GRAPHS

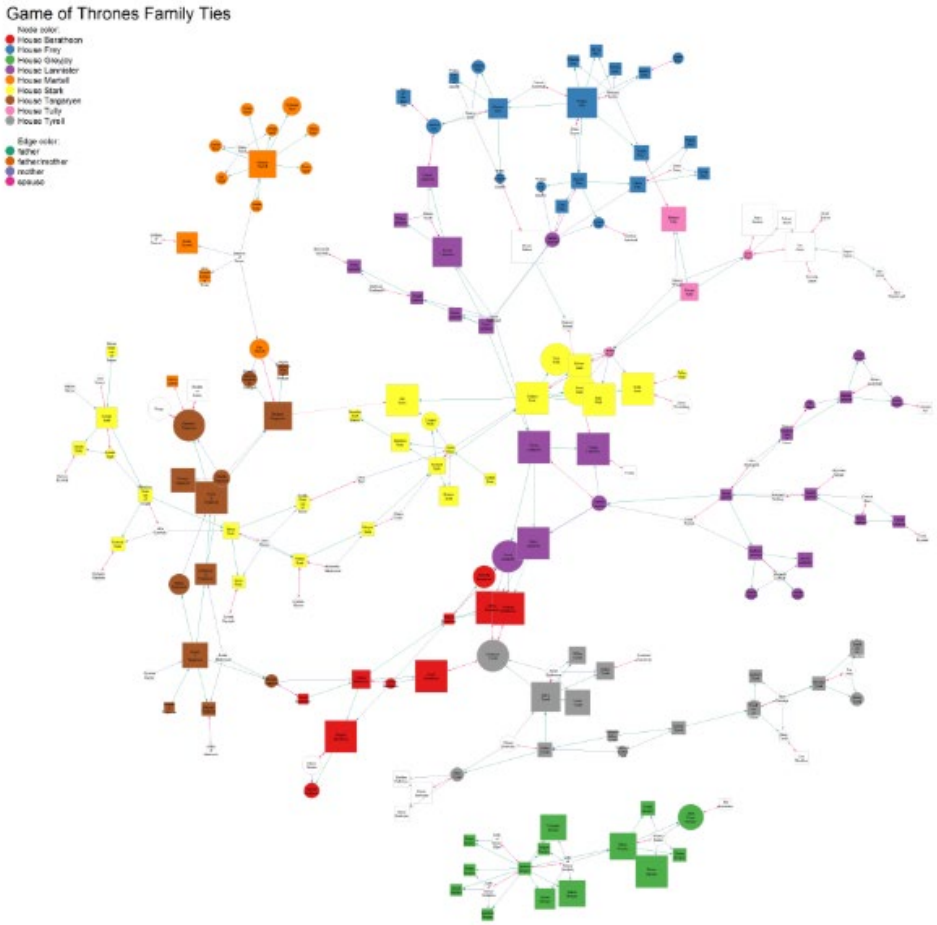
PRESENTATION AND VISUALIZATION – MIREIA RIBERA

DATA SCIENCE MASTER DEGREE

2 GRAPHS

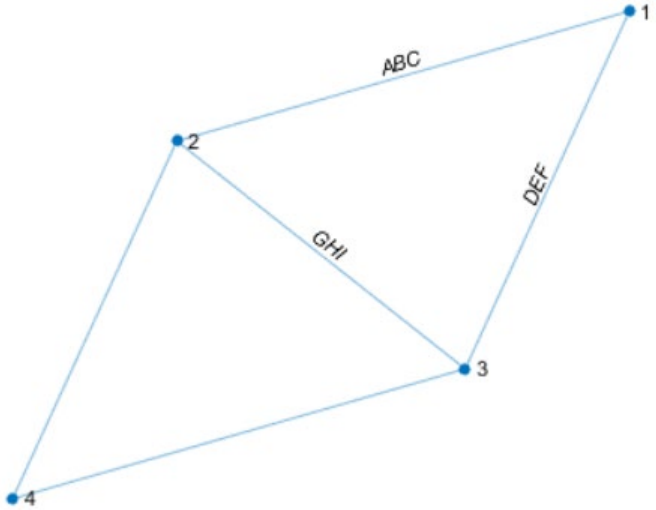
- What is a graph?
  - a **graph** is a structured representation of connected things and how they are related
  - **nodes** are entities (or essentially "things") that have relations between them. Nodes are often represented visually by a circle.
  - a **link/edge** is a relationship between **nodes** and is typically drawn as a line.
- Why graphs?
  - Graph analysis brings complex relationships to light

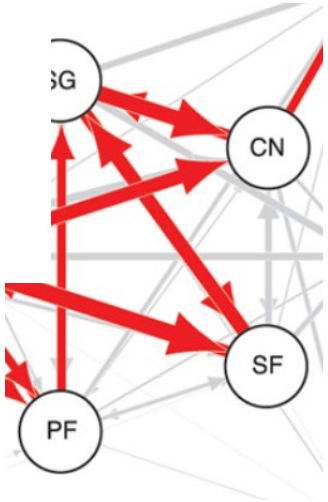
3 ENCODING AND INTERACTIVITY: NODES



- Category: colour / shape / image
- Attributes: tooltip with "key : value" pairs; Label
- Quantity: size, colour
- Highlight: border, colour (luminance, saturation), mark

4 ENCODING AND INTERACTIVITY: EDGES

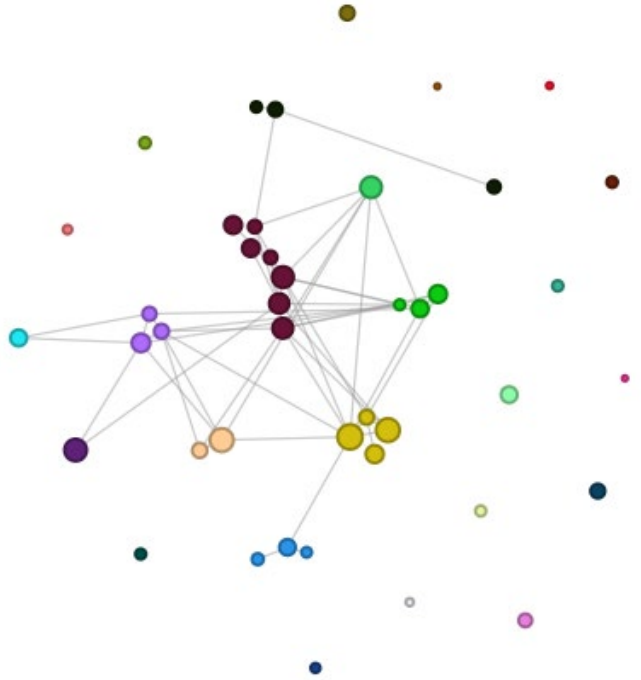




- Links/Edges:
  - Direction:
    - Arrows (directed graphs)
  - Category: colour / dash-array
  - Attributes:
    - Tooltip with "key : value" pairs ;
    - Label (weighted graphs)
  - Quantity: size
  - Highlight: border, colour (selected link or the rest)

5 ENCODING AND INTERACTIVITY: GROUPING

- Grouping
  - Similarity
  - Connectedness
  - Clustering (physical proximity)
  - Closure
- Interaction
  - Collapse / Expand
  - Select and filter: by attribute, by connection
  - Move nodes or edges



6 TYPES OF GRAPHS

- Spatial networks
- Flow diagrams
- Hierarchies
- Relationships
- Communities

7 SPATIAL NETWORKS

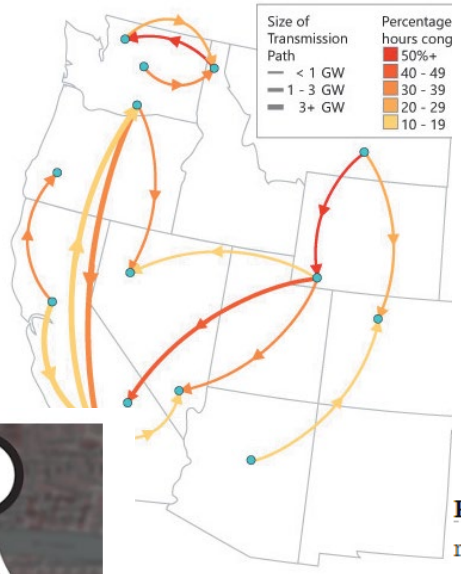


Figure 2-16: Schematic network diagrams abstract spatial layout to optimize legibility, as in the route map shown here.

The seven bridges of EULER

Figure 1-1: In the seven bridges of Königsberg problem, Leonhard Euler explored whether each bridge could be crossed only once. On the left is a map showing the seven bridges, and on the right is the graph equivalent.

Supply chain optimization

8 SPATIAL NETWORKS

- Def: Graphs based on spatial data
- The data can be plotted directly based on the spatial coordinates associated with the nodes and links.
- Because the relative position of nodes is predetermined, there can be challenges for working with this data.
  - metro maps have changed it for the sake of simplicity!

9 FLOW DIAGRAMS

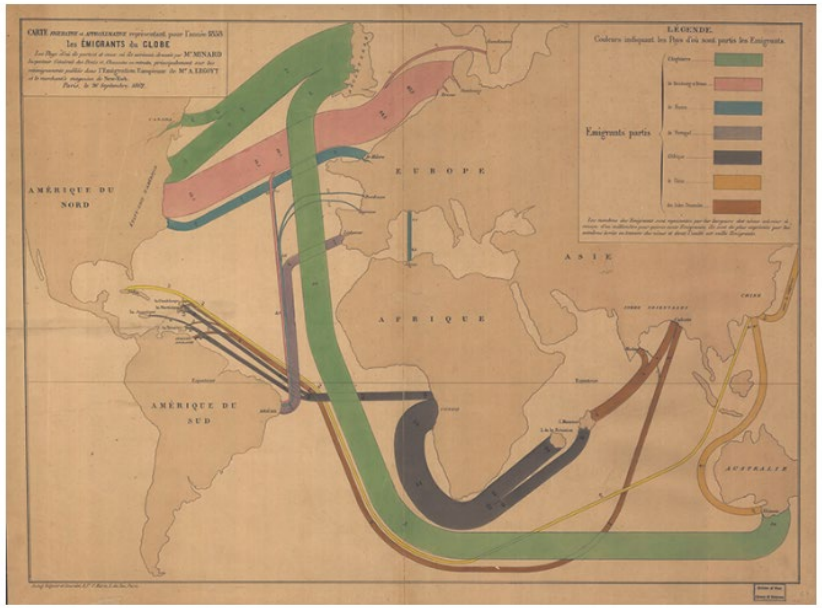


Figure 1-3: Joseph Minard's flow graph shows emigrants worldwide in 1858.

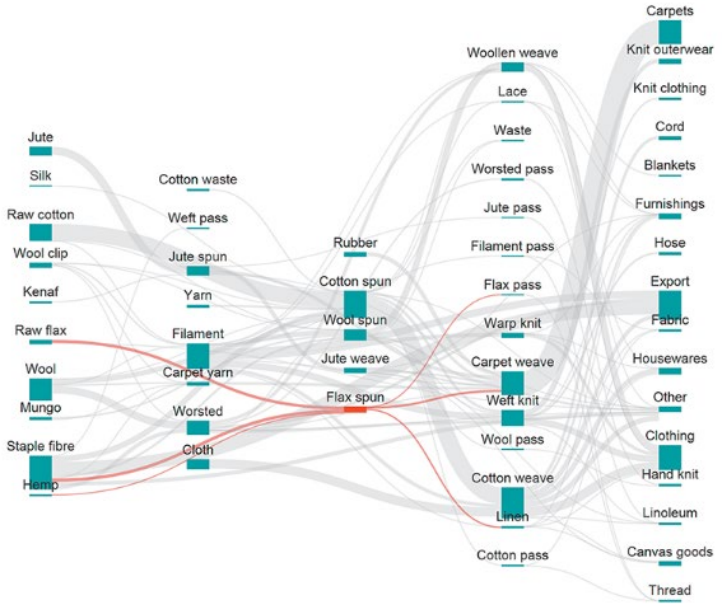


Figure 2-11: Sankey diagrams are an ideal graph technique for showing flow. Here, flow of materials in textile production is shown, where width indicates volume.

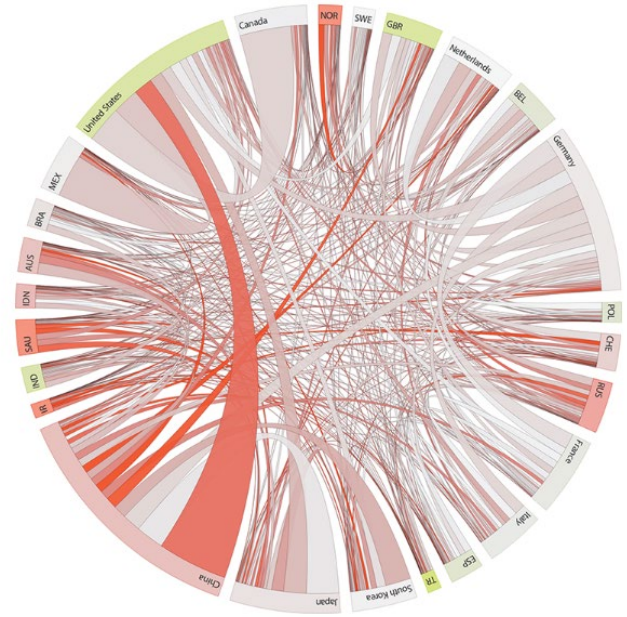
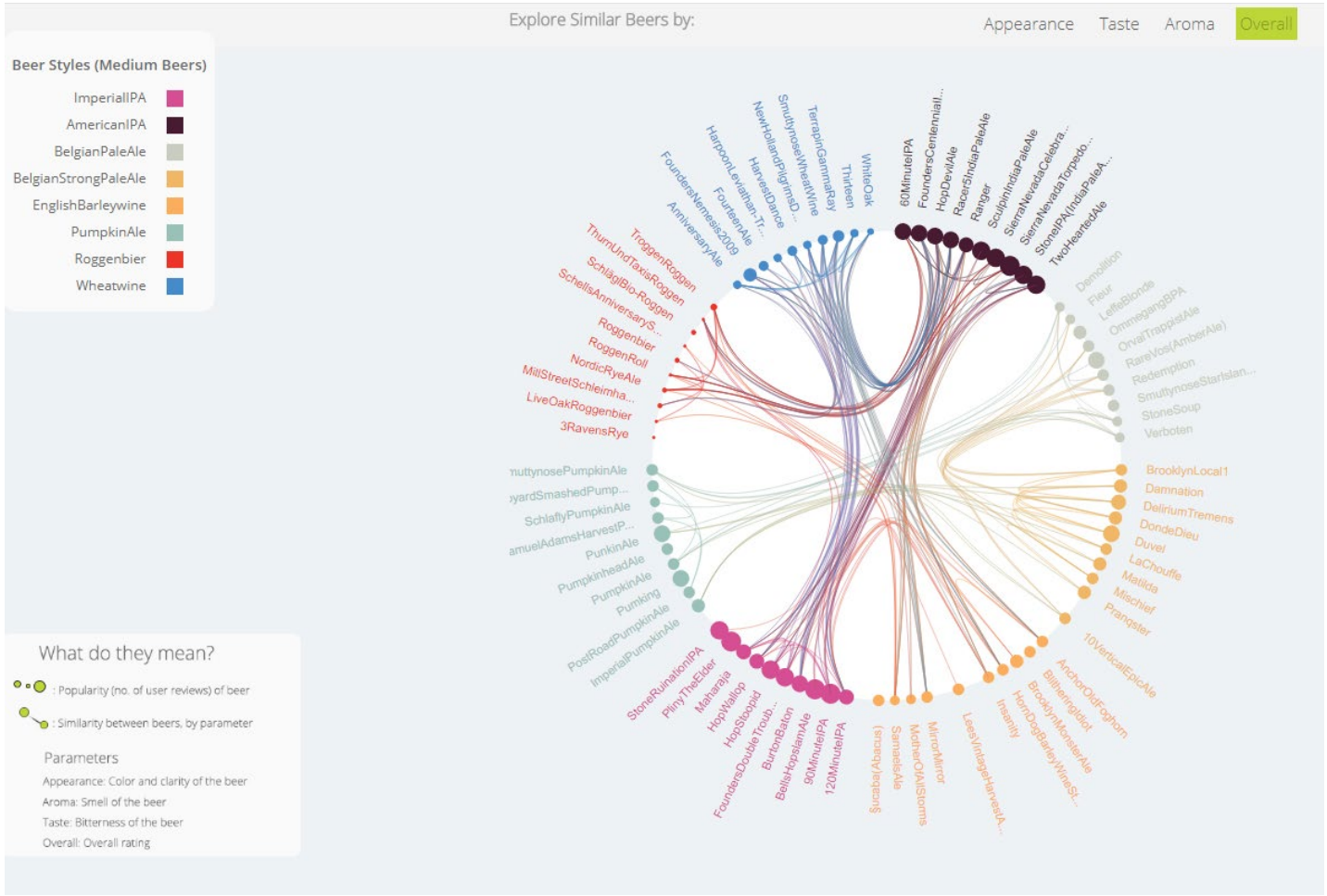


Figure 2-15: Chord diagrams show reciprocal flow between entities. Here, all reported trade of goods between countries is represented by width of link at the exporting country. Color flags trade imbalance.

10 FLOW DIAGRAMS

- Def: Analyse the movement of people, goods, money through the world, through processes or through websites
- Typical charts for flow visualization are Sankey diagrams and Chord diagrams.
- Selecting relationships and moving nodes in Sankey diagrams are typical analysis possibilities

11 SOME EXAMPLES



12 HIERARCHIES

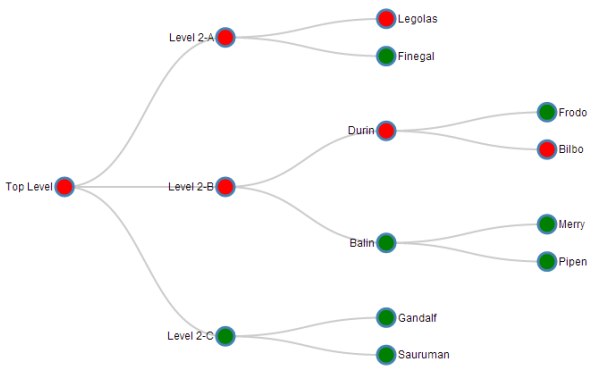
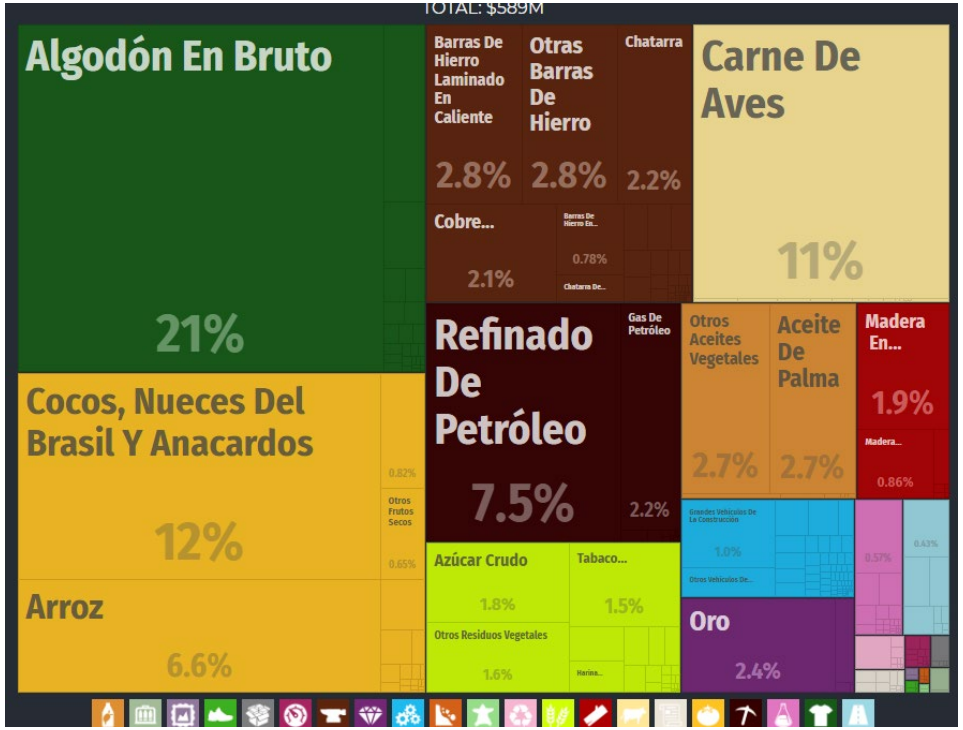


Figure 2-5: An org chart reflects organizational structure using a tree. Corporate performance-related characteristics such as department size and affirmative action employment classification can be mapped to size and color of each node.

Figure 2-6: A sunburst chart provides an alternate representation of hierarchy appropriate for viewing organizational subdivisions by proportion of the whole. Profit and loss are shown in degrees of green and red, revealing roots of overall corporate performance.



13 HIERARCHIES

- Def: tree graphs (non cyclic graphs), suited for organizational charts, taxonomies, decision trees.
- They are particularly useful to organize data and to navigate big spaces of concepts

14 GRAPHS FOR RELATIONSHIPS

Cross-Sell Patterns		First Device Purchased				
		Smart Phone	Music Device	Tablet Computer	Laptop Computer	Desktop Computer
Additional Device Purchased	Smart Phone	-	63%	12%	7%	28%
	Music Device	4%	-	1%	3%	2%
	Tablet Computer	19%	18%	-	18%	19%
	Laptop Computer	11%	6%	11%	-	3%
	Desktop Computer	4%	9%	8%	4%	-

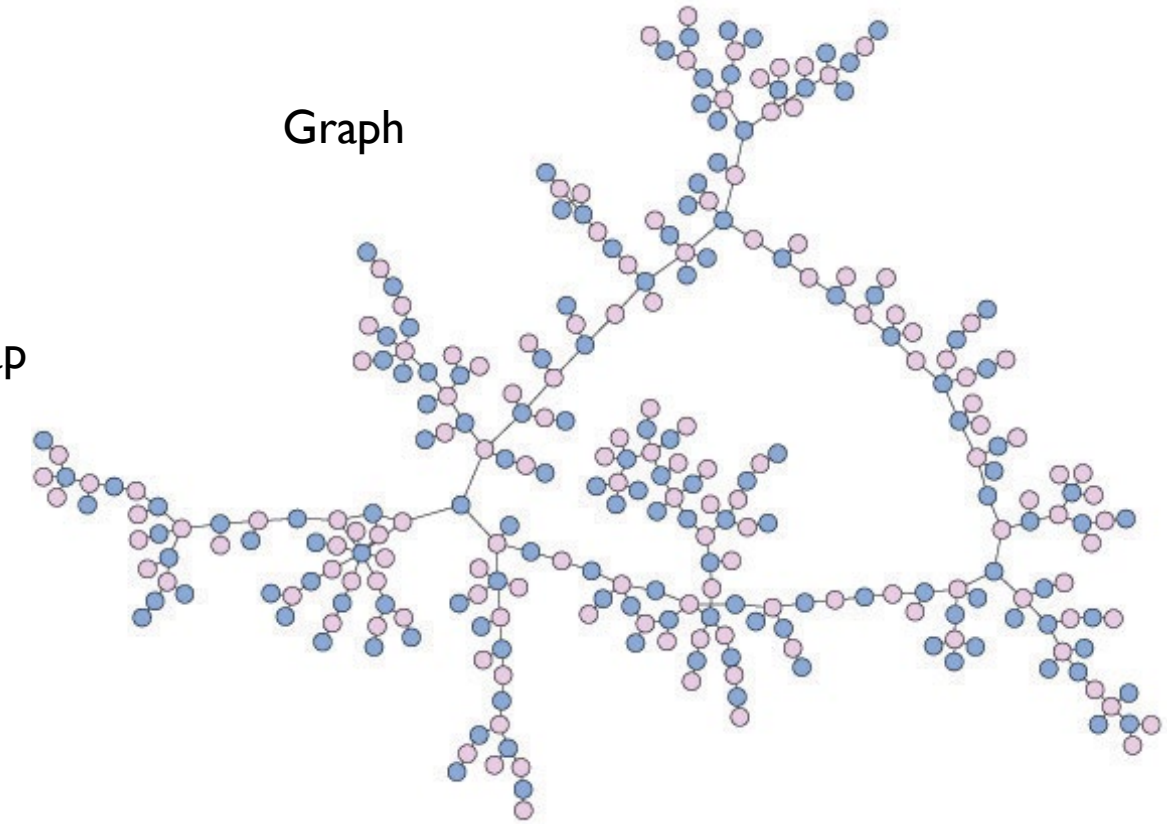
Adjacency matrix

Sub-Industry	2011	2012	2013	2014	2015	2016(YTD)
Healthcare						
Advertising, Sales & Marketing						
Business Intelligence						
Security						
Finance						
IoT/Wearables						
Education						
Customer Relationship Management						
Personal Assistants/Productivity						
E-Commerce						
Robotics						

CB INSIGHTS Min Max www.cbinsights.com

Heat map

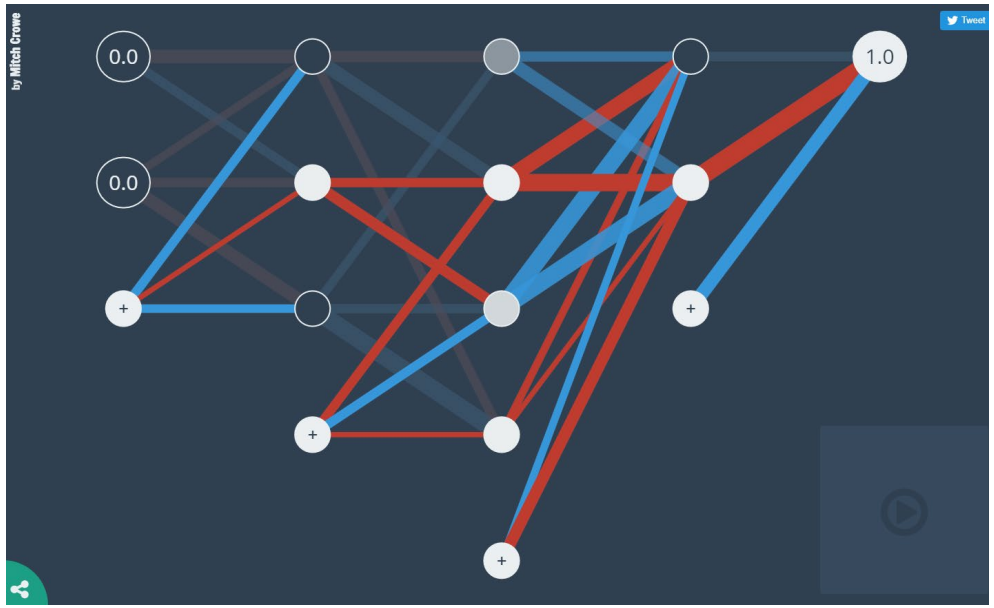
Graph



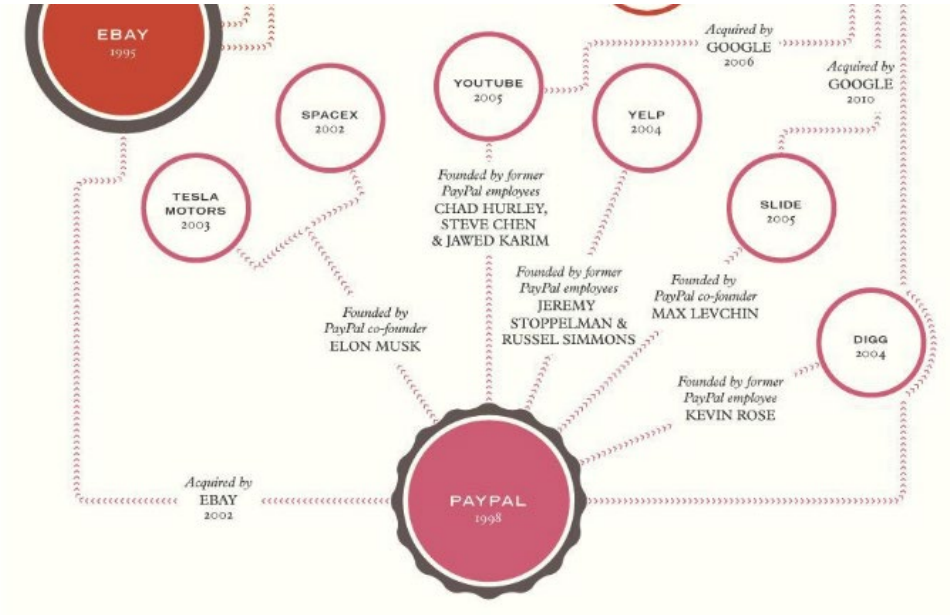
15 RELATIONSHIPS

- Def: Analyse relationships and identify clusters, on geographical or logical connections between things.
- Graphs may have any number of links between a pair of nodes.
- For some types of applications (such as fraud analysis), it is important to keep all links between these nodes and have techniques to analyze the many different connections.

16 SOME LIVE EXAMPLES

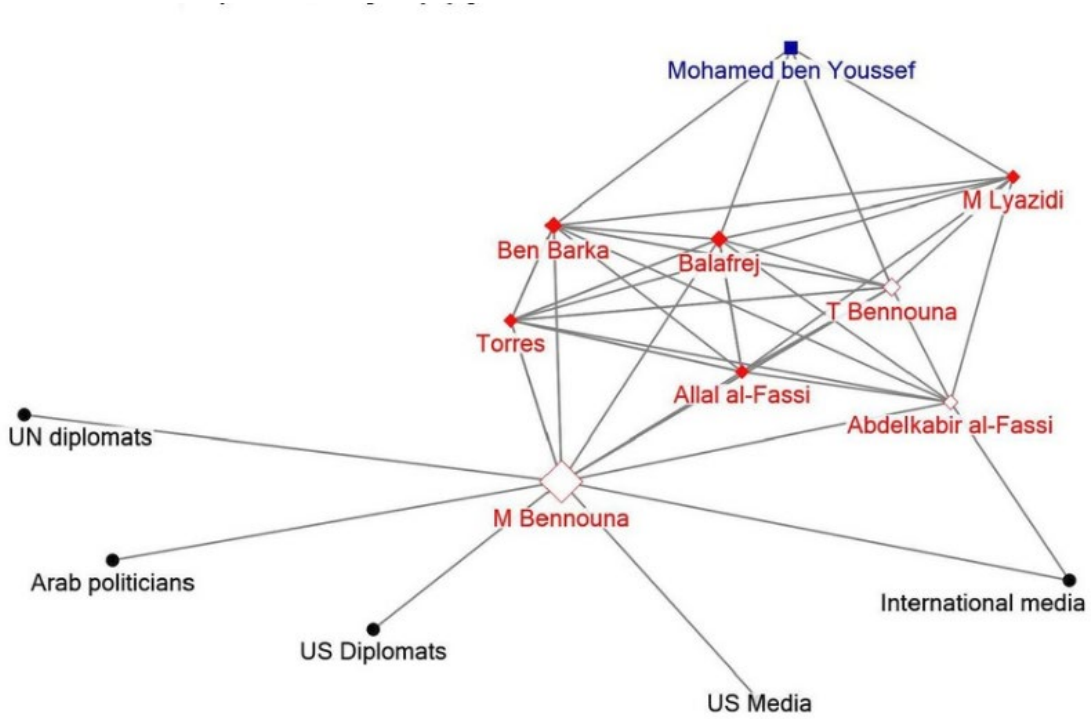


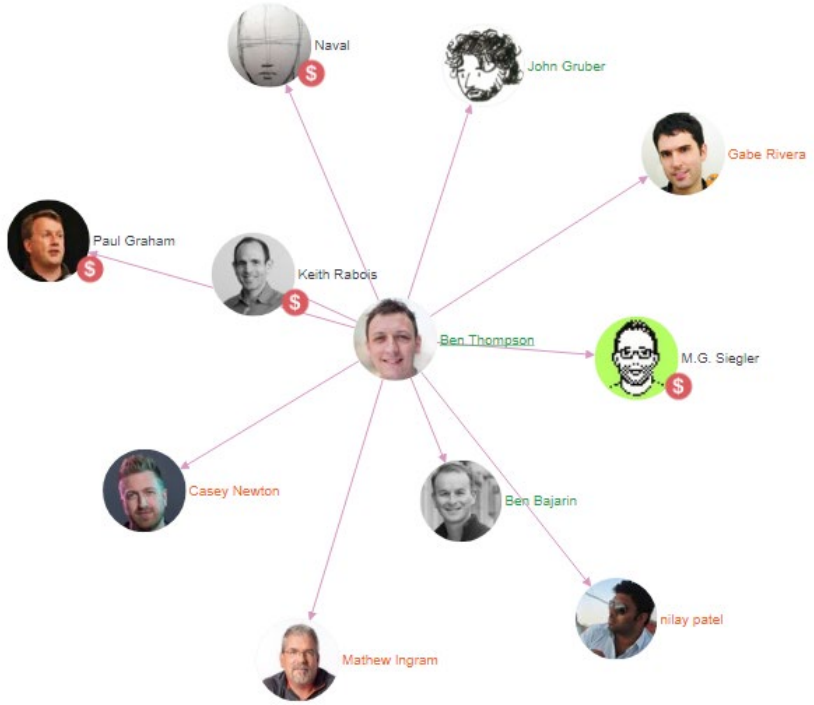
Neural networks



Tech companies relations

17 COMMUNITIES



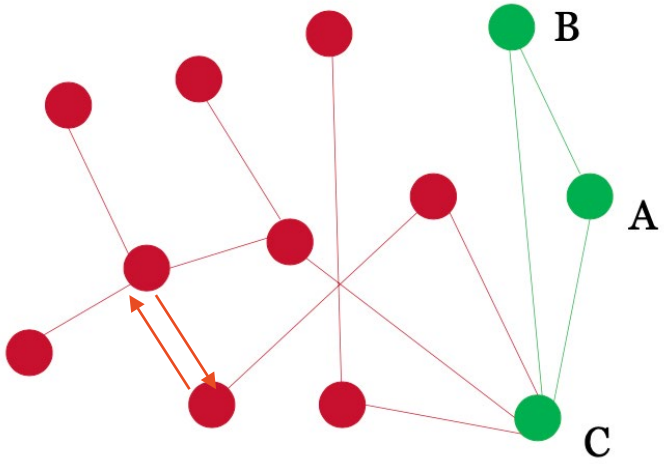


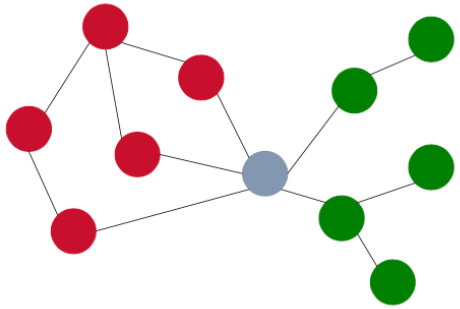
18 COMMUNITIES

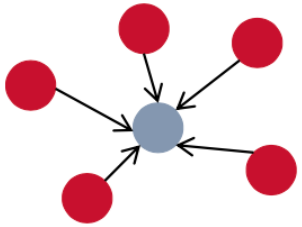
- Def: analyze large spaces of data, mainly to identify groups
- The clustering of nodes in graphs reveals communities.
- Enhancing node and link data can help refine the qualities to define these communities
- Filtering, grouping, and additional analytic techniques make the communities visually apparent
- Specific field: social network analysis (interaction, ties)

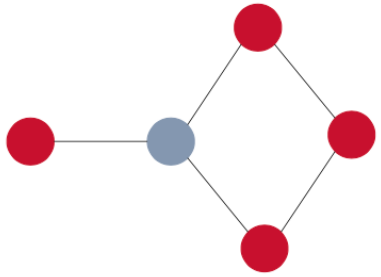
19 SNA (SOCIAL NETWORK ANALYSIS)

- Analyse the links between nodes
  - Interaction patterns
  - Power concentration
    - Density: # edges / potential edges ex. 12/139
    - Transitivity: if A is connected to C and B is connected to C, it is probable that B is connected to A
    - Reciprocity: reflects power relationships
- Position of the nodes:
  - Centrality (in or out-degree), betweenness centrality, eigenvector centrality

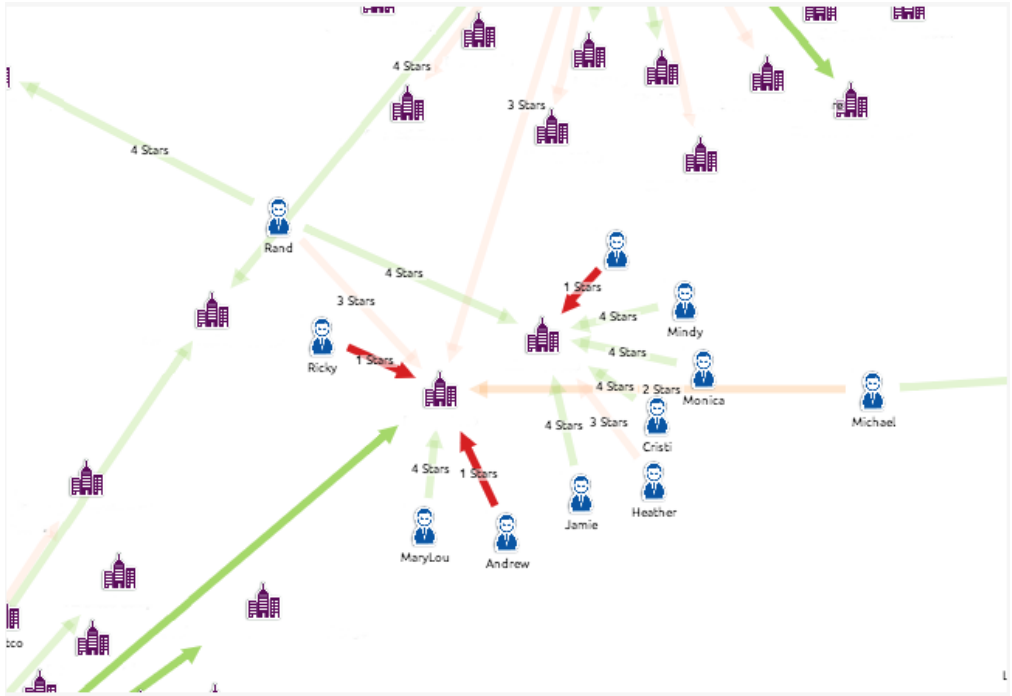




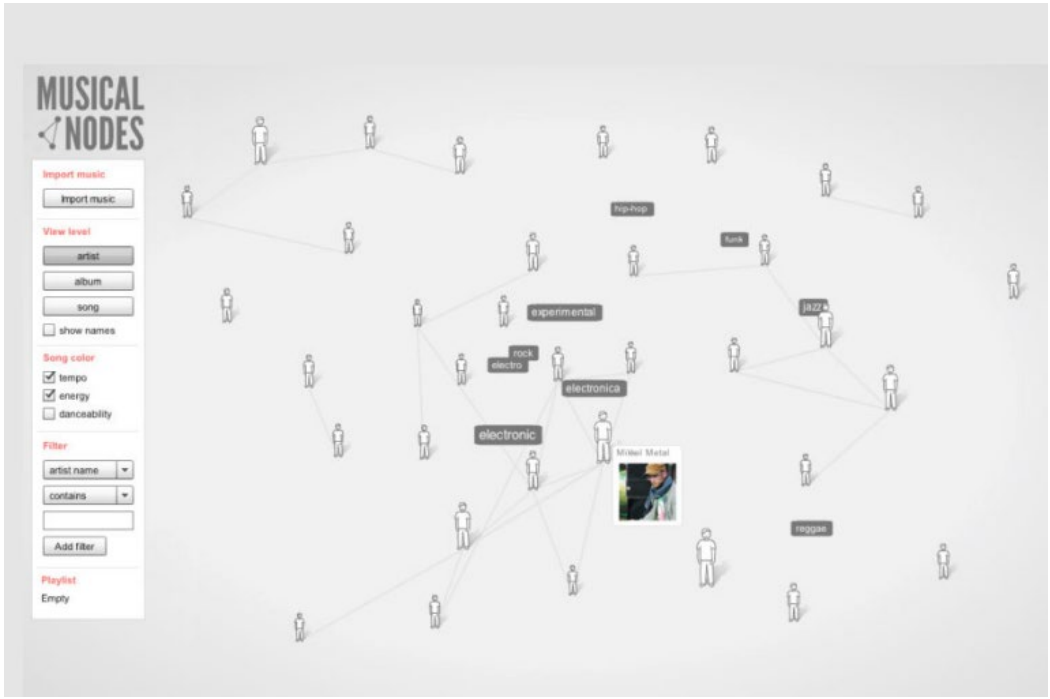




20 SOME EXAMPLES



Fraud analysis



Music networks

21 BIBLIOGRAPHY

- David Jonker, Richard Brath Graph analysis and visualization: discovering business opportunity in linked data. John Wiley and Sons, 2015 ISBN 9781118845844
- Edwards, G & Crossley, N. (2009). Measures and Meanings: Exploring the Ego-Net of Helen Kirkpatrick Watts, Militant Suffragette, *Methodological Innovations Online*, 4, p. 37-61