Stanford University

2015 Autumn CS224w

Social and Information Network Analysis

Final Report

Due on Dec 8th, 2015

Author:
Bojiong Ni (bojiong)
Tianjiao Zhang (tjzhang)
Zhiming Shi (zhimings)

Instructor:
Lada Adamic

By turning in this assignment, I agree by the Stanford Honor Code
and declare that all of this is my own work.

Dec. 8th, 2015

REVIEWER EXPERTISE RANKING USING LINK ANALYSIS

BOJIONG NI, TIANJIAO ZHANG, ZHIMING SHI

Abstract

User generated online reviews on products or services are valuable to other users to make informed decisions. Thus, identifying experts in reviewers becomes important. We tackle the problem of reviewer expertise ranking in online review datasets by analyzing the bipartite graphs of reviewers and reviewed items, an approach that has not been considered in previous studies of the problem. We use Co-HITS as our primary ranking algorithm, and compare it with other ranking metrics such as degree and pagerank in reviewer graphs. We evaluate the rankings using Kendall rank correlation coefficient, and validate our results against the real datasets.

1. Introduction

Many e-commerce websites provide channels for users to contribute in the evaluation of product by tagging, commenting, and rating. Online reviews play a pivotal role in users' decision making processes today. Rich user reviews become one of the key factors behind the success of websites like Amazon, Yelp, and IMDB.

Identifying domain experts through user behavior becomes increasingly important, since the experts play an important role in the online community. They tend to discover new media or products before other users, and provide high-quality product reviews that can influence other users' purchasing decisions. Therefore, these domain experts are of great value to merchandisers and online platforms. Merchandisers can target these experts in their marketing campaigns. Online platforms can highlight experts' product reviews to improve user experience. Therefore, the problem of ranking the underlying usefulness of reviews or the expertise level of reviewers has received increasing research attention.

This project tackles the problem of ranking the domain expertise level of reviewers using their reviews in online communities. Traditional approaches include rating the usefulness of reviews, or the expertise level of reviewers, by mining the review text, extracting features from reviewer profiles and their review activities with natural language processing techniques, and carrying out machine learning algorithms. New approaches involve using network analysis on the social network of reviewers. We examine several approaches in the latter category.

2. Related Work

2.1. A Generalized Co-HITS Algorithm and Its Application to Bipartite Graphs (2009).

In [1], the Deng et al. proposed a model for web query suggestion. They modeled the queries and URLs as a bipartite graph where queries and URLs form the two corresponding sets of nodes. Instead of using HITS or personalized PageRank algorithm directly, they devised the Co-HITS algorithm which incorporates the content information from both sides of the graph. The initial hub and authority scores are precomputed by the Statistical Language Model. This is to customize the initial score for reinforcement through a random walk on the bipartite graph. In addition, the score evolution can take flexible weight on the original (current) score and the scores from the connecting hubs. The recurrence of updated score is $x_{new} = (1-\lambda)x + \lambda Mx$. Thus by adjusting λ, one can put more or less weight on the initial score from the Statistical Language Model. HITS algorithm and personalize page rank algorithm correspond to two special cases of λ values.

This paper provides an interesting approach to relevance ranking, which is to use bipartite graph and a mutual-reinforcement algorithm. This approach is very relevant to the reviewer expertise ranking problem, as we could also define a bipartite graph on a typical online review dataset where reviewers and the reviewed items constitute the two node sets. The network modeling, and the intuition and applicability of Co-HITS algorithm to this model will be discussed in later sections.

2.2. Finding Trendsetters in Information Networks (2012).

Date: December 9, 2015.

There are several ways to measure the importance of people in an information network - one definition is to identify the people that first adopt a new idea, i.e. the "trendsetters." Trendsetters play an important role in the process of information diffusion.

In [2], Saez-Trumper et al. presented a method to identify the "trendsetters" by analyzing the information network for any topic. Given a topic (e.g. news events), the information network is modeled as a directed graph $G(N, E)$, where nodes are people, and two people share an edge if they follow the same trend (e.g. hashtag, word, phrase, meme). Each edge is directed from the early adopter to the latecomer. With each edge, the paper also defines an influence score measuring the influence of the early adopter over the latecomer. Finally, the paper produces a trendsetters ranking of nodes by a Pagerank-like algorithm.

The paper tested the proposed algorithm on Twitter dataset, where the concepts of trendsetters, topics etc. are very naturally mapped. In the evaluation section, the paper compared the proposed ranking algorithm against several other ranking algorithms using Kendall Rank Correlation Coefficient, and presented several interesting observations, such as, users with high-indegree do not propose ideas that became popular.

2.3. **Exploiting Social Context for Review Quality Prediction (2010).**

The problem studied in [3] is the automatic determination of review quality. Traditionally this problem is tackled by treating each review as standalone text and using natural language processing techniques together with machine learning algorithms. Lu et al. incorporated the social context by building a directed graph on reviewers, with a directed edge representing trust relationship. From this network, and a set of hypothesis like trust consistency and link consistency, they extracted some regularization term into the machine learning model. They've also studied the approach of using the property extracted from the network as features. They applied their algorithm on data set from Ciao UK, and concluded that the addition of this social context improves the review quality predication results.

The strength and novelty of this paper is that it is among the earliest research work tackling the problem of review quality rating with consideration of the social network structure and reviewer behaviors. The shortcoming of this paper is that, they've made some strong assumptions in their hypothesis. Moreover, they did not consider meaningful network metric like centrality measure or link analysis.

2.4. **Robust Expert Ranking in Online Communities - Fighting Sybil Attacks (2012).**

The problem [4] tackles is the expert ranking problem in online review media. In particular, it addresses the problem of Sybil attack, which characterizes the behavior of malicious users spamming reviews with a large number of accounts. The approach used in this paper is link analysis by treating reviewers and review items as a bipartite graph, i.e. only considering link structure, not the review contents. Rashed et al. proposed an eigenvalue based ranking algorithm to rank users' expertise level in online media communities. They employed the MHITS algorithm, another variation of the classical HITS algorithm in link structure analysis, and they incorporated SumUp algorithm to achieve robustness in defending against Sybil attacks. They applied their algorithm to a synthetic dataset corresponding to fake media detection system. They compared the expertise ranking against ranking obtained using just HITS algorithm.

The novelty of this paper is that it uses link analysis under network analysis setting to perform expertise ranking, whereas a large number of pre-existing efforts use machine learning and natural language processing approaches. It develops a new variation of HITS algorithm, and incorporate that with SumUp to address the Sybil attack problem. However, the shortcoming of this paper is that it only experiments its algorithm on a fake dataset, without investigating the performance on real world data sets.

2.5. **Further discussion.**

[4] inspires us to use bipartite graphs to model the user-review datasets, and convinces us that link analysis is a promising approach for expertise ranking. However, the iterative equation of the MHITS algorithm in paper 4 is actually a special case of the CO-HITS algorithm we reviewed in [1]. As such, in our project we shall implement and experiment the more general Co-HITS algorithm.

[3] convinces us that social context improves the review quality predication results. The users on Yelp essentially form a social community. Two users rating the same restaurant are likely to share similar experiences (i.e they may live in similar area, share similar tastes in food). This intuition provides the basis for analyzing user-user graph to obtain the expertise ranking.

The trendsetter problem discussed in [2] provides us with an alternative way of looking at the expertise ranking problem. In the Yelp Challenge, one of the questions raised is very similar to the problem this paper solved - whether we can identify the "trendsetters", e.g. who found the best waffle joint before waffles were cool. We could build a similar model from the Yelp dataset, implement the several ranking methods mentioned in this paper, and evaluate these ranking algorithms. We could study whether the "trendsetters" are also the authoritative reviewers on a specific topic, e.g. would the people who first find new good Chinese restaurants also be the experts in Chinese food?

3. Problem Definition

A typical online review dataset consists of 3 classes of entities. U, the set of reviewers; I, the set of items being reviewed; and R, the set of review texts, which conceptually could be regarded as $\{(u,i) \mid u \in U,\, i \in I\}$. The reviewer expertise ranking problem takes these 3 sets of data as input, and outputs a vector $\mathbf{e} \in \mathbb{R}^{|U|}$, where each entry of e_u represents the latent domain expertise score corresponding to user $u \in U$.

In the data set we use, there is a fourth entity set, C, the category of items. Each item could belong to several categories. In this case, the subsets of users, review texts and items pertaining to a single category $c \in C$ is regarded as an independent instance of expertise ranking problem.

We propose to obtain the expertise score using link analysis algorithms. Under such type of algorithms, the expertise score effectively is the eigenvector centrality measure. But we also experiment other expertise measures, and evaluate the ranking outputs by these different approaches.

4. Data

Our data is from the Yelp Dataset Challenge[1]. It is publicly released data in Yelp website. We are interested in 3 objects in the data set: user, business, review. The businesses are each tagged with several predefined "categories", such as *Restaurants*, *Bar*, *Chinese*. In addition, we have the location for each business, its average star ratings and the total reviews it has. For each user, we have the average star rating of the reviews they have given, the number of reviews they have written, when they opened their accounts, and their social graph (friends). For each review, we have the business being reviewed, the user reviewing it, the review text and the time the review is written.

The entire data set contains 61,184 businesses, 366,715 users, 1,569,264 reviews and 783 different categories.

[1] `https://www.yelp.com/dataset_challenge/dataset`

In our investigation, we attempt to identify domain experts in one category. Therefore, we extract a subset of data based on business category in the following way. First, we identify all the businesses having the tag we are interested in. Then for each business in the category, we find out all the reviews for it. From the reviews data, we capture all the users who rated this category.

5. Models and Algorithms

We consider several different ways to model the user and review data. We could use both graph-based algorithms and non-graph-based algorithms. Different models allow different algorithms to be used to rank user expertise. As discussed, the Yelp dataset could be partitioned by *business categories*. We examined the ranking of reviewers using data within some specific categories.

5.1. Ranking with Bipartite Graph.

Our primary approach to the expertise ranking problem is to model the review data for a given category of business as a bipartite graph $G(U \cup V, E)$, where the two node sets are users U and businesses V. An edge (u, v) exists when user u has reviewed business v. With this model, we could use Co-HITS algorithm proposed in [1] to compute the expertise score for each user.

For a given business category $c \in C$ of review data set, we use the following algorithm to obtain the reviewers' expertise ranking in this category:

Algorithm Ranking reviewer using Co-HITS

Input: Users U, business V, reviews R, category c

Perform:

1. Construct the bipartite graph using data only pertaining to c.
2. Initialize transition probability Matrix, and starting vectors $\mathbf{x}^0 \in \mathbb{R}^{|U|}$ and $\mathbf{y}^0 \in \mathbb{R}^{|V|}$.
3. Run Co-HITS algorithm to obtain expertise score $\mathbf{e} \in \mathbb{R}^{|U|}$.

Output: Expertise ranking according to $\mathbf{e}$.

The Co-HITS algorithm is a very generalized iterative algorithm that propagates scores on nodes according to a random walk defined by the transition probability matrix. We denote the scores for nodes in U by $\mathbf{x} \in \mathbb{R}^{|U|}$ and scores for nodes in V by $\mathbf{y} \in \mathbb{R}^{|V|}$. The transition probability matrix from U to V is W_{uv}, a $|U| \times |V|$ matrix, and that from V to U is W_{vu}, a $|V| \times |U|$ matrix. They satisfy normalization condition for transition probability, that all the row vectors are sum-normalized, i.e. $\sum_{j \in V} w_{ij}^{uv} = 1$ and $\sum_{i \in U} w_{ji}^{vu} = 1$. We also normalize the initial vectors $\mathbf{x}^0$ and $\mathbf{y}^0$, such that the iterations could be carried out without the need for further normalization. The iterative formula in matrix form is written explicitly below:

$$\mathbf{x}^{i+1} = (1 - \lambda_u)\mathbf{x}^0 + \lambda_u W_{vu}^T \mathbf{y}^i$$
$$\mathbf{y}^{i+1} = (1 - \lambda_v)\mathbf{y}^0 + \lambda_v W_{uv}^T \mathbf{x}^i$$

The physical meaning for the node score in U is the latent domain expertise of each user for this category of businesses. The intuition behind the score propagation is the mutual reinforcement of authorities for the linked entities.

We initialize W_{uv}, W_{vu}, $\mathbf{x}^0$, and $\mathbf{y}^0$ using the dataset. First, the transition probabilities intuitively corresponds to the review frequencies between users and businesses:

$$w_{ij}^{uv} = \frac{\text{\# of reviews user } i \text{ writes for business } j}{\text{total \# of reviews user } i \text{ writes}}$$
$$w_{ji}^{vu} = \frac{\text{\# of reviews user } i \text{ writes for business } j}{\text{total \# of reviews business } j \text{ receives}}$$

In the Yelp dataset, there is a usefulness vote on each review by other users. We use this to initialize the starting scores as follows:

$x_i^0 =$ total usefulness votes user i receives

$y_j^0 =$ total usefulness votes in the reviews of business j

We iterate Co-HITS update until convergence. The runtime complexity for each iteration is the complexity for the multiplication of transition probability matrix and score vector, $O(|U||V|)$. In practice we observe that the algorithm takes roughly 10 iterations to converge. The final node scores for U, $\mathbf{x}$ is outputted as expertise scores for users in this category for ranking.

5.2. **Ranking with User Graph.**

The second way of modeling the data set would be to construct a user graph. We represent a network as a undirected graph $G(N, E)$, where N is the set of nodes and E the set of edges. The nodes in this graph corresponds to all the users who have reviewed at least one business in the given category. For users u, v, $(u, v) \in E$ if they have reviewed the same business.

Since the reviews are dated, we could build a directed graph $G'(N, E')$. For users u, v, $(u, v) \in E$ if they have reviewed the same business, and u reviewed the business before v.

Given the above graphs, we can run several naive ranking algorithms on the users, such as degree centrality ranking and PageRank.

Finally, we could adapt trendsetter ranking algorithm in [2] to our data set. Here, we map the "trend" as reviewing a particular business. Let B be the set of businesses, and $t_i(u)$ be the time that user u reviewed business $b_i \in B$. If the user has never reviewed b_i, then $t_i(u) = 0$. We define two vectors: $s_1(u)_i$ and $s_2(u, v)_i$.

$$s_1(u)_i = \begin{cases} 1, & \text{if } t_i(u) > 0 \\ 0, & \text{otherwise} \end{cases}$$
$$s_2(u, v)_i = \begin{cases} e^{-\frac{\Delta}{\alpha}}, & \text{if } t_i(v) > 0 \text{ , } t_i(v) < t_i(u) \\ 0, & \text{otherwise} \end{cases}$$

Here $\Delta = t_i(u) - t_i(v)$ and $\alpha > 0$. Conceptually, vector $s_1(u)$ informs if node u reviewed (or not) each business (trend), while $s_2(u, v)$ shows if u adopted these trends after v and weights the relation as a function of the period of time between $t_i(u)$ and $t_i(v)$.

From these vectors, we can determine the influence from user u over v as:

$$I(u, v) = \left(\frac{s_1(u) \cdot s_2(u, v)}{||s_1(v)|| \times ||s_2(u, v)||} \right) \times \left(\frac{L(s_2(u, v))}{|B|} \right)$$

Here $L(s_2(u, v))$ is the number of components of $s_2(u, v)$. $I(u, v) = 0$, if $||s_2(u, v)|| = 0$.
Given the above setup, we can run a page-rank like algorithm. We define the "trendsetter" rank as:

$$TS(u) = \frac{1 - d}{|N|} + d \sum_{w \in N} TS(w) I(w, v)$$

where d denotes the damping factor.

5.3. **Ranking with Deviation.**

For each user, we compute how "close" the user's rating is to each business' actual rating. In aggregation of a "category", a user whose prediction is more accurate can be deemed more authoritative in the "category". Intuitively, it reflects how accurately a user rates a business compared to the majority in the online community. An alternative perspective could be how likely a user agrees with the authority defined by majority.

We define the deviation of a user from average ratings using a root-mean-square like metric. For user

u, let B_u be the set of business in the category of interest that u has reviewed, $\overline{r}^i$ be the final rating for business i, and r_u^i be u's rating to i. The deviation d_u is defined to be

$$d_u \equiv \sqrt{\frac{\sum_{i \in B_u} (r_u^i - \overline{r}^i)^2}{|B_u|}}$$

The users are then ranked in order of increasing deviation score.

In the actual implementation, we found that there are many users with only one review. If this single review happens to agree with the business' aggregated review, this user would become highly ranked. Also, we realized that users with fewer reviews are more likely to get smaller deviation score. These observations contradict with the intuition that users with more reviews are likely to exhibit more expertise. In order to account for this discrepancy, we first set a threshold on review counts, i.e. we only rank users with at least 5 reviews in the category. Second, we divide the deviation score defined above by the user's review count $|B_u|$ once again. Thus our final deviation score becomes:

$$d'_u \equiv \frac{d_u}{|B_u|} = \sqrt{\frac{\sum_{i \in B_u} (r_u^i - \overline{r}^i)^2}{|B_u|^3}}$$

Note that this is a very empirical modeling of expertise, and could serve as our baseline.

6. Results and Analysis

We studied four restaurant categories: "Chinese", "Indian", "American (Traditional)" and "Fast Food".

6.1. **Rank with CO-HITS.**

The following table shows the sizes of the bipartite graphs we built from the 4 chosen categories.

	Chinese	Indian	Fast Food	American (Trad)
Num. Bus.	1,489	374	2,373	2,112
Num. Users	33,359	9,549	20,013	68,042
Num. Edges	50,196	12,670	32,527	116,861

The following table shows the ranking we obtained using the Co-HITS algorithm.

Rank	Chinese	Indian	Fast Food	American(Trad)
1	Rand	Rand	Georgie	J
2	Tony	Albert	J	Georgie
3	Norm	Aileen	Rand	Rand
4	Teri	Jenny	Scott	Lindsey
5	Jade	Evan	Christie	Norm

We manually examined the top users, and they seem to be elite Yelp members who have many friends and fans, and who have written many reviews (at least in hundreds across all categories). Many of their reviews receive a lot of "useful" votes. The user "Rand" is ranked amongst the top 5 of all the categories. Rand wrote thoughtful reviews for a variety of restaurants. He has 110 friends and has been an elite member since 2009. However, Rand is not the most prolific reviewer in the dataset, he has written 805 reviews, whereas the top reviewer in the dataset has written 8843 reviews. Rand reviews mostly buffets, as the word "[Bb]uffet" appear in 646 of his reviews. Out of these reviews, a significant portion is about Chinese buffets and Indian buffets. This is consistent with the fact that he is the top reviewer for both Chinese and Indian categories.

6.2. **Rank with User graph.**

The following table shows the sizes of the user graphs we built from the 4 chosen categories.

	Chinese	Indian	Fast Food	American (Trad)
Num. Nodes	33,359	9,549	20,013	68,042
Num. Edges	3,908,271	778,198	1,285,278	18,470,162

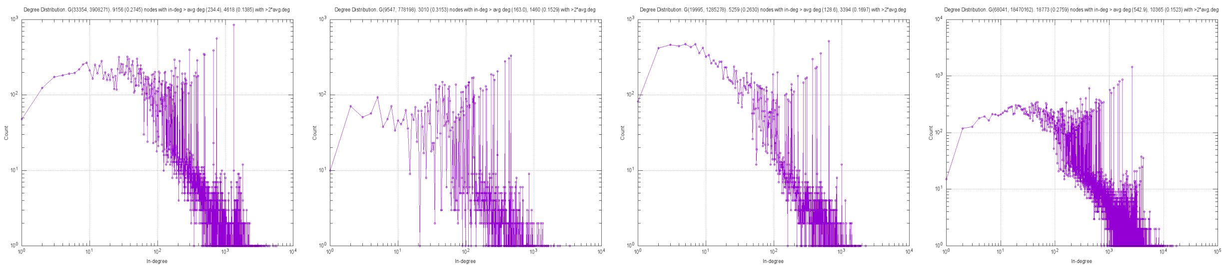
Figure 1 shows the degree distributions. We could see that the degree distributions from the 4 categories are very similar. Each of them follows a power law distribution, suggesting that there are a few users that reviewed lots of businesses.

Table 1 shows the top 5 users for each category. Users that appear in both ranking methods are highlighted. Many of the top reviewers are consistent with the result from the Co-HITS algorithm. We manually examined the remaining top reviewers, and they share the same attributes as the top reviewers found from the Co-HITS algorithm.

Finally, we tried the trendsetter approach. However, we cannot interpret the result very well. Our hypothesis is that we need to a better way of capturing "influence" than the ordering of review dates. Hence we shall exclude this ranking method in our evaluation.

6.3. **Rank with Deviation.**

The resulting rank with deviation is interesting. Users with more reviews do emerge on top of the expert ranking, but it is not the case that a user with more reviews has strictly more expertise. From the ranking results of the four categories (Table 2), we can clearly see that there are cases in which users with fewer reviews are ranked higher. We spot checked these top ranked users and realized that they have a large number of fans, and are typically elite members. Note that user "Rand" is again ranked as the



(A) Chinese (B) Indian (C) Fast Food (D) American (Traditional)

Figure 1. Degree distribution of user graphs

Rank	Degree	PageRank
1	**Jade**	Judy
2	**David**	**David**
3	Emily	Alexander
4	Teri	**Jade**
5	Cathy	Rand

(A) Chinese

Rank	Degree	PageRank
1	**Albert**	**Rand**
2	**Mike**	**Albert**
3	**Rand**	**Mike**
4	Philip	Evan
5	Norm	Jennifer

(B) Indian

Rank	Degree	PageRank
1	Emily	Rand
2	Demitri	**J**
3	Nikki	Karen
4	Niki	Bruce
5	**J**	Georgie

(C) Fast Food

Rank	Degree	PageRank
1	**Norm**	**J**
2	Blair	Michael 2
3	Michael	Jennifer
4	Samantha	**Norm**
5	**J**	Brad

(D) American (Traditional)

Table 1. User graph rankings

Rank	Deviation	reviews	Usernames	Elite years	Fans
1	0.0871	212	Rand	2009-2015	110
2	0.0878	18	Lora	2012-2015	12
3	0.0972	36	Teri	2006-2015	219
4	0.1	5	Brandon	2009-2010	10
5	0.1	5	Nachiket	2014	2

(A) Chinese

Rank	Deviation	reviews	Usernames	Elite years	Fans
1	0.0874	93	Rand	2009-2015	110
2	0.1	5	Meredith	2009-2015	34
3	0.1	5	Christopher	2010-2011	22
4	0.1384	30	Gibson		1
5	0.1414	5	Mason	2012-2015	16

(B) Indian

Rank	Deviation	reviews	Usernames	Elite years	Fans
1	0.0832	103	J	2010	164
2	0.1	10	James	2013-2014	11
3	0.1	5	Kay		1
4	0.1016	22	Michael	2008	62
5	0.1033	51	Gibson		1

(C) Fast Food

Rank	Deviation	reviews	Usernames	Elite years	Fans
1	0.0727	95	Michael 2	2008	62
2	0.0743	113	J	2010	164
3	0.0786	9	Jason	2011-2015	15
4	0.0833	6	Roy	2014-2015	5
5	0.0833	6	Katie	2008-2013	40

(D) American (Traditional)

Table 2. Deviation Rankings

top expert in both "Chinese" and "Indian" category again. However, this is not a necessary condition for users to be ranked towards the top by deviation, like in the case of Gibson.

7. Evaluation

We've already discussed our spot checking on the rankings we've obtained, and the top ranked users do show expertise in each category. To further compare and evaluate these different rankings, we use the Kendall rank correlation coefficient τ:

$$\tau = \frac{\#\text{ concordant pairs} - \#\text{ discordant pairs}}{\#\text{ pairs}}$$

It takes value in interval $[-1, 1]$, where 1 means two rankings are in total agreement and -1 means one is the reverse of the other.

7.1. Real Dataset.

The difficulty in performing good quantitative evaluation in a real dataset is the lack of a golden standard. We thus look at the Kendall's τ between each pair of ranking algorithms in Table 3.

The Co-HITS ranking from the bipartite graph is positively correlated to RageRank result generated using user graph. This could be due to the fact that both rankings are eigenvector based. Degree ranking is correlated with the two eigenvector based results for "Chinese" and "American (Traditional)", while negatively correlated in the other two. Conceptually, degree ranking captures the prolificacy of the reviewers. Those with more reviews tend to have more edges; and we observe preferential attachment as one tends to connect to a prolific reviewer reviewing popular restaurants. Although many reviewers appear on both degree ranking and PageRank ranking, they are not always concordant. Finally, deviation ranking seems to agree with the eigenvector based approaches

	Co-HITS	PageRank	Degree
Co-HITS	-	-	-
PageRank	0.0222	-	-
Degree	0.6	0.156	-
Deviation	0.244	0.333	0.2

(A) Chinese

	Co-HITS	PageRank	Degree
Co-HITS	-	-	-
PageRank	0.156	-	-
Degree	-0.022	-0.244	-
Deviation	0.467	0.333	-0.289

(B) Indian

	Co-HITS	PageRank	Degree
Co-HITS	-	-	-
PageRank	0.022	-	-
Degree	-0.156	-0.244	-
Deviation	-0.556	-0.2	0.422

(C) Fast Food

	Co-HITS	PageRank	Degree
Co-HITS	-	-	-
PageRank	0.156	-	-
Degree	0.067	0.289	-
Deviation	-0.2	-0.244	-0.067

(D) American (Traditional)

Table 3. Kendall Rank Correlation Coefficient

for "Chinese" and "Indian", but disagree for the rest two. The intuition behind deviation ranking is that authority is usually recognized by majority. It is expected that deviation ranking and the other approaches may or may not have correlation, depending on the specific category. In addition, users that have only written a handful of reviews may appear on deviation ranking.

The results above for Co-HITS is obtained from using $\lambda_u = 0.4$ and $\lambda_v = 0.8$. We discovered that the user expertise ranking is rather insensitive to the variation of λ_v. This might be the reason [4] uses MHITS algorithm which is equivalent to $\lambda_v = 1$. We experimented with different λ_u values. Note that $\lambda_u = 0$ would correspond to the initial values (user usefulness votes), and $\lambda_u = 1, \lambda_v = 1$ would correspond to the original HITS [1]. Figure 2(A) shows the rank correlation change varying λ_u. We picked a λ_u value that optimizes the correlation between the empirical standard (deviation) and Co-HITS across all categories.

7.2. **Synthetic Dataset.**

To compensate for the lack of a golden standard, we adopt the approach of generating a synthetic dataset in the modeling of a real problem, as in [4].

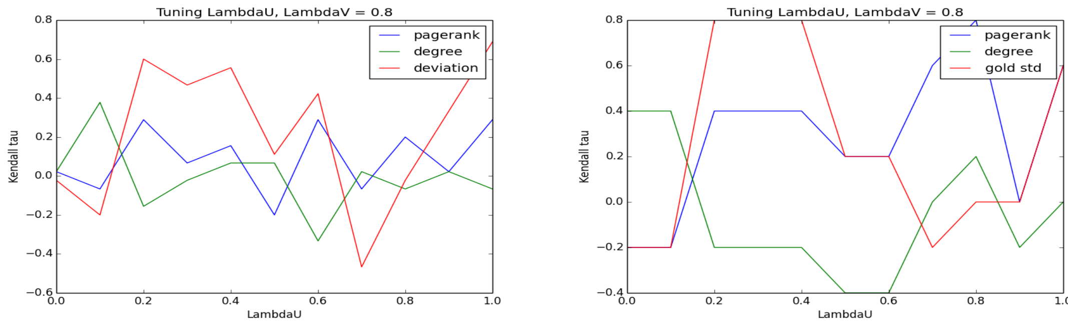
Guillaume and Latapy showed in [6] and [7] that the degree distributions of both sets of nodes in a bipartite graph originating from complex networks tend to follow a power law, and a random bipartite graph generated using their proposed algorithm can achieve this exact property. We've also observed this in our dataset, i.e. the degree distributions of both user nodes and business nodes in the bipartite graph follow power law, see Figure 3. As such, we first generate a random bipartite graph using Guillaume-Latapy algorithm with parameters (distribution characteristics) drawn from our real datasets. This should simulate the underlying dynamics of the online reviews. Since our algorithm considers only link structure, not other features, we would not need to generate ratings. Next, we simulate users voting usefulness according to a probability distribution that is roughly proportional to the review counts of the other users (with random noise). We take the resulting voted usefulness ranking as the true expertise rank. We replace the deviation ranking with this ranking as the golden standard in our analysis.

Figure 2(B) shows the tuning results using the synthetic dataset, and the resulting rank correlation coefficients are tabulated below. We see that Co-HITS is able to achieve a high rank correlation with the gold standard.

	Co-HITS	PageRank	Degree
Co-HITS	-	-	-
PageRank	0.4	-	-
Degree	-0.2	0.2	-
Gold	0.8	0.3	-0.2

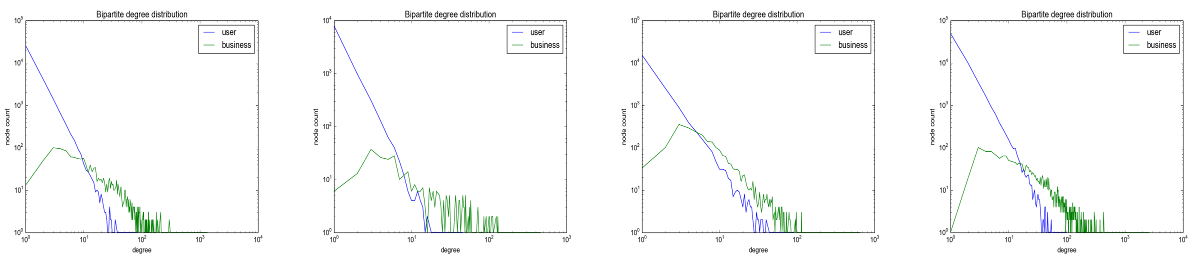
8. Discussion and Future work

In this project, we've tackled the problem of reviewer expertise ranking in online review datasets by analyzing the bipartite graphs of reviewers and reviewed items. This approach has not been used in previous studies of the same problem. We considered Co-HITS as our primary ranking algorithm, and compared the results with other ranking metrics such as degree and pagerank in reviewer graphs, and review rating deviation. We evaluated the rankings on both the real datasets and a synthetic dataset



(A) Co-HITS param tuning: real dataset (B) Co-HITS param tuning: fake dataset

FIGURE 2. Co-HITS Parameter Tuning



(A) Chinese (B) Indian (C) Fast Food (D) American (Traditional)

FIGURE 3. Degree distribution of bipartite graphs

generated using a reasonable model.

This study could be extended to problems like ranking robustness when spam attack is present. We could also incorporate machine learning techniques and natural language processing into the ranking framework.

9. Acknowledgment

We would like to thank Professor Lada Adamic and TA Nihit Himanshu Desai for their ideas and feedback during our discussion.

References

[1] Deng, Hongbo and Lyu, Michael R and King, Irwin. A generalized Co-HITS algorithm and its application to bipartite graphs. Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining. 2009.

[2] Saez-Trumper, Diego and Comarela, Giovanni and Almeida, Virgílio and Baeza-Yates, Ricardo and Benevenuto, Fabrício. Finding trendsetters in information networks. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. 2012.

[3] Lu, Yue and Tsaparas, Panayiotis and Ntoulas, Alexandros and Polanyi, Livia. Exploiting social context for review quality prediction. Proceedings of the 19th international conference on World wide web. 2010.

[4] Rashed, Khaled AN and Balasoiu, Cristina and Klamma, Ralf. Robust expert ranking in online communities-fighting Sybil Attacks Collaborative Computing: Networking, Applications and Worksharing (CollaborateCom), 2012 8th International Conference on. 2012.

[5] Newman, Watts and Strogatz. Random graph models of social networks Proceedings of the National Academy of Sciences 99.suppl 1 (2002): 2566-2572.

[6] Guillaume and Latapy. Bipartite Graphs as Models of Complex Networks Physica A: Statistical Mechanics and its Applications 371.2 (2006): 795-813.

[7] Guillaume and Latapy. Bipartite structure of all complex networks Information Processing Letters (IPL), 90(5):215221, 2004.

Google Inc.
E-mail address: bojiong@stanford.edu, tjzhang@stanford.edu, zhimings@stanford.edu