Breaking Cycles in Noisy Hierarchies

Jiankai Sun [1]
Deepak Ajwani [2] Patrick Nicholson [2] Alessandra Sala [2]
Srinivasan Parthasarathy[1]

[1]The Ohio State University

[2]Bell Labs, Nokia, Ireland

WebSci'17, June 26 -28, 2017

NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

Outline

1. Motivation
2. Related Work
3. Our Framework: Breaking Cycles via Graph Hierarchies
4. Experiments
5. Conclusion

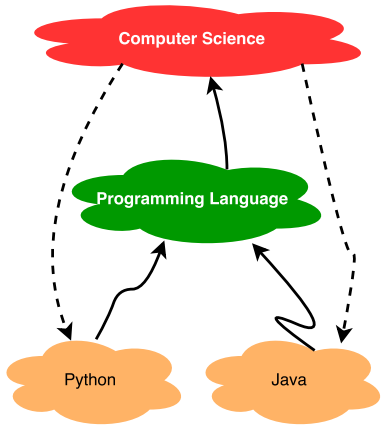
NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

[0] Pic from Web

Motivation

- Taxonomy graphs that capture ”has a” or ”is a” relationships should be acyclic
- Ontological knowledge bases such as Wikipedia categories, created in crowd-sourced way, cause errors (cycles)
- Breaking Cycles to get a Directed Acyclic Graph (DAG) can benefit other applications such as job/dataflow scheduling



NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

Related Work

- Simple Heuristic Based on BFS or DFS
- Minimum Feedback Arc Set
- Domain-specific Algorithms

NOKIA Bell Labs

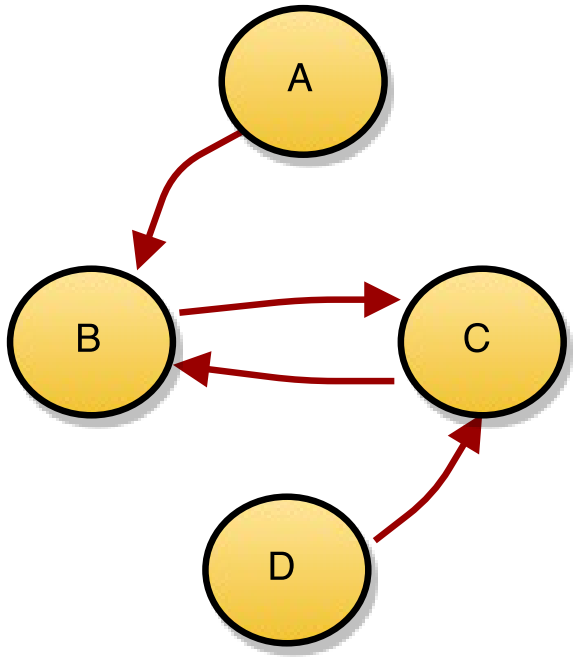
THE OHIO STATE UNIVERSITY

[0] Pic from Web

DFS & BFS: simple, domain independence

Depth-first Search

- detect and remove back edges randomly (un-deterministic)



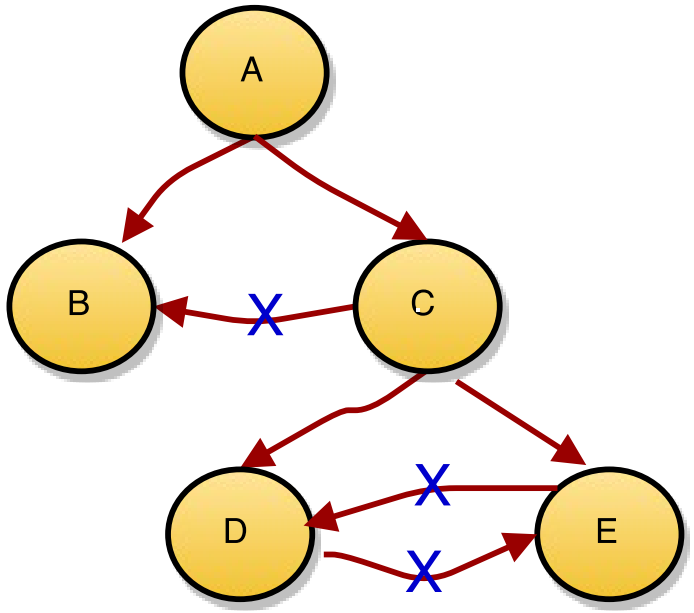
NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

DFS & BFS: simple, domain independence

Breadth-first Search

- can remove non-cycle edges

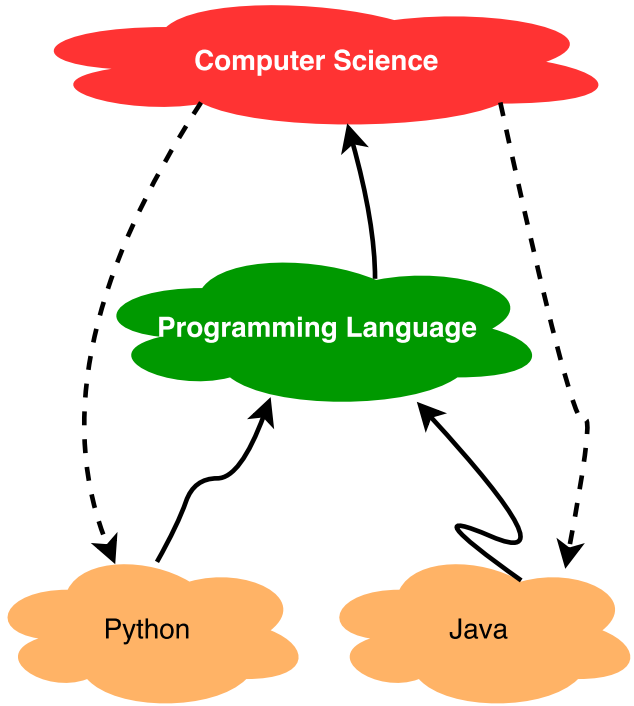


NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

Minimum Feedback Arc Set

- Remove the least number of edges to break cycles
- NP-Hard Problem
- Cannot guarantee it preserves the logical hierarchy structure while minimizing the edges to remove



NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

Graph Hierarchy Based Framework

Goal: break cycles from a directed graph, while preserving the underlying hierarchy of the relationships as much as possible

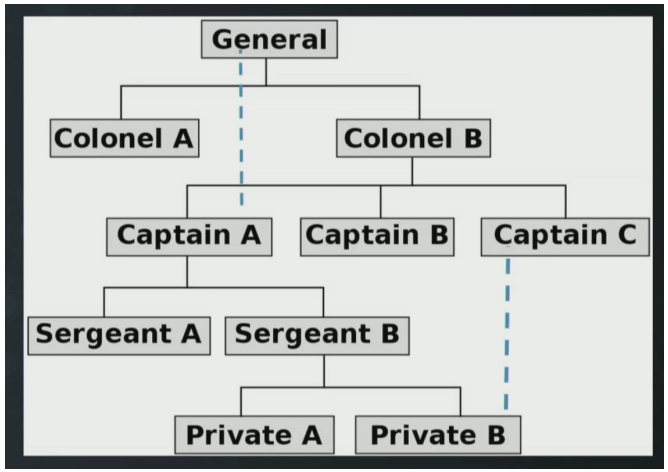
1. Inferring graph hierarchy
   - TrueSkill
   - SocialAgony
2. Proposing strategies to select violation edges as candidates for removal based on graph hierarchy
   - Forward
   - Backward
   - Greedy

NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

Finding a ranking function to infer graph hierarchy

- f assigns a ranking score to each node in the graph
- A higher ranking score indicates the corresponding node is higher up (or more general) in the hierarchy
- Edges violate the hierarchy (edges from a higher/general group to a lower/specific group) are potential edges for removal



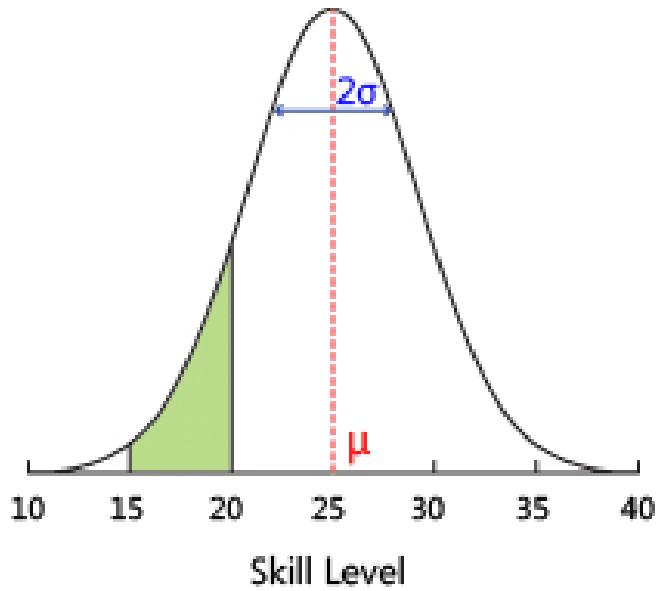
NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

[0]Figure: `http://bit.ly/2sZTH3c`

Inferring Graph Hierarchy by TrueSkill

- TrueSkill ranking system is a skill based ranking system to rank Xbox players, developed by Microsoft Research
- Each player has two numbers
  - μ: average skill of the player
  - σ : degree of uncertainty in the player's skill



NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

View it as a competition graph

- a directed graph $G = (V, E) \Rightarrow$ a multi-player tournament with $|V|$ players and $|E|$ competitions
- an edge $(u, v) \in E \Rightarrow u$ loses the game between u and v

Updates of skill levels given an edge (u, v)

- If player v has a higher skill level than u, then the outcome of edge (u, v) is expected $\Rightarrow$ small updates in skill level μ and σ.
- If player u has a higher skill level than v, then the outcome of edge (u, v) is unexpected $\Rightarrow$ large updates in skill level μ and σ.

NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

Inferring Graph Hierarchy by TrueSkill

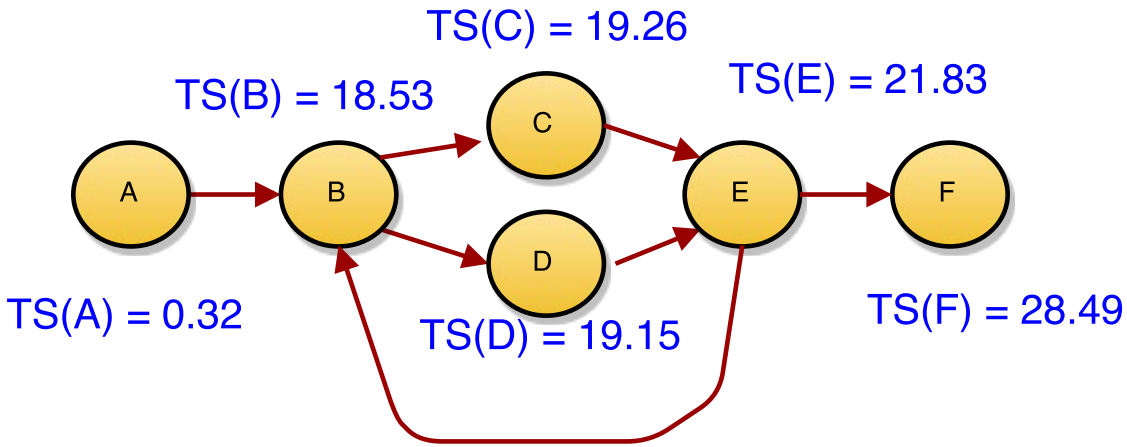


Figure: TrueSkill Compuation Demo

- As far as we know, we are the first researchers to consider **graph hierarchy inference as a competition problem**
- A node v's ranking score in the graph hierarchy: $f_{ts}(v) = \mu_v - 3\sigma_v$

Inferring Graph Hierarchy by Social Agony

- Social agony proposed by Gupte et al. assumes the existence of a link indicates a rank recommendation
  - A link $u \Rightarrow v$ indicates a recommendation of v from u
  - If there is no reverse link from v to u, it could indicate that v is higher up in the hierarchy than u
- In social networks such as Twitter, agony can be caused when people follow other people who are lower in the hierarchy

NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

[0]Figure: `http://bit.ly/2r7afHV`

Computation of Graph Agony

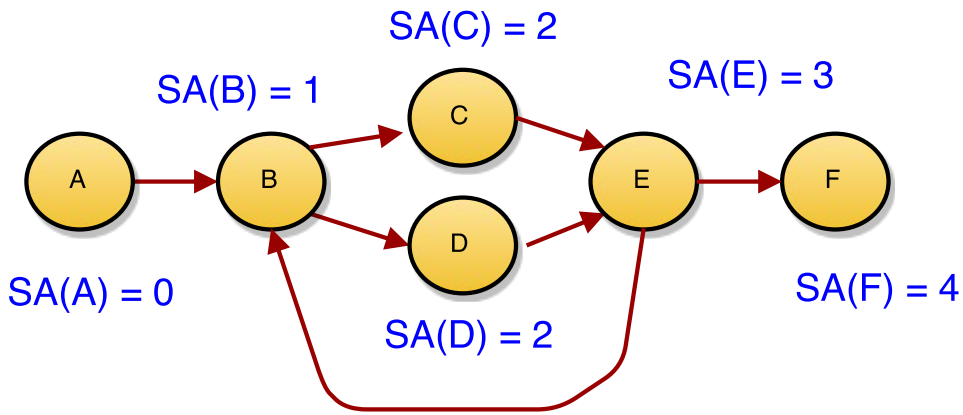
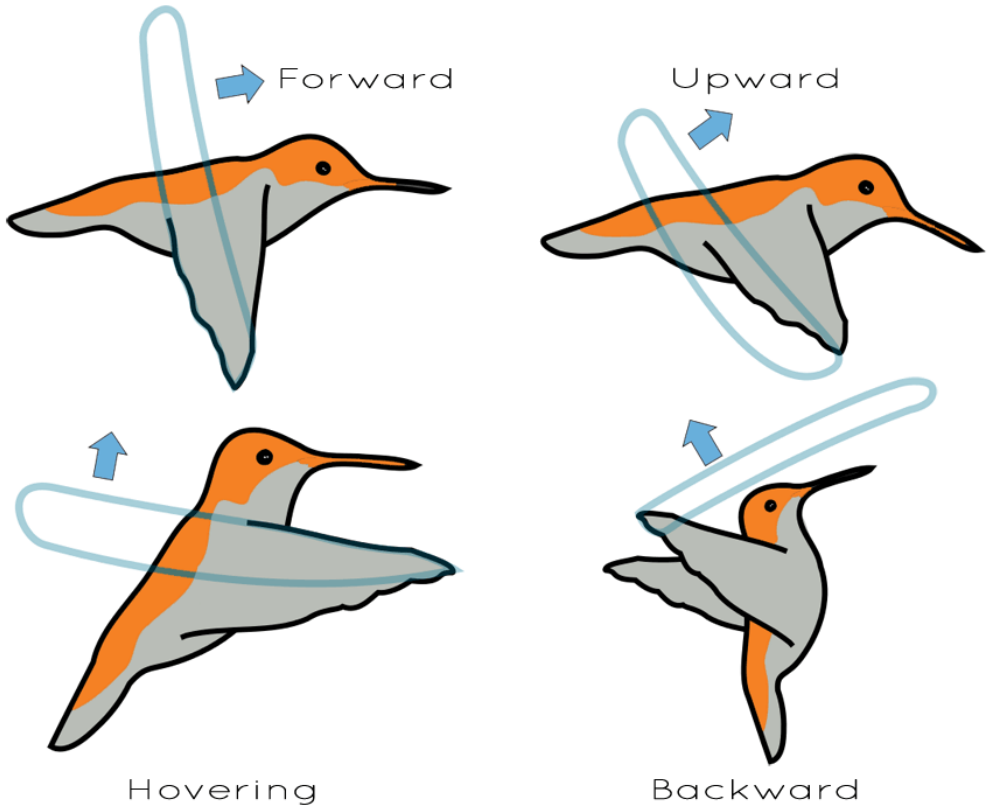


Figure: SocialAgony Compuation Demo

- Gupte et al., Tatti et al. proposed efficient algorithms to find a ranking r to minimize the agony of the graph
- A node v's ranking score in the graph hierarchy inferred by social agony: $f_{agony}(v) = r(v)$

We provide 3 solutions to select violation edges

- Forward
- Backward
- Greedy



NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

[0]Figure: http://bit.ly/2sCJNrf

Forward to select edges to remove and break cycles

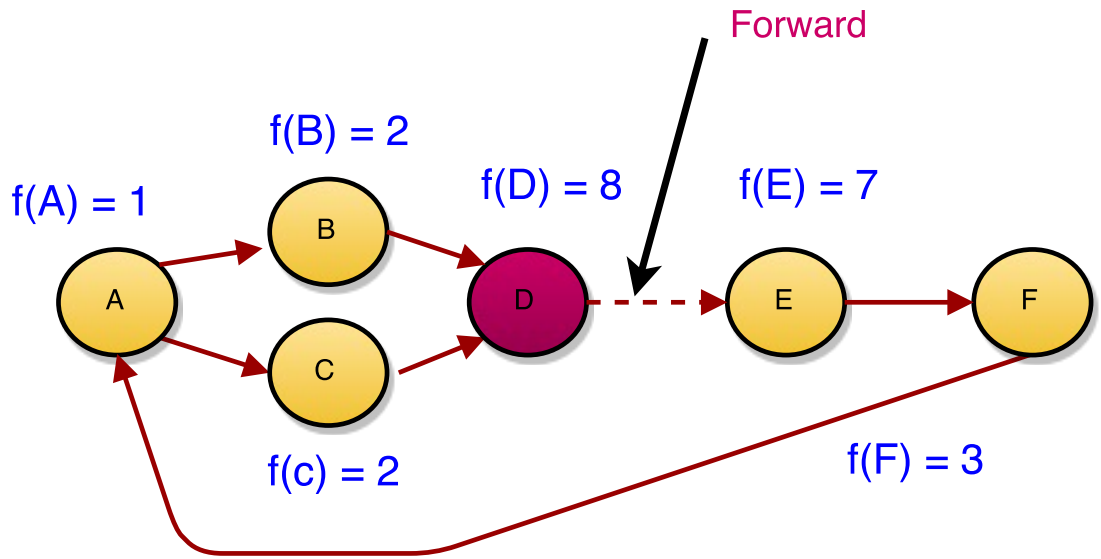


Figure: Strategy Forward to select violation edges

- *Forward*: Select the node which has the *highest* ranking score in the SCC and then remove its all *out* edges.

Backward to select edges to remove and break cycles

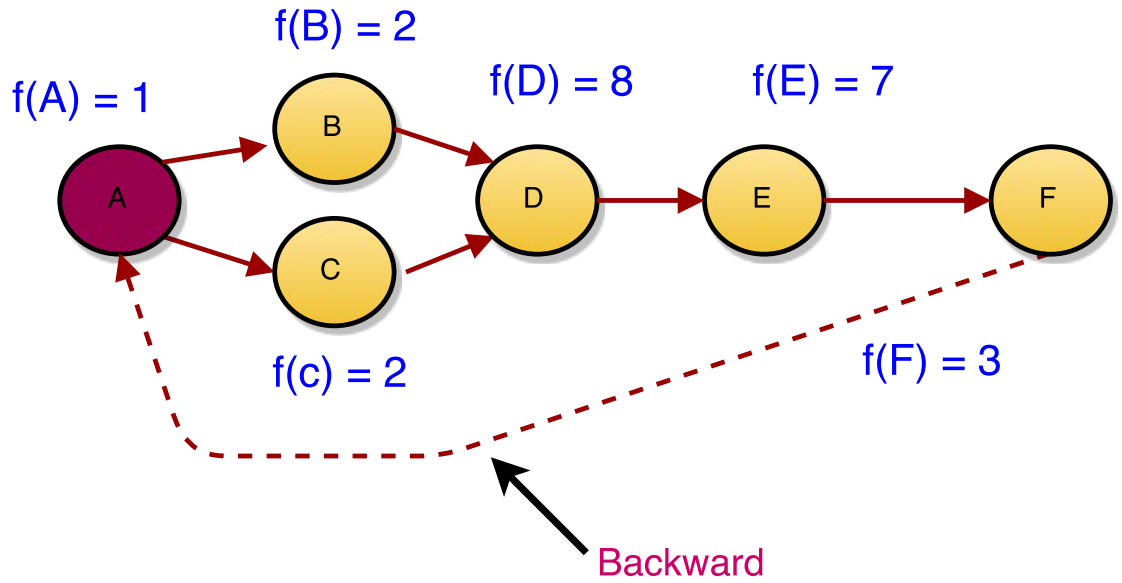


Figure: Strategy Forward to select violation edges

- *Backward*: Select the node which has the *lowest* ranking score in the SCC and then remove its all *in* edges.

Greedy to select edges to remove and break cycles

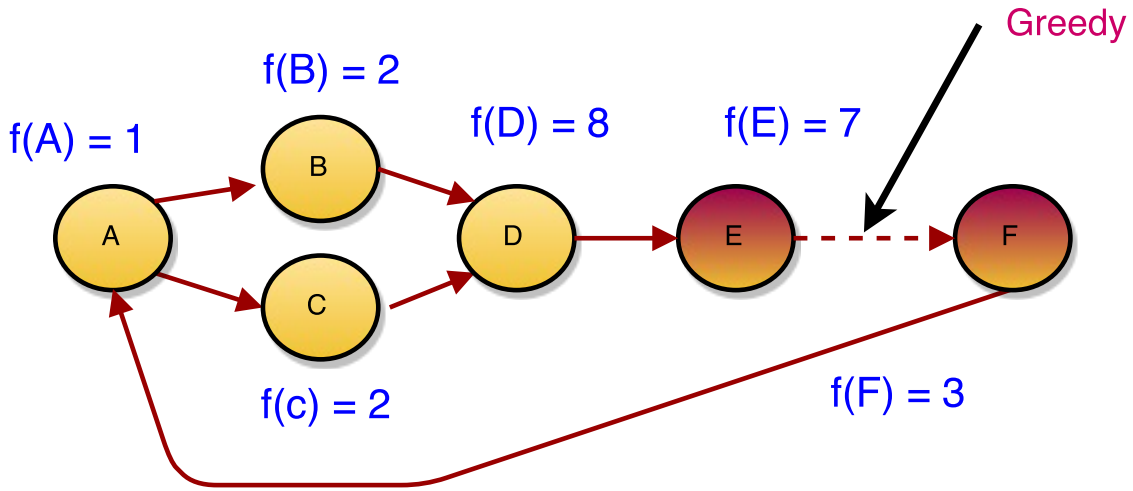


Figure: Strategy Forward to select violation edges

- *Greedy*: Select the edge which violates the hierarchy the *most* to remove.

Combine Them Toghether

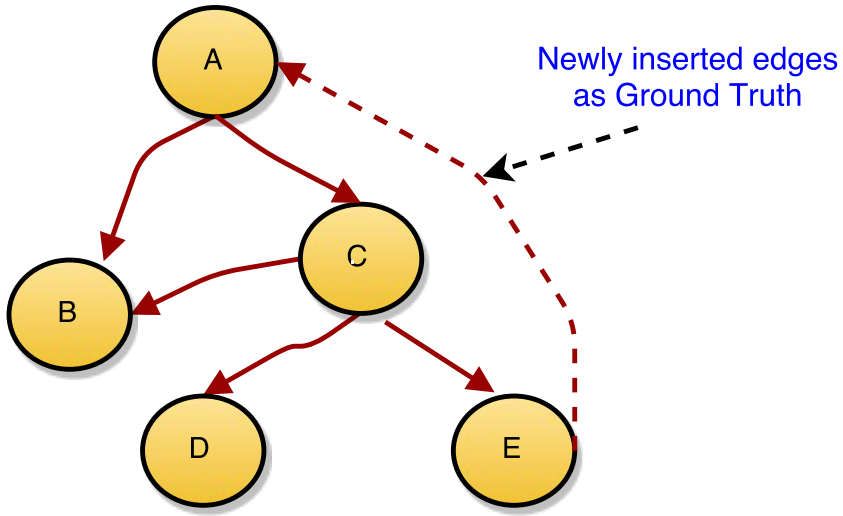
- 2 ways to infer graph hierarchy: TrueSkill and SocialAgony
- 3 solutions to select edges: *Forward, Backward, Greedy*
- $\Rightarrow 6$ strategies to break cycles
  - TS_G, TS_B, TS_F
  - SA_G, SA_B, SA_F
- Assembled together: *H_Voting* selects the edge with the highest voting score for removal
  - voting score for an edge e: $\sum_m (I_m(e))$
  - $m \in \{TS\_G, TS\_F, TS\_B, SA\_G, SA\_F, SA\_B\}$
  - if edge e is removed by method m, $I_m(e) = 1$, otherwise $I_m(e) = 0$
  - remove the edge with the highest voting score first

NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

Experimental Setup

- Few large real taxonomy graphs have ground truth (edges are labeled as errors)
- Introduce cycles (randomly) to real and synthetic DAG
  - insert edges that violate the partial order

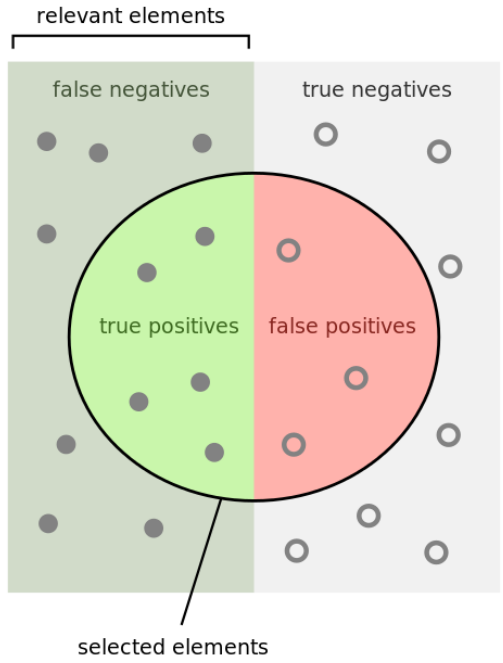


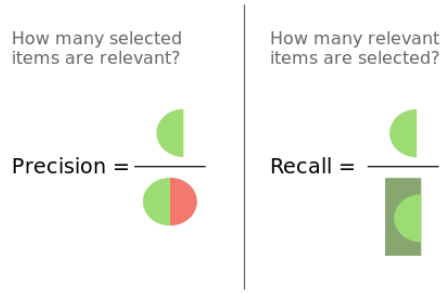
NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

Evaluation Measures

- Ground truth edges T, edges removed by an approach T'
- Precision: $\frac{|T \cap T'|}{|T'|}$
- Recall: $\frac{|T \cap T'|}{|T|}$
- F Measure: $2*$(precision $*$ recall)/(precision + recall)



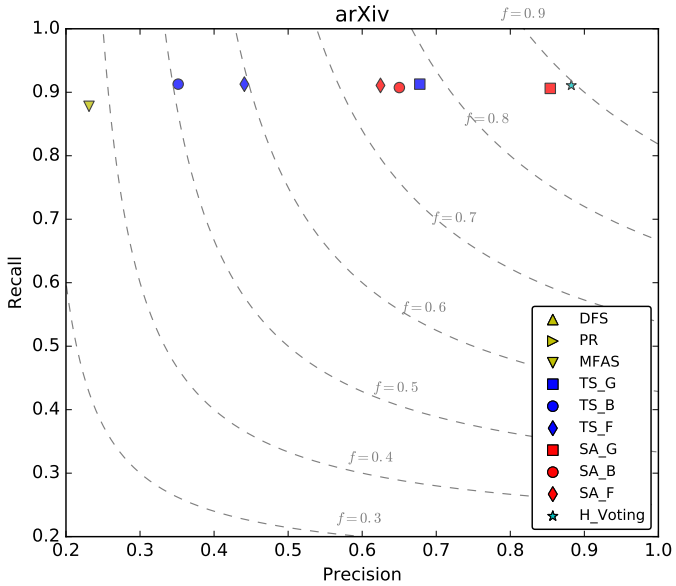


NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

[0]Figure: http://bit.ly/2piTCZv

Performance on Real Graphs

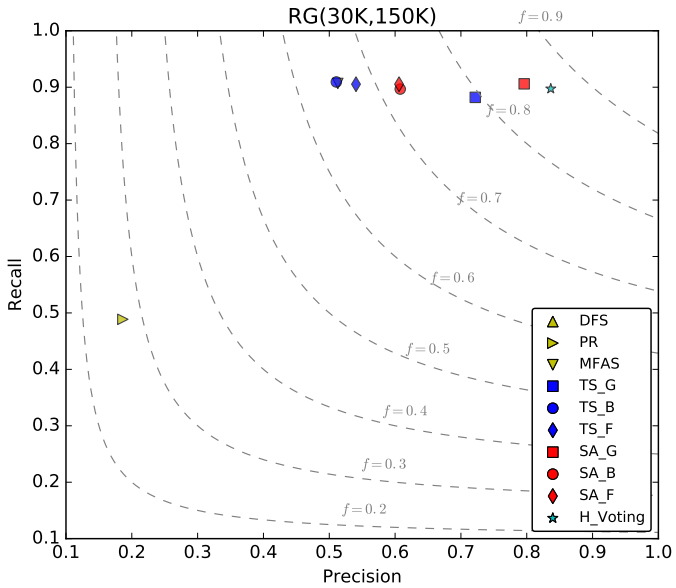


- Results on more real datasets showing comparable results are available on our paper

NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

Performance on Synthetic Graphs

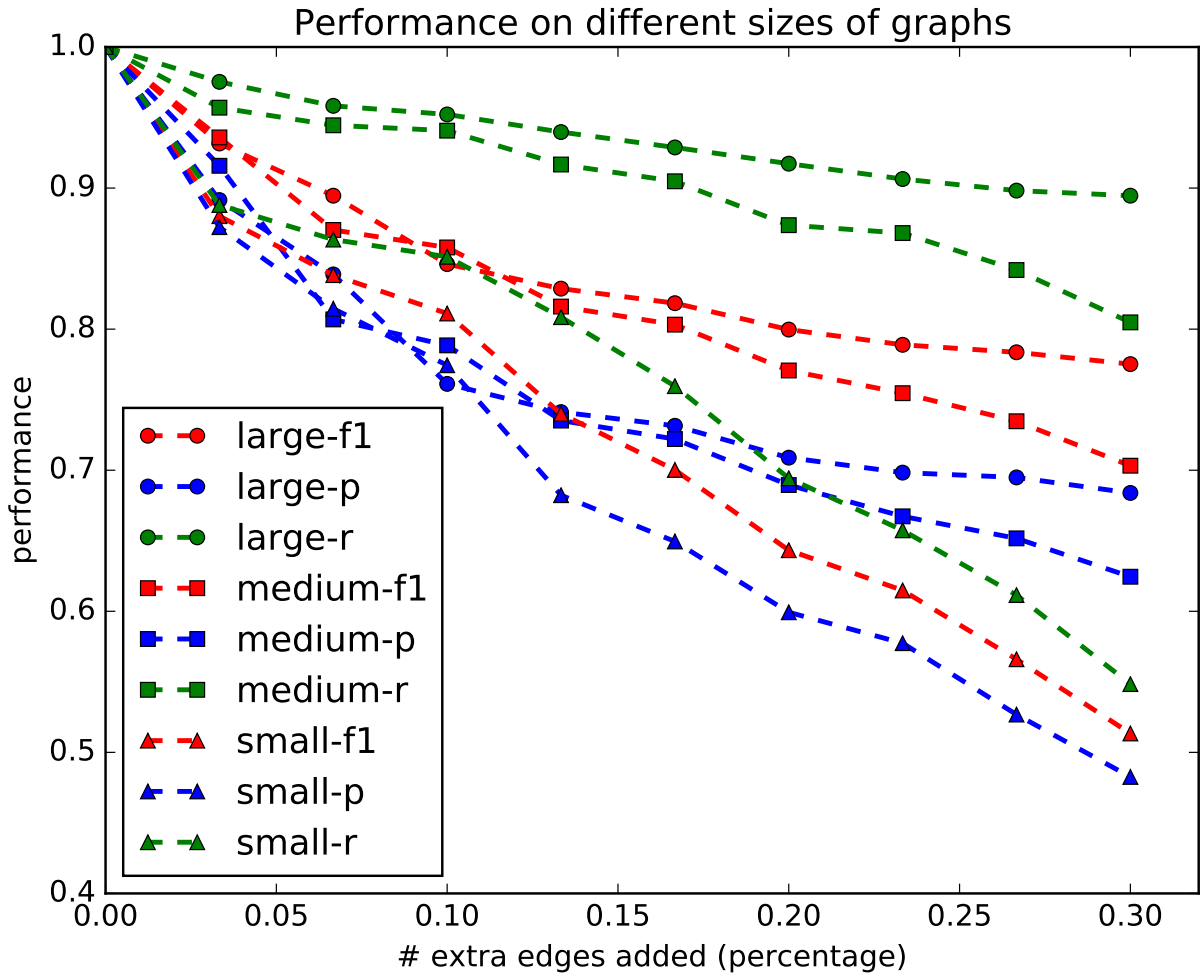


- Results on more synthetic datasets showing comparable results are available on our paper

NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

Sensitivity to Number of Noisy Edges



NOKIA

THE OHIO STATE UNIVERSITY

Conclusion & Future Work

- Main Contribution
  - our approach addresses the problem of breaking cycles while preserving the graph hierarchy
  - we are the first researchers to infer graph hierarchy by viewing it as a competition problem
  - we propose several strategies and an ensemble approach to identify edges that should be removed
- Future Work
  - propose a model-based approach to predict which edge should be removed
- **Code is available on GitHub** [1]

NOKIA Bell Labs

THE OHIO STATE UNIVERSITY

[1] https://goo.gl/491v7q

Q & A
Thanks

NOKIA Bell Labs

THE OHIO STATE UNIVERSITY