Node sampling by partitioning on graphs via convex optimization

Cristian Rusu and John Thompson
Institute for Digital Communications, University of Edinburgh
Email: {c.rusu, john.thompson}@ed.ac.uk

***Abstract*—In this paper we deal with the problem of efficiently and accurately reconstructing a complete graph signal from partially observed noisy measurements. Given a graph structure, we propose a solution based on convex optimization techniques to partition the nodes of the graph into subsets such that sampling a graph signal from any of these subsets provides an accurate, low mean squared error for example, reconstruction of the original complete graph signal. We show how the proposed sampling set construction approach relates to optimal experimental design, sensor management, positioning and selection problems and provide numerical simulation results on synthetic and real-world graphs.**

***Index Terms*—graph signal processing, sampling on graphs, sensor management, convex optimization, binary optimization.**

I. Introduction

Modern signal processing has extended the notion of time varying signals to also cover signals on irregular domains like graphs [1]. Graphs naturally model a large number of very diverse real world applications such as transport, energy, neural, social and sensor networks [2]. By successfully combining ideas from algebraic and spectral graph theory together with applied computational harmonic analysis, these structures can be studied and their behavior analyzed and controlled.

Given a graph we consider that each node is capable of performing a measurement or sample in time (for example, in a sensor network each node could measure some environment characteristic such as temperature). We call this collection of samples a graph signal. The key idea of graph signal processing is to take into account the additional information of the structure (geometry) of the graph when processing signals originating from its nodes. For instance, in a weighted graph, the weights of the edges between two nodes describe in some sense (that is defined in and dependent on the concrete application) the “connection” or “correlation” between these nodes. Therefore, this information can be exploited when processing a graph signal. We show in Fig. 1 a few examples of synthetic and real-world graphs.

Just as sampling theory plays a crucial role in classic signal processing (when dealing with time-varying signals) by providing a connection between the analog and digital domains, the same concepts can be extended to graph signals [3]. There the problem becomes one of reconstructing a full graph signal from only a subset of node values given that we have some information about the graph and its structure. In this paper we describe and use the previously introduced concept of bandlimited signals over the graph, which is a formalization of the smoothness assumption of graph signals using the graph Fourier transform, and we discuss the reconstruction of graph signals from noisy measurements [5].

In this paper we deal with the following question: given a graph structure, what is the best way to select a subset of the nodes where we perform noisy measurements such that we can estimate the full graph signal (the measurements in all the nodes of the graph) with high accuracy, on average. In this paper we propose a selection/partitioning algorithm to construct subsets of nodes such that estimation of the full graph signal is performed with similar accuracy across all partitions. This approach is particularly useful in practical scenarios when the graph cannot be queried completely or it is resource intensive to do so. Also, previous sampling set construction methods (like for example the recently greedy method introduced in [5] or the one in [6]) do not allow us to construct partitions of the node set (or to add some other application dependent constraints), they just select a subset of nodes for best performance. From this perspective, this work follows the line of work started in [13], [14] that deals with sensor scheduling problems under energy and time constraints.

We test our proposed method on both synthetic and real-world graphs. We then finally discuss connections between this work and experimental design [7], [8], [9] and sensor management problems [10] where selection/scheduling/partitioning solutions have been proposed and analyzed in the past [11]–[14] with some also focusing on resource management [15], [16], [17].

II. Brief background on graph signal processing

First, we introduce some notation, which is consistent to the one in [5]. Bold lowercase letters define vectors such as $\mathbf{f} \in \mathbb{R}^m$, uppercase letters define matrices such as $\mathbf{L} \in \mathbb{R}^{m \times m}$ and calligraphic letters such as $\mathcal{S}$ denote sets where $|\mathcal{S}|$ defines the size of the set. It is convenient to be able to talk about subsets of vectors and matrices and therefore we introduce the notation $\mathbf{f}_\mathcal{S} \in \mathbb{R}^{|\mathcal{S}|}$ where $\mathcal{S}$ is a set containing the indices of the vector $\mathbf{f}$ that are kept while $\mathbf{U}_{\mathcal{S}\mathcal{R}} \in \mathbb{R}^{|\mathcal{S}| \times |\mathcal{R}|}$ defines a sub-matrix of $\mathbf{U}$ where only the rows indexed in the set $\mathcal{S}$ and the columns indexed in the set $\mathcal{R}$ are kept.

We call the graph $G = (\mathcal{V}, \mathcal{E})$ a collection of nodes given by the set $\mathcal{V} = \{1, \ldots, m\}$ and connected by links indexed in the set $\mathcal{E} = \{w_{ij}\}_{\forall i,j \in \mathcal{V}, i \neq j}$ where each $w_{ij} \in \mathbb{R}_+$ denotes the weight of the link between node i and j – with the associated symmetric (because we assume G is an undirected graph)

978-1-5386-1663-5/17/$31.00 ©2017 IEEE

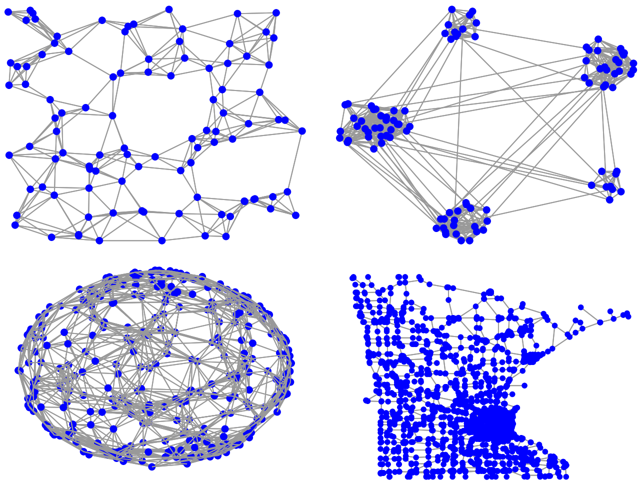

Figure 1: Examples of graph structures: a sensor network (upper left), a community graph (upper right) and a spherical-shaped graph (lower left) all with 100 nodes and the Minnesota road network with 2642 nodes (lower right). The synthetic examples are generated by the GSPBOX toolbox [4].

adjacency matrix $\mathbf{W} \in \mathbb{R}^{m \times m}$. We also introduce the degree matrix $\mathbf{D} = \text{diag}(d_1, d_2, \dots, d_m)$, where we have denoted the degree of node i as $d_i = \sum_j w_{ij}$, and the real symmetric $m \times m$ positive semidefinite Laplacian matrix

$$\mathbf{L} := \mathbf{D} - \mathbf{W} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T, \tag{1}$$

where the second equality highlights its eigenvalues decomposition [18, Chapter 4], with an orthonormal $\mathbf{U}$ (we also call $\mathbf{U}^T$ the graph Fourier transform) and the diagonal $\mathbf{\Lambda}$ containing the positive eigenvalues λ_k, $k = 1, \dots, m$, in ascending order, i.e., $0 \leq \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$. Operators such as the Laplacian $\mathbf{L}$ measure of level of smoothness in the graph and describe some of its properties.

A graph signal is a vector $\mathbf{f} \in \mathbb{R}^m$ defined over all the m nodes of the graph G, i.e., each entry is a graph function $f_i : \mathcal{V} \to \mathbb{R}$ defined with a scalar real value on each node of the graph. We call $\tilde{\mathbf{f}} = \mathbf{U}^T\mathbf{f}$ the graph Fourier transform of a signal $\mathbf{f}$. We say a signal $\mathbf{f}$ is $\omega-$bandlimited if $\tilde{f}_i = 0$ for all $\lambda_i > \omega$, i.e., it is supported only on the eigenvectors of $\mathbf{L}$ whose associated eigenvalues are strictly smaller than ω, i.e., there is a number $r \leq m$ such that

$$\mathbf{f} = \sum_{i=1}^{r} \tilde{f}_i \mathbf{u}_i = \mathbf{U}_{\mathcal{VR}} \tilde{\mathbf{f}}_{\mathcal{R}}, \tag{2}$$

where $\mathbf{U} = \begin{bmatrix}\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m\end{bmatrix} \in \mathbb{R}^{m \times m}$, $\mathcal{R} \in \{1, \dots, r\}$ and therefore $\mathbf{U}_{\mathcal{VR}} \in \mathbb{R}^{m \times r}$ represents the first r columns of $\mathbf{U}$.

Consider now the following problem: we are given a graph G with Laplacian $\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ and the noisy measurements $\mathbf{y}_\mathcal{S} = \mathbf{f}_\mathcal{S} + \mathbf{n}, \mathbf{y}_\mathcal{S} \in \mathbb{R}^{|\mathcal{S}|}$, from a subset $\mathcal{S}$ of the nodes of G and we are asked to estimate the graph signal in all the other nodes of the graph. We have denoted here the noise vector $\mathbf{n} \in \mathbb{R}^{|\mathcal{S}|}$. It is known that for consistent reconstruction [19] we follow

$$\begin{aligned}\hat{\mathbf{f}} =& \mathbf{U}_{\mathcal{VR}}(\mathbf{U}_{\mathcal{SR}}^T\mathbf{U}_{\mathcal{SR}})^{-1}\mathbf{U}_{\mathcal{SR}}^T\mathbf{y}_\mathcal{S} \\ =& \mathbf{U}_{\mathcal{VR}}(\mathbf{U}_{\mathcal{SR}}^T\mathbf{U}_{\mathcal{SR}})^{-1}\mathbf{U}_{\mathcal{SR}}^T\mathbf{f}_\mathcal{S} \ \dots \\ & + \mathbf{U}_{\mathcal{VR}}(\mathbf{U}_{\mathcal{SR}}^T\mathbf{U}_{\mathcal{SR}})^{-1}\mathbf{U}_{\mathcal{SR}}^T\mathbf{n}.\end{aligned} \tag{3}$$

The reconstruction error is given by $\mathbf{e} = \hat{\mathbf{f}} - \mathbf{f} = \mathbf{U}_{\mathcal{VR}}(\mathbf{U}_{\mathcal{SR}}^T\mathbf{U}_{\mathcal{SR}})^{-1}\mathbf{U}_{\mathcal{SR}}^T\mathbf{n}$ and we assume that $\mathbf{n}$ is a Gaussian i.i.d. vector with zero mean and identity covariance. Then the covariance matrix of $\mathbf{e}$ is given by

$$\mathbb{E}[\mathbf{e}\mathbf{e}^T] = \mathbf{U}_{\mathcal{VR}}(\mathbf{U}_{\mathcal{SR}}^T\mathbf{U}_{\mathcal{SR}})^{-1}\mathbf{U}_{\mathcal{VR}}^T. \tag{4}$$

With this definition, and assuming that the columns of $\mathbf{U}_{\mathcal{VR}}$ are orthogonal (as it is the case with any symmetric Laplacian matrix $\mathbf{L}$), in the spirit of optimal experimental design, we can quantify in several ways the reconstruction error in (3):

1) A-optimality, mean-squared error (MSE):

$$\text{MSE}(\mathbf{U}_{\mathcal{SR}}) = \text{tr}((\mathbf{U}_{\mathcal{SR}}^T\mathbf{U}_{\mathcal{SR}})^{-1}). \tag{5}$$

2) E-optimality, worst-case error variance (WCE):

$$\text{WCE}(\mathbf{U}_{\mathcal{SR}}) = \lambda_{\text{max}}((\mathbf{U}_{\mathcal{SR}}^T\mathbf{U}_{\mathcal{SR}})^{-1}). \tag{6}$$

3) D-optimality, volume of the confidence ellipsoid (VCE):

$$\text{VCE}(\mathbf{U}_{\mathcal{SR}}) = \log\det(\mathbf{U}_{\mathcal{SR}}^T\mathbf{U}_{\mathcal{SR}}). \tag{7}$$

We are interested in minimizing the MSE and WCE or maximizing the VCE in order to obtain accurate reconstructed graph signals $\hat{\mathbf{f}}$ in the presence of noise $\mathbf{n}$. We do this over the available degrees of freedom: the choice of the set $\mathcal{S}$ (the nodes to be selected from the graph G) which is up to us. Our goal is to partition the full set of nodes $\mathcal{V}$ into $P \in \mathbb{N}^*$ (where this value is given) subsets such that each one of these achieves low MSE and WCE or high VCE. This opens the possibility of operating only in the subsets, sequentially, such that full graph signals are reconstructed with high accuracy, on average, with less effort.

III. The proposed node partitioning method

For the clarity of exposition, we focus our development in this section on the MSE criterion (5) (but the algorithm presented here extends naturally to the WCE and VCE). We begin by noting that

$$\mathbf{U}_{\mathcal{SR}}^T\mathbf{U}_{\mathcal{SR}} = \mathbf{U}_{\mathcal{R}}^T\text{diag}(\mathbf{z})\mathbf{U}_{\mathcal{R}} \in \mathbb{R}^{|\mathcal{R}| \times |\mathcal{R}|}, \tag{9}$$

where $\mathbf{z} \in \{0, 1\}^m$ is a binary vector such that $z_i = 1$ when $i \in \mathcal{S}$ and $z_i = 0$ otherwise (when $i \notin \mathcal{S}$).

Given a graph with m nodes and a fixed number of partitions P we define the binary scheduling table

$$\mathbf{Z} = \begin{bmatrix}\mathbf{z}_1 & \mathbf{z}_2 & \dots & \mathbf{z}_P\end{bmatrix} \in \{0, 1\}^{m \times P}, \tag{10}$$

and we denote the scheduler for partition p by $\mathbf{z}_p \in \{0, 1\}^m$, i.e., the columns of $\mathbf{Z}$, we denote z_{ij} the $(i, j)^{\text{th}}$ entry of $\mathbf{Z}$ and we denote z_i the entries of $\mathbf{z}_p$. We interpret $z_{ij} = 1$ as activating (or performing a measurement) in partition j at node i. Also, since we are dealing with a partitioning problem, the scheduling problem is constrained to have a single value "one"

```
Algorithm 1 – Node partitioning by $\ell_1/\ell_\infty$ minimization.
Input: The graph sensing matrix $\mathbf{U}_\mathcal{R}$ of the network with $m$ sensors, the total number of partitions $P$ and the regularization parameter $\lambda > 0$.
Output: The scheduling table $\mathbf{Z} \in \{0,1\}^{m \times P}$ for the node activations for each partition such that the signal recovery performance is similar across the $P$ partitions.

  Initialization:
    1. Set initial weights $\mathbf{w}_p = \mathbf{1}_{m\times 1}$ and the initial all-zero solution $\mathbf{z}_p = \mathbf{0}_{m\times 1}$ for $p = 1,\ldots,P$, i.e., $\mathbf{Z} = \mathbf{0}_{m\times P}$.
    2. Initialize sets $\mathcal{N} = \emptyset$ indexing nodes that are not selected and $\mathcal{K} = \emptyset$ indexing nodes that are selected.

  Iterations:
    1. Set $\mathbf{Z}^{(\text{prev})} \leftarrow \mathbf{Z}$.
    2. Update weights according to $w_{ij} = (z_{ij}^{(\text{prev})} + \epsilon)^{-1}$.
    3. Solve (12) with current weights to obtain the new estimate $\mathbf{Z}$ via a standard solver like [22].
    4. Update the sets $\mathcal{N} = \{(i,j) \mid z_{ij} \leq \epsilon\}$ and $\mathcal{K} = \{(i,j) \mid z_{ij} \geq 1-\epsilon\}$ while maintaining the row and column constraints in $\mathbf{Z}$ (only one activation per row, minimum $|\mathcal{R}|$ activations per column).
    5. If the iterative process has converged, i.e., $\|\mathbf{Z} - \mathbf{Z}^{(\text{prev})}\|_F^2 \leq \epsilon$, then we update
          $\mathcal{K} \leftarrow \mathcal{K} \cup \{\arg\max_{(i,j)} z_{ij},\ (i,j) \notin \mathcal{K}\}.$     (8)
    6. If the solution for $\mathcal{K}$ and $\mathcal{N}$ is binary, i.e., $|\mathcal{N}|+|\mathcal{K}| = mP$, then stop otherwise go to step 1 of the iterative process.
```

in each row of $\mathbf{Z}$ and since we have to ensure full rank (in order to have invertibility in (4)) we need to ensure that for each partition p at least $|\mathcal{R}|$ nodes are selected.

To partition the nodes such that the estimation accuracy of the full graph signal is high over all P partitions we would like to solve the following NP-hard mixed-integer semidefinite optimization problem

$$\begin{aligned} \underset{y,\ \mathbf{Z}\in\{0,1\}^{m\times P}}{\text{minimize}} \quad & y \\ \text{subject to} \quad & \sum_{p=1}^{P} \mathbf{z}_p = \mathbf{1}_{m\times 1} \\ & \mathbf{1}_{1\times m}\mathbf{z}_p \geq |\mathcal{R}|, p = 1,\ldots,P \\ & \text{tr}((\mathbf{U}_\mathcal{R}^T \text{diag}(\mathbf{z}_p)\mathbf{U}_\mathcal{R})^{-1}) \leq y, p = 1,\ldots,P, \end{aligned} \tag{11}$$

The first constraint ensures the partitioning of the nodes, i.e., each node is activated in only one of the P partitions, while the second constraint explicitly ensures that at each partition p the matrix $\mathbf{U}_\mathcal{R}^T\text{diag}(\mathbf{z}_p)\mathbf{U}_\mathcal{R}$ is full rank, i.e., at least $|\mathcal{R}|$ nodes are selected. The third constraint, which in this case is for the MSE, is in the style of ℓ_∞ norm minimization (using the slack variable $y \in \mathbb{R}_+$), i.e., we want to minimize the worst case MSE over all P partitions. We thus ensure that across all partitions the MSE reconstruction performance is similar on average. Extensions to the WCE (6) and VCE (7) are trivial and immediate since they are also convex constraints, and are therefore omitted for brevity. Notice that this approach does not deal explicitly with how many nodes are activated in each partition; we rather impose the reconstruction accuracy constraint and let the optimization process establish the number of nodes to be activated.

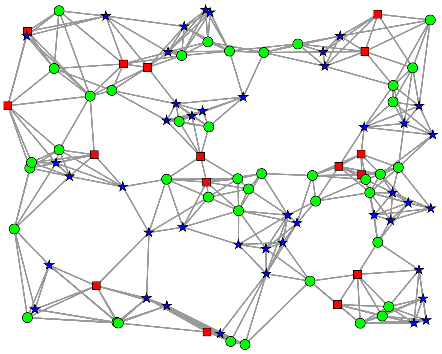

Figure 2: Result of applying Algorithm 1 for a network with $m = 100$ sensors and with $P = 3$ partitions (the three partitions are shown in red squares, green circles and blue stars, respectively).

We relax the binary constraints and propose to solve the following tractable, convex optimization [20] problem

$$\begin{aligned} \underset{y,\ \mathbf{Z}\in[0,1]^{m\times P}}{\text{minimize}} \quad & y + \lambda \sum_{p=1}^{P} \mathbf{w}_p^T \mathbf{z}_t \\ \text{subject to} \quad & \sum_{p=1}^{P} \mathbf{z}_p = \mathbf{1}_{m\times 1} \\ & \mathbf{1}_{1\times m}\mathbf{z}_p \geq |\mathcal{R}|, p = 1,\ldots,P \\ & \text{tr}((\mathbf{U}_\mathcal{R}^T \text{diag}(\mathbf{z}_p)\mathbf{U}_\mathcal{R})^{-1}) \leq y, p = 1,\ldots,P, \end{aligned} \tag{12}$$

where λ is the given regularization parameter, the weight vectors $\{\mathbf{w}_p\}_{p=1}^{P} \in \mathbb{R}^m$ are given as $w_{ij} = (z_{ij} + \epsilon)^{-1},\ i = 1,\ldots,m,\ j = 1,\ldots,P$ in the style of the iterative reweighted ℓ_1 approach [21]. We use this optimization problem in the proposed Algorithm 1 where we detail the procedure of allocating the entries in $\mathbf{Z}$ to binary values.

In Algorithm 1, we build the binary solution by keeping track of two sets: $\mathcal{K}$ (containing the "one" entries) and $\mathcal{N}$ (containing the "zero" entries). We build up these sets iteratively by allowing the ℓ_1 optimization procedure to push the entries to binary values. If the procedure converges while the solution is still not binary we set to "one" the largest entry in $\mathbf{Z}$ that does not yet belong to $\mathcal{K}$ and then proceed with the iterative process. In this manner we ensure that most of the entries in $\mathbf{Z}$ are allocated to the set $\mathcal{K}$ or $\mathcal{N}$ by the optimization procedure and not by a thresholding operation.

From a computational perspective, Algorithm 1 is efficient. At each step, once an entry has been allocated to either the set $\mathcal{K}$ or $\mathcal{N}$ it is removed from the optimization problem. Therefore, as the algorithm proceeds, the dimension of the

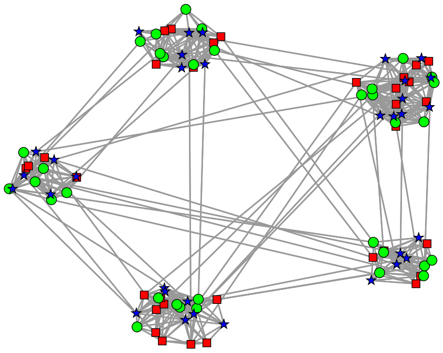

Figure 3: Analogous simulation scenario as in Fig. 3 for a community network with $m = 100$ nodes and 5 groups.

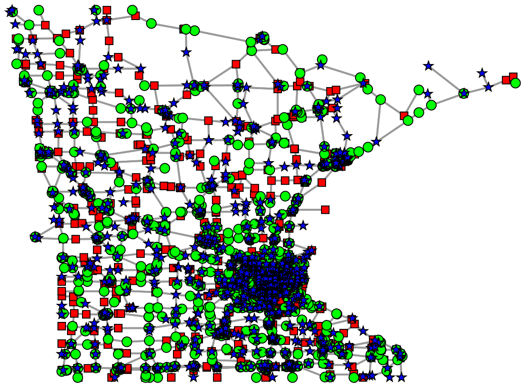

Figure 4: The result of applying Algorithm 1 on the Minnesota road map graph with $P = 3$ partitions.

optimization problem that is solved at Step 3 decreases significantly (to the extent that the last iterations involve only a few optimization variables).

IV. Results

In this section we test our proposed approach to partition graph nodes such that signal can be reconstructed from a subset of noisy measurement in a few nodes with high accuracy.

In Fig. 2 and 3 we show partitioning results on synthetic sensor and community graphs of $m = 100$ elements each. In each case we construct the Laplacian matrices $\mathbf{L}$ as in (1) and we keep only the low-pass components of the graph: the eigenvectors associated with the ten smallest eigenvalues of $\mathbf{L}$, i.e., $\mathbf{U}_{\mathcal{V}\mathcal{R}} \in \mathbb{R}^{100\times 10}$ with $|\mathcal{R}| = 10$. We always partition the graphs into $P = 3$ groups (for ease of exposition in the figures). In Fig. 2 notice that the sensors are partitioned such that each group closely covers all regions of the graph while in Fig. 3 each one of the five communities themselves have members in all the partitions. The observations are intuitive and they suggests that for a specific type of graph some preprocessing (like clustering) that takes into account its structure could help the partitioning procedure produce better results (or at least perform the partitioning more efficiently by splitting the overall problem is smaller sub-problems).

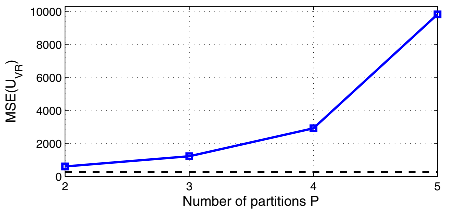

Figure 5: The result of applying Algorithm 1 on the Minnesota road map graph with varying partition size P. We show the best case performance (query the full network, no partitioning) for the MSE, which is at 264. These results show that the performance degrades as the number of partitions increases.

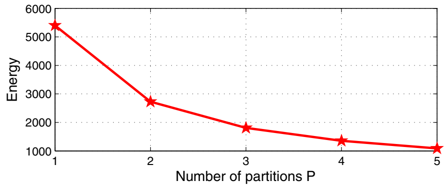

Figure 6: Trade-off between the number of partitions P (when $P = 1$ no partitioning takes place) and the energy consumption for acquiring the samples of the Minnesota graph network according to the energy cost (13).

In the next experimental setup we use a real-world graph, the Minnesota road map graph[1]. The graph has $m = 2642$ nodes and we keep 10% of the low-pass components, i.e., $|\mathcal{R}| = 264$. In Fig. 4 we show the result of partitioning the graph in $P = 3$ groups while in Fig. 5 we show how the MSE scales with the size of the partition P. Better results are obtained for smaller P, as expected since more nodes are activated per each partition. The best MSE is obtained when $\mathcal{S} = \mathcal{V}$ (there is only one partition) and we show in the figure this best case value of $\text{MSE}(\mathbf{U}_{\mathcal{V}\mathcal{S}}) = |\mathcal{R}| = 264$. Comparisons with other methods, like the greedy sampling set selection proposed in [5] is difficult since these are selection procedures that cannot be easily extended to partitioning tasks.

The benefits of partitioning for sampling any graph are clear when we also consider an energy model attached to the model. For example, consider that all samples are collected in a node of the graph (which we call in this paper the central node) where the data are processed. For example, let us consider again the previously mentioned Minnesota road graph where we now pick a central node whose coordinates are $\mathbf{c} \in \mathbb{R}^2$. The other nodes of the graph are positioned at coordinates $\{\mathbf{v}_i\}_{i=1}^{2641} \in \mathbb{R}^2$ and their cost of transmitting the measurements to the central node is

$$s_i = O(\|\mathbf{c} - \mathbf{v}_i\|_2), \tag{13}$$

i.e., proportional to the euclidean distance. We show in Fig. 6 and 7 how this cost depends on the partitioning achieved,

[1] Part of he MatlabBGL Matlab library, by D. Gleich, available at http://www.cs.purdue.edu/homes/dgleich/packages/matlab_bgl/index.html

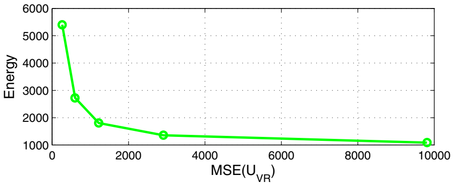

Figure 7: Trade-off between MSE and energy consumption in an simulation scenario analogous to Fig. 6.

as a function of the MSE and the number of partitions P respectively. We observe that the partitioning procedure brings high energy gains at the cost of slight increases in MSE, especially for low P, and therefore balances the cost of using the network with the accuracy of its measurements.

In all runs of Algorithm 1 we set the regularization parameter $\lambda = 1$ and the internal parameter $\epsilon = 10^{-5}$.

From an analysis perspective, some theoretical results are possible in our partition application. Notice that given an orthonormal $\mathbf{U}$ then any subset $\mathbf{U}_{\mathcal{VR}}$ has orthonormal columns and therefore it is a tight frame, i.e., $\mathbf{U}_{\mathcal{VR}}^T\mathbf{U}_{\mathcal{VR}} = \mathbf{I}$. The problem of node selection (or partitioning) on the graph is equivalent to a sensor placement problem in the case where the sensor measurement matrix is tight. In this case, theoretical analysis based on the solution to the Kadison-Singer problem [23], as explained in detail in [14], shows that

$$\mathrm{MSE}(\mathbf{U}_{\mathcal{SR}}) \geq \begin{cases} |\mathcal{R}|P^2, & \text{if } P \ll m|\mathcal{R}|^{-1} \\ \frac{m}{(1+\sqrt{c_1})^4}P, & \text{if } P \approx m|\mathcal{R}|^{-1}, \end{cases} \qquad (14)$$

for a constant c_1 that depends on the entries in $\mathbf{U}_{\mathcal{VR}}$. The result puts a limit on how good the MSE can be depending on the number of partitions: more partitions lead to worse accuracy per partition (fewer nodes are selected, of course). If we are in a regime when $P \ll m|\mathcal{R}|^{-1}$ then each partition activates a large number of nodes, i.e., significantly larger than $|\mathcal{R}|$, and then even a random selection of the nodes performs very well, see [14, Result 4]. Otherwise, if $P \approx m|\mathcal{R}|^{-1}$ then each partition contains a number of nodes close to $|\mathcal{R}|$ and therefore invertibility and rank issues might occur when constructing $\mathbf{U}_{\mathcal{R}}\mathrm{diag}(\mathbf{z}_p)\mathbf{U}_{\mathcal{R}}$ leading to large values of the MSE. This is the case where an approach like Algorithm 1 can bring the highest benefit as compared to a more simplistic partitioning method.

V. Conclusions

In this paper we have presented a graph sampling algorithm based on an convex programming approach in the spirit of ℓ_1/ℓ_∞ optimization to partition the nodes of a graph in such a way as to guarantee that all subsets of the partition perform similarly in graph signal reconstruction operations. The method is numerically efficient and easily applicable to any numerical scenarios by using off-the-shelf standard optimization libraries. Furthermore, since it is based on a convex optimization formulation, extra constraints can be accommodated depending on the application at hand and its real-world constraints.

Acknowledgment

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/K014277/1 and the MOD University Defence Research Collaboration (UDRC) in Signal Processing.

References

[1] D. I Shuman, S. K. Narang, P. Frossard, A. Ortega and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Sig. Process. Mag.*, vol. 30, no. 3, pp. 83-98, 2013.

[2] A. Sandryhaila and J. Moura, "Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure," *Sig. Process. Mag.*, IEEE, vol. 31, no. 5, pp. 80-90, 2014.

[3] I. Pesenson, "Sampling in Paley-Wiener spaces on combinatorial graphs," *Trans. American Math. Soc.*, vol. 360, no. 10, pp. 5603–5627, 2008.

[4] N. Perraudin, J. Paratte, D. Shuman, V. Kalofolias, P. Vandergheynst and D. K. Hammond, "GSPBOX: A toolbox for signal processing on graphs," *arXiv:1408.5781*, 2014.

[5] A. Anis, A. Gadde and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *Trans. Sig. Proc.*, vol. 64, no. 14, pp. 3775–3789, 2016.

[6] A. Sakiyama, Y. Tanaka, T. Tanaka1 and Antonio Ortega, "Efficient sensor position selection using graph signal sampling theory," *in Proc. ICASSP*, 2017.

[7] R. John and N. Draper, "D-optimality for regression designs: A review," *Technometrics*, vol. 17, no. 1, pp. 15–23, 1975.

[8] W. Welch, "Branch-and-bound search for experimental designs based on D-optimality and other criteria," *Technometrics*, vol. 24, no. 1, pp. 41–48, 1982.

[9] N. Nguyen and A. Miller, "A review of some exchange algorithms for constructing discrete D-optimal designs," *Comput. Statist. Data Anal.*, vol. 14, pp. 489–498, 1992.

[10] A. O. Hero III and D. Cochran, "Sensor management: Past, present, and future," *IEEE Sensors Journal*, vol. 11, no. 12, pp. 3064–3075, 2011.

[11] H. Jamali-Rad, A. Simonetto and G. Leus, "Sparsity-aware sensor selection: centralized and distributed algorithms," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 217–220, 2014.

[12] M. P. Vitus, W. Zhang, A. Abate, J. Hu and C. J. Tomlin, "On efficient sensor scheduling for linear dynamical systems," *Automatica*, vol. 48, no. 10, pp. 2482–2493, 2012.

[13] C. Rusu, J. Thompson and N. M. Robertson, "Balanced sensor management across multiple time instances via l-1/l-infinity norm minimization," *in Proc. ICASSP*, 2017.

[14] C. Rusu, J. Thompson and N. M. Robertson, "Sensor management with time, energy and communication constraints,"*arXiv:1702.04927*, 2017.

[15] Y. Chen, C. Chuah and Q. Zhao, "Sensor placement for maximizing lifetime per unit cost in wireless sensor networks," *in Proc. IEEE Military Commun. Conf. (MILCOM)*, 2005, pp. 1097–1102.

[16] S. Liu, A. Vempaty, M. Fardad, E. Masazade and P. K. Varshney, "Energy-aware sensor selection in field reconstruction," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1476–1480, 2014.

[17] S. Liu, F. Chen, A. Vempaty, M. Fardad, L. Shen and P. K. Varshney, "Sparsity-promoting sensor management for estimation: an energy balance point of view," *18th International Conference on Information Fusion*, 2015, pp. 6–9.

[18] R. Horn and C. Johnson, "Matrix Analysis," *Cambridge University Press*, 1990.

[19] Y. C. Eldar, "Sampling with arbitrary sampling and reconstruction spaces and oblique dual frame vectors," *Journal Fourier Analysis and Applications*, vol. 9, no. 1, pp. 77–96, 2003.

[20] S. Joshi and B. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, 2009.

[21] M. B. Wakin, E. Candes and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877–905, 2008.

[22] M. Grant and S. Boyd, "Matlab Software for Disciplined Convex Programming," Software available at http://cvxr.com/, 2011.

[23] A. W. Marcus, D. A. Spielman, and N. Srivastava, "Interlacing families II: Mixed characteristic polynomials and the Kadison-Singer problem," *Annals of Math.*, vol. 182, no. 1, pp. 327–350, 2015.