

K-way Regularized Spectral Clustering in Human Graphs

Raymundo Duran
rdurarl@mit.edu

Abstract—I examine the clustering properties of two regularized spectral clustering methods in three empirical datasets with distinct graph structures. Recent literature [1] explains the role of the said methods in creating balanced partitions when performing spectral clustering under the presence of "small trees" sets, however little more is known: cluster number asymptotics, performance under general graph structures and comparisons between the two regularization approaches are open questions. I find strict conductance improvements under regularization on all 3 graphs for k clusters, but mixed, graph-dependant results in the remaining inquiries.

I. INTRODUCTION

GRAPH partitioning methods are increasingly paramount to numerous domains. Be them as a means to reduce relational complexity within a system, or as a standalone tool to gain qualitative insight of intricate communities, clustering tasks involve finding a number of optimal node-edge sets according to predetermined criteria on graph structure and possibly a set of observables per node.

While various analytical frameworks such as Stochastic block Models (SBM) or Graph Neural Networks (GNNs) exist for graph partitioning in the type of contexts studied by this paper, I'm chiefly concerned with Spectral Clustering (SC) due to its advantages: firstly, as shown by Fiedler, 1973 [2], there exists an analytically tractable framework, based upon the *algebraic connectivity* properties of graphs, relating eigendecomposition with observed structure. This approach –compared to most others– favors intuitive understanding of underlying edge relationships; in particular, it facilitates intuitive interpretations of the nodes' adjacency structure, making it a good candidate for applications in the social sciences. Secondly, SC warrants relatively limited computational complexity [3] and for the most part, stable results:

SC generally involves solving eigenproblems on the $n \times n$ normalized Laplacian matrix of a graph. Clustering algorithms using the eigenvectors of said matrix often yield good approximations in $\mathcal{O}(n^3)$ time to what otherwise would be NP-hard combinatorial optimization problems, and they do so with only mild assumptions on the graph's connectedness (mostly overcome by regularization). The same cannot be said of most other methods which either rely on a model with strong assumptions or are not robust to small input changes; for instance, SBM assumes an expected degree branch-out of stochastic magnitude $\mathbf{O}(\log n)$ [4], thus relying on the likelihood of said process generating the graph to cluster adequately; this approach may not hold in common human-originated graphs failing to follow ergodic generation (e.g. Non-homophilous graphs). And granted GNNs are well-suited

for flexible clustering given their graph structure agnosticism, input growth limitations [5] in addition to weak interpretability are inconveniences that make them less relevant in disciplines where the output warrants some notion of equilibrium thus requiring consistent properties.

Consistency and interpretability notwithstanding, the foremost motivation for SC in this paper is that of the clusters' "qualitative properties" when applied to human-originated graphs: be them social networks or produced indirectly by human activity. Admittedly, the goodness of said properties is domain-conditioned. Still, a relevant observation constantly arising in various fields ([6] community detection, [7] bio statistics, or [8] in social network analysis) when evaluating SC for applied research states clusters produced by unregularized –Vanilla– SC are "uninteresting", and often capture core-periphery patterns where, for a relevant range of desired clusters e.g. $k = 2, \dots, 5$ the graph is partitioned in a very large group containing the most central nodes (the core), and $k-1$ non-informative small groups covering highly peripheral, "irrelevant" nodes. Particularly undesirably, the biggest set changes little as k increases. These, of course, are not defects in Fiedler's conception; rather, they are a natural consequence of the features exhibited by a growing set of empirical graphs. Modern analytical tasks in human graphs contrive the need for less exact clusters: "suboptimal" partitions robust to relatively few isolated nodes, still sparsely although not necessarily uniquely linked.

To the best of my knowledge, the aforementioned observation has escaped formal and unambiguous definition, possibly due to the difficulty in stating what a desirable partition of the core is, and the challenge of finding a parsimonious way to classify distinct graph structures and their "expected cluster quality". Surprisingly, even in the absence of a problem statement, satisfactory solutions were –somewhat incidentally– proposed a decade ago, in the form of regularization. *Regularized SC* (RSC) aims to find "more balanced" partitions –more on conductance and its equivalence to constrained Rayleigh quotient problems in section 2. Doing so simply entails adding or adjusting arbitrarily "weak edges" to particular nodes. I will herein discuss the results of two approaches: The first one (*standard* RSC) suggests linking the whole graph equally. A second approach (henceforth *Chaudhuri* RSC) instead proposes a form of degree correction meant simply as a means to control for high degree heterogeneity. Having a flimsy question comes at the cost of being unable to determine which approach performs better on a particular case and why;

in the last contribution to this literature, Zhang et al., 2019 [1] explain poor performance of Vanilla SC in terms of a subgraph topology they name *g-dangling sets* and show how regularization increases the conductance of small partitions. Unfortunately, they only studied the case of bipartite graphs, and the instance they argue is empirically limited.

Several related RSC matters remain unstudied: what is the effect of k-way RSC on conductance? Do results converge to Vanilla SC as the number of partitions grows? If so, how fast does convergence occur given a graph structure? Is there a sensible way to separate the effects of increasing k and regularization on conductance? Zhang et al. (*ibid.*) also argue Chaudhuri and Standard RSC are roughly equivalent for g-dangling-sets graphs but Standard is efficient; are there graph structures where Chaudhuri generally performs better? Are there sensible choices for the regularization parameter or is blind hyperparameter tuning warranted? In attempt to elucidate these questions, I carry out k-way clustering, show metrics and argue on the results of 3 distinct empirical human graphs intended to serve as ideal cases where either Vanilla and RSC may either fail or succeed for different reasons.

The rest of this paper is organized as follows: section 2 reviews spectral graph partitioning methods and literature starting from Fiedler's (1973) approach. Section 3 introduces the datasets used in this paper, describes their differences through summary statistics and provides Vanilla SC results. Section 4 examines RSC and hyperparameter tuning results, as well as provides some analysis for k-partitions asymptotics. Section 5 concludes.

II. LITERATURE REVIEW

Given the undirected graph $G = (V, E)$ with $V = \{v_1, v_2, \dots, v_n\}$ the set of vertices, each associated to an indicative vector $e_i \in E$, $i = 1, 2, \dots, n$ of edges, with each vector element $e_i^j = 1$ if node i is connected to node $j \neq i$ and $e_i^i = 0$. For ease of discussion, multigraphs and directed graphs are excluded. I discuss weighted graphs when degree normalization and regularization are introduced.

Define the oriented incidence matrix $\mathcal{E} = -[e_1; e_2; \dots; e_n]$. Because G is undirected, this is also the negative adjacency matrix ($A = -\mathcal{E}$) between nodes of G . Further define $L \equiv \mathcal{E}^T \mathcal{E} = D - A$ the graph's Laplacian matrix. Said matrix can be expressed as the difference of D , a diagonal matrix with typical element $d_{i,i}$ the degree of node v_i and A , the unweighted adjacency matrix of the graph. Given $\mathcal{E}^T \mathcal{E}$ is Hermitian, it admits the eigendecomposition $L = Q \Lambda Q^T$, with $q_i, i = 1, \dots, n$ the eigenvectors (and orthonormal base) of L , and Λ a diagonal matrix with typical element λ_{ii} , the eigenvalues such that $L q_i = \lambda_{ii} q_i$ for all i . Without loss of generality assume all eigenvectors/values are ordered $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$.

A. Rayleigh quotient optimizations and eigenvalue problems

I introduce key results linking clustering to Rayleigh quotient optimization problems and graph structures described by eigenpairs, used abundantly in this paper.

The Generalized Rayleigh quotient (GRQ) of the Laplacian $R(L, M, x) = \frac{x^T L x}{x^T M x}$ where M is a $n \times n$ matrix and x a vector of length n , are essential tools for the discussion of both clustering performance metrics and graph structures (Note, if $M = I_n$; $R^*(L, x) = \frac{x^T L x}{x^T x}$, the Rayleigh quotient [RQ]). Admittedly the most immediate optimization result is that of eigenvectors. For any eigenvector q_i :

$$R^*(L, q_i) = \frac{\lambda_i x^T x}{x^T x} = \lambda_i$$

For a Hermitian PSD matrix (discussed in next section), the Fischer-Courant min-max principle implies the biggest and smallest eigenvectors are the RQ's global maximizers and minimizers respectively:

$$\min_x R^*(L, x) = \lambda_1 \quad \max_x R^*(L, x) = \lambda_n$$

And importantly for algebraic connectivity, given linear restrictions $\|q_k\| = 1 \cap C q_k = 0^{k-1}$ where C is a $(k-1) \times n$ matrix, the constrained minimizer is the feasible eigenvector associated with the smallest eigenvalue:

$$\min_{x \in S} \max_{S = \{x \in \mathbb{R}^n | Cx = 0, \|x\| = 1\}} R^*(L, x) = \lambda_k$$

Finally, many clustering methods don't involve examining the quotient of particular eigenvectors, but instead evaluate cluster-indicator vectors associated with their elements. For instance, discretization through a cutoff $\kappa : q_{i,j} \rightarrow \{-1, 1\}$, $j = 1, \dots, n$. For values of q_i less than or equal, and greater than some constant α , respectively such that node j gets allocated to one of the two clusters. In those cases $R^*(L, x_\kappa)$ of some cluster-binary vector x_κ won't be a constrained optimum of the RQ. But as [9] show, there exists some GRQ for which this vector is a constrained minimizer –more on this in subsection C– e.g.:

$$R(L, D, x_\kappa) = \frac{x_\kappa^T L x_\kappa}{x_\kappa^T D x_\kappa}$$

$$s.t. x_\kappa^T D \mathbf{1} = 0, \quad \|x\| = 1$$

The "normalized NCut" for Fiedler's indicative vector.

B. Graph structure and eigenpairs; Fiedler's, 1975 [10] setup

Unbeknownst to them, several authors (Fiedler, 1973 [2], [Doob, 1970], [Anderson & Morely, 1971]) simultaneously advanced the same characterizations of a graph's structure from its eigenpairs, as well as found bounds for the eigenvalues in terms of its nodes' degree. I will review Fiedler's graph partition idea and briefly motivate its network formalization by Shi-Malik, 2000 [9].

As stated before, the largest eigenvalue is also the global maximum of the RQ: $R^*(L, q_n)$.

L is positive-semidefinite, since for $q_1 = \mathbf{1}$: $L \mathbf{1} = (D - A) \mathbf{1} = 0^n \implies \lambda_1 = 0$. Of course, the global minimum of $R^*(L, \mathbf{1}) = 0$. If a graph has m subsets of vertices $V_i, V_j \subset V$ $i \neq j$ sharing no edges between them (*components*), then

$\lambda = 0$ is of multiplicity m with eigenvectors equal to the indicative vector of V_i ; $q = \mathbf{1}_{V_i}$ and $\sum_{i=1}^m \mathbf{1}_{V_i} = \mathbf{1}$. For ease of exposition and adequate comparison with RSC I will henceforth refer to single component graphs only.

The eigenvector associated with the second smallest eigenvalue, known as the *Fiedler vector* is the constrained minimum of the RQ with constraint $q_1^T q_2 = \mathbf{1}^T q_2 = 0$, a true assertion by orthonormality of Q . As expected $R(L, q_2) = \lambda_2$, often named the *algebraic connectivity* of the graph. Perhaps the most important result in Fiedler, 1975 [10] is related to the elements of the Fiedler vector and the subgraphs they induce based on their value:

Let $G(V, E)$ be a connected graph.

- For real number $\alpha^+ \geq 0$ define the set of vertices $V_1(\alpha^+) = \{v \in V | q_{2,j} \geq -\alpha^+\}$ and its induced edges E^+ . The subgraph $G(V_1, E^+)$ is connected.
- For real number $\alpha^- \leq 0$ define the set of vertices $V_2(\alpha^-) = \{v \in V | q_{2,j} \leq |\alpha^-|\}$ and its induced edges E^- . The subgraph $G(V_2, E^-)$ is connected.

Note in particular, provided $q_{2,j} \neq 0 \forall j$, we can have $V_1 \cap V_2 = \emptyset$ if we set $\alpha^+ = \alpha^- = 0$, but the special aspect about these non-overlapping sets of vertices is they form connected (single-component) graphs. The qualitative finding is that Fiedler's vector describes a spectrum of connectedness in the real line, and places each vertex at said spectrum in a separate graph due to the fact that connectedness does not "abruptly" change, except maybe at 0, where this quality turns problematic and the subgraphs partition is only approximate.

Two-way clustering in this context consists of discretizing the Fiedler vector into indicative vector y , with elements

$$y_j = \begin{cases} 1 & q_{2,j} > 0 \\ -1 & q_{2,j} < 0 \end{cases}$$

For $j = 1, \dots, n$. The subgraphs induced by the indicator vector $G_1(V_1(y), E_1), G_2(V_2(y), E_2)$ have the feature that they (approximately) share the least amount of edges between them. Later network science work would coin the *cut* function to describe this feature:

$$\text{cut}(V_1, V_2) = \sum_{\{(e_1, e_2) \in V | e_1 \in V_1, e_2 \in V_2\}} E(e_1, e_2) \quad (1)$$

Note equation 1 is equivalent to $y^T L y$, but the metric network-wise is easily generalizable to more than 2 subgraphs, while the result of $(y_1 y_2)^T L (y_1 y_2)$ is ambiguous. Shi & Malik [9] would show Fiedler's method minimizes the subgraph's "average cut". Define the volume as the sum of the degrees in nodes of a graph $\text{vol}(V) = \sum_{v \in V} d_{v,v} = \sum_{i=1}^N d_{i,i}$:

$$\min_{V_1 \subset V} \text{cut}(V_1, V_1 \setminus V) \cdot \left(\frac{1}{\text{vol}(V_1)} + \frac{1}{\text{vol}(V \setminus V_1)} \right) \quad (2)$$

But this notion and its equivalence with a GRQ minimization would only be brought upon much later in the search for generalized metrics in k-way clustering and network theory developments.

C. K-partitions in spectral graph theory and GRQ-based metrics

Within spectral graph theory, k-partitioning methods can be divided in hierarchical and k-way. The former are mainly found in early work by Hagen & Kahng [11], and Spielman et al. [12], consisting of recursively partitioning the sorted values of the Fiedler vector such that equation 2 is minimized for each iterated subpartition. As shown below in algorithm 1, Hagen et al.'s innovation is iteratively checking for the locally optimal vertex allocation with respect to the average cut by step-wise passing of the Fiedler values between the induced graphs. Analogous algorithmic contributions based on other criteria include Dasgupta et al. [13] by minimizing intra-partition variance on the Fiedler vector's values, or Newman [14] by subgraph modularity (not a fully spectral method).

Algorithm 1 Recursive value partitioning (Hagen et al., 1992)
partition j of subvector $q_{1,j-1}$ or $q_{2,j-1}$

```

INPUT( $q_{1,j-1}$ )
DECLARE( $V_{1,j}, V_{2,j}$ ),  $n = \text{length}(q_{1,j-1})$ ,  $\text{avgcut} = \dim(n-1)$ 
SORT( $q_{1,j-1}$ ) =  $q_{1,j-1}^*$ 
for  $i$  in  $1 : (n-1)$ 
   $V_{1,j} \leftarrow q_{1,j-1}[1 : n-i]$ ,  $V_{2,j} \leftarrow q_{1,j-1}[n-i : n]$ 
   $\text{avgcut}[i] = \frac{\text{cut}(V_{1,j}, V_{2,j})}{\min\{\text{vol}(V_{1,j}), \text{vol}(V_{2,j})\}}$ 
return  $i^* = \arg \min_i \text{avgcut}[i]$ 
 $V_{1,j} \leftarrow q_{1,j-1}[1 : i^*-1]$   $V_{2,j} \leftarrow q_{1,j-1}[i^* : n]$ 
return  $q_{1,j} = q_{1,j-1}[1 : i^*-1]$ ,  $q_{2,j} = q_{1,j-1}[i^* : n]$ 

```

K-way methods rely on the informativeness of several eigenvectors to cluster vertices together by conventional metric partitioning such as k-means. Under this perspective, low-rank approximations of the Laplacian can create meaningful (orthogonal) bases upon which informative groupings of nodes would be revealed. Unlike principal components approaches, in $L \approx \sum_{i=1}^k q_i q_i^T \lambda_i$, the eigenpairs associated to the –potentially– most informative vectors will be smaller than the rest in magnitudes bounded by the degree of the graphs themselves; generally, the eigenvalues will be negligible and the eigenvector's elements smaller for large human graphs. This problem is well-recognized by the first contributions in this direction. Shi-Malik, 2000 [9], Ng et al., 2002 [15] and most contributions hereafter adopt solutions that involve *normalizing* the graph edges by pre/post multiplying the Laplacian by each node's total degree. For instance, Given $L = D - A$, the degree can be normalized by $D = \text{diag}\{\text{Vol}(v_1), \text{Vol}(v_2), \dots\}$:

$$L_N \equiv D^{-1/2} L D^{-1/2} = I_n - D^{-1/2} A D^{-1/2} \quad (3)$$

Where, as explained in the previous subsection, $\lambda_i \in [0, 2]$, all node's degree is equal to one, and the normalized eigenvectors are approximately standardized. Similar transforms such as $L_{NR} \equiv D^{-1} L = I_n - D^{-1} A$ exist, where L_{NR} loses symmetry but it has well-known equivalences (e.g. $L_{NR} u = \lambda u \iff L_N(D^{-1/2} u) = \lambda(D^{-1/2} u)$) making the normalization of eigenvectors more apparent. After this

transformation, most literature reviewed across several fields of application obeys the following structure:

K-way clustering general procedure :

- 1) Normalize L through some transformation: L_N
- 2) Justify and select a number of eigenvectors of L_N : Q^N
- 3) Use some technique to find k groups of nodes from Q^N .
- 4) (opt) If subgraphs disconnected, use method to relink
- 5) (opt) Report/compare clustering metrics empirically

For instance, Binkiewicz et al., 2017 [7] propose a way to mix weighted covariates for all graph elements: given X a $n \times R$ matrix of R covariates per node and edge, they compute the k largest eigenpairs of the weighted Laplacian $L_N + \alpha X X^T$ where $\alpha \in [0, +\infty)$ is a tuning parameter; other works [15] [8] instead employ the largest k eigenvectors in combination with extensions specific to empirical datasets or particular applications.

The literature corpus on SC is not cohesive. Contributions do not follow up on developments outside their respective fields, and there is a generalized lack of theoretical treatment/analysis of the proposed techniques; few efforts have been made to make systematic comparisons between methods. Not to mention theoretical guarantees for said methods are extremely scarce and limited in scope; as a result, there's a lack of frameworks to describe, compare and categorize clustering properties or performance problems for a given graph structure. This is a documented fact in the field (take for instance Zhang et al., 2019 [1] in their abstract). Other authors more amicably highlight the lack of guarantees and metrics generalization: Dall'Amico, 2018 [22], 2019 [16] whom even went on to propose a unified framework for SC in sparse graphs in 2021 [17]. Among the scarce attempts to develop theoretically sound results in SC are Leskovec et al., 2008 [6] and Zhang et al., 2019 [1]; in both cases, the authors transcend general sparsity of the Laplacian, and successfully link specific topological cases to concrete clustering problems. In my view this approach can plausibly provide domain-robust property comparisons of the many methods for Vanilla (and consequently, regularized) SC in arbitrary graphs.

D. Clustering metrics

A closely linked problem is defining what constitutes cluster quality in an unlabeled graph, and the metrics through which it is evaluated. Perhaps the most generalizable precept is that of *conductance* as justified by Kannan et al., 2004 [18] (Shi-Malik, 2000 [9] make an equivalent formulation up to a constant, but with another focus). They define cluster quality in terms of the weighted dissimilarity of the resulting subgraphs such that connected nodes are highly similar and disconnected ones highly dissimilar. Given graph $G(V, e)$, define a subgraph's $A(V_1, a)$ *association* as the total number of edges from all nodes incident in A :

$$\text{assoc}(A, V) = \sum_{a \in V_1} \sum_{e \in V} E(e, a) \quad (4)$$

Note the GRQ-equivalent expression for $k = 2$ is simply $y^T D y$. The conductance of cut A , $\phi(A)$ is therefore defined as the association-weighted cut:

$$\phi(A) = \frac{\text{cut}(A, V \setminus A)}{\min\{\text{assoc}(A, V \setminus A), \text{assoc}(V \setminus A, A)\}} \quad (5)$$

Where $\phi(A) \in (0, 1]$. A low conductance would indicate either a very within-connected subgraph or a partition in a graph with high algebraic connectivity; in either case, the general implication is relatively high intra-cluster connectedness as a way to compensate for an "unclean" cut with multiple edges connecting different clusters. Shi-Malik also knew conductance (*Ncut*, as they called it) was further bounded by the Cheeger inequality, such that

$$\frac{1}{2} \lambda_2 \leq \phi(A_k) \leq \sqrt{2\lambda_2}$$

For any cluster A_k in a k-partition. This is also a quality constraint on k-clusters. For k arbitrarily large, clusters will have a dissimilarity bounded by graph structure. Part of this paper's contribution is illustrating the convergence to said limitation as k grows, also for RSC (regularization deflates general sparsity, hence inflates association).

In k-partitions $A_1 \cap A_2 \cap \dots \cap A_k = \emptyset$ total conductance is typically defined as the minimal conductance:

$$\Phi(A_1, \dots, A_k) = \min\{\phi(A_1), \phi(A_2), \dots, \phi(A_k)\} \quad (6)$$

This is a stringent although intuitively and technically justified criterion; as preambled in subsections 2A-2B, [9] show equivalence between the problem of finding a bipartition $(V \setminus A, A)$ such that $\phi(A)$ is minimized and the constrained minimizer for the GRQ objective $\min_{\{0,1\}^n} R(L_N, D; y \in [-1, 1]^n \mid y^T D 1 = 0)$, but it's easy to see the problem's generalization to multiple indicative vectors is ambiguous. Furthermore, the minimal conductance result is a GRQ constrained minimum by the Courant-Fisher minmax theorem motivated by the (regularized) Laplacian's positive -semidefiniteness, hence other criteria like maximal or median conductance don't have constrained optima properties, which makes them not-suitable for regularization, and have difficult intuitive interpretations. Aggregated metric objectives such as mean conductance may fall under the minmax theorem, but generally fail to translate robustly into clustering quality ($k - 1$ low-conductance clusters may be offset by one high-conductance cut).

The vulnerabilities of minimal conductance are observed but barely researched in the literature. Only recently, [1] show a consistent cause for bad partition quality are *g-dangling sets*, these are subgraphs $A_g \subset G$ forming trees with $g - 1$ branches and one edge connecting them to the rest of the graph. The regularized graph conductance of this partition $\phi(A_g) = \frac{\text{cut}(A_g, G \setminus A_g)}{\text{Assoc}(A_g, G \setminus A_g)} = \frac{1}{2g-1} \approx (2g)^{-1}$ is relatively low and g plausibly large in empirical settings. They show these sets grow $\Theta(n)$ in non-homogeneous random graphs, with the consequence that, for $\lambda_2 \leq (g - 1)^{-1} \leq \frac{1}{n} \sum_{i=1}^n d_i \equiv \bar{d}$, there will be a dense proportion of eigenvalues in $[\lambda_2, (g - 1)^{-1}]$ where partitioning eigenvectors will fall, concealing more balanced partitions in $[(g - 1)^{-1}, \bar{d}]$. Regularization increases the conductance of these sets.

E. Regularization(s)

The term *regularization* has a degree of polysemy in the literature. In this work I refer explicitly to the set of techniques addressing sparsity in degree-heterogeneous, weighted and non-directed graphs, be them (semi) labeled, or unlabeled. These contributions appeared chiefly in stochastic model-based approaches to clustering but are also relevant to SC as they address the same type of sparsity.

In this context, regularization of degree-heterogeneous graphs entails artificially reducing its sparsity while changing its relative degree composition: firstly by adding low-weight edges between previously disconnected nodes, and secondly by changing the relative weights of previously connected nodes. While there are copious ways one can attempt to vary weights to a graph and reduce its sparsity, I examine two methods:

Amini et al., 2013 [19] (RSC): Adding $\frac{\tau}{n}$ throughout, where n is the number of nodes and $\tau \in [0, +\infty)$. Laplacian:

$$D_\tau = D + \tau I_n \quad (7a)$$

$$A_\tau = A + \frac{\tau}{n} \quad (7b)$$

$$L_\tau = D_\tau - A_\tau \quad (7c)$$

Then if k-way clustering, simply normalizing appropriately $L_\tau^N = I_n - \sqrt{D_\tau^{-1}} A_\tau \sqrt{D_\tau^{-1}}$. Node degrees are normalized to 1, and the typical edge in the adjacency matrix is

$$a_{i,j}^\tau = \frac{e_{i,j} + \frac{\tau}{n}}{\sqrt{(d_i + \tau)(d_j + \tau)}}$$

Previously disconnected nodes ($e_{i,j} = 0$) are added an edge of magnitude $\frac{\tau}{\bar{d} + \tau}$. The lesser degree nodes previously had, the larger the new normalized edge with $a_{i,j}^\tau \rightarrow \frac{1}{n}$. Importantly, nodes with many prior edges are made weaker (e.g. for a g-dangling set with single previous outside edge: $a_{ij}^\tau - a_{ij} = \frac{n+\tau}{n \cdot (1+\tau)} < 0$ for $\tau > 0$), where as τ increases, the weight also converges to $\frac{1}{n}$. Therefore, hyperparameter τ acts as an uneven equalizer of edge weights, with convergence rate partially determined by the previous degrees. Note as $\tau \rightarrow \infty$ the regularized graph will converge to \mathbf{K}_n , the complete n-node graph with edge $\frac{1}{n}$, which by definition has conductance equal to one for any k-partition $A_1, \dots, A_k \subset G$:

$$\Phi(A_1, \dots, A_k) \xrightarrow{\tau \rightarrow \infty} \phi(\mathbf{K}_n \setminus K^j, K^j) = \frac{n_k(n - n_k)n^{-1}}{(n - n_k)n_k n^{-1}} = 1$$

It is not generally known whether this convergence is uniform or if τ is bounded around the regularized minimal conductance, which complicates hyperparameter tuning greatly both in parameter space size and inference when a minimum is reached. However, Zhang et al., 2019 [1] do prove, for a very restricted case of g-dangling sets, that RSC increases the conductance of said sets for bounded tau only. In addition to Qin & Rohe, 2013 [20] showing the associated eigenvalues change to be bounded by equation 8

lower conductance and more balanced regularized cuts rely, to the best of my knowledge, solely on these results.

$$\|\Lambda - \Lambda_\tau\|_F \leq \frac{1}{\lambda_2} \sqrt{\frac{\ln(4n)}{\underline{d} + \tau}} \quad (8)$$

Chaudhuri et al., 2012 [21] (Chaudhuri RSC): Normalizing the edges by $\frac{1}{\underline{d} + \tau}$ only, as a way to control sparsity:

$$D_\tau = D + \tau I_n \quad (9a)$$

$$L_\tau = I_n - D_\tau^{-\frac{1}{2}} A D_\tau^{-\frac{1}{2}} \quad (9b)$$

There are two differences with respect to RSC: firstly note the typical edge is $a_{i,j}^\tau = \frac{e_{i,j}}{\sqrt{(d_i + \tau)(d_j + \tau)}}$ therefore weak links are not added. Secondly $a_{i,j}^\tau \rightarrow 0$ as $\tau \rightarrow \infty$, which, on principle, would imply conductance is undefined (approaches zero) as $L \rightarrow I_n$. Note however, for finite τ edges are $\mathcal{O}(\frac{1}{\tau + d})$, which would, on principle, preserve the Vanilla SC result because the order between eigenvalues would also be preserved.

Chaudhuri –unlike Standard– RSC preserves the Laplacian's sparsity, hence sparse `linalg` libraries compute faster when tuning τ . Further, given diverse graph structures, an additional sparsity-based technique can enrich the analysis.

III. GRAPHS DESCRIPTION, SUMMARY STATISTICS AND VANILLA-SC RESULTS

I compare SC results on three human-originated graph datasets exhibiting distinct graph structures. The datasets' generating processes depict a spectrum of applications covering clustering cases where graph structure warrants close to ideal Vanilla results, as well as instances in which even Regularized SC may fail. Differences among datasets are described both qualitatively and through a set of normalized graph measures I summarize in Table I.

All three datasets are plotted in Figure 1. The Facebook Egonets dataset (henceforth *Facebook*) shown in panel 1a is part of the SNAP social network datasets; introduced by Leskovec et al., 2014 [22]. It compiles the friend lists of facebook users who authorized an application with a network mining tool to retrieve said lists. Each node represents a user along with its features (omitted), and edges a friendship between users on the platform. It originally contains 4039 nodes and 88,234 edges –I removed 32 nodes in 7 components disconnected from the main 4037-node one. This is a highly homophilous graph with evident clusters (It is likely we're not observing links between clusters) fostered by the limited data generating mechanism, which only maps the direct friends list. Its normalized algebraic connectivity is 0.0038; along with the observed structure, this should be a canonical case for Vanilla SC.

The CoRA Citations dataset (*Cora*) introduced by McCalum et al., 2000 [23] is shown in panel 1b. It gathers a collection of scientific publications classified into seven fields (omitted), which create a directed network of citations – each node represents a publication along with an indicative vector of

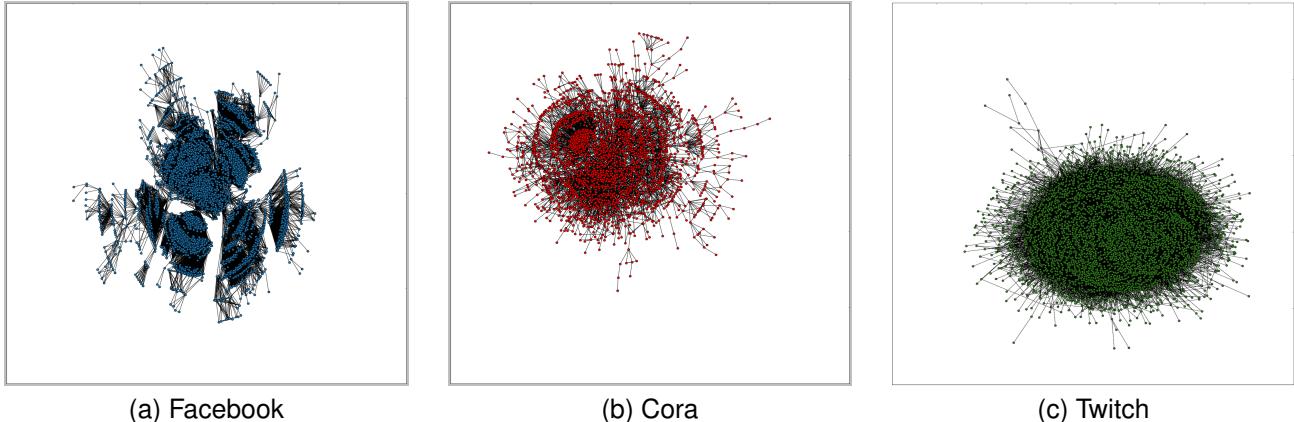


Fig. 1. Kamada-Kawai layouts of the graph datasets used in the analysis

words (omitted), and each edge represents a citation. It consists of 2708 nodes and 10,556 edges out of which I removed 203 nodes in 28 components (along with 348 edges). This graph exhibits a clearer core-periphery structure [24] presumably with several d-dangling sets –weakly linked peripheral sets with tree-like topologies. As discussed in the next subsection, the core is relatively sparse: it is neither very hierarchical (as is Facebook) nor abundantly degree-heterogeneous. Vanilla SC is expected to perform poorly and partition a d-dangling set; regularization should help create more central cuts, hopefully resemblant of academic fields.

The Twitch Streamers Multiscale Attributed dataset (*Twitch*), by Rozemberczki et al., 2019 [25] 2021 [26] is part of a large non-homophily benchmark assembly of datasets. ‘Twitch’ is the name of a famous livestreaming website. Although the dataset is freely available upon API request, this version is cleaned and controlled by the authors. It shows the reciprocal followage (edges) between accounts (nodes) registered in the platform. Because of size and processing considerations the analysis includes only a subset of the data for accounts registered in Spain. Panel 1c shows said subset; because of the graphs’ structure and despite attempting several plotting techniques, this is the most insightful 2-D visualization I could obtain. My subset contains 4,638 nodes and 66,322 edges.

It is a highly non-homophilous and non-hierarchical graph that reflects the cheap costs of highly central accounts with degree ≈ 300 (famous streamers) incur in following much less famous ones. Because of its non-homophilous generation, it exhibits many d-dangling sets directly stemming from the highly connected nodes, warranting very poor Vanilla results, but good regularization prospects.

In terms of analytical comparisons, Table I reports select normalized graph metrics. Column one reports the eigengap bounds. Columns 3 and 2 report the empirical degree variance of the graphs and the standardized average degree respectively. The degree distribution of a graph is usually highly conditioned by its size. While degree normalization as in equation 3 aids (eigen) degree comparatives, heterogeneity metrics are still highly conditioned by both size and (normalized) degree.

TABLE I
COMPARISON OF THE DATASETS’ NORMALIZED INDICATORS

Graph/ Metric	λ_2, λ_n	$\frac{\bar{d}}{\sigma_d}$	$\hat{\sigma}_d$	$\frac{\hat{\sigma}_d - \sigma}{\hat{\sigma}_d \cdot \sigma} \bar{d}$	eigctr(λ)
Facebook	1.89, 0.004	0.93	46.3	5.65	0.08
Cora	1.98, 0.017	0.75	5.4	1.26	0.65
Twitch	1.91, 0.203	0.51	49.3	4.5	0.15

In an attempt to disentangle the effect of average degree in the variance, Column 4 reports the *degree heterogeneity in excess of uniformity* by subtracting the theoretical variance an Erdős–Rényi random graph of the same standardized degree would have. Since all edges would be drawn from a binomial with probability $p \equiv \frac{\bar{d}}{\hat{\sigma}_d}$, the theoretical variance σ would be $\sigma = np(1 - p)$. Such that

$$\frac{\bar{d}}{\sigma} - \frac{\bar{d}}{\hat{\sigma}_d} = \frac{\hat{\sigma}_d - \sigma}{\hat{\sigma}_d \cdot \sigma} \bar{d}$$

Would represent the average degrees any node could have in excess of its variance in the instance the graph was generated randomly. Finally, Column 5 reports the eigencentrality of the largest eigenvector as a measure of hierarchy in the graph’s core. The number steps required for a walk from the center to the periphery decreases with larger eigencentrality.

Among the most relevant observations, note Twitch’s large algebraic connectivity, with a degree heterogeneity similar to Facebook’s. Still it would appear it is much more hierarchical than it initially seems, and close to Facebook’s. Cora has a peculiar decentralized yet non-random (large eigen gap: 1.96) core that is highly non-hierarchical (0.65 eigencentrality) such that distant nodes are propagated quickly; this fact is confirmed by an unexpectedly low degree heterogeneity. Facebook is simply the baseline case.

A. Vanilla results

Figure 2 reports Vanilla clustering results for $k = 2, 3, 4$ (horizontal direction) per Shi-Malik’s, 2000 [9] algorithm in all three graphs (vertical direction). Clustering colors are also

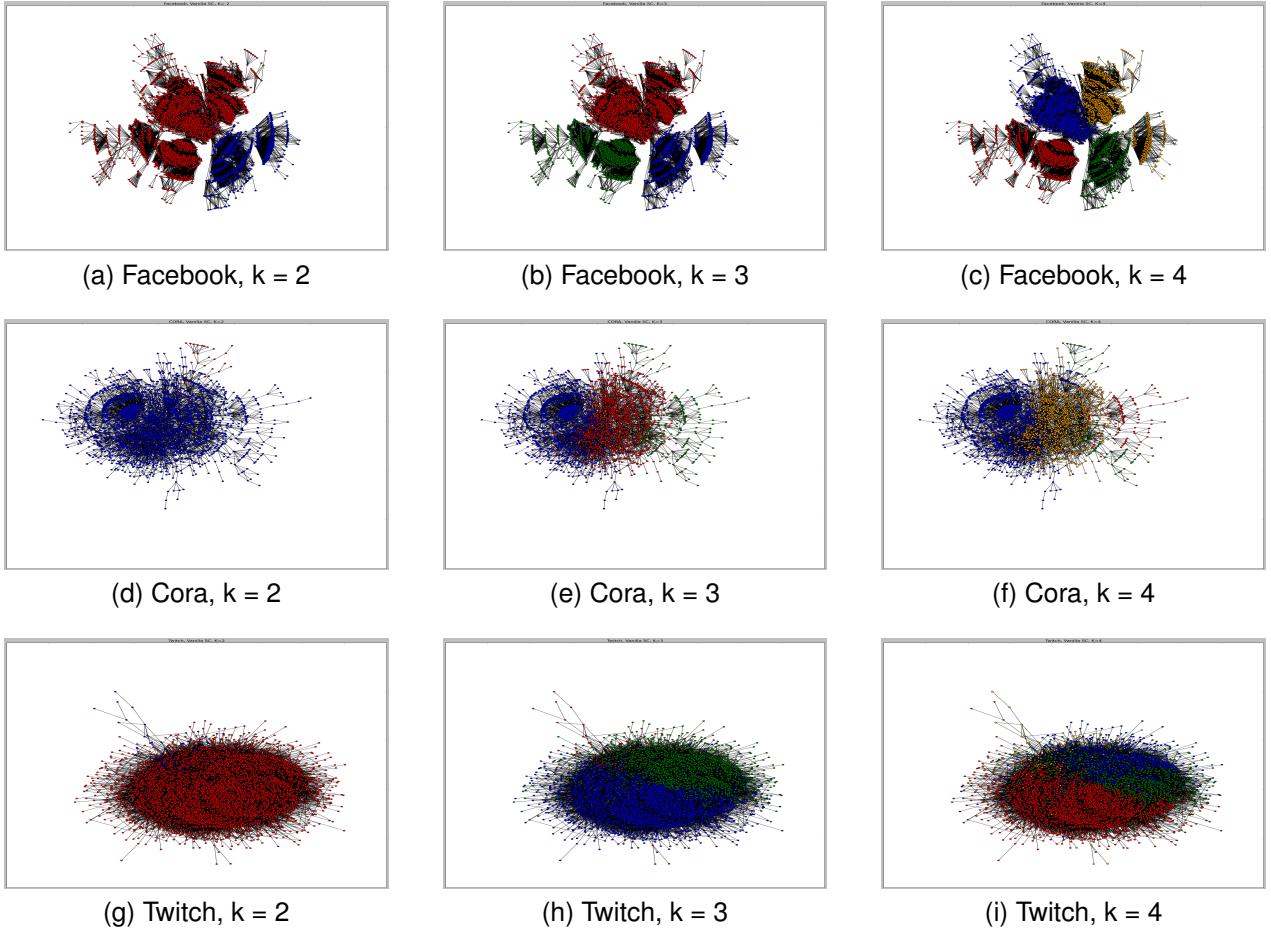


Fig. 2. Kamada-Kawai layouts of the Vanilla results for all analyzed datasets. Node color-ordered conductance from smallest to largest: blue, red, green, yellow.

sorted from lowest to highest conductance in this order: blue, red, green, yellow.¹ ². As preamble, Facebook is naturally a low-algebraically-connected, highly hierarchical graph, hence lowest conductance cuts are simply unions of the 6-9 visible conglomerates, divided in the number of specified clusters. Whether these already low-conductance allocations can be improved upon by regularization remains a question. Cora Vanilla partitions a very small d-dangling set for $k = 2$ (panel 2d) with a conductance almost identical to the minimal one. However, note in panels 2e, 2f once this crucial d-dangling set is accounted for, a large core cut is produced in an effect akin to regularization (further discussion in next section); said right component, however, seems to be quite vulnerable to further d-dangling sets partition (as shown in Figure 6 minimal Vanilla conductance increases for $k = 3, 5$; further discussion in the next section). Twitch results are analogous to Cora; the first partition is a d-dangling set and further partition reveals slightly more robust cuts (d-dangling sets are less visually identifiable because of how Twitch is plotted).

IV. HYPER PARAMETER TUNING, REGULARIZATION RESULTS, SIZE AND NUMBER EFFECTS

Before addressing the regularization results for both methods along with its implications for k-clustering and cluster size, I deem it important to outline my reasoning for the selection of τ and the scope of external validity it may have.

Both Amini et. al., 2013 [19] and Chaudhuri et. al., 2012 [21] limit themselves to provide the regularization magnitudes that work for the specific applications they were working on. And although posterior work [20], [1] attempts to provide general guidance on initial guesses and bounds for τ , the suggestions are effectively unusable in empirical graphs because they require specific model assumptions in one hand, and uniform topology on the other (g-dangling sets need to be all the same). Additionally, as explained in Section 2E we don't have guarantees on the rate, smoothness or uniformity of convergence for either algorithm.

This creates two practical problems when hyperparameter tuning τ for extrema of minimal conductance:

- Extrema may be a near-spike in between searched values of τ if $\Phi(\tau)$ is not well-behaved.
- Because we have no bounds for τ in most situations, nothing precludes extrema from existing "too close to

¹Space and clarity constraints prevent larger plots. High-quality figures can be zoomed, and are also available via the supplements.

²Full conductance Vanilla and regularized results for up to $k = 7$ reported in section 4.

the convergence result”, such that computational and time constraints may effectively bias the methods upwards.

In both cases, my best palliative, especially for low k was to conduct extensive, mostly unreported parameter searches, particularly until I could observe convergence to $\Phi(\cdot) = 1$ for RSC. Omitted extrema unfortunately can’t be ruled out. Hence in the following figures I limit myself to reporting the domain in which the chosen τ value was found.

Figure 3 shows my hyperparameter tuning in all three datasets (both RSC and Chaudhuri RSC) for $k = 2$ (Results for larger k included in the Appendix; in such cases convergence is faster and graphs seem generally smoother) with parameter search interval $[e^{-7}, e^{12}]$ divided in 400 log-linearly spaced values. The gray line denotes the Vanilla minimal conductance for the graph.

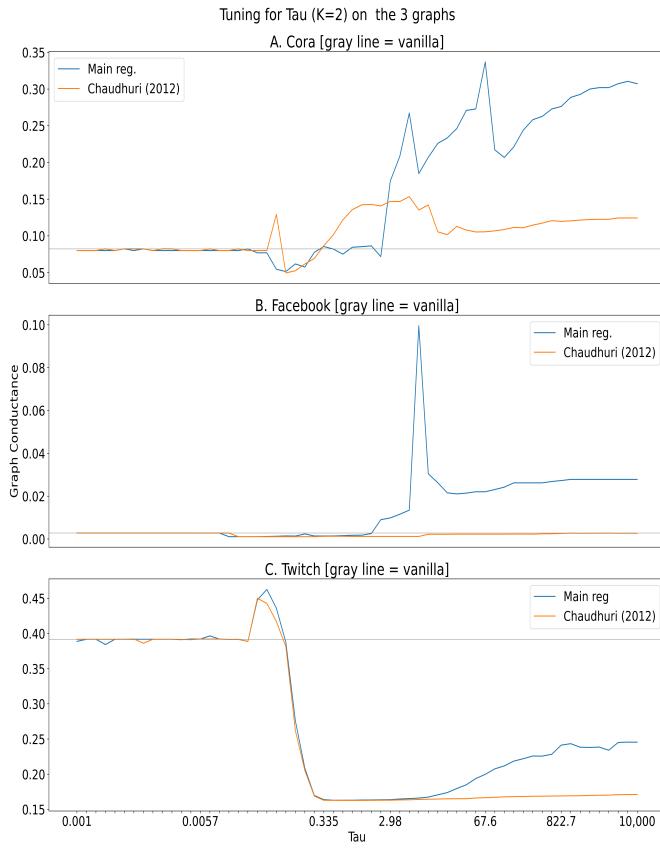


Fig. 3. Regularized minimal conductance (both algorithms) in all graphs for a range of values of hyperparameter τ

While it need not be the same, the regularized conductance for both methods shows a degree of correlation , especially around the regions where minima lie. Expectedly, some clustering results (especially those very established such as Facebook) are robust to wide variations in τ . Although it may appear minimal conductance τ is very similar in both methods, this is not the case for larger k , where they may fall several log points apart. Unfortunately, I don’t have clear evidence to state anything about the smoothness of $\Phi(\tau)$ or how the existence of an extrema may configure.

As regards the external validity of tuning tau, finding a graph sampling method for generating a comparable test-train

split on each dataset (especially for Twitch and Cora) is a nontrivial subject that likely exceeds the scope of this paper; distinct methods would favor a graph structure over others, thus creating a difficulty well beyond the subject I cover in this short work. Furthermore, regularizing a single sample is not moot: in the unsupervised and semi-supervised graph learning settings, correctly clustering a concrete graph is itself a valuable task where tuning as in this paper is an accepted practice. Hence I decided to restrict the analysis to self-tuned outcomes.

A. Facebook

Figure 4 summarizes the regularization outcomes of both Chaudhuri and RSC tuned as previously explained, for multiple ($k = 2, \dots, 7$) partitions of Facebook. Vanilla outcomes are also reported for comparison. To further illustrate the results, Figure 5 plots the regularized Facebook clusters in a fashion analogous to Figure 1, except exclusively for Facebook and one regularization approach is reported per row. Analogous reporting is made also for Cora and Twitch.

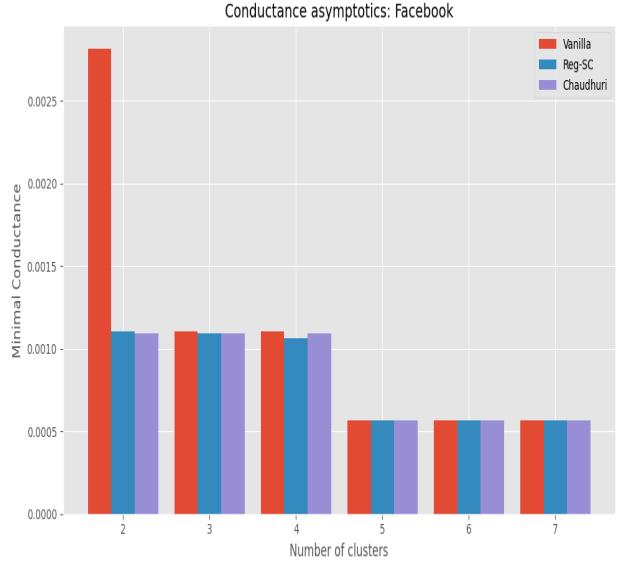


Fig. 4. Bar plot of the minimal conductance outcomes of Facebook, separated by number of partitions and colored by method.

Note regularization can greatly improve outcomes even in ideal cases such as this one. Both methods reduce conductance by roughly 66% for the bipartition instance. In both cases, a lower conductance combination of conglomerates is found in the south west quadrant of the plot (panels 5a , 5d) as opposed to the south east Vanilla allocation (2a). Small conductance differences between regularizations for $k = 2, 3$ are due to slightly different allocations of nodes in the margins of the clusters; however, from panels 5c, 5f one can also appreciate substantial differences between regularized allocations, even if they have almost the same conductance. Further note, even in this ideal case, examples where Chaudhuri performs better than RSC as vice-versa can both be found for $k = 4$ and $k = 2$ respectively. Finally, for $k > 4$ all three allocations are exactly the same.

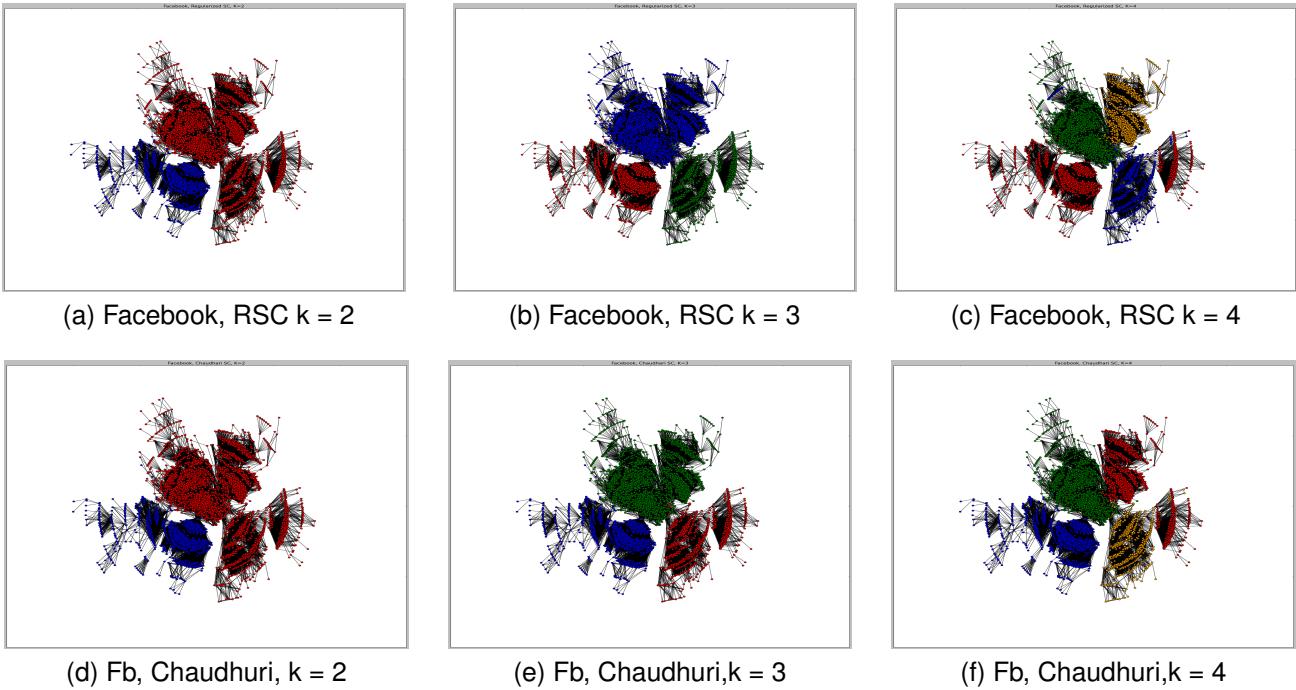


Fig. 5. Kamada-Kawai layouts of regularized clustering results for Facebook. Node color-ordered conductance from smallest to largest: blue, red, green, yellow.

B. Cora

Analogous to section 4A, figures 6 and 7 report full metrics and plot node allocations up to 4 partitions respectively.

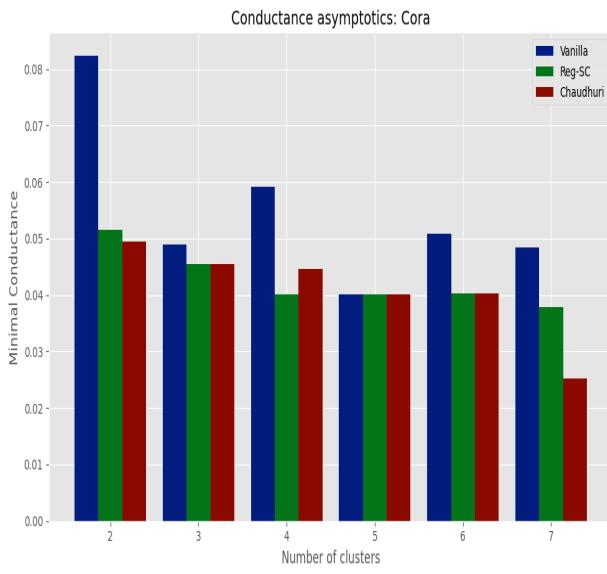


Fig. 6. Bar plot of the minimal conductance outcomes of Cora, separated by number of partitions and colored by method.

Generally speaking, this is an ambiguous instance of regularization improvement on outcomes. While minimal conductance of the regularized clustering methods is strictly improved for all k , the quality and informativeness of the allocations is –particularly for 2 and 3 partitions– perhaps lower than that of Vanilla.

Note from panel 7a that, in addition to the d-dangling set clustered in Vanilla (panel 2d), this cluster includes an outermost red set in the eastern part of the network; upon analysis in Python, this set includes 13 distinct disconnected components, 4 of which are trees with upwards of 10 nodes, and the rest are small branchouts of 2-6 sequential nodes. Apart from the obvious imbalance of the cut, its disconnected configurations are not particularly informative of the graph's structure. Rather, the addition of loose nodes at the margin likely reflects the bound in equation 8 and the impossibility of the two methods in increasing conductance enough. Chaudhuri in panel 7d is even slightly worse as it doesn't contain the Vanilla set.

With small differences, results for $k = 3$ are all qualitatively the same, and as discussed in the previous section, once those large quasi d-dangling sets are removed, a core partition is indeed produced. Admittedly, from panels 2f -7c, the pattern is repeated for $k = 4$; no qualitative change and a non-informative set is added. But if we look at panel 7f, we find an additional core partition that in one hand is larger in conductance than the combination of disconnected sets from Vanilla and RSC, but in the other produces the type of qualitative result both types of regularization aim for in most contexts. More generally, regularization for $k > 4$ doesn't appear to correct sparseness at all. For instance, while conductance for $k = 5$ is identical, regularized sets contain strictly more disconnected components. The same happens larger k even though conductances are lower.

Also note, given Cora and Twitch graph structures, increasing the number of partitions may increase conductance. In Cora's case, only Vanilla conductance and regularizations are non-increasing, but for Twitch we also find strictly in-

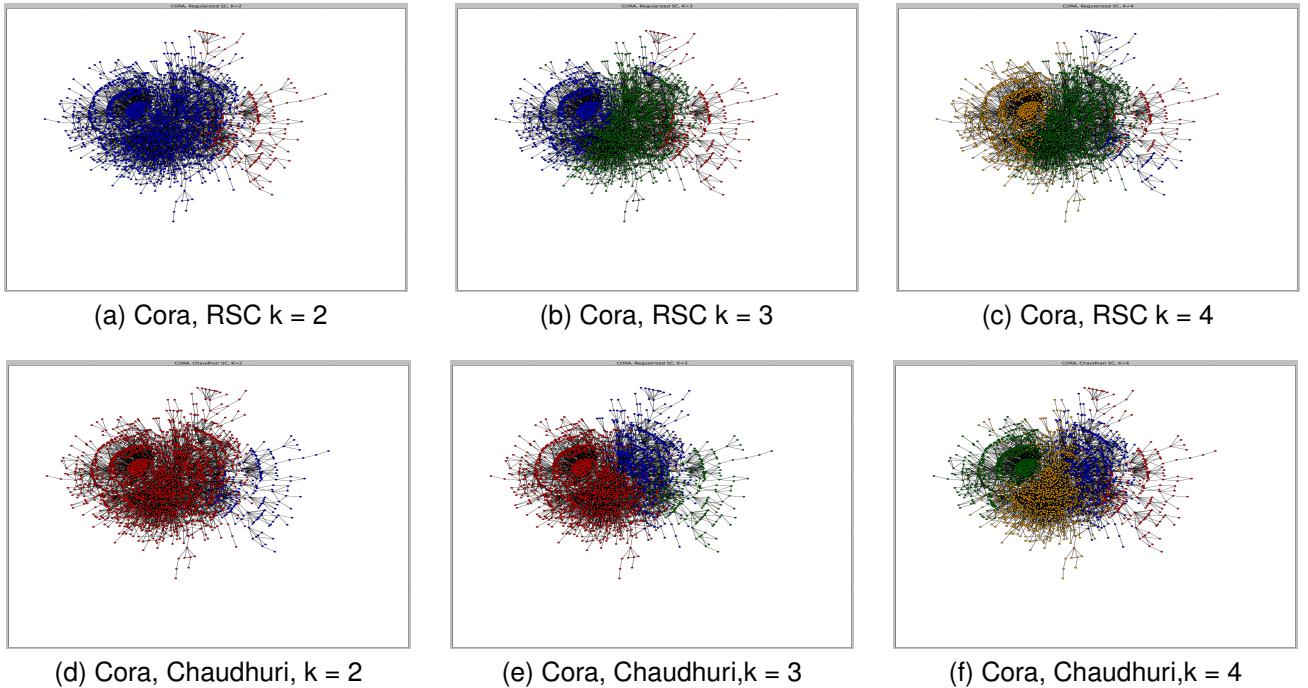


Fig. 7. Kamada-Kawai layouts of the regularized clustering results for Cora. Node color-ordered conductance from smallest to largest: blue, red, green, yellow.

creasing conductance. Particularly in the regularized cases, this shouldn't happen as long as separable d-dangling sets remain in the graph, although I cannot exclude other unstudied topologies having unexpected counteracting effects.

C. Twitch

Analogous to section 4B, figures 8 and 9 report full metrics and plot node allocations up to 4 partitions respectively.

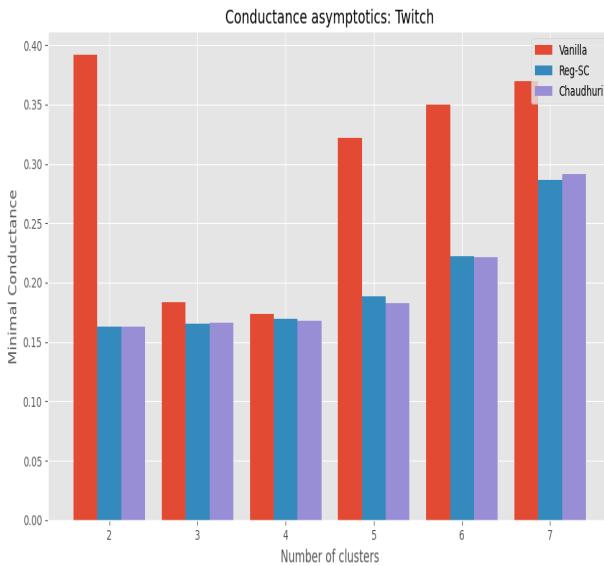


Fig. 8. Bar plot of the minimal conductance outcomes of Twitch, separated by number of partitions and colored by method.

Contrary to Cora, regularization on Twitch achieves the desired low-conductance core cut (Panels 9a, 9d vis-à-vis 2g)

for $k = 2$. Said cut in Vanilla was only possible for $k = 3$ after previously partitioning a sparse set. Note however, that regularized results for 3 partitions include said sparse set back almost identically (2h, 9b, 9e), hinting to the Zhang et al.'s (*ibid*) eigenvalues correction. Still, an interesting question remains on the differences between Cora and Twitch that allowed for the core cut to be prioritized in Twitch but not in Cora. Besides this observation and despite conductance, results are qualitatively the same for all k in this graph. Disconnected sets upon further partitions are the expected result in this type of high-algebraic-connectivity graph, hence the inevitable increase in conductance.

D. Effects of τ and k on conductance and cluster size considerations

Throughout this section (especially on Twitch and Facebook), it is shown there exist persistent low-conductance partitions of a graph. Although said partitions may appear in different order given regularization, an interesting phenomenon occurs whereby pairs (τ, k) of the parameters may produce quite similar and sometimes identical cuts: e.g. $(0, 3)$ and $([0.335, 3.15], 3)$ on Twitch produce roughly the same sets. Although I have no formal proof, I'd like to submit a good avenue of research would be to find direct and indirect effects of both k and τ on conductance, such that we can formulate the total change as

$$\Delta_\phi = a\Delta_\tau + b\Delta_k - \gamma(G) \cdot \sqrt{\Delta_\tau \Delta_k}$$

$a, b \in \mathbb{R}$, where $\Delta_\phi, \Delta_\tau, \Delta_k$ are the changes in conductance, regularization and number of clusters respectively, and $\gamma(\cdot)$ is a function of graph structure; effectively describing

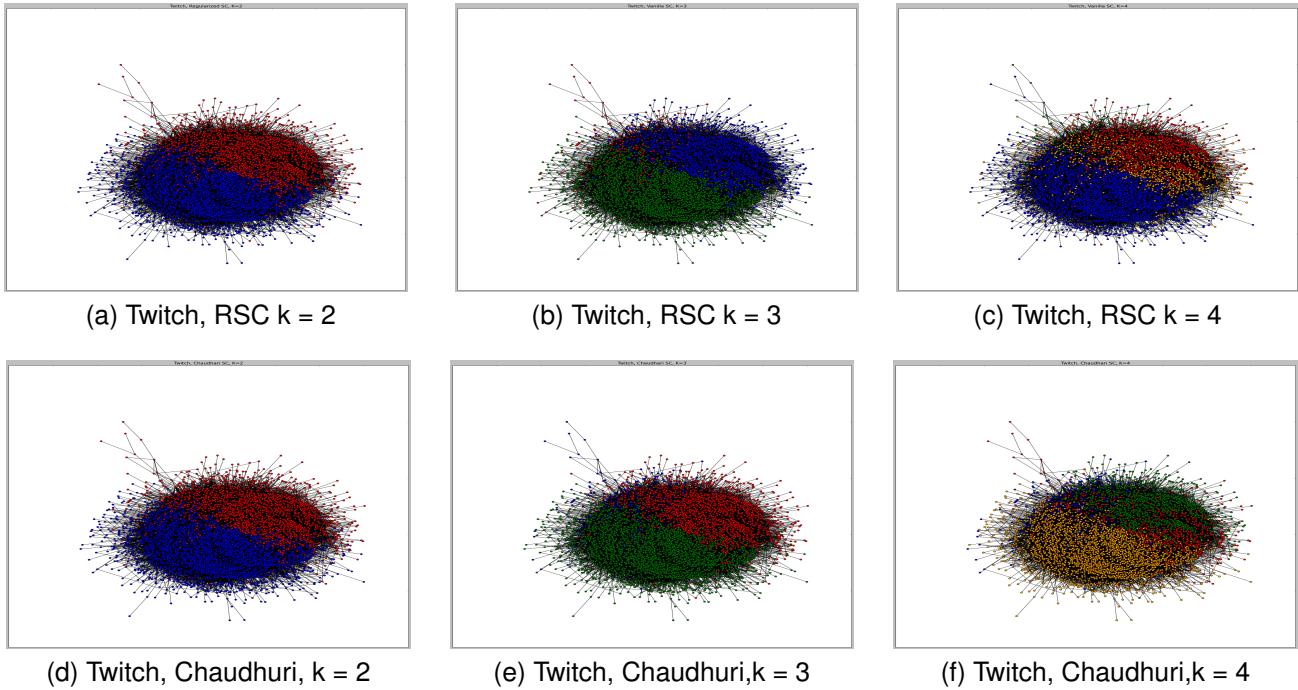


Fig. 9. Kamada-Kawai layouts of the regularized clustering results for Twitch. Node color-ordered conductance from smallest to largest: blue, red, green, yellow.

the covariance of k, τ . Although statistical identification of the changes in k would likely require being able to correctly validate the model to create an acceptable counterfactual.

Another important consideration in the recent literature, is that of the size for minimal conductance clusters under regularization. Figure 10 shows a scatter plot with relative cluster size for Vanilla clusters in the x-axis, and relative cluster size for regularized clusters in the y-axis.

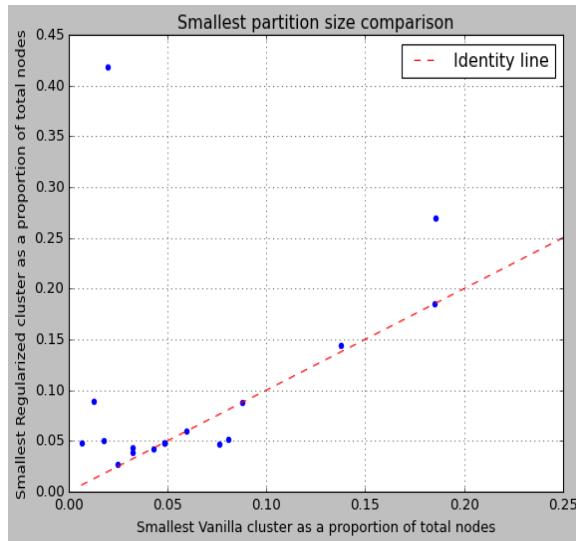


Fig. 10. Scatter plot of minimal cluster sizes (Vanilla vs regularizations)

Contrary to findings of related work usually affirming regularization increases relative size, I find no clear relationship (perhaps a weak grouping around the identity line) between

sizes. This is not surprising as my choice of datasets aims precisely at instances where the so-far-conventional results don't apply.

V. CONCLUSIONS

After detailed examination of two regularization methods for the popular k -partitions spectral clustering algorithm in datasets akin to resemble general graph structures, I intend to fill knowledge gaps that, in my view, have not been either theoretically or empirically answered:

Even though regularization methods are weakly better-performing than Vanilla SC, none of the regularization methods is strictly better than the other. Both best-perform contingent on graph structure and number of partitions. Based on my datasets, the method based on Chaudhuri et al., 2012 [21] seems to perform better in less hierarchical, more algebraically connected graphs, while the method by Amini et al., 2013 [19] seems to perform better in hierarchical graphs with sharp rather than smooth solutions.

As the number of partitions grows, I do not find exact convergence to Vanilla conductance or clusters, yet admittedly as k grows I find substantial qualitative similarity in the appearance of the clusters, with some punctual deviations of this pattern, especially for Chaudhuri et al.'s method. While regularization sometimes generates "novel" cuts that wouldn't exist for large k in Vanilla, my general observation is these methods aid in "prioritizing" persistent low-conductance cuts that already show when we "subtract" d-dangling sets and other non-informative components by increasing the number of partitions in Vanilla. I use a simple yet incomplete way to compare the effect of increasing the number of partitions with

that of regularization by comparing the Vanilla counterfactual for $k + 1$ with the regularization result τ for k , but I don't manage to answer what the marginal effect of either parameter in conductance is.

In terms of the choice of τ , I find, in absence of theoretical guarantees, simple hyperparameter tuning is vulnerable to non-smooth validation or lack of uniform convergence to large values of the parameter; in fact I still sustain doubts about my choice of regularization for the results in Section 4C. Validation of the choice of τ for generalizability of a specific type of graph should not be problematic, but it is still unclear whether choosing the minimal conductance minimizer is a proper choice, particularly because some of my results show regularizations where, despite improvements in conductance, the resulting clusters consist of many disconnected, sparse sets with no intuitive or heuristic significance. Therefore regularization can worsen clustering quality, particularly in non-homophilous graphs.

APPENDIX HYPERPARAMETER TUNING FOR K = 4 PARTITIONS

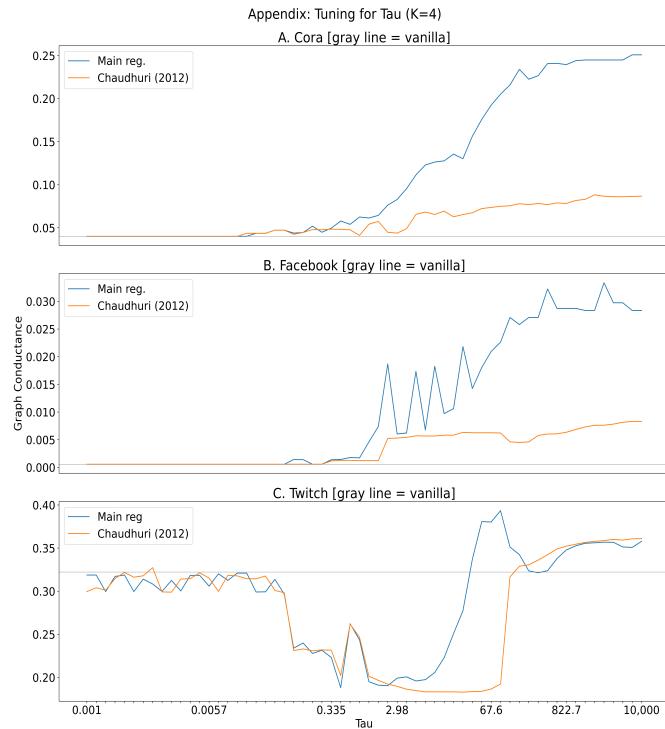


Fig. 11. Simulation results for the network.

REFERENCES

- [1] Y. Zhang and K. Rohe, "Understanding regularized spectral clustering via graph conductance," 2018.
- [2] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Mathematical Journal*, vol. 23, no. 2, pp. 298–305, 1973.
- [3] B. Arsić, D. Cvetković, S. K. Simić, and M. Škarić, "Graph spectral techniques in computer sciences," *Applicable Analysis and Discrete Mathematics*, vol. 6, no. 1, pp. 1–30, 2012. [Online]. Available: <http://www.jstor.org/stable/43666153>
- [4] P. W. Holland, K. B. Laskey, and S. Leinhardt, "Stochastic blockmodels: First steps," *Social Networks*, vol. 5, no. 2, pp. 109–137, 1983. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0378873383900217>
- [5] J. Vatter, J. Vatter, R. Mayer, and H.-A. Jacobsen, "The evolution of distributed systems for graph neural networks and their origin in graph processing and deep learning: A survey," *ACM computing surveys.*, vol. 56, no. 1, pp. 1–37, 2024.
- [6] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, "Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters," 2008.
- [7] N. Binkiewicz, J. T. Vogelstein, and K. Rohe, "Covariate-assisted spectral clustering," *Biometrika*, vol. 104, no. 2, pp. 361–377, mar 2017. [Online]. Available: <https://doi.org/10.1093%2Fbiomet%2Fax008>
- [8] Y. Zhang, M. Poux-Berthe, C. Wells, K. Koc-Michalska, and K. Rohe, "Discovering political topics in facebook discussion threads with graph contextualization," 2018.
- [9] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, Aug 2000.
- [10] M. Fiedler, "A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory," *Czechoslovak Mathematical Journal*, vol. 25, no. 4, pp. 619–633, 1975. [Online]. Available: <http://eudml.org/doc/12900>
- [11] L. W. Hagen and A. B. Kahng, "New spectral methods for ratio cut partitioning and clustering," *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*, vol. 11, pp. 1074–1085, 1991. [Online]. Available: <https://api.semanticscholar.org/CorpusID:17757903>
- [12] D. A. Spielman and S.-H. Teng, "Spectral partitioning works: Planar graphs and finite element meshes," EECS Department, University of California, Berkeley, Tech. Rep. UCB/CSD-96-898, Feb 1996. [Online]. Available: <http://www2.eecs.berkeley.edu/Pubs/TechRpts/1996/5359.html>
- [13] A. Dasgupta, J. Hopcroft, R. Kannan, and P. Mitra, "Spectral clustering by recursive partitioning," pp. 256–267, 09 2006.
- [14] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006. [Online]. Available: <https://www.pnas.org/doi/abs/10.1073/pnas.0601602103>
- [15] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems*, T. Dietterich, S. Becker, and Z. Ghahramani, Eds., vol. 14. MIT Press, 2001. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2001/file/801272ee79cfde7fa5960571fee36b9b-Paper.pdf
- [16] L. Dall'Amico, R. Couillet, and N. Tremblay, "Revisiting the bethe-hessian: Improved community detection in sparse heterogeneous graphs," 2019.
- [17] ———, "A unified framework for spectral clustering in sparse graphs," 2021.
- [18] R. Kannan, S. Vempala, and A. Vetta, "On clusterings: Good, bad and spectral," *J. ACM*, vol. 51, no. 3, p. 497–515, may 2004. [Online]. Available: <https://doi.org/10.1145/990308.990313>
- [19] A. A. Amini, A. Chen, P. J. Bickel, and E. Levina, "Pseudo-likelihood methods for community detection in large sparse networks," *The Annals of Statistics*, vol. 41, no. 4, aug 2013. [Online]. Available: <https://doi.org/10.1214%2F13-aos1138>
- [20] T. Qin and K. Rohe, "Regularized spectral clustering under the degree-corrected stochastic blockmodel," 2013.
- [21] K. Chaudhuri, F. Chung, and A. Tsitsias, "Spectral clustering of graphs with general degrees in the extended planted partition model," JMLR Workshop and Conference Proceedings, pp. 35–1, 2012.
- [22] J. Leskovec and A. Krevl, "SNAP Datasets: Stanford large network dataset collection," <http://snap.stanford.edu/data>, Jun. 2014.
- [23] A. K. McCallum, K. Nigam, J. Rennie, and K. Seymore, "Automating the construction of internet portals with machine learning," *Information Retrieval*, vol. 3, no. 2, pp. 127–163, 07 2000, copyright - Copyright (c) 2000 Kluwer Academic Publishers; Last updated - 2023-08-17. [Online]. Available: <https://www.proquest.com/scholarly-journals/automating-construction-internet-portals-with/docview/230645231/se-2>
- [24] S. P. Borgatti and M. G. Everett, "Models of core/periphery structures," *Social Networks*, vol. 21, no. 4, pp. 375–395, 2000. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0378873399000192>
- [25] B. Rozemberczki, C. Allen, and R. Sarkar, "Multi-scale attributed node embedding," 2019.
- [26] B. Rozemberczki and R. Sarkar, "Twitch gamers: a dataset for evaluating proximity preserving and structural role-based node embeddings," 2021.