Mieczysław A. Kłopotek[1]
ORCID: 0000-0003-4685-7045

Institute of Computer Science of Polish Academy of Sciences
ul. Jana Kazimierza 5, 01-248 Warszawa, Poland

[1] klopotek@ipipan.waw.pl

Spectral Clustering Versus Watanabe Theorem

DOI: 10.34739/si.2024.30.02

Abstract. We devote this paper to a special case of Graph Spectral Clustering of graphs with identical distances between nodes. This study is motivated by the special theorem presented by Watanabe which claims that given all derivable attributes are taken into account, all distinct objects are at the same distance. As the multi-view clustering becomes popular, the mentioned Watanabe theorem may imply serious problems for recovering the intrinsic structure of the collection of objects. We show that Graph Spectral Clustering should not be affected in the most favourable case that is block structure of similarity matrix in theory, but in practice the underlying k-means algorithm introduces up to 20% error rate in assignment of elements to clusters.

Keywords: Ugly Duckling Theorem, Graph Spectral Clustering, k-means clustering algorithm, Fiedler vector

1 Introduction

Watanabe [30,31] presented in 1969 and reiterated in 1987 an interesting theorem, called Ugly Duckling Theorem. We provide here a bit elaborated version.

Theorem 1 *Assume that a set of binary attributes $A_1, \ldots, A_n$ are relevant for the description of objects O_k in some domain. Assume that if two attributes A_i, A_j are relevant, then their logical conjunction $A_i \wedge A_j$, disjunction $A_i \vee A_j$ and negation $\neg Aj$ are also relevant. Assume further that for two objects O_k, O_l their similarity $sim(O_k, O_l)$ is measured as the share of relevant attributes that have the same value for both objects. Then for distinct objects (ones with at least one not agreeing value for attributes $A_1, \ldots, A_n$), the similarity is equal 0.5.*

This theorem has a multitude of implications in the domain of data clustering. In [15] it has been shown that under these circumstances the Fisher's Flexible Clustering algorithm COBBWEB [9] produces a flat hierarchy, and under careful selection of a subset of relevant attributes any hierarchical clustering can be obtained (see also [19]). In [17], a further analysis showed that attributes both weakly and strongly related with other attributes have deteriorating impact onto the overall classification. Proper construction of derived attributes as well as selection of scaling of individual attributes strongly influences the obtained hierarchical clustering. [22] shows that the Ugly Duckling effect studied by Watanabe affects both clustering and metaclustering. In [16], it is shown that also consensus between document similarity measures ([7,27]) is also affected by the phenomenon. We observe in recent years growing interest in manipulation of document similarity, [36,37,8,32] which may introduce the effects of Ugly Duckling to the results of clustering.

The Ugly Duckling problem itself attracted for years the attention of researchers in various theoretical and practical domains, just see papers [28] (mathematical theories), [2] (biology), [3,4] (philosophy), [26] (psychology), [10] (business), [13] (cellular automata simulation), [11] (modern GPT problems), [14] (recommender systems), and many more.

These and other observations prompted us to study the impact of Ugly Duckling effect on the results of the now quite popular Graph Spectral Clustering (GSC) [34,23,24,5]. This is especially important in the light of new approaches to spectral clustering, the aforementioned multi-view approaches [29,20] that manipulate the similarity matrices based on derived attributes.

2 Graph Spectral Clustering

Spectral Clustering methods draw much attention over many years due to their capability to identify clusters with non-typical shapes [35,34,21] and constantly attract researchers to improve their performance [12,1,18].

Spectral clustering methods are deemed to be a relaxation of cut based graph clustering methods. They start with transforming a similarity matrix to a graph Laplacian, for which lowest eigenvectors are taken as input to a conventional clustering method. Let S be a similarity matrix between pairs of items (e.g. documents). It induces a graph whose nodes correspond to the items. A combinatorial Laplacian L corresponding to this matrix is defined as

$$L = D - S, \tag{1}$$

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

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

If one now wants to cluster into k clusters, one takes k eigenvectors of one of the above Laplacians, related to k lowest eigenvalues of the Laplacian as representation of the objects (documents) and clusters with this representation using k-means [21]. If you want to cluster just into two clusters, you take the so-called Fiedler vector, that is the vector associated with the second lowest eigenvalue. Then the elements associated with positive elements of the Fiedler vector will belong to one cluster, and the elements associated with negative elements of the Fiedler vector will belong to the second one. The tricky issue is what to do if the vector elements are equal zero. For details see e.g. [34].

3 Proof of Watanabe's Ugly Duckling Theorem

Proof (Proof of Theorem 1). The proof is quite simplistic, but we will present it for completeness of presentation. Let two objects O_k, O_l have attribute values $A_1 = a_{k1}, \ldots, A_n = a_{kn}$ and $A_1 = a_{l2}, \ldots, A_n = a_{ln}$ resp. Recall that all of the relevant attributes can be constructed as follows: Create all attributes of the form $B_p = *A_1 \wedge, \ldots, \wedge * A_n$ where * stands for either nothing or $\neg$. p would range from 1 to 2^n. Let B_k be the expression true for O_k and B_l the expression true for O_l. Note that there is only one expression B_p true for each of them and they are different. All other are false for both objects.

Let $\mathbb{B}$ be the set of all B_p. $2^{\mathbb{B}}$ be the set of all subsets of $\mathbb{B}$. $\mathbb{C}$ be the set of all expressions c derived as follows: for any set $b \in 2^{\mathbb{B}}$ let c be the disjunction $\vee$ of all elements from b. Then the complete set of all derived attributes of $A_1, \ldots, A_n$ is the set $\mathbb{C}$.

Consider a subset $\mathbb{C}_0$ of $\mathbb{C}$ of expressions containing neither B_k nor B_l, a subset $\mathbb{C}_k$ of $\mathbb{C}$ of expressions containing B_k but not B_l, a subset $\mathbb{C}_l$ of $\mathbb{C}$ of expressions not containing B_k but containing B_l, a subset $\mathbb{C}_2$ of $\mathbb{C}$ of expressions containing both B_k and B_l as elementary conjunctions. It is easily seen that $\mathbb{C}_0$, $\mathbb{C}_k$, $\mathbb{C}_l$, $\mathbb{C}_2$ are of the same cardinality.

Obviously, both objects have the same value for attributes from $\mathbb{C}_0$ (false) and $\mathbb{C}_2$ (true), but different for $\mathbb{C}_k$ and $\mathbb{C}_l$. Due to mentioned cardinalities $sim(O_k, O_l) = 0.5$. □

4 GSC Ugly Duckling Issues for Combinatorial Laplacians

For a set of m objects, let us consider a similarity matrix $S_{0.5,m}$ with values 0.5 at all off-diagonal elements. In a combinatorial Laplacian $L_{0.5,m}$, all off-diagonal elements will be -0.5, and all diagonal equal $0.5(m-1)$.

Theorem 2 *The system of eigenvectors of the combinatorial Laplacian $L_{0.5,m}$ can be obtained as follows. One eigenvalue, λ_1, of L will be equal 0, with eigenvector $\mathbf{v_1} = (-1, \ldots, -1)^T$. Other eigenvectors $i = 2, \ldots, m$ will have the form: $v_{i,j} = -1$ for $j \geq i$, $v_{i,j} = 0$ for $j \leq i-2$, $v_{i,i-1} = m - (i-1)$. Their corresponding eigenvalues are all equal $m/2$.*

Proof. The eigenvalue λ_1 and eigenvector $\mathbf{v_1}$ are common for all combinatorial Laplacians.

Consider the eigenvector $\mathbf{v_i}$ for $i = 2, \ldots, m$ and its multiplication with the $i-1^{st}$ row of $L_{0.5,m}$. $v_{i,i-1} \cdot L_{0,5;i-1,i-1} = (m-(i-1)) \cdot (m-1)/2$. For $j \geq i$ we have $v_{i,j} \cdot L_{0,5;i-1,i-1} = -1 \cdot (-0.5) = 0.5$. For $j \leq i$ we have $v_{i,j} = 0$, so it is not contributing. So in all $L_{0,5;i-1,*} = (m-(i-1)) \cdot (m-1)/2 + (m-i+1) \cdot 0.5 = (m-(i-1)) \cdot m/2 = m/2 \cdot v_{i,i-1}$.

Consider the eigenvector $\mathbf{v_i}$ for $i = 2, \ldots, m$ and its multiplication with the j^{th} row ($j \geq i$) of $L_{0.5,m}$. $v_{i,i-1} \cdot L_{0,5;j,i-1} = (m - (i-1)) \cdot (-1/2)$. $v_{i,j} \cdot L_{0,5;j,j} = (-1) \cdot (m-1)/2$. For $k < i-1$ $v_{i,k} = 0$, so no contribution. For $k \geq i, k \neq j$ $v_{i,k} \cdot L_{0,5;j,k} = (-1) \cdot (-1/2)$. So in all $(m-(i-1)) \cdot (-1/2) + (-1) \cdot (m-1)/2 + (m-i) \cdot 1/2 = (-1/2) \cdot (m-i+1+m-1-m+i) = -m/2 = m/2 \cdot v_{i,j}$.

Consider the eigenvector $\mathbf{v_i}$ for $i = 2, \ldots, m$ and its multiplication with the j^{th} row ($j \leq i-2$) of $L_{0.5,m}$. $v_{i,i-1} \cdot L_{0,5;j,i-1} = (m-(i-1)) \cdot (-1/2)$. For $k < i-1$ $v_{i,k} = 0$, so no contribution. For $k \geq i$ $v_{i,k} \cdot L_{0,5;j,k} = (-1) \cdot (-1/2)$. So in all $(m-(i-1)) \cdot (-1/2) + (m-i+1) \cdot 1/2 = (-1/2) \cdot (m-i+1-m+i-1) = 0 = m/2 \cdot v_{i,j}$. So we see that $\mathbf{v_i}$ is an eigenvector with the eigenvalue equal $m/2$.

We have still to demonstrate that this is a system, that is there are m distinct eigenvectors (which is obvious) and that they are orthogonal.

The orthogonality of $\mathbf{v_i}$ for $i \geq 2$ to $\mathbf{v_1}$ is obvious because the sum of elements of $\mathbf{v_i}$ is equal zero.

Consider $\mathbf{v_i}, \mathbf{v_j}$ for $i > j \geq 2$. $\mathbf{v_i}^T\mathbf{v_j} = 0$ because all elements of $\mathbf{v_i}$ with indices below $i-1$ are equal to zero and all elements of both $\mathbf{v_i}, \mathbf{v_j}$ with indices equal or bigger than $i-1$ are equal to minus one (thus their product sums up to $m-(i-1)$), while $v_{i,i-1} = m-(i-1)$ and $v_{j,i-1} = -1$ (with product $-(m-(i-1))$). □

Note that any vector $\mathbf{v}$ with sum of elements equal to zero is an eigenvalue of $L_{0.5,m}$ (see Theorem 5). This means, according to Fiedler's rule, clustering into two clusters can yield any result. This is obvious if all similarities are the same.

By the way the fact that there are multiple possible eigenvector systems if some eigenvalues are identical was already pointed at by us in [33] that if the eigenvalues are equal to each other then clustering should always take into account all the eigenvectors with the same eigenvalue.

For a set of m objects, let us consider a generalization for document similarity equal to p, and not just 0.5. Then we obtain similarity matrix $S_{p,m}$ with values p at all off-diagonal elements. In a combinatorial Laplacian $L_{p,m}$, all off-diagonal elements will be -p, and all diagonal equal $p(m-1)$.

Theorem 3 *The system of eigenvectors of the combinatorial Laplacian $L_{p,m}$ can be obtained as follows. One eigenvalue, λ_1, of L will be equal 0, with eigenvector $\mathbf{v_1} = (-1, \ldots, -1)^T$. Other eigenvectors $i = 2, \ldots, m$ will have the form: $v_{i,j} = -1$ for $j \geq i$, $v_{i,j} = 0$ for $j \leq i-2$, $v_{i,i-1} = m-(i-1)$. Their corresponding eigenvalues are all equal mp.*

Proof. Apparently, $L_{p,m} = 2pL_{0.5,m}$. Therefore, for an non-univalued eigenvector $\mathbf{v^{(0.5)}}$ for $L_{0.5,m}$

$$L_{p,m}\mathbf{v^{(0.5)}} = 2pL_{0.5,m}\mathbf{v^{(0.5)}} = 2pm/2\mathbf{v^{(0.5)}} = pm\mathbf{v^{(0.5)}}$$

that is each such vector is also an eigenvector of $L_{p,m}$ □

5 GSC Ugly Duckling Issues for Normalized Laplacian

For a similarity matrix $S_{p,m}$, its diagonal matrix $D_{p,m}$ will have all diagonal elements equal to $p(m-1)$. Hence

$$\mathcal{L}_{p,m} = D_{p,m}^{-1/2} L_{p,m} D_{p,m}^{-1/2} = D_{p,m}^{-1} L_{p,m} = L_{p,m}/(p(m-1))$$

This means that: In a normalized Laplacian $\mathcal{L}_{p,m}$, all off-diagonal elements will be -1/(m-1), and all diagonal equal 1. So $\mathcal{L}_{p,m}$ does not actually depend on p.

Therefore

Theorem 4 *The system of eigenvectors of the normalized Laplacian $\mathcal{L}_{p,m}$ can be obtained as follows. One eigenvalue, λ_1, of $\mathcal{L}$ will be equal 0, with eigenvector $\mathbf{v_1} = (-1, \ldots, -1)^T$. Other eigenvectors $i = 2, \ldots, m$ will have the form: $v_{i,j} = -1$ for $j \geq i$, $v_{i,j} = 0$ for $j \leq i-2$, $v_{i,i-1} = m-(i-1)$. Their corresponding eigenvalues are all equal $m/(m-1)$.*

Proof. For an non-univalued eigenvector $\mathbf{v^{(p)}}$ for $L_{p,m}$

$$\mathcal{L}_{p,m}\mathbf{v^{(p)}} = L_{p,m}/(p(m-1))\mathbf{v^{(p)}} = (pm)/(p(m-1))\mathbf{v^{(p)}} = m/(m-1)\mathbf{v^{(p)}}$$

□

We claim that

Theorem 5 *Any non-zero vector $\mathbf{v}$ with the sum of elements $\sum \mathbf{v} = 0$ is an eigenvector of $L_{p,m}$ and $\mathcal{L}_{p,m}$*

Proof. In any combinatorial Laplacian, in each row, the sum of off-diagonal elements is equal to negated diagonal element. The off-diagonal elements in $L_{p,m}$ are all equal p, and the diagonal element is equal to $(m-1)p$. Consider the i^{th} element v_i of $\mathbf{v}$. Let us multiply the i^{th} row of $L_{p,m}$. The sum of off-row element products amounts to $-(-v_i)p$ and that of ith element equals $v_i(m-1)p$. So the scalar product equals $v_i mp$. As this holds for each row, $L_{p,m}\mathbf{v} = mp\mathbf{v}$ which was to be demonstrated.

In the special case of $\mathcal{L}_{p,m}$, also the sum of off-diagonal elements is equal to negated diagonal element. So we conclude the same as in combinatorial Laplacian. □

6 Laplacians of blockdiagonal relations

So far we have considered the case when all attributes participate in the Ugly Duckling effect. For some reasons, however, one can introduce some restrictions on generalizations so that it applies for blocks of data only. In this case, the similarity matrix would have the form of a blockdiagonal matrix.

Assume now that we have blockdiagonal similarity matrix of the form

$$S_{p_1,p_2,m_1,m_2} = \begin{bmatrix} S_{p_1,m_1} & 0 \\ 0 & S_{p_2,m_2} \end{bmatrix}$$

representing similarities between documents in sets C_1, C_2 (S_{p_1,m_1} for C_1 blockdiagonal and S_{p_2,m_2} for C_2 blockdiagonal).

It is well-known that the eigenvalues of a blockdiagonal matrix are the eigenvalues of the individual blocks, and eigenvectors are the eigenvectors extended with zeros appropriately. So consider now the combinatorial Laplacian L_{p_1,p_2,m_1,m_2} of S_{p_1,p_2,m_1,m_2}. It will have the form

$$L_{p_1,p_2,m_1,m_2} = \begin{bmatrix} L_{p_1,m_1} & 0 \\ 0 & L_{p_2,m_2} \end{bmatrix}$$

Therefore

Theorem 6 *The system of eigenvectors of the combinatorial Laplacian L_{p_1,p_2,m_1,m_2} can be obtained as follows. Two eigenvalues, $\lambda_{1,1}, \lambda_{1,2}$, of L will be equal 0, with eigenvectors* $\mathbf{v_{1,1}} = (-1, \ldots, -1, 0, \ldots, 0)^T$. $\mathbf{v_{1,2}} = (0, \ldots, 0, -1, \ldots, -1)^T$. *Other eigenvalues are equal to $m_1 p_1$ ($m_1 - 1$ of them) and $m_2 p_2$ ($m_2 - 1$ of them). The corresponding eigenvectors $i_1 = 2, \ldots, m_1$ $i_2 = 2, \ldots, m_2$ will have the form: $v_{1,i_1,j} = -1$ for $j \geq i_1$, $v_{1,i_1,j} = 0$ for $j \leq i_1 - 2$ and $j \geq m_1$, $v_{1,i_1,i_1-1} = m_1 - (i_1 - 1)$. and $v_{2,i_2,j} = -1$ for $j \geq i_2 + m_1$, $v_{2,i_2,j} = 0$ for $j \leq i_2 - 2 + m_1$ $v_{2,i_2,i_2-1} = m_2 - (i_2 - 1)$.*

If we apply Fiedler's vector as a criterion for separating clusters now, we have a tricky task to distinguishing where elements with zero values belong to.

Consider now the normalized Laplacian $\mathcal{L}_{p_1,p_2,m_1,m_2}$ of S_{p_1,p_2,m_1,m_2}. It will have the form

$$\mathcal{L}_{p_1,p_2,m_1,m_2} = \begin{bmatrix} \mathcal{L}_{p_1,m_1} & 0 \\ 0 & \mathcal{L}_{p_2,m_2} \end{bmatrix}$$

Therefore

Theorem 7 *The system of eigenvectors of the normalized Laplacian $\mathcal{L}_{p_1,p_2,m_1,m_2}$ can be obtained as follows. Two eigenvalues, $\lambda_{1,1}, \lambda_{1,2}$, of $\mathcal{L}$ will be equal 0, with eigenvectors* $\mathbf{v_{1,1}} = (-1, \ldots, -1, 0, \ldots, 0)^T$. $\mathbf{v_{1,2}} = (0, \ldots, 0, -1, \ldots, -1)^T$. *Other eigenvalues are equal to $m_1/(m_1 - 1)$ ($m_1 - 1$ of them) and $m_2/(m_2 - 1)$ ($m_2 - 1$ of them). The corresponding eigenvectors $i_1 = 2, \ldots, m_1$ $i_2 = 2, \ldots, m_2$ will have the form: $v_{1,i_1,j} = -1$ for $j \geq i_1$, $v_{1,i_1,j} = 0$ for $j \leq i_1 - 2$ and $j \geq m_1$, $v_{1,i_1,i_1-1} = m_1 - (i_1 - 1)$. and $v_{2,i_2,j} = -1$ for $j \geq i_2 + m_1$, $v_{2,i_2,j} = 0$ for $j \leq i_2 - 2 + m_1$ $v_{2,i_2,i_2-1} = m_2 - (i_2 - 1)$.*

If we apply Fiedler's vector as a criterion for separating clusters, we have again a tricky task to distinguishing where elements with zero values belong to.

The analytical determination of eigenvectors and eigenvalues are easily extended to $k > 2$ blocks in the diagonal structure, that is to $S_{\mathbf{p},\mathbf{m}}$ and its Laplacians $L_{\mathbf{p},\mathbf{m}}$ and $\mathcal{L}_{\mathbf{p},\mathbf{m}}$, where $\mathbf{m}$ is a vector of block cardinalities and $\mathbf{p}$ is the vector of similarities within each block.

Then, however, the identifying of clusters requires application of k-means to k lowest eigenvalue eigenvectors as representation of the elements [21].

In theory, the lowest value eigenvectors (with eigenvalue equal to zero) mentioned in the above theorems are sufficient to identify uniquely the blocks.

7 Impact of noise

Let us consider the impact of fixed noise - adding the same fixed value $p^{[n]}$ to all elements of similarity matrix. This situation can be represented as a sum of two similarity matrices $S_{\mathbf{p},\mathbf{m}} + S_{p^{[n]},\sum \mathbf{m}}$ It is easily seen that also the combinatorial Laplacian can be represented analogously as $L_{\mathbf{p},\mathbf{m}} + L_{p^{[n]},\sum \mathbf{m}}$ It does not, however, apply to normalized Laplacian.

So consider the combinatorial Laplacian $L_{\mathbf{p},\mathbf{m}} + L_{p^{[n]},\sum \mathbf{m}}$. Consider a vector $\mathbf{v}$ consisting of $card(\mathbf{m})$ blocks each of length m_j with entries equal $1/m_i$ for the i block and $1/(\sum \mathbf{m} - m_i)$ for any other block $j \neq i$. Clearly $\sum \mathbf{v} = 0$, so according to Theorem 5, it is an eigenvector of $L_{p^{[n]},\sum \mathbf{m}}$ with eigenvalue $p^{[n]}(\sum \mathbf{m})$. But also due to its construction, it is an eigenvector of $L_{\mathbf{p},\mathbf{m}}$ with eigenvalue equal to zero. So in all, $\mathbf{v}$ is an eigenvector of $L_{\mathbf{p},\mathbf{m}} + L_{p^{[n]},\sum \mathbf{m}}$ with eigenvalue $p^{[n]}(\sum \mathbf{m})$. Though the vectors of this form do not constitute the set of orthogonal eigenvectors, they are all with the second lowest eigenvalue so that an orthogonal system of the space spanned by them can be found and used to clearly identify the blocks with a clustering algorithm.

8 Experimental evaluation

We investigated experimentally the capability of k-means algorithm used within GSC to properly identify the blocks in datasets with blockdiagonal similarity matrices in the category of Ugly Duckling similarity matrices $S_{\mathbf{p},\mathbf{m}}$.

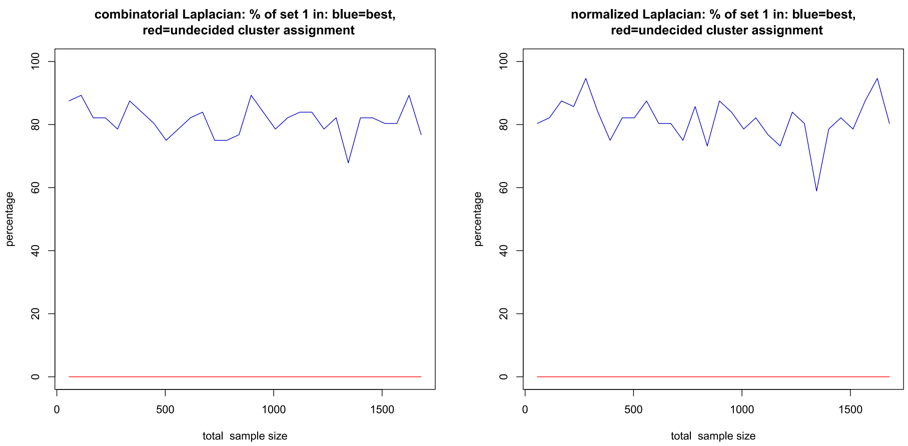

Figure 1. Percentage of correct clustering - combinatorial (to the left) and normalized (to the right) Laplacian based GSC for the block structured similarity matrices with block similarities **p**= (0.4, 0.7,0.4, 0.7,0.4, 0.7,0.4, 0.7,0.4, 0.7,0.4, 0.7,0.4, 0.7,0.4, 0.7) and proportions of block cardinalities $\mathbf{m}^{[ground]}$=(3,4,3,4,3,4,3,4,3,4,3,4,3,4,3,4)

We used the vector **p** listed in Fig. 1. As k-means algorithm is sensitive to unbalanced classes, we used block size equal to $j \cdot \mathbf{m}^{[ground]}$ for $j = 1, \ldots, 30$, where $\mathbf{m}^{[ground]}$ is given in caption of Fig. 1. Fig. 1 shows how successful the algorithm was in finding proper block structure for the various implied sample sizes. It shows the percentage of correctly clustered elements of the sample (the blue line). The red line shows the percentage of undecided sample elements which turned out to be equal zero everywhere[1]. Both combinatorial and normalized Laplacian based GSC had success rate between 80 and 100 %.

Next we performed similar experiments, but by adding noise. We did not use, however, fixed noise, discussed in Section 7, but rather a random noise, uniformly sampled for each

[1]Undecided cases are frequent if we would use Fiedler vector based decision making, which applies to two clusters only

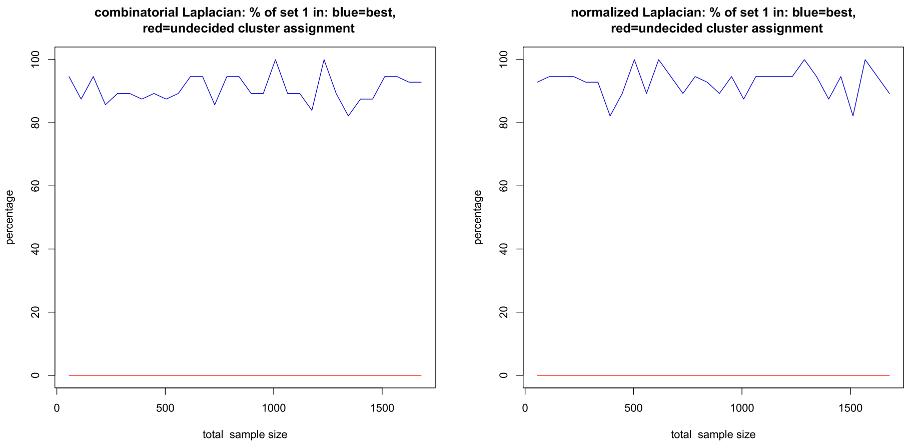

Figure 2. Same as in Fig. 1 with addition of noise 0.001.

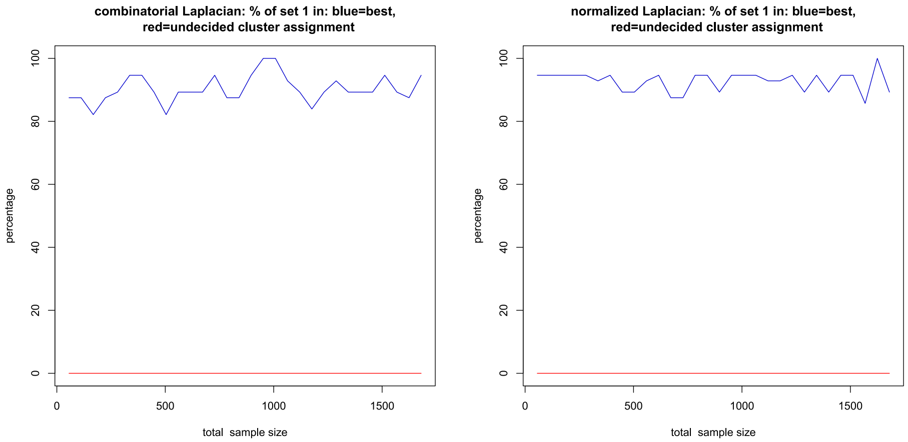

Figure 3. Same as in Fig. 1 with addition of noise 0.01.

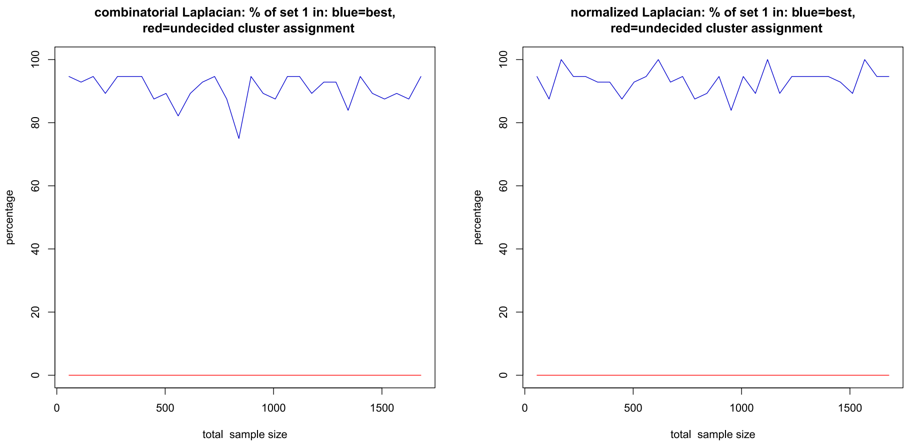

Figure 4. Same as in Fig. 1 with addition of noise 0.1.

similarity cell from the range 0 to 0.001 (Fig. 2), from the range 0 to 0.01 (Fig. 3), and from the range 0 to 0.1 (Fig. 4),

As could be expected from Section 7, the noise does not deteriorate clustering effectiveness, and in some cases even improves the clustering performance (by introducing more stability to eigenvectors).

We investigated the impact of k, the number of blocks, on the efficiency of clustering. The experiment was performed as follows: For the series of p, m values shown in caption of Fig. 5, first k were taken, for $k = 2, \ldots, 16$, the similarity block matrix was generated and then GSC was performed. Fig. 5 shows the results. For $k \leq 8$ in case of combinatorial Laplacians and for $k \leq 6$ in case of normalized Laplacians, the block structure was recovered correctly. For bigger values, the correctness level dropped. Same experiment was repeated in the presence of noise, from the range 0 to 0.001 (Fig. 6), from the range 0 to 0.01 (Fig. 7), and from the range 0 to 0.1 (Fig. 8), As the figures show, apparently the noise contributed to correctness improvement.

9 Conclusions

We have shown analytically that the Ugly Duckling effect identified by Watanabe, trivializes the Graph Spectral Clustering.

Furthermore, the experiments indicate that k-means seems to be a handicap of the Graph Spectral Clustering. We have pointed at this behaviour already in [25] for real world data. Here we have shown that the problem persists even for ideal blockdiagonal structures of the datasets.

As already indicated in [6], in our opinion it is not the blockdiagonal matrix that is essential for spectral clustering success but rather a different similarity structure within each block.

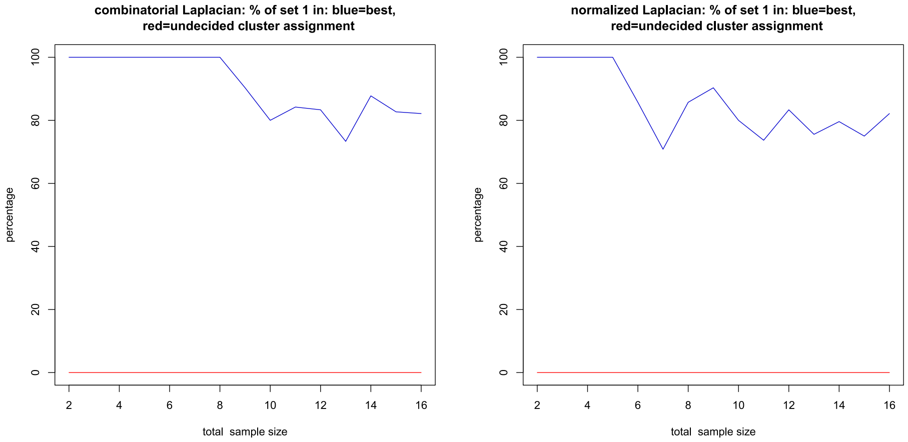

Figure 5. Percentage of correct clustering - combinatorial (to the left) and normalized (to the right) Laplacian based GSC for the block structured similarity matrices with block similarities **p**= (0.4, 0.7,0.4, 0.7,0.4, 0.7,0.4, 0.7,0.4, 0.7,0.4, 0.7,0.4, 0.7,0.4, 0.7) and proportions of block cardinalities $\mathbf{m} = 30 \cdot (3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4, 3, 4)$

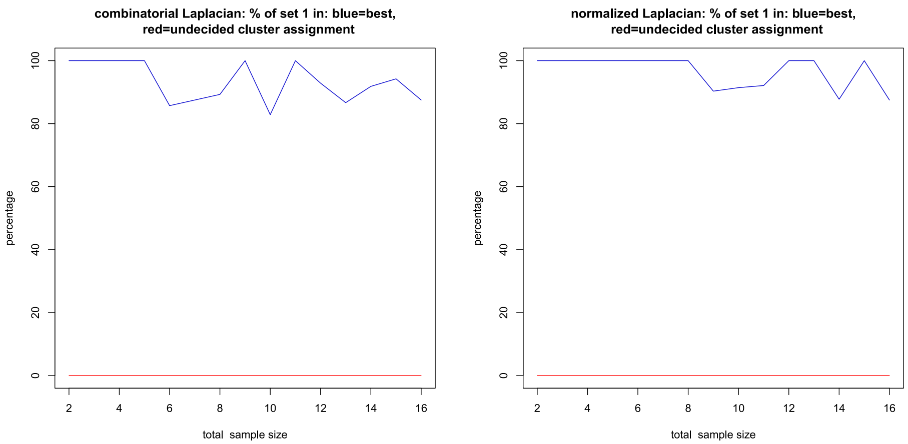

Figure 6. Same as in Fig. 5 with addition of noise 0.001.

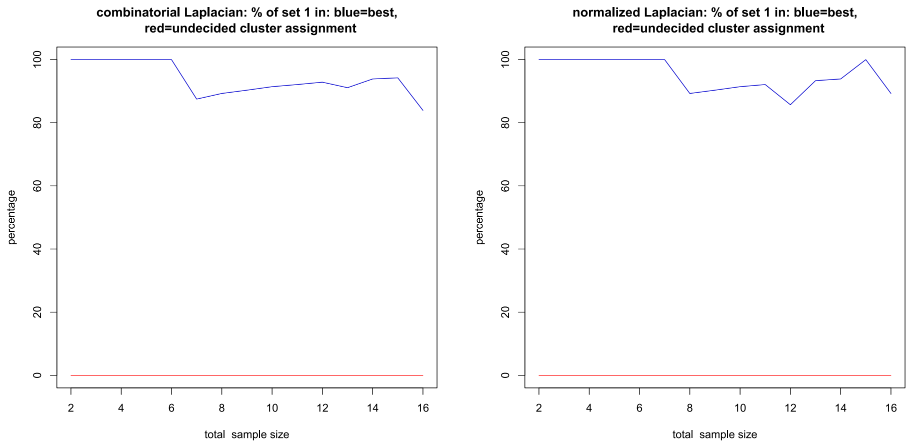

Figure 7. Same as in Fig. 5 with addition of noise 0.01.

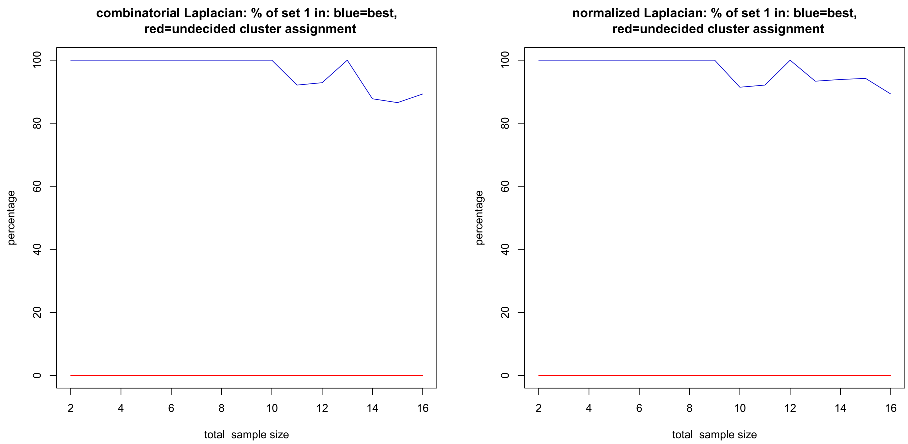

Figure 8. Same as in Fig. 5 with addition of noise 0.1.

This insight may be used to correct the way of thinking about spectral clustering as such and help find new application areas for it.

References

1. Alshammari, M., Takatsuka, M.: Approximate spectral clustering with eigenvector selection and self-tuned k. Pattern Recognition Letters **122**, 31–37 (may 2019). https://doi.org/10.1016/j.patrec.2019.02.006
2. Bartlett, S.J.: The species problem and its logic: Inescapable ambiguity and framework-relativity (2015)
3. Blumson, B.: Does everything resemble everything else to the same degree (2022), National University of Singapore; https://ap5.fas.nus.edu.sg/fass/phibrkb/uglyduckling.pdf
4. Blumson, B.: The metaphysical significance of the ugly duckling theorem. In: Australian National University, School of Philosophy Lectures (2014), https://philosophy.cass.anu.edu.au/events/ben-blumson-singapore-metaphysical-significance-ugly-duckling-theorem
5. Boedihardjo, M., Deng, S., Strohmer, T.: A performance guarantee for spectral clustering (2020)
6. Borkowski, P., Kłopotek, M.A., Starosta, B., Wierzchoń, S.T., Sydow, M.: Eigenvalue based spectral classification. PLoS ONE **18**(4), e0283413 (2023). https://doi.org/https://doi.org/10.1371/journal.pone.0283413
7. Caruana, R., Elhawary, M.F., Nguyen, N., Smith, C.: Meta clustering. In: Proceedings of the 6th IEEE International Conference on Data Mining (ICDM 2006), 18-22 December 2006, Hong Kong, China. pp. 107–118. IEEE Computer Society (2006). https://doi.org/10.1109/ICDM.2006.103, https://doi.org/10.1109/ICDM.2006.103
8. Chao, G., Sun, S., Bi, J.: A survey on multi-view clustering. IEEE Transactions on Artificial Intelligence **2**(2), 146–168 (2021)
9. Fisher, D.H.: Noise-tolerant conceptual clustering. In: Proceedings of the 11th International Joint Conference on Artificial Intelligence - Volume 1. p. 825–830. IJCAI'89, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (1989)
10. Fujimoto, T., Ikuine, F.: Industrial Competitiveness and Design Evolution. Springer Tokyo (01 2018). https://doi.org/10.1007/978-4-431-55145-4
11. Hatakeyama-Sato, K., Watanabe, S., Yamane, N., Igarashi, Y., Oyaizu, K.: Using gpt-4 in parameter selection of materials informatics: Improving predictive accuracy amidst data scarcity and 'ugly duckling' dilemma. ChemRxiv. Cambridge: Cambridge Open Engage (2023)
12. Hong, X., Gao, J., Wei, H., Xiao, J., Mitchell, R.: Two-step scalable spectral clustering algorithm using landmarks and probability density estimation. Neurocomputing **519**, 173–186 (2023). https://doi.org/https://doi.org/10.1016/j.neucom.2022.11.063
13. Ilachinski, A.: Cellular Automata: A Discrete Universe. World Scientific, Singapore (2001)
14. Kamishima, T., Akaho, S.: Considerations on recommendation independence for a find-good-items task. In: Proc. Fairness, Accountability and Transparency in Recommender Systems (08 2017). https://doi.org/10.18122/B2871W
15. Klopotek, M.A.: On the phenomenon of flattening "flexible prediction" concept hierarchy. In: Jorrand, P., Kelemen, J. (eds.) Fundamentals of Artificial Intelligence Research, International Workshop FAIR '91, Smolenice, Czechoslovakia, September 8-13, 1991, Proceedings. Lecture Notes in Computer Science, vol. 535, pp. 99–111. Springer (1991). https://doi.org/10.1007/3-540-54507-7_9, https://doi.org/10.1007/3-540-54507-7_9
16. Klopotek, M.A.: On seeking consensus between document similarity measures. Fundam. Informaticae **156**(1), 43–68 (2017). https://doi.org/10.3233/FI-2017-1597, https://doi.org/10.3233/FI-2017-1597

17. Klopotek, M.A., Matuszewski, A.: On irrelevance of attributes in flexible prediction. In: Proc. 2nd Int. Conf. on New Techniques and Technologies for Statistics (NTTS'95), vol. abs/2005.11979, pp. 282–293. GMD Sankt Augustin, Bonn, Germany (1995), https://arxiv.org/abs/2005.11979
18. Klus, S., Djurdjevac Conrad, N.: Koopman-based spectral clustering of directed and time-evolving graphs. J Nonlinear Sci **33**(8) (2023). https://doi.org/https://doi.org/10.1007/s00332-022-09863-0
19. Kłopotek, M.: Zależność funkcji oceny od współczynnika korelacji w metodzie formowania pojęć "flexible prediction". In: P. Sienkiewicz, J. Tchórzewski Eds.: Sztuczna Inteligencja i Cybernetyka Wiedzy (cybernetyka - inteligencja - rozwój, CIR'91), PTC, WSRP w Siedlcach, pp. 37–42. Siedlce (1991)
20. Liang, W., Zhou, S., Xiong, J., Liu, X., Wang, S., Zhu, E., Cai, Z., Xu, X.: Multi-view spectral clustering with high-order optimal neighborhood laplacian matrix. CoRR **abs/2008.13539** (2020), https://arxiv.org/abs/2008.13539
21. von Luxburg, U.: A tutorial on spectral clustering. Statistics and Computing **17**(4), 395–416 (2007). https://doi.org/https://doi.org/10.48550/arXiv.0711.0189
22. M.A.Kłopotek, S.T.Wierzchoń, K.Ciesielski, M.Dramiński, D.Czerski: Klasteryacja i metaklasteryzacja w świetle twierdzenia watanabe. In: Systemy Wspomagania Decyzji, pp. 83–97. Wydawnictwo Instytut Informatyki Uniwersytetu S'la'skiego (2012)
23. Ng, A.Y., Jordan, M.I., Weiss, Y.: On spectral clustering: Analysis and an algorithm. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS. pp. 849–856. MIT Press (2001), http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.19.8100
24. Peng, P., Yoshida, Y.: Average sensitivity of spectral clustering (2020)
25. Starosta, B., Kłopotek, M., Wierzchoń, S., Czerski, D.: Hashtag discernability – competitiveness study of graph spectral and other clustering methods. In: accepted for the 18th Conference on Computer Science and Intelligence Systems FedCSIS 2023 (IEEE #57573) Warsaw, Poland, 17–20 September, 2023 (2023)
26. Stone, J.V.: Vision and Brain: How We Perceive the World. MIT Press (2012)
27. Strehl, A., Ghosh, J.: Cluster ensembles — A knowledge reuse framework for combining multiple partitions. J. Mach. Learn. Res. **3**, 583–617 (2002), http://jmlr.org/papers/v3/strehl02a.html
28. Towster, E.: Two ugly duckling theorems for concept-formers. Information Sciences **8**(4), 359–368 (1975). https://doi.org/https://doi.org/10.1016/0020-0255(75)90047-X, https://www.sciencedirect.com/science/article/pii/002002557590047X
29. Wang, H., Zong, L., Liu, B., Yang, Y., Zhou, W.: Spectral perturbation meets incomplete multi-view data. In: Proc. of the 28-th Intl, Joint Conference on Artificial Intelligence (IJCAI-19). pp. 3677–3683 (2019)
30. Watanabe, S.: Theorem of the ugly duckling. In: Knowing and Guessing: A Quantitative Study of Inference and Information, pp. 376–377+. Wiley (1969)
31. Watanabe, S.: Pattern Recognition, Human and Mechanical. John-Willey and Sons, New York (1987)
32. Wen, J., Zhang, Z., Fei, L., Zhang, B., Xu, Y., Zhang, Z., Li, J.: A survey on incomplete multi-view clustering. IEEE Transactions on Systems, Man, and Cybernetics: Systems (Early Access) (2022)
33. Wierzchon, S.T., Klopotek, M.A.: Spectral cluster maps versus spectral clustering. In: Computer Information Systems and Industrial Management. LNCS, vol. 12133, pp. 472–484. Springer (2020). https://doi.org/10.1007/978-3-030-47679-3_40, https://doi.org/10.1007/978-3-030-47679-3_40
34. Wierzchoń, S., M.A.Kłopotek: Modern Clustering Algorithms, Studies in Big Data, vol. 34. Springer Verlag (2018)

35. Xu, Y., Srinivasan, A., Xue, L.: A Selective Overview of Recent Advances in Spectral Clustering and Their Applications, pp. 247–277. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-72437-5_12
36. Yang, Y., Wang, H.: Multi-view clustering: A survey. Big Data Mining and Analytics **1**(2), 83–107 (2018). https://doi.org/10.26599/BDMA.2018.9020003, https://www.sciopen.com/article/10.26599/BDMA.2018.9020003
37. Zhao, J., Xie, X., Xu, X., Sun, S.: Multi-view learning overview: Recent progress and new challenges. Information Fusion **38**, 41–54 (2017)