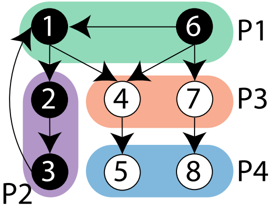
Regularities in bisimulation reductions of big graphs

Yongming Luo, George Fletcher, Jan Hidders,
Paul De Bra and Yuqing Wu

TU Eindhoven, TU Delft, IU Bloomington
Fourth TUC Meeting, April 2014, Amsterdam

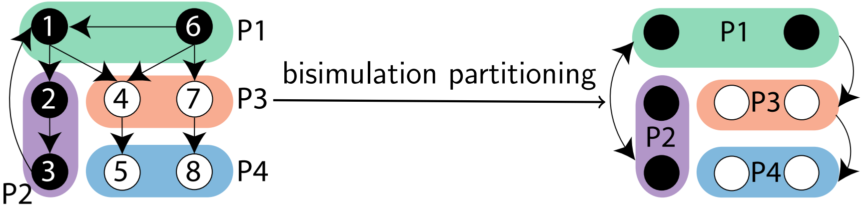
Bisimulation reduction of graphs

- Bisimulation partitioning is an important concept in many fields (computer science, modal logic, etc.), in DB research as well (structural index, graph reduction)
- It can be seen as a way of clustering nodes



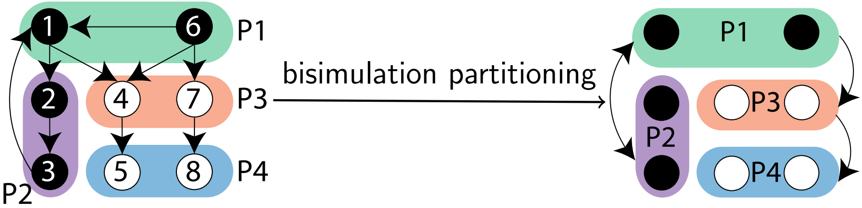
Bisimulation reduction of graphs

- Bisimulation partitioning is an important concept in many fields (computer science, modal logic, etc.), in DB research as well (structural index, graph reduction)
- It can be seen as a way of clustering nodes



Bisimulation reduction of graphs

- Bisimulation partitioning is an important concept in many fields (computer science, modal logic, etc.), in DB research as well (structural index, graph reduction)
- It can be seen as a way of clustering nodes



- Reduce graph size while preserving structural properties (e.g., reachability)
- Result can be seen as a (PB) graph
- What properties does the PB graph have?

Questions

Regularities, such as power-law distribution exists in real graphs.

Questions

Regularities, such as power-law distribution exists in real graphs.

- Do results under bisimulation reduction (e.g., PB graph) also have such properties?

Questions

Regularities, such as power-law distribution exists in real graphs.

- Do results under bisimulation reduction (e.g., PB graph) also have such properties?
- How would that knowledge help us?

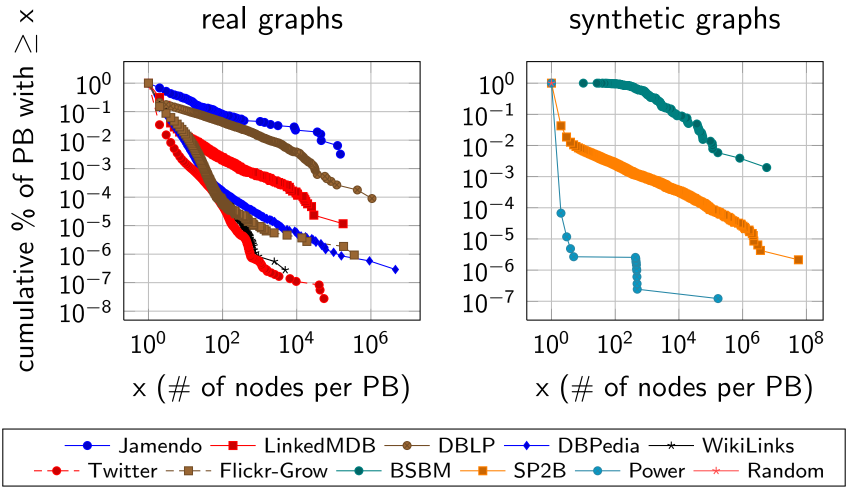
Experimental setup for investigation

- Big graphs, from 1 Million to 1.4 Billion edges (Twitter, DBPedia, etc.)
- State-of-the-art external-memory algorithm for computing bisimulation reductions
- We use cumulative distribution function (CDF) to present distributions

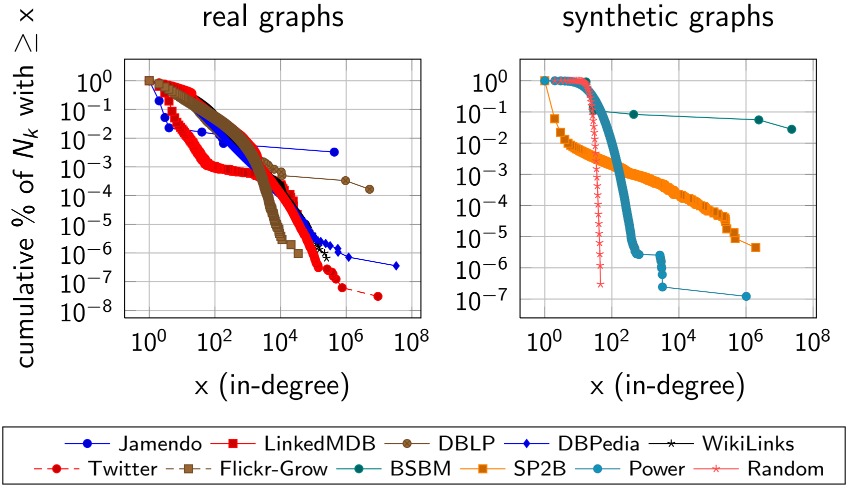
Regularities - bisimulation result

Power-law also exists in many attributes for bisimulation partition results for *real graphs*. But this is not the case for *synthetic graphs*.

Regularities - partition block size distribution



Regularities - PB graph in-degree distribution



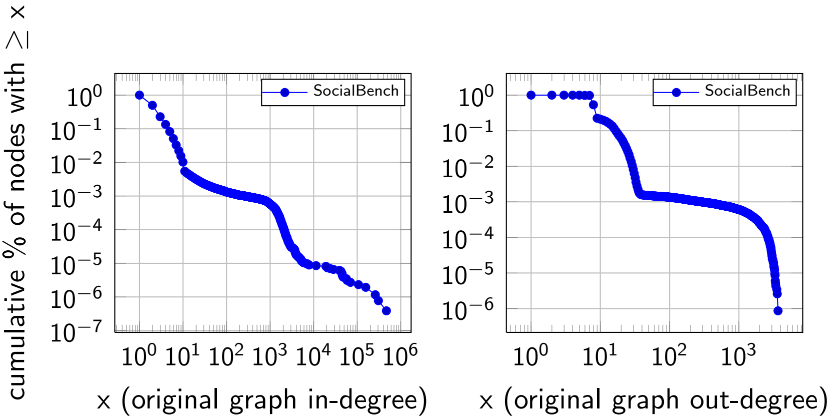
A close look at Social Intelligence Benchmark (old)

- What structure is exhibited by graphs generated by the Social Intelligence Benchmark?

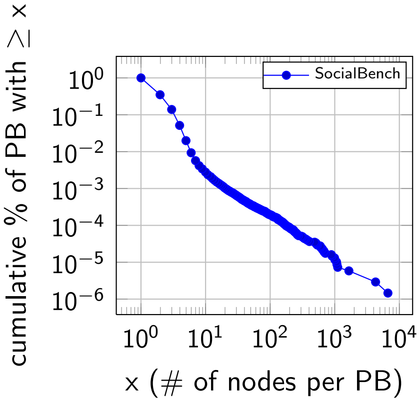
A close look at Social Intelligence Benchmark (old)

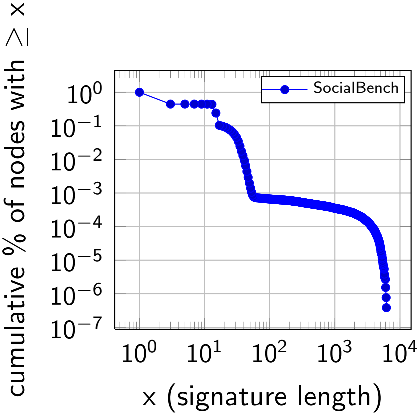
- What structure is exhibited by graphs generated by the Social Intelligence Benchmark?
- Use s3g2130313.tar, downloaded from `sourceforge.net/projects/sibenchmark/` (thanks to Minh-Duc Pham)
- Number of nodes: 2.6M, Number of edges: 12.6M
- Configuration: numtotalUser: 10000, 2010-1-1 to 2012-1-1

In-degree and out-degree of original graph

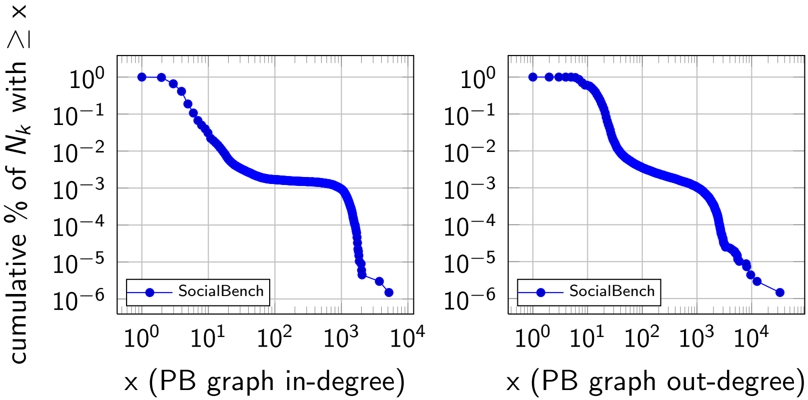


Partition block size & signature length distribution





In-degree and out-degree of PB graph



Insights

- Power-law distributions in bisimulation results emphasize the fact that data skews are expected in applications (indexes, data partitioned among machines, . . .)

Insights

- Power-law distributions in bisimulation results emphasize the fact that data skews are expected in applications (indexes, data partitioned among machines, . . .)
- Some more work remains to be done for synthetic graph generators towards exhibiting the reduction properties of real graphs.

Insights

- Power-law distributions in bisimulation results emphasize the fact that data skews are expected in applications (indexes, data partitioned among machines, ...)
- Some more work remains to be done for synthetic graph generators towards exhibiting the reduction properties of real graphs.
- Bisimulation result/graph grows when original graph grows, which calls for scalable/adaptive algorithms (e.g., choose different k for different parts of the graph, different node/edge labeling)

Thank you! Q&A

For more information, just google **seeqr project** or

visit: `bit.ly/seeqr`

Definition of *k-bisimilar*

Definition

Let k be a non-negative integer and $G = \langle N, E, \lambda_N, \lambda_E \rangle$ be a graph. Nodes $u, v \in N$ are called *k-bisimilar* (denoted as $u \approx^k v$), iff the following holds:

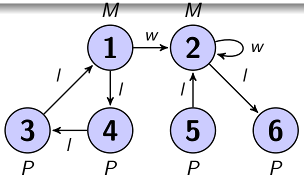
1. $\lambda_N(u) = \lambda_N(v)$,
2. if $k > 0$, then for any edge $(u, u') \in E$, there exists an edge $(v, v') \in E$, such that $u' \approx^{k-1} v'$ and $\lambda_E(u, u') = \lambda_E(v, v')$, and
3. if $k > 0$, then for any edge $(v, v') \in E$, there exists an edge $(u, u') \in E$, such that $v' \approx^{k-1} u'$ and $\lambda_E(v, v') = \lambda_E(u, u')$.

Definition of *k-bisimilar*

Definition

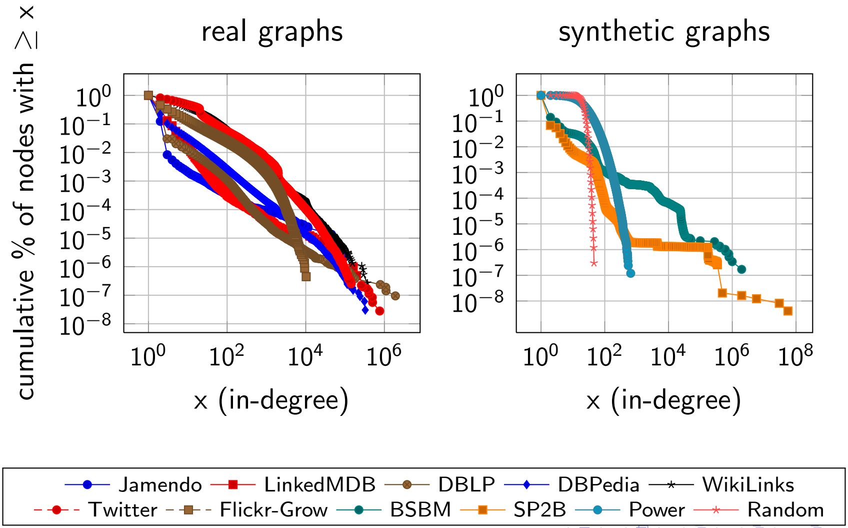
Let k be a non-negative integer and $G = \langle N, E, \lambda_N, \lambda_E \rangle$ be a graph. Nodes $u, v \in N$ are called *k-bisimilar* (denoted as $u \approx^k v$), iff the following holds:

1. $\lambda_N(u) = \lambda_N(v)$,
2. if $k > 0$, then for any edge $(u, u') \in E$, there exists an edge $(v, v') \in E$, such that $u' \approx^{k-1} v'$ and $\lambda_E(u, u') = \lambda_E(v, v')$, and
3. if $k > 0$, then for any edge $(v, v') \in E$, there exists an edge $(u, u') \in E$, such that $v' \approx^{k-1} u'$ and $\lambda_E(v, v') = \lambda_E(u, u')$.

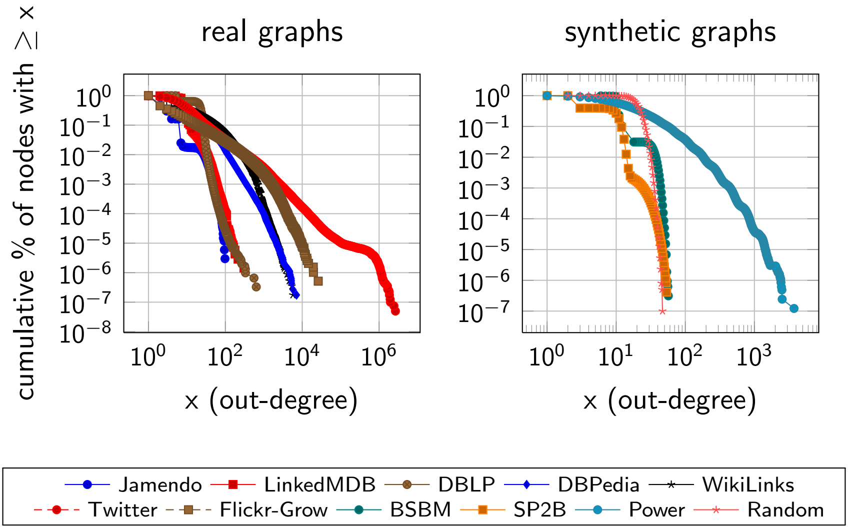


In this example graph, nodes 1 and 2 are 0- and 1- bisimilar but not 2-bisimilar.

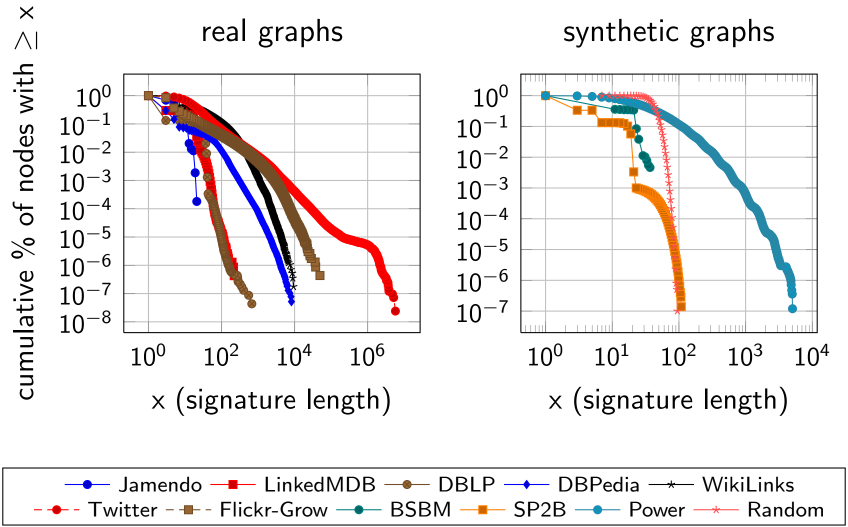
In-degree distribution of original graphs



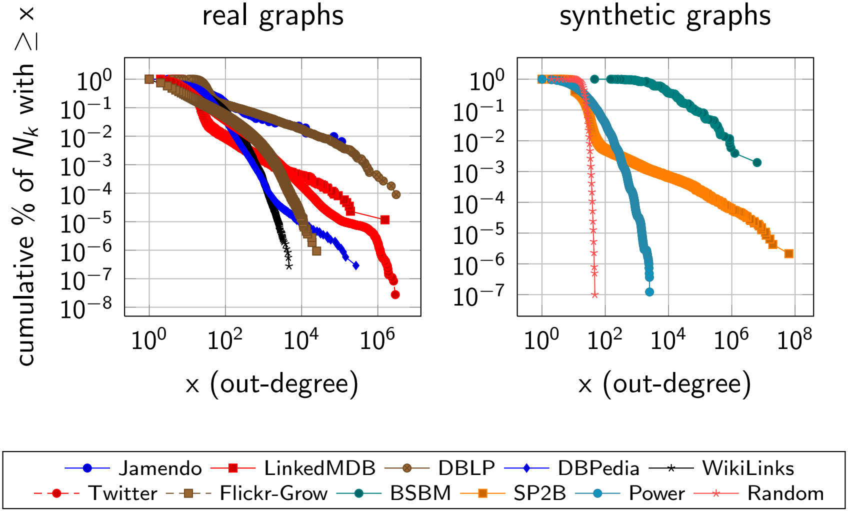
Out-degree distribution of original graphs



Signature length



Out-degree of PB graph



Dynamics - a real growing social graph

- Dynamic social graph, from 17 Million to 33 Million edges (Flickr-grow)

Dynamics - a real growing social graph

- Dynamic social graph, from 17 Million to 33 Million edges (Flickr-grow)
- Does the bisimulation result grow when the original graph grows?

Dynamics - a real growing social graph

- Dynamic social graph, from 17 Million to 33 Million edges (Flickr-grow)
- Does the bisimulation result grow when the original graph grows?
    - Yes.

Dynamics - a real growing social graph

- Dynamic social graph, from 17 Million to 33 Million edges (Flickr-grow)
- Does the bisimulation result grow when the original graph grows?
  - Yes.
- How fast does it grow?

Dynamics - a real growing social graph

- Dynamic social graph, from 17 Million to 33 Million edges (Flickr-grow)
- Does the bisimulation result grow when the original graph grows?
  - Yes.
- How fast does it grow?
  - Linearly with respect to the original graph.