HUST

ĐẠI HỌC BÁCH KHOA HÀ NỘI

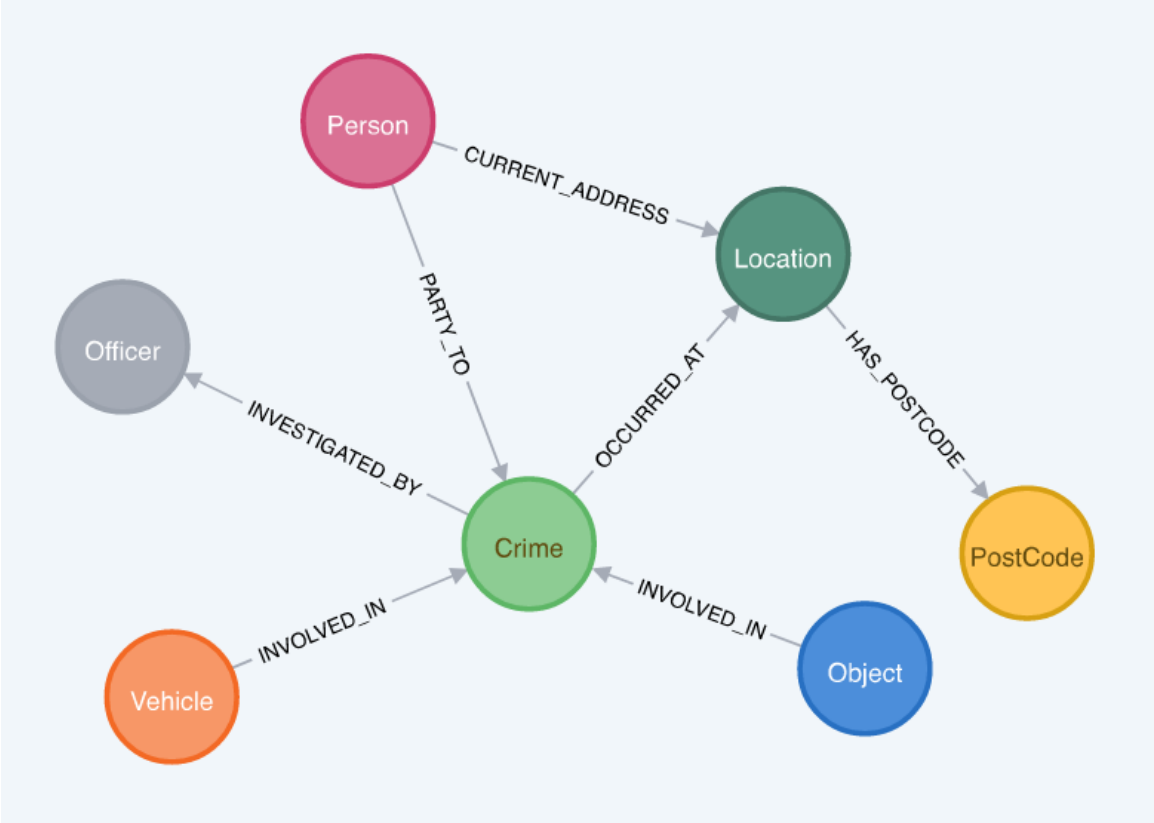
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.

Graph data visualization

What is graph data

- Data representing relationships between entities
  - Entities are represented by nodes (vertices or points)
  - Relationships are represented by edges (lines or arcs)
  - The graph G(V.E) includes the set of vertices V and the set of edges E

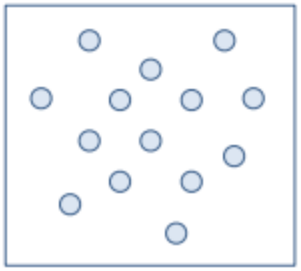


Properties of Graph Data

- Sparsity
  - Not all nodes are connected by edges
  - Number of edges is often much less than the number of possible connections between nodes
- Directionality
  - **Directed edges:** Show a one-way relationship between nodes (e.g., following on Twitter)
  - **Undirected edges:** Show a two-way relationship between nodes (e.g., friendship on Facebook)
- Weights
  - Weights represent the strength or importance of the relationship between nodes
  - Can be numerical values, distances, times, or other quantities

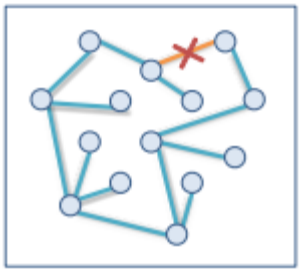
Basic concepts

- **An independent set, stable set, coclique or anticlique** is a set of vertices in a graph, no two of which are adjacent.
- **Clique** is a subset of vertices of an undirected graph such that every two distinct vertices in the clique are adjacent.
- **A graph** is said to be **connected** if every pair of vertices in the graph is connected.
- **A tree** is an undirected graph in which any two vertices are connected by exactly one path, or equivalently a connected acyclic undirected graph.

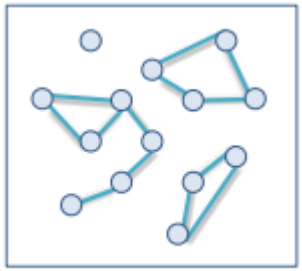


Independent Set

Clique



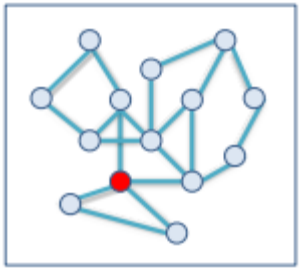
Tree



Unconnected Graph

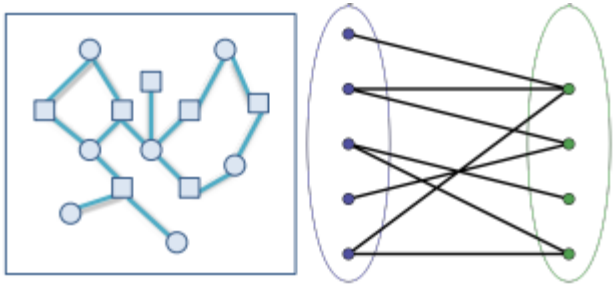
Basic concepts

- **An articulation point (or cut vertex) (Điểm khớp)** is defined as a vertex which, when removed along with associated edges, makes the graph disconnected
- **A biconnected graph (đồ thị nối đôi)** is a connected and "nonseparable" graph, meaning that if any one vertex were to be removed, the graph will remain connected.
  - A biconnected graph has no articulation vertices.
- **A bipartite graph (or bigraph) (đồ thị hai phần)** is a graph whose vertices can be divided into two disjoint and independent sets U and V such that every edge connects a vertex in U to one in V



Articulation Point (red)

Biconnected Graph

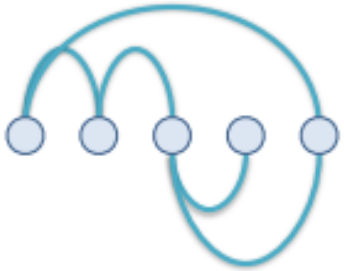


Bipartite Graph

Basic concepts

- Degree (bậc) of a vectex
  - deg(n) is the number of edges that are incident to the vertex
- Graph diameter (đường kính)
  - diam(G) is the greatest distance (shortest parh) between any pair of vertices

Graph visualization

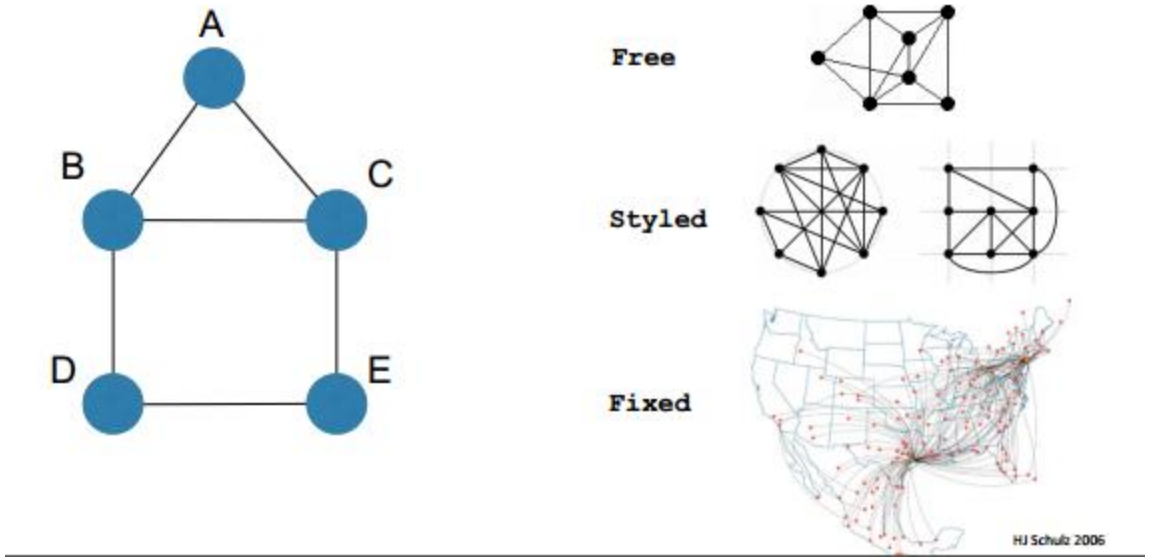


Explicit
(Node-Link)

Matrix

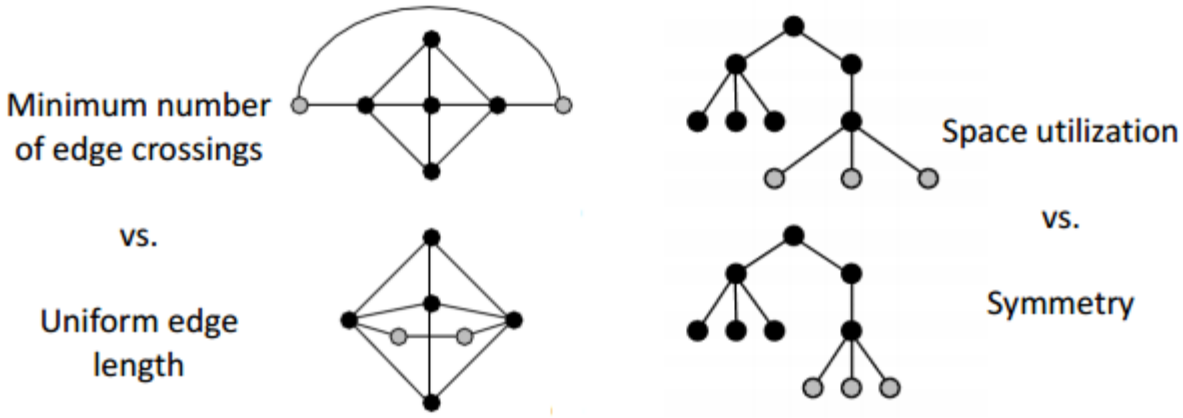
Implicit

Explicit graph visualization (node-link)



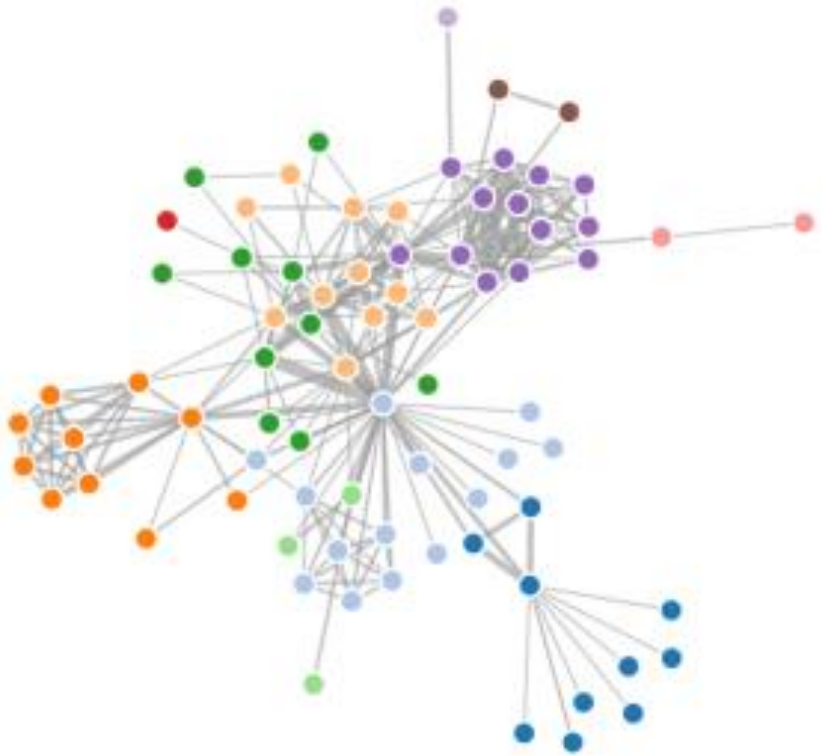
Criteria for node-link representation

- Minimizing the intersection edges
- Minimize the distance between the vertices
- Minimize drawing area
- Edges of similar length
- Maximum angle between different edges
- Symmetry (graphs with the same structure must look the same)



Force-directed graph

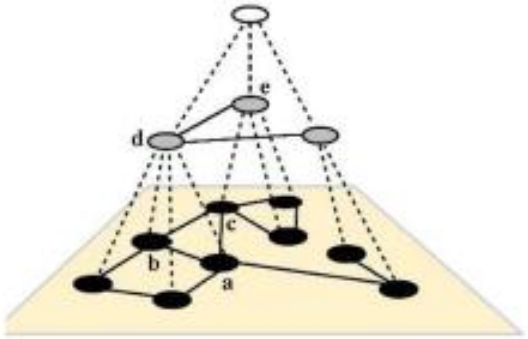
- The forces are applied to the nodes, pulling them closer together or pushing them further apart.
- This is repeated iteratively until the system comes to a mechanical equilibrium state

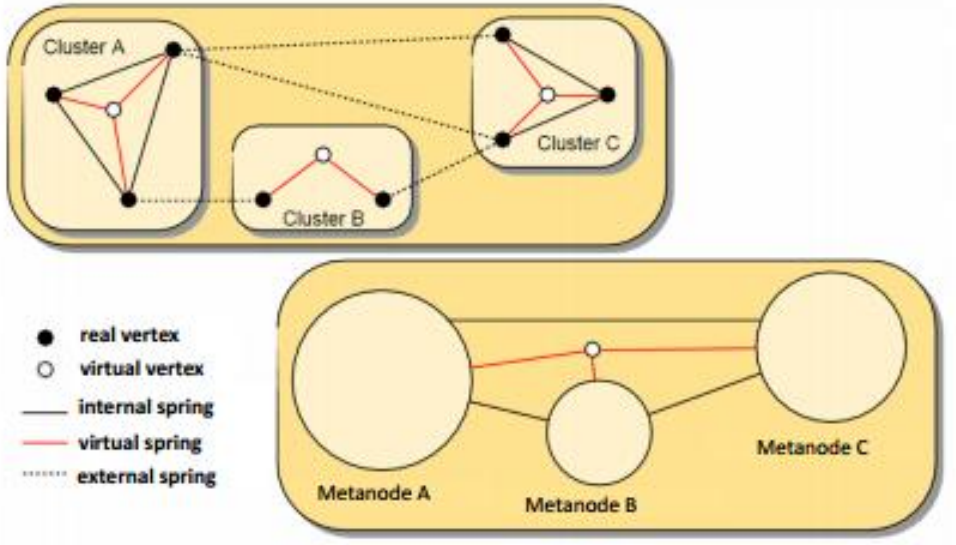


https://observablehq.com/@d3/force-directed-graph

Multi-level technique

- Reduce the magnitude of a graph by merging vertices together, compute a partition on this reduced graph, and finally project this partition on the original graph.

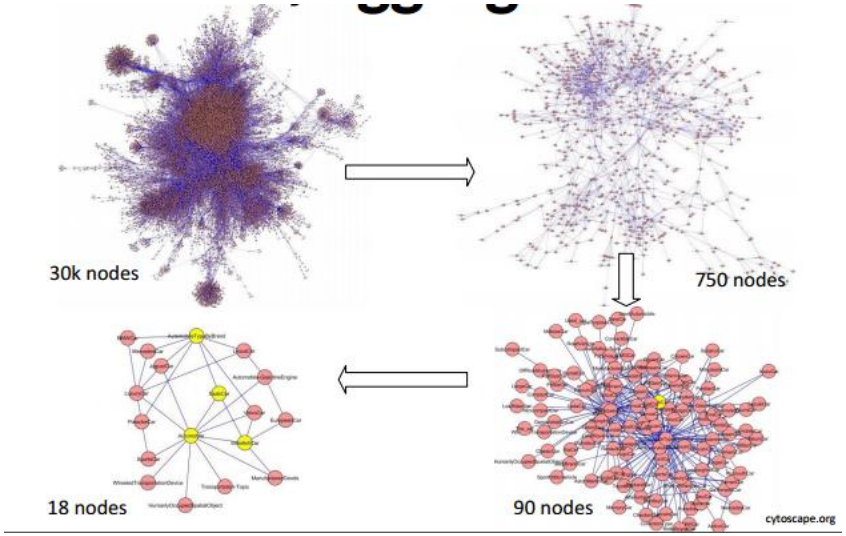




[Schulz 2004]

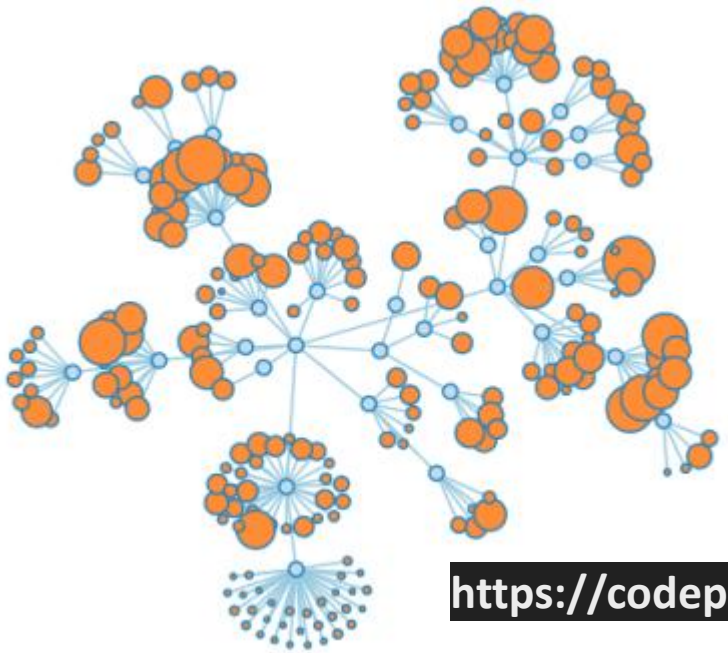
Sampling

- Large graphs can overwhelm visualization tools and viewers.
- Sampling allows us to explore a manageable subset of the data.
- Different sampling techniques highlight various aspects of the graph.
- Key sampling techniques
  - Random Sampling
  - Degree-Based Sampling
  - Ego-Network Sampling
  - Community Sampling



Collapse/Expand

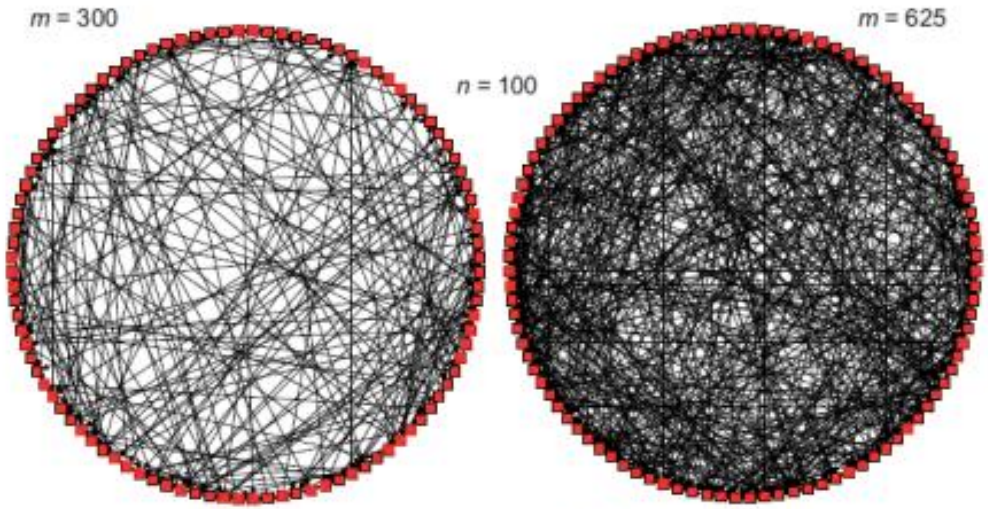
- Particularly useful for hierarchical graphs, where sub-trees can be collapsed or expanded as needed.
- **Clarity:** Reduces visual clutter, making it easier to focus on specific areas of interest.
- **Scalability:** Handles large datasets efficiently by allowing parts of the graph to be hidden.



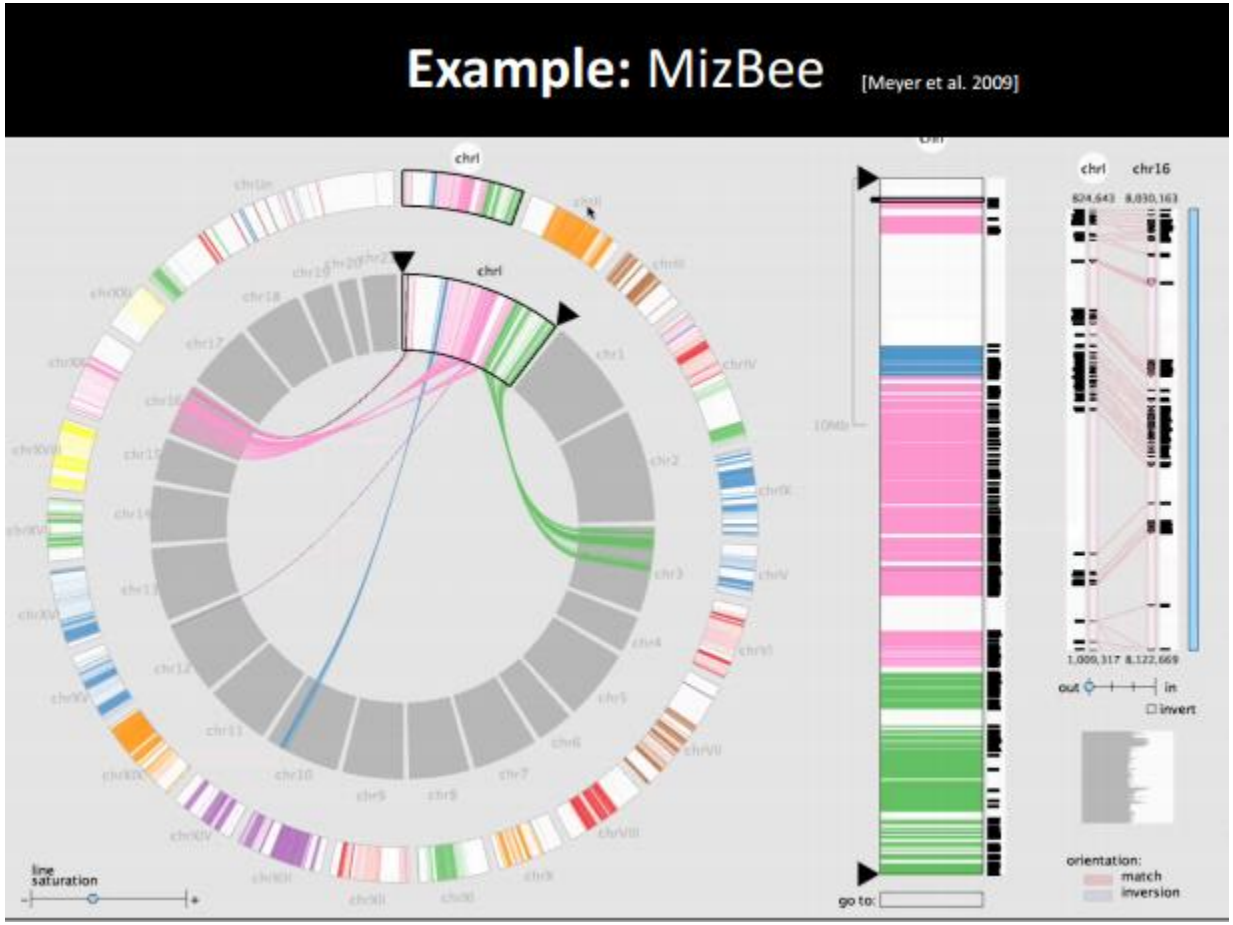
https://codepen.io/gyunee/pen/oNvEoOb

Fixed arrangement for graphs

- Circos Layout
  - **Compact Visualization:** Efficiently uses space to show large, complex datasets.
  - **Pattern Recognition:** Facilitates the identification of clusters, patterns, and relationships.
  - **Aesthetic Appeal:** Attractive and engaging, enhancing the visual storytelling of data.

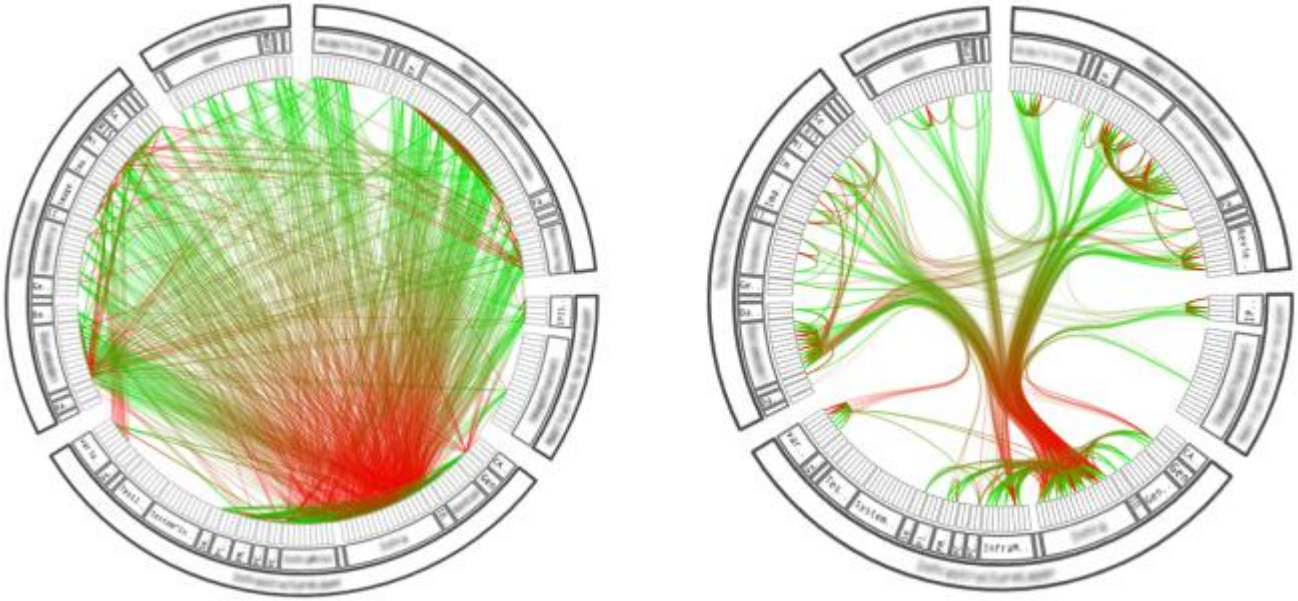


Fixed arrangement for graphs



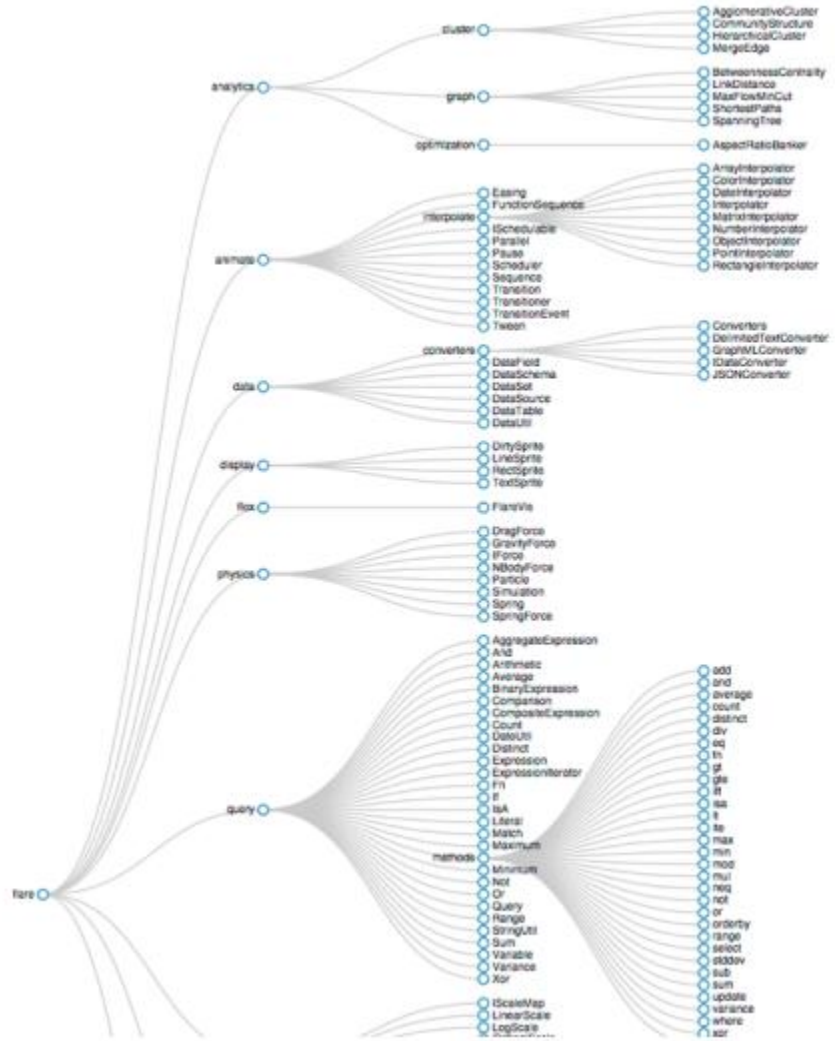
Fixed arrangement for graphs

- Edge bundling



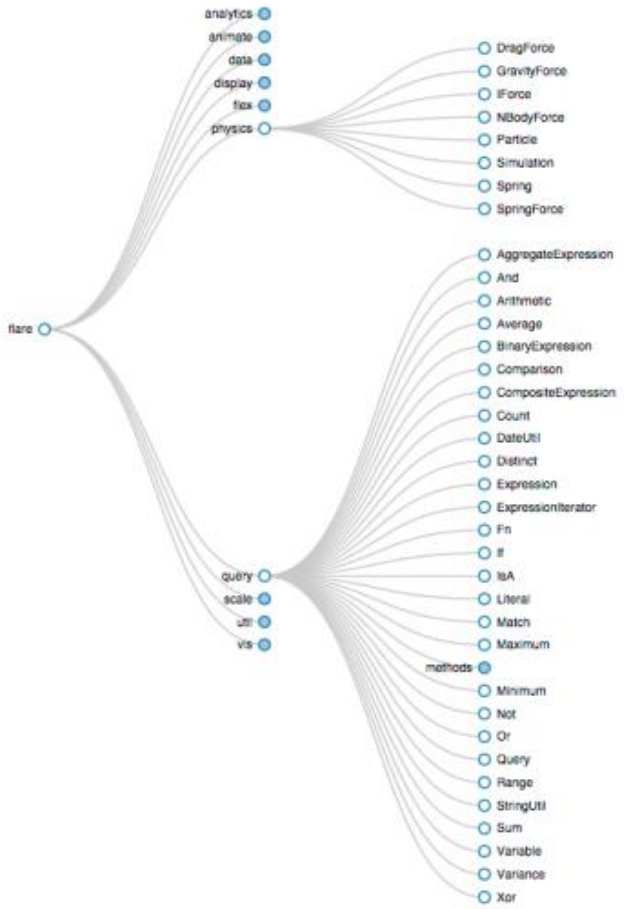
https://observablehq.com/@d3/hierarchical-edge-bundling

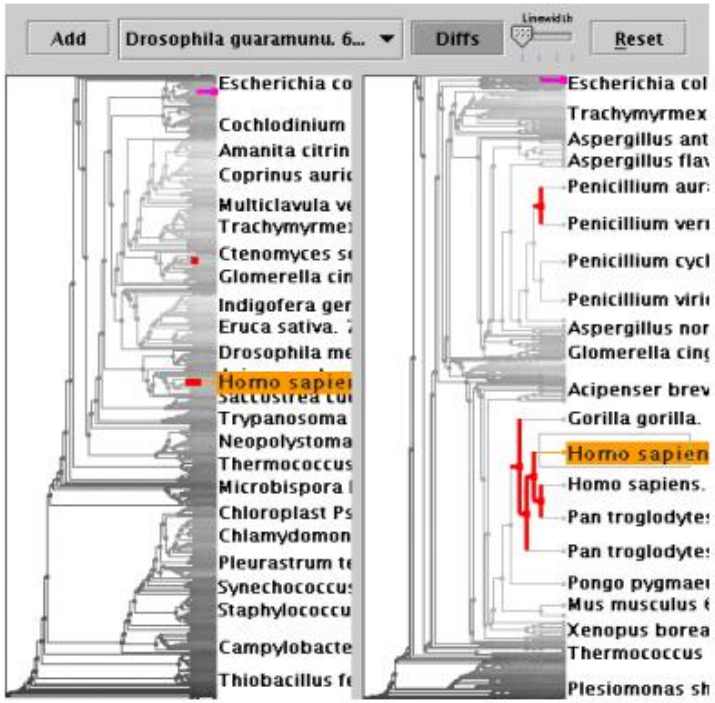
Tree layout



https://observablehq.com/@d3/tree

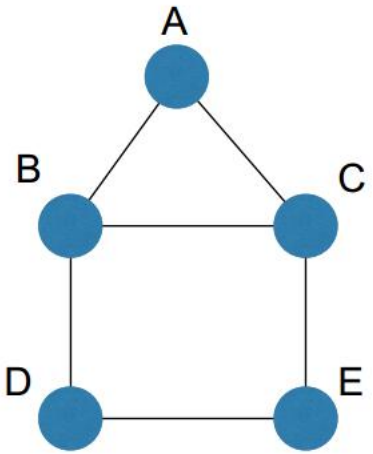
Tree layout





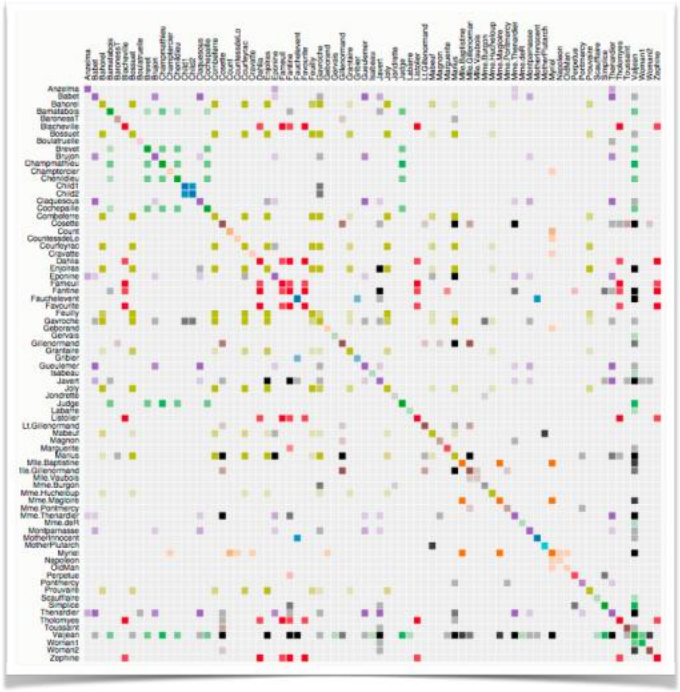
http://bl.ocks.org/mbostock/4339083

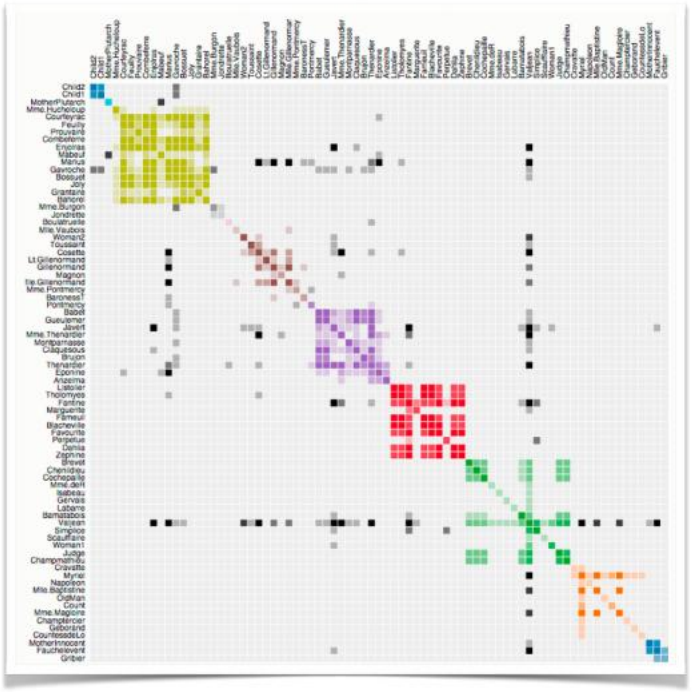
Adjacency matrix



	A	B	C	D	E
A		■	■		
B	■		■	■	
C	■	■			■
D		■			■
E			■	■	

Adjacency matrix

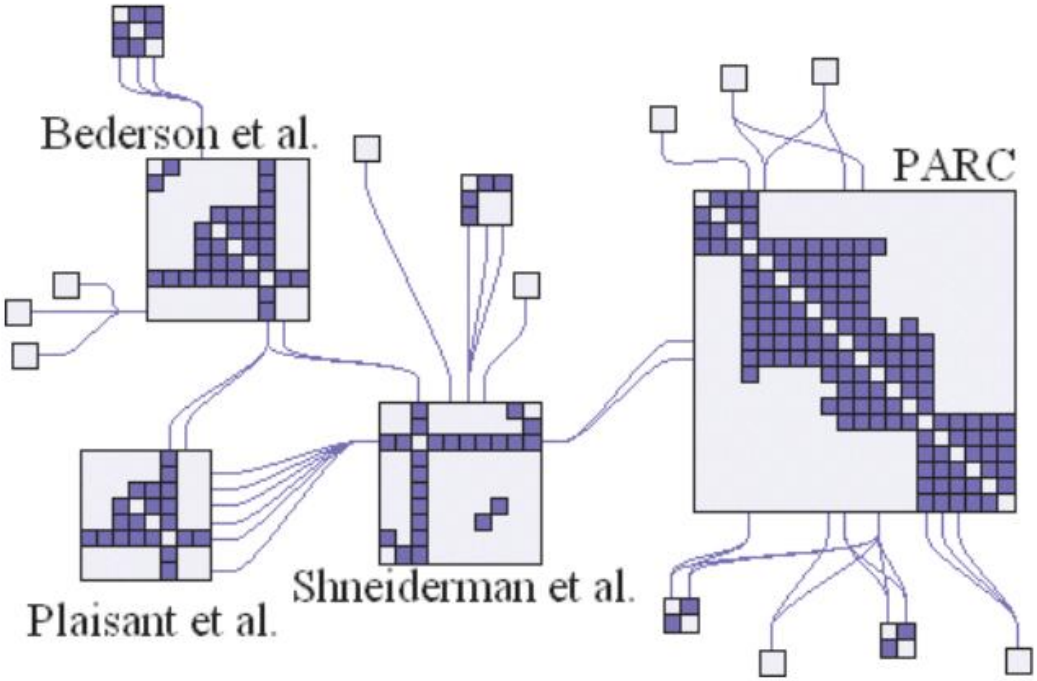




Adjacency matrix

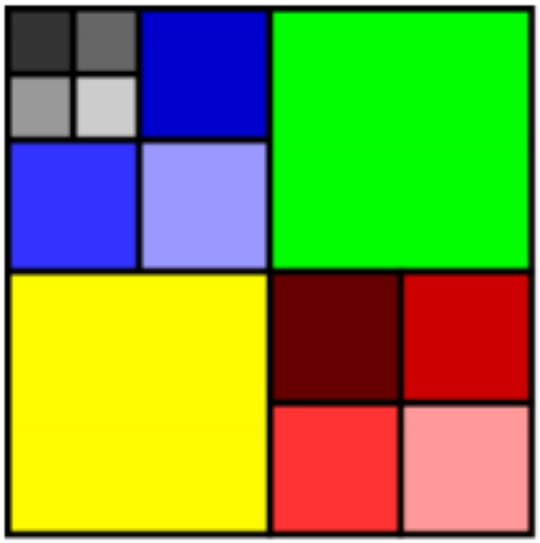
- Advantage
  - Can represent most graphs (except hypergraphs)
  - Focus on edges instead of vertices
  - No need to care about layout
- Disadvantage
  - Difficult to detect relationships such as paths, cycles, etc.

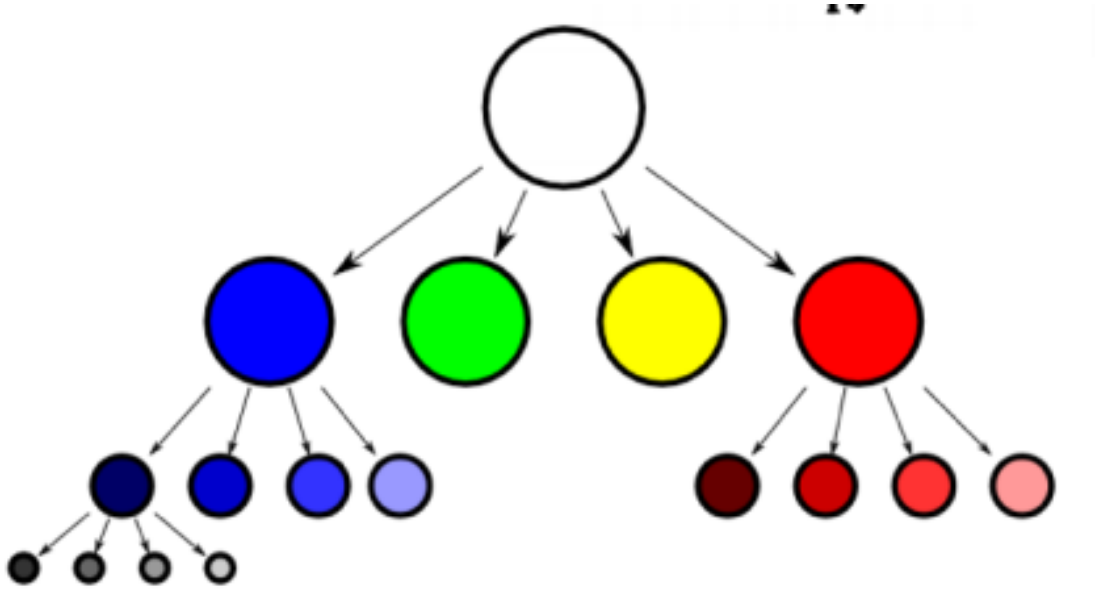
Hybrid layout



Implicit presentation

- A treemap chart is a visualization that displays hierarchical data using nested rectangles.
- Structure
  - The entire chart area represents the root of the hierarchy.
  - Each rectangle (or cell) represents a branch or leaf node.
  - Rectangles are sized and ordered by a quantitative variable.
  - Colors can be used to represent different categories or to encode additional data dimensions.



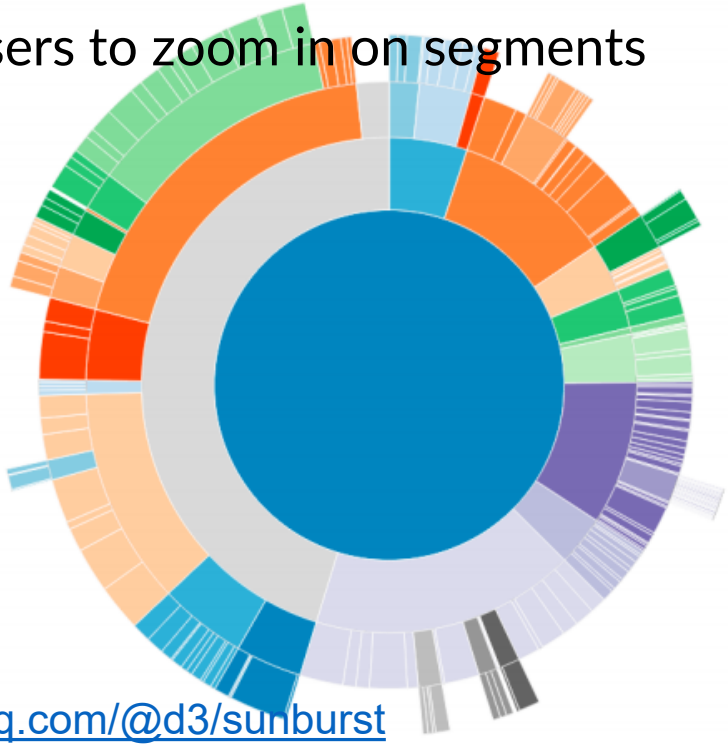


http://www.nytimes.com/packages/html/newsgraphics/2011/0119-budget/


TECHNOLOGY
SOFTWARE - INFRASTRUCTURE
MSFT +0.88%
ORCL +0.66%
ADBE +0.20%
PANW +1.64%
SNPS +0.33%
FTNT +1.64
FLT
SEMICONDUCTORS
NVDA +1.95%
AVGO +2.39%
AMD +0.43%
QCOM +0.02%
TXN -0.41%
INTC +1.30%
ADI +0.27%
NXPI +1.82
MU +1.78%
ON
MPWR
CONSUMER ELECTRONICS
AAPL +0.74%
SOFTWARE - APPLICATION
CRM +0.79%
NOW +1.66%
CDNS +0.18%
ROP +0.34
INTU +0.42%
ADSK +0.08%
FICO
INFORMATION TEC
ACN +0.64%
IBM +1.09%
FI +0.71
CTSH -0.13
FIS +0.03
IT +1.34
BR
SEMICONDU
AMAT -0.45%
LRCX -0.69
KLAC
COMMUNIC
CSCO +0.25%
MSI +0.36%
COMPUT
ANET
STX
ELECT
APH
TEL +0.72
SCIENTIFIC
TDY
FSLR
FINANCIAL
CREDIT SERVICES
V -0.03%
MA -0.24%
AXP -0.11%
PYPL +0.79%
COF +2.35
DFS
SYF
BANKS - DIVERSIFIED
JPM +1.10%
BAC +0.98%
WFC +1.54%
C +0.80%
INSURANCE - DIVERSIFIED
BRK-B +0.11%
FINANCIAL DATA
SPGI -0.48%
CME +0.42%
MCO +0.53%
MSCI
ICE
FDS
CAPITAL MAR
MS +2.40%
GS +1.80%
SCHW +1.76%
RJF
INSURANCE -
PGR -0.33%
CB -0.10%
TRV -0.36%
HIG
ALL
L
ASSET MANAGEM
BX -1.13%
BK
AMP
STT
BLK -0.10%
PFG
INSURAN
MMC
AON +0.75
AJG -0.04
WTW
BRO
BANKS - REGION
USB
TFC
PNC +0.64%
RF
MTB
INSURANCE - L
AFL +0.67
MET +0.99
PRU +1.21
COMMUNICATION SERVICES
INTERNET CONTENT & INFORMATION
GOOG -1.31%
META +1.89%
TELECOM SERVICES
TMUS +0.03%
VZ -1.03%
CMCSA -0.17%
T -0.99%
ENTERTAINMENT
NFLX +0.39%
DIS +0.51%
EA
LYV
IPG
CONSUMER CYCLICAL
INTERNET RETAIL
AMZN +0.37%
AUTO MANUFACT
TSLA +0.49%
F
GM
RESTAURANTS
MCD -0.44%
SBUX +0.18%
CMG +0.75
YUM +0.06
DRI
TRAVEL SERVICES
BKNG +1.13%
ABNB +0.60
CCL
EXPE
RCL
APPAREL RET
TJX +0.27%
LULU +5.37%
ROST +0.32%
SPECIALTY R
ORLY -0.82%
AZO
RESIDEN
DHI
LEN
LODGI
MAR +0.66
HOME IMPROVEMEN
HD +0.09%
LOW -0.23
FOOTWEAR
NKE +0.96%
PACKAGI
AUTO
RESORTS
LVS
CONSUMER DEFENSIVE
DISCOUNT STORES
WMT -1.05%
COST -0.03%
TGT
DG
BEVERAGES - NO
KO -0.22%
PEP -1.06%
KDP -0.77
HOUSEHOLD &
PG -0.92%
CL -0.58%
KMB -1.28
EL
CHD
CLX
TOBACCO
PM +0.21%
MO -0.05
CONFECT
MDLZ -0.73%
STZ -0.60
BF-B +1.03
PACKAGED FOOD
K
KHC
GIS
FARM
SYY
BG
KR
HEALTHCARE
DRUG MANUFACTURERS - GENERAL
LLY +1.66%
JNJ -0.63%
MRK -0.13%
ABBV +0.89%
PFE +0.52%
AMGN -0.81%
BMY +0.30%
GILD +1.24%
BIIB
HEALTHCARE PLAN
UNH +0.27%
ELV +0.21%
CI +0.33%
HUM
CVS +1.47%
CNC -0.63%
MEDICAL DEVICES
ABT +0.44%
SYK -0.10%
MDT -0.25%
EW
ZBH
STE
BSX -0.14%
DIAGNOSTICS &
TMO -0.88%
DHR -0.65%
IDXX +1.58
A -1.15
LH
IQV +0.14
MTD -0.86
MEDICAL INST
ISRG -0.37%
BDX -0.53
RMD
BAX
MEDICA
MCK +0.49
CAH
BIOTECHN
VRTX
REGN -1.06%
ZTS +1.52%
MEDICA
HCA
GEHC
INDUSTRIALS
SPECIALTY INDUSTRI
GE +0.93%
ETN +1.92%
ITW +0.82
EMR +0.94%
AME
CMI -0.07
IR
XYL
PH +1.12
ROK +0.23
DOV
IEX
AEROSPACE & DEFE
BA +3.10%
RTX -0.22%
LMT +0.20%
GD +0.32%
TDG +1.13
NOC -0.26%
LHX
FARM & HEAVY
CAT +0.81%
DE +0.68%
PCAR
CONGLOME
HON -1.61%
STAFFING
ADP -0.31%
RAILROADS
UNP -0.74%
CSX -1.61%
NSC -2.81
BUILDING P
CARR +4.50%
TT
JCI
SPECIALTY
CPRT -0.61%
GPN
WASTE
WM -0.20
RSG
AIRLI
INTEGRATED FR
UPS -0.13%
FDX +1.71%
INDUST
FAST -0.11
ODFL -1.48%
URI
J
REAL ESTATE
REIT - SPECIALTY
AMT -2.03
EQIX -1.03%
DLR
CCI
IRM
REIT - INDU
PLD +0.76%
PSA +0.26
AVB
EXR
ESS
REIT -
SPG +1.17%
O
REIT -
CSGP -0.23
CBRE
VICI
ENERGY
OIL & GAS INTEGRA
XOM +1.15%
CVX +1.33%
OIL & GAS E
COP
EOG +0.52%
PXD +1.11%
OXY +0.12%
HES +1.72%
EQT
OIL & GAS
MPC
OIL & GAS
PSX
VLO
OIL & GAS
SLB +0.70
HAL
BKR
OIL & GAS
WMB
KMI
OKE
UTILITIES
UTILITIES - REGULATED
NEE -0.25%
SO -0.21%
DUK -0.07%
SRE -0.90
AEP -0.70
PCG +1.38
ED -0.96
EIX
CEG
D
XEL -1.04
DTE
ES
AEE
FE
PPL
CMS
EXC -0.33
PEG +0.24
ETR
AWK
BASIC MATERIALS
SPECIALTY CHEMICALS
LIN +0.66%
SHW -0.60%
ECL -0.16
APD +0.48%
LYB
DD
IFF
ALB
CTVA +1.46
CF
STEEL
BUILDI
COPPE
CHEMI

Implicit presentation

- A sunburst chart is a radial visualization that represents hierarchical data using concentric circles.
- Structure
  - The center circle represents the root of the hierarchy.
  - Each subsequent ring represents a level in the hierarchy.
  - Slices are divided into segments representing subcategories.
- Interactive Exploration: Often allows users to zoom in on segments for detailed views.



https://observablehq.com/@d3/sunburst
https://observablehq.com/@d3/zoomable-sunburst

- https://observablehq.com/@3cd7d5ec89dc7f68/graph-visualization-introduction

HUST

ĐẠI HỌC BÁCH KHOA HÀ NỘI

HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.