Tag Recommendation in Stack Overflow

Mentor:
Suman Kalyan Maity
Prithwish

Submitted By:
Group 26
[Shreya Chakraborty(15IT60R13)
Nitya Tandon(15IT60R01)
Gyanendra Singh(15IT60D03)
Pooja Kokane(15IT60R02)
Priya Shree(15IT60R19)
Chandra Bhanu Jha(15BM6J12)]

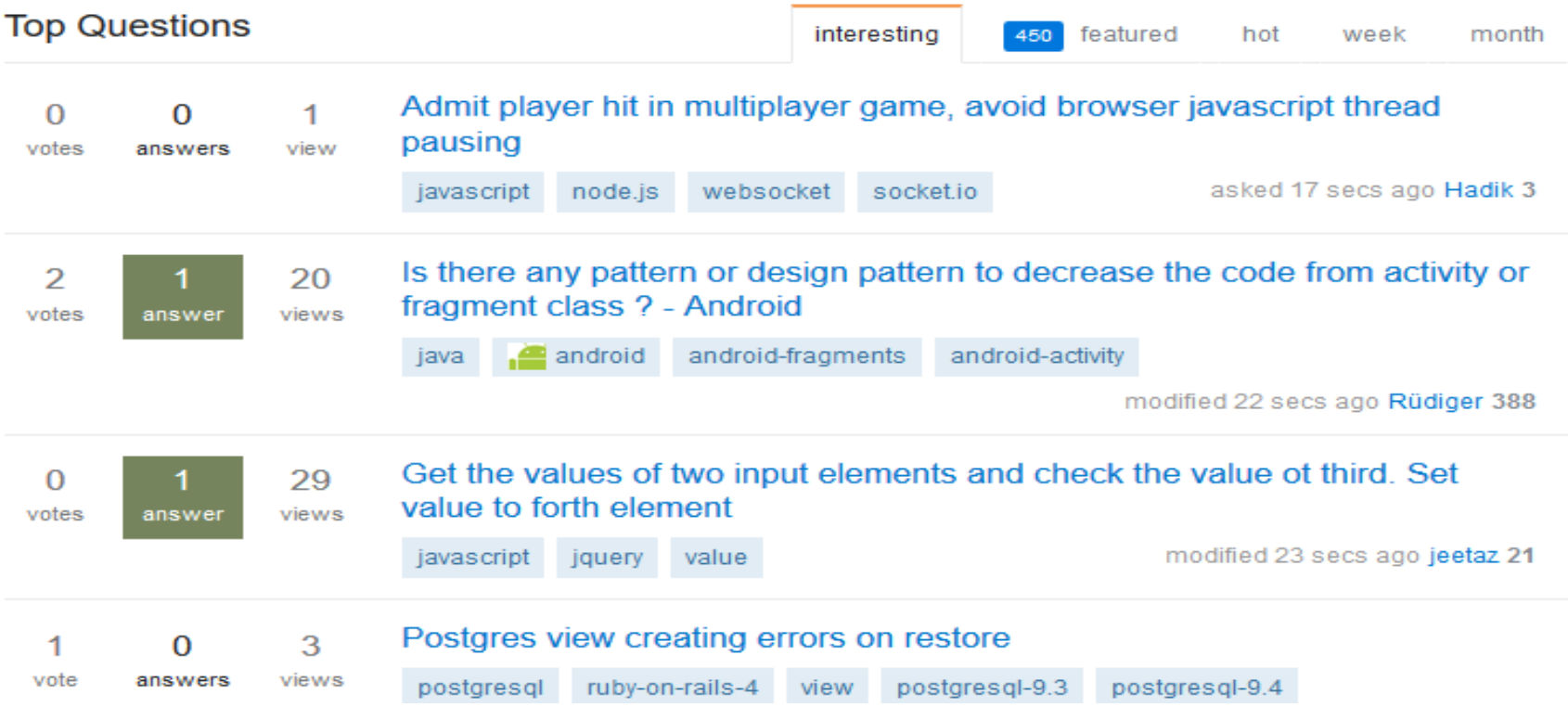
INTRODUCTION

- **People share ideas and experiences through sites like Stack Overflow, Ask Ubuntu, Ask Different and Free Code**
- **Tags help in searching this information (posts) in software information sites**
- **To improve quality of tags, related tags can be recommended to users**
- **Dataset from Stack Overflow has been used for this project to implement EnTagRec algorithm**

MOTIVATION

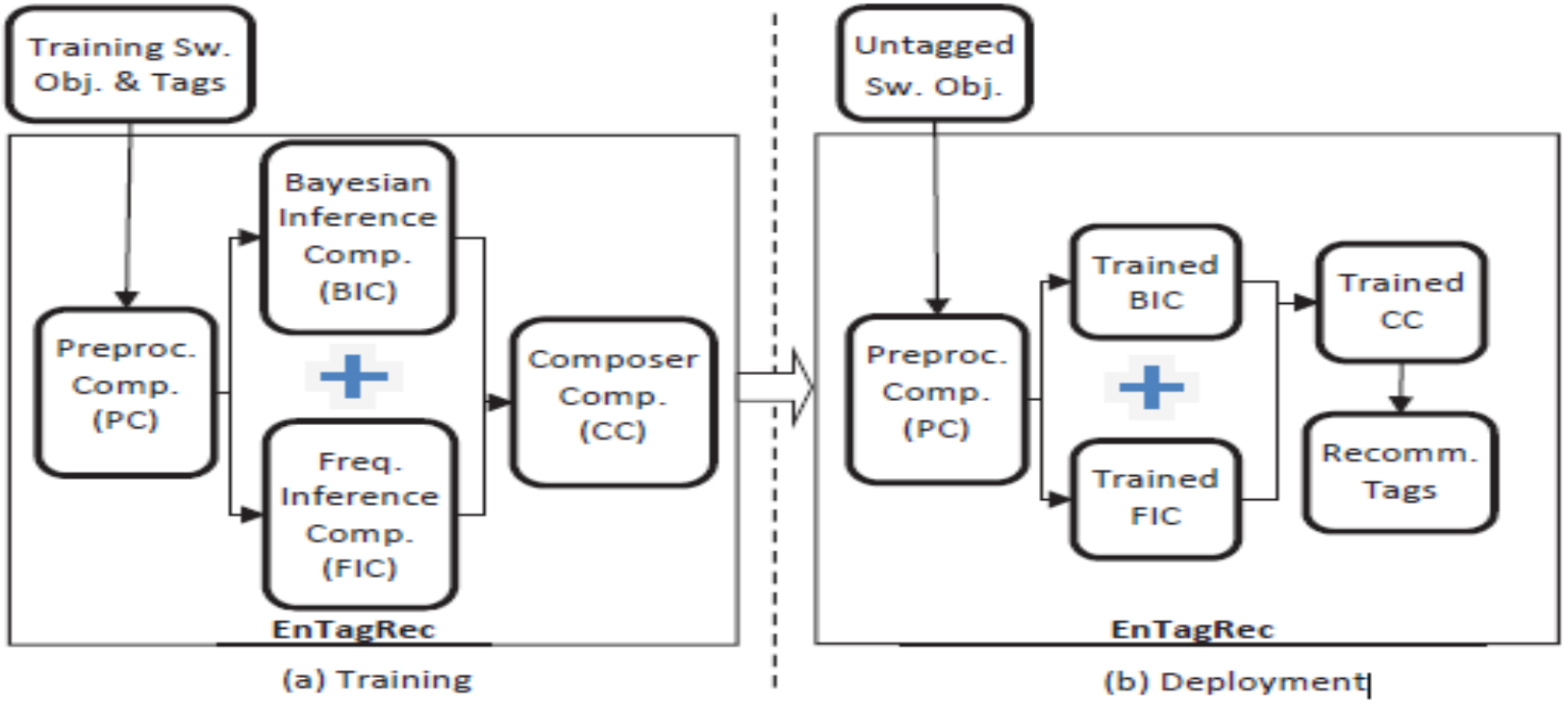
- **Tag Recommendation system**
  - Help users select appropriate tags easily and quickly
  - In-time help homogenize the entire collection of tags such that similar objects are linked together by common tags more frequently

DATASET



Posts and corresponding tags on Stack Overflow

APPROACH



Two Phases of the EnTagRec Algorithm

PREPROCESSOR COMPONENT (PC)

- **Each software object (questions, answers) are converted into a bag of words**

- **This bag of word then goes through the following process:**
  - Tokenization
  - Stop Word Removal
  - Stemming: Snowball Stemmer
  - All words having occurrences less than 20 and tags having occurrences less than 50 were excluded from the dataset

BAYESIAN INFERENCE COMPONENT (BIC)

- **Models software objects as a probability distribution of tags**

$$P(t|w_1 \dots w_n) = \frac{P(w_1 \dots w_n|t) * p(t)}{P(w_1 \dots w_n)}$$

- **Also, a tag is represented as a probability distribution of words appearing in the software objects**

- **BIC uses Labelled Latent Dirichlet Allocation (L-LDA)**

- **L-LDA is a supervised learning technique which works on the topical model**

- **It gives the probability distribution of topics (tags) for a software object (post)**

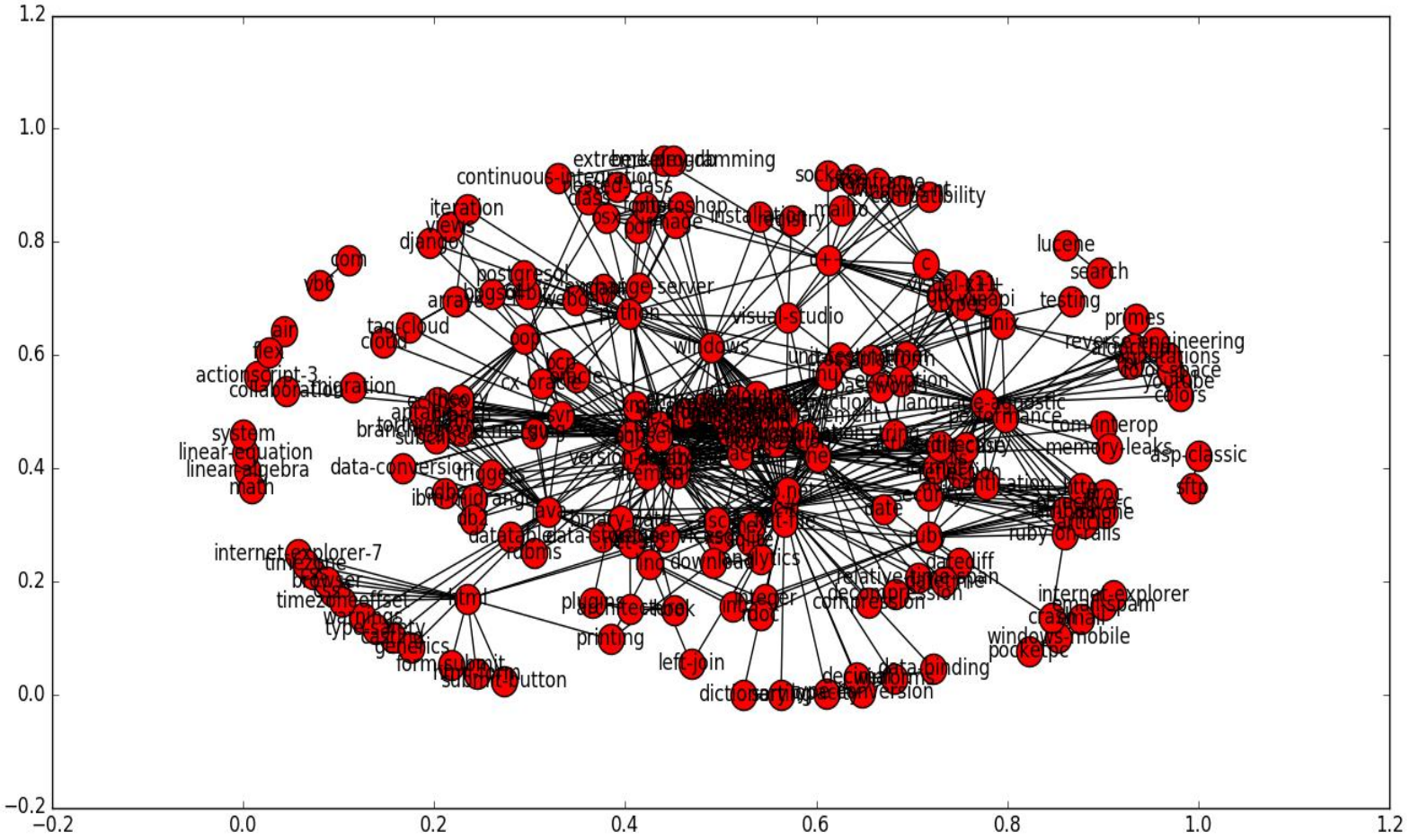
FREQUENTIST INFERENCE APPROACH (FIC)

- **Computes the probability that a software object is assigned a particular tag**

- **Considers the number of words that appear along with a tag in software objects**

- **Steps:**
- **POS Tagging**
- **Assigning a weight to each tag:** $W(o,t) = \sum_{w_i \in o} P(t|wi)$
- **Spreading Activation: Making a network of tags using Jaccard similarity between tags**

SPREADING ACTIVATION (CONTD.)



Tag-Tag Association

COMPOSER COMPONENT (CC)

- **Both BIC and FIC produce a list of tags along with their probabilities**
- **CC combines the two list of tags into one with an updated set of probabilities**
- $EnTagRec_o(t) = a * Bo(t) + b * Fo(t)$

IMPROVING NETWORK FEATURES

- **Network based similarity between the users based on their tags**
- **Users tend to behave similarly to the users having similar interests**
- **Information in the user-user similarity graph propagates to only two hops**
- **Jaccard similarity measure is used to track user similarity**

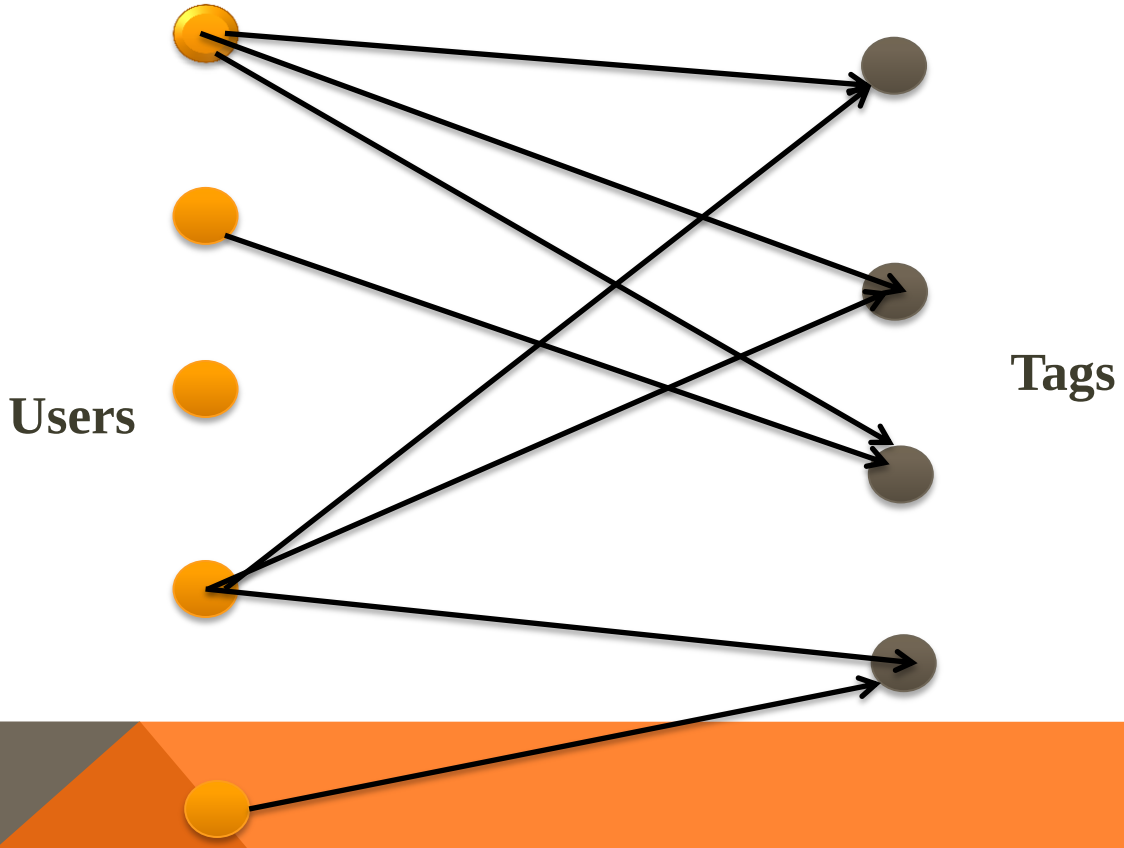
IMPROVING NETWORK FEATURES (CONTD.)

- **P(Rahul chooses Java) =**

w1*P(Rahul chooses Java with random chance) +
w2*P(Users similar to Rahul choose Java)+
w3*P(Rahul's second degree neighbours chose Java),

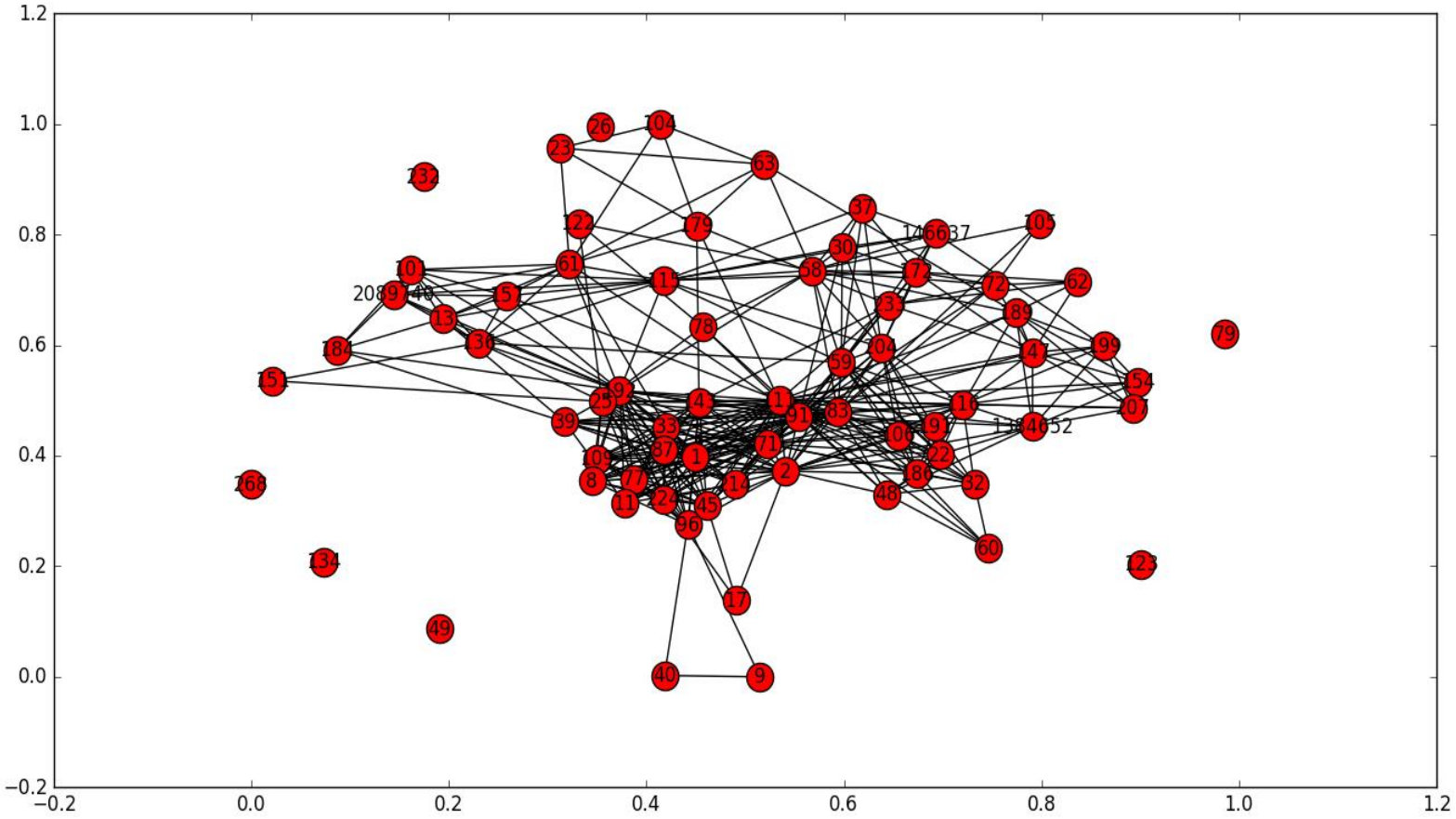
where, w1 + w2 + w3 = 1.

Improving Network Features (Contd.)



User Tag Relationship Bipartite Graph

IMPROVING NETWORK FEATURES (CONTD.)



RESULTS

- **Recall@k values were calculated separately for tags obtained from BIC, FIC and CC for k=10**

	Bayesian (BIC)	Frequentist (FIC)	Composer (CC)
Recall@10	0.31	0.65	0.71

References

[1]Wang, Shaowei, et al. "EnTagRec: An enhanced tag recommendation system for software information sites." *2014 IEEE International Conference on Software Maintenance and Evolution (ICSME)*. IEEE, 2014.

[2]Short, Logan, Christopher Wong, and David Zeng. "Tag recommendations in stackoverflow." (2014).

[3]Ramage, Daniel, et al. "Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora." Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1. Association for Computational Linguistics, 2009.

THANK YOU!