White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/146970/

Version: Accepted Version

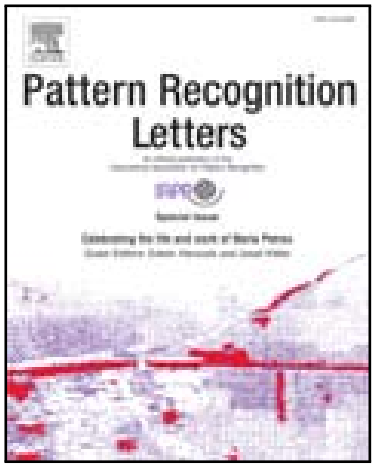
Article:
Curado, Manuel, Lozano, Miguel A., Escolano, Francisco et al. (2019) Dirichlet Densifier Bounds: Densifying Beyond the Spectral Gap Constraint. Pattern Recognition Letters. pp. 425-431. ISSN: 0167-8655

https://doi.org/10.1016/j.patrec.2019.06.001

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Accepted Manuscript



Dirichlet Densifier Bounds: Densifying Beyond the Spectral Gap Constraint

Manuel Curado, Miguel A. Lozano, Francisco Escolano,
Edwin R. Hancock

PII: S0167-8655(19)30169-2
DOI: https://doi.org/10.1016/j.patrec.2019.06.001
Reference: PATREC 7534

To appear in: *Pattern Recognition Letters*

Received date: 28 January 2019
Revised date: 27 April 2019
Accepted date: 2 June 2019

Please cite this article as: Manuel Curado, Miguel A. Lozano, Francisco Escolano, Edwin R. Hancock, Dirichlet Densifier Bounds: Densifying Beyond the Spectral Gap Constraint, *Pattern Recognition Letters* (2019), doi: https://doi.org/10.1016/j.patrec.2019.06.001

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Highlights

- We study the universal bounds of our densifier through new bounds for CTs estimation
- We explain how works our method where constrain the spectral gap cannot fully explain
- We show the deep implications of graph densification in commute times estimation

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Pattern Recognition Letters
journal homepage: www.elsevier.com

Dirichlet Densifier Bounds: Densifying Beyond the Spectral Gap Constraint

Manuel Curado[a,**], Miguel A. Lozano[b], Francisco Escolano[b], Edwin R. Hancock[c]

[a]*Department of Technology, Catholic University of Avila, 05005 Avila, Spain*
[b]*Department of Computer Science and AI, University of Alicante, 03690 Alicante, Spain*
[c]*Department of Computer Science, University of York, York YO10 5GH, UK*

ABSTRACT

In this paper, we characterize the universal bounds of our recently reported Dirichlet Densifier. In particular we aim to study the impact of densification on the bounding of intra-class node similarities. To this end we derive a new bound for commute time estimation. This bound does not rely on the spectral gap, but on graph densification (or graph rewiring). Firstly, we explain how our densifier works and we motivate the bound by showing that implicitly constraining the spectral gap through graph densification cannot fully explain the cluster structure in real-world datasets. Then, we pose our hypothesis about densification: a graph densifier can only deal with a moderate degradation of the spectral gap if the inter-cluster commute distances are significantly shrunk. This points to a more detailed bound which explicitly accounts for the shrinking effect of densification. Finally, we formally develop this bound, thus revealing the deeper implications of graph densification in commute time estimation.

© 2019 Elsevier Ltd. All rights reserved.

1. Introduction

Graphs have been used in a variety of different problems pattern recognition fields (31)(16)(4)(30)(28). However, some graph analysis problems pose significant problems due to their excessive sparsity, i.e. low edge density. One way to overcome these problems is to apply edge densification as preconditioning operation before subsequent pattern recognition tasks are attempted.

Graph Densification is a technique from extremal graph theory which was originally formalized by Hardt and coworkers (17) as a means of ruling out non-trivial graph embeddings. Here the problem was posed as a constrained optimization problem driven by cut preservation. They proved that a graph can be densified if and only if it cannot be embedded under a weak notion of embeddability. This formally poses densification as the principled study of how to significantly increase the number of edges of a given input graph $G = (V, E)$ by generating a new graph $H = (V, E')$, where $E \subset E'$, which approximates G with respect to a given test function. One concrete example is whether there exists a given cut within the two graphs, and the cuts in G are preserved (or bounded) to some extent in H. Thus it is possible to generate an input graph so that the subsequent pattern recognition task is better conditioned. Moreover, the problem is also of interest in graph-based manifold learning where the input graphs (typically kNN or Gaussian) are very sparse.

Originally, this characterization was motivated by the need to understand structural differences between sparse graphs and dense graphs in order to reduce the complexity of certain combinatorial problems. The aim here is to take advantage of the fact that certain NP-hard problems have a Polynomial Time Approximation Scheme when their associated graphs are dense. This is the case for the MAX-CUT problem (3). Frieze and Kannan (15) raise the question of whether this computational "easiness" can be explained by the Szemerédi Regularity Lemma, which states that very large dense graphs have many of the properties of random graphs (18). Moreover, any sufficiently large (dense) graph can almost entirely be partitioned into a bounded number of random-like graphs, which are bipartite. In this case, there are procedures (algorithms) that can be used to test whether a graph can be partitioned (2). However, they are usually conditioned by a tower-exponential condition. This method represents a link between *extremal graph theory* and structural pattern recognition. Extremal graph theory concerns

**Corresponding author: Tel.: +34 653164473;
e-mail: manuel.curado@ucavila.es (Manuel Curado)

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

the *existence* of particular graphs satisfying certain test functions or properties (5). Despite the fact that extremal graph theory contains many interesting combinatorial tools, such as the Ramsey Theory, it is typically axiomatic, i.e. non-procedural.

Existing graph densification procedures such as the construction of anchor graphs (19), rely on semi-definite programming (SDP) and they can only deal with very small graphs in practice (17). Their formulation is thus too simple to preserve global information in realistic situations, and SDP solvers are polynomial in the number of unknowns (21).

However, the link between densification and Commute Times was firstly explored in (12), where we highlighted the fact that densification leads to a shrinkage of the inter-cluster distances, thus making Commute Times meaningful in large graphs. Later on, in (13), we further highlighted the fact that state-of-the-art densifiers rely on semi-definite programming and motivate a novel algorithm, which is more scalable and robust. The core of this algorithm is harmonic analysis. To develop the mathematical machinery for this study, we commence by exploring the link between the Cheeger constant (7), the spectral gap (9), the heat kernel trace on the Laplacian matrix (29), and the commute distances (27). The solution is to introduce the concept of graph densification, and specifically its formulation as a constrained optimization problem in which cuts are to some extent preserved in the densified graph. This exploits the fact that densification often requires cut preservation, in order to conjecture that densified graphs can be better conditioned for spectral clustering than their un-densified counterparts. In (12) we highlighted the fact that densification leads to a shrinkage of the inter-cluster distances, thus making Commute Times meaningful in large graphs. Later on, in (13), we highlighted the fact that state-of-the-art densifiers rely on semi-definite programming and motivate a novel algorithm, which is more scalable and robust. The core of this algorithm is harmonic analysis. Moreover, in (14) we applied the resulting densification technique to preprocessing large graphs so that they become better conditioned and more tractable for compression and decompression.

More recently, and guided by the insights provided by this initial work, we developed a densifier which minimises the combinatorial Dirichlet integral (8). The so-called Dirichlet densifier further exploits the link between densification and Commute Times, and highlights the fact that densification leads to a shrinkage of the inter-cluster distances, thus making Commute Times meaningful in large graphs. This method increases the edge density in undirected graphs, which are more suitable for estimating meaningful commute times by minimizing the Cheeger constant (and thus the spectral gap). It is both a more scalable and a more effective method than that based on semidefinite programming (22), and it is completely unsupervised. However, the relationship between densification and the spectral gap constraint is not thoroughly explored.

To develop the mathematical machinery to study the relationship in more depth, in this paper we commence by exploring the link between: a) the Cheeger constant (6), b) the spectral gap (10) and c) commute distances (26). In this regard, it is well known that commute times suffer from the problem of *global information loss*. More precisely, von Luxburg et al. (24) showed that commute times are diffused through a graph in such a way that the local part of the diffusion (in the neighbourhood of both the origin and destination nodes) dominates the global one (inside the graph). Since this behaviour is consistent with the preservation of bottlenecks, we establish a link with the minimization of graph conductance Φ (or Cheeger constant). Minimizing or constraining the graph conductance leads us to constrain the spectral gap λ_2, since $\lambda_2 \leq 2\Phi$.

Based on our prior work, our working hypothesis in this paper is that densification provides an effective way to obtain more clustered subgraphs so that the commute times can be shrunk for inter-cluster nodes. This allows a more effective estimation of the intra-cluster distances, so that they cannot be confused with larger inter-cluster distances. However to achieve this goal in a controlled manner, we need tighter bounds for commute time estimation than that which relies on the usual bound which is based on constraining the Cheeger constant and thus the spectral gap.

The outline of this paper is as follows. We commence by reviewing our earlier Dirichlet densification algorithm, which typically doubles the number of edges with respect to the original graphs. We then analyze von Luxburg et al.'s bound, which relies on the spectral gap (and Cheeger constant) and presents some practical limitations. This observation motivates a more detailed analysis, and this allows us to introduce a novel bound for commute time estimations which we refer to as scaled effective resistance estimation which we study in depth.

2. Contributions

In this paper, we characterize the universal bounds of our Dirichlet Densifiers (upper and lower bounds), studying the impact of the densification in the bounding of intra-class node similarities. To commence in Section 3 we review our updated Dirichlet Densifier algorithm. In Section 4, we derive a new bound for commute time estimation. This bound does not rely on the spectral gap but on graph densification (or graph rewiring). Firstly, we explain how our densifier works and we motivate the bound by showing that implicitly constraining the spectral gap through graph densification cannot fully explain some estimation effects in real-world datasets. We present different experiments which compare the densifications their corresponding spectral gaps for several datasets. Then, we present and analyse our hypothesis concerning densification: if our densifier can deal with a moderate degradation of the spectral gap, then this is due to the fact that the inter-cluster commute distances are significantly shrunk. This points to a more detailed bound which explicitly accounts for the shrinking effect of densification. Finally, we formally develop this bound, thus uncovering the deeper implications of graph densification in commute time estimation, and to lead to a change of concept in densification. We present our conclusions in Section 5.

3. Dirichlet Densifiers

In (13) we develop a novel densifier, which infers new intra-class edges while minimizing the number of new inter-class

ACCEPTED MANUSCRIPT

edges. To this end, we proceed to design a structural filter, using Return Random Walks (RRW), and then we build the line graph and run a Dirichlet process on it. Our algorithm is updated, showing that the RRWs implement a weighted diffusion process, and this process minimizes the probability that a random walk starting and ending at a given node traverses the inter-class links. The resulting weighting matrix W' is denser and more clustered than that associated with the input graph. The so called Dirichlet approach consists of the following steps:

1. **Generate a Knn-graph**: Given a data set $\chi = \{\vec{x}_1, ..., \vec{x}_n\} \subset \mathbb{R}^d$, we map the $\vec{x}_i$ to the vertices V of an undirected weighted graph $G(V, E, W)$ with $W_{ij} = e^{-\|\vec{x}_i - \vec{x}_j\|^2/\sigma^2}$ and $(i, j) \in E$ if $W_{ij} > 0$ and $j \in N_k(i)$.
2. **Apply Return Random Walk algorithm:** Given $G = (V, E, W)$ reformulate W in terms of W' so that
$$W'_{ij} = \max_k \max_{\forall l \neq k} \{p_{v_k}(v_j|v_i)p_{v_l}(v_i|v_j)\} , \quad (1)$$
where $p_{v_k}(v_j|v_i) = \frac{W_{ik}W_{kj}}{d(v_i)d(v_j)}$, $p_{v_l}(v_i|v_j) = \frac{W_{jl}W_{li}}{d(v_j)d(v_i)}$ (*go* and *return* probabilities, respectively) and $d(.)$ is the degree function. Therefore, W'_{ij} relies on maximizing the probability that a random walk goes from i to j through l and then returns through a different vertex k. This strategy minimizes the weight of spurious inter-class links. Our strategy includes a filtering of W' to reduce inter-class noise, considering the relationship between the shortest path and the sum of different weights of the algorithm.
3. **High-level Edge Selection:** Given $G' = (V, E, W')$, select the highest weighted edges $E'' \subset E$, with $|E"| \ll |E|$ as follows:
   a) $\mathcal{S} = \textbf{sort}(E, W'_e, descend)$.
   b) $\mathcal{S}' = \mathcal{S} \sim \{e \in \mathcal{S} : W'_e < \delta_1\}$ where δ_1 is set so that $|\mathcal{S}'| = \alpha|\mathcal{S}|$.
4. **Construct the Line Graph** Given $G'' = (V, \mathcal{S}', W')$ generate the corresponding graph as follows:
$$Line = (\mathcal{S}', Line_E, Line_W)$$
where
   a) The nodes of $e_i \in Line$ are the edges in $\mathcal{S}'$.
   b) The weight function $Line_W$ is defined as follows:
$$Line_W(e_a, e_b) = \sum_{k=1}^{|E"|} p_{e_k}(e_b|e_a)p_{e_k}(e_a|e_b) , \quad (2)$$
   i.e. we use *go* and *return* probabilities.
   c) $Line_E = \{(e_a, e_b) : Line_W(e_a, e_b) > 0\}$
5. **Dirichlet Densification Process:** Given the *Line* graph, we proceed as follows:
   a) $\mathcal{SB} = \textbf{sort}(\mathcal{S}', Line_W, descend)$.
   b) $\mathcal{SB}' = \mathcal{SB} \sim \{e \in Line_E : Line_W < \delta_2\}$ where δ_2 is set so that $|\mathcal{SB}'| = \beta|\mathcal{SB}|$.
   c) Consider $\mathcal{SB}'$ as the boundary B (known labels) of a Dirichlet process driven by the Laplacian $Line_{\mathcal{L}} = Line_D - Line_W$. Then, finding an harmonic function, i.e. a function $u(.)$ satisfying $\nabla^2 u = 0$ consists of minimizing:
$$D_{Line}[u] = \frac{1}{2}u^T Line_{\mathcal{L}} u \quad (3)$$
   where $u = [u_B, u_I]$ and $Line_{\mathcal{L}}$ are re-ordered so that the boundary nodes (edges in *Line*) come first. Then, minimizing $D_{Line}[u]$ with respect to u_I leads to the labels of the unknown nodes (edges in *Line*) u_I as the solutions to the following linear system:
$$L_I u_I = -K^T u_B , \quad (4)$$
   where the u_B are all set to the unit, L_I is the sub-Laplacian of $Line_{\mathcal{L}}$ for the nodes u_I, and K is a $|\mathcal{SB}'| \times |\mathcal{SB}'|$ block of the re-ordered Laplacian.
6. **Relabelling:** We relabel the edges in the original graph with the information coming from the Dirichlet process in the line graph, since there is a bijection between the nodes in the line graph and the edges in the original graph.

4. Understanding Dirichlet Densifiers Bounds

After reviewing the Dirichlet densification algorithm, which typically doubles the number of edges with respect to the original graph, we want to determinate a new bound for Commute Time estimation. This bound does not rely on the spectral gap but on Graph Densification (or graph rewiring). Firstly, we motivate the bound by showing that implicitly constraining the spectral gap through Graph Densification cannot fully explain some estimations in real datasets, where graphs with an important degradation of the spectral gap are better densified. Then, we set our working hypothesis: if densification can deal with a small/moderate degradation of the spectral gap, this is due to the fact that inter-cluster commute distances are considerably shrunk (these values are in a shrinking range). This suggests a more detailed bound which explicitly accounts for the shrinking effect of densification. Finally, we formally develop this bound, thus uncovering the deep implications of Graph Densification in Commute Times estimation.

Moreover, we can interpret the graph $G = (V, E, W)$ as a resistor networks, where the *resistance* of an edge $e = (i, j)$ is defined as $r_e = 1/W_{ij}$, i.e. the weights W_{ij} define the conductance of the edges. To define the Commute Times estimation, we need the concept of *effective resistance* R_{ij} (24) (23). In general, we have $R_{ij} \neq r_{ij}$ even if i and j are linked by an edge. Conceptually, the effective resistance is more global and encodes the resistance of the graph as a whole if we inject a unit current into i and it diffuses until reaching j. We can see effective resistance as current or electrical flow, which can be seen as a scaled Commute Times since $CT_{ij} = vol(G)R_{ij}$. This link characterizes the diffusive nature of Commute Times, however CTs are globally meaningless, unless we re-scale or re-define them (1) (25).

4.1. The von Luxburg et al. bound and Cheeger constant

Given a connected graph $G = (V, E)$ that is not bipartite, we can define the following bound derived by the approach of von Luxburg et al. (24):

$$\left| \frac{1}{vol(G)} CT_{st} - \left(\frac{1}{d_s} + \frac{1}{d_t} \right) \right| \leq 2 \left(\frac{1}{\lambda_2} + 2 \right) \frac{w_{max}}{d^2_{min}} \quad (5)$$

where $CT_{st} = R_{st} vol(G)$ is the commute time between the nodes s and t, R_{st} is the effective resistance, $vol(G)$ is the volume of the graph, λ_2 is the *spectral gap* and d_{min} is the minimum node

degree in G. The spectral gap λ_2 is the second eigenvalue of the normalized graph Laplacian $L = I - D^{-1}W$ where $D = diag(d_1, \ldots, d_n)$ is the degree matrix and W is the (symmetric) weighted adjacency matrix, with $w_{ij} > 0$ if $(i, j) \in E$. Then w_{max} is the maximal affinity.

The above equation explains why commute times are meaningless in large graphs. These graphs tend to have large spectral gaps due to the existence of inter-cluster links (noise). As a result, we have $R_{st} \approx \frac{1}{d_s} + \frac{1}{d_t}$, i.e. commute times only depend on their local degrees and not the path between them. Consequently they are meaningless for measuring distances between nodes in large graphs.

Conversely, a way of making $R_{st} \approx \frac{1}{d_s} + \frac{1}{d_t}$ diverge (and thus make commute times meaningful) is to reweight/rewire the edges in E so that $\lambda_2 \to 0$. This task is *partially due* to graph densification, which implicitly constrains the spectral gap as much as possible.

The existence of a small bottleneck is also compatible with the minimization of the graph conductance or Cheeger constant Φ (6):

$$\Phi \triangleq \min_{S \subseteq V} \frac{cut(S)}{\min(vol(S), vol(\bar{S}))}, \tag{6}$$

then, we have the following upper bound for λ_2:

$$\lambda_2 \leq 2\Phi, \tag{7}$$

where Φ is the Cheeger constant. This bound suggests that the spectral gap λ_2 is minimized when: a) the cut is minimized, and b) $min(vol(S), vol(\bar{S}))$ is as large as possible. It is well known that for two cliques of size n linked by r edges, we have $\Phi = \frac{r}{n(n-1)}$, i.e. $\lim_{n\to\infty} \Phi = 0$. However, if $r = n$ we need larger cliques for constraining the spectral gap. This rationale opens the door to modify the set of edges E, by adding and/or reweighting edges so that $min(vol(S), vol(\bar{S}))$ is maximized for all $S \subset V$. However, we must take into account the fact that the Cheeger constant relies on the worst case.

Our preliminary experiments show that Dirichlet densifiers (algorithm described in Section 2) lead to improve the Adjusted Rand Index (ARI) obtained from commute times *after densification* in a variety of datasets (NIST[1], COIL-20[2] and FlickrLOGOs-32 [3].

To motivate our discussion, in Tables 1, 2 and 3 we show the ARIs obtained for the NIST, COIL and LOGO datasets in several scenarios. Each scenario is characterized by: (i) a value k for building the k-NN, (ii) the fraction $|E"|$ of dominating edges chosen for building the line graph, and (iii) the fraction of dominating $|E_B|$ edges chosen as seeds for the harmonic analysis (Dirichlet process). In all scenarios, the ARIs *before densifying the datasets* is below 70%, 90% and 62% in NIST, COIL and LOGO datasets respectively (decreases as k increases). The question addressed by densification is whether this performance can be improved by rewiring/densifying the similarity graphs. Our analysis shows that for a small fraction of $|E"|$ (typically 0.35 in all of our datasets) and a tiny fraction of $|E_B|$ (around 0.05), densification significantly improves the commute times of the input graphs (best result in NIST, COIL and LOGO are 74.4%, 95.44% and 62.96%, respectively).

A detailed interpretation of the above ARIs leads us to evaluate the bound in Eq. 5 from the perspective of the spectral gap λ_2. In other words, we want to quantify the real effect of constraining the spectral gap on improving the commute time estimates. In Tables 4, 5 and 6, we show the spectral gaps for each of the scenarios, corresponding to the previous tables. As expected, the larger the spectral gap the poorer the performance. We remove from the analysis disconnected graphs ($\lambda_2 = 0$) arising when $k = 15$ since they are not accommodated by the bound. However, as k increases ($k = 25$, $k = 35$), we find some anomalies. In some densified graphs with larger spectral gaps we get better ARI's than for the corresponding un-densified graph (especially for optimal configurations).

Table 1. NIST: Adjusted Rand Index for different thresholds and number of k

		kNN 15			kNN 25			kNN 35		
		E_B								
		0.05	0.25	0.5	0.05	0.25	0.5	0.05	0.25	0.5
E''	0.05	37.3	41.88	40.62	57.23	54.33	52.26	27.12	30.88	43.49
	0.15	66.9	63.52	61.64	70.87	70.84	57.65	69.51	68.54	67.42
	0.25	71.78	69.15	65.01	71.05	70.4	70.21	69.95	71.6	70.51
	0.35	**74.4**	71.06	70.08	71.02	71.51	70.42	70.55	71.23	70.49
No dense			69.25			65.62			63.74	

Table 2. COIL: Adjusted Rand Index for different thresholds and number of k

		kNN 15			kNN 25			kNN 35		
		E_B								
		0.05	0.25	0.5	0.05	0.25	0.5	0.05	0.25	0.5
E''	0.05	55.17	57.99	33.51	54.31	51.03	30.94	72.66	71.68	67.85
	0.15	73.16	72.04	72.69	63.33	64.26	74.11	90.96	84.57	71.13
	0.25	93.69	83.68	82.98	92.09	91.09	64.32	91.01	91.99	90.27
	0.35	**95.44**	94.54	83.01	92.41	92.81	90.55	90.53	91.01	92.11
No dense			89.75			89.65			85.42	

The above results suggest that von Luxburg et al.'s bound (Eq. 5) does not fully characterize the real effect of densification. Our working hypothesis is that constraining the spectral gap (and Cheeger constant) is only part of the process of re-estimating commute times for mid-size/large-size graphs. Of course, the spectral gap has to be kept as small as possible for a reliable estimation of commute times. However, this becomes more and more difficult as k grows due to the appearance of inter-cluster links. Thus, if densification can deal with a small to moderate degradation of the spectral gap, and this is due to the fact that inter-cluster commute distances are exhibited sig-

[1] http://yann.lecun.com/exdb/mnist/
[2] http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php
[3] http://www.multimedia-computing.de/flickrlogos/

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Table 3. LOGO dataset: Adjusted Rand Index for different thresholds and number of k

		kNN 15			kNN 25			kNN 35		
		E_B								
		0.05	0.25	0.5	0.05	0.25	0.5	0.05	0.25	0.5
E''	0.05	20.21	18.11	22.56	45.59	42.99	40.2	19.94	14.01	16.69
	0.15	60.55	58.81	56.03	57.68	47.21	48.71	52.77	14.43	51.54
	0.25	61.77	60.58	59.81	59.24	59.21	47.56	54.65	53.33	53.29
	0.35	**62.96**	61.75	61.39	60.65	59.7	58.73	57.23	55.11	53.71
No dense		61.92			59.82			54.11		

nificant shrinkage. This suggests that a more detailed bound is needed which can explicitly account for the effect of densification.

4.2. The proposed bound

Given a graph $G = (V, E)$ and two nodes $s, t \in V$, the commute times CT_{st} is the expected time it takes a random walk to travel from s to t and back (11)(20)(27). The diffusive nature of commute times is characterized by the link with resistance distance $R_{st} = \frac{1}{vol(G)} CT_{st}$, for which the effective resistance is

$$R_{st} \triangleq \arg\min_{\mathbf{f}} \sum_{e \in E} r_e |\mathbf{f}_e|^p \,, \tag{8}$$

with $p = 2$, and $\mathbf{f} \triangleq \{\mathbf{f}_e\}_{e \in E}$ is the unit flow from s to t. In other words, we inject a unit current at s, extract it at t and observe the flow traced across the edges $e \in E$. Unit flows have two interesting properties: a) they are quite scattered along the edges (even in moderate size graphs), and b) the bulk of their magnitude is confined to the neighbourhood of both s and t.

Effective resistances also satisfy the Rayleigh monotonicity principle: given G with adjacency/similarity matrix W, let G' with adjacency/similarity W' which is identical to W except for the increase in the weight of one arbitrary edge (i, j), so that $W'_{ij} = W_{ij} + \delta$. Then, for arbitrary vertices s and t, we have

$$R^G(s, t) \geq R^{G'}(s, t) \,, \tag{9}$$

i.e. introducing new edges (or reweighting them incrementally) does not increase the effective resistance between any pair of nodes s and t in the graph. Thus, in order to quantify the effect of densification in bounding the effective resistance, we will exploit this principle as follows.

4.3. Upper bound

Let $G = (V, E)$ be an undirected and unweighted graph ($r_e = 1$ for $e \in E$), with $n = |V|$ and average degree $\tau = \Theta(d)$. Given any pair of nodes, s and t, let $\mathbf{f} \triangleq \{\mathbf{f}_e\}_{e \in E}$ be any unit flow between these nodes, and $\mathbf{f}^* \triangleq \{\mathbf{f}^*_e\}_{e \in E}$ the minimal flow that represents the effective resistance $R^G(s, t) = \sum_{e \in E} |\mathbf{f}^*_e|^2$. As a result: $R^G(s, t) \leq \sum_{e \in E} |\mathbf{f}_e|^2$. Consequently, we will obtain a compacted upper bound for $R^G(s, t)$ (as in (1)) and then we will show that when G is densified, leading to the new graph $H = (V, E')$ with $E \subset E'$, the bound connected with $R^H(s, t)$ is even tighter.

The flow $\mathbf{f} \triangleq \{\mathbf{f}_e\}_{e \in E}$ is constructed as follows:

1) Start at s by injecting a unit flow. The local flow transmitted to any of the N_1 neighbours of s is $1/d_s$. Their contribution to $\mathbf{f}$ is $1/d_s$.

2) The flow must be *unitary* (input flow equal to output flow for each node, until arriving to destination t). Thus, any of the N_2 neighbours of N_1 must diffuse a flow $1/(N_2 d_s)$. Then, let S be the number of *layers* with successive neighbours $N_1, N_2, \ldots, N_S$. Since $N_k = \tau k$, we have that, if any neighbour diffuses $1/N_k$ then

$$R^G(s, t) \leq \frac{1}{d_s} + \frac{1}{\tau^2} \sum_{k=1}^{S} \frac{1}{k} \,. \tag{10}$$

The value of S depends on the graph and it is not constant but for balanced trees (see Figure 1 for more clarity). Thus, the bound in Eq. 10 is an upper bound derived from setting S as the maximum reachable neighbourhood according to unitary diffusion. This indicates that there exists a symmetric process starting from the destination node t. Without loss of generality (for the definition of a bound), we can assume that this symmetric process has also S layers. Then:

$$R^G(s, t) \leq \frac{1}{d_s} + \frac{1}{d_t} + 2\frac{1}{\tau^2} \sum_{k=1}^{S} \frac{1}{k} \,. \tag{11}$$

3) Finally, to have a unit flow, we must link the two last layers (the one coming from s and that coming from t) through some of the existing edges between the nodes of these finals layers so that only a flow of $1/N_S$ per node is transferred in order to ensure unitarity. Then:

$$R^G(s, t) \leq \frac{1}{d_s} + \frac{1}{d_t} + 2\frac{1}{\tau^2} \sum_{k=1}^{S} \frac{1}{k} + \frac{1}{\tau^2} \cdot \frac{1}{S} \tag{12}$$

At this point, it is unclear what happens after densification. To resolve this question, we can characterise it as a process that modifies the average degree, τ to give a revised value $q\tau$. In particular, Dirichlet densifiers operate with $q = 2$ (two transitive edges are linked by an additional one). For a densified graph H obtained using a Dirichlet process, the bound in 12 can be reformulated as

$$R^H(s, t) \leq \frac{1}{d_s} + \frac{1}{d_t} + \frac{1}{2\tau^2} \sum_{k=1}^{S} \frac{1}{k} + \frac{1}{4\tau^2} \cdot \frac{1}{S} \,, \tag{13}$$

which reduces the bound for G by a factor of at least $1/4$ of the flow propagated through the S layers in one direction (either from s to t or viceversa).

ACCEPTED MANUSCRIPT

Table 4. NIST: Spectral gaps for different thresholds and number of k

		kNN 15			kNN 25			kNN 35		
		E_B								
		0.05	*0.25*	*0.5*	*0.05*	*0.25*	*0.5*	*0.05*	*0.25*	*0.5*
E''	*0.05*	0.0	0.0	0.0	0.0209	0.0251	1.9561	0.0498	0.0478	0.0395
	0.15	0.0049	0.0	0.0	0.0310	0.0275	0.0233	0.0778	0.0714	0.0630
	0.25	0.0097	0.0	0.0	0.0446	0.0356	0.0290	0.1043	0.0899	0.0732
	0.35	**0.0176**	0.0130	0.0073	0.0632	0.0478	0.0323	0.1337	0.1120	0.0865
No dense			0.0192			0.0481			0.0775	

Table 5. COIL: Spectral gaps for different thresholds and number of k

		kNN 15			kNN 25			kNN 35		
		E_B								
		0.05	*0.25*	*0.5*	*0.05*	*0.25*	*0.5*	*0.05*	*0.25*	*0.5*
E''	*0.05*	0.0	0.0	0.0	0	0	0	0	0	1.974
	0.15	0	0	0	0.0383	0.0006	0.0009	0.0085	0.0001	1.8446
	0.25	0.0738	0.0575	0.0342	0.0033	0.0017	0.0076	0.0184	0.0136	0.0069
	0.35	**0.0962**	0.0706	0.0542	0.0022	0.0007	0.0054	0.0004	0.0065	0.0043
No dense			0.0029			0.0022			0.0004	

4.4. Lower bound

We now turn our attention to the lower bound, $R^G(s,t)$, making use of the Rayleigh principle to construct a graph G' as follows (see also (1)(11)).

Graph G' is a linear contracted graph following the line connecting any pair of nodes s and t. We start with node s and add different edges of resistance 0 between all the neighbours of s and merge all these nodes in a single node N_1, and these edges form a *slice* as we can see in Fig 1 (top-left, in orange). We iterate this process for nodes $N_2, \ldots, N_S$ where E_j are the edges associated with the slice between N_j and N_{j+1} (top-right and bottom-left). To end we add a final slice between N_S and t. This construction is useful because it is ideal for a lower bound since removing edges in the original graph increases the effective resistance, and the flow between N_j and N_{j+1} is always unitary (bottom-right). Moreover, the edges E_j lead to an inverse parallel resistance according to the law $1/r = 1/r_1 + 1/r_2$.

More precisely, we can formulate the lower bound as follows:

$$R^{G'}(s,t) = \sum_{e\in E} i_e^2 = \frac{1}{d_s} + \sum_{j=0}^{S}\sum_{k=1}^{E_j} i_k^2 + \frac{1}{d_t}\,, \tag{14}$$

According to the generalized mean inequality we have:

$$\sum_{k=1}^{E_j} i_k^2 \geq \sum_{j=0}^{S} \frac{1}{E_j} \underbrace{\left(\sum_{k=1}^{E_j} i_k\right)}_{1} = \sum_{j=0}^{S} \frac{1}{E_j}\,,$$

For this reason, since G' has fewer edges than G, then $R^{G'}(s,t) \geq R^{G'}(s,t)$ and we have the following bound for a undensified graph:

$$R^G(s,t) \geq \frac{1}{d_s} + \frac{1}{d_t} + \sum_{j=0}^{S} \frac{1}{E_j} \geq \frac{1}{d_s} + \frac{1}{d_t} + \frac{S-1}{E_{max}}\,, \tag{15}$$

where E_{max} the maximal number of edges in a slice.

4.5. The proposed bound

With the above obtained bounds, for a densified graph H we have the following bounds (lower and upper) for any effective resistance:

$$R_{app} + \frac{\mathbf{1}}{\mathbf{2}} \cdot \frac{S-1}{E_{max}} \leq R^H(s,t) \leq R_{app} + \frac{1}{2}\cdot\frac{1}{\tau^2}\sum_{k=1}^{S}\frac{1}{k} + \frac{\mathbf{1}}{\mathbf{4}\tau^2}\cdot\frac{1}{S}\,, \tag{16}$$

where $R_{app} = 1/d_s + 1/d_t$ with respect to the same bound for the not-densified graph G:

$$R_{app} + \frac{S-1}{E_{max}} \leq R^G(s,t) \leq R_{app} + 2\frac{1}{\tau^2}\sum_{k=1}^{S}\frac{1}{k} + \frac{1}{\tau^2}\cdot\frac{1}{S}\,, \tag{17}$$

Summarizing, we have shown that the Dirichlet densification significantly reduces (by a half, **1/2**) the lower bound and also reduces by a quarter (**1/4**) the upper bound associated with undensified graphs. This is because of $q = 2$ for Dirichlet densifiers. Moreover, since the Dirichlet process minimizes inter-cluster links, we have that the commute time shrinkage is confined to intra-cluster nodes. This leads to the best values of the Adjusted Rand Index after commute times are estimated in densified graphs.

ACCEPTED MANUSCRIPT

Table 6. LOGO: Spectral gaps for different thresholds and number of k

		kNN 15			kNN 25			kNN 35		
		E_B								
		0.05	*0.25*	*0.5*	*0.05*	*0.25*	*0.5*	*0.05*	*0.25*	*0.5*
E''	*0.05*	0	0	0	0	0	0	0	0	0
	0.15	0.0091	0.009	2	1.9054	1.9054	1.9054	0.0535	0.0501	0.04
	0.25	0.013	0.0095	0.0085	0.039	0.039	0.039	0.1099	0.0974	0.0749
	0.35	**0.014**	0.0133	0.0089	0.0574	0.05	0.0356	0.1497	0.1356	0.1018
No dense			0.011			0.0481			0.0311	

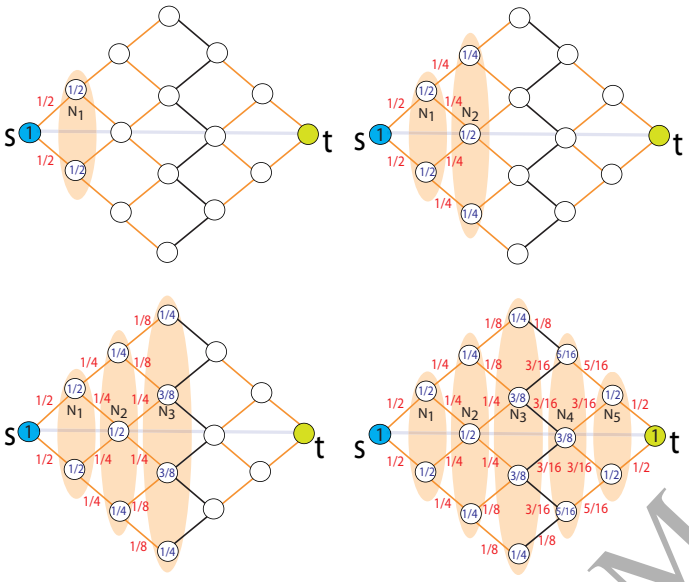


Fig. 1. Toy example of sub-optimal unit flows for bounding. From Top to bottom and left to right, we show the unit flow between nodes s and t with the layers (upper bound) in orange. Inter-layer links are in black. In this example there are $S = 3 + 2$ layers.

4.6. Results in densification

Once we have studied the limitations or bounds before and after Dirichlet densification, we resume the link between gap spectral and Cheeger constant (see Equation 7) with respect to the constraining of the spectral gap through two ways: i) reducing the number of inter-class edges, and ii) adding intra-class edges. This hypothesis is confirmed by our preliminary experiments and the defined bounds indicate that the inter-class commute distances are affected more by shrinkage after densification. We can be also improve the results by increasing the number of intra-class edges. We test the bounds as follows: we randomly add a few new intra-class edges for one class of the NIST dataset. We obtain a better result (74.92%) with respect to the original densification (72.52%), by only increasing the global density of the input graph by 0.03%.

5. Discussion and Conclusion

In this paper we have defined new theoretical bounds for the effective resistance in densified graphs, experimentally analyzing the impact of graph densification in bounding effective resistances (in other words, scaled commute times). We contribute a novel bound, which is more detailed in its predictions than simply relying on the spectral gap λ_2. Although the spectral gap is linked with the edge density of the graph (it is upper bounded by the Cheeger constant), the analysis based on λ_2 only addresses the ratio between the smallest cut and the graph density. However, the re-formulation of von Luxburg et al.'s bound requires us to estimate the impact of densification on the shrinkage of the inter-cluster commute distances, thus leading to better estimates than those provided by the original graph.

As a result, we prove that for our Dirichlet densification, the lower bound for CTs reduces significantly (1/2) the CTs bound compared to that for un-densified graphs, while the upper bound gives a reduction of 1/4 of the CTs bound for un-densified graphs. This means that we can better discriminate between the distributions of intra-class and inter-class commute times in densified graphs. Moreover, since the Dirichlet procedure minimizes the number of inter-cluster links, we have that the shrinkage of commute distances is confined to the intra-cluster nodes. This leads to the best ARIs (Adjusted Rand Indices) after commute times are estimated in densified graphs. This fully explains our experiments with real-world datasets. These bounds have not changed with respect to our previous works, but through this paper we can demonstrate that they are a useful tool to understand the benefits of our Dirichlet Densification.

Our formal development of this bound reveals important implications for graph densification in commute times estimation. In particular the bounds open the possibility of densification beyond simply increasing the volume of the graph. It thus allows to achieve less dense graphs with better estimates of the commute time due to the fact that the Dirichlet principle leads to an intelligent (minimum energy) diffusion, whereas the spectral gap is kept close to zero.

Densification is a different way of link prediction, and the optimization criterion must be included in networks (e.g. improvement of the quality of different graphs through learning). In future work, we are going to study these bounds in CNNs, which have to filter structural noise as in image denoising.

References

[1] Alamgir, M., von Luxburg, U.: Phase transition in the family of p-resistances. In: Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011. Proceedings of a meeting held 12-14 December 2011, Granada, Spain. pp. 379–387 (2011)

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

[2] Alon, N., Duke, R.A., Lefmann, H., Rödl, V., Yuster, R.: The algorithmic aspects of the regularity lemma. J. Algorithms 16(1), 80–109 (1994)

[3] Aroraa S., Kargerb, D., Karpinskic, M.: Polynomial time approximation schemes for dense instances of np-hard problems. Journal of Computer and System Sciences 58(1), 193 – 210 (1999)

[4] Bai, X., Yan, C., Yang, H., Bai, L., Zhou, J., Hancock, E.R.: Adaptive hash retrieval with kernel based similarity. Pattern Recognition 75, 136–148 (2018)

[5] Bollobás, B.: Extremal Graph Theory. Dover Publications, New York (2009)

[6] Chung, C.: Spectral graph theory, conference board of the mathematical sciences, no. 92. American Mathematical Society, Providence, RI (1997)

[7] Chung, F.R.K.: Spectral Graph Theory. Conference Board of the Mathematical Sciences (CBMS), number 92, American Mathematical Society (1997)

[8] Curado, M., Escolano, F., Lozano, M.A., Hancock, E.R.: Dirichlet densifiers for improved commute times estimation. Pattern Recognition 91, 56 – 68 (2019)

[9] Diaconis, P., Stroock, D.: Geometric bounds for eigenvalues of markov chains. Ann. Appl. Probab. 1(1), 36–61 (02 1991)

[10] Diaconis, P., Stroock, D., et al.: Geometric bounds for eigenvalues of markov chains. The Annals of Applied Probability 1(1), 36–61 (1991)

[11] Doyle, P.G., Snell, J.L.: Random Walks and Electric Networks, vol. 22. Mathematical Association of America, 1 edn. (1984)

[12] Escolano, F., Curado, M., Hancock, E.R.: Commute times in dense graphs. In: Structural, Syntactic, and Statistical Pattern Recognition - Joint IAPR International Workshop, S+SSPR 2016, Proceedings. pp. 241–251 (2016)

[13] Escolano, F., Curado, M., Lozano, M.A., Hancock, E.R.: Dirichlet graph densifiers. In: Structural, Syntactic, and Statistical Pattern Recognition - Joint IAPR International Workshop, S+SSPR 2016, Proceedings. pp. 185–195 (2016)

[14] Fiorucci, M., Torcinovich, A., Curado, M., Escolano, F., Pelillo, M.: On the interplay between strong regularity and graph densification. In: Graph-Based Representations in Pattern Recognition - 11th IAPR-TC-15 International Workshop, GbRPR 2017, Anacapri, Italy, May 16-18, 2017, Proceedings. pp. 165–174 (2017)

[15] Frieze, A.M., Kannan, R.: The regularity lemma and approximation schemes for dense problems. In: 37th Annual Symposium on Foundations of Computer Science, FOCS '96, Burlington, Vermont, USA, 14-16 October, 1996. pp. 12–20 (1996)

[16] Hancock, E.R., Wilson, R.C.: Pattern analysis with graphs: Parallel work at bern and york. Pattern Recognition Letters 33(7), 833–841 (2012)

[17] Hardt, M., Srivastava, N., Tulsiani, M.: Graph densification. In: Innovations in Theoretical Computer Science 2012, Cambridge, MA, USA, January 8-10, 2012. pp. 380–392 (2012)

[18] Komlós, J., Shokoufandeh, A., Simonovits, M., Szemerédi, E.: The regularity lemma and its applications in graph theory. In: Theoretical Aspects of Computer Science, Advanced Lectures (First Summer School on Theoretical Aspects of Computer Science, Tehran, Iran, July 2000). pp. 84–112 (2000)

[19] Liu, W., Wang, J., Kumar, S., Chang, S.: Hashing with graphs. In: Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011. pp. 1–8 (2011)

[20] Lovász, L.: Random walks on graphs: A survey. In: Miklós, D., Sós, V.T., Szőnyi, T. (eds.) Combinatorics, Paul Erdős is Eighty, vol. 2, pp. 353–398. János Bolyai Mathematical Society, Budapest (1996)

[21] q. Luo, Z., k. Ma, W., c. So, A.M., Ye, Y., Zhang, S.: Semidefinite relaxation of quadratic optimization problems. IEEE Signal Processing Magazine 27(3), 20–34 (2010)

[22] Luo, Z., Ma, W., So, A.M., Ye, Y., Zhang, S.: Semidefinite relaxation of quadratic optimization problems. IEEE Signal Processing Magazine 27(3), 20–34 (May 2010)

[23] von Luxburg, U., Radl, A., Hein, M.: Getting lost in space: Large sample analysis of the resistance distance. In: Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems 2010. Proceedings of a meeting held 6-9 December 2010, Vancouver, British Columbia, Canada. pp. 2622–2630 (2010)

[24] von Luxburg, U., Radl, A., Hein, M.: Hitting and commute times in large random neighborhood graphs. Journal of Machine Learning Research 15(1), 1751–1798 (2014)

[25] Nguyen, C.H., Mamitsuka, H.: New resistance distances with global information on large graphs. In: Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, AISTATS 2016, Cadiz, Spain, May 9-11, 2016. pp. 639–647 (2016)

[26] Qiu, H., Hancock, E.R.: Clustering and embedding using commute times. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(11), 1873–1890 (2007)

[27] Qiu, H., R., E.: Clustering and embedding using commute times. IEEE Trans. Pattern Anal. Mach. Intell. 29(11), 1873–1890 (2007)

[28] Torsello, A., Hancock, E.R.: Graph embedding using tree edit-union. Pattern recognition 40(5), 1393–1405 (2007)

[29] Xiao, B., Hancock, E.R., Wilson, R.C.: Graph characteristics from the heat kernel trace. Pattern Recognition 42(11), 2589–2606 (2009)

[30] Zhang, F., Hancock, E.R.: Graph spectral image smoothing using the heat kernel. Pattern Recognition 41(11), 3328–3342 (2008)

[31] Zhang, H., Bai, X., Zhou, J., Cheng, J., Zhao, H.: Object detection via structural feature selection and shape model. IEEE transactions on image processing 22(12), 4984–4995 (2013)

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Conflict of Interest

We have not conflict of interest.

ACCEPTED MANUSCRIPT