

# UCS-WN: An Unbiased Compressive Sensing Framework for Weighted Networks

Hamidreza Mahyar\*, Hamid R. Rabiee<sup>†</sup>, Zakieh S. Hashemifar\* and Payam Siyari\*

Department of Computer Engineering, Sharif University of Technology (SUT)

Email: \*{hmahyar, hashemifar, siyari}@ce.sharif.edu, <sup>†</sup>rabiee@sharif.edu

**Abstract**—In this paper, we propose a novel framework called UCS-WN in the context of compressive sensing to efficiently recover sparse vectors representing the properties of the links from weighted networks with  $n$  nodes. Motivated by network inference, we study the problem of recovering sparse link vectors with network topological constraints over weighted networks. We take sufficient number of collective additive measurements using this framework through connected paths for constructing a feasible measurement matrix. We theoretically show that only  $O(k \log(n))$  path measurements via UCS-WN are sufficient for uniquely recovering any  $k$ -sparse link vector with no more than  $k$  non-zero elements. Moreover, we demonstrate that this framework would converge to an accurate solution for a wide class of networks by experimental evaluations on both synthetic and real datasets.

## I. INTRODUCTION

Compressive Sensing (CS) [1–5] is a new paradigm in signal processing for efficient sparse signal recovery which aims to simultaneously sample and compress sparse signals. It tries to recover high dimensional signals from a total number of measurements much less than their dimensions. This is possible if we have prior knowledge about some properties (*i.e.* sparsity) of the signal to be recovered.

Compressive sensing has drawn much attention in signal processing for the last couple of years, but its role in communication networks is still in its early stages due to the existing challenges. One of these challenges is the construction of measurement matrix that must be feasible (with non-negative integer entries) considering the topology of network (graph), and it should also incorporate the required properties for uniquely recovering any  $k$ -sparse link vector. In other words, only links that induce a connected sub-graph in the underlying topology can be aggregated together in the same measurement. There have only been a few recent works which consider network topological constraints in order to design a feasible measurement matrix over networks (graphs) using compressive sensing [6–10].

To the best of our knowledge, all of the past studies for CS applications in networking rely on the assumption that the underlying topology of network is a simple graph and the measurements do not impose any considerable running cost. But, this assumption is not necessarily true. Most of the networked structures in the real world not only have often a complex topological structure, but also demonstrate a heterogeneous intensity and capacity over their connections. Hence, modelling of such networks should go beyond sim-

ple graphs with links states representing either existence or absence of an edge (link) in the graph of a network. In practice, those links must be assigned a given weight (capacity, distance, etc.) and such models are generally called “*weighted networks (graphs)*”. For example in transportation systems and computer networks, the distance of the roads and bandwidth of the links may represent the weight of the connections in these networks, respectively. In weighted networks, each constructed measurement would impose a cost equal to the sum of the weights of traversed links in the same measurement. Although, the total cost would be the sum of the costs of all measurements.

Two main topological characterizations of networks are the probability distribution  $p(d)$  that a node has degree  $d$ , and the distribution  $p(w)$  that any given link has weight  $w$ . To the best of our knowledge, these features have been completely ignored in the past studies of network monitoring via compressive sensing. For example, one of the state-of-the-art CS-based algorithms for sparse recovery in networks, uses Random Walk (RW) to construct the measurements [10]. However, RW is significantly biased towards high-degree nodes and high-weighted links. So, it may be inapplicable in many networks.

Therefore, due to the broad node degree distribution and topological weighted characterization of networks, the constructed measurement matrix has to provide the maximum information coverage using the minimum number of measurements while considering the needs for unbiasedness and lower total cost. For instance, the bandwidth of the links in computer networks represent the cost of the links and the CS algorithms should try to avoid any biasedness towards high-bandwidth links, although they should recover the sparse vectors with the minimum possible bandwidth usage (cost).

In this paper, for the first time, we introduce a general framework called “UCS-WN” (Unbiased Compressive Sensing for Weighted Networks) that considers the aforementioned challenges in the context of compressive sensing. Our specific works in this paper are summarized as follows:

- To the best of our knowledge, this is the first work that provides an unbiased compressive sensing framework for weighted network inference. We theoretically prove that the UCS-WN is unbiased towards high-degree nodes (section III and IV).
- We provide a feasible measurement construction with sufficient information coverage over weighted networks by using the UCS-WN framework. The number of required

measurements to recover any  $k$ -sparse link vector in the UCS-WN is only  $O(k \log(n))$  (section V).

- We theoretically prove the null-space property for the constructed measurement matrix which guarantees the correctness of the matrix (section V).
- We have also evaluated the UCS-WN performance by extensive simulations. The results demonstrate that our approach can recover sparse vectors with lower costs in comparison with the RW method (section VI).

## II. MODEL AND PROBLEM FORMULATION

We consider a weighted network, represented by an undirected weighted graph  $\mathcal{G} = (V(\mathcal{G}), E(\mathcal{G}))$ , where  $V(\mathcal{G}) = \{v_1, v_2, \dots, v_n\}$  denotes the set of nodes (vertices) with cardinality  $|V| = n$ , and  $E(\mathcal{G}) = \{e_1, e_2, \dots, e_N\}$  is the set of links with cardinality  $|E| = N$ . For a node  $v \in V$ , we denote its degree by  $deg(v)$  and the list of its neighbors by  $nbr(v) \subset V$ . The weight of link  $(u, v) \in E$  is denoted by  $w(u, v)$  and the weight of node  $u \in V$  is given by:

$$w(u) = s(u) = \sum_{v \in nbr(u)} w(u, v), \quad (1)$$

and the vector of link weights is represented by  $W(\mathcal{G}) = [w(u, v)]_N$ . A very significant measure of the network properties in terms of the actual weights is obtained by looking at the node strength  $s(u)$  defined in Eq. (1). We investigated some weighted networks (such as C.elegans and USTop500 in section VI), and found that in general, the links with larger weights are pointing to the neighbors of larger degree nodes. Therefore, we may assume  $d_{nn,u}^w > d_{nn,u}$  holds, where  $d_{nn,u}$  is the *average nearest neighbors degree*, and  $d_{nn,u}^w$  denotes the *weighted average nearest neighbors degree* which is calculated by:

$$d_{nn,u}^w = \frac{1}{s(u)} \sum_{v=1}^n a(u, v) \times w(u, v) \times deg(v), \quad (2)$$

where  $a(u, v)$  indicates the existence of a link between nodes  $u$  and  $v$ .

Suppose every link  $i$  has a real value  $x_i$ , and vector  $\mathbf{x} = (x_i, i = 1, 2, \dots, N)$  is associated with  $E(\mathcal{G})$ . Then,  $\mathbf{x}$  is a  $k$ -sparse link vector if  $\|\mathbf{x}\|_0 = k$ , where  $\|\cdot\|_0$  denotes the number of non-zero elements of  $\mathbf{x}$ . Suppose that we have  $m$  end-to-end (aggregate) measurements over the network ( $m \ll N$ ) which are some connected routes (paths) over  $\mathcal{G}$ . We would like to identify these  $k$  links (*i.e.* congested links with larger delays) from these measurements.

Let  $\mathbf{y} \in \mathcal{R}^m$  be the vector of  $m$  measurements whose each entry represents the total link delays of a connected path in  $\mathcal{G}$ , and  $\mathcal{A}$  be an  $m \times N$  measurement matrix where its  $i$ -th row corresponds to the  $i$ -th measurement. Here,  $\mathcal{A}_{ij} = 1$  ( $i = 1, \dots, m, j = 1, \dots, N$ ) if and only if the  $i$ -th measurement includes link  $j$ , and zero otherwise. Hence, in the compact form we can write  $\mathbf{y} = \mathcal{A}\mathbf{x}$ .

The cost of each measurement  $i$  (*i.e.* bandwidth usage) is calculated by the sum of the weights of visited links in that

measurement, which is denoted by  $C_{m_i}$ , and the total cost for recovering sparse vector  $\mathbf{x}$  would be:

$$C_{total} = \sum_{i=1}^m C_{m_i} \quad (3)$$

In this paper, we will discuss how to construct a feasible measurement matrix under network topological constraints with respect to node degree distribution and total links weights. In other words, walking on the lower-degree nodes and lower-weight links are considered as the factors for optimizing measurement constructions. Therefore, the constructed measurements will be completely unbiased and relatively low cost.

## III. THE PROPOSED FRAMEWORK: UCS-WN

In this section, we propose an Unbiased Compressive Sensing framework for Weighted Networks (UCS-WN) which is an efficient sparse recovery algorithm to recover any  $k$ -sparse link vector with relatively low cost in a sufficiently connected weighted network. In this approach, we construct a feasible measurement matrix  $\mathcal{A}$  to infer the link parameters (such as delay) inside a weighted network through end-to-end probing between nodes along some random routes (paths) on the network. The constructed matrix  $\mathcal{A}$  from the UCS-WN should hold the following properties, and we will theoretically prove the last two, in section V. These conditions are:

- 1) Each measurement (with non-negative integer entries) has to be feasible in the sense that the links of the same measurement should correspond to connected sub-graph. This condition represents sparse recovery with network topological constraints.
- 2) This measurement matrix has to satisfy the null-space property for uniquely recovering sparse link vectors.
- 3) The generated measurement matrix  $\mathcal{A}$  from the UCS-WN will be able to recover any  $k$ -sparse link vector using only  $O(k \log(|V|))$  path measurements.

In the proposed method, Every row of the measurement matrix  $\mathcal{A}$  is constructed from a walk based on the UCS-WN framework. In order to construct each row of  $\mathcal{A}$  (UCS-WN walk), the following steps are iteratively performed: (i) A node is selected relative to  $P_u$  as the current node. (ii) The transition probabilities of current node and its neighbors are calculated. (iii) Then the next node is selected under different conditions. (iv) The last step is repeated  $t$  times which is the length of a walk. More details can be seen in Algorithm 1.

In this algorithm,  $P_v$  is the probability of moving from current node ( $Cn$ ) to node  $v$  if link  $(Cn, v) \in E$ , where  $\forall \{Cn, v\} \in V$ . The basic idea is inspired by the Metropolis-Hasting Markov Chain Monte Carlo (MCMC) technique [11]. In the proposed method, we can avoid biasedness towards high-degree nodes and high-weighted links by selecting a "good start" node for every  $m$  measurements (walks) in lines (5)-(9) of Algorithm 1, and also assigning different probabilities to the neighbors of current node to find the best next node for every walk of length  $t$  in line (13). Choosing a good start node has three phases, that are: scoring, normalizing, and finding probability.

---

**Algorithm 1** Proposed Framework: UCS-WN

**Input:** graph  $\mathcal{G}(V, E)$ ,  $m$ ,  $t$

- 1:  $m$ : number of measurements
- 2:  $t$ : number of measurement steps
- 3:  $P(Cn) = \text{NULL}$  /\*Current-node\*/
- 4: **for**  $i = 1 \rightarrow m$  **do**
- 5:     **foreach** node  $u \in V$  **do**
- 6:          $\text{Score}_u = \frac{\text{deg}(u)}{2|E|} + \frac{w(u)}{\sum_{i \in V} w(i)}$  /\*Scoring\*/
- 7:          $T = \frac{\text{Score}_u}{\sum_{j \in V} \text{Score}_j}$  /\*Normalization\*/
- 8:          $P_u = \frac{1}{|V|-1} \times (1 - T)$  /\*Prob. of Start-node\*/
- 9:     **end for**
- 10:  $Cn = \text{Select a node relative to } P_u$
- 11: **for**  $j = 1 \rightarrow t$  **do**
- 12:     **foreach** node  $v \in \text{nbr}(Cn)$  **do**
- 13:          $P_v = \frac{1}{\text{deg}(Cn)} \times \min(1, \frac{\text{deg}(Cn)}{\text{deg}(v)})$
- 14:     **end for**
- 15:     **if**  $P(Cn) = \text{NULL}$  **then**
- 16:          $P(Cn) = 1 - \sum_{v \in \text{nbr}(Cn)} P_v$
- 17:     **end if**
- 18:     **if**  $\exists v \in \text{nbr}(Cn); P_v \geq P(u = Cn)$  **then**
- 19:         Next-node = Select randomly one of these  $v$
- 20:     **else if**  $\forall v \in \text{nbr}(Cn); P_v < P(u = Cn)$  **then**
- 21:         Next-node = Select  $v$  relative to  $P_v$
- 22:     **else**
- 23:         Next-node = Back track to the previous node
- 24:     **end if**
- 25:     Remove the link between  $Cn$  and Next-node
- 26:      $Cn = \text{Next-node}$
- 27:     **end for**
- 28: **end for**

**Output:** feasible measurement matrix  $\mathcal{A}$

---

As a result, we have a uniform stationary distribution  $\pi_v^{UCSWN} = \frac{1}{|V(\mathcal{G})|}$  that leads to construct an efficient measurement matrix for unbiased compressive sensing with relatively low cost. Note that these measurements (walks) through the connected paths, according to the assumptions in section II, evidence feasibility of the measurement matrix  $\mathcal{A}$  (the first condition). Furthermore, it satisfies sparse recovery with network topological constraints.

#### IV. ANALYSING THE NODE DEGREE BIAS

A basic, but very important property of networks (graphs) is their node degree distribution  $p(d)$ , *i.e.*, the fraction of nodes with degree equal to  $d$ , for all  $d \geq 0$ . We assume that the graph  $\mathcal{G}$  is drawn uniformly at random from the set of all networks with fixed  $p(d)$ . We designate this model by  $\mathbb{R}\mathbb{G}(p(d))$  [12]. In this part, we analyse the observed node degree bias for constructed measurements (walks), in particular, for Random Walk (RW) used in [10] and UCS-WN framework used in this paper, over the random graph  $\mathbb{R}\mathbb{G}(p(d))$ . Typically, the expected observed degree distribution in the raw measurements is different from the original one,  $q(d) \neq p(d)$ , with higher average value  $\langle q(d) \rangle > \langle p(d) \rangle$ . Below, we derive  $q(d)$  as a function of  $p(d)$ . We have summarized some of the notations in Table I.

1) *Random Walk (RW)*: Random walks have been widely studied; see [13] for an excellent survey. RW is also used in compressive sensing, [10], to generate a measurement matrix

**TABLE I:** Notation Summary.

$\mathcal{G} = (V, E)$	graph $\mathcal{G}$ with nodes $V$ and links $E$
$\text{deg}(v)$	degree of node $v$
$p(d) = \frac{1}{ V } \sum_{v \in V} 1_{\{\text{deg}(v)=d\}}$	degree distribution in $\mathcal{G}$
$\langle d \rangle = \langle p(d) \rangle = \sum_d d p(d)$	average node degree in $\mathcal{G}$
$q(d)$	expected observed degree distribution
$\langle q(d) \rangle = \sum_d d q(d)$	expected observed average node degree

in order to recover any  $k$ -sparse link vector. In RW, for any given connected and aperiodic graph, the probability of being at a particular node  $v$  converges to the stationary distribution  $\pi_v^{RW} = \frac{\text{deg}(v)}{2|E(\mathcal{G})|}$ . Therefore, the expected observed degree distribution  $q^{RW}(d)$  is calculated by:

$$\begin{aligned}
 q^{RW}(d) &= \sum_v \pi_v^{RW} \cdot 1_{\{\text{deg}(v)=d\}} \\
 &= \frac{d}{2|E|} p(d) |V| \\
 &= d p(d) \frac{|V|}{2|E|} \\
 &= \frac{d p(d)}{\langle d \rangle}. \tag{4}
 \end{aligned}$$

Consequently, the expected observed average node degree is:

$$\begin{aligned}
 \langle q^{RW}(d) \rangle &= \sum_d d q^{RW}(d) \\
 &= \frac{\sum_d d^2 p(d)}{\langle d \rangle} \\
 &= \frac{\langle d^2 \rangle}{\langle d \rangle}, \tag{5}
 \end{aligned}$$

where  $\langle d^2 \rangle$  denotes the average squared node degree in  $\mathcal{G}$ . According to Eq. 5, we can easily see the biasedness of RW towards higher degree nodes, because the expected observed average node degree is greater than the average node degree in random graph  $\mathbb{R}\mathbb{G}(p(d))$  ( $\langle q^{RW}(d) \rangle > \langle d \rangle$ ).

2) *UCS-WN*: As we showed in Algorithm 1, the transition probability  $P_{u,v}^{UCSWN}$  leads to a uniform stationary distribution  $\pi_v^{UCSWN} = \frac{1}{|V(\mathcal{G})|}$ , and consequently:

$$\begin{aligned}
 q^{UCSWN}(d) &= \sum_v \pi_v^{UCSWN} \cdot 1_{\{\text{deg}(v)=d\}} \\
 &= \frac{1}{|V|} p(d) |V| \\
 &= p(d) \tag{6}
 \end{aligned}$$

and the expected observed average node degree is given by:

$$\begin{aligned}
 \langle q^{UCSWN}(d) \rangle &= \sum_d d q^{UCSWN}(d) \\
 &= \sum_d d p(d) \\
 &= \langle d \rangle \tag{7}
 \end{aligned}$$

Based on Eq. (7), the UCS-WN framework estimates the true mean and it is unbiased for compressive sensing, because the expected observed average node degree is equal to the

average node degree distribution in random graph  $\mathbb{RG}(p(d))$  ( $\langle q^{UCSWN}(d) \rangle = \langle d \rangle$ ).

As we mentioned, RW is clearly biased towards high-degree nodes, and it is not efficient to be used for design of the measurement matrix to infer networks with diverse degree distributions, ranging from constant-degree (e.g., in regular graphs), a distribution concentrated around the average value (e.g., in Erdős-Rényi random graphs, or in well-balanced peer-to-peer networks), to heavily right-skewed distributions (as the case in World Wide Web, unstructured P2P, Internet at the IP and Autonomous System level, and Online Social Networks). Because in these networks, the congested links are mostly located on the links pointing to lower degree nodes. On the contrary, the proposed UCS-WN framework is the proper approach to solve the aforementioned problem. We offer the UCS-WN framework to infer a wide range of complex networks.

## V. THEORETICAL ANALYSIS

In this section, we theoretically prove the last two properties mentioned in section III for the constructed measurement matrix  $\mathcal{A}$  using the UCS-WN framework. First we state a general condition for the correctness of the measurement matrix  $\mathcal{A}$  to indicate how it satisfies the null-space property of  $\mathcal{A}$ . Second we calculate the minimum required number of UCS-WN measurements which guarantees the unique recovery of any up to  $k$ -sparse link vector.

The following theorem shows a sufficient condition for the correctness of the measurement matrix  $\mathcal{A}$ .

**Theorem 1.** *Let  $y_{m \times 1} = \mathcal{A}_{m \times N} x_{N \times 1}$ . Suppose that for every subset of columns (say  $S$ ) of the measurement matrix  $\mathcal{A}$  from the UCS-WN with  $|S| \leq s$  (i.e. the number of columns is no more than  $s$ ), the corresponding sub-matrix  $\mathcal{A}_S$  includes at least one row with a single non-zero element. Then the unique recovery of any link vector  $x$  with regarding to  $y = \mathcal{A}x$  is conceivable, if  $k < \frac{s+1}{2}$ .*

*Proof:* Suppose that we adopt an  $m \times N$  measurement matrix  $\mathcal{A}$  generated by  $m$  independent walks from the UCS-WN framework. see Algorithm 1 for more detailed explanation.

To relax the condition of recovering an arbitrary  $k$ -sparse vector  $x$ , [10] introduces the following lemma which is easy to prove.

**Lemma 1.** *The link vector  $x$  is the unique sparsest vector from  $y = \mathcal{A}x$ , if and only if  $x$  is a  $k$ -sparse link vector with  $k < \min_{w \in \mathcal{N}(\mathcal{A}), w \neq 0} \frac{\|w\|_0}{2}$ , where  $\mathcal{N}(\mathcal{A})$  denotes the null-space of  $\mathcal{A}$  and  $\|w\|_0$  is the number of non-zero elements in  $w$ .*

According to lemma 1, we only need to show that  $\|w\|_0 > s$  for  $\forall w \in \mathcal{N}(\mathcal{A})$ . Suppose that there is a vector in  $\mathcal{N}(\mathcal{A})$  denoted by  $w'$  which  $\|w'\|_0 \leq s$ . The support set of  $w'$  is indexed by  $I \in \{1, 2, \dots, |E(\mathcal{G})|\}$ . Since the corresponding sub-matrix  $\mathcal{A}_I$  from the UCS-WN has at least one row with a single non-zero element,  $\mathcal{A}_I w'_I$  will be non-zero which is unacceptable because  $w'$  is from  $\mathcal{N}(\mathcal{A})$ . So for  $\forall w \in \mathcal{N}(\mathcal{A})$ ,

we definitely have  $\|w\|_0 > s$ . Therefore, based on lemma 1, any  $k$ -sparse link vector  $x$  with  $k < \frac{s+1}{2}$  can be uniquely recovered from the constructed measurement matrix  $\mathcal{A}$  via UCS-WN satisfying  $y = \mathcal{A}x$ . ■

Now, we want to indicate how many path measurements are sufficient to recover any up to  $k$ -sparse link vector. To the rest of this section, we theoretically analyse the special class of networks called uniform networks. Some other types of networks such as Barabási-Albert, Small-World, C.elegans and USTop500 will be experimentally investigated in section VI. An undirected and sufficiently connected graph  $\mathcal{G}$  is called a  $(D, c)$  uniform graph whenever for some constant  $c$  and any node  $v \in V$ ,  $D < \deg(v) < cD$ . Suppose that a walk based on the UCS-WN framework over the graph  $\mathcal{G}$  has a stationary distribution  $\mu$ . The  $\delta$ -mixing time of  $\mathcal{G}$  is defined as the smallest  $t$  such that any walk of length  $t$  starting at a ‘‘good start’’ node in our framework ends up with a distribution  $\mu'$  such that  $\|\mu - \mu'\|_\infty \leq \delta$  where  $\|\cdot\|_\infty$  denotes the supremum norm. We consider  $\delta = \frac{1}{(2cn)^2}$  and define  $T(n)$  as the  $\delta$ -mixing time of  $\mathcal{G}$ .

The following theorem represents the minimum number of measurements constructed from the UCS-WN to insure the unique recovery of any  $k$ -sparse link vector.

**Theorem 2.** *Let  $m = O(k \log(n))$  be the number of required measurements generated from the UCS-WN framework to recover any  $\Theta(k)$ -sparse link vector  $x$ . Then with probability  $1 - o(1)$ , any link set with no more than  $k$  links would have a row in  $\mathcal{A}$  with a single non-zero element. Hence, the constructed measurement matrix  $\mathcal{A}$  from the UCS-WN guarantees recovering any  $\Theta(k)$ -sparse link vector using only these  $m = O(k \log(n))$  UCS-WN measurements.*

*Proof:* According to the previously mentioned assumptions for the uniform graph  $\mathcal{G}$ , we have the following lemma from [14].

**Lemma 2.** *Consider a degree  $D_0 = O(c^2 k T^2(n))$ . Whenever  $D \geq D_0$ , if we set the length of a walk to  $t = O(\frac{nD}{c^3 k T(n)})$ , the following holds. Let  $B$  be a set of links in  $\mathcal{G}$  with  $|B| = (k-1)$ , and suppose that  $e$  is a link which is not included in the set  $B$ . we define  $\pi_{e,B}$  as the probability that a walk traverses  $e$ , but misses all the links of  $B$ . With the stated assumptions,  $\pi_{e,B} = \Omega(\frac{1}{c^4 k T^2(n)})$ .*

Suppose we take  $m$  independent measurements based on the UCS-WN framework over  $\mathcal{G}$  with regarding to graph topological constraints. Then, for every arbitrary link set  $\Phi$  with cardinality  $|\Phi| = k'$  ( $1 \leq k' \leq k$ ), we express the probability that there does not exist any UCS-WN measurement with a single non-zero element in the  $k'$  columns corresponding to the links from  $\Phi$ , by:

$$P = (1 - \pi_\Phi)^m, \quad (8)$$

where  $\pi_\Phi$  is the probability that a measurement traverses only one single link of  $\Phi$  and misses all the others. Therefore, as defined by lemma 2,  $\pi_\Phi = \Omega(\frac{1}{c^4 k T^2(n)}) \times k'$ , since the

events of having a single non-zero element can be divided into  $k'$  disjoint events, where each of them is the event of traversing only one of the  $k'$  distinct links belonging to  $\Phi$  in each measurement.

In order that there are  $\binom{|E(\mathcal{G})|}{k'}$  different ways of choosing  $k'$  links from the links of  $\mathcal{G}$ , the probability of having one such combination of links with  $|\Phi| = k'$  and without any row having a single nonzero element is:

$$P_{k',k} \leq \binom{|E|}{k'} (1 - \pi_\Phi)^m \quad (9)$$

$$\leq \binom{n^2}{k'} \left(1 - \Omega\left(\frac{k'}{c^4 k T^2(n)}\right)\right)^m \quad (10)$$

$$\leq e^{k'(1 + \log(\frac{n^2}{k'})) + m \log(1 - \Omega(\frac{k'}{c^4 k T^2(n)})})} \quad (11)$$

We want  $P_{k',k}$  to be smaller than 1, therefore according to Eq. (11),  $e^{k'(1 + \log(\frac{n^2}{k'})) + m \log(1 - \Omega(\frac{k'}{c^4 k T^2(n)})})}$  should be less than 1. Thus as long as the required number of measurements  $m > \max_{1 \leq k' \leq k} \left( -\frac{k'(1 + \log(\frac{n^2}{k'}))}{\log(1 - \Omega(\frac{k'}{c^4 k T^2(n)})})} \right)$ , with probability  $1 - o(1)$ , the constructed measurement matrix  $\mathcal{A}$  via the UCS-WN framework guarantees recovering up to  $\frac{k}{2}$ -sparse link vectors according to Theorem 1. In fact,  $m = O(k \log(n))$  path measurements from the UCS-WN framework suffice to recover any  $\Theta(k)$ -sparse link vector. ■

## VI. EXPERIMENTAL EVALUATION

In this section, we experimentally evaluate the performance of the proposed approach. We do the evaluation on two synthetic networks: First, the Barabási-Albert (BA) network (Graphs with extreme degree distributions, also known as power-law or scale-free graphs), with 2485 links and each new node is preferentially connected to 5 existing nodes. Second, the Watts-Strogatz (SW) network (Small-world graphs with high clustering and low path lengths), with 2490 links. The rewiring probability is 0.5 and the number of initial closest neighbors is 5. We make these networks weighted by distributing the weights on the links of the network according to zero-mean Gaussian distribution with unit-variance.

In addition, we consider two real-world networks: First, the weighted network of 500 busiest commercial airports in the United States (USTop500), [15], with 500 nodes and 2980 links and the neural weighted network of the *Caenorhabditis elegans* worm (C.elegans), [16], with 306 nodes and 2345 links. According to Eq. (2) in section II, we analysed these weighted networks and found that the links with larger weights are often pointing to the neighbors of larger degree nodes ( $d_{nn,u}^w > d_{nn,u}$ ). For example, in USTop500 with 500 nodes, 64% of these nodes follow this property and in C.elegans with 306 nodes, 67.97% of its nodes conform it.

In each of these test cases, we generated 10 set of walks by using the UCS-WN framework. All sets contain  $\frac{|E|}{5}$  walks with length of  $\frac{|E|}{5}$ . The points in the figures represent the mean value of the tests for all sets. Moreover, we consider the relative error as the measure of performance. We also set the

sparsity (the number of non-zero elements) of the unknown vector to 20% of the number of links in each network.

For the optimization step, we use MATLAB and the SPAMS package [17], and try to minimize the objective function that follows the LASSO model [18],[19]. This function has  $\ell_1$  norm as the regularization term:

$$\min_x \|x\|_1 + \|\mathcal{A}x - y\|_2^2. \quad (12)$$

In all of the test cases, we compare our method with the work in [10] which we call RW in short. We consider a constant positive value for the non-zero elements and spread them in the unknown vector in the following way:

- 1) Set  $N = \text{Sort the nodes by their degrees in ascending order.}$
- 2) Set  $l = \text{NULL.}$
- 3) Traverse  $N$  in ascending order and for each node, add the connecting links to  $l$ .
- 4) Set  $s = \text{number of non-zero elements for the network (from the sparsity percentage).}$
- 5) Set the value of the corresponding element of the first  $s$  links in  $l$  to a non-zero constant value.

Fig. 1 shows the performance evaluation of our approach compared to RW, in terms of the recovery error. As it is shown, in all networks our approach outperforms RW: In synthetic networks by 23.99 and 12.92 percent improvement in BA and SW, respectively and in real world networks by 14.09 and 19.62 percent improvement in C.elegans, and USTop500 network, respectively.

Meanwhile, in Fig. 2, it is shown that our approach achieves this by preserving the measurements weight much lower than RW. There are 22.98 and 19.46 percent improvements in measurements total weight in BA and SW, respectively. There are also 51.79 and 48.75 percent improvements in C.elegans and USTop500, respectively.

As the final result, it can be seen that the UCS-WN framework fulfils the requirements that we mentioned for an unbiased compressive sensing approach over general type of weighted networks. Therefore, the UCS-WN is an accurate solution to efficiently recover any  $k$ -sparse link vectors with relatively low cost in weighted networks.

## VII. CONCLUSION

To the best of our knowledge, this is the first paper to introduce a novel and general framework, called UCS-WN, for the problem of recovering sparse vectors representing certain parameters of the links in weighted networks. We used this framework in the context of compressive sensing to construct a feasible measurement matrix under network topological constraints. Our theoretical analysis showed that only  $O(k \log(n))$  UCS-WN measurements are sufficient to uniquely recover any  $k$ -sparse link vector with no more than  $k$  non-zero elements. In addition, we experimentally evaluated the performance of our approach for some networks and simulation results demonstrate that this framework can be employed to efficiently infer weighted networks.

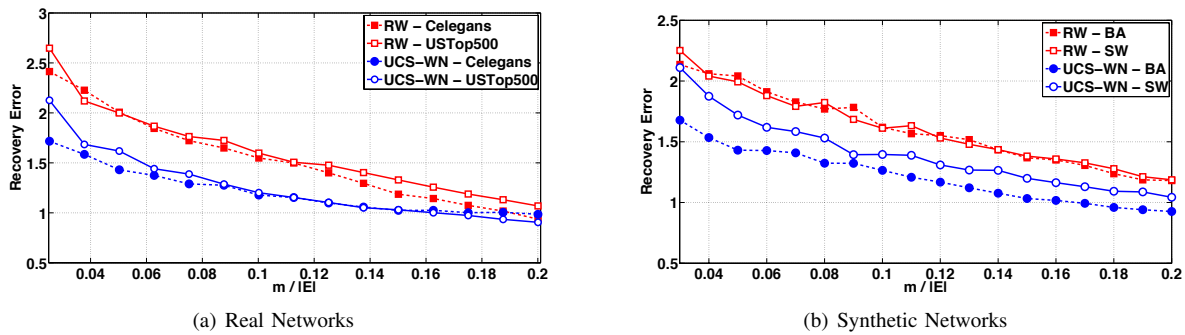


Fig. 1: Recovery error for comparison of the UCS-WN and RW

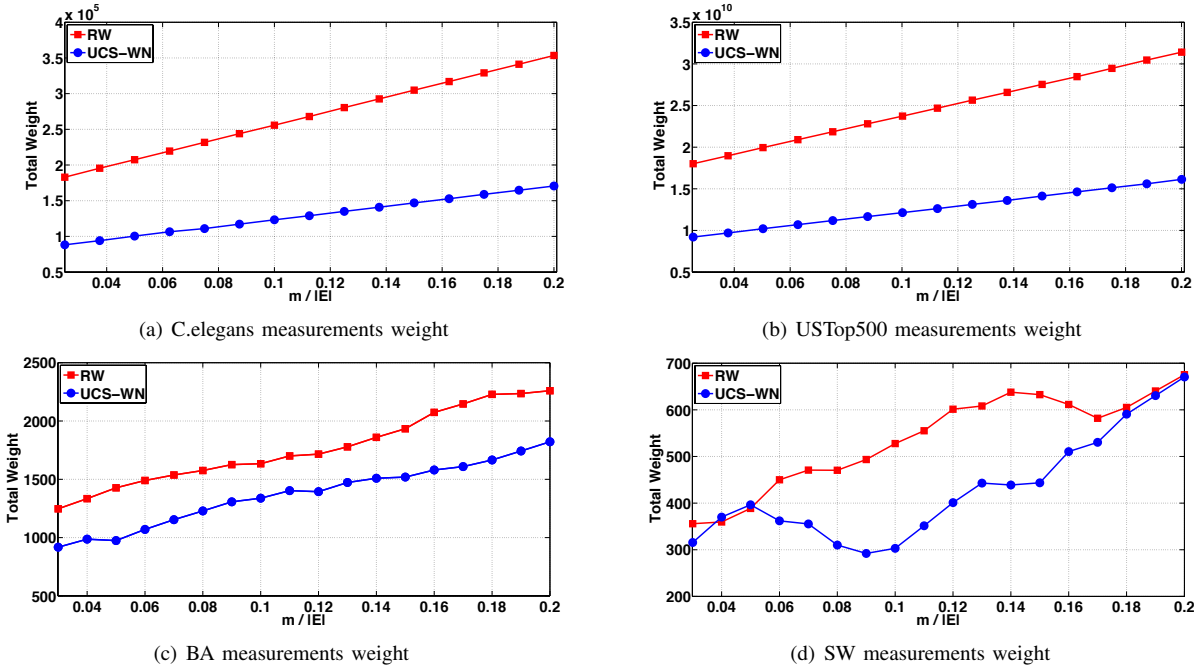


Fig. 2: Measurements weight for comparison of the UCS-WN and RW

## REFERENCES

- [1] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss, "Combining geometry and combinatorics: a unified approach to sparse signal recovery," in *46th Annual Allerton Conference on Communication, Control, and Computing*, Sep. 2008, pp. 798–805.
- [2] E. J. Cands and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
- [3] E. J. Cands, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [4] J. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
- [5] J. Donoho and J. Tanner, "Sparse nonnegative solution of underdetermined linear equations by linear programming," *Natl. Acad. Sci. U.S.A.*, vol. 102, no. 27, pp. 9446–9451, Mar. 2005.
- [6] M. Coates, Y. Pointurier, and M. Rabbat, "Compressed network monitoring for ip and all-optical networks," in *ACM SIGCOMM IMC*, Oct. 2007, pp. 241–252.
- [7] M. Firooz and S. Roy, "Link delay estimation via expander graphs," *arxiv:1106.0941*, 2011.
- [8] M. Wang, W. Xu, E. Mallada, and A. Tang, "Sparse recovery with graph constraints: Fundamental limits and measurement construction," in *IEEE INFOCOM*, Mar. 2012, pp. 1871–1879.
- [9] J. Haupt, W. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Processing Magazine*, vol. 52, no. 2, pp. 92–101, Mar. 2008.
- [10] W. Xu, E. Mallada, and A. Tang, "Compressive sensing over graphs," in *IEEE INFOCOM*, Apr. 2011, pp. 2087–2095.
- [11] W. K. Hastings, "Monte carlo sampling methods using markov chains and their applications," *Biometrika*, vol. 57, no. 1, pp. 97–109, 1970.
- [12] M. Kurant, A. Markopoulou, and P. Thiran, "On the bias of BFS," in *International Teletraffic Congress (ITC)*, Sep. 2010, pp. 1–8.
- [13] L. Lovasz, "Random walks on graphs: A survey," *Combinatorics, Paul Erdos is Eighty*, vol. 2, no. 1, pp. 1–46, Jan. 1993.
- [14] M. Cheraghchi, A. Karbasi, S. Mohajer, and V. Saligrama, "Graph constrained group testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 248–262, Jan. 2012.
- [15] V. Colizza, R. Pastor-Satorras, and A. Vespignani, "Reaction-diffusion processes and metapopulation models in heterogeneous networks," *Nature Physics*, vol. 3, pp. 276–282, 2007.
- [16] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, 1998.
- [17] Sparse modeling software (spam). [Online]. Available: <http://spams-devel.gforge.inria.fr/index.html>
- [18] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society B*, vol. 58, pp. 267–288, 1994.
- [19] E. J. Candes, M. Rudelson, T. Tao, and R. Vershynin, "Error correction via linear programming," in *46th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, Oct. 2005, pp. 668–681.