



HAL
open science

Expert finding in Citizen Science platform via weighted PageRank algorithm

Zakaria Saoud, Colin Fontaine

► **To cite this version:**

Zakaria Saoud, Colin Fontaine. Expert finding in Citizen Science platform via weighted PageRank algorithm. Advances in Intelligent Data Analysis XVII, Oct 2018, Hertogenbosch, Netherlands. hal-02393811

HAL Id: hal-02393811

<https://hal.science/hal-02393811>

Submitted on 4 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Expert finding in Citizen Science platform via weighted PageRank algorithm

Zakaria SAOUD¹, Colin Fontaine¹

1) Centre d'Ecologie et des Science de la Conservation, UMR 7204 CNRS-MNHN-SU,
Muséum national d'Histoire naturelle, 61 rue Buffon, 75005 Paris, France

zakaria.saoud@mnhn.fr

colin.fontaine@mnhn.fr

Received: Date Month 20XX / Accepted: Date Month 20XX

© Springer 2018

Abstract. Several citizen science platforms aiming at monitoring biodiversity have emerged in the recent years. These platforms collect biodiversity data from participants and allow them to increase their scientific knowledge and share it with other participants, experts and scientists. One key aspect of such platforms is quality control on the **data, a task usually** performed by a limited number of co-opted experts. With the amount of data collected increasing steeply, finding new experts is needed. In this paper we propose a new graph-based expert finding approach for the citizen science platform SPIPOLL, aiming at collecting data on pollinator diversity across France. We exploit both users' comments quality and users' social relations to calculate users' expertise for specific insect family. Experimental results show that the proposed method performs better than the state-of-the-art expert finding algorithms.

Keywords: Expert finding, PageRank algorithm, Citizen Sciences.

1 Introduction

Citizen science (CS) platforms represent a powerful tool allowing participants to contribute to research and increase their scientific knowledge. Furthermore, CS platforms help scientists in their research projects, by collecting more data and analyzing it. Generally, the **primary** goal of CS platforms is connecting many participants, experts, and researchers towards a common scientific goal. Nowadays, numerous CS platforms have emerged and can be classified **according to** their scientific objectives such as: medicine, ecology, astronomy, computer science, psychology, etc. Many popular CS platforms with large communities of participants exist today, such as Zooniverse¹, Foldit², Eyewire³, and eBird⁴. Zooniverse benefits from the collaboration from more than 1 million registered users to analyze pictures of distant galaxies. Foldit allows users to fold the structures of selected proteins as correctly as possible, by playing an online puzzle video game. Eyewire challenges players to map neurons in 3D, by solving 2D puzzles, thereby helping researchers to model information processing circuits. eBird collects birds pictures from many volunteers, to provide data about bird distribution and abundance in real-time. Similarly to eBird, SPIPOLL⁵ allows users to take photos of flowering plants and their pollinating insects to study changes in pollinator assemblages across space and time. However, most of the existing CS platforms still lack an expert finding (EF) mechanism, which could improve the quality of collected data and optimize data evaluation time. EF approaches aim to extract a list of experts with high knowledge and expertise in a specific domain, to produce high quality answers to questions from online communities. Most of these approaches were focused on communities question answering (CQA) websites. Unlike the existing EF approaches, our study deals with the problem of EF in online CS platform on biodiversity, with the SPIPOLL as a study case.

In the SPIPOLL, after taking pictures of pollinators on flowers, the users give a name to each photographed insect from 600 possibilities and share their photos and associated insect names on the platform. While users can comment on each other observations and identifications, experts validate or correct the pollinator identifications. In our approach, we analyze the users' comments and extract the comments that contain precise identifications. The extracted comments will be

¹ <https://www.zooniverse.org/>

² <http://fold.it/portal/>

³ <https://eyewire.org/explore>

⁴ <https://ebird.org/home>

⁵ <http://www.spipoll.org/>

considered as answers and will be used to construct the users' social network. A weighted PageRank algorithm will be applied on the obtained network, to calculate the users' expertise for a specific insect family. This paper is organized as follows: Section 2 provides an overview of the related work in the area of EF in CQA websites. Section 3 presents the general structure of the SPIPOLL website. Section 4 introduces the details of our proposed EF approach. Section 5 describes the experimental setup and obtained results. Finally, we provide some concluding remarks in Section 6.

2 RELATED WORK

CQA websites represent a powerful tool of knowledge mining on specific topics which can't be extracted easily from general web search engines. CQA websites allow online users to post and answer questions and exchange knowledge among them. Nowadays, several CQA platforms have emerged, such as Quora⁶, Yahoo Answers⁷, Blurtit⁸ and Stack Overflow⁹. With the increase of these platforms, the task of EF has received significant attention in the literature. EF aims to find the appropriate users or experts who can provide good quality answers for posted questions. Many research fields can benefit from EF techniques, such as questions recommendation [1] and spam detection [2] [3]. For CQA websites, several approaches have been proposed, which can be classified into three main categories: 1- graph-based EF approaches, 2- content-based EF approaches and 3- competition-based EF approaches.

In graph-based EF approaches, the users' network is represented by a directed graph, where nodes represent users, and edges represent the relationship among them. A

link from user A to user B is drawn, if the user B answers for question posted by

user A . The user expertise score can be estimated from the number of edges pointing to him. Most of existing works in this category have adopted link analysis algorithms like PageRank [4] or Hits [5], to calculate the users' expertise scores. We provide in what follows a brief review of such approaches: Zhang et al. [6] proposed a new experts ranking algorithm, named ExpertiseRank. This algorithm is based on PageRank algorithm and calculates the expertise of each user according to the

⁶ <https://fr.quora.com/>

⁷ <https://fr.answers.yahoo.com/>

⁸ <https://www.blurtit.com/>

⁹ <https://stackoverflow.com/>

expertise of others related users to him. Li et al. [7] combined documents quality, documents topic-focus degree and users' activities to calculate the users' expertise rank. A social network analysis (SNA) algorithm has been used to analyze the links between the discovered experts, to obtain the specific experts for a specific topic. Zhao et al. [8] exploited the online social relations between users via graph regularized matrix to find experts in CQA systems. Zhao et al. [9] proposed a novel ranking metric network learning framework for EF by exploiting both the social interactions between users and users' relative quality rank to given questions. Rafiei et al. [10] proposed a hybrid method for EF based on content analysis and SNA. The content analysis is based on concept map and SNA is based on PageRank algorithm. Wei et al. [11] proposed the ExpRank algorithm, an extension of the PageRank algorithm. In this algorithm, the negative and the positive agreements relations between users have been both exploited to calculate their expertise. Yeneterzi et al. [12] exploited topic-relevant users and the interactions between them, to construct topic specific authority graph, called Topic-Candidate (TC) graph. This graph has been used to estimate the topic-specific authority scores for each user. Zhu et al. [13] exploited the information in both relevant and target categories, to improve the quality of authority ranking. Procaci et al. [14] proposed a new approach for EF in online communities based on graph ranking algorithm and information retrieval approach. In this approach, two machine learning techniques, artificial neural network, and clustering algorithm have been exploited for EF. Dom et al. [15] applied a graph-based algorithm to rank email correspondents according to their degree of expertise on specific topics. Their results showed that PageRank algorithm performs better than all other algorithms. Jie et al. [16] used a weighted HITS algorithm for computing users' reputation and recommending the obtained experts to the users who have posted questions. Content-based EF approaches analyze the extracted information from the users' answers to predict their expertise. User expertise score can be estimated from his Z-score [6], his answers' quality [17], his expertise domains [18] or his answers voted score [19]. Competition-based approaches suppose that the best answerer has higher expertise than other answerers for a question. To achieve that, they explore the pairwise comparisons between users (players) deduced from best answer selections, to estimate user expertise score. The resulting pairwise comparisons can be considered as two-players competition. Liu et al. [20] applied two-players competition models to determine the relative expertise score of users. Aslya et al. [21] proposed a novel community expertise network structure, by creating relations among the best answerer and other answerers they have beaten. The EF process is based on the principle of competition among the answerers of a question. In this work, unlike the existing graph-based EF approaches, we take into account the relationship degrees between users. We represent the interactions between users by a weighted graph. Then, we apply a weighted PageRank algorithm on this graph to

estimate the users' expertise. Details of the proposed method will be described in Section 4.

3 The general structure of the SPIPOLL

SPIPOLL is an SC platform created by the National Museum of Natural History (MNHN) and the Office for Insects and their Environment (Opie), to collect data on flowers and their insect pollinators within metropolitan France. The collected data improve the users' knowledge about insect pollinators and allow scientists to assess the abundance variations of pollinator communities. In the SPIPOLL, each user (observer) is asked to take pictures of all insects visiting chosen flowering plant, for a certain period of time. The collected pictures of insects and flowering plant are then uploaded on the SPIPOLL website to form a photographic collection. Nowadays, the SPIPOLL database contains more than 31329 photographic collections and 307719 insects' pictures. After data collection, observers are asked to identify insects and flowering plants, using an online identification key. Finally, the identifications will be validated by a small group of entomologists from the OPIE. In the SPIPOLL, users can also comment pictures and collections, and add doubts in the identified photos if they aren't sure about identifications. Figure 1 represents the general process of the SPIPOLL.

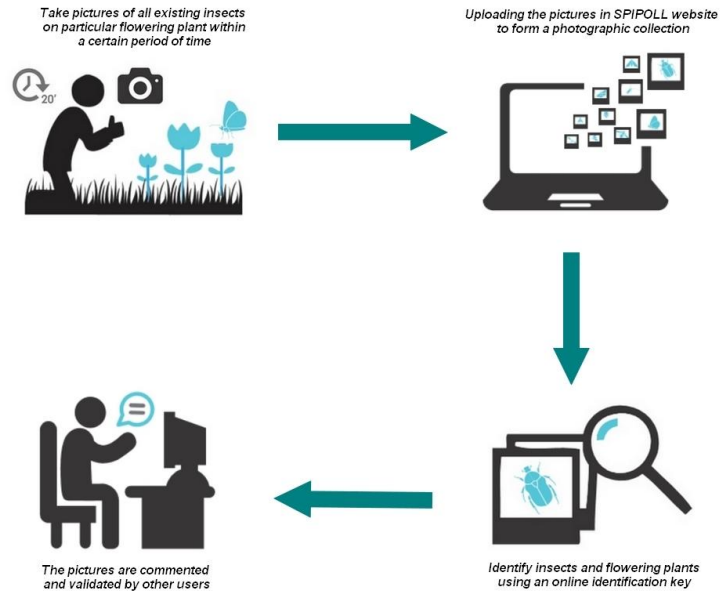


Figure 1. The general process of the SPIPOLL.

However, with the increase of collected pictures in the SPIPOLL, the limited number of entomologists is insufficient to validate all identifications. Therefore, we propose a novel approach to identify expert within the users for specific insect family based on the users' comments. The comments which contain precise identifications will be considered as answers. Each answer will be compared to corresponding validation (the correct identification validated by experts) to verify its reliability. In other word, we know what the true identification is and we then search for comment that gave the right answer with no ambiguity. In the SPIPOLL, all data will be eventually validated as correct identification which is a prerequisite for ecological analysis. This validation is performed by few co-opted experts that do not manage to cope with the amount of data to validate. The aim of this work is to find some new experts among the participants. As a consequence, the selection of these new experts is performed on a database that is validated.

4 The proposed approach

In our approach, we exploit both comments (answers) quality and social interactions between users to predict their expertise. In our weighted graph model, users are

represented as nodes, related among them by weighted directed edges. Each edge points from the questioner (the observer) to the answerer (the commentator). The edges weights are calculated according to the reliability of exchanged answers between users. We consider the comments that contain a precise identification (the exact name of the insect) as answers and the posted pictures as questions which wait for identifications (answers). An answer is considered correct if it's identical to the validation. Finally, we estimate the users' expertise, by applying a weighted PageRank algorithm on the graph representing the network of questions and answers among users. Our proposed approach can be summarized as follows:

- 1- Merging users' comments on pictures and collections.
- 2- Extracting precise identifications from comments, using text analysis technique.
- 3- Extracting the comments with precise identifications (CPIs).
- 4- Comparing the extracted CPIs with the corresponding validations (the true identifications) and calculate a score for each user and for each insect family.
- 5- Calculating the relationship degree between users and constructing the users' social network graph.
- 6- Apply a weighted PageRank algorithm on the obtained graph and determine the expert users.

Figure 2 represents the overall process of our proposed EF approach.

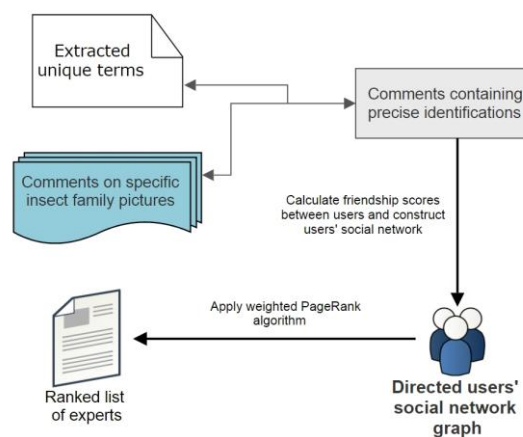


Figure 2. The overall process of the proposed approach.

4.1 Merging users' comments

The comments posted on collections represent 90% of the whole comments on the SPIPOLL website. This is due to the fact that most users prefer to add comments directly to collections rather than on the insect pictures as it avoids several clicks. This situation, prevent us from knowing the precise pictures that users' refer in their comments. As a solution for this, we compare the validation of each picture belonging to a collection, with its collection' comments. Each comment will be attributed to the corresponding picture if this comment contains identification identical to one of the validated picture of the collection.

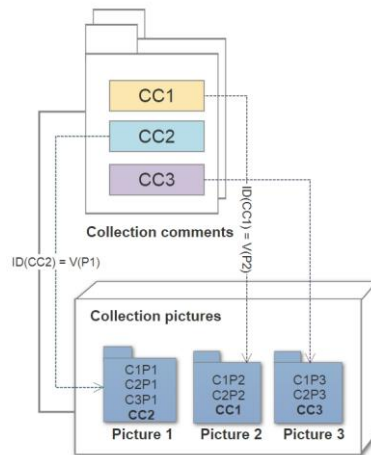


Figure 3. Example of the comments merging process.

Comments without any identical identification to any pictures' validations will be attributed randomly to any picture without comment from the collection. In the end, collections' comments will be merged with pictures' comments. Figure 3 shows an example of the comments merging process.

4.2 Extracting precise identifications from comments

In the SPIPOLL, each user can add comments on pictures or collection, to great other observers, to comment the picture esthetics, or to comment identifications. Users can also add identifications in comments if they think that posted identifications are false. Usually, the proposed identifications in comments are used by observers to change old identifications. In some case, users propose wrong identifications which can push the observer to change their correct identifications. For this reason, the comments represent an important key for obtaining reliable identifications. Hence, comments can be used to calculate users' expertise. In one hand, we suppose that users with high expertise in specific insect family are more likely to add comments with true and precise identification. In other hand, users with low expertise are likely to add comments with wrong identifications. However, some comments can contain an imprecise identification and can't be used to judge users' answers. Identification is considered imprecise when it doesn't contain a term or term combination that correspond unequivocally to a single insect name. On the contrary, comments with precise identification contain a term or combination of terms that correspond unequivocally to a single insect name and can be defined as follows:

$$CPI = \{term | \exists term \in Unique_{terms}\}$$

With:

term : is a comment term.

Unique_{terms} : is the set of existing unique terms. To obtain the set of unique terms, we apply a text analysis technique on the SPIPOLL' insect names. First, we transform each insect name to a list of tokens, we then eliminate the stopwords. Table 1 shows the extracted unique terms from three different insects.

We mention that unigram unique terms (with one word) which have ambiguous meanings (like brown, garden, day,..etc.) have been deleted, because they have insufficient meanings to describe the insects.

| Insect names | Uniques terms | | |
|--------------|---------------|---------|--------|
| | | Unigram | Bigram |

| | | | |
|--|--|---|---|
| Les Ammophiles (Ammophila, Hoplammophila) | Ammophiles Ammophila Hoplammophila | ('Ammophiles', 'ammophila') (Ammophila, 'Hoplammophila') | ('Ammophiles', 'Ammophila', 'Hoplammophila') |
| Les Rayures (Aplocera) | Rayures Aplocera | ('Rayures', 'Aplocera') | \ |
| Les Moirés sans ocelles (Erebia) | \ | ('Moirés', 'Ocelles') (Ocelles', 'Erebia') | ('Moirés', 'Ocelles', 'Erebia') |

Table 1. Some insects names with their unique terms.

4.3 Calculating relationship degree between users

The extracted CPIs will be used to calculate the relationship degree between users. These comments will be considered as answers, and the posted pictures will be considered as questions which wait for good identifications (answers). In our case, we use only CPIs that have been posted on insects' pictures of the same family. The relationship between two users will be calculated for one target insect family, using their average answers' scores of insects which belong to the target insect family. The relationship strength between two users will increment if they exchange good answers (i.e. if their answers are identical with the validations). The relationship degree between two users will decrement if they exchange wrong answers. The difficulty of identification of insect can affect on answers' gained score. The user will earn more score if he gives good answers for a difficult insect to identify, and will earn less score if he gives good answers for an easy insect to identify. On the other hand, the users will lose fewer score if he gives wrong answers for difficult insect, and will lose more score if he gives wrong answers for an insect easy to identify. The length of the answer can also affect on answers' gained score. Expert users are expected to give long answers with more unique terms. The relationship degree should be calculated from each user side. Thus, we can calculate the relationship degree between two users

A (the commentator) and B (the observer) for specific insects family (insects set)

f , as follow :

$$relationship_f(A, B) = \sum_{taxon \in f} \frac{score_{answers_{ex}}(A, B)}{|f|}$$

$|f|$: is the number of existing insects in the f insect family.

$score_{answers_{tx}}(A, B)$ represents the score of posted answers of user A on the pictures of the user B , for a specific insect tx . This score is calculated using the following formula:

$$score_{answers_{tx}}(A, B) = \frac{\sum_{R \in Answers_{tx}(A, B)} \begin{cases} \frac{1}{ease(tx)} * |R| & , R = V \\ - ease(tx) * |R| & , R \neq V \end{cases}}{\sum_{R \in Answers_{tx}(A, B)} |R|}$$

With:

$ease(tx)$: represent the ease score of the insect tx . This score is high when the insect is easy to identify and is low when it's hard to identify. This score is calculated as follows:

$$ease(tx) = \frac{\text{Number of } tx \text{ pictures with true identifications}}{\text{Total number of } tx \text{ validated pictures}}$$

$Answers_{tx}(A, B)$: is the set of posted answers of user A on the pictures of the user B for the insect tx .

R : is one answer from the set of answers $Answers_{tx}(A, B)$.

$|R|$: is the length of the answer, i.e. the size of the largest existing unique term on the answer.

V : is the corresponding picture validation.

In our study, each insect with score higher than 0.65 (the average ease score of all insects), will be considered easy for identification. On the other hand, an insect with a score lower than 0.65, will be considered hard for identification.

4.4 Construct the users' social network

When users (observers) post pictures on the SPIPOLL website, some other users can comment on his pictures. Connecting observers to commentators by directional weighted arrows from observers to commentators, allows us to create the users' social

network. Hence, the SPIPOLL' users can be organized in a weighted and directed graph $G(V, E)$, Where:

V : is the set of users who share or comment pictures of one specific insect family.

E : is the set of directed edges, where $e_{i,j}$ indicates that user u_j has commented on one or more pictures of user u_i . These edges are weighted using the friendship degree formula (see section 4.3).

4.5 Calculate users expertise using a weighted PageRank algorithm

Nowadays, PageRank algorithm has proven its efficiency not only on web pages ranking but also on EF field. Many PageRank-based EF algorithms [6][10][11][15]. have proved that PageRank outperforms other algorithms like HITS and Z_scores [6] for EF. However, these studies have applied PageRank only on non-weighted graphs. In our case, we use a weighted PageRank algorithm to extract experts from a weighted graph. Several Weighted PageRank algorithms have been proposed [22][23][24][25] to improve the performance of original PageRank. The weighted PageRank consists of adding weights to different parts of PageRank formula. According to [24][26], weighted PageRank performs better than traditional PageRank. In our approach, we use the proposed weighted PageRank algorithm by Mihalcea [27]. In this algorithm, the PageRank score of target vertice V_a is calculated using the weights of coming edges from of its predecessors' vertices $In(V)_a$ and the weights of destined edges to the successors of its predecessors' vertices $Out(V)_b$.

In our approach, we calculate the weighted PageRank score for a user A as follows:

$$WP(A) = (1 - d) + d \sum_{B \in In(A)} \frac{relationship_f(B, A)}{\sum_{C \in Out(B)} relationship_f(B, C)} * WP(B)$$

With:

B : is a user who has received at least a comment from user A .

$In(A)$: is the list of users who have received comments from user A .

C : is a user who has commented on pictures or collections of user B .

$Out(B)$: is the list of users who have commented on the pictures or collections of user B .

$WP(B)$: is the PageRank score of the user B .

d : is a damping facto which can be set between 0 and 1. Similar to the previous studies, we will set the damping factor to 0.85.

5 Experiments

In this section, we evaluate the performance of our proposed approach using a set of observers and commentators from the SPIPOLL. The collected comments are posted on the insects' pictures of the same family. In our study, we choose the "Apidae" insect family because it contains the most observed insects in SPIPOLL. To show the effectiveness of our proposed approach, we compare it with 2 state-of-the-art methods: the Z-score [6] and ExpertiseRank [6].

To generate the ground truth ranking scores, we use the set of added identifications on the pictures. We calculate for each commentator, his ground truth expertise score for specific insect, by comparing his identifications with the corresponding validations. The ground truth expertise of the user U_n for the insect tx_m can be defined as follows:

$$Expertise(U_n, tx_m) = \frac{\text{Number of correct identifications posted on } tx_m \text{ by } U_n}{\text{Totale number of identifications posted on } tx_m \text{ by } U_n}$$

The obtained expertise will be used to calculate the user ground truth expertise score for specific insect family. The ground truth expertise of the user U_n for the insect family f_m can be defined as follows:

$$Expertise(U_n, f_m) = \frac{\sum_{tx_m \in f_m} Expertise(U_n, tx_m)}{|f_m|}$$

$|f_m|$: is the number of existing insects in the f_m insect family.

5.1 Data Preparation

The dataset is obtained from a sample of the SPIPOLL database. We collected the information from all posted pictures and comments from April 2010 to October 2017. In total, we extracted 31329 collections, 307719 pictures, 76288 comments and 1455 users. Among these comments, 28% contain precise identifications. In our case, we use only the posted comments on the insect pictures of the “Apidae” insect family, which represent 12% of all comments. Thus, we obtain a sample which contains 1844 validated pictures, 252 users, and 1866 CPIs. Figure 4 shows the obtained social network using this sample. In this graph, the node size represents the number of connections of the node with the other nodes. Largest nodes have a higher degree of connections than others.

5.2 Evaluation Criteria

We evaluate the performance of each algorithm under investigation based on three evaluation metrics: Precision at K ($P@K$), Kendall’s tau and Spearman’s rho. The first metric measures the proportion of the best commentators (best experts) ranked in the top K results. In our evaluation, each commentator with higher ground truth expertise than 0.4 (the average ground truth expertise of all users), will be considered as best expert. The second and the third metric measure the correlation between the ideal ranking (the ground truth ranking) and the obtained ranking. We calculate Kendall’s Tau and Spearman’s rho for the 10, 20 and 40 top ranked commentators.

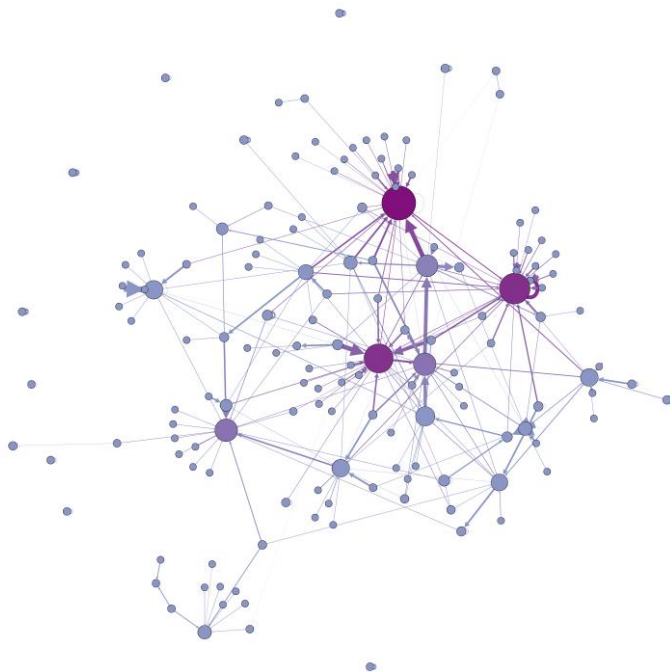


Figure 4. The obtained social network.

5.3 .Results

Figure 5 shows the obtained precision from the top 10, 20, 30, 40 commentators respectively. We can see that graph-based algorithms perform better than the Z-score algorithm. This result proves that the exploiting of relations among users can improve the performance of experts' identification. As figure 5 shows, our weighted PageRank algorithm also outperforms the ExpertiseRank algorithm, especially in the top 10, 20 and 30 users. The precision of the weighted PageRank algorithm reduces when the number of users increases and is equal to the ExpertiseRank algorithm on the top 40 users.

To measure the performance of the three algorithms, we calculated the correlation between each algorithm and the ground truth ratings. Fig. 6 illustrates the statistical results regarding Spearman's rho. From this figure, we can see that for all algorithms, the correlation decreases when the number of users increases. We can see also that our weighted PageRank algorithm gives a relatively higher correlation than other algorithms, which show that our approach is useful to rank experts than other algorithms.

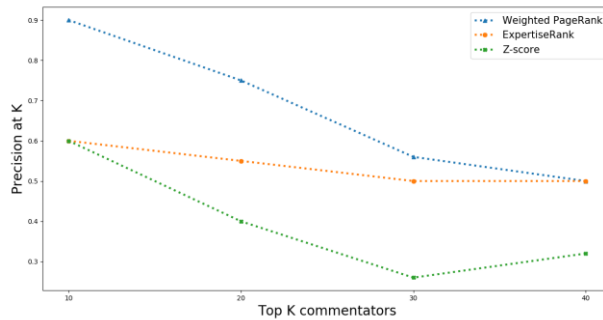


Figure 5. Precision at top K commentators.

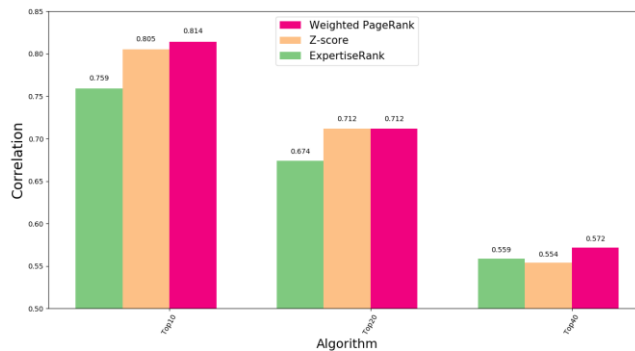
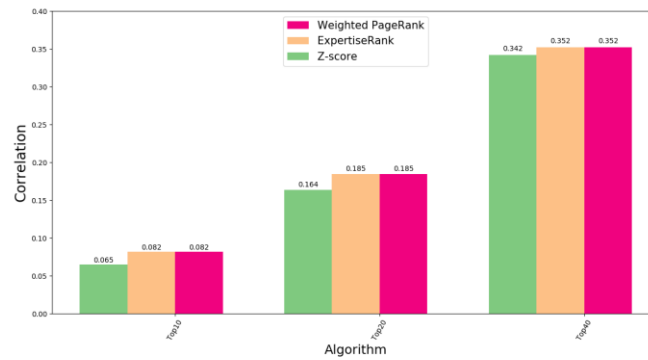


Figure 6. The performance of three algorithms in Spearman's rho distance.

Figure 7. The performance of three algorithms in Kendall tau distance.

From figure 7, we can see that for all algorithms, the Kendall tau increase when the number of users increases. Our weighted PageRank algorithm gives a relatively higher correlation than the ExpertiseRank algorithm and an equal correlation to the Z-score algorithm. From the obtained results, we can see that our weighted PageRank algorithm outperforms the other EF algorithms.



6 Conclusions

In this paper, we proposed a new graph-based EF approach for the citizen science platform, the SPIPOLL. This approach exploits users' comments and users' social relations to predict their expertise on a specific insect family. The relationship between users is calculated according to the exchanged comments. Depending on the insects' identification ease and the length of comments, the relationship between users can increase or decrease. These relationships have been used to construct a weighted graph. Then, a weighted PageRank algorithm has been applied on the obtained graph to rank the users according to their expertise. We evaluated the performance of our method using a dataset from the SPIPOLL database. Experimental results showed that our method could achieve better performance than the state-of-the-art EF algorithms.

References

1. San Pedro, J., & Karatzoglou, A. (2014, October). Question recommendation for collaborative question answering systems with RankSLDA. In Proceedings of the 8th ACM Conference on Recommender systems (pp. 193-200). ACM.
2. Becchetti, L., Castillo, C., Donato, D., Leonardi, S., & Baeza-Yates, R. A. (2006, August). Link-based characterization and detection of web spam. In AIRWeb (pp. 1-8).
3. Gyöngyi, Z., Garcia-Molina, H., & Pedersen, J. (2004, August). Combating web spam with trustrank. In Proceedings of the Thirtieth international conference on Very large data bases-Volume 30 (pp. 576-587). VLDB Endowment.
4. Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). The PageRank citation ranking: Bringing order to the web. Stanford InfoLab.
5. Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5), 604-632.
6. Zhang, J., Ackerman, M. S., & Adamic, L. (2007, May). Expertise networks in online communities: structure and algorithms. In Proceedings of the 16th international conference on World Wide Web (pp. 221-230). ACM.
7. Li, Y., Ma, S., Zhang, Y., & Huang, R. (2012, July). Expertise network discovery via topic and link analysis in online communities. In *Advanced Learning Technologies (ICALT), 2012 IEEE 12th International Conference on* (pp. 311-315). IEEE.
8. Zhao, Z., Zhang, L., He, X., & Ng, W. (2015). Expert finding for question answering via graph regularized matrix completion. *IEEE Transactions on Knowledge and Data Engineering*, 27(4), 993-1004.
9. Zhao, Z., Yang, Q., Cai, D., He, X., & Zhuang, Y. (2016, July). Expert Finding for Community-Based Question Answering via Ranking Metric Network Learning. In *IJCAI* (pp. 3000-3006).
10. Rafiei, M., & Kardan, A. A. (2015). A novel method for expert finding in online communities based on concept map and PageRank. *Human-centric computing and information sciences*, 5(1), 10.
11. Wei, C. P., Lin, W. B., Chen, H. C., An, W. Y., & Yeh, W. C. (2015). Finding experts in online forums for enhancing knowledge sharing and accessibility. *Computers in Human Behavior*, 51, 325-335.
12. Yeniterzi, R., & Callan, J. (2014, July). Constructing effective and efficient topic-specific authority networks for expert finding in social media. In Proceedings of the first international workshop on Social media retrieval and analysis (pp. 45-50). ACM.
13. Zhu, H., Chen, E., Xiong, H., Cao, H., & Tian, J. (2014). Ranking user authority with relevant knowledge categories for expert finding. *World Wide Web*, 17(5), 1081-1107.

14. Procaci, T. B., Siqueira, S. W. M., Braz, M. H. L. B., & de Andrade, L. C. V. (2015). How to find people who can help to answer a question?—Analyses of metrics and machine learning in online communities. *Computers in Human Behavior*, 51, 664-673.
15. Dom, B., Eiron, I., Cozzi, A., & Zhang, Y. (2003, June). Graph-based ranking algorithms for e-mail expertise analysis. In *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery* (pp. 42-48). ACM.
16. Shen, J., Shen, W., & Fan, X. (2009, May). Recommending experts in Q&A communities by weighted HITS algorithm. In *Information Technology and Applications, 2009. IFITA'09. International Forum on* (Vol. 2, pp. 151-154). IEEE.
17. Zhou, Z. M., Lan, M., Niu, Z. Y., & Lu, Y. (2012, April). Exploiting user profile information for answer ranking in cqa. In *Proceedings of the 21st international conference on World Wide Web* (pp. 767-774). ACM.
18. Huang, C., Yao, L., Wang, X., Benatallah, B., & Sheng, Q. Z. (2017, June). Expert as a Service: Software Expert Recommendation via Knowledge Domain Embeddings in Stack Overflow. In *Web Services (ICWS), 2017 IEEE International Conference on* (pp. 317-324). IEEE.
19. Cai, Y., & Chakravarthy, S. (2013, April). Expertise ranking of users in QA community. In *International Conference on Database Systems for Advanced Applications* (pp. 25-40). Springer, Berlin, Heidelberg.
20. Liu, J., Song, Y. I., & Lin, C. Y. (2011, July). Competition-based user expertise score estimation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval* (pp. 425-434). ACM.
21. Aslay, Ç., O'Hare, N., Aiello, L. M., & Jaimes, A. (2013, July). Competition-based networks for expert finding. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (pp. 1033-1036). ACM.
22. Radicchi, F., Fortunato, S., Markines, B., & Vespignani, A. (2009). Diffusion of scientific credits and the ranking of scientists. *Physical Review E*, 80(5), 056103.
23. Życzkowski, K. (2010). Citation graph, weighted impact factors and performance indices. *Scientometrics*, 85(1), 301-315.
24. Aktas, M. S., Nacar, M. A., & Menczer, F. (2004, August). Using hyperlink features to personalize web search. In *International Workshop on Knowledge Discovery on the Web* (pp. 104-115). Springer, Berlin, Heidelberg.
25. Liu, X., Bollen, J., Nelson, M. L., & Van de Sompel, H. (2005). Co-authorship networks in the digital library research community. *Information processing & management*, 41(6), 1462-1480.
26. Xing, W., & Ghorbani, A. (2004, May). Weighted pagerank algorithm. In *Communication Networks and Services Research, 2004. Proceedings. Second Annual Conference on* (pp. 305-314). IEEE.
27. Mihalcea, R. (2005, October). Unsupervised large-vocabulary word sense disambiguation with graph-based algorithms for sequence data labeling. In *Proceedings of the conference*

on Human Language Technology and Empirical Methods in Natural Language Processing
(pp. 411-418). Association for Computational Linguistics.