**Web Intelligence**

28/09/2021

Lecturer: Dr. Charlie Abela

Focus:

- Collective Intelligence: Knowledge and Information agents

* Data Science: Big data infrastructure and applications, behaviour modelling, social group dynamics
* Human-centric computing: Intelligent human-web interaction, recommender systems
* Knowledge Management: Semantic Web, ontology engineering, linked data, knowledge graphs
* Network Science: Complex network analysis, dynamic networks, sensor networks, internet of things

NLP – Natural Language Processing

Connected World

We live in a deeply connected world

* Even though some connections are not obvious

All sorts of complex systems, whether natural or artificial, tend to exhibit some form of connectivity:

* Interacting molecules
* Genes
* Neurons
* Humans and social systems
* Words
* Sensors

Data Stored in Databases

mongoDB

* Cross-platform document-oriented database
* Container for collections
  + Each collection is a group of documents
* A document is a set of key-value pairs defined in JSON
  + Documents have a dynamic schema
    - Documents in the same collection do not need to have the same set of fields or structure
    - Common fields in the documents may hold different types of data

Sharding – Distributing amounts of data

Data as Graphs

A graph *G* is a tuple (*V,E*) of vertices *V* and edges *E*.

* Each edge *e* = *E* is said to connect two vertices *u, v* and is denoted as *e* = (*u, v*)
* V(G) -> collection of vertices
* E(G) -> collection of edges

In order to help us understand graphs more visual aids are important, i.e. colour, key, etc.

Other Graph Applications

* Traversal and path exploration
  + Planning and routing of internet traffic and transportation
    - Shortest paths, centrality measures
* Bipartite matching
  + Item recommendations
  + Mapping Protein-Molecule bindings
    - Maximum Matching
* Identify “important” nodes and communities
  + Spreading of information in social networks and/or spread of an epidemic
    - PageRank, Connected Components, Label Propagation
* Pattern mining
  + Finding frequent subgraphs in IoT data

**Graph Theory**

Types of Graphs

* A graph which does not have loops or multiple edges such as graph *G* is referred to as simple.
* *G*’ is referred to as complex graph.
* *G* and *G’* are both undirected, since edges do not exhibit a direction.
* A graph in which the edges exhibit a direction is called a directed graph
* *G* is a simple, directed graph.
* A graph can have edges that contain a certain weight to represent an arbitrary value such as cost, distance, quantity etc.
* In this case
* A simple graph that has *n* vertices, with ach vertex being adjacent to every other vertex is known as a complete graph.
* A simple graph that has *n* vertices, is acyclicand exhibits a hierarchical structure, is called a tree.
* A tree with a designated root node is a rooted tree.
* A DAG is a directed acyclic graph no cycles.
* These graphs are important when representing structures with dependencies.
* All out-trees are DAGs but not all DAGs are out-trees.