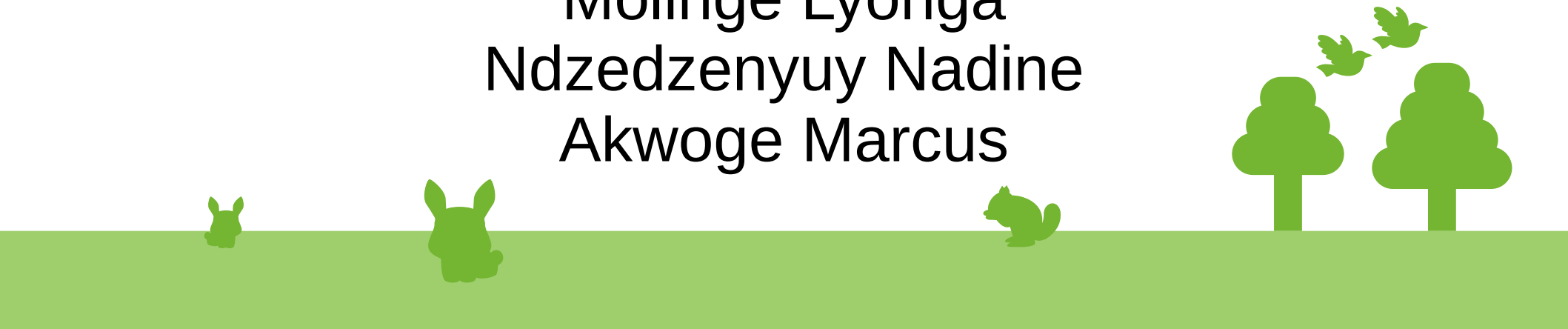
Visualizing Networks and Graphs

By
Molinge Lyonga
Ndzedzenyuy Nadine
Akwoge Marcus



Outline

- Definitions of graphs and network
- Key concepts in network graphs
- Why visualize network graphs
- Graph representation
- Examples of Graph and Network
- Issues to consider when representing structured data through a graph
- Process of Visualization
- Case Study

Definition of graphs and networks

- **Graph:**
  - A graph is a visual representation of data, particularly used to show the relationship between different elements.
  - It’s a diagram or chart that uses points, lines, and other shapes to depict how data changes or relates to other data points.
- **Network :**
  - Network (also called a network diagram or graph visualization) is a visual representation of a network, showing how different entities or objects are connected.

Key concepts in network graphs

- **Nodes:** Represent entities or objects in the network, such as individuals, companies, or cities.
- **Edges:** Represent the connections or relationships between nodes, such as friendships, collaborations, or transactions.
- **Directed Graphs:** In directed graphs, edges have a direction, representing one-way relationships (e.g, follower-follower relationships on Twitter).
- **Undirected Graphs:** In undirected graphs, edges represent mutual relationships (e.g, friendships on Facebook).

Key concepts in network graphs

- **Degree:** The number of connections (edges) a node has. A higher degree indicates a more connected or influential node.
- **Centrality:** A measure of how central or important a node is within the network. Popular centrality measures include degree centrality, between centrality, and closeness centrality.
- **Clustering:** Detecting clusters or groups of nodes that are more densely connected with each other than with other nodes in the network.
- **Path:** A sequence of edges that connects two nodes.

Why Visualize Network Graphs?

- Network graphs are a great way to visualize relationships and interactions between entities such as people, organizations, or objects.
- **Reveal Patterns:** Discover hidden patterns, clusters, and connections within large datasets.
- **Identify Key Entities:** Highlight important nodes (entities) that play a central role in the network (e.g, influencers in social networks).

Why Visualize Network Graphs?

- **Understand Connectivity:** By representing connections through nodes and edges, network graphs can reveal insights into how entities are linked, clusters form, and information flows through the network thereby facilitating the analysis of information flow or relationship.
- **Analyze Communities:** Detect groups or communities within a network that share common characteristics.
- For example, in a social network, network graphs can help identify influential users, communities, and the spread of information.

Graph representation

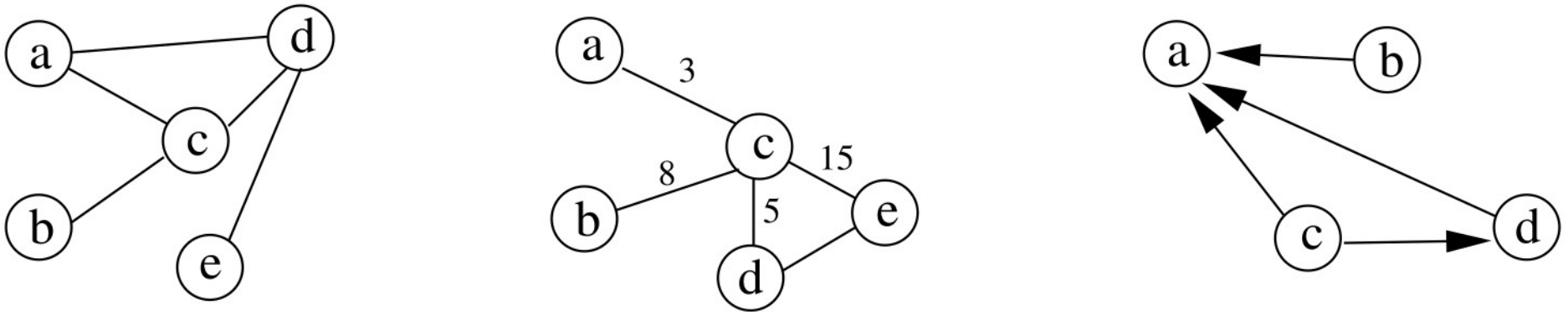


Fig 1: On the left is a normal graph, in the center is a graph in which each edge is given a numerical value, and to the right is a directed graph.

Graph representation

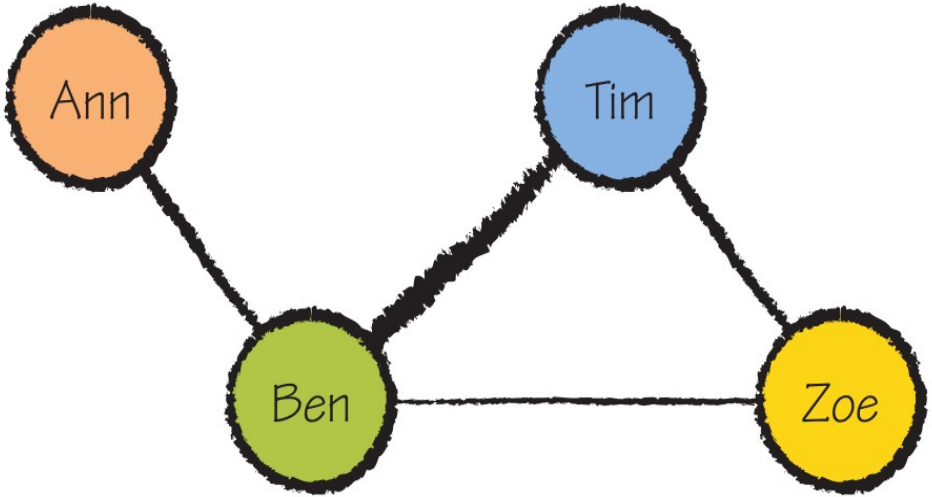


Fig 2: In this graph, lines indicate the relationships between people. The color and width of the line add extra information.

Examples of Graph and Network

- Social Networks
- Transport networks
- 3D Graphs
- Organization communication newtorks (e.g emails, meetings).
- Student learning path graphs
- etc

Issues to consider when representing structured data through a graph

1. *Positioning of nodes:*
   - Graphs often represent abstract data types that don’t have a natural spatial location.
   - It is necessary to decide on which criteria to arrange the nodes in the space (layout).
   - *Multidimensional scaling* can convert a collection of multivariate data (of any dimension), mapping them into one, two, or three dimensions, from which the Cartesian coordinates are derived to position the nodes in the space.
   - Can map some attributes to the shape, color, and dimension of the nodes.

Issues to consider when representing structured data through a graph

2. Representation of the edges

- A relationship between two nodes has to be represented by an edge.
- They can have associated weights.
- Can be direct or indirect.
- The weights can be indicated next to the edges or mapped to the color or to the width of the edge.

Issues to consider when representing structured data through a graph

3. Dimensioning:

- Some datasets can have thousands or even millions of records.
- These records can’t be directly represented by graphs in a one-to-one relationship with nodes and edges.
- It is necessary to find solutions in such cases.

Issues to consider when representing structured data through a graph

4. Interaction with graphs

- Modern visual representation software allows the user to interact with the general view.
- Such interaction is very helpful when we deal with complex graphs with a high number of nodes and edges.
- The user can manipulate the representation of a graph to move the nodes, zoom in on a part of the graph, and hide or show edges or even a part of the graph.

Process of Visualization

- Data:
  - Collect and clean the data.
  - Prepare it for use as a graph dataset with explicit nodes and edges and associated attributes for each.
  - For example, for e-mail data, this involves creating nodes and edges out of To, From, and Cc fields.

Process of Visualization

- Layout:
  - Review statistics and arrange the nodes and edges in a way that reveals insightful patterns such as components, clusters, and so on.
  - Many different kinds of layout are possible.
  - For example, in an e-mail graph, this may mean identifying nodes that act as bridges between different groups of people.

Process of Visualization

- Add Visual Attributes
  - Adjust labels, sizes, colors, and line thickness to enhance understanding.
  - For example, with an e-mail dataset, you can use additional data such as the number of messages, message size, and how recent the e-mail is to adjust visual attributes such as node size and color.
  - This helps differentiate particular individuals of interest.

Process of Visualization

- Interact, Explain, Further Analysis
  - Zoom, select, filter, annotate, and explain.
  - In a social dataset, zooming, filtering, and drilling down all provide ways to isolate data of interest and to identify particular individuals.
  - You can then present or publish the results or export/integrate the result with other software for further analysis.

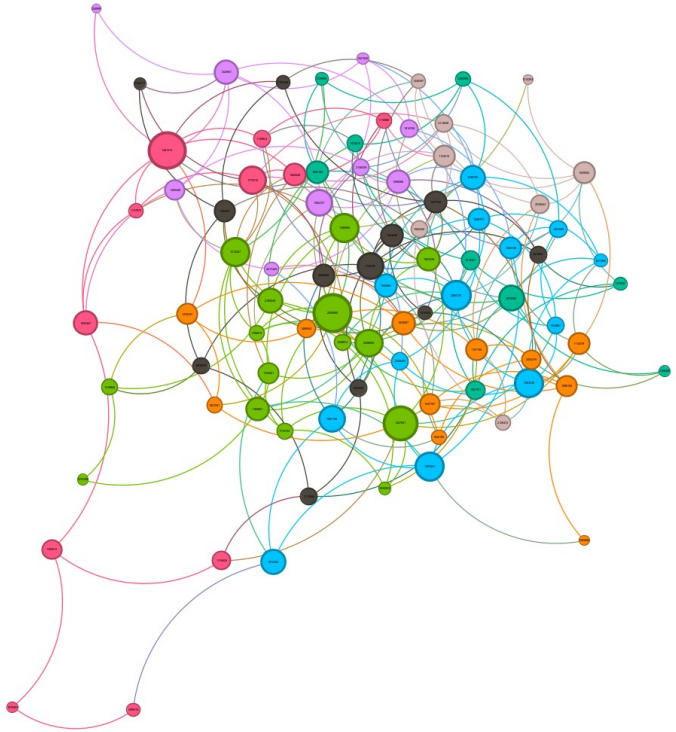
Case Study

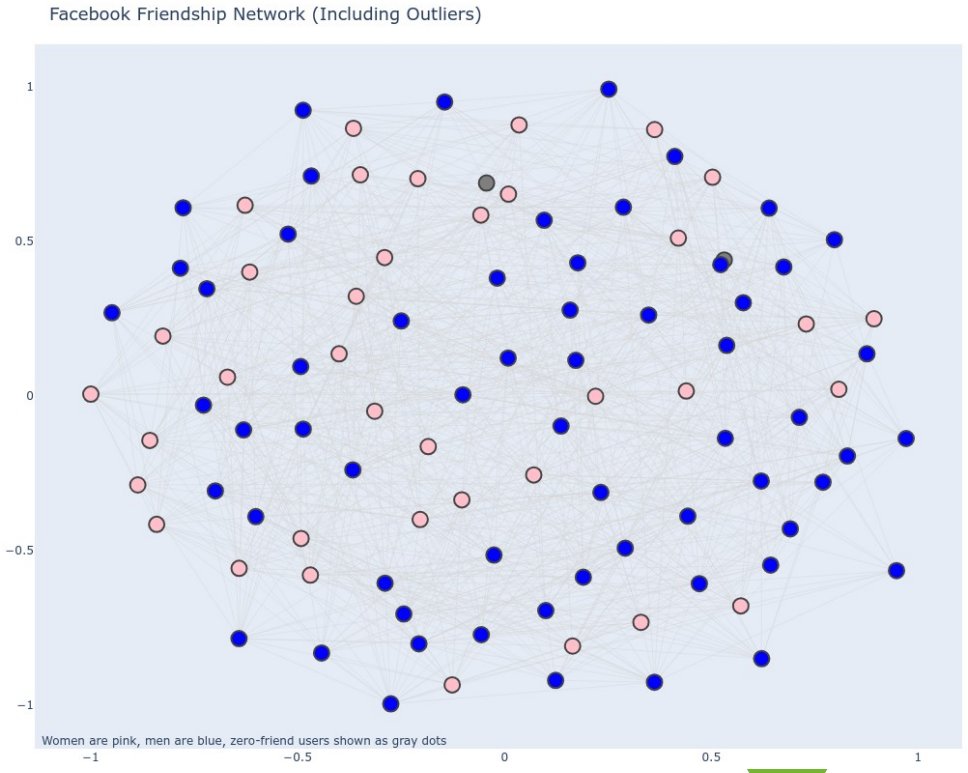
- Visualizing Social Network (Pseudo Facebook) dataset.
- This data analysis gives insights from Facebook dataset which consists of identifying users that can be focused more to increase the business. These valuable insights should help Facebook to take intelligent decision to identify its useful users and provide correct recommendations to them.
- Tools used are:
  - **Python** for data and network preparation with graph display.
  - **Gephi** for network visualization.

Case Study

- Questions to answer
  - How does age affect number of friends?
  - What makes users "popular" (high friend counts)? Gender? Age? Activity?
  - Do males and females behave differently in terms of friend count or activity?
  - Which users are "influencers" based on their network centrality?

Sample diagrams (Gephi, Plotly)





Reference Texts

- Graph Analysis and Visualization: Discovering Business Opportunity in Linked Data - by Richard Brath and David Jonker
- Introduction to Information Visualization – by Riccardo Mazza