Persistent Roles in Online Social Networks
ECML-PKDD 2016, M. Revelle, C. Domeniconi, and A. Johri

Contents

Introduction

Roles

Roles are a method us to analyse networks

- Graph similarity
- Network modelling
- Visualization

Problem of Roles

- Roles are network-specific giving it limited comparative options with other networks with another set of roles
- In dynamic networks it is possible for roles to change or disappear over time
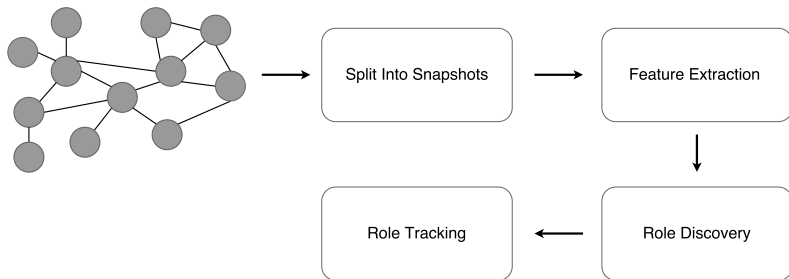
The Goal of the Article

The purpose of this paper is to present a novel method that uses features based on interaction patterns and structural positions of a users to:

- Find roles independent of dataset
- Find roles that persists over time

This method should utilize the latent factors to identify the feature structure of the roles.
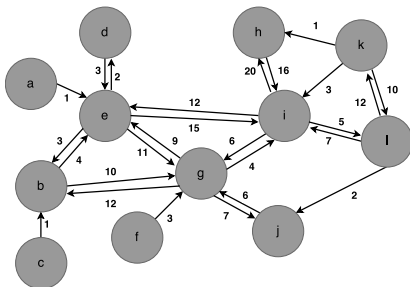
Methodology

The Approach

The Data

Datasets:

- Facebook - Wall posts from one user to another
- Scratch - Comments on uploaded programming projects

Snapshots

The datasets are split into a total of 26 snapshots:

- 7 from Facebook
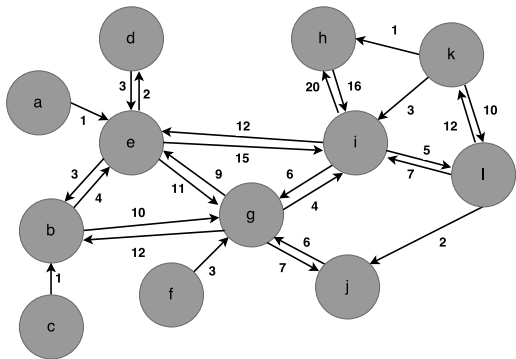- 19 from Scratch

Social Network Graph
$D = (N, E)$

Snapshot
$S_t = (N_t, E_t)$

Feature Selection

- In-degree
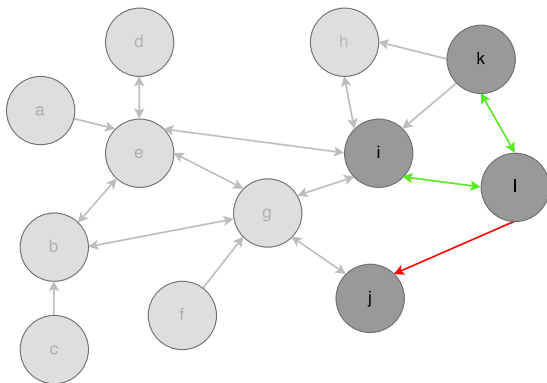- Out-degree
- Weighted in-degree
- Weighted out-degree

Feature Selection

Reciprocity is the rate a user is replayed.

Reciprocity

$$r = \frac{\text{Out-degree}^{<->}}{\text{Out-degree}}$$

Feature Selection

The *new activity count* feature is the number of new outgoing edges based on the difference of snapshot S_t and S_{t-1}

> **New Activity Count**
>
> $$c = \text{Out-degree}_t - \text{Out-degree}_{t-1}$$

The feature *social strategy* is the ratio of new outgoing edges over all outgoing edges
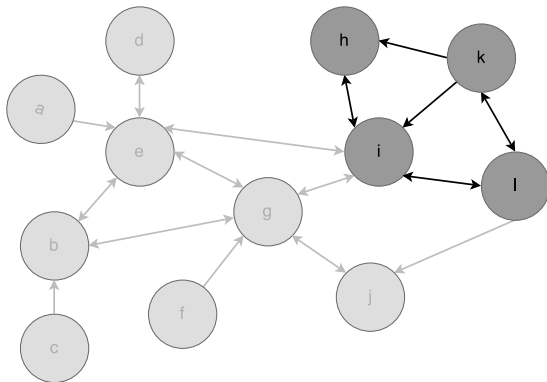
> **Social Strategy**
>
> $$s = \frac{\text{New Activity Count}}{\text{Out-degree}}$$

Feature Selection

The value of the feature *betweenness centrality* is the product of the number of shortest paths passing through a vertex

Betweenness Centrality

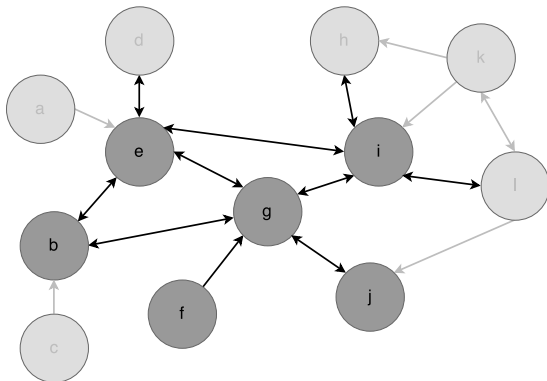$$b(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Feature Selection

The feature *PageRank* measures centrality based on ingoing edges and their influence

PageRank

$$pr(g) = 1 - d + d\left(\frac{pr(b)}{2} + \frac{pr(e)}{4} + \frac{pr(f)}{1} + \frac{pr(i)}{4} + \frac{pr(j)}{1}\right)$$
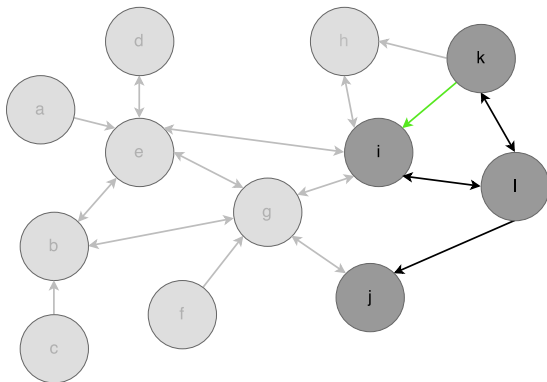
Feature Selection

Transitivity measures is the local clustering coefficient which gives the probability of a vertexes neighbours being connected

Transitivity

$$C(v_i) = \frac{|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i-1)}$$

Selected Features
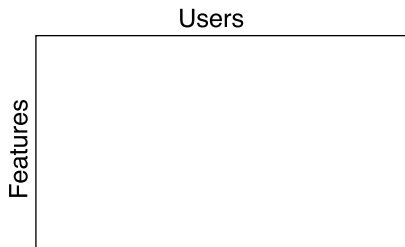
Summary of the features

- In-degree
- Out-degree
- Weighted in-degree
- Weighted out-degree
- Reciprocity
- New activity count
- Social strategy
- Betweenness centrality
- PageRang
- Transitivity

Features omitted from the walk-though

- Weighted PageRank
- Weighted transitivity

Feature Extraction

The result of feature extraction is a features × users feature matrix

Users

Features

The features in the matrix is normalized by the use of feature scaling so that all values is within the interval of 1 and 0
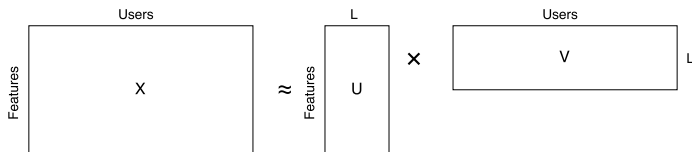
Role Discovery

Roles are discovered by decomposing the feature matrix using NMF. This gives us:

- Basis matrix U, containing the feature characteristic for the roles
- Coefficient matrix V, containing role membership weights for each user

Non-negative Matrix Factorization

$$X \approx UV$$

Selection of L

Root Mean Squired Error

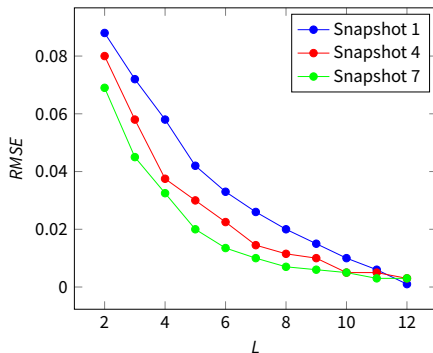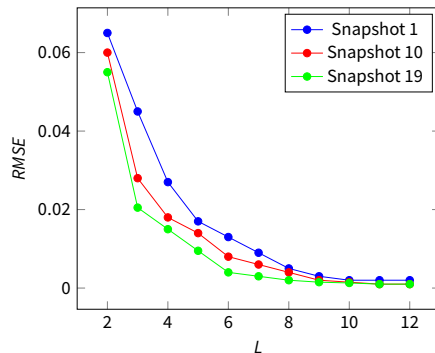$$RMSE = \sqrt{\frac{1}{|X|} \sum_{(u,f) \in X} (X_{u,f} - X'_{u,f})^2}$$

Figure: Facebook

Figure: Scratch

Tracing Roles

Cosine Similarity

$$sim(A_i, B_i) = \frac{\displaystyle\sum_{i=1}^{n} A_i B_i}{\sqrt{\displaystyle\sum_{i=1}^{n} A_i^2} \sqrt{\displaystyle\sum_{i=1}^{n} B_i^2}}$$

Cosine similarity returns a value
between -1 and 1

- -1 means that the vectors are
  opposites
- 1 means they are exactly the
  same

The article sets a threshold on 0.75
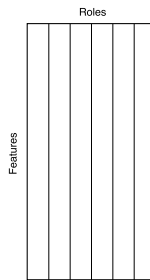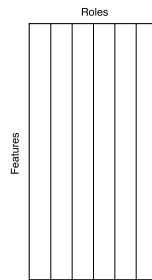that roles from S_t and S_{t+1} must have
similarity measure above

Figure: U_t

Figure: U_{t+1}

Results

The Persistent Roles

| Role	Feature Characteristics
Popular	In-degree, Betweenness, PageRank
Friendly	Out-degree
Reciprocated	Reciprocity
Explorer	Social Strategy
Community Member	Transitivity
Active Community Member	Weighted transitivity

Feature Representation of the Roles

Role Features
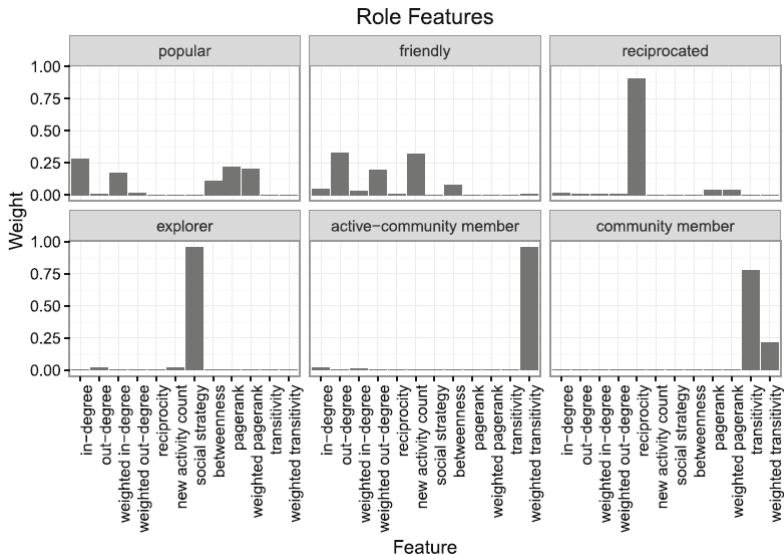
Interaction Preferences

Role Preferences
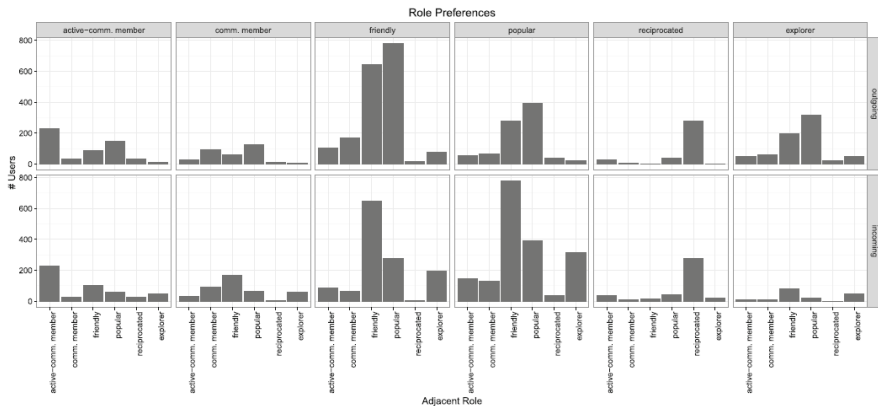
Evidence of Roles Dependence on Network Structure
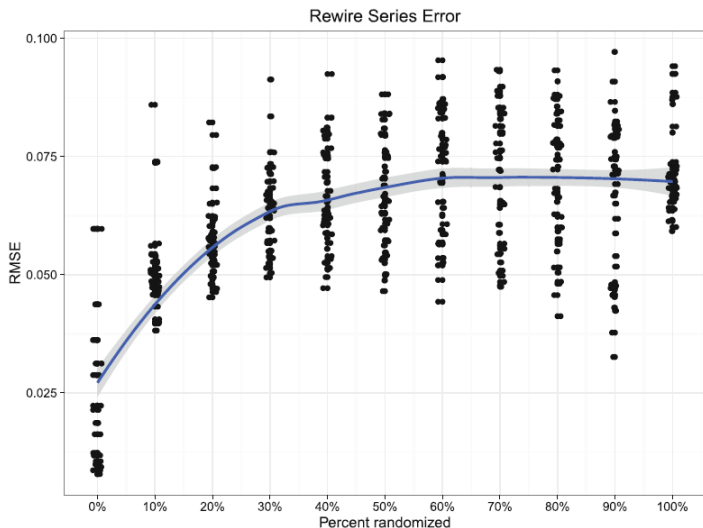
Conclusion

Conclusion

This paper presents a feature-based methodology which identify six different roles that are:

- Persistent throughout the timespan of the datasets
- Independently derived from different datasets

Shortcomings

- The authors does not argue for their selection of features or give a reference to an article that does.
- I would have appreciated some more examples or are more detailed description of some of their approaches.
- The distribution of snapshots are not consistent.