Reproducing the results of Understanding Regularized Spectral Clustering via Graph Conductance (Zhang and Rohe, 2018)

Cristian Bodnar
University of Cambridge
`cb2015@cam.ac.uk`
`Word count: 2407`

Abstract

Spectral clustering is a popular clustering algorithm used for community detection, manifold learning or image segmentation. However, spectral clustering can perform poorly in the absence of regularisation. Zhang and Rohe (2018) try to explain the impact of regularisation by exploiting the relationships between graph conductance and spectral clustering. This report reviews the main theoretical contributions of their paper and successfully reproduces their experiments. Further experiments are carried out to examine the importance of regularisation when partitioning graphs in more than two clusters and to observe the effects of the regularisation parameter τ.

1 Introduction

Spectral clustering is an algorithm for partitioning in k clusters the vertices $V = \{1, \dots, n\}$ of an undirected graph $\mathcal{G}(V, E)$ using leading eigenvectors of the Laplacian matrix of the graph. The properties of this matrix contain information about the structure of the graph that can be used to find a k-way cut inside it. The edges of the graph are given by a similarity matrix S describing how similar any two pairs of nodes are. More formally, $E = \{(i,j), S_{ij} > 0\} \subset V \times V$.

Before reviewing the paper of Zhang and Rohe [2018], a few useful definitions have to be introduced. $d_i = \sum_{j \in V} S_{ij}$ represents the *degree of node* i. The *volume* of a set of nodes $A \subseteq V$ is given by $vol(A) = \sum_{i \in A} d_i$. The *cut* for a non-empty partition $A \subset V$ is $cut(A, \mathcal{G}) = \frac{1}{2} \sum_{i \in A, j \in A^c} S_{ij}$. These notations are used to define the key concept of *conductance*.

Definition 1.1 (Conductance) *Let $A \subset V$ be a non empty subset of V with $vol(A, \mathcal{G}) < vol(A^c, \mathcal{G})$. The graph conductance of the subset A is $\phi(A, \mathcal{G}) = \frac{cut(A,\mathcal{G})}{vol(A,\mathcal{G})}$.*

Vanilla Spectral Clustering (VSC) is the algorithm which attempts to find an approximate set of partitions with minimal graph conductance. However, VSC was shown empirically to perform poorly in many cases [Amini et al., 2012]. On graphs with a dense core and many "dangling sets" at the periphery, VSC overfits and fails to find a cut at the core of the graph. To overcome this problem, Amini et al. [2012] add a regularisation term τ/n to the weights of the graph and consequently use a similarity matrix $[S_\tau]_{ij} = S_{ij} + \tau/n$. Regularised Spectral Clustering (RSC) partitions the nodes of the graph using the leading k eigenvectors of the (regularised) normalised Laplacian $L_\tau = I - D_\tau^{-\frac{1}{2}} S_\tau D_\tau^{-\frac{1}{2}}$ where D_τ is a diagonal matrix containing the degrees of the nodes of the regularised graph. A similar regularisation technique was proposed by Chaudhuri et al. [2012] and it uses a different Laplacian $L'_\tau = D_\tau^{-\frac{1}{2}} S D_\tau^{-\frac{1}{2}}$, where S_τ is replaced by S. This alternative method is also briefly compared to RSC in Section 4.3.

Zhang and Rohe [2018] exploit the relationships between graph conductance and spectral clustering to analyse the effects of RSC theoretically. Section 2 reviews the contribution of their paper. Section 3 reproduces their experiments and analyses the outcomes. Section 4 contains extended experiments that examine RSC with $k > 2$ and the effects of the parameter τ.

2 Review

The paper of Zhang and Rohe [2018] identifies "dangling sets" as the root cause for the problems of Vanilla Spectral Clustering (VSC) and shows how Regularised Spectral Clustering (RSC) tackles this issue. Formally, g-dangling sets are defined as follows:

Definition 2.1 (g-dangling set) *A g-dangling set is a subset $S \subset V$ such that the subgraph induced by S is a tree with g nodes and the tree is connected to S^{c} by exactly one edge.*

An immediate corollary of this definition is that the conductance of these sets $\phi(S, S^{\mathrm{c}}) = \frac{1}{2g-1} \approx (2g)^{-1}$ is low, since $volume(S) = 2(g-1) + 1$ and $cut(S, S^{\mathrm{c}}) = 1$. Because VSC minimises for the conductance of the partitions, this simple result hints at the sensitivity of VSC to g-dangling sets. The authors start from this insight to prove a series of theorems that show why g-dangling sets are a common vulnerability for VSC:

1. For any fixed g, the expected number of g-dangling sets of a random graph with N nodes and independent edges (also called an inhomogeneous random graph) is $\Theta(N)$. In other words, these kinds of graphs naturally posses a number of g-dangling sets proportional to N.
2. Because of the numerous g-dangling sets, there are $\Theta(N)$ eigenvectors with values less than $(g-1)^{-1}$ which conceals more balanced partitions at the core of the graph.
3. There are $\Theta(N)$ eigenvalues smaller than g^{-1} and this reduces the eigengap. Consequently, the numerical convergence for the computation of the eigenvalues and eigenvectors is slowed down.

It is important to note that inhomogenous random graphs are a generalisation of the Stochastic Block Model (SBM). SBM has been used extensively in recent research for studying community detection methods [Holland et al., 2003, Amini et al., 2012, Fishkind et al., 2013, Rohe et al., 2010, Karrer and Newman, 2011] even though it is not an accurate model for some real-world graphs. Most of these works use even more unrealistic assumptions such as the fact that the minimum degree of the graph grows by a polynomial power of $\log(n)$. A more relaxed assumption is used by Joseph and Yu [2014]. The advantage of the presented theorems is that they do not use the SBM assumption and therefore the results apply to a broader category of real-world graphs.

The last important proof in the paper shows that RSC significantly increases the conductance of small partitions in the regularised graph. This prevents RSC to overfit and suffer from the sensitivities of VSC. The experiments in the paper also support these theoretical considerations. The reproduced experiments are described in detail in the next section.

3 Reproducing the results

Section 3.1 compares the eigenvectors of the vanilla graph and the regularised graph. Section 3.2 looks at the effects of regularisation on a random graph with dangling sets. Section 3.3 empirically verifies the theoretical claims of the paper on real datasets.

3.1 Analysis of the eigenvectors

Very often, the eigenvectors corresponding to the smallest eigenvalues of the vanilla graph are not able to distinguish between core clusters as the magnitude of the eigenvector is concentrated around a few entries of the vector. However, in the regularised graphs, the magnitude of the eigenvector is spread over multiple components. This is illustrated in Figure 1.

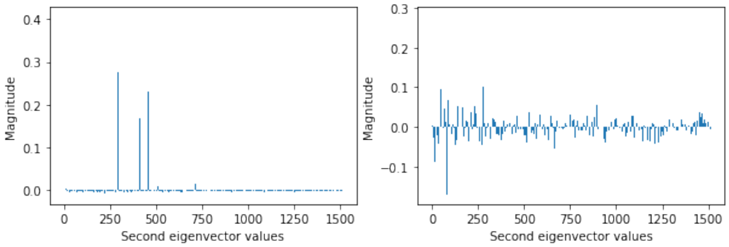
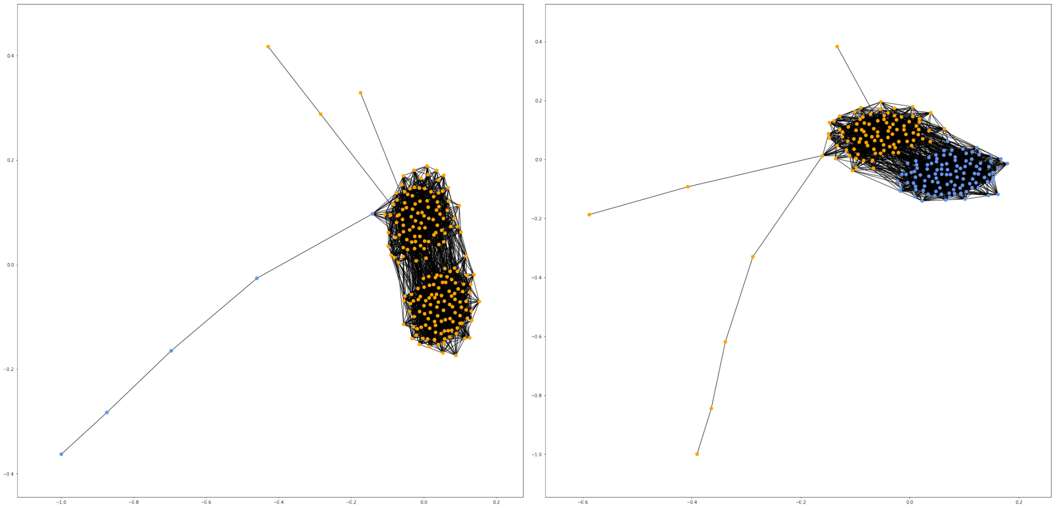


Figure 1: Eigenvectors of the vanilla graph (left) and regularised graph (right)



(a) Vanilla

(b) Regularised

Figure 2: Spectral clustering of a random graph with three dangling sets

3.2 A random graph with dangling sets

For directly observing the effects of regularisation, this experiment uses a random graph with two partitions whose size follow a Gaussian distribution. The probability of an edge inside the cluster is set to 0.4 and the probability of an edge between nodes in different clusters is set to 0.05. Three dangling sets are manually added to the graph. As it can be seen in Figure 2, only RSC is capable of finding a cut through the core of the graph.

3.3 Experiments on real graphs

To reproduce the results of Zhang and Rohe [2018], I used 16 real-world graphs containing up to 6000 nodes from the Stanford Large Network Dataset Collection [Leskovec and Krevl, 2014]: email-Eu-core, Facebook subgraph, gnutella06, gnutella08, gnutella09, 5 graphs from as733, ca-GrQc, twitter, Facebook TV Shows, Facebook Politicians, Facebook government, wiki-Vote.

For each of the graphs, the largest connected component is chosen. As in the paper, I randomly select half of the edges of the largest connected component and place them in a testing graph, while the rest of them are used in a training graph. On the training graph, the largest connected component is identified again, and this subgraph is used in the experiments. All the figures have on their x axis values corresponding to RSC and on the y axis values corresponding to VSC. The line $x = y$ is also plotted in black. The plot markers have sizes proportional to the number of nodes in the dataset.

One of the theoretical results of the paper is that RSC does not assign low conductance to dangling sets in the regularised graph. This is also supported by empirical evidence. As Figure 3 shows, for the vast majority of the graphs used in the evaluation, RSC finds for almost all graphs clusters that are more balanced. This is consistent with the findings of Zhang and Rohe [2018].

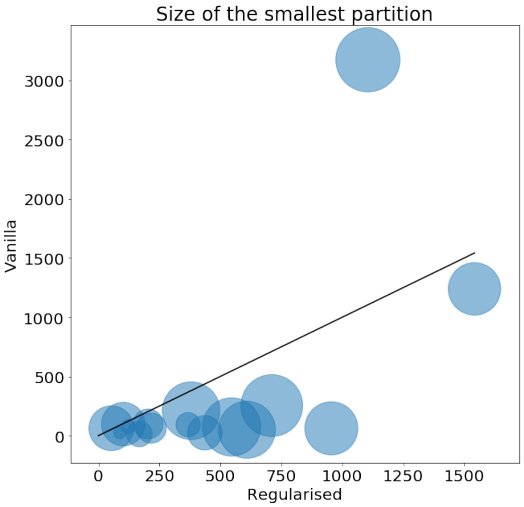
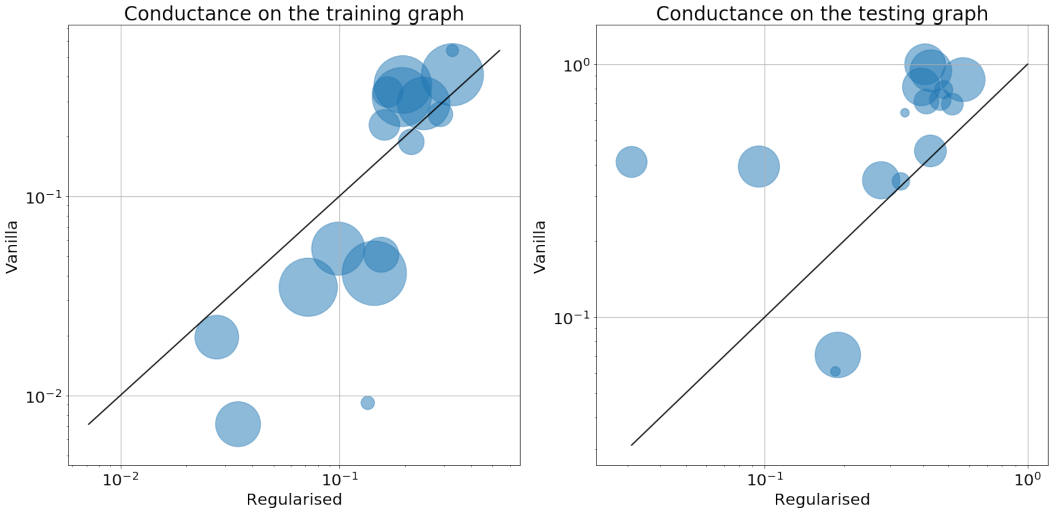


Figure 3: RSC finds more balanced partitions than VSC.

The conductance of the partitions found by VSC and RSC on the training and testing datasets are a good indicator of the tendency of VSC to overfit by finding small partitions at the periphery of the network. As figure 8 shows, VSC generally finds partitions with a lower conductance than RSC on the training graph. This is to be expected since VSC directly minimises for the conductance. However, on the testing graph, the conductance of the partitions found by VSC is much higher than of the RSC partitions. Furthermore, the phenomenon described by the authors as “catastrophic overfitting” is encountered. Many of the VSC partitions on the testing graph have conductance one or very close to one. This means that these partitions have (almost) no internal edges.



(a) Conductance on training graph

(b) Conductance on testing graph

Figure 4: VSC overfits on the training graphs.

The last experiment in the paper aims to demonstrate the speed up in the computation of the leading eigenvectors of the regularised Laplacian. For this experiment, I use the *scipy.sparse.linalg.eigsh* method that is based on the ARPACK package. The processing time for all the graphs is recorded and plotted in Figure 5. The library is faster by a constant factor on the regularised Laplacian.

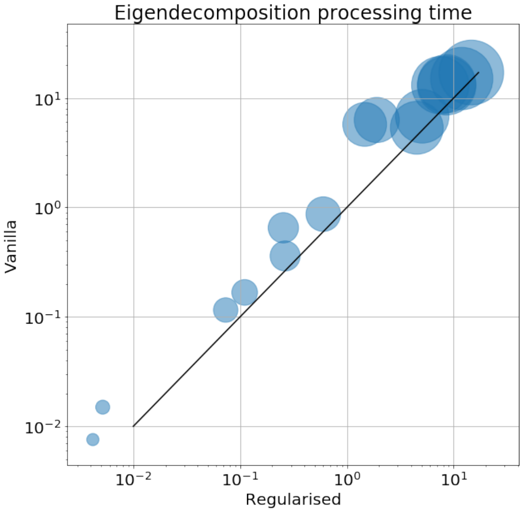


Figure 5: Computing the eigenvalues and eigenvectors is faster in the regularised graph than in the vanilla graph.

4 Further experiments

In addition to the reproduced experiments, this section includes further experiments on regularised spectral clustering. Section 4.1 examines the effect of regularisation when partitioning a graph in more than two clusters. Section 4.2 analyses the impact of the regularisation parameter τ. Section 4.3 examines τ in the context of classification and also evaluates the RSC method of Chaudhuri et al. [2012]. The last section of this part looks at an example where the default initialisation of τ with the average degree of the graph does not perform well.

4.1 Regularisation on more than two clusters

For these experiments, I use a power-law cluster model to generate a graph with many clusters at the core and many dangling sets at the periphery. This model is an extension of the Barabási-Albert (BA) [Albert, 2001]. It generates scale-free networks (networks whose node degrees follow a power law distribution). Many real-world graphs such as the Internet, citation networks or social networks are scale-free networks, and therefore, this model is a useful choice as it is representative of many real-world graphs. The network used for evaluation has 2000 nodes. During the generation process, each new node adds two random edges to existing nodes and the probability of forming a triangle for each of the new edges is 0.05. This graph is illustrated in Figure 6.

To investigate the balance of the clusters for $k > 2$, VSC and RSC are applied repeatedly on the random graph with $k \in \{2, \dots, 20\}$ and the variance of the partition sizes are recorded. Figure 7 plots the standard deviation of the cluster sizes for the VSC and RSC clusters. For small values of k, there is a very high gap between these variations, but the gap gradually reduces as the number of clusters increases. This demonstrates that the RSC clusters are significantly more balanced up until some k and the difference reduces with larger k. This is not surprising given that the clusters become smaller and smaller as k increases.

The second experiment examines the overfitting of VSC. VSC and RSC are applied repeatedly on the training and testing graphs for the same range of k and the average conductance of the returned partitions is recorded. This data is plotted in Figure 8. As in the particular case $k = 2$, on the training graph, VSC generally finds partitions with lower conductance, but these partitions usually have higher conductance on the testing dataset. This discrepancy is evident for smaller k. As in the previous experiment, the gap between the VSC and RSC testing conductance gradually reduces as k increases. For some values of k such as 12, RSC has slightly higher conductance on the test graph.

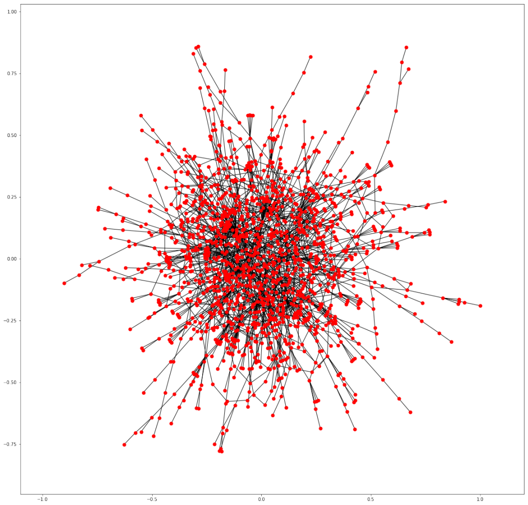


Figure 6: Graph generated by a powerlaw cluster model.

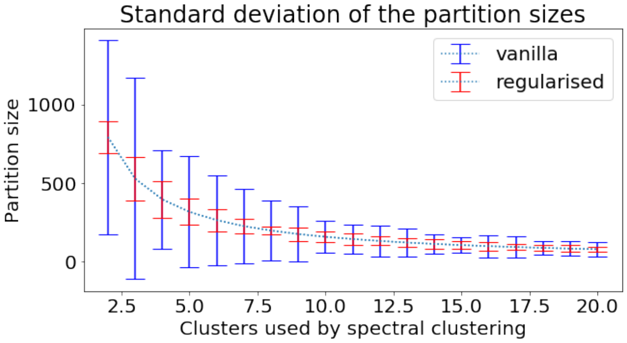


Figure 7: The partitions generated by RSC are more balanced than the ones generated by VSC.

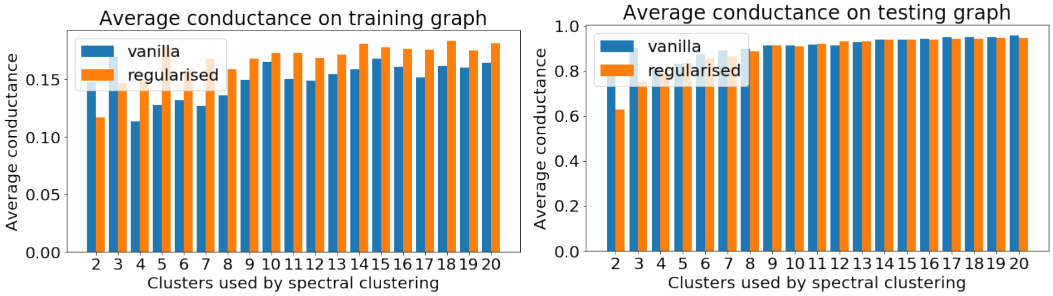
4.2 The effects of the regularisation parameter

The paper of Zhang and Rohe [2018] doesn't examine the tuning of the parameter τ and, as in the experiments described so far, τ is initialised with the average degree of the nodes of the graph. The following experiments analyse the effects of τ on the size of the smallest partition and the conductance when $k = 2$. The same power law cluster graph as in the previous subsection is used. The behaviour described in this subsection was similar on the other (random and real) graphs I experimented with.

A common pattern in all these experiments is that the effects of τ are roughly linear in the regions with extreme values, but nonlinear for the intermediate values. Figure 9a shows that for intermediate values of τ there are multiple peaks for the size of the smallest partition. In the limit, the size of the smallest partition flattens, but it remains higher than the one of the vanilla graph. The minimum conductance is also achieved for an intermediate value as shown by Figure 9b. The vertical red line represents the heuristic value of τ.

4.3 Regularisation and classification performance

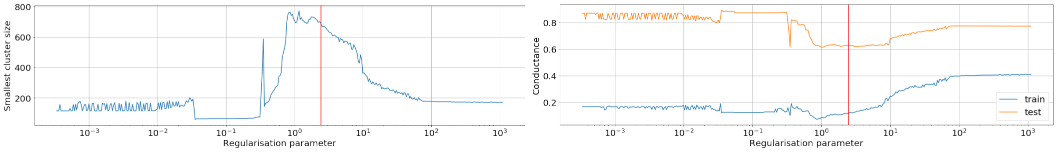
Another interesting experiment is to test the effects of the regularisation parameter for spectral clustering based classification. To do this, I use the political blogs dataset from the 2004 US elections [Adamic and Glance, 2005] where each of the nodes represents blogs with either conservative or liberal views and the email-Eu-core dataset that contains emails from 42 departments of a research institute. For both datasets, I also compare the results of RSC with the alternative method proposed by Chaudhuri et al. [2012] (referred to as RSC Chaudhuri in the figures).



(a) Average conductance on training graph

(b) Average conductance on testing graph

Figure 8: The benefits of regularisation gradually reduce as k increases.



(a) The size of the smallest cluster as a function of τ

(b) Conductance as a function of τ

Figure 9: The behaviour of the partition size and conductance is linear for the extreme values of τ and non linear for the intermediate value

Figure 10a plots the accuracy on the political blog dataset as a function of τ. Without regularisation, VSC achieves an accuracy of 52%, while RSC with $\tau = 0.5$ achieves 95%. It is also important to note that the heuristic value of τ achieves 80% accuracy. For large values of τ, the accuracy stabilises above 70% as the Laplacian becomes more and more similar to a constant matrix. This is consistent with the findings of Joseph and Yu [2014]. The step function behaviour of the accuracy in the first half of the plot is also interesting. This shows that tiny fluctuations in τ can drastically affect performance. Figure 10a also plots the accuracy obtained by the regularisation method of Chaudhuri et al. [2012]. This method seems to be more robust in changes of the parameter τ as it maintains an accuracy of 95% in the limit.

Figure 10b plots the Normalised Mutual Information (NMI) between the actual 42 labels of the dataset and the labels found by RSC and the algorithm of Chaudhuri et al. [2012]. In this case, regularisation brings marginal improvements over VSC and the NMI score is lower compared to VSC when the heuristic value of τ is used (indicated by the red line in the figures). This is consistent with the finding from the previous section that the benefits of regularisation diminish with the increase in k. As in the previous experiment, the regularisation method of Chaudhuri et al. [2012] has better performance in the limit as it stabilises around a higher NMI value.

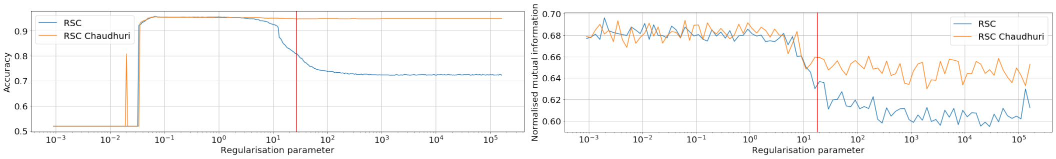
4.4 The danger of the heuristic regularisation parameter

On the political blogs dataset, τ initialised with the average degree of the graph brings a 28% accuracy boost over VSC. However, this is not always the case. Figure 11 compares the NMI score obtained by VSC and RSC (with heuristic τ) when partitioning the email-Eu-core dataset in clusters ranging from 2 to 42. For all the values of k, VSC obtains a higher NMI score.

5 Conclusion

This report reviews the contributions in the paper of Zhang and Rohe [2018] and successfully reproduces their results. Moreover, the extended experiments empirically demonstrate that:

- The benefits of regularisation diminish as the number of clusters increases.
- The effects of the regularisation parameter are nonlinear for intermediate values of τ, but the observed behaviour stabilises for the extreme values.



(a) Accuracy on the political blog dataset

(b) NMI on the email-Eu-core dataset

Figure 10: The impact of τ for two classification tasks: 2 clusters (left) and 42 clusters (right)

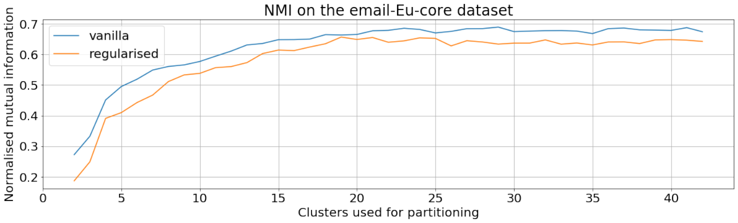


Figure 11: Comparison of the NMI score for VSC and RSC at various values of k. The curve of RSC is always below the one of VSC.

- Small variations of τ can cause large changes in performance.
- The regularisation method of Chaudhuri et al. [2012] empirically has better performance than RSC for large values of τ on two classification tasks.
- The initialisation of τ with the average degree of the graph might not always be good enough.

References

Lada A. Adamic and Natalie Glance. The political blogosphere and the 2004 u.s. election: Divided they blog. In *Proceedings of the 3rd International Workshop on Link Discovery*, LinkKDD '05, pages 36–43, New York, NY, USA, 2005. ACM. ISBN 1-59593-215-1. doi: 10.1145/1134271.1134277. URL `http://doi.acm.org/10.1145/1134271.1134277`.

Reka Zsuzsanna Albert. *Statistical Mechanics of Complex Networks*. PhD thesis, Notre Dame, IN, USA, 2001. AAI3000268.

Arash A. Amini, Aiyou Chen, Peter J. Bickel, and Elizaveta Levina. Pseudo-likelihood methods for community detection in large sparse networks. 2012.

Kamalika Chaudhuri, Fan Chung Graham, and Alexander Tsiatas. Spectral clustering of graphs with general degrees in the extended planted partition model. In *COLT*, 2012.

Donniell E. Fishkind, Daniel L. Sussman, Minh Tang, Joshua T. Vogelstein, and Carey E. Priebe. Consistent adjacency-spectral partitioning for the stochastic block model when the model parameters are unknown. *SIAM J. Matrix Analysis Applications*, 34:23–39, 2013.

Paul W. Holland, Seroice, Kathryn Blackmond, and Samuel Leinhardt. Stochastic blockmodels: First steps *. 2003.

Antony Joseph and Bin Yu. Impact of regularization on spectral clustering. *2014 Information Theory and Applications Workshop (ITA)*, pages 1–2, 2014.

Brian Karrer and Mark E. J. Newman. Stochastic blockmodels and community structure in networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 83 1 Pt 2:016107, 2011.

Jure Leskovec and Andrej Krevl. SNAP Datasets: Stanford large network dataset collection. `http://snap.stanford.edu/data`, June 2014.

Karl Rohe, Sourav Chatterjee, and Bin Yu. Spectral clustering and the high-dimensional stochastic blockmodel. 2010.

Yilin Zhang and Karl Rohe. Understanding regularized spectral clustering via graph conductance. 06 2018.