

View

Edit

Home

Research

People

Demos

Software/Datasets

Publications

Dataset: FastPPV-DblpLjGraphs-Aug2013

1. Overview

This dataset consists of two real-world graphs. One is an undirected bibliographic network based on DBLP, and the other is a directed social network based on LiveJournal.

2. Bibliographic Reference

If you are using this dataset, please cite our work:

```
@article{DBLP/pvldb/ZhuFCY13,
  author    = {Fanwei Zhu and Yuan Fang and Kevin C.-C. Chang and Jing Ying},
  title     = {Incremental and Accuracy-Aware Personalized PageRank through Scheduled Approximation.},
  booktitle = {PVLDB},
  year      = {2013},
  volume    = {6},
  number    = {6},
  pages     = {481-492}
}
```

3. Source Publications

F. Zhu, Y. Fang, K. C.-C. Chang and J. Ying. Incremental and Accuracy-Aware Personalized PageRank through Scheduled Approximation. In PVLDB 6(6), pp. 481–492. For presentation at VLDB 2013.

4. Creation

We created/adapted two real-world graphs for our work.

1) Bibliographic network (Dblp)

- 2 million nodes and 8.8 million edges
- Based on [DBLP](#) collected in 2009, courtesy of [Yizhou Sun](#).
- Models bi-directional relationship between authors, papers and venues.

2) LiveJournal friendship network (Lj)

- 1.2 million nodes and 4.8 million edges
- Sampled from the [LiveJournal](#) dataset from [Stanford Network Analysis Project](#) (SNAP)
- Models directed friendship declarations between users.

5. Data Files

Data File	Download	Subject	Description
FastPPV-DblpLjGraphs-Aug2013-Dblp.zip	link	Bibliographic network (Dblp)	Network of authors, papers and venues
FastPPV-DblpLjGraphs-Aug2013-Lj.zip	link	LiveJournal friendship network (Lj)	Network of LiveJournal users

1) Description of FastPPV-DbpLjGraphs-Aug2013-Dbp.zip

This file contains the graph for the bibliographic network (Dbp) as well as our test queries, which is described by 3 text files inside: dblp-nodes.txt, dblp-edges.txt, dblp-queries.txt.

a) dblp-nodes.txt

This file stores the 3 types of nodes of DBLP graph with each line representing a single node by its ID:

- Paper nodes: ID 1 ~ 1243232
- Author nodes: ID 10000001 ~ 10749731
- Venue nodes: ID 20000001 ~ 20005011

b) bibnet-edges.txt

This file stores the undirected edges of the graph. These edges represent two kinds of relationships:

- Author - Paper (Author writes Paper)
- Paper - Venue (Paper published in Venue)

Note: even though we only have undirected edges in this dataset, each undirected edge is represented by two directed edges of opposite directions. This is to be consistent with a single data format with the directed graph LiveJournal (see below).

Each line is tab-delimited, and encodes one **directed** edge (i.e., one direction of an undirected edge), according to the following format (without the spaces around \tab).

```
[From] \tab [To]
```

where:

[From] ID of the starting node of the directed edge

[To] ID of the ending node of the directed edge

Some examples (without the spaces around \tab)

```
659387 \tab 10020732
10020732 \tab 659387
659387 \tab 10060110
10060110 \tab 659387
```

Note: You may assign different weights to edges depending on your application. In our work we simply use a uniform weight of 1 across all edges.

c) dblp-quereis.txt

This file stores our test queries, with each line representing a single query node by its ID.

2) Description of FastPPV-DbpLjGraphs-Aug2013-Lj.zip

This file contains the graph for the LiveJournal friendship network (Lj) as well as our test queries, which is described by 3 text files inside: lj-nodes.txt, lj-edges.txt, lj-queries.txt.

a) lj-nodes.txt

This file stores the user nodes of LiveJournal graph. Each line is a single node represented by its ID.

b) lj-edges.txt

This files store the directed edges of LiveJournal graph. Each edge represents a user1 -> user2 relationship (user1 declares user2 as a friend, which is not necessarily reciprocal).

Each line is tab-delimited, and represents one directed edge according to the following format (without the spaces around \tab).

```
[From] \tab [To]
```

where:

```
[From] ID of the starting node of the directed edge  
[To] ID of the ending node of the directed edge
```

Some examples (without the spaces around \tab)

```
593640 \tab 31270  
593646 \tab 243330  
593646 \tab 438064  
593640 \tab 31270  
593646 \tab 243330  
593646 \tab 438064
```

Note: You may assign different weights to edges depending on your application. In our work we simply use a uniform weight of 1 across all edges.

c) lj-queries.txt

This file stores our test queries, with each line representing a single query node by its ID.

6. Owners of the Processed Dataset

[Kevin Chen-Chuan Chang](#) 

Department of Computer Science

University at Illinois at Urbana-Champaign