

Article

Localization with Graph Diffusion Property

Pengpeng Chen, Yuqing Yin, Shouwan Gao, Qiang Niu * and Jun Gu

School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China; chenp@cumt.edu.cn (P.C.); yinyuqing@cumt.edu.cn (Y.Y.); gaoshouwan@cumt.edu.cn (S.G.); gujcumt@163.com (J.G.)

* Correspondence: niuq@cumt.edu.cn; Tel.: +86-516-8359-1705

Received: 10 May 2017; Accepted: 14 July 2017; Published: 15 July 2017

Abstract: Node localization is an essential issue in wireless sensor networks (WSNs). Many range-free localization methods have been proposed to satisfy the requirement of low-system cost. However, some range-free methods only depend on network connectivity, and others only utilize the proximity information attached in neighborhood ordering. To employ the strength of the above two aspects, this paper introduces a new metric system called Combined and Weighted Diffusion Distance (CWDD). CWDD is designed to obtain the relative distance among nodes based on both graph diffusion property and neighbor information. We implement our design by embedding CWDD into two well-known localization algorithms and evaluate it by extensive simulations. Results show that our design improves the localization performance in large scale and non-uniform sensor networks, which reduces positioning errors by as much as 26%.

Keywords: node localization; graph diffusion property; wireless sensor networks

1. Introduction

Node localization is a fundamental issue in wireless sensor networks (WSNs) that has also been studied for a long time and has received many classic and improved theories [1–4] for practical applications, e.g., border security and surveillance [4], item tracking [5,6], road monitoring [7], or environment data collection [8]. Two basic methods, namely, range-based and range-free schemes, are proposed for node localization. Although the range-based localization can provide high-accuracy ranging, range-free localization may decrease system cost. Thus, the range-free method is a better choice compared with the former one.

The range-free approach uses simple sensing to localize nodes; this approach is based on the information of wireless networks, such as connectivity and proximity information. Connectivity-based localization, e.g., DV-Hop [2], may decrease system cost at the expense of accuracy. Additionally, connectivity-based localization does not perform well in unevenly distributed networks. On the other side, some ideas collect the Received Signal Strength (RSS) to obtain the proximity information. The RSS values become uncertain and irregular because of some factors, such as radio path loss, random noise, and complex terrain. This phenomenon may also increase localization errors.

Our present work is motivated by the finding that the graph diffusion property [9] of network connectivity can be applied to calculate a new metric for relative distance among node pairs. This new concept “Graph diffusion property” was developed in the bioinformatics field for protein function prediction. In protein to protein interaction (PPI) networks, the simple shortest-path metric is not suitable to apply to the fine-grained PPI graph for similarity measurement. In view of this, Lenore J. Cowen et.al., proposed diffusion state distance (DSD), a new metric based on graph diffusion property, designed to capture finer-grained distinction in proximity for transfer of functional annotation in PPI networks. The innovation point of DSD is to apply random walk to network structure, which exactly brings out the theme “graph diffusion”. Accordingly, DSD can make full

use of the important knowledge encoded in the structure of the network and can tell the difference between the adjacent of two nodes and that of two other nodes. Furthermore, few methods combine both connectivity and proximity information in WSNs to design new localization methods. Hence, the need for a novel method that compensates each other's shortcomings arises.

In this paper we first propose a network connectivity-based diffusion distance (DD), which is a new metric to indicate the relative distance between sensor pairs. With the idea of random walk [9,10], we suspect that one node can reach the remaining neighboring nodes with the same probability. Moreover, one of the neighboring nodes will continue on repeating this section to the next neighbor node within a preset step. Subsequently, we accumulate the probabilities produced by each step among random walks and extract the dissimilarity of probabilities that serves as a relative distance. Second, this study improves the existing method and introduces combined and weighted DD (CWDD). According to RSS value, the far-and-near relation between two neighbor nodes can be determined. This relationship can be converted to numbers as the weight of connected lines. Therefore, we modify the process in random walks [11]. In terms of different weight values attached on neighbor lines, one node can reach the remaining neighbor nodes with different probabilities. When two neighbor nodes are far from each other, the probability value of reaching each other will be small. Afterward, we obtain the probabilities produced by each step among random walks. We also extract the dissimilarity of probabilities to serve as a relative distance. Finally, we embed CWDD into two classic localization algorithms, i.e., DV-Hop [2] and RPA [12], to replace hops with relative distances. Extensive simulations prove that CWDD-embedded methods perform better than hop-based methods in dense and non-uniform networks and reduce the localization errors by approximately 26%. Specifically, our main contributions in this paper are as follows.

- (1) Given the graph diffusion property of network connectivity, a new metric called DD is designed to indicate the relative distance between sensor node pairs.
- (2) To further improve the localization performance, a refined version named CWDD is proposed by adding nodes' proximity information that serves as the weight on connected lines in sensor networks.
- (3) Extensive simulations are conducted to prove that CWDD-embedded method outperforms original methods with high accuracy.

The remaining parts of this paper are organized as follows. Section 2 briefly surveys the related work. Section 3 explains the objective of this study. The main design is elaborated in Section 4. Section 5 embeds the main design into classic algorithms, and Section 6 discusses the results from simulations. Finally, conclusions are provided in Section 7.

2. Related Work

Localization in WSNs is regarded as an emerging technology for numerous systems [13]. Range-based [14,15] and range-free [16–20] are two main types of existing methods on localization work. Lately, other methods which combine range-based or range-free principle are also available [21].

Range-based localization requires costly ranging hardware [22] and system calibration [23] support to measure the physical distance among sensor nodes precisely and then locate unknown nodes with geometric approaches, such as TOA [24], TDOA [14], and AOA [15]. The RSS localization scheme has low hardware cost which only relies on the communication module, but it has relatively poor localization performance. In all, although most of range-based localization achieves high-localization accuracy, its hardware is expensive and limits the scope of application, which is not appropriate for large-scale WSNs.

Conversely, low-system cost is researched by using range-free localization methods. On early stage, range-free localization utilizes the advantages of anchor proximity information to locate nodes, like centroid [25]. To obtain high accuracy, the network requires high anchor density, but it is not suitable for large-scale networks because of high cost of anchors. Afterward, connectivity-based

range-free methods, including DV-Hop [2], RPA [12], and MDS-MAP [26], are designed to reduce the proportion of anchors and obtain a relatively high accuracy, which reduces the system cost to some extent. Moreover, these methods construct global coordinates by creating a metric for hop-based virtual distances via neighborhood sensing, which is practical for large-scale networks.

However, these methods may suffer from hop-distance ambiguity problem. Therefore, a method based on regulated neighborhood distance (RND) is proposed to address this problem by relating the proximity of two neighbors to their neighbor partitions [13]. Supposing that node a and node b are neighbors, A is the neighbor set of node a and B is the neighbor set of node b . It defines ND as neighborhood distance, where $ND(a, b) = |A - A \cap B|/|A|$ and $ND(b, a) = |B - A \cap B|/|B|$. Then, $RND(a, b)$ is denoted as the average of two NDs. This method achieves a high correlation coefficient between RND and physical distance, and performs well in uniform and dense networks. Besides, a range-free localization method beyond connectivity is proposed in [27], which excavates further information regarding neighborhood sensing. Each node can obtain a neighbor ordering that ranges from the large values of neighborhood RSS to small ones. A signature distance between a pair of nodes can be subsequently measured by comparing the neighbor sequences of node pairs. Considering non-uniform networks, the signature distance is regulated as regulated signature distance (RSD), which is a novel metric of proximity among neighbor nodes. This design remarkably improves the localization accuracy, but increases additional time complexity responsively.

Unlike the previously presented ideas, this paper introduces the CWDD, a novel distance estimation metric based on both graph diffusion property of network connectivity and neighbor information. Given that CWDD is embedded into classic range-free localization methods, our design performs well in accuracy and can be applied for large-scale and non-uniform networks.

3. Motivation

We first elaborate the drawbacks of connectivity-based localization and then show the opportunity for our new method.

3.1. Limitations of Hop-Based Localization

Three typical connectivity-based localization approaches, namely, DV-Hop [2], MDS-MAP [12], and RPA [26], are available. The common idea of these approaches is the usage of the shortest path based on hops among node pairs. Afterward, the shortest paths are transferred into relative distances, and coordinates are estimated by the least square method. Considering DV-Hop as an example, distance vector routing and theory of GPS localization are used. The algorithm is divided into the following steps:

- (1) In all nodes, the shortest hops reaching each anchor are calculated. $Hops(u_i, v_j)$ represents the least number of hops between an unknown node u_i and an anchor v_j .
- (2) Utilizing the anchors' coordinates, the algorithm can estimate the average physical distances for one hop using the following Equation (1):

$$HopSize = \frac{\sum_{i \neq j} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{\sum_{i \neq j} Hops(v_i, v_j)} \quad (1)$$

where $v_i(x_i, y_i)$ and $v_j(x_j, y_j)$ are anchors, and $Hops(v_i, v_j)$ is the number of the least hops between v_i and v_j . For each unknown node u_i , the distances of reaching each anchor v_j can be estimated using the formula $HopSize \times Hops(u_i, v_j)$.

- (3) Along with anchors' coordinates and distances to all anchors, an unknown node will finally determine its coordinate using maximum likelihood estimation method.

The drawback of this algorithm is that hop size only stands for a mean distance for one hop, which may get fair precision in uniform networks because of similar distances among hops. However, this hop is not appropriate especially for non-uniform networks. Hop-based algorithm will cause large localization errors, because one hop may corresponds several distances.

3.2. Opportunity for Our New Method

Considering the limitation investigated in the last section, we bypass the hop-distance model by analyzing some other useful information, such as graph diffusion property from network connectivity. As illustrated in Figure 1, two remarkable observations are obtained.

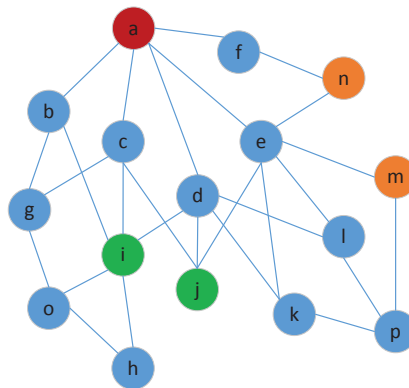


Figure 1. Different number of common neighbors showing different distances.

Remark 1. In wireless sensor networks, two nodes with many neighbors in common are closer than nodes with few neighbors in common.

Remark 2. Such two nodes with many common low-degree neighbors are closer than nodes with high-degree common neighbors.

As shown in Figure 2, there are three nodes with radio range R . The common neighbors can only exist in the overlapping region of two large circles. Therefore, the overlapping region is large with the presence of many common neighbors in a node pair, and the distance within a node pair is close when the two centers of circles are close to each other, which explains why Remark 1 makes sense.

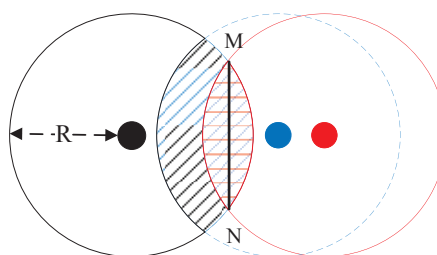


Figure 2. The diagrammatic sketch of Remarks 1 & 2.

Remark 2 leads us to think about the following question. Under the condition of same quantity of common neighbor nodes between two nodes, which types of common neighbor nodes the two nodes have more, low-degree nodes or high-degree nodes, can shorten the distance between these two nodes? Assuming that there are fixed number of common neighbors in the overlapping region. One common neighbor can treat all other common neighbors in the overlapping region as its neighbor, when the intersecting chord of two circles like segment MN in Figure 2 (the longest connection range

of two common neighbors) is shorter than radio range R . On the other hand, when the intersecting chord of two circles is longer than radio range R , there may exist that one common neighbor cannot connect with some other nodes in the overlapping region, which may relatively reduce the degree of this common neighbor compared with the former case to some extent. To sum up, low-degree common neighbors appear in the larger overlapping region than relatively high-degree common neighbors do, and according to the explanations of Remark 1, larger overlapping region draw two nodes more closely.

A connectivity graph of a network is observed by comparing the nodes in Figure 1. As concluded from Remark 1, the distance between nodes a and n is smaller than that of nodes a and m because nodes a and n exhibit two common neighbors, namely, nodes f and e . Nodes a and m share node e only. Furthermore, the degrees of nodes c , d , and e are 4, 5, and 6, respectively. According to Remark 2, the distance between nodes a and i is smaller than that of nodes a and j because nodes a and i display relatively low-degree common neighbors. By contrast, nodes a and j show relatively high-degree common neighbors.

According to the property of network connectivity mentioned above, we first propose a DD, which is a new metric to indicate the relative distance between node pairs. With the limitation of hop-based localization, we regard proximity information as a weight attached on connective line for one hop. Therefore, we prove the method and introduce CWDD by combining the properties of network connectivity and proximity information.

4. System Design

4.1. Definition of a New Distance Metric

On a 2D plane, we consider locating a network of wireless nodes with connectivity information as the undirected graph $G(V, E)$, where $V = \{v_1, v_2, \dots, v_n\}$ is the vertex set, and E is the adjacency matrix of this graph.

Given the adjacency matrix of graph G , a probability matrix for one-hop neighbors P can be obtained with the following steps, where the (i, j) th entry is given by Equation (2).

- (1) The value of each row in adjacency matrix E is summed as the degree of each node.
- (2) Each row of E is divided with the corresponding degree.

$$P(i, j) = \begin{cases} \frac{1}{\sum_{l=1}^n E(v_i, v_l)} & \text{if } E(v_i, v_j) = 1 \\ 0 & \text{if } E(v_i, v_j) = 0 \end{cases} \quad (2)$$

Considering the concept of one-hop probability matrix P , any node v_i will reach its next one-hop neighbors with the same probability, which is determined by the degree of node v_i . Subsequently, $p^{(l)}(v_i, v_j)$ is defined as the probability from node v_i proceeding l steps toward node v_j in a random walk. Due to each step of random walk is an independent event, the probability of each path from node v_i proceeding l steps toward node v_j can be expressed as the product of all one-hop probabilities along this path. Furthermore, $p^{(l)}(v_i, v_j)$ comes out by adding all probabilities of each path. $Ap^{(m)}(v_i, v_j)$ is denoted as the accumulated probability produced from one step to m steps in a random walk starting at node v_i and finishing at node v_j , where

$$Ap^{(m)}(v_i, v_j) = \sum_{l=1}^m p^{(l)}(v_i, v_j) \quad (3)$$

We further define a n -dimensional vector $Apv^{(m)}(v_i), \forall v_i \in V$, where

$$Apv^{(m)}(v_i) = \{Ap^{(m)}(v_i, v_1), Ap^{(m)}(v_i, v_2), \dots, Ap^{(m)}(v_i, v_{n-1}), Ap^{(m)}(v_i, v_n)\} \quad (4)$$

A DD between two vertices v_i and v_j , $\forall v_i, v_j \in V$ is also denoted as follows:

$$DD^{(m)}(v_i, v_j) = \left\| Apv^{(m)}(v_i) - Apv^{(m)}(v_j) \right\|_1 \quad (5)$$

where $\|\dots\|_1$ means the L_1 norm of two vectors' difference.

Figure 3 shows a simple example, including a topological graph G_0 of a wireless network on the left and the corresponding adjacency matrix E on the right. We use the connectivity information illustrated in Figure 3 to calculate the value of $DD^{(2)}(v_1, v_4)$ and $DD^{(2)}(v_1, v_6)$ with the assumption $m = 2$.

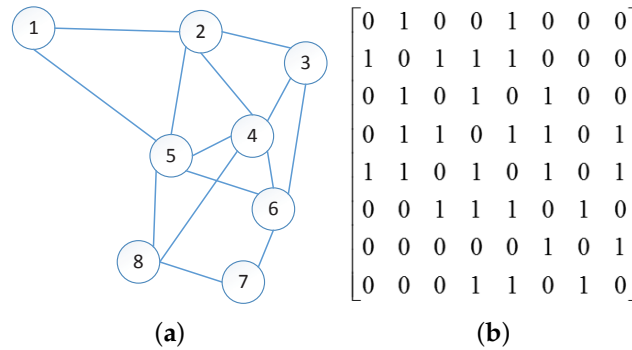


Figure 3. Diffusion Distance (DD): (a) Topological graph G_0 , (b) Adjacency matrix E .

First, the result of one-hop probability matrix P is the following matrix.

$$P = \begin{bmatrix} 0 & 1/2 & 0 & 0 & 1/2 & 0 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 & 1/4 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 1/3 & 0 & 1/3 & 0 & 0 \\ 0 & 1/5 & 1/5 & 0 & 1/5 & 1/5 & 0 & 1/5 \\ 1/5 & 1/5 & 0 & 1/5 & 0 & 1/5 & 0 & 1/5 \\ 0 & 0 & 1/4 & 1/4 & 1/4 & 0 & 1/4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 1/3 & 1/3 & 0 & 1/3 & 0 \end{bmatrix}$$

Second, the accumulated probability can be calculated as follows:

$$\begin{aligned} Ap^{(2)}(v_1, v_4) &= p^{(1)}(v_1, v_4) + p^{(2)}(v_1, v_4) \\ &= 0 + 9/40 \approx 0.225, \end{aligned}$$

where $9/40 = 1/2 \times 1/4 + 1/2 \times 1/5$, means that two paths exist from v_1 to v_4 through v_2 and v_5 during the two steps. Similarly,

$$Ap^{(2)}(v_1, v_6) = p^{(1)}(v_1, v_6) + p^{(2)}(v_1, v_6) = 0 + 0.1 = 0.1.$$

Third, we obtain three n -dimensional vectors, namely, $Apv^{(2)}(v_1)$, $Apv^{(2)}(v_4)$, and $Apv^{(2)}(v_6)$.

$$\begin{aligned} &Apv^{(2)}(v_1) \\ &= \{Ap^{(2)}(v_1, v_1), Ap^{(2)}(v_1, v_2), \dots, Ap^{(2)}(v_1, v_8)\} \\ &= \{0.225, 0.6, 0.125, 0.225, 0.625, 0.1, 0, 0.1\} \end{aligned}$$

$$\begin{aligned}
& Apv^{(2)}(v_4) \\
&= \{Ap^{(2)}(v_4, v_1), Ap^{(2)}(v_4, v_2), \dots, Ap^{(2)}(v_4, v_8)\} \\
&= \{0.09, 0.31, 0.3, 0.27, 0.37, 0.31, 0.12, 0.24\} \\
& Apv^{(2)}(v_6) \\
&= \{Ap^{(2)}(v_6, v_1), Ap^{(2)}(v_6, v_2), \dots, Ap^{(2)}(v_6, v_8)\} \\
&= \{0.05, 0.18, 0.3, 0.38, 0.3, 0.31, 0.25, 0.225\}
\end{aligned}$$

Finally, $DD^{(2)}(v_1, v_4)$ and $DD^{(2)}(v_1, v_6)$ are counted as follows.

$$\begin{aligned}
DD^{(2)}(v_1, v_4) &= \|Apv^{(2)}(v_1) - Apv^{(2)}(v_4)\|_1 \\
&= |0.225 - 0.09| + |0.6 - 0.31| \\
&\quad + |0.125 - 0.3| + |0.225 - 0.27| \\
&\quad + |0.625 - 0.37| + |0.1 - 0.31| \\
&\quad + |0 - 0.12| + |0.1 - 0.24| \\
&= 1.36 \\
DD^{(2)}(v_1, v_6) &= \|Apv^{(2)}(v_1) - Apv^{(2)}(v_6)\|_1 \\
&= |0.225 - 0.05| + |0.6 - 0.18| \\
&\quad + |0.125 - 0.3| + |0.225 - 0.38| \\
&\quad + |0.625 - 0.3| + |0.1 - 0.31| \\
&\quad + |0 - 0.25| + |0.1 - 0.225| \\
&= 1.88
\end{aligned}$$

L_1 norm counts the absolute value of two vectors' difference. Thus, the definition of this new distance metric guarantees $DD^{(m)}(v_i, v_j) \equiv DD^{(m)}(v_j, v_i)$, assuming that the steps of random walks m are fixed. Moreover, we can conclude that $DD^{(m)}(v_i, v_i) \equiv 0$ from Equation (5).

Algorithm 1: Diffusion Distance.

Input: $adj_m, i, j = \{1, 2, \dots, N\}$
Output: DD_m
1: $degree(i) = \text{sum}(adj_m(i, :))$;
2: $P(i, :) = adj_m(i, :) / degree(i)$;
3: $Ap_m^{(m)} = P + P^2 + \dots + P^m$;
4: $DD_m(i, j) = \|Ap_m^{(m)}(i, :) - Ap_m^{(m)}(j, :)\|_1$.

Algorithm 1 shows the entire process for computing a DD matrix, where the (i, j) th entry represents the relative distance between nodes v_i and v_j . Initially, the adjacency matrix of graph G_0 shortened as adj_m is needed as an input. For $i = \{1, 2, 3, \dots, N\}$, $degree(i) = \text{sum}(adj_m(i, :))$ counts the degree of each node with the complexity $O(N)$, where N is the number of all nodes. Subsequently, for $i = \{1, 2, \dots, N\}$, one-hop probability matrix P consists of $P(i, :) = adj_m(i, :) / degree(i)$ with the cost of $O(N)$. The accumulated probability matrix $Ap_m^{(m)}$ can be calculated by formula $Ap_m^{(m)} = P + P^2 + \dots + P^m$, where the i th row or column is coincident with the n -dimensional vector $Apv^{(m)}(v_i)$. Finally, the DD matrix DD_m is composed of $N \times N$ entries. The (i, j) th entry is the result of L_1 norm of the difference between the i th and j th rows of $Ap_m^{(m)}$ based on Equation (5) with complexity $O(N^3)$, where the values on diagonal line of DD_m are all zeros.

4.2. Insight into the DD

This section discusses how DD can serve as relative distance and what DD is going to be when m tends to infinity.

Recalling Remark 1 from Section 3, two nodes with many neighbors in common are closer than nodes with few neighbors in common. On the basis of the example above, node pair (v_1, v_4) presents two common neighbors, which is more than one common neighbor between node pair (v_1, v_6) . Thus, v_4 is closer to v_1 than v_6 . Furthermore, the DDs between (v_1, v_4) and (v_1, v_6) node pairs, which are counted as following formulas, are 1.36 and 1.88, respectively. These two numbers reflecting far and near relation among node pairs are also coincident with this remark.

$$\begin{aligned} DD^{(2)}(v_1, v_4) &= \left\| Ap^{(2)}(v_1) - Ap^{(2)}(v_4) \right\|_1 \\ &= |0.225 - 0.09| + \underline{|0.6 - 0.31|} \\ &\quad + |0.125 - 0.3| + |0.225 - 0.27| \\ &\quad + |0.625 - 0.37| + |0.1 - 0.31| \\ &\quad + |0 - 0.12| + |0.1 - 0.24| \\ &= 1.36 \end{aligned}$$

$$\begin{aligned} DD^{(2)}(v_1, v_6) &= \left\| Ap^{(2)}(v_1) - Ap^{(2)}(v_6) \right\|_1 \\ &= |0.225 - 0.05| + \underline{|0.6 - 0.18|} \\ &\quad + |0.125 - 0.3| + |0.225 - 0.38| \\ &\quad + |0.625 - 0.3| + |0.1 - 0.31| \\ &\quad + |0 - 0.25| + |0.1 - 0.225| \\ &= 1.88 \end{aligned}$$

The underlined items (the second ones) mean the absolute values of formulas:

$$Ap^{(2)}(v_1, v_2) - Ap^{(2)}(v_4, v_2),$$

$$Ap^{(2)}(v_1, v_5) - Ap^{(2)}(v_4, v_5);$$

The bold items (the fifth ones) mean the absolute values of formulas:

$$Ap^{(2)}(v_1, v_2) - Ap^{(2)}(v_6, v_2),$$

$$Ap^{(2)}(v_1, v_5) - Ap^{(2)}(v_6, v_5);$$

where v_2, v_5 are the common neighbors of node pair (v_1, v_4) and v_5 is the common neighbor of node pair (v_1, v_6) . When the same common neighbor v_5 is shared, the values of two bold items (the fifth ones) are almost same. However, two underlined items (the second ones) show different values according to the existence of a common neighbor v_2 . The difference value of sharing v_2 in common is smaller than that of not having v_2 in common. Consequently, difference calculation of the accumulated probability vectors can shorten the DD of node pairs with many common neighbors, thereby the conclusion is shown as Equation (6) based on Remark 1:

$$DD^{(m)}(v_i, v_j) \propto PD(v_i, v_j) \quad (6)$$

where PD represents the physical distance. This new distance metric, DD, can be a relative distance because of this approximately proportional relation. Further information on how to convert relative distance to physical distance is introduced in Section 5.

Furthermore, we are going to discuss the DD value under the condition that random walk step m tends to infinity. There is a remarkable characteristic of the random walk model, that is, the result of each iteration is only related to the previous one, and it has nothing to do with the earlier results. This process is also known as Markov process or Markov chain. Thus, there exists a positive integer $k(k < m)$, such that matrix $P^k = P^{k+1} = \dots = P^m$ when m becomes infinity in a random walk. These same parameters may be removed when making difference calculation of accumulated probability vectors. Consequently, the DD value may converge to the extreme value as m goes to infinity. Let $M = \lim_{m \rightarrow \infty} P^m$ is the constant matrix in which each row is a copy of the unique steady state distribution [28]. According to Algorithm 1, we can derive the stable value of DD between node v_i and node v_j as follows:

$$\begin{aligned} & \lim_{m \rightarrow \infty} DD^{(m)}(v_i, v_j) \\ &= \left\| \lim_{m \rightarrow \infty} Apv^{(m)}(v_i) - Apv^{(m)}(v_j) \right\|_1 \\ &= \left\| \lim_{m \rightarrow \infty} (\mathbf{b}_i^T - \mathbf{b}_j^T)(P + P^2 + \dots + P^m) \right\|_1 \end{aligned} \quad (7)$$

$$= \left\| \lim_{m \rightarrow \infty} (\mathbf{b}_i^T - \mathbf{b}_j^T) \frac{P(1 - P^m)}{1 - P} \right\|_1 \quad (8)$$

$$= (\mathbf{b}_i^T - \mathbf{b}_j^T)P(1 - M)(1 - P)^{-1} \quad (9)$$

where $\mathbf{b}_i$ is the i th basis vector, i.e., the row vector of all zeros except for a 1 in the i th position [28]. For deeper interpretation, Equation (7) holds by step 3 in Algorithm 1, Equation (8) utilizes the summation formula of geometric series, and Equation (9) forms the final invariable value of DD.

4.3. CWDD Design

In the above analysis, we only make use of network connectivity for the random walk. However, the neighborhood proximity information behind RSS values can also affect the distance between two nodes. In addition, it is not accurate to consider that one node can reach to each of its next-hop nodes with same probabilities. In view of these reasons, we propose a refined version of DD, that is, combined and weighted diffusion distance (CWDD). Just as the name says, CWDD combines network connectivity and neighborhood proximity these two factors by taking proximity information as a weight value attached on connective lines. The weight values can distinguish the probabilities from one node to its neighbor nodes, and then CWDD can tell the tiny differences among several distances, which can improve the positioning accuracy to some extent. We use the modified RSD [27] as the weight. Recalling the introduction from Section 2, a signature distance can be first obtained from RSS values and then regulated to RSD with Equation (10):

$$RSD(v_i, v_j) = SD(S_i, S_j) \frac{\sqrt{K}}{K(K-1)/2} \quad (10)$$

where S_i is the neighbor ordering of node v_i , and $K = |S_i \cup S_j|$ is the total number of nodes in the neighborhood of nodes v_i and v_j .

Subsequently, we reconstruct the undirected graph $G(V_{comb}, E_{comb}, W)$, where $V_{comb} = V$ is the vertex set, and $E_{comb} = E$ is the adjacency matrix. W is the weight matrix, where the (i, j) th entry is described as follows:

$$w_{i,j} = 1 - SD(S_i, S_j) \frac{2}{K(K-1)} \quad (11)$$

Equation (11) is a modified RSD, which satisfies the condition that the weight value is inversely proportional to physical distance. This process is in preparation for one-hop probability matrix P_{comb} , where the (i, j) th entry is given by the following:

$$P_{comb}(i, j) = \begin{cases} \frac{w_{ij}}{\sum_{l=1}^n w_{il}} & \text{if } E(v_i, v_j) = 1 \\ 0 & \text{if } E(v_i, v_j) = 0 \end{cases} \quad (12)$$

Equation (12) implies the small distance between v_i and its one-hop neighbor v_j and the high probability proceeding from v_i to v_j in a random walk. Accordingly, $p_{comb}^{(m)}(v_i, v_j)$ is constructed as the probability from node v_i proceeding m steps toward node v_j , and $Ap_{comb}^{(m)}(v_i, v_j)$ is the accumulated probability. Therefore, the n -dimensional vector $Apv_{comb}^{(m)}(v_i), \forall v_i \in V$ can be obtained, and the definition of CWDD is shown as Equation (13).

$$CWDD^{(m)}(v_i, v_j) = \left\| Apv_{comb}^{(m)}(v_i) - Apv_{comb}^{(m)}(v_j) \right\|_1 \quad (13)$$

Thus, we introduce the new metric for relative distance and its combined version applying for node localization in WSNs.

5. Embedding Design

The design of new metrics for relative distance can be embedded into two classical connectivity-based schemes for localizing the coordinates of unknown nodes. These nodes are compared with real coordinates to evaluate the system accuracy, and the two schemes are as follows:

- DV-Hop [2] by D. Niculescu and B. Nath.
- RPA [12] by C. Savarese, J.M. Rabary et al.

DV-Hop is described in Section 3, and a brief description of RPA algorithm is described in this section. RPA is separated into two periods, namely, start-up and refinement. For the first phase, Hop-TERRAIN, an in-house algorithm similar to DV-Hop, is used for allowing all nodes to arrive at initial position estimates by cooperatively spreading awareness of the anchor nodes' positions throughout the network. The second phase proposes an iterative refinement algorithm for position adjustment based on the results of first step because of the low accuracy of initial estimates. In addition, at iteration k , refinement is concerned only with nodes existing within a one-hop neighborhood, and it is recomputed based on the estimated positions from iteration $k - 1$. RPA aims to increase the system accuracy. However, the high number of iterations will result in further time complexity in the system costs.

Considering the challenge of hop-based design, the value of new metric for relative distance can be used to replace the shortest-path hop distance. For DV-Hop, the expected physical distance for one unit relative distance (DD or CWDD) can be modified based on Equation (1) in Section 3, which is given by Equation (14).

$$RD_{unitSize} = \frac{\sum_{i \neq j} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{\sum_{i \neq j} RD(v_i, v_j)} \quad (14)$$

For RPA, Equation (14) is also suitable for the revision of the first phase. Moreover, the refinement step also benefits from embedding DD or CWDD for adjusting iteration position.

6. Simulation Evaluation

In this section, we report different simulation results about several aspects, such as the correlation between relative distance and physical distance, discussion of the effects with different random-walk steps, and localization performance under different settings with and without CWDD embedding.

6.1. Default Simulation Configurations

In the simulation, we create a WSN as a square map without holes, which means that radio signal can reach anywhere in its area of interest. Further complicated maps can be used with works [29–31]. Unless otherwise specified, default simulation configurations are listed in Table 1 for the entire section.

Table 1 shows that noise model is a widely used logarithmic attenuation model [32,33] for RSS sensing, where the sensing result between two nodes is given by Equation (15):

$$P_{i,j}(t) = P_t - P_l(d_0) - 10\beta \log\left(\frac{PD(v_i, v_j)}{d_0}\right) + X_i(t) \quad (15)$$

where $P_{i,j}(t)$ stands for the sensing result starting from node v_i to node v_j at time instance t , P_t is the transmit power at a short reference distance d_0 , $P_l(d_0)$ is the pass loss for the reference distance d_0 , β means the pass loss exponent or fading factor and $PD(v_i, v_j)$ is the physical distance between sender node v_i and receiver node v_j , and $X_i(t)$ means that for node v_i it follows $X_i(t) \sim N(0, \sigma_X^2)$ at time t .

Table 1. Default simulation configurations.

Parameter	Default Values	Description
System Scale	700 m × 700 m	
Noise Model	Logarithmic Attenuation Model	
Number of Sensor Nodes	300, randomly deployed with uniform distribution	
Proportion of Anchor Nodes	20%, randomly deployed	
Radio Range	150 m	

Furthermore, we define system localization error LE as follows:

$$LE = \frac{\sum_{i=1}^n \sqrt{(x_{ie} - x_{ir})^2 + (y_{ie} - y_{ir})^2}}{N_{non_anc} \times R} \quad (16)$$

where n is the number of nodes, (x_{ie}, y_{ie}) is the estimated coordinate of node v_i , and (x_{ir}, y_{ir}) is the real coordinate of node v_i , N_{non_anc} means the number of non-anchor nodes, and R represents the radio range. For high confidence, all following statistics of localization errors are averaged over 50 runs.

6.2. Distance Correlations

Figure 4 shows the correlations between hop-based distance, DD, and the physical distance. Figure 4a,b plot the node pairs within one-hop connectivity in networks. Figure 4c,d plot all the node pairs, including one-hop and multi-hop node pairs and their shortest-path hops and DD distance against physical distance. In these four figures, X axis represents the physical distance among node pairs, and Y axis represents the hop distance and DD.

Figure 4a indicates that node pairs within one-hop connectivity exhibit an identical hop distance “1”, regardless of the corresponding physical distances. Consequently, correlation coefficient ρ is 0. By contrast, as illustrated in Figure 4b, a roughly linear correlation exists between DD value and physical distance, where the correlation coefficient is $\rho = 0.77$. The new metric for relative distance provides a sub-hop resolution that is not available in hop-based methods. Comparing Figure 4c,d, the correlation coefficients are close, where $\rho = 0.90$ is for hop-based distance, and $\rho = 0.92$ is for DD. The latter figure shows that DD provides improved resolution. Figure 4c indicates that only discrete

integer hop distances exist among node pairs, and Figure 4d shows that physical distances can be mapped to continuous values.

In addition, Figure 4b,d confirm the remark analyzed in Section 4 that DD is approximately proportionally related to physical distance.

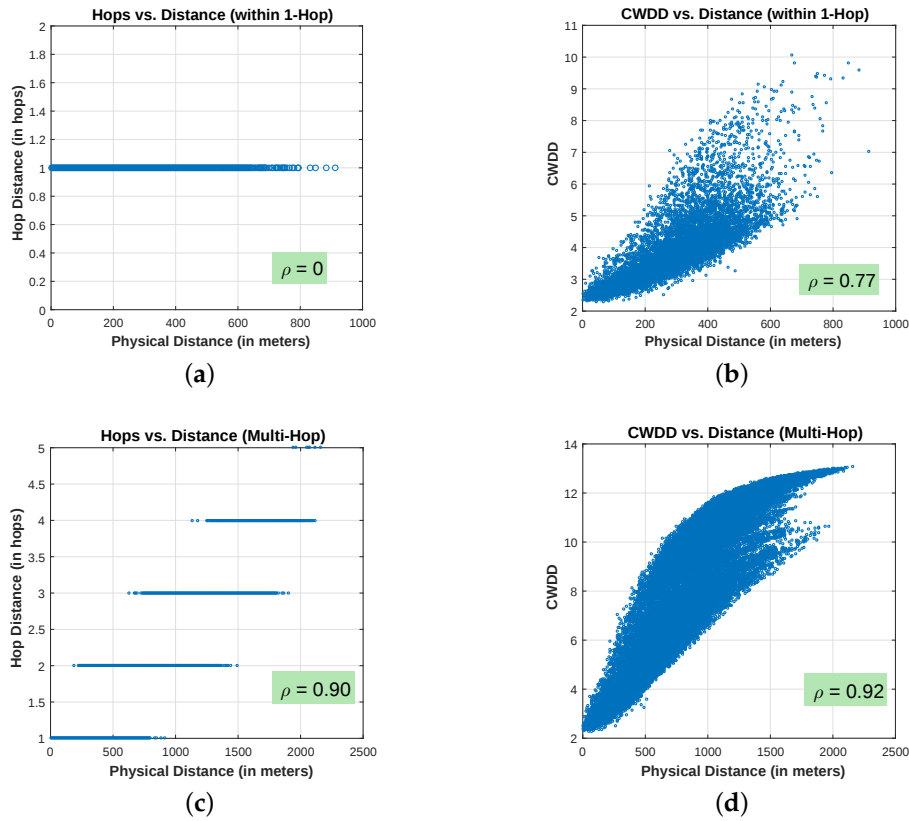


Figure 4. Distance correlation comparison: Hop vs. DD: (a) hops vs. distance (within 1-hop), (b) CWDD vs. distance (within 1-hop), (c) hops vs. distance (multi-hop), (d) CWDD vs. distance (multi-hop).

6.3. DD vs. CWDD

Table 2 is obtained from the simulation with default configurations and compares the localization performances with DD and CWDD serving as a relative distance. Node deviation is the Euclidean distance between expected nodes' coordinate and real nodes' coordinate. CWDD proves localization precision, and it is a better metric of relative distance than DD.

Table 2. Comparison: DD vs. CWDD.

	DD	CWDD
Localization Error (LE)	0.2118	0.186
Maximum Node Deviation	179.1175	144.8799
Minimum Node Deviation	1.1007	1.6635
Mean Node Deviation	27.5344	24.1856

6.4. Localization Performance

This subsection provides the localization results by comparing the images and statistics between two types of localization methods (without and with CWDD embedding).

For improved vision effect, we reduce the scale of the network in this simulation as a $200\text{ m} \times 200\text{ m}$ map. As shown in Figure 5a, 100 nodes (30 red asterisks for the anchors, and 70 blue dots for the nodes to be located) are randomly deployed in this non-uniform network. With the radio range $R = 50$, Figure 5b presents the connectivity of the system. Connected lines are clustered together in the upper right quarter and lower left quarter parts of Figure 5b. In addition, nodes in Figure 5a are considerably clustered around these two areas.

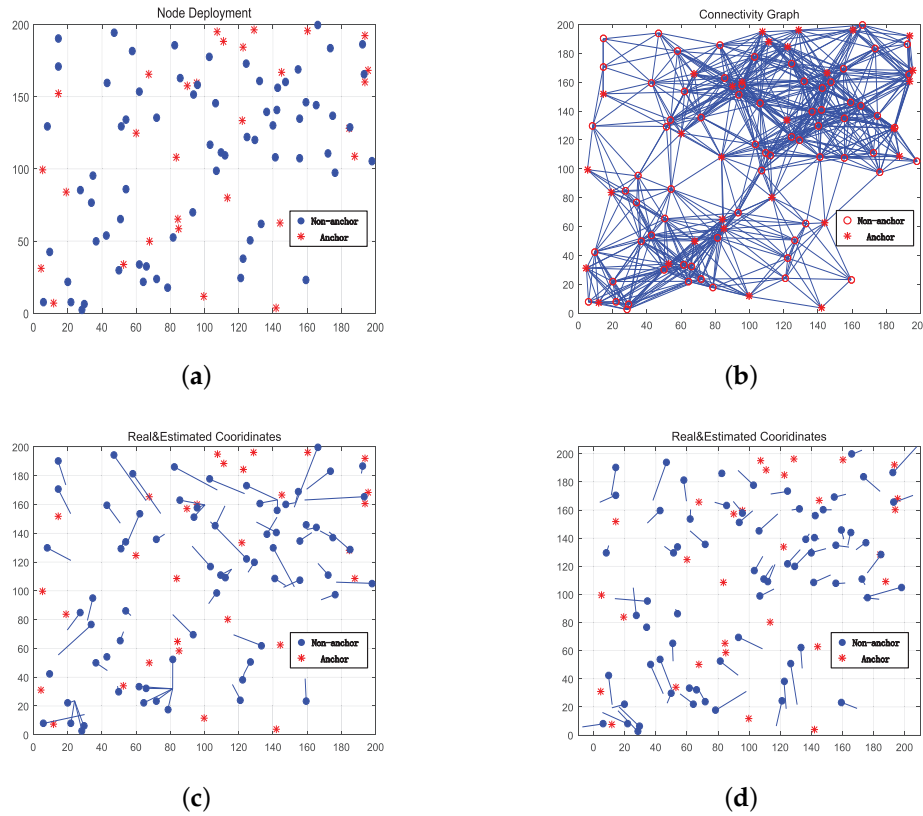


Figure 5. Localization results of DV-Hop and DV-CWDD: (a) node deployment, (b) connectivity graph, (c) DV-Hop, (d) DV-CWDD.

Figure 5c,d display the localization results from DV-Hop and DV-CWDD, respectively. In both figures, the red asterisks are anchors' positions, blue dots are the real positions of unknown nodes, and the blue line segments are directing to the estimated positions. Long segment results in large prediction error. According to Figure 5c,d, it can be shown that:

- (1) DV-CWDD performs better than DV-Hop as most line segments in Figure 5d are shorter than those in Figure 5c.
- (2) DV-Hop may cause clustered mapping. Therefore, some nodes are mapped to the same estimated position. The different distances between node pairs are mapped to the identical hop distance, thereby causing an ambiguity problem. However, this type of issues cannot be observed evidently in DV-CWDD.
- (3) In the upper right quarter of Figure 5d, CWDD obtains higher localization accuracy than that in other areas.

Figure 6 shows the maximum, minimum, and mean deviations from the results of simulations for DV-Hop and DV-CWDD. All statistics in this bar chart are averaged by over 100 runs for high reliability. Three red bars are a third less than blue bars. This finding is a good confirmation that CWDD offers an effective improvement in localization accuracy.

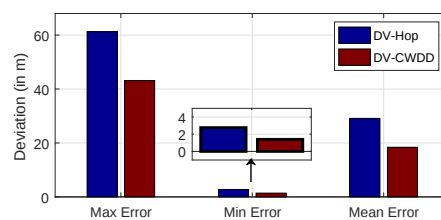


Figure 6. Comparison among maximum, minimum, and mean deviations.

Localization results prove that CWDD method provides an effective resolution for localization precision and ambiguity problem. Moreover, this method performs well in non-uniform networks, especially in dense areas.

6.5. Effectiveness of CWDD

This subsection evaluates the new distance metric by comparing localization errors among hop-based methods (DV-Hop and RPA-Hop), CWDD-embedded methods (DV-CWDD and RPA-CWDD) under different configurations.

6.5.1. Effect of Random-Walk Steps

DD is not constant because of random walk step m . When m is fixed, CWDD is considered a relative distance proportional to real distance. Nevertheless, localization accuracy may change with variable m . Figure 7 shows the localization errors of DV-CWDD varying with m under three system scales (500×500 , 600×600 , and 700×700) with 130 m radio range.

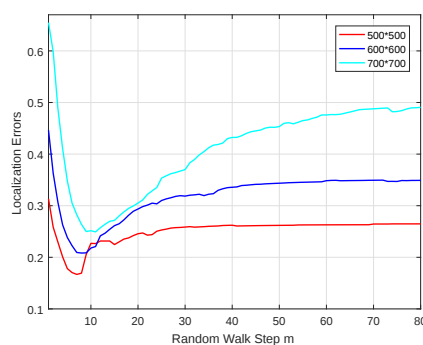


Figure 7. Localization errors with variable m .

As illustrated in Figure 7, each line exhibits a valley near the front part, and the corresponding m is marked as m_{best} . Localization accuracy achieves the most suitable result at m_{best} . Prior to m_{best} , the lines decline when m_{best} lines present the upward trend and stabilize gradually. Numerous simulations concluded that m_{best} is related with the size of networks and system connectivity.

When nodes v_i and v_j are deployed at diagonal vertex of the square area, the farthest distance between a pair of nodes is obtained. The longest path proceeding from v_i to v_j is constant. Therefore, most hops from v_i to v_j in a random walk, which are approximately similar to m_{best} , are likely the longest path divided by radio range. With Algorithm 1, at m_{best} , our design considers the full connectivity of system, including the farthest situation in node deployment. Thus, the accumulated probability matrix may calculate all the node pairs' walking probabilities, thereby maximizing the effectiveness of the new metric for relative distance. Consequently, the system may achieve the most desirable localization performance. This result may explain the reason why the lines in

Figure 7 decline at the beginning. With the increase of m from one to m_{best} , this phenomenon fully utilizes system connectivity information stepwise. However, if m continues to increase, many repeat random walking may be considered into the accumulated probability matrix calculation. In such case, the system may cause an opposite effect in localization performance. The lines also present an upward trend.

Figure 7 also shows that lines go linearly when m becomes large, which confirms our conclusion that DD value may converge to the extreme value as m goes to infinity in Section 4.

6.5.2. Effect of Anchor Density

In this simulation, we increase the anchor numbers from 20 to 100 in steps of 20 in the network, whereas other system configurations remain invariant. From Figure 8, we can see that minimum and median localization errors decline by the increase of anchor numbers for all methods. Therefore, a large number of anchors help improve system accuracy. From the Figure 8, we can see that CWDD-embedded methods show better performance than hop-based methods. The localization errors fall by an average of approximately 20% for DV-Hop and approximately 15% for RPA-Hop by applying CWDD. The maximum reduction of localization error is approximately 26% among all statistics.

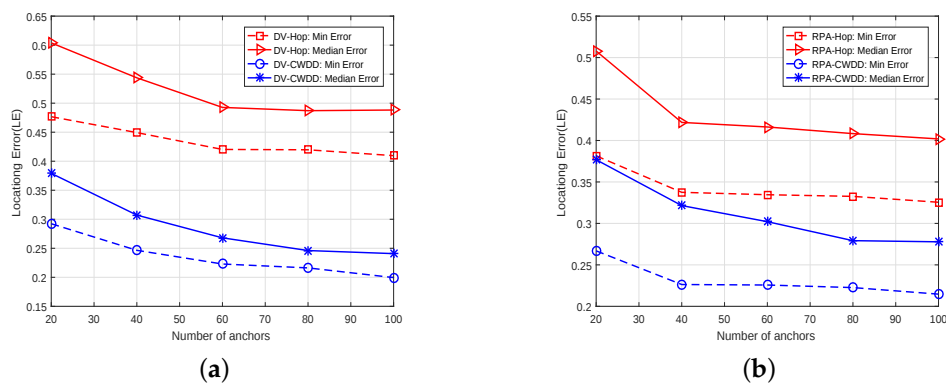


Figure 8. Different numbers of anchors: (a) Dv-based, (b) RPA-based.

6.5.3. Effect of Node Density

In this experiment, we increase the number of nodes from 100 to 300 in steps of 50 without changing any configurations in a network map. Figure 9 shows the localization errors under different node densities for DV-based and RPA-based methods. CWDD-applied methods achieve approximately 22.7% and 15.8% performance gain from DV-Hop and RPA-Hop, respectively. It can be seen from Figure 9 that:

- (1) For DV-Hop and RPA-Hop, the curve presents an increasing trend. That's because the increased node density may cause the ambiguity problem in estimating coordinates, where nodes are mapped to the same positions.
- (2) For CWDD-embedded methods, the accumulated probability matrix based on random-walk steps m can be applied to expand the discrimination between node pairs' relative distance, where the shortest-path hops among node pairs are identical. More neighbors means richer network structure and higher connectivity, which can help to improve the accuracy of CWDD. Thus, the localization performance can be improved.

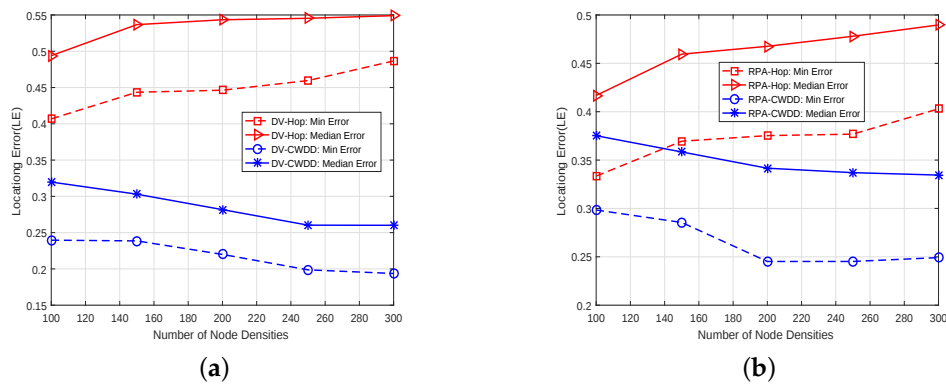


Figure 9. Different node densities: (a) Dv-based, (b) RPA-based.

6.5.4. Effect of System Scale

In this experiment, we expand the dimension of the map from 200 m (in length and width) to 600 m in steps of 100 m. Note that the number of nodes in the network was increased proportionally to maintain the same node density. And we keep the number of anchors constant (20). Figure 10 shows that:

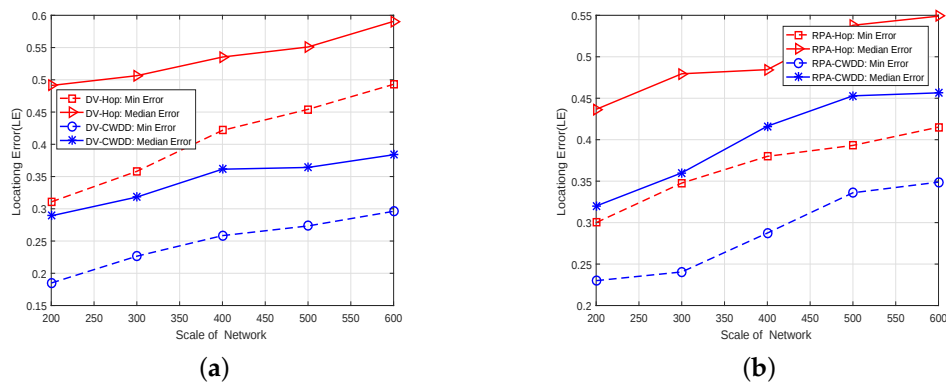


Figure 10. Different system scales: (a) Dv-based, (b) RPA-based.

- (1) CWDD-embedded approaches can achieve higher localization accuracy than that of original methods based on hops.
- (2) With the enlargement of the system's scale, all methods perform unfitting in localization accuracy because the number of anchors is not increased proportionally. Another reason is that the information of neighbor nodes is removed, due to the fixed communication range.

7. Conclusions

This paper mainly proposes a new metric for relative distance and a refined version, which are two range-free methods aiming at proving the node localization accuracy. Connectivity information is optimized, and our DD, which is a value of accumulated probabilities among node pairs in a random walk, is obtained as a new metric for relative distance. For further expansion, we provide a modified version called CWDD by adding proximity information. This new metric can be embedded into two classical hop-based localization algorithms for direct evaluation. Numerous simulations prove that CWDD methods improve system accuracy and provide a resolution for ambiguity problem based on several perspectives.

Acknowledgments: This work was supported by the Fundamental Research Funds for the Central Universities (Grant No. 2017QNA20).

Author Contributions: Pengpeng Chen, Yuqing Yin and Qiang Niu designed the study and wrote the paper; Shouwan Gao performed the simulations; Jun Gu wrote the paper.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Niu, Q.; Huan, T.; Chen, P. NMCT: A novel Monte Carlo-based Tracking algorithm using potential proximity information. *Int. J. Distrib. Sens. Netw.* **2016**, *2016*, 1–10.
2. Niculescu, D.; Nath, B. DV Based Positioning in Ad Hoc Networks. *Telecommun. Syst.* **2003**, *22*, 267–280.
3. Psarakis, Z.; Toumpakaris, D. A sparsity-based algorithm for power-efficient node localization. In Proceedings of the 2016 IEEE Wireless Communications and Networking Conference, Doha, Qatar, 3–6 April 2016; pp. 1–6.
4. Hammoudeh, M. Putting the lab on the map: A wireless sensor network system for border security and surveillance. In Proceedings of the International Conference on Internet of Things and Cloud Computing, Cambridge, UK, 22–23 March 2016; p. 4.
5. Zhang, S.; Gong, J.; Chen, K.T.; Xiao, W.; Ang, V.M.H.; Tham, R.Q.F. Mobile sensing and simultaneously node localization in wireless sensor networks for human motion tracking. In Proceedings of the International Conference on Control, Automation, Robotics and Vision, Icarcv 2010, Singapore, 7–10 December 2010; pp. 367–374.
6. Chen, P.; Ma, H.; Gao, S.; Huang, Y. Modified Extended Kalman Filtering for Tracking with Insufficient and Intermittent Observations. *Math. Probl. Eng.* **2015**, *2015*, 1–9.
7. Niu, Q.; Yang, X.; Gao, S.; Chen, P.; Chan, S. Achieving Passive Localization with Traffic Light Schedules in Urban Road Sensor Networks. *Sensors* **2016**, *16*, 1662.
8. Werner-Allen, G.; Johnson, J.; Ruiz, M.; Lees, J. Monitoring volcanic eruptions with a wireless sensor network. In Proceedings of the Second European Workshop on Wireless Sensor Networks, Istanbul, Turkey, 2 February 2005; pp. 108–120.
9. Cao, M.; Zhang, H.; Park, J.; Daniels, N.M.; Crovella, M.E.; Cowen, L.J.; Hescott, B. Going the Distance for Protein Function Prediction: A New Distance Metric for Protein Interaction Networks. *PLoS ONE* **2013**, *8*, e76339.
10. Yazdani, M.; Popescu-Belis, A. A Random Walk Framework to Compute Textual Semantic Similarity: A Unified Model for Three Benchmark Tasks. In Proceedings of the IEEE Fourth International Conference on Semantic Computing, Pittsburgh, PA, USA, 22–24 September 2010; pp. 424–429.
11. Cao, M. New directions for diffusion-based network prediction of protein function: Incorporating pathways with confidence. *Bioinformatics* **2014**, *30*, 219–227.
12. Savarese, C. Robust Positioning Algorithms for Distributed Ad-Hoc Wireless Sensor Networks. In Proceedings of the General Track of the Conference on Usenix Technical Conference, Monterey, CA, USA, 10–15 June 2002; pp. 317–327.
13. Chowdhury, T.J.S.; Elkin, C.; Devabhaktuni, V.; Rawat, D.B.; Oluoch, J. Advances on localization techniques for wireless sensor networks: A survey. *Comput. Netw.* **2016**, *110*, 284–305.
14. Savvides, A. Dynamic fine-grained localization in Ad-Hoc networks of sensors. In Proceedings of the International Conference on Mobile Computing and Networking, Rome, Italy, 16–21 July 2002; pp. 166–179.
15. Niculescu, D.; Nath, B. Ad hoc positioning system (APS) using AOA. In Proceedings of the IEEE INFOCOM 2003, Twenty-Second Annual Joint Conference of the IEEE Computer and Communications Societies, San Francisco, CA, USA, 30 March–3 April 2003; Volume 3, pp. 1734–1743.
16. Safa, H. A novel localization algorithm for large scale wireless sensor networks. *Comput. Commun.* **2014**, *45*, 32–46.
17. Lasla, N.; Younis, M.F.; Ouadjaout, A.; Badache, N. An Effective Area-Based Localization Algorithm for Wireless Networks. *IEEE Trans. Comput.* **2015**, *64*, 2103–2118.
18. Tomic, S.; Beko, M.; Rui, D. RSS-Based Localization in Wireless Sensor Networks Using Convex Relaxation: Noncooperative and Cooperative Schemes. *IEEE Trans. Veh. Technol.* **2015**, *64*, 2037–2050.

19. Wang, X.; Liu, Y.; Yang, Z.; Lu, K. Robust Component-Based Localization in Sparse Networks. *IEEE Trans. Parallel Distrib. Syst.* **2014**, *25*, 1317–1327.
20. Zaidi, S.; Assaf, A.E.; Affes, S.; Kandil, N. Range-Free Nodes Localization in Mobile Wireless Sensor Networks. In Proceedings of the IEEE International Conference on Ubiquitous Wireless Broadband, Montreal, QC, Canada, 4–7 October 2015; pp. 1–6.
21. Quattrone, A.; Kulik, L.; Tanin, E. Combining range-based and range-free methods: A unified approach for localization. In Proceedings of the Gis'15 Sigspatial International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, 3–6 November 2015; pp. 1–4.
22. Cheng, X.; Shu, H.; Liang, Q.; Du, H.C. Silent Positioning in Underwater Acoustic Sensor Networks. *IEEE Trans. Veh. Technol.* **2008**, *57*, 1756–1766.
23. Whitehouse, K.; Karlof, C.; Culler, D. A practical evaluation of radio signal strength for ranging-based localization. *ACM SIGMOBILE Mob. Comput. Commun. Rev.* **2007**, *11*, 41–52.
24. Lanzisera, S.; Lin, D.T.; Pister, K.S.J. RF Time of Flight Ranging for Wireless Sensor Network Localization. In Proceedings of the 2006 International Workshop on Intelligent Solutions in Embedded Systems, Vienna, Austria, 30–30 June 2006; pp. 759–769.
25. Chen, W.; Li, W.; Shou, H.; Yuan, B. Weighted Centroid Localization Algorithm Based on RSSI for Wireless Sensor Networks. *J. Wuhan Univ. Technol.* **2006**, *30*, 265.
26. Shang, Y.; Ruml, W.; Zhang, Y.; Fromherz, M.P.J. Localization from mere connectivity. In Proceedings of the ACM International Symposium on Mobile Ad Hoc Networking & Computing, Annapolis, MD, USA, 1–3 June 2003; pp. 201–212.
27. Zhong, Z.; He, T. RSD: A Metric for Achieving Range-Free Localization beyond Connectivity. *IEEE Trans. Parallel Distrib. Syst.* **2011**, *22*, 1943–1951.
28. Madras, N. A Note on Diffusion State Distance. *Mathematics* **2015**, arXiv: 1502.07315.
29. Lederer, S.; Wang, Y.; Gao, J. Connectivity-based localization of large-scale sensor networks with complex shape. *ACM Trans. Sens. Netw.* **2009**, *5*, 789–797.
30. Li, M.; Liu, Y. Rendered Path: Range-Free Localization in Anisotropic Sensor Networks With Holes. *IEEE ACM Trans. Netw.* **2010**, *18*, 51–62.
31. Wang, C.; Xiao, L. Locating Sensors in Concave Areas. In Proceedings of the IEEE INFOCOM 2006, IEEE International Conference on Computer Communications, Barcelona, Spain, 23–29 April 2006; pp. 1–12.
32. Costa, J.A.; Patwari, N.; Hero Iii, A.O. Distributed weighted-multidimensional scaling for node localization in sensor networks. *ACM Trans. Sens. Netw.* **2006**, *2*, 39–64.
33. Patwari, N.; Iii, A.O.H.; Costa, J.A. *Learning Sensor Location from Signal Strength and Connectivity*; Springer: New York, NY, USA, 2007; pp. 57–81.



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).