# **Assumptions and Considerations**

While a conscious effort is made to make the solution as complete as possible, there are certain assumptions and considerations that have been made. The reasons for making such assumptions stem from various constraints such as time, resource availability and current technology. The following are the assumptions and considerations made and the future scope in which these can be removed.

The primary source of data for the system is the DBLP dataset. It has been assumed that the data contained within it is valid and verified. However, while all the publications can thought of as valid publications, there exists ambiguity about authors with similar names and initials. The xml schema is such that there are optional attributes which are otherwise extremely important. This some entries to miss out on information which would be extremely useful to certain use cases. Unique Ids assigned to authors, institutions and publication channels would also have removed ambiguity and made the end result more complete.

As pointed out in the previous case, with regards to missing attributes, there is a lack of overall data as well. The dataset worked with is assumed to be a complete list of all publications in computer science. While it is very significant but misses out on many publication channels due to unavailability of data from a few important publication channels like ACM. This normally would be a minor problem, however, as the trust score is being calculated, missing data could affect this final output.