Topology Maps and Distance-Free Localization from Partial Virtual Coordinates for IoT Networks

Anura P. Jayasumana
Electrical & Computer Engineering
Colorado State University
Fort Collins, CO 80525

Randy Paffenroth
Mathematical Sciences
Data Science Program
Worcester Polytechnic Institute
Worcester, MA 01609

Sridhar Ramasamy
Electrical & Computer Engineering
Colorado State University
Fort Collins, CO 80525

***Abstract*—While physical coordinates are useful for IoT and sensor network operations, physical localization is not a viable option for large-scale networks of simple devices in complex or harsh environments. Topology Preserving Maps (TPM) extracted from anchor-based Virtual Coordinates (VCs) are an attractive localization free alternative for physical maps. We present an approach, based on the theory of low-rank matrix completion, to extract TPMs with only partial information about VCs. Evaluation using 2D and 3D networks with random anchors shows that accurate TPMs can be obtained even when up to 40% to 60% of random coordinates are missing. Coordinate generation and communication cost thus may be reduced significantly. TPM generation can now also be based on a different VC system as long as it characterizes each node with distances to a small set of random nodes instead of a global set of anchors.**

Keywords: Localization, virtual coordinates, topology preserving maps, IoT, sensor networks

I. Introduction

Emerging IoT infrastructure will include large-scale networks of inexpensive wireless devices such as smart RFIDs and self-powered sensor nodes. These networks need to rely on geometric properties of the network layout for functions such as ensuring area or volume coverage, topology control, sensing and routing. The traditional way to capture the geometric features of networks embedded in 2D and 3D spaces is to characterize each node by its physical, aka geometric, coordinates. While physical localization, using techniques such as RSSI and time delay measurements, have been proposed to obtain geographical coordinate information, in practice these localization methods are hampered or made ineffective by issues such as multipath interference, reflections, shadowing and clock synchronization, thus restricting their use only to very limited cases [11]. For instance, RSSI does not work in environments where there is no line-of-sight, or where there are reflecting surfaces, which is typically the case for many environmental monitoring applications as well as with many 3D deployments. Estimating distance using time delay is affected by issues such as clock drift and accuracy of clock synchronization. GPS is not applicable indoors, and even in cases where it is applicable it adds a significant cost to the individual nodes, preventing its use in large-scale networks of inexpensive nodes. Consequently, much research has focused on localization-free techniques for carrying out networking functions in sensor networks embedded in physical spaces.

Virtual-Coordinate (VC) based characterization of nodes provide an alternative to physical localization for many networking algorithms, but without the need for node localization. Prominent among these are algorithms using anchor-based VC Systems (VCS), in which each node is represented by a VC vector corresponding to its minimum hop-distances to a set of anchor nodes [5, 6, 20, 24, 26]. A VCS therefore is an M-dimensional abstraction of the network connectivity, where M is the number of anchors. A key question therefore is the number (M) and the placement of anchors. If an adequate number of anchors are not appropriately deployed, it may also cause the network to suffer from identical coordinates and local minima [9], resulting in logical/virtual voids. Too many anchors increases the cost of VCS generation as well as the address lengths. The difficulty of determining the optimal anchor set is compounded by the fact that the number of anchors and its optimal placement are dependent on each other. Current approaches select anchors randomly or by selecting nodes with specific properties, e.g., by selecting all the perimeter nodes [20], or extreme nodes on internal and external boundaries as anchors [8]. The anchor coordinates are generated by one flooding message from each of the anchors, with a counter that gets incremented as it is relayed by a node. Random anchor selection typically requires a significantly larger number of anchor nodes compared to algorithmic approaches. The time and energy consumed for coordinate generation, lengths of address fields, and memory requirement for algorithm are few examples of the cost associated with large number of anchors. This is an added motivation for us to seek a solution that does not rely on the full anchor-based VC set.

Success of VC based techniques such as routing is due to the fact that the connectivity based VCs implicitly overcomes the features such as concave voids present in geographic domain. However, VCs are not orthogonal to each other, resulting in gross errors in distance estimations. Furthermore, the directional information such as left/right, north/south are completely lost, and geometric and layout features of the networks such as voids and shapes are completely invisible in the VC domain. These are major disadvantages of VCS, which makes them a less attractive solution for large-scale networks.

Topology Preserving Maps (TPMs) derived from VCS are maps that are nearly homeomorphic to physical maps [11]. In case of a 2-D (3-D) sensor network with M anchors, VCS is a mapping from the 2-dimensional (3-dimensional) network layout to an M-dimensional space. TPMs recover a 2-D (3-D) projection from this M-dimensional representation. A TPM is a rotated and distorted version of the physical node layout in such a way that the distortion accounts for network connectivity. It preserves the main features such as boundaries and shapes of voids of the network. Topology Coordinates (TCs) refer to the coordinates of a node on a TPM, and they can serve as an effective alternative for physical coordinates in geographic algorithms. TPMs preserve much of the geometric and neighborhood relationships present in the physical domain, but without the need for distance measurements. TC based schemes have demonstrated better or comparable performance compared to their geographic coordinate based counterparts in 2D and 3D networks [10, 12, 16], and are also applicable in mobile networks [23].

While the algorithms that are based on VCS and its derivatives such as TCs [11] provide a viable, competitive and robust alternative to traditional geographic coordinate based methods, these techniques have so far required the complete set of VCs in order to extract the geometric information. The first contribution of the paper is the demonstration of recovery of TPMs and geometric properties of a network without the need for complete knowledge of the virtual coordinates. Missing VCs are recovered by the use of principles from low-rank matrix completion. We use three networks, with 500 to 1600 nodes, representing deployments in 2-D and 3D spaces, with different shapes and voids to demonstrate the recovery of TPMs. The second and major contribution of this paper follows from the above result and the observation that a random subset of VCs really is a set of random elements from the distance matrix. Therefore TPM and TC derivation no longer need to be tied to anchor-based VCs, but it can be based on any measurement technique that allows the observation of an adequate number of elements from the distance matrix. The impact is that any number of strategies now can be used to sample minimum hop distances between random node pairs in the network to obtain TPMs.

Section II presents the theory and the algorithm for VC completion. Section III contains results and evaluation, followed by the conclusion in Section IV.

II. Theory and Algorithm

Consider a network with N nodes, which we represent by an undirected graph G defined by $G = \{V, E\}$. V is the set of nodes or vertices and E is a set of unordered pairs of nodes or edges corresponding to communication links. For mathematical analysis, the graph is often represented by an *adjacency matrix* A [13]. As networks of interest to us are connected, instead of A, we consider a *hop distance* matrix $D \in \mathbb{Z}^{N\times N}$, where

$$\begin{aligned} D_{ij} = &\text{ the length of the shortest path (in hops)} \\ &\text{ from node } i \text{ to node } j. \end{aligned}$$

A key question then is "what information about the network geometry and topology and encoded in such hop distance matrices?" In particular, our work is aimed at networks where,

- it is not possible to obtain the complete information, i.e., D, which may be due to complexities associated with measurement or the size of the graph, or
- the size of the graph is so large that it is only feasible to manipulate a small subset of information appropriately sampled.

This is, in general, the case for large-scale wireless sensor networks (WSNs) and emerging IoT subnets. Our research is aimed at the case where it is not feasible to have this complete information, i.e, when a significant fraction, and possibly even a vast majority of elements, of these matrices are not available.

To wit, consider a sensor network with N nodes and M anchors. With an anchor-based Virtual Coordinate System (VCS), each node is characterized by a VC vector of length M. Let $P \in \mathbb{N}^{N\times M}$ be the matrix containing the VCs of all the nodes, e.g., the i-th row corresponds to the $\mathbb{N}^{1\times M}$ VC vector of the i-th node, and j-th column corresponds to the j-th virtual ordinate of all the nodes in the network with respect to j-th anchor. Therefore, we consider

$$P = \begin{bmatrix} h_{n_1A_1} & \dots & h_{n_1A_M} \\ \vdots & \ddots & \vdots \\ h_{n_NA_1} & \dots & h_{n_NA_M} \end{bmatrix} \tag{1}$$

where $h_{n_1A_1}$ is the hop distance from node n_i to anchor A_j. P is precisely a subset of the full hop distance matrix D derived by selecting just a few anchor nodes and constructing P from the columns corresponding to those anchor nodes. For sensor network and IoT applications, it is generally desirable to have only a small subset of nodes as anchors, i.e., $M \ll N$.

A. Low Rank Matrices

As can be inferred from prior work on Topology Preserving Map (TPM) generation [11], and as we further demonstrate here, hop distance matrices D for many interesting and realistic networks are, somewhat surprisingly, *approximately low-rank*. It is this empirical observation that inspires our work.

Note, there is a rich literature related to techniques for analyzing such low rank structures, including Spectral Graph Theory [2, 7, 17, 25] that relates the properties of a graph to the characteristic polynomial, eigenvalues, and eigenvectors of these matrices.

A widely used tool for analyzing low rank structure in matrices is the SVD [14] and the closely related idea of Principal Component Analysis (PCA) [1]. Given a matrix $P \in \mathbb{R}^{N\times M}$ one can write P as

$$P = U\Sigma V^T$$

where $U \in \mathbb{R}^{N\times \min(N,M)}$, $\Sigma \in \mathbb{R}^{\min(N,M)\times\min(N,M)}$, and $V \in \mathbb{R}^{M\times\min(N,M)}$. U and V are *unitary*) and Σ is diagonal. Finally, the diagonal entries of Σ are called the singular values of P and the rank of A is precisely the number of non-zero singular values. Accordingly, given a low rank matrix P, the entire matrix can be approximated by using just a few columns of U and V, and just a few entries of Σ.

B. Topology Coordinates and Topology Preserving Maps

The mathematical foundation of our previous work in TPM generation from a VCS follows from the above formulation [11]. Consider the principle components of P given by,

$$P_{SVD} = \Sigma V^T$$

In TC domain, each node in a 3D network is characterized by a triple of Cartesian coordinates $(x_T(i), y_T(i), z_T(i))$. Let $[X_T, Y_T, Z_T]$ be the matrix of TCs for the entire set of nodes, i.e., the ith row is the TCs of node i. Then from [11],

$$[X_T, Y_T, Z_T] = [P^{(2)}_{SVD}, P^{(3)}_{SVD}, P^{(4)}_{SVD}] \tag{2}$$

where, $P^{(j)}_{SVD}$ is the j-th column of P_{SVD}. In case of 2D networks Z_T is not used.

C. Matrix Completion

Prior work on TPM generation is based on the case where entire columns are taken from D and used to construct P. However, in the current work, we consider the more interesting, and practically important case, where each anchor node only has a *partial* set of measurements to the rest of the network. Accordingly, some entries in P are *not observed* and the matrix P is therefore incomplete. Predicting the unobserved entries in P can be phrased as a low-rank *matrix completion* problem. In particular, we have leveraged modern ideas in low-rank *matrix completion* [3, 4, 19, 21, 22]

The idea of such methods can be phrased as the following optimization problem

$$L = \arg\min_{L_0} \rho(L_0), \tag{3}$$
$$\text{s.t. } P_\Omega(P) = P_\Omega(L_0)$$

where ρ is the rank operator and Ω is the set of observed entries in P, so that P_Ω is a projection operator onto this set. In other words, we seek to find a matrix L such that the rank of L (denoted $\rho(L)$) is minimized while enforcing the constraint that the matrix we construct matches our observed hop distances $P_\Omega(P)$, but L is free to take on any values it likes outside of Ω to minimize its rank.

Unfortunately, as stated, (3) is an NP-hard optimization problem, and can only be solved for small networks. Recent results [3, 4, 19, 21, 22] allow, under mild assumptions, for the NP-hard optimization in (3) to be recast as a convex optimization problem

$$L = \arg\min_{L_0} \|L_0\|_*, \tag{4}$$
$$\text{s.t. } P_\Omega(M) = P_\Omega(L_0)$$

where $\|L_0\|_*$ is the nuclear norm, or sum of the singular values of L_0. The optimization problem in (4) is convex and can easily be solved for millions of nodes on commodity computing hardware using splitting techniques and iterative matrix decomposition algorithms [19, 21, 22].

While we study *hop* distance matrices here, we would be remiss not to observe the deep connection between the results herein the rich theory and many practical algorithms for analyzing distance matrices that arise in other contexts. In particular, there is a substantial literature on *Euclidean Distance Matrices* (EDM) [15, 18]. While a detailed discussion of connection between VCs and EDMs is beyond the scope of this document, we merely observe that there is a strong connection between the two domains, which we plan to more fully explore in future work.

III. Results

This section evaluates the effectiveness of the proposed approach in constructing topology preserving maps from only a fraction of VCs, compared to that required by existing TPM techniques, and the accuracy of such TPMs. Three representative networks are used for the evaluation. First is an odd shaped 2-D network with 550 nodes the physical layout of which is shown in Fig. 1(a) [11]. The second is a 2-D circular network with multiple circular voids of 496 nodes as shown in Fig. 2(a) [11]. Third is the 3-D network shown in Fig. 3(a) consisting of 1640 nodes, which occupies a cube shaped volume with a hollow region in the shape of an hourglass devoid of nodes. For a given network we generate VCs and randomly drop a certain percentage of VCs. Then we complete the missing entries by solving Eqn.(4) by using the matrix completion procedure outlined in Section II.C. Equation (2) is then used to solve for the TCs, which are used to generate the corresponding TPM.

First we examine the rank of the VC sets. Twenty random anchors were selected in each case, i.e., $M = 20$. The singular values of the full VC matrices of the three networks are shown in Fig.4, which indicates that the VC matrix is relatively low rank, thus lending support to the use of low-rank matrix completion techniques to recover missing information, albeit with some error that increases with the number of missing coordinates. Our interest is in the recovery of layout information and geometric relationships such as the general shapes of boundaries and voids of the network, and node neighborhood preservation, and thus the question is whether such information is preserved and can be extracted from this partial information.

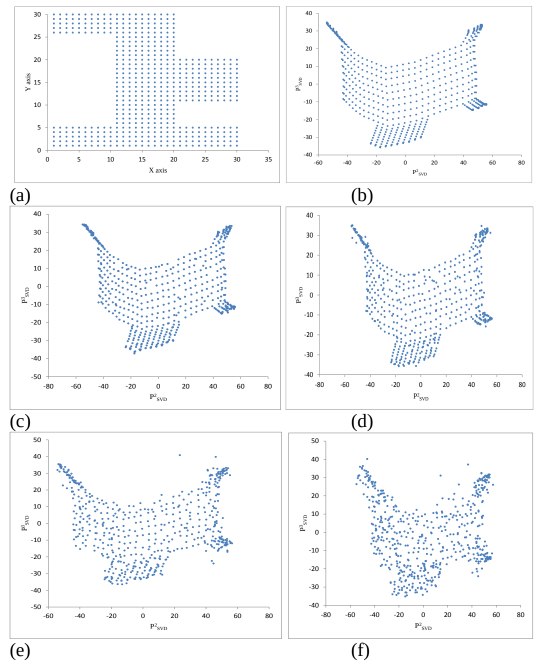

Fig. 1: Odd-shaped network: (a) Physical layout, and (b) TPM recovered from full set of VCs with 20 random anchors; Recovered TPM with (c) 10%, (d) 20%, (e) 40%, and (f) 60% of random coordinates missing.

TPMs extracted using the full set of VCs following [11] are shown in Fig. 1(b), 2(b) and Fig. 3(b) respectively for the three networks. It is important to note that even the full set of VCs corresponding to 20 anchors contains only $3.6\%, 4\%$, and 1.2% of elements in the corresponding distance matrix. Next we discard randomly 10%, 20%, 40% and 60% respectively of these VCs. The TPMs recovered using low-rank matrix completion followed by TPM extraction are shown in Fig. 1(c)-(f) for the odd network, in Fig. 2(c)-(f) for the circular network, and Fig. 3(c)-(f) for the 3-D network. The results indicate that layout maps of networks are obtainable with only a fraction of virtual coordinates.

To quantify the error introduced to the TPM due to missing VCs, we define the mean error (E) as follows:

$$E = \left[\sum_{i,j=1}^{N} |d_{ij}(f) - d_{ij}(0)| \right] / \left[\sum_{i,j=1}^{N} d_{ij}(0) \right] \quad (5)$$

where, $d_{ij}(f)$ refers to the distance between nodes i and j when f fraction of random anchor coordinates are missing. The variation of the mean percentage error with percentage of missing VCs for the three networks

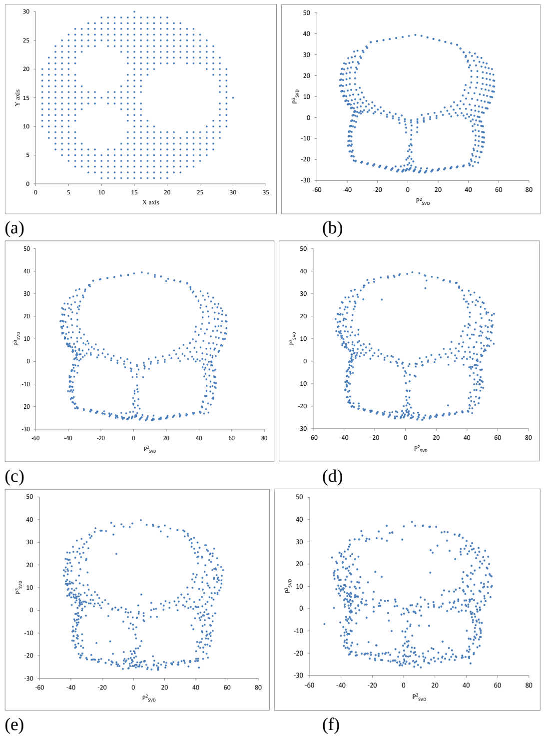

Fig. 2: Circular network: (a) Physical layout, and (b) TPM recovered from full set of VCs with 20 random anchors; Recovered TPM with (c) 10%, (d) 20%, (e) 40%, and (f) 60% of random coordinates missing.

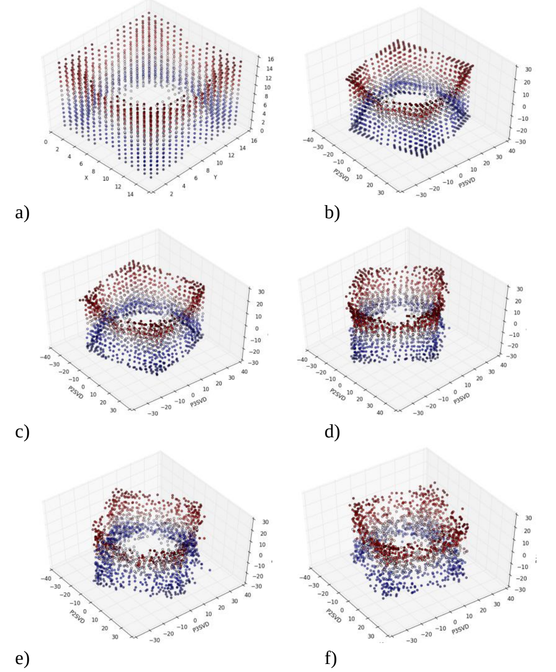

Fig. 3: Cube with hourglass shaped void: (a) Physical layout, and (b) TPM recovered from full set of VCs with 20 random anchors; Recovered TPM with (c) 10%, (d) 20%, (e) 40%, and (f) 60% of random coordinates missing.

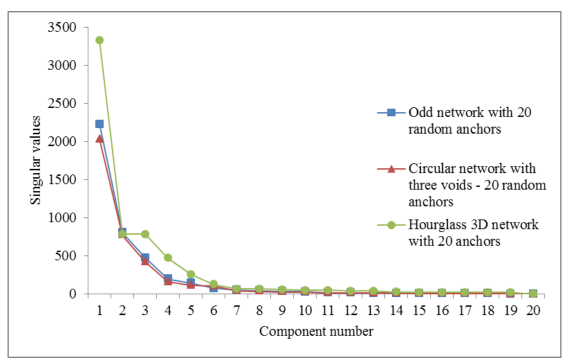

Fig. 4: Singular values of VC matrix for Circular Network, Odd-Shaped Network, and the 3-D network with void indicating the low-rankness of VCS data

are shown in Fig. 5. It is important to note that even when mean error is high, much of the local neighborhood and shape information are preserved. It is also noteworthy that the 3D network with 1640 nodes is less sensitive to loss of VC information compared to the 2D networks with 500 nodes each. We also note that change of scale between two maps contribute to an increase in error, but still preserves the neighborhood relationships which is what is important from the point of TPMs.

IV. Conclusion

This paper addresses the question of recovering geometric and physical layout information in large-scale networks such as wireless sensor networks and IoT deployed in 2-D and 3-D spaces. The approach starts with anchor-based VCs. Unlike prior approaches we require only a fraction of the virtual coordinate set. A technique based on the theory of low-rank matrix completion was proposed that reconstructs missing VCs to recover topology coordinates and layout maps. Results show that topology maps with reasonable accuracy can be recovered even when up to 60% of the VCs are missing.

The VC matrix is a random subset of columns, corresponding to selected anchors, of the hop-based distance matrix D of the network. TPM was generated using a subset of those elements. The fact that we recovered the network layout from a small subset of elements of D points to the existence of other virtual coordinate

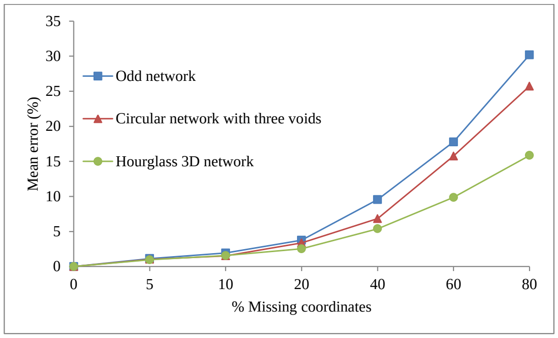

Fig. 5: Mean error vs. the percentage of missing virtual coordinates

systems, beyond just anchor based VCs, from which the physical and geometrical properties of network layout can be obtained. An example is a coordinate system where a node is characterized by its distances to its own randomly selected set of nodes, but not a global set of anchors. This is an exciting possibility, especially for mapping large-scale complex networks while keeping virtual coordinate generation manageable. The results presented here allows the reduction of cost (communication, power, etc.) in cases where a centralized or a mobile node with appropriate computation capability carries out network mapping or topology coordinate generation. Importantly, it also opens the possibility of using topology coordinate based techniques for networks involving 'soft-state' coordinate systems, where some of the coordinate values may be allowed to expire, thus allowing for more resilient network operations.

References

[1] C.M. Bishop. *Pattern recognition and machine learning*, volume 4. Springer New York., 2006.

[2] K. Börner, S. Sanyal, and A. Vespignani. Network science. *Annual Review of Information Science & Technology*, 41:537–607, 2007.

[3] E.J. Candes and Y. Plan. Matrix completion with noise. *Proceedings of the IEEE*, 98(6):11, 2009.

[4] E.J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, 2009.

[5] Q. Cao and T Abdelzaher. Scalable logical coordinates framework for routing in wireless sensor networks. In *ACM Transactions on Sensor Networks,Vol. 2,pp. 557-593, Nov.*, 2006.

[6] A. Caruso, S. Chessa, S. De, and A. Urpi. GPS free coordinate assignment and routing in wireless sensor networks. In *Proc. INFOCOM*, 2005.

[7] F.R.K. Chung. Spectral Graph Theory. *ACM SIGACT News*, 30:14, 1999.

[8] D.C. Dhanapala and A.P. Jayasumana. Anchor placement and topology preserving maps in wsns - a directional virtual coordinate based approach. In *Proc. IEEE Conf. Local Computer Networks (LCN)*, Oct, 2011.

[9] D.C. Dhanapala and A.P. Jayasumana. Convex subspace routing (CSR): Routing via anchor-based convex virtual subspaces in sensor networks. *Computer Communications. Jan. 2011*, 2011.

[10] D.C. Dhanapala and A.P. Jayasumana. Geo-logical routing in wireless sensor networks. In *Proc. IEEE Conf. Sensor, Mesh and Ad Hoc Communications and Networks (SECON)*, 2011.

[11] D.C Dhanapala and A.P. Jayasumana. Topology Preserving MapsExtracting Layout Maps of Wireless Sensor Networks From Virtual Coordinates. *IEEE/ACM Transaction on Networking*, 22(3):784–797, 2014.

[12] D.C. Dhanapala, S. Mehta, and A.P. Jayasumana. On boundary detection of 2-d and 3-d wireless sensor networks. In *Proc. IEEE Global Communications Conference (GLOBECOM)*, 2011. Dec 2011.

[13] Reinhard Diestel. *Graph Theory*, volume 173. Springer Verlag, 2005.

[14] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218, 1936.

[15] J.C. Gower. Euclidean distance geometry. *Mathematical Scientist*, 7(1):1–14, 1982.

[16] Y. Jiang and A. P. Jayasumana. Anchor selection and geo-logical routing in 3d wireless sensor networks,. In *Proc. 39th IEEE Local Computer Networking Conference (LCN)*, 2014.

[17] R. Kannan and S. Vempala. Spectra of graphs. *Foundations and Trends in Theoretical Computer Science*, 4(3-4):132–288, 2008.

[18] N Krislock and H Wolkowicz. *Euclidean distance matrices and applications*. 2012.

[19] Z. Lin, M. Chen, and Y. Ma. The Augmented Lagrange Multiplier Method for Exact Recovery of Corrupted Low-Rank Matrices. *arXiv:1009.5055*, page 23, 2013.

[20] K. Liu and N. Abu-Ghazaleh. Aligned virtual coordinates for greedy routing in wsns,. In *Proc. IEEE Int. Conf. on Mobile Adhoc and Sensor Systems*, 2006. 2006.

[21] R. C. Paffenroth, P. Du Toit, R. Nong, L. L. Scharf, A. P. Jayasumana, and V. Bandara. Space-time signal processing for distributed pattern detection in sensor networks. *IEEE Journal of Selected Topics in Signal Processing*, 7(1):38–49, 2013.

[22] R.C. Paffenroth, R. Nong, and P. Du Toit. On covariance structure in noisy, big data. *SPIE Optical Engineering+ Applications*, pages 88570E–88570E, 2013.

[23] K. R. Rahmatizadeh, S. A. Khan, A. P. Jayasumana, D. Turgut, and L. Boloni. Routing towards a mobile sink using virtual coordinates in a wireless sensor network,. In *Proc. IEEE Int. Conf. on Communications (ICC)*, 2014.

[24] A. Rao, S. Ratnasamy, C. Papadimitriou, S. Shenker, and I. Stoica. Geographic routing without location information. In *Proc. 9th Int. Conf. on Mobile Computing and Networking*, 2003.

[25] D. A. Spielman. Spectral graph theory. *in Combinatorial Scientific Computing*, page Chapter 6, 2012. CRC Press.

[26] M. J. Tsai, H. Y. Yang, and W. Q. Huang. Axis based virtual coordinate assignment protocol and delivery guaranteed routing protocol in wireless sensor networks. In *Proc. INFOCOM*, 2007.