**Motivation**

People can showcase their knowledge within a particular field in [www.stats.stackexchange.com](http://www.stats.stackexchange.com/)

The quality of the answers can easily be seen from the rating given by other users. Users ‘up vote’ an answer if they find it helpful or interesting.

Almost all the papers we read were focused on the users – how to find out who are the most influential users in network based on HIST or PageRank or how to predict who is going to be influential user in network based on first few months’ performance in network. Overall, the main idea of the papers is to identifying experts in network – that is what we did already (We used degree of users in order to identify this users – this approach is not in the papers as well (I think because it is too easy for papers)).

**PLAN A**

We decided to look on the problem from different perspective, instead of looking at users we focus on the answers they provide and we would like to describe how good answer look like. Good answer is defined as accepted answer and answer with high votes.

Why does it matter?

One very obvious benefit is that all those high-quality answers are archived in the website. Users can use those answers as real proofs that they have expertise in the field. Thus, it can improve their career prospect by showing those proofs in their CV. Besides that, people in the community will also recognize you as an expert within the community. (need reference of this assumption?)

However, it is not easy to provide a high rating answers. And it’s getting more difficult for the answer to be the accepted ones. In this study, we want to understand the factors behind an accepted answer with high rating.

**Methodology**

Part 1: Users Sampling

* Accepted answer = an answer that is accepted for a question
* Answer provider = user who has at least one accepted answer
* Answer seeker = user who asks at least one question with an accepted answer
* Out-degree = measurement for number of accepted answers of a user
* We look for top 10% answer providers with highest out-degree values as the sample. (need reference of this method?)

Part 2: Feature Selection (sentiment analysis, specific words, time when the answer)

* We concentrate to work only with the answers from the sample.
* We list all possible features that are reasonable to drive the rating of an accepted answer
* We perform regression to understand how significant those features are in driving the rating
* For all significant features, we only care about features with positive effect

**PLAN B**

We will focus on the answers and not users again. We would like to find out whether the answer will be accepted or not.

Features: length of answer, reputation of answer provider, responsiveness, users’ accepted answer ratio, sentiment of answer, out-degree of the user, in-degree of the user, sentiment of question, votes of answer, votes of question

Why to look at this?