Evaluating Reachability Queries over Path Collections

P. Bouros$^1$, S. Skiadopoulos$^2$, T. Dalamagas$^3$, D. Sacharidis$^3$, T. Sellis$^{1,3}$

$^1$National Technical University of Athens
$^2$University of Peleponnese
$^3$Institute for Management of Information Systems – R.C. Athena
Motivating example

Q: How can I go from Academy to Zappeion?
A: Academy → University → Parliament → Zappeion

SSDBM'09
Problem definition

- Large path collections
- Frequent updates
  - New paths
- Evaluating reachability queries
  1. Is there a path from node $F$ to $C$?
  2. Find a path from node $F$ to $C$

Path collection

$p_1 \quad (A, B, C, D, J)$

$p_2 \quad (A, F, D, N, B, T)$

$p_3 \quad (N, L, M)$

$p_4 \quad (D, N, B, F, K)$

$p_5 \quad (A, F, K)$
Solving the problem

• Using a graph
  – Merge paths to build underlying graph
  – Apply graph search algorithm for queries

• Our approach
  – Algorithm pfs
    • Visit nodes in paths in depth-first way
    • Push to dfs stack set of nodes at once
  – Index path collections, P-Index, pfsP
  – Index path connections, H-graph & H-Index, pfsH
Thank you!