algorithm approximation association betweenness bound collection data database dataset definition different distribution error estimation fi frequency frequent graph itemsets large lemma log mining number obtain pages pair Parma paths presented probability problem Queries random range relative results rules Sample selectivity shortest Size space tables transactions values vc VC-dimension vertices work