Lesson 7

Graph Analytics

Data store Graph model

- Graph data store provides additional information about associations
- Graph analytics thus enables analysis of additional properties, besides the ones using standard RDBMS or data warehouse framework

Centralities Parameters in Graphs

- Computations of the centralities of entities in terms of
- in-degree in nodes of directed graphs
- out-degrees
- distribution of degrees among the nodes
- Refer Example 8.7 for understanding computations

Node Centrality of a node

- Defined in reference to other nodes using the metrics
- Metrics for centrality: Degree, closeness, betweenness or other characteristic of the node, such as rank, belief, expectation, evidence, reputation or status.

Path and Flow Analysis Algorithms

- Analyze the shapes (triangles, hexagons, trees and junction trees), shortest paths and top-K shortest paths
- Algorithm for triangles counting finds the number of triangular relationships among the nodes

Path and Flow Analysis Algorithms

- Top-K shortest paths means finding distances of the multiple paths which connect the vertices and which have the shortest paths among top K
- Value of K is 2, 3, 4 and so on for top-2, top-3, top-4 and so on.

Matching and Search Analytics in Graphs

- Matches the graphs and subgraphs after a graph search on path traversals
- A filter algorithm uses the label, vertex-property, edge-property or geographical location for filtering the graph vertices

Collaborative Filtering Algorithms

- Does the searches
- Finds matches in a bipartite weighted graph
- Used by recommender systems (Section 6.4.3)

Detection of Clusters

- Identifies the groups of special cases
- Example: Study of effect of discount offered in versus sales, the identification of clusters enables the price discovery

Detection of Clusters

- Example: A system identifies a cluster of students with deep interest in Big Data analytics
- This enables a department to start additional new course in that area

Detection and Analysis of Patterns

- Required in many applications, for example, identifying patterns of sales increase after an advertisement of a car model, for planning future advertisement strategies
- Finding new opportunities (for example in education, business or health care)
- Knowledge Discovery

Anomaly Detection

- Find the abnormal behaviour, structure, feature, content or semantic features
- Help in its usability and summarising anomaly attributes
- Enables identification of spam source
- Detection of frauds related to credit card, medical claim or fictitious transactions.

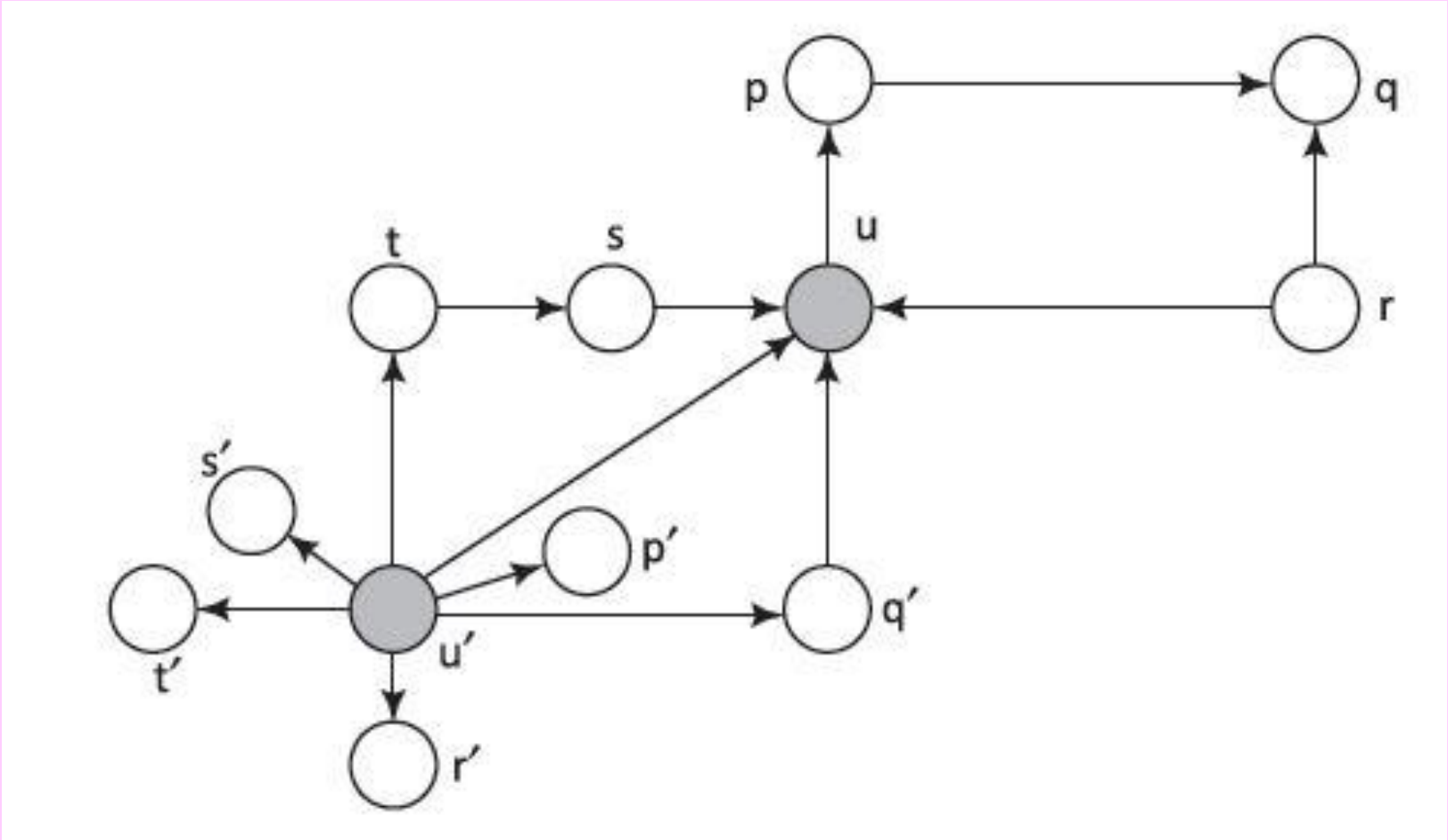
Community And Network Analysis

- Analyses the close-by entities, and fully mesh-like connected sets
- The network graph analysis besides the centralities analysis, also find page rank of the links

K-neighbourhood and K-Core analysis

- Finding the number of 1st neighbour nodes, 2nd neighbour nodes and so on. (K = 1, 2, 3, 4 and so on)
- Finding core, which may consist of a triangle of connected vertices
- A rectangle with interconnected edges and diagonals

Figure 8.7: A multi-directed graph with 12 vertices (p, q, r, s, t, u, p′, q′, r′, s′, t′ and u′).



StatsModel and Probabilistic Distribution analysis

- Inter-relationships between the entities
- Learning of graph structure
- Computations of conditional independence, and probabilistic inference, evidence, belief, expectations or influence as a function of distances

Use of Junction Trees Graph (JTG) formed for the Bayesian

- Bayesian network graph reduced to a JTG using a junction tree algorithm (JTAl)
- JT has a root node at a junction, which has number of directed edges to a set of junctions (daughter nodes)

Junction Tree

- Each JT root daughter is again a junction, which has number of directed edges to a next set of junctions
- Thus, a tree-like structure exists starting from the root

JTAl (Clique tree method)

- A machine learning algorithm
- Extracts marginalization in general graphs
- Propagates belief (evidence collected at the junction from number of connected nodes in the junction tree, also called inference from that tree).

JTAl

- Does the belief computations for each junction tree
- Usage of JTG eliminates the cycles in traversals also
-

Choosing Graph Analytics

1. Finding Connectivities from the number of relationships and association types
2. Finding undirected unidentified patterns, undirected graph path traversal, discovering new unidentified pattern (for example, finding students of new batches opting for Natural Language Processing subject

Choosing Graph Analytics

3. Missing pattern (for example, finding none of computer science student interested in financial analysis in an organization)
4. Discovering new knowledge [for example, evolving interest in Big Data Analytics course in students of UG Computer Science (Example 8.5)

- on events

Choosing Graph Analytics

5. Analysis using the ad hoc queries for finding reasons
6. Predicting interactive performance, which means predicting interactions among entities on events

Graph Analytics Use Cases

1. Monitoring and analysis of social media
2. Analysis of an enterprise social network: The analytics enables search of expertise, knowledge, recommending expertise, experts location, and detecting of spam and anomaly

Graph Analytics Use Cases

3. Financial analysis for undertaking the financial decisions, new knowledge discovery, security analysis, anomaly and fraud detection, such as credit card frauds and fake bill payments, account manipulations
4. Commerce and trade analysis for customer behaviour prediction

Graph Analytics Use Cases

5. Planning and strategies for sales promotion, price discovery, detecting weak supply chain links and commerce frauds
6. Analysis of an enterprise social network for finding expertise, knowledge recommending expertise, experts location

Graph Analytics Use Cases

7. Detecting of spam and anomaly
8. Disease diagnostics, patient and disease analytics in healthcare studies, health care quality analysis, medicine research and development in genomics

Summary

We learnt:

- StatsModel and Probability Based Graph Analytics
- Centralities Parameters based Analytics
- Path and flow analysis algorithms
- Propagation of inference, evidence, belief, expectation or influence as a function of distances

Summary

We learnt:

- Matching and Search Analytics
- Collaborative Filtering
- Anomaly Detection
- Detection and Analysis of Patterns
- Community, Micro-communities, triangles in Network Graphs

Summary

We learnt:

- Choosing Graph Analytics
- Graph Analytics Use Cases
-

End of Lesson 7 on

Graph Analytics