Higher-order Graphical Models for Classification in Social and Affiliation Networks

Elena Zheleva
Computer Science Dept.
University of Maryland, College Park
elena@cs.umd.edu

Lise Getoor
Computer Science Dept.
University of Maryland, College Park
getoor@cs.umd.edu

Sunita Sarawagi
Computer Science and Engineering
IIT Bombay, India
sunita@cse.iitb.ac.in

Abstract

In this work we explore the application of higher-order Markov Random Fields (MRF) to classification in social and affiliation networks. We consider both friendship links and group membership for inferring hidden attributes in a collective inference framework. We explore different ways of using the social groups as either node features or to construct the graphical model structure. The bottleneck in applying higher-order MRFs to a domain with many overlapping large cliques is the complexity of inference which is exponential in the size of the largest clique. To circumvent the slow inference problem, we borrow recent advancements in the computer vision community to achieve fast approximate inference results. We provide preliminary results using a dataset from Facebook which suggest that our higher-order MRF models are capturing the structural dependencies in the networks and they yield higher accuracy than linear classifiers.

1 Introduction

A common assumption in social network analysis is that one can infer a lot about people from their social environment. A useful task in this type of analysis is collective classification where the goal is to infer hidden attributes of the nodes, and the classification algorithm considers not only the local features of each node in the network but also the features and the class labels of its network neighbors [6]. In a social network, where nodes represent actors, the actor-actor links are used to boost the accuracy of local classifiers or even provide classification labels in the absence of local features. While most collective classification algorithms take advantage of the statistical dependencies induced by the actor-actor links, very little work has been on using actor groups of size larger than two.

Online affiliation networks contain information about groups that actors have formed over time. Unlike the clusters resulting from automatic graph clustering of the social network which make certain assumptions about what constututes a cluster, online groups are observed affiliations in which the actors have participated. They provide a clustering of the actors that is completely data driven and perhaps more informative than inferring groups based on actor-actor links. Affiliation networks have been shown to have a strong signal for predicting actor attributes [9].

The goal of our work is to provide a principled approach to classification using the available data in a model which overlays information from the social network and the affiliation network. We investigate the use of higher-order Markov Random Field models which exploit the structure of both

social and affiliation networks to perform better classification. Our contributions include identifying an approach for defining higher-order MRFs based on multi-modal social networks and proposing a model selection method informed by the existing structure in the network.

Relational data, such as social networks, can be modeled as a pairwise Markov Random Field [7]. In particular, each actor's attribute in the social network is a random variable in the MRF, and each actor-actor link is considered as a pairwise dependency between two random variables. Inference on the MRF can be used for classification of the missing attributes in the data. To the best of our knowledge, MRFs which use not only the dependencies coming from the observed friendship links but also from the observed affiliations have not been applied to classification tasks in social networks. Yet, the affiliation network structure provides rich dependencies which go beyond pairwise.

One way of including information from the affiliation network is to introduce a clique for each group. This approach has a number of challenges. First, in online social networks both the number and the size of groups can be very large, and inference on a dense network can be prohibitively slow. Therefore, it becomes extremely important to learn which groups should participate in the MRF structure. Second, many MRFs rely on approximate inference algorithms which have to be tailored to the domain of interest in order to perform well.

Within the computer vision community, there has been a growing body of work on higher-order MRFs [3]. For example, in image analysis, segmentation is an important task in which given pixel information, such as color and location, an algorithm aims to classify each pixel to one of a number of classes, e.g., tree, sky, ground. Rather than taking the picture as a vector of pixels, pairwise MRFs encode structural information by considering the dependencies between neighbouring pixels. This has been shown to improve classification because classes of neighboring pixels are often dependent on each other. However, pairwise MRFs tend to make mistakes on pixels that are on the edges of image segments, e.g., the border pixels separating tree and sky. Incorporating longer-range dependencies between the pixels leads to better solutions. Higher-order MRFs take care of such dependencies by considering overlapping segments from different segmentation algorithms as cliques [3]. Recently, the computer vision community has developed inference algorithms that are extremely efficient and can find optimal solutions for a class of models in polynomial time. We discuss them in Section 4.

2 Preliminaries

Online social networks, such as Facebook, Flickr, LinkedIn, etc, allow individuals to create a profile and link, e.g. "friend" each other, or "affiliate" by joining groups of interest. They include online services which allow users to set their preferences to online content, such as tagging articles, commenting on photos and rating movies, to mention a few.

2.1 Data model and graphical model

We distinguish between two types of graphs: 1) the data graph, which we refer as the network G, and 2) the graph of random variables which represents the graphical model. The social and affiliation network data $G = (\mathbf{V}, \mathbf{E}_v, \mathbf{H}, \mathbf{E}_h)$ consists of n actors $\mathbf{V}$ with attributes $\mathbf{V}.\mathbf{A}$, and two types of commonly occurring links - actor-actor links, $\mathbf{E_v}$, and actor-group links, $\mathbf{E_h}$. The groups can be overlapping and of various sizes. The graphical model consists of a vector of discrete random variables $\mathbf{X} = \{X_1, X_2, ..., X_n\}$ for the actor attribute we aim to classify. Each variable X_i can take on a number of class labels. A Markov Random Field model is an undirected graphical model which represents a family of probability distributions for a random variable vector $\mathbf{X}$ given by

$$Pr(\mathbf{x}) = \frac{1}{Z} exp(-\sum_{c \in C} \phi_c(\mathbf{x}_c))$$

where C is a set of cliques, ϕ_c is the potential function for clique c, and Z is the normalizing constant, known as the partition function. $E(\mathbf{x}) = \sum_{c \in C} \phi_c(\mathbf{x}_c)$ corresponds to the Gibbs energy of a possible variable assignment. Each clique consists of a fully connected set of variables which are statistically dependent. A potential function is a function which assigns a positive real number to each possible variable vector assignment in the clique, and we discuss specific potential functions in Section 3.3. In pairwise MRFs, the clique potentials are over pairs of variables whereas higher-order MRFs can have cliques of arbitrary size.

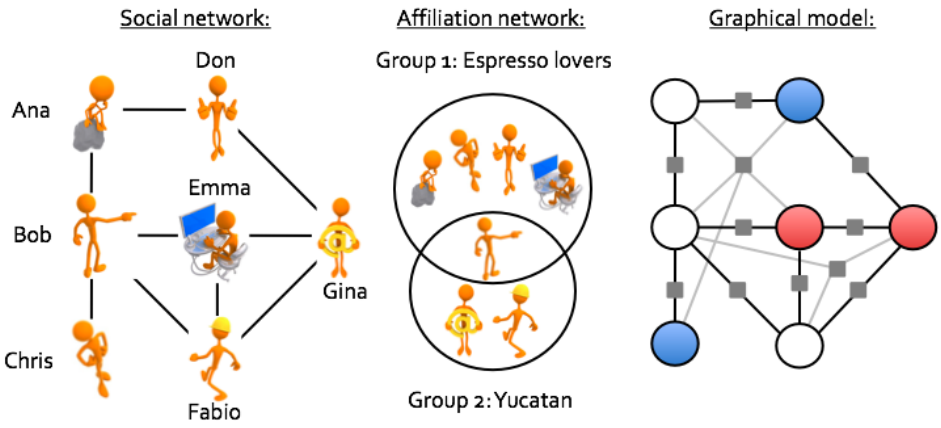

Figure 1: A toy social and affiliation network and its corresponding higher-order MRF.

2.2 Problem description

Given a network G in which the values of attribute a are given for some observed nodes $\mathbf{V}_o$, we would like to find the hidden attribute for the rest of the nodes in the network, $\mathbf{V}_h$. We concentrate on the case where the group memberships and friendship links are given for all nodes, and there are no other node attributes. The incentive for this is to evaluate the worth of the dependencies expressed in the network structure alone.

To make this problem more concrete, we construct the graphical model. First, we partition the random variables into $\mathbf{X}_o$ and $\mathbf{X}_h$, corresponding to the nodes $\mathbf{V}_o$ and $\mathbf{V}_h$. We would like to find the most probable assignment of $\mathbf{X}_h$, given the assignment of $\mathbf{X}_o$. This corresponds to the maximum a posteriori (MAP) estimation of $\mathbf{X}_h$:

$$\hat{\mathbf{X}}_h = \underset{\mathbf{x_h}}{\operatorname{argmax}}\, Pr(\mathbf{X}_h|\mathbf{x}_o) = \underset{\mathbf{x_h}}{\operatorname{argmin}}\, E(\mathbf{x_h}, \mathbf{x}_o)$$

Next we discuss how to construct the cliques in the graphical model.

3 Graph structure and potentials

Our solution first selects the MRF structure as discussed in Section 3.1, then it bootstraps the MRF model by computing informative node potentials described in Section 3.2. Then, using the node and clique potentials (in Section 3.3), it performs efficient MRF inference which we overview in Section 4.

3.1 MRF structure

There are different ways in which one can construct the graphical model by incorporating the structure of the network data. Here, we propose four different constructions. The most naïve way is to include all the friendship links as pairwise dependencies in the MRF. Each link $e_v(v_i, v_j)$ in the data corresponds to a clique of size two, $c(x_i, x_j)$. This incorporates the idea of homophily in social networks, or the tendency of individuals to associate with similar others, by making class labels of friends dependent on each other. This creates the pairwise Markov Random Field model, *pMRF*.

Another possibility is to include the affiliation network by treating each social group $h(\mathbf{V}^*)$ as a clique in the MRF $c(\mathbf{X}^*)$, where the random variables in the clique, $\mathbf{X}^*$, correspond to the group members $h(\mathbf{V}^*)$. This leads to our second model, the higher-order MRF with all groups, *hoMRF-AG*.

While including all groups may be an enticing idea, some of the social groups are more informative about certain actor attributes than others, e.g. women may be more likely to join a social group for

breastfeeding advice than men. Following this idea, we look at group properties and select informative groups, which leads to our third model, higher-order MRF with selected groups (*hoMRF-SG*). We select the set of informative groups in the network based on their observed properties, such as size and entropy of the nodes with observed class labels. Our last model constructs the MRF by using both the pairwise dependencies from the friendship links and the higher-order cliques from selected social groups (*hoMRF-SG-AL*).

Figure 1 shows an example social and affiliation network, together with its corresponding graphical model. The graphical model is presented as a factor graph to make the cliques (grey rectangles) over which the potentials are defined explicit. There are 7 actors with 9 friendship links and 2 social groups. The two social groups correspond to the two cliques of size 3 and 5 in the graphical model. Each link has a potential associated with it as well. The class labels of some of the actors are observed (shaded circles in the graphical model), while the labels of others are unknown (unshaded circles).

3.2 Node potentials

Each node in the MRF is a clique of size one, and it has a unary node potential. For each $X \in \mathbf{X}_h$, we compute the potential for each class value to be the negative log likelihood of the class value according to a linear classifier. The classifier, such as logistic regression or Naïve Bayes, uses the friendship links and group memberships as node features. Besides computing the node potentials, this classifier provides the baseline method in our experiments.

3.3 Clique potentials

Possible potentials for cliques of size larger than one are functions of the counts of class labels, such as majority and sum, negative/reciprocal entropy. We adopt the Robust P^n Potts clique potential of Kohli et al.[3] because it is intuitive and it allows efficient inference. This potential is defined as:

$$\phi_c(\mathbf{x}_c) = \begin{cases} \gamma_k + \frac{N_i(\mathbf{x}_c)}{Q}(\gamma_{max} - \gamma_k) & \text{if } N_i(\mathbf{x}_c) \leq Q \\ \gamma_{max} & \text{else} \end{cases}$$

where γ_k is the minimum possible potential value if all labels in the clique are the same, and γ_{max} is the maximum possible potential value if the number of node labels that are different from the majority class label, $N_i(\mathbf{x}_c)$, is larger than a pre-specified threshold, called truncation ratio Q. For pairwise MRFs, this potential simplifies to the Potts potential. The intuition behind the Robust P^n Potts potential is that it allows disagreement between class labels inside each clique to a certain extent. Besides being intuitive, this potential is important for efficient inference using graph-cut based methods which we discuss next.

4 Inference and energy minimization

Exact inference in higher-order MRF models is exponential in the size of the largest clique. There are a number of approximate inference algorithms, e.g., belief propagation, variational inference, MCMC-based techniques, which aim to alleviate the complexity burden [2]. In the computer vision community, graph cut based methods have gained popularity because they have a polynomial complexity when assuming certain potential functions, such as Robust P^n Potts potential, and they work efficiently in practice [3]. Kohli et al. [3] compare the running time and accuracy of tree-reweighted message passing (TRW-S) [4] with move-making inference algorithms which use graph cuts for models with large cliques. They find that the move-making algorithms are faster and yield better solutions.

A move-making algorithm starts from an initial solution and it makes a series of moves leading to lower energy solutions. At each step, it searches for the best possible move within its allowed range and then makes that move. The algorithm converges when it reaches a state from which it cannot find a lower energy solution.

Two move-making inference algorithms are α-expansion and α-β swap [1]. A move can be encoded as a vector of binary variables $\mathbf{t}$, one for each unobserved random variable in the hoMRF, $X_i \in \mathbf{X_h}$.

In an α-expansion move, each random variable X_i either retains its current label if $t_i = 1$, or changes it to α if $t_i = 0$. In an α-β swap move, each random variable $X_i \in \mathbf{X_h}$ with a current label of α or β can either retain/change to a label of α if $t_i = 0$, or retain/change to a label of β if $t_i = 1$. One iteration of the algorithm searches through the space of possible move vectors to find the one that would lead to the lowest energy solution and then it makes that move.

Finding the optimal move vector for both the expansion and swap algorithms can be computed in polynomial time using graph cuts. For details, we refer the reader to Kohli et al. [3]. According to the same authors, the best ordering of moves is an ongoing research topic.

5 Experiments

5.1 Data description

For our evaluation, we studied a dataset from the social network Facebook[1], available for research purposes [5]. Facebook allows users to communicate with each other, to form undirected friendship links and participate in groups and events. The dataset contains all $1,225$ profiles of first-year students in a small college who share at least one interest group with another first-year student according to their Facebook profiles. The interest groups are the favorite books, music and movies of the users. There are $2,932$ groups, and the largest one has 290 members. There are $51,389$ friendship links in the data. The attribute we are trying to predict is the gender of each student. Half of the students are female, so a random guess would achieve an accuracy of 50%.

5.2 Experimental setup

We provide results for two-fold cross validation. The node potentials are computed using the java version of the liblinear logistic regression classifier [8]. For the move-making inference, we adapt the implementation of Kohli et al. [3]. For selecting the groups to be included as cliques in the MRF, we vary the allowed size, entropy and percent of observed nodes per group. First we performed a coarse-grained search through the space of parameters by setting the minimum group size to $\{2, 4, 6, 10\}$, maximim group size to $\{10, 50, 290\}$, maximum entropy of the nodes with observed class labels to $\{0, 0.5, 0.7, 0.9, 1\}$, and the minimum percentage of nodes with observed class labels in the group to $\{0, 0.5\}$. This yields a space of 120 experiment points, e.g. point $(10, 290, 0, 0.5)$ means all groups of size between 10 and 290 with entropy of 0 and at least 50% of node labels observed. To obtain further improvement, we performed a fine-grained search around the parameters that yielded the best accuracy in the coarse-grained search.

We set γ_{max} to 10, γ_k to 0 for all possible labels, and the truncation ratio Q to 0.3 after some limited exploration of the parameter space. We set the node potentials to the negative log probabilities of the class labels coming out of the linear classifier. In the case of probability of 0, we set it to 10 (which is close to the negative log of the smoothed out probability). We report on three types of node features: friendship link vector, group membership vector and a vector which includes both. We compare the results for the linear classifier (*LR*), the pairwise MRF (*pMRF*) and the variants of the higher-order MRF: *hoMRF-AG*, *hoMRF-SG* and *hoMRF-SG-AL*.

5.3 Results

Table 1 summarizes the results from our experiments. The baseline linear classifier which uses the friendship link vector as features to classify nodes yielded an accuracy of 64.06%. Using the group memberships, this accuracy increases to 71.67%. Using both types of features, the accuracy is the highest, 75.75%. Our observations on the comparison between the linear classifier and our proposed models can be summarized as follows:

1) Using all groups naïvely as the cliques in the hoMRF (*hoMRF-AG*) improves performance only when the node potentials are bootstrapped with the friendship links as features alone. In the other two cases, where the node potentials use the group memberships as features, *hoMRF-AG* is not able to exploit the affiliation network structure further and it even hurts performance.

[1] At `http://www.facebook.com`.

Table 1: Accuracy of the logistic regression baseline (LR), the pairwise MRF (*pMRF*), the higher-order MRF with all groups as cliques (*hoMRF-AG*), with selected groups as cliques (*hoMRF-SG*) and with selected groups and all friendship links as cliques (*hoMRF-SG-AL*).

FEATURES	LR	PMRF	HOMRF-AG	HOMRF-SG	HOMRF-SG-AL
Friendship links	64.06%	64.31%	69.13%	69.22%	69.22%
Group memberships	71.67%	71.83%	69.80%	74.12%	74.53%
Both	75.75%	75.84%	69.63%	77.39%	78.37%

2) *pMRF* improves accuracy only marginally $(0.09 - 0.25\%)$ compared to the baseline.

3) Adding selected groups as cliques in the MRF increases the prediction accuracy in all cases $(1.64 - 5.16\%)$. Moreover *hoMRF-SG* consistently outperforms *LR* for the different folds of the cross validation. This means that the higher-order MRF is able to exploit the affiliation structure twice, once as features in the node potentials, and a second time by using informative groups as cliques. We report on the group selection experiment with the highest average accuracy in the *hoMRF-SG* column of Table 1.

4) *hoMRF-SG-AL* which adds the friendship links as pairwise cliques to *hoMRF-SG* did not improve the accuracy of *hoMRF-SG* when using friendship links as features. However, when the node potentials were using group memberships or both types of features, the accuracy of *hoMRF-SG-AL* is higher than *hoMRF-SG*, by 0.41% and 0.98%, respectively.

The common theme in the parameter values for the best performing *hoMRF-SG* is that the selection criteria based on entropy are irrelevant to accuracy and that very small groups are uninformative. More concretely, for the friendship link features, the experiment points of group selection which yielded the highest accuracy were $(5, \{30, 40, 50\}, any, 0)$. In the strictest case (one with smallest number of groups as cliques), $(5, 30, 0, 0)$, with only 290 cliques out of the $2,932$ groups, it is possible to achieve 5.16% improvement from the baseline. Similarly, the experiment points with the highest accuracy (74.12%) for the group membership features is $(6, \{40, 50\}, any, 0)$. In its strictest case, $(6, 40, 0, 0)$, this includes an average of 201 cliques out of the $2,932$ groups. Lastly, the highest accuracy using both types of features in the *hoMRF-SG* was at experiment points $(8, 30, any, 0)$. In its strictest case, $(8, 30, 0, 0)$, this includes an average of only 100 cliques. Learning the best parameters from data is left for future work.

We also experimented with setting weights for the clique potentials based on the feature weights of the linear classifier since the node features and the graphical model cliques have a one-to-one correspondence. However, this did not provide any increase in accuracy.

The approximate inference in the hoMRF is very fast, and it takes less than 2 seconds to run on our dataset using a machine with 3.2 GHz processor and 3 Gb of RAM.

6 Conclusion

This is a preliminary study on the application of higher-order MRFs to classification in social and affiliation networks. We used recent advances in the computer vision community to ensure fast and accurate approximate inference results. In this study, we were relying on the given, noisy structure of the data to find the graphical model structure using feature selection criteria based on the group properties. In our future work, we would like to explore principled structure learning algorithms which incorporate the knowledge of the existing structure in the network data in various ways. In addition, we would like to apply this method to other real-world and synthetic datasets, in order to understand its properties better.

Acknowledgments

This work was supported in part by NSF Grant #0746930.

References

[1] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23:1222–1239, November 2001.

[2] M. I. Jordan and Y. Weiss. Graphical models: Probabilistic inference. In M. Arbib, editor, *The Handbook of Brain Theory and Neural Networks*, pages 490–496. MIT Press, 2 edition, 2002.

[3] P. Kohli, L. Ladicky, and P. Torr. Robust higher order potentials for enforcing label consistency. *IJCV*, 82:302–324, May 2009.

[4] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583, October 2006.

[5] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis. Tastes, ties, and time. *hdl:1902.1/11827*, 2008.

[6] P. Sen, G. M. Namata, M. Bilgic, L. Getoor, B. Gallagher, and T. Eliassi-Rad. Collective classification in network data. *AI Magazine*, 29(3):93–106, 2008.

[7] B. Taskar, P. Abbeel, and D. Koller. Discriminative probabilistic models for relational data. In *UAI*, 2002.

[8] G.-X. Yuan, K.-W. Chang, C.-J. Hsieh, and C.-J. Lin. A comparison of optimization methods and software for large-scale l1-regularized linear classification. *JMLR*, To appear.

[9] E. Zheleva and L. Getoor. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In *WWW*, 2009.