COL776:Probabilistic graphical models
Assignment 1

Summary on Bayes-Ball: The Rational Pastime by Ross D.Shachter

1. **Brief Summary**

   Paper discusses about a new algorithm developed to find the requisite information needed to calculate probabilities of inference queries.The name of the algorithm to determine requisite information and irrelevant information is given as bayes ball as the algorithm composes of running a ball from bayesian graph nodes and the time complexity of this algorithm is linear in the size of bayesian graph.

2. **Introduction**

   Main advantage of representing a problem in bayesian network is we can recognize many important properties of the graph easily.To fully specify a network, the possible states and probability distributions must be obtained for all variables. For a particular inference query or decision problem, only some of that information is needed.Bayes ball algorithm helps in identifying conditional independencies(irrelevance) and requisite information easily and efficiently.

3. **Notation**(only Important ones are described)
   a. **Structured belief network B** = (N; A,F) consists of nodes N and directed arcs A which together form a directed acyclic graph G(N,A).F is set of deterministic nodes
   b. **Directed Factorization:**A joint probability distribution over X_N is said to admit a directed factorization with respect to a structured belief network B = (N, A, F) if X_j is a deterministic function of $X_{Pa(j)}$ for all j $\in$F and $Pr(X_N = \prod_{j \in N} Pr(X_j | X_{Pa(j)})$
   c. **Irrelevance**:There are two type of irrelevancies are there one **probabilistically irrelevant** (independent) and **irrelevant**.The main difference is first is from the mathematical view and second will follow from belief network.It is important to understand that probabilistic irrelevance is not the same as the irrelevance represented by a belief network
   d. Requisite Nodes
      i. Requisite Probability Nodes:The requisite probability nodes for J given K, denoted $N_p(J|K)$, are those nodes for which conditional probability distributions (and possible states) might be needed to compute $Pr(X_J | X_K)$.

      ii. Requisite Observed Nodes:The requisite observations for J given K, $N_e(J|K) \subseteq K$, are those observed nodes for which observations(and hence the possible states which might be observed) might be needed to compute $Pr(X_J | X_K)$
   e. **Active Path**:An active path from J to L given K is a simple trail (or undirected path) between i $\in$ L and j $\in$ J, such that every node with two incoming arcs on

the trail is or has a descendant in K; and every other node on the trail is not functionally determined by K.

f. **D-separated**:Given sets of nodes, J, K, and L from belief network B, K is said to D-separate J from L in B if there is no active path from J to L given K.

4. **Bayes Ball Algorithm**:

Statement: Given K finding requisite and d-separated nodes for J.

Data Structure:Node j,with boolean variables **visited**,marked on top($\mathbf{j_{top}}$), bottom($\mathbf{j_{bottom}}$)(all initialized to false),Schedule queue **S**.

Algo:

I. Put all nodes in J to S as visited from children

II. While there are still nodes scheduled to be visited:

A. Remove any node $j \in S$

B. Mark j as visited.

C. If $j \in K$ and the visit to j is from a child:

1. if j_{top} not marked, then mark j_{top} & Keep j's parents in S;

D. If the visit to j is from a parent:

1. If $j \in K$ & j_{top} not marked then mark j_{top} & Keep j's parents in S;
2. If $j \notin K$ & j_{bottom} not marked then mark j_{bottom} & Keep j's children in S;

III. **irrelevant nodes** = j_{bottom} not marked nodes;
requisite probability nodes= j_{top} marked nodes;
requisite observation nodes = nodes in K and marked as visited.

5. **complexity:** As we are marking visited from child and from parent i.e at top and bottom at max the step of marking it again adding it's parents/child will take max 2 times.So, the time complexity of the algorithm is $O(n+A_v)$ where n=number of nodes,A_v are edges incident on nodes marked during algorithm .So, this A_v can be at max number of edges So, at max O(n+m) where m is number of edges.