Algorithmic Aspects of Information Visualization

Debajyoti Mondal
Department of Computer Science

Summer Symposium, IEEE Computer Society Bangladesh Chapter, June 4-5, 2021
INFORMATION

Algorithms and Geometry

VISUALIZATION
A little about my Journey...

But I want to tell you more...
How about an infographic?
Is it better? More organized? If so, then why?
Is it better? More organized? If so, then why?

- Number of Crossings (2)
- Number of Slopes (5)
- Number of Bends (0)

- Number of Crossings (0)
- Number of Slopes (2 – hori. & vert.)
- Number of Bends (6)
How can we model the problem with algorithms?

If so, then we can transform a resume into an infographic visualization.
Task: Find crossing-free connection from each point to its corresponding rectangle minimizing the total number of bends.
Task: Find crossing-free connection from each point to its corresponding rectangle minimizing the total number of bends.

Can you solve it in polynomial time?
We Live in an Era of Automated Visualization
Automated Visualization & Big Data
Visualization helps to cope with data volume, variety, velocity and furthermore to extract the value from Big Data.
Networks (Graphs)

Opte project available at http://opte.org/maps/
Networks (Graphs)
Layout of Large Networks

Opte project available at http://opte.org/maps/
Computing a Network Layout [Eades 1984]

Place the nodes in random position

Move all the nodes over a certain number of iterations

Such that the layout converges to the minimum energy state
Computing a Network Layout: Defining Energy Function

The key idea is to define the energy of the model in such a way such that minimizing energy would give us the output we want.

Non adjacent nodes repel each other
Model nodes as positive electrical charges

Adjacent nodes attract each other
Model edges as springs
At the minimum energy state the node movement is zero.
At the minimum energy state the node movement is zero.

\[
\frac{\partial (\text{Aggregated Force at } a)}{\partial x} = 0
\]

\[
\frac{\partial (\text{Aggregated Force at } a)}{\partial y} = 0
\]

This rate of change gives us the direction to move the node.
We got the layout, but ...
Layout of Large Networks

Visualize in Layers: Zoom in → → →
“Overview First, Zoom and Filter, Details on Demand” - by B. Shneiderman 1996
Trade Relation 2015 (Gemstone)

Bhutan

Cuba
Geospatial Data (Backbone of Flight Network)
All of these are powered by Algorithms and Geometry.

Step 1

**Rank nodes** to distribute them to different zoom layers.

**Relevant Algorithms**
- Google PageRank
- Centrality, Betweenness
All of these are powered by Algorithms and Geometry

Step 2
Compute node positions and create a mesh (skeleton) for routing the edges

Relevant Algorithms
Node Position: Force Layout, MDS, t-SNE
Creating Mesh: Emanation Graph, Delaunay Triangulation
Routing Edges: Shortest paths or Greedy Search
All of these are powered by Algorithms and Geometry

Step 3
Load balance the tiles so that the display never contains too many nodes

Relevant Algorithms
Minimum cost
maximum flow, Quadtree
based data structure
...All of these are powered by Algorithms and Geometry

Step4
Enable Interaction

Relevant Algorithms
Dynamic Data Structures
Data Size is Relative
Push the boundaries ...

What if the networks are so big that just reading the edge list takes a long time?

Even linear-time algorithms may be slow due to hidden constants!
Push the boundaries …

What if the networks are so big that just reading the edge list takes a long time?

Even linear-time algorithms may be slow due to hidden constants!

Ideally, we want a one-pass algorithm, that never looks back at any edge twice.
Streaming or One-pass or Few-Pass Algorithms

Input

Output
Streaming or One-pass orFew-Pass Algorithms

Input

Output
Streaming or One-pass or Few-Pass Algorithms
Streaming or One-pass or Few-Pass Algorithms

Input

Output
Streaming or One-pass or Few-Pass Algorithms
Streaming or One-pass or Few-Pass Algorithms

Input

Output
A webpage-hyperlink graph with over a half-million vertices and 7 million edges.
A GPU accelerated Force Layout in 3 minutes [BRT 2017]

We can only see some community structures or clusters

GPU Accelerated Force Layout
A webpage-hyperlink graph with over a half-million vertices and 7 million edges. A GPU accelerated Force Layout in 3 minutes [BRT 2017].

Detecting communities and merging them into supernodes (7 seconds).

A probabilistic algorithm based on the idea that when reading the edge list in random, edges that are within a community appears earlier than the edges that connect two different communities [HMBL 2017].
A webpage-hyperlink graph with over a half-million vertices and 7 million edges. A GPU accelerated Force Layout in 3 minutes [BRT 2017]

Detecting communities and merging them into supernodes (7 seconds)

Coloring Force Layout using the community detection algorithm
Few-Pass Algorithm + GPU Acceleration [MM 2020]

What is the trade-off?
How reliable are these visualizations?

The communities detected by the few-pass algorithm exist also in the Detailed visualization.
Network Algorithms are not only for Networks!

Table Cartogram Always Exists! [EFKKMNV 2018]
Network Algorithms are not only for Networks!

Table Cartogram Always Exists! [EFKMKMV 2018]
Network Algorithms are not only for Networks!

\[
\text{minimize} \sum_{i=1}^{4} \left( \left( \frac{a_i^2 x_i^2}{a_i^2 + b_i^2} \right)^2 \cdot x^2 + \left( \frac{b_i^2 y_i^2}{a_i^2 + b_i^2} \right)^2 \cdot y^2 + \left( \frac{a_i b_i}{\sqrt{a_i^2 + b_i^2}} \right) \cdot xy \right).
\]

\[
+ \left( \frac{a_i \cdot (c_i - h'_i)}{\sqrt{a_i^2 + b_i^2}} \right) \cdot x + \left( \frac{b_i \cdot (c_i - h'_i)}{\sqrt{a_i^2 + b_i^2}} \right) \cdot y
\]
Network Algorithms are not only for Networks!

Cartography [EFKKMNV 2018]
Network Algorithms are not only for Networks!

Cartography [EFKKMNV 2018]

minimize \[ \sum_{j=1}^{r} h_j \cdot b_{j,c} \]
s.t. \[ a_{j,1} a_{j+1,1}^{-1} = 1, \]
\[ b_{j,c} b_{j+1,c}^{-1} = 1, \]
\[ w_{j,k} h_j b_{j,k}^{-1} + a_{j,k} b_{j,k}^{-1} \leq 1, \]
\[ a_{j,k} b_{j,k}^{-1} \leq 1, \]
\[ a_{j,k} b_{j+1,k}^{-1} \leq 1, \]
\[ a_{j+1,k} b_{j,k}^{-1} \leq 1, \]
\[ h_1 H^{-1} + \ldots + h_r H^{-1} \leq 1, \]
\[ W^{-1} b_{1,c} \leq 1, \]
\[ a_{j,k}, b_{j,k}, h_j \geq 0 \]

<table>
<thead>
<tr>
<th>WA 6.725</th>
<th>MT 0.989</th>
<th>ND 0.673</th>
<th>MN 5.304</th>
<th>WI 5.687</th>
<th>NY 19.378</th>
<th>VT 0.626</th>
<th>ME 1.328</th>
</tr>
</thead>
<tbody>
<tr>
<td>OR 3.831</td>
<td>ID 1.568</td>
<td>SD 0.814</td>
<td>IA 3.046</td>
<td>IL 12.831</td>
<td>IN 6.484</td>
<td>OH 11.537</td>
<td>CT 3.574</td>
</tr>
<tr>
<td>NV 2.701</td>
<td>WY 0.564</td>
<td>NE 1.826</td>
<td>MO 5.989</td>
<td>KY 4.339</td>
<td>WV 1.853</td>
<td>MD 5.774</td>
<td>NJ 8.792</td>
</tr>
<tr>
<td>UT 2.764</td>
<td>CO 5.029</td>
<td>KS 2.853</td>
<td>AR 2.916</td>
<td>TN 6.346</td>
<td>SC 4.625</td>
<td>VA 8.001</td>
<td>DE 0.898</td>
</tr>
<tr>
<td>CA 37.254</td>
<td>NM 2.059</td>
<td>OK 3.751</td>
<td>AR 2.916</td>
<td>TN 6.346</td>
<td>SC 4.625</td>
<td>VA 8.001</td>
<td>DE 0.898</td>
</tr>
</tbody>
</table>
Network Algorithms are not only for Networks!

StreamTable [EM 2021]
Network Algorithms are not only for Networks!

Infographics for Geospatial Correlation

ALBEDO
reflection of solar radiation

ALBEDO is high

ALBEDO is low

Transform by Soil Moisture SH2O

SH2O is high

SH2O is low

Potential positive relation
Network Algorithms are not only for Networks!

Creating Visual Effects in Images

Increase in light illumination
Network Algorithms are not only for Networks!

Digital Art (mosaic art effect)
Summary

To truly leverage the full power of visual analytics and data science we need to seamlessly integrate Algorithms, Visualization, and Geometry.
To truly leverage the full power of visual analytics and data science we need to seamlessly integrate Algorithms, Visualization, and Geometry.

Thank You!