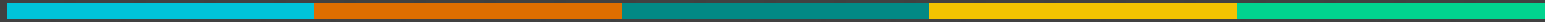


BIG DATA



DOWN THE RABBIT HOLE WE GO

OVERVIEW



NETWORKS



NODES



EDGES



HYPERLINKS



WEB
CRAWLERS



BIG DATA NETWORKS

NETWORKS

TYPES OF BIG DATA NETWORKS

EGOCENTRIC NETWORK

Network focused on one node

COMPLETE NETWORK

All nodes and edges are equal

PARTIAL NETWORK

Partially observed sampling of nodes

UNIMODAL NETWORK

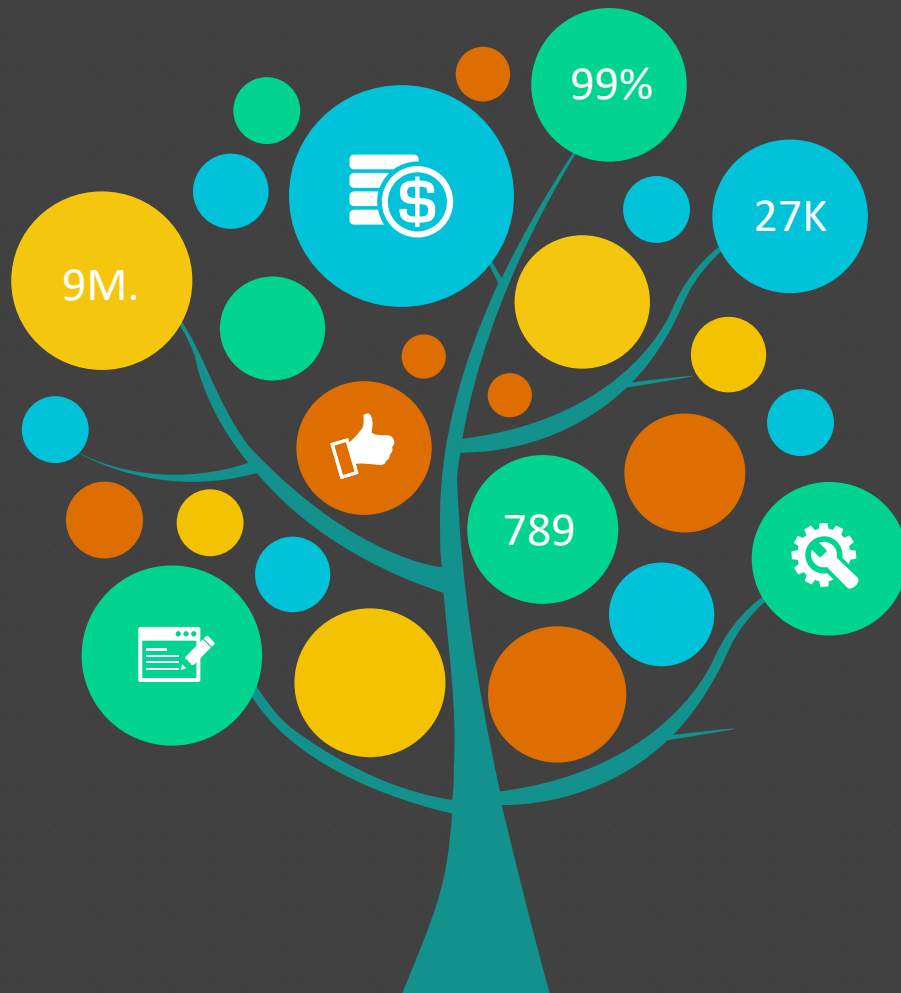
Contains only one type of node

BIMODAL NETWORK

Contains exactly two types of nodes

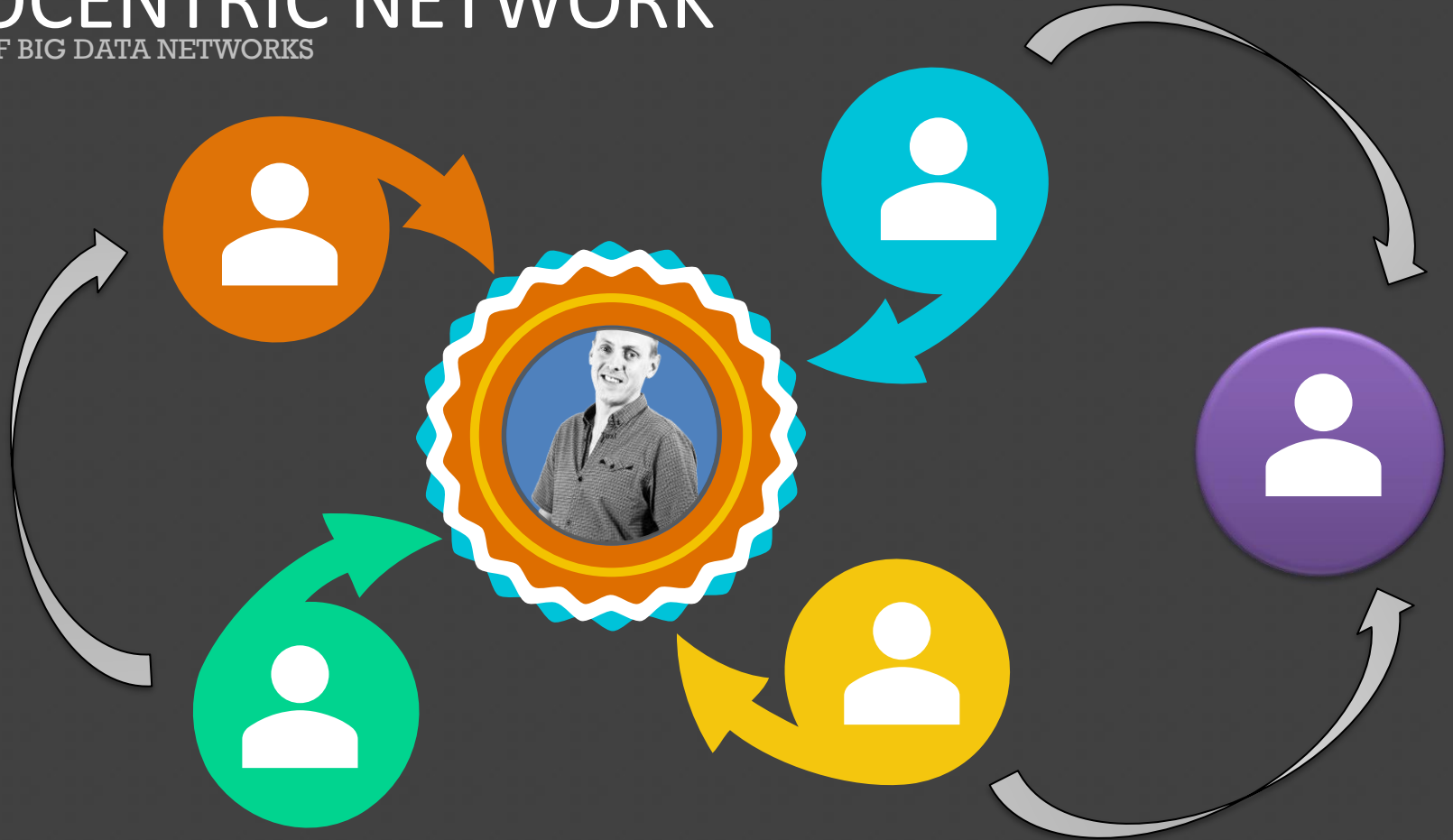
MULTIPLEX NETWORK

Contains multiple differing nodes



EGOCENTRIC NETWORK

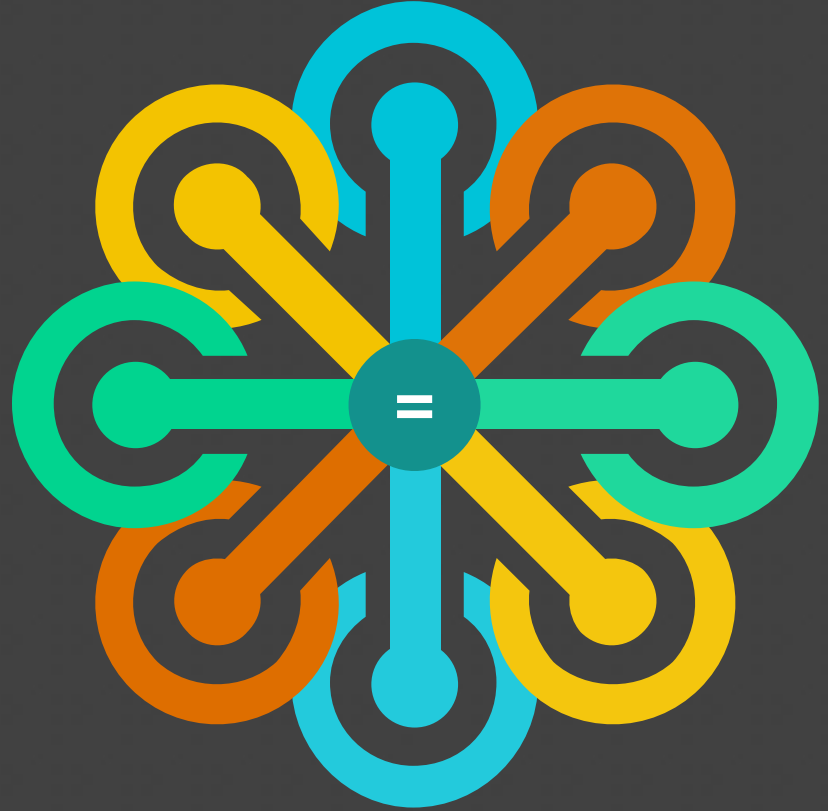
TYPES OF BIG DATA NETWORKS



COMPLETE NETWORK

BIG DATA NETWORKS

ALL NODES & EDGES ARE EQUAL



PARTIAL NETWORK

BIG DATA NETWORKS

ACCESSES ONLY PART OF THE NETWORK
(“SNOWBALL” SAMPLING)



UNIMODAL NETWORK

BIG DATA NETWORKS

ALL NODES ARE THE SAME DATA TYPE



BIMODAL NETWORK

BIG DATA NETWORKS

THERE ARE PRECISELY TWO NODE TYPES



MULTIPLEX NETWORK

BIG DATA NETWORKS

THERE ARE MULTIPLE NODE TYPES





BIG DATA NODES

NODES

BIG DATA NODES



MASTER NODE

- Hadoop File System (HDFS)
- Distributed data storage
- Map Reduce (Job Tracker)



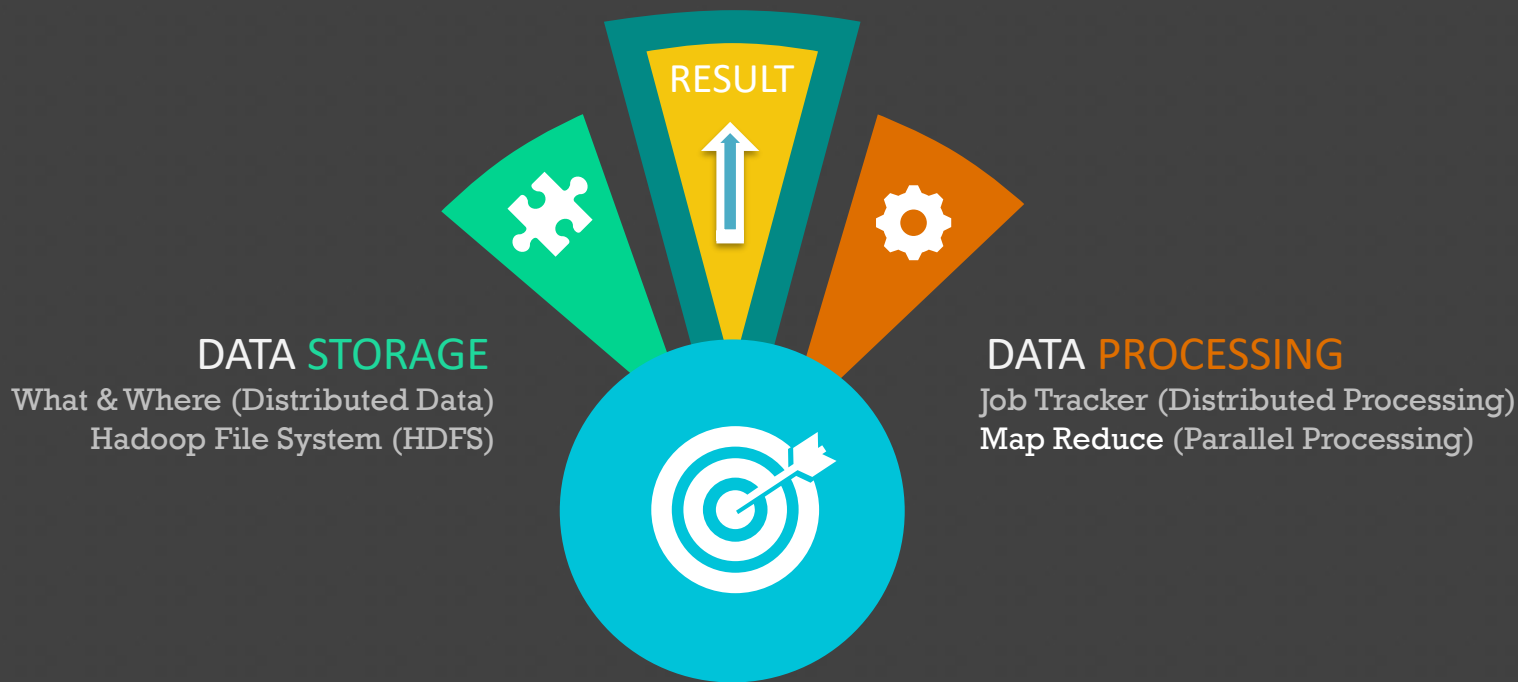
SLAVE NODE

- Holds unstructured data
- Task Tracker



MASTER NODE

BIG DATA NODES



TRADITIONAL PROCESSING

BIG DATA NODES



THE QUERY

Query sent to server /
database



THE PROCESS

Sorting & filtering of
data occurs at origin



THE RESULT

Extraneous data
discarded



THE OUTPUT

Results of data
processed by software

FASTER (PARALLEL) PROCESSING

BIG DATA NODES



THE QUERY

Loads data (query) to a data node



THE PROCESS

Data analysed (Map Reduce) at data node



THE RESULT

Stores result to data node



THE OUTPUT

Reads the result from data node to origin

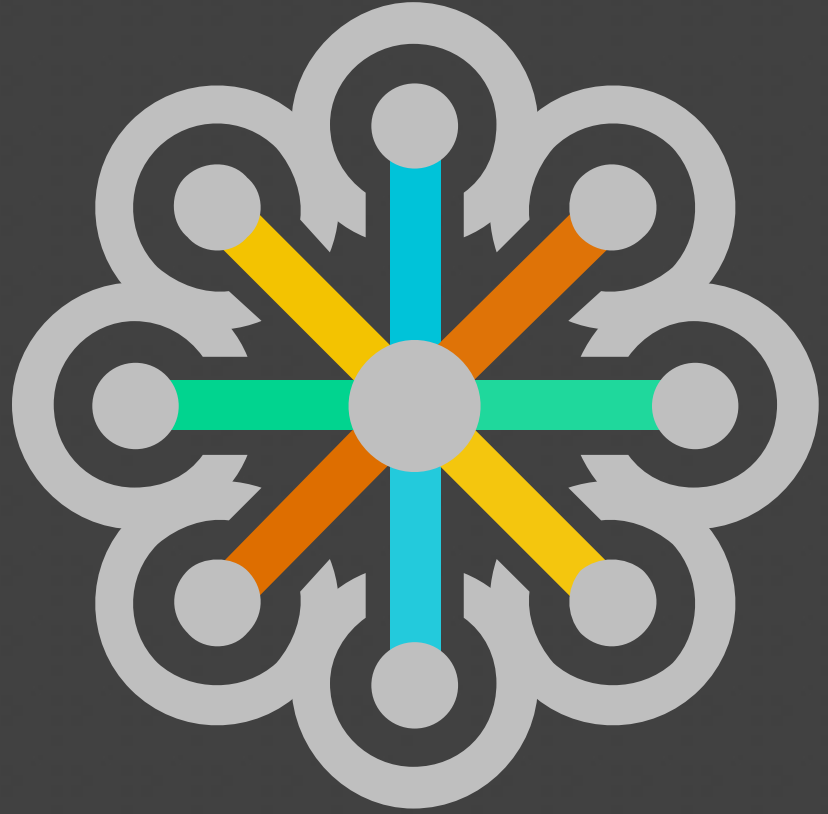


BIG DATA EDGES

EDGES

BIG DATA EDGES

EDGES ARE THE **LINKS** BETWEEN NODES



EDGES

BIG DATA EDGES

DIRECTED

Data is passed in a specific direction

WEIGHTED

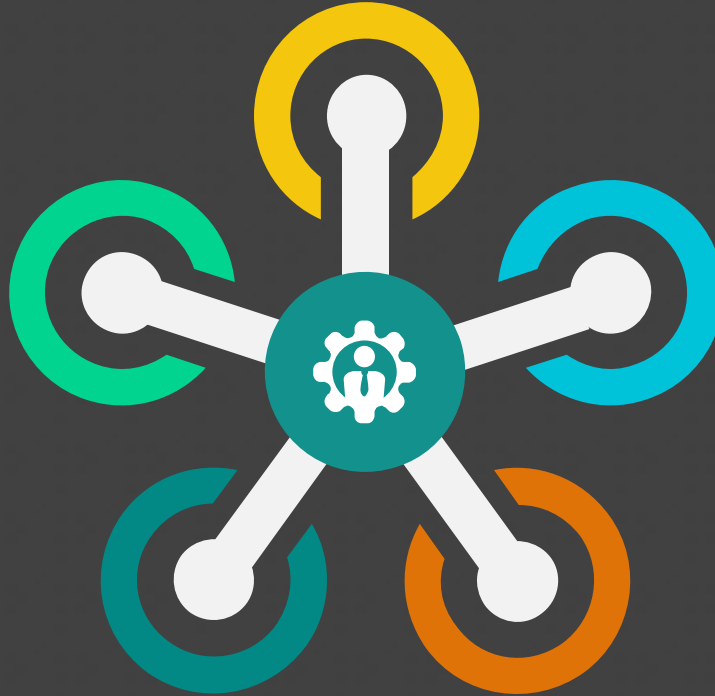
A value (weight) is assigned to the data

LABELLED

Data is given an identifier (label)

HOMOGENOUS

Data is familiar / common structure



UNDIRECTED

Data can travel in any direction

UNWEIGHTED

The data has no value (weight) assigned

UNLABELLED

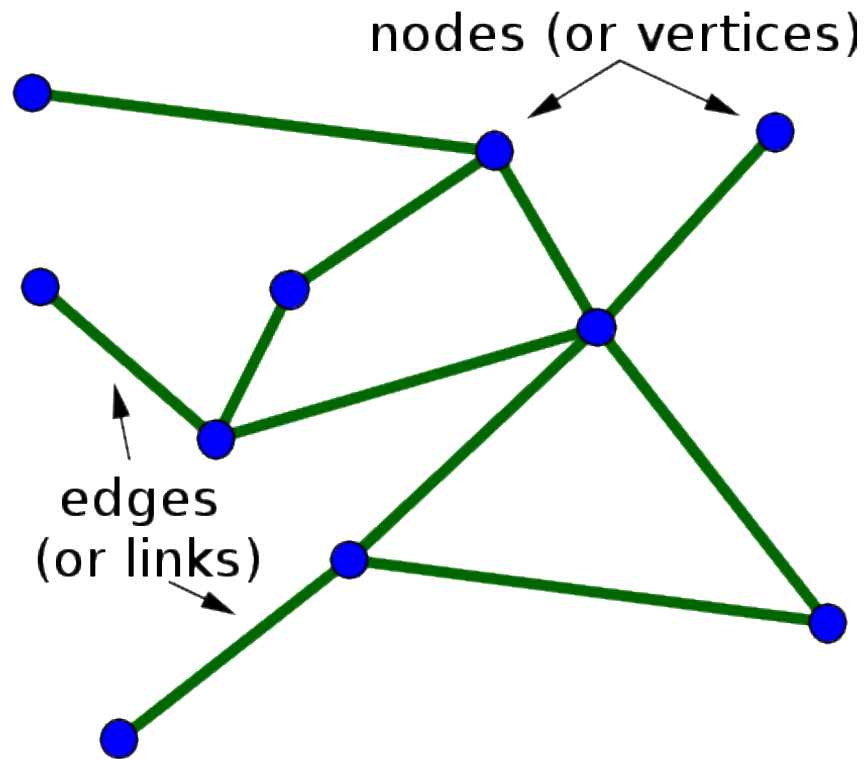
Data has no unique identifier assigned

HETEROGENOUS

Data relies on semantic analysis

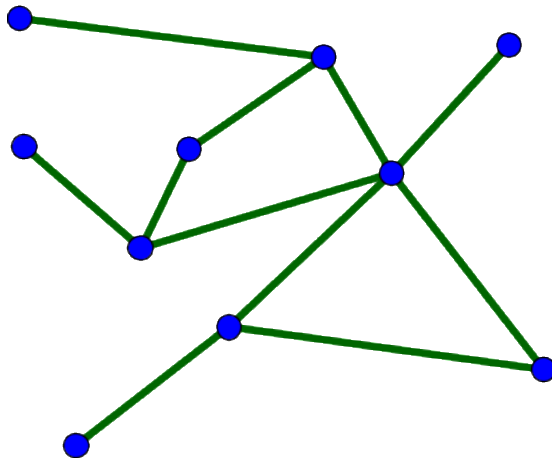
EDGES

BIG DATA EDGES



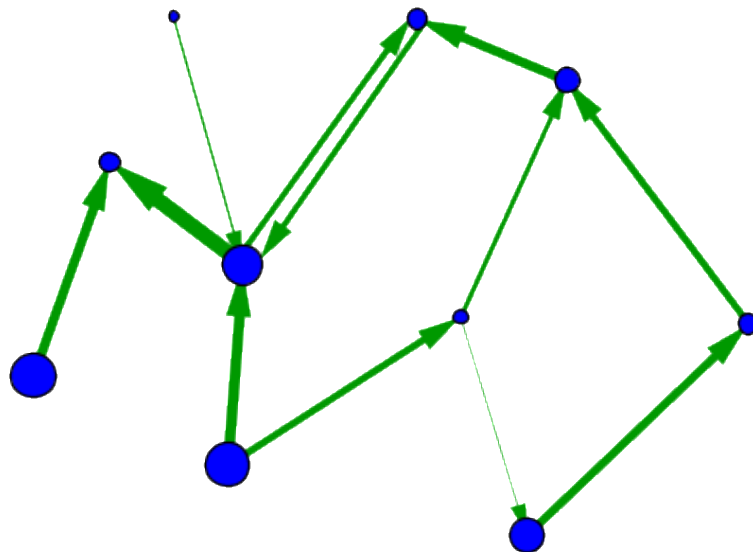
UNDIRECTED

BIG DATA EDGES



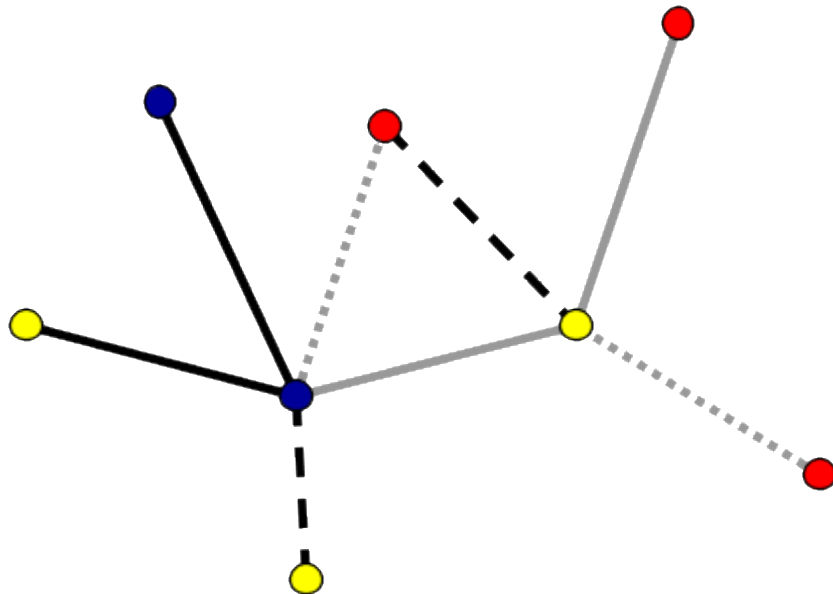
DIRECTED, WEIGHTED

BIG DATA EDGES



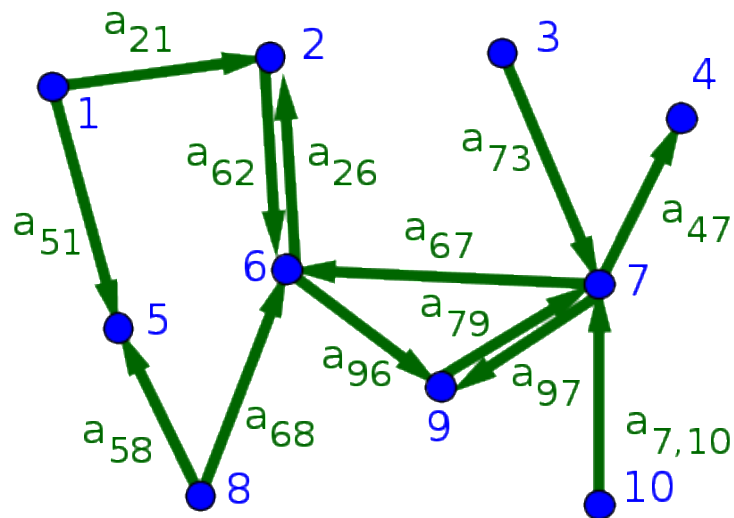
UNDIRECTED, LABELLED

BIG DATA EDGES



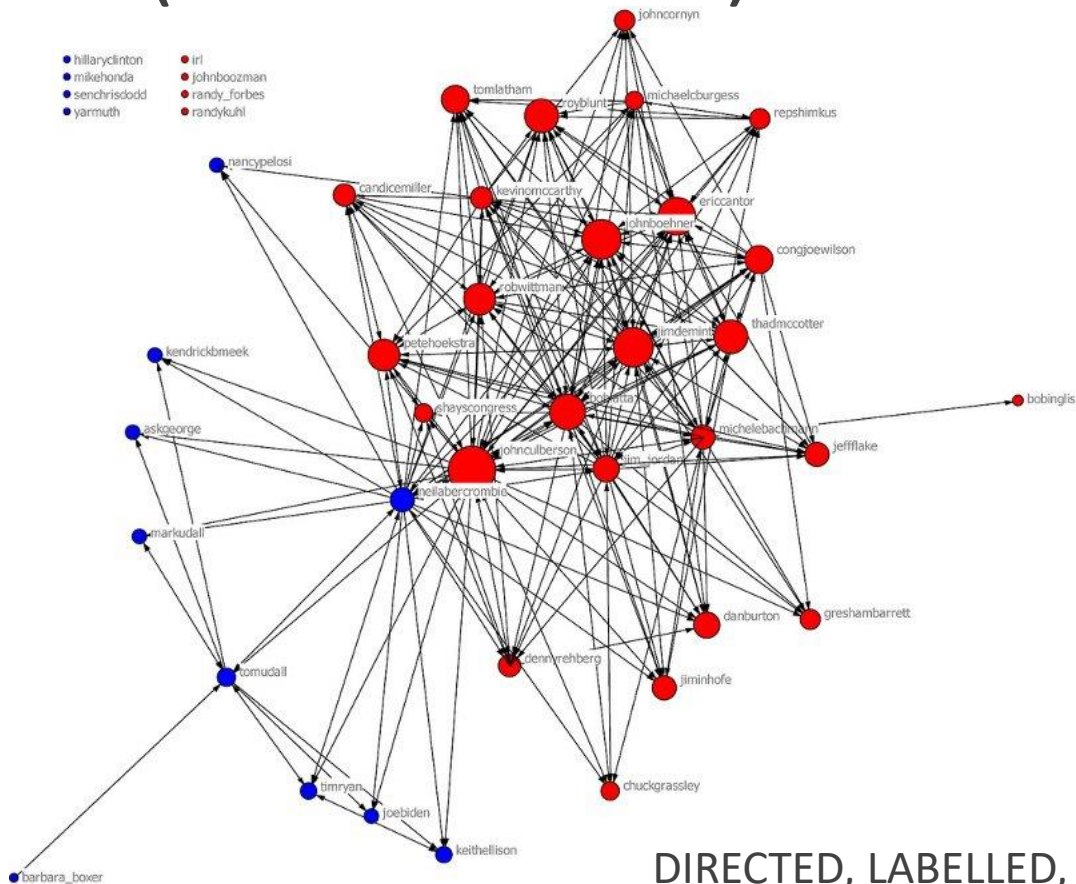
DIRECTED, LABELLED, WEIGHTED

BIG DATA EDGES



TWITTER USERS (US CONGRESS)

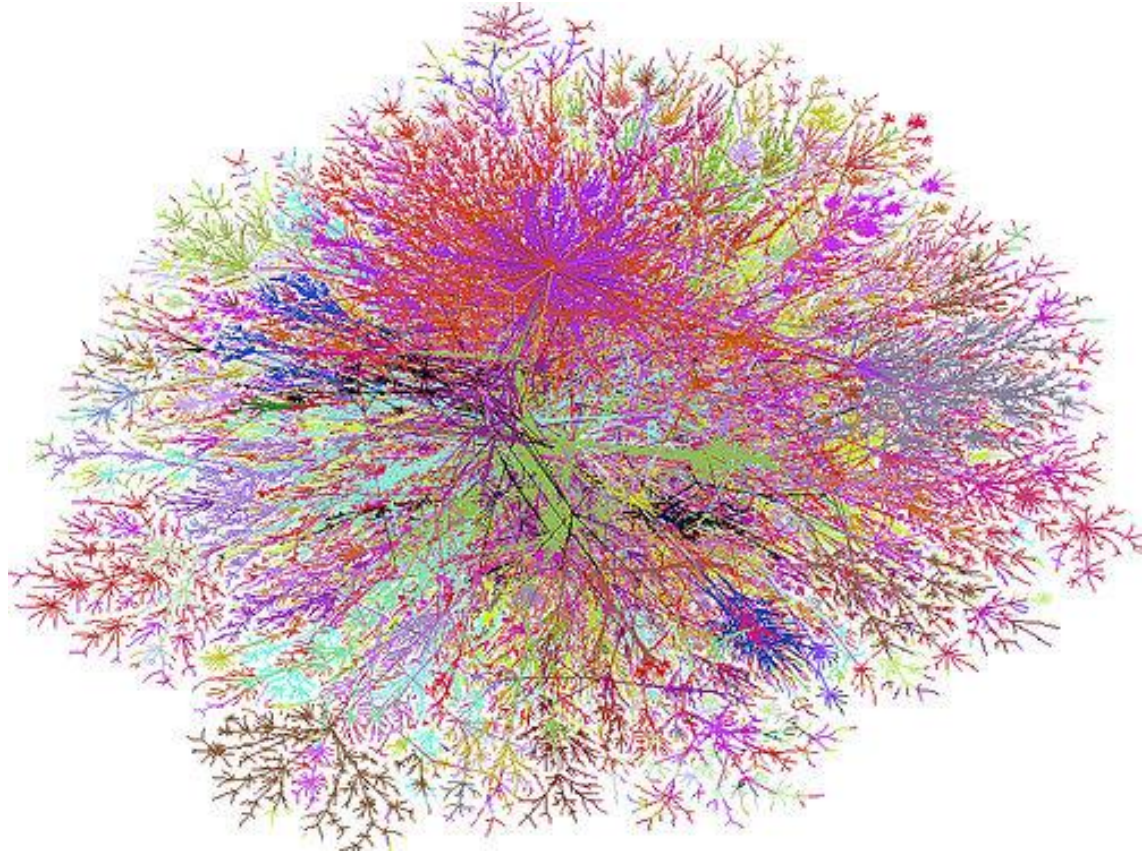
BIG DATA EDGES



DIRECTED, LABELLED, WEIGHTED

INTERCONNECTED DEVICE MAP (2004)

BIG DATA EDGES



INTERCONNECTED DEVICE MAP (2024)

BIG DATA EDGES





BIG DATA HYPERLINKS

HYPERLINK NETWORKS

BIG DATA HYPERLINKS



CITATION

Assesses inbound
links (regression
analysis)



SOURCE



ISSUE

Represent public
debate and links to
support arguments



SENTIMENT



SOCIAL

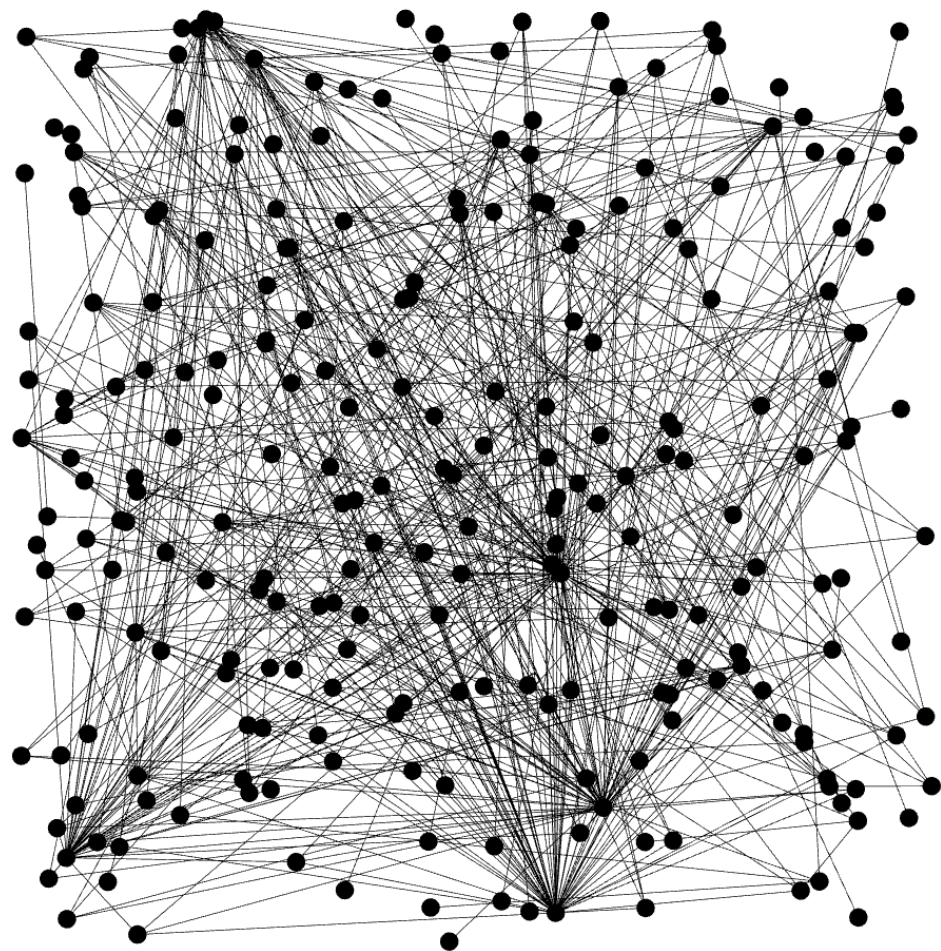
Interdependent
relational analysis



SEMANTICS



BIG DATA WEB CRAWLERS



Internal External Protocol Response Codes URI Page Titles Meta Description Meta Keywords H1 H2 Images Directives AJAX Custom

Filter: All Export

View: Tree Search

Path	Address	Content	Status Code
▼ http/			
▼ www.murdoch.edu.au/	http://www.murdoch.edu.au/	text/html; charset=iso-8859-1	301 Mov
▼ study/			
▼ courses/			
▶ course-details/			
▼ research-students/			
▶ fees-scholarships/			
why-choose-murdoch?cid=MY2...	http://www.murdoch.edu.au/study/why-choose-murdoch?cid=MY2018_HP_BRN_...	text/html; charset=iso-8859-1	301 Mov
▼ undergraduate-students/			
▶ entry-requirements/			
contact-us/	http://www.murdoch.edu.au/contact-us/	text/html; charset=iso-8859-1	301 Mov
▼ Support-Murdoch/	http://www.murdoch.edu.au/Support-Murdoch/	text/html; charset=UTF-8	200 OK
▼ Major-projects/			
Geoffrey-Bolton/	http://www.murdoch.edu.au/Support-Murdoch/Major-projects/Geoffrey-Bolton/	text/html; charset=UTF-8	200 OK
▼ life-at-murdoch/			
?pid=cl_hp_sl1_bb1	http://www.murdoch.edu.au/life-at-murdoch/?pid=cl_hp_sl1_bb1	text/html; charset=iso-8859-1	301 Mov
▼ perth-campus/			
emergencies	http://www.murdoch.edu.au/life-at-murdoch/perth-campus/emergencies	text/html; charset=iso-8859-1	301 Mov
Learning-and-Teaching/	http://www.murdoch.edu.au/Learning-and-Teaching/	text/html; charset=UTF-8	200 OK
▼ Work-Integrated-Learning/			
?pid=bau_hp_sl3_wl2	http://www.murdoch.edu.au/Work-Integrated-Learning/?pid=bau_hp_sl3_wl2	text/html; charset=UTF-8	200 OK
New-students/	http://www.murdoch.edu.au/New-students/	text/html	301 Mov
Kulbardi/	http://www.murdoch.edu.au/Kulbardi/	text/html; charset=UTF-8	200 OK
contact-us	http://www.murdoch.edu.au/contact-us	text/html; charset=iso-8859-1	301 Mov
▼ Contact-us/			
Feedback/	http://www.murdoch.edu.au/Contact-us/Feedback/	text/html; charset=UTF-8	200 OK
▼ docs/			

Filter Total: 415

Overview

Summary

- Total URI Encountered: 500
- Total Blocked by robots.txt: 21
- Total URI Crawled: 479
- Total Internal URI: 415
- Total External URI: 64

SEO Elements

Internal

All (415) (100.00%)
HTML (388) (93.49%)
JavaScript (6) (1.45%)
CSS (1) (0.24%)
Images (12) (2.89%)
PDF (6) (1.45%)
Flash (0) (0.00%)
Other (2) (0.48%)

External

All (64) (100.00%)
HTML (50) (78.13%)
JavaScript (2) (3.13%)
CSS (1) (1.56%)
Images (0) (0.00%)
PDF (0) (0.00%)
Flash (0) (0.00%)
Other (11) (17.19%)

Protocol

All (429) (100.00%)
HTTP (82) (19.11%)

Internal



Name Value

URL Info

Inlinks

Outlinks

Image Info

SERP Snippet



Murdoch
UNIVERSITY

CURRENT STUDENTS [↗](#) STAFF [↗](#) CONTACT US GLOBAL

STUDY

LIFE AT MURDOCH

RESEARCH

INDUSTRY & COMMUNITY

ALUMNI & GIVING

ABOUT US

MENU
≡



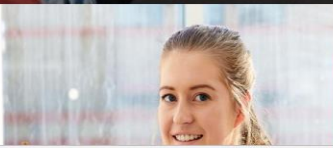
Discover campus life

Explore our unique campus setting and immerse yourself in the vibrant community here at Murdoch.

LEARN MORE

STUDY AT MURDOCH

Our courses combine cutting edge theory with real-world



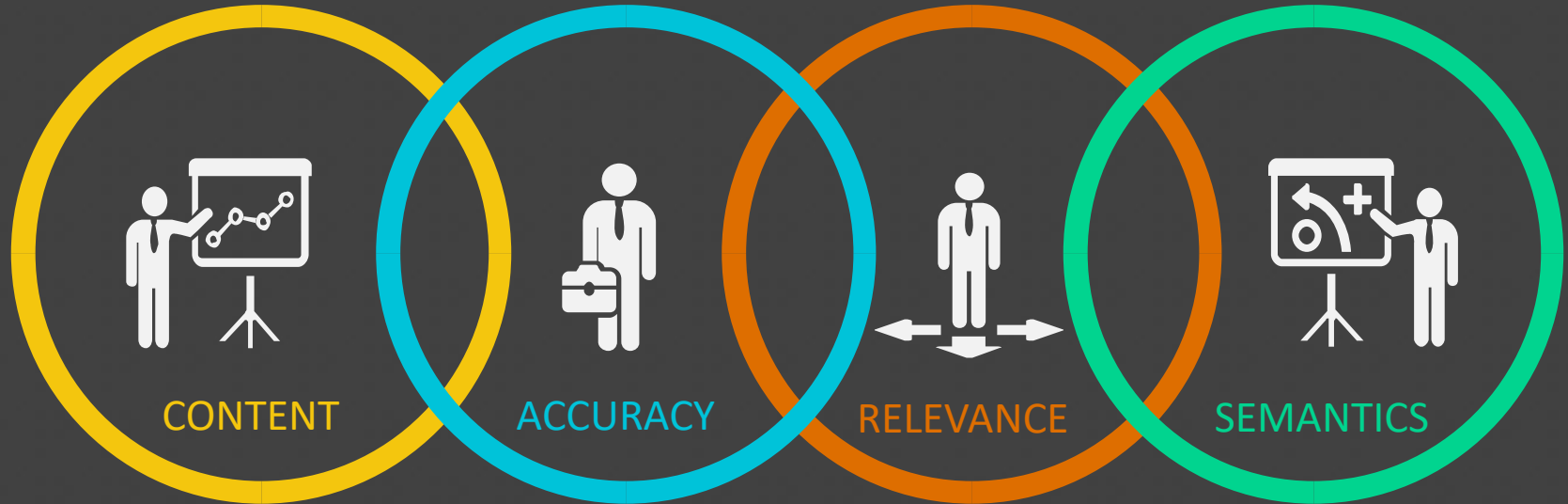
Events for future students

Thinking of starting your first degree?

LIVE CHAT

WEB CRAWLING

BIG DATA WEB CRAWLERS



WEB CRAWLING

BIG DATA WEB CRAWLERS

ACQUISITION

Tools used to acquire datasets (web crawler)

CURATION

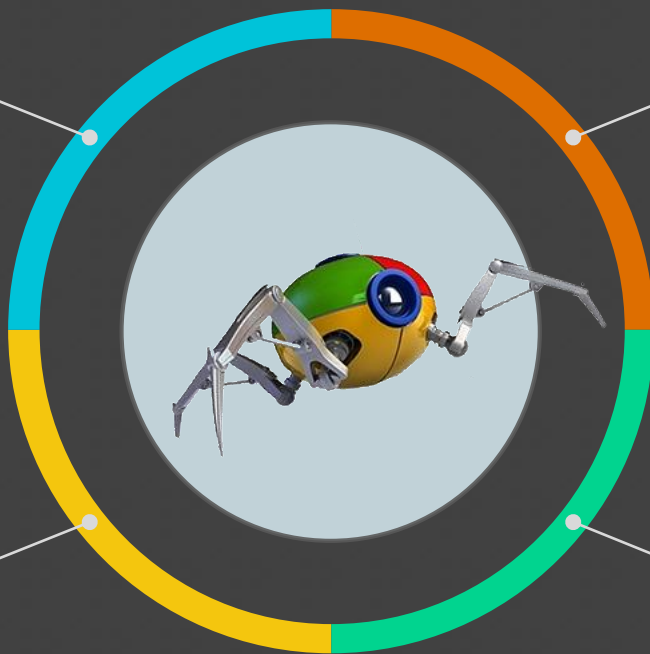
Archival & categorization of collected data (ontology)

MAINTENANCE

Format of crawl data and currency of information

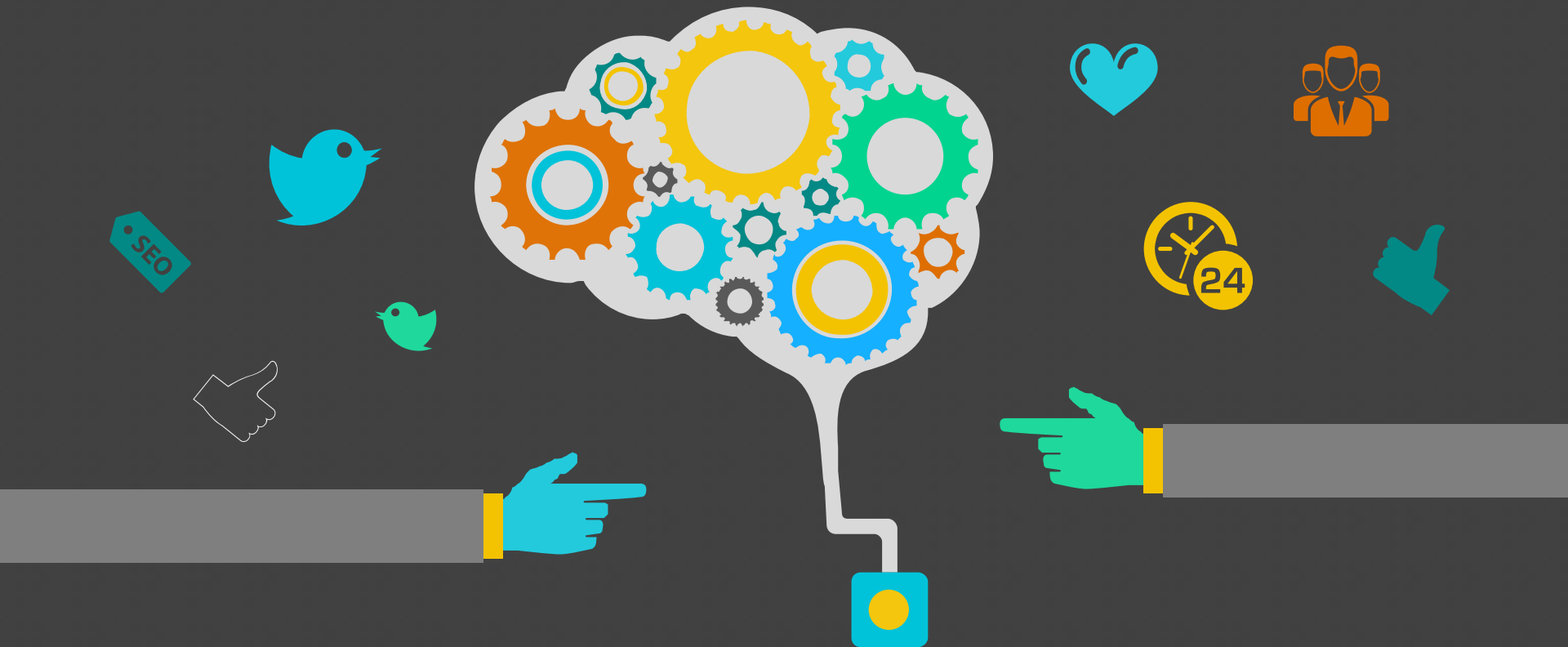
ACCESSIBILITY

Navigation and presentation of data output



WELCOME TO THE INTERNET

BIG DATA: IT'S ONLY THE BEGINNING



RESOURCES

BIG DATA RESOURCES



DATA SCIENCE

Hyperlinks to Data Science resources
you may find useful

VOSON

<http://vosonlab.net/>

KNIME

<https://www.knime.com/>

Wolfram Alpha

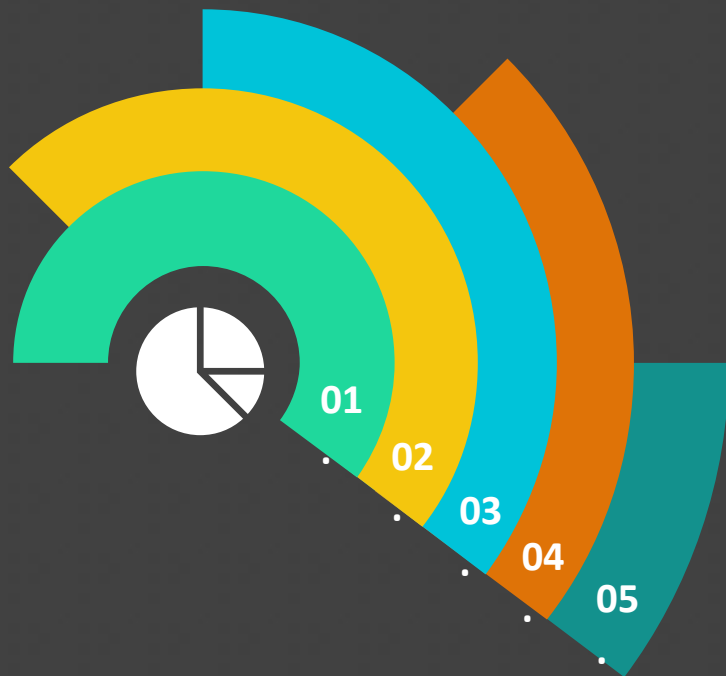
<http://www.wolframalpha.com/>

Metabase

<https://www.metabase.com/>

RESOURCES

BIG DATA RESOURCES



DATA VISUALISATION

Hyperlinks to Data Visualisation resources you may find useful

Chart.js – Javascript Library

<http://www.chartjs.org/>

Data Driven Documents (D3)

<https://d3js.org/>

Raw Graphs

<https://rawgraphs.io/>

BIRT

<http://www.eclipse.org/birt/>

Gephi

<https://gephi.org/>

RESOURCES

BIG DATA RESOURCES



BUSINESS INTELLIGENCE

Hyperlinks to Business Intelligence resources you may find useful

Microsoft Power BI

<https://powerbi.microsoft.com/en-us/pricing/>

Qlik Personal Edition

<https://www.qlik.com/us/products/qlikview/personal-edition>

Tableau Community Edition

<https://public.tableau.com/s/>

THANK YOU!



HAPPY EXPLORING...