Maverick: Discovering Exceptional Facts from Knowledge Graphs

Gensheng Zhang, Damian Jimenez, Chengkai Li

SIGMOD, Houston, June 14, 2018

UNIVERSITY OF
TEXAS
ARLINGTON

iDiR
idir.uta.edu

Exceptional Facts

abc NEWS

Denzel Washington followed Sidney Poitier as only the second black to win the Best Actor award.

Entity of interest	Denzel Washington
Context	Best Actor award winners
Attributes	Ethnicity
Peculiar value	African American (only two satisfy)

iDiR

idir.uta.edu

Exceptional Facts

abc NEWS

Denzel Washington followed Sidney Poitier as only the second black to win the Best Actor award.

Entity of interest	Denzel Washington	Given an entity **x** **find**
Context	Best Actor award winners	A context
Attributes	Ethnicity	A set of attributes (subspace)
Peculiar value	African American (only two satisfy)	

Exceptional Facts

abc NEWS

Denzel Washington followed Sidney Poitier as only the second black to win the Best Actor award.

Entity of interest	Denzel Washington	Given an entity **x**	
		find	**such that**
Context	Best Actor award winners	A context	the context has many entities, including **x**
Attributes	Ethnicity	A set of attributes (subspace)	
Peculiar value	African American (only two satisfy)		**x** bears a peculiar value w.r.t. the subspace (few in the context have the value)

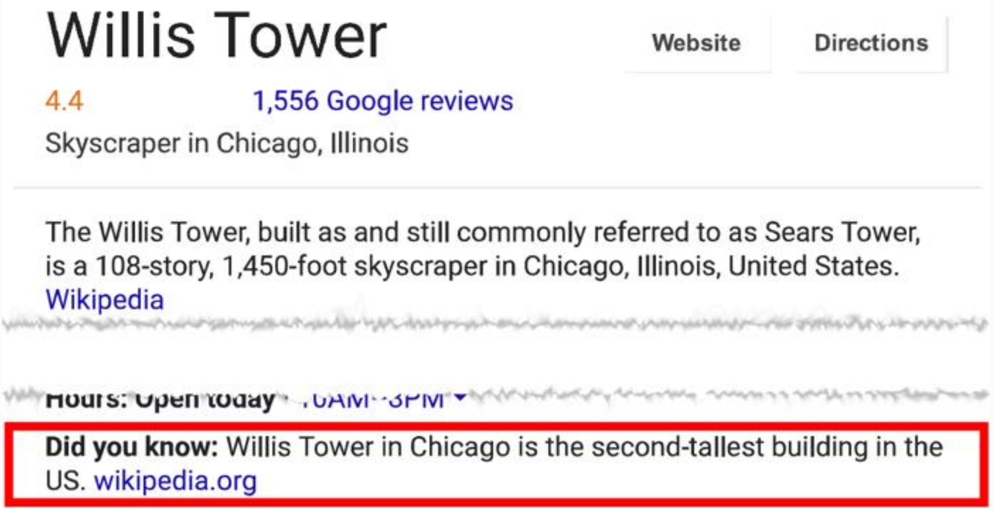
Exceptional Facts

abc NEWS — Denzel Washington followed Sidney Poitier as only the second black to win the Best Actor award.

YAHOO! SPORTS — This was Brazil's first own goal in World Cup history.

Chicago Tribune — Hillary Clinton becomes first female presidential nominee.

Applications



Willis Tower

4.4 1,556 Google reviews

Skyscraper in Chicago, Illinois

The Willis Tower, built as and still commonly referred to as Sears Tower, is a 108-story, 1,450-foot skyscraper in Chicago, Illinois, United States. Wikipedia

Hours: Open today · 10AM–8PM

Did you know: Willis Tower in Chicago is the second-tallest building in the US. wikipedia.org

Computational Journalism

- Fact-finding
- Fact-checking
  - The first female presidential nominee was Victoria Woodhull, not Hillary Clinton (snopes.com)

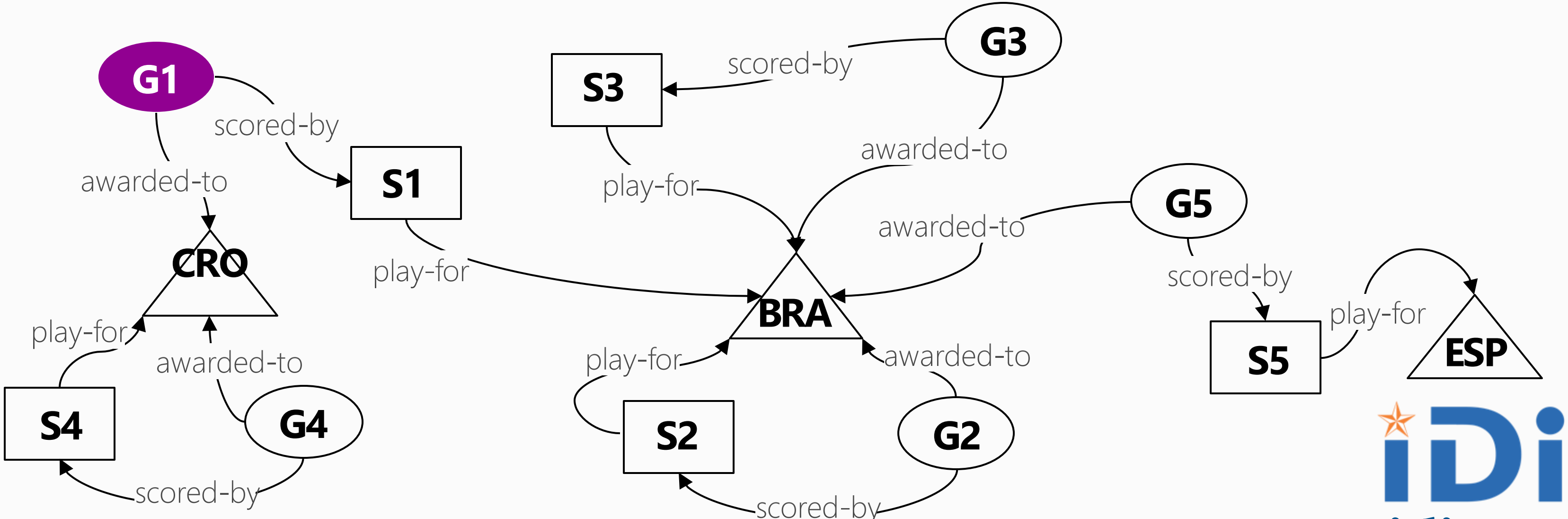
Data Cleaning

Recommendation Systems

- Friends, news, and product promotion

Exceptional Facts from Knowledge Graphs

What is exceptional about G1?



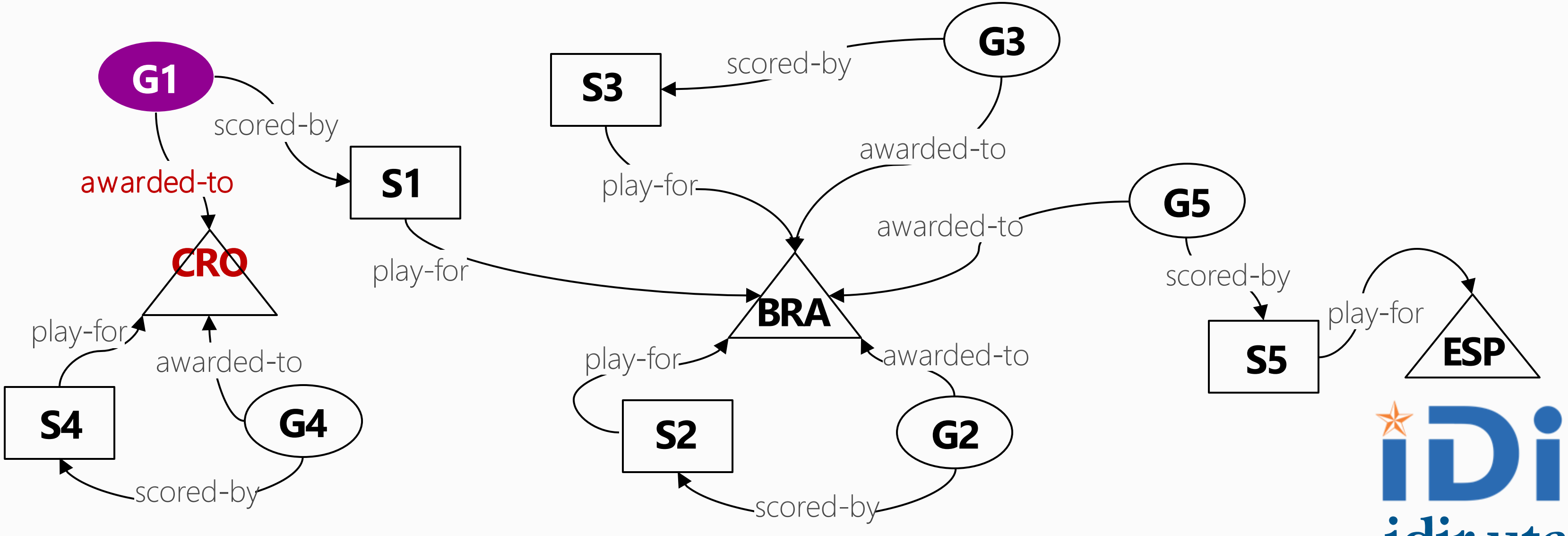
idir.uta.edu

Modeling

Attributes: labels of incoming/outgoing edges

Subspace: a subset of attributes

G1. awarded-to = CRO

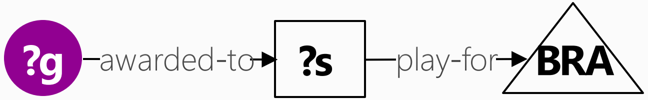


idir.uta.edu

Modeling

Context: entities sharing some common characteristics

Defined by a pattern-variable pair

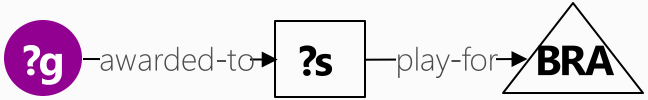


Goals scored by Brazilian players

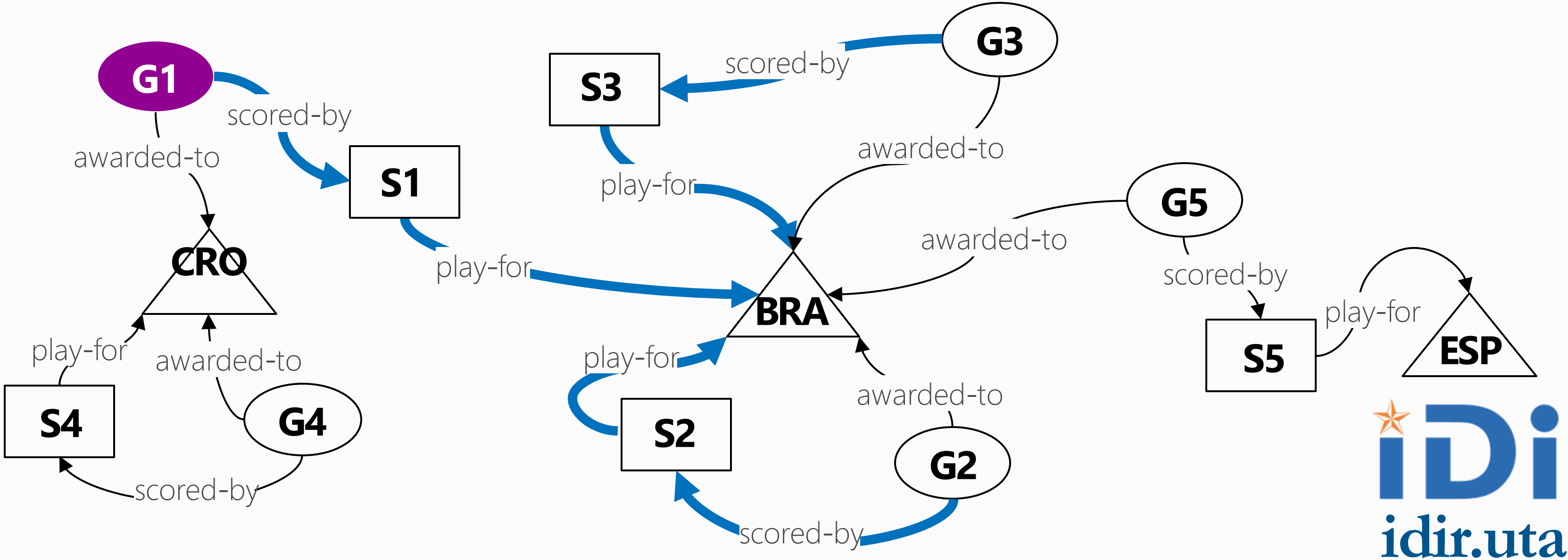
Modeling

Context: entities sharing some common characteristics

Defined by a pattern-variable pair



Goals scored by Brazilian players

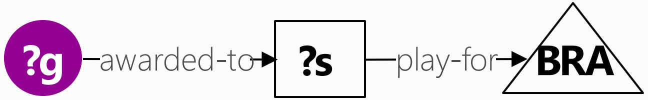


idir.uta.edu

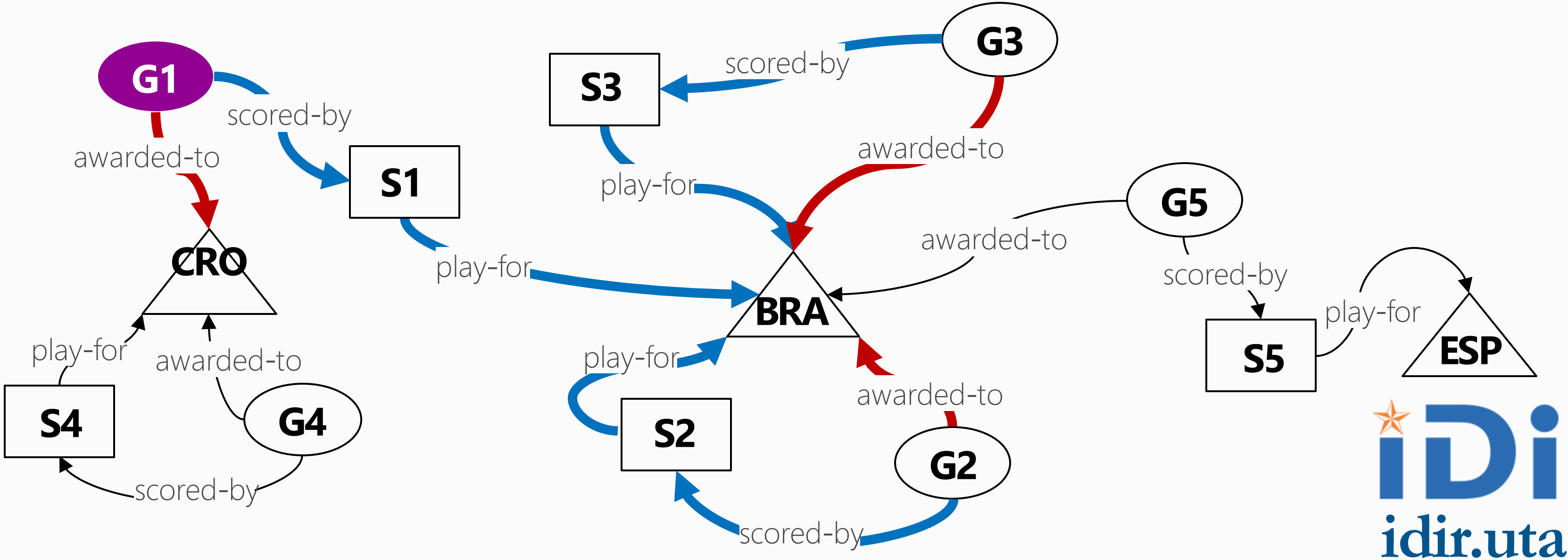
Modeling

Context: entities sharing some common characteristics

Defined by a pattern-variable pair



Goals scored by Brazilian players

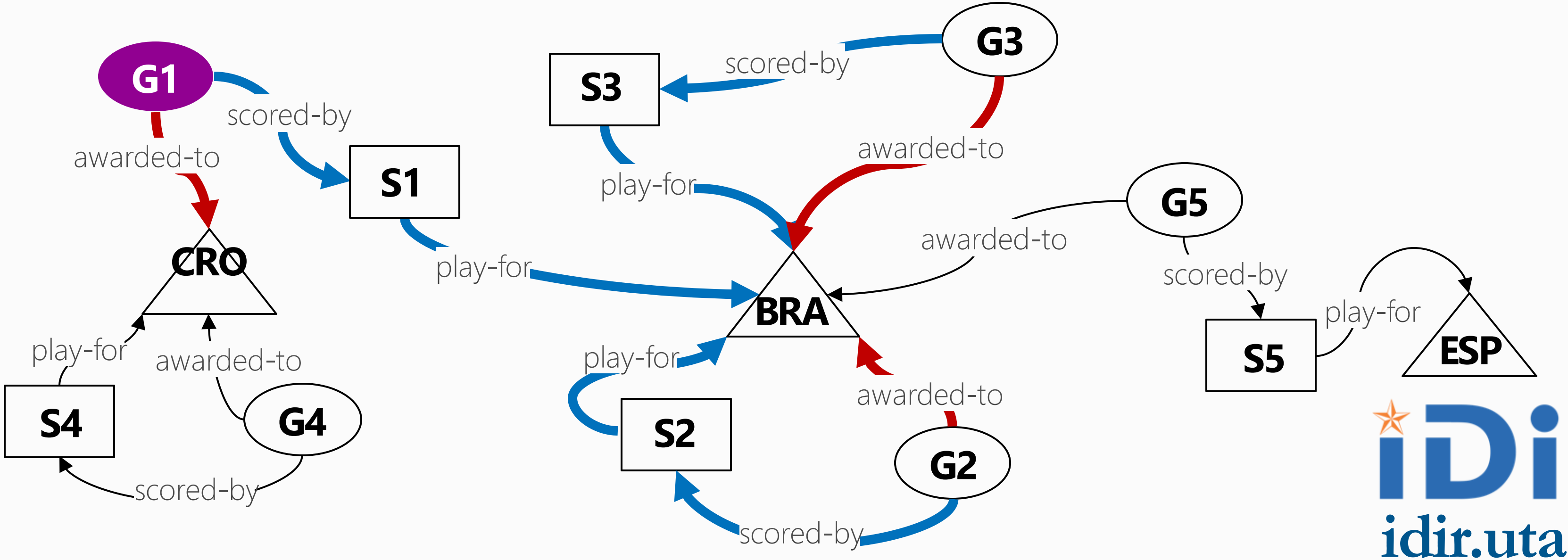


idir.uta.edu

Modeling

What is exceptional about G1?

Among all the goals scored by BRA players, G1 is the only own goal.



idir.uta.edu

Problem Formulation

Input

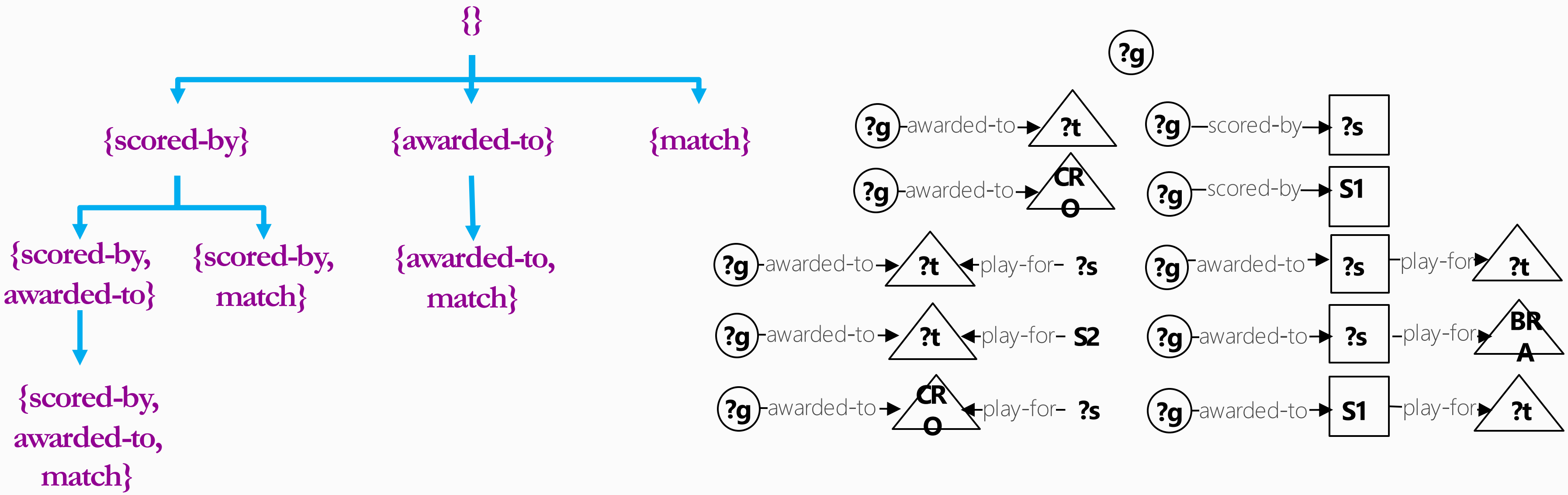
- Entity of interest v_0
- Exceptionality function χ
- Result size k

Output

- Top-k (context, subspace) pairs with regard to χ, in which v_0 stands out

Challenges

- Number of attribute subspaces: $O(2^{|A_{v_0}|})$
- Number of patterns (contexts): $\Omega(2^{|V_G|})$


{}
{scored-by}
{awarded-to}
{match}
{scored-by, awarded-to}
{scored-by, match}
{awarded-to, match}
{scored-by, awarded-to, match}
?g
?g awarded-to ?t
?g awarded-to CRO
?g scored-by ?s
?g scored-by S1
?g awarded-to ?t play-for ?s
?g awarded-to ?t play-for S2
?g awarded-to CRO play-for ?s
?g awarded-to ?s play-for ?t
?g awarded-to ?s play-for BRA
?g awarded-to S1 play-for ?t

Related Work

Outlier detection

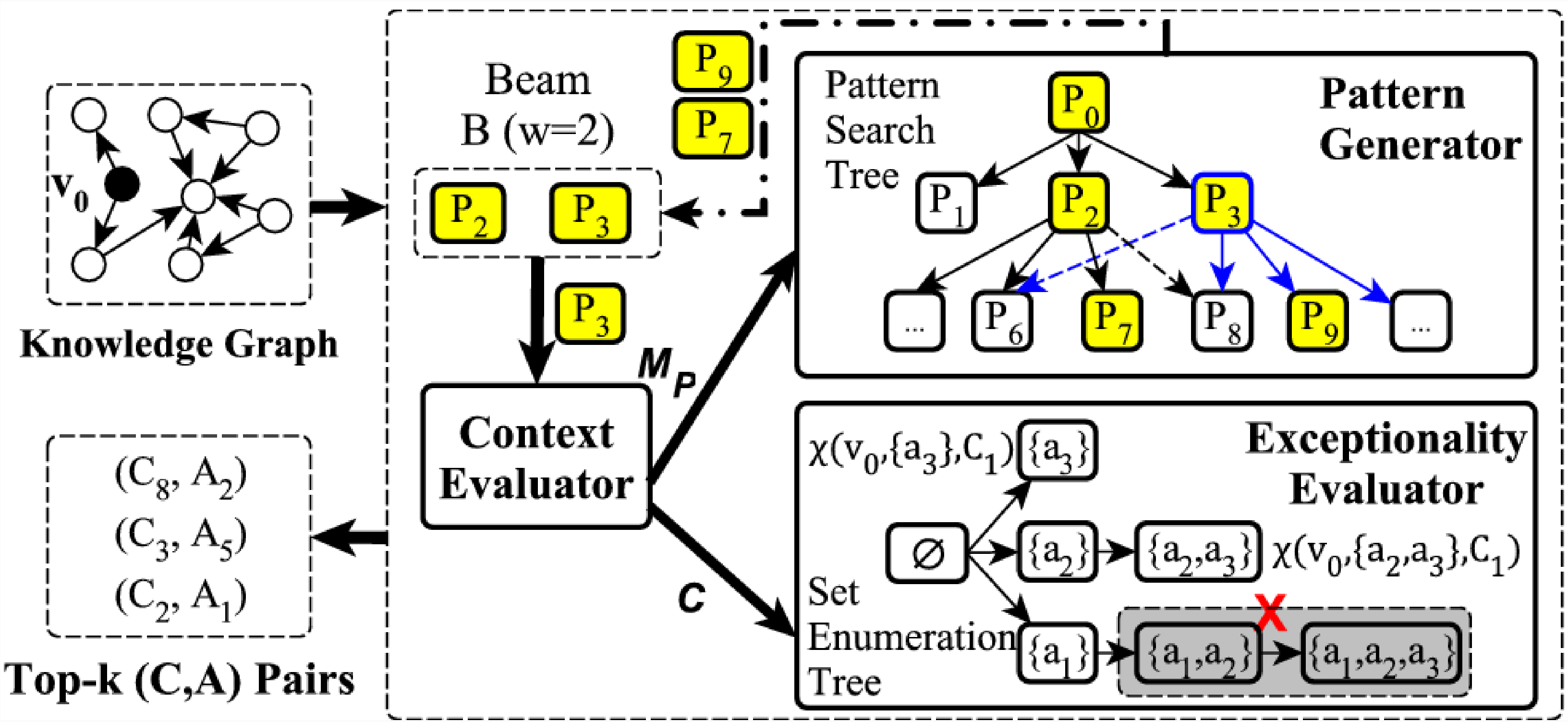
- Maverick finds conditions that make an object stand out, although the object may not necessarily be an outlier.

Outlying aspect mining

Challenges in adopting existing algorithms:

- Many assume a single-table model: a graph can be an extremely large and sparse table
- Conjunctive queries on a single table ≠ pattern queries
- Multiple tables: unclear how to handle joins
- Unclear how to handle set values

Maverick



v_0
Knowledge Graph
Beam
B (w=2)
P_2
P_3
P_9
P_7
P_3
Context Evaluator
M_P
C
Pattern Search Tree
Pattern Generator
P_0
P_1
P_2
P_3
...
P_6
P_7
P_8
P_9
...
Exceptionality Evaluator
$\chi(v_0,\{a_3\},C_1)$
$\{a_3\}$
∅
$\{a_2\}$
$\{a_2,a_3\}$
$\chi(v_0,\{a_2,a_3\},C_1)$
Set Enumeration Tree
$\{a_1\}$
$\{a_1,a_2\}$
X
$\{a_1,a_2,a_3\}$
(C_8, A_2)
(C_3, A_5)
(C_2, A_1)
Top-k (C,A) Pairs

Exceptionality Function χ

$\chi(v, A, C) \in \mathcal{R}$

outlierness (χ_o) [Angiuli2009TODS], one-of-the-few(χ_f) [Wu2012KDD], isolation scores (χ_i) [Liu2008ICDM]

Upper bound function

Theorem 4.2

$$upper_o(v, A, C) = \sum_{S \in \mathcal{S}_A} (p_S)^2 - \frac{(2\,p_{v.A} + 1) \times |\,C\,| - 2}{|\,C\,|^2}$$

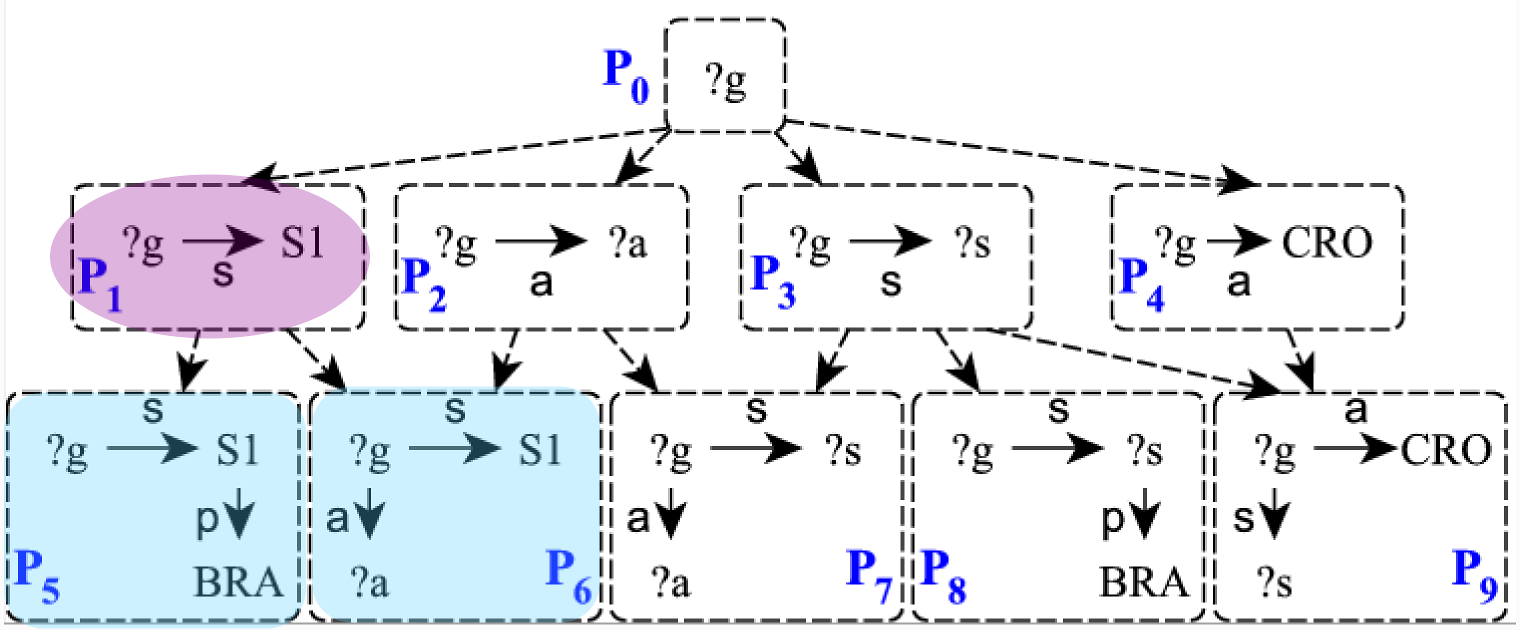
Theorem 4.3

$$upper_f(v, A, C) = |\,\{u \mid u \in \overline{C_v},\ p^A_{u.A} > 1/|\,C\,|\}\,| \Big/ |\,C\,|$$

Theorem 4.4

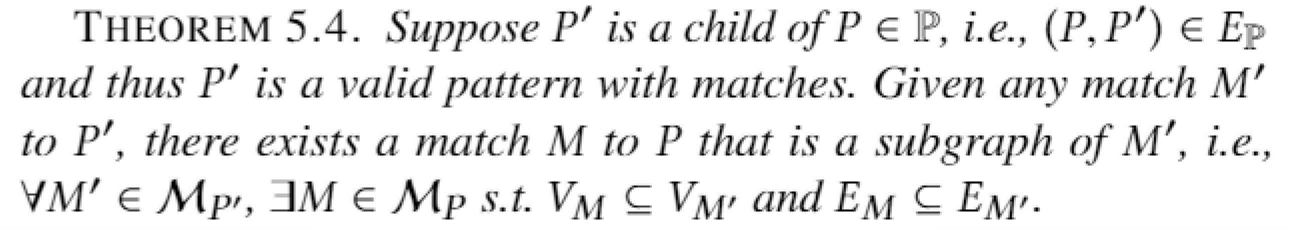
$$upper_i(v, A, C) = 1 - 2^{-\frac{-\log_2 \frac{1}{|\,C\,|}}{-q_{v.A} - sum_{S \in \mathcal{S}_A \setminus \{v.A\}}\, p_S \times \log_2 p_S}}$$

Pattern Generator (PG)

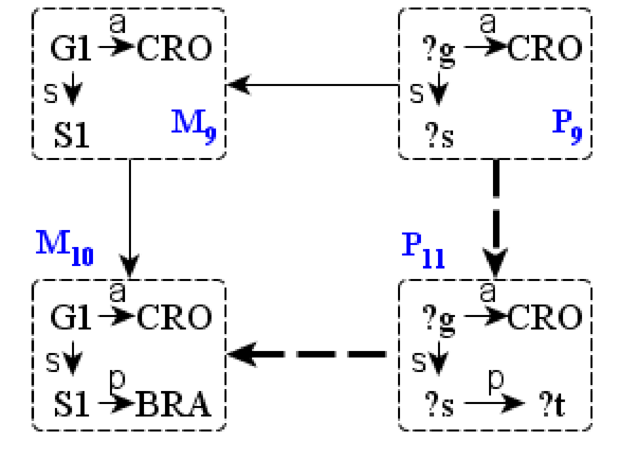


Match–based Pattern Generation

- Construct Partial Order of Valid Patterns



THEOREM 5.4. *Suppose P' is a child of $P \in \mathbb{P}$, i.e., $(P, P') \in E_{\mathbb{P}}$ and thus P' is a valid pattern with matches. Given any match M' to P', there exists a match M to P that is a subgraph of M', i.e., $\forall M' \in \mathcal{M}_{P'}, \exists M \in \mathcal{M}_{P}$ s.t. $V_M \subseteq V_{M'}$ and $E_M \subseteq E_{M'}$.*



idir.uta.edu

Datasets and Experiments

WCGoals

Created based on FIFA.com

11 node types, 13 edge types

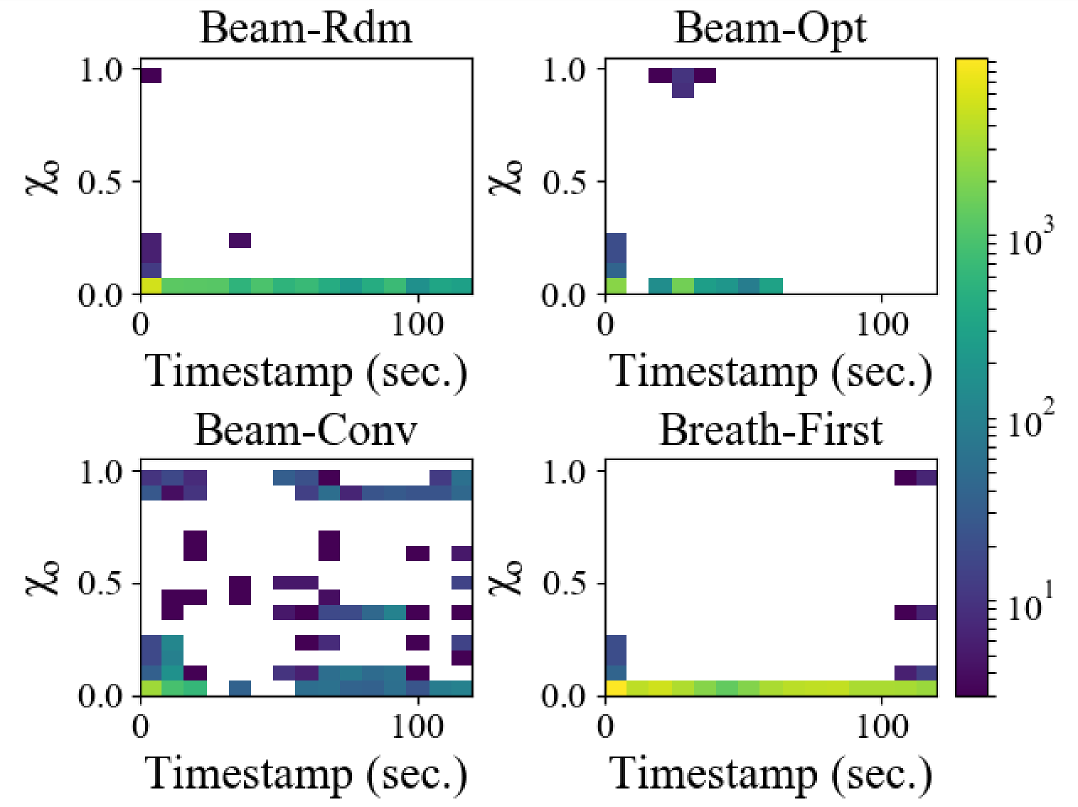
49,078 nodes, 158,114 edges

Film-Award

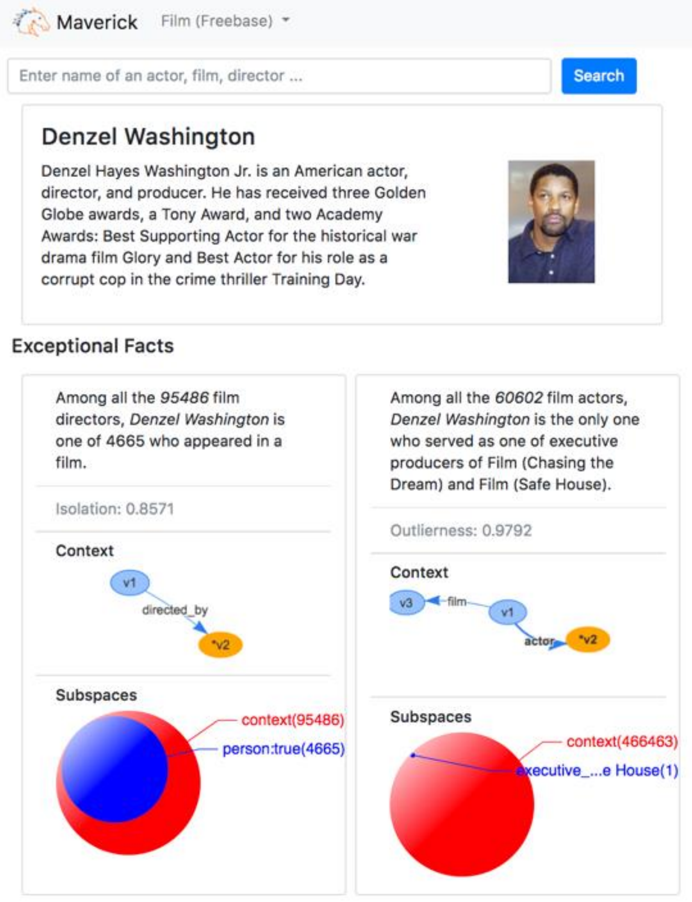
A subgraph of Freebase

95 node types, 117 edge types

5,437,628 nodes, 10,879,448 edges



See you in Rio



Denzel Washington

Denzel Hayes Washington Jr. is an American actor, director, and producer. He has received three Golden Globe awards, a Tony Award, and two Academy Awards: Best Supporting Actor for the historical war drama film Glory and Best Actor for his role as a corrupt cop in the crime thriller Training Day.

Exceptional Facts

Among all the *95486* film directors, *Denzel Washington* is one of 4665 who appeared in a film.

Isolation: 0.8571

Context

Subspaces

Among all the *60602* film actors, *Denzel Washington* is the only one who served as one of executive producers of Film (Chasing the Dream) and Film (Safe House).

Outlierness: 0.9792

Context

Subspaces

VLDB2018 demo

Maverick: A System for Discovering Exceptional Facts from Knowledge Graphs

idir.uta.edu