VeTo-web: A Recommendation Tool for the Expansion of Sets of Scholars

Serafeim Chatzopoulos
Dep. of Informatics & Tel/tions
Univ. of the Peloponnese
Tripoli, Greece
schatzop@uop.gr

Thanasis Vergoulis
IMSI
ATHENA RC
Athens, Greece
vergoulis@athenarc.gr

Theodore Dalamagas
IMSI
ATHENA RC
Athens, Greece
dalamag@athenarc.gr

Christos Tryfonopoulos
Dep. of Informatics & Tel/tions
Univ. of the Peloponnese
Tripoli, Greece
trifon@uop.gr

***Abstract*—Expanding a set of known experts with new ones that share similar expertise is a problem that emerges in various real-life applications. We demonstrate VeTo-web, an open source, publicly available tool that deals with this problem in the context of searching for academic experts. VeTo-web exploits analysis techniques for scholarly knowledge graphs to identify scholars that share similar research activities with a given expert group and offers a Web-based user interface to assist its users in expanding a set of academic experts with additional scholars with similar expertise.**

***Keywords*-expert finding, scholarly knowledge graphs**

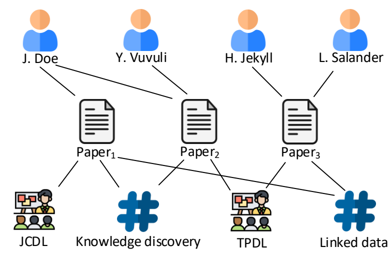

Figure 1. An example SKG including scholars, papers, venues, and topics.

I. Introduction

Searching for individuals that share similar expertise with a set of known experts is a problem with various practical applications, many of which coming from the field of academia. For example, consider a conference planner searching for additional reviewers. Although there is a significant amount of work in the broad field of *expert finding* [1], this exact problem, known as *expert set expansion*, has only recently received more attention [2], [3].

Focusing on uses for the academic world, in a previous work we introduced VeTo [3], an expert set expansion approach that leverages the information included in *Scholarly Knowledge Graphs (SKGs)*. SKGs are graphs comprising different types of entities (nodes) and relationships (edges). Figure 1 illustrates an example SKG capturing scholarly data that consists of entities for `Scholars` (or `S`, for brevity), `Papers` (`P`), `Venues` (`V`), and `Topics` (`T`) and the relationships between them. There are three types of relationships in this network: between authors and papers (denoted as `SP` or `PS`), papers and venues (denoted as `PV` or `VP`) and, finally, between papers and topics, denoted as `PT` or `TP`.

Contrary to the majority of works in the field of expert finding that utilise text corpora, VeTo exploits the structured information of SKGs to identify similarities between scholars based on their research activity. In particular, VeTo leverages latent patterns in the way that scholars select to publish their work, i.e., the venues they choose to publish and the topics of their papers. Our experiments [3] showed that VeTo outperformed other relevant approaches in terms of recommendation accuracy, giving more precise suggestions for the expansion of conference program committees.

To find similar scholars, VeTo leverages the *metapath-based similarities* of SKG nodes. In SKGs (and KGs, in general), each path represents a complex relationship between two nodes (the first and the last one) having a very specific interpretation. In fact, the interpretation of such relationships is determined by the sequence of node and edge types in the corresponding path. Thus, all SKG paths with the same node and edge type sequence share common semantics. In the literature, the sequences of nodes and edges are known as *metapaths*. Here, we follow a common notation simplification assumption: we assume that no multiple edge types exist between the same pair of node types; hence, we denote each metapath by the sequence of node types involved. For instance, in the SKG of Figure 1, the `SPVPS` metapath relates scholars that have published papers in the same venues, while `SPTPS` relates those scholars with papers in the same topics. According to the concept of metapath-based similarity, two SKG nodes are similar if they are strongly connected based on a given metapath of interest. Indicatively, in Figure 1, "Y. Vuvuli" and "L. Salander" are similar based on `SPVPS` as they both have a paper in the TPDL conference but they have no similarity based on `SPTPS` as their papers are in different topics. VeTo combines `SPVPS`- and `SPTPS`-based similarities (i.e., venue- and topic-based) for scholars to identify matches between an already known expert group and a group of candidates for expansion. The candidates with the largest similarity are those provided as expansion recommendations.

In this work, we demonstrate VeTo-web, an open source[1], Web-based tool that leverages VeTo to provide expansion recommendations for a given expert (scholar) set. A prototype version of the tool, built on top of AMiner's DBLP Citation Network [4][2], has been deployed and is publicly available[3] to demonstrate VeTo's effectiveness in expert set expansion applications for scholars.

II. System Overview

VeTo-web consists of four components: (a) the metapath-based transformator, (b) the entity similarity calculator, (c) the VeTo engine, and (d) the Web UI. The *metapath-based transformator* calculates all pairs of nodes connected through metapaths `SPVPS` and `SPTPS` along with the number of paths connecting each pair. It essentially produces two 'views' over the initial SKG: one connecting scholars with the topics of their papers and another one with relations between authors and the venues they have published. The component uses the adjacency matrix representation to encode SKGs, thus the transformation is implemented as a matrix multiplication operation [5]. Since adjacency matrices are inherently sparse, the component also utilises sparse matrix representations[4] to speedup the computations. The *entity similarity calculator* takes as input the SKG views from the previous component and for each expert in the initial expert set produces two lists of similar experts, based on metapaths `SPVPS` and `SPTPS`, respectively (JoinSim [6] similarity measure is used). The *VeTo engine* is the core of our tool, implementing the algorithm presented in [3]. It combines the similarity-based ranked lists of experts from the previous component using the borda count rank aggregation scheme. The final two lists, based on topic and venue similarities are further multiplied with user-defined weights before being sorted to produce the final unified list. Finally, the *Web UI*, implemented using React JS, provides the functionalities described in Section III.

III. Functionalities & Demonstration Scenario

Figure 2 presents a screenshot of VeTo-web's UI. To perform a new expert set expansion, the user first selects the scholars they want to constitute the initial expert set. To select the desired scholars, the user can either give individual expert names in the input element on the top left corner of the page (which supports auto-completion) or upload a file containing multiple author names (one per line). Next, the user can adjust the significance weights of the metapaths (used to determine similarity between scholars) using the appropriate slider and clicks the "execute" button. A progress bar appears, indicating the step of the process being executed and after the process is completed, the

[1]https://github.com/schatzopoulos/veto
[2]The exact details for the used SKG dataset can be found in [3].
[3]http://veto.imsi.athenarc.gr
[4]Eigen linear algebra library: http://eigen.tuxfamily.org

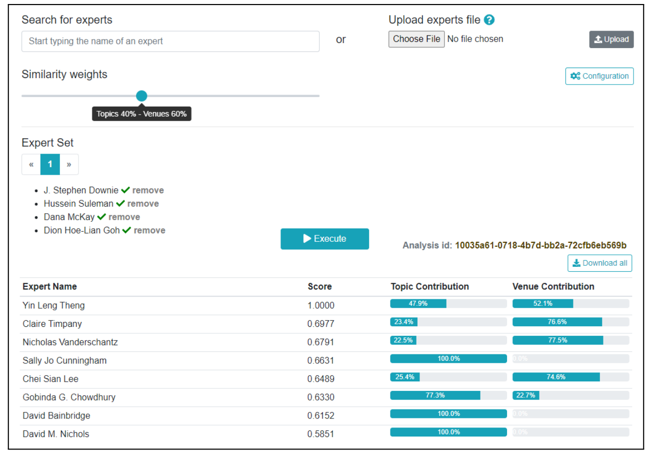

Figure 2. Screenshot from VeTo-web's Web UI.

results appear in a tabular form sorted based on VeTo's normalised score. The contribution of each metapath in the final similarity score is also indicated.

During the demonstration session, we will run the following scenario: A user is interested in adding new members in the Organizing Committee of JCDL 2021. Therefore, they provide the four names depicted in Figure 2 as the input expert set. Additionally, they adjust the similarity weights to 40% for topics and 60% for venues. Interestingly, VeTo-web returns "Claire Timpany" and "David M. Nichols" as suggested expert expansions, that are indeed members of the Organising Committee of JCDL 2021. In addition, we will be ready to run any other scenario requested by the audience.

Acknowledgment

This work was partially funded by the EU H2020 project SmartDataLake (825041).

References

[1] R. Gonçalves and C. F. Dorneles, "Automated expertise retrieval: a taxonomy-based survey and open issues," *ACM Computing Surveys (CSUR)*, vol. 52, no. 5, pp. 1–30, 2019.

[2] K. Balog and M. De Rijke, "Finding similar experts," in *Proc. of the 30th ACM SIGIR*, 2007, pp. 821–822.

[3] T. Vergoulis, S. Chatzopoulos, T. Dalamagas, and C. Tryfonopoulos, "Veto: Expert set expansion in academia," in *Proc. of TPDL*. Springer, 2020, pp. 48–61.

[4] J. Tang, J. Zhang, L. Yao, J. Li, L. Zhang, and Z. Su, "Arnetminer: extraction and mining of academic social networks," in *Proc. of the 14th ACM SIGKDD*, 2008, pp. 990–998.

[5] C. Shi, Y. Li, S. Y. Philip, and B. Wu, "Constrained-meta-path-based ranking in heterogeneous information network," *Knowledge and Information Systems*, vol. 49, no. 2, pp. 719–747, 2016.

[6] Y. Xiong, Y. Zhu, and S. Y. Philip, "Top-k similarity join in heterogeneous information networks," *IEEE TKDE*, vol. 27, no. 6, pp. 1710–1723, 2014.