Correlated Compressive Sensing for Networked Data

Tianlin Shi
IIIS
Tsinghua University
Beijing, China

Da Tang
IIIS
Tsinghua University
Beijing, China

Liwen Xu
IIIS
Tsinghua University
Beijing, China

Thomas Moscibroda
Microsoft Research Asia
Beijing, China

Abstract

We consider the problem of recovering sparse correlated data on networks. To improve accuracy and reduce costs, it is strongly desirable to take the potentially useful side-information of network structure into consideration. In this paper we present a novel correlated compressive sensing method called `CorrCS` for networked data. By naturally extending Bayesian compressive sensing, we extract correlations from network topology and encode them into a graphical model as prior. Then we derive posterior inference algorithms for the recovery of jointly sparse and correlated networked data. First, we design algorithms to recover the data based on pairwise correlations between neighboring nodes in the network. Next, we generalize this model through a diffusion process to capture higher-order correlations. Both real-valued and binary data are considered. Our models are extensively tested on several real datasets from social and sensor networks and are shown to outperform baseline compressive sensing models in terms of recovery performance.

1 INTRODUCTION

Networked data, from domains such as social network of friends, hyper-linked networks of webpages and distributed network of sensors, are becoming increasingly pervasive and important in modern signal processing and machine learning. Recent research has demonstrated compelling approaches to extract useful information from these networked data, including latent structure of social links (Kemp et al., 2004), community detection (Fortunato, 2010), etc. However, with the massive amount of data generated at an exploding rate, conventional ways of collecting networked data are being challenged, particularly when measurements are expensive and/or data are redundant.

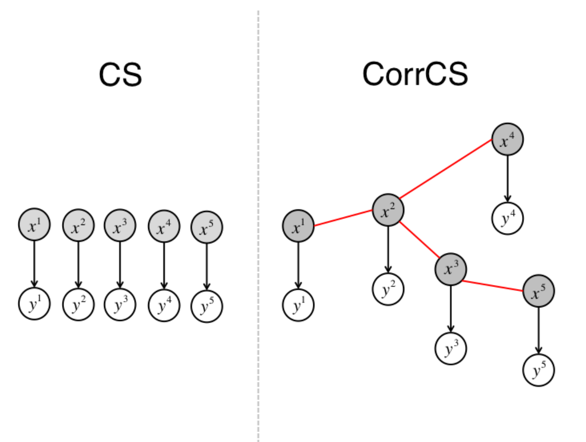


Figure 1: A graphical illustration of Compressive Sensing (`CS`) and Correlated Compressive Sensing (`CorrCS`) on networked data. In `CS`, each sparse signal $\boldsymbol{x}^i$ is recovered independently as $\boldsymbol{y}^i$; in `CorrCS`, they are recovered jointly with the network structure.

A significant finding of a large class of high-dimensional data over the last two decades is their inherent sparsity (Candès and Wakin, 2008). As a robust tool to leverage the sparsity, Compressive Sensing (`CS`) (Donoho, 2006) has been developed to collect high-dimensional sparse data from their low-dimensional projections. With sufficiently sparse signals, compressive sensing is guaranteed to recover the original signal from fewer samples required by Shannon-Nyquist limit (Candès, 2006). Compressive sensing has therefore been sucessfully applied to attack various problems of data collection in a wide range of fields, such as medical imaging (Lustig et al., 2008), low-level vision (Yang et al., 2008), etc.

Successful attempts have been made to apply compressive sensing to collect and analyze sparse networked data. For example, in studying large-scale sensor networks, the pioneering work by Luo et al. (2009) shows a successful scheme to efficiently gather spatially sparse sensor readings. For social networks, Compressive Network Analysis (Jiang et al., 2011) proposes a novel framework for clique detection. These approaches typically require finding a proper basis over the network topology so that data could be sparsely represented.

In many cases, however, it is unclear how the network topology could imply the sparse structure of the data, and therefore difficult to identify the sparse basis. Take the social network for example. It has been known that social influence and correlation exist at a large statistical level (Backstrom et al., 2006), but it turns out hard to directly model due to unobserved latent factors (Anagnostopoulos et al., 2008). An interesting question would be, under uncertainty about correlation of sparse data across the network, is it possible to seriously incorporate the network structure into compressive sensing and hopefully to improve recover performance?

In this paper, we present Correlated Compressive Sensing (`CorrCS`) to solve this problem. In particular, the setting in Figure 1 is considered. Each node i is equipped with a sensor, and we aim to recover the original high-dimensional data $\boldsymbol{x}^i$ from its low-dimensional measurements $\boldsymbol{y}^i$. Instead of independently recover sparse signals, `CorrCS` incorporates side-information of the network structure and build correlation into signal modeling jointly with the inherent sparsity. By adopting a probabilistic approach, we show that it is possible to exploit the flexibility of graphical models to improve compressive sensing. Our approach is extensively tested on several real datasets, including product review data from social trust networks, social polling data and Air Quality Index from distributed sensors. The results show that `CorrCS` outperforms `CS` in terms of recover performance and demonstrates the usefulness of correlation in sensing networked data.

2 PRELIMINARY

In a typical sensing problem, the data of interest is regarded as a signal, which is a vector $\boldsymbol{x}$ in a high-dimensional space $\mathbb{R}^r$. A measurement of $\boldsymbol{x}$ is a low-dimensional vector $\boldsymbol{y} \in \mathbb{R}^m$ $(m \leq r)$ from which the information of $\boldsymbol{x}$ can be extracted. From a Bayesian point of view, this corresponds to inferring $p(\boldsymbol{x} \mid \boldsymbol{y}) \propto p(\boldsymbol{y} \mid \boldsymbol{x})p(\boldsymbol{x})$. The likelihood term $p(\boldsymbol{y} \mid \boldsymbol{x})$ describes a sensing model, which is the noisy measurement process, and the term $p(\boldsymbol{x})$ corresponds to a signal model, which represents the prior knowledge.

2.1 THE SENSING MODEL

A large body of sensing methods focus on the linear system

$$\boldsymbol{y} = V\boldsymbol{x} \tag{1}$$

with the goal to recover $\boldsymbol{x}$ from $\boldsymbol{y}$ accurately. However, since $r \gg m$, the inversion problem is highly ill-posed due to the undetermined solutions. One way to deal with the uncertainty is to adopt a Gaussian generative model for $\boldsymbol{y}$ as follows,

$$\boldsymbol{y} \mid \boldsymbol{x} \sim \mathcal{N}(V\boldsymbol{x}, \beta), \tag{2}$$

where β is the variance controlling the precision of measurement.

2.2 THE SIGNAL MODEL

Many natural signals $\boldsymbol{x} \in \mathbb{R}^r$ can be sparsely represented under some basis $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, ..., \boldsymbol{\phi}_K]$ as

$$\boldsymbol{x} = \boldsymbol{\Phi}\boldsymbol{z} \tag{3}$$

where $\boldsymbol{z}$ is the sparse coefficients such that $||\boldsymbol{z}||_0 = S \ll K$. Compressive sensing (Donoho, 2006; Candès and Wakin, 2008) shows that if the signal $\boldsymbol{x}$ is sufficiently sparse, one can recover it effectively through minimizing the number of non-zero components in $\boldsymbol{z}$:

$$\begin{aligned} \min \quad & ||\boldsymbol{z}||_0 \\ \text{s.t.} \quad & \boldsymbol{y} = M\boldsymbol{z} \end{aligned} \tag{4}$$

where $M = V\boldsymbol{\Phi}$. In practice, it is often hard to solve the non-convex objective in (4) exactly, and a ℓ_1 relaxation is usually adopted, which corresponds to the basis pursuit (BP) algorithm (Chen et al., 1998). Candès et al. (2006) have proved that, under certain isometry properties, one can recover $\boldsymbol{x}$ perfectly from $m = \Omega(S \log r)$ observations $\boldsymbol{y}$ through BP.

To allow extra flexibility that we would exploit later, the framework of CS could be reformulated approximately as a Bayesian inference problem (Ji et al., 2008), and its goal is to design a signal model $p(\boldsymbol{z}; \boldsymbol{\Gamma})$ with some parameter $\boldsymbol{\Gamma}$ so that $\boldsymbol{z}$ is controlled to be sufficiently sparse. Below we describe two common sparse signal models in social and sensor networks.

ℓ_1 prior for real-valued $\boldsymbol{z}$. Using sparsity-favoring Laplace priors on the coefficients (Babacan et al., 2010), one could use the following signal model:

$$p(\boldsymbol{z}; \lambda) = \frac{\lambda^{K/2}}{2^K} \exp\Big(-\sqrt{\lambda}||\boldsymbol{z}||_1\Big) \tag{5}$$

In practice, it is often inconvenient that the Laplace prior is not conjugate to the Gaussian signal model. However, one can show that (5) is equivalent to a hierarchical conjugate model parametrized by $\boldsymbol{\Gamma} = (\{\gamma_k\}_{k=1}^K, \lambda)$ (Seeger and Nickisch, 2008).

$$\begin{aligned} p(\boldsymbol{z}; \boldsymbol{\Gamma}) &= \prod_{k=1}^K \mathcal{N}(z_k; 0, \gamma_k) \\ p(\boldsymbol{\Gamma}) &= \text{Gamma}(\gamma_k; 1, \frac{\lambda}{2}) \end{aligned} \tag{6}$$

The inference of $\boldsymbol{z}$ for (6) can be efficient via the EM algorithm (Dempster et al., 1977).

Beta process for binary $\boldsymbol{z}$. An efficient way to characterize the sparsity of binary-valued coefficients is the Beta process (Paisley and Carin, 2009). In practice, a finite truncation of the process is used and leads to the following hierarchical conjugate model:

$$p(\boldsymbol{z};\boldsymbol{\Gamma}) = \prod_{k=1}^{K} \text{Bernoulli}(z_k;\pi_k) \tag{7}$$
$$p(\boldsymbol{\Gamma}) = \prod_{k=1}^{K} \text{Beta}(\pi_k; \frac{a}{K}, b(1-\frac{1}{K}))$$

where $\boldsymbol{\Gamma} = (\{\pi_k\}_{k=1}^K, a, b)$ and a, b are hyper-prior affecting sparsity. The exact posterior inference of (7) is intractable, but can be approximated through MCMC (Mohamed et al., 2011) or mean-field variational inference (Paisley and Carin, 2009).

3 CORRELATED COMPRESSIVE SENSING

Now, consider the problem of collecting data distributed on a network of n nodes. When the network structure is known, it can be described by a graph $G(V,E)$, where the edges have weight

$$E_{ij} = \begin{cases} w_{ij}, & \text{node } i \text{ and } j \text{ are adjacent} \\ 0, & \text{otherwise} \end{cases}, \tag{8}$$

where w_{ij} encodes the side-information about the correlation between the node i and j. In practice, such weight can either be collected directly from network or be computed through some metrics such as Pearson correlation and some function of geographic distance. When this weight is not exactly available, it is convenient to set $w_{ij} = 1$ for all edges uniformly. Let $(\boldsymbol{Z},\boldsymbol{X},\boldsymbol{Y}) = \{\boldsymbol{z}^i,\boldsymbol{x}^i,\boldsymbol{y}^i\}_{i=1}^n$, the Bayesian formulation of sensing is generalized as the following principle

$$p(\boldsymbol{Z}|\boldsymbol{Y}) \propto p(\boldsymbol{Z}) \prod_{i=1}^{N} \mathcal{N}(M^i\boldsymbol{z}^i, \beta). \tag{9}$$

Instead of applying compressive sensing independently to each node (i.e. $p(\boldsymbol{Z}) = \prod_i p(\boldsymbol{z}^i)$), Correlated Compressive Sensing (`CorrCS`) fuses the network structure G into recovery as side-information. To fulfill this goal, a joint distribution $p(\boldsymbol{Z})$ is explored in this section to capture the notion of joint sparsity and correlation.

3.1 PAIRWISE CORRELATION

The simplest form of correlation among networked data is pairwise according to the edge connecting neighboring nodes. Inspired by graphical models, we consider a range of pairwise Correlated Compressive Sensing (`CorrCS-Pair`) that can be formulated as the Gibbs distribution

$$p(\boldsymbol{Z} \mid \boldsymbol{\Gamma}) \propto \exp\Big(-\sum_i \mathcal{S}^i - c \sum_{(i,j)\in E} \mathcal{C}^{ij}(\boldsymbol{z}^i,\boldsymbol{z}^j)\Big). \tag{10}$$

where $\mathcal{S}^i$ is the sparsity of individual node $\boldsymbol{z}^i$ controlled by hyper-parameter $\boldsymbol{\Gamma}$ and $\mathcal{C}_{ij}$ models the pairwise correlation between two neighboring signal coefficients $\boldsymbol{z}_i$, $\boldsymbol{z}_j$. The parameter c represents our prior on how strong this correlation should be.

Notice when $c = 0$, equation (10) reduces to independently applying BCS to each node. And the bigger c is, more correlation between neighboring nodes is favored over sparsity. On networked data with inherent correlation, we would expectedly improve the recovery performance with proper choice of positive c. However, if $c \to \infty$, the model would totally neglect sparsity, and therefore be undesirable. This variation of the recovery performance happens in actual experiments, as will be discussed later.

Below we consider two specific forms of `CorrCS` for real-valued and binary networked data. We discuss their inference algorithm in section 3.3.

Laplace-GRF Model. Assume $\boldsymbol{Z} = \mathbb{R}^{k\times n}$. Often we have the prior knowledge that neighboring sparse coefficient $\boldsymbol{z}^i, \boldsymbol{z}^j$ are close. This intuition leads to combining Laplace prior and Gaussian Random Field (GRF). Let

$$\begin{aligned} \mathcal{S}^i &= ||\boldsymbol{z}^i||_1 \\ \mathcal{C}^{ij} &= w_{ij}||\boldsymbol{z}^i - \boldsymbol{z}^j||^2 \end{aligned} .$$

The distribution $p(\boldsymbol{Z}\ ;\ \boldsymbol{\Gamma})$ is jointly Gaussian.

Beta-Ising Model. Assume $\boldsymbol{Z} = \{0,1\}^{k\times n}$. This case is appealing for potential social network applications. For example, z_{ij} could be a latent feature indicating whether user i likes the product j. The Beta process (7) enforces the sparsity of binary coefficients. And based on similar idea of closeness, the Ising model shows a way to incorporate pairwise correlation into the model:

$$\mathcal{C}^{ij} = \sum_{k=1}^{K} w_{ij}(2z_{ik}-1)(2z_{jk}-1). \tag{11}$$

3.2 DIFFUSION PROCESS

We show that the pairwise correlation for real-valued signals can be generalized through a Diffusion Process (DP) on the graph G. The Correlated Compressive Sensing with Diffusion Process (`CorrCS-DP`) characterizes the covariant structure of the latent signals with a generative model, whose zeroth-order approximation is compressive sensing, and first-order approximation is pairwise `CorrCS`.

Diffusion Process. For any graph $G(V,E)$, a value function $f : V \to \mathbb{R}$ can be defined. Diffusion Process (DP) is a natural class of stochastic processes on graphs that yields covariance structure of the function f (Kondor and Lafferty, 2002). First, we extend our value function as a function of time t: define $\boldsymbol{f}[t]$ be the snapshot vector of $[f(v_1), f(v_2), ..., f(v_n)]$ at time t. Next, a diffusion generator H is defined as a Laplacian matrix of graph G as:

$$H_{ij} = \begin{cases} w_{ij} & \text{for } i \neq j \text{ and } j \in \mathcal{A}(i) \\ -\sum_{i'} w_{ii'} & \text{for i = j} \\ 0 & \text{otherwise.} \end{cases} \tag{12}$$

Then H is applied to the value function in the following way

$$\frac{\partial \boldsymbol{f}[t]}{\partial t} = \alpha H \boldsymbol{f}[t]. \tag{13}$$

Solving (13), we obtain

$$\boldsymbol{f}[t] = K\boldsymbol{f}[0], \tag{14}$$

where $K = \exp(\alpha t H)$ is the *Diffusion Kernel*. Notice that heat kernel is always invertible, which means given $\boldsymbol{f}[t]$ at any time t it is easy to compute $\boldsymbol{f}[0] = \exp(-\alpha t H)\boldsymbol{f}[t]$. Notice that when $t = 0$, $K = I$; when t is small, we have the first-order approximation $K = I - \alpha t H$.

Correlation via Diffusion. In `CorrCS`, we can define $f_k : V \to \mathbb{R}$ for each dimension of the features as $f_k^i = z_k^i$, so the snapshot $\boldsymbol{f}_k$ is a vector. By studying the statistical characteristics of f_k, $k = 1, 2, ..., K$, we can then build a correlated sparse signal model $p(\boldsymbol{Z})$, which is the core of `CorrCS-DP`.

Imagine that $\boldsymbol{f}_k$ is generated through the following process: Initially $\boldsymbol{f}_k[0]$ is sparse. This means each entry $f_k^i[0]$ is distributed i.i.d as

$$f_k^i[0] \sim \mathcal{N}\Big(0, \gamma_k^i\Big), \quad \forall v_i \in V \tag{15}$$

Hyper-parameters γ_k^i control the sparsity of each entry. They are both drawn from hyper-priors according to equation (6).

A diffusion process is then run on the graph $G(V, E)$ and stops at some time t producing $\boldsymbol{f}_k[t] = K\boldsymbol{f}_k[0]$. Using the diffusion-based generative model, the inference problem (9) for real-valued signals becomes

$$p(\boldsymbol{Z} \mid \boldsymbol{Y}; \boldsymbol{\Gamma}) \propto \Pr(\boldsymbol{\Gamma}) \exp\Big(-\frac{1}{2}\sum_k (\boldsymbol{f}^k)^\top (K^{-1})^\top D_k K^{-1} \boldsymbol{f}^k \\ -\frac{1}{2}\beta \sum_i (\boldsymbol{y}_i - M_i \boldsymbol{x}_i)^\top (\boldsymbol{y}_i - M_i \boldsymbol{x}_i)\Big) \tag{16}$$

where $D_k = \text{diag}(\boldsymbol{\gamma}^k)^{-1}$. Notice that in this formulation, we no longer need constant c to control the extent of correlation, since it is directly induced by the prior uncertainty γ_k^i. A few observations can be made about the connection of `CorrCS-DP` to other compressive sensing methods summarized as follows.

Proposition 3.1. *Using zeroth-order approximation $K := I$, where I is the identity matrix,* `CorrCS-DP` *subsumes BCS.*

Proof. Straightforward. By replacing $K := I$ in (16), the posterior $p(\boldsymbol{X}|\boldsymbol{Y})$ is exactly the same as BCS. □

Proposition 3.2. *Using first-order approximation $K := I - \alpha t H$,* `CorrCS-DP` *reduces to* `Laplace-GRF` *model.*

Proof. This claim is induced by the general property of Laplacian H that $\boldsymbol{f}_k^\top H \boldsymbol{f} = \sum_{ij} w_{ij}(f_k^i - f_k^j)^2$. Using the first order approximation, we have

$$-\log p(\boldsymbol{Z}|\boldsymbol{Y}; \boldsymbol{\Gamma}) = \frac{1}{2}\sum_{ik} (\frac{\boldsymbol{f}_i^k}{\gamma_i^k})^2 - \alpha t \sum_k (\boldsymbol{f}^k)^\top D_k H \boldsymbol{f}^k \\ + Const \tag{17}$$

Let $d_i^k = 1/(\gamma_i^k)^2$. Notice H is a Laplacian matrix,

$$\begin{aligned} -(\boldsymbol{f}^k)^\top D_k^H \boldsymbol{f}^k &= \sum_{ij} w_{ij} \Big(d_i^k (f_i^k)^2 + d_j^k (f_j^k)^2 \\ &\quad - (d_i^k + d_j^k) f_i^k f_j^k \Big) \\ &= \sum_{ij} w_{ij} (\frac{d_i + d_j}{2})(f_i^k - f_j^k)^2 \\ &\quad - \frac{\alpha t}{2} \sum_i (H\boldsymbol{d}^k)_i \end{aligned}$$

Let

$$\mathcal{S} = \frac{1}{2}\sum_{ik} (f_i^k)^2 ((I - \frac{\alpha t}{2}) d^k)_i$$

and

$$\mathcal{C}_{ij} = w_{ij}(\frac{d_i + d_j}{2})(f_i^k - f_j^k)^2,$$

then $p(\boldsymbol{X}|\boldsymbol{Y}; \boldsymbol{\Gamma}) \propto \exp(-\mathcal{S} - \sum_{ij} \mathcal{C}_{ij})$ shows that `Corr-DP` reduces to a pairwise correlation model. □

3.3 INFERENCE ALGORITHM

The exact inference of `CorrCS` is largely intractable due to two reasons. First, the signal model and the sensing model is not in same conjugate family. Second, even if $p(\boldsymbol{X}; \boldsymbol{\Gamma})$ is jointly Gaussian, in real applications either the number of nodes or the dimension of features is big.

Instead, we resort to approximation methods and develop the posterior inference based on Variational Bayes EM (Bernardo et al., 2003). In particular, we use the mean-field approximation $p(\boldsymbol{X}|\boldsymbol{Y}; \boldsymbol{\Gamma}) = \prod_{ik} q(z_{ik}; \boldsymbol{\Gamma})$ and perform the following two-step scheme. In the E-step, we propagate information across nodes to spread correlation, which can be related to a message-passing process (Donoho et al., 2009); in the M-step, we update $\boldsymbol{\Gamma}$ to enforce sparsity. The details are outlined in Algorithm 1.

Algorithm 1 Correlated Compressive Sensing

Input: Network $G(V, E)$, $\boldsymbol{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2, ..., \boldsymbol{y}_n]$, basis $\boldsymbol{\Phi}$, measurement matrices V^i and `iter`.
for $i = 1 \rightarrow n$ **do**
    compute $M^i = V^i \boldsymbol{\Phi}$ for all $i = 1, 2, ..., n$.
    initialize $\boldsymbol{z}^i = (M^i)^\top \boldsymbol{y}_i$.
end for
for $j = 1 \rightarrow$ `iter` **do**
    for $i = 1 \rightarrow n$ **do**
        for $k = 1 \rightarrow k$ **do**
            Update factor $q_k^i(\boldsymbol{z}_k^i)$ using equation (19), (20), (21).
        end for
    end for
    for $i = 1 \rightarrow n$ **do**
        For binary case, use equation (22) to update $\boldsymbol{\pi}^i$; for real-valued case, use equation (24) to update γ^i.
    end for
end for

E-Step: Spread Correlation.

In the pairwise case, the update algorithm in general is

$$q_i(z_k^i) \propto \exp\Big(\mathbb{E}_{q(\boldsymbol{z}_{\neg k}^i)}[\tfrac{1}{2}\beta||\boldsymbol{y}^i - M_i \boldsymbol{z}_i||^2 + \mathcal{S}_i] + c \sum_{j \in \mathcal{A}(i)} \mathbb{E}_{q_j(\boldsymbol{z}^j)}[\mathcal{C}_{ij} + \mathcal{C}_{ji}]\Big). \quad (18)$$

where $\boldsymbol{z}_{\neg k}^i$ denotes all variables in $\boldsymbol{z}^i$ except z_k^i. Intuitively, the first expectation in (18) propagates information across dimensions of $\boldsymbol{z}^i$, while the second expectation in (18) spreads correlation among different nodes on the graph via the edges in between. Notice that for directed networks $w_{ij} \neq w_{ji}$, the information propgates forward and backward the edge in the same way, due to the symmetry of $\mathcal{C}_{ij} + \mathcal{C}_{ij}$.

Specifically, for `Beta-Ising` model,

$$\begin{aligned} q(z_k^i = 1) \; &\propto \pi_k^i \exp\Big(- \tfrac{1}{2}\beta(M_k^i)^\top (M_k^i) \\ &\quad -2(M_k^i)^\top(\boldsymbol{y}^i - M_{\neg k}^i \mathbb{E}[\boldsymbol{z}_{\neg k}^i]) \\ &\quad + c \sum_{j \in \mathcal{A}(i)} (w_{ij} + w_{ji})(2z_k^i - 1)\Big) \\ q(z_k^i = 0) \; &\propto 1 - \pi_k^i. \end{aligned} \quad (19)$$

Similarly for `Laplace-GRF` model, the mean-field update for each factor is $q_k^i(z_k^i) = \mathcal{N}(\mu_k^i, \sigma_k^i)$, where

$$\begin{aligned} \sigma_k^i &= (\beta(M_k^i)^\top M_k^i + 1/\gamma_k^i)^{-1} \\ \mu_k^i &= \sigma_k^i \cdot \Big(\beta(M_k^i)^\top(\boldsymbol{y}^i - M_{\neg k}^i \boldsymbol{\mu}_{\neg k}^i) \\ &\quad + c \sum_{j \in \mathcal{A}(i)} (w_{ij} + w_{ji})\mu_k^j\Big). \end{aligned} \quad (20)$$

where $\boldsymbol{\mu}_{\neg k}^i$ denotes all entries in $\boldsymbol{\mu}^i$ except μ_k^i. For the extention of `Laplace-GRF` model, `CorrCS-DP` contains long-range interaction among the node, so all other nodes contribute to the distribution of the current node being updated. As in `Laplace-GRF`, we still have $q_k^i(z_k^i) = \mathcal{N}(\mu_k^i, \sigma_k^i)$, but instead

$$\begin{aligned} \sigma_k^i &= \Big(\beta(M_k^i)^\top M_k^i + U_{ii}^k\Big)^{-1} \\ \mu_k^i &= \sigma_k^i \cdot \Big(\beta(M_k^i)^\top(\boldsymbol{y}^i - M_{\neg k}^i \boldsymbol{\mu}_{\neg k}^i) \\ &\quad + \sum_{j \in V} (U_{\neg i}^k + (U^k)_{\neg i}^\top)^\top \boldsymbol{u}_k^j\Big). \end{aligned} \quad (21)$$

where $U^k = K^{-T} D_k K^{-1}$ and $\boldsymbol{u}_k = [\mu_k^1, \mu_k^2, ..., \mu_k^n]$.

Iteratively updating the factors according to equation (20), (19) and (21) guarantees convergence (Wainwright and Jordan, 2008).

M-Step: Update hyper-parameters. With the expectation of current belief about the signal to recover, we can further update the hyper-parameters $\boldsymbol{\Gamma}$ to enforce sparsity based on EM algorithm (Dempster et al., 1977). For `Beta-Ising`, we update the parameters of the Bernoulli prior π_k^i as follows

$$\pi_k^i \sim \frac{a/K + \mathbb{E}[z_k] - 1}{a/K + b(1 - 1/K) - 1}. \quad (22)$$

For `Laplace-GRF`, we update the global parameters

$$\gamma_k^i = -\frac{1}{2\lambda^2} + \sqrt{\frac{1}{4\lambda^2} + \frac{(\sigma_k^i + (\mu_k^i)^2)}{\lambda}}. \quad (23)$$

The update of γ_k^i in `CorrCS-DP` is similar to `Laplace-GRF`. Specifcally, let Q^k be a diagonal matrix at time t such that $Q_{ii}^k = \sigma_k^i$, compute $\widetilde{Q}^k = K^{-1} Q^k K^{-\top}$, which can be regarded as the uncertainty about $\boldsymbol{f}_k[0]$. Then we modify (24) as

$$\gamma_k^i = -\frac{1}{2\lambda^2} + \sqrt{\frac{1}{4\lambda^2} + \frac{\widetilde{Q}_{ii}^k}{\lambda}}. \quad (24)$$

Combining E-step and M-step, we can jointly optimize $\boldsymbol{\Gamma}$ and infer $\boldsymbol{Z}$, which eventually recovers the networked data on the graph G.

4 EXPERIMENT

We evaluate Correlated Compressive Sensing (`CorrCS`) empirically on real datasets from social and sensor networks with pairwise or Diffusion-like correlation.

4.1 SOCIAL NETWORK DATA WITH PAIRWISE CORRELATION

Using compressive sensing with pairwise correlation, we test the two recovery models `Laplace-GRF` and `Beta-Ising` on two datasets: product review on Epinion [1] and consumer polling data in Michigan.

4.1.1 Performance on different datasets

Michigan Polling Data. The social polling data is collected from a survey of consumers in Michigan with 500 monthly telephone calls from January, 1978 to December 2012. The data is real-valued aggregation of four hundred economic indices. It has been known that pairwise correlation exists in these indices. Using half of the dataset as past history, this pairwise correlation is computed through Pearson correlation and taken as the weight w_{ij}. Therefore, we establish a graph of features representing their inherent correlation. Furthermore, the data is continuous real values and to sparsify it, we use online sparse matrix factorization (Mairal et al., 2010) to find a set of sparse basis.

We test `Laplace-GRF` model on the Michigan polling data with $a = 1, b = 0.1$ and $c = 1$. As measurements, we randomly select a fraction of the polling data for each feature. We refer to the dimension of the selected data versus the dimension of the original data as measurement ratio. The performance is evaluated via Mean Squared Error (MSE) of the recover signal with respect to the original one. The MSE is normalized by the 2-norm of the original signal. For independent compressive sensing, we include two popular choices – Bayesian Compressive Sensing (Ji et al., 2008)(BCS) and Orthogonal Matching Pursuit (Tropp and Gilbert, 2007)(OMP) – as baseline algorithms. Notice that some baseline implementations of compressive sensing, such as Basis Pursuit (Chen et al., 1998)(BP), are too slow and impractical for networked data.

The result is reported in Figure 2. As we can see, `Laplace-GRF` outperforms BCS and OMP largely for small measurement ratios (less than 0.6). When the measurement ratio is large (exceed 0.6), `Laplace-GRF` will have a performance similar to BCS, and it still outperform OMP to a big extent. Correlation is less useful for large measurement ratios, mainly because the numbers of observations are already sufficient for sparse recovery. However, since in practice, people usually use a measure ratio in the rang $[0.2, 0.5]$, and rarely measure more than 60% of the data for recovering. Hence, this will not have strong influence on `Laplace-GRF`'s practical use.

To measure the superiority of `Laplace-GRF` for small measurement ratio, we could compare the minimum measurement ratio that is needed to achieve a fixed accuracy level (i.e. MSE level) for different models. Like in this dataset, we compare the minimum measurement ratio for different models to achieve a MSE not more than 0.2: our `Laplace-GRF` needs a measurement ratio of 0.23, BCS needs 0.32, while the most common baseline algorithm OMP needs 0.56. We could see that, to achieve this same recovery effect, `Laplace-GRF` needs 21% less measurements than BCS, and 59% less measurements than OMP. This result may imply that `Laplace-GRF` model could have valuable practical use since we could use a small number of measurements to recover a polling result which is very close to the original ones. And by considering the inherent correlation between these indices, we could reduce the number of measurements by a ratio of 21% to recover a result with MSE 0.2 on this data, which means we could save a great deal of costs for this poll.

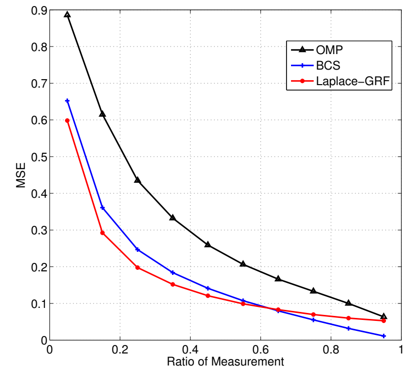


Figure 2: Performance of `Laplace-GRF` model versus the BCS model and OMP on the polling data.

Epinion Data. The Epinion data is derived from the social product review network Epinions with 17,022 customers and 139,738 products. The graph G is built from the trust-list of all users: $w_{ij} = 1$ if and only if user i trusts user j, and therefore it is directed. To reduce the dimensionality of features, we select a subset of the most 100 popular products. Then z_{ij} represents whether customer i liked product j. The data $\boldsymbol{Z}$ is inherently sparse with only 5 to 10 nonzero per column, because the fraction of products rated by each customer is small. As measurements, each column $\boldsymbol{z}^i$ is projected to a low-dimensional space.

We test `Beta-Ising` model on the Epinion dataset against Bayesian Compressive Sensing (BCS) with beta prior. We choose $\lambda = 1, c = 0.3$. For the binary Epinion data, MSE is not a good choice because the data is zero almost everywhere. Instead, we regard the recovery as predicting label z_{ij} and use F1 score from classifier evaluations, which is the harmonic mean of precision (ratio of number of correct 1's we recover over the total number of 1's in our recovery result) and recall (ratio of number of correct 1's we recover over the total number of 1's in the ground truth).

[1] Available at trustlet.org.

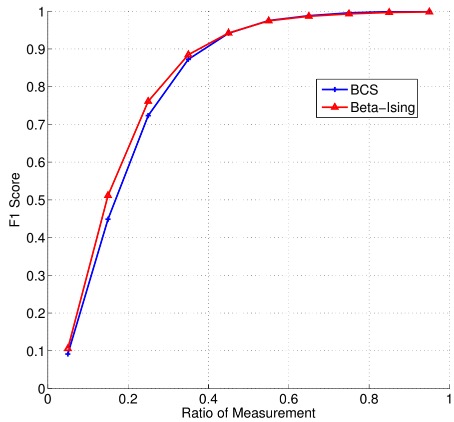


Figure 3: Performance of `Beta-Ising` model versus the Bayesian-Compressive Sensing on the Epinion data.

Figure 3 shows the result under F1 score. It can be seen that with correlation, the performance of BCS can be improved with a varying number of measurement ratios.

4.1.2 Sensitivity evaluation

Impact of parameters. In these two datasets, all parameters we could set are the parameter of the hierarchical conjugate prior a, b, λ and weight c. As has been discussed in BCS (Ji et al., 2008), our models are not sensitive the parameters a, b, λ. So in this paper, we focus on the impact of the weight parameter c on the performance of the `CorrCS` models.

The choice of c is the key to `CorrCS`, which can be viewed as a regularization parameter controlling the tradeoff with sparsity. To evaluate the impact of the c on the performance, we test the variation of the behavior of the `CorrCS` model on different scenario. For example, consider the `Beta-Ising` model on the Epinion data, we compare the variation of the recovery F1 score corresponding to the change of c when some reasonable measurement ratios are selected (15%, 25% and 35%). Figure 4 shows that among all these 3 measurement ratios, the performance of this model will be improved when c starts to grow from 0, and will be demoted when c pass some specific values. This result accords with our discussion about c in Section 3.1. This kind of variation of the performance according to c's variation is reasonable since different values of c imply different extents that we care about the inherent correlation between nodes in the network structure. This experiment show us that the performance will have the same trends of variation on different measurement ratios when c is changing.

It is worth noting that among the 3 reasonable measurement ratios, the best c's that will induce an optimal performance

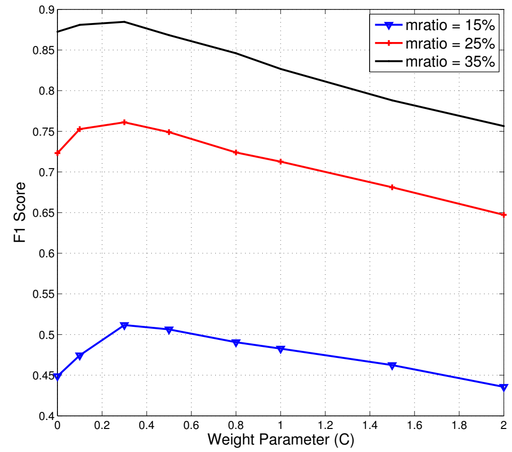


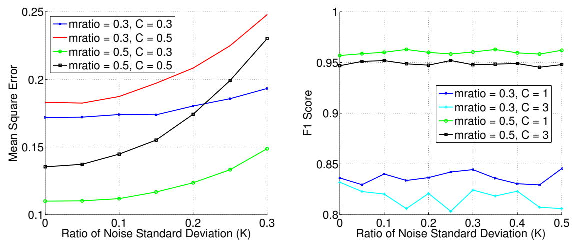
Figure 4: Relationship between the performance of the Beta-Ising model and the parameter c on different measure ratios.

are very close. As shown in Figure 4, the model will possess an optimal performance when c is some value among 0.3. If $c = 0.3$, then the model will always have a nearly optimal recovery result as long as the measurement ratio is in a reasonable range. Therefore, in this model, we could choose an optimal choice of c that works well on all reasonable cases.

Impact of noise on the measured data. To test the robustness of these two models, we test the performance of our model on the two datasets when noises are added. More precisely, we add a Gaussian noise at each dimension of the observation $\boldsymbol{y}$, where the standard deviation of this noise at each dimension is κ times the original value of this dimension, where κ is Signal-to-Noise Ratio. We test the performance of the two models as κ increases from 0 to 0.5 on different scenarios (i.e. different measure ratios and weight parameters). As can be seen in Figure 5, this two models possess strong robustness on the two datasets since even if κ goes to 0.5 the recovery result will not vary too much. The `Beta-Ising` model on the Epinion dataset has a slightly better robustness than that of the `Laplace-GRF` model on the polling dataset since it deals with binary variables, whose robust recovery turns out to be easier.

4.2 POLLUTION DATA FROIM SENSOR NETWORK

The Beijing pollution data includes a network of 22 monitoring stations collecting data in the same time window from Feb. 8th 2013 to Dec. 17 2013. The Air Quality Indexes (AQI) PM2.5 is recorded at an interval of 1 hour. The geography information of the sensors are available as GPS coordinates (g_i, l_i), which we use to compute the edge weight of the sensor network through their euclidean dis-



(a) (b)

Figure 5: Robustness of `CorrCS` on the 2 datasets when some noise is added to the measure. **(a)**. Robustness of the Laplace-GRF model on the Polling dataset. **(b)**. Robustness of the Beta-Ising model on the Epinion dataset

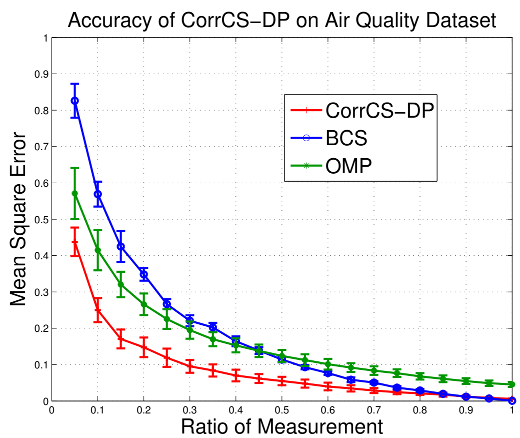


Figure 6: Performance of CorrCS-DP on Air Quality Dataset. Recovery accuracy of CorrCS-DP compared with BCS and OMP.

tance $e_{ij} = \exp(-\theta\sqrt{(g_i - g_j)^2 + (l_i - l_j)^2})$.

The time sequence data is divided into 22 chunks, with 2 weeks of pollution data each chunk. The data is then split into two parts. Then we use 11 chunks to train a set of sparse basis using online sparse matrix factorization (Mairal et al., 2010), and also as cross-validation to find the best choice of diffusion time $t = 0.1$ and $c = 5$. The rest 11 chunks are used to test `CorrCS` and its counterparts. To simulate a real setting of measurement, we randomly select a portion of the samples as measurement and try to recover the rest. To test recovery accuracy in various situations, we change the ratio of measurement from 0 to 1 and compute the mean square error of the recovered signals.

Figure 6 and 7 shows the recovery accuracy and convergence rate of `CorrCS-DP` on the pollution dataset with comparison to the compressive sensing counterparts. The results are averaged over 10 independent runs. From Figure 6 we can see that `CorrCS` largely improves the recovery

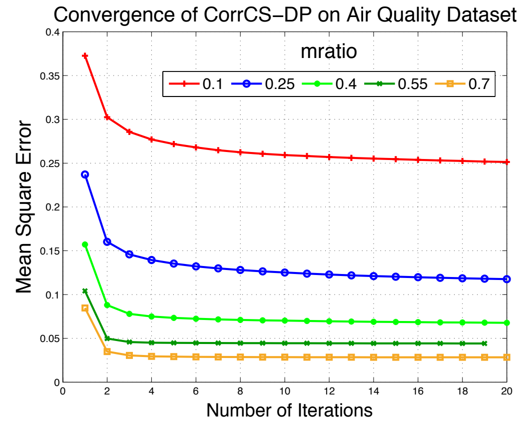


Figure 7: Convergence of recovery accuracy with number of iterations.

Objective MSE	`CorrCS-DP`	OMP	BCS
0.1	0.30	0.60	0.54
0.15	0.18	0.41	0.43
0.2	0.13	0.29	0.35

Table 1: Minimum measurement ratio to reach some particular values of MSE on the AQI dataset on different models

accuracy for various ratio of measurement, due to exploiting the correlation among different nodes. To measure the improvement of `CorrCS-DP` comparing to the other two models precisely, we could again compare the necessary minimum measurement ratio to reach some particular values of MSE on this data set on different models. As shown in Table 1, to reach the same MSE, `CorrCS-DP` could measure at least 40% less data than the other two models. Furthermore, as shown in Figure 7, `CorrCS` converges in about 3-4 steps.

5 RELATED WORK

Last two decades have witnessed significant advances in the theory and application of sensing sparse signals. Compressive sensing exploited the fact that natural signals are sparse and compressive under proper basis and designed sampling algorithms beyond the Nyquist-Shannon limit (Candès and Wakin, 2008). The theory of compressive sensing was developed by Candès et al. (2006) to explain this novel recovery performance. This theory was further improved by Candès and Tao (2006) to account for noisy measurements. The underlying property empowering sparse recovery is Restricted Isometry Property (RIP) of measurement matrices (Candès, 2008). On the algorithmic side, the first attempts to solve compressive sensing problems rely on ℓ_1 minimization under linear programming (Chen et al., 1998; Candès and Tao, 2005). Instead of optimizing with a large number of constraints, Orthogonal Matching Pursuit (Tropp and Gilbert, 2007) used a greedy heuristic to find solutions close to ℓ_0 optimum. To facilitate large-scale applications, Donoho et al. (2009) borrowed ideas from graphical models and derived a message passing algorithm for compressive sensing.

The Bayesian formulation of compressive sensing (BCS) is first proposed by Ji et al. (2008), which used a tractable conjugate Gamma prior on signal precision to enforce sparsity. It was shown that Bayesian compressive sensing allows uncertainty estimate and adaptive sampling. Based on BCS, Ji et al. (2009) developed a multi-tasking compressive sensing algorithm that allows simultaneously data collection from multiple sensors. Babacan et al. (2010) showed that stronger sparsity can be achieved for BCS with an conjugate prior on signal variance that is equivalent to the Laplace prior. As a counterpart of Laplace prior, beta prior is also commonly used (Paisley and Carin, 2009), with an additional latent variable controlling the support of signals. By comparing Laplace and Beta priors for sparse representation, Mohamed et al. (2011) concluded that Beta prior enforced stronger sparsity than the Laplace prior.

Real data is typically not sparse and therefore one must take effort in finding the appropriate basis. With a data-drive approach, dictionary learning for sparse basis originated from efforts in reproducing V1-like visual neurons through sparse coding (Olshausen and Field, 1997). Aharon et al. (2006) generalized the K-means clustering algorithm, and computed sparse decomposition by iteratively updating sparse coefficients and dictionary items. Mairal et al. (2009) proposed online dictionary learning methods, which leads to efficient computation of sparse coding.

Compressed sensing has find great applications in sensor networks. It was first successfully applied to network monitoring for optical and all-IP networks (Coates et al., 2007). In terms of data gathering, Luo et al. (2009) constructed a sensor network with a sink collecting compressed measurements, which is equivalent to a random matrix projection. Xu et al. (2013) considered more general compressed sparse functions for sparse representation of signals over graphs. Other than collecting data, compressive sensing was also used as a network analysis tool to identify social community on graphs (Jiang et al., 2011).

Diffusion process has long been used as a general tool to capture correlation among data (Kondor and Lafferty, 2002). Ma et al. (2008) used diffusion process to model marketing candidate selection in social networks. The diffusion process may also be utilized to produce an representing wavelet basis on graphs and manifolds (Bremer et al., 2006). Using diffusion wavelets, it is possible to sparsify signals on different graph topologies and allow compressive sensing (Haupt et al., 2008). However, our correlated compressive sensing does not rely on the strong assumption that data on the network should be sparse under some basis, but rather weakly correlated.

6 CONCLUSIONS AND FUTURE WORK

In this paper, we present Correlated Compressive Sensing (`CorrCS`) to leverage correlation among networked data and to empower better sparse recovery. Using a Bayesian approach, `CorrCS` allows flexible representation of prior knowledge about correlation via a graphical model. Two common types of correlation of networked data are considered: pairwise and diffusion-based. We have shown the diffusion-based formulation subsumes the pairwise case via a low-order approximation. Through extensive empirical evaluation on real data on social and sensor networks, we have demonstrated the advantage of correlated compressive sensing over its counterparts.

As future work, we are interested in showing bounds in its recovery performance to better understand its properties. Also we are interested in developing nonparametric extensions of the current approach to allow adaptive inference of key parameters and the basis for sparse representation.

References

Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322.

Anagnostopoulos, A., Kumar, R., and Mahdian, M. (2008). Influence and correlation in social networks. In *SIGKDD*, pages 7–15. ACM.

Babacan, S. D., Molina, R., and Katsaggelos, A. K. (2010). Bayesian compressive sensing using laplace priors. *Image Processing, IEEE Transactions on*, 19(1):53–63.

Backstrom, L., Huttenlocher, D., Kleinberg, J., and Lan, X. (2006). Group formation in large social networks: membership, growth, and evolution. In *SIGKDD*, pages 44–54. ACM.

Bernardo, J., Bayarri, M., Berger, J., Dawid, A., Heckerman, D., Smith, A., West, M., et al. (2003). The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. *Bayesian statistics*, 7:453–464.

Bremer, J. C., Coifman, R. R., Maggioni, M., and Szlam, A. D. (2006). Diffusion wavelet packets. *Applied and Computational Harmonic Analysis*, 21(1):95–112.

Candès, E. J. (2006). Compressive sampling. In *Proceedings of the international congress of mathematicians*, volume 3, pages 1433–1452. Madrid, Spain.

Candès, E. J. (2008). The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*, 346(9):589–592.

Candès, E. J., Romberg, J., and Tao, T. (2006). Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *Information Theory, IEEE Transactions on*, 52(2):489–509.

Candès, E. J. and Tao, T. (2005). Decoding by linear programming. *Information Theory, IEEE Transactions on*, 51(12):4203–4215.

Candès, E. J. and Tao, T. (2006). Near-optimal signal recovery from random projections: Universal encoding strategies? *Information Theory, IEEE Transactions on*, 52(12):5406–5425.

Candès, E. J. and Wakin, M. B. (2008). An introduction to compressive sampling. *Signal Processing Magazine, IEEE*, 25(2):21–30.

Chen, S. S., Donoho, D. L., and Saunders, M. A. (1998). Atomic decomposition by basis pursuit. *SIAM journal on scientific computing*, 20(1):33–61.

Coates, M., Pointurier, Y., and Rabbat, M. (2007). Compressed network monitoring for IP and all-optical networks. In *SIGCOMM*, pages 241–252. ACM.

Dempster, A. P., Laird, N. M., Rubin, D. B., et al. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal statistical Society*, 39(1):1–38.

Donoho, D. L. (2006). Compressed sensing. *Information Theory, IEEE Transactions on*, 52(4):1289–1306.

Donoho, D. L., Maleki, A., and Montanari, A. (2009). Message-passing algorithms for compressed sensing. *PNAS*, 106(45):18914–18919.

Fortunato, S. (2010). Community detection in graphs. *Physics Reports*.

Haupt, J., Bajwa, W. U., Rabbat, M., and Nowak, R. (2008). Compressed sensing for networked data. *Signal Processing Magazine, IEEE*, 25(2):92–101.

Ji, S., Dunson, D., and Carin, L. (2009). Multitask compressive sensing. *Signal Processing, IEEE Transactions on*, 57(1):92–106.

Ji, S., Xue, Y., and Carin, L. (2008). Bayesian compressive sensing. *Signal Processing, IEEE Transactions on*, 56(6):2346–2356.

Jiang, X., Yao, Y., Liu, H., and Guibas, L. (2011). Compressive network analysis. *arXiv preprint arXiv:1104.4605*.

Kemp, C., Griffiths, T. L., and Tenenbaum, J. B. (2004). Discovering latent classes in relational data.

Kondor, R. I. and Lafferty, J. (2002). Diffusion kernels on graphs and other discrete input spaces. In *ICML*, volume 2, pages 315–322.

Luo, C., Wu, F., Sun, J., and Chen, C. W. (2009). Compressive data gathering for large-scale wireless sensor networks. In *MobiCom*, pages 145–156. ACM.

Lustig, M., Donoho, D. L., Santos, J. M., and Pauly, J. M. (2008). Compressed sensing MRI. *Signal Processing Magazine, IEEE*, 25(2):72–82.

Ma, H., Yang, H., Lyu, M. R., and King, I. (2008). Mining social networks using heat diffusion processes for marketing candidates selection. In *CIKM*, pages 233–242. ACM.

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2009). Online dictionary learning for sparse coding. In *ICML*, pages 689–696. ACM.

Mairal, J., Bach, F., Ponce, J., and Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *JMLR*, 11:19–60.

Mohamed, S., Heller, K., and Ghahramani, Z. (2011). Bayesian and L1 approaches to sparse unsupervised learning. *arXiv preprint arXiv:1106.1157*.

Olshausen, B. A. and Field, D. J. (1997). Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision research*, 37(23):3311–3325.

Paisley, J. and Carin, L. (2009). Nonparametric factor analysis with beta process priors. In *ICML*, pages 777–784. ACM.

Seeger, M. W. and Nickisch, H. (2008). Compressed sensing and Bayesian experimental design. In *ICML*, pages 912–919. ACM.

Tropp, J. A. and Gilbert, A. C. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *Information Theory, IEEE Transactions on*, 53(12):4655–4666.

Wainwright, M. J. and Jordan, M. I. (2008). Graphical models, exponential families, and variational inference. *Foundations and Trends in Machine Learning*, 1(1-2):1–305.

Xu, L., Qi, X., Wang, Y., and Moscibroda, T. (2013). Efficient data gathering using compressed sparse functions. In *INFOCOM, 2013 Proceedings IEEE*, pages 310–314. IEEE.

Yang, J., Wright, J., Huang, T., and Ma, Y. (2008). Image super-resolution as sparse representation of raw image patches. In *CVPR*, pages 1–8. IEEE.