

A Clusterized WLS Localization Algorithm for Large Scale WSNs

Giuseppe Destino, Davide Macagnano, Giuseppe Abreu
Center for Wireless Communications
University of Oulu, P.O. box 4500, 90014-Oulu, Finland
E-mails: [destino, macagnan, giuseppe]@ee.oulu.fi

Abstract—We present a low-complexity, accurate and robust localization algorithm suitable for large scale Wireless Sensor Networks (WSNs). The algorithm is a clusterized version of the Weighted Least-Squares (WLS) localization technique which we recently introduced in [1]. The WLS algorithm is a low-complexity localization technique that owes its high-accuracy to the ability to complete and approximate the Euclidean Distance Matrix (EDM) samples constructed from incomplete and error-disturbed ranging information collected from the sensors. The performance of this algorithm is, however, known to decrease sharply [1] when the completeness is not sufficient to ensure the uniqueness of the network (Graph) realization [2]. The clusterization procedure is based on recent Graph-theoretical results [3] showing that the elements of the second smallest eigenvector of the Laplacian matrix of a Graph are strongly correlated with the proximity of its vertices. This Graph-spectrum analytical tool is utilized here to separate the network into sub-groups that satisfy the completeness constraints of the WLS technique. The resulting clusterization procedure, which relies solely on connectivity information, allows the WLS to be applied into smaller parts of the network, each exhibiting a prescribed completeness level, leading simultaneously to a significant improvement in accuracy and to a reduction in the computational demand of the WLS optimization.

I. INTRODUCTION

Large-scale Wireless Sensor Networks (WNSs) are typically composed by power-limited, unevenly spread, low-cost devices. The localization problem in such WSNs is challenged by two conflicting conditions, namely, the requirement for low-complexity techniques and the frequent occurrence of errors and erasures on ranging measurements.

In [4], a solution for the source localization problem from imperfect and incomplete ranging was presented, which made use of the Semi-definite Programming-based Euclidean Distance Matrix completion technique proposed in [5]. Later, we introduced a low-complexity alternative to that technique [1], where the EDM completion/approximation problem was reformulated as a Weighted Least-Squares optimization, which benefits from closed-form gradient and Hessian functions similar to those derived in [6] to provide for highly accurate localization at low computational costs.

While both the WLS and the SDP-based localization algorithms are effective in mitigating the incompleteness of EDM samples, both these techniques are ultimately limited by the conditions for the unique realization of the Graph corresponding to the network [2].

In fact, even the WLS algorithm can become prohibitively complex when the location of a large number of sensors is to be estimated from an even larger number of ranging measurements. Indeed, if N and η denote the number of nodes and the dimension of the Euclidean space, respectively, then the order of the complexity is estimated to be $\mathcal{O}(N\eta)^3$, when a Levenberg-Marquardt algorithm [7], specially designed for nonlinear Least Squares problems, is used. Due to the reasons outlined above, clusterization is imperative in order to further reduce the computational complexity of the WLS algorithm to the level required for coping with large-scale WSNs.

In so doing, however, it is desirable not to cause deterioration on the accuracy of the localization algorithms. Adding to these conflicting requirements for low complexity and high accuracy are further challenges inherent to the clusterization technique in itself. Specifically, it is desirable that the clusterization procedure be independent of error-perturbed quantities such as ranging and sensor location estimates, so as to prevent the induction of errors in the localization algorithm and vice-versa.

In this work, we present a clusterized version of the WLS sensor localization algorithm that satisfies the latter requirement in which it relies solely on connectivity information. The clusterization technique is based on recent Graph-theoretical results [3] which establish the relationship between the elements of the second smallest eigenvector of the Laplacian matrix of a Graph and the proximity of its vertices. This Graph-spectrum analytical tool is utilized here to separate the network into sub-groups, hereafter referred to as *clusters*, that satisfy the completeness constraints of the WLS technique.

The remainder of the paper is structured as follows. In section II, terminology and mathematical notations, used throughout the paper, are defined. In section III the proposed algorithm is described in detail. The performance of the technique is studied in section IV and a conclusion is given in section V.

II. PRELIMINARIES

We begin with some notations. Let a network be represented by a weighted graph $G_{\eta,N}(V, E, W)$, with the vertices $V = \{v_1, \dots, v_N\}$ defined in the Euclidean space of dimension η , with edges $e_{ij} \in E$ and weights $w_{ij} \in W$ where the subscript ij indicates the link between the vertices i and j . Edges represent communication links, while weights correspond to the squared Euclidean distance between devices.

The matrix containing the measured distances will be referred to as a *Euclidean Distance Matrix* (EDM) sample, and will be denoted by $\tilde{\mathbf{D}}$. As defined in [4], a sample $\tilde{\mathbf{D}}$ is said to be an *imperfect* and *incomplete* EDM. Let the matrix $\mathbf{X} \in \mathbb{R}^{N \times \eta}$ be a realization of the graph where the i th-row vector indicates the coordinates of the i th-node.

Let \mathbf{C} be the binary-valued adjacent matrix of the graph $G_{\eta,N}(V, E, W)$, with $c_{ij} = 1$ and $c_{ij} = 0$ indicating the existence and the non-existence of a link between v_i and v_j . Hereafter, the matrix \mathbf{C} will be referred to as the *connectivity matrix*. The connectivity of a network will be measured in terms of its *completeness*, which is defined as:

$$\varrho \triangleq \frac{2}{N(N-1)} \cdot \sum_i^N \sum_{j=i+1}^N c_{ij}. \quad (1)$$

We shall also make use of the notion of Laplacian matrix \mathbf{L} of the graph $G_{\eta,N}(V, E, W)$, which is defined as,

$$\mathbf{L} = \mathbf{M} - \mathbf{C}, \quad (2)$$

where \mathbf{M} is a diagonal matrix whose i -th diagonal element m_{ii} is equal to the number of connections per corresponding node, hereafter referred to as *node-degree*, which can be computed directly from \mathbf{C} using

$$\mathbf{M} = \text{diag}([1, \dots, 1] \cdot \mathbf{C}). \quad (3)$$

While the focus of this work is on robust clusterization methods, the general localization problem underlining the discussion, is formulated as a non linear weighted least square optimization [1]:

$$\min_{\hat{\mathbf{X}} \in \mathbb{R}^{N \times n}} \left\| \mathcal{H}(\{\tilde{\mathbf{D}}\}) \circ \left(\mathcal{A}(\{\tilde{\mathbf{D}}\}) - \mathcal{D}(\hat{\mathbf{X}}) \right) \right\|_F^2, \quad (\text{P} - 1)$$

where $\{\tilde{\mathbf{D}}\}$ denotes a set of EDM samples, \mathcal{A} is a function that aggregates the multiple samples, \mathcal{H} is a weighting function that returns a confidence measure of the aggregated distance estimates and, finally, \mathcal{D} is a function that gives the EDM associated to a coordinate matrix $\hat{\mathbf{X}}$.

III. CLUSTERIZED WLS ALGORITHM

A. Spectral Analytical Cluster Identification

In general, a cluster identification problem is a classification problem in which similar objects are grouped or, more precisely, in which a complete data set is partitioned into subsets (clusters), so that the data in each subset share some common trait, such as proximity, as defined by some metric. In our context of localization for WSNs, a cluster can be defined as a group of nodes sufficiently connected (meshed) to one another.

This differs somewhat from classical clusterization problems [8], which often involve Euclidean metrics, such as the Euclidean distance. Indeed, since distance measurements available in WSNs localization applications are affected by errors, classic clusterization algorithms may not be effective.

Therefore, the clusterization method here considered is independent on the noise-effected distance estimates and, instead,

is based on graph-spectra analysis. In particular, we exploit a well known property of the spectrum of the Laplacian matrix of the graph associated with a meshed network, which establishes a correlation between the eigenspace and the structural property of the graph [9].

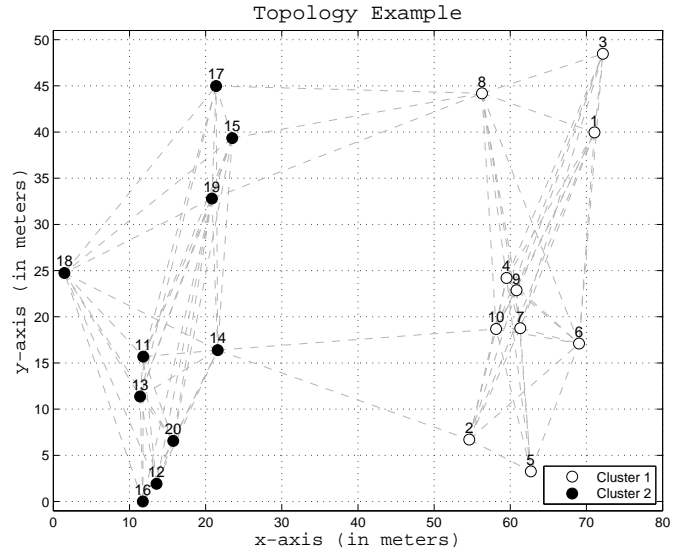


Fig. 1. Example of a network with 20 nodes, located within a squared area of 100×100 meters.

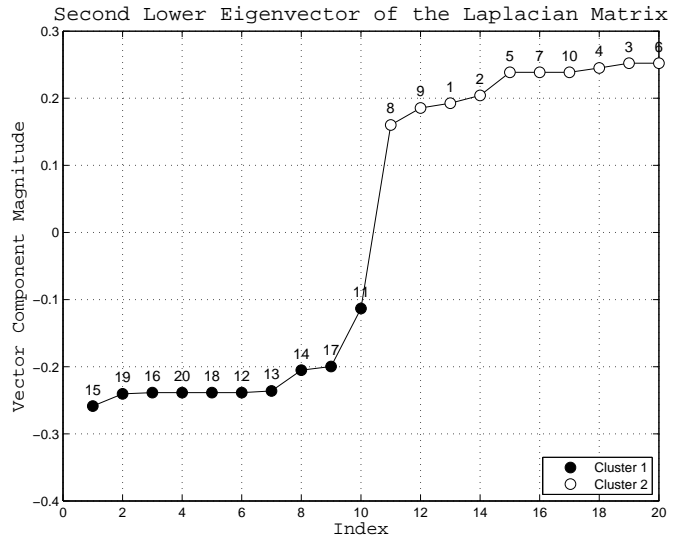


Fig. 2. Sorted components of the 2nd smaller eigenvector of the Laplacian matrix for the graph in 1.

It was recently shown [3] that clusters in a given Graph can be identified by sorting the elements of the second smallest eigenvector the Laplacian matrix \mathbf{L} against the vertices (nodes) indexes. In this method, the nodes forming a cluster are associated with the elements of the second smallest eigenvector having similar magnitudes.

As an example, consider a network with 20 nodes, clusterized into two groups of 10 nodes each, which can be easily identified visually, as shown in figure 1. The figure shows also the connections amongst nodes, from with a connectivity matrix \mathbf{C} is constructed, and the node-degrees of each sensor,

from which a diagonal node-degree \mathbf{M} is built. From \mathbf{C} and \mathbf{M} , the Laplacian matrix \mathbf{L} can then be computed.

Consider the elements b_i of the second smallest eigenvector \mathbf{b} of \mathbf{L} . Each element b_i is related to one and only one sensor v_i in the network. Figure 2 shows a plot of b_i , sorted by magnitude and labeled by corresponding node indexes.

Notice how the two clusters seen in figure 1 can be clearly identified in figure 2, due to the large gap between b_i 's with low (negative) and high (positive) magnitudes.

From this exemplary result, the power and the simplicity of the spectral-analytical cluster identification technique of [3] can be appreciated. Notice, however, that the method offers little clue as to the connectivity between nodes belonging to different clusters. For example, the nodes 8 and 11 appear adjacently in figure 2, but are in fact not directly connected to one another, as can be seen from figure 1.

B. Hierarchical Clusterization Method

The spectral-analytical cluster identification technique of [3] can be iterated so as to yield a hierarchical, multi-level clusterization map of an entire network, as follows. Starting from the graph \mathcal{G} , the connectivity matrix $\mathbf{C}^{(k)}$, the corresponding Laplacian matrix $\mathbf{L}^{(k)}$, and its second smallest eigenvector $\mathbf{b}^{(k)}$ are computed, where the superscript indicates the level of the hierarchical tree.

The vector $\mathbf{b}^{(k)}$ is then partitioned into 2 complementary parts $\{\{b_i^{(k)}\}, \{b_j^{(k)}\}\}$ accordingly to the following criteria:

$$\begin{aligned} \{b_i^{(k)}\} &= \{b_r^{(k)} \mid b_r^{(k)} \leq \gamma^{(k+1)}\} \\ \{b_j^{(k)}\} &= \mathbf{b}^{(k)} \setminus \{b_i^{(k)}\} \end{aligned} \quad (4)$$

where $\cdot \setminus \cdot$ denotes set-difference and the threshold $\gamma^{(k)}$ is given by

$$\gamma^{(k)} = \begin{cases} 0 & \text{if } \exists b_r < 0 \text{ and } b_r > 0; \\ \frac{\max(\mathbf{b}^{(k)}) + \min(\mathbf{b}^{(k)})}{2} & \text{if } b_r \leq 0, \text{ or } b_r \geq 0 \forall r. \end{cases} \quad (5)$$

Once the eigenvector is partitioned the two sub-graphs \mathcal{G}_1 and \mathcal{G}_2 are uniquely identified by their connectivity matrix $\mathbf{C}_1^{(k+1)}$ and $\mathbf{C}_2^{(k+1)}$, which are obtained from the original connectivity matrix $\mathbf{C}^{(k)}$ as follows:

$$\mathbf{C}_1^{(k+1)} = \mathbf{C}^{(k)}(\{i\}, \{i\}), \quad (6)$$

$$\mathbf{C}_2^{(k+1)} = \mathbf{C}^{(k)}(\{j\}, \{j\}). \quad (7)$$

Once the network has been clusterized, the WLS algorithm [1] can be applied to each cluster independently. The increased completeness of each cluster and the reduced size of the sub-graph results into a lower complexity optimization and a highly accurate estimation of the corresponding nodes.

Let us assume, for instance, that the initial network can be partitioned into Q clusters, each with N_q nodes. The optimization of the q -th cluster will require $\mathcal{O}(N_q \cdot \eta)^3$ operations. Therefore, the complexity reduction for the localization of all nodes in such a network is proportional to

$$\beta = \frac{\sum_{q=1}^Q \mathcal{O}(N_q \cdot \eta)^3}{\mathcal{O}(N \cdot \eta)^3}. \quad (8)$$

Notice, however, that the clusters (nodes) associated with the connectivity matrices $\mathbf{C}_1^{(k+1)}$ and $\mathbf{C}_2^{(k+1)}$ do not share any common nodes. This prevents the WLS solutions obtained based on the corresponding set of distance estimates to be merged back together.

We add also the constraint that such a division occurs only if each sub-cluster includes a minimum number of nodes n_{MIN} .

Notice, however, that localization algorithms relying on ranging information can only estimate node locations up to, *i.e.*, rotations, translations, reflections and scaling. After the nodes in each cluster has been estimated using the WLS algorithm, the different clusters must be re-connected. To this end, an overlap between adjacent clusters is added and the Procrustes transformation [10] is applied using such common nodes as ‘‘pseudo-anchors’’.

The selection of the pseudo-anchors is not trivial. Firstly, the general behavior of any localization algorithm indicates the quality of the node location estimate degrades at those nodes with lower node-degree and moreover at those nodes located at perimetrical border of the network. Secondly, a more challenging condition to be required is that the pseudo-anchors have to ensure the uniqueness realization of the graph [2] after merging.

For simplicity, we use a more practical criteria to select the pseudo anchors, which verifies that a partition can occur only if each new sub-clusters have nodes connected with at least $\eta + 1$ links to the other subcluster and the union of such sets consists of I nodes at least.

Finally, after each cluster is localized independently and adjacent clusters are reconnected via Procrustes, the reconnected clusters are then brought into an absolute orientation using the Procrustes transformation computed with An anchor nodes located in the network.

In general, the procrustes transformation is performed as follows:

$$\hat{\mathbf{X}}_A = \alpha \cdot \mathbf{Y} \cdot \mathbf{Z} + \mathbf{c} \otimes \mathbf{e}_{\{(N) \times 1\}}, \quad (9)$$

where the scalar α , the n -by- n matrix \mathbf{Z} and the 1-by- n vector \mathbf{c} are the coefficients of this transformation. A is a subscript that indicates the anchors (or pseudo-anchors when two subclusters needs to be merged), N is the number of nodes and \mathbf{Y} is the location estimated to be transformed.

Summarizing the clusterized WLS localization algorithm is performed as follows:

- 1) Consider a network and model it with a graph \mathcal{G} .
- 2) Compute the completeness of the graph and the number of nodes.
- 3) If the completeness is lower than a given threshold OR the number of nodes is larger than a given value, then continue to split the graph, otherwise go to 8.
- 4) Compute the 2nd lower eigenvector of the spectrum of considered graph and sort it with respect to the magnitude.
- 5) Use the partitioning criteria to detect two sub-clusters.
- 6) If the number of nodes per subcluster is larger than the minimum required, then compute the intersection nodes,

otherwise re-connect the subclusters into the original and go to 8.

- 7) Consider one subcluster at time and go to 2.
- 8) Store the discovered minimal cluster and check that all the subclusters has been analyzed.
- 9) If all the subclusters are minimal, then apply the WLS localization technique at each one.
- 10) Start merging by procrustes transformation the subclusters accordingly to their hierarchical generation.
- 11) Apply the final procrustes transformation on the overall network by using the knowledge of the true position of few nodes.

IV. SIMULATION RESULTS

To show the performance of the proposed clustered WLS algorithm, we consider a typical 2D-network, which represents a small warehouse of size 30×30 meters. The area includes 4 sub-areas of size 9×9 meters located at the corner of the warehouse. In each sub-area 10 target nodes (T) are randomly distributed, while in the interspace amongst them, 5 anchor nodes (A) are positioned forming a cross. It assumed that the radio connectivity for each node is limited up to 15 meters and the measurable distances are affected by error, generated from a Gamma distribution [11] with 0 mean and standard deviation σ .

In figure 3, the clusterization of the network is illustrated. Two clusters are highlighted, and only for one the connectivity is shown, and it is verified that the algorithm found a group of node that are highly meshed. The figure also shows the overlapping between the considered clusters, which has been imposed during the clusterization identification.

Another qualitative result is shown in figure ?? where we can appreciate the performance of the localization algorithm. The clustered WLS performs as good as the no-clusterized, with the further advantage of a lower complexity. It is also interesting to denote that the selection of those pseudo-anchors did not introduce a drastic error propagation, so that the final result does not degrade too much.

A second analysis is performed with respect to the noise σ and two metric are used to evaluate the performance.

The first one, shown in figure 5, concerns the local error introduced at each cluster and it is evaluated by comparing the error between the reconstructed partial EDM and the partial true EDM $\mathbf{D} = \mathcal{D}(\mathbf{X})$. In other words, such a metric is defined as follows:

$$\xi = \frac{1}{L} \sum_l^L (d_l - \hat{d}_l)^2, \quad (10)$$

where L is the overall number of links in all the clusters and \hat{d} is the Euclidean distance computed from the node estimates.

The result, illustrated in 5, indicates that the clustered WLS has a tiny better performance than the other algorithm at a local level. This is due to the high completeness of each cluster which allows to achieve lower error as shown in [1].

The second metric, instead, evaluates the error after the merging procedure. It is foreseen that the selection of pseudo-

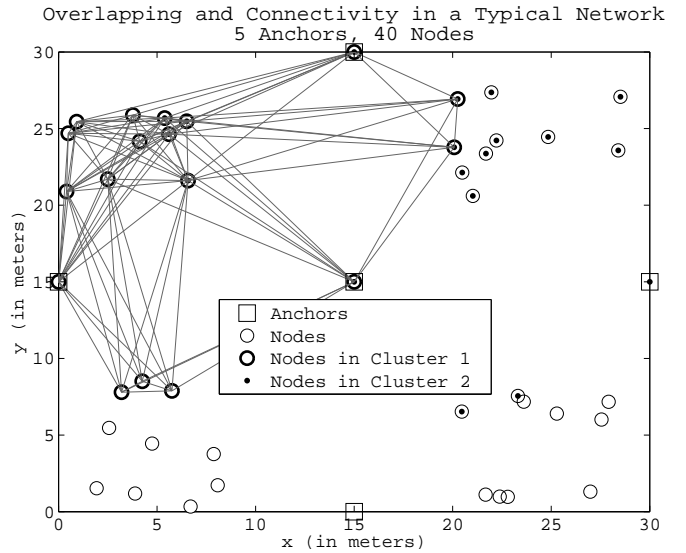


Fig. 3. Typical network with 5 anchors and 40 nodes. The connectivity is shown only for 1 cluster.

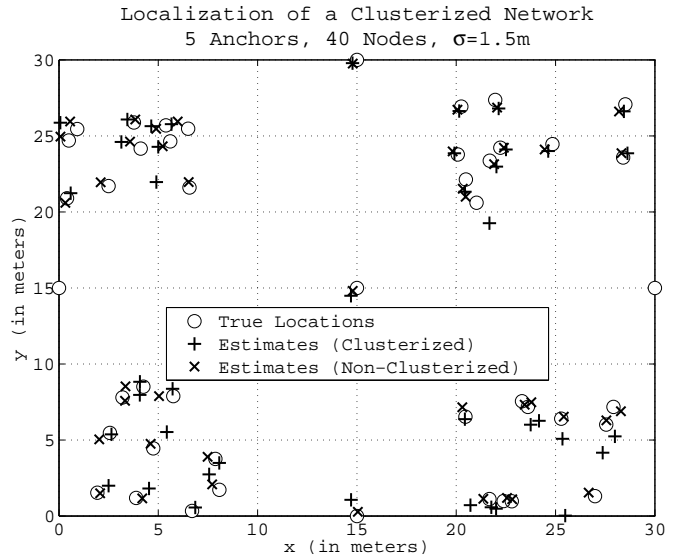


Fig. 4. Typical network with 5 anchors and 40 nodes. Clusterized and non-clusterized WLS are compared.

anchors amongst nodes that are estimates allows an error propagation. Indeed, such an effect is shown in figure 6 where the previous local solutions are merged together. The metric used is the Root-Mean-Square-Error (RMSE), given by

$$\zeta^{(\theta)} \triangleq \left\| [\mathbf{X}]_{UL:(N-A) \times \eta} - [\hat{\mathbf{X}}^{(\theta)}]_{UL:(N-A) \times \eta} \right\|_F / \sqrt{N-A}, \quad (11)$$

The result, illustrated in 6, indicates that the error propagation has, indeed, deteriorated the performance of the clustered WLS but, nevertheless, the a good trade-off with between complexity and accuracy is met.

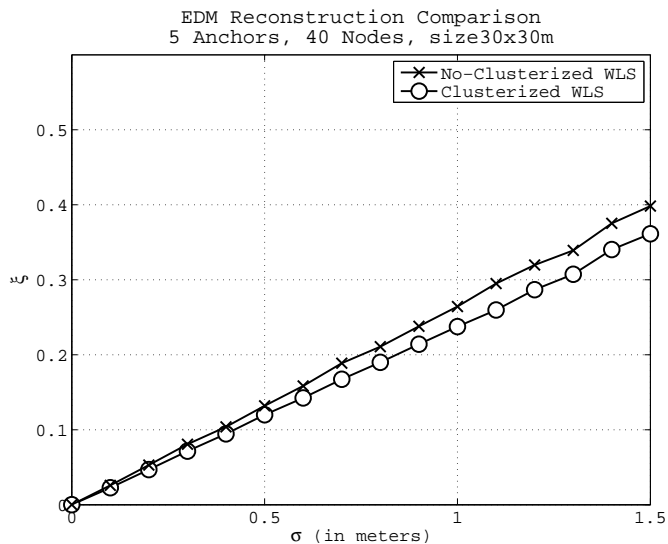


Fig. 5. Comparison between the clustered and no-clusterized WLS with respect to error over the EDM.

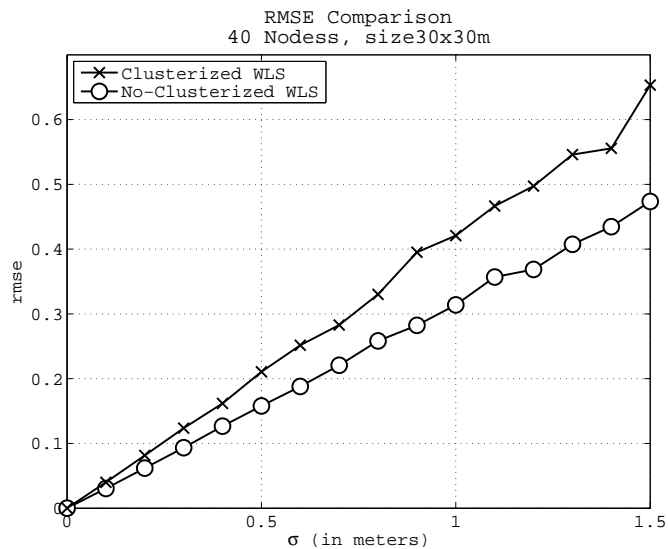


Fig. 6. Comparison between the clustered and no-clusterized WLS with respect to rmse.

V. COMMENTS AND CONCLUSIONS

An efficient and robust clustered WLS localization algorithm for large scale WSNs is presented in this work. It is based on a cluster identification method, localized WLS optimization and sequential merging.

The cluster identification method is based on recent results on Graph-spectrum theory and it only relies on the connectivity information. In the context of localization for WSNs, such a property allows to the cluster identifications without the influence of the error in the distance measurement. Since only the connectivity is required, the cluster identification method can be performed before a priori and it can be used to find a logic position of the device, such as: close to anchor A, in the room X, etc.

The main problem of the proposed algorithm, as discussed in the work, concerns the merging procedure, since it needs to be performed over nodes that are estimated. In order to mitigate the error propagation, further studies in graph rigidity need to be carried out.

REFERENCES

- [1] G. Destino and G. T. F. de Abreu, "Sensor localization from wls optimization with closed-form gradient and hessian," in *Proc. IEEE 49th Annual Globecom Conference (GLOBECOM'06)*, 2006.
- [2] B. Hendrickson, "Conditions for unique graph realization," *SIAM Journal on Comp.*, vol. 21, no. 1, pp. 65–84, 1992.
- [3] K. B. O. Krishnadev and S. Vishveshwara, "A graph spectral analysis of the structural similarity network of protein chains," *PROTEINS: Structure, Function, and Bioinformatics*, vol. 61, pp. 152–163, 2005.
- [4] G. Destino and G. T. F. de Abreu, "Localization from imperfect and incomplete ranging," in *Proc. IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'06)*, 2006.
- [5] A. Y. Alfakih, H. Wolkowicz, and A. Khandani, "Solving euclidean distance matrix completion problems via semidefinite programming," *Journal on Computational Optimization and Applications*, vol. 12, no. 1, pp. 13 – 30, 1999. [Online]. Available: <http://portal.acm.org/citation.cfm?id=316254/#>
- [6] M. Chu and G. Golub, *Inverse Eigenvalue Problems Theory, Algorithms, and Applications*, ser. Numerical Mathematics and Scientific Computation. Oxford University Press, 2005. [Online]. Available: <http://www4.ncsu.edu/~mtchu/Research/Papers/distance03.pdf>
- [7] J. Nocedal and S. J. Wright, *Numerical Optimization*. Springer, 2000.
- [8] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Computing Surveys*, vol. 31, no. 3, pp. 264–323, 1999.
- [9] N. Biggs, *Algebraic Graph Theory*. Cambridge University Press, 1994.
- [10] P. Fiore, "Efficient linear solution of exterior orientation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 2, pp. 140–148, February 2001.
- [11] G. Destino and G. T. F. de Abreu, "MDS-WLS optimization for sensor localization," in *Proc. IST Mobile & Wireless Communications Summit*, 2006.