VeTo: Expert Set Expansion in Academia

Thanasis Vergoulis[1], Serafeim Chatzopoulos[1,2], Theodore Dalamagas[1], and Christos Tryfonopoulos[2]

[1] IMSI, "Athena" Research Center, Athens, 15125, Greece
{vergoulis,schatz,dalamag}@athenarc.gr
[2] Univ. of the Peloponnese, Dep. of Informatics & Tel/tions, Tripoli, 22100, Greece
trifon@uop.gr

Abstract. Expanding a set of known domain experts with new individuals, that have similar expertise, is a problem with many practical applications (e.g., adding new members to a conference program committee). In this work, we study this problem in the context of academic experts and we introduce VeTo, a novel method to effectively deal with it by exploiting scholarly knowledge graphs. In particular, VeTo expands the given set of experts by identifying researchers that share similar publishing habits with them, based on a graph analysis approach. Our experiments show that VeTo is more effective than existing techniques that can be applied to deal with the same problem.

Keywords: Expertise retrieval · Scholarly knowledge graphs.

1 Introduction

Expanding a set of known domain experts with new individuals, that have similar expertise, is a problem that emerges in many real-life applications in academia and industry. For instance, consider a conference organiser that attempts to add new members to the program committee of the conference, since some old members have retired; or consider an officer in a funding agency that seeks new referees to review funding proposals, since some of the current ones are not available. Problems like these motivated the work in the broad area of *expert finding* [10].

Early works in this field assume that the person seeking for experts provides a set of keywords describing the desired topics of expertise. Thus, the proposed expert finding approaches (e.g., [3]) attempt to match these topics to experts by utilising the co-occurrences of topic keywords with person names in text corpora (e.g., websites, publications). However, in many cases it is difficult to explicitly define the desired topics as concrete sets of keywords. To overcome this issue various approaches (e.g., [4,8]) support querying by example: the seeker provides the name of a known expert of the desired field and the approach seeks individuals that seem to have a similar expertise profile. In most cases, the aforementioned profiles are constructed based on analysing the existing text corpora (e.g., applying linguistic processing or topic modeling techniques). Although most such

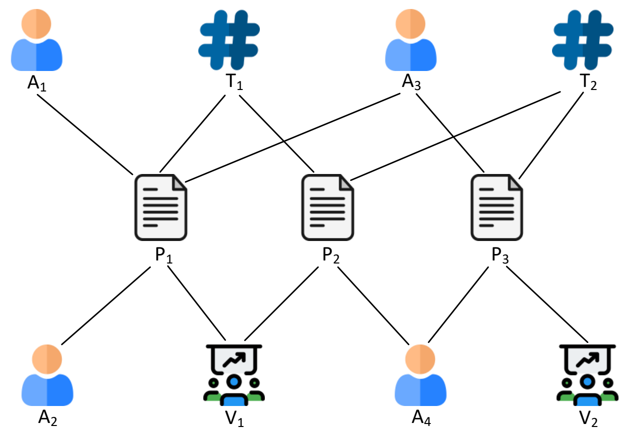


Fig. 1: An example scholarly knowledge graph including academics, papers, venues and topics.

methods search for individuals that are similar to a single expert, some of them are also capable to identify similarities to groups of experts, as well [2].

As it is evident, the effectiveness of all previously described approaches depends on the availability of concrete text corpora that contain information about the expertise of the individuals. In the context of academia this means that these approaches require as input a large set of scientific publications. However, the full texts of publications are often restricted behind paywalls and, thus, it is practically impossible to construct a concrete set of the relevant texts. Moreover, even if it was possible to construct a corpus containing an adequate number of relevant publications, its size would be vast and, thus, gathering and preprocessing it in a regular basis would be a tedious and time-consuming task. This problem motivated the introduction of alternative methods that, instead, utilise *scholarly knowledge graphs* (e.g., [9]). In late years, due to the systematic effort of various developing teams, a variety of large scholarly knowledge graphs has been made available (e.g., the AMiner's DBPL-based datasets [20], the Open Research Knowledge Graph [11], the OpenAIRE Research Graph [12,13]). These heterogeneous graphs consist a very rich and relatively clean source of information about academics, their publications and relevant metadata. Figure 1 presents an illustrative example of such a graph comprising academics, papers, venues, and topics.

In this context, we introduce *VeTo*, a novel, knowledge graph-based approach to deal with the problem of expanding a set of known experts with new individuals with similar expertise. Our approach exploits recent developments in techniques to analyse heterogeneous graphs to identify similarities between researchers based on their publishing habits. In particular, VeTo takes advantage of latent patterns in the way academics select the venue to publish and in the topics of their respective publications.

Our main contributions could be summarised in the following:

- We introduce VeTo, a novel approach that effectively deals with the expert set expansion problem in academia by exploiting scholarly knowledge graphs (Section 3).
- We propose an evaluation framework that could be used to assess the effectiveness of expert set expansion approaches in a fairly objective way (Section 4).
- We exploit the developed framework using as expert sets the lists of program committees of known data management conferences to evaluate the effectiveness of VeTo against competitor methods that could be used to solve the same problem (Section 5).
- We provide the expert sets used for our experiments as open datasets so other researchers could use them as benchmarks following the same framework to evaluate the effectiveness of their own approaches (Section 5).

2 Background

The focus of this work is on a specific expert finding problem applied in the academic world: to reveal, among a set of candidate researchers C, the n most suitable of them, to extend a set of known experts E_{kn}. We refer to this problem as the *expert set expansion* problem, however it is also known as the *finding similar experts* problem (e.g., in [2]). This is a problem with various real-life applications like reviewer recommendation, collaborator seeking, new hire recommendation, etc.

In addition, for reasons elaborated in Section 1, we focus on approaches that exploit scholarly knowledge graphs to deal with the problem. *Knowledge graphs*, also known as *heterogeneous information networks* [19], are graphs that contain nodes and edges of multiple types capturing knowledge about entities (nodes) and the different types of relationships between them (edges). For instance, consider the scholarly knowledge graph illustrated in Figure 1. This graph contains information about 3 papers (P_1, P_2, P_3), their venues (V_1, V_2), their topics (T_1, T_2), and the academics that have authored them ($A_1, \ldots, A_4$). Of course, real-life scholarly knowledge graphs contain a larger variety of entity types (e.g., academic institutions, funding organisations, research projects, as well).

Knowledge graphs capture rich information about their respective domains encoding not only direct relationships of the involved entities, but also more complex ones that correspond to larger paths in the graph. In particular, all paths that involve the same sequence of entity and edge types capture relationships of exactly the same semantics between their first and last nodes. These generalised path patterns are widely known as *metapaths* and we refer to the paths that follow these patterns as their *instances*. For example, in the graph of Figure 1, the paths $A_1 - P_1 - T_1$ and $A_3 - P_3 - T_2$ both are instances of the metapath `Academic`

`- Paper - Topic` (or `APT`, for brevity) and both have the same interpretation: they relate an academic with a topic through a paper authored by her.

In recent literature, the similarity of two entities (nodes) of the same type according to the semantics of a particular metapath is measured using the number of instances of this metapath that connect these nodes with nodes of the last node type of the metapath. For example, academics A_1 and A_2 seem similar based on the topics of their published papers (i.e., based on the semantics of the `APT` metapath) since they both have only one paper connecting them to the T_1 topic (i.e., 1 `APT` instance) and no paper connecting them to the T_2 topic (i.e., 0 `APT` instances).

A well-known metapath-based similarity measure that follows the previous intuition is JoinSim [21]. This measure calculates cosine similarity on node feature vectors based on the relationships indicated by a given metapath. For instance, given the metapath `APT`, JoinSim first constructs for each academic a feature vector with the topics related to the papers she has authored, and then calculates cosine similarity scores between the academics based on these vectors.

3 Our approach

3.1 The intuition

The main intuition behind VeTo, our approach, is that it deals with the expert set expansion problem by considering the similarities of academics to the experts based on a metapath-based similarity of academics, according to 2 particular metapaths, `APT` and `APV`, that capture interesting "publishing habits". In particular, the former considers the venues in which the compared academics select to publish their articles, while the latter the topics of their published articles.

3.2 Formal description

Given a set of known experts E_{kn}, a set of candidates C, and the number of expansions that need to be performed n, VeTo performs the following steps:

1. For each expert $e \in E_{kn}$ its similarity scores to all candidates $c \in C$ according to the `APV` metapath are calculated and C^{e}_{APV}, the ranked list of all candidates based on these scores is produced.
2. A rank aggregation algorithm is applied on the C^{e}_{APV} for all $e \in E_{kn}$ to produce C_{APV}, the aggregated ranked list that ranks all candidates considering their similarities to all experts according to `APV`.
3. A procedure similar to the one performed in Steps 1 & 2 is followed to produce the ranked list C_{APT} that ranks all candidates according to their "aggregated" similarity based on the `APV` metapath.
4. A rank aggregation algorithm is applied on C_{APV} and C_{APV} to produce an aggregated ranked list C_{fin} that takes into account the similarities between experts and other academics based on both metapaths.

5. The n most similar items of C_{fin} as an answer to the given expert set expansion problem instance.

All metapath-based similarities required by our approach are calculated using the JoinSim [21] algorithm (see also Section 2). Regarding the rank aggregation algorithm, required in many steps of the approach, in this work, we select to use the *Borda Count* approach. Based on this approach, the aggregation of two ranked lists of size n is performed as follows: A score of n is assigned to the first element of each list, $n-1$ to the second, and so on. Then, for each of the n elements, the two scores (one for each list) are added to produce the final score for this element. Finally, the elements are being sorted in descending order based on the aggregated scores.

It should be noted that although the proposed approach is tailored to the problem of expert finding in academia, it could also be adapted and applied in other domains given the appropriate knowledge graphs and metapaths.

4 Proposed Evaluation Framework

A common issue in various expert finding problems is that it is not easy to evaluate the effectiveness of a given approach, since it is impossible to construct an objective ground truth. Luckily, in the context of expert set expansion, it was possible for us to develop an evaluation framework that can be used to assess the effectiveness of an approach based on a fairly objective ground truth. To the best of our knowledge, this is the first time that this evaluation framework is used for this problem.

The intuition behind this framework is to gather available expert lists from real-life applications (e.g., the PC members of a conference, editorial boards of journals) and, then, use each of them as dataset for a k-fold cross validation process. This means that, for each expert list E, a given approach is assessed as follows:

1. E is shuffled and, then, split in k disjoint sets $E_1, \ldots, E_k$, all of equal size[3] $n = \lfloor |E|/k \rfloor$.
2. For each E_i (with $i \in [1, k]$), a pair of training and testing set $\{E_i^{train}, E_i^{test}\}$ is constructed, where $E_i^{train} = \bigcup_{j \neq i} E_j$ and $E_i^{test} = E_i$.
3. For each $\{E_i^{train}, E_i^{test}\}$ pair:
   - we use E_i^{train} as the set of known experts (i.e., $E_{kn} = E_i^{train}$)
   - we apply the expert set expansion approach on E_i^{train} and get O_i, its output
   - we examine false & true positives and negatives in the top-x items of O_i based on E_i^{test} and we calculate proper information retrieval measures based on them, for $x \in [1, n]$ (where, $n = |O_i| = |E_i^{test}|$).

[3] The last one may be larger than the others, however it is easy to take this into consideration.

Regarding the information retrieval measures that are suitable to be used in Step 3 of the aforementioned process, we propose the use of top-x precision, recall, and F_1 score that can be defined as follows:

$$Precision = \frac{|O_i \cap E_i^{Test}|}{x}, \quad Recall = \frac{|O_i \cap E_i^{Test}|}{n},$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

The larger the values of these measures are, the better the effectiveness of the method based on the given list E at the corresponding measuring point x is. The values of all measuring points could be used to construct a line plot.

Moreover, after completing the previous process for E, we propose to also calculate, for the same expert set, the mean reciprocal rank (MRR) based on all outputs O_i (for all $i \in [1, k]$) which can be calculated as follows:

$$MRR = \frac{1}{k} \sum_{i=1}^{k} \frac{1}{rank_i}$$

where, $rank_i$ refers to the rank position of the first true positive element in the output O_i.

The described evaluation framework was used for the experiments presented in Section 5. In particular, we gathered the list of program committee members for two well-known data management conferences (SIGMOD & VLDB) and applied the process of the framework on both of them using the aforementioned information retrieval measures. Gathering the program committee data was relatively easy by applying a semi-automatic approach that utlises Web page scrapping tools. In fact, our collected data could be used by third parties as benchmarks to evaluate the effectiveness of their own expert set expansion approaches. This is why we provide them as open datasets (more details in Section 5.1).

5 Evaluation

In this section we describe the experiments we have conducted to evaluate the effectiveness of our approach. In Section 5.1 we describe the experimental setup and in Section 5.2 we present our findings.

5.1 Setup

Approaches The evaluation involves four different approaches to provide answer to the expert set expansion problem.

- *VeTo*, our proposed approach which exploits academics similarity according to the `APV` and `APT` metapaths (see also Section 3).

- *Baseline*, an approach that counts the number of papers an academic has published in the corresponding conference, ranks academics based on this number, and provides the top academics as the most suitable expansions.
- *ADT*, the best performing graph-based approach proposed in [9], that attempts to capture the association strength between two academics by considering the paths that relate them to topics (based on their papers) according to the ProductPaths technique.
- *WG*, a graph-based approach proposed in [2], which exploits working groups to capture similarity; in our context working groups correspond to co-authorship relations among academics[4].

The basic implementations of all approaches were written in Python, however, part of the preprocessing was implemented in C++ for improved efficiency. In addition, JoinSim [21] scores were calculated using the open entity similarity Java library HeySim[5].

Datasets For our experiments, we used the following sets of data:

- *DBLP Scholarly Knowledge Graph (DSKG) dataset.* It contains data for approximately 1.5M academics, their papers in the period 2000-2017, the corresponding venues and the involved topics. DSKG is based on the AMiner's DBLP citation network [20], enriched with topics assigned to papers by the CSO Classifier [16,15] (based on their abstracts). Finally, DSKG contains approximately 3.9M and 34.1M `APV` and `APT` metapath instances, respectively.
- *Program Committees (PC) dataset.* It contains program committee data from two well-known conferences from the field of data management: the ACM SIGMOD conference and the VLDB conference. The data were gathered by scrapping the official Web pages of these conferences for the years $2007-2017$ and, then, applying a semi-automatic cleaning process to properly map the PC members to academics in the DSKG dataset.

The DSKG dataset was used as a knowledge base that the various approaches could take advantage of. The PC dataset, on the other hand, was used to create the required training and testing sets for the evaluation based on the framework described in Section 4). This latter dataset was also made openly available at Zenodo[6] so other researchers could use it as benchmark to assess the effectiveness of their own approaches.

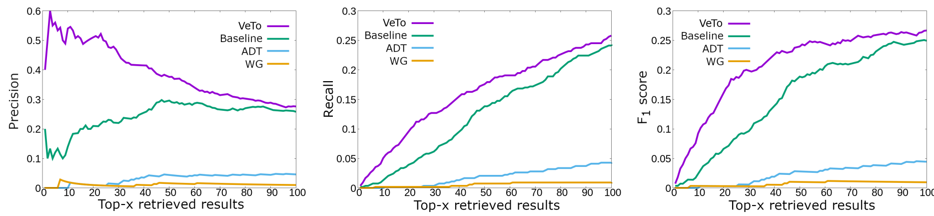
5.2 Evaluation of VeTo against competitors

In this experiment, we compare the effectiveness of our approach against its rivals based on the framework discussed in Section 4 by using both expert sets in the PC dataset (SIGMOD and VLDB).

[4] We have also conducted experiments using *DOC*, the alternative graph-based approach proposed in the same paper, however it performed worse in all cases and its results were omitted from the experimental section for presentation reasons.

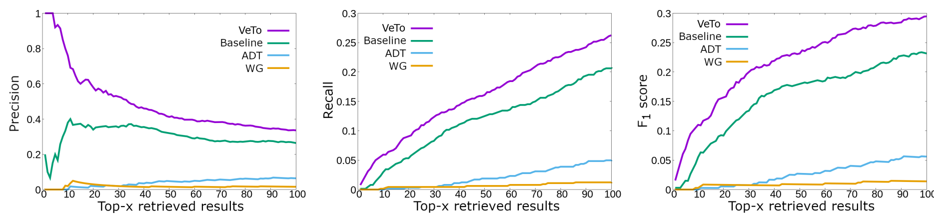
[5] `https://github.com/schatzopoulos/HeySim`

[6] `https://doi.org/10.5281/zenodo.3739316`



(a) Precision (b) Recall (c) F_1 score

Fig. 2: Evaluation against competitors for SIGMOD conference.



(a) Precision (b) Recall (c) F_1 score

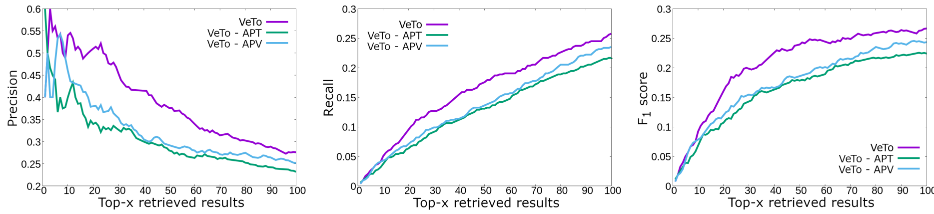
Fig. 3: Evaluation against competitors for VLDB conference.

Top-x precision, recall & F_1-score Figures 2 and 3 present the precision, recall and F_1 score of all compared approaches for SIGMOD and VLDB expert sets, respectively. Larger values for all measures indicate superior effectiveness. It is evident that VeTo clearly outperforms its competitors in all scenarios. More importantly, in both datasets, it achieves notably higher precision in comparison to all other approaches for at least the top-40 results. The latter fact is really important since, in practice, for most real-life applications of the expert set expansion problem, usually n is relatively small.

Furthermore, it should be noted that the baseline approach seems to work pretty well (but, at the same time, significantly worse than VeTo) in most cases. This result indicates that there is a correlation between the academics that publish articles in a conference and its program committee members. On the other hand, both ADT and WG do not perform well.

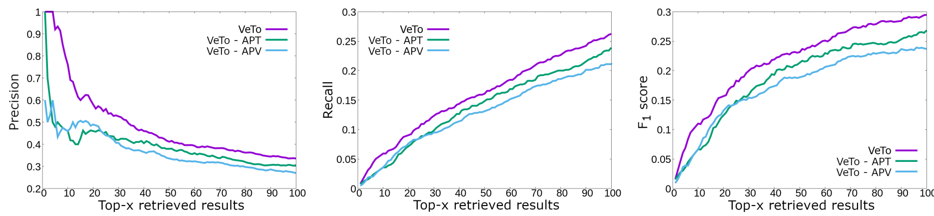
Table 1: MRR based on the folds of each dataset

	Baseline	**ADT**	**WG**	**VeTo**
SIGMOD	0.323	0.043	0.039	0.8
VLDB	0.357	0.046	0.061	1
Total	0.34	0.0445	0.05	0.9



(a) Precision (b) Recall (c) F_1 score

Fig. 4: Comparison of different variants of our method for SIGMOD conference.



(a) Precision (b) Recall (c) F_1 score

Fig. 5: Comparison of different variants of our method for VLDB conference.

MRR per conference Table 1 includes the assessment of all approaches based on the mean reciprocal rank (MRR) for both expert sets (SIGMOD and VLDB) separately and in total (if we use simultaneously all their folds). Larger values of MRR indicate better approach effectiveness. The results are in compliance with the previous experiment: since VeTo achieves significantly larger precision for small values of x, it performs significantly better than its competitors in terms of MRR (see also MRR definition in Section 4). Again ADT and WG perform significantly worse than the baseline.

5.3 Studying & configuring VeTo

In this section, we examine different configurations of our approach and we investigate the effect they have in its effectiveness.

Table 2: MRR of different variants based on the folds of each dataset

	VeTo-APT	**VeTo-APV**	**VeTo**
SIGMOD	0.766	0.766	0.8
VLDB	0.8	0.8	1
Total	0.783	0.783	0.9

Table 3: Top-10 recommendations per configuration (1^{st} fold)

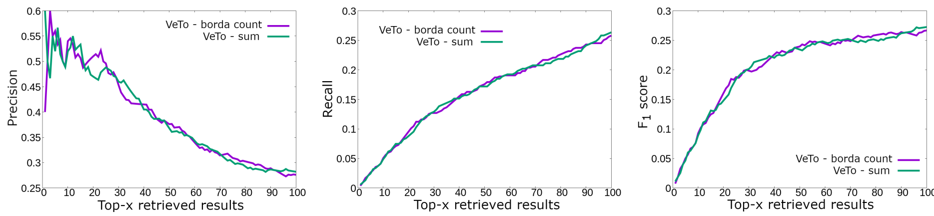
	SIGMOD		
	VeTo-APT	**VeTo-APV**	**VeTo**
1	Jeffrey F. Naughton	Dong Deng	Jeffrey F. Naughton
2	Beng Chin Ooi*	Jeffrey F. Naughton	Beng Chin Ooi*
3	Neoklis Polyzotis*	Ihab F. Ilyas*	Ihab F. Ilyas*
4	Guoren Wang	Jennifer Widom	Neoklis Polyzotis*
5	Ihab F. Ilyas*	Beng Chin Ooi*	Jennifer Widom
6	Dongqing Yang	Philip Bohannon*	David J. DeWitt*
7	Wolfgang Lehner*	David J. DeWitt*	Volker Markl*
8	Stéphane Bressan	Michael J. Carey	Raghu Ramakrishnan
9	Ge Yu	Neoklis Polyzotis*	Michael J. Carey
10	Marios Hadjieleftheriou*	Lijun Chang	Ashraf Aboulnaga*

	VLDB		
	VeTo-APT	**VeTo-APV**	**VeTo**
1	Dan Suciu*	Yannis Papakonstantinou*	Yannis Papakonstantinou*
2	Guoren Wang	Dong Deng	Christoph Koch*
3	Christoph Koch*	Jiannan Wang	Jennifer Widom*
4	Dongqing Yang	Jennifer Widom*	Volker Markl*
5	Timos K. Sellis*	Mourad Ouzzani*	Dan Suciu*
6	Ge Yu*	Renée J. Miller	Shivnath Babu*
7	Vassilis J. Tsotras	Philip Bohannon	Bolin Ding*
8	Xiaofeng Meng	Bolin Ding*	Renée J. Miller
9	Nikos Mamoulis*	Paolo Papotti*	Mourad Ouzzani*
10	Tengjiao Wang	Lu Qin	Marios Hadjieleftheriou

Studying the effect of the used metapaths VeTo's approach considers similarities of academics based on two criteria: their similarity based on the venues they prefer to publish (captured by the `APV` metapath) and on the topics of their published papers (captured by the `APT` metapath). In this experiment we examine the effect of each of these metapaths by implementing two VeTo's variants: one that considers only the `APV` metapath (called VeTo-APV) and a second one that considers only the `APT` metapath (called VeTo APT).

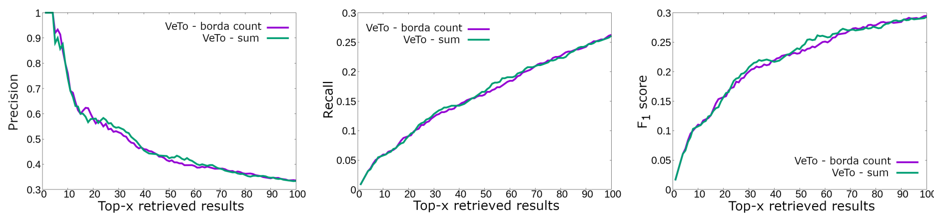
Figures 4 and 5 illustrate the measured top-x precision, recall and F_1 score of VeTo, VeTo-APV, and VeTo-APT for SIGMOD and VLDB, respectively, while Table 2 summarizes the corresponding MRR scores. It is evident that VeTo outperforms its two variants in all cases, however the variants usually achieve comparable (but always worse) effectiveness.

It should be noted that VeTo-APT achieves slightly higher precision and recall in most cases in the SIGMOD dataset, while the other variant is usually slightly better for VLDB. Moreover, in Table 3 we present the top-10 results provided by VeTo, VeTo-APV, and VeTo-APT based on the first fold of each expert set experiment. With asterisk we indicate all true positives. It is evident that, although both metapaths provide some common top suggestions (e.g., Ihab F. Ilyas in SIGMOD), they also identify some unique correct results (e.g., David



(a) Precision (b) Recall (c) F_1 score

Fig. 6: Comparison of different rank aggregation methods for SIGMOD conference.



(a) Precision (b) Recall (c) F_1 score

Fig. 7: Comparison of different rank aggregation methods for VLDB conference.

J. DeWitt provided by VeTo-APV in SIGMOD) that the other metapath fails to bring. These findings indicate that both metapaths are capable to identify some unique good results, thus VeTo's approach to combine both of them has a potential to achieve improved performance (as is confirmed by our experiments).

Studying the effect of different rank aggregations Part of VeTo's approach consists of using a rank aggregation algorithm. Our default selection in our implementation is Borda Count (see also Section 3). In this section, we examine the effect that the use of an alternative rank aggregation algorithm would have. We do so by implementing a variant that instead computes the similarity of a candidate as the sum of its similarities with the experts in the test set. This is a common rank aggregation algorithm used in various works (e.g., it is also used for JoinSim [2]). In Figures 6 and 7 we present the top-x precision, recall, and F1-measure of this variant in comparison to the same measurements for the basic VeTo implementation that uses Borda Count. It is evident that no significant differences are observable.

6 Related Work

Expertise retrieval consists an interesting field of research in many disciplines like digital libraries, data management, information retrieval, and machine learning.

A wide range of problems, ranging from expert finding to expert profiling, belong in this field and there are many related real-time applications (e.g., collaboration recommendation, reviewer recommendation). A detailed review of the field is beyond the scope of the current work. The reader interested could refer to the excellent survey in [10]. In the next we will focus on the variations of the *expert finding* problem.

Finding experts for a given topic in the industry has been a relatively well-studied problem. Initial approaches relied on manually curated databases of skills and knowledge (e.g., [6]), however the interest quickly shifted to approaches that extract employee's expertise from document collections that could be found within corporate intranets or the Web [3,5]. A common platform to empirically assess such approaches has been developed by the TREC community[7] facilitating the development of various relevant methods [1,7,14,17]. Apart from details about the exact expert finding problems solved by each of the previous methods, VeTo significantly differs from these works in principle, since it is tailored for academic experts and since it does not rely on document collections because such collections are often available due to the existing paywalls.

Finding experts in academia, where the experts are researchers with knowledge and interests in a given topic, has also been an important field (e.g., [22,18]). However, most of these methods also rely on scientific text corpora which are often limited behind paywalls. Most relevant to VeTo are methods that try to exploit scholarly knowledge graphs to perform the same tasks (e.g., [9]). However, in contrast to them, VeTo takes advantage of recent developments in the field of heterogeneous information networks and knowledge graphs.

7 Conclusions

In this work, we study the expert set expansion problem for academic experts, i.e., given a set of known experts to find the n most suitable candidates to expand this set. In this context, we introduced VeTo, a set expert expansion approach for academic experts that exploits information from a given scholarly knowledge graph to estimate similarities between academics based on their publishing habits. Moreover, we introduce a new evaluation framework that can assess the effectiveness of such approaches in a fairly objective way. Finally, we utilise the developed framework to compare VeTo against a set of competitors showing that it is superior in terms of effectiveness.

Acknowledgments

We acknowledge support of this work by the project "Moving from Big Data Management to Data Science" (MIS 5002437/3) which is implemented under the Action "Re-inforcement of the Research and Innovation Infrastructure", funded

[7] `https://trec.nist.gov/`

by the Operational Programme "Competitiveness, Entrepreneurship and Innovation" (NSRF 2014-2020) and co-financed by Greece and the European Union (European Regional Development Fund).

Icons in Figure 1 were collected from www.flaticon.com and were made by Freepik, Good Ware and Pixel perfect .

References

1. Balog, K., Azzopardi, L., de Rijke, M.: Formal models for expert finding in enterprise corpora. In: SIGIR 2006: Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Seattle, Washington, USA, August 6-11, 2006. pp. 43–50 (2006)
2. Balog, K., de Rijke, M.: Finding similar experts. In: Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. p. 821–822. SIGIR '07, Association for Computing Machinery, New York, NY, USA (2007). https://doi.org/10.1145/1277741.1277926, `https://doi.org/10.1145/1277741.1277926`
3. Cao, Y., Liu, J., Bao, S., Li, H.: Research on expert search at enterprise track of TREC 2005. In: Proceedings of the Fourteenth Text REtrieval Conference, TREC 2005, Gaithersburg, Maryland, USA, November 15-18, 2005 (2005), `http://trec.nist.gov/pubs/trec14/papers/microsoft-asia.ent.pdf`
4. Chen, H.H., Gou, L., Zhang, X., Giles, C.L.: Collabseer: A search engine for collaboration discovery. In: Proceedings of the 11th Annual International ACM/IEEE Joint Conference on Digital Libraries. p. 231–240. JCDL '11, Association for Computing Machinery, New York, NY, USA (2011). https://doi.org/10.1145/1998076.1998121, `https://doi.org/10.1145/1998076.1998121`
5. Craswell, N., Hawking, D., Vercoustre, A.M., Wilkins, P.: P@noptic expert: Searching for experts not just for documents. In: Ausweb Poster Proceedings, Queensland, Australia. vol. 15, p. 17 (2001)
6. Davenport, T.H., Prusak, L.: Working knowledge: how organizations manage what they know. Ubiquity **2000**(August), 6 (2000)
7. Fang, H., Zhai, C.: Probabilistic models for expert finding. In: Advances in Information Retrieval, 29th European Conference on IR Research, ECIR 2007, Rome, Italy, April 2-5, 2007, Proceedings. pp. 418–430 (2007)
8. Gollapalli, S.D., Mitra, P., Giles, C.L.: Similar researcher search in academic environments. In: Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries. p. 167–170. JCDL '12, Association for Computing Machinery, New York, NY, USA (2012). https://doi.org/10.1145/2232817.2232849, `https://doi.org/10.1145/2232817.2232849`
9. Gollapalli, S.D., Mitra, P., Giles, C.L.: Ranking experts using author-document-topic graphs. In: 13th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '13, Indianapolis, IN, USA, July 22 - 26, 2013. pp. 87–96 (2013). https://doi.org/10.1145/2467696.2467707
10. Gonçalves, R., Dorneles, C.F.: Automated expertise retrieval: A taxonomy-based survey and open issues. ACM Comput. Surv. **52**(5), 96:1–96:30 (2019)
11. Jaradeh, M.Y., Oelen, A., Farfar, K.E., Prinz, M., D'Souza, J., Kismihók, G., Stocker, M., Auer, S.: Open research knowledge graph: Next generation infrastructure for semantic scholarly knowledge. In: Proceedings of

the 10th International Conference on Knowledge Capture, K-CAP 2019, Marina Del Rey, CA, USA, November 19-21, 2019. pp. 243–246 (2019). https://doi.org/10.1145/3360901.3364435

12. Manghi, P., Atzori, C., Bardi, A., Shirrwagen, J., Dimitropoulos, H., La Bruzzo, S., Foufoulas, I., Löhden, A., Bäcker, A., Mannocci, A., Horst, M., Baglioni, M., Czerniak, A., Kiatropoulou, K., Kokogiannaki, A., De Bonis, M., Artini, M., Ottonello, E., Lempesis, A., Nielsen, L.H., Ioannidis, A., Bigarella, C., Summan, F.: Openaire research graph dump (Dec 2019). https://doi.org/10.5281/zenodo.3516918, `https://doi.org/10.5281/zenodo.3516918`
13. Manghi, P., Bardi, A., Atzori, C., Baglioni, M., Manola, N., Schirrwagen, J., Principe, P.: The openaire research graph data model (Apr 2019). https://doi.org/10.5281/zenodo.2643199, `https://doi.org/10.5281/zenodo.2643199`
14. Petkova, D., Croft, W.B.: Hierarchical language models for expert finding in enterprise corpora. International Journal on Artificial Intelligence Tools **17**(1), 5–18 (2008)
15. Salatino, A., Osborne, F., Thanapalasingam, T., Motta, E.: The CSO Classifier: Ontology-Driven Detection of Research Topics in Scholarly Articles, pp. 296–311 (08 2019)
16. Salatino, A.A., Thanapalasingam, T., Mannocci, A., Osborne, F., Motta, E.: The computer science ontology: a large-scale taxonomy of research areas. In: International Semantic Web Conference. pp. 187–205. Springer (2018)
17. Serdyukov, P., Hiemstra, D.: Modeling documents as mixtures of persons for expert finding. In: Advances in Information Retrieval , 30th European Conference on IR Research, ECIR 2008, Glasgow, UK, March 30-April 3, 2008. Proceedings. pp. 309–320 (2008). https://doi.org/10.1007/978-3-540-78646-7_29, `https://doi.org/10.1007/978-3-540-78646-7_29`
18. Sfyris, G.A., Fragkos, N., Doulkeridis, C.: Profile-based selection of expert groups. In: International Conference on Theory and Practice of Digital Libraries. pp. 81–93. Springer (2016)
19. Shi, C., Li, Y., Zhang, J., Sun, Y., Yu, P.S.: A survey of heterogeneous information network analysis. IEEE Trans. Knowl. Data Eng. **29**(1), 17–37 (2017). https://doi.org/10.1109/TKDE.2016.2598561, `https://doi.org/10.1109/TKDE.2016.2598561`
20. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., Su, Z.: ArnetMiner: Extraction and Mining of Academic Social Networks. In: Proceedings of the 14th ACM SIGKDD. pp. 990–998. ACM (2008)
21. Xiong, Y., Zhu, Y., Yu, P.S.: Top-k similarity join in heterogeneous information networks. IEEE Transactions on Knowledge and Data Engineering **27**(6), 1710–1723 (2015)
22. Yang, K., Kuo, T., Lee, H., Ho, J.: A reviewer recommendation system based on collaborative intelligence. In: 2009 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2009, Milan, Italy, 15-18 September 2009, Main Conference Proceedings. pp. 564–567 (2009). https://doi.org/10.1109/WI-IAT.2009.94, `https://doi.org/10.1109/WI-IAT.2009.94`