A Method for Handling Negative Similarities in Explainable Graph Spectral Clustering of Text Documents *

Mieczysław A. Kłopotek[1 [0000−0003−4685−7045]],
Sławomir T. Wierzchoń[1 [0000−0001−8860−392X]],
Bartłomiej Starosta[1 [0000−0002−5554−4596]],
Dariusz Czerski[1 [0000−0002−3013−3483]], and
Piotr Borkowski[1 [0000−0001−9188−5147]]

Institute of Computer Science, Polish Academy of Sciences, ul. Jana Kazimierza 5, 01-248 Warsaw, Poland
{klopotek,barstar,stw,dcz,pbr}@ipipan.waw.pl, http://www.ipipan.waw.pl

Abstract. This paper investigates the problem of Graph Spectral Clustering (GSC) with negative similarities, resulting from document embeddings different from the traditional Term Vector Space (like doc2vec, GloVe, etc.). Solutions for combinatorial Laplacians and normalized Laplacians are discussed. An experimental investigation shows the advantages and disadvantages of solutions proposed in the literature and in this research. The research demonstrates that GloVe embeddings frequently cause failures of normalized Laplacian based GSC due to negative similarities. Application of methods curing similarity negativity leads to accuracy improvement for both combinatorial and normalized Laplacian based GSC. It also leads to applicability for GloVe embeddings of explanation methods developed for Term Vector Space embeddings.

Keywords: Artificial Intelligence · Machine Learning · Graph spectral clustering · document embeddings · negative similarities

1 Introduction

Graph Spectral Clustering (GSC) is known as an effective method of clustering when data are available in the form of a similarity matrix. As the method relies on Laplacians of the similarity matrix, non-negative similarities are required. However, there exist multiple applications where non-negativity is not guaranteed, which leads to numerous formal and numerical problems, as pointed e.g. in [4]. Although solutions have been proposed for various domains, they have not been discussed for text document clustering. In this paper, we attempt to address them.

Originally, the similarity of text documents was computed as a cosine of the angle between documents embedded in the Term Vector Space (TVS for

* Supported by Polish Ministry of Science

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_1

short). These similarities were non-negative by definition. However, the emergence of new and more efficient embedding methods for textual documents such as Word2vec [12], Doc2Vec [7], GloVe [9], BERT [2] based on transformers and others [8] gave rise to the problem of the emergence of negative similarities. This fact causes formal, theoretical, and computational problems for GSC, as computational efficiency and accuracy deteriorate. In addition, normalized Laplacians may not be computable, and the procedure developed to explain clustering results, as described in [13] will fail.

In this paper, we address the clustering of tweets. Their sheer volume, noise, and dynamics impose challenges that hinder the effectiveness of observing clusters with high intra-cluster and low inter-cluster similarity, see e.g. [10].

The paper is organized as follows: Section 2 gives an overview of previous work on related topics. Section 3 contains our proposed solution to the negative similarity problem, and Section 4 illustrates the effectiveness of the proposed method. A summary of the article is given in Section 5. Due to space limitations, only relevant comments are presented here. The reader will find more details in the extended version [6].

2 Previous work

Graph Spectral Clustering is a methodology for low-complexity approximation of graph clustering based on graph cut criteria. The best-known criteria are RCut (ratio-cut) and NCut (normalized cut) defined as follows:

$$RCut(\Gamma) = \sum_{j=1}^{k} \frac{cut(C_j, \bar{C}_j)}{|C_j|} = \sum_{j=1}^{k} \frac{1}{|C_j|} \sum_{i \in C_j} \sum_{\ell \notin C_j} s_{i\ell} \tag{1}$$

$$NCut(\Gamma) = \sum_{j=1}^{k} \frac{cut(C_j, \bar{C}_j)}{\mathcal{V}_j} = \sum_{j=1}^{k} \frac{1}{\mathcal{V}_j} \sum_{i \in C_j} \sum_{\ell \notin C_j} s_{i\ell} \tag{2}$$

Here $s_{i\ell}$ stands for the similarity between objects i and ℓ, (usually it is a number between 0 and 1), Γ is the partition of objects, $\bar{C}_j$ denotes the complement of the cluster C_j, $|C_j|$ stands for the cardinality of C_j, and $\mathcal{V}_j = \sum_{i \in C_j} \sum_{\ell} s_{i\ell}$ is the volume of j-th cluster. The elements s_{ij} form a similarity matrix S, and by convention $s_{ii} = 0$ as acyclic graphs are used in GSC.

A *combinatorial Laplacian* is defined as

$$L = D - S, \tag{3}$$

where D is the diagonal matrix with $d_{ii} = \sum_{\ell=1}^{n} s_{i\ell}$ for each $i \in [n]$. A *normalized Laplacian* $\mathcal{L}$ of the graph represented by S is defined as

$$\mathcal{L} = D^{-1/2} L D^{-1/2} = I - D^{-1/2} S D^{-1/2}. \tag{4}$$

There exist numerous application areas where it is convenient to use negative similarity measures. They include, but are not limited to, studies based on correlations [3], investigations of electric networks [16], and others. As mentioned

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_1

in the Introduction, such similarity measures constitute various problems both for the graph cut criteria and the GSC clustering methods if we want to extend them into such a realm.

To overcome the problems with negative similarities, several proposals were elaborated. One can eliminate negative similarities setting them to zero[1]

$$s_{ik}^{(pZ)} = \begin{cases} s_{ik} \text{ if } s_{ik} > 0 \\ 0 \quad \text{otherwise} \end{cases} \tag{5}$$

Other simple possibilities include taking absolute values, or adding a positive constant to all edge weights. Approaches depend on the application, i.e. why some weights are negative and what the negativity means.

3 Our approach to technical problems

A deeper discussion of these topics can be found in the extended version [6].

3.1 The problem of combinatorial Laplacians

A simple transformation relies upon adding a positive constant c to all off-diagonal similarities. New matrix $\tilde{S}$ takes the form $\tilde{S} = S + c(J - I)$, where I is the identity matrix, $J = \mathbf{1}\mathbf{1}^T$ is the matrix with all elements equal to one, and $\mathbf{1}$ is the vector with all entries equal to one. Thus the entries of $\tilde{S}$ are

$$s_{ik}^{(pA)} = s_{ik} + c \tag{6}$$

and the degree matrix $\tilde{D}$ induced by $\tilde{S}$ is

$$\tilde{D} = \text{diag}(\tilde{S}\mathbf{1}) = \text{diag}(S\mathbf{1} + c(n-1)\mathbf{1}) = D + c(n-1)I \tag{7}$$

where $\text{diag}(v)$ is a diagonal matrix with v as its diagonal. Laplacian of $\tilde{S}$ is

$$\tilde{L} = \tilde{D} - \tilde{S} = L - cJ + cnI \tag{8}$$

Let (λ, v) be an eigenpair of the Laplacian L. Then

$$\tilde{L}v = Lv - cJv + cmIv = (\lambda + cn)v \tag{9}$$

since $Jv = \mathbf{0}$. This shows that $(\lambda + cn, v)$ is an eigenpair of $\tilde{L}$. So, RCut minimizing clustering remains unchanged under such an operation.

Alternatively, we can add a positive constant α to the diagonal elements of the degree matrix, that is, $\hat{D} = D + \alpha I$. Then for any eigenpair of L

$$(\hat{D} - S)v = Lv + \alpha v = (\lambda + \alpha)v \tag{10}$$

[1] The proposal of signed cuts in [5] ignores in fact negative weights, see [4].

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_1

Although the matrix $(\hat{D} - S)$ does not fulfill the requirements of being a combinatorial Laplacian, it still belongs to a large family of generalized graph Laplacians, [1]. Rocha and Trevisan call such matrices perturbed Laplacians and develop their theory in [11].

The similarities $s_{ik}^{(pA)}$ will get out of the range $[0, 1]$ for large enough c. To get them again into this range, we can divide them by $c+1$, leading to elements of the matrix $\bar{S} = \frac{S+c(J-I)}{1+c}$ of the form:

$$s_{ik}^{(pN)} = \frac{s_{ik} + c}{1 + c} \tag{11}$$

which will lead to the same eigenvectors of the resulting combinatorial Laplacian $\bar{L} = \bar{D} - \bar{S}$ with $\bar{D} = \mathrm{diag}(\bar{S}\mathbf{1})$ as for the original L.

Conclusion: The calculation of Laplacians $\tilde{L}, \hat{L}, \bar{L}$ is not necessary, because the eigenvectors of the original L will not differ. Hence, also the clustering based on the lowest eigenvectors will yield the same results.

Interestingly, the formula (11) can be assigned a geometric interpretation if we compute the similarities as cosines between the document embedding vectors in an N-dimensional space, such as the doc2vec or GloVe space, upon extending this space with an additional dimension, with a constant coordinate, see [6].

3.2 The problem of normalized Laplacians

Two types of problems with computation of normalized Laplacian $\mathcal{L} = D(S)^{-1/2} L D(S)^{-1/2}$ may occur. First type of problems is faced, when L contains positive off-diagonal elements while all diagonal elements are positive, so that $\mathcal{L}$ is computable, but some off-line elements remain positive. Curing this situation is analogous to combinatorial Laplacian and shall not be detailed here. Second type of problems occurs, as described among others in [4], when some elements of D are negative. This is a more profound problem than just square rooting negative numbers, [4]. The NCut criterion refers to the cluster volume that may turn out to be negative. A cluster with negative volume – that is with strongly dissimilar documents – has a chance to minimize the NCut criterion. Instead of clusters with strongly similar documents one gets ones with strongly dissimilar ones. This issue strongly resembles the problems with kernel k-means which may not reach the minimum of k-means criterion. Therefore, the NCut criterion must be addressed at the very beginning.

We will consider several proposals.

- Adding a constant to the diagonal of the matrix D,
- Adding a constant to each element of the similarity matrix S,
- Manipulating similarity computation by taking not the cosine of the angle between documents, but half of this angle.
- Replacing similarity with the exponent of the negated distance between documents on a unit sphere.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_1

Consider adding a positive number to all similarities. The clustering taking similarities into account, that is RCut, will not change. Adding a sufficiently large constant will make Normalized Laplacian computable. But will the NCut change under such an operation? Define

$$s_{ik}^{(pD)}(x) = \frac{s_{ik}}{\sqrt{d_{ii}+x}\sqrt{d_{kk}+x}} \tag{12}$$

$s_{ik}^{(pD)}(0)$ is the negated off-diagonal element of normalized Laplacian.

It can be shown that if $s_{ik}^{(pD)}(0) > s_{i\ell}^{(pD)}(0)$, then $s_{ik}^{(pD)}(c) > s_{i\ell}^{(pD)}(c)$, for $c > 0$ and $s_{ik} > s_{i\ell}$. Consider three documents, i, k, ℓ and let

$$s_{ik}^{(pD)}(c) > s_{i\ell}^{(pD)}(c) \text{ which means } \frac{s_{ik}}{\sqrt{d_{kk}+c}} > \frac{s_{i\ell}}{\sqrt{d_{\ell\ell}+c}}$$

If we are in the realm of non-negative similarities (other cases can be handled similarly)

$$\frac{s_{ik}^2}{s_{i\ell}^2} > \frac{d_{kk}+c}{d_{\ell\ell}+c}$$

With an increase of c, the expression on the right-hand side grows/decreases towards one. Therefore, if originally $s_{ik} > s_{i\ell}$, then the expression is true.

This means that adding a constant to the normalized Laplacian diagonal keeps to a great extent the ordering of similarities, so that the results of clustering may be similar, unless the normalization changes proportions between similarities in the original Laplacian. In case of some negative d_{ii}, adding an appropriate constant may turn the Laplacian into a computable one, resulting in clustering similar to the one originally intended.

However, the problem is that this solution tends to be in fact a version of the newly introduced NRCut [14], and not NCut. So another approach is needed.

Another solution would be adding a constant c to each similarity ($s_{ik}^{(pA)}(c) = s_{ik} + c$. Again, no warranty that the ordering of all normalized similarities will be the same and hence that clustering result is the same.

One solution could be to transform the similarity matrix S into a positive one $S^{(pQ)}$ as follows:

$$s_{ik}^{(pQ)} = \cos\left(\frac{\pi}{2}\frac{\arccos s_{ik}}{\max_{i,k\in[n],i\neq k}\arccos s_{ik}}\right) \tag{13}$$

whereby max is computed over all off-diagonal elements of the S matrix. cos is non-negative in the range $[0, \frac{\pi}{2}]$ while it is negative for greater angles. By dividing the actual angles between documents by the maximal angle, and multiplying with $\frac{\pi}{2}$ we scale all the angles into the non-negative cosine range. Now, the traditional normalized Laplacian is applicable. The ranking of similarities of combinatorial Laplacian is preserved completely, but again no warranty for the normalized similarities. The above formula can be generalized to:

$$s_{ik}^{(pC)} = \cos\left(\frac{\arccos s_{ik}}{1+c}\right) \tag{14}$$

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_1

$c = 1$ is for sure a reasonable choice, because *arccos* returns values in the range $[0, \pi]$ and dividing this result by 2 scales them into the required range $[0, \frac{\pi}{2}]$.

If the graph has isolated nodes, we already get into trouble with normalized Laplacian because of division by zero; the same applies to the NCut criterion. Therefore, a change in understanding NCut is needed in such a way that a cluster with all nodes isolated has a non-zero volume. So, the similarity needs to be transformed. As the Euclidean distance between two normalized vectors x_i, x_j equals to $\|x_i - x_j\|^2 = 2(1 - s_{ij})$, where $s_{ij} = cos(x_i, x_j)$, our proposal is

$$s_{ik}^{(pE)} = e^{-(1-s_{ik})/2} \tag{15}$$

This should be applied to get a new similarity matrix S' as well as a redefinition of NCut to $\text{NCut}^{(pE)}$ (based on the new similarities). There is no need to worry about isolated nodes. If we generalize the transformation $s_{ik}^{(pE)}$ to

$$s_{ik}^{(pE)}(c) = e^{-(1-(s_{ik}+c))/2} \tag{16}$$

the normalized Laplacian will remain the same for all values $c \geq 0$ because adding a constant in the exponent is the same as multiplying the similarity with another constant.

3.3 Negativity versus Explainability

GSC result explanation procedure elaborated in [13] encounters serious problems as it is based on the products of word embedding vectors and cluster center vectors which would lead to meaningless negative word importance. The correction proposed for combinatorial Laplacian based GSC keeps the spirit of [13]. As normalized Laplacian is concerned, we show in [13] that additive corrections of similarity measure does not disturb the explanation bridge.

4 Experiments

We conducted experiments on the effectiveness of GSC methods to predict hashtags for a large set of **short** tweets using different methods to deal with negative similarities, as mentioned in the formulas (6), (11), (13), (14), (16), (5) for $c = 0, 1, 2, 3$. Note that $c = 0$ means that no correction of negative similarity was performed. For the modified similarity matrices, both combinatorial and normalized Laplacians were used in GSC. The computations were performed for the traditional Term Vector Space (TVS, tf, tfidf) as well as for the GloVe based embeddings: TweetGlove (trained on Twitter data) and WikiGlove (trained on Wikipedia Data). The results can be accessed via the link https://github.com/ipipan-barstar/ICCS25.MfHNSiEGSCoTD.

As expected, the Term Vector Space embeddings have no negative similarity problems. TweetWiki embedding leads to numerous negative similarity matrix entries, but no problem with row sums occurs for our samples. The most difficult problems occur for the WikiGloVe embedding, as there are many more

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_1

negative similarities and there are multiple rows with negative entries in three of the samples. When the correction of negative similarities is based on zeroing them, Normalized Laplacian based clustering could be executed. We see that GloVe based do not have a big advantage over TVS based embeddings. The results are the worst compared to other methods. When the correction of negative similarities is based on adding a constant to all off-diagonal similarities, with or without dividing for normalization, Normalized Laplacian-based clustering could be executed except for $c = 0$ in WikiGlove embedding because the diagonal of D contained negative entries. GloVe based GSC does not have any TVS based embeddings. At the same time, adding the constant $c = 1$ significantly improves the performance, while higher constants do not contribute much to the results. When normalizing over the largest angle between document vectors, the results are worse for TVS embeddings, and slightly worse for GloVe embeddings. When dividing the angle between document vectors, the results constitute an improvement when dividing by at least two, but dividing by higher values does not contribute anything. When replacing primary similarities with their exponential variants, the variants do not differ much, but replacement of negative similarities with exponential ones helps the GloVe based embeddings, and also the TVS embeddings benefit from this transformation. The GSC results for combinatorial Laplacians are significantly worse, and the effects of transformations are generally marginal, as expected.

Detailed results for all samples are available at https://github.com/ipipan-barstar/ICCS25.MfHNSiEGSCoTD.

5 Conclusions

In this paper we discussed the issues in graph spectral clustering of documents resulting from growing popularity of embeddings different from the traditional Term Vector Space. The major problem is the negative cosine similarities between documents under these embeddings. We have studied six different methods for overcoming negative similarities. Essentially, the combinatorial Laplacian-based clusterings seem to be unaffected by negative similarities, as demonstrated by theoretical arguments. In case of normalized Laplacians, the method of setting negative similarities to zero yields the worst results. The other methods perform similarly. Interestingly, it turns out that for Term vector Space embeddings there may be an improvement of performance when the similarity correction is applied. We were also able to provide a geometric interpretation of one of the studied methods [6]. Note however other approaches to use similarities, e.g. [15]. This study was limited to two GloVe type embeddings, based on Wiki training data and Tweeter training data.

References

1. Biyikoglu, T., Leydold, J., Stadler, P.F.: Laplacian Eigenvectors of Graphs. Perron-Frobenius and Faber-Krahn Type Theorems, Lecture Notes in Mathematics,

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_1

vol. 1915. Springer-Verlag, Berlin Heidelberg (2007), https://doi.org/10.1007/978-3-540-73510-6
2. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding (2019). https://doi.org/10.48550/ARXIV.1810.04805
3. Knyazev, A.: Edge-enhancing filters with negative weights. In: 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP). pp. 260–264 (2015). https://doi.org/10.1109/GlobalSIP.2015.7418197
4. Knyazev, A.: Signed Laplacian for spectral clustering revisited. arXiv (Jan 2017), https://arxiv.org/abs/1701.01394
5. Kunegis, J., Schmidt, S., Lommatzsch, A., Lerner, J., Luca, E.W.D., Albayrak, S.: Spectral Analysis of Signed Graphs for Clustering, Prediction and Visualization, pp. 559–570. Society for Industrial and Applied Mathematics. (2010). https://doi.org/10.1137/1.9781611972801.49
6. Kłopotek, M.A., Wierzchoń, S.T., Starosta, B., Czerski, D., Borkowski, P.: A method for handling negative similarities in explainable graph spectral clustering of text documents – Extended Version. CoRR **abs/2504.12360** (2025), https://arxiv.org/abs/2504.12360
7. Lau, J.H., Baldwin, T.: An empirical evaluation of doc2vec with practical insights into document embedding generation (2016). https://doi.org/10.48550/ARXIV.1607.05368
8. Lin, T., Wang, Y., Liu, X., Qiu, X.: A survey of transformers. AI Open **3**, 111–132 (2022). https://doi.org/https://doi.org/10.1016/j.aiopen.2022.10.001
9. Pennington, J., Socher, R., Manning, C.: GloVe: Global vectors for word representation. In: Moschitti, A., Pang, B., Daelemans, W. (eds.) Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). pp. 1532–1543. Association for Computational Linguistics, Doha, Qatar (Oct 2014). https://doi.org/10.3115/v1/D14-1162
10. Ravi, J., Kulkarni, S.: Text embedding techniques for efficient clustering of Twitter data. Evol. Intel. **16**, 1667–1677 (2023). https://doi.org/https://doi.org/10.1007/s12065-023-00825-3
11. Rocha, I., Trevisan, V.: A Fiedler-like theory for the perturbed laplacian. Czech Math J **66**, 717–735 (2016), https://doi.org/10.1007/s10587-016-0288-4
12. Rong, X.: word2vec parameter learning explained (2014). https://doi.org/10.48550/ARXIV.1411.2738, arXiv:1411.2738 [cs.CL]
13. Starosta, B., Kłopotek, M.A., Wierzchoń, S.T., Czerski, D., Sydow, M., Borkowski, P.: Explainable graph spectral clustering of text documents. PLoS One **20(2):e0313238** (February 2025). https://doi.org/10.1371/journal.pone.0313238
14. Starosta, B., Kłopotek, M.A., Wierzchoń, S.T.: Approaches to explainability of output of graph spectral clustering methods. to appear in monograph "Design and Implementation of Artificial Intelligence Systems", published by University of Siedlce (2025)
15. Stec, H., Ekanadham, C., Kallus, N.: Is cosine-similarity of embeddings really about similarity? arXiv (2024). https://doi.org/https://doi.org/10.1145/3589335.3651526
16. Zelazo, D., Buerger, M.: On the definiteness of the weighted Laplacian and its connection to effective resistance. In: 53rd IEEE Conference on Decision and Control. p. 2895–2900. IEEE (Dec 2014). https://doi.org/10.1109/cdc.2014.7039834

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97635-3_1