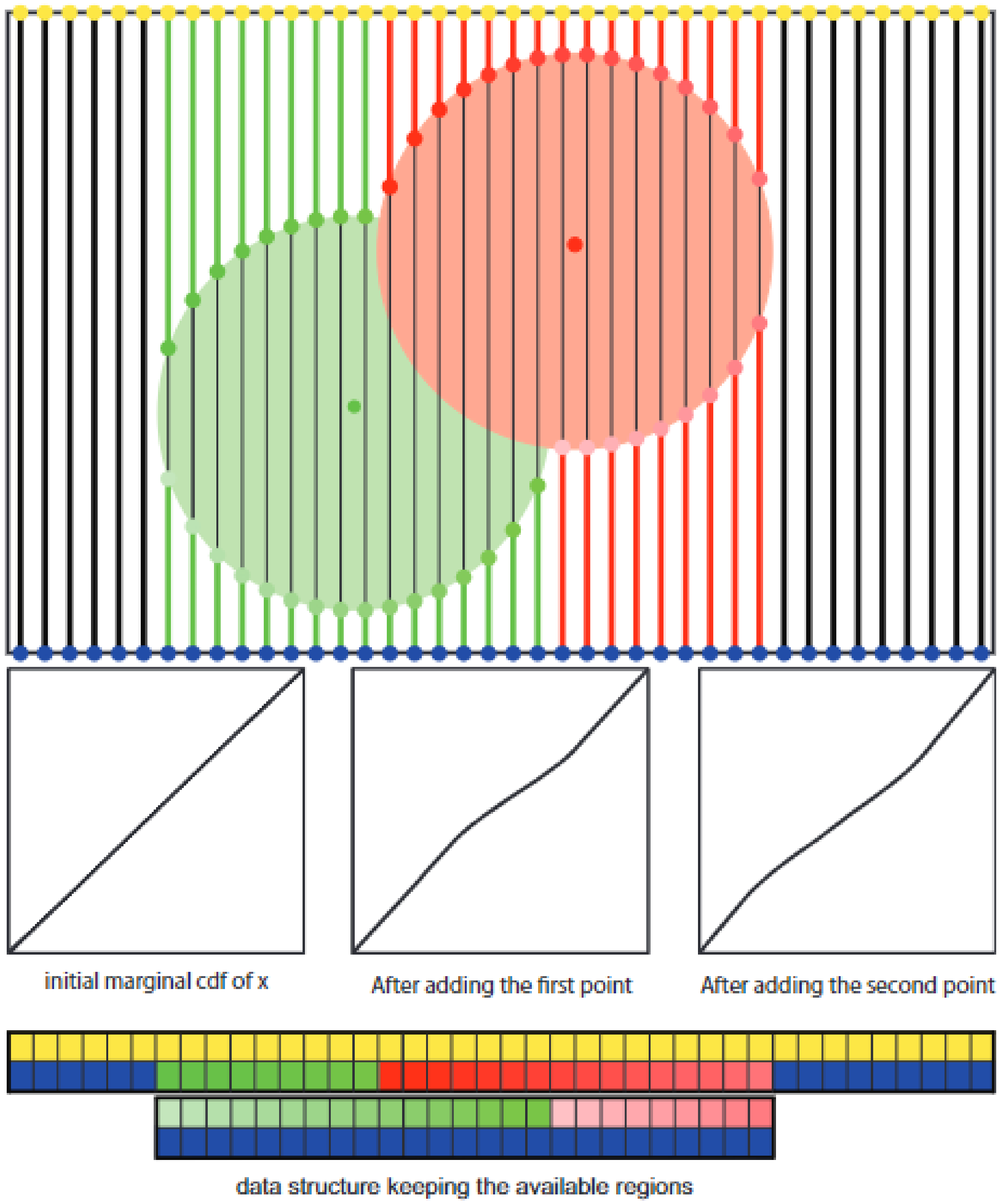
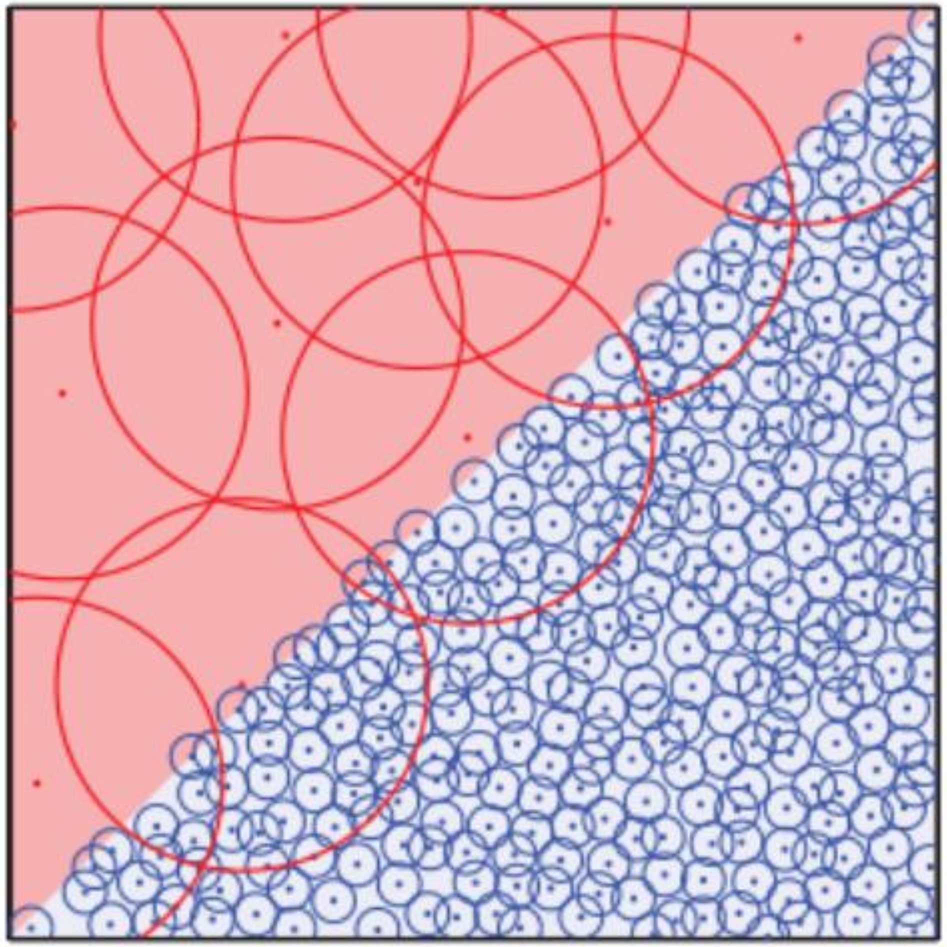
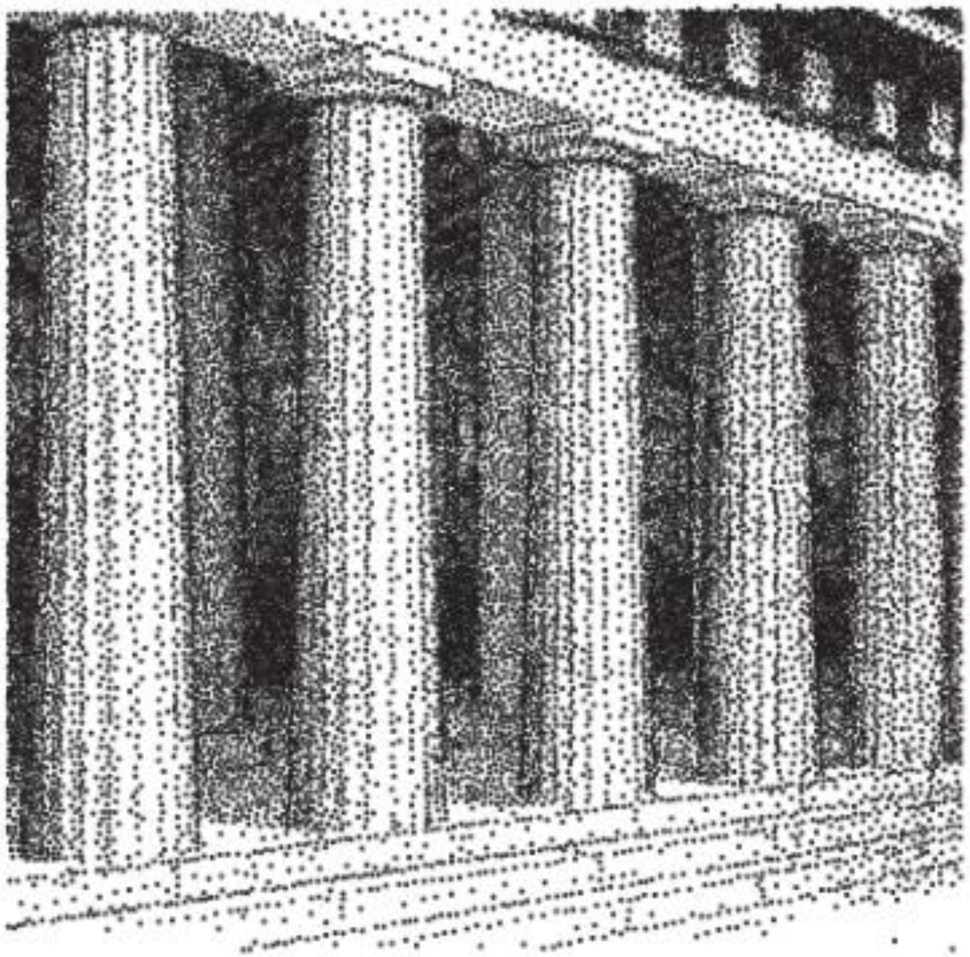


Figure 1: ***Visualization of the sampling process.*** *The first sample is shown with a red dot with the rejection disk around it in light red. The black lines inside the red disk represent regions where the pdf has been zeroed out and are subtracted from the marginal cdf. The red lines indicate available regions which are added to our availability data structure representing $f_{Y|X}(y|x)$, shown at the bottom. We have color coded the start and end points of each span to show the entry in the data structure. For the next sample, shown with a green dot, we update the cdf by using the black lines that are inside the green circle but outside the red one. The available span data structure now contains two spans in the middle positions. The cdf plot has been exaggerated for illustration purposes so the effect of subtracting the regions inside the circles can be seen.*

Variable-density Sampling

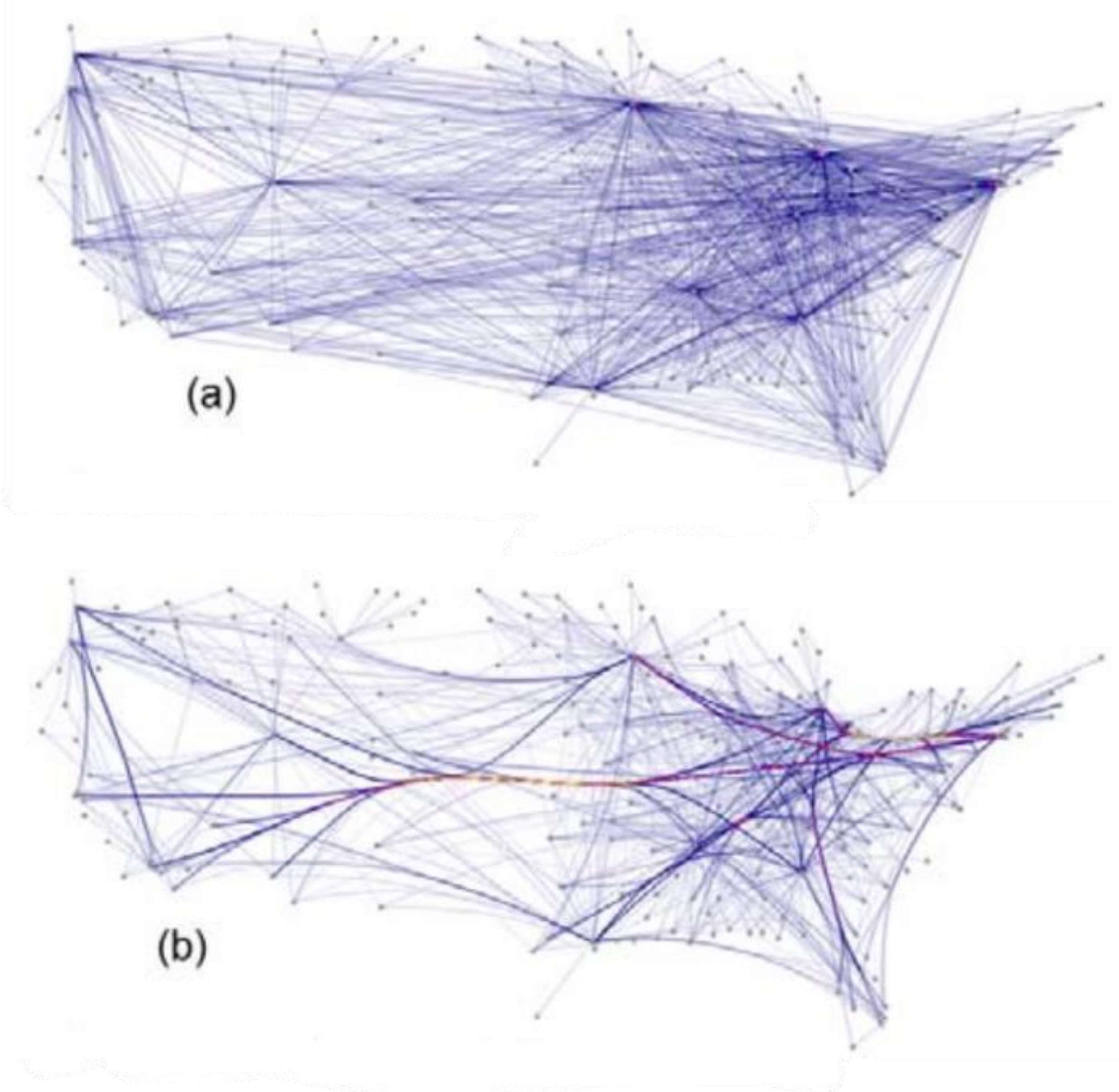




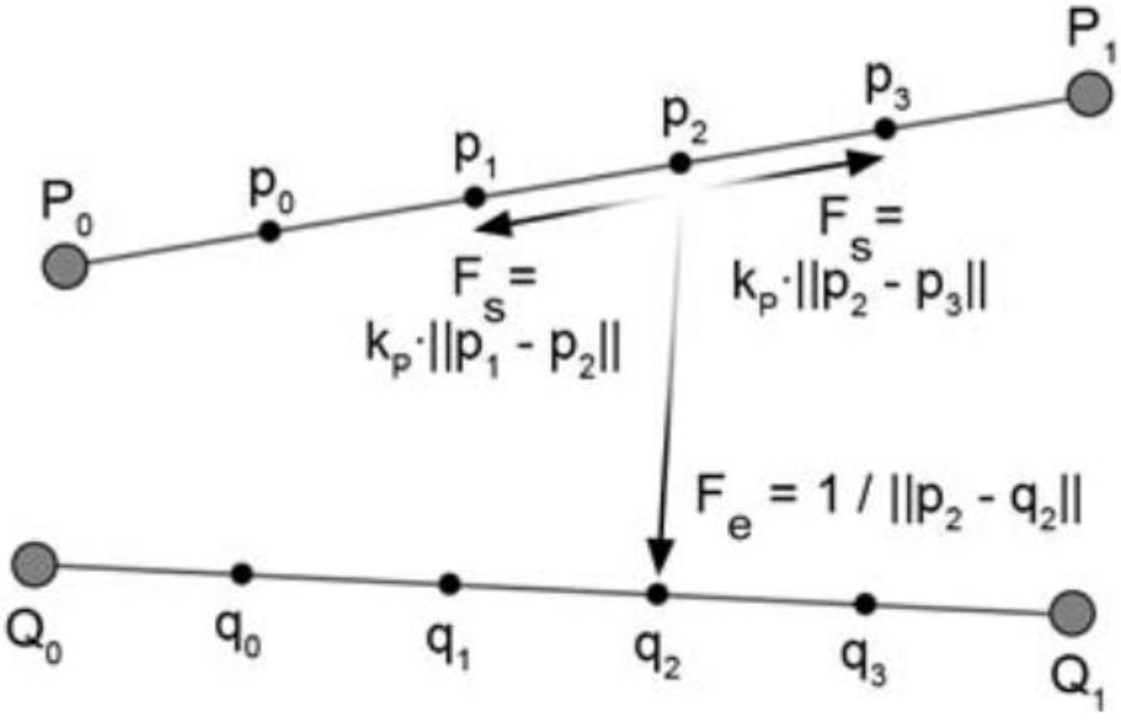
Blue noise data sampling

1. Randomly sample an item, (supposedly) located at x, from the dataset.
2. Calculate the density $D(x)$ around x.
3. Set the radius r based on $D(x)$.
   - High density -> small r
   - Low density -> large r
4. Remove all data items if their distances to x are smaller than r

Force-Directed Edge Bundling for Graph Visualization



US airlines graph (235 nodes, 2101 edges)



$$\mathbf{F}_{\mathbf{p}_i} = k_P \cdot (\|p_{i-1} - p_i\| + \|p_i - p_{i+1}\|) + \sum_{Q \in E} \frac{1}{\|p_i - q_i\|},$$

with

k_p	:	spring constant for each segment of edge P,
E	:	set of all interacting edges except edge P.

Angle compatibility

$C_a(P,Q) \in [0,1]$ as

$$C_a(P,Q) = |\cos(\alpha)|,$$

with

$$\alpha \quad : \quad \arccos\left(\frac{P \cdot Q}{|P||Q|}\right).$$

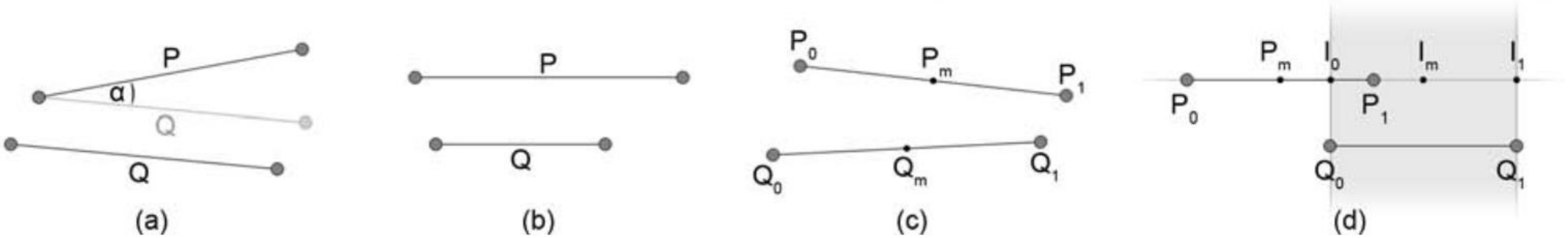
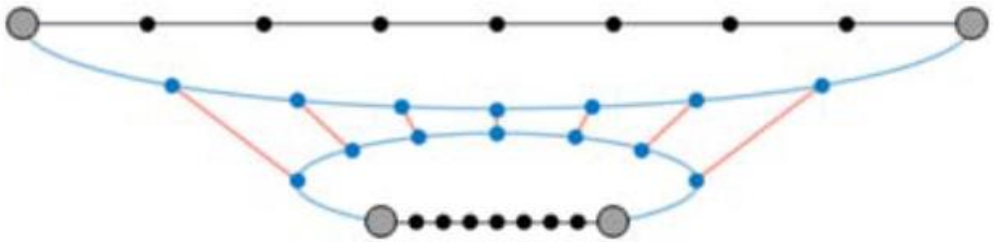


Figure 3: *Geometric concepts and situations necessary to illustrate the edge compatibility measures (a) angle compatibility C_a, (b) scale compatibility C_s, (c) position compatibility C_p, and (d) visibility compatibility C_v.*



Scale compatibility

$C_s(P,Q) \in [0,1]$ as

$$C_s(P,Q) = \frac{2}{l_{avg} \cdot \min(|P|,|Q|) + \max(|P|,|Q|)/l_{avg}},$$

with

$$l_{avg} \quad : \quad \frac{|P|+|Q|}{2}.$$

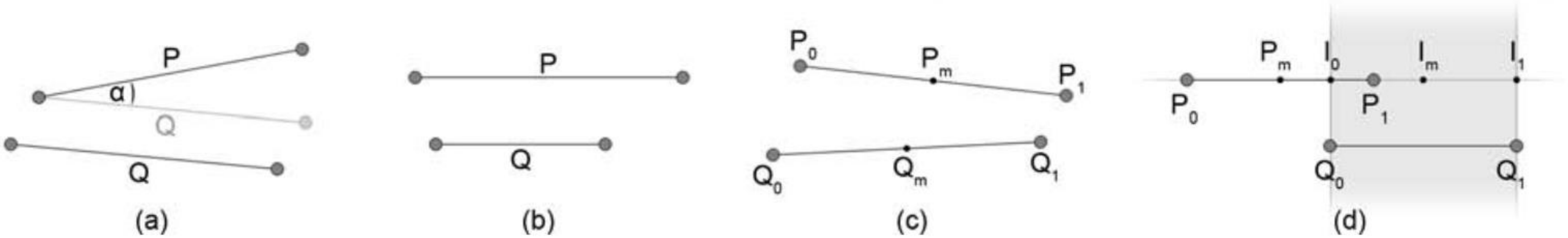


Figure 3: *Geometric concepts and situations necessary to illustrate the edge compatibility measures (a) angle compatibility C_a, (b) scale compatibility C_s, (c) position compatibility C_p, and (d) visibility compatibility C_v.*

Position compatibility

$C_p(P,Q) \in [0,1]$ as

$$C_p(P,Q) = l_{avg}/(l_{avg} + \|P_m - Q_m\|),$$

with

P_m and Q_m : midpoints of edges P and Q.

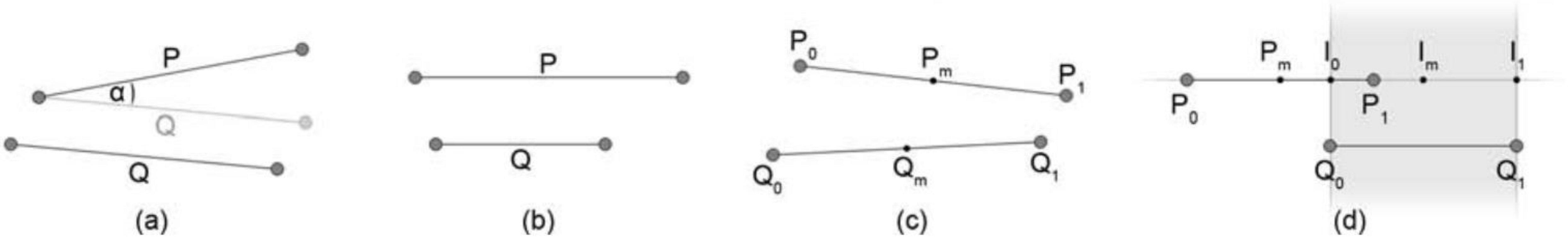


Figure 3: *Geometric concepts and situations necessary to illustrate the edge compatibility measures (a) angle compatibility C_a, (b) scale compatibility C_s, (c) position compatibility C_p, and (d) visibility compatibility C_v.*

Visibility compatibility

$C_v(P,Q) \in [0,1]$ as

$$C_v(P,Q) = \min(V(P,Q), V(Q,P)),$$

with

$$V(P,Q) \quad : \quad \max\left(1 - \frac{2\|P_m - I_m\|}{\|I_0 - I_1\|}, 0\right),$$
$$I_m \quad : \quad \text{midpoint of } I_0 \text{ and } I_1.$$

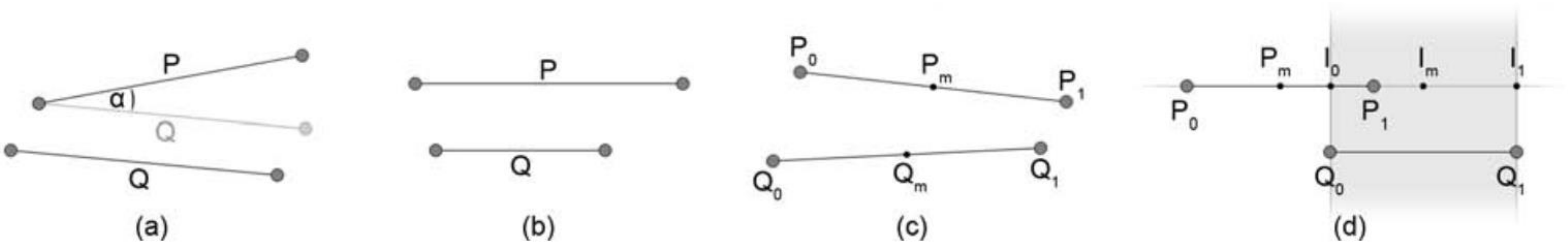


Figure 3: *Geometric concepts and situations necessary to illustrate the edge compatibility measures (a) angle compatibility C_a, (b) scale compatibility C_s, (c) position compatibility C_p, and (d) visibility compatibility C_v.*

We define the total edge compatibility $C_e(P,Q) \in [0,1]$ between two edges P and Q as

$$C_e(P,Q) = C_a(P,Q) \cdot C_s(P,Q) \cdot C_p(P,Q) \cdot C_v(P,Q).$$

The combined force $\mathbf{F_{p_i}}$ is now redefined as

$$\mathbf{F_{p_i}} = k_P \cdot (\|p_{i-1} - p_i\| + \|p_i - p_{i+1}\|) + \sum_{Q \in E} \frac{C_e(P,Q)}{\|p_i - q_i\|}.$$

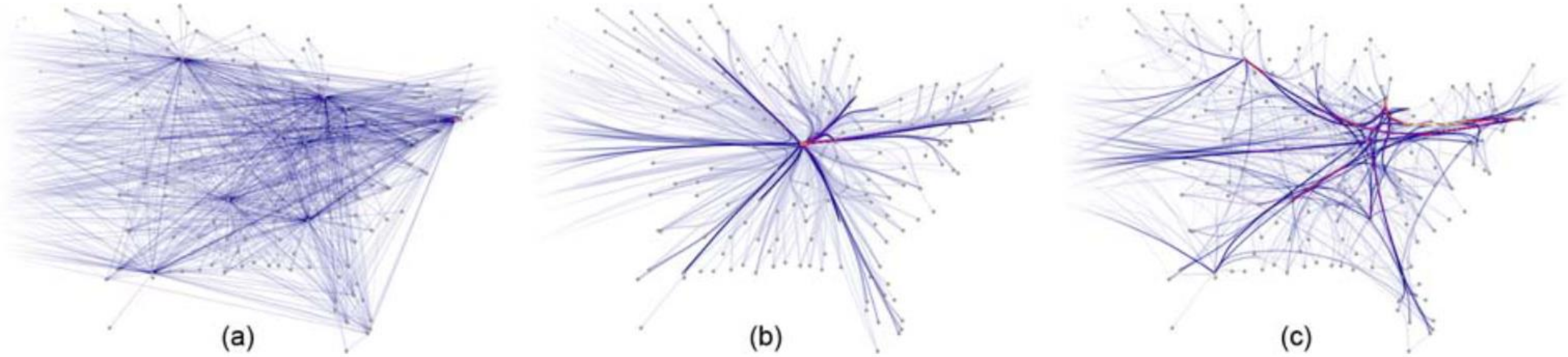
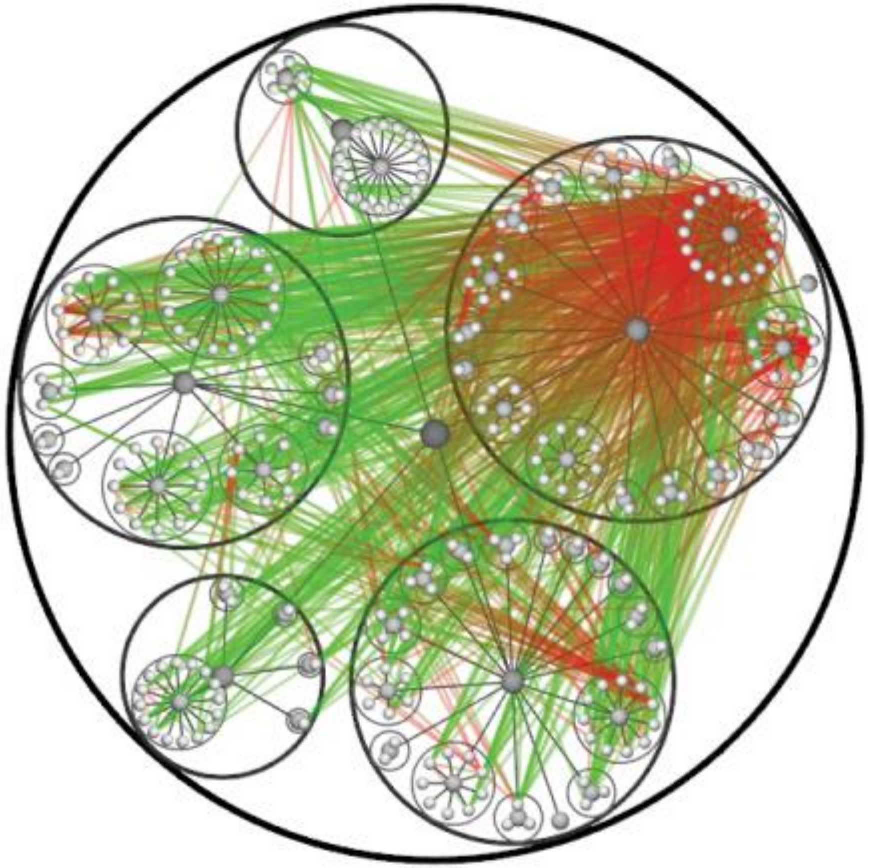


Figure 2: *Part of (a) a straight-line graph that is bundled (b) without and (c) with edge compatibility measures. These measures reduce the amount of bundling between incompatible edges while retaining it in parts of the graph where this is desirable.*

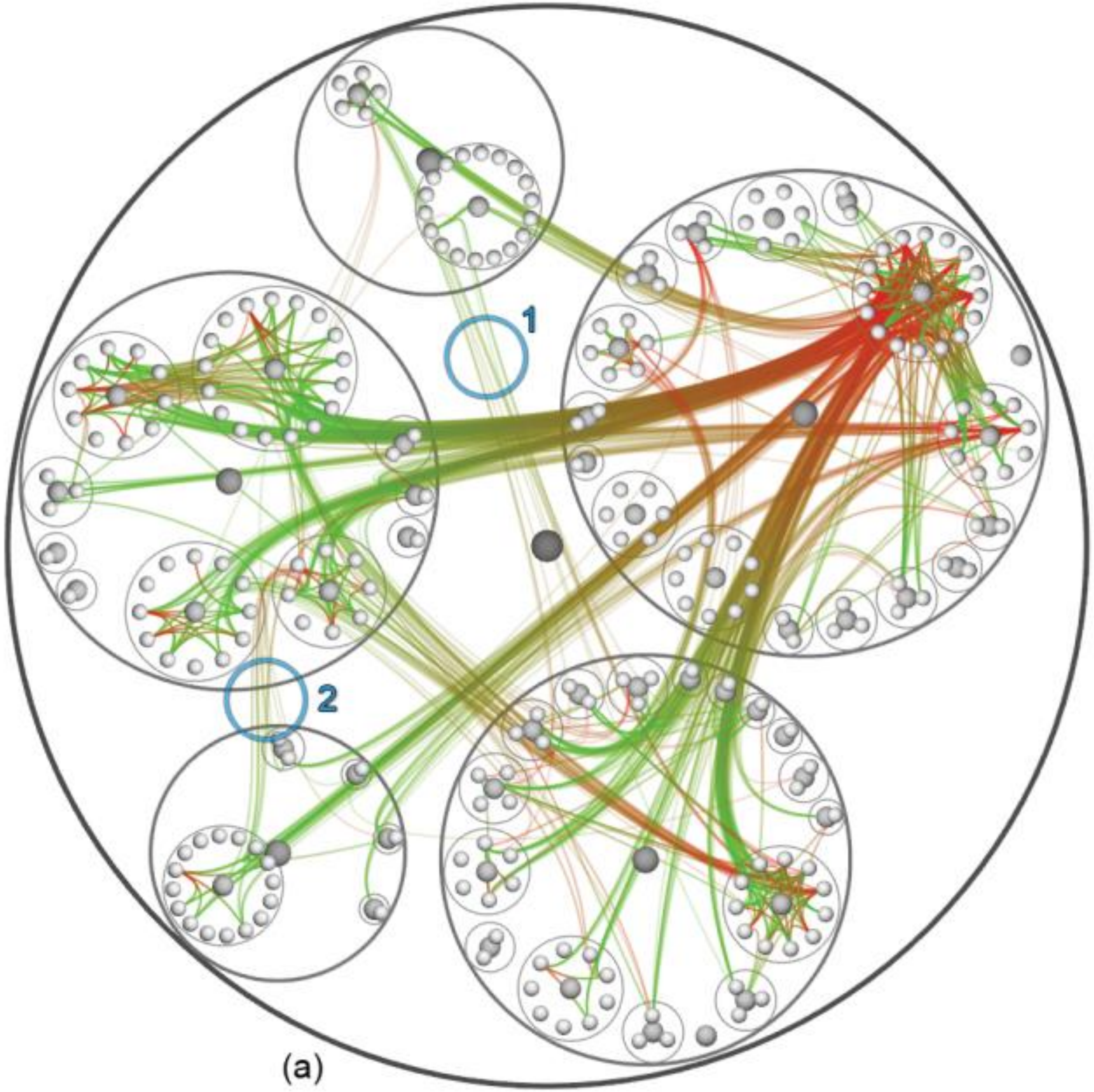
Iterative refinement scheme

- The simulation starts with an initial number of subdivision points P_0 for each edge and an initial step size S_0.
- A fixed number of simulation cycles C is performed.
- In each cycle, a specific number of iteration steps I is performed.

cycle	0	1	2	3	4	5
P	1	2	4	8	16	32
S	.04	.02	.01	.005	.0025	.00125
I	50	33	22	15	9	7

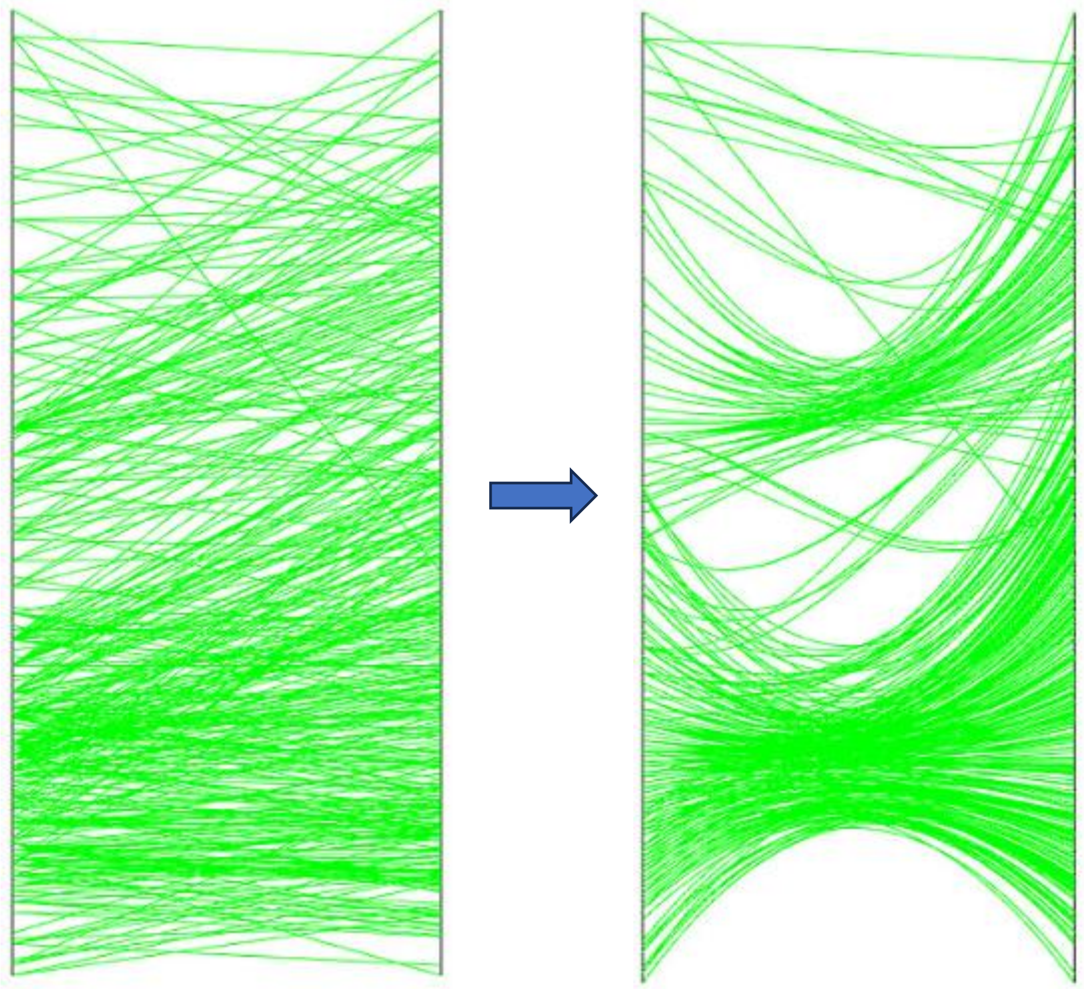


(a)



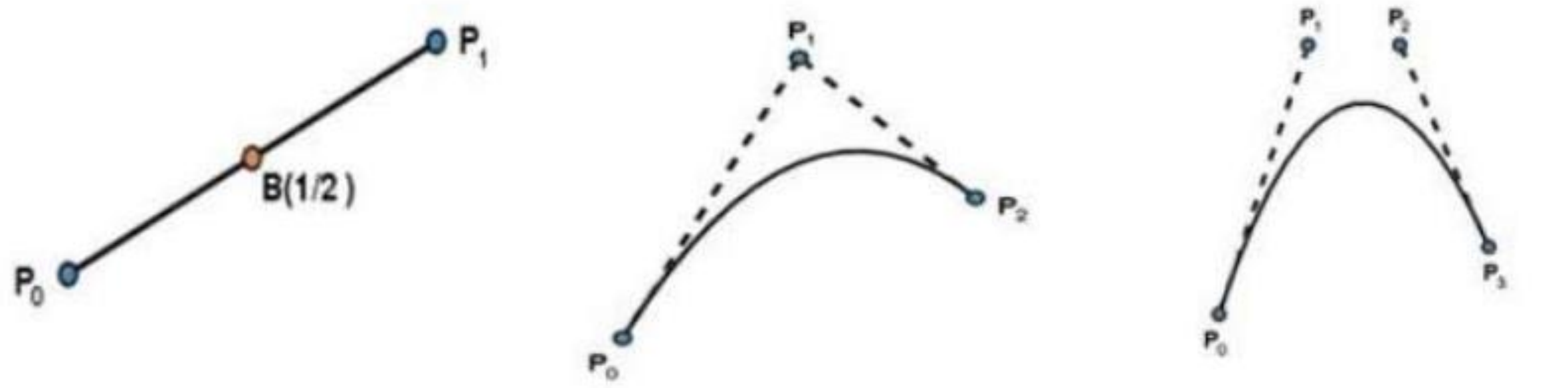
(a)

Energy-based Visual Clustering in Parallel Coordinates



Bezier Curves

$$\sum_{k=0}^{n} P_i B_i^n(t) \qquad B_i^n(t) = \binom{n}{i}(1-t)^{n-i}t^i$$



Simple Bezier Curve Quadratic Bazier Curve Cubic Bazier Curve

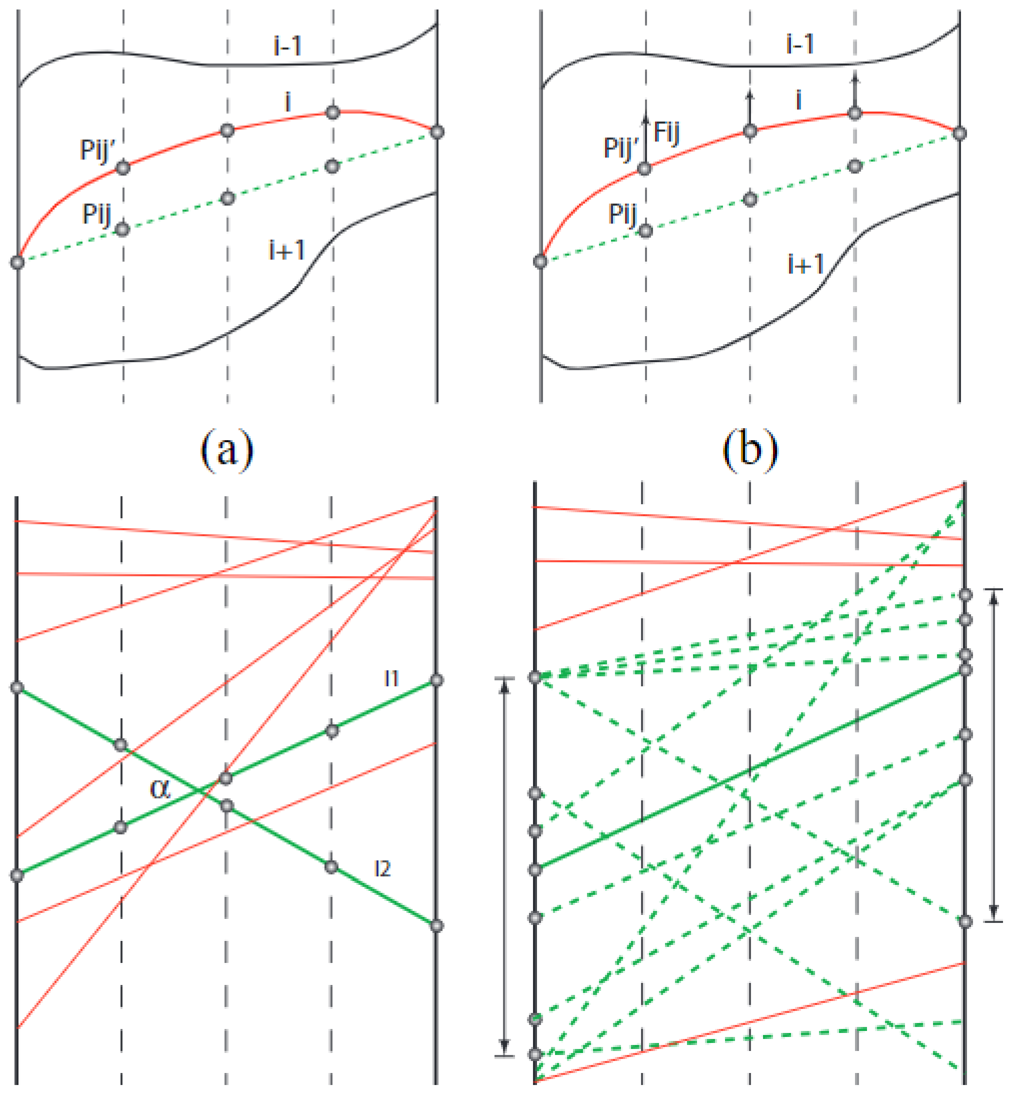
$$\mathbf{E} = \alpha_c \mathbf{E}_{curvature} + (1-\alpha_c)\mathbf{E}_{gravitation}$$

$$\mathbf{E}_{curvature} = \sum_{i=1}^{n}\sum_{j=1}^{m} |P'_{ij} - P_{ij}|$$

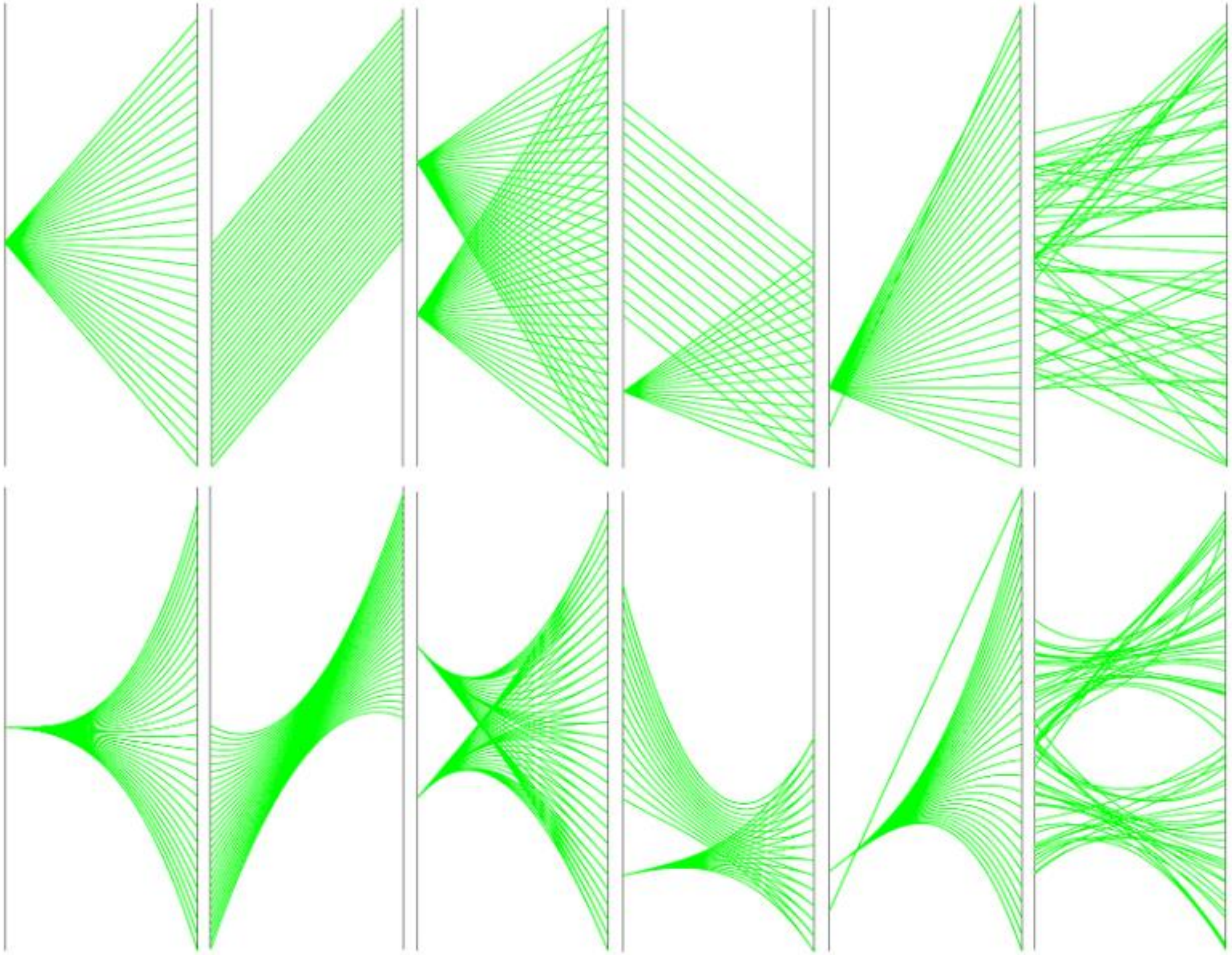
$$\mathbf{E}_{gravitation} = \sum_{i=1}^{n}\sum_{j=1}^{m} -F_{ij}\cdot(P'_{ij} - P_{ij}) + E_{ij}$$

$$\mathbf{F}_{ij} = \sum_{k=1}^{n'} f_{(l_i,l_k)_j} \qquad f_{(l_i,l_k)_j} = \frac{1}{\alpha_{l_i,l_k}^{q\alpha}} \cdot \frac{1}{D_{(l_i,l_k)_j}^{qd}}$$

$$D_{(l_i,l_k)_j} = |P_{ij} - P_{kj}|$$



(a) (b) (c) (d)



(a) (b) (c) (d) (e) (f)

Figure 3.4: Visual clustering on synthesized datasets.

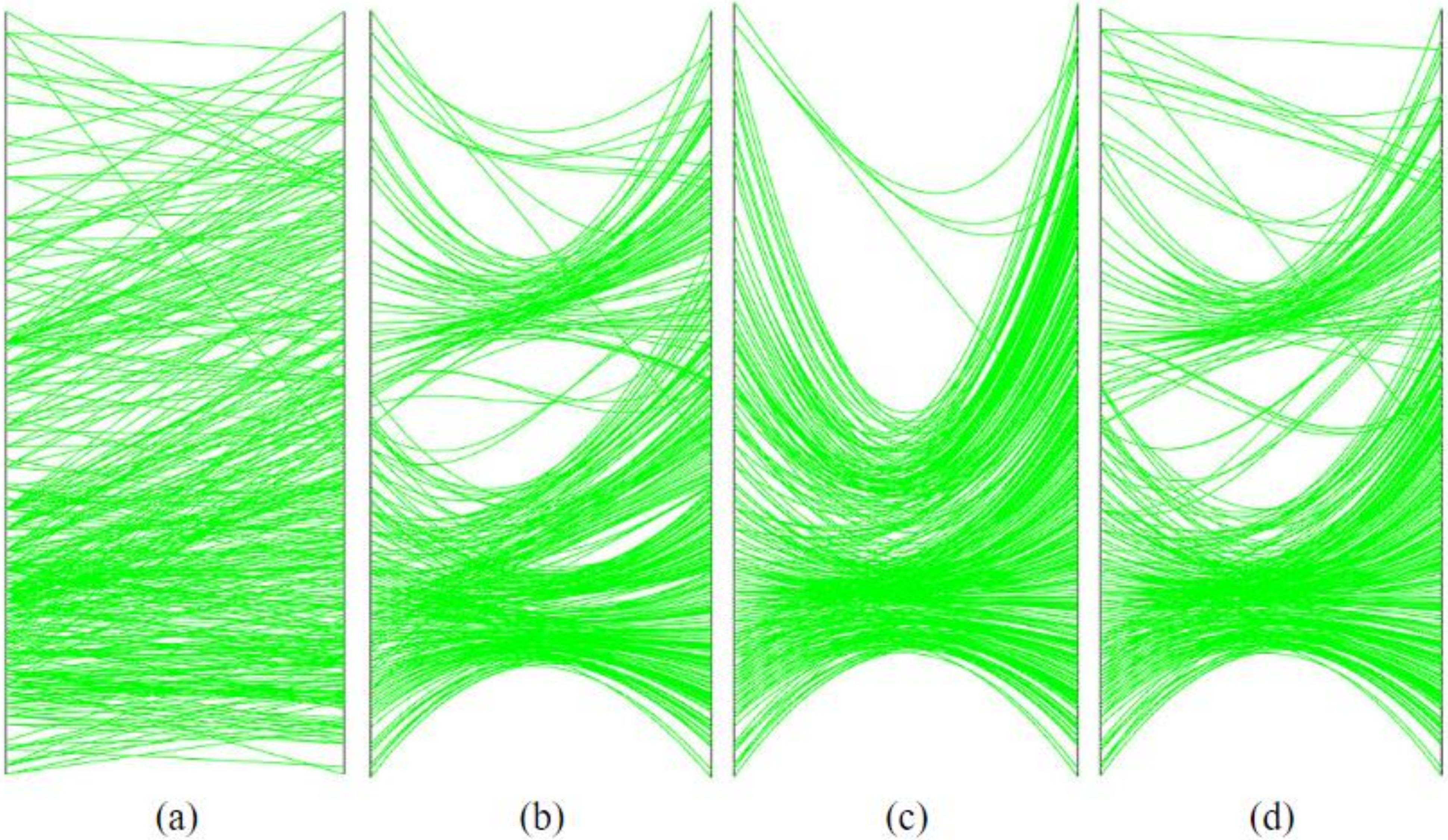
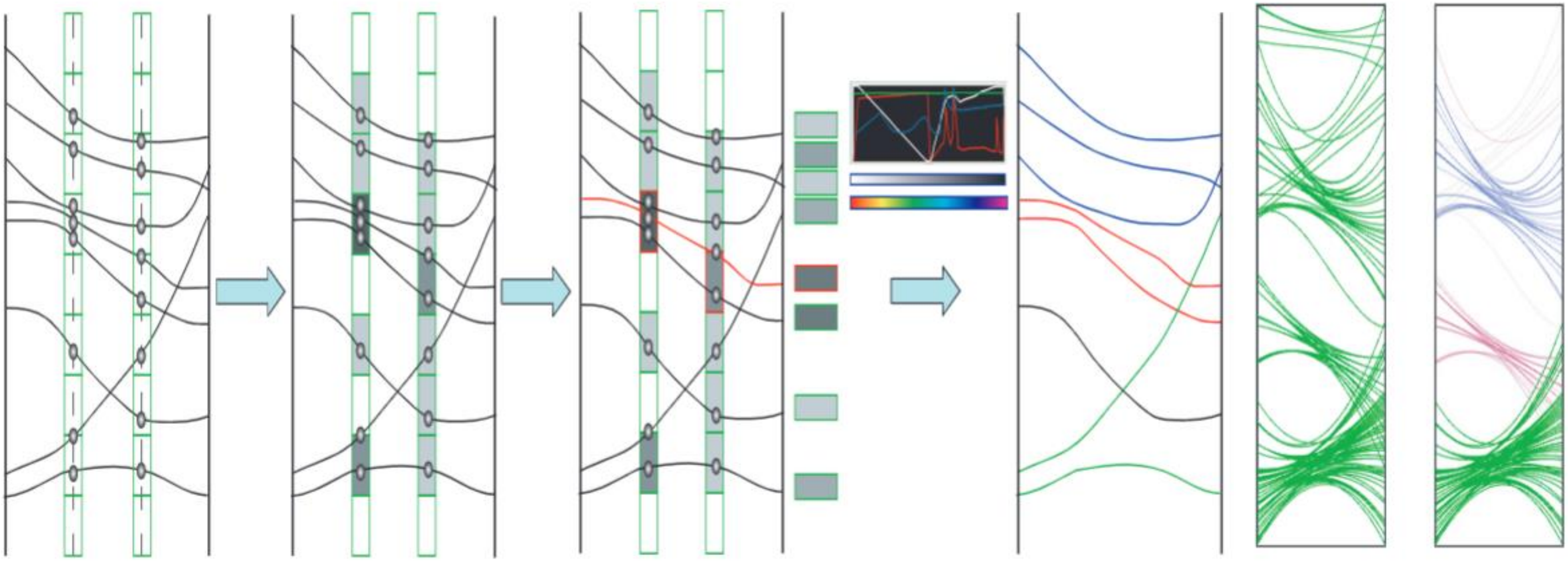
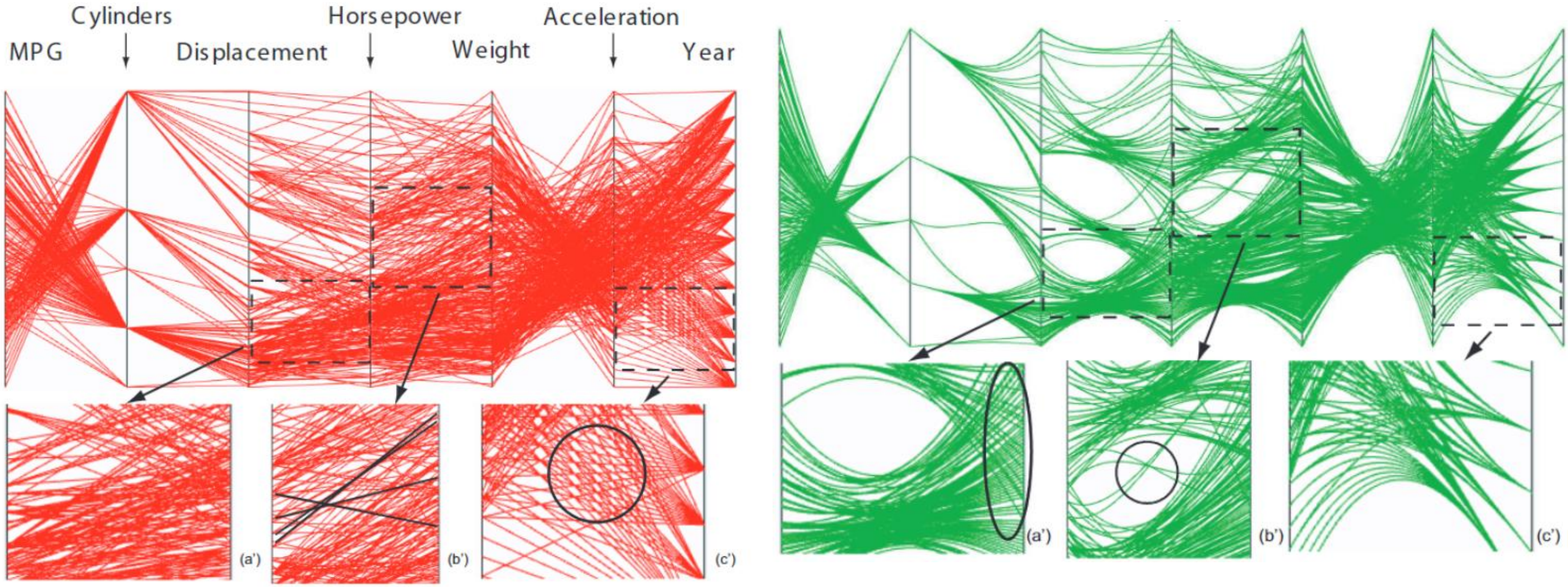


Figure 3.2: The effect of energy term weighting on visual clustering: (a) No visual clustering; (b) $\alpha_c = 0$, $q_\alpha = q_d = 15$; (c) $\alpha_c = 0$, $q_\alpha = q_d = 30$; (d) $\alpha_c = 0.15$, $q_\alpha = q_d = 30$.

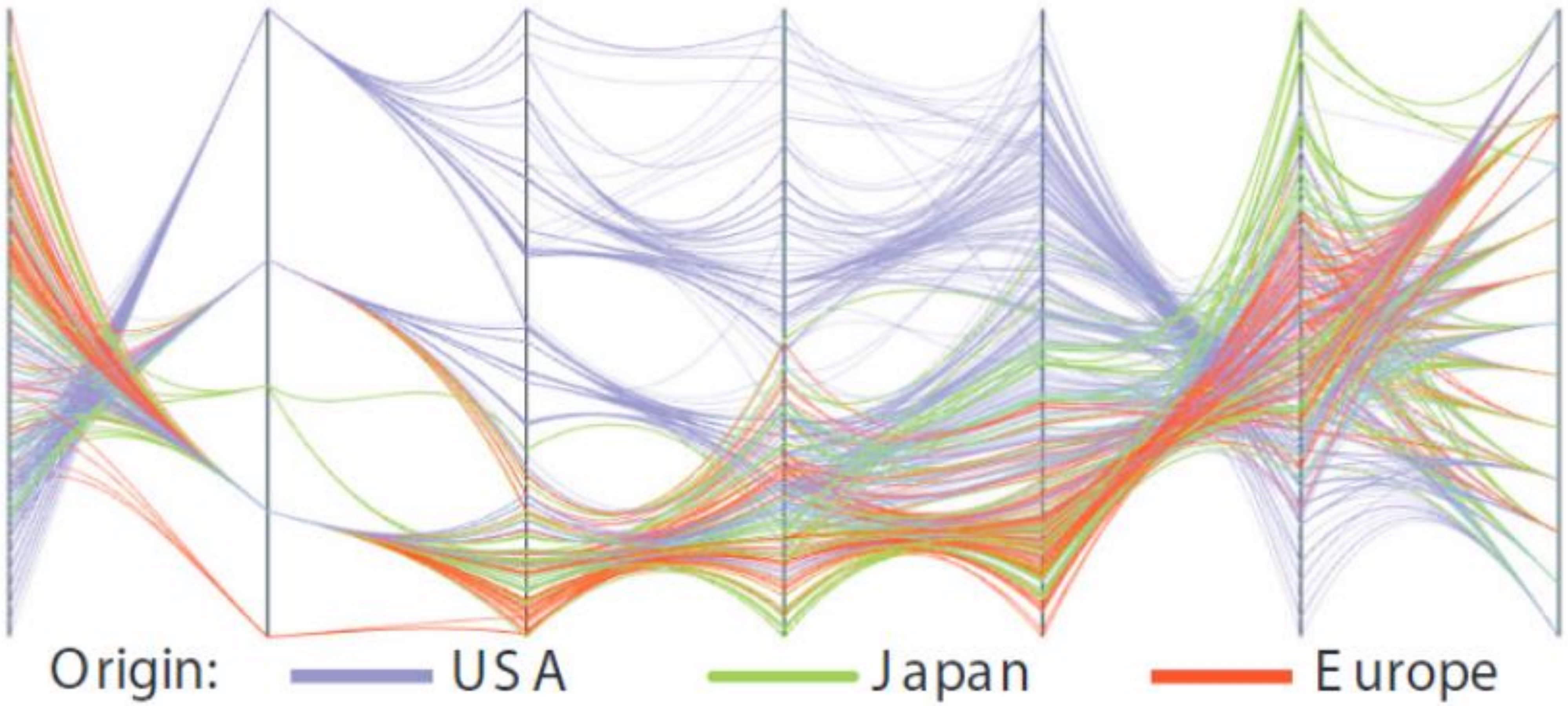
Applying color and opacity based on line density



Car evaluation dataset (7 variables, 392 items)

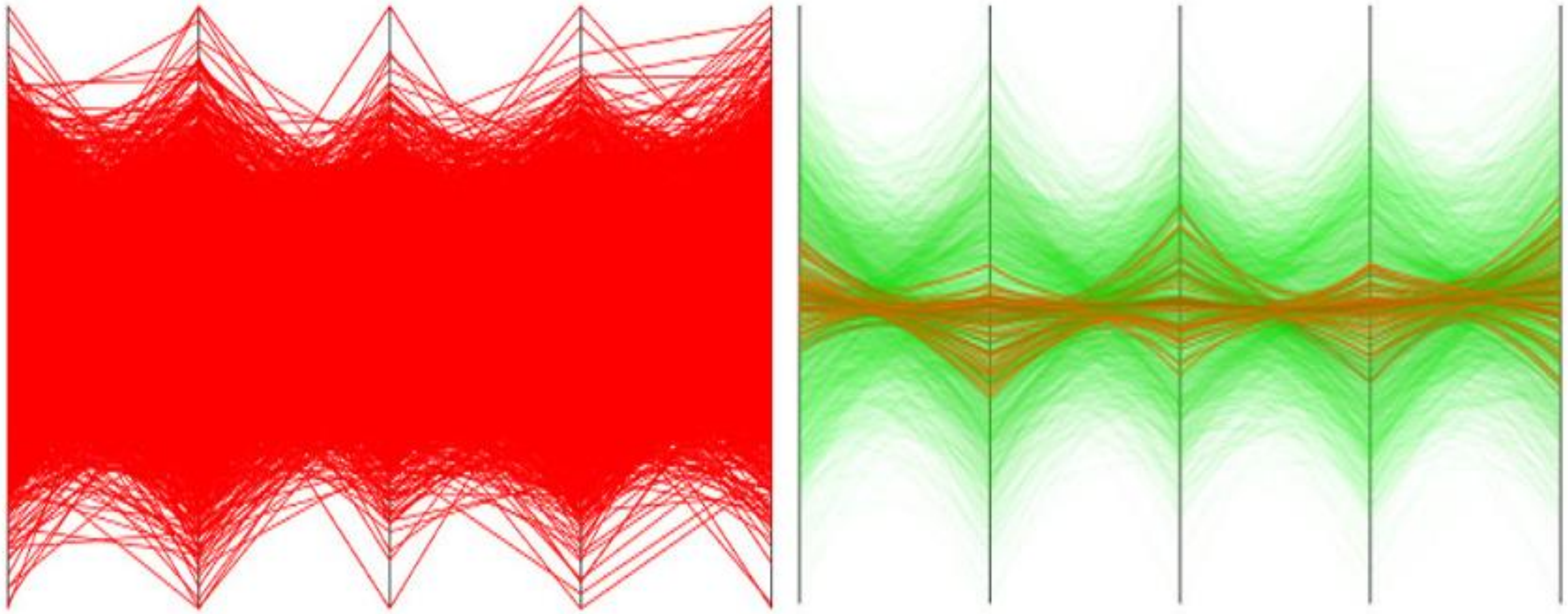


Car evaluation dataset (7 variables, 392 items)

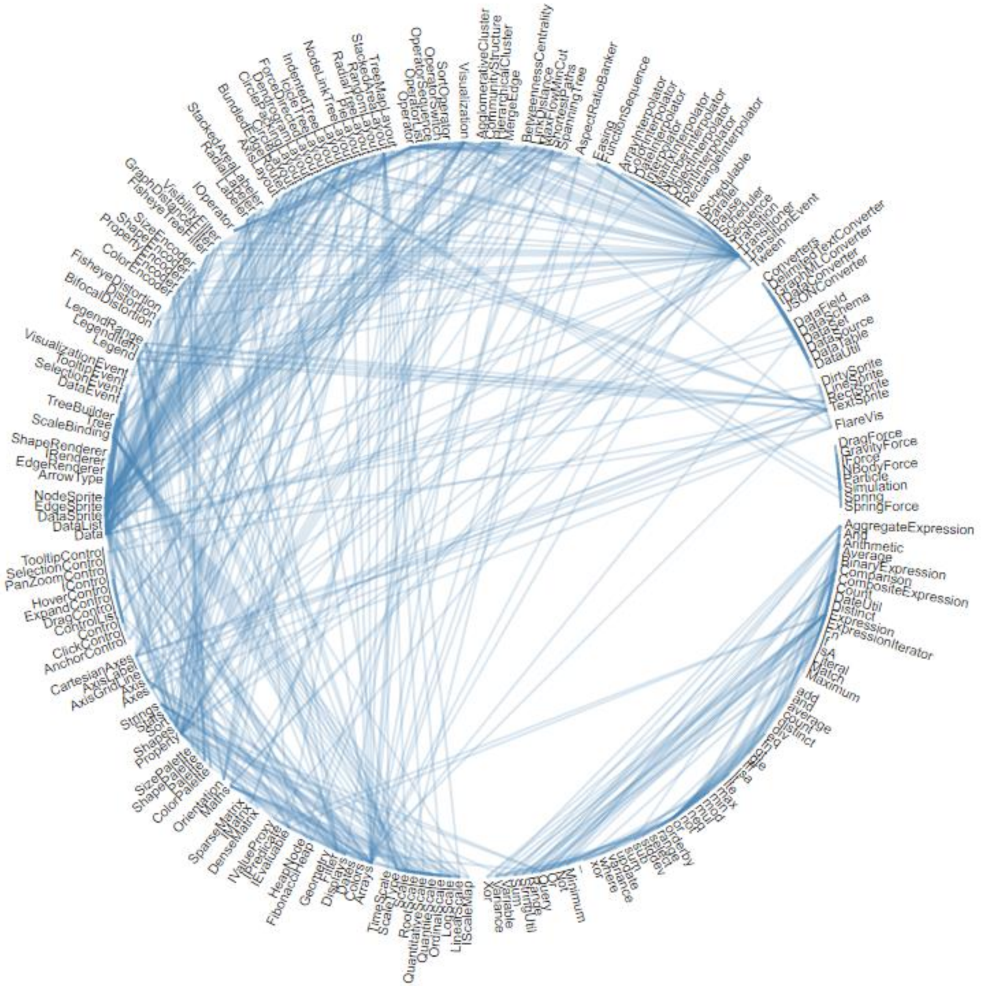


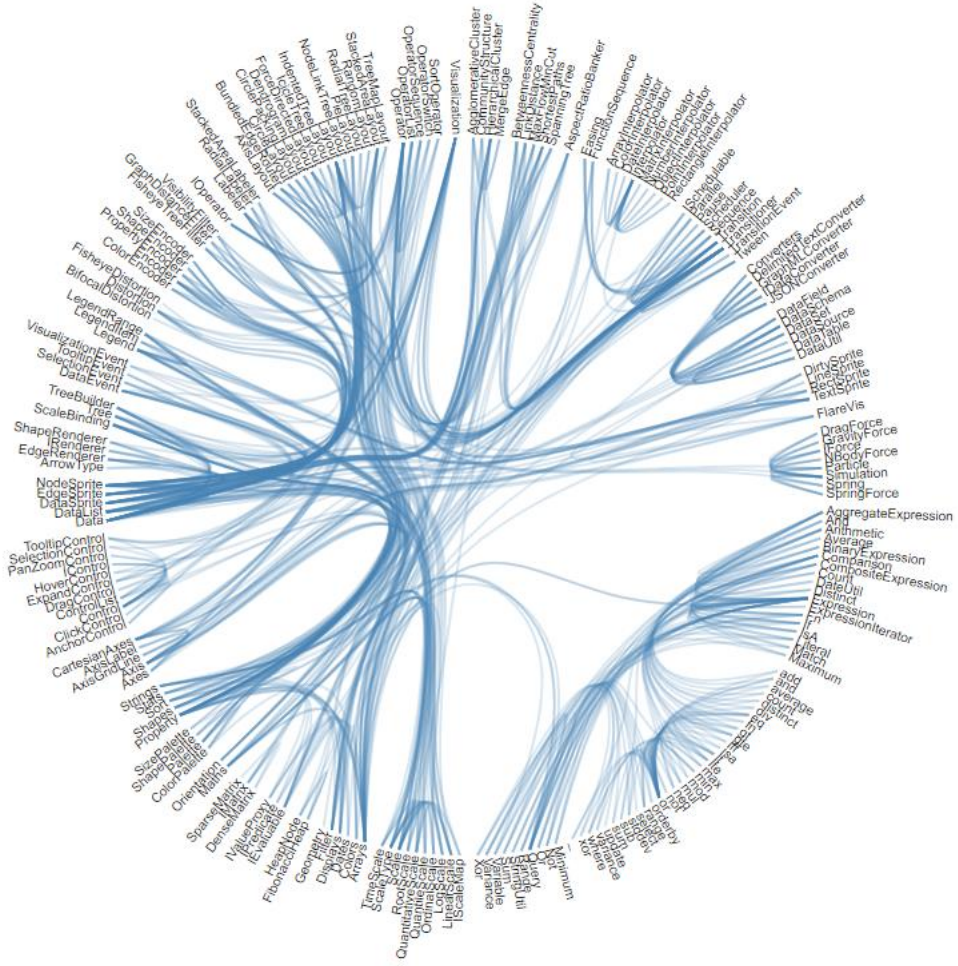
Color Coding: the original of cars

Pollen dataset (5 variables, 3848 items)



Edge bundling for circular graphs





https://vega.github.io/vega/examples/edge-bundling/