**PRINCEx: Provider-side Interpretablility with Counterfactual Explanations in Recommender Systems**

Azin Ghazimatin1, Oana Balalau2, Rishiraj Saha Roy4, and Gerhard Weikum4

1 Max Planck Institute for Informatics, Germany  2 Inria and Ecole Polytechnique, France

**MOTIVATION**

- Explain why I received this recommendation!
- Setup: Heterogeneous information networks, model-aware
- Explanations are usually path-based
- But need to be actionable, concise and scrutuable!
- Our solution: PRINCE

**What is new in PRINCE?**

- **Actionable:** User’s own actions
- **Concise:** Use minimal sets
- **Scrubtable:** Use counterfactual setup

**Formally:** PRINCE finds a minimal set of actions whose removal displaces the top recommendation.

**PRINCE METHOD**

**SwapOrder FUNCTION**

Input: \( G = (V, E, B), u, rec, l \)
Output: \( A', rec' \)
Initialize: \( A' = \emptyset \), \( rec' = rec \)
for each \( i \in I \)
- if \( |A'| < |A| \)
  - \( A' = \text{SwapOrder}(G, u, rec, i) \)
- \( A' = A', rec' = i \)
return \( A', rec' \)

**Model:** RecWalk [Nikolaopoulos and Karypis WSDM 2019]

**Representative Example**

Your recommendation: Baby stroller

**PRINCE produces more concise explanations**

**Graph baselines noisy approximations of PRINCE**

Paths potentially violate privacy

**User Study**

- **PRINCE vs. CredPaths** [Yang et al. ICDM 2018]
- Rigorous presentation bias and spam control
- PRINCE is judged more useful at all explanation sizes

Action-based explanations in recommenders are both feasible and useful!

Contact: aghazima@mpi-inf.mpg.de and rishiraj@mpi-inf.mpg.de

**Graph Results**

- **Datasets:** Amazon and Goodreads
- **PRINCE is difficult to approximate**
- **Explanations shrink with increasing k**

<table>
<thead>
<tr>
<th>k</th>
<th>Amazon PRINCEx HC</th>
<th>Goodreads PRINCEx HC</th>
<th>Amazon HC</th>
<th>Goodreads HC</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5.09*</td>
<td>6.87</td>
<td>5.20</td>
<td>7.57</td>
</tr>
<tr>
<td>5</td>
<td>3.41*</td>
<td>4.62</td>
<td>3.01</td>
<td>5.01</td>
</tr>
<tr>
<td>10</td>
<td>2.66*</td>
<td>3.66</td>
<td>1.66</td>
<td>4.15</td>
</tr>
<tr>
<td>15</td>
<td>2.13*</td>
<td>3.00</td>
<td>1.11</td>
<td>3.68</td>
</tr>
<tr>
<td>20</td>
<td>1.80*</td>
<td>2.39</td>
<td>1.12</td>
<td>3.28</td>
</tr>
</tbody>
</table>

**Mean**

<table>
<thead>
<tr>
<th>Method</th>
<th>Mean</th>
<th>Std.Dev.</th>
<th>#Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRINCE</td>
<td>1.91*</td>
<td>0.66</td>
<td>200</td>
</tr>
<tr>
<td>CredPaths</td>
<td>1.78</td>
<td>0.63</td>
<td>200</td>
</tr>
<tr>
<td>PRINCE (Size=1)</td>
<td>1.87</td>
<td>0.66</td>
<td>154</td>
</tr>
<tr>
<td>PRINCE (Size=2)</td>
<td>1.88*</td>
<td>0.70</td>
<td>28</td>
</tr>
<tr>
<td>PRINCE (Size=3)</td>
<td>2.21*</td>
<td>0.52</td>
<td>18</td>
</tr>
</tbody>
</table>