Regularities and dynamics in bisimulation reductions of big graphs

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

GRADES 2013 • SIGMOD/PODS 2013

The research of YL, GF, JH and PD is supported by the Netherlands Organisation for Scientific Research.
The research of YW is supported by Research Foundation Flanders during her sabbatical visit to Hasselt University, Belgium.

TU/e Technische Universiteit **Eindhoven** University of Technology

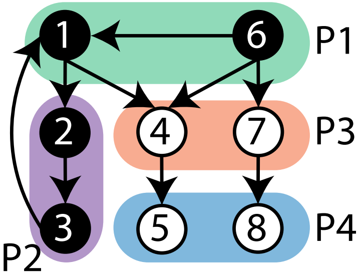
Outline

Motivation

Experimental setup

Results

Insights



An example of bisimulation reduction

Bisimulation reduction

- 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

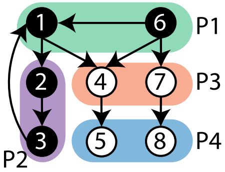


Figure: Bisimulation partition example, partition block graph (reduction graph) $\{P2 \leftrightarrow P1 \rightarrow P3 \rightarrow P4\}$

Bisimulation reduction

- 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

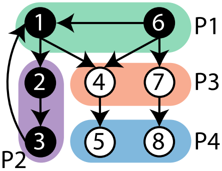


Figure: Bisimulation partition example, partition block graph (reduction graph) $\{P2 \leftrightarrow P1 \rightarrow P3 \rightarrow P4\}$

- Reduce graph size while preserving structural properties (e.g., reachability)
- Result can be seen as a graph
- Many algorithms, no work on analyzing the results

Questions

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

Technische Universiteit Eindhoven University of Technology

Questions

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

- Do graphs under bisimulation reduction also have such properties?

Questions

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

- Do graphs under bisimulation reduction also have such properties?
- How would that knowledge help us?

Experimental setup for investigation

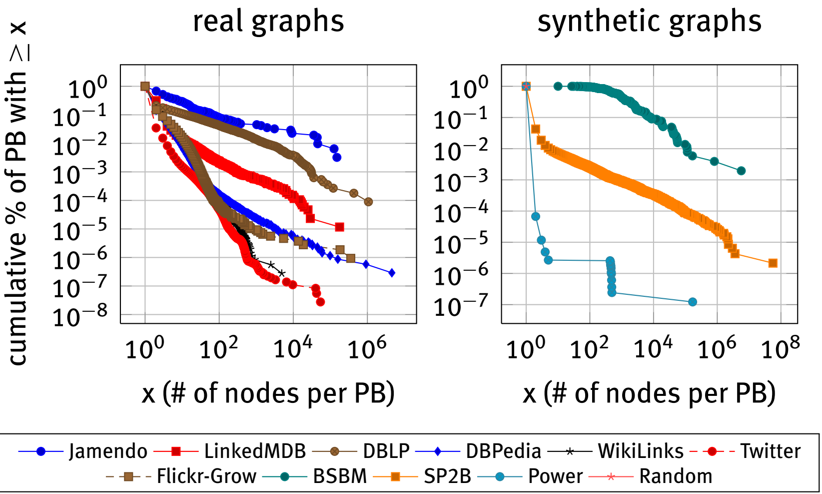
- Big graphs, from 1 Million to 1.4 Billion edges (Twitter, DBPedia, etc.)
- One dynamic social graph, from 17 Million to 33 Million edges (Flickr-grow)
- State-of-the-art I/O efficient algorithm for computing bisimulation reductions (k-bisim, $k = 10$)
- We use cumulative distribution function (CDF) 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*.

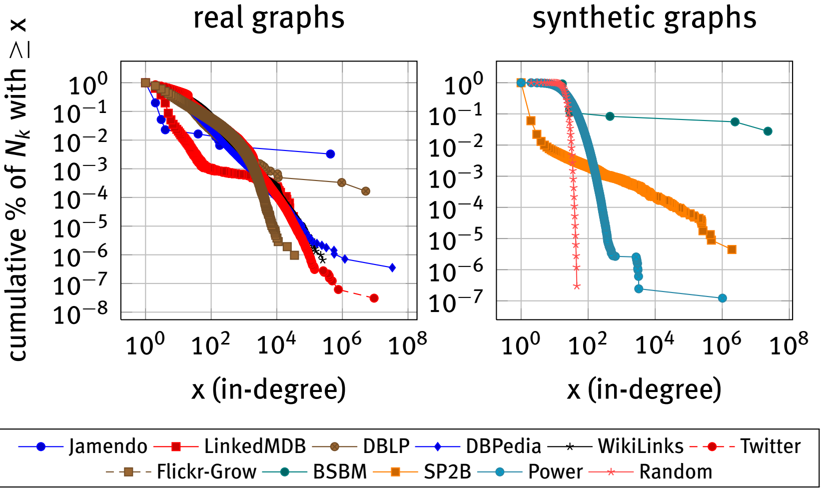
Regularities - bisimulation result

Partition block size distribution



Regularities - bisimulation result

Bisimulation graph in/out-degree distribution



Dynamics - a real growing social graph

- Does the bisimulation result grow when the original graph grows?

Dynamics - a real growing social graph

- Does the bisimulation result grow when the original graph grows?
  - Yes.

Dynamics - a real growing social graph

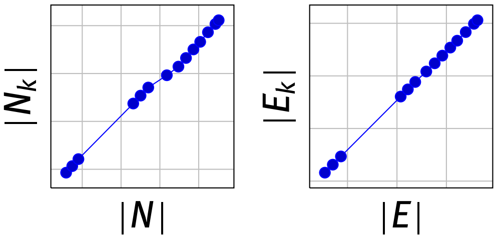
- Does the bisimulation result grow when the original graph grows?
  - Yes.
- How fast does it grow?

Technische Universiteit Eindhoven University of Technology

Dynamics - a real growing social graph

- Does the bisimulation result grow when the original graph grows?
  - Yes.

- How fast does it grow?
  - Linearly with respect to the original graph.



Insights

- Power-law distributions in bisimulation results $\Rightarrow$ skew expected in applications (indexes, data partitioned among machines, . . .)

Insights

- Power-law distributions in bisimulation results $\Rightarrow$ skew expected in applications (indexes, data partitioned among machines, ...)
- Behaviors of graph generators $\Rightarrow$ some more work needs to be done for graph generators

Insights

- Power-law distributions in bisimulation results $\Rightarrow$ skew expected in applications (indexes, data partitioned among machines, ...)
- Behaviors of graph generators $\Rightarrow$ some more work needs to be done for graph generators
- Bisimulation result/graph grows $\Rightarrow$ lower k or other adaptations (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 $\lambda_E(v, v') = \lambda_E(u, u')$.

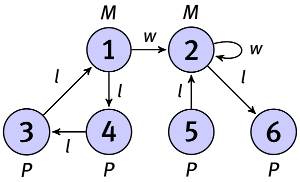
Technische Universiteit Eindhoven University of Technology

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:

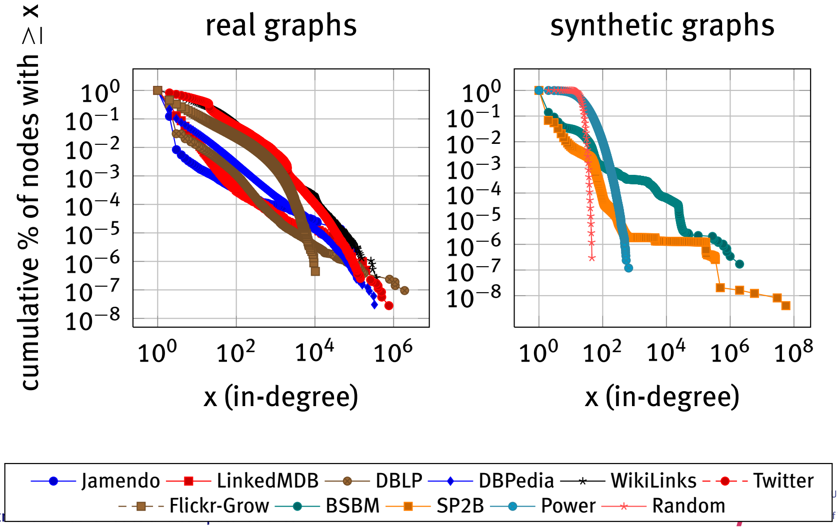
1. $\lambda_N(u) = \lambda_N(v)$,
2. if $k > 0$, then for any edge $(u, u') \in E$, there exists an edge $(v, v') \in E$, such that $u' \approx^{k-1} v'$ and $\lambda_E(u, u') = \lambda_E(v, v')$, and
3. if $k > 0$, then for any edge $(v, v') \in E$, there exists an edge $(u, u') \in E$, such that $v'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



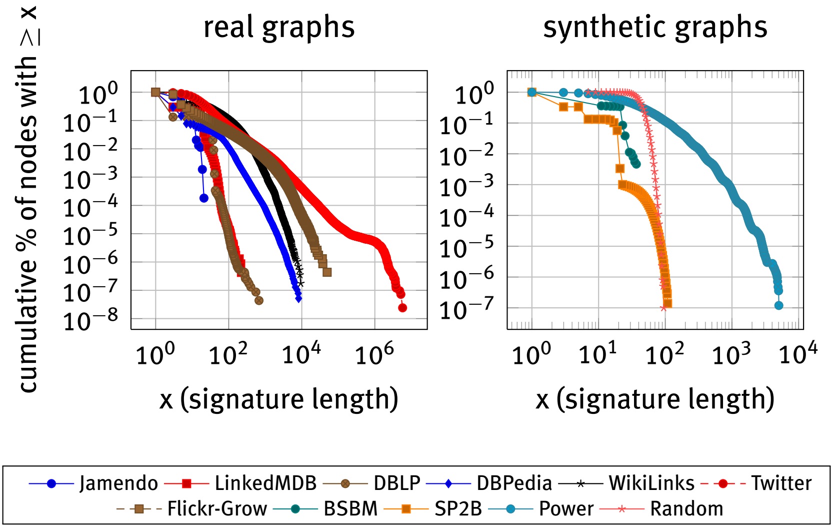
In this example graph, nodes 1 and 2 are 0- and 1- bisimilar but not 2-bisimilar.

Regularities - original graphs

Power-law exists in in/out-degree distribution for most of the examined graphs.

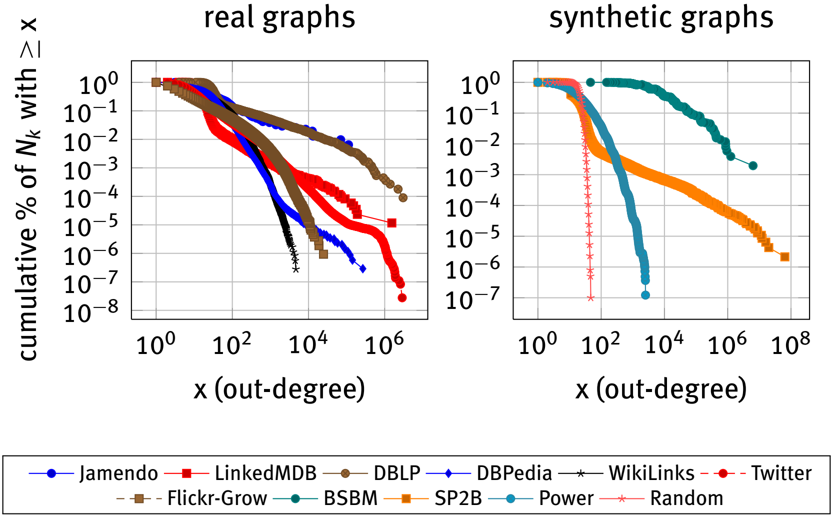


Signature length



TU/e Technische Universiteit Eindhoven University of Technology

Out-degree


real graphs
synthetic graphs
cumulative % of N_k with $\geq$ x
10^0
10^{-1}
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
10^{-7}
10^{-8}
10^0
10^2
10^4
10^6
10^8
x (out-degree)
x (out-degree)
Jamendo
LinkedMDB
DBLP
DBPedia
WikiLinks
Twitter
Flickr-Grow
BSBM
SP2B
Power
Random