# UNIT-8 [Around 5 marks]

## Graph Mining and Social Network Analysis:

### ⊛Graph Mining:

Graph mining is the process of extracting patterns (sub-graphs) of interest from graphs, that describe the underlying data and could be used further, e.g, for classification or clustering. Graph mining has vast number of applications such as circuits, chemical compounds, protein structures, social networks, Web, and XML documents. Graph mining is used for Fraud Detection, Community/Cluster detection, Recommending friends, Finding Influential Nodes.

### ⊛Graph Mining Algorithms: [Imp]

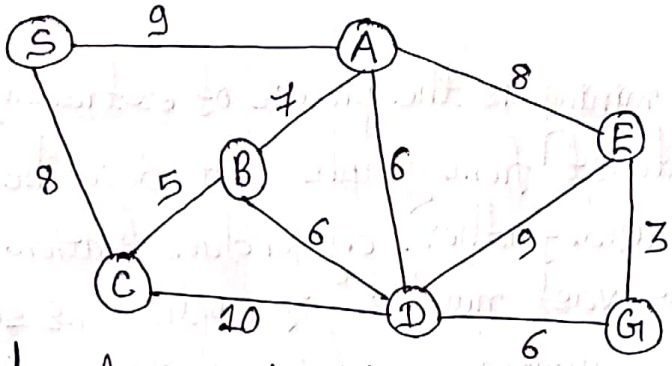
**1) Beam Search:** Beam Search was developed in an attempt to achieve the optimal solution without consuming too much memory. Beam Search is heuristic approach where only the most promising $\beta$ nodes (instead of all nodes) at each step of the search are retained for further branching. $\beta$ is called Beam Width. Beam search is an optimization of best-first search that reduces its memory requirements.

**Algorithm:**

Open = {initial state}
while Open is not empty do
1. Remove the best node from Open, call it n.
2. If n is the goal state, backtrace path to n and return path.
3. Create n's successors.
4. Evaluate each successor, add it to Open.
5. If |Open| > $\beta$, take the best $\beta$ nodes and remove the others from the Open.

done.

**Example:** Trace beam search for the following graph. Assume beam width=2.




Heuristic value to G1 (goal node) from other nodes as given below:

S→G1=12 A→G1=6 B→G1=8 C→G1=10 D→G1=4 E→G1=5

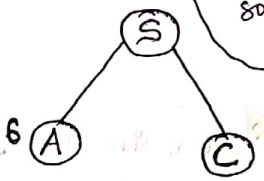
**Solution:**

**Step1:** Open={S}

Start node S लाई लिएको S मा A र B connected छन् so, heuristic value सहित लेखेको

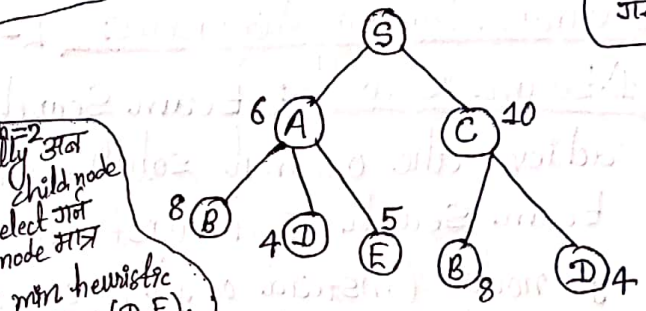
**Step2:** Open={A,C}

S लाई its child nodes ले replace गरेको

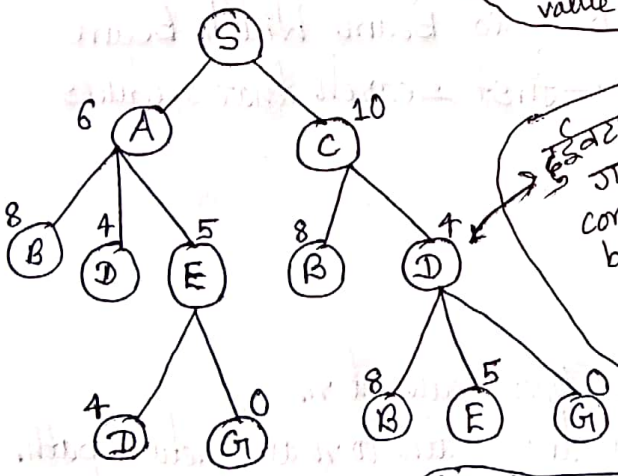



Beam width 3 भनेको भए 3 वटा लिन्थ्यौं।

Since Beam width=2 अब Beam Search actually लाग्छ, 2 भन्दा बढि child node भएकोले 2 मात्र select गर्ने 2 most promising node मात्र select हुने भए min heuristic value को आधारमा so(D,E).




**Step3:** Open={D,E}




हुँदैवटा D द्वय जुलसुकेमा गर्दा कि हुन्छ। D सँग connected A,B,C,E,G छन् but A र C already as parent node आएकोले B, E, G मात्र लिन मिल्यो Similarly for E.

G goal node भएकोले यसको heuristic value 0 हुन्छ

**Step4:** Open={G,D}

most promising nodes are G,D from Step3

Goal Found

most promising node मा G आइसक्यो goal node so stop

Solution Path = S→C→D→G

Path Cost = 8+10+6 = 24

Path को value question को fig मा हेर्ने

Most promising nodes मा (G,D) मा D द्वय so, S,A,E,G path मा D भन्दा neglect गरि D भएको S,C,D,G path मै choose गरेको

## 2) Inductive Logic Programming:

Inductive Logic Programming (ILP) is a research area formed at the intersection of Machine Learning and Logic Programming. ILP systems develop predicate descriptions from examples and background knowledge. Induction is reasoning from the specific to the general whereas deduction is reasoning from general to specific. In deduction we use rules and some facts to deduce more facts whereas in induction we derive rules from set of facts. Thus we can say that goal of ILP is to find hypotheses expressed in terms of logic programming clauses from a set of positive and negative examples and domain knowledge.

**Example:** Given the following facts:

→ read as a is father of c

```
parent(a,c)    parent(b,c)    father(a,c)
mother(b,c)    male(a)        female(b)
```

Goal of ILP is to learn following types of rules from the given dataset.

```
father(x,y) ← parent(x,y) & male(x)
mother(x,y) ← parent(x,y) & female(x)
```

→ दिएको dataset अनुसार ILP ले यसरि rule हरु learn गर्छ भनेर देखाउन मात्र खोजेको

We can represent graphs in catalog facts, then we can use ILP to identify interesting patterns in the graph.

## ⊛ Social Network Analysis:

Social network analysis (SNA) is the process of investigating social structures through the use of networks and graph theory. It characterizes networked structures in terms of nodes (individual actors) and the edges or links (relationships) that connect them. Data Mining techniques can assist effictively in dealing with the three primary challenges with social media data:

→ First, social media data sets are large.
→ Second, Social media site's data sets can be noisy.
→ Third, data from online social media platforms are dynamic.

Recently, link mining is becoming a very popular research area not only for data mining and web mining but also in the field of social network analysis. Many researches are focusing on

developing new link mining techniques and algorithms. By considering links, more information is made available to the mining process. This brings about several new tasks as discussed below:

Link-based object classification: Link-based classification predicts the category of an object based not only on its attributes, but also on its links, and on the attributes of linked objects.

Object type prediction: This predicts the type of an object, based on its attributes and its links, and on the attributes of objects linked to it.

Predicting link existence: We may want to predict whether a link exists between two objects.

Link cardinality estimation: We may wish to predict number of in-links and out-links in a node.

## ⊛Friends of Friend:

Friends of friend are indirect connections in graph representation of a social networks. This type of analysis is done in social networks to recommend friends. Person C is a friend of a friend of person A when there is a person B that is a friend of both A and C.

## ⊛Degree Assortativity:

Assortativity, or assortative mixing is a preference for a network's nodes to attach to others that are similar in some way. For instance, in social networks, nodes tend to be connected with other nodes with similar degree values. This tendency is referred to as assortative mixing, or assortativity. This means, degree assortativity characterizes the tendency for large-degree nodes to connect to other large-degree nodes and low-degree to low-degree. Social networks are typically thought to be distinct from other networks in being assortative.

## ⊛Signed Network:

In a social network analysis, a positive or a negative friendship can be established between two nodes in a network; this results in a signed network. As, social interaction between people can be positive or negative, so can be links between the nodes. When a positive or a negative value is attributed on the relationship between the two nodes, it is called a user evaluation. In social groups, people can like or dislike, respect or disrespect other people in their social groups.

[Imp] **Structural Balance Theory:** Structural balance considers the possible ways in which triangles on three individuals can be signed. This theory suggests that triangles with three positive signs (three mutual friends) and those with one positive sign (two friends with a common enemy) are more possible and hence should be more prevalent in real networks. Such triangles are called balanced. Triangles with two positive signs (two enemies with a common friend) or none (three mutual enemies) are not possible. Such triangles are called unbalanced.

[Imp] **Theory of Status:** A signed link from A to B can have more than one possible interpretation, as higher status and lower status. In this theory of status, we consider a positive directed link as having higher status; and a negative directed link as having lower status. These relative levels of status can then be propagated along multi-step paths of signed links, often leading to different predictions than balance theory.

**Conflict between the theory of balance and status:** [Imp]

Consider the situation in which user A links positively to a user B, and B in turn links positively to a user C. If C then forms a link to A, what sign should we expect this link to have?

→ Balance theory predicts that since C is a friend of A's friend, we should see a positive link from C to A.

→ Status theory on the other hand, predicts that A regards B as having higher status, and B regards C as having higher status, so C should regard A as having low status and hence be inclined to link negatively to A.

In other words, the two theories suggest opposite conclusions in this case.

## ⊛ Trust in a Network: [Imp]

Web of trust is used in network to express or predict trust/distrust between users. The problem is to determine trust values for the remaining user pairs using only those which are explicitly specified. There are various ways to infer trust in networks: Atomic propagation, Propagation of Distrust, and Iterative propagation.

**Atomic Propagation:** The atomic propagations are a "basis set" of techniques by which the system may infer that one user should trust or distrust another. The basis set is as follows:

**Direct propagation:** If i trusts j and j trusts k, then we could conclude that i trusts k.

**Co-citation:** If i1 trusts j1 and j2, and i2 trusts j2, then we can conclude that i2 should also trust j1.

**Transpose trust:** Here i's trust of j causes j to develop some level of trust towards i.

**Trust coupling:** When both i and j trust k, this implies that i trusts j.

**Propagation of Distrust:**

The following are three models for the propagation of trust and distrust, given initial trust and distrust matrices T and D respectively:

**Trust only:** By completely disregarding distrust scores and propagating only the trust scores.

**One-step distrust:** When a user distrusts someone, they discount all their judgements. Therefore, distrust propagate only one step.

**Propagated distrust:** When trust and distrust both propagate together, we get: $B = T - D$. Here, B is set of trust and distrust.

## Iterative Propagation:

The end goal is to produce a final matrix F that has the trust or distrust between any pair of users in this universe. There are two approaches to computing F from the sequence of propagations: Eigenvalue Propagation and Weighted Linear Combination.

## ⊛ Predicting Positive and Negative Links:

In social networks relationships can be either positive or negative. Many researchers have proposed algorithm based on the trust and distrust propagation to predict trust and distrust relations. Machine learning algorithm for predicting positive and negative links in social networks inspired by structural balance theory and social status theory. A low-rank matrix factorization approach with generalized loss functions is proposed to predict trust and distrust relations.