Link Prediction

Latha Pemula
Theoretical Justification to Popular Link Prediction Heuristics
What is Link Prediction?

• How does Facebook, LinkedIn suggest friends?
• Applications in recommender systems
Link Prediction

- Which pair of nodes \{i,j\} should be connected?
- Variant: node \(i\) is given

Should Facebook suggest Alice to Bob as a future friend?
Popular Approaches

- Predicting links between pair of nodes with most common neighbours
- Weighing the common neighbours according to a function of their degrees
- Consider number of paths between two neighbours. Weigh short paths more than long paths
- This paper provides intuitions to heuristics
Latent Space

• Nodes connect based on homophily (similar characters). Eg: geographic location, work place, school

• Each node in D dimensional Latent space where the distance between two nodes defines the closeness of two nodes

• Nodes are distributed uniformly in the unit volume hyper sphere

• Link prediction problem is reduced to estimating distances between pair of nodes
Link Prediction – Latent Space

\[(\text{modified}) \text{ RHH model: } P(i \sim j|d_{ij}) = \frac{1}{1 + e^{\alpha(d_{ij} - r)}}\]

Higher probability of linking

$\alpha$ determines the steepness

radius $r$
Deterministic Model with same radii

• All nodes have same radii $r$ and probability is a step
• Implies if $d_{ij} < r$, they form a link
• Close to a regular graph
• Degree are distributed as $\text{bin}(N,V(r))$
Common Neighbors

\[ \Pr_2(i, j) = \Pr(\text{common neighbor} \mid d_{ij}) \]

\[ \Pr_2(i, j) = \int \Pr(i \sim k \mid d_{ik}) \Pr(j \sim k \mid d_{jk}) P(d_{ik}, d_{jk} \mid d_{ij}) \partial d_{ik} \partial d_{jk} \]

Product of two logistic probabilities, integrated over a volume determined by \( d_{ij} \)

As \( \alpha \to \infty \) Logistic \( \to \) Step function
Much easier to analyze!
Common Neighbors

Everyone has same radius $r$

$\Pr_2(i, j) = A(r, r, d_{ij})$

$\eta$=Number of common neighbors

Unit volume universe

Empirical Bernstein

Bounds on distance

$V(r)$=volume of radius $r$ in $D$ dims
Common Neighbors

• OPT = node closest to i
• MAX = node with max common neighbors with i

• Theorem:

\[ d_{\text{OPT}} \leq d_{\text{MAX}} \leq d_{\text{OPT}} + 2[\varepsilon/V(1)]^{1/D} \]

\[ \varepsilon = c_1 \left( \text{var}_N/N \right)^{1/2} + c_2/(N-1) \]

D = dimensionality

Common neighbors is an asymptotically optimal heuristic as \( N \to \infty \)
Deterministic Model with Distinct Radii

• Each node has different radius, $r_i$
• Degree is proportional to $N V(r_i)$
• Number of common nodes:
  • Types
• Type2 gives us bound \( d_{ij} \leq d_{ik} + d_{jk} \leq 2r_k \)
  if \( r_k \) is big, it is not good enough bound

• So common neighbours should be weighed differently, depending on their radii
Type 2 common neighbors

\[ w(r) = \frac{\partial A}{\partial d_{ij}} \frac{A(1 - A)}{A} \]

**Variance**

Small variance \( \rightarrow \) Presence is more surprising

Small variance \( \rightarrow \) Absence is more surprising

**Jacobian**

\[ w(r) \approx \frac{\text{const}}{r} \approx \frac{\text{const}}{\deg^{1/D}} \]

- \( r \) is close to max radius
- Real world graphs generally fall in this range
Observations

• The weights are high for both nodes of very low radius and for high radius
• Presence of a low radius common neighbor indicates that $d_{ij}$ is small.
• Absence of very high degree node in common neighbors suggests that $d_{ij}$ is very large
• Real graphs have high degree nodes 10-20% N so real graphs $V(r) << 1$
Adamic/Adar model justification

• Similarity between webpages

\[
\text{Adamic/Adar} = \sum_{k \in \mathcal{N}(i) \cap \mathcal{N}(j)} \frac{1}{\log(\text{deg}(k))}
\]
Take aways

• Number of common neighbors between a pair of nodes gives bounds on distance between them

• Upper bound decreases quickly as the number of common neighbors increases

• Justifies popular heuristic of predicting links based on number of common neighbors
Estimators using Longer Paths in Deterministic Model

- There may not be any common neighbors of $i$ and $j$ if $d_{ij} > r$ or No points fall in the intersection of $A(r,r,d_{ij})$ due to small sample size $N$.
- In such cases we consider paths of length $l>2$.
- Even if common neighbors exist, this can provide tight bounds on $d_{ij}$. Specially if there are many longer paths and fewer common neighbors.
$\ell$ hop Paths

- Common neighbors = 2 hop paths

- Analysis of longer paths: two components
  1. Bounding $E(\eta_\ell | d_{ij})$. [\(\eta_\ell = \# \ell\) hop paths]
     - Bounds $Pr_\ell(i,j)$ by using triangle inequality on a series of common neighbor probabilities.
  2. $\eta_\ell \approx E(\eta_\ell | d_{ij})$
     Bounds of form
     - $d_{ij} \leq lr(1 - g(\beta_\ell(i,j), \varepsilon))$
Conclusions

• When short paths are non existent or rare, bounds obtained on $d_{ij}$ are loose
• Longer paths can yield tighter bounds
• As number of hops increases, more long paths are required to obtain a tighter bound
• The existence of a short path can improve upper bounds obtained from longer paths
• Number of paths is important to the bound
Non Deterministic case

- $\alpha$ is finite
- $\alpha$ is small yields close to random graphs. Link prediction does not work
- For high $\alpha$, bounds become looser but results are qualitatively similar
Revisiting Raftery et al.’s model

\[
\frac{1}{4} A + \frac{1}{2} e^{-\alpha d_{ij}} (V - A) \leq \text{Pr}_2(i, j) \leq A + c(r, \alpha, D)
\]

Factor ¼ weak bound for Logistic

✓ Can be made tighter, as logistic approaches the step function.
Summary

• Simple heuristic of counting common neighbors performs well
• Weighting common neighbors performs better
• Longer paths can be used to obtain bounds when shorter paths are absent
• Bounds obtained are tighter if short paths are known to exist
Inferring Social Ties across Heterogeneous Networks
Key Contributions

• Often people do not create labels for their relation with other users
• Develop a framework for classifying the type of social relationships by learning across heterogeneous networks
• Infers type of social relationships in a target network by borrowing knowledge from a difference source network
• Eg: Inferring social ties in a communication network using a reviewer network
• Propose a Transfer based factor graph (TanFG) model

Figure 1: Example of inferring social ties across two heterogeneous networks: a reviewer network and a mobile communication network.
Datasets

• Undirected:
  • Epinions : Network of product reviewa - Trust/Distrust
  • Slashdot:Network of friends for Technology related news– friends/foes
  • Mobile: Network of mobile users logs of calls, Bluetooth scanning, Cell tower IDs

• Directed:
  • Coauthor: Network of authors – advisor - advisee
  • Enron: email communication network – Manager-subordinate
Problem Statement

• $G=(V, E_l, E_u, X)$ denote partially labeled social network, where $E_l$ is labeled relations, $E_u$ is unlabeled relations, $X$ is attribute matrix of the edges

• Task is to predict labels $y_i$ for edges $e_i$

• Input is $G_s$ and $G_t$ with $|E^S_l|>>|E^T_l|$

• Consider undirected networks with static relations
Factors that influence social ties

Social psychological theories

1. Social balance: People in a social network tend to form into a balanced network structure

How is social balance property satisfied in different networks.
2. Structural hole: A person who is linked to people in parts of network that are otherwise not well connected

- Would structural holes have a similar behavior pattern in different networks?

![Figure 4: Structural hole. Probabilities that two connected (or disconnected) users (A and B) have the same type of relationship with user C, conditioned on whether user C spans a structural hole or not. It is clear that (1) users are more likely (averagely +70% higher than chance) to have the same type of relationship with C if C spans a structural hole; and (2) disconnected users are more likely than connected users to have the same type of relationship with a user who spans a structural hole (except the mobile network).](image)
3. Social status

- A triad of all positive edges (directed edges indicating source has higher status) should be acyclic.
- Distribution of triads.
- How do different networks satisfy the properties of social status.

Figure 5: Illustration of status theory. (A) and (B) satisfy the status theory, while (C) and (D) do not satisfy the status theory. Here positive “+” denotes the target node has a higher status than the source node; and negative “−” denotes the target node has a lower status than the source node. In total there are 16 different cases.

Figure 6: Social status. Distribution of five most frequent formations of triads with social status. Given a triad \((A, B, C)\), let us use 1 to denote the advisor-advisee relationship and 0 colleague relationship. Thus the number 011 to denote \(A\) and \(B\) are colleagues, \(B\) is \(C\)’s advisor and \(A\) is \(C\)’s advisor.
4. Two step flow

- Ideas flow first to opinion leaders and then to a wider population
- How do different networks follow the “two-step flow” of information propagation?

![Bar chart showing network flows](image)

Figure 7: Opinion leader. OL - Opinion leader; OU - Ordinary user. Probability that two types of users have a directed relationship (from higher social status to lower status, i.e., manager-subordinate relationship in Enron and advisor-advisee relationship in Coauthor. It is clear that opinion leaders (detected by PageRank) are more likely to have a higher social-status than ordinary users.
Summary

• Probabilities of balanced triads based on communications are different in different networks while probabilities based on friendships are similar

• Users are more likely to have same type of relation with a structural hole

• Most triads satisfy social status theory

• Opinion leaders are more likely to have higher social status than ordinary users
Model Framework

• Bayes rule on Labels Y, attributes X and network G

\[ P(Y|X, G) = \frac{P(X, G|Y)P(Y)}{P(X, G)} \propto P(X|Y) \cdot P(Y|G) \]

• Attributes are independent

\[ P(Y|X, G) \propto P(Y|G) \prod_i P(x_i|y_i) \]
• Each attribute is modeled as

\[ P(x_i | y_i) = \frac{1}{Z_1} \exp \left\{ \sum_{j=1}^{d} \alpha_j g_j (x_{ij}, y_i) \right\} \]

\[ P(Y | G) = \frac{1}{Z_2} \exp \left\{ \sum_c \sum_k \mu_k h_k(Y_c) \right\} \]

• Maximize log-likelihood objective function

\[ O(\theta) = \log P_{\theta}(Y | X, G) \]

\[ \theta^* = \text{argmax} \ O(\theta) \]
Learning across Heterogenous networks

• Social theories are general for all networks
• Triad based features to denote the proportion of balanced triangles
• For structural holes, define edge correlation based features
• For social status, features over triads to represent features of seven most frequent formations of triads
• Opinion leaders, define features over each edge
Tran FG Objective function

\[ \mathcal{O}(\alpha, \beta, \mu) = \mathcal{O}_S(\alpha, \mu) + \mathcal{O}_T(\beta, \mu) \]
\[ = \sum_{i=1}^{\mid V_S \mid} \sum_{j=1}^{d} \alpha_{ij} g_j(x_{i,j}^S, y_i^S) + \sum_{i=1}^{\mid V_T \mid} \sum_{j=1}^{d'} \beta_{ij} g_j'(x_{i,j}^T, y_i^T) \]
\[ + \sum_{k} \mu_k \left( \sum_{c \in G_S} h_k(Y_c^S) + \sum_{c \in G_T} h_k(Y_c^T) \right) \]
\[ - \log Z \]

- Gradient descent to solve for $\theta^*$
Table 2: Performance comparison of different methods for inferring friendships (or trustful relationships). (S) indicates the source network and (T) the target network. For the target network, we use 40% of the labeled data in training and the rest for test.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Method</th>
<th>Prec.</th>
<th>Rec.</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epinions (S) to Slashdot (T) (40%)</td>
<td>SVM</td>
<td>0.7157</td>
<td>0.9733</td>
<td>0.8249</td>
</tr>
<tr>
<td></td>
<td>CRF</td>
<td>0.8919</td>
<td>0.6710</td>
<td>0.7658</td>
</tr>
<tr>
<td></td>
<td>PFG</td>
<td>0.9300</td>
<td>0.6436</td>
<td>0.7607</td>
</tr>
<tr>
<td></td>
<td>TranFG</td>
<td>0.9414</td>
<td>0.9446</td>
<td>0.9430</td>
</tr>
<tr>
<td>Slashdot (S) to Epinions (T) (40%)</td>
<td>SVM</td>
<td>0.9132</td>
<td>0.9925</td>
<td>0.9512</td>
</tr>
<tr>
<td></td>
<td>CRF</td>
<td>0.8923</td>
<td>0.9911</td>
<td>0.9393</td>
</tr>
<tr>
<td></td>
<td>PFG</td>
<td>0.9954</td>
<td>0.9787</td>
<td>0.9870</td>
</tr>
<tr>
<td></td>
<td>TranFG</td>
<td>0.9954</td>
<td>0.9787</td>
<td>0.9870</td>
</tr>
<tr>
<td>Epinions (S) to Mobile (T) (40%)</td>
<td>SVM</td>
<td>0.8983</td>
<td>0.5955</td>
<td>0.7162</td>
</tr>
<tr>
<td></td>
<td>CRF</td>
<td>0.9455</td>
<td>0.5417</td>
<td>0.6887</td>
</tr>
<tr>
<td></td>
<td>PFG</td>
<td>1.0000</td>
<td>0.5924</td>
<td>0.7440</td>
</tr>
<tr>
<td></td>
<td>TranFG</td>
<td>0.8239</td>
<td>0.8344</td>
<td>0.8291</td>
</tr>
<tr>
<td>Slashdot (S) to Mobile (T) (40%)</td>
<td>SVM</td>
<td>0.8983</td>
<td>0.5955</td>
<td>0.7162</td>
</tr>
<tr>
<td></td>
<td>CRF</td>
<td>0.9455</td>
<td>0.5417</td>
<td>0.6887</td>
</tr>
<tr>
<td></td>
<td>PFG</td>
<td>1.0000</td>
<td>0.5924</td>
<td>0.7440</td>
</tr>
<tr>
<td></td>
<td>TranFG</td>
<td>0.7258</td>
<td>0.8599</td>
<td>0.7872</td>
</tr>
</tbody>
</table>
Table 3: Performance comparison of different methods for inferring directed relationships (the source end has a higher social status than the target end). (S) indicates the source network and (T) the target network. For the target network, we use 40% of labeled data in training and the rest for test.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Method</th>
<th>Prec.</th>
<th>Rec.</th>
<th>F1-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coauthor (S) to</td>
<td>SVM</td>
<td>0.9524</td>
<td>0.5556</td>
<td>0.7018</td>
</tr>
<tr>
<td>Enron (T) (40%)</td>
<td>CRF</td>
<td>0.9565</td>
<td>0.5366</td>
<td>0.6875</td>
</tr>
<tr>
<td></td>
<td>PFG</td>
<td><strong>0.9730</strong></td>
<td>0.6545</td>
<td>0.7826</td>
</tr>
<tr>
<td></td>
<td>TranFG</td>
<td>0.9556</td>
<td><strong>0.7818</strong></td>
<td><strong>0.8600</strong></td>
</tr>
<tr>
<td>Enron (S) to</td>
<td>SVM</td>
<td>0.6910</td>
<td>0.3727</td>
<td>0.4842</td>
</tr>
<tr>
<td>Coauthor (T) (40%)</td>
<td>CRF</td>
<td><strong>1.0000</strong></td>
<td>0.3043</td>
<td>0.4666</td>
</tr>
<tr>
<td></td>
<td>PFG</td>
<td>0.9916</td>
<td>0.4591</td>
<td>0.6277</td>
</tr>
<tr>
<td></td>
<td>TPFG</td>
<td>0.5936</td>
<td><strong>0.7611</strong></td>
<td>0.6669</td>
</tr>
<tr>
<td></td>
<td>TranFG</td>
<td>0.9793</td>
<td>0.5525</td>
<td><strong>0.7065</strong></td>
</tr>
</tbody>
</table>
Factor contribution analysis

Figure 8: Factor contribution analysis. TranFG-SH denotes our TranFG model by ignoring the structural hole based transfer. TranFG-SB stands for ignoring the structural balance based transfer. TranFG-OL stands for ignoring the opinion leader based transfer and TranFG-SS stands for ignoring social status based transfer.
(c) Epinions-to-Mobile

(d) Slashdot-to-Mobile
Figure 10: Performance of inferring directed relationship with and w/o the status based transfer by varying the percent of labeled data in the target network.