p-Norm Flow Diffusion for Local Graph Clustering

Kimon Fountoulakis [*1] **Di Wang** [*2] **Shenghao Yang** [*1]

Abstract

Local graph clustering and the closely related seed set expansion problem are primitives on graphs that are central to a wide range of analytic and learning tasks such as local clustering, community detection, semi-supervised learning, nodes ranking and feature inference. Prior work on local graph clustering mostly falls into two categories with numerical and combinatorial roots respectively, in this work we draw inspiration from both fields and propose a family of convex optimization formulations based on the idea of diffusion with p-norm network flow for $p \in (1, \infty)$. In the context of local clustering, we characterize the optimal solutions for these optimization problems and show their usefulness in finding low conductance cuts around input seed set. In particular, we achieve quadratic approximation of conductance in the case of $p = 2$ similar to the Cheeger-type bounds of spectral methods, constant factor approximation when $p \to \infty$ similar to max-flow based methods, and a smooth transition for general p values in between. Thus, our optimization formulation can be viewed as bridging the numerical and combinatorial approaches, and we can achieve the best of both worlds in terms of speed and noise robustness. We show that the proposed problem can be solved in strongly local running time for $p \geq 2$ and conduct empirical evaluations on both synthetic and real-world graphs to illustrate our approach compares favorably with existing methods.

[*]Equal contribution [1]School of Computer Science, University of Waterloo, Canada [2]Google Research, USA. Correspondence to: Kimon Fountoulakis <kimon.fountoulakis@uwaterloo.ca>, Di Wang <wadi@google.com>, Shenghao Yang <shenghao.yang@uwaterloo.ca>.

Proceedings of the 37^{th} International Conference on Machine Learning, Online, PMLR 119, 2020.

1. Introduction

Graphs are ubiquitous when it comes to modeling relationships among entities, e.g. social (Traud et al., 2012) and biology (Tuncbag et al., 2016) networks, and graphs arising from modern applications are massive yet rich in small-scale local structures (Leskovec et al., 2009; Jeub et al., 2015; Fortunato & Hric, 2016). Exploiting such local structures is of central importance in many areas of machine learning and applied mathematics, e.g. community detection in networks (Ng et al., 2001; White & Smyth, 2005; Leskovec et al., 2009; Jeub et al., 2015) and PageRank-based spectral ranking in web ranking (Page et al., 1999; Gleich, 2015). Somewhat more formally, we consider local graph clustering as the task of finding a community-like cluster around a given set of seed nodes, where nodes in the cluster are densely connected to each other while relatively isolated to the rest of the graph. Moreover, an algorithm is called *strongly local* if it runs in time proportional to the size of the output cluster rather than the size of the whole graph.

Strongly local algorithms for local graph clustering are predominantly based on the idea of diffusion, which is the generic process of spreading mass among vertices by sending mass along edges. This connection has been formalized in many previous results (Spielman & Teng, 2013; Andersen et al., 2006; Wang et al., 2017). Historically, the most popular diffusion methods are spectral methods (e.g. random walks) based on the connection between graph structures and the algebraic properties of the spectrum of matrices associated with the graph (Lovász & Simonovits, 1990; 1993; Chung, 2007b; Tsiatas, 2012). Spectral diffusion methods are widely applied in practice due to the ease of implementation, efficient running time and good performances in many contexts. However, it is also known in theory and in practice that spectral methods can spread mass too aggressively and not find the right cluster when structural heterogeneities exist, and thus are not robust on real-world graphs constructed from very noisy data (Guattery & Miller, 1998; von Luxburg, 2006; Leskovec et al., 2009; Jeub et al., 2015). A more recent line of work on diffusion is based on the combinatorial idea of max flow exploiting the canonical duality between flow and cut (Wang et al., 2017; Fountoulakis et al., 2017a; Orecchia & Zhu, 2014). These methods offer improved theoretical guarantees in terms of locating local cuts, and have been shown to be important for pathological cases where

spectral methods do not perform well in practice. However, combinatorial methods are generally accepted to be more difficult to understand and implement in practice due to the more complicated underlying dynamics.

In this paper, we propose and study a family of primal and dual convex optimization problems for local graph clustering. We call the primal problem p-norm flow diffusion, parameterized by the ℓ_p-norm used in the objective function, and the problem defines a natural diffusion model on graphs using network flow. We refer the dual problem as the q-norm local cut problem where q is the dual norm of p (i.e. $1/p + 1/q = 1$). The optimal solution to the q-norm local cut problem can be used to find good local clusters with provable guarantees. Throughout our discussion, we use p-norm diffusion to refer both the primal and the dual problems since our main inspiration comes from diffusion, even though our results are technically for the q-norm local cut.

We note that almost all previous diffusion methods are defined with the dynamics of the underlying diffusion procedure, i.e. the step-by-step rules of how to send mass, and the analysis of these methods is based on the behaviors of the algorithm. Although there are some efforts on interpreting the algorithms with optimization objectives (Fountoulakis et al., 2017a;b; Mahoney et al., 2012), this line of research remains predominantly bottom-up starting with the algorithmic operations. On the other hand, our work starts with a clear optimization objective, and analyze the properties of the optimal solution independent from what method is used to solve the problem. This top-down approach is distinct in theory, and is also very valuable in practice since the de-coupling of objective and algorithm gives the users the freedom at implementation to choose the most suitable solver based on availability of infrastructure and code-base.

1.1. Our Main Contributions

We refer readers to Section 2 for formal discussion. Our first main result is a novel theoretical analysis for local graph clustering using the optimal solution of p-norm diffusion. In particular, suppose there exists a cluster B with conductance $\phi(B)$, and we are given a set of seed nodes that overlaps reasonably with B. Then the optimal solution of p-norm diffusion can be used to find a cluster A with conductance at most $\mathcal{O}(\phi(B)^{1/q})$. For $p = 2$, this result resembles the Cheeger-type quadratic guarantees that are well-known in spectral-based local graph clustering literature (Spielman & Teng, 2013; Andersen et al., 2006). When $p \to \infty$, our conductance guarantee approaches a constant factor approximation similar to max flow methods, while achieving a smooth transition for general p values in between. We observe in practice that our optimization formulation can achieve the best of both worlds in terms of speed and robustness to noise for when p lies in the range of small constants, e.g. $p = 4$.

On the algorithm side, we show that a randomized coordinate descent method can obtain an ϵ accurate solution of p-norm diffusion for $p \geq 2$ in strongly local running time. The running time of our algorithm captures the effect that it takes longer for the algorithm to converge for larger p values. Although the iteration complexity analysis is not entirely new, we show a crucial result on the monotonicity of the dual variables, which establishes the strongly local running time of the algorithm.

Our analysis illustrates a natural trade-off as a function of p between robustness to noise and the running time for solving the dual of the p-norm diffusion problem. In particular, for $p = 2$ the diffusion problem can be solved in time linear in the size of the local cluster, but may have quadratic approximation error $\mathcal{O}(\sqrt{\phi(B)})$. On the other hand, the approximation error guarantee improves when p increases, but it also takes longer to converge to the optimal solution. We believe the regime of p being small constants offer the best trade-offs in general.

1.2. Previous Work

The local clustering problem is first proposed and studied by (Spielman & Teng, 2013). Their algorithm Nibble is a truncated power method with early termination, and their result were later improved to $\tilde{\mathcal{O}}(\sqrt{\phi(B)})$ conductance approximation and $\tilde{\mathcal{O}}\left(\frac{\text{vol}(B)}{\phi(B)}\right)$ time using approximate personalized PageRank (Andersen et al., 2006), which is one of the most popular local clustering methods. The EvoCut algorithm (Andersen & Peres, 2009) is the fastest spectral based local clustering method with running time $\tilde{\mathcal{O}}(\frac{\text{vol}(B)}{\sqrt{\phi(B)}})$. In (Zhu et al., 2013) the authors analyzed the behavior of approximate personalized PageRank under certain intra-cluster well-connected conditions to give strengthened results. There are many other spectral based diffusions, examples include local Lanczos spectral approximations (Shi et al., 2017), evolving sets (Andersen et al., 2016), seed expansion methods (Kloumann & Kleinberg, 2014) and heat-kernel PageRank (Chung, 2009; Kloster & Gleich, 2014). Note that the above methods based on spectral diffusion are subject to the quadratic approximation error in worst case, informally known as the Cheeger barrier.

Combinatorial methods for local graph clustering are mostly based on the idea of max flow. Some examples include the flow-improve method (Andersen & Lang, 2008), the local flow-improve method (Orecchia & Zhu, 2014) and the capacity releasing diffusion (Wang et al., 2017). The former relies on black-box max flow primitives, while the latter two require specialized max flow algorithms with early termination to to achieve running time that is nearly linear in the size of the local cluster. These algorithms achieve constant

approximation error, i.e., the output cluster has conductance $\mathcal{O}(\phi(B))$, as opposed to quadratic approximation error for spectral methods.

The first graph clustering method to explicitly incorporate norms beyond $p = 2$ is the p-spectral clustering proposed by (Bühler & Hein, 2009), based on the notion of graph p-Laplacians initially studied in (Amghibech, 2003). It generalizes the standard spectral approach for global graph partitioning and achieves tighter approximations beyond the Cheeger barrier. Similar ideas are later extended to the context of hypergraphs (Li & Milenkovic, 2018). All these methods rely on global analysis and computation of eigenvalues and eigenvectors, and thus do not enjoy the same properties that local methods have.

2. Preliminaries and Notations

We consider un-directed connected graph G with V being the set of nodes and E the set of edges. For simplicity we focus on un-weighted graphs in our discussion, although our result extends to the weighted case in a straightforward manner. The degree $\deg(v)$ of a node $v \in V$ is the number of edges incident to it, and we denote $\vec{\boldsymbol{d}}$ as the vector of all nodes' degrees and $D = \text{diag}(\vec{\boldsymbol{d}})$. We refer to $\text{vol}(C) = \sum_{v \in C} \deg(v)$ as the *volume* of $C \subseteq V$. We use subscripts to indicate what graph we are working with, while we omit the subscripts when the graph is clear from context.

A *cut* is treated as a subset $S \subset V$, or a partition $(S, \overline{S})$ where $\overline{S} = V \setminus S$. For any subsets $S, T \subset V$, we denote $E(S,T) = \{\{u,v\} \in E \mid u \in S, v \in T\}$ as the set of edges between S and T. The *cut-size* of a cut S is $\delta(S) = |E(S, \overline{S})|$. The *conductance* of a cut S in G is $\Phi_G(S) = \frac{\delta(S)}{\min(\text{vol}_G(S), \text{vol}_G(V \setminus S))}$. Unless otherwise noted, when speaking of the conductance of a cut S, we assume S to be the side of smaller volume. The conductance of a graph G is $\Phi_G = \min_{S \subset V} \Phi_G(S)$.

A *routing* (or *flow*) is a function $f : E \to \mathbb{R}$. For each undirected edge e with endpoints u, v, we arbitrarily orient the edge as $e = (u, v)$, i.e. from u to v. The magnitude of the flow over e specifies the amount of mass routed over e, and the sign indicates whether we send flow in the forward or reverse direction of the orientation of edge e, i.e. $f(u, v)$ is positive if mass is sent from u to v and vice versa. We abuse the notation to also use $f(v, u) = -f(u, v)$ for an edge $e = (u, v)$. We denote B as the signed incidence matrix of the graph of size $|E| \times |V|$ where the row of edge $e = (u, v)$ (again, using the arbitrary orientation) has two non-zeros entries, -1 at column u and 1 at column v. Throughout our discussion we refer to a function over edges (or nodes) and its explicit representation as an $|E|$-dimensional (or $|V|$-dimensional) vector interchangeably.

3. Diffusion as Optimization

Given a graph G with signed incidence matrix B, and two functions $\Delta, T : V \to \mathbb{R}_{\geq 0}$, we propose the following pair of convex optimization problems, which are the p-norm flow diffusion

$$\begin{aligned} \text{minimize } & \|f\|_p \\ \text{s.t. } & B^T f + \Delta \leq T \end{aligned} \tag{1}$$

and its dual formulation with q such that $1/q + 1/p = 1$

$$\begin{aligned} \text{maximize } & (\Delta - T)^T x \\ \text{s.t. } & \|Bx\|_q \leq 1 \\ & x \geq 0. \end{aligned} \tag{2}$$

The solution to the dual problem $x \in \mathbb{R}^{|V|}_{\geq 0}$ gives an embedding of the nodes on the (non-negative) real line. This embedding is what we actually compute in the context of local clustering, and we use the primal problem and its flow solution $f \in \mathbb{R}^{|E|}$ mostly for insights and analysis purposes. We discuss the interpretation of the primal problem as a diffusion next.

The Primal Problem. As mentioned earlier, we consider a diffusion on a graph $G = (V, E)$ as the task of spreading mass from a small set of nodes to a larger set of nodes. More formally, the function Δ will specify the amount of initial mass starting at each node, and the function T will give the sink capacity of each node, i.e. the most amount of mass we allow at a node after the spreading. We denote the *density* (of mass) at a node v as the ratio of the amount of mass at v over $\deg(v)$, and when we use density without any specific node, we mean the maximum density at any node. Naturally in a diffusion, we start with Δ having small support and high density, and the goal is to reach a state with bounded density enforced by the sink function. This gives a clean physical interpretation where paint (i.e. mass) spills from the source nodes and spreads over the graph and there is a sink at each node where up to a certain amount of paint can settle[1].

In our work, we will always use the particular sink capacity function where $T(v) = \deg(v)$ for all nodes, i.e. density at most 1. We extend the notation to write $\Delta(S) = \sum_{v \in S} \Delta(v)$ and $T(S) = \sum_{v \in S} T(v)$ for a subset of nodes S, and for our particular choice of sink capacity we have $T(S) = \text{vol}(S)$. We also write $|\Delta| = \Delta(V)$ as the total amount of mass we start with, which remains constant throughout the diffusion as flow routing conserves mass.

Given initial mass Δ and routing f, the vector $\vec{m} = B^T f + \Delta$ gives the amount of mass $m(v)$ at each node v after the

[1] There is also a "paint spilling" interpretation for personalized PageRank where instead of sinks holding paint, the paint dries (and settles) at a fixed rate when it pass through nodes. These are two very different mechanisms on how the mass settles.

routing f, i.e. $m(v) = \Delta(v) + \sum_u f(u,v)$ is the sum of initial mass and the net amount of mass routed to v. We say f is a *feasible routing* for a diffusion problem when $m(v) \leq T(v)$ for all nodes, i.e. the mass obeys the sink capacity at each node. We say v's sink is *saturated* if $m(v) \geq T(v)$ and $\text{ex}(v) = \max(m(v) - T(v), 0)$ the *excess* at v. Note there always exists some feasible routing as long as the total amount of mass $|\Delta|$ is at most $\text{vol}(G)$, i.e. there is enough sink capacity over the entire graph to settle all the mass. This will be the case in the context of local clustering, and we will assume this implicitly through our discussion.

The goal of our diffusion problem is to find a feasible routing with low cost. We consider the p-norm of a routing $\|f\|_p = (\sum_e f_e^p)^{1/p}$ as its cost. For example, when $p = 2$ we can view the flow as electrical current, then the cost is the square root of the energy of the electrical flow; when $p = \infty$ the cost corresponds to the most congested edge's congestion of the routing. For $p < \infty$, the cost will naturally encourage the diffusion to send mass to saturate nearby sinks before expanding further, and thus our model inherently looks for local solutions.

For reader familiar with the network flow literature, in the canonical p-norm flow problem, we are given the exact amount of mass required at each sink, i.e. the inequality constraint is replaced by equality, and the high level question is *how* to route mass efficiently from given source(s) to destination(s). In our diffusion problem, as we have the freedom to choose the destination of mass as long as we obey the sink capacities, the essential question becomes *where* to route the mass so the spreading can be low cost. Despite their similarity and close connection, we believe the distinct challenge of our p-norm diffusion problem poses a novel and meaningful direction to the classic problem of network flow.

The Dual Problem. It is straightforward to check problem (2) is the dual of p-norm flow diffusion, and strong duality holds for our problems so they have the same optimal value. For the dual problem, we view a solution x as assigning heights to nodes, and the goal is to separate the nodes with source mass (i.e. seed nodes) from the rest of the graph. This is reflected in the objective where we gain by raising nodes with large source mass higher but loss by raising nodes in general. If we consider the height difference between an edge's two endpoints as the *length* of an edge, i.e. $|x(u) - x(v)|$, we constrain the separation of nodes with a budget for how much we can stretch the edges. More accurately, the q-norm of the vector of edge lengths (i.e. $\|Bx\|_q$) is at most 1. This naturally encourages stretching edges along a bottleneck (i.e. cut with small number of edges crossing it) around the seed nodes, since we can stretch each edge more when the number of edges is smaller (and thus raise seed nodes higher). The dual problem also inherently looks for local solutions as raising nodes far away from the source mass only hurts the objective.

In contrast to random walk based linear operators such as personalized PageRank and heat kernel, even 2-norm diffusion is non-linear due to the combinatorial constraint $B^T f + \Delta \leq T$. More generally, introducing non-linearity has proved to be very successful in machine learning, most notably in the context of deep neural networks. This may offer some intuition why 2-norm diffusion achieves better results empirically comparing to personalized PageRank despite the connection between 2-norm diffusion and spectral methods.[2]

4. Local Graph Clustering

In this section we discuss the optimal solutions of our diffusion problem (and its dual) in the context of local graph clustering. At a high level, we are given a set of seed nodes S and want to find a cut of low conductance close to these nodes. Following prior work in local clustering, we formulate the goal as a promise problem, where we assume the existence of an unknown good cluster $C \subset V$ with $\text{vol}(C) \leq \text{vol}(V)/2$ and conductance $\phi_G(C) = \phi^*$. We consider a generic fixed $G = (V, E)$ and $p \in (1, \infty)$ through our discussion. We keep our discussion at a high level, and defer missing proofs and technical details to the supplementary material.

4.1. Diffusion Setup

To specify a particular diffusion problem and its dual, we need to provide the source mass Δ, and recall we always set the sink capacity $T(v) = \deg(v)$. Given a set of seed nodes S, we pick a scalar δ and let

$$\Delta(v) = \begin{cases} \delta \cdot \deg(v) & \text{if } v \in S, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Note this gives the total amount of mass $|\Delta| = \delta \cdot \text{vol}(S)$. We will discuss the choice of δ shortly, but we start with a simple lemma on the locality of the optimal solutions for the primal and dual problems.

Lemma 1. *Let f^* and x^* be optimal solutions of (1) and (2) respectively, $\text{supp}(f^*)$ be the subset of edges with non-zero mass crossing them (i.e. the support of vector f^*), and $\text{supp}(x^*)$ be the subset of nodes with strictly positive dual value. We have*

1. $|\text{supp}(f^*)| \leq \delta \cdot \text{vol}(S)$,

[2] We note that algorithms (e.g. (Andersen et al., 2006)) based on random walks nonetheless introduce non-linearity (and also strong locality) to the underlying linear model in the form of approximation or regularization, whereas our model is intrinsically non-linear and strongly local.

2. $\text{vol}_G(\text{supp}(x^*)) \leq \delta \cdot \text{vol}(S)$, *and*

3. $x^*(u) > 0$ *only if* $(B^T f^* + \Delta)(u) = \deg(u)$.

Proof sketch. For $p < \infty$, the optimal routing will never push mass out of a node u unless u's sink is saturated, i.e. $f^*(u, v) > 0$ for u, v only if $(B^T f^* + \Delta)(u) = \deg(u)$, since otherwise we can reverse the mass routed out of the un-saturated node to reduce the cost of the routing. The total amount of mass $\delta \cdot \text{vol}(S)$ upper-bounds the total volume of the saturated nodes since it takes $\deg(v)$ amount of mass to saturate the node v. This observation proves the first claim. The second claim also follows from the same observation and the additional property of complementary slackness, which is exactly the third claim. □

Now we discuss how to set δ. The intuition is that we want the total amount of source mass starting in C to be a constant factor larger than the volume of C, say $\Delta(C) \geq 2\text{vol}(C)$ (any constant reasonably larger than 1 would work). The reason is that in such scenario, at least $\Delta(C) - \text{vol}(C) \geq \text{vol}(C)$ amount of mass has to be routed out of C since the nodes in C have total sink capacity $\text{vol}(C)$. When C is a cut of low conductance, any feasible routing must incur a large cost since $\text{vol}(C)$ amount of mass has to get out of C using a relatively small number of discharging edges. In this case, the optimal dual solution x^* will certify the high cost of any feasible primal solution. Naturally, the appropriate value of δ to get $\Delta(C) \geq 2\text{vol}(C)$ depends on how well the seed set S overlaps with C. Suppose $\text{vol}(S \cap C) \geq \alpha\text{vol}(C)$, then we can set $\delta = 2/\alpha$. Without loss of generality, we assume the right value of δ is known since otherwise we can employ binary search to find a good value of δ.

More formally, we derive a low conductance cut from x^* using the standard sweep cut procedure. In our case, because x^* has bounded support, the procedure can be implemented in $O(\delta \cdot \text{vol}(S))$ total work.

1. Sort the nodes in decreasing order by their values in x^*.
2. For each $i \geq 1$ such that the i-th node still has strictly positive dual value, consider the cut containing the first i nodes. Among all these cuts (also called *level cuts*) output the one with the smallest conductance.

Figure 1. **The Sweep Cut Procedure.**

4.2. Local Clustering Guarantee

Theorem 2. *Given a set of seed nodes S, suppose there exists a cluster C such that*

1. $\text{vol}(S \cap C) \geq \alpha\text{vol}(C)$ *for some* $\alpha \in (0, 1]$,

2. $\text{vol}(S \cap C) \geq \beta\text{vol}(S)$ *for some* $\beta \in (0, 1]$.

Then the cut $\tilde{C}$ returned by the sweep cut procedure on the optimal dual solution x^* satisfies

$$\phi(\tilde{C}) \leq O\left(\frac{\phi(C)^{1/q}}{\alpha\beta}\right)$$

where $q \in (1, \infty)$ is the norm used in (2)

Note the sweep cut computation only requires the dual solution x^*, while the primal solution f^* and the values of α, β are only for analysis. Recall we want to set $\delta = 2/\alpha$ in (3) to formulate the dual problem, but we assume δ is known via binary search. We also assume the entire graph is larger than the total amount of source mass so the primal is feasible and the dual is bounded. As summarized below,

Assumption 1. *The source mass function Δ in our problem formulation as specified in (3) satisfies $\delta = 2/\alpha$, which gives $\Delta(C) \geq 2\text{vol}(C)$ and $|\Delta| = \Delta(S) \leq 2\text{vol}(C)/\beta < \text{vol}(G)$.*

Not surprisingly, our theorem is more meaningful when the given seed set S has a good overlap with some low conductance cut C, i.e. when α, β are bounded away from 0. In particular, suppose α, β are both at least $\frac{1}{\log^t(\text{vol}(C))}$ for some constant t, then the bound in our theorem becomes $\widetilde{O}(\phi(C)^{1/q})$, where we follow the tradition of using $\widetilde{O}$ to hide poly-logarithmic factors. In particular, for 2-norm diffusion (i.e. $p = q = 2$) this matches the bound achieved by spectral and random walk based methods in this setting, and for p-norm diffusion with p approaches ∞ (i.e. q tends to 1), this matches the guarantees of previous flow based methods in this setting, e.g. (Wang et al., 2017; Orecchia & Zhu, 2014). We prove Theorem 2 in the rest of the section. We will keep the discussion at high-level and defer details to the supplementary material.

We start with the simple lemma stating that the objective value of the optimal dual (and primal) solution must be large.

Lemma 3. *Suppose $k = |E(C, V \setminus C)|$ is the cut-size of C, then*

$$(\Delta - \vec{\boldsymbol{d}})^T x^* = \|f^*\|_p \geq \text{vol}(C)/k^{1/q}.$$

Proof sketch. This directly follows from Assumption 1 that at least $2\text{vol}(C)$ amount of source mass is trapped in C at the beginning, so $\text{vol}(C)$ amount of excess needs to get out of C using the k cut edges, and we focus on the cost of f^* restricted to these edges alone. Since $p > 1$, the cost is the smallest if we distribute the routing load evenly on the k edges, and it is simple to see this incurs cost $\text{vol}(C)/k^{(p-1)/p}$. The total cost of f^* must be at least the cost incurred just by routing the excess out of C.□

Recall that we define the length of an edge $e = (u, v)$ to be $l(e) = |x^*(u) - x^*(v)|$. The actual dual solution may incur edges with tiny non-zero length which causes difficulties in the analysis. Thus, we define the following perturbed edge length so that any non-zero edge length is at least $1/\text{vol}(C)^{1/q}$. Note this is only for analysis purpose and doesn't require changing x^*.

$$\tilde{l}(e) = \begin{cases} \max\left(\frac{1}{\text{vol}(C)^{1/q}}, l(e)\right) & \text{if } l(e) > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Note the constraint in the dual problem gives

$$\sum_{e=(u,v)} |x^*(u) - x^*(v)| \cdot l(e)^{q-1} = \sum_{e=(u,v)} |x^*(u) - x^*(v)|^q = \|Bx^*\|_q^q \leq 1.$$

The next lemma states that the perturbations on edges lengths are small enough so the above quantity remains small.

Lemma 4. $\sum_{e=(u,v)} |x^*(u) - x^*(v)| \cdot \tilde{l}(e)^{q-1} \leq 1 + \frac{2}{\beta}$.

Proof sketch. This follows from Assumption 1 that the total amount of mass is at most $\frac{2}{\beta}\text{vol}(C)$, which also upper-bounds $\text{vol}_G(\text{supp}(x^*))$ by Lemma 1. Thus, the number of edges with positive $l(e)$ cannot be too large, and our perturbation only increases the lengths for these edges by at most $\frac{1}{\text{vol}(C)^{1/q}}$. The lemma follows from these two facts. □

Consider the sweep cut procedure where we order the nodes by their dual values in x^*, and for any $h > 0$ denote the cut $S_h = \{u | x^*(u) \geq h\}$ to be the set of nodes with dual value larger than h. We only need to consider S_h when h equals to the strictly positive dual value of some node in the support of x^*, and the sweep cut procedure will examine all such cuts. We proceed to argue that among these level cuts, there must exists some h where $\phi(S_h)$ satisfies the bound in Theorem 2, and thus prove the main theorem.

We start with rewriting the dual objective and constraint using the level cuts.

Claim 1. *With level cuts S_h as defined above, we have*

$$\int_{h=0}^{\infty} (\Delta(S_h) - \text{vol}(S_h))\, dh \geq \text{vol}(C)/k^{1/q},$$

and

$$\int_{h=0}^{\infty} \sum_{e \in E(S_h, V \setminus S_h)} \tilde{l}(e)^{q-1} dh \leq 1 + \frac{2}{\beta}.$$

Proof Sketch. Both claims follow from changing the order of summation to get

$$(\Delta - \vec{d})^T x^* = \int_{h=0}^{\infty} (\Delta(S_h) - \text{vol}(S_h))\, dh,$$

and

$$\sum_{e=(u,v)} |x^*(u) - x^*(v)| \cdot \tilde{l}(e)^{q-1} = \int_{h=0}^{\infty} \sum_{e \in E(S_h, V \setminus S_h)} \tilde{l}(e)^{q-1} dh,$$

and then invoke Lemma 3 and Lemma 4 respectively. To see the change of summation, pick any node v in the first equation. v contributes $(\Delta(v) - \deg(v)) \cdot x^*(v)$ to the left hand side, and the same amount to the right hand side as v is in the level cuts for all $h \in (0, x^*(v)]$. The second equation follows similar reasoning.□

Using Claim 1, we take the ratio to get

$$\frac{\int_{h=0}^{\infty} \sum_{e \in E(S_h, V \setminus S_h)} \tilde{l}(e)^{q-1} dh}{\int_{h=0}^{\infty} (\Delta(S_h) - \text{vol}(S_h))\, dh} \leq \frac{3k^{1/q}}{\beta \text{vol}(C)},$$

which means there must exist some h with S_h non-empty and

$$\frac{\sum_{e \in E(S_h, V \setminus S_h)} \tilde{l}(e)^{q-1}}{\Delta(S_h) - \text{vol}(S_h)} \leq \frac{3k^{1/q}}{\beta \text{vol}(C)}. \tag{5}$$

All it remains is to connect the left hand side in the above inequality to the conductance of S_h. For the denominator, since the source mass has density at most $\delta = 2/\alpha$ at any node, we get

$$\Delta(S_h) - \text{vol}(S_h) \leq \frac{2}{\alpha} \text{vol}(S_h). \tag{6}$$

For the numerator, any edge $e = (u, v)$ crossing a level cut S_h must have dual values $x^*(u), x^*(v)$ on different sides of h, thus having non-zero length $l(e)$, which means $\tilde{l}(e)$ is at least $1/\text{vol}(C)^{1/q}$. This gives

$$\sum_{e \in E(S_h, V \setminus S_h)} \tilde{l}(e)^{q-1} \geq \frac{|E(S_h, V \setminus S_h)|}{\text{vol}(C)^{(q-1)/q}}. \tag{7}$$

Put (5), (6) and (7) together we get

$$\phi(S_h) = \frac{|E(S_h, V \setminus S_h)|}{\text{vol}(S_h)} \leq \frac{6k^{1/q}}{\alpha\beta \text{vol}(C)^{1/q}} = \frac{6\phi(C)^{1/q}}{\alpha\beta},$$

which proves our main theorem.

5. Strongly Local Algorithm

For general constrained convex optimization problem, the state of the art solvers are by interior point techniques. However, these methods start with a fully dense initial point and then iteratively solve a linear system to obtain the update direction, which quickly become prohibitive when the input size is only moderately large. For example, the iterates of CVX (Grant & Boyd, 2014) fail to converge when solving

Algorithm 1 Coordinate solver for smoothed dual problem

Initialize: $x_0 = 0$
For $k = 0, 1, 2, \ldots,$ **do**
  Set $S_k = \{i \in V \mid \nabla_i F_\mu(x_k) < 0\}$.
  Pick $i_k \in S_k$ uniformly at random.
  Update $x_{k+1} = x_k - \frac{\mu^{2-q}}{\deg(i_k)} \nabla_{i_k} F_\mu(x_k) e_{i_k}$.
  If $S_k = \emptyset$ **then return** x_k.

the dual problem (2) on a 60×60 grid graph with initial mass $|\Delta| = 12000$ and $p = 4$.

As noted in Lemma 1, the support of optimal primal and dual solutions are bounded by the total amount of initial mass $|\Delta|$. Hence, we exploit locality and propose a randomized coordinate descent variant that solves an equivalent regularized formulation[3] of (2) (for a formal argument on the equivalence see supplementary material):

$$\min_{x \geq 0} \; F(x) := \tfrac{1}{q}\|Bx\|_q^q - x^T(\Delta - \vec{\boldsymbol{d}}). \tag{8}$$

Our algorithm is strongly local, that is, its running time depends on $|\Delta|$ rather than the size of the whole graph.

Because the function $F(x)$ in the objective of (8) has non-Lipschitz gradient for any $q < 2$, directly minimizing $F(x)$ requires step-sizes that go to zero to guarantee convergence, which leads to slow practical and worst-case convergence rate. To cope with this, we smooth $F(x)$ by replacing $\|Bx\|_q^q$ with $\|Bx\|_{\mu,q}^q$, where

$$\|y\|_{\mu,q}^q := \sum\nolimits_i \; (y_i^2 + \mu^2)^{q/2}.$$

The proposed numerical scheme in Algorithm 1 solves the smoothed problem

$$\min_{x \geq 0} F_\mu(x) := \tfrac{1}{q}\|Bx\|_{\mu,q}^q - x^T(\Delta - \vec{\boldsymbol{d}}). \tag{9}$$

In the context of p-norm flow diffusion, coordinate method enjoys a natural combinatorial interpretation[4]: each coordinate update corresponds to sending mass from a node to its neighboring nodes along incident edges. This combinatorial characteristic of coordinate method distinguishes it from other gradient type methods and ensures the solution path is strongly local. Note that the iterative construction of S_k in Algorithm 1 guarantees that the algorithm always produces monotonic iterates that satisfy $x_k \leq x_{k+1}$ and $\text{supp}(x_k) \subseteq \text{supp}(x_{k+1})$ for all k. In the supplementary material we show that the iterates indeed converge to a unique optimal solution x_μ^* of (9). We state the following lemma which justifies the definition of S_k in Algorithm 1. The claim follows a straightforward gradient computation and Lipschitz continuity, and we include details in the appendix.

Lemma 5. *For any iteration k and node i in Algorithm 1, $\nabla_i F_\mu(x_k) \leq 0$, $\forall i \in \text{supp}(x_k)$.*

Lemma 5 guarantees that coordinates with strictly positive partial derivatives all have value 0 in x (i.e. not in the support), and thus should not be selected for an update due to the $x \geq 0$ constraints.

Note that $F_\mu(x)$ is not strongly convex in general, but locality gives us strong convexity: when restricting $F_\mu(x)$ to the iterates generated by Algorithm 1, $F_\mu(x)$ has a strong convexity parameter γ which is the minimum eigenvalue of the sub-matrix of the Laplacian defined by the nodes in $\text{supp}(x_\mu^*)$ (Chung, 2007a), multiplied by a positive weight constant, and satisfies

$$\gamma > \frac{1}{(p-1)|\Delta|^p}, \tag{10}$$

where p is such that $1/p + 1/q = 1$. The result is pessimistic because we do not make any assumption about the internal conductance of the target cluster. Lower bounding γ under stronger assumptions are beyond the scope of this paper. In practice, we observe much better performance, because most clusters are well connected internally.

Finally, the optimal solution x_μ^* of (9) also minimizes a locally smoothed function $F_\mu^l(x)$ subject to $x \geq 0$, where

$$F_\mu^l(x) := \frac{1}{q} \sum_{(i,j) \in \text{supp}(x_\mu^*)} ((x(i) - x(i))^2 + \mu^2)^{q/2} + \frac{1}{q} \sum_{(i,j) \notin \text{supp}(x_\mu^*)} |x(i) - x(j)|^q - x^T(\Delta - \vec{\boldsymbol{d}}).$$

Since $1 < q < 2$, the scalar mapping $s \mapsto s^{q/2}$ is concave on $\mathbb{R}_{\geq 0}$, and so $|s|^q \leq (|s|^2 + \mu^2)^{q/2} \leq |s|^q + \mu^q$. Therefore, for all x,

$$\begin{aligned} F(x) \leq F_\mu^l(x) &\leq F(x) + \tfrac{1}{q}\mu^q |\text{supp}(Bx_\mu^*)| \\ &\leq F(x) + \tfrac{1}{q}\mu^q |\Delta|. \end{aligned} \tag{11}$$

Details on the equivalent minimization of $F_\mu(x)$ and $F_\mu^l(x)$ and the above inequality are provided in the supplementary material.

Lipschitz continuity and strong convexity give us linear convergence rate to the smoothed problem (9), and with locality and error bound (11) we obtain the following running time guarantee for solving the original regularized problem (8).

[3] One can easily verify that the regularized formulation (8) is the dual of our primal p-norm flow diffusion problem with the objective $\|f\|_p$ replaced by $\frac{1}{p}\|f\|_p^p$. Moreover, (8) and (2) have identical solutions, up to a constant scaling of variable values.

[4] From primal-dual optimality condition, for any node v, the coordinate gradient $\nabla_v F$ measures the difference between the current mass at node v and its sink capacity $\deg(v)$, i.e., $-\nabla_v F(x) = m(v) - \deg(v)$. Hence S_k can be viewed as the set of nodes that have positive excess mass (cf. Section 3) at iteration k.

Theorem 6 (Running time). *If we pick* $\mu = \mathcal{O}\big(\big(\frac{\epsilon}{|\Delta|}\big)^{1/q}\big)$, *the total running time of Algorithm 1 to obtain an* ϵ *accurate solution of* (8) *is* $\mathcal{O}\big(\frac{|\Delta|\bar{d}^2}{\gamma}\big(\frac{|\Delta|}{\epsilon}\big)^{2/q-1}\log\frac{1}{\epsilon}\big)$, *where* $\bar{d} = \max_{i\in\text{supp}(x^*_\mu)}\deg(i)$, *and* γ *is the strong convexity parameter that satisfies* (10).

We remark that Algorithm 1 is easily parallelizable, as all coordinates in S_k can be updated at the same time, without sacrificing locality.

6. Empirical Results

We implemented Algorithm 1 in Julia[5]. Implementation details are given in the supplementary material. The goal of our experiments is two-fold. First, we carry out experiments on various LFR synthetic graphs (Lancichinetti et al., 2008) and demonstrate that the theoretical guarantees of p-norm flow diffusion are well reflected in practice. Second, we show the advantage of p-norm diffusion for local graph clustering tasks on four real datasets. We compare the performance of p-norm flow diffusion with ℓ_1-regularized PageRank (Fountoulakis et al., 2017b) and nonlinear diffusion under power transfer (Ibrahim & Gleich, 2019). Although our theoretical analysis holds for $p \in (1, \infty)$ and Algorithm 1 converges linearly for any $p \geq 2$, we think in practice the most interesting regime is when p is a small constant, e.g. $p \in [2, 8]$, as our theory suggests that the marginal gain in terms of conductance guarantee diminishes as p grows large. We elaborate this in the synthetic experiment by comparing marginal improvements in the performance by raising p from 2 to 4, and from 4 to 8.

The LFR model (Lancichinetti et al., 2008) is a widely used benchmark for evaluating community detection algorithms. It is essentially a stochastic block model with the additional property that nodes' degrees follow the power law distribution, and there is a parameter μ controlling what fraction of a node's neighbours is outside the node's block. Our theory (and also experiment on LFR graph in the supplementary material) indicates (not surprisingly) that better overlap of the input seed set and a target cluster will result in output cluster having better conductance and F1 measure. In all subsequent experiments, however, when we compare recovery results with ℓ_1-regularized PageRank and nonlinear power diffusion, we always start the diffusion process from one seed node, as this is the most common practice for semi-supervised local clustering tasks. We set the parameter for p-norm diffusion so $|\Delta|$ is a constant factor of the volume of some target cluster (recall from Assumption 1 this is WLOG). We use the same parameter setting of nonlinear power diffusion as what the authors suggested (Ibrahim & Gleich, 2019). For ℓ_1-regularized PageRank, we allow it to "cheat" in the sense that we use ground truth to choose the parameter giving the best conductance result.

[5]The code is available at `github.com/s-h-yang/pNormFlowDiffusion`.

We compare these methods on 5 datasets. First on LFR synthetic graphs with different mixing parameter μ. We pick μ between 0.1 and 0.4 as this range gives rise to graphs that contain reasonable noisy (but not completely noise) clusters. For each graph, we start from a random seed node and we repeat the experiment 100 times. Figure 2 shows the mean and the variance for the conductance and the F1 measure while varying μ. Notice that p-norm diffusion behaves similarly to the tuned ℓ_1-regularized PageRank in both conductance and F1 measure when $p = 2$, whereas it significantly outperforms other methods when $p = 4$ and $p = 8$. We observe a slight gain in terms of conductance by raising $p = 4$ to $p = 8$, but such improvement is marginal.This is not really surprising, since qualitatively the 4-norm unit ball is already very close to the ∞-norm unit ball (i.e. the box).

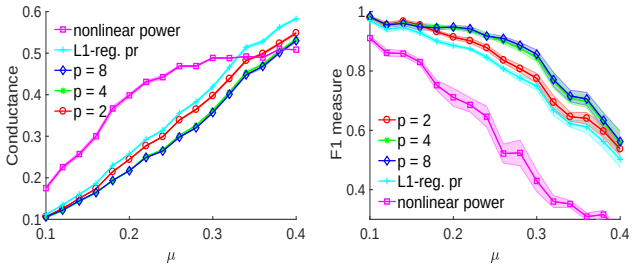

Figure 2. Conductance and F1 measure of various models on LFR synthetic datasets. The bands show the variation over 100 trials.

The other four datasets are real-world graphs: the Facebook college graphs of John Hopkins (FB-Johns55) and Colgate (Colgate88) (Traud et al., 2012)), the social network Orkut (Yang & Leskovec, 2012)) and the biological network Sfld (Brown et al., 2006). For each dataset, we filter the ground truth clusters by setting reasonable thresholds on volume, conductance, and the ratio between its internal connectivity and cut conductance. We include the statistics of the ground truth clusters in the supplementary material. For each dataset, we run algorithms starting at each vertex in each cluster and report the average conductance and F1 measure. We omit nonlinear power diffusion for Orkut, as it does not scale well to reasonably large graphs.

The Facebook datasets contain ground truth clusters ranging from low to medium conductance. For almost all clusters, p-norm flow diffusion methods have both the best F1 measure and conductance result. The six clusters in the biological dataset Sfld are very noisy, having median conductance around 0.8. The performance of the algorithms is mixed. The nonlinear power diffusion yields the lowest conductance but at the same time the lowest F1 measure for most cases. On the other hand, local methods, i.e., p-norm flow diffusion and ℓ_1-regularized PageRank, are more robust and give rise

to the best F1 measures. On Orkut dataset, 4-norm diffusion gives best result on all clusters. We defer more details and discussions to the supplementary material.

Table 1. Results for real-world graphs

dataset	feature	measure	$p=2$	$p=4$	ℓ_1-reg. pr	nonlinear
FB-Johns55	year 2006	f1	**0.36**	0.35	0.31	0.31
		cond	**0.34**	**0.34**	0.50	0.40
	year 2007	f1	**0.39**	**0.39**	0.38	0.36
		cond	0.31	**0.30**	0.45	0.41
	year 2008	f1	**0.51**	**0.51**	**0.51**	0.37
		cond	**0.34**	**0.34**	0.44	0.41
	year 2009	f1	0.84	**0.85**	0.83	0.49
		cond	0.23	**0.22**	0.24	0.40
	major index 217	f1	0.85	**0.87**	0.83	0.75
		cond	0.23	**0.22**	0.25	0.29
Colgate88	year 2004	f1	0.50	**0.51**	0.43	0.25
		cond	0.66	0.66	0.71	**0.36**
	year 2005	f1	**0.45**	**0.45**	0.41	0.37
		cond	0.51	0.51	0.53	**0.37**
	year 2006	f1	**0.45**	**0.45**	0.43	0.39
		cond	**0.37**	**0.36**	0.50	0.38
	year 2007	f1	0.49	0.49	**0.51**	0.45
		cond	**0.34**	**0.34**	0.45	0.39
	year 2008	f1	0.76	**0.80**	0.74	0.55
		cond	0.31	**0.30**	0.35	0.40
	year 2009	f1	0.96	**0.97**	0.96	0.82
		cond	0.13	**0.12**	0.13	0.24
Sfld	urease	f1	0.74	**0.76**	0.72	0.63
		cond	**0.44**	0.45	**0.44**	0.48
	AMP	f1	**0.83**	**0.83**	**0.83**	**0.83**
		cond	**0.41**	**0.41**	0.42	0.43
	phosphotriesterase	f1	0.93	0.93	**1.00**	0.13
		cond	0.81	0.81	0.81	**0.49**
	adenosine	f1	**0.44**	**0.44**	**0.44**	0.34
		cond	0.46	0.46	0.46	**0.44**
	dihydroorotase3	f1	**0.96**	**0.96**	**0.96**	0.07
		cond	0.84	0.84	0.84	**0.44**
	dihydroorotase2	f1	**0.39**	**0.39**	0.20	0.20
		cond	0.77	0.78	0.85	**0.48**
Orkut	A	f1	0.56	**0.58**	0.49	
		cond	0.48	**0.47**	0.51	
	B	f1	0.71	**0.73**	0.66	
		cond	0.35	**0.33**	0.37	
	C	f1	0.63	**0.64**	0.57	
		cond	0.33	**0.32**	0.35	
	D	f1	0.73	**0.76**	0.72	
		cond	0.49	**0.48**	0.51	
	E	f1	0.61	**0.62**	0.56	
		cond	0.52	**0.51**	0.54	
	F	f1	0.79	**0.81**	0.76	
		cond	0.52	**0.51**	0.54	
	G	f1	0.72	**0.73**	0.68	
		cond	0.52	**0.50**	0.53	
	H	f1	0.68	**0.70**	0.67	
		cond	0.52	**0.51**	0.54	
	I	f1	0.60	**0.62**	0.56	
		cond	0.49	**0.48**	0.52	
	J	f1	0.52	**0.54**	0.47	
		cond	0.53	**0.52**	0.56	
	K	f1	0.54	**0.56**	0.51	
		cond	0.57	**0.56**	0.60	

7. Conclusion

In this work we draw inspiration from spectral and combinatorial methods for local graph clustering. We propose a new method that is naturally non-linear and strongly local, and offers a spectrum of clustering guarantees ranging from quadratic approximation error (typically obtained by spectral diffusions) to constant approximation error (typically obtained by combinatorial diffusions). We note that the proposed diffusion is different from previous methods, since we analyze the optimal solution of a clearly defined family of optimization problems, rather than the output of an algorithmic procedure with no optimization objective, e.g. methods based on random walk. Furthermore, we provide a strongly local algorithm that solves the optimization formulation efficiently, which enables our method to scale to real-world large graphs with billions of nodes and edges.

We point out that a major advantage of this work is the simplicity of the proposed optimization formulation, clustering analysis, and algorithm design. Our model requires very few parameter tuning, and hence, it is extremely easy to use in practice and delivers consistent results. The algorithm is simple and has a very intuitive combinatorial interpretation, which facilitates possible future extension of p-norm flow diffusion to other applications and in much broader contexts, e.g., in community detection and graph semi-supervised learning (Fountoulakis et al., 2020), in defining localized network centrality measures (Yang et al., 2020).

Acknowledgements

Kimon Fountoulakis would like to acknowledge NSERC for providing support for this work.

References

Amghibech, S. Eigenvalues of the discrete p-laplacian for graphs. *Ars Comb.*, 67:283–302, April 2003.

Andersen, R. and Lang, K. J. An algorithm for improving graph partitions. *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pp. 651–660, 2008.

Andersen, R. and Peres, Y. Finding sparse cuts locally using evolving sets. In *STOC 2009*, pp. 235–244, 2009.

Andersen, R., Chung, F., and Lang, K. Local graph partitioning using pagerank vectors. *FOCS '06 Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science*, pp. 475–486, 2006.

Andersen, R., Gharan, S. O., Peres, Y., and Trevisan, L. Almost optimal local graph clustering using evolving sets. *Journal of the ACM*, 63(2), 2016.

Brown, S. D., Gerlt, J. A., Seffernick, J. L., and Babbitt, P. C. A gold standard set of mechanistically diverse enzyme superfamilies. *Genome biology*, 7(1):R8, 2006.

Bühler, T. and Hein, M. Spectral clustering based on the graph p-laplacian. In *Proceedings of the 26th Annual International Conference on Machine Learning*, ICML 2009, pp. 81–88, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605585161.

Chung, F. Random walks and local cuts in graphs. *Linear Algebra and its Applications*, 423(1):22–32, 2007a. ISSN 0024-3795. Special Issue devoted to papers presented at the Aveiro Workshop on Graph Spectra.

Chung, F. The heat kernel as the PageRank of a graph. *Proceedings of the National Academy of Sciences*, 104 (50):19735–19740, 2007b.

Chung, F. R. K. A local graph partitioning algorithm using heat kernel pagerank. *Internet Mathematics*, 6(3):315–330, 2009.

Fortunato, S. and Hric, D. Community detection in networks: A user guide. *Physics Reports*, 659(11):1–44, 2016.

Fountoulakis, K., Gleich, D. F., and Mahoney, M. W. An optimization approach to locally-biased graph algorithms. *Proceedings of the IEEE*, 105(2):256–272, 2017a.

Fountoulakis, K., Roosta-Khorasani, F., Shun, J., Cheng, X., and Mahoney, M. W. Variational perspective on local graph clustering. *Mathematical Programming*, 174:553–573, 2017b.

Fountoulakis, K., Liu, M., Gleich, D. F., and Mahoney, M. W. Flow-based algorithms for improving clusters: A unifying framework, software, and performance. arXiv:2004.09608 [cs.LG], 2020.

Gleich, D. F. PageRank beyond the web. *SIAM Review*, 57 (3):321–363, 2015.

Grant, M. and Boyd, S. CVX: Matlab software for disciplined convex programming, version 2.1. `http://cvxr.com/cvx`, March 2014.

Guattery, S. and Miller, G. On the quality of spectral separators. *SIAM Journal on Matrix Analysis and Applications*, 19:701–719, 1998.

Ibrahim, R. and Gleich, D. Nonlinear diffusion for community detection and semi-supervised learning. *WWW'19: The World Wide Web Conference*, pp. 739–750, 2019.

Jeub, L. G. S., Balachandran, P., Porter, M. A., Mucha, P. J., and Mahoney, M. W. Think locally, act locally: Detection of small, medium-sized, and large communities in large networks. *Physical Review E*, 91:012821, 2015.

Kloster, K. and Gleich, D. F. Heat kernel based community detection. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1386–1395, 2014.

Kloumann, I. M. and Kleinberg, J. M. Community membership identification from small seed sets. *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 1366–1375, 2014.

Lancichinetti, A., Fortunato, S., and Radicchi, F. Benchmark graphs for testing community detection algorithms. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 78:046110, 11 2008. doi: 10.1103/PhysRevE.78.046110.

Leskovec, J., Lang, K., Dasgupta, A., and Mahoney, M. Community structure in large networks: Natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1):29–123, 2009.

Li, P. and Milenkovic, O. Submodular hypergraphs: p-laplacians, Cheeger inequalities and spectral clustering. In Dy, J. and Krause, A. (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 3014–3023, Stockholmsmässan, Stockholm Sweden, 10–15 Jul 2018. PMLR.

Lovász, L. and Simonovits, M. The mixing rate of Markov chains, an isoperimetric inequality, and computing the volume. *Proceedings [1990] 31st Annual Symposium on Foundations of Computer Science*, pp. 346–354, 1990.

Lovász, L. and Simonovits, M. Random walks in a convex body and an improved volume algorithm. *Random Struct. Algorithms*, 4(4):359–412, 1993.

Mahoney, M. W., Orecchia, L., and Vishnoi, N. K. A local spectral method for graphs: with applications to improving graph partitions and exploring data graphs locally. *Journal of Machine Learning Research*, 13:2339–2365, 2012.

Ng, A., Jordan, M., and Weiss, Y. On spectral clustering: Analysis and an algorithm. In *NIPS '01: Proceedings of the 15th Annual Conference on Advances in Neural Information Processing Systems*, 2001.

Orecchia, L. and Zhu, Z. A. Flow-based algorithms for local graph clustering. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms*, pp. 1267–1286, 2014.

Page, L., Brin, S., Motwani, R., and Winograd, T. The pagerank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

Shi, P., He, K., Bindel, D., and Hopcroft, J. Local Lanczos spectral approximation for community detection. In *Proceedings of ECML-PKDD*, September 2017.

Spielman, D. A. and Teng, S. H. A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning. *SIAM Journal on Scientific Computing*, 42(1):1–26, 2013.

Traud, A. L., Mucha, P. J., and Porter, M. A. Social structure of facebook networks. *Physica A: Statistical Mechanics and its Applications*, 391(16):4165–4180, 2012.

Tsiatas, A. Diffusion and clustering on large graphs. *PhD Thesis, University of California, San Diego*, 2012.

Tuncbag, N., Milani, P., Pokorny, J. L., Johnson, H., Sio, T. T., Dalin, S., Iyekegbe, D. O., White, F. M., Sarkaria, J. N., and Fraenkel, E. Network modeling identifies patient-specific pathways in glioblastoma. *Scientific Reports*, 6:28668, 2016.

von Luxburg, U. A tutorial on spectral clustering. Technical Report 149, Max Plank Institute for Biological Cybernetics, August 2006.

Wang, D., Fountoulakis, K., Henzinger, M., Mahoney, M. W., and Rao, S. Capacity releasing diffusion for speed and locality. *Proceedings of the 34th International Conference on Machine Learning*, 70:3598–3607, 2017.

White, S. and Smyth, P. A spectral clustering approach to finding communities in graphs. In *SDM '05: Proceedings of the 5th SIAM International Conference on Data Mining*, pp. 76–84, 2005.

Yang, J. and Leskovec, J. Defining and evaluating network communities based on ground-truth. In *IEEE 12th International Conference on Data Mining*, pp. 745–754, 2012.

Yang, S., Senapati, P., Wang, D., Bauch, C. T., and Fountoulakis, K. Targeted pandemic containment through identifying local contact network bottlenecks. arXiv:2006.06939 [cs.SI], 2020.

Zhu, Z. A., Lattanzi, S., and Mirrokni, V. S. A local algorithm for finding well-connected clusters. In *ICML 2013*, pp. 396–404, 2013.