Reviewer #1

Thank you for your affirmation and encouragement. Your suggestions are of great help to the improvement of the paper.

Issue 1: While the authors claim that the connection between k-means and ratio-cut is the basis for their algorithm, the justification for this motivation does not come through. The objective is clear enough without this motivation.

Reply 1: As you said, the objective can indeed be clearly expressed without this motivation. However, our idea of using k-nearest neighbor graph and Euclidean distance metric in the objective function comes from the observation of the unified view. Therefore, we take it as the motivation of our model. I am sorry the paper is not clearly written.

specific comments/queries: 9. In order to distinguish it from the distance matrix involved in k-means, I expressed the degree matrix of the graph as $\mathbf{\Delta}$. I will also modify the degree matrix involved in Section 1 to $\mathbf{\Delta}$. **10.** Vector $\mathbf{p}(\mathbf{p} \geq 0, \mathbf{p}^T\mathbf{1} = 1)$, used to represent the weight of each item. **12.** As you said, my statement does have some improprieties. I will separate the model with constraint from the final model we employed so that readers can clearly understand the concessions made by the model. **15.** $\tilde{\mathbf{Y}}$ represents the indicator matrix $\mathbf{Y}$ but the i-th row of $\tilde{\mathbf{Y}}$ does not need to be optimized. **18.** As mentioned in Section 1, SC can usually obtain satisfactory performance with high complexity. Therefore most studies are devoted to reducing its complexity, resulting in the performance of the proposed algorithm is usually not as good as SC. So, in the tables, we have ignored the cases where SC performs best. **19.** Accuracy and Precision are two different metrics. The calculation of Accuracy needs the help of bestmap that is very time-consuming. Precision often appears together with recall, is a metric that is more suitable for large-scale data. The clustering performance in terms of precision, recall, and F_1 score is shown in the supplementary material.

In addition, I will seriously consider the suggestions in **1, 2, 3, 4, 5, 6, 7, 8, 11, 13, 14, 16, 17, and 20**, and carefully modify the submission, regardless of whether the submission is accepted or not.

Reviewer #2

Issue 1: There are several assumptions made during the text, for example, the distance measure is doubly stochastic, the clusters are of similar size. These assumptions might not always hold.

Reply 1: (1) Indeed, the matrix $\mathbf{W}$ cannot be guaranteed to always be a double random matrix. But all we need to know is that the equivalence between KM and Ratio-cut holds under some conditions. (2) As the reviewer said, this assumption (the clusters are of similar size) is a concession that makes the problem easy to solve.

Issue 2: It's unclear what is the novelty of the work. The optimization problem presented is not stated as a contribution.

Reply 2: The proposed model is indeed one of the important contributions of this article. As you said, this was ignored when listing contributions to this article. We will declare this contribution in the list in subsequent versions.

Issue 3: The algorithm adds a new hyper-parameter to the problem (i.e. the number of nearest nighbors).

Reply 3: You are right. Almost all graph-based methods involve this parameter, which is difficult to avoid. Fortunately, the performance of our model is not sensitive to this parameter. Satisfactory performance is achieved with K is fixed at 20, which is also mentioned in Section 4.2.

Issue 4: The accuracy of the algorithms is calculated using a reference clustering. What is this reference?

Reply 4: Our model was evaluated on datasets with labels, where the labels serve as a proxy of the optimal clustering.

Reviewer #3

First of all, thank you for your valuable comments. However, this paper does not actually have the problems described in suggestions 2 and 3. The main reason for this misunderstanding is that the paper is not clearly written. To this end, I will carefully revise my paper based on the reviewers' suggestions.

Issue 1: Deriving algorithms from the relationship between k-means and graph-cut is not surprising, and the novelty of the proposed work seems might be challenged.

Reply 1: As the reviewer #1 said, the relationship between k-means and ratio-cut is the basis of our algorithm. It is used to illustrate the motivation of the proposed model, and this relationship is not a contribution of this article.

suggestion 2: it is suggested to include the clustering performance of k-d tree as a baseline to compare since the proposed method uses k-d tree to construct a k-nearest neighbor graph as an intermediate step.

Reply 2: Sorry, I don't get the point of this suggestion. How do I compare the clustering algorithm with the k-d tree? The k-d tree seems to be an algorithm used to quickly find the neighbors of samples. I don't know how to get the cluster structure from this algorithm.

suggestion 3: Why not include the performance of ratio-cut or normalized-cut algorithms as a comparison baseline?

Reply 3: The problems of both ratio-cut and normalized-cut are NP-hard. The normalized spectral clustering (SC) considered in our experiments can be seen as a relaxed version of normalized-cut. And the unnormalized SC derived from ratio-cut have a similar performance with normalized SC.

Reviewer # 4

Thank you for your affirmation and encouragement. As you said, the most attractive contribution of this article is the performance in terms of computational overhead. To the best of our knowledge, the proposed method is the first graph-based algorithm with the computational complexity that is independent of the product of n (the number of samples) and c (the number of clusters). In addition, I will seriously consider your suggestions and revise this paper.