

Breaking Cycles in Noisy Hierarchies

Jiankai Sun ¹

Deepak Ajwani ²

Patrick Nicholson ²

Alessandra Sala ²

Srinivasan Parthasarathy ¹

¹The Ohio State University

²Bell Labs, Nokia, Ireland

WebSci'17, June 26 -28, 2017

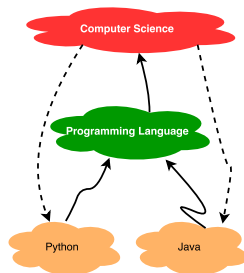
Outline

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



Motivation

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



Related Work

- Simple Heuristic Based on BFS or DFS
 - DFS: un-deterministic
 - BFS: remove more edges, even non-cycle edges
- Minimum Feedback Arc Set
 - NP-Hard
 - Cannot preserve graph hierarchy
- Domain-specific Algorithms



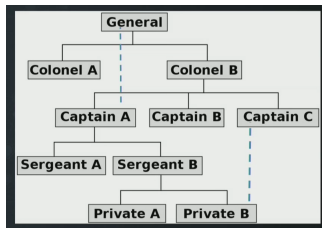
Graph Hierarchy Based Framework

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

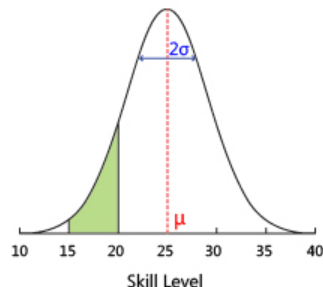
Finding a ranking function to infer graph hierarchy

- 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



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



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 .

Inferring Graph Hierarchy by TrueSkill

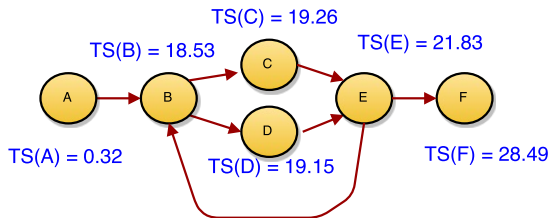


Figure: TrueSkill Computation Demo

- A node v 's ranking score in the graph hierarchy: $f_{ts}(v) = \mu_v - 3\sigma_v$
- As far as we know, **graph hierarchy inference as a competition problem** has not been researched yet

Inferring Graph Hierarchy by Social Agony

- In social networks such as Twitter, people are **not likely** to follow people who are **lower** in the hierarchy
- **lower** can be caused when people follow other people who are lower in the hierarchy
- Social agony proposed by Gupte et al. assumes the existence of a link indicates a **rank recommendation**
 - An edge $u \Rightarrow v$ indicates a recommendation of v from u
 - If there is no reverse edge from v to u , it could indicate that v is higher up in the hierarchy than u



NOKIA Bell Labs


THE OHIO STATE
UNIVERSITY

⁰Figure: <http://bit.ly/2r7afHV>

Computation of Graph Agony

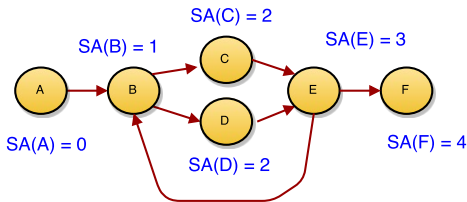
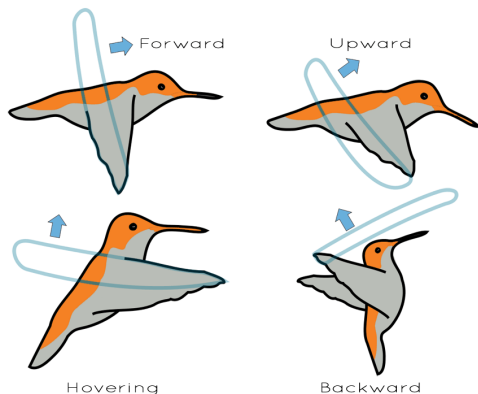


Figure: SocialAgony Computation Demo

- 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



Forward to select edges to remove and break cycles

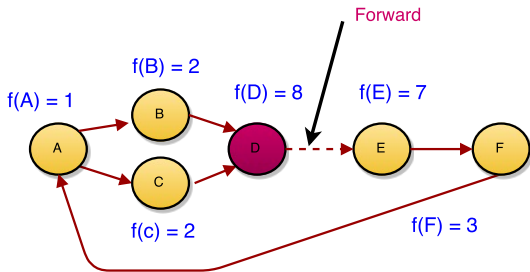


Figure: Strategy Forward to select violation edges

- *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

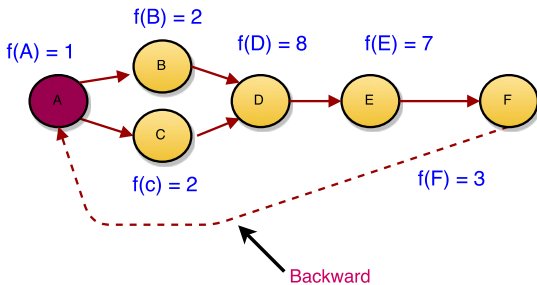


Figure: Strategy Forward to select violation edges

- **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

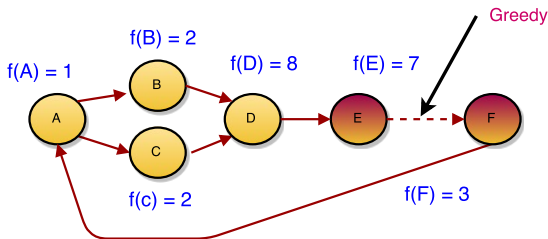


Figure: Strategy Forward to select violation edges

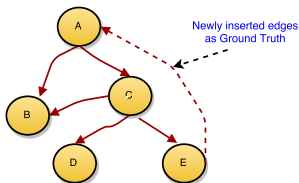
- *Greedy*: Select the edge which violates the hierarchy the *most* to remove.

Combine Them Together

- **Two** ways to infer graph hierarchy: TrueSkill and SocialAgony
- **Three** solutions to select edges: *Forward, Backward, Greedy*
- \Rightarrow **Six** 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

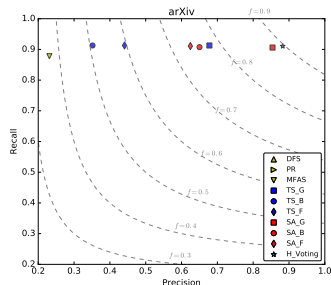
Experimental Setup

- 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



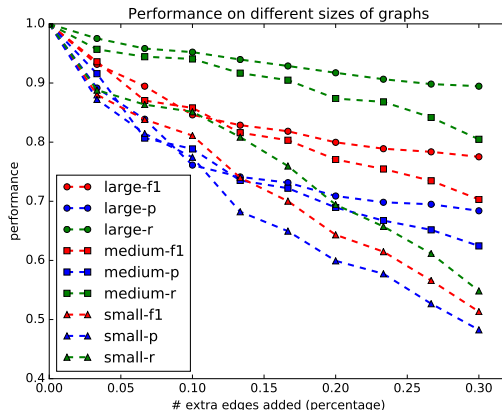
- Evaluation Measures: precision, recall, and f-measure

Performance



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

Sensitivity to Number of Noisy Edges



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

¹<https://goo.gl/491v7q>

Q & A
Thanks