Clustering Nodes with Attributes via Graph Alignment

Dimitrios Rafailidis
Department of Informatics
Aristotle University of Thessaloniki, Greece
draf@csd.auth.gr

ABSTRACT

We propose a two-stage clustering method based on a graph Alignment technique, and a spectral clustering approach with Multiple Kernel Learning, namely AMKL. Our method is capable of performing clustering in attributed graphs, that is, graphs with multiple attributes per node. As attributes have different meanings, we propose a graph alignment technique for each attribute, aiming at computing an aligned matrix with the rest of the attributes. In the second step, we use an unsupervised multiple kernel learning technique, to automatically weigh the importance of edges, and the aligned matrices of the first stage, to generate the final clusters based on a spectral approach. In our experiments with five real-world attributed graphs, we evaluate each stage of AMKL, separately, and we demonstrate that the combination of the two-stages in the proposed AMKL method significantly boosts the clustering accuracy. In addition, we show that AMKL outperforms state-of-the-art methods.

CCS Concepts

•**Information systems** → **Data mining;**

Keywords

Graph clustering; attributed graphs; alignment

1. INTRODUCTION

With the advent of applications on the web, nodes in information networks can be connected based on a type of link, such as friendship, co-authorship, etc., while the nodes may have multiple attributes; for example, in a bibliography network, authors are connected based on their co-authorship, while each author has her personal attributes, such as research interests, age, affiliation; or in a social media platform, users are connected if they are friends, or have similar tastes, while for each user her/his personal information is available, like age, gender, hobbies, and so on. To model such information networks, the structure of attributed graphs has been introduced [1, 10]. Attributed graph clustering has many real-world applications, such as community detection, recommendation, targeted advertisement, and so on. As plain graph clustering methods ignore multiple attributes per node [3], several clustering methods have been proposed, such as statistical models which enforce the intra-cluster similarity by modeling the values of the node attributes and the edges in a cluster by various distributions [13]; or graph transformation [10], and spectral subspace clustering [1].

However, not all the node attributes are relevant to the clustering task, an issue that is not handled by the aforementioned clustering methods. In [11], a modularity-aware driven clustering method has been introduced, namely Maximization of Attribute-Aware Modularity (MAM). The key idea in this work is to perform local attribute selection, by identifying the most important attributes for each cluster, demonstrating that the clustering accuracy is increased, when considering an attribute selection strategy. What is missing from all the aforementioned works is that the meanings of the attributes are different, and thus there is a semantic gap when comparing them. Meanwhile, although each attribute represents one characteristic of the node, the attribute values are not independent for the same node. This is because an individual's various attributes are related to each other; for example, two co-authors in a bibliography network are connected, if they have similar research interests, affiliations, and so on.

In the effort to bridge the semantic gap between the different node attributes, we propose a graph alignment strategy. Graph alignment has several real-world applications, such as network management, or cross-network search in the multiple profiles that users have in different social media platforms [4]; however, these works focus on multi-layer networks with different types of edges. This differs from our problem, since we consider different types of node attributes, following [1, 11]. The aforementioned attributed graph clustering methods do not consider an alignment strategy over the clustering, thus ignoring the semantic gap between the node attributes, when comparing them. In this study, we propose the two-stage AMKL method. In the first stage, we

SAC 2016,April 04-08, 2016, Pisa, Italy

 ISBN 978-1-4503-3739-7/16/04...$15.00
DOI: http://dx.doi.org/10.1145/2851613.2851886

generate aligned graphs based on the attributes. By performing alignment, we obtain a set of aligned graph matrices. In the second stage, we perform an unsupervised multiple kernel learning technique, to find the optimal weighting values to aggregate the existing edges with the aligned graph matrices into a unified measure. The final clusters are generated by a spectral clustering approach, widely used for detecting clusters with arbitrary shapes and sizes [7]. As we will experimentally show, we can significantly boost the accuracy by considering both stages into the clustering task.

2. PROPOSED APPROACH

2.1 Problem Formulation

Given the sets $\mathcal{E}$ and $\mathcal{N}$ of edges and nodes, respectively, in an attributed graph, each node has a set $\mathcal{A}$ of attributes, denoted by an attribute matrix $F \in \Re^{|\mathcal{N}|\times|\mathcal{A}|}$. In addition, the edges are weighted, denoted by an adjacency matrix $B \in \Re^{|\mathcal{N}|\times|\mathcal{N}|}$, with B_{ij} in the range of $[0, 1]$, and $B_{ij} > 0$ if there is an edge $< i, j >$ in the set $\mathcal{E}$. The goal of the proposed approach is to *group the $|\mathcal{N}|$ nodes into k disjoint clusters.*

The proposed clustering method consists of two stages. Firstly, we construct $|\mathcal{A}|$ different similarity graphs based on the attributes, and for the a-th graph, with $a \in 1, \ldots, |\mathcal{A}|$, we generate an aligned graph matrix $\Phi_a \in \Re^{|\mathcal{N}|\times|\mathcal{N}|}$ with all the attributes. The goal of this stage is to minimize the gap between the $|\mathcal{A}|$ different attribute-based graphs, using the aligned matrices Φ_a. In the second stage, we use an unsupervised multiple kernel learning strategy, to combine the $|A|$ different aligned matrices with the adjacency matrix B into a unified measure, that is, a global kernel matrix $\Phi(\cdot) \in \Re^{|\mathcal{N}|\times|\mathcal{N}|}$, by weighting the importance of the edges and the aligned matrices. Finally, having computed Φ, we follow a spectral clustering approach to generate the final k clusters.

According to [7], spectral clustering methods differ in how they define and construct the Laplacian matrix L of the kernel matrix Φ and, thus, which eigenvectors are selected to represent the unified graph, aiming to exploit special properties of different matrix formulations. For example, Ratio Cut tries to minimize the total cost of the edges crossing the cluster boundaries, normalized by the size of the k clusters, to encourage balanced cluster sizes. Normalized Cut (NCut) uses the same objective criterion as Ratio Cut, normalized by the total degree of each cluster, making thus the clusters having similar degrees. In our approach, we apply the NCut method, as it generates more accurate clusters than Ratio Cut. The remainder of this Section details each step of our approach.

2.2 Graph Alignment Strategy

As aforementioned, we construct a similarity graph, denoted by a matrix $S^{(a)} \in \Re^{|\mathcal{N}|\times|\mathcal{N}|}$ for the a-th attribute, with the numerical attributes being normalized to $[0, 1]$. In the case of categorical values in the a-th attribute, the entries in matrix $S^{(a)}$ are binary, that is, 1 in case that the i-th node has the categorical value and 0 otherwise, with $i \in 1 \ldots |\mathcal{N}|$. By calculating the eigen-decomposition of $S^{(a)}$, we have:

$$S^{(a)} \approx U^{(a)} \cdot {V^{(a)}}^T \tag{1}$$

where $U^{(a)} \in \Re^{|\mathcal{N}|\times r}$, $V^{(a)} \in \Re^{|\mathcal{N}|\times r}$, and r denotes the number of latent factors. The goal is to minimize the gap between the graph $S^{(a)}$ and the rest of $|A|$-1 graphs. In particular, given the attribute matrix F (Section 2.1), the goal is to find a matrix $W^{(a)} \in \Re^{|\mathcal{A}|\times r}$, such as the aligned graph matrix Φ_a is equal to:

$$\Phi_a = F \cdot W^{(a)} \cdot {W^{(a)}}^T \cdot F^T \tag{2}$$

Therefore, the goal is to minimize the following optimization problem:

$$\min_{W^{(a)}} ||\Phi_a - S^{(a)}||_F^2 \Leftrightarrow \min_{W^{(a)}} ||F \cdot W^{(a)} \cdot {W^{(a)}}^T \cdot F^T - S^{(a)}||_F^2 \tag{3}$$

where $||\cdot||_F$ is the Frobenius norm. The aligned graph matrix Φ_a, denoted by the inner product in Eq. (3), makes the dimensions too high for the optimization problem. Given the decomposition matrix $V^{(a)}$ in Eq. (1), according to [6], the minimization problem of Eq. (3) can be approximated and rewritten as follows:

$$\min_{W^{(a)}} ||F \cdot W^{(a)} - V^{(a)}||_F^2 \tag{4}$$

To avoid overfitting, we add a regularization term in Eq. (4), resulting in the following optimization problem:

$$\min_{W^{(a)}} ||F \cdot W^{(a)} - V^{(a)}||_F^2 + \lambda ||W^{(a)}||_{2,1} \tag{5}$$

where $||\cdot||_{2,1}$ is the $\ell_{2,1}$-norm [14], and λ is a regularization parameter. Note that the regularization term $||W^{(a)}||_{2,1}$ forces $W^{(a)}$ to be sparse. The formulation in Eq. (5) is a convex problem and $W^{(a)}$ can be solved iteratively [6].

Summarizing, for each a-th node attribute, we calculate the matrix $W^{(a)}$ by solving the optimization problem in Eq. (5). Then, the a-th aligned graph matrix Φ_a is computed according to Eq. (2). Without loss of generality, in our approach, the $|A|$ aligned graph matrices are considered as linear kernel matrices [2]. The adjacency matrix B, that contains the weights of the edges of set $\mathcal{E}$, is also considered as a linear kernel matrix, having in total $|A| + 1$ kernel matrices, with $\Phi_{|A|+1} = B$. At this point, we must mention that other types of kernel functions could be used, such as Gaussian and Polynomial, to generate the kernel matrices, thoroughly examined in [2].

2.3 Kernel Matrices Weighting & Clustering

Given the $|A| + 1$ matrices, we can construct a linear combined kernel based on Multiple Kernel Learning (MKL) techniques [5, 15] as follows:

$$\Phi = \sum_{v=1}^{|A|+1} c_v \cdot \Phi_v \tag{6}$$

where c_v is the v-th weight of the kernel matrix Φ_v. Provided that, in our setting, node labels-clusters are not available prior to the clustering task, in our approach, we calculate the optimal weights c_v, using the unsupervised MKL algorithm of [15], which assumes the kernel evaluation result of the nodes as its local coding, while computing the final kernel matrix Φ based on an efficient iterative algorithm. After calculating each weight c_v in order to find the importance of each kernel matrix, we compute the combined kernel matrix Φ based on Eq. (6). Then, we calculate the normalized Laplacian of Φ:

$$L = I - D^{-1/2}\Phi D^{-1/2} \tag{7}$$

where $I \in \Re^{|N|\times|N|}$ is the identity matrix, $D \in \Re^{|N|\times|N|}$ is a diagonal degree matrix with $D_{ii} = \sum_{j=1}^{|\mathcal{N}|} \Phi_{ij}$, and $D^{-1/2}$ indicates the inverse square root of D. According to the spectral clustering approach of NCut, we compute the final k clusters by solving the following eigen-decomposition problem

$$\underset{U\in\mathbb{R}^{|\mathcal{N}|\times k}}{\operatorname{argmax}}\ tr\left(U^T \cdot L \cdot U\right),\ \text{s.t. } U^T \cdot U = I \tag{8}$$

where $tr(\cdot)$ is the trace operator, that is, the sum of the diagonal entries of a matrix. The columns of $U \in \mathbb{R}^{|\mathcal{N}|\times k}$ are the top-k eigenvectors, and its rows are the $|\mathcal{N}|$ embeddings in the eigenspace. The final clusters are generated by applying k-means in the $|\mathcal{N}|$ embeddings.

3. EXPERIMENTS

Evaluation Graphs: In our experiments, we evaluate the performance of the proposed AMKL method on five data sets of real-world graphs, publicly available at [16], also used in [11]. Table 1 summarizes the characteristics of the evaluation graphs. *Disney* is a subgraph of the Amazon co-purchase network, where each product-node has 28 attributes, such as prices or reviews ratings [8]. *DFB* consists of soccer players-nodes, and the edges represent if they have played in the same club, with each player having 4 attributes, such as the number of goals or games [1]. *ARXIV* contains papers-nodes and their citations-edges, where for each paper the 30 node attributes denote how often 30 specific keywords appear in the abstract of a paper [1]. *IMDB* consists of movies-nodes, with 21 attributes, corresponding to 21 movie genres, extracted by the Internet Movie Database, and each relation-edge between the movies is established if they share actors or there is a reference to each other via spoofs or follow ups [1]. *DBLP* contains authors-nodes with co-authorship information-edges, and the 46 node attributes denote publication ratios at certain conferences [12].

Table 1: Evaluation Graphs

Graph	$\lvert\mathcal{N}\rvert$	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{A}\rvert$
Disney	124	333	28
DFB	100	1106	5
ARXIV	856	2660	30
IMDB	862	2660	21
DBLP	28112	95410	46

Evaluation Protocol: Provided that no groundtruth is available for the evaluation graphs, following the evaluation strategy of [11], we measure the performance of the examined methods in terms of *attribute-aware modularity* (AQ). For k clusters, AQ is defined as:

$$AQ = \sum_k AC(k) \cdot Q(k) \tag{9}$$

where $Q(k)$ is the widely known *modularity* [9], and $AC(k)$ is the *attribute compactness*, that is, the sum of the relevances of all attributes $a \in \mathcal{A}$ in the k-th cluster [11], which is calculated as follows:

$$AC(k) = \frac{1}{|\mathcal{A}|}\sum_{a\in\mathcal{A}} max\big(R_a(k), 0\big) \tag{10}$$

where $R_a(k)$ is the *relevance* function [11], measuring the relevance of attribute a within the cluster k, that is, the local variance of attribute a of the nodes in cluster k divided by the global variance of a in all nodes. Although, several methods measure the performance of attributed graph clustering methods in terms of Normalized Mutual Information and the attribute Entropy, focusing on the graph structure and the attribute values, respectively, AQ is a unified measure [11]. This means that the higher the value of AQ, the better the result is with respect to both the graph structure and the attribute values, which is the main reason that we select the unified measure AQ in our evaluation strategy, following the evaluation strategy of [11], as well.

Compared Methods: To evaluate each stage of the proposed AMKL method, we consider the following baselines: *Alignment* is the proposed method, without using the weighting of the aligned matrices of Section 2.3, by setting equal weights c_v to the kernel matrices Φ_v in Eq. (6); *MKL* denotes the proposed AMKL method without the graph alignment technique of Section 2.2, by setting the kernel matrices equal to the initially computed similarity graph matrices, that is, $\Phi_a = S^{(a)}$, thus measuring the performance of the second stage of the proposed AMKL method; *CODICIL* [10] is a graph transformation strategy, which performs clustering by firstly fusing different types of links-edges into a unified graph, and then generates the final clusters. In this method, we consider the different types of similarity graphs, generated by the $|A|$ attributes, as well as the set $\mathcal{E}$ of edges. As CODICIL selectively retains edges of high relevancy within local neighborhoods from the unified graph, we determine the top-k neighbors by following the heuristic strategy of [10]. After having constructed the unified graph, following [10], the final clusters are generated based on the METIS software [3], publicly available at [17]. *MAM* [11] is the Maximization of Attribute-Aware Modularity method, a recently proposed modularity-driven clustering strategy of attributed graphs, which aims at identifying the irrelevant node-attributes to the clustering task (Section 1). In our evaluation, we use the implementation of MAM at [16]. For each evaluation graph, we report the best strategy among the five strategies proposed in [11], that is, *LM* for Disney, DFB, ARXIV, and *Coarse* for IMDB and DBLP. As MAM automatically identifies the number of clusters, we set the same number of clusters in all the examined clustering methods to make fair comparison.

Results: In Figure 1, we report the influence of the regularization parameter λ of Eq. (5) on the AMKL's performance. We conclude to 10^{-1} for Disney, ARXIV, IMDB; 10^{-2} and 10^{0} for DBLP and DFB, respectively. We observe that in the graphs with a high number $|A|$ of attributes, lower values of the λ parameter are selected. This happens because the regularization term of Eq. (5) has more impact on the optimization algorithm, as the number $|\mathcal{A}|$ of attributes increases, thus requiring a more conservative selection of λ.

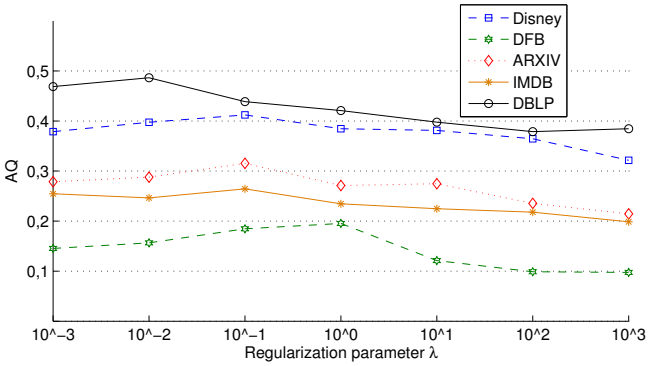

Figure 1: The impact of the regularization parameter λ of Eq. (5) on the performance of AMKL.

In Figure 2, we compare the performance of the examined methods. Alignment and MKL have limited clustering accuracy, because they either ignore the kernels weighting, or the graph alignment of the node attributes, respectively. The proposed AMKL method, achieves significantly higher accuracy than Alignment and MKL, as AMKL combines the graph alignment with the multiple kernel learning in the two-stage clustering procedure. In addition, AMKL outperforms the competitive CODICIL and MAM methods. Similar to the second stage of AMKL, that is the MKL method, CODICIL follows a weighting strategy to determine the importance of the node attributes and edges in a unified graph, whereas the MAM method performs attribute selection to identify the most important attributes for the clustering. However, CODICIL and MAM do not perform alignment, thus ignoring the different meanings of the node attributes, when comparing them, which is the main reason that the proposed AMKL method is superior over these methods.

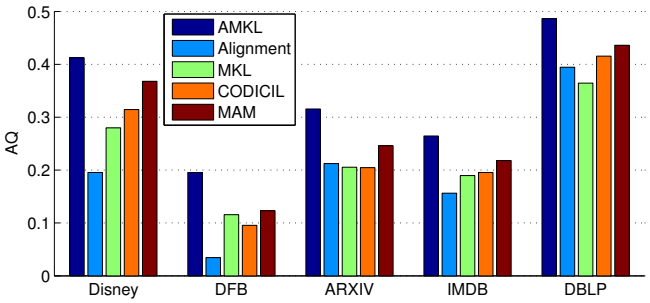

Figure 2: Performance evaluation of the examined methods, in terms of AQ.

4. CONCLUSION

In this paper, we presented AMKL, an efficient two-stage clustering method for attributed graphs. We experimentally showed that we achieve high clustering accuracy by considering two important key factors: (1) we propose a graph alignment strategy based on the node attributes to bridge the semantic gap that different attributes have; (2) we apply an unsupervised multiple kernel learning method to weigh the importance of the edges and the attribute-based graphs in a unified kernel matrix. It is the combination of the two key factors that allows the proposed AMKL method to outperform the baseline methods. Interesting future directions are to extend our method for generating multiple non-redundant clusterings, as well as to consider an incremental strategy for dynamic attributed graphs, where new edges, nodes, or the node attributes may be added or evolve over time.

5. REFERENCES

[1] S. Günnemann, I. Färber, S. Raubach, and T. Seidl. Spectral subspace clustering for graphs with feature vectors. In *IEEE International Conference on Data Mining*, ICDM '13, pages 231–240, 2013.

[2] T. Hofmann, B. Schölkopf, and A. J. Smola. Kernel methods in machine learning. *Ann. Statist.*, 36(3):1171–1220, 2008.

[3] G. Karypis and V. Kumar. Multilevel algorithms for multi-constraint graph partitioning. In *ACM/IEEE Conference on Supercomputing*, SC '98, pages 1–13, 1998.

[4] D. Koutra, H. Tong, and D. Lubensky. BIG-ALIGN: fast bipartite graph alignment. In *IEEE International Conference on Data Mining*, ICDM '13, pages 389–398, 2013.

[5] G. R. G. Lanckriet, N. Cristianini, P. L. Bartlett, L. E. Ghaoui, and M. I. Jordan. Learning the kernel matrix with semidefinite programming. *Journal of Machine Learning Research*, 5:27–72, 2004.

[6] X. Liu, L. Wang, J. Zhang, J. Yin, and H. Liu. Global and local structure preservation for feature selection. *IEEE Transactions on Neural Networks and Learning Systems*, 25(6):1083–1095, 2014.

[7] U. Luxburg. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[8] E. Müller, P. I. Sanchez, Y. Mülle, and K. Böhm. Ranking outlier nodes in subspaces of attributed graphs. In *IEEE International Conference on Data Engineering*, ICDE '13, pages 216–222, 2013.

[9] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.

[10] Y. Ruan, D. Fuhry, and S. Parthasarathy. Efficient community detection in large networks using content and links. In *International World Wide Web Conference*, WWW '13, pages 1089–1098, 2013.

[11] P. I. Sánchez, E. Müller, K. Böhm, A. Kappes, T. Hartmann, and D. Wagner. Efficient algorithms for a robust modularity-driven clustering of attributed graphs. In *SIAM International Conference on Data Mining*, SDM '15, 2015.

[12] P. I. Sanchez, E. Müller, O. Irmler, and K. Böhm. Local context selection for outlier ranking in graphs with multiple numeric node attributes. In *Conference on Scientific and Statistical Database Management*, SSDBM '14, pages 16:1–16:12, 2014.

[13] Xu, Zhiqiang and Ke, Yiping and Wang, Yi and Cheng, Hong and Cheng, James. GBAGC: A general bayesian framework for attributed graph clustering. *ACM Trans. Knowl. Discov. Data*, 9(1):5:1–5:43, 2014.

[14] Y. Yang, H. T. Shen, Z. Ma, Z. Huang, and X. Zhou. $l_{2,1}$-norm regularized discriminative feature selection for unsupervised learning. In *International Joint Conference on Artificial Intelligence*, IJCAI '11, pages 1589–1594, 2011.

[15] J. Zhuang, J. Wang, C. H. Hoi, and X. Lan. Unsupervised multiple kernel learning. *Journal of Machine Learning Research*, 20:129–144, 2011.

[16] http://ipd.kit.edu/~muellere/mam/.

[17] http://glaros.dtc.umn.edu/gkhome/metis/metis/overview.