IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 6, 2018, accepted January 8, 2019, date of publication January 14, 2019, date of current version March 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2892757

Second-Order Markov Assumption Based Bayes Classifier for Networked Data With Heterophily

SA DONG[1,2], DAYOU LIU[1,2], RUOCHUAN OUYANG[3], YUNGANG ZHU[1,2], LINA LI[1,2], TINGTING LI[1], AND JIE LIU[1,2]

[1]College of Computer Science and Technology, Jilin University, Changchun 130012, China
[2]Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun 130012, China
[3]Big Data and Network Management Center, Jilin University, Changchun 130012, China

Corresponding author: Jie Liu (liu_jie@jlu.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 61502198, Grant 61572226, and Grant 61472161.

ABSTRACT The classification of networked data is an interesting and challenging problem. Most traditional relational classifiers that are based on the principle of homophily have an unsatisfactory classification performance in networks with heterophily. This is because these methods treat inhomogeneous networks homogeneously. A progression of a network-only Bayes-classifier-based second-order Markov assumption is proposed for heterophilous networks in this paper to address this problem. First, we estimate the class distribution of an unlabeled node according to the class distribution of its neighbors' neighbors. In this process, we perform this computation on the known and unknown neighbors separately. Second, we combine the two parts using multinomial naïve Bayesian classification. Meanwhile, we pair a relaxation labeling collective inference method (which imports simulated annealing) with this new method to update the class distributions at each iteration. Comparisons of the experimental results demonstrate that the proposed method performs better when the networks are heterophilous.

INDEX TERMS Artificial intelligence, data mining, heterophilous networks, machine learning, networked data classification, relational classifier.

I. INTRODUCTION

Networked data are exploited to model entities that are interconnected such that nodes represent entities with local attributes, labels represent topics or classes of nodes, and edges represent the connections between them. They are different from conventional data which are independent and identically distributed. Networked data present complications and the potential relationships between entities can be used to help with the classification. The main task and major challenge in analyzing networked data is the classification problem, which is to find the best matching labels of unlabeled nodes according to the network and the categories of labeled nodes [1]. The effectiveness of classification depends on the distinct datasets and classifiers. Many network classifiers are based on the principle of homophily [2]–[4], [15]–[18]. Homophily assumes that similar nodes (nodes with the same labels) tend to be connected or that interconnected nodes are more likely to possess the same labels [5], [6]. This phenomenon has been revealed in many social networks [7], [8]. Homophily-based methods predict the class of an unlabeled node using the classes of its direct neighbor nodes. Direct neighbor is also called immediate neighbor which directly links with the unlabeled node through an edge. Therefore, they can achieve high accuracy when the homophily degree of the networked data is high. However, when the labels of most interconnected nodes are diverse, in other words, the homophily degree of the networked data is low, and most of the previous methods face a decline in performance. In this case, the networked data are heterophilous [9].

A. RELATED WORK

A large set of relational approaches for network classification [22]–[26] has been developed in recent years. There are some typical relational classifiers take advantage of the direct neighbors of unlabeled nodes to classify. Macskassy and Provost [10] proposed a simple classifier, they termed a weighted-vote relational neighbor classifier (WVRN) that computes the class probabilities of unlabeled nodes as the weighted mean of the class probabilities of the direct neighbor nodes. WVRN performs relational classification in an

iterative manner without a training process. Their experimental results demonstrated that WVRN performed quite well on homophilous networks, and the structure of the network itself contained much more information, which is helpful for network classification; WVRN should be regarded as a baseline for networked data classifiers [1]. Following Perlich and Provost [11], [12], and based on Rocchio's method [13], Macskassy and Provost [1] defined a class-distribution relational neighbor classifier (CDRN) which estimates the probable label probabilities according to the normalized vector similarity between an unlabeled node's class vector and corresponding label's reference vector . CDRN is more flexible and may get better discrimination on the homophilous networks than WVRN but it is more complex. Chakrabarti *et al*. [14] applied a naïve Bayes model to deal with local attributes and the neighbor nodes of the unlabeled node. Based on this algorithm, Macskassy and Provost [1] ignored the node's local attributes and the direction of links. Therefore, they called it a network-only Bayes classifier (NBC). The experimental results showed that NBC was almost always worse than WVRN and CDRN. Lu and Getoor [15] used logistic regression on the aggregations of local attributes and the neighbor labels respectively, from which Macskassy and Provost [1] derived their network-only link-based classification (NLB). NLB did not consider the local attributes. NLB's count aggregation got best performance among all the aggregations and it is the same as the computation of CDRN's class vector.

These methods classify an unlabeled node by relying on the immediate neighbors' labels of this node in the network. We can also state that these methods are homophily-based and make the first-order Markov assumption [2]–[4], [15]. That means the neighbor nodes set comprises only the direct neighbors of the unlabeled node in the network. They can achieve an accurate classification when the networked data are highly homogeneous. However, heterophily is ubiquitous, occurring in such situations as web pages that are linked via hyperlinks. In this situation, most interlinked nodes do not have the same classes or topics and the above methods cannot classify correctly.

To deal with the classification of networked data with heterophily, Wang *et al.* [9] presented a classification algorithm based on a class propagating distribution (CPD). CPD utilizes an adjacency matrix to compute the influence of the neighbor nodes in an iterative manner like MultiRankWalk (MRW) [16]. CPD performs better on heterophilous networks, but it needs more storage space and time. Dong *et al.* [19] improved CDRN by computing the propagating class vector and propagating reference vector separately, and then comparing the similarities between the two (PCDRN). PCDRN also shows better performance when the networks are of heterophily and needs less storage space than CPD. Dong *et al.* [20] also proposed a relational logistic regression classifier (UNLB) based on the second-order Markov assumption for heterophilous networked data classification. UNLB is a generalized linear regression which is more efficient. Gupta *et al.* [21] proposed a novel meta-path based framework, HeteClass, for transductive classification. HeteClass can incorporate the knowledge of a domain expert and be applied to heterogeneous networks. Experimental results show that these methods perform better when networked data are heterophilous.

As previously discussed, networked data contain interconnected nodes which obscure interdependencies among them. This means that the estimate of the label of one node may influence the estimate of another node which is connected to it and vice versa. Therefore, the nodes in networks should be simultaneously inferred. Collective inference is applied under this circumstance, which infers the interrelated nodes at the same time. Several research studies indicate that collective inference shows advantages for sparsely labeled network classification [27]. Combining a relational classifier with the collective inference method can achieve more reasonable classification results. Collective inference methods update the label estimations of unlabeled nodes continuously until they satisfy the convergence condition or maximum iteration number. Typical collective inference methods include Gibbs sampling (GS) [28], relaxation labeling (RL) [14], and iterative classification (IC) [15] etc.

B. CONTRIBUTIONS

Given that homophily-based classifiers can easily lead to a complete misclassification of heterophilous networks, our contribution is to propose an improved network-only Bayes classifier which can achieve a better classification performance of networked data with heterophily. It is based on the second-order Markov assumption. Specifically, it estimates the class distributions of unlabeled nodes using the label estimations of the neighbors' neighbors of this node which are also called radius-two neighbors. In addition, we propose the concept of radius-two heterophily degree, which describes the heterophily level of networks. The experiments show that the proposed network classifier performs better on networks with heterophily.

C. ORGANIZATION

The rest of this paper is organized as follows. Section II describes the proposed method in detail. Section III covers the experimental setting and data used and subsequently discusses the results. Section IV summarizes the conclusions.

II. METHODS

Compared with the networked data of homophily, most of the interlinked nodes in a heterophilous network have different labels. Therefore, the traditional relational classifiers cannot work reliably in heterophilous networks. For instance, homophily-based method WVRN assumes that neighbor nodes' labels might be the same. An elaborate adaption of NBC is presented in this paper for the classification of networked data with heterophily.

As mentioned earlier, links or edges between nodes in the networked data pose new problems not addressed in traditional classification. Links contain high-quality

semantic clues; however, it is difficult to exploit information among them, because of noise. It is impossible to explore all the neighborhoods of unlabeled nodes through links in a network. To enhance tractability, relational learning often makes a Markov assumption. Most of the relational classifiers are based on a first-order Markov assumption [2]–[4], [15]. However, the classification information in the direct neighbors of unlabeled nodes is limited for heterophilous network classification. For the purposes of obtaining more useful information and avoiding too much noise or incorrect signals, we start with an obvious idea: making a second-order Markov assumption. This means that we use the neighbor nodes of unlabeled nodes' neighbors also called radius-two neighbors to perform the classification.

Chakrabarti *et al.* [14] demonstrate that simply using neighbor node's local attributes could degrade classification accuracy; therefore, only the labels of neighboring nodes are crucial for classification. Furthermore, if we use the local attributes of nodes, the dimensionality of the feature may be overly large compared to the training set sizes. In the proposed method, we only use the edges in the network and the class labels of the nodes to classify. Without local attributes, this setting needs only a tiny amount of storage space and is very fast. This type of simplification is practical in data gathering, processing, and storage [1], [14].

Networked data can be described by a graph which is defined by the nodes and links in network. Considering websites as an example, web pages are nodes, topics are labels, and hyperlinks among the web pages are edges of the graph. Given a network G, v_i is any node in the graph, x_i denotes some (estimated) label of node v_i for m classes, where $x_i \in \{c_1, \ldots, c_m\}$, c is a non-specified label. We use N_i to represent the set of immediate neighbors of v_i in the graph that nodes immediate connect with v_i using an edge. N_i^K represents the set of immediate neighbors of v_i whose labels are known, and N_i^U represents the set of immediate neighbors of v_i whose labels are not known, N_i^U and N_i^K are two disjoint sets, so $N_i = \{N_i^U, N_i^K\}$. w_{ij} denotes the edge weight between node v_i and v_j, because we ignore the directionality of the edge which is different from NBC, so $w_{ij} = w_{ji}$.

When we are training a classifier, we input not only the set of nodes, but also the graph and the labels of known nodes. When the classifier classifies a new node, it has as input not only the node, but also some neighbor nodes of that node. Given a single node v_i in a graph, v_j represents any immediate neighbor of v_i, and v_k represents any immediate neighbor of v_j (i.e., $v_j \in N_i$ and $v_k \in N_j$), so v_k is the radius-two neighbor of v_i. For the second-order Markov assumption, we use $P(x_i = c|\ N_j)$ to represent the class distribution of v_i, where $v_j \in N_i$. The class distribution is used to describe the probabilities that values of v_i belong to each class. When we estimate the class distribution of v_i, the relational classifier and the collective inference method must be able to use the class distribution of v_k (i.e., $P(x_k = c|\ N_l)$, $v_l \in N_k$). The proposed method, called UNBC, uses a multinomial naïve Bayesian classification like NBC. We chose c to maximize $P(x_i = c|N_j)$ based on the labels of v_j's neighbors which is radius-two neighbors set of v_i as in (1).

$$P(x_i = c|N_j) = \frac{P(c) \cdot P(N_j|c)}{P(N_j)} \tag{1}$$

where $v_j \in N_i$, $P(c)$ is the frequency of label c in the training datasets. Given that $P(N_j)$ is not a function of c, normally we do not need to compute it explicitly. Obtaining a known function for $P(N_j|c)$ is extremely unlikely, so we need to estimate it. In order to simplify the computation and estimation we make independence assumptions like NBC. This kind of approximation is not correct but practical, and the validity can be judged by the classification accuracy [14]. Under the naïve Bayes approximation, the neighbor nodes' labels of any node are independent. Since we use second-order Markov assumption, we make use of the radius-two neighbor sets N_j^K and N_j^U of v_i. In the realistic setting, some neighbor nodes are unknown, and we do not wish to ignore all these unknown neighbors. In those cases, we first use neighbors' priori to classify the nodes. The known and unknown neighbors need to be computed in different formulas, so we deal with them separately. As illustrated above, $N_j = \{N_j^K, N_j^U\}$, and we rewrite equation (1) as (2):

$$P(x_i = c|N_j) = P(c) \cdot P(N_j^K|c) \cdot P(N_j^U|c) \tag{2}$$

For all the known labels of all the radius-two neighbor nodes in N_j, we have the independence assumptions of all the neighbor labels. That is:

$$P(N_j^K|c) = \prod_{v_j \in N_i, v_k \in N_j^K} P(x_k = \tilde{x}_k|x_i = c)^{w_{jk}} \tag{3}$$

where $\tilde{x}_k$ is the label observed at node x_k.

For the unknown labels of N_j, according to the total probability formula, $P(N_j^U|c)$ can be expressed as:

$$P(N_j^U|c) = \prod_{v_j \in N_i, v_k \in N_j^U, v_l \in N_k} \{\sum_{h=1}^{m} [P(x_k = c_h|N_l) \cdot P(x_k = c_h|x_i = c)^{w_{jk}}]\} \tag{4}$$

where $P(x_k = c_h|N_l)$ represents the current probability estimations for v_k, and based on the labeled nodes, $P(x_k = c_h|x_i = c)$ and $P(x_k = \tilde{x}_k|x_i = c)$ in equation (3) can be computed during training. To summarize, equation (2) can be obtained from equations (3) and (4).

As we discussed above, the collective inference method is able to infer a set of labels for unlabeled nodes simultaneously. We use relaxation labeling with simulated annealing for simultaneous inferring which is different from NBC's relaxation labeling. This technique has been applied to the computer vision and image processing fields [29], [30]. Rather than assigning each unlabeled node a fixed label as iterative classification does, relaxation labeling retains the ''current'' uncertainty of the nodes. The class distribution at step $t + 1$ will be updated based on the label estimations

TABLE 1. Pseudo-code for algorithm UNBC.

Input: Graph G, training data set, testing data set
Output: Final class distributions for all unlabeled nodes
1. Initialized the prior for each unlabeled node v_i using known nodes in training data
$\boldsymbol{e} \leftarrow prior(v_i)$ $\Delta \boldsymbol{e}$[1…m] is a vector of estimates of class distribution $P(x_i = c| N_j)$, m is the number of labels, v_j is the direct neighbor of v_i. $\boldsymbol{e}$ [k] is the kth value in $\boldsymbol{e}$, which represents $P(x_i = c_k| N_j)$
2. Induce a relational classification model using training data
for $v_k \in N_j$ Δ v_k is the radius-two neighbor of v_i
(a) if $v_k \in N_j^K$ then
for a←0 to m do ΔFor m labels
$\boldsymbol{e}$[a]←$\boldsymbol{e}$[a]* count[h][a]$^{w_{jk}}$
Δcount[1…m][1…m] is a two dimensional array, and count[j][i] represents how many time is label j a radius-two neighbor of label i; h represents x_k 's index; count[h][a] represents $P(x_k = \tilde{x}_k \mid x_i = c)$ in equation (3)
(b) if $v_k \in N_j^U$ then
for a←0 to m do
for b←0 to m do
sum←sum+d[b]* count[b][a]$^{w_{jk}}$
Δd[1…m] is v_k's current probability estimations, which is $P(x_k = c_h \mid N_l)$, and v_l is the radius-two neighbor of v_j, count[b][a] is $P(x_k = c_h \mid x_i = c)$ in equation (4)
$\boldsymbol{e}$[a]←$\boldsymbol{e}$[a]*sum
3. Apply collective inferencing as equation (5) using the relational classification model result in step 2
$\boldsymbol{e}^{(t+1)} \leftarrow \beta^{(t+1)} * \boldsymbol{e}^{(t)} + (1-\beta^{(t+1)}) * \boldsymbol{e}^{(t+1)}$
Δt is the iteration count. Repeat step 2 and 3 for 99 iterations. $\boldsymbol{e}$ will comprise the final class distributions.

from step t. To guarantee convergence, we use the improved relaxation labeling of Macskassy and Provost [1]:

$$P(x_i = c|N_j)^{(t+1)} = \beta^{(t+1)} \cdot P(x_i = c|N_j)^{(t)} + (1 - \beta^{(t+1)}) \cdot P(x_i = c|N_j)^{(t+1)} \quad (5)$$

where $\beta^0 = k$, $\beta^{(t+1)} = \beta^{(t)} \cdot \alpha$, in which k is set to 1, and α is a decay constant, which is set to 0.99. The algorithm is given in TABLE 1.

III. EXPERIMENTS AND RESULTS

Experimental analysis is included in this section to present the resulting improvements.

A. EXPERIMENT SETUP

We vary the percentage r of labeled nodes initially in the network from 10% to 90%. Based on this percentage, the training data set is selected by choosing a class-stratified random sample of the nodes in the network. The rest of the nodes are applied as the testing data set that needs to be labeled according to classification methods. We follow the standard class-stratified 10-fold cross-validation. That is, for a given data set and label ratio r, each experiment for the classification consists of 10 random train/test splits. We keep the partitions of training and testing disjoint as much as possible, because there is dependence in network which is different from the traditional data. Training and Testing are also conducted 10 times at each r. In addition, the accuracy is averaged 10 times. We remove any disconnected nodes in the network. Moreover, we apply Laplace smoothing [1] to avoid computing possible zeros for training.

During training, the nodes whose labels are not known are ignored. The classifier develops classification models for each label, and this process is also called learning. For testing, the labeled nodes can serve as background information and the unlabeled nodes are assigned class distributions, depending upon the classifier.

TABLE 2. The information about the experimental data.

Network	Number of nodes	Number of edges
Cora	4240	22516
Imdb	1441	20317
Cornell	351	1393
Texas	338	1002
Washington	434	1941
Wisconsin	354	1155

B. DATA

We experimented with 6 real networks from 3 domains to examine the performance of the proposed UNBC methods. The datasets are Cora [15], [31], [32], Imdb [10], [33], [34], and four computer science departments' web pages in the WebKB project [15], [34]–[36]. The WebKB datasets include Texas, Cornell, Wisconsin, and Washington. Each page of the four universities is labeled by one of the labels from "course, department, faculty, project, staff and student"; hyperlinks are the edges between them. Cora is comprised of computer science research papers and the citation relationships between them. The labels in Cora are the paper's topics. In Imdb, the nodes are movies and the links are whether they share a production company. We will predict whether the opening weekend box-office receipts will exceed $2 million. The data is listed in TABLE 2.

According to Sen *et al.* [36], the homophily degree of a network can be expressed by the mean percentage of the same labels in a node's direct neighbors. We use the following formula to calculate the radius-two heterophily degree (*HED*):

$$HED = \frac{\sum_{v_j \in N_i} (|D_j| / |N_j|)}{n} \quad (6)$$

where $|D_j|$ denotes the number of nodes in radius-two neighbors of v_i and have labels different to v_i, $D_j \subseteq N_j$, while n denotes the number of nodes in the network. The homophily degree and the radius-two heterophily degree of the networks in TABLE 2 are calculated and listed in TABLE 3.

TABLE 3 shows that the homophily degrees of Cora and Imdb are very high. In contrast, the homophily degrees of the four networks of WebKB are low, so Texas, Cornell,

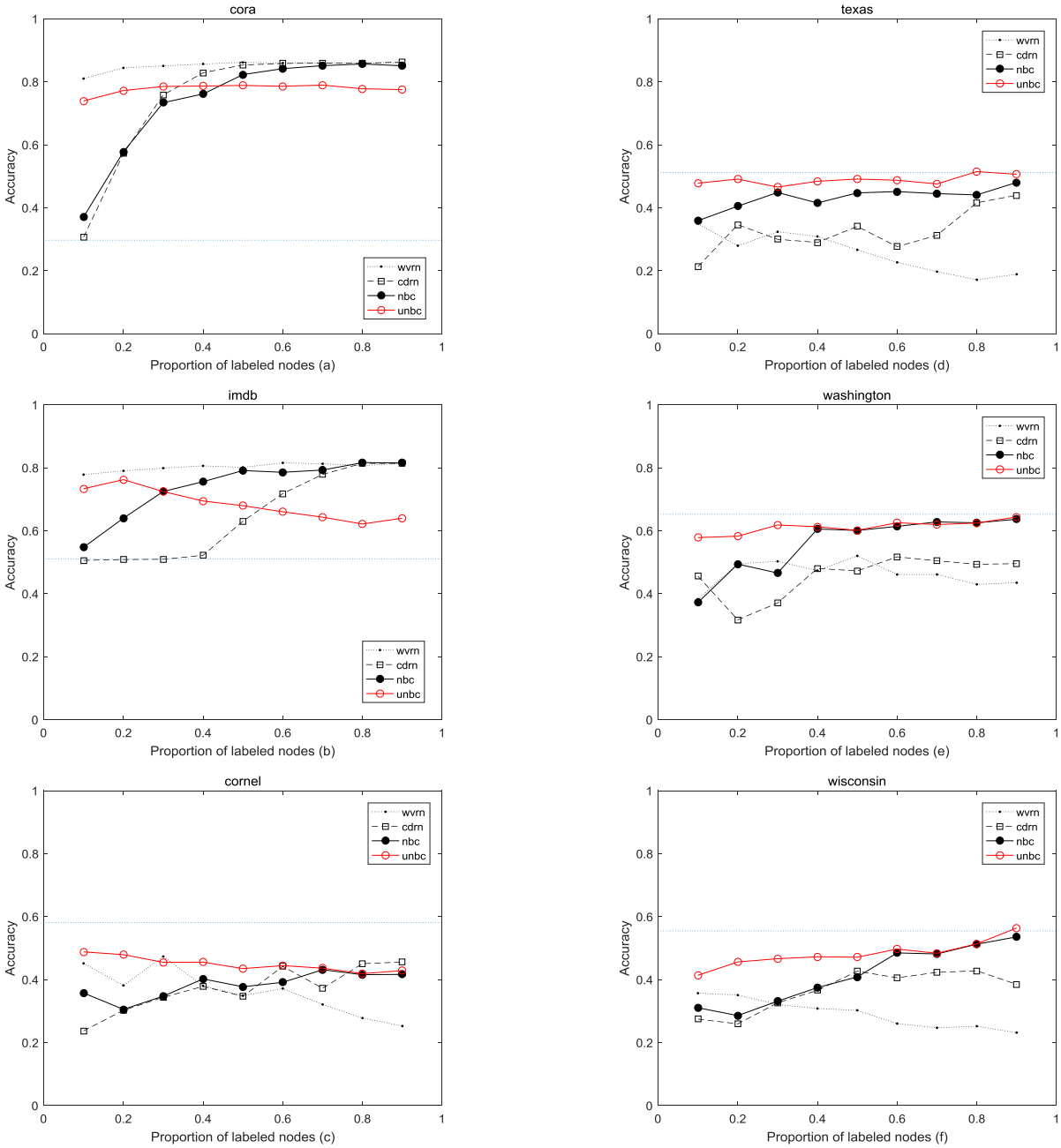


FIGURE 1. Classification accuracies of homophily-based methods on the six datasets. The horizontal axis plots the proportion of a network's nodes for which the label is known. Datasets are tagged above in each graph. The four methods are WVRN, CDRN, NBC, and UNBC. The horizontal line is the most prevalent class rate.

TABLE 3. The homophily degree and radius-two heterophily degree of the networks in TABLE 2.

Network	Homophily degree	Radius-two heterophily degree
Cora	0.805703	0.254091
Imdb	0.738337	0.292562
Cornell	0.265515	0.457993
Texas	0.165911	0.472201
Washington	0.409598	0.390425
Wisconsin	0.231258	0.477433

Wisconsin, and Washington are heterophilous networks. Meanwhile their radius-two heterophily degrees are relatively higher than the above two sets.

C. EXPERIMENTAL RESULTS

The proposed method is not only compared to homophily-based relational classifiers as shown in FIGURE 1 but also compared to heterophily-based network classification methods, as shown in FIGURE 2; this systematic comparison has not been made previously.

FIGURE 1 shows the classification accuracies of the four network classification methods across the six datasets as the proportion of labeled nodes increases from 0.1 to 0.9. Each graph is for a particular data set. The three homophily-based classifiers are WVRN, CDRN, and NBC.

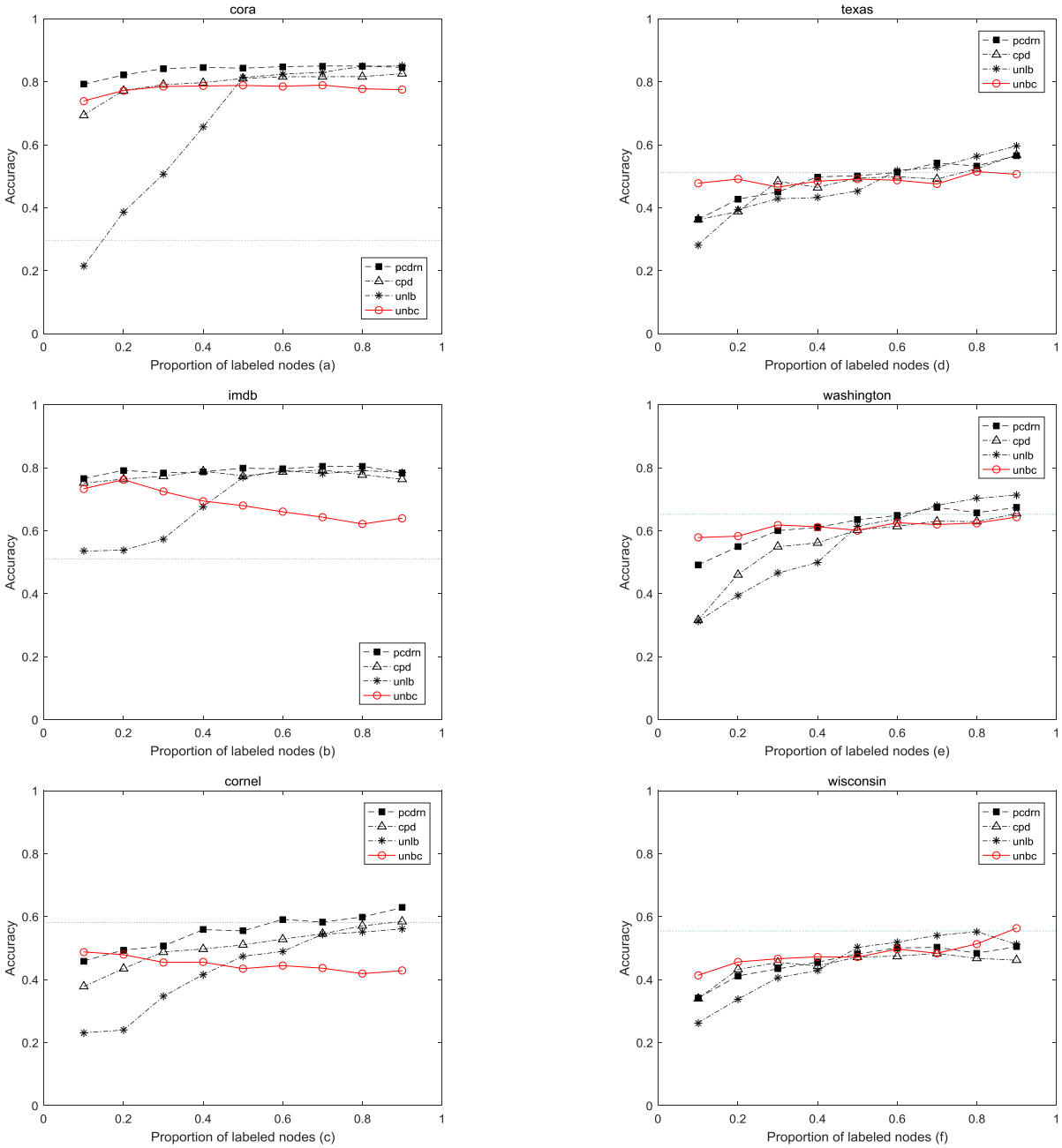


FIGURE 2. Comparison of heterophily-based methods on the six datasets. The four methods are PCDRN, CPD, UNLB, and UNBC.

As is clear from FIGURE 1, WVRN, CDRN and NBC perform worse than UNBC on the last four datasets. It is worth mentioning that, generally, WVRN is used as a baseline classifier for relational classification. From the experimental results, we can discern that homophily-based methods perform better on networks of homophily (Cora and Imdb), but worse on heterophilous networks (Cornel, Texas, Washington, and Wisconsin). The experimental results can be explained by the homophily degree in TABLE 3. Notably, the radius-two heterophily degree declines in Washington, so UNBC and NBC are roughly comparable to Washington. There must be some relationship among homophily degree, heterophily degree and radius-two heterophily degree that need further exploration.

FIGURE 1 also shows that NBC outperforms the other two homophily-based methods on heterophilous networks. However, UNBC is significantly better than NBC on the four heterophilous datasets especially when the proportion of labeled nodes is less than 0.6 because it abandons the homophily assumption. NBC performs relatively poorly with small numbers of labeled nodes owing to the lack of training data. Conversely, UNBC indicates an advantage at low

TABLE 4. p-values for the wilcoxon signed-Rank test, comparing the accuracies between pairs of four relational methods in figure 1 across four heterophilous data sets.

r	**UNBC** v WVRN	**UNBC** v CDRN	**UNBC** v NBC
0.1	**0.068**	**0.068**	**0.068**
0.2	**0.068**	**0.068**	**0.068**
0.3	**0.144**	**0.068**	**0.068**
0.4	**0.068**	**0.068**	**0.068**
0.5	**0.068**	**0.068**	**0.068**
0.6	**0.068**	**0.068**	**0.068**
0.7	**0.068**	**0.068**	**0.465**
0.8	**0.068**	**0.144**	**0.144**
0.9	**0.068**	**0.144**	**0.144**

TABLE 5. p-values for the wilcoxon signed-rank test, comparing the accuracies between pairs of four heterophily-based methods in Figure 2 across four heterophilous data sets.

r	**UNBC** v PCDRN	**UNBC** v CPD	**UNBC** v UNLB
0.1	**0.068**	**0.068**	**0.068**
0.2	**0.144**	**0.068**	**0.068**
0.3	*0.715*	*1.000*	**0.068**
0.4	*0.715*	*0.465*	**0.068**
0.5	*0.068*	*0.068*	*0.465*
0.6	*0.068*	*0.715*	*0.068*
0.7	*0.068*	*0.144*	*0.068*
0.8	*0.273*	*0.465*	*0.068*
0.9	*0.273*	*0.465*	*0.144*

sample ratios. Experiments show that UNBC performs better than homophily-based methods on networked data with heterophily.

TABLE 4 shows the p-values for a Wilcoxon signed-rank test of the four methods in FIGURE 1 assessing whether UNBC is significantly better than the other three homophily-based methods across four heterophilous data sets. Bold text means that UNBC was better than the second method and italics means it was worse. For each pair, averaging the accuracies of the 10 splits gives one average accuracy score for each heterophilous data sets. The results show clearly that UNBC outperforms the other three homophily-based methods across the board.

FIGURE 2 shows, for six of the datasets, the comparative performance of four heterophily-based methods: PCDRN, CPD, UNLB, and UNBC. UNBC performs relatively well at low sample ratios. There is no significant difference on performance between this new heterophily-based method and the other three methods for a high proportion of labeled nodes. The worst relative performance is on Cora and Imdb which are the homophilous networks. The foregoing analysis provides some evidence that UNBC performs better than the homophily-based methods when the networks are heterophilous.

TABLE 5 shows statistical results of p-values for the Wilcoxon signed-rank test across four heterophilous data sets that is corresponding to the four heterophilous methods in FIGURE 2. As discussed above, UNBC often substantially worse than the other three methods.

PCDRN, UNLB, and UNBC compute only one node at a time and get the neighbor nodes through edges. CPD uses the adjacency matrix to compute the class distribution, so CPD consumes much more storage space than other three heterophily-based methods. According to the previous variable definition, m represents the number of labels, while n represents the total number of nodes. The time complexity of PCDRN and UNBC is $O(mn^3)$ and $O(m^2n^3)$ separately in the worst case when the out degree of each unlabeled node is $n-1$. Although the time complexity of UNLB is $O(mn^2)$, UNBC performs better than UNLB when the proportion of labeled nodes is less than 0.5.

IV. CONCLUSION AND FUTURE WORKS

There is a large quantity of networked data with heterophily in the real world. In these networks, most of the interconnected nodes have distinct labels. In addition, the homophily degrees of these types of networks are low. Many homophily-based relational classifiers perform poorly on heterophilous networks. In this paper, we proposed a novel probabilistic network classifier based on a second-order Markov assumption. The proposed method uses multinomial naïve Bayesian classification. Based on the independence hypothesis, it computes the class distribution of each unlabeled node separately according to the known and unknown radius-two neighbors. Finally, it combines relaxation labeling with simulated annealing for simultaneous inferring. The experiments demonstrate that this proposed method outperforms other network classifiers on heterophilous network datasets. The proposed method is applicable to the heterophilous networks and the performance depends on the heterophily degree of the network. In future, we plan to explore the relationship between homophily degree or radius-two heterophily degree and further link distance of the nodes in the network. We will also attempt some improvements of other network classifiers for heterophilous network data.

REFERENCES

[1] S. A. Macskassy and F. Provost, "Classification in networked data: A toolkit and a univariate case study," *J. Mach. Learn. Res.*, vol. 8, pp. 935–983, May 2007.

[2] J. Neville and D. Jensen, "Iterative classification in relational data," in *Proc. 15th AAAI Workshop Learn. Stat. Models Relational Data*, Menlo Park, CA, USA, 2000, pp. 42–49, doi: 10.1.1.23.2875.

[3] V. R. Carvalho and W. W. Cohen, "On the collective classification of email 'speech acts,'" in *Proc. 28th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, New York, NY, USA, Aug. 2005, pp. 345–352, doi: 10.1145/1076034.1076094.

[4] J. Neville and D. Jensen, "Collective classification with relational dependency networks," in *Proc. ACM 2nd Workshop Multi-Relational Data Mining KDD*, New York, NY, USA, Aug. 2003, pp. 77–91.

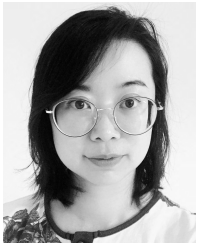
[5] M. McPherson, L. Smith-Lovin, and J. M. Cook, "Birds of a feather: Homophily in social networks," *Annu. Rev. Sociol.*, vol. 27, no. 1, pp. 415–444, 2001, doi: 10.1146/annurev.soc.27.1.415.

[6] D. Jensen and J. Neville, "Linkage and autocorrelation cause feature selection bias in relational learning," in *Proc. 19th ICML*, Jul. 2002, pp. 259–266.

[7] L. Tang and H. Liu, "Leveraging social media networks for classification," *J. Data Mining Knowl. Discovery*, vol. 23, no. 3, pp. 447–478, Nov. 2011, doi: 10.1007/s10618-010-0210-x.

[8] H. M. Richardson, "Community of values as a factor in friendships of college and adult women," *J. Social Psychol.*, vol. 11, no. 2, pp. 303–312, 1940, doi: 10.1080/00224545.1940.9918751.

[9] Z. Wang, F. Yin, W. Tan, and W. Xiao, "Classification in networked data with heterophily," *Sci. World J.*, vol. 2013, Apr. 2013, Art. no. 236769. [Online]. Available: https://www.hindawi.com/journals/tswj/2013/236769/. doi: 10.1155/2013/236769.
[10] S. A. Macskassy and F. Provost, "A simple relational classifier," in *Proc. MRDM 9th ACM SIGKDD*, 2003, pp. 64–76.
[11] C. Perlich and F. Provost, "Aggregation-based feature invention and relational concept classes," in *Proc. 9th ACM SIGKDD*, Washington, DC, USA, Aug. 2003, pp. 167–176, doi: 10.1145/956750.956772.
[12] C. Perlich and F. Provost, "Distribution-based aggregation for relational learning with identifier attributes," *Mach. Learn.*, vol. 62, nos. 1–2, pp. 65–105, Feb. 2006, doi: 10.1007/s10994-006-6064-1.
[13] J. J. Rocchio, "Relevance feedback in information retrieval," in *The SMART Retrieval System: Experiments in Automatic Document Processing*. Upper Saddle River, NJ, USA: Prentice-Hall, 1971, ch. 14, pp. 313–323.
[14] S. Chakrabarti, B. Dom, and P. Indyk, "Enhanced hypertext categorization using hyperlinks," in *Proc. ACM SIGMOD*, Seattle, WA, USA, vol. 27, Jun. 1998, pp. 307–318, doi: 10.1145/276304.276332.
[15] Q. Lu and L. Getoor, "Link-based classification," in *Proc. 20th ICML*, Washington, DC, USA, 2003, pp. 496–503.
[16] F. Lin and W. W. Cohen, "Semi-supervised classification of network data using very few labels," in *Proc. ASONAM*, Odense, Denmark, Aug. 2010, pp. 192–199, doi: 10.1109/ASONAM.2010.19.
[17] D. Zhou, O. Bousquet, T. N. Lal, J. Weston, and B. Schölkopf, "Learning with local and global consistency," in *Proc. 16th Neural Inf. Process. Syst.*, Whistler, BC, Canada, Dec. 2004, pp. 321–328.
[18] J. He, J. G. Carbonell, and Y. Liu, "Graph-based semi-supervised learning as a generative model," in *Proc. 20th Int. Joint Conf. Artif. Intell. (AAAI)*, Menlo Park, CA, USA, Jan. 2007, pp. 2492–2497.
[19] S. Dong, D. Y. Liu, L. N. Li, R. C. Ouyang, and X. L. Chai, "Relational neighbor algorithm based on class propagation distributions for classification in networked data with heterophily," *J. Jilin Univ., Eng. Technol. Ed.*, vol. 46, no. 2, pp. 522–527, Mar. 2016, doi: 10.13229/j.cnki.jdxbgxb201602029.
[20] S. Dong, D. Y. Liu, R. C. Ouyang, Y. G. Zhu, and L. N. Li, "Logistic regression classification in networked data with heterophily based on second-order Markov assumption," *J. Jilin Univ., Eng. Technol. Ed.*, vol. 48, no. 5, pp. 1571–1577, Sep. 2018, doi: 10.13229/j.cnki.jdxbgxb20170717.
[21] M. Gupta, P. Kumar, and B. Bhasker, "HeteClass: A meta-path based framework for transductive classification of objects in heterogeneous information networks," *Expert Syst. Appl.*, vol. 68, pp. 106–122, Feb. 2017, doi: 10.1016/j.eswa.2016.10.013.
[22] F. Serafino, G. Pio, and M. Ceci, "Ensemble learning for multi-type classification in heterogeneous networks," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 12, pp. 2326–2339, Dec. 2018, doi: 10.1109/TKDE.2018.2822307.
[23] Y. Sun, Y. Yuan, and G. Wang, "An on-line sequential learning method in social networks for node classification," *Neurocomputing*, vol. 149, pp. 207–214, Feb. 2015, doi: 10.1016/j.neucom.2014.04.074.
[24] S. Wang, Y. Ye, X. Li, X. Huang, and R. Y. K. Lau, "Semi-supervised collective classification in Multi-attribute network data," *Neural Process. Lett.*, vol. 45, no. 1, pp. 153–172, Feb. 2017, doi: 10.1007/s11063-016-9517-y.
[25] G. Pio, F. Serafino, D. Malerba, and M. Ceci, "Multi-type clustering and classification from heterogeneous networks," *Inf. Sci.*, vol. 425, pp. 107–126, Jan. 2018, doi: 10.1016/j.ins.2017.10.021.
[26] D. Huang, G. Guan, J. Zhou, and H. Wang, "Network-based naive Bayes model for social network," *Sci. China-Math.*, vol. 61, no. 4, pp. 627–640, Apr. 2018, doi: 10.1007/s11425-017-9209-6.
[27] M. B. Hu, W. Tang, and C. H. Cai, "Application of a new subspace updating algorithm in DOA estimation," *J. Huaqiao Univ., Natural Sci.*, vol. 33, no. 4, pp. 375–379, 2012.
[28] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 6, pp. 721–741, Nov. 1984, doi: 10.1109/TPAMI.1984.4767596.
[29] R. A. Hummel and S. W. Zucker, "On the foundations of relaxation labeling processes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-5, no. 3, pp. 267–287, May 1983, doi: 10.1109/TPAMI.1983.4767390.
[30] L. Pelkowitz, "A continuous relaxation labeling algorithm for Markov random fields," *IEEE Trans. Syst., Man and*, vol. 20, no. 3, pp. 709–715, May 1990, doi: 1.01109/21.57279.
[31] A. K. McCallum, K. Nigam, J. Rennie, and K. Seymore, "Automating the construction of Internet portals with machine learning," *Inf. Retr.*, vol. 3, no. 2, pp. 127–163, Jul. 2000, doi: 10.1023/A:1009953814988.
[32] B. Taskar, E. Segal, and D. Koller, "Probabilistic classification and clustering in relational data," in *Proc. 17th IJCAI*, Seattle, WA, USA, vol. 2, Aug. 2001, pp. 870–878.
[33] D. Jensen and J. Neville, "Data mining in networks," in *Proc. Symp. Dyn. Social Netw. Modeling Anal. Nat. Acad. Sci.*, Washington, DC, USA: Academy, Nov. 2002.
[34] J. Neville, D. Jensen, L. Friedland, and M. Hay, "Learning relational probability trees," in *Proc. 9th ACM SIGKDD*, Washington, DC, USA, Aug. 2003, pp. 625–630, doi: 1145/956750.956830.
[35] M. Craven, D. Freitag, A. McCallum, T. M. K. Nigam, and C. Y. Quek, "Learning to extract symbolic knowledge from the World Wide Web," in *Proc. 15th AAAI*, Madison, WI, USA, 1998, pp. 509–516.
[36] P. Sen, G. Namata, M. Bilgic, L. Getoor, B. Gallagher, and T. Eliassi-Rad, "Collective classification in network data," *AI Mag.*, vol. 29, no. 3, pp. 93–106, Sep. 2008, doi: 10.1609/aimag.v29i3.2157.



SA DONG was born in Liaoyang, Liaoning, China, in 1985. She received the professional master's degree in technologies for e-government from the College of Computer, University of Trento, in 2010, and the M.S. degree from the College of Computer Science and Technology, Jilin University, Changchun, China, in 2011, where she is currently pursuing the Ph.D. degree in computer science and technology.

Her research interests include artificial intelligence, statistical machine learning, and data mining.

DAYOU LIU was the Dean of the College of Computer Science and Technology, Jilin University, China, from 1993 to 2004, where he was the Dean of the Faculty of Informatics, from 2006 to 2011, and has been a Professor with the College of Computer Science and Technology, since 1990. He has authored eight books and has published more than 300 articles which have more than 4000 citations by other colleagues. His research interests include artificial intelligence, knowledge engineering, statistical learning, data mining, and expert systems.

Dr. Liu is a Distinguished Member of the China Computer Federation (CCF), the Honorary Chair of the Computer Federation of Jilin Province of China, and an Honorary Committee Member of the CCF Technical Committee on Artificial Intelligence and Pattern Recognition. He received the State Scientific and Technological Progress Award of China, in 2006.

RUOCHUAN OUYANG was born in Changchun, Jilin, China, in 1985. He received the professional master's degree in technologies for e-government from the College of Computers, University of Trento, in 2010, and the M.S. degree from the College of Computer Science and Technology, Jilin University, in 2011. He served as an Electronic and Electrical Developing Engineer with the R&D Center, FAW, from 2011 to 2018. He is currently an Engineer with the Big Data and Network Management Center, Jilin University.

YUNGANG ZHU received the Ph.D. degree in computer science from Jilin University, China, in 2012, where he is currently an Assistant Professor with the College of Computer Science and Technology.

He was a Visiting Research Fellow or a Postdoctoral Fellow with the Vienna University of Technology, Austria, the Dresden University of Technology, Germany, and the University of Trento, Italy. In recent years, he has published more than ten articles in international journals or conferences. His current research interests include probabilistic graphical models, information fusion, statistical machine learning, and data mining, with applications to knowledge engineering.

Dr. Zhu is a Committee Member of the CCF Computer Applications Technical Committee and the CAAI Intelligent Service Technical Committee. He served in the program committees for several IEEE international conferences. He serves as an Associate Editor for the IEEE CANADIAN JOURNAL OF ELECTRICAL AND COMPUTER ENGINEERING.

LINA LI was born in Jilin, China, in 1982. She received the Ph.D. degree in computer science from Jilin University, in 2012, where she is currently a Lecturer with the College of Computer Science and Technology. Her research interests include social network analysis, recommendation systems, statistical machine learning, and data mining.

TINGTING LI was born in Jilin, China, in 1986. She received the M.S. degree from the College of Business School, Jilin University, in 2011, where she is currently an Assistant Researcher with the Administrative Office, School of Computer Science. Her current research interest includes administration.

JIE LIU was born in 1973. She received the Ph.D. degree in computer science from Jilin University, China, in 2007, where she is currently an Associate Professor with the College of Computer Science and Technology. Her research interests include data mining and pattern recognition.