CALL FOR PAPERS (2nd call)

IEEE Transactions on Big Data (TBD): http://www.computer.org/web/tbd
Special Issue on Big Scholar Data Discovery and Collaboration
----

Important Dates:
** July. 31, 2015: Paper Submission Due
** Oct. 31, 2015: Author Notification
** Dec. 15, 2015: Revision Submission Due
** Jan. 15, 2016: Camera-Ready Due

Editor-in-Chief
Qiang Yang, Hong Kong University of Science and Technology

Guest Editors
Yu-Ru Lin, University of Pittsburgh (yurulin@pitt.edu)
Hanghang Tong, Arizona State University (hanghang.tong@asu.edu)
Jie Tang, Tsinghua University (jietang@tsinghua.edu.cn)
K. Selçuk Candan, Arizona State University (candan@asu.edu)

Overview
Academics and researchers worldwide continue to produce large numbers of scholarly documents including papers, books, technical reports, etc. and associated data such as tutorials, proposals, and course materials. The abundance of data sources enables researchers to study scholarly collaboration at a very large scale. The ever increasing diversity of disciplines and complexity of research problems, particularly multi-disciplinary research, requires collaboration. Besides the traditional venues of collaboration where scholars typically meet annually at conferences or meetings, the Internet provides a wide range of platforms for scholars to engage with other scholars. These new platforms include academic search-oriented Web engines such as Google Scholar, social media sites such as Academia.edu, ResearchGate and Mendeley, more interactive social sites such as Twitter and Facebook, and Wiki-style virtual collaboration sites. These services allow scholars to share academic resources, exchange opinions, follow each other's research, keep up with current research trends, and build their professional networks. Researchers increasingly realize that scholarly achievements should not merely be the final published articles. The datasets used in study and many other intermediary results are equally important for supporting research. Therefore, a set of rapidly developing research topics, research data management, data curation/stewardship, data sharing policy, etc. are becoming important issues for research communities. This special issues aims at bringing together researchers with diverse interdisciplinary backgrounds interested in scholarly big data.

Topics
The topics of interest include, but are not limited to:

- Data analytical tools for studying scholarly discovery and collaboration, including:
  - New approaches to measure and predict the impact of research and researchers in a particular fields of study;
  - Searching and mining large digital libraries, repositories for scholarly publications and patents and linking to other data sources such as funded proposals and patents;
  - Novel data search and mining tools for studying scholarly collaboration structure using big data, including scalable graph mining, etc.;
  - Data infrastructure that supports scalable computation, e.g., document indexing with cloud computing services;
  - Algorithms for accessing, extracting and recommending scholarly articles, experts and findings.
- Online scholar data platforms and systems consideration for scholarly discovery and collaboration, including:
  - Heterogenous data source integration, especially with open-access, novel datasets (e.g., Wikipedia, government census data, patent data, etc.);
  - Storage, indexing and query processing for research data;
  - Design considerations for effectively support scholars' engagement in using online and social platforms;
  - Social and collaborative support for scholarly discovery and collaboration;
  - Privacy and security issues and management in online scholarly collaboration.
- Digital data curation and management for scholarly discovery and collaboration, including:
  - Issues and solutions to data curation, management, and archival;
  - Existing practices for managing research data;
  - Scalability and usability of managing research data
  - Research reproducibility and data sharing policy.
- Other aspects of scholarly discovery and collaboration, including:
  - Design of next generation collaboration platforms;
  - Information professionals' role in engaging in online scholarly collaboration;
  - Cultural and community acceptance and evaluation of activities in online scholarly collaboration.

Scholar Data Challenge
One of the greatest challenges is the difficulty of collecting massive research dataset in the public domain. Hence, in addition to the above topics, this special issue will feature a "Scholar Data Challenge" associated with a dataset consisting over 2 million papers and more than 8 million citation relationships. The goal is to encourage approaches from different fields to explore the disparate facets of the same datasets in order to stimulate an interdisciplinary yet more focused research discussion. Submissions with a use of this dataset are optional. Information about the dataset can be found at: http://aminer.org/big-scholar-challenge/

Submission Information

All manuscripts must be directly submitted via the IEEE TBD submission web site: https://mc.manuscriptcentral.com/tbd-cs

Submissions must follow instructions for formatting and length as regular paper described at: http://www.computer.org/web/tbd/author

Scholar Data Challenge: Data Description

AMiner Paper Citation (APC): This dataset includes paper information, paper citation, author information, and author collaboration. In particular, there are three files in this dataset:

- **AMiner-Paper.rar:** It records meta information (title, author list, author affiliation, published year, published venue, abstract, and the list of references) of 2,092,356 papers.
- **AMiner-Author.zip:** It records meta information (name, affiliation, #publication-paper, #citation, H-index, expertise-by-keywords) of 1,712,433 authors.
- **AMiner-Coauthor.zip:** It records the collaboration relationships between authors recorded in AMiner-Author.zip. The collaboration relationships can be also derived from the authorship information recorded in AMiner-Paper.rar.

The dataset can be used for multiple different tasks such as:

- **Expert finding:** Given a query (consisting of one or multiple keywords), to find experts for this query. For evaluation, one can find a collection of groud truth here http://arnetminer.org/lab-datasets/expertfinding/
- **Profile extraction:** Given an author in this dataset, to automatically populate his/her profile meta information (name, affiliation, #publication-paper, #citation, H-index, expertise-by-keywords). Information recorded in the AMiner-Author.zip file can be used as ground truth to evaluate different methods.
- **Prediction:** Several subtasks can be explored for this purpose: (1) Given the collaboration network at time t, to predict who will create new collaboration relationships at time (t+1); (2) Given an author and his publication list at time t, to predict H-index of the author at time (t+1).
- **Citation recommendation:** Given a paper (title, author list, abstract, publication venue, and publication year), to recommend top-k papers that should be cited by this paper.

The dataset can be downloaded here http://arnetminer.org/AMinerNetwork