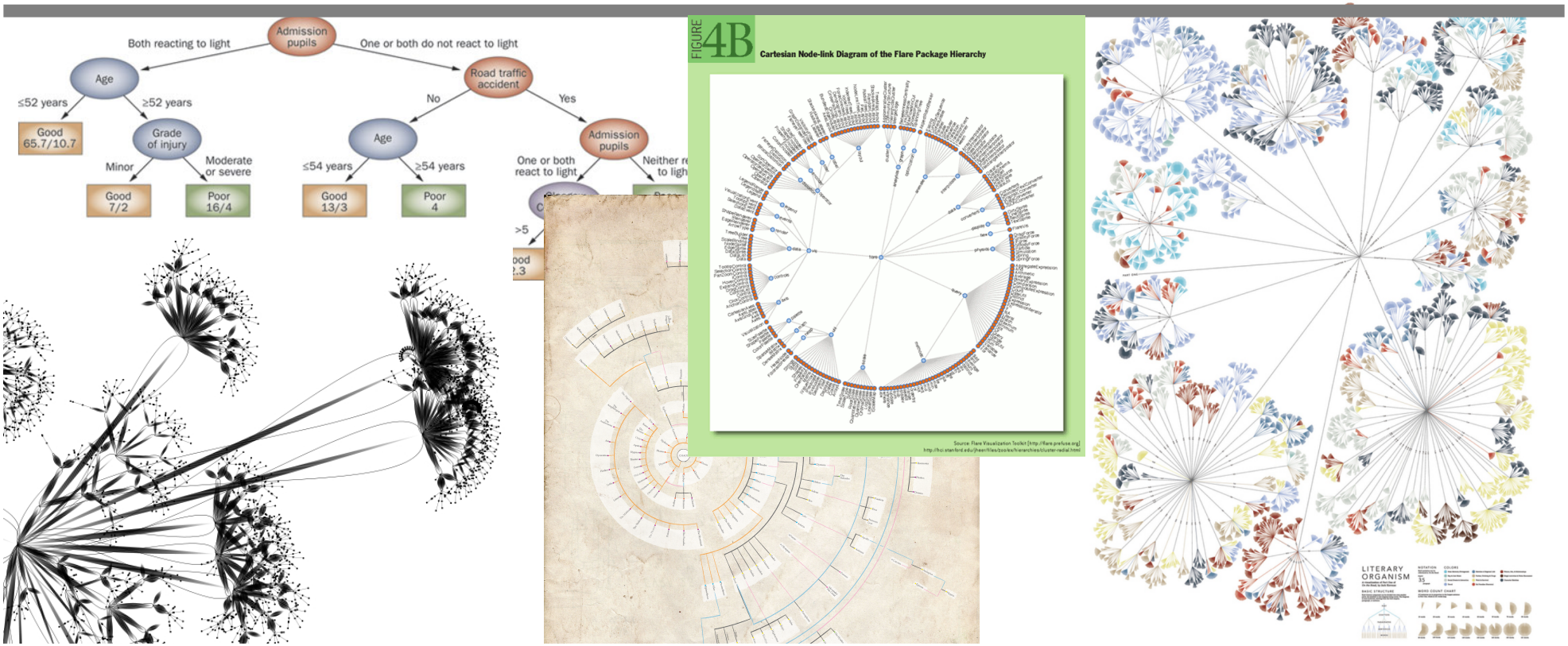
Introduction

Course : Data Visualization
Lecturer : Tamara Mchedlidze
Utrecht University, Dept. of Information and Computing Sciences



Lecture Overview

- Why Data visualization?
- Data types and their models
- Networks and their visualizations (gallery)
- Let’s draw some networks
- Basis for algorithm design: formalization of network visualization problem
- Evaluation of Network Visualization
- Types of algorithms (mind map)
- Some basic notions in Algorithm Complexity Theory

Why Data Visualization?

We live in the era of data

Well-defined questions can be solved by purely computational methods (statistics, algorithms, machine learning)

Many problems are ill-specified – people do not know how to approach the problem, which questions to ask $->$ human in the loop

Vis tools are appropriate when we want to augment human capabilities

Visualization allows people to offload internal cognition and memory usage to the perceptual system (goes beyond classical data vis)

We need computer and algorithms to construct visualizations

Data Types

We collect and process tremendous amounts of complex data sets

To study data visualization systematically it helps to organize data by types

Data Types

We collect and process tremendous amounts of complex data sets

To study data visualization systematically it helps to organize data by types

Question: What data types do you know?

Data Types

We collect and process tremendous amounts of complex data sets

To study data visualization systematically it helps to organize data by types

Question: What data types do you know?

- High-dimensional data
- Relational data
- Spatial data

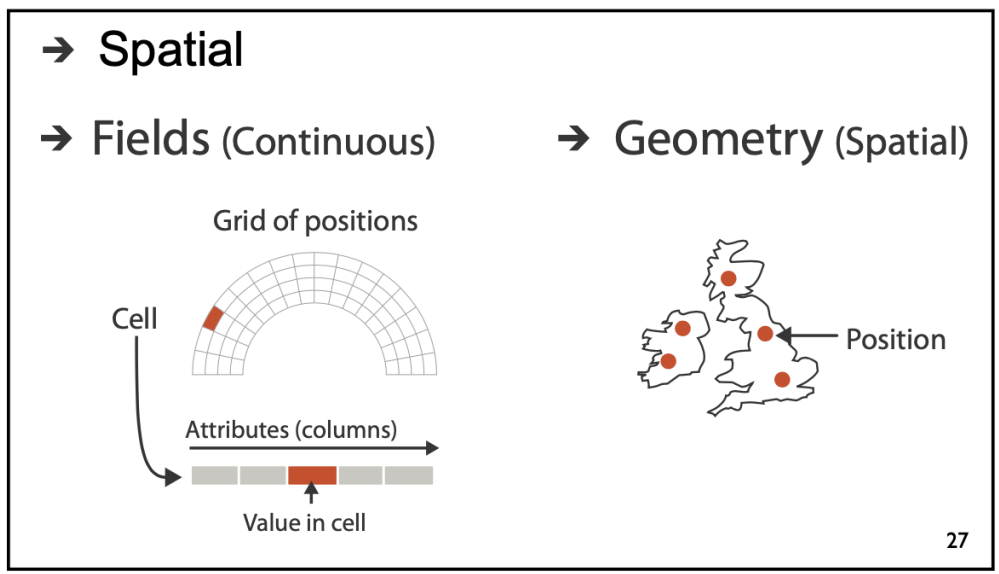
Data Types

We collect and process tremendous amounts of complex data sets

To study data visualization systematically it helps to organize data by types

Question: What data types do you know?

- High-dimensional data
- Relational data
- Spatial data



from the book of T. Muzner

Data Types

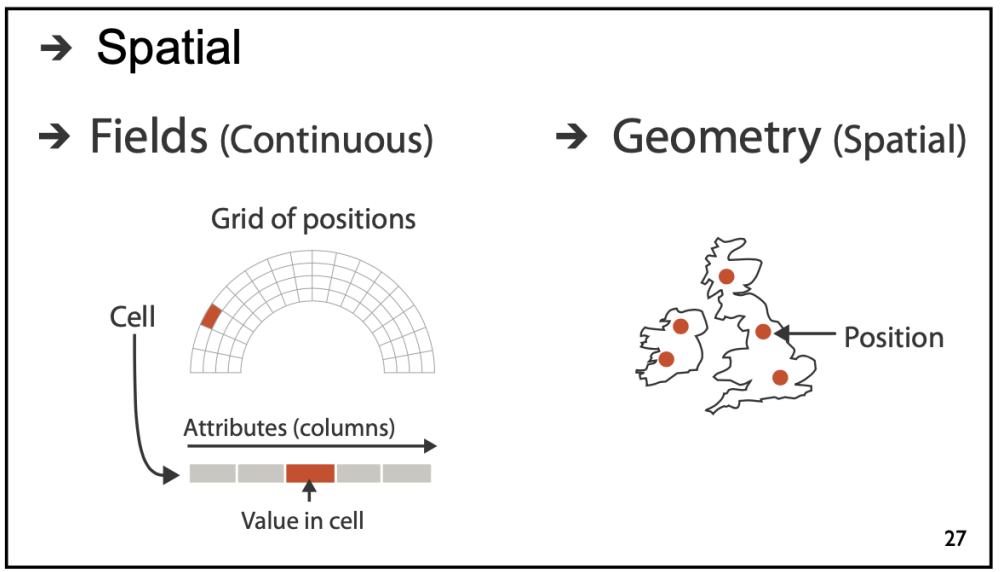
We collect and process tremendous amounts of complex data sets

To study data visualization systematically it helps to organize data by types

Question: What data types do you know?

In this course

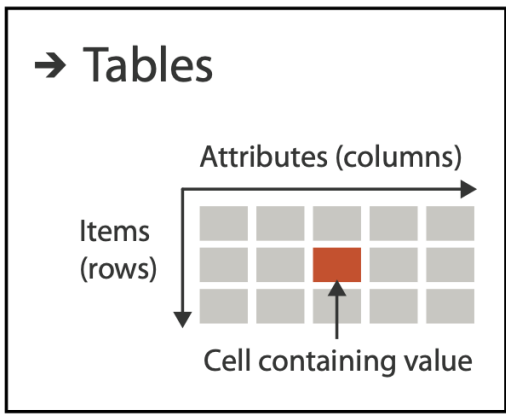
- High-dimensional data
- Relational data
- Spatial data



from the book of T. Muzner

High-dimensional data

High-dimensional data are modeled as tables



Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

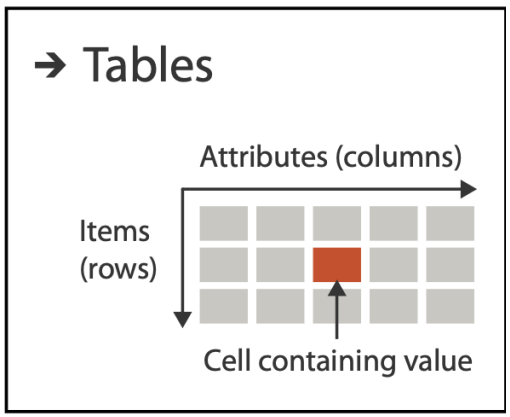
Cell

Item

Column/attribute/dimension

High-dimensional data

High-dimensional data are modeled as tables



Name	Age	Shirt Size	Favorite Fruit
Amy	8	S	Apple
Basil	7	S	Pear
Clara	9	M	Durian
Desmond	13	L	Elderberry
Ernest	12	L	Peach
Fanny	10	S	Lychee
George	9	M	Orange
Hector	8	L	Loquat
Ida	10	M	Pear
Amy	12	M	Orange

Cell

Item

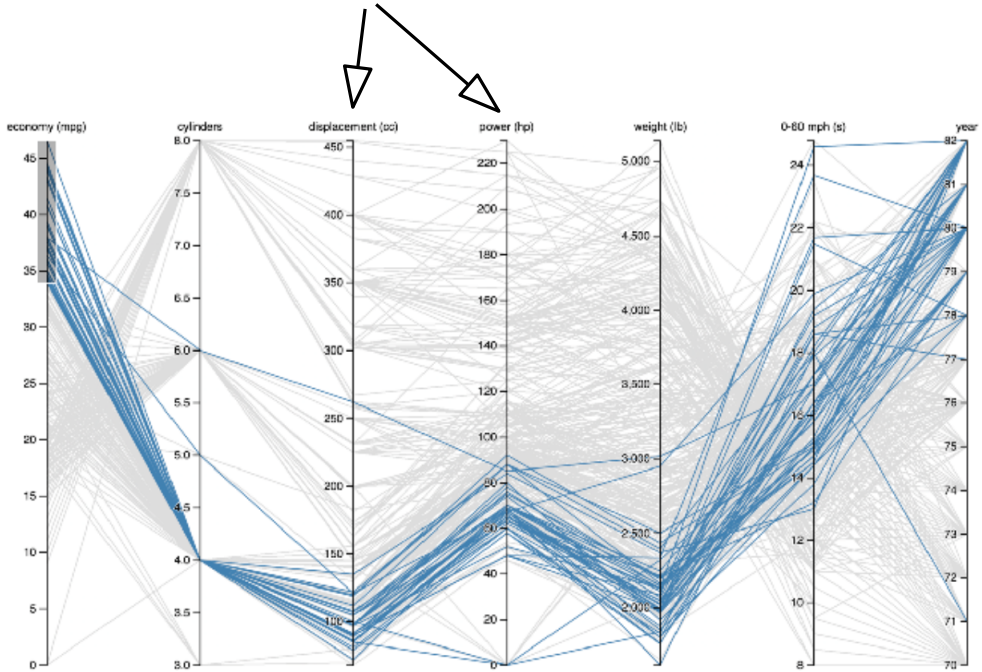
Column/attribute/dimension

Question: Examples of high-dimensional data?

How to visualize high-dimensional data

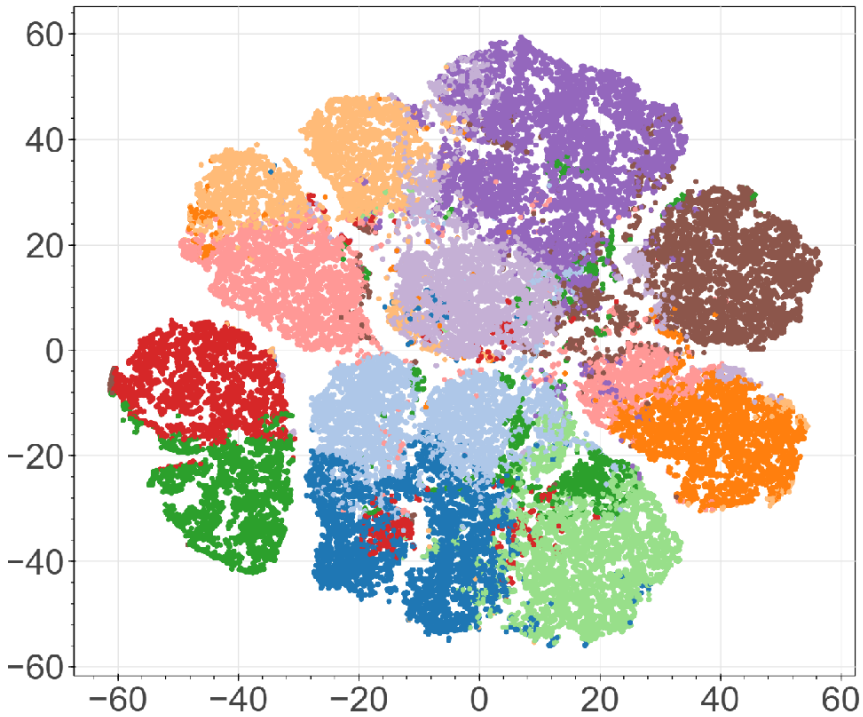
explicit dimension representation

dimensions



Parallel coordinates

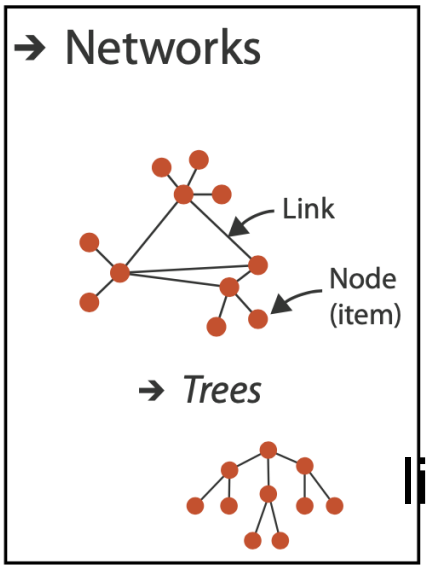
n initial dimensions are “squeezed” into two

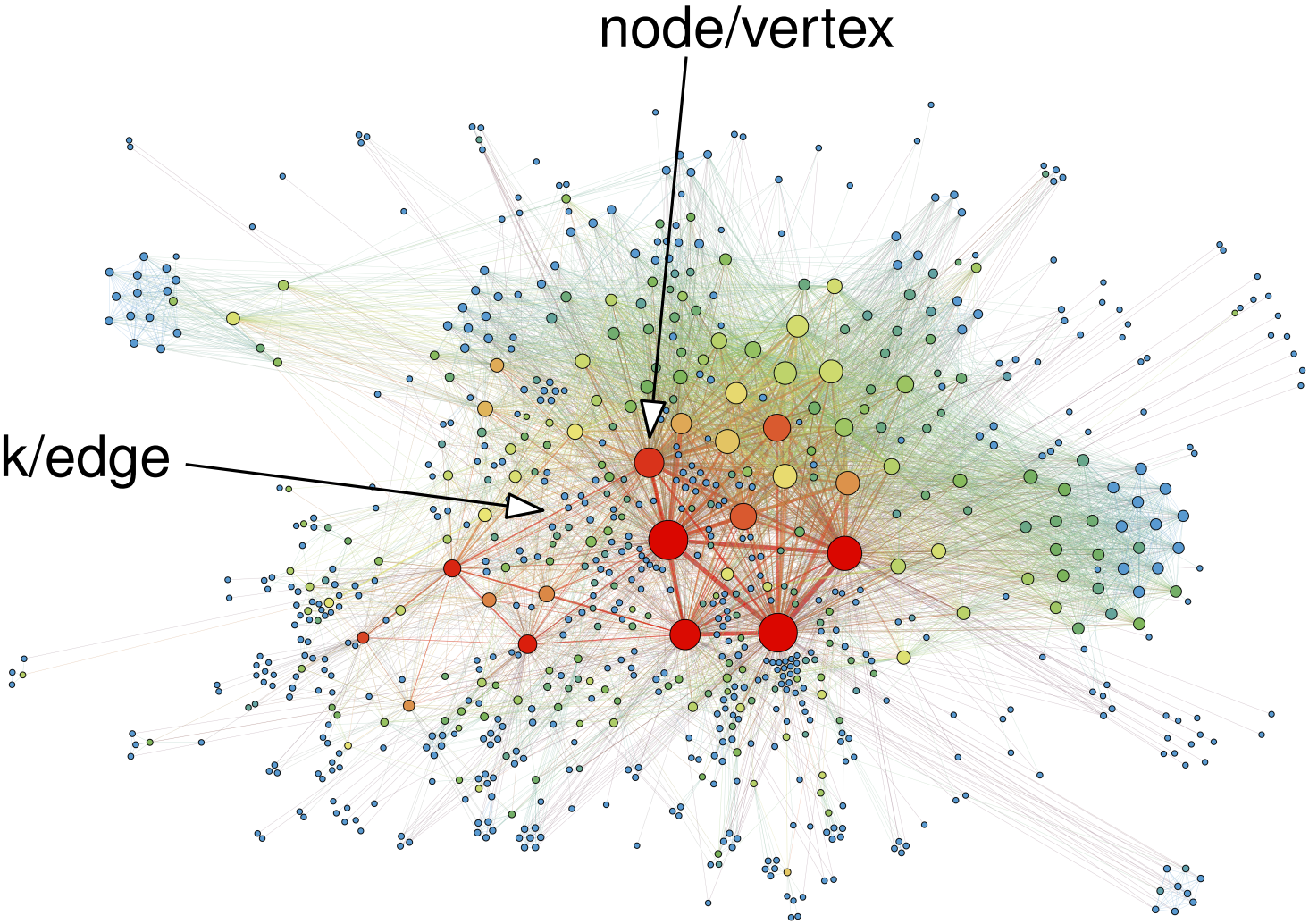


Dimensionality reduction

Relational data

Relational data are modeled as networks

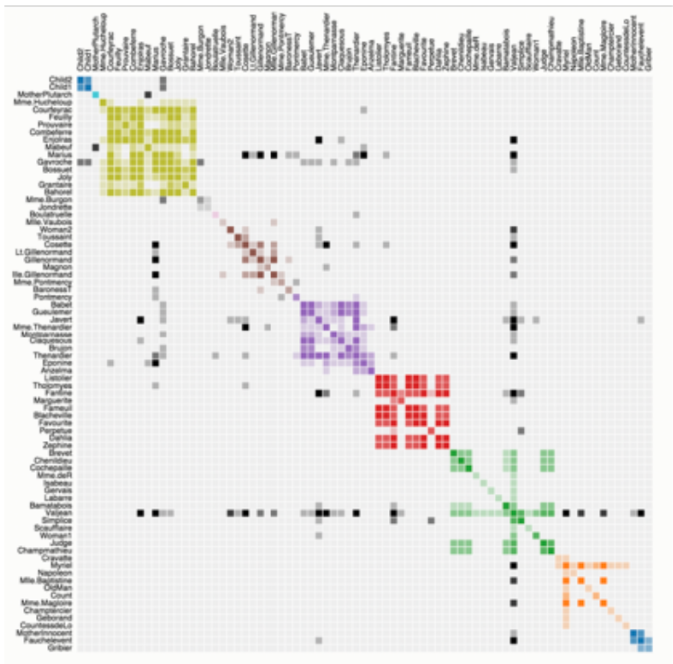




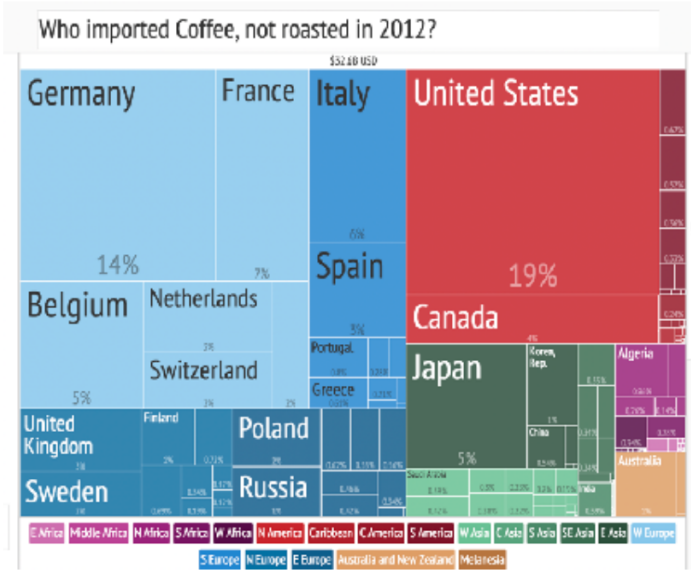
Question: Examples of relational data?

How to visualize relational data

Node-link diagram



Adjacency Matrix

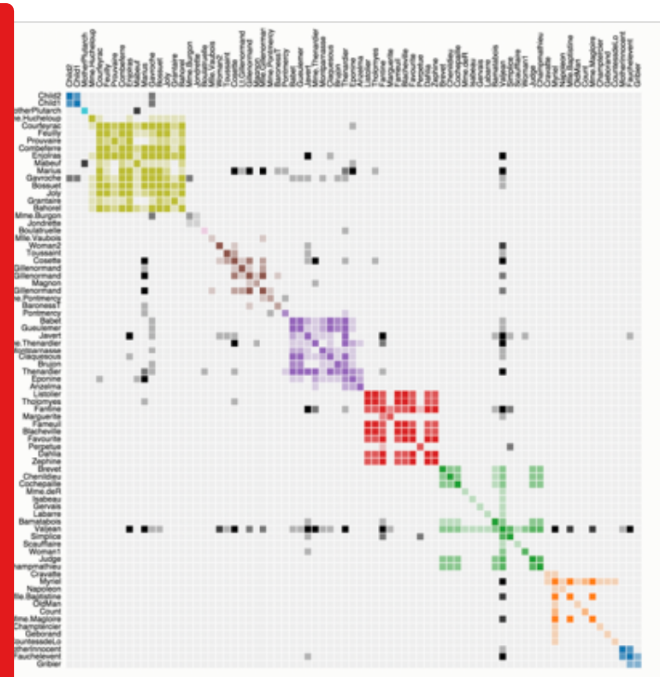


Treemaps

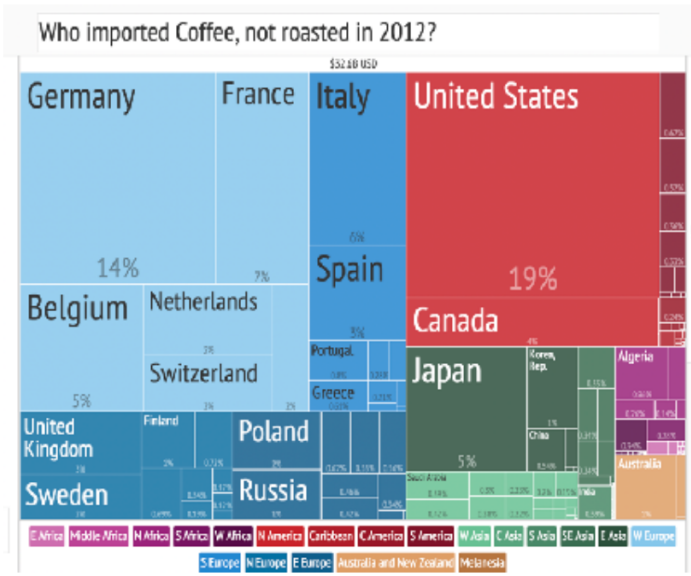
How to visualize relational data

In this course

Node-link diagram

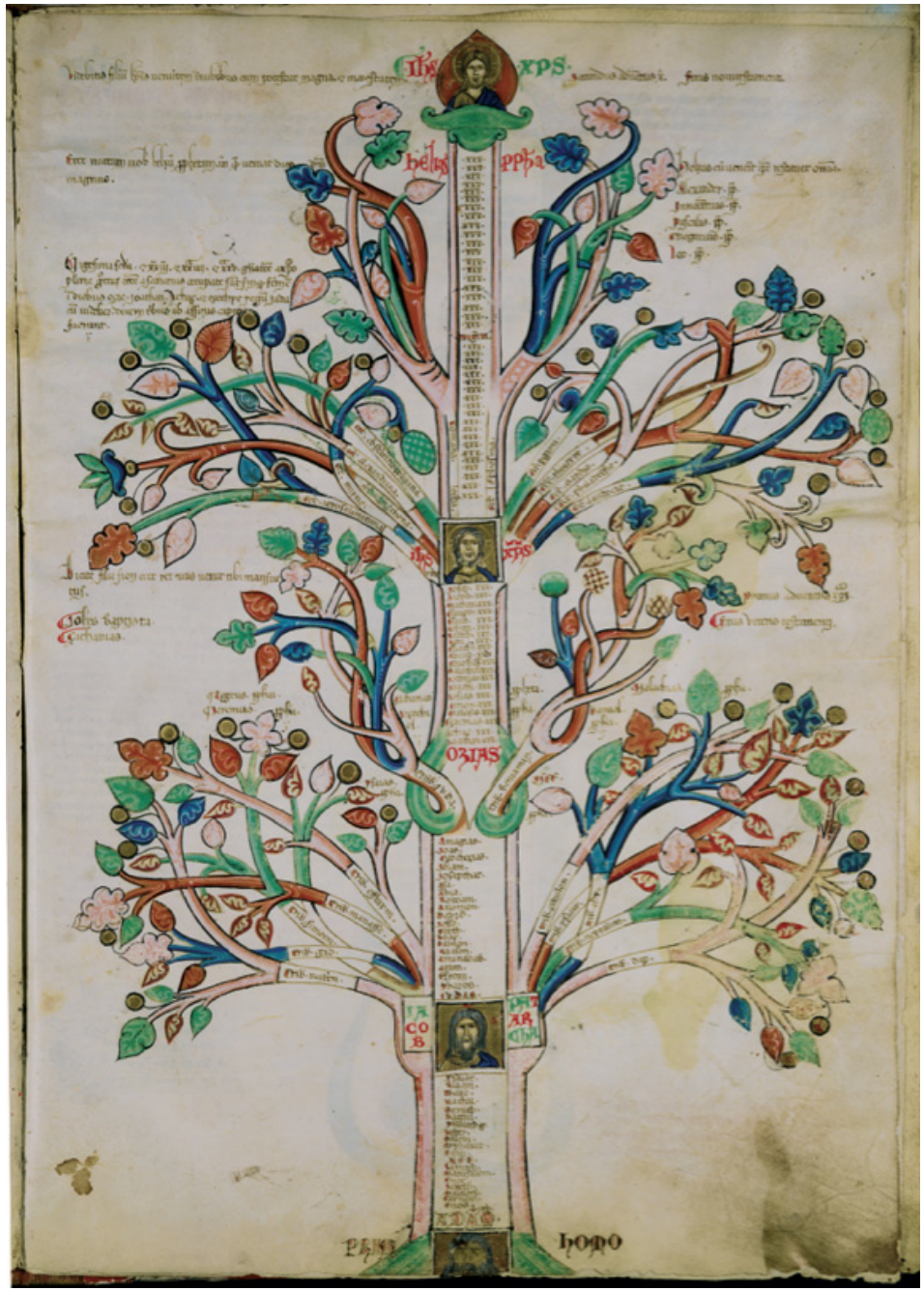


Adjacency Matrix



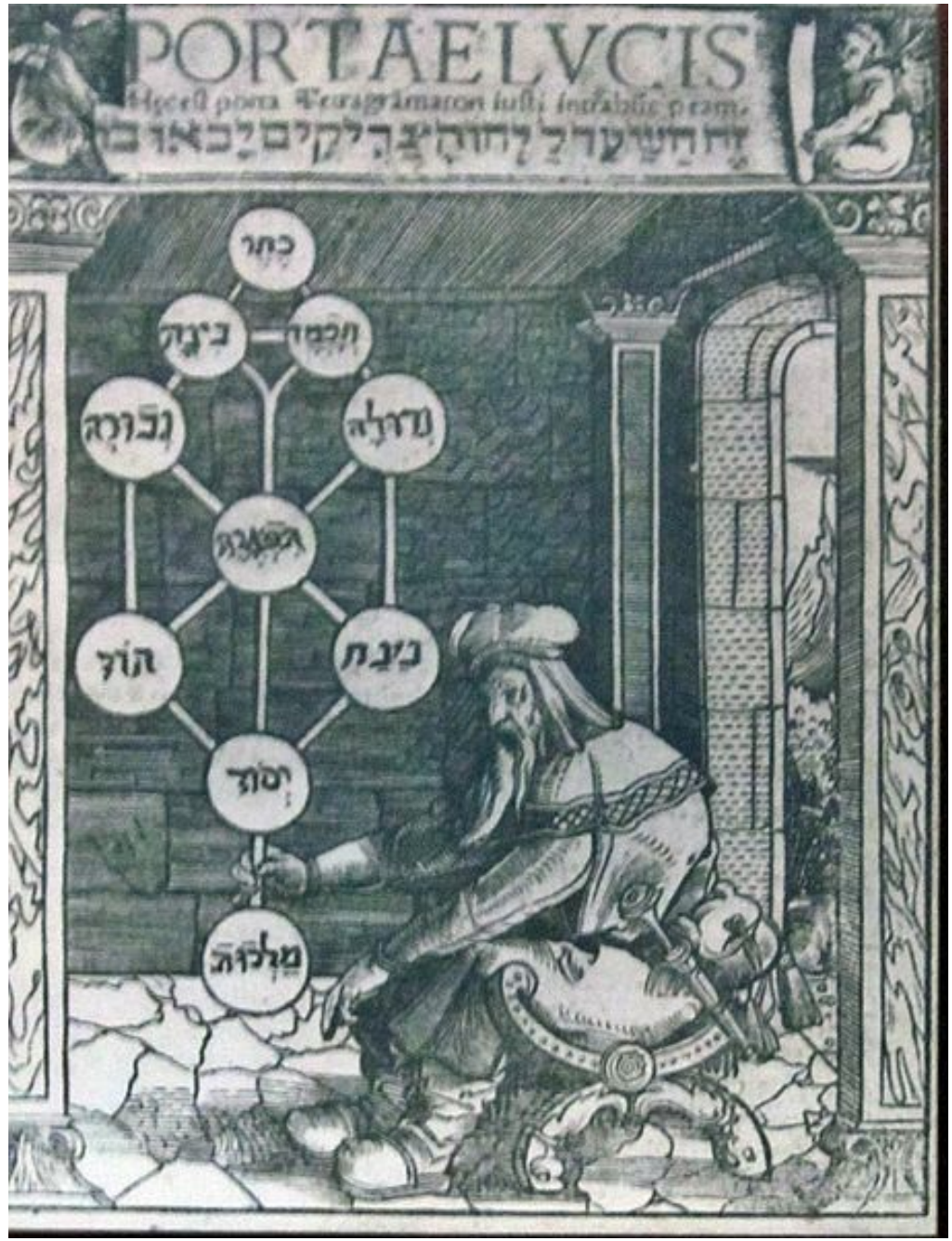
Treemaps

Biblical characters and events (1202)



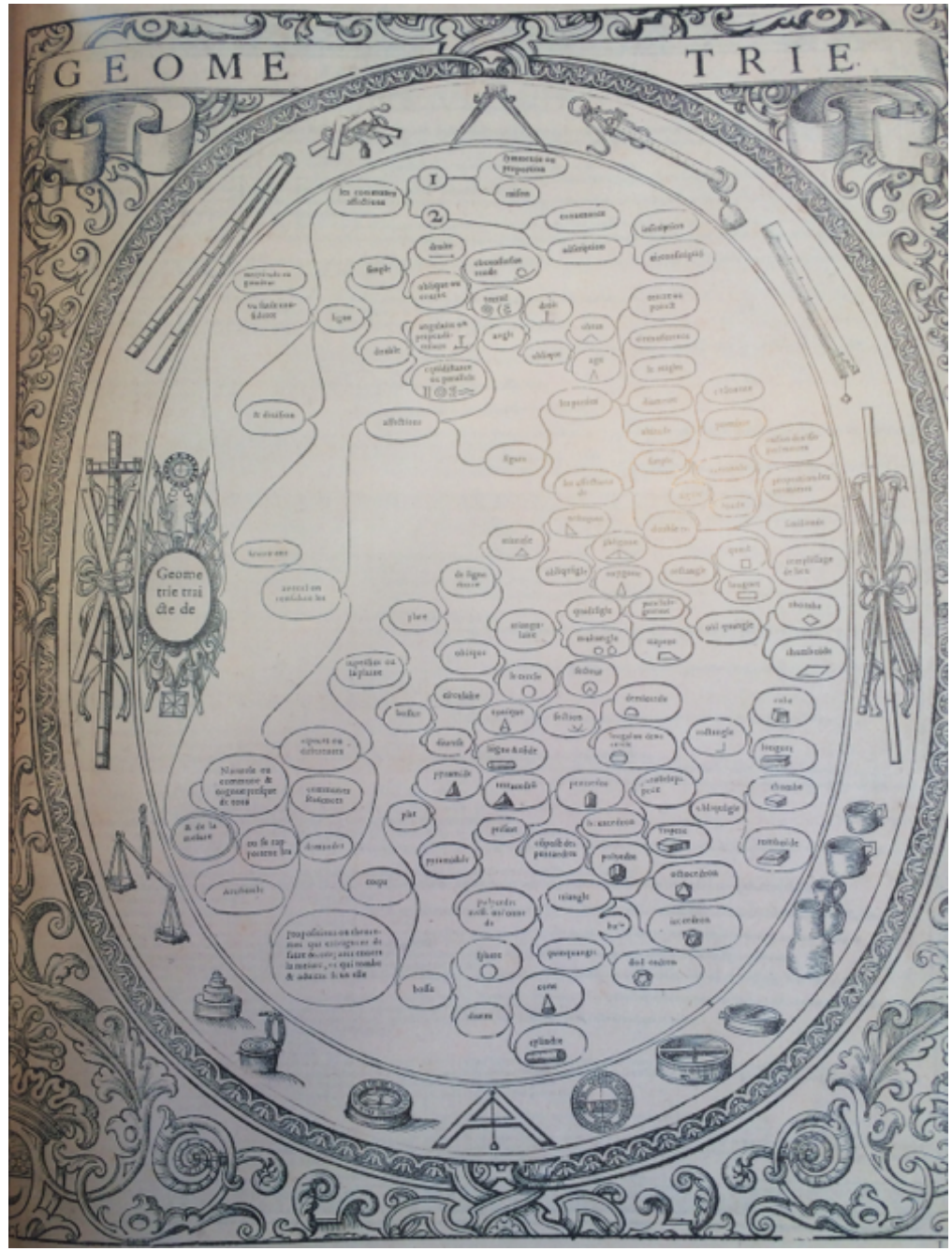
Source: Joachim de Fiore

”Tree of Life”’ (1516)



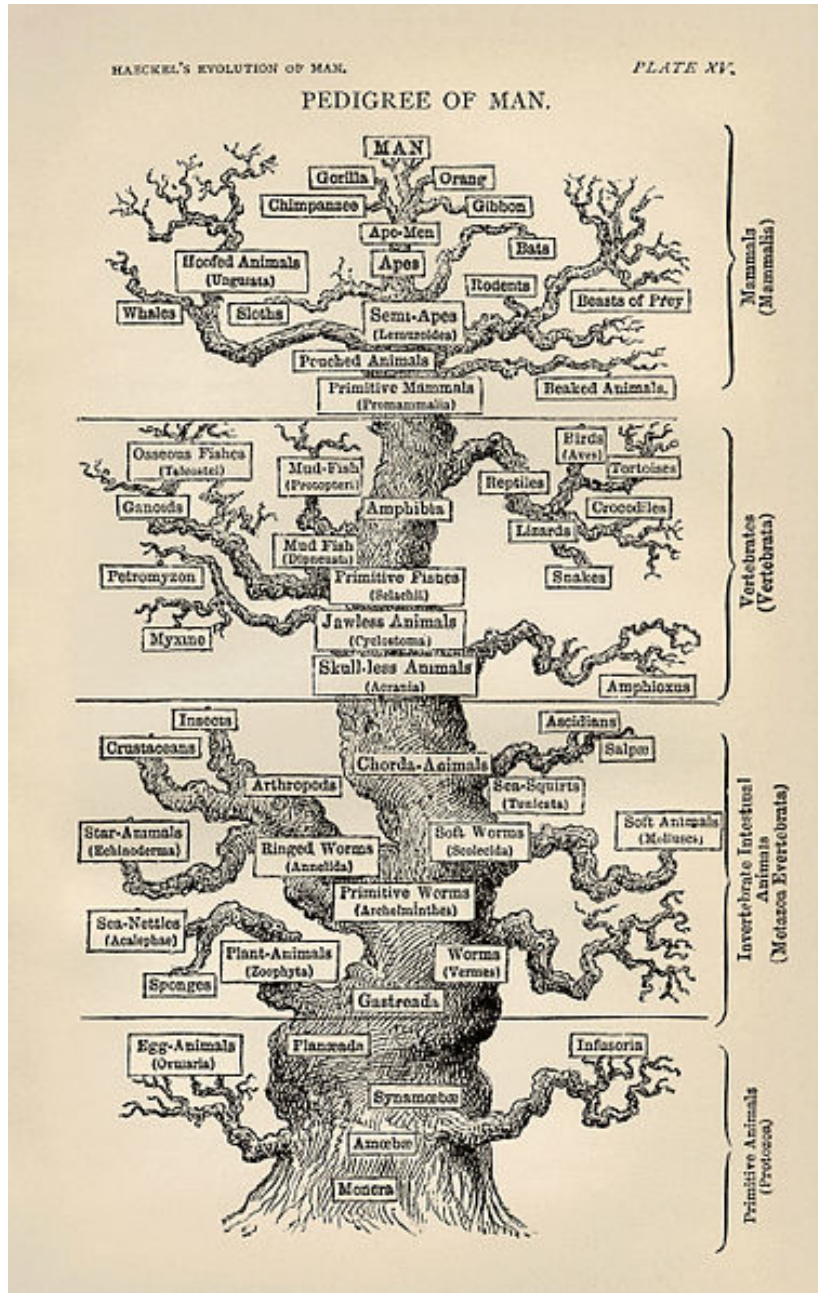
Source: Paul Riccius, Portae Luci

Geometrical Concepts (1587)



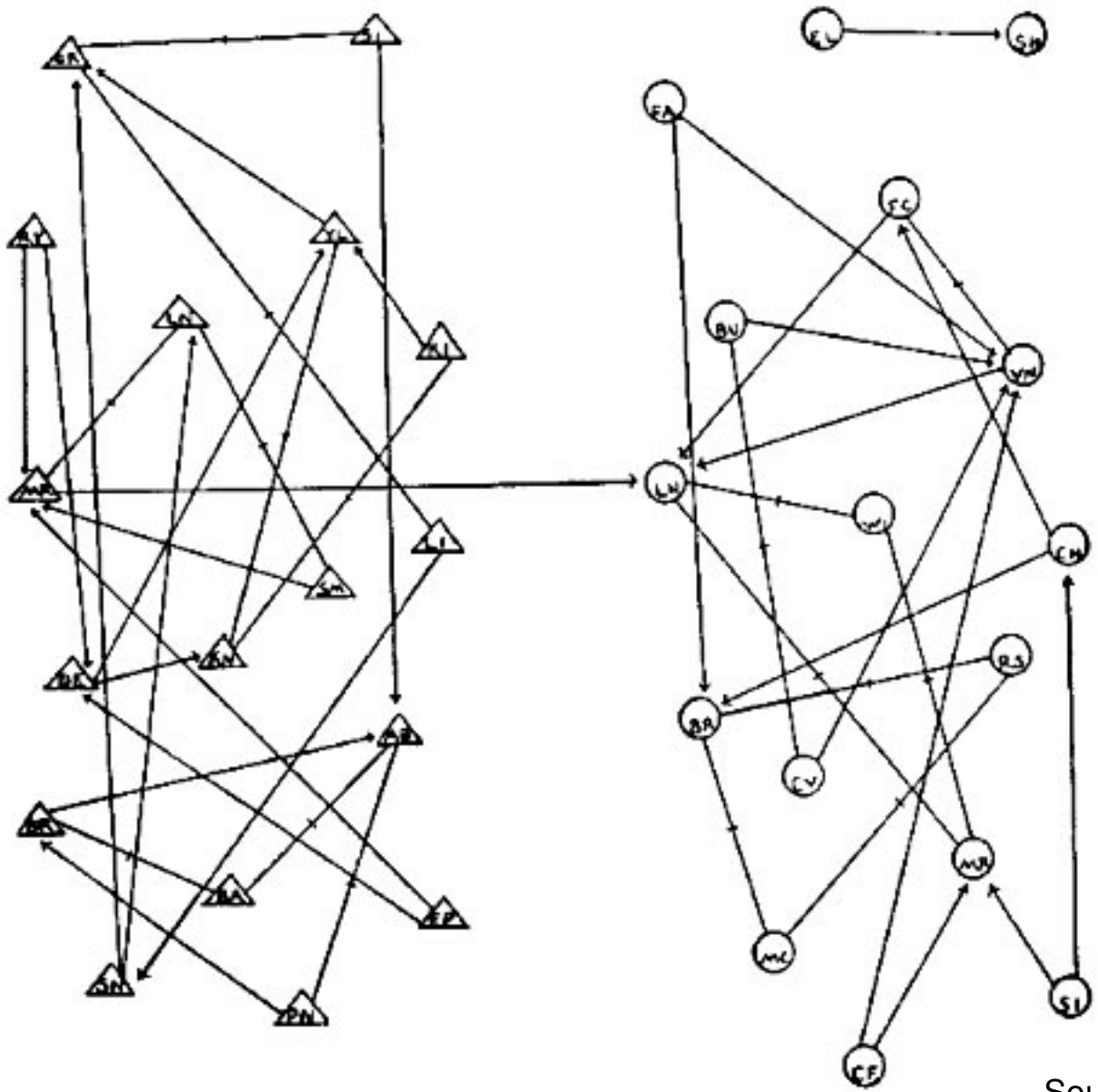
Source: Christophe de Savigny

Genealogical Tree (1879)



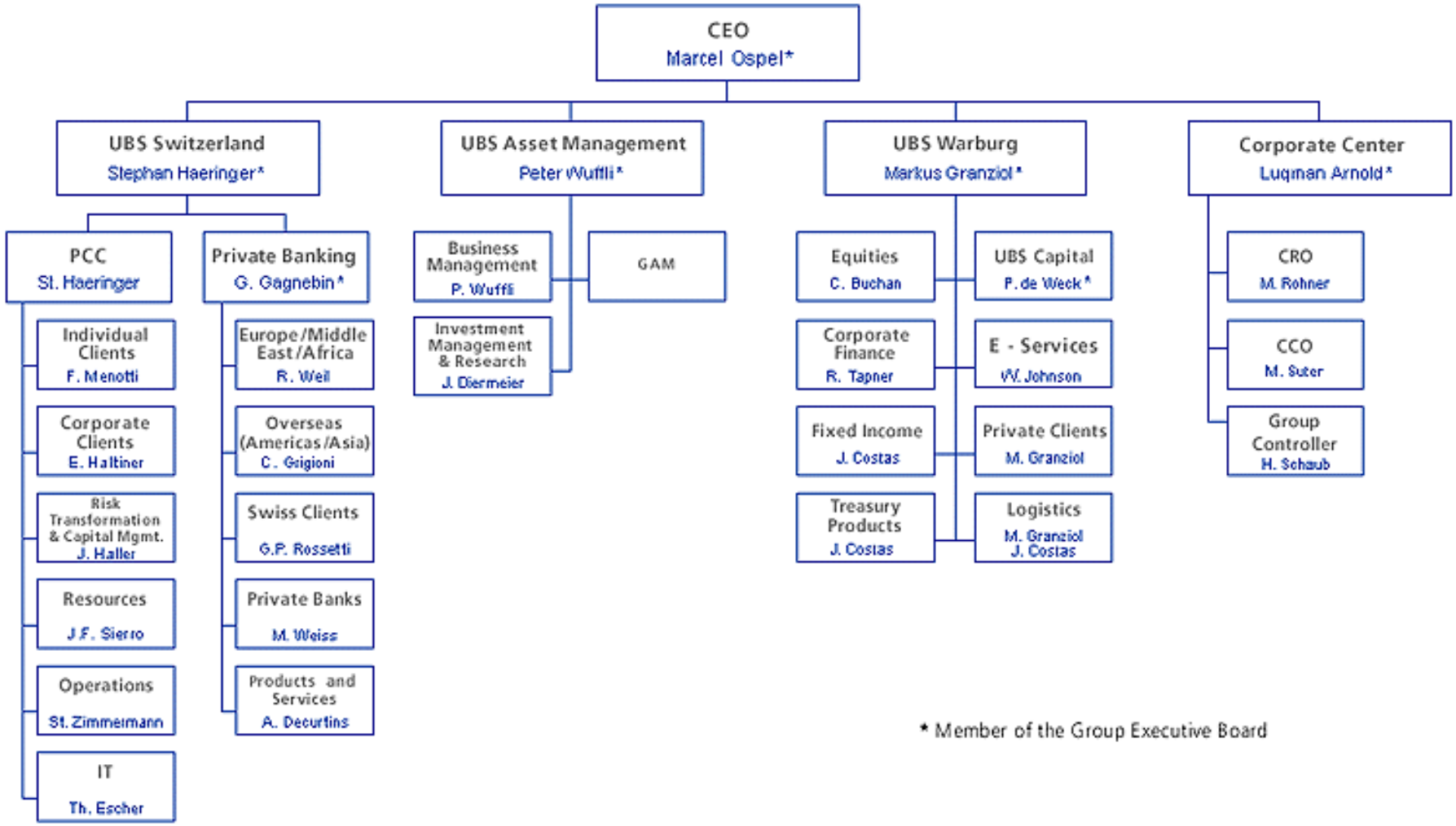
Source: Ernst Haeckel

Sociogramm (1933)



Source: Moreno, 1933

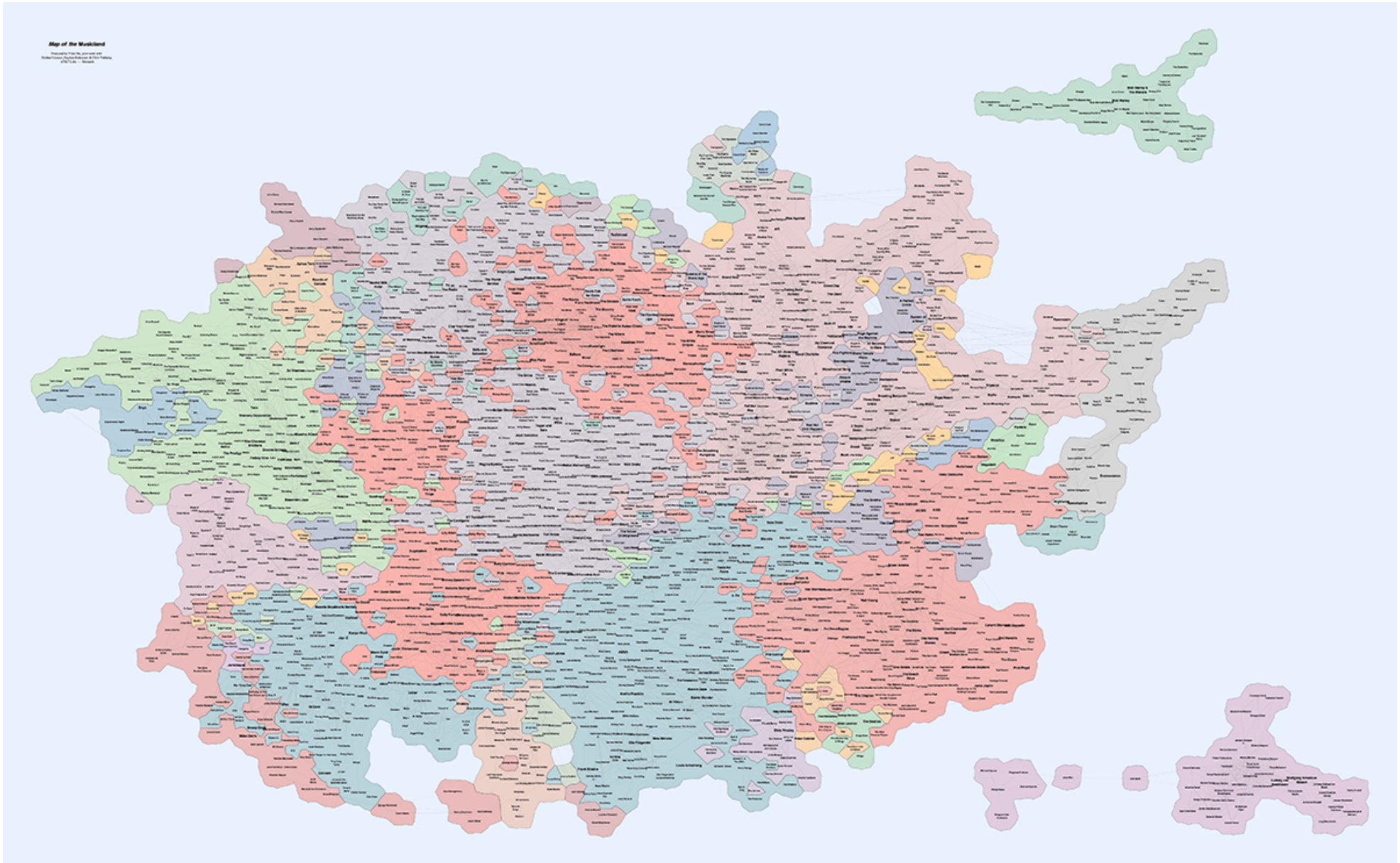
Social Network – Organization within UBS



CPAN Developer-Graph

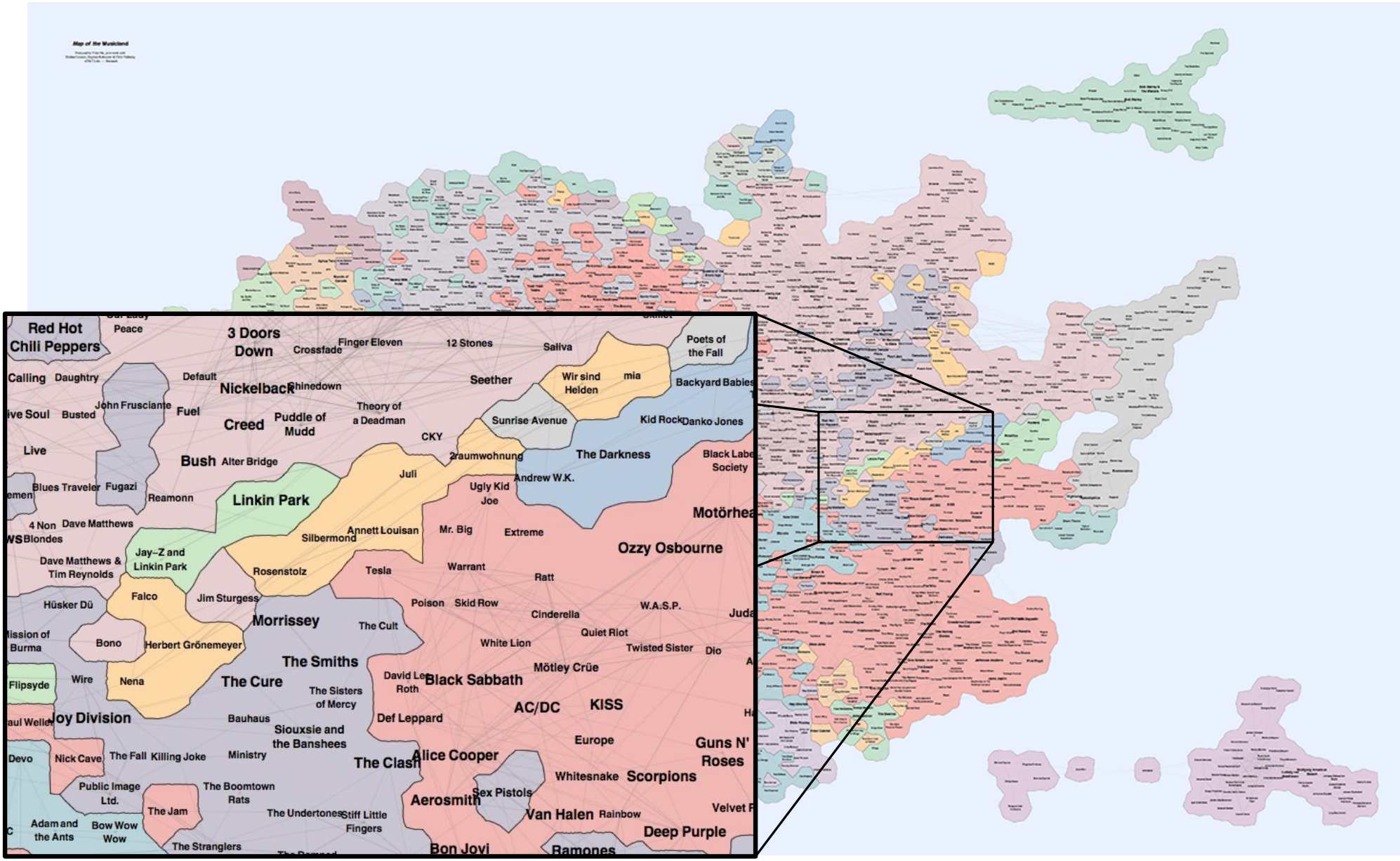
Source: cpan-explorer.org

last.fm Graph of musics as political map



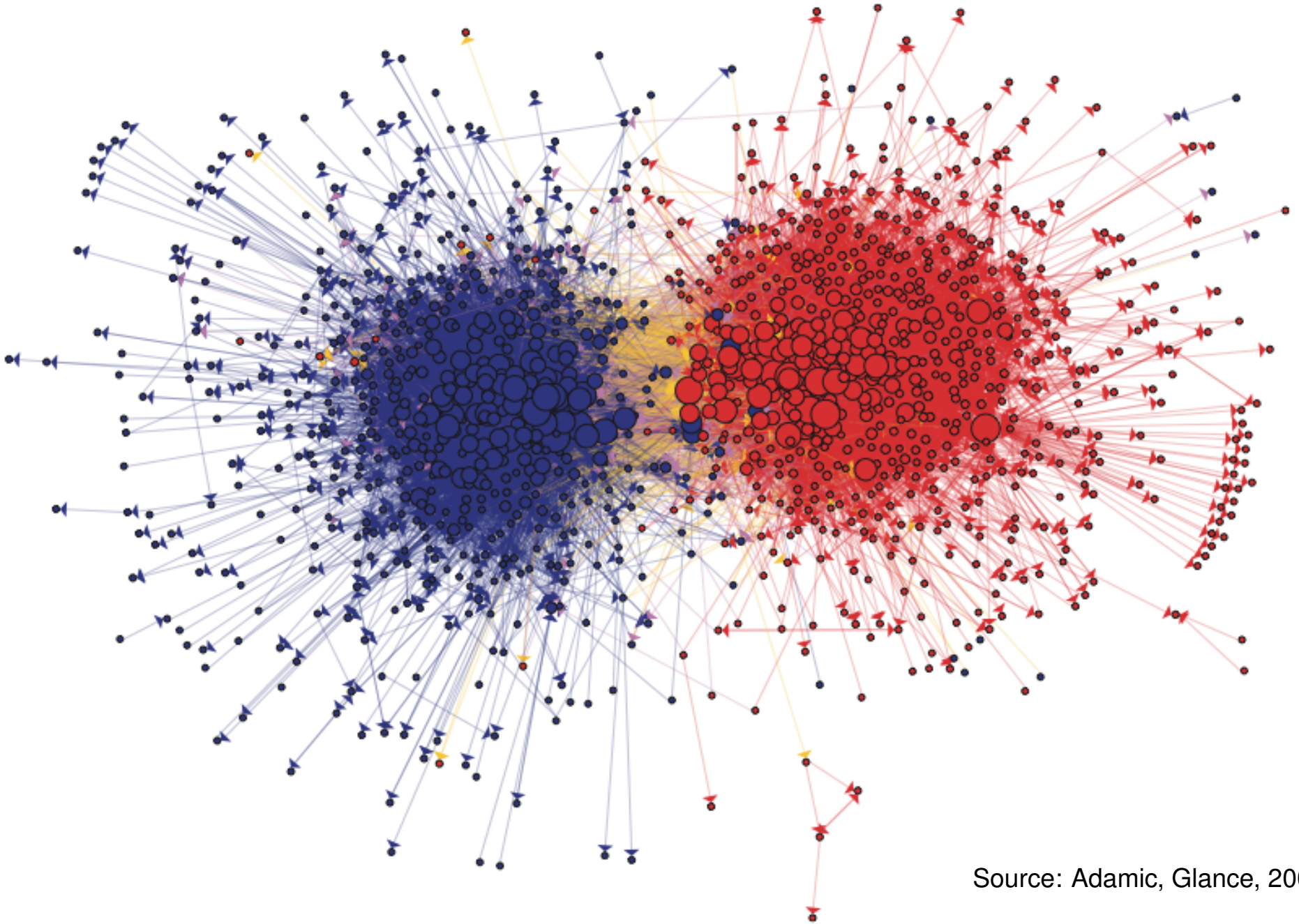
(Gansner, Hu, Kobourov: GMap, 2009

last.fm Graph of musics as political map



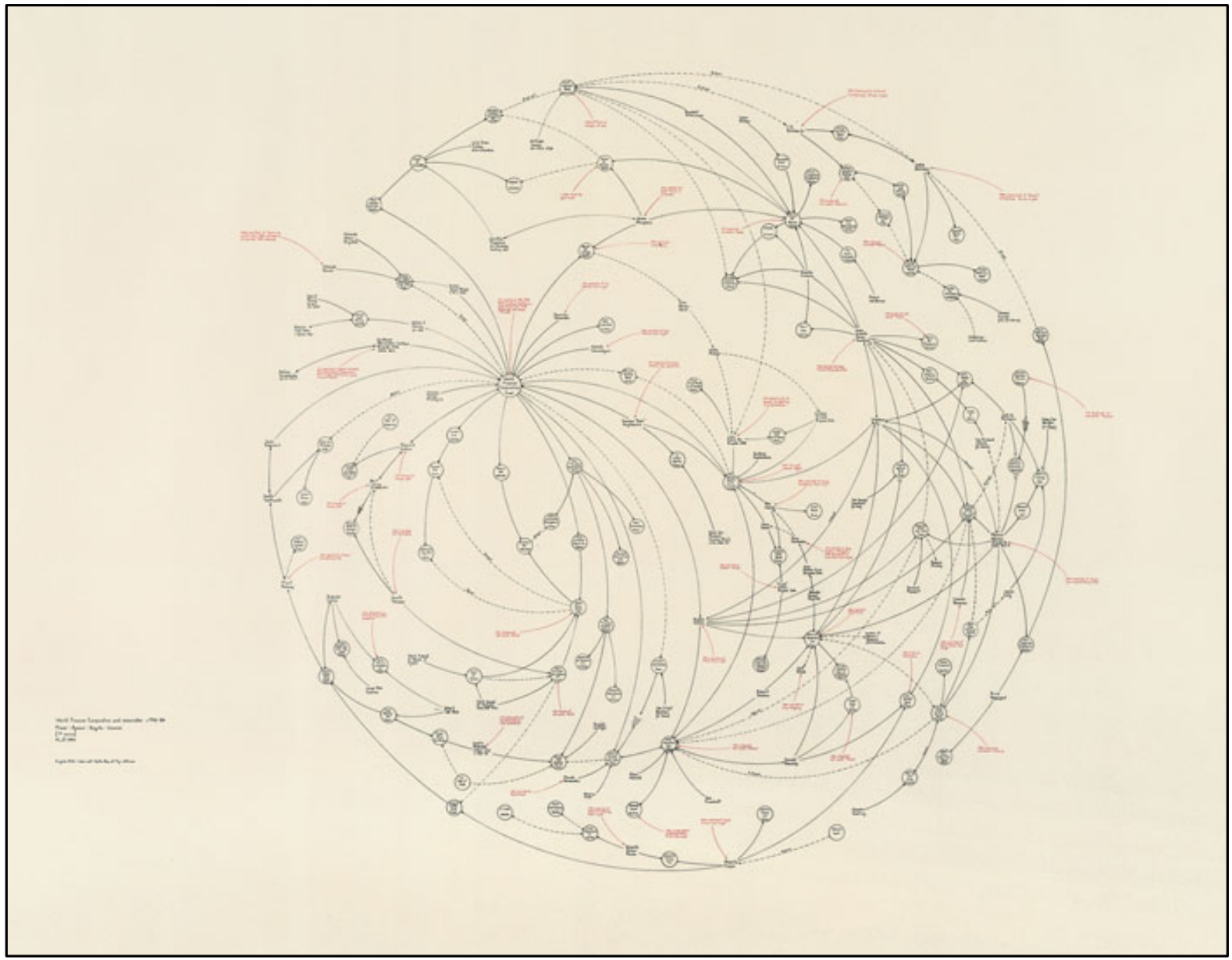
(Gansner, Hu, Kobourov: GMap, 2009

Blogosphere 2004 Elections USA



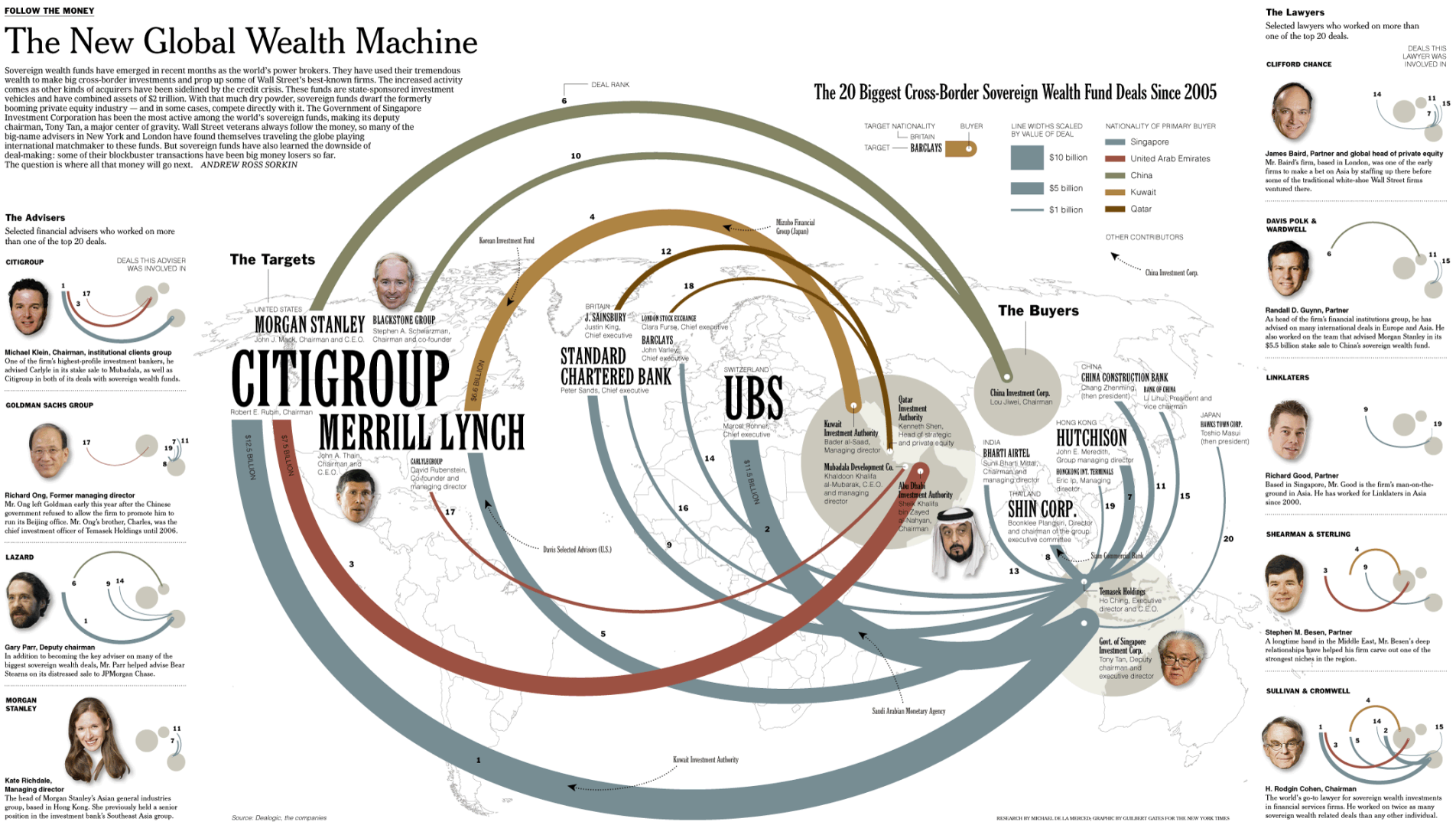
Source: Adamic, Glance, 2005

Social Network – World Finance System

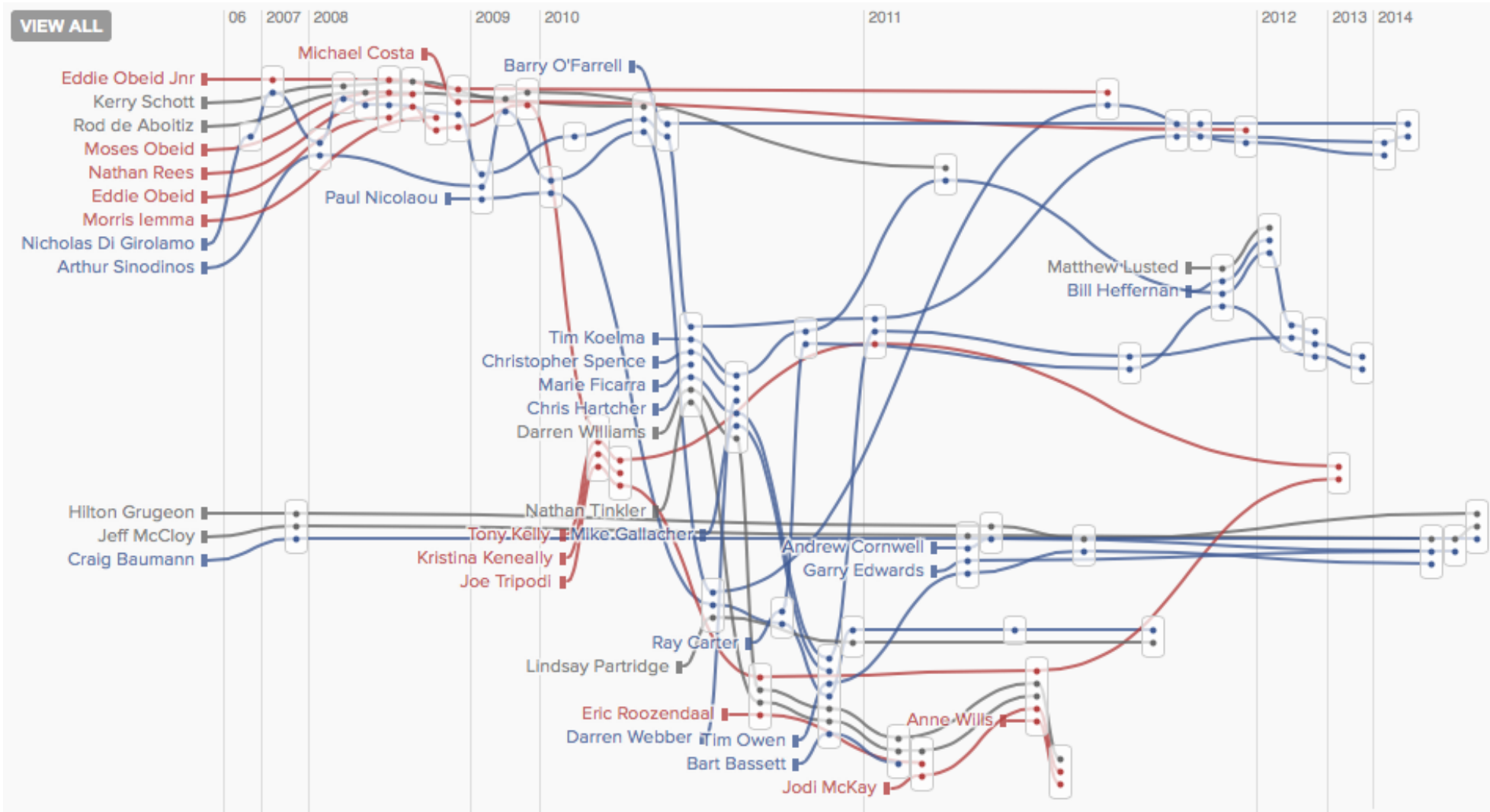


World Finance Corporation
© Mark Lombardi

Social Networks – State Funds

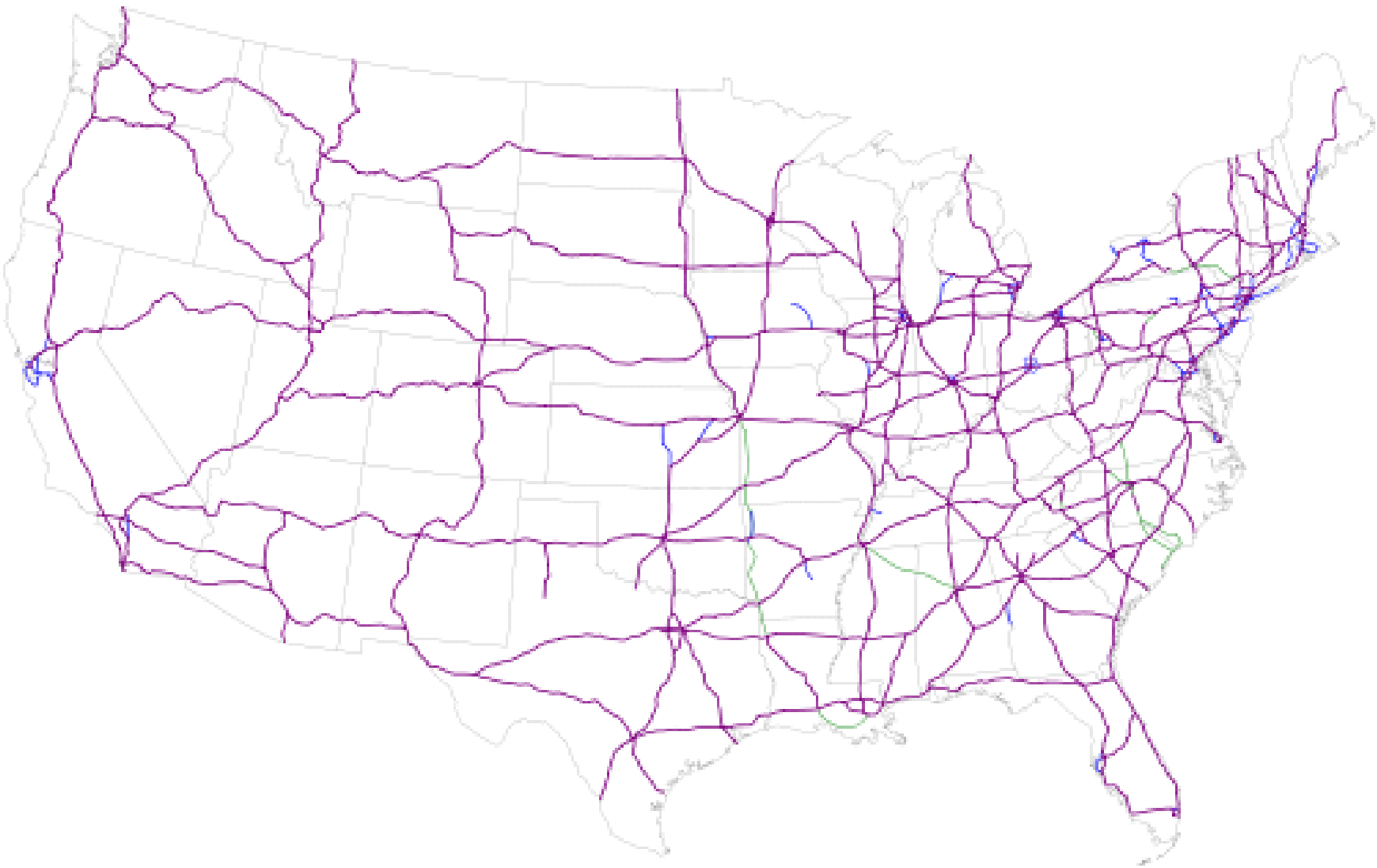


Temporal Graph Layout: Storylines

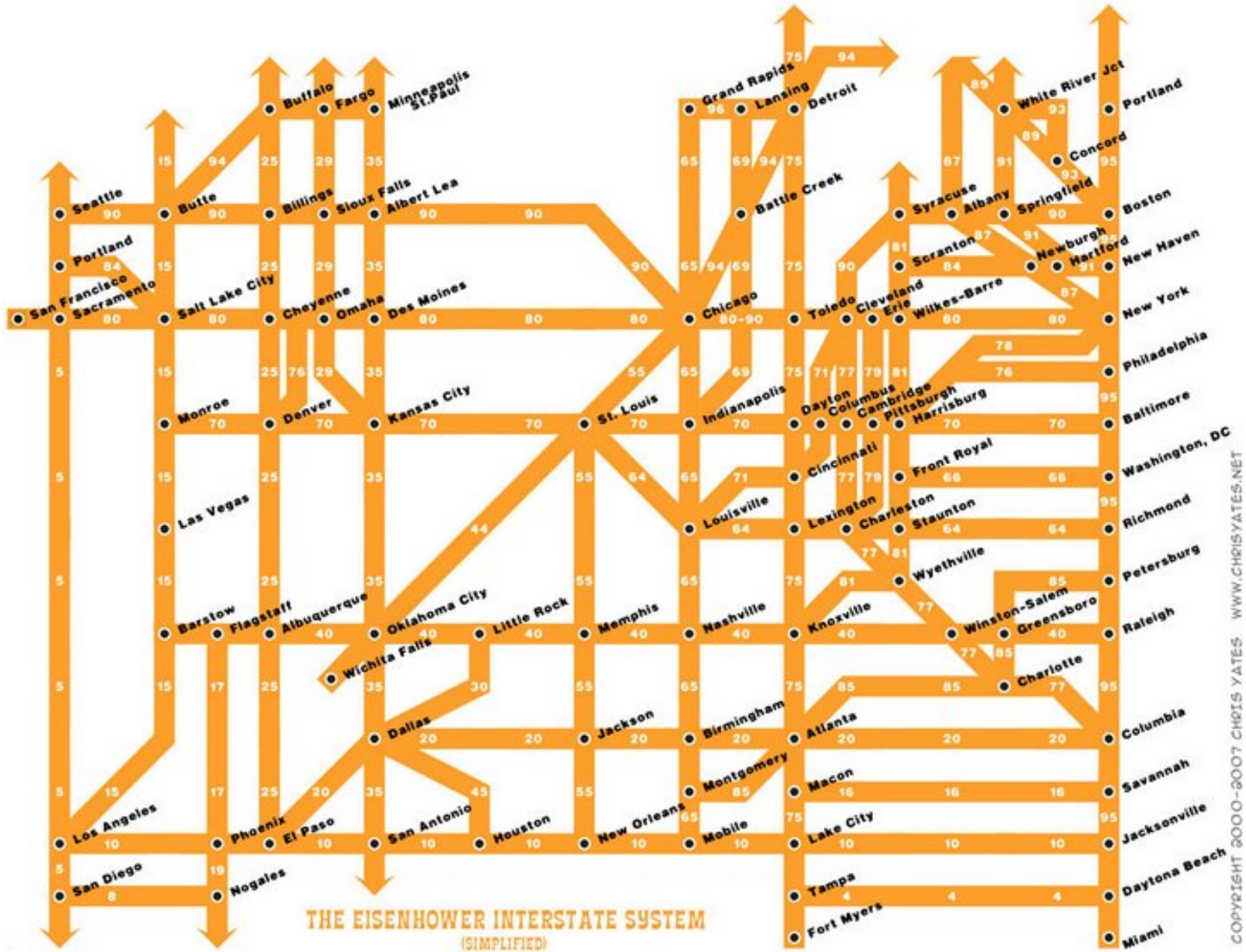


Source: ABC news, Australia

Traffic network – Highways USA



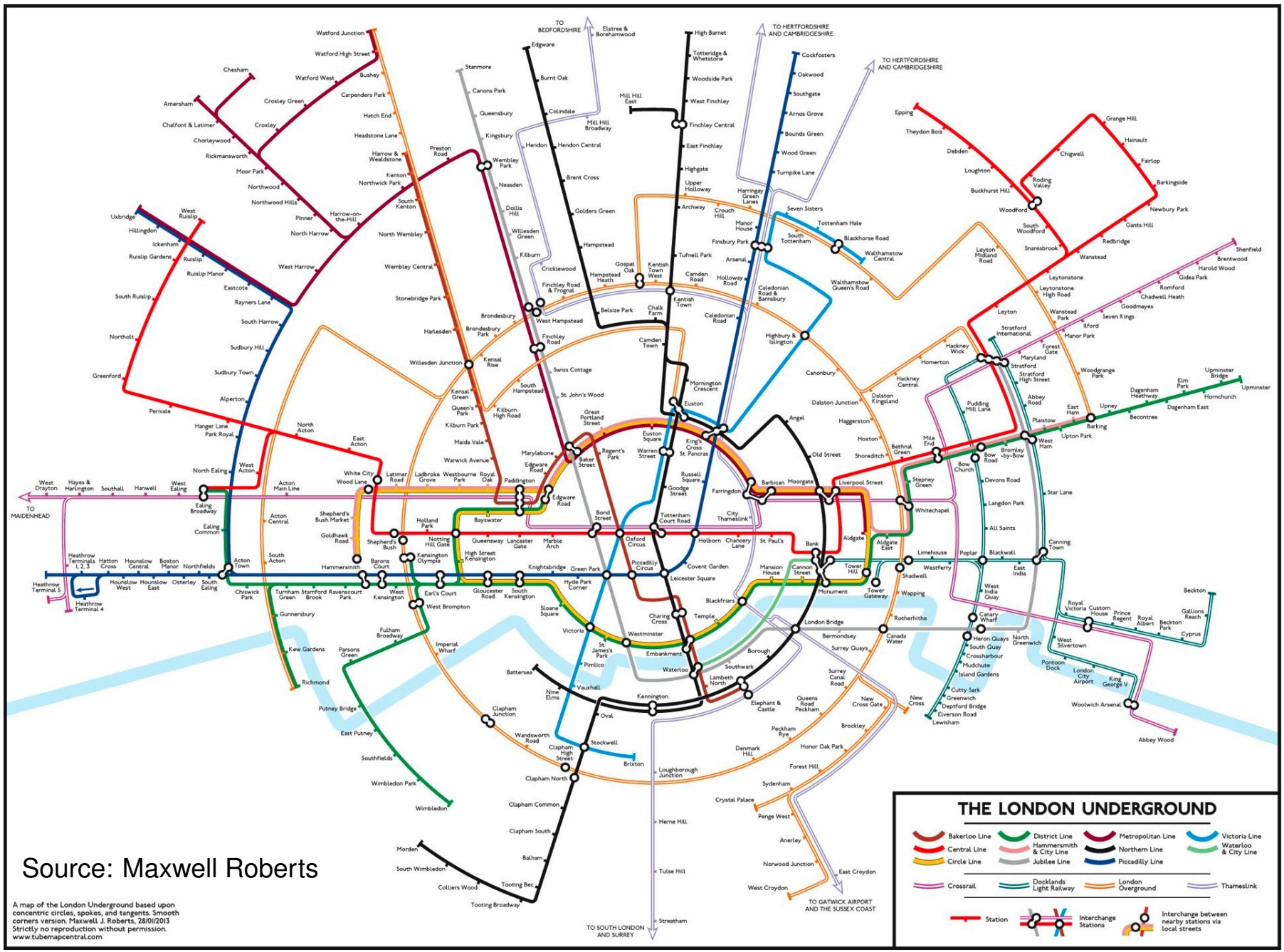
Traffic network – Highways USA



London Tube Map (1933)

Source: Henry Beck

Co-centric Tube Map

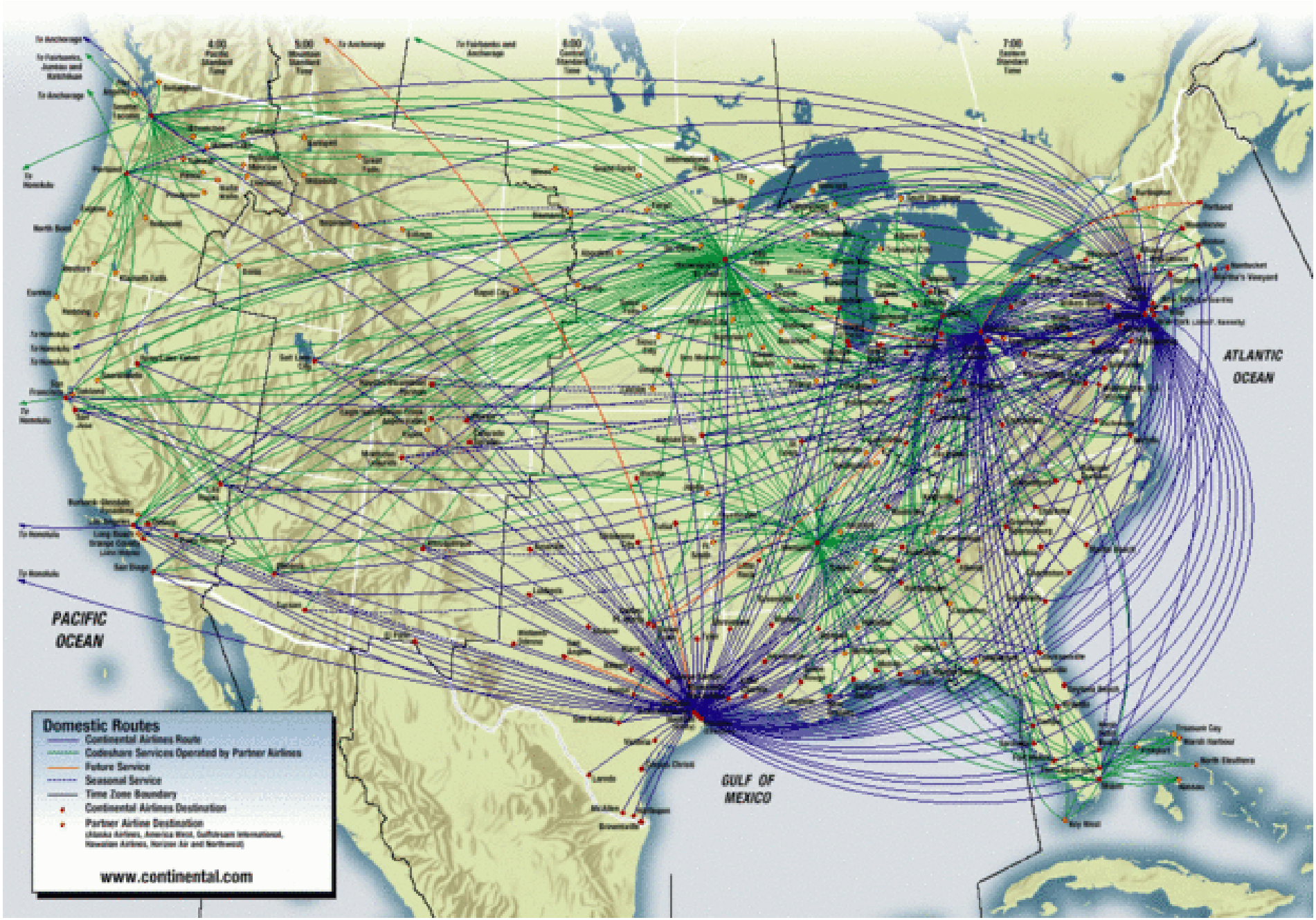


Source: Maxwell Roberts

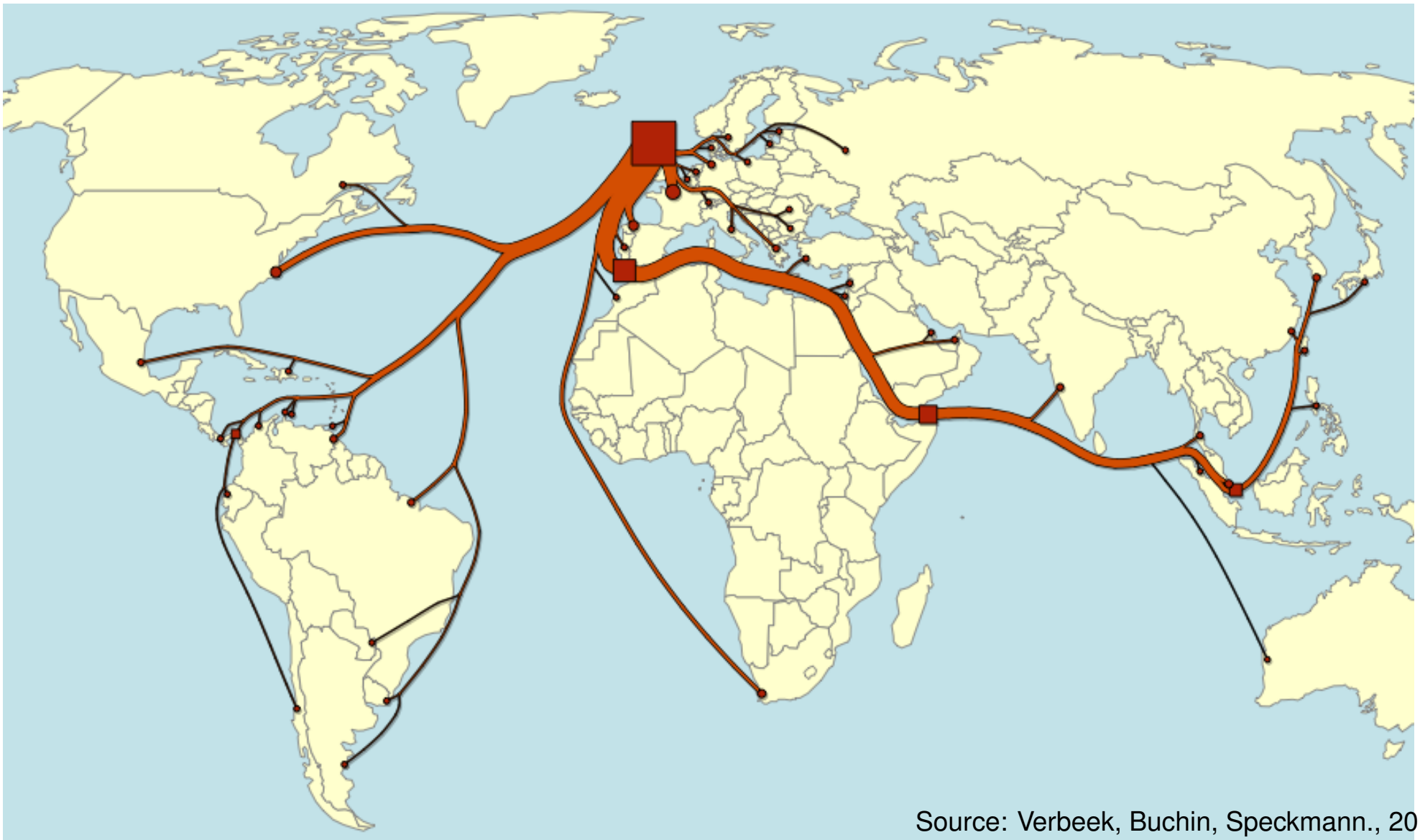
Curvilinear S/U-bahn map

Source: KVV and Maxwell Roberts

Flight Connections

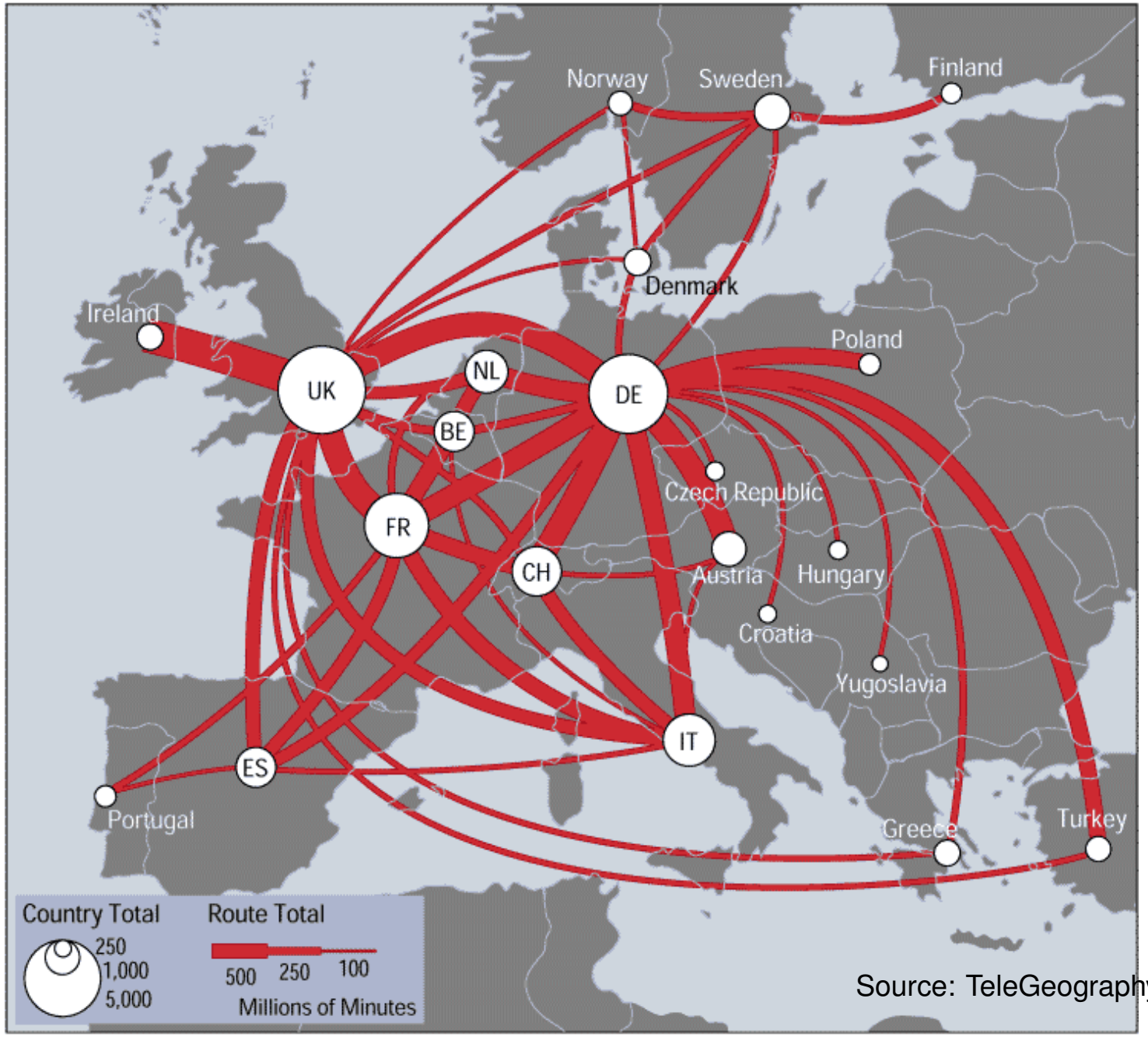


Flow-Map: Whiskey Export



Source: Verbeek, Buchin, Speckmann., 2011

Telephony Map

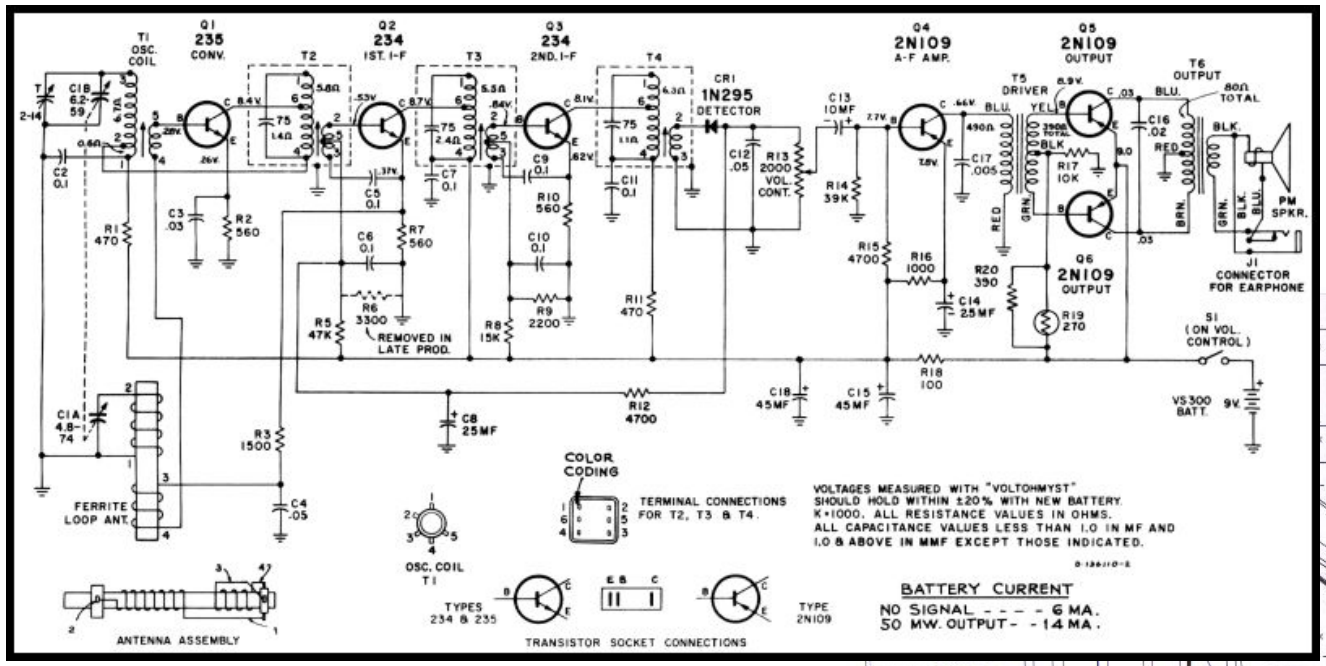


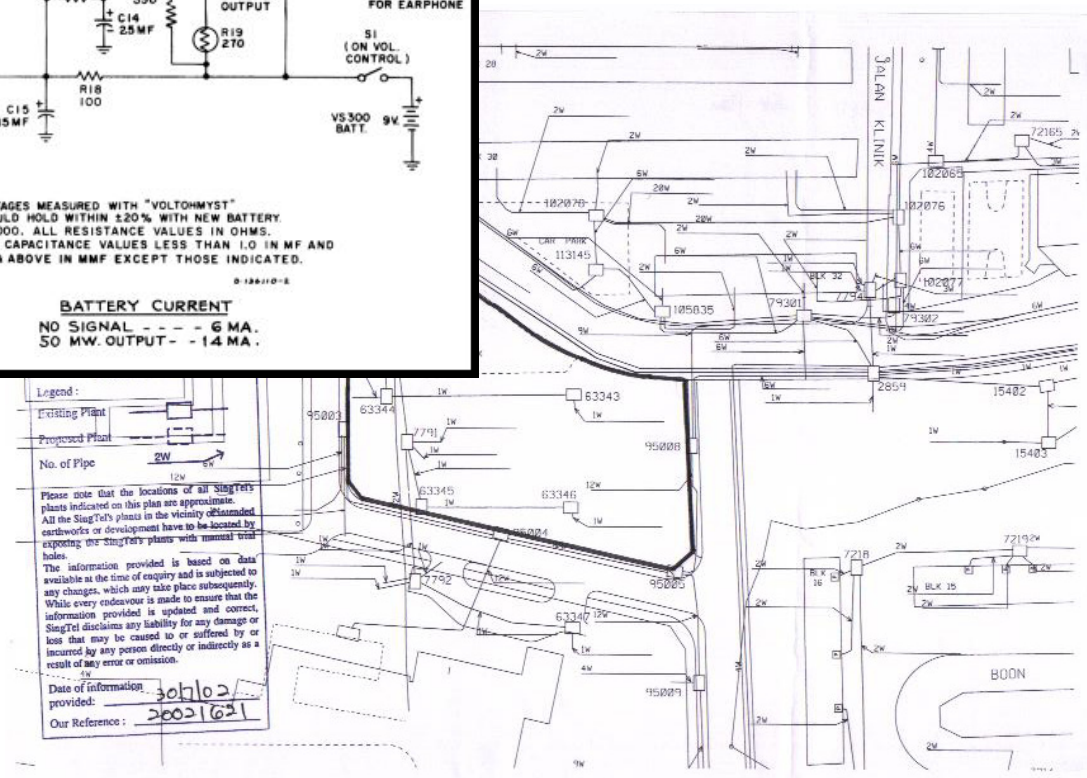
Source: TeleGeography

Monitoring of Energy Network

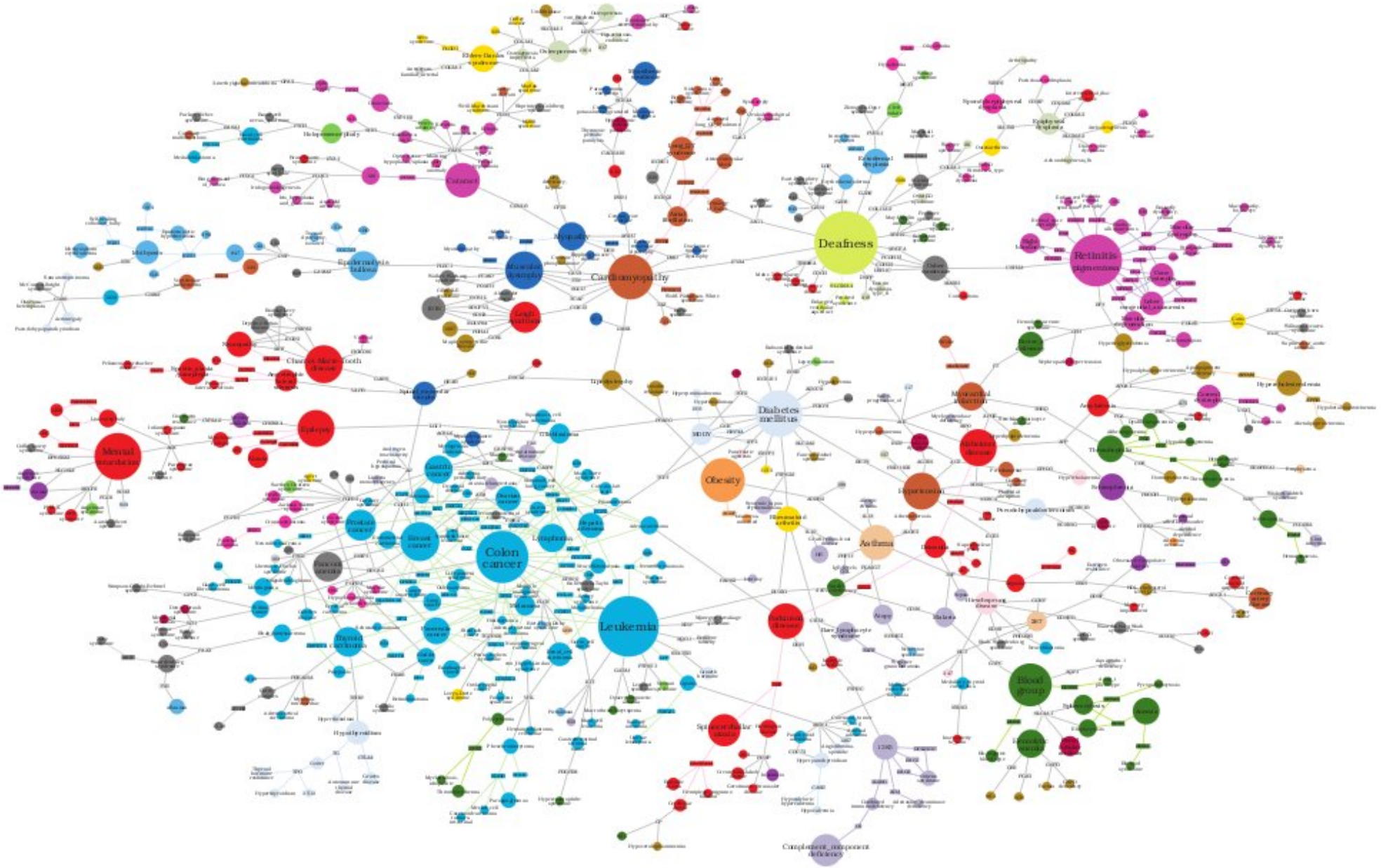
Source: Eir Grid, Irlan

Wiring plan/ Cable plan

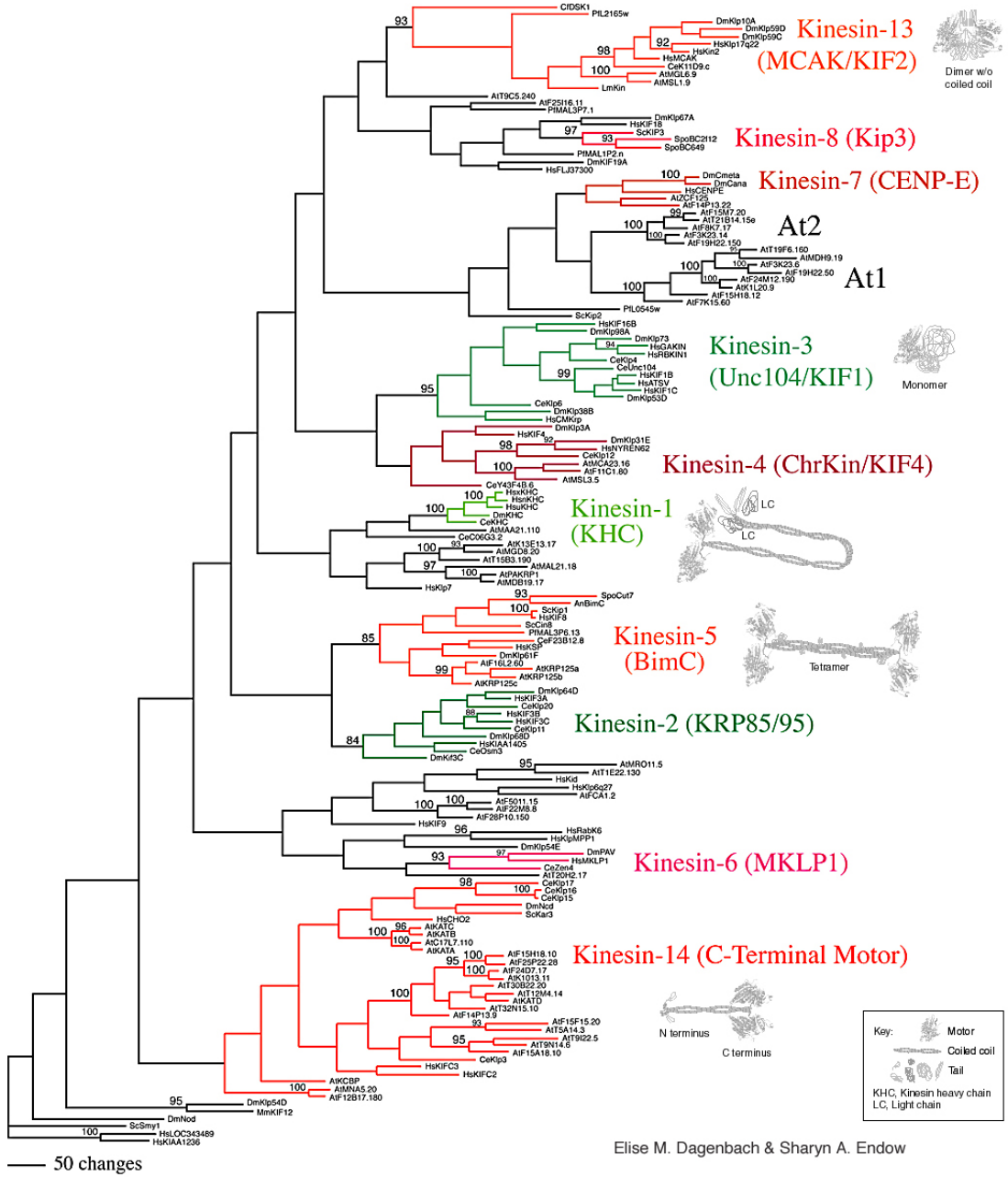




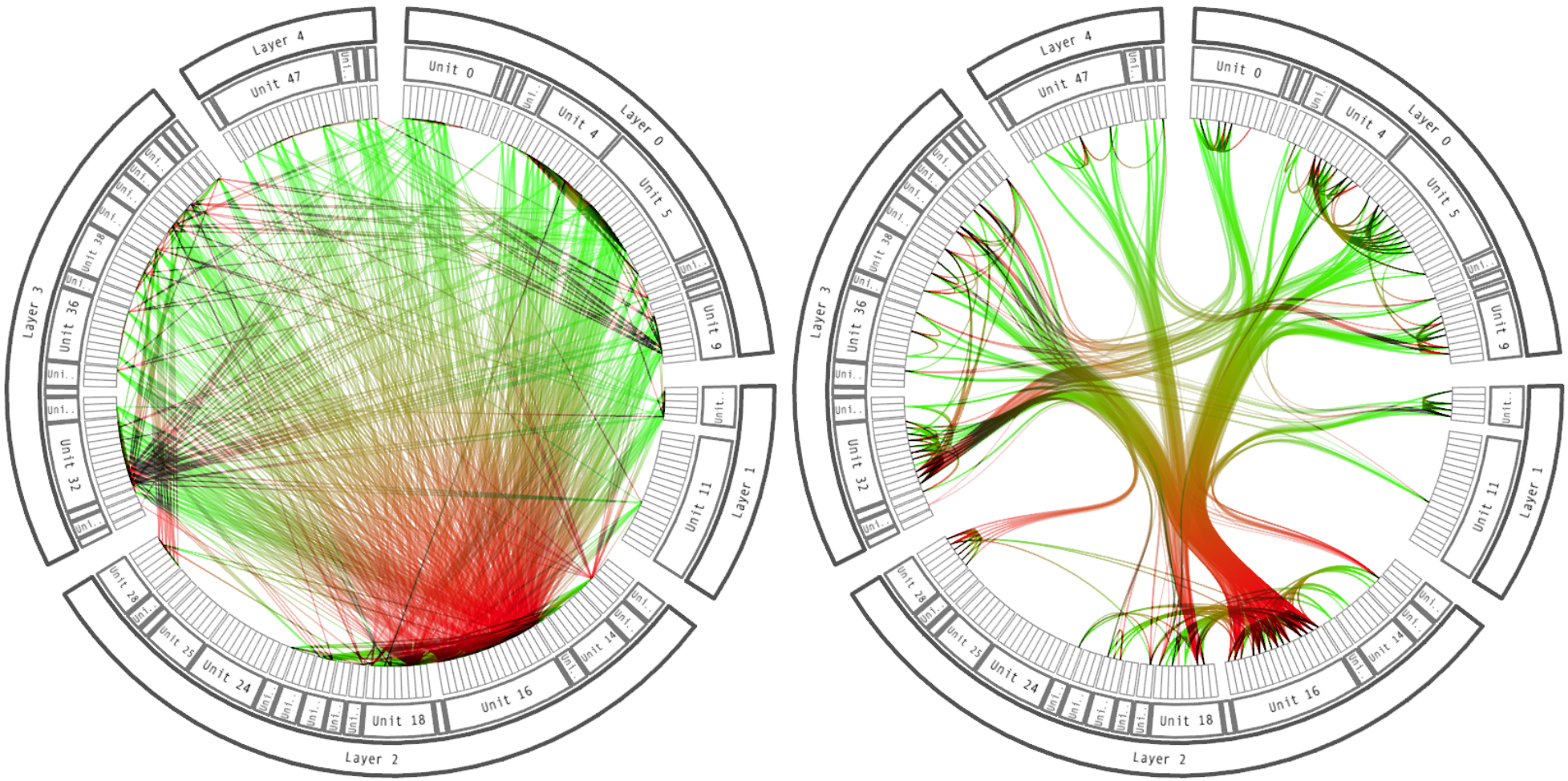
Medicine – Deseases



Medicine – phylogenetic Tree

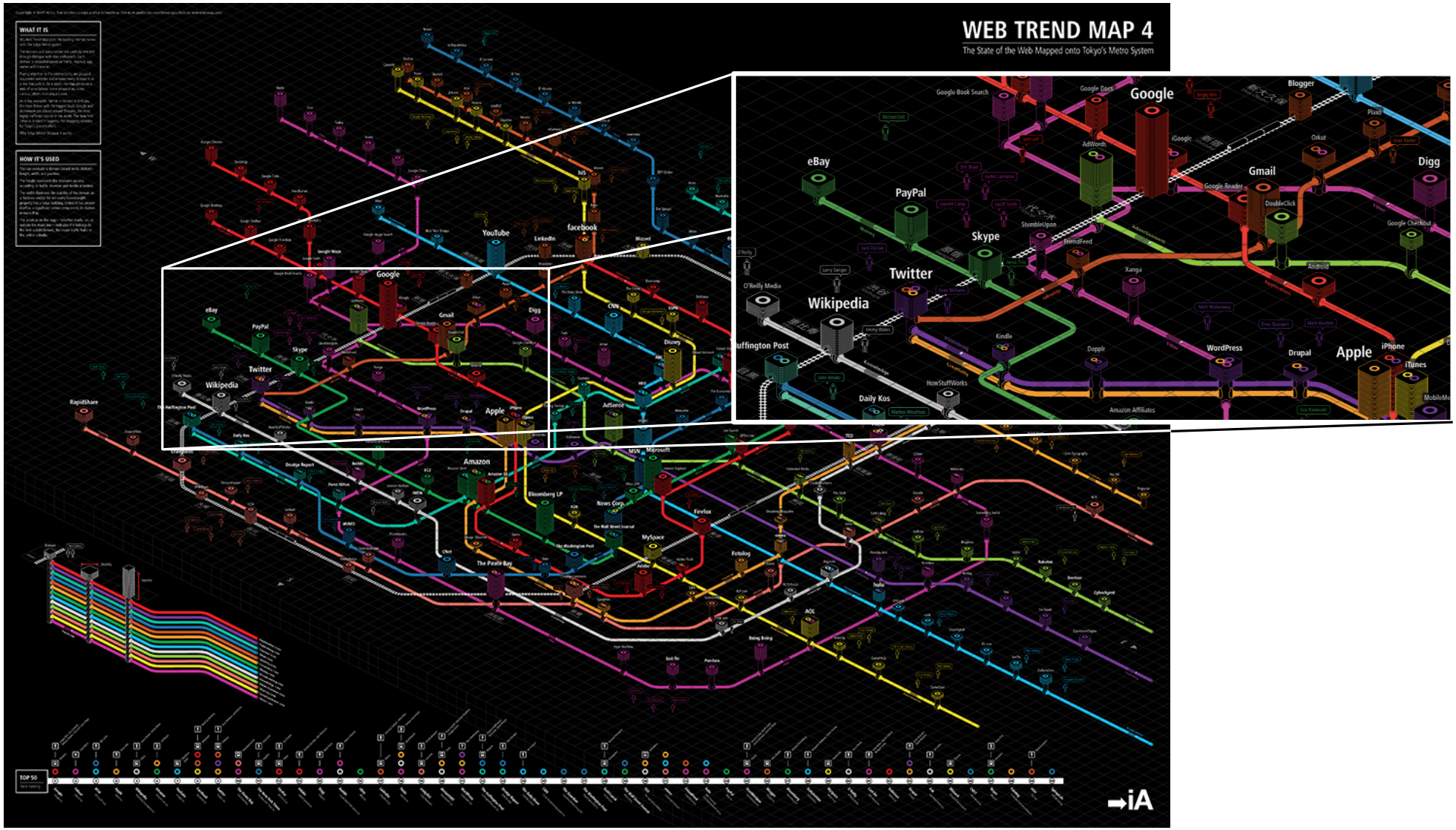


Software Call-Graph with edge-bundling



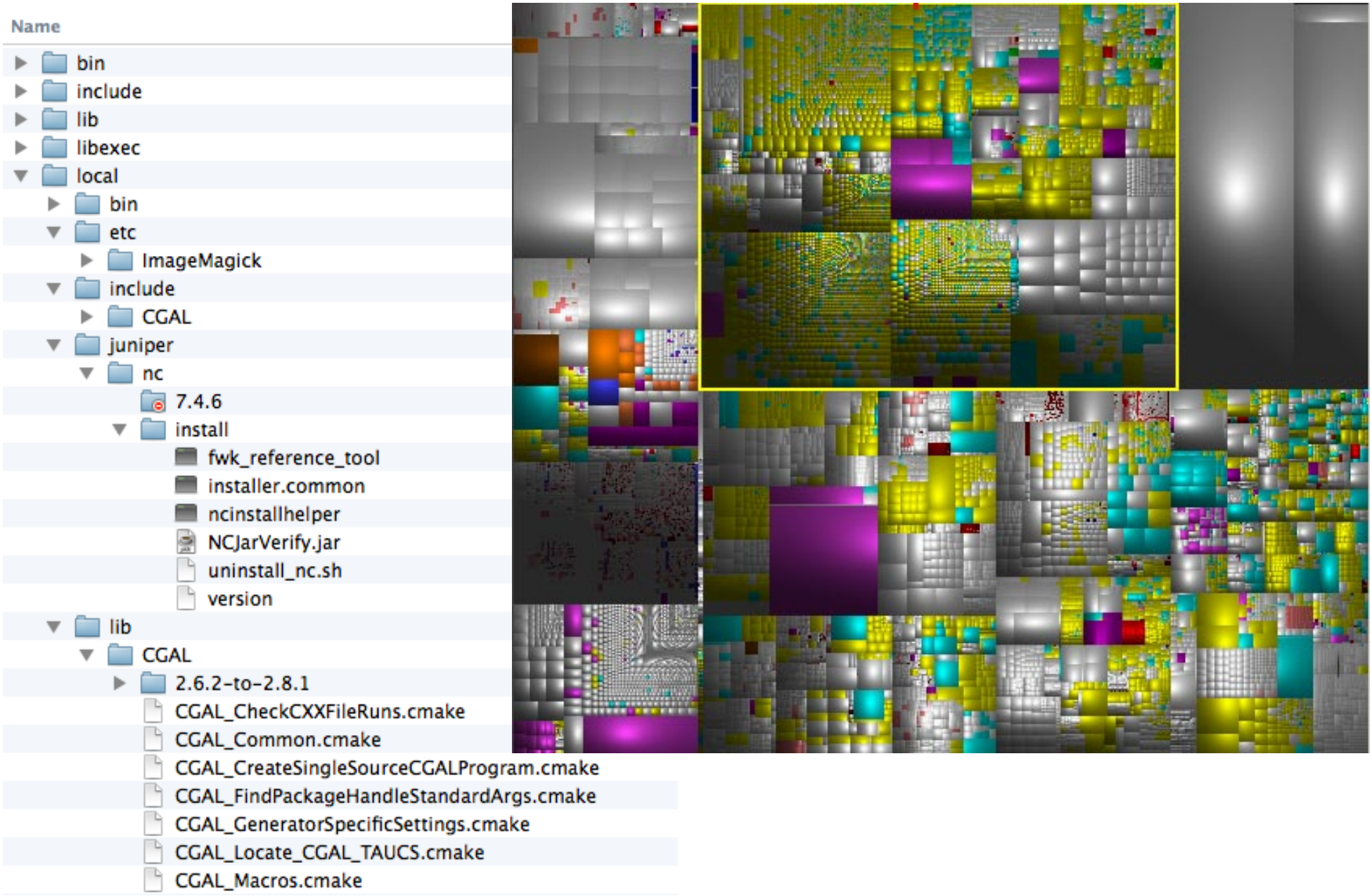
Source: Danny Holten, 2011

Web Trend Map

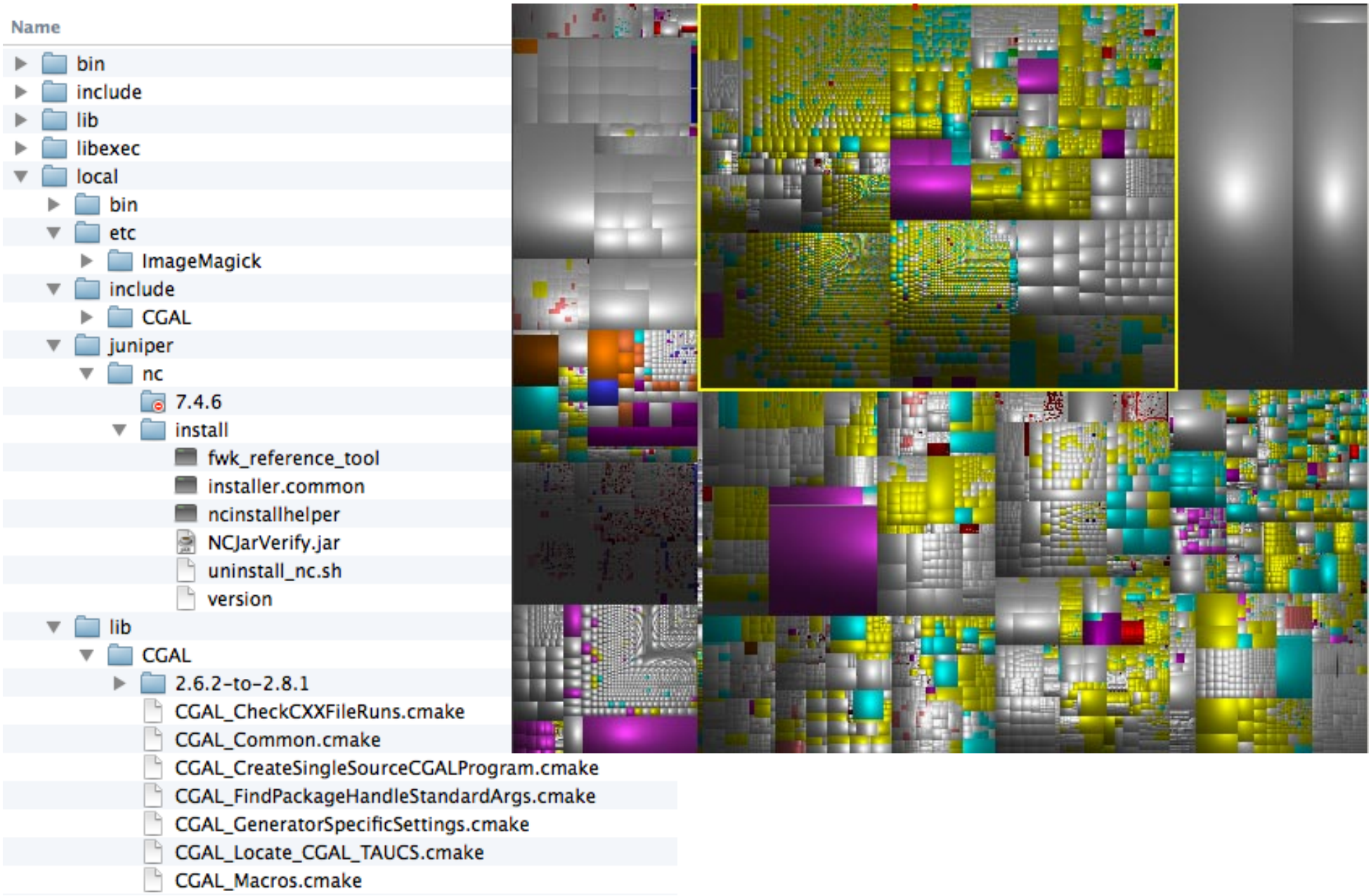


Source: information Architects, 2009

Alternative Visualizations: Explorer vs Treemap



Alternative Visualizations: Explorer vs Treemap



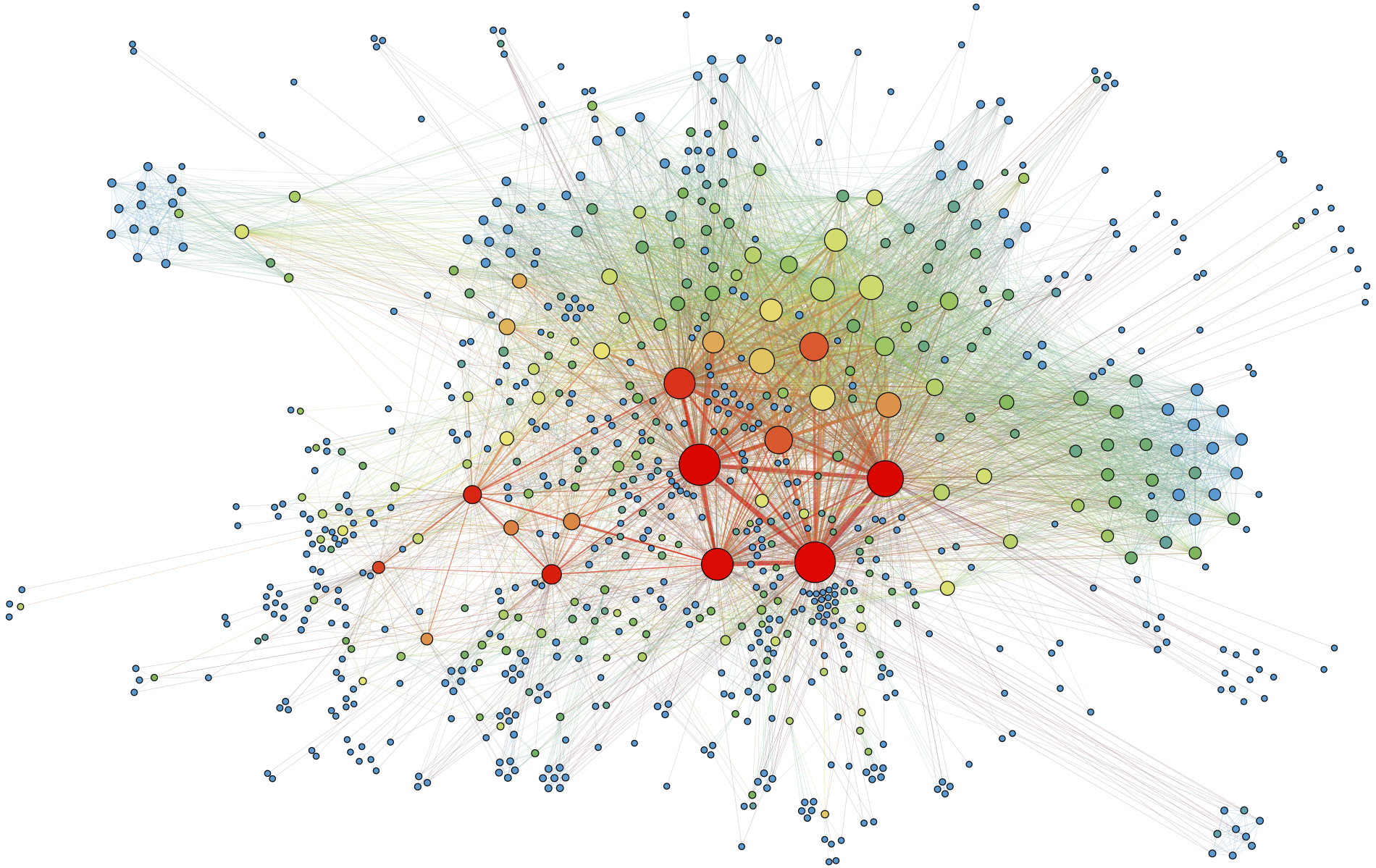
For more: `http://www.visualcomplexity.com/`

Graph and its Representation

What is a Network/Graph?

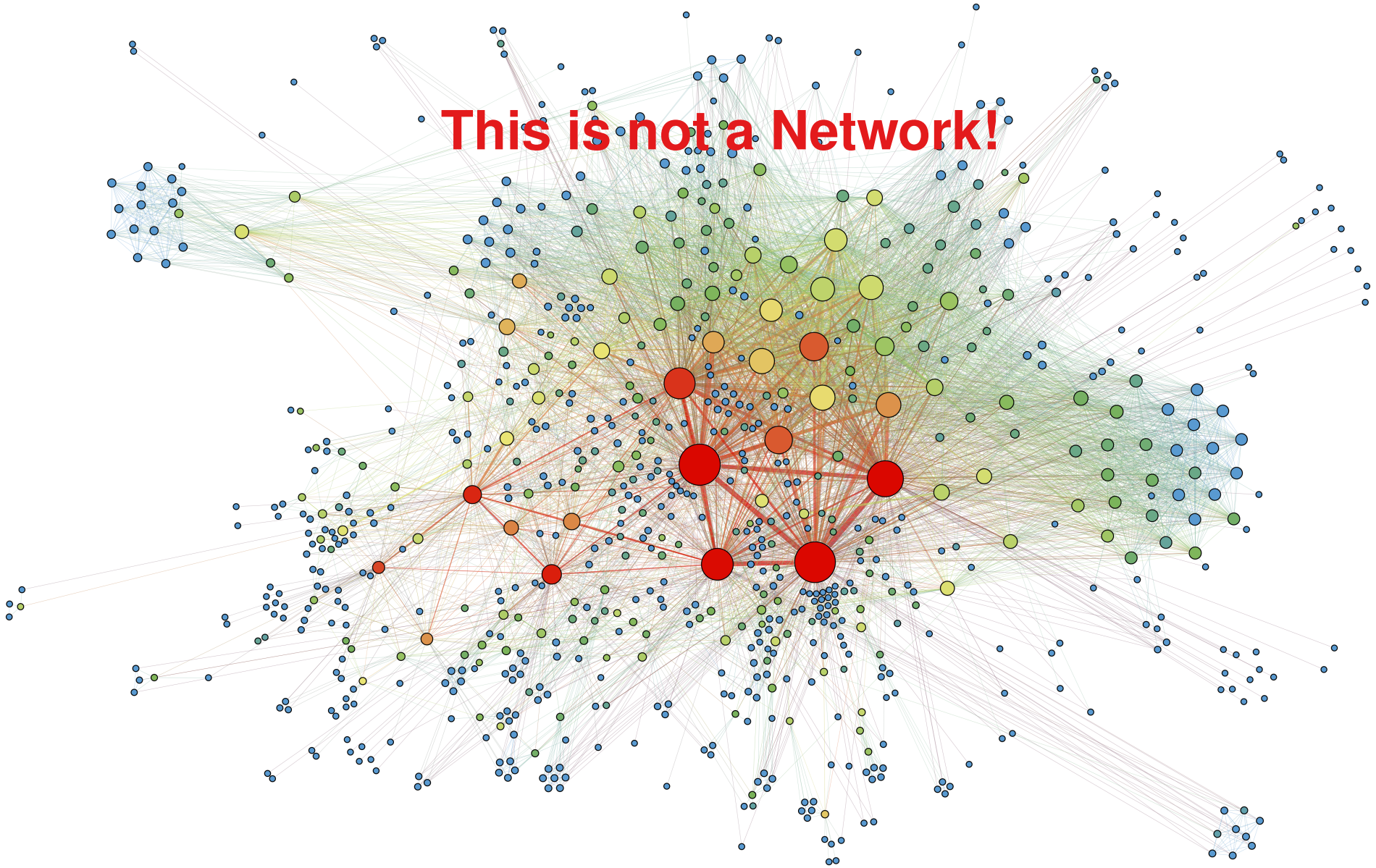
Graph and its Representation

What is a Network/Graph?



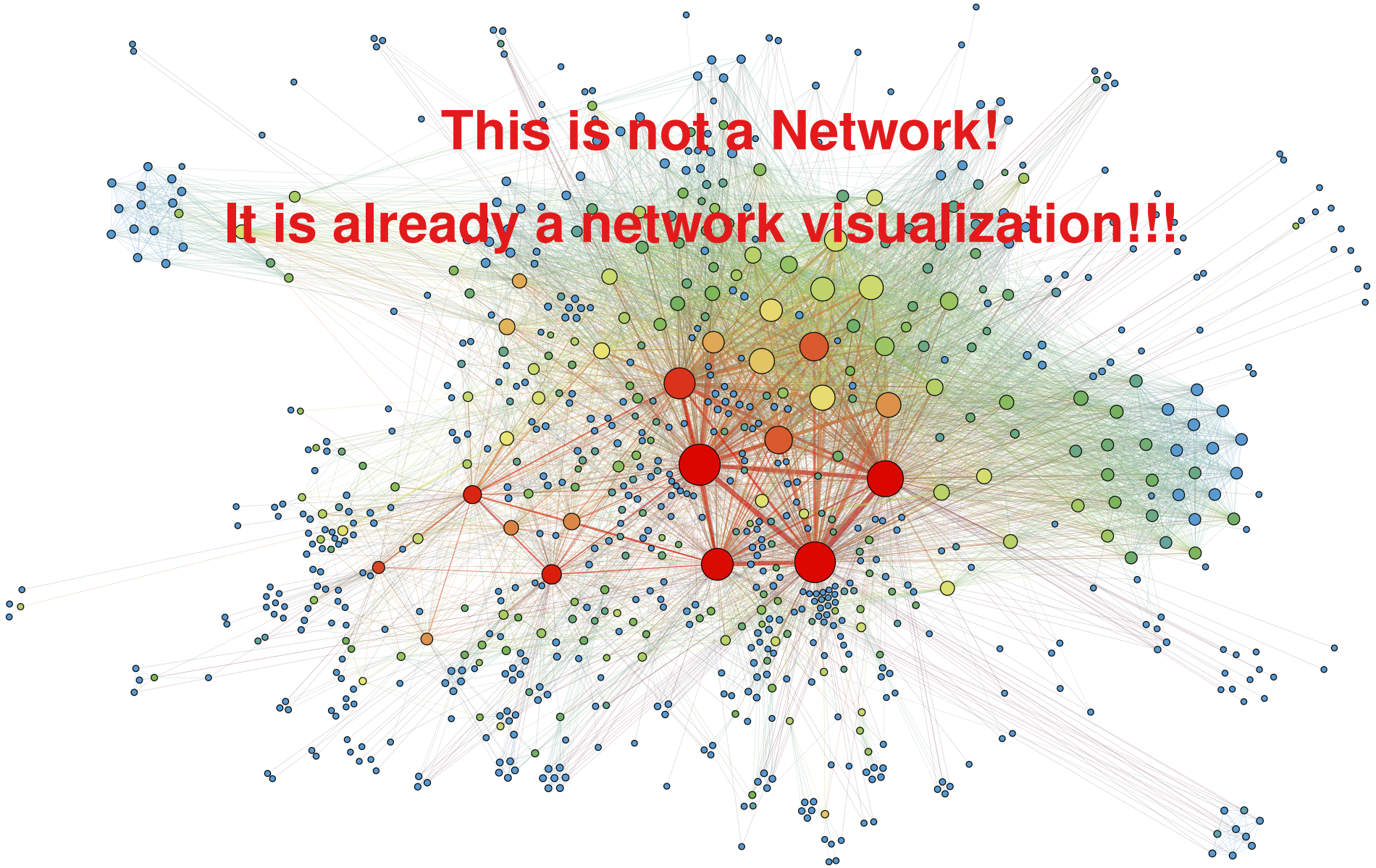
Graph and its Representation

What is a Network/Graph?



Graph and its Representation

What is a Network/Graph?



Graph and its Representation

What is a Network/Graph?

Tuple $G = (V, E)$
Set of vertices/nodes $V = \{v_1, \ldots, v_n\}$
Set of edges/links $E = \{e_1, \ldots, e_m\}$,
$e_i = \{v_j, v_k\},\ 1 \leq i \leq m,\ 1 \leq j, k \leq n$

Graph and its Representation

What is a Network/Graph?

Tuple $G = (V, E)$
Set of vertices/nodes $V = \{v_1, \ldots, v_n\}$
Set of edges/links $E = \{e_1, \ldots, e_m\}$,
$e_i = \{v_j, v_k\},\ 1 \leq i \leq m,\ 1 \leq j, k \leq n$

Representations?

Graph and its Representation

What is a Network/Graph?

Tuple $G = (V, E)$
Set of vertices/nodes $V = \{v_1, \ldots, v_n\}$
Set of edges/links $E = \{e_1, \ldots, e_m\}$,
$e_i = \{v_j, v_k\},\ 1 \leq i \leq m,\ 1 \leq j, k \leq n$

Representations?

Set representation:

$$V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}\}\ E = \{\{v_1, v_2\}, \{v_1, v_8\}, \{v_2, v_3\}, \{v_3, v_5\}, \{v_3, v_9\} \{v_3, v_{10}\}, \{v_4, v_5\}, \{v_4, v_6\}, \{v_4, v_9\}, \{v_5, v_8\} \{v_6, v_8\}, \{v_6, v_9\}, \{v_7, v_8\}, \{v_7, v_9\}, \{v_8, v_{10}\} \{v_9, v_{10}\}\}$$

Graph and its Representation

What is a Network/Graph?

Tuple $G = (V, E)$
Set of vertices/nodes $V = \{v_1, \ldots, v_n\}$
Set of edges/links $E = \{e_1, \ldots, e_m\}$,
$e_i = \{v_j, v_k\},\ 1 \leq i \leq m,\ 1 \leq j, k \leq n$

Representations?

Set representation
Adjacency list

v_1 : v_2, v_8
v_2 : v_1, v_3
v_3 : v_2, v_5, v_9, v_{10}
v_4 : v_5, v_6, v_9
v_5 : v_3, v_4, v_8
v_6 : v_4, v_8, v_9
v_7 : v_8, v_9
v_8 : $v_1, v_5, v_6, v_7, v_9, v_{10}$
v_9 : $v_3, v_4, v_6, v_7, v_8, v_{10}$
v_{10} : v_3, v_8, v_9

Graph and its Representation

What is a Network/Graph?

Tuple $G = (V, E)$
Set of vertices/nodes $V = \{v_1, \ldots, v_n\}$
Set of edges/links $E = \{e_1, \ldots, e_m\}$,
$e_i = \{v_j, v_k\},\ 1 \leq i \leq m,\ 1 \leq j, k \leq n$

Representations?

Set representation
Adjacency list
Adjacency matrix

$$\begin{pmatrix}
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 1 \\
0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0
\end{pmatrix}$$

Graph and its Representation

What is a Network/Graph?

Tuple $G = (V, E)$
Set of vertices/nodes $V = \{v_1, \ldots, v_n\}$
Set of edges/links $E = \{e_1, \ldots, e_m\}$,
$e_i = \{v_j, v_k\},\ 1 \leq i \leq m,\ 1 \leq j, k \leq n$

Representations?

We will use terms Graph and Network interchangeably, but commonly Network assumes a more complex data set, attributes on nodes and links, clusters, etc.

Let’s draw some graphs

Work with your neighbour or in groups of three **15 min**

- go to Teams/General/Class Materials/Graphs to draw
- graphs in form of adjacency matrix/list
- Use yEd: https://www.yworks.com/products/yed/download
- draw one or more graphs as nice and as readable as possible
- export to PNG/PDF and upload to Teams/Network Visualization/Drawn Graphs
- keep the file name, add name of the team

We will show and discuss the results afterwards

Layout Problem

Graph visualization problem

given : Graph $G = (V, E)$
find: good drawing Γ of G

- $\Gamma : V \to \mathbb{R}^2$, nodes $v \to$ point $\Gamma(v)$
- $\Gamma : E \to$ curves in $\mathbb{R}^2$, edge $\{u, v\} \to$ simple open curve $c_{uv} : [0, 1] \to \mathbb{R}^2$ where $c_{uv}(0) = \Gamma(u)$ and $c_{uv}(1) = \Gamma(v)$

Layout Problem

Graph visualization problem

given : Graph $G = (V, E)$

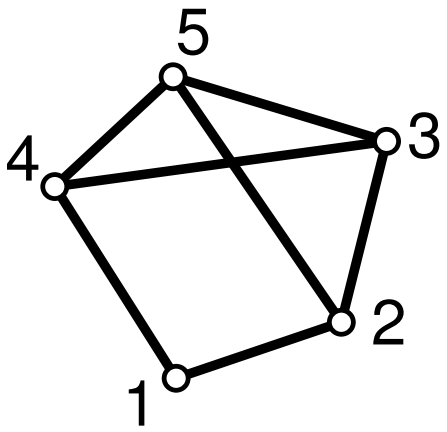
find: **good** drawing Γ of G

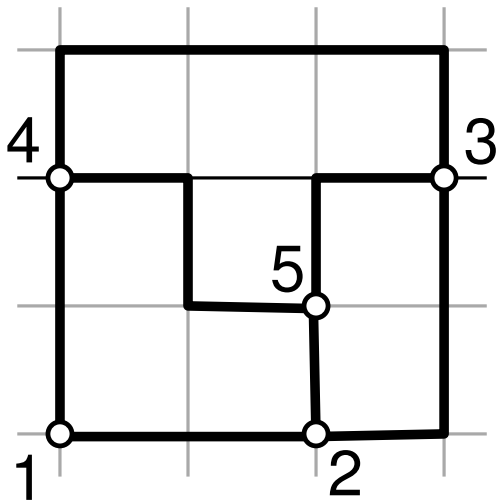
- $\Gamma : V \to \mathbb{R}^2$, nodes $v \to$ point $\Gamma(v)$
- $\Gamma : E \to$ curves in $\mathbb{R}^2$, edge $\{u, v\} \to$ simple open curve $c_{uv} : [0, 1] \to \mathbb{R}^2$ where $c_{uv}(0) = \Gamma(u)$ and $c_{uv}(1) = \Gamma(v)$

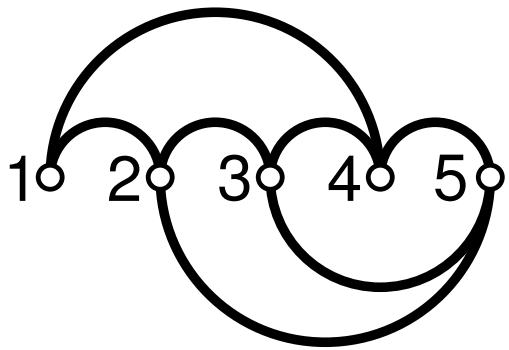
Layout Problem

1) **Drawing conventions :** required properties, for example:

- straight-line edges
- orthogonal edges (polylines with 90º bends)
- drawing on a grid
- crossing-free





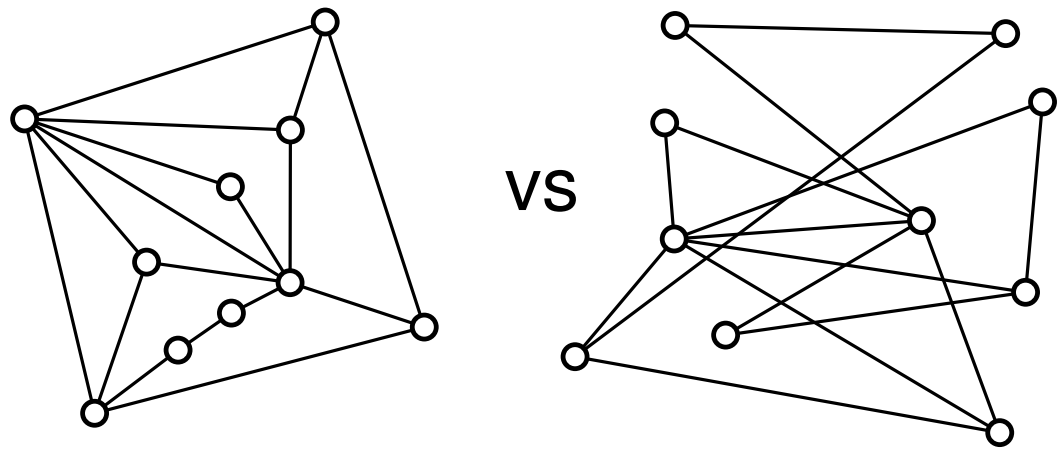


Layout Problem

1) **Drawing conventions :** drawing rules that stem from applications

2) **Quality metrics :** criteria to be optimized that make drawing more readable⋆

- number of crossing
- number of bends
- area/length
- crossing resolution
- stress of layout



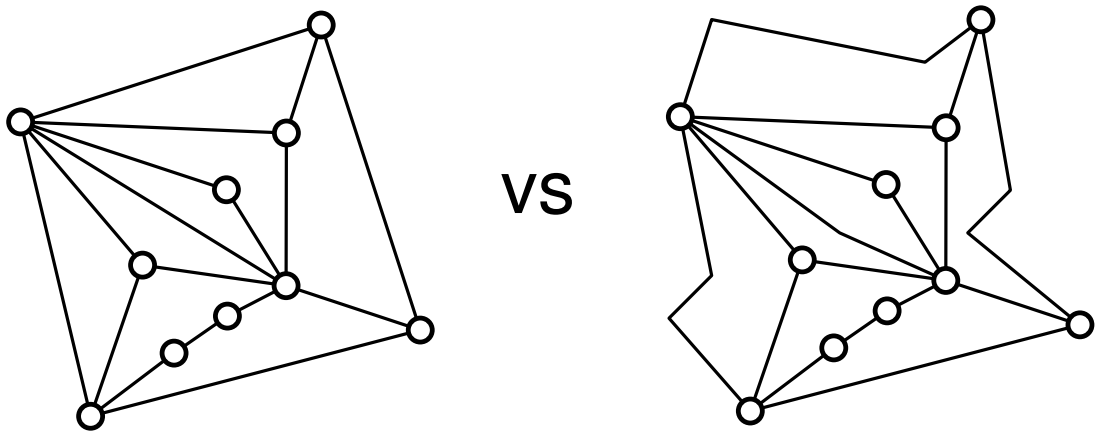
⋆ insight from cognitive psychology

Layout Problem

1) **Drawing conventions :** drawing rules that stem from applications

2) **Quality metrics :** criteria to be optimized that make drawing more readable⋆

- number of crossing
- number of bends
- area/length
- crossing resolution
- stress of layout



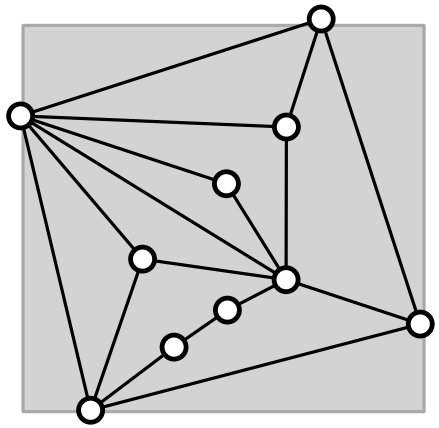
⋆ insight from cognitive psychology

Layout Problem

1) **Drawing conventions :** drawing rules that stem from applications

2) **Quality metrics :** criteria to be optimized that make drawing more readable⋆

- number of crossing
- number of bends
- area/length
- crossing resolution
- stress of layout



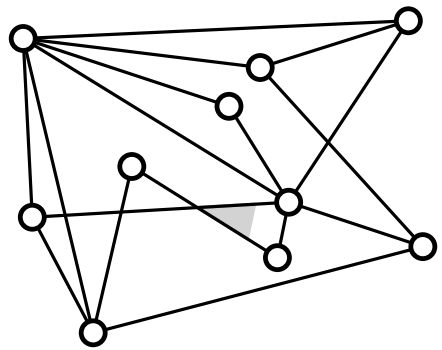
⋆ insight from cognitive psychology

Layout Problem

1) **Drawing conventions :** drawing rules that stem from applications

2) **Quality metrics :** criteria to be optimized that make drawing more readable⋆

- number of crossing
- number of bends
- area/length
- crossing resolution
- stress of layout



the smallest angle formed by two crossing edges

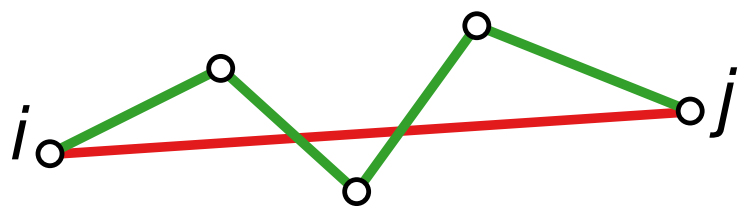
⋆ insight from cognitive psychology

Layout Problem

1) **Drawing conventions :** drawing rules that stem from applications

2) **Quality metrics :** criteria to be optimized that make drawing more readable*

- number of crossing
- number of bends
- area/length
- crossing resolution
- stress of layout



$$\sum_{i<j} \frac{1}{d_{ij}^{\alpha}} (\|X_i - X_j\| - d_{ij})^2$$

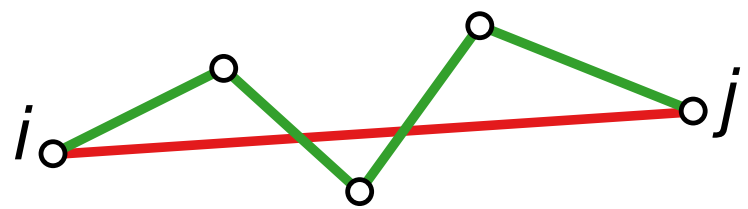
$\star$ insight from cognitive psychology

Layout Problem

1) **Drawing conventions :** drawing rules that stem from applications

2) **Quality metrics :** criteria to be optimized that make drawing more readable*

- number of crossing
- number of bends
- area/length
- crossing resolution
- stress of layout



$$\sum_{i<j} \frac{1}{d_{ij}^{\alpha}} (\|X_i - X_j\| - d_{ij})^2$$

3) **Partial constraints :** rules to be applied on a part of the drawing

$\star$ insight from cognitive psychology

Layout Problem – Second Attempt

Graph visualization problem

given: Graph $G = (V, E)$
find: a drawing Γ of G, that

- complies with drawing conventions
- optimizes aesthetics
- satisfies local/partial constraints

Layout Problem – Second Attempt

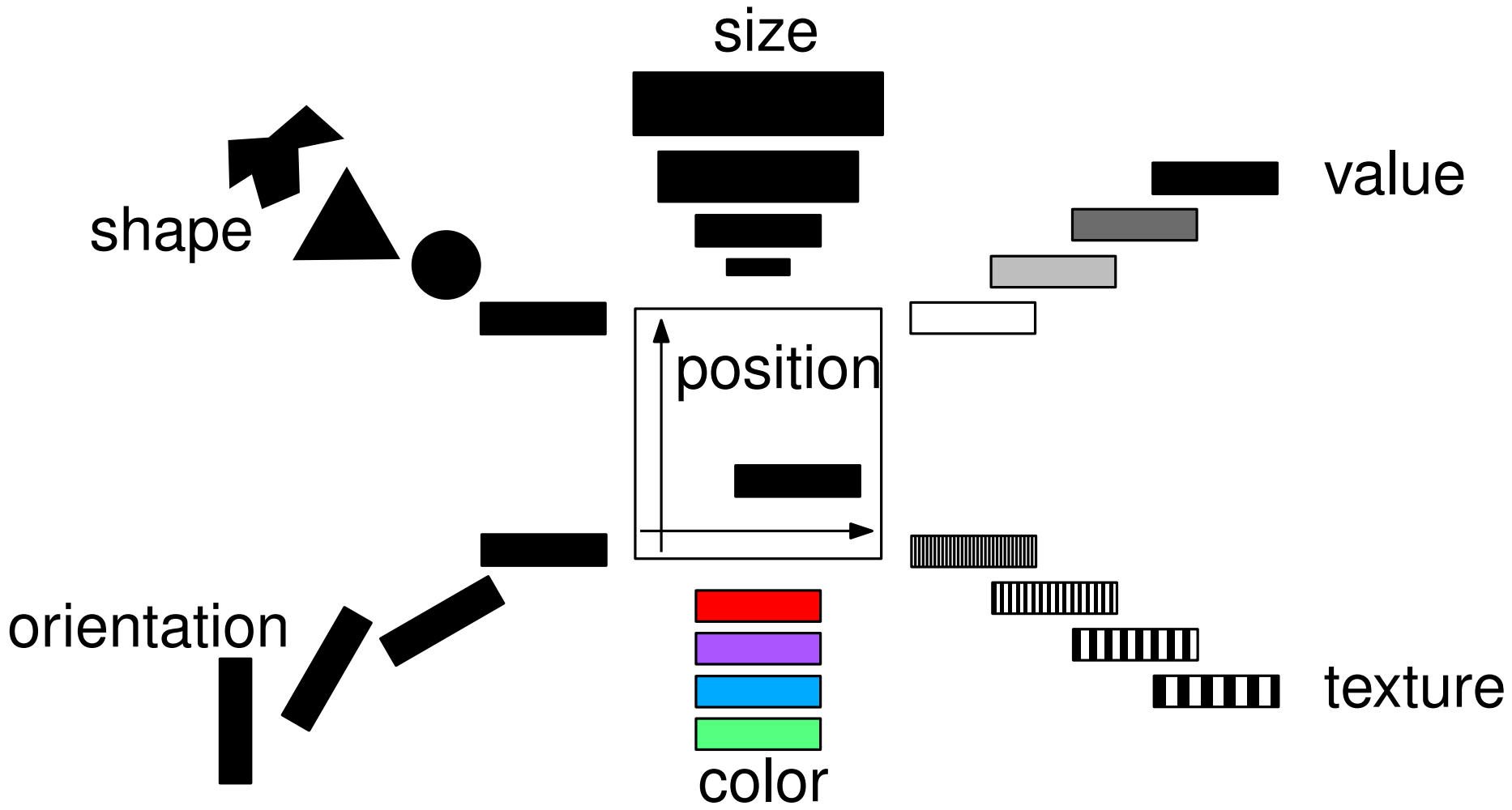
Graph visualization problem

given: Graph $G = (V, E)$
find: a drawing Γ of G, that

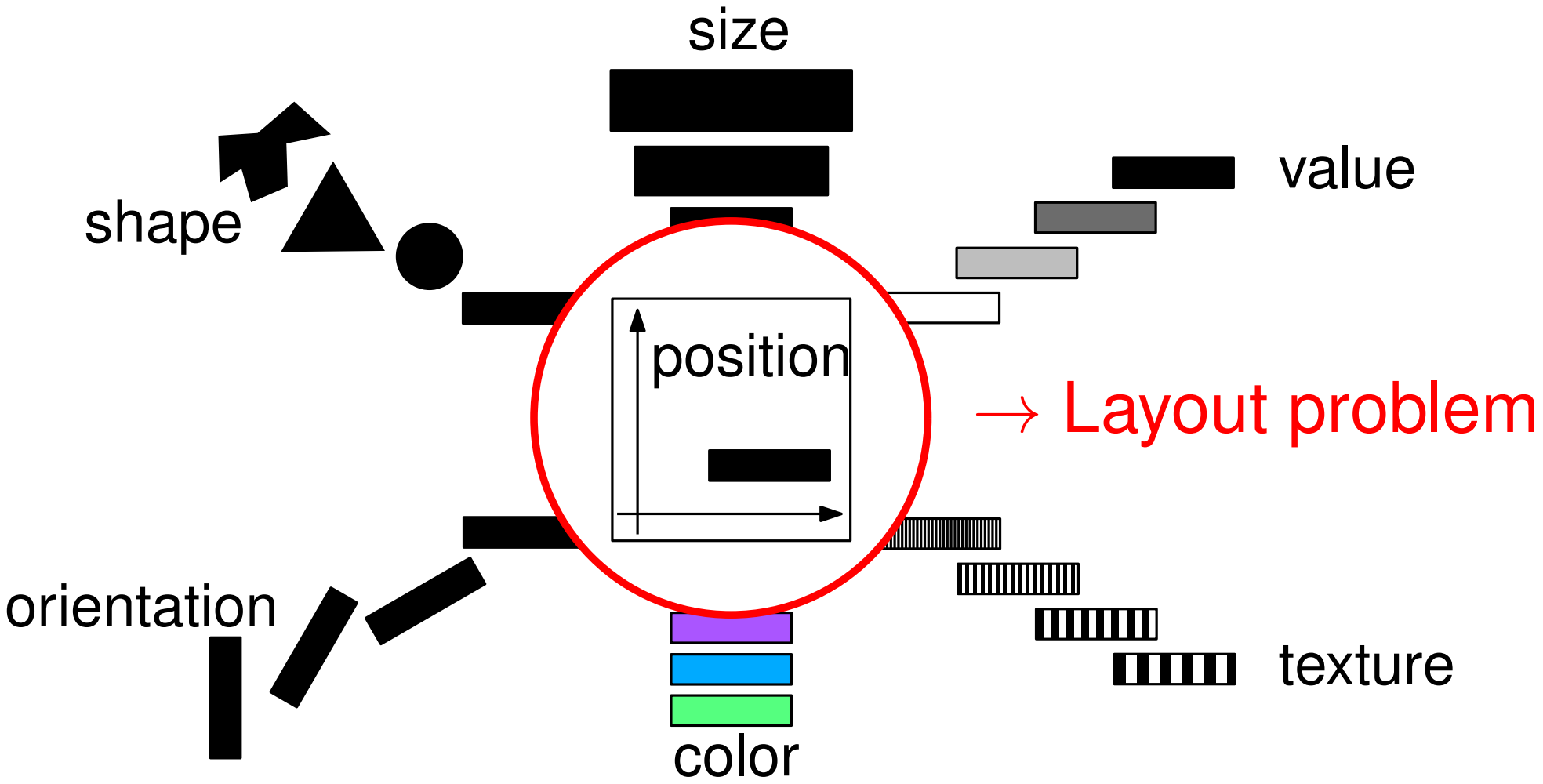
- complies with drawing conventions
- optimizes aesthetics
- satisfies local/partial constraints

$\rightarrow$ often lead to NP-hard optimization problems!
$\rightarrow$ often several competing criteria: small area and number of crossings

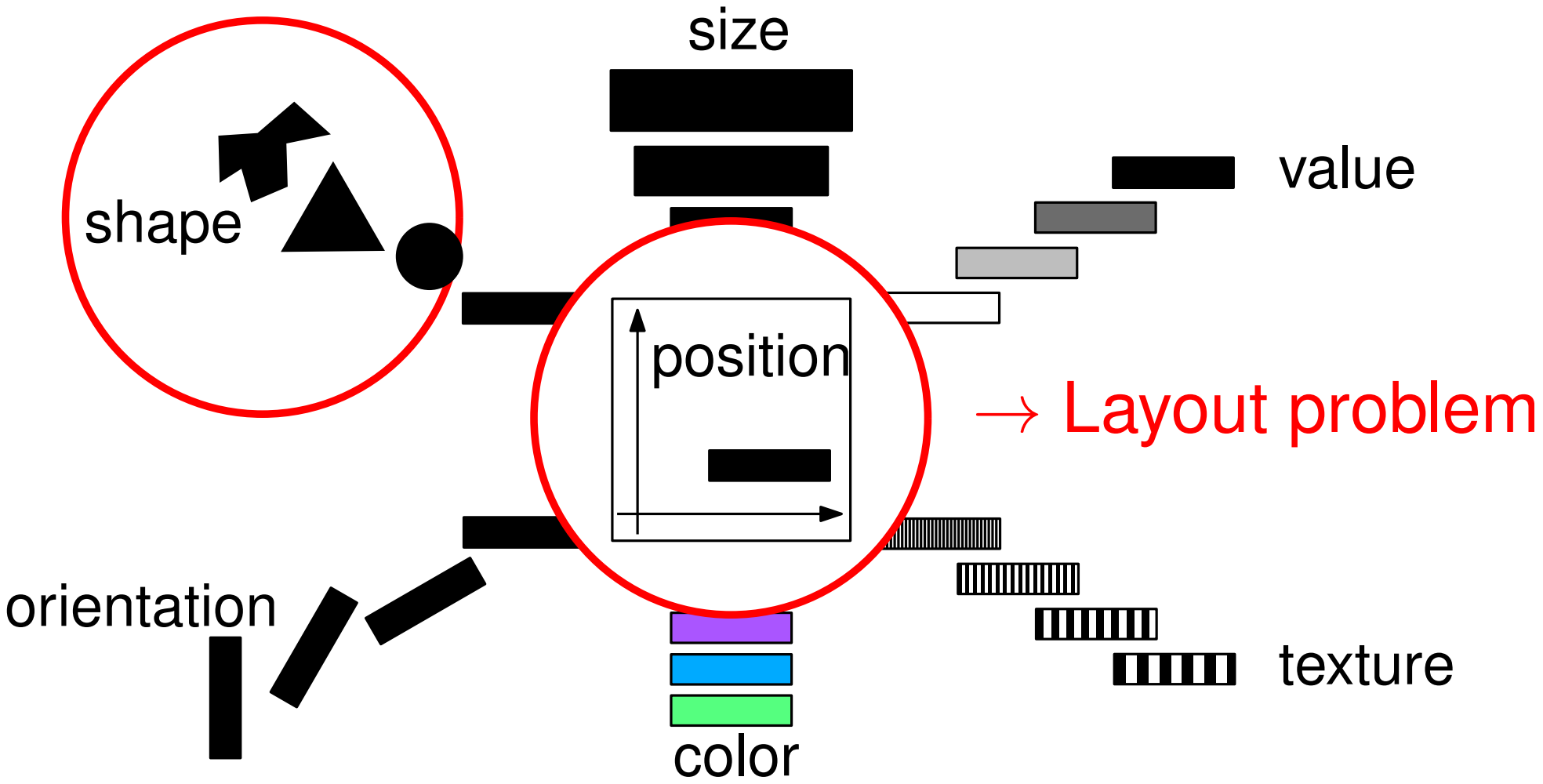
Visual Variables according to Bertin (1967)



Visual Variables according to Bertin (1967)

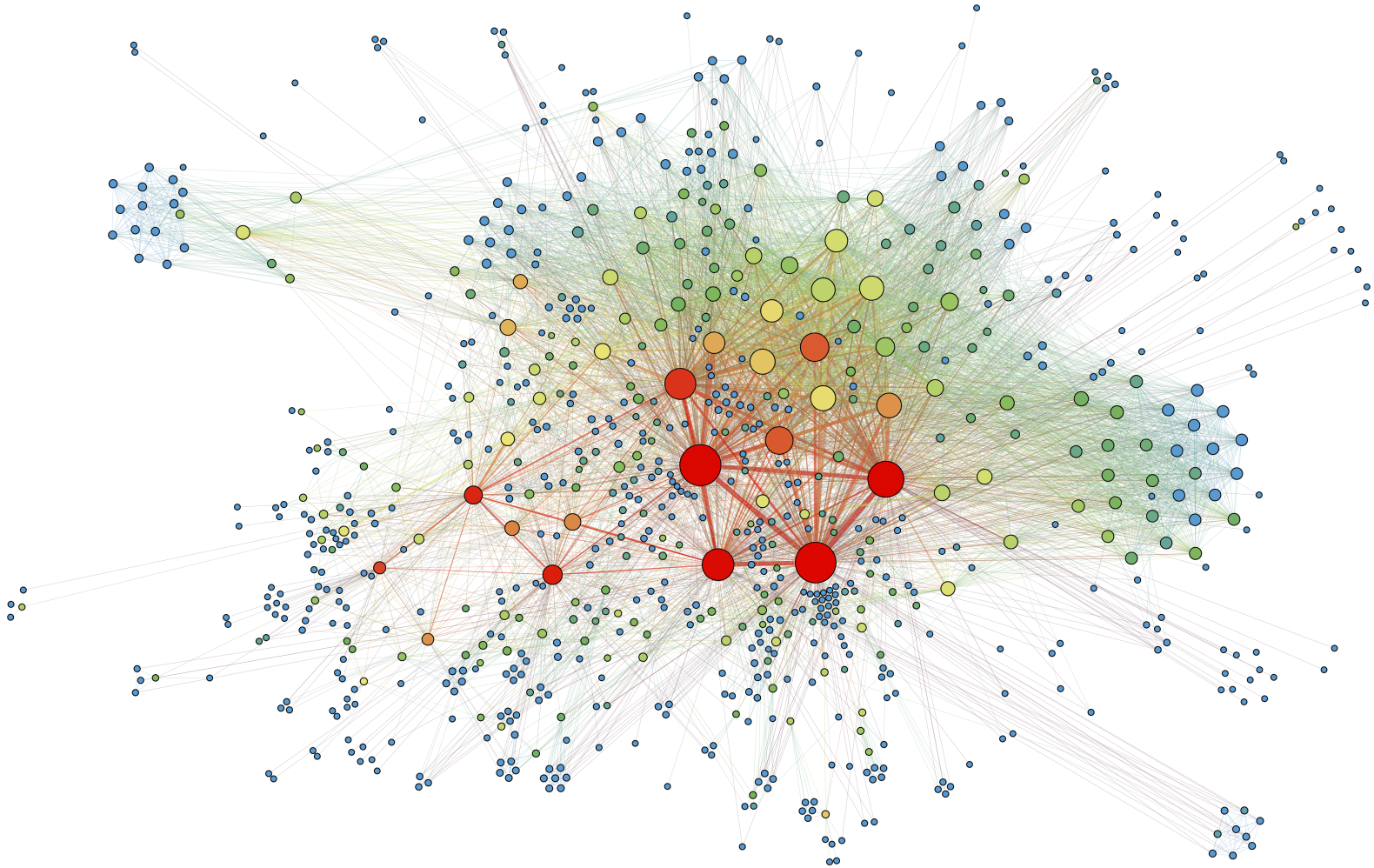


Visual Variables according to Bertin (1967)



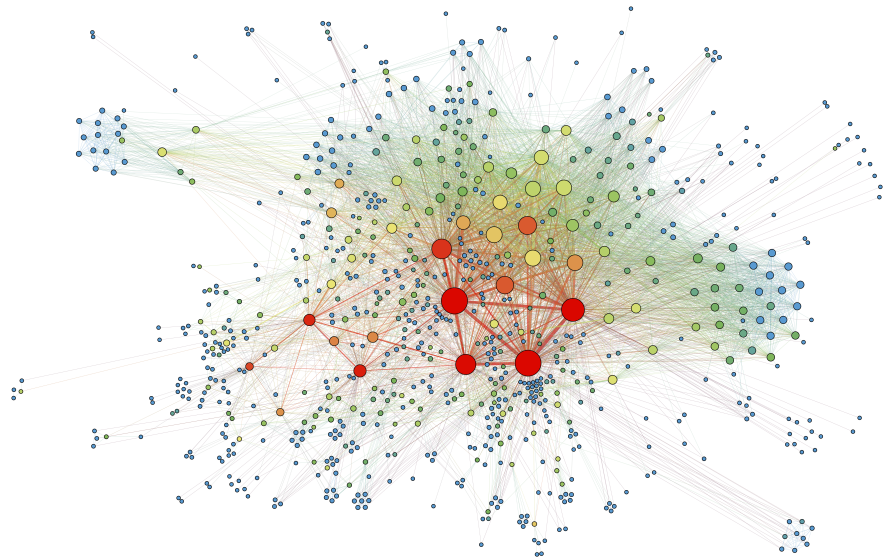
Evaluation of Network Visualization

Question: How can we evaluate quality of a network visualization?



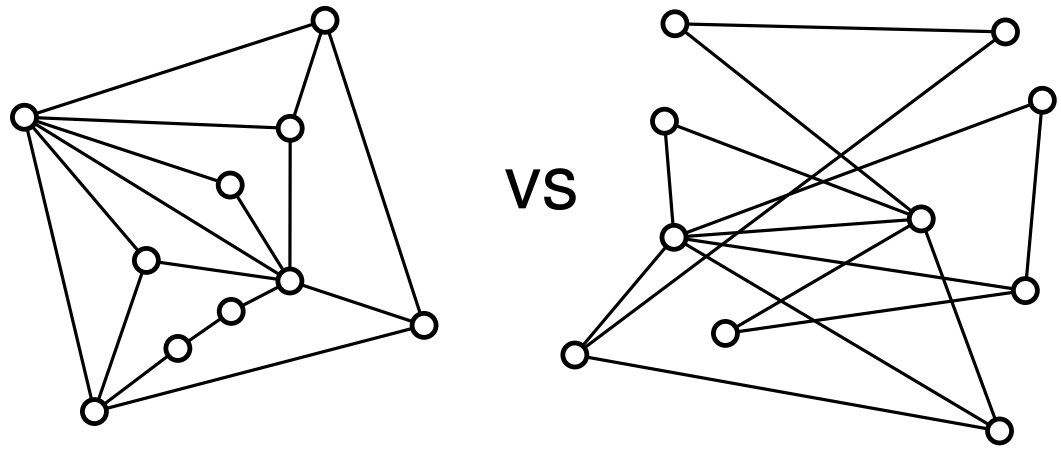
Evaluation of Network Visualization

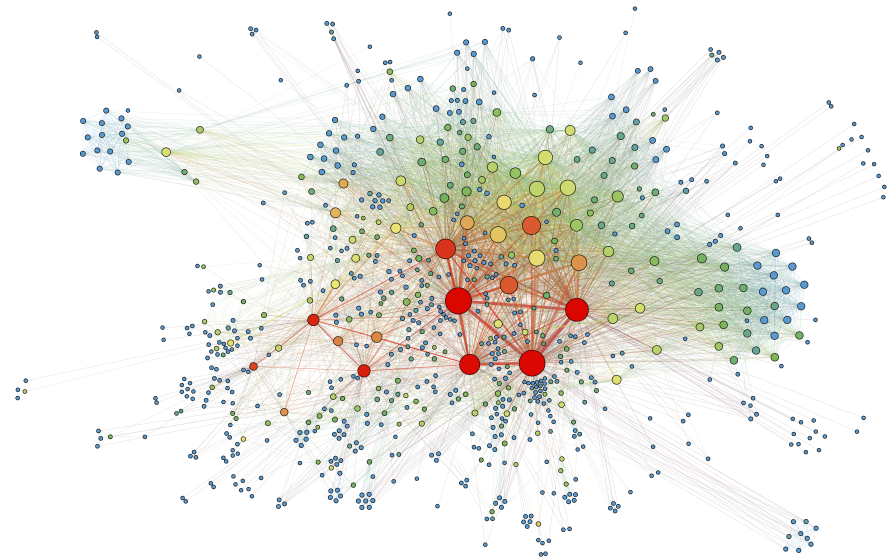
Automatic methods $->$ use quality metrics, quantitative evaluation



Evaluation of Network Visualization

Automatic methods $->$ use quality metrics, quantitative evaluation

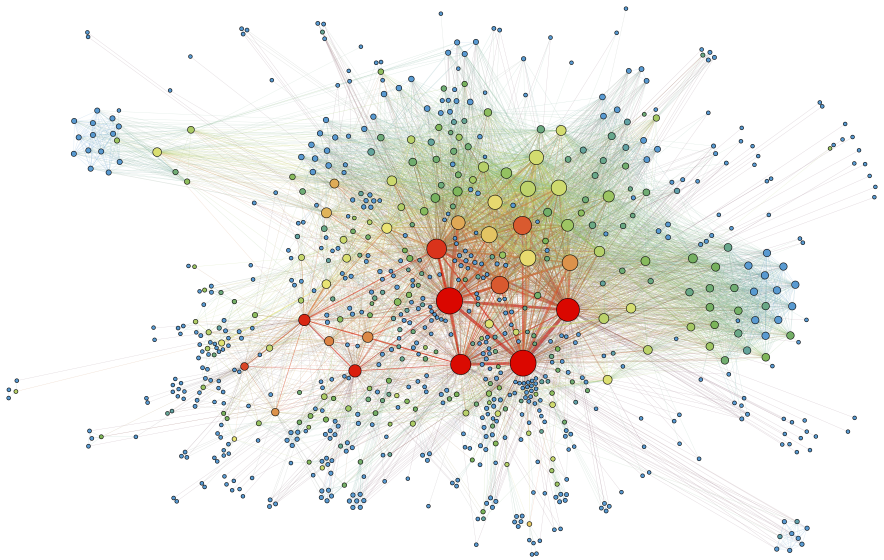




Evaluation of Network Visualization

Automatic methods $->$ use quality metrics, quantitative evaluation

User studies $->$ both qualitative & quantitative evaluation

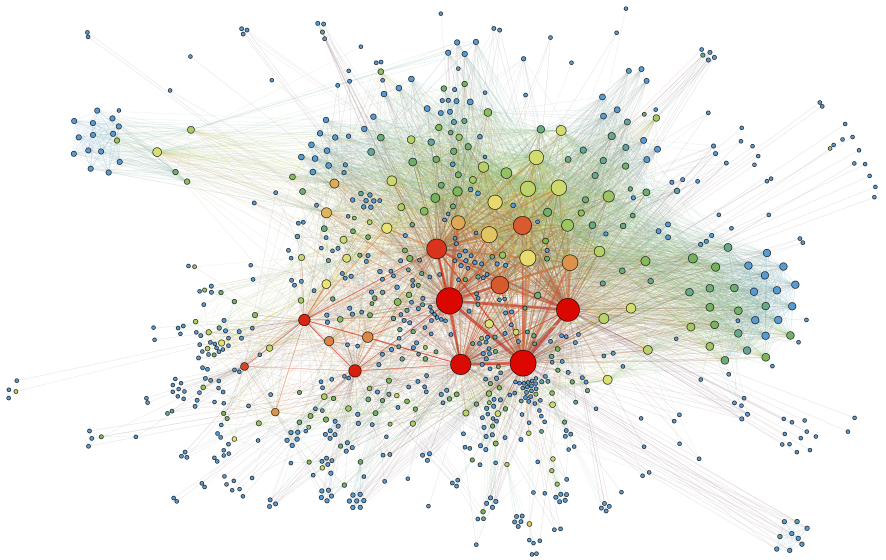


Evaluation of Network Visualization

Automatic methods $->$ use quality metrics, quantitative evaluation

User studies $->$ both qualitative & quantitative evaluation

Qualitative User Evaluation $->$ e.g. observe user interacting with a net.vis. and draw conclusion from this



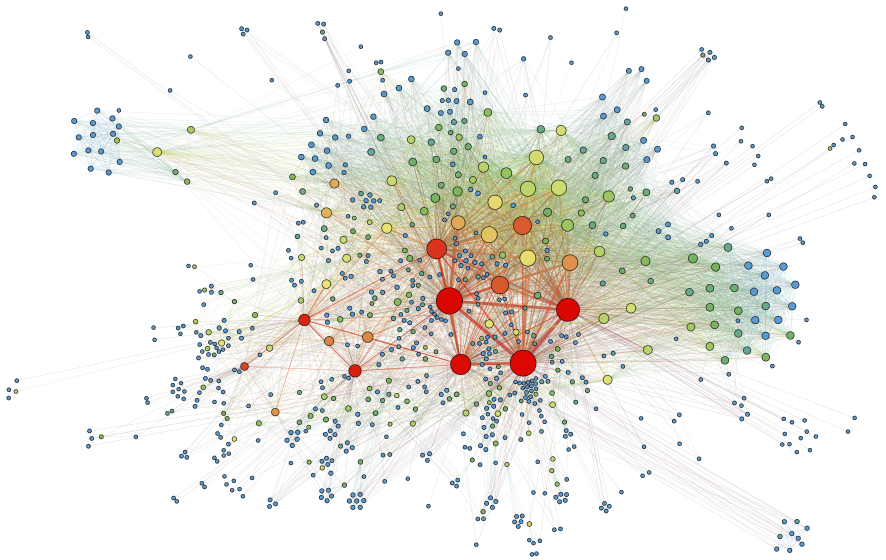
Evaluation of Network Visualization

Automatic methods $->$ use quality metrics, quantitative evaluation

User studies $->$ both qualitative & quantitative evaluation

Qualitative User Evaluation $->$ e.g. observe user interacting with a net.vis. and draw conclusion from this

Quantitative User Evaluation $->$ Tasks, subjective preference



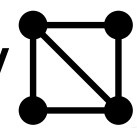
Evaluation of Network Visualization

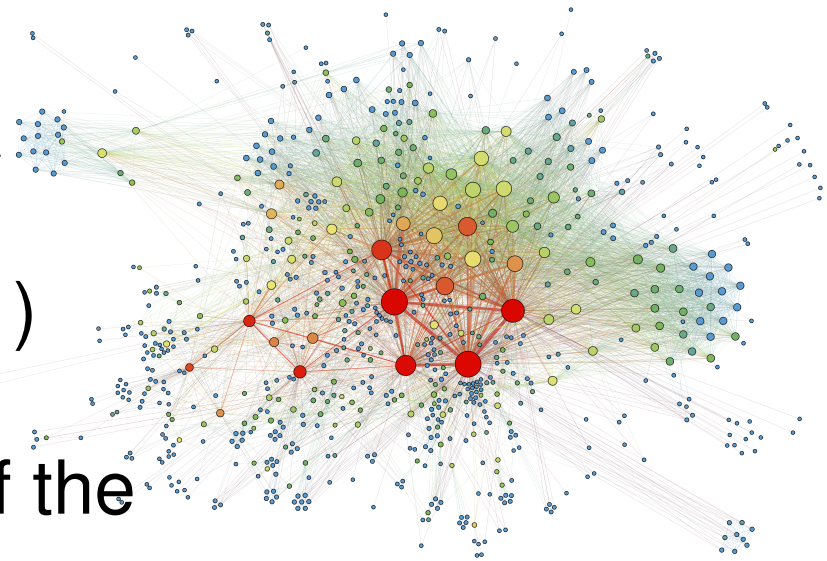
Automatic methods $->$ use quality metrics, quantitative evaluation

User studies $->$ both qualitative & quantitative evaluation

Qualitative User Evaluation $->$ e.g. observe user interacting with a net.vis. and draw conclusion from this

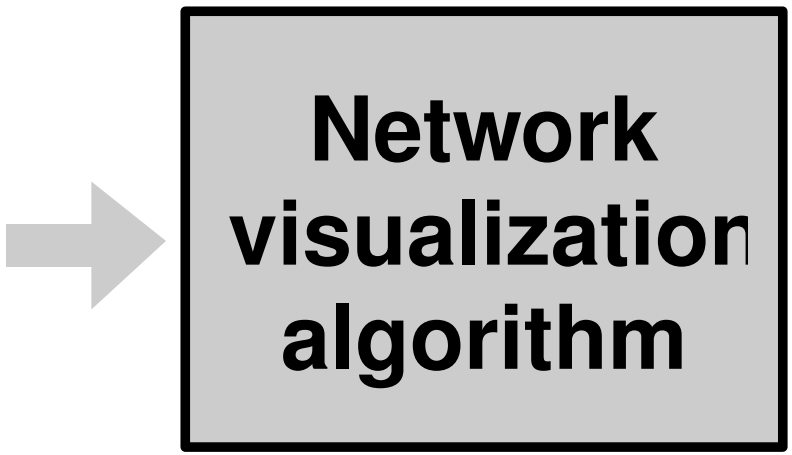
Quantitative User Evaluation $->$ Tasks, subjective preference

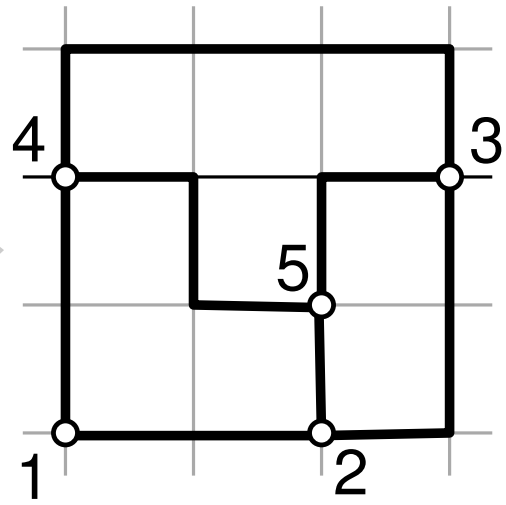
- What is the length of the shortest path between nodes A and B?
- What is the degree of a node A (number of incident edges)?
- Does certain structure (say ) appears as a subgraph?
- Subjective preference: which of the two layouts you prefer?



Types of Algorithms

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

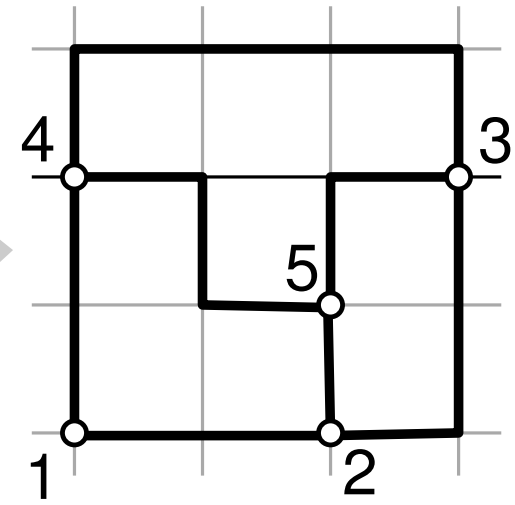




Types of Algorithms

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

Network visualization algorithm



Quality metrics, drawing conventions and constraints

Let's build the mental map: What types of algorithms have you heard about?

Computational Complexity

How can we investigate how efficient is an algorithm?

Computational Complexity

How can we investigate how efficient is an algorithm?

Common measures: the quality of the layout and the running time

Computational Complexity

How can we investigate how efficient is an algorithm?

Common measures: the quality of the layout and the running time

Measure time running the algorithms for multiple inputs, compute e.g. average

Computational Complexity

How can we investigate how efficient is an algorithm?

Common measures: the quality of the layout and the running time

Measure time running the algorithms for multiple inputs, compute e.g. average

Or/and compute the worst-case computational complexity of the algorithm

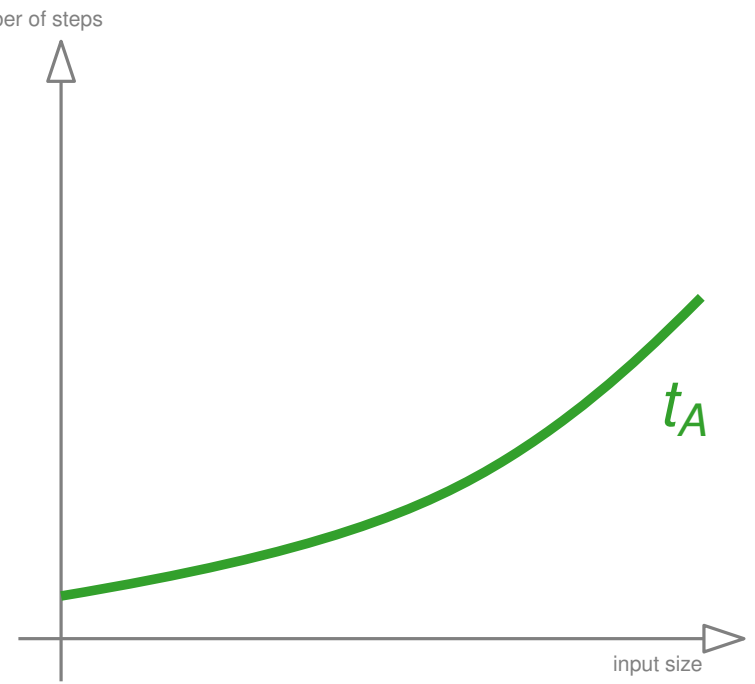
Computational Complexity

How can we investigate how efficient is an algorithm?

Common measures: the quality of the layout and the running time

Measure time running the algorithms for multiple inputs, compute e.g. average

Or/and compute the worst-case computational complexity of the algorithm



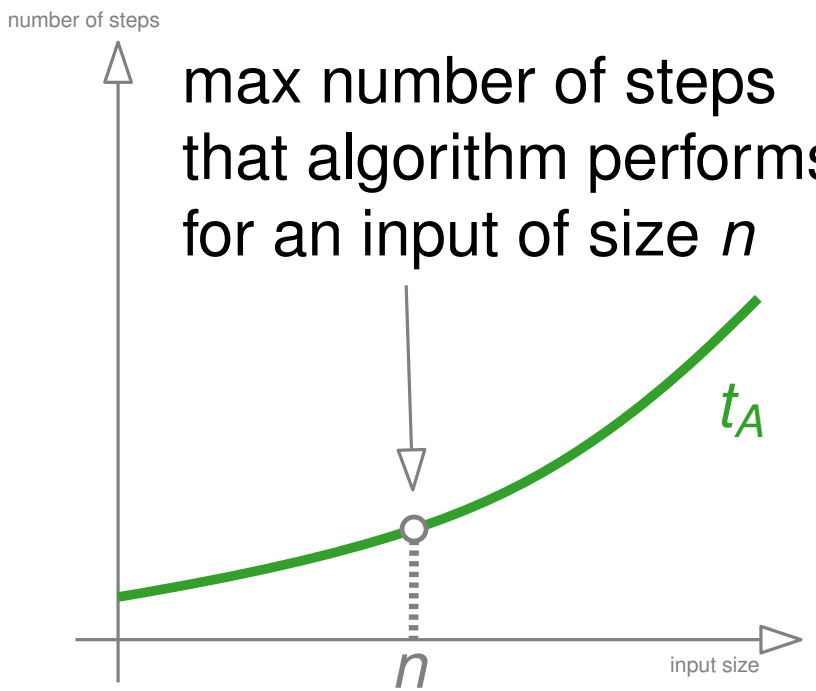
Computational Complexity

How can we investigate how efficient is an algorithm?

Common measures: the quality of the layout and the running time

Measure time running the algorithms for multiple inputs, compute e.g. average

Or/and compute the worst-case computational complexity of the algorithm



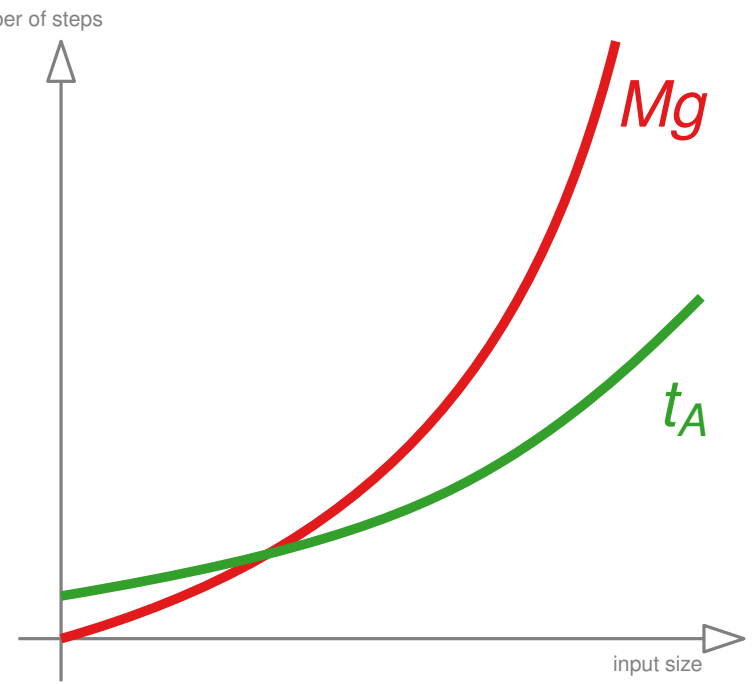
Computational Complexity

How can we investigate how efficient is an algorithm?

Common measures: the quality of the layout and the running time

Measure time running the algorithms for multiple inputs, compute e.g. average

Or/and compute the worst-case computational complexity of the algorithm



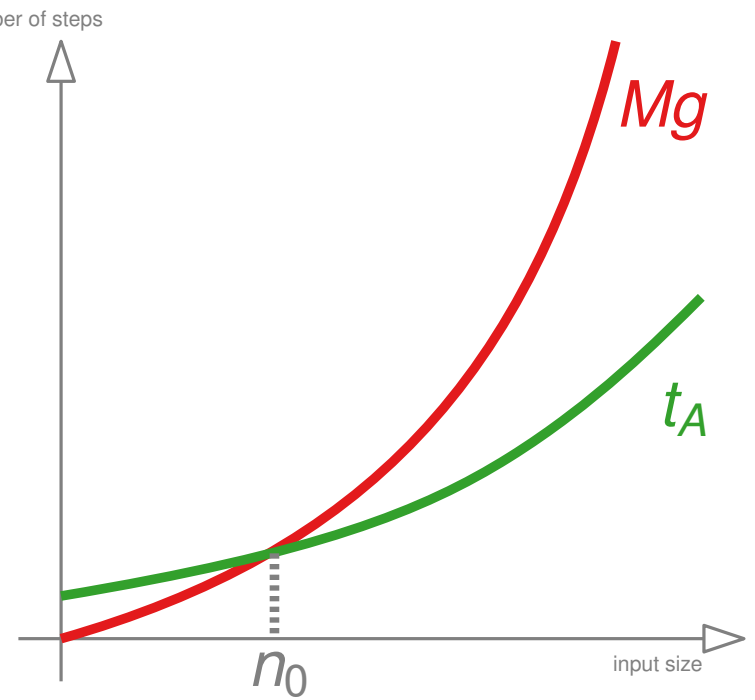
Computational Complexity

How can we investigate how efficient is an algorithm?

Common measures: the quality of the layout and the running time

Measure time running the algorithms for multiple inputs, compute e.g. average

Or/and compute the worst-case computational complexity of the algorithm



Computational Complexity

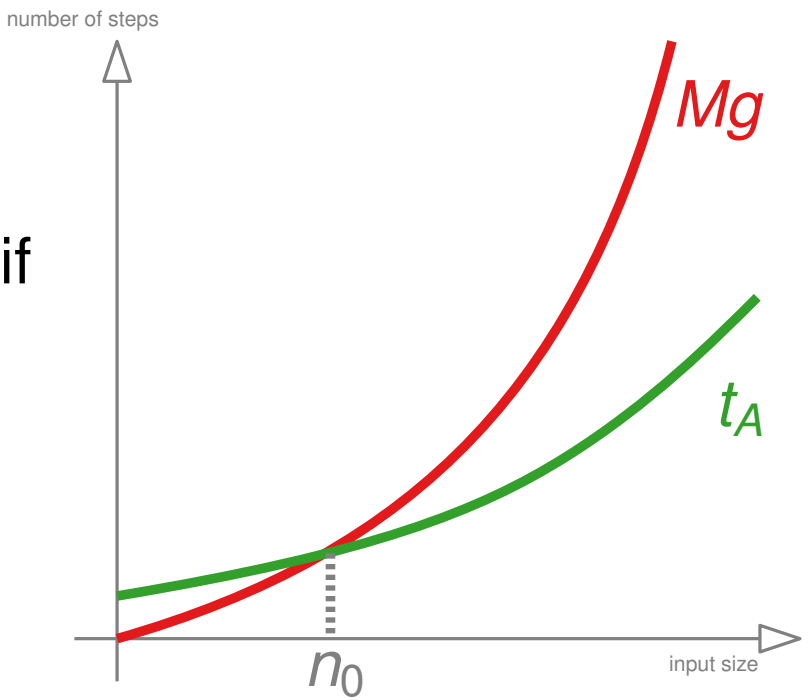
How can we investigate how efficient is an algorithm?

Common measures: the quality of the layout and the running time

Measure time running the algorithms for multiple inputs, compute e.g. average

Or/and compute the worst-case computational complexity of the algorithm

We say that the worst-case computational complexity (or shortly **running time**) of an algorithm A is $g(n)$ if the number of steps that A performs for an input of size n (denote by $t_A(n)$) is upper bounded by $Mg(n)$, for some $n > n_0$ and constant M: $\exists M, \forall n > n_0$ $t_A(n) < Mg(n)$



Lecture Overview

- Why Data visualization?
- Data types and their models
- Networks and their visualizations (gallery)
- Let’s draw some networks
- Basis for algorithm design: formalization of network visualization problem
- Evaluation of Network Visualization
- Types of algorithms (mind map)
- Some basic notions in Algorithm Complexity Theory

Reading and Next

Recall

Types of algorithms

How to compute running time of algorithms

Reading and Next

Recall

Types of algorithms

How to compute running time of algorithms

Next

Tree drawing algorithms:

Layered layout, radial layout and bubble layout