WSDM Cup 2016 – Entity Ranking Challenge

Alex D. Wade
Microsoft Research
One Microsoft Way
Redmond, WA USA
awade@microsoft.com

Kuansan Wang
Microsoft Research
One Microsoft Way
Redmond, WA USA
kuansanw@microsoft.com

Yizhou Sun
Northeastern University
360 Huntington Avenue
Boston, MA USA
yi.sun@neu.edu

Antonio Gulli
Elsevier B.V.
Radarweg 29, 1043 NX
Amsterdam The Netherlands
a.gulli@elsevier.com

ABSTRACT
In this paper, we describe the WSDM Cup entity ranking challenge held in conjunction with the 2016 Web Search and Data Mining conference (WSDM 2016). Participants in the challenge were provided access to the Microsoft Academic Graph (MAG), a large heterogeneous graph of academic entities, and were invited to calculate the query-independent importance of each publication in the graph. Submissions for the challenge were open from August through November 2015, and a public leaderboard displayed teams' progress against a set of training judgements. Final evaluations were performed against a separate, withheld portion of the evaluation judgements. The top eight performing teams were then invited to submit papers to the WSDM Cup workshop, held at the WSDM 2016 conference.

Keywords
Heterogeneous graphs; Microsoft Academic; Microsoft Academic Graph

1. INTRODUCTION
The recent explosive growth of data recording our daily online activities are often manifested as heterogeneous graphs, ranging from Facebook's Open Graph that records our social and communication activities to the graphs gathered by major search engine companies that represent a snapshot of our collective knowledge. As demonstrated in many web search and data mining applications (e.g., [1] [2] [3]), a critical element to make the best use of the data is the ability to assess the relative importance of the nodes in these networks.

The sphere of scholarly research outputs and related entities can similarly be manifested as a heterogeneous graph. In assessing and ranking scholarly entities, citation metrics have long reigned as the primary means of measuring relative impact and importance. In the 1950's Eugene Garfield established the Journal Impact Factor (JIF)

[4] to aide librarians in evaluating which journal subscriptions to preserve in the face of rising subscription costs and static library budgets. A decade ago, Jorge Hirsch created the h-index, a scholar-centric metric based on the distribution of one's publications to the number of citations received, in order to quantify "the cumulative impact and relevance of an individual's scientific research output" [5]. And indeed for individual publications themselves, many academic discovery services today rely upon cumulative citation counts as a key ranking feature.

However, the overall reliance upon citation metrics is not without disadvantages, including, but not limited to the notion that individual citations may carry uneven weights or semantic meanings [6], and that citation metrics inherently take years to emerge following the publication of the paper being cited. The drawbacks are well known, and the Weighted PageRank [7] and the emerging alt-metrics movement [8] are examples of responses to these limitations.

2. DATA
In 2015, Microsoft Research released a static snapshot of the Microsoft Academic Graph (MAG), a large heterogeneous graph comprised of publications, authors, venues (e.g., journals or conferences), organizations, and the fields of study [9] in order to enable the community to jointly work on a modern, principled and scalable way for measuring scholarly impacts and influences. The relationships and the inherent heterogeneity of objects in this graph provide new opportunities to evaluate and rank objects. For instance, the relative importance of new publications that have yet to receive any citations might be inferred from the venues, authors, or author affiliations of the publications.

The Microsoft Academic Graph used in this challenge contains approximately 140 million publication records, 650 million citation relationships, 40 million authors, 3.5 million institutions, 60,000 journals and conference "venues" and 55,000 fields of study. As a reference, a static rank score based on the PageRank algorithm [10] was provided as a part of the MAG data for papers. This data was made available as a series of text files on Azure blob storage, and could be downloaded or copied to a personal Azure account for use in the project. Challenge participants were asked to use the 2015-08-20 version of the MAG for development.

3. CHALLENGE TASK
The stated goal of the 2016 WSDM Cup challenge was to advance the start-of-the-art of data mining algorithms in quantifying the importance of nodes in a heterogeneous graph. For the purpose of

the challenge, the importance is defined in the context of search, i.e., the participants were asked to compute the query-independent static ranks for the graph. To make the task more manageable, the challenge focused on the publications in the MAG, even though the algorithms may be applicable to authors, venues and other types of nodes as well. More specifically, the participants were asked to assign a single numerical value to each paper, with a larger value representing a more important paper. Participants were also encouraged to utilize any additional public information that is not already included in the MAG (e.g., information about best paper or test-of-time awards and other honors, abstract and full text body of each paper).

4. EVALUATION

The evaluation used Pairwise Correctness as the metric to evaluate the submitted entries. The challenge organizers employed scholars of good academic standing with extensive experience serving on conference program committees or journal review boards in the field of Computer Science to perform pairwise judgments on publications. These judges were presented with pairs of articles they have previously cited or read, and asked to make a binary assessment on which article from each pair has or is likely to have more impact on their fields of study. These judgements were divided into two sets, the first used to calculate the score on the Leaderboard during the development phase. The second set was withheld as the evaluation data, and was used to calculate the final results.

The eight participant teams whose entries had the highest pairwise correctness scores were invited to the 2016 WSDM Cup workshop. and qualified for a "bonus" round of evaluation where the result from each qualifying team will replace the static rank of an existing web search engine and be deployed as online experiments. For this portion, the Challenge effectively evaluates the entries in their abilities to predict the ranking order the search engine users prefer.

5. RESULTS

Over 30 teams actively submitted results for evaluation. Of these, 19 teams performed better than PageRank.

Table 1. Final scores for top eight teams (plus *PageRank*)

Score	Team Name	Institutions
0.684	NTU_Pseudo-Tripartite	National Taiwan University National Chengchi University
0.676	Eigenfactor	University of Washington
0.671	ufmglatin	U Federal de Minas Gerais
0.664	NTU_WeightedPR	Academica Sinica National Taiwan University
0.659	bletchleypark	Open University Mendeley
0.656	teambuaa	Beihang University
0.651	sapirank	Sapienza University of Rome
0.642	NTU_Ensemble	Academica Sinica National Taiwan University National Chengchi Univeristy
0.618	*PageRank*	*n/a*

The winning team was NTU_Pseudo_Tripartite, made up of Ming-Han Feng, Kuan-Hou Chan, and Huan-Yuan Chen of National Taiwan University, and Ming-Feng Tsai of National Chengchi University, Taiwan.

The top eight teams were subsequently invited to re-submit ranking results based on an updated instance of the MAG data (version 2015-11-06) to be used in a special 'bonus round' online evaluation. The results of this online evaluation will be presented at the 2016 WSDM Cup workshop, to be held in conjunction with the WSDM Conference in February 2016. Table 1 contains the top eight team scores based on the final evaluations, as well as the PageRank performance over the same evaluation judgements.

6. REFERENCES

[1] Aral, S., and Walker, D. 2012. Identifying influential and susceptible members of social networks, *Science*, pp. 337-341, 2012. http://doi.org/10.1126/science.1215842

[2] Wang, Y., Cong, G., Song, G., and Xie, K. 2010. Community-based greedy algorithm for mining top-k influential nodes in mobile social network, in *Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining*. http:// doi.org/10.1145/1835804.1835935

[3] Sun, Y., Yu, Y. and Han, J. 2009. Ranking-based clustering of heterogeneous information networks with star network schema, in *Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining*. http://doi.org/10.1145/1557019.1557107

[4] Garfield, E. 2006. The history and meaning of the journal impact factor, *Journal of American Medical Association*, vol. 295, no. 1. http://doi.org/ 10.1001/jama.295.1.90

[5] Hirsch, J. E. 2005. An index to quantify an individual's scientific research output, *Proceedings of the National Academy of Sciences*, vol. 102, no. 46, pp. 16596-16572. http://doi.org/10.1073/pnas.0507655102

[6] Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B., Wang, K. 2015. An Overview of Microsoft Academic Service (MAS) and Applications. In *Proceedings of the 24th International Conference on World Wide Web (WWW '15)*. http://research.microsoft.com/apps/pubs/default.aspx?id=246609

[7] Yan, E. and Ding, Y. 2010. Measuring scholarly impact in heterogeneous networks," in *Proceedings of the 73rd ASIST annual meeting*, Pittsburgh PA. http://doi.org/10.1002/meet.14504701033

[8] Priem, J., Groth, P., and Taraborelli, D. 2012. The Altmetric collection, *PLOS ONE*, vol. 7, no. 11. http://doi.org/10.1371/journal.pone.0048753

[9] Microsoft Academic Graph. In http://aka.ms/academicgraph, November 2015.

[10] Brin, S., Page, L, Motwani, R., Winograd, T. 1999. The PageRank Citation Ranking: Bringing Order to the Web. Technical Report. Stanford InfoLab. http://ilpubs.stanford.edu:8090/422/