**Social Media Data Types**

* Behavioral Data
* Engagement Data
* Personal Data
* Attitudinal Data
* Preference Data

**5 trait of data quality:**

* Accuracy
* Completeness
* Reliability
* Relevance
* Timeliness

Diagram

Description automatically generated with medium confidence

**Degree Distribution**

P(d)= n(d) / d

N(d) is the no of node with degree d.

**Type of graph:**

* Null Graph or Empty Graph - A null graph is a graph in which there are no edges between its vertices. A null graph is also called an empty graph.
* Directed and Undirected
* Weighted
* Signed – Weights are in binary like 0 or +1/-1 or +/-
* Web graphs – Directed graph, node represent as html page. Edge is like we can go from one back to another page.

Walk – Traverse via edge and node/edge can be visited more than once.

Open: Start and destination are not same

Close: Start and destination are same

Trail – Traverse via edge but edge can be only visited once.

Close trail is tour/circuit.

Path – Traverse via edge and edge & node can be only visited once.

Close path is cycle.

Graph diameter is the largest shortest path.

BFS, DFS and Dijkstra

**Network Density:**

Undirected:

(2\* no of e)/((no of node) \* (no of node -1))

Directed:

(no of e)/((no of node) \* (no of node -1))

**Centrality Measures**

* Degree Centrality. – Aju baju wale node. Number of neighbor node.
* Closeness Centrality.

C(x) = N – 1 / Sum(d(y,x))

N = no of node

D(y,x) = the shortest distance

Imagine we have a road network where cities are your nodes and roads are the edges. you own a clothes business and you want to install a new storage facility to supply your clients in all the cities. Trivially you will choose to place it in a city that is closer to all the other ones, in order to reduce your transportations costs.

* Betweeness Centrality.

Count no of shortest path of ever node to every node.

Count no of shortest path of ever node to every node which includes node for which we are looking for.

Sum (Include/All)

A picture containing graphical user interface

Description automatically generated

Thus a node with high betweenness centrality may have considerable influence over the information passing between other nodes.

Clustering coefficient

Diagram

Description automatically generated