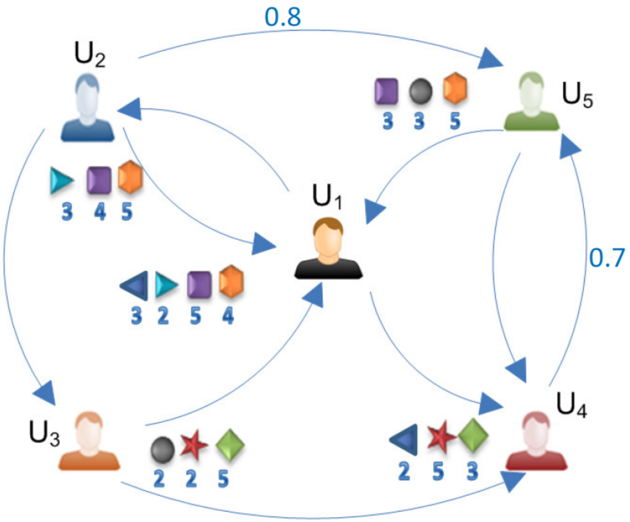
Social Recommendation, Predicting Reciprocity

Pawan Goyal

CSE, IITKGP

November 17, 2014

Recommendation in Social Networks



Effects in Social Networks

Social Influence

Ratings are influenced by ratings of friends, i.e. friends are more likely to have similar ratings than strangers

Effects in Social Networks

Social Influence

Ratings are influenced by ratings of friends, i.e. friends are more likely to have similar ratings than strangers

Benefits

- Can deal with cold-start users, as long as they are connected to the social network
- Exploit social influence, correlational influence, transitivity
- Are more robust to fraud, in particular to profile attacks

Memory Based Approaches

- Explore the network to find raters in the neighborhood of the target user
- Aggregate the ratings of these raters to predict the rating of the target user
- Different methods to calculate the “trusted neighborhood” of users

TidalTrust; Goldbeck (2005)

- Modified breadth-first search in the network
- Consider all raters v at the shortest distance from the target user u
- Trust between u and v:

$$t_{u,v} = \frac{\sum_{w \in N_u} t_{u,w} t_{w,v}}{\sum_{w \in N_u} t_{u,w}}$$

  where N_u denotes the set of (direct) neighbors (friends) of u
- Trust depends on all connecting paths

TidalTrust; Goldbeck (2005)

- Modified breadth-first search in the network
- Consider all raters v at the shortest distance from the target user u
- Trust between u and v:

$$t_{u,v} = \frac{\sum\limits_{w \in N_u} t_{u,w} t_{w,v}}{\sum\limits_{w \in N_u} t_{u,w}}$$

where N_u denotes the set of (direct) neighbors (friends) of u

- Trust depends on all connecting paths

Trust between direct neighbors

Can be based on profile similarity or a value provided by the users themselves.

TidalTrust

Predicted Rating

$$\hat{r_{u,i}} = \frac{\sum_{v \in raters} t_{u,v} r_{v,i}}{\sum_{v \in raters} t_{u,v}}$$

$r_{v,i}$ denotes rating of user v for item i

TidalTrust

Predicted Rating

$$\hat{r_{u,i}} = \frac{\sum_{v \in raters} t_{u,v} r_{v,i}}{\sum_{v \in raters} t_{u,v}}$$

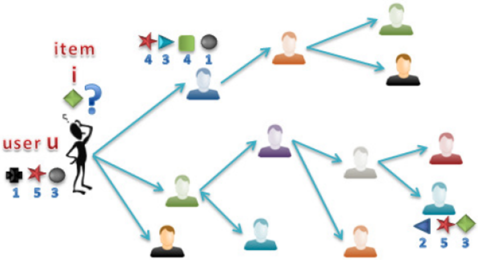
$r_{v,i}$ denotes rating of user v for item i

Shortest distance?

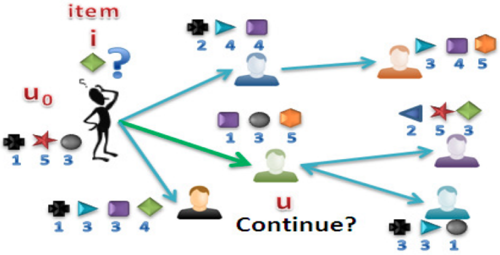
- Efficient
- Taking a short distance gives high precision and low recall
- One can consider raters up to a maximum-depth d, a trade-off between precision (and efficiency) and recall

TrustWalker

- How far to explore the network?: trade-off between precision and coverage
- Instead of far neighbors who have rated the target item, use near neighbors who have rated similar items



Random Walk Starting from a Target User u_0



At step k, at node u

- If u has rated i, return $r_{u,i}$, otherwise
- With probability $\phi_{u,i,k}$, stop random walk, randomly select item j rated by u and return $r_{u,j}$
- With probability $1 - \phi_{u,i,k}$, continue the random walk to a direct neighbor of u

Selecting $\phi_{u,i,k}$

- $\phi_{u,i,k}$ gives the probability of staying at u to select one of its items at step k, while we are looking for a prediction on target item i
- This probability should be related to the similarities of the items rated by u and the target item i, consider the maximum similarity
- The deeper we go into the network, the probability of continuing random walk should decrease, so $\phi_{u,i,k}$ should increase with k

$$\phi_{u,i,k} = max_{j \in RI_u} sim(i,j) \times \frac{1}{1+e^{-\frac{k}{2}}}$$

where RI_u denotes the set of items rated by user u

Selecting $\phi_{u,i,k}$

Selecting $sim(i,j)$

Let $UC_{i,j}$ be the set of common users, who have rated both items i and j, we can define the correlation between items i and j as:

$$corr(i,j) = \frac{\sum_{u \in UC_{i,j}} (r_{u,i} - \overline{r_u})(r_{u,j} - \overline{r_u})}{\sqrt{\sum_{u \in UC_{i,j}} (r_{u,i} - \overline{r_u})^2} \sqrt{\sum_{u \in UC_{i,j}} (r_{u,j} - \overline{r_u})^2}}$$

Selecting $\phi_{u,i,k}$

Selecting $sim(i,j)$

Let $UC_{i,j}$ be the set of common users, who have rated both items i and j, we can define the correlation between items i and j as:

$$corr(i,j) = \frac{\sum_{u \in UC_{i,j}} (r_{u,i} - \overline{r_u})(r_{u,j} - \overline{r_u})}{\sqrt{\sum_{u \in UC_{i,j}} (r_{u,i} - \overline{r_u})^2} \sqrt{\sum_{u \in UC_{i,j}} (r_{u,j} - \overline{r_u})^2}}$$

Taking the effect of common users

The size of the common users is also important. For the same value of $corr(i,j)$, if number of common users, $|UC_{i,j}|$, is higher, the similarity should be higher

$$sim(i,j) = \frac{1}{1 + e^{-\frac{|UC_{i,j}|}{2}}} \times corr(i,j)$$

When does a random walk terminate?

Three alternatives

- Reaching a node which has expressed a rating on the target item i
- At some user node u, decide to stay at the node and select one of the items rated by u and return the rating for that item as result of the random walk
- The random walk might continue forever, so terminate when it is very far ($k > max - depth$). What value of k ?

When does a random walk terminate?

Three alternatives

- Reaching a node which has expressed a rating on the target item i
- At some user node u, decide to stay at the node and select one of the items rated by u and return the rating for that item as result of the random walk
- The random walk might continue forever, so terminate when it is very far ($k > max - depth$). What value of k ?
- "six-degrees of separation"

How to recommend a rating?

Perform several random walks, as described before and the aggregation of all ratings returned by different random walks are considered as the predicted rating $\hat{r_{u_0,i}}$.

How to recommend a rating?

Perform several random walks, as described before and the aggregation of all ratings returned by different random walks are considered as the predicted rating $\hat{r_{u_0,i}}$.

Estimated rating for source user u on target item i:

$$\hat{r_{u_0,i}} = \sum_{\{(v,j)|R_{v,j}} P(XY_{u,i} = (v,j))r_{v,j}$$

- $XY_{u,i}$ is the random variable for stopping the random walk at node v and selecting item j rated by v

Social Matrix Factorization

Intuition

Can we incorporate the Social information in the matrix factorization methods?

Social Matrix Factorization

Intuition

Can we incorporate the Social information in the matrix factorization methods?

Recollect the Matrix factorization problem

$$min_{p^*,q^*} \sum_{(u,i)\in K} (r_{ui} - \hat{r_{ui}})^2 + \lambda(||q_i||^2 + ||p_u||^2)$$

where r_{ui} is the actual rating given by user u to item i, $\hat{r_{ui}}$ approximates user u's rating of item i, simplest of the expression being ${q_i}^T p_u$, though other biases can also be incorporated.

Social Matrix Factorization

Basic Idea

Neighbors in the social network may have similar interests.

Social Matrix Factorization

Basic Idea

Neighbors in the social network may have similar interests.

Incorporating social factors

- Let the social network information be represented by a matrix $S \in R^{u_0 \times u_0}$, where u_0 is the number of users.
- $S_{u,v} \in (0,1]$ denotes the directed and weighted social relationship of user u with use v
- Each of the rows of the social matrix S is normalized to 1, resulting in the new matrix S^*, such that $\sum_v S^*{}_{u,v} = 1$ for each user u

Social Matrix Factorization

Basic Idea

Neighbors in the social network may have similar interests.

Incorporating social factors

- Let the social network information be represented by a matrix $S \in R^{u_0 \times u_0}$, where u_0 is the number of users.
- $S_{u,v} \in (0,1]$ denotes the directed and weighted social relationship of user u with use v
- Each of the rows of the social matrix S is normalized to 1, resulting in the new matrix S^*, such that $\sum_v S^*{}_{u,v} = 1$ for each user u

Modified objective function

$$min_{p^*,q^*} \sum_{(u,i)\in K} (r_{ui} - \hat{r_{ui}})^2 + \beta \sum_{\text{all } u} ((q_u - \sum_v S^*{}_{u,v} q_v)(q_u - \sum_v S^*{}_{u,v} q_v)^T)$$

$$+\lambda(||q_i||^2 + ||p_u||^2)$$

Circle-based Social Recommendation

Basic Idea

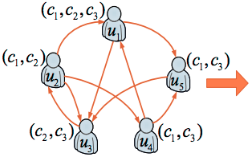
A user may trust different subsets of friends regarding different domains.

Circle-based Social Recommendation

Basic Idea

A user may trust different subsets of friends regarding different domains.

Inferring circles based on categories

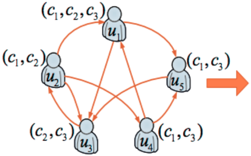


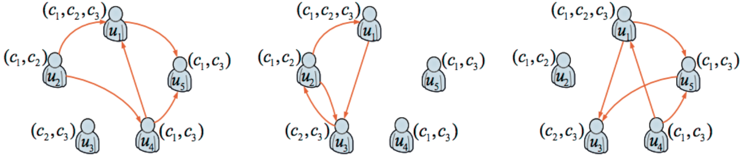
Circle-based Social Recommendation

Basic Idea

A user may trust different subsets of friends regarding different domains.

Inferring circles based on categories





Circle-based Social Recommendation

v is in inferred circle c of u iff u connects to v and both are interested in the category c.

Circle-based Social Recommendation

v is in inferred circle c of u iff u connects to v and both are interested in the category c.

Example Categories

- Videos and DVDs
- Books
- Music
- Toys
- Software
- Cars
- ...

Circle-based Social Recommendation

Using the nomalized trust matrix $S^{(c)*}$, a separate matrix mactorization model is trained for each category c.

Circle-based Social Recommendation

Using the nomalized trust matrix $S^{(c)*}$, a separate matrix mactorization model is trained for each category c.

Modified Objective function

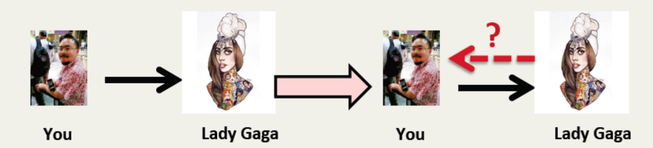
$$L^{(c)}(r^{(c)}, q^{(c)}, p^{(c)}, S^{(c)}) = min_{p^*, q^*} \sum_{(u,i) \in K} ({r^{(c)}}_{ui} - \hat{r_{ui}}^{(c)})^2$$

$$+\beta \sum_{\text{all } u} (({q_u}^{(c)} - \sum_{v} {S^{(c)*}}_{u,v} {q^{(c)}}_{v})({q_u}^{(c)} - \sum_{v} {S^{(c)*}}_{u,v} {q^{(c)}}_{v})^T)$$

$$+\lambda(||{q^{(c)}}_i||^2 + ||{p^{(c)}}_u||^2)$$

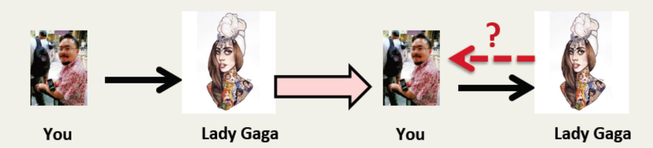
Reciprocity and Triadic Closure

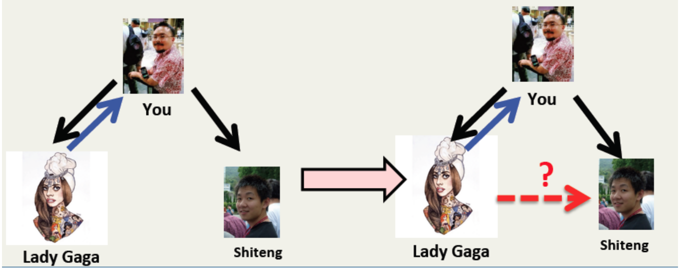
How to tackle the problem of reciprocity prediction?



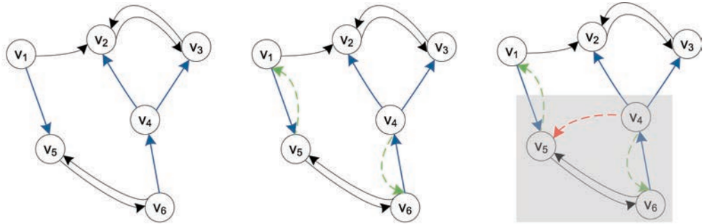
Reciprocity and Triadic Closure

How to tackle the problem of reciprocity prediction?



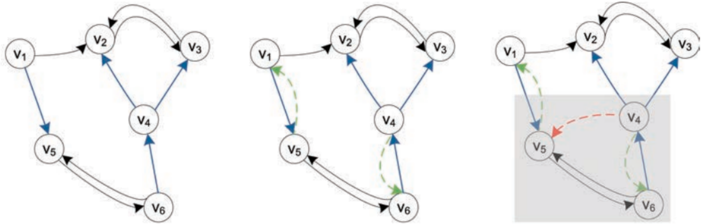


Predicting Reciprocity and Triadic Closure



(a) Following network at time t (b) Follow back at $(t+1)$ (c) Closure triadic at $(t+2)$

Predicting Reciprocity and Triadic Closure



(a) Following network at time t (b) Follow back at $(t+1)$ (c) Closure triadic at $(t+2)$

Reciprocity and Triadic Closure

- (a) A following network, where the blue arrows indicate new following relationship created at time t
- (b) Network with follow-back relations, green dash arrows indicate the follow back relationships developed at time $(t+1)$
- (c) is the network with a closure triad, where a new follow relationship denoted as a red dash arrow is created at time $(t+2)$, forming a directed closure triad.

Interesting Observations

Elite users tend to follow each other

The likelihood of an elite user following back another elite user is nearly 8 times higher than that of two ordinary users and 30 times that of an elite user and an ordinary user.

Interesting Observations

Elite users tend to follow each other

The likelihood of an elite user following back another elite user is nearly 8 times higher than that of two ordinary users and 30 times that of an elite user and an ordinary user.

Effect of location

The number of reciprocal relationships between users within the same time zone is 20 times higher than the number of users from different time zones

Problem Definition

Modeing Twitter Network as a directed graph $G = (V,E)$

- $V = \{v_1, \ldots, v_n\}$ be the set of users
- $E \subseteq V \times V$ be the set of directed links between users
- Each directed link $e_i \in E$ can be written with its two end-users ${v_i}^s$ and ${v_i}^u$ such that ${v_i}^s$ follows ${v_i}^u$
- ${v_i}^s$ is called the follower of ${v_i}^u$ and ${v_i}^u$ is the followee.

Problem Definition

Modeing Twitter Network as a directed graph $G = (V, E)$

- $V = \{v_1, \ldots, v_n\}$ be the set of users
- $E \subseteq V \times V$ be the set of directed links between users
- Each directed link $e_i \in E$ can be written with its two end-users ${v_i}^s$ and ${v_i}^u$ such that ${v_i}^s$ follows ${v_i}^u$
- ${v_i}^s$ is called the follower of ${v_i}^u$ and ${v_i}^u$ is the followee.
- A new link results when a user performs a behavior of following another user in Twitter (95% of the changes to link are adding new links)

New Follow and Follow Back

- Suppose at time t, user v_i creates a link to v_j, who has no previous link to v_i, then v_i performs a **new-follow** behavior on v_j.
- When user v_i creates a link to v_j at time t, who already has a link to v_i at time t, then v_i performs a **follow-back** behavior on v_j
- New-follow behavior corresponds to the one-way (parasocial) relationship
- Follow-back behavior corresponds to two-way (reciprocal) relationship

Follow Back Prediction

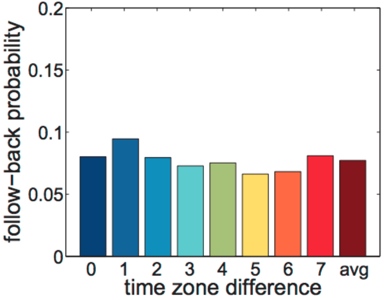
- Let $y_i{}^t = 1$ denote that user $v_i{}^s$ follows back $v_i{}^u$ at time t and
- $y_i{}^t = 0$ denote that user $v_i{}^s$ does not follow back.

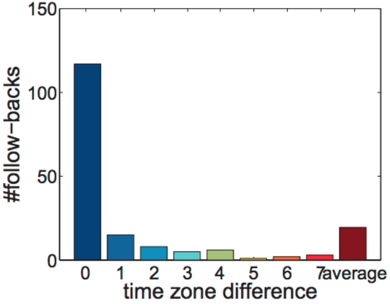
Prediction Problem

- Let $< 1, \ldots, t >$ be a sequence of timesstamps with a particular time granularity (e.g., day, week, etc.)
- Given Twitter networks from time 1 to t, $\{G^t = (V^t, E^t, Y^t)\}$, where Y^t is the set of follow-back behaviors at time t, the task is to find a predictive function

$$f : (\{G^1, \ldots, G^t\}) \rightarrow Y^{(t+1)}$$

Geographic distance correlation





Effect of Homophily

Principle of homophily

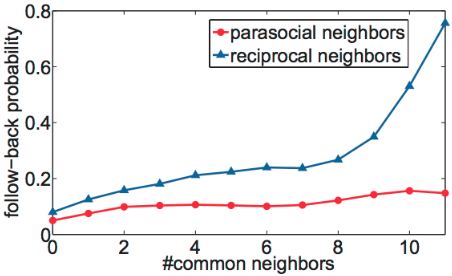
Users with similar characteristics (e.g., social status, age) tend to associate with each other

Effect of Homophily

Principle of homophily

Users with similar characteristics (e.g., social status, age) tend to associate with each other

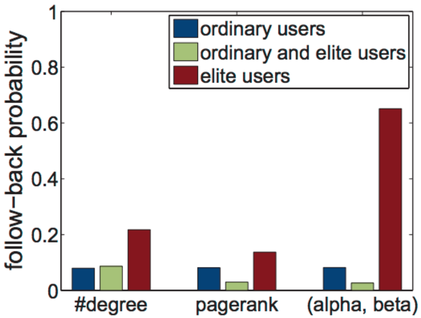
Link Homophily: Common Neighbors



Effect of Homophily: ordinary and elite

Eliteness based on

- Top 1% users having the highest pageRank
- Top 1% users with the highest number of indegree
- (α, β): users selected in the core community C of size 200 (every vertex in C connected to at-least β vertices of C and every vertex outside C is connected to atmost α vertices inside C).



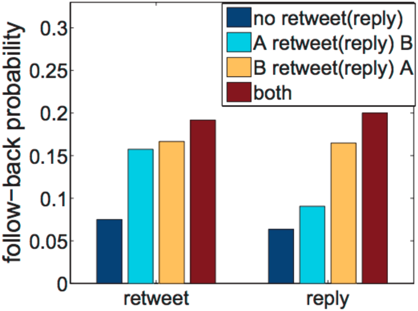
Effect of retweet (reply)

- User A mentioning user B in his tweet, @B, is called a **reply link**
- User A forwarding user B's tweet, is called a **retweet link**

Effect of retweet (reply)

- User A mentioning user B in his tweet, @B, is called a **reply link**
- User A forwarding user B's tweet, is called a **retweet link**

Probability that B follows A back



Structural balance

Structural Balance Property

For every group of three users (triad), balance property implies that

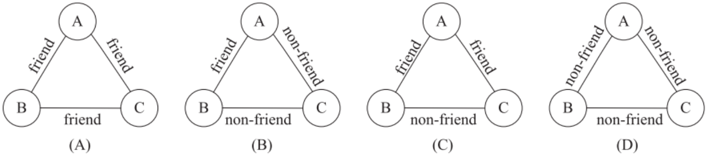
- Either all three of the users are friends, or
- Only one pair of them are friends

Structural balance

Structural Balance Property

For every group of three users (triad), balance property implies that

- Either all three of the users are friends, or
- Only one pair of them are friends

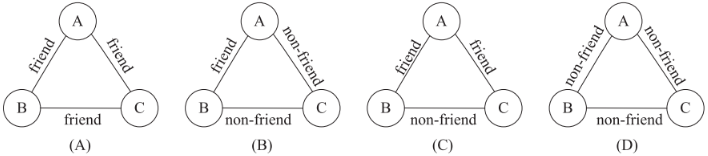


Structural balance

Structural Balance Property

For every group of three users (triad), balance property implies that

- Either all three of the users are friends, or
- Only one pair of them are friends



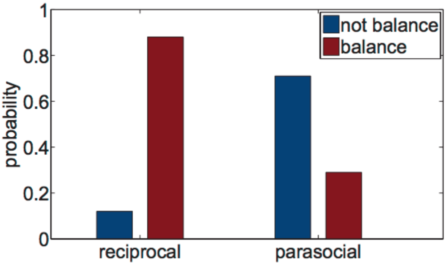
A and B are balanced, while C and D are not.

Verifying Structural balance theory for Twitter

- Either reciprocal or parasocial relations can be mapped on the friendship

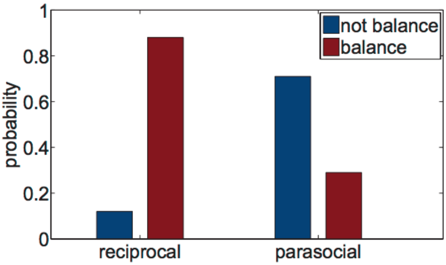
Verifying Structural balance theory for Twitter

- Either reciprocal or parasocial relations can be mapped on the friendship



Verifying Structural balance theory for Twitter

- Either reciprocal or parasocial relations can be mapped on the friendship



Why so unbalanced for parasocial?

Two users may not know each other but may follow a same movie star with a high probability, which results in a unbalanced triad

Model Framework

- For an edge $e_i \in E$, if user ${v_i}^s$ follows ${v_i}^u$ at time t, our task is to predict whether user ${v_i}^u$ will follow ${v_i}^s$ back at time $(t+1)$.

Model Framework

- For an edge $e_i \in E$, if user ${v_i}^s$ follows ${v_i}^u$ at time t, our task is to predict whether user ${v_i}^u$ will follow ${v_i}^s$ back at time $(t+1)$.
- Each edge can be described using various attributes, denoted as x_i
- If d is the number of attributes, the $|E| \times d$ attribute matrix X describes edge-specific characteristics
- **Example attributes:** whether the two end-users are from the same time zone

Triad Factor Graph (TriFG)

- We need the posterior probability of $P(Y|X,G)$
- Using Bayes' theorem, $P(Y|X,G) \propto P(X|Y)P(Y|G)$
- Assuming the generative probability of attributes given the label of each edge is conditionally independent, we get

$$P(Y|X,G) \propto P(Y|G)\prod_i P(x_i|y_i)$$

- $P(x_i|y_i)$ and $P(Y|G)$ can be instantiated using a Markov random field

Instantiating the two probabilities

$$P(\mathbf{x}_i|y_i) = \frac{1}{Z_1}\exp\left\{\sum_{j=1}^{d}\alpha_j f_j(x_{ij}, y_i)\right\}$$

$$P(Y|G) = \frac{1}{Z_2}\exp\left\{\sum_{c}\sum_{k}\mu_k h_k(Y_c)\right\}$$

Instantiating the two probabilities

$$P(\mathbf{x}_i|y_i) = \frac{1}{Z_1}\exp\left\{\sum_{j=1}^{d}\alpha_j f_j(x_{ij}, y_i)\right\}$$
$$P(Y|G) = \frac{1}{Z_2}\exp\left\{\sum_{c}\sum_{k}\mu_k h_k(Y_c)\right\}$$

- $f_j(x_{ij}, y_i)$ is a feature function, defined for each attribute x_{ij} associated with edge e_i
- α_j is the weight of the j^{th} attribute
- A set of correlation feature functions $\{h_k(Y_c)\}_k$ is defined over each triad Y_c in the network
- μ_k is the weight of the k^{th} correlation feature function

Model Learning

Log-likelihood objective function

$$O(\theta) = logP_\theta(Y|X,g)$$

$$\mathcal{O}(\theta) = \sum_{i=1}^{|E|}\sum_{j=1}^{d}\alpha_j f_j(x_{ij}, y_i) + \sum_c \sum_k \mu_k h_k(Y_c) - \log Z$$

Model Learning

Log-likelihood objective function

$$O(\theta) = logP_\theta(Y|X,g)$$

$$\mathcal{O}(\theta) = \sum_{i=1}^{|E|}\sum_{j=1}^{d}\alpha_j f_j(x_{ij}, y_i) + \sum_{c}\sum_{k}\mu_k h_k(Y_c) - \log Z$$

- Learning the TriFG model is to estimate a parameter configuration $\theta = (\{\alpha\},\{\mu\})$ to maximize $O(\theta)$, i.e. $\theta^* = \arg\max O(\theta)$

Model Learning

Log-likelihood objective function

$$O(\theta) = logP_\theta(Y|X,g)$$

$$\mathcal{O}(\theta) = \sum_{i=1}^{|E|}\sum_{j=1}^{d}\alpha_j f_j(x_{ij}, y_i) + \sum_c \sum_k \mu_k h_k(Y_c) - \log Z$$

- Learning the TriFG model is to estimate a parameter configuration $\theta = (\{\alpha\}, \{\mu\})$ to maximize $O(\theta)$, i.e. $\theta^* = \arg\max O(\theta)$
- A gradient descent method can be used to solve this objective function

$$\frac{\mathcal{O}(\theta)}{\mu_k} = \mathbb{E}[h_k(Y_c)] - \mathbb{E}_{P_{\mu_k}(Y_c|\mathbf{X},G)}[h_k(Y_c)]$$