Discovering Roles and Anomalies in Graphs: Theory and Applications

Part 1: Theory

Tina Eliassi-Rad (Rutgers)
Christos Faloutsos (CMU)

SDM'12 Tutorial
Overview

Roles

Features

Anomalies

Patterns

= rare roles
Overview

Roles

Features

Anomalies

Patterns

= rare roles
Roadmap

• What are roles
• Roles and communities
• Roles and equivalences (from sociology)
• Roles (from data mining)
• Summary
What are roles?

• “Functions” of nodes in the network
  – Similar to functional roles of species in ecosystems
• Measured by structural behaviors
• Examples
  – centers of stars
  – members of cliques
  – peripheral nodes
Example of Roles

- Centers of stars
- Members of cliques
- Peripheral nodes
### Why are roles important?

#### Role Discovery

- **Automated discovery**
- **Behavioral roles**
- **Roles generalize**

<table>
<thead>
<tr>
<th>Task</th>
<th>Use Case</th>
</tr>
</thead>
<tbody>
<tr>
<td>Role query</td>
<td>Identify individuals with similar behavior to a known target</td>
</tr>
<tr>
<td>Role outliers</td>
<td>Identify individuals with unusual behavior</td>
</tr>
<tr>
<td>Role dynamics</td>
<td>Identify unusual changes in behavior</td>
</tr>
<tr>
<td>Identity resolution</td>
<td>Identify known individuals in a new network</td>
</tr>
<tr>
<td>Role transfer</td>
<td>Use knowledge of one network to make predictions in another</td>
</tr>
<tr>
<td>Network comparison</td>
<td>Determine network compatibility for knowledge transfer</td>
</tr>
</tbody>
</table>
Roadmap

• What are roles
• Roles and communities
• Roles and equivalences (from sociology)
• Roles (from data mining)
• Summary
Roles and Communities

• Roles group nodes with similar structural properties

• Communities group nodes that are well-connected to each other

• Roles and communities are complementary
Roles and Communities

Roles and Communities

Consider the social network of a CS dept

• Roles
  – Faculty
  – Staff
  – Students
  – ...

• Communities
  – AI lab
  – Database lab
  – Architecture lab
  – ...

SDM'12 Tutorial
T. Eliassi-Rad & C. Faloutsos
Roadmap

• What are roles
• Roles and communities
• Roles and equivalences (from sociology)
• Roles (from data mining)
• Summary
Equivalences

- Deterministic
- Automorphic
- Structural
- Stochastic
- Probabilistic
- Regular
Deterministic Equivalences

Regular

Automorphic

Structural
Structural Equivalence

- [Lorrain & White, 1971]
- Two nodes $u$ and $v$ are structurally equivalent if they have the same relationships to all other nodes.
- Hypothesis: Structurally equivalent nodes are likely to be similar in other ways – i.e., you are your friend.
- Weights & timing issues are not considered.
- Rarely appears in real-world networks.
Structural Equivalence: Algorithms

- CONCOR (CONvergence of iterated CORrelations) [Breiger et al. 1975]

- A hierarchical divisive approach
  1. Starting with the adjacency matrix, repeatedly calculate Pearson correlations between rows until the resultant correlation matrix consists of +1 and -1 entries
  2. Split the last correlation matrix into two structurally equivalent submatrices (a.k.a. blocks): one with +1 entries, another with -1 entries

- Successive split can be applied to submatrices in order to produce a hierarchy (where every node has a unique position)
Structural Equivalence: Algorithms

- STRUCUTRE [Burt 1976]
- A hierarchical agglomerative approach
  1. For each node $i$, create its ID vector by concatenating its row and column vectors from the adjacency matrix
  2. For every pair of nodes $(i, j)$, measure the square root of sum of squared differences between the corresponding entries in their ID vectors
  3. Merge entries in hierarchical fashion as long as their difference is less than some threshold $\alpha$
Structural Equivalences: Algorithms

• Combinatorial optimization approaches
  – Numerical optimization with tabu search [UCINET]
  – Local optimization [Pajek]

• Partition the sociomatraces into blocks based on a cost function that minimizes the sum of within block variances
  – Basically, minimize the sum of code cost within each block
Deterministic Equivalences

Regular

Automorphic

Structural
Automorphic Equivalence

• [Borgatti, et al. 1992; Sparrow 1993]

• Two nodes $u$ and $v$ are automorphically equivalent if all the nodes can be relabeled to form an isomorphic graph with the labels of $u$ and $v$ interchanged
  
  – Swapping $u$ and $v$ (possibly along with their neighbors) does not change graph distances

• Two nodes that are automorphically equivalent share exactly the same label-independent properties
Automorphic Equivalence: Algorithms

- Sparrow (1993) proposed an algorithm that scales linearly to the number of edges.
- Use numerical signatures on degree sequences of neighborhoods.
- Numerical signatures use a unique transcendental number like $\pi$, which is independent of any permutation of nodes.
- Suppose node $i$ has the following degree sequence: 1, 1, 5, 6, and 9. Then its signature is $S_{i,1} = (1 + \pi)(1 + \pi)(5 + \pi)(6 + \pi)(9 + \pi)$.
- The signature for node $i$ at $k+1$ hops is $S_{i,(k+1)} = \Pi(S_{i,k} + \pi)$.
- To find automorphic equivalence, simply compare numerical signatures of nodes.
Deterministic Equivalences

Regular

Automorphic

Structural
Regular Equivalence

- [Everett & Borgatti, 1992]
- Two nodes $u$ and $v$ are regularly equivalent if they are equally related to equivalent others

Regular Equivalence (continued)

- Basic roles of nodes
  - source
  - repeater
  - sink
  - isolate
Regular Equivalence (continued)

• Based solely on the social roles of neighbors

• Interested in
  – Which nodes fall in which social roles?
  – How do social roles relate to each other?

• Hard partitioning of the graph into social roles

• A given graph can have more than one valid regular equivalence set

• Exact regular equivalences can be rare in large graphs
Regular Equivalence: Algorithms

- Many algorithms exist here

- Basic notion
  - Profile each node’s neighborhood by the presence of nodes of other "types"
  - Nodes are regularly equivalent to the extent that they have similar "types" of other nodes at similar distances in their neighborhoods
Equivalences

- Deterministic
- Probabilistic
  - Stochastic
- Automorphic
- Structural
  - Regular
Stochastic Equivalence

- Two nodes are stochastically equivalent if they are “exchangeable” w.r.t. a probability distribution
- Similar to structural equivalence but probabilistic
Stochastic Equivalence: Algorithms

• Many algorithms exist here
• Most recent approaches are generative [Airoldi, et al 2008]
• Some choice points
  – Parametric vs. non-parametric models
Roadmap

• What are roles
• Roles and communities
• Roles and equivalences (from sociology)
  • Roles (from data mining)
• Summary
RolX: Role eXtraction

- Introduced by Henderson, et al. 2011b
- Automatically extracts the underlying roles in a network
  - No prior knowledge required
- Assigns a mixed-membership of roles to each node
- Scales linearly on the number of edges
RolX: Flowchart

**Input**

Node × Node Matrix → Recursive Feature Extraction → Node × Feature Matrix

**Output**

Role Extraction

Role × Feature Matrix → Node × Role Matrix
**RolX: Flowchart**

- **Input**
  - Node $\times$ Node Matrix
  - **Recursive Feature Extraction**
  - **Node $\times$ Feature Matrix**
  - **Role Extraction**
  - **Role $\times$ Feature Matrix**
  - **Node $\times$ Role Matrix**

- **Output**
  - Example: degree, avg weight, # of edges in egonet, mean clustering coefficient of neighbors, etc
Recursive Feature Extraction

- ReFeX [Henderson, et al. 2011a] turns network connectivity into recursive structural features

- Neighborhood features: What is your connectivity pattern?
- Recursive Features: To what kinds of nodes are you connected?
Propositionalisation (PROP)

- From multi-relational data mining with roots in Inductive Logic Programming (ILP)
- Summarizes a multi-relational dataset (stored in multiple tables) into a propositional dataset (stored in a single “target” table)
- Derived attribute-value features describe properties of individuals
- Related more to recursive structural features than structural roles
Role Extraction

Recursively extract roles

Automatically factorize roles
Automatically Discovered Roles

Network Science
Co-authorship Graph
[Newman 2006]
Making Sense of Roles

<table>
<thead>
<tr>
<th>NodeSense</th>
<th>NeighborSense</th>
</tr>
</thead>
<tbody>
<tr>
<td>Degree</td>
<td>Role 1</td>
</tr>
<tr>
<td>Avg. Weight</td>
<td>Role 2</td>
</tr>
<tr>
<td>Clustering Coeff.</td>
<td>Role 3</td>
</tr>
<tr>
<td>Eccentricity</td>
<td>Role 4</td>
</tr>
<tr>
<td>PageRank</td>
<td>Role 4</td>
</tr>
<tr>
<td>GateKeeper</td>
<td>Role 3</td>
</tr>
<tr>
<td>GateKeeper-Local</td>
<td>Role 2</td>
</tr>
<tr>
<td>Pivot</td>
<td>Role 1</td>
</tr>
<tr>
<td>Structural Hole</td>
<td>Role 2</td>
</tr>
</tbody>
</table>

Roles:
- Role 1: clique
- Role 2: bridge
- Role 3: periphery
- Role 4: isolated
- Default
Mixed-Membership over Roles

- Bright red nodes are locally central nodes
- Bright blue nodes are peripheral nodes

Amazon Political Books Co-purchasing Network
[V. Krebs 2000]
Role Query

Node Similarity for M.E.J. Newman (*bridge*)

Node Similarity for J. Rinzel (*isolate*)

Node Similarity for F. Robert (*cliquey*)
Roles Generalize across Disjoint Networks
Roles Generalize across Networks

Discovery Stage

- Feature Discovery
  - V₁
  - L

- RolX
  - G₁
  - F

- Learning
  - M

V: (node × feature) matrix
G: (node × role) matrix
F: (role × feature) matrix
L: List of feature names
C: Class labels
M: model

E.g., degree, avg wgt, etc

SDM'12 Tutorial
Roles Generalize across Networks

E.g., degree, avg wgt, etc

V: (node × feature) matrix
G: (node × role) matrix
F: (role × feature) matrix
L: List of feature names
C: Class labels
M: model

Discovery Stage  Application Stage
## Roles: Regular Equivalence vs. RolX

<table>
<thead>
<tr>
<th></th>
<th>RolX</th>
<th>Regular Equivalence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mixed-membership over roles</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Fully automatic</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Uses structural features</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Uses structure</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Generalizable across disjoint</td>
<td>✓</td>
<td>?</td>
</tr>
<tr>
<td>networks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scalable (linear on # of edges)</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>
Roadmap

• What are roles
• Roles and communities
• Roles and equivalences (from sociology)
• Roles (from data mining)
• Summary
Summary

• Roles
  – Structural behavior (“function”) of nodes
  – Complementary to communities
  – Previous work mostly in sociology under equivalences
  – Recent graph mining work produces mixed-membership roles, is fully automatic and scalable
  – Can be used for many tasks: transfer learning, re-identification, node dynamics, etc
Acknowledgement

• LLNL: Brian Gallagher, Keith Henderson
• IBM Watson: Hanghang Tong
• Google: Sugato Basu
• CMU: Leman Akoglu, Danai Koutra, Lei Li

Thanks to: LLNL, NSF, IARPA.
References

Deterministic Equivalences


References


References


• D.R. White, K. Reitz: Graph and Semi-Group Homomorphism on Networks and Relations. *Social Networks*, 5:143-234, 1983.
References

Stochastic blockmodels


SDM'12 Tutorial T. Eliassi-Rad & C. Faloutsos
References

Role Discovery


References

Community Discovery


Propositionalisation


Back to Overview

Roles

Features

Anomalies

= rare roles

Patterns