Guided Learning for Role Discovery (GLRD): Framework, Algorithms, and Applications

Presented by: Sean Gilpin*

Tina Eliassi-Rad†, Ian Davidson*

*University of California, Davis, Department of Computer Science
† Rutgers University, Department of Computer Science
Roadmap

• Motivation / Previous Work / Contributions
• Guidance / Formulation
• Experiments
What is Role Discovery?

What is Role Discovery?

**Input:**
- Graph

**Guidance:**
- Sparsity
- Diversity
- Alternative

**Outputs:**
- Assignment
- Role Definitions

- Bridge
- Tightly Knit
- Main Stream
- Pathy

Sean Gilpin - University of California Davis
Our Contribution

ReFeX
It’s Who You Know: Graph Mining Using Recursive Structural Features
In KDD 2011

GLRD
Our Work

RolX
RolX: Structural Role Extraction & Mining in Large Graphs,
In KDD 2012.

Sean Gilpin - University of California Davis
Roadmap

• Motivation / Previous Work / Contributions
• **Guidance / Formulation**
• Experiments
NMF Interpretation

Role assignment vector.

\[ V \approx G \]

Soft role assignments for individual user.

Role definition / explanation.

Sean Gilpin - University of California Davis
# Guidance Overview

<table>
<thead>
<tr>
<th>Guidance</th>
<th>Effect of increasing guidance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>on role assignment</td>
</tr>
<tr>
<td></td>
<td>on role definition</td>
</tr>
<tr>
<td>Sparsity</td>
<td>Reduces the number of nodes with minority memberships in roles</td>
</tr>
<tr>
<td></td>
<td>Decreases likelihood that features with small explanatory benefit are included</td>
</tr>
<tr>
<td>Diversity</td>
<td>Limits the amount of allowable overlap in assignments</td>
</tr>
<tr>
<td></td>
<td>Roles must be explained with completely different sets of features</td>
</tr>
<tr>
<td>Alternative</td>
<td>Decreases the allowable similarity between the two sets of role assignments</td>
</tr>
<tr>
<td></td>
<td>Ensures that role definitions are very dissimilar between the two sets of role assignments</td>
</tr>
</tbody>
</table>

Sean Gilpin - University of California Davis
Sparsity

\[
\underset{G,F}{\text{argmin}} \quad \|V - GF\|_2
\]

subject to: \quad G \geq 0, F \geq 0

\forall i \quad \|G_{i*}\|_1 \leq \epsilon_G

\forall i \quad \|F_{i*}\|_1 \leq \epsilon_F

where \quad \epsilon_G \text{ and } \epsilon_F \text{ define upperbounds for the sparsity constraints (amount of allowable density).}
Diversity

\[
\begin{align*}
\text{argmin}_{G,F} & \quad ||V - GF||_2 \\
\text{subject to:} & \quad G \geq 0, F \geq 0 \\
& \quad \forall i, j \quad G^{T}_{i,j} G_{i,j} \leq \epsilon_G \quad i \neq j \\
& \quad \forall i, j \quad F_{i,j} F^{T}_{i,j} \leq \epsilon_F \quad i \neq j
\end{align*}
\]

where \( \epsilon_G \) and \( \epsilon_F \) define upper bounds on how angularly similar role assignments and role definitions can be to each other.

Sean Gilpin - University of California Davis
Alternative Role Discovery

\[
\text{argmin}_{G,F} \quad ||V - GF||_2 \\
\text{subject to:} \quad G \geq 0, \quad F \geq 0 \\
\forall i, j \quad G_{i}^* G_{j} \leq \epsilon_G \\
\forall i, j \quad F_i^* F_j^T \leq \epsilon_F \\
\text{where} \quad \epsilon_G \text{ and } \epsilon_F \text{ define upperbounds on how similar the results can be to } G^* \text{ and } F^*.
\]
Roadmap

• Motivation / Previous Work / Contributions
• Guidance / Formulation
• Experiments
Identity Resolution Across Graphs

• **Question**: Can diversity and sparsity constraints create better role definitions?

• **Conjecture**: Better role definitions will better facilitate other problems such as identity resolution across graphs.

• **Experiment**: Compare graph mining results using multiple methods of role discovery.

• Conferences: KDD, CIKM, SDM, ICDM, SIGMOD, VLDB.
Identity Resolution Across Graphs

Sean Gilpin - University of California Davis
Alternative Roles

• Question: Do alternative sets of roles exist in graphs and can they be discovered?
Alternative Roles

Sean Gilpin - University of California Davis
Conclusion and Future Work

• Showed how guidance could be encoded in NMF formulation of role discovery.
• Empirically showed that guidance could improve the quality of learned role definitions.
• Showed that useful alternative role definitions exist in graphs.

• Future:
  • Additional types of guidance
  • More complex settings.
Questions

- Thank You.

- Supported in part by grants from ONR, NSF, DARPA SMISC, DTRA, and LLNL.