Quantification of Node Significance Based on Overall Connectivity and Relative Position

Waltenegus Dargie, *Senior Member, IEEE*

***Abstract*—In multi-hop wireless networks, such as wireless sensor networks, some nodes are more important than others due to their physical and topological significance, their degree of connectivity, communication load, or simply physical placement. Expressing their significance in a measurable way can be useful for communication protocols to assign particular roles to them. In this paper we propose a model for measuring node significance. The model modifies and extends Katz's *Status Index* which was first proposed to rank the status of people in a social network. In its original form, the *Status Index* is inadequate for our purpose, as it fails to take some of the key features of wireless sensor networks into consideration. Our modification addresses its limitations and puts the significance of each node in its proper context.**

***Index Terms*—Multi-hop communication, node significance, status index, wireless links, wireless sensor networks**

I. Introduction

A wireless sensor network (WSN) consists of a large number of nodes integrating sensing, processing, and wireless communication capabilities [1]. For communication, the nodes share the ISM band [2] with other wireless technologies. As a result, both to reduce interference and utilise energy efficiently, their communication range is typically less than $100\,\mathrm{m}$ [3], [4]. When the network size is large, multi-hop communication is used for packet forwarding.

Maintaining a fully connected network is critical in fulfilling the purpose for which the network is deployed. However, there are some challenges associated with this task. To begin with, the nodes operate with exhaustible batteries and some of them may exhaust their batteries faster than others. These nodes may cause the network to fragment prematurely. Secondly, field deployments show that most wireless links are lossy due to several physical factors [5]. Thirdly, the availability of nodes during packet forwarding is highly dependent on their communication load, which in turn depends on their relative position in the network and their degree of connectivity. If packet forwarding protocols do not take these aspects into account, they may overwhelm critical nodes with requests and a large amount of packets may be dropped. Fourthly, often nodes define duty cycles to save energy, the premise being interesting events occur infrequently and, therefore, nodes should switch off their radios when idle [6], [7]. Sleeping schedules, however, should take the relative significance of nodes into account in order to avoid packet transmission delay. For example, nodes which are placed near a base station should sleep shorter than leaf or edge nodes.

Manuscript first received on 28 April 2020, revised on 19 August 2020.
W. Dargie is with the Technische Universität Dresden, Chair of Computer Networks, 01062, Dresden, Germany.
E-mail: waltenegus.dargie@tu-dresden.de
This work has been partially funded by the German Research Foundation (DFG) in the context of the RoReyBan project (DA 1121/7-1).

In order to address these and similar concerns, it is useful to assign a measure of significance to nodes. This quantity should encode the nodes degree of connectivity and relative position and express how important nodes are, in their presence as well as in their absence. Some of the nodes, when present, are critical for robust message dissemination and packet forwarding. On the other hand, when they fail, they potentially cause the network to fragment prematurely. Hence, different communication protocols may use knowledge of node significance to achieve different goals. For example, a routing protocol may use it to identify nodes which are efficient in disseminating a message or a command. Similarly, a power management protocol may use it to estimate the lifetime of nodes and the rate at which nodes utilise energy.

The purpose of this paper is to propose a measure of significance to the nodes of a wireless sensor network. Our model relies only on a binary adjacency matrix encoding the topology (connectivity) of the network. It modifies and extends Katz's *Status Index* [8] which was first proposed to rank the status of people in a social network. Nevertheless, in its original form, the *Status Index* is inadequate to measure significance, as it fails to take some of the unique features of wireless sensor networks into consideration. The contributions of the proposed model are summarised as follows:

- It ranks nodes according to their relevance (position, degree) in the network.
- Associates rank (significance) with various network assignments.
- Identifies vulnerable or critical nodes which may cause the network to fragment prematurely.

We demonstrate the usefulness of our approach by:

- Discovering reliable routes.
- Defining the sleeping schedules of individual nodes.
- Estimating the degree of network fragmentation when critical nodes fail.

The remaining part of this paper is organised as follows: In Section II, we motivate the usefulness of quantifying the relative significance of nodes. In Section III, we review Related Work. In Section V, we introduce Katz's *Status Index* and highlight the underlying assumptions. In Section VI, we extend the *Status Index* and give justification for doing so. In Section VII, we provide quantitative evaluation using two different networks. In Section VIII we discuss potential applications for the proposed approach. Finally, in Section IX, we provide concluding remarks.

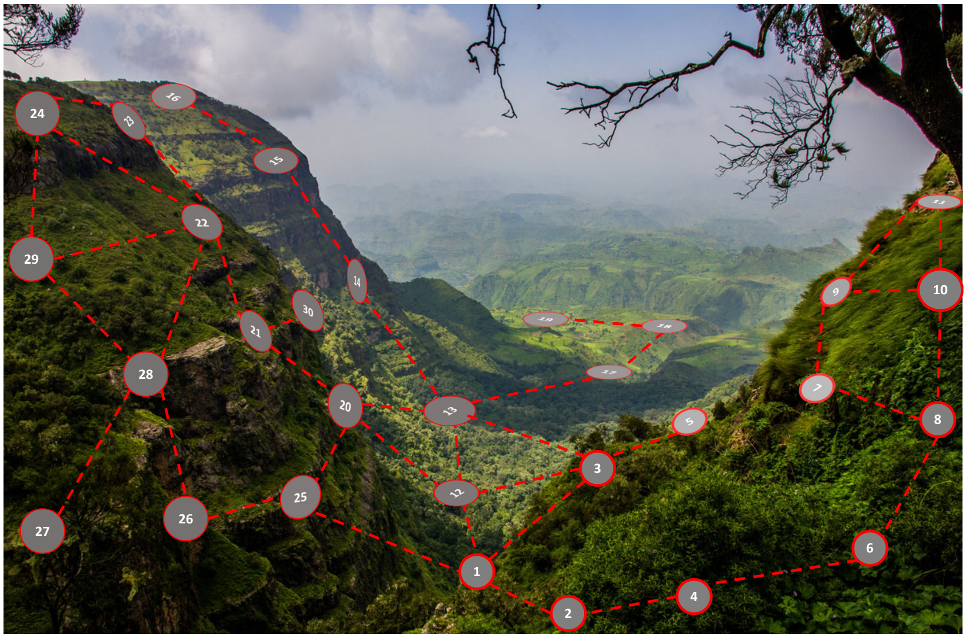

Fig. 1: A wireless sensor network for monitoring wildlife.

II. Motivation

Suppose we are given the network shown in Fig. 1, which consists of wireless cameras monitoring the presence and distribution of certain migrant insects believed to be associated with climate change. The dashed lines indicate direct wireless links and the topology of the network is determined by the transmission range of the nodes as well as the deployment (geographical) setting. The predominant traffic flow is towards a designated base station, even though the base station may occasionally send commands to individual nodes in order to adjust the orientation and field-of-view of the cameras, the sampling rate, and the sampling duration.

Once the network is established, some of the tasks carried out by communication and network management protocols are the following:

1) Identify strategic nodes which efficiently disseminate commands to the rest of the nodes.
2) Locate vulnerable nodes which may cause the network to fragment prematurely.
3) Identify rendezvous nodes which facilitate data collection.
4) Define the sleeping schedule of the nodes.
5) Determine reliable routes.

At first sight, it may appear that, depending on the task at hand, different nodes may be viewed as most suitable. For example, the nodes which are the best candidate for command dissemination are those which are well-connected. In this regard, the node degree can be taken as a measure of significance. On the other hand, the nodes which are vital for keeping the network connected may not necessarily have high degrees. In the figure, node 2 and 13 are equally significant to keep the network connected, but they have different degrees. By contrast, Nodes 3 and 22 have the highest degrees in the network and yet the network remains, by and large, intact if they fail. Similarly, when the predominant traffic flow converges towards a designated base station, the nodes which are highly relevant are those located near the base station. This is because they have to aggregate or forward the packets coming from the leaf nodes. If the location of the base station is not fixed, this means that the significance of the nodes changes. Suppose we deploy a drone to collect data. If the drone hovers above the second mountain on the left side, then node 16 will be the best to assume the role of a base station, in which case, nodes 15, 14, and 13 are vital links to all the other nodes in the network. The situation changes, however, if we decide to hover the drone in the valley, above node 1, in which case, nodes 2,3, 12, and 25 become more relevant than the other nodes.

An interesting research question to ask (and answer) is whether it is possible to assign a measure of significance to the nodes which can be useful for carrying out the above and similar tasks objectively. We answer this question affirmatively in Sections V and VI, but first we shall review state-of-the-art.

III. Related Work

Measuring the significance of nodes is critical in a wide range of networks. In social networks, it is employed to target nodes which are efficient to diffuse information [9]. In psychometry and neurology, it has been proposed to determine causal dependencies and information flow efficiency [10]. In general-purpose computer networks, it has been used to identify nodes which can be the target of attacks and to estimate the magnitude and extent of coordinated attacks [11]. In power-grids, it has been proposed to estimate the magnitude and extent of cascaded failure, both random and systematic [12].

A long list of metrics have been proposed to measure the significance (centrality) of nodes including closeness, betweenness, ties, link importance (weight) [13], [14], [15], [16]; information index [17], lobbying index [18], and cascade depth [12]. Other metrics take the peculiarities of the networks into account [19]. Thus, for example, in neurology, the absence of an edge encodes conditional independence between two nodes [10]. In the evaluation section we shall quantitatively demonstrate that a substantial number of these metrics do not fully quantify the significance of nodes in wireless sensor networks due to the fact that they do not take into consideration the unique features of these networks. In the remaining part of this section, we limit ourselves to the reviewing of Related Work pertaining to wireless sensor networks only.

In wireless sensor networks, knowledge of the significance of nodes has been used for various purposes [20]. Dong et al. [21] propose a data collection (routing) algorithm which aims at guaranteeing reliable event detection and maximising network lifetime. The algorithm takes the placement of nodes (hotspot) into account to decide on their candidacy in a specific data collection assignment. Event detection is maximised by minimising information distortion [22] whereas network lifetime is maximised by maximising the residual energy of nodes participating in the event detection process. The probability of event occurrence in the network and the distance of nodes from the base station (in terms of hop count) determine their likelihood of participating in a data collection assignment.

Dobslaw et al. [23] employ a non-binary adjacency matrix to describe the link quality between any pair of nodes in a wireless sensor network based on which a reliable end-to-end route between a source and a sink is determined. The values assigned to the elements of the matrix correspond to the likelihood of communicating a packet successfully using a direct link. Furthermore, the authors define a column vector expressing the state of the network. The elements of the column vector describe the number of packets at each node waiting in a buffer to be transmitted. Based on these specifications, the authors assign a quality index to each configuration of the network using a constrained optimisation.

Similarly, Silva et al. [24] quantitatively express the reliability of a wireless sensor network as a whole in terms of the reliability and availability of individual nodes as well as the collective reliability of the wireless links. The first two are expressed in terms of the time to failure and the time to recover probability distributions. In order to measure link dependability, the authors employ fault tree analysis [25], [26]. Fault trees are graphical models representing the combination of events that potentially lead to a system failure. Given a topology, the proposed model assigns a probability of dependability to the entire network.

In general, the proposed approaches attempt to characterise the significance or relevance of nodes in the presence of lossy links and unpredictable operation conditions. However, instead of regarding nodes independently, they model relevance as a global aspect, referring to a network as a whole, a given end-to-end route, a pre-defined data gathering assignment, or a system configuration. In the final analysis, however, it is individual nodes which fail, exhaust batteries, or overwrite packets on account of congestion and overflowing buffers. Likewise, it is individual nodes which should aggregate or forward packets or whose configuration (such as duty-cycle) should be adjusted to guarantee high availability. In this paper we complement proposed approaches by providing a mechanism to rank the contribution of nodes to the overall relevance of a wireless sensor network.

IV. Centrality Metrics

Most of the proposed metrics to account for the significance of nodes in a network can be evaluated from two complementary perspectives. The first is the scope of the metrics. Some of them can be regarded as local metrics whereas others are global. The computational cost of the latter is considerably higher than the computational cost of the former but their expression power is also higher. When the nodes are relatively free to make their own decision based on the local dynamics of the network (for example, during route selection), local metrics can be more useful than global metrics. The second perspective is the purpose of the metrics. Specifically, this refers to whether a metric accounts for message propagation efficiency or network resilience.

A. Closeness

Closeness is a measure of the centrality of a node in a network. A node ranking the highest in this regard is the one achieving the smallest hop count on average in establishing routes with all the nodes in the network. It is a global parameter and useful for efficient message dissemination. Mathematically it is expressed as follows:

$$\sigma_i = \frac{N-1}{\sum_{j=1}^{N} d_{ij}} \tag{1}$$

where σ_i is the closeness index of node $i : 1 \leq i \leq N$ and d_{ij} is the hop distance between Nodes i and j, taking the shortest route between i and j.

B. Betweenness

Betweenness is a measure of the significance of a node in maintaining a fully connected network. Hence, it is expressed as:

$$\beta_i = \sum_{i \neq p \neq q} \frac{\Gamma_{pq}(i)}{\Gamma_{pq}} \tag{2}$$

where $\Gamma_{pq}(i)$ refers to the number of shortest routes connecting nodes p and q requiring node i as their intermediate node and Γ_{pq} refers to the overall number of shortest routes connecting nodes p and q. Whether or not two routes are equally short is determined by the number of intermediate nodes they involve in connecting nodes p and $q : 1 \leq p, q \leq N$. For example, in Fig. 1, there are four 5-hop routes connecting Nodes 24 and 1 ($\Gamma_{24,1} = 4$. Ref. to Tab. I). Of these, three of them require Node 22 as their intermediate node (i.e., $\Gamma_{24,1}(22) = 3$). Hence, as far as Nodes 24 and 1 are concerned, Equation 2 for Node 22 yields 3/4.

Route ID	Participating Nodes
1	24-29-28-26-25-1
2	24-22-28-26-25-1
3	24-22-21-20-25-1
4	24-22-21-20-12-1

TABLE I: Four 5-hop routes connecting Nodes 24 and 1 in Fig.1

C. Information Index

The above two metrics consider the shortest routes connecting any pair of nodes in the network. The *Information Index* by contrast considers all possible routes but assigns different weights of significance to them. The basic idea is as follows: Node i may transmit a message to Node j via all the possible routes available to it. As the message propagates, it undergoes attenuation. Hence, the longer the path of propagation, the more it is attenuated and the less significant it becomes to Node j. Hence, the remnant information extracted from a route connecting nodes i and j is given as:

$$q_{ij} = (r_{ii} - r_{jj} - 2r_{ij})^{-1} \tag{3}$$

where r_{ij} are the elements of the matrix:

$$\mathbf{R} = (\mathbf{D} - \mathbf{C} + \mathbf{U})^{-1} \tag{4}$$

where $\mathbf{D}$ is a diagonal matrix consisting of the degree of each node in the network, $\mathbf{C}$ is the adjacency matrix signifying the topology of the network, and $\mathbf{U}$ is an N by N matrix of unit elements. Finally, the *Information Index* of Node i is given as:

$$\gamma_i = \left[\frac{1}{n} \sum_{j=1}^{N} q_{ij} \right]^{-1} \tag{5}$$

The inverse relationships in the above equations signify the attenuation effect of message propagation.

As we shall demonstrate in Section VII, regardless of what they wish to achieve, the above metrics – and, indeed, many of the proposed metrics which we did not explicitly address here – assume that the traffic in the network is omnidirectional (i.e., it originates anywhere and propagates to anywhere), which is not the case in wireless sensor networks. Hence, our aim is to define a *Significance Index* which takes this particular aspect into consideration.

V. Katz's Status Index

Katz represents the relationship between people in a social circle using a binary adjacency matrix $\mathbf{C}$ in which 1 and 0 represent the presence and absence of a direct link, respectively. The diagonal elements of C and all the powers of $\mathbf{C}$ are set to zero or simply ignored. The column sums of $\mathbf{C}$ yield the number of people with which each person is directly linked. Squaring $\mathbf{C}$ yields a square matrix of the same dimension as $\mathbf{C}$[1] expressing the number of two-hop links the people form in the network. This operation is computationally tractable even for a large network, since $c_{ij}^2 = \sum_x c_{ix} c_{xj}$ and each multiplication operation yields either 0 or 1. Similarly, $\mathbf{C}^3$ reveals how the people are interlinked with one another through two intermediate persons, and so on.

[1]The dimension of $\mathbf{C}$ is $N \times N$, where N is the number of people in the network.

In Katz's social circle, a link is not necessarily symmetric (i.e., confidence is not necessarily reciprocated). Assuming that the probability of propagating information between any two directly linked individuals is p, Katz asserts that the summation of all the powers of $p\mathbf{C}$ encodes how well-connected a person is in the social network. Hence:

$$\mathbf{T} = p\mathbf{C} + (p\mathbf{C})^2 + ... + (p\mathbf{C})^k + ... = (\mathbf{I} - p\mathbf{C})^{-1} - \mathbf{I} \tag{6}$$

Note that the summation term is a geometric series. $\mathbf{T}$ is a square matrix and summing its columns yields a row vector revealing the total number of direct and indirect links by which information reaches the people:

$$\mathbf{t}^\intercal = \mathbf{u}^\intercal \left[(\mathbf{I} - p\mathbf{C})^{-1} - \mathbf{I} \right] \tag{7}$$

where $\mathbf{u}$ is a column vector consisting of N unit elements. Multiplying both sides of Equation 7 by $(\mathbf{I} - p\mathbf{C})$ and rearranging terms yields [8]:

$$\mathbf{t} = \left(\frac{1}{p} \mathbf{I} - \mathbf{C}^\intercal \right)^{-1} \mathbf{d} \tag{8}$$

where $\mathbf{d} = \mathbf{u}^\intercal \mathbf{C}$. In other words, $\mathbf{d}$ encodes the number of direct (single-hop) links the people establish with their peers. Equation 8 is not normalised by the maximum number of available links (single as well as multi-hop) given the topology. The normalisation term is proposed by Katz himself, which is given as:

$$m = \sum_{j=1}^{\infty} p^j (N-1)^j \cong (N-1)! p^{N-1} e^{1/p} \tag{9}$$

Consequently, the *Status Index* of each person expressed as an element of a column vector is determined as follows:

$$\mathbf{k} = \frac{1}{m} \mathbf{t} \tag{10}$$

$$\mathbf{C} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & \cdots \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \end{bmatrix} \tag{11}$$

A. The Status Index as a Measure of Node significance

In [27], we argue that there are many parallels between the social networks Katz sets out to model and wireless sensor networks. Hence, it is possible to

- employ an adjacency matrix to represent the topology of a wireless sensor network; and,

- apply the *Status Index* to rank the significance of nodes in the network.

Equation 11 displays a partial adjacency matrix constructed from the topology of the network shown in Fig. 1.

Some of Katz's assumptions, nevertheless, are not applicable to wireless sensor networks and require reconsideration. This is because, in Katz's social circle:

- The nodes' physical location does not have any bearing on their social status whereas this is not the case in wireless sensor networks.
- The probability of two people sharing information with one another is assumed to be dependent on the adjacency matrix. Indeed, Katz's approach breaks in case $1/p$ is not greater than the largest characteristic root of $\mathbf{C}$ – This can easily be verified from Equation 8. In wireless sensor networks, the probability of two nodes communicating with one another at any given time is independent of the topology of the network (i.e., the adjacency matrix).
- Direct links are typically asymmetric whereas in wireless sensor networks this may not be the case. Indeed, most short-range links are rather symmetric.
- All multi-hop links of the same order or length are equally valid; this cannot be asserted in wireless sensor networks.

The last item can be illustrated by considering $\mathbf{c}^3$ for the network shown in Fig. 1. Equation 12 displays a portion of $\mathbf{C}^3$ – the first five rows and five columns:

$$\mathbf{C}^3 = \begin{array}{c} \\ N1 \\ N2 \\ N3 \\ N4 \\ N5 \end{array} \begin{array}{c} \begin{array}{ccccc} N1 & N2 & N3 & N4 & N5 \end{array} \\ \left[\begin{array}{ccccc} 0 & 6 & 2 & 2 & 2 \\ 6 & 0 & 7 & 7 & 2 \\ 2 & 7 & 0 & 5 & 6 \\ 2 & 7 & 5 & 0 & 6 \\ 2 & 2 & 6 & 6 & 0 \end{array} \right] \end{array} \quad (12)$$

Accordingly, there are 5 three-hop routes connecting Node 3 and Node 4 whereas in reality there are none. Nevertheless, the effect of this error is marginal, since the probability that the two neighbour nodes communicate with each other using three-hop routes (or, in general, long routes) is very small (for details, ref. to [27]).

B. Node Placement and In-Network Processing

In a wireless sensor network, there is typically a single node – the base station – to which all the sensed data are destined. The base station occasionally sends commands and queries but the predominant traffic flow is from the child nodes to the base station. The nodes participate in packet forwarding and aggregation. In this respect, the computational and communication burden of the nodes which are placed near the base station is heavier than the burden of those which are farther away. Furthermore, in the multi-hop communication chain, the significance of the role played by each node depends on:

1) its relative position and,
2) the availability of nodes in its proximity which can provide alternative routes in case the node becomes unavailable. In general, nodes with fewer number of direct links may cause the entire network to fragment in case they fail. The nearer they are to the base station, the more critical their failure become.

VI. Significance Index

In order to accommodate the aspects we discussed above, we modify and extend the *Status Index*. Firstly, we normalise the adjacency matrix by $(N-1)$, as this will relax the assumption that p should depend on the topology of the network (refer to Equations 9 and 10):

$$\mathbf{H} = \frac{\mathbf{C}}{(N-1)} \quad (13)$$

Notice that for a fully meshed network, the column sum of $\mathbf{H}$ yields 1. Likewise:

$$\mathbf{H}^k = \frac{\mathbf{C}^k}{(N-1)^k} \quad (14)$$

Thus, the total number of normalized links available to each node can be approximated by:

$$\mathbf{L} \approx \sum_{k=1}^{\infty} (p\mathbf{H})^k = p\mathbf{H}\left(\mathbf{I} - p\mathbf{H}\right)^{-1} \quad (15)$$

The reason why we say "approximated" is that some of the links are invalid in the higher powers of $\mathbf{H}$, as we already pointed out. The error arising in Equation 15 is minimised, however, on account of p, since p^k approaches zero as k becomes large. Indeed, we do not need to worry about node significance unless p is small. If p is large, it means that shorter routes are sufficient to guarantee an end-to-end communication, in which case, $\mathbf{H}^K$ encodes, by and large, the number of valid multi-hop routes. In case p is small, we need longer routes to guarantee communication. At the same time, however, the error in Equation 15 becomes very small. The normalization factor $(N-1)^k$ further suppresses the error in $\mathbf{H}^K$.

If we multiply Equation 15 by a column vector of unit elements, we get the modified *Staus Index*:

$$\mathbf{k}_m = \mathbf{u}^\intercal \mathbf{L} \quad (16)$$

Equation 16 frees the calculation of the significance of nodes from being dependent on the topology of the network. This is because, the explicit normalisation factor, m, in Equation 10 is no longer required. Nevertheless, Equation 16 still suffers from the following shortcomings:

- It considers all direct (single-hop) links to be equally significant whereas this is not the case in wireless sensor networks, where the links near the base station are more significant than the links farther away from the base station and towards the edge of the network.
- It assumes that information propagates in all directions with equal significance and probability. In wireless sensor networks, however, the predominant traffic flow is towards the base station. Hence, the gradient formed towards the base station is the most significant gradient.

In order to encode these two essential aspects, we introduce an additional term to Equation 15. Let h_i be the minimum number of hops the base station needs to communicate with node i. Similarly, let d_i be the degree of node i. The weight of significance we attach to node i according to its placement in the network and its distance from the base station can be expressed as:

$$w_i = \begin{cases} \frac{h_{max}+1-h_i}{h_{max}+1} \cdot \frac{1}{p\ d_i} & d_i > 1 \\ \frac{h_{max}+1-h_i}{h_{max}+1} & d_i = 1 \end{cases} \tag{17}$$

where h_{max} is the number of hops the base station uses in order to communicate with the farthest node in the network, using the shortest route possible. The coefficients $w_i ... w_n$ make up a row vector $\mathbf{w}$.

For $d_i > 1$ (intermediate nodes), the first part of the right-hand term gives more weight to nodes which are near to the base station whilst the second gives more weight to nodes which potentially cause the network to fragment in case of failure. If there are many nodes which are connected to node i (i.e., if $d_i > 2$), it means that there can be as many alternative relay nodes at that distance to forward packets towards the base station, thereby reducing the significance of the node by $1/d_i$ in case it fails or becomes unavailable. Similarly, the smaller p is in the network – notice that p is a global variable[2] –, the less reliable are all links and the more significant is each node's contribution to the overall reliability of the network, and, therefore, more weight should be given to the node. If, on the other hand, p is big, then each link is equally significant and the reward of having multiple links is not so great. For $d_i = 1$ (leaf nodes), we remove the second weight term to exclude leaf nodes from being rewarded for having a small degree. Subsequently, the extended *Status Index* – i.e., our *Significance Index*, is gives as:

$$\mathbf{s} = \mathbf{wL} \tag{18}$$

where $\mathbf{s}$ is a column vector consisting of the *Significance Index* of individual nodes. The advantage of Equation 18 is its ability to capture a node's relationship with its neighbours in a wider sense. A node's significance is determined not merely based on its link with its immediate neighbours but also based on the significance of those with which it is linked, whose significance, in turn, is determined by the significance of those with which they are connected, and so on. Table II displays the list of symbols we used to describe the significance of nodes in a network. **Algorithm 1** summarises the computation of the *Significance Index*.

Symbol	Description
$\mathbf{C}$	A binary matrix describing a topology
N	Number of nodes in a network
p	The probability of establishing a link between two nodes
m	A factorisation term
$\mathbf{T}$	Number of single- and multi-hop links connecting nodes (a matrix)
$\mathbf{d}$	Node degree (a column vector)
$\mathbf{t}$	Nodes overall degree of connectivity (a column vector)
$\mathbf{k}$	Katz *Status Index* (a column vector)
$\mathbf{H}$	A normalised adjacency matrix
$\mathbf{L}$	Normalised number of single- and multi-hop links connecting nodes (a matrix)
$\mathbf{k}_m$	Modified *Status Index* (column vector)
h_{max}	The minimum number of hops connecting the base station with the farthest node in the network
h_i	The minimum number of hops connecting the base station with node i
d_i	The degree of node i
$\mathbf{w}$	Weight of significance (row vector)
$\mathbf{s}$	Significance Index

TABLE II: Notations to describe the significance of nodes in a network

Algorithm 1 Procedure for Computing the Significance Index

procedure SINDEX($\mathbf{C}, N, p, \mathbf{h}, \mathbf{d}$)
 Initialize:
 $w_i \leftarrow 0,\ i = 1, \ldots, N$
 $\mathbf{H} \leftarrow \left(\frac{1}{N-1}\right) \mathbf{C}$
 $\mathbf{L} \leftarrow p\mathbf{H}\left(\mathbf{I} - p\mathbf{H}\right)^{-1}$
 for $i \leftarrow 1$ to N **do**
  if $d_i > 1$ **then**
   $w_i \leftarrow \left(\frac{h_{max}+1-h_i}{h_{max}+1}\right)\left(\frac{1}{pd_i}\right)$
  else if $d_i = 1$ **then**
   $w_i \leftarrow \frac{h_{max}+1-h_i}{h_{max}+1}$
  else
   break
  end if
 end for
 $\mathbf{s} \leftarrow \mathbf{wL}$
 return $\mathbf{s}$
end procedure

VII. Numerical Evaluation

For the numerical evaluation, we take the network of Fig. 1 as a reference throughout this section. The topology of this network is partially expressed by Equation 11. The network has a flat topology, meaning, all nodes play the same roles (sensing and packet forwarding) and Node 1 is the designated base station. Hence, the predominant traffic of the network flows towards this node. To illustrate the usefulness of our approach to hierarchical topologies, we shall consider another topology in Subsection VIII.

[2]This is based on the assumption that the deployment setting affects all nodes equally and that the wireless channels are, statistically speaking, independent.

A. Node Ranking

For a small network and a small p, Katz's *Status Index* and its modified version rank the significance of nodes in the same way (as can be seen in Fig. 2 (top)). The modified version, as we indicated in Equation 15, frees the *Status Index* from being dependent on the magnitude of p. When p is large, the *Status Index* breaks even for a small network. Similarly, for a small network, the modified *Status Index* and the *Significance Index* rank nodes, by and large, in the same way, regardless of the value of p. This is because the impact of $\mathbf{w}$ on $\mathbf{s}$ in Equation 18 is not appreciable when h_{max} is small.

Fig.2 (bottom) compares the ranking of nodes using the modified *Status Index* and the *Significance Index* for $p = 0.7$. Even though the plots appear similar at first sight, there is a subtle difference. For example, the modified *Status Index* gives

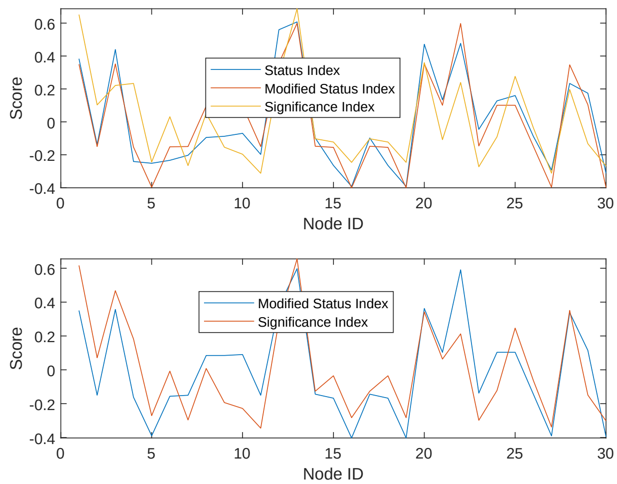

Fig. 2: Ranking nodes according to their significance using Katz's original *Status Index*, the modified *Status Index*, and the *Significance Index*. Top: $p = 0.2$. Bottom: $p = 0.7$

Rank	k	k$_m$	s	Energy
1	13	13	13	25
2	12	22	20	2
3	22	12	12	13
4	20	20	25	12
5	3	3	22	20

TABLE III: The top five significant nodes for the network of Fig. 1

higher scores to Nodes 9, 10, and 22 compared to the scores they are given by the *Significance Index*. By contrast, the *Significance Index* gives higher scores to Nodes 2, 6, 15, 18, and 25 compared to the scores they are given by the modified *Status Index*. These differences arise from the fact that the *Significance Index* gives emphasis to critical nodes whereas the modified *Status Index* to nodes' degree of connection. In summary, for a large p and a large network, the *Significance Index* modifies the *Status Index* in three ways:

- Nodes are not regarded as significant simply because they are well-connected.
- Critical nodes are identified as significant.
- Leaf nodes are downgraded even though they have relatively high node degree.

B. Energy Consumption

The relative significance of the nodes can be determined by closely examining their energy consumption characteristics. In order to carry out this task, we model energy consumption as follows. Each node generates a packet every second using a Poisson distribution and sends it to its immediate neighbour, setting the base station as the destination node. The size of a packet is 28 Byte and packet transmission is unicast. Intermediate nodes forward the packet without modifying it. In other words, they do not perform in-network processing (such as data aggregation or filtering). In addition, intermediate nodes transmit the packets they themselves generate. An intermediate node makes up the shortest path connecting

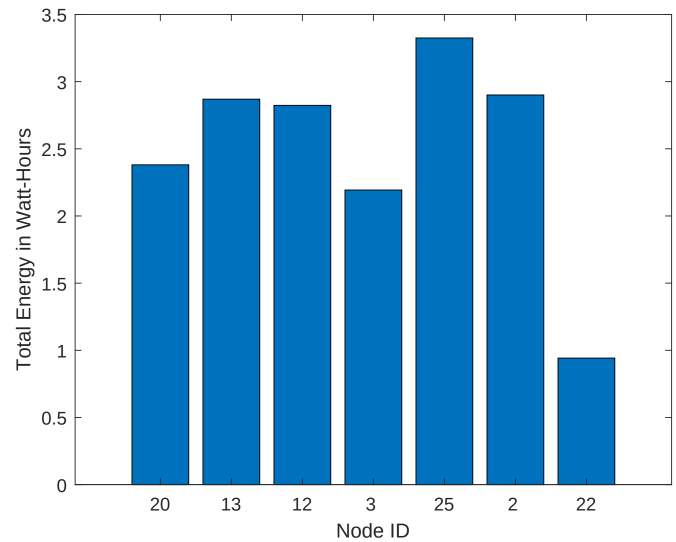

Fig. 3: The overall energy consumption (in Watts-Hours) of the 7 most critical nodes.

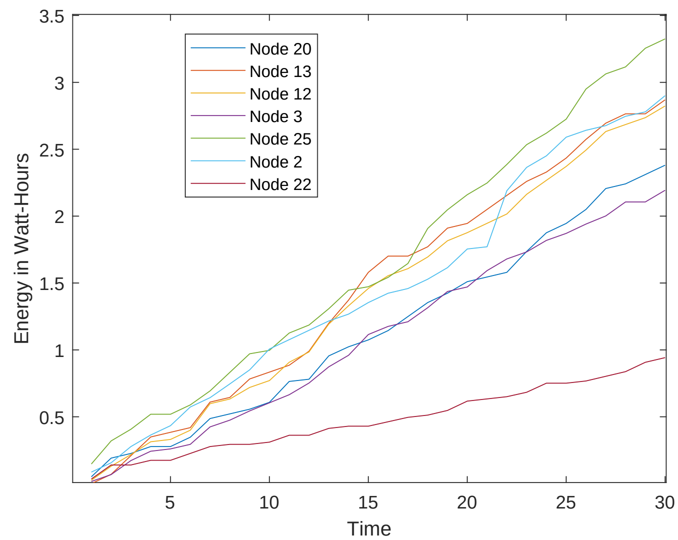

Fig. 4: The energy depletion profile (in Watts-Hours) of the 7 most critical nodes.

a source node with the base station. The cost of packet transmission and receiving is taken from the CC2420 datasheet[3]. Accordingly, a node consumes 18.8 mA when receiving and 17.4 mA when transmitting. The transceiver's supply voltage can be varied between 2.1 and 3.6 V. We set it to 3.6 V. The nominal transmission rate of the CC2420 radio is 240 Kbps. Hence, the minimum amount of energy a node consumes to receive a single packet from a neighbour is: $(28 \times 8 \times 18.8)/(240 \times 10^6) = 17.5\ \mu W\ s$. Likewise, the minimum packet transmission cost is $16\ \mu W\ s$.

The energy consumption characteristic can be evaluated in two complementary ways. Firstly, we can evaluate the amount of energy a node consumes in a given time interval. This is

[3]https://www.ti.com/lit/ds/symlink/cc2420.pdf?ts=1591251841135 (last access on 04 June 2020, 08:24 o'clock CET).

useful for estimating its energy reserve (lifetime). Secondly, we can evaluate the rate at which a node exhausts its energy. This can be useful for routing protocols to make decision whether the participation of a node in a routing assignment leads to a premature network partitioning. If, for instance, a critical node exhausts its energy reserve at a high rate, it is likely that a part of the network partitions prematurely. In both cases, the higher the amount, the more significant a node is as its peers rely on it to forward packets to the base station.

Fig. 3 shows the 30 minute energy consumption of the top 7 nodes for the deployment of Fig. 1. From their energy consumption we can deduce their communication burden. The energy profiles of Node 12, 13, and 20 are consistent with their *Status Index*, *Modified Status Index* and *Significance Index*. When it comes to Node 3, 22, and 25, however, the ranking strategies produce different results. Katz's *Status Index* identifies Node 22 and Node 3 as the third and fifth most significant nodes, respectively. The *Modified Status Index* ranks them as the second and the fifth most significant nodes, respectively. These aspects are not reflected in their energy profile (they are seventh and sixth, respectively). The *Significance Index*, on the other hand, ranks Node 25 as the fourth most significant node and does not rank Node 3 as one of the five most significant nodes.

An interesting aspect which has not been captured by any of the ranking strategies is the case of Node 2, which has consumed a significantly large amount of energy. The significance of this node can be confirmed by a visual inspection, for if this node fails, the entire right wing of the network would fragment. None of the ranking strategies was able to foresee this, because the node is not well connected. Similarly, according to the energy profile, Node 25 appears to be the most significant node. Even though its *Significance Index* suggests the significance of Node 25, it is, nonetheless, seen as the fourth most significant node, not the most significant node.

Fig. 4 shows the rate at which the 7 most significant nodes utilised energy. This figure highlights the burden of Node 25. Whereas Node 12 and Node 13, which are well-connected and near to the base station, have other nodes which can alleviate their packet forwarding burden, Node 25 does not have. As a result, even if its node degree is less than that of Node 12 and Node 13, its energy consumption is significant.

C. Network Partitioning

One way to demonstrate the significance of nodes in a network is to remove the most significant nodes and observe how the network survives. We can characterise the fragment in two ways, namely, in terms of:

- the number of "islands" created; and,
- the population of each island.

The bigger the number of islands and the fewer the relative population of the islands, the worst is the fragmentation. Using these two concepts, we can evaluate how well a given approach measures the relevance of nodes in a wireless sensor network.

Rank	**Closeness** (σ)	**Betweenness** (β)	**I** (γ)	**s**
1	12	2	13	25
2	20	13	24	2
3	13	4	6	13
4	25	20	3	12
5	2	6	12	20

TABLE IV: The top five significant nodes. **I**: Information Index. **s**: Status Index.

Route ID	$\mathbf{k}_m$	**s**	γ
R1	Node 26 (-0.14)	Node 29 (-0.15)	Node 25 (0.06)
R2	Node 26 (-0.14)	Node 26 (-0.06)	Node 25 (0.06)
R3	Node 21 (0.10)	Node 21 (0.06)	Node 21 (0.01)
R4	Node 21 (0.10)	Node 21 (0.06)	Node 21 (0.01)
`max(min)`	**R3**, **R4**	**R3**, **R4**	**R1**, **R1**

TABLE V: Vulnerable intermediate nodes and their normalised (between -0.5 and 0.5) scores.

Hence, given a significance metric, its resilience score can be expressed as:

$$\Phi_j(x) = \frac{1}{K} \sum_{k=1}^{K} \frac{p_k}{N} \tag{19}$$

where $\Phi_j(x)$ refers to the resilience score of metric x when j of the most significant nodes are removed from the network; K is the number of islands after fragmentation, N is the number of Nodes in the fully connected network, and p_k is the population of the k-th island. A fully connected network has a resilience score of 1. The smaller Φ is, the worst is the fragmentation of the network as a result of the failure of its j most significant nodes.

In order to evaluate the expression power of our significance metric in this regard, we implemented the centrality metrics we discussed in Section IV – *closeness*, *betweenness*, and *information index* – for comparison. Tab. IV lists the top most significant nodes identified by these metrics in the network of Fig. 1. If we remove these nodes from the network, five islands will be formed in the case of *closeness* and *Information Index*; four islands, in the case of *betweenness*, and six islands, in the case of *Status Index* (ref. to fig. 5). The corresponding resilience scores are: $\Phi_5(\sigma) = 0.17$, $\Phi_5(\beta) = 0.2$, $\Phi_5(\gamma) = 0.17$, $\Phi_5(S) = 0.14$. Subsequently, both in terms of the number of islands and the resilience score, our metrics correctly predicted the most relevant nodes in the network.

D. Reliable Route

In wireless networks different metrics are used to discover efficient routes. One of these evaluates the minimum amount of available power of eligible nodes [1] using the `maximum(minimum(` **available power** `))` operation. It works as follows: The nodes with the minimum available power will be selected from each potential route and the route with the highest minimum available power will be selected as a winner. The goal is to ensure that nodes are exhausting their energy evenly. One can take a similar approach to identify the most reliable route. Thus, vulnerable nodes are selected from all eligible routes and the route with the least vulnerable node will be selected.

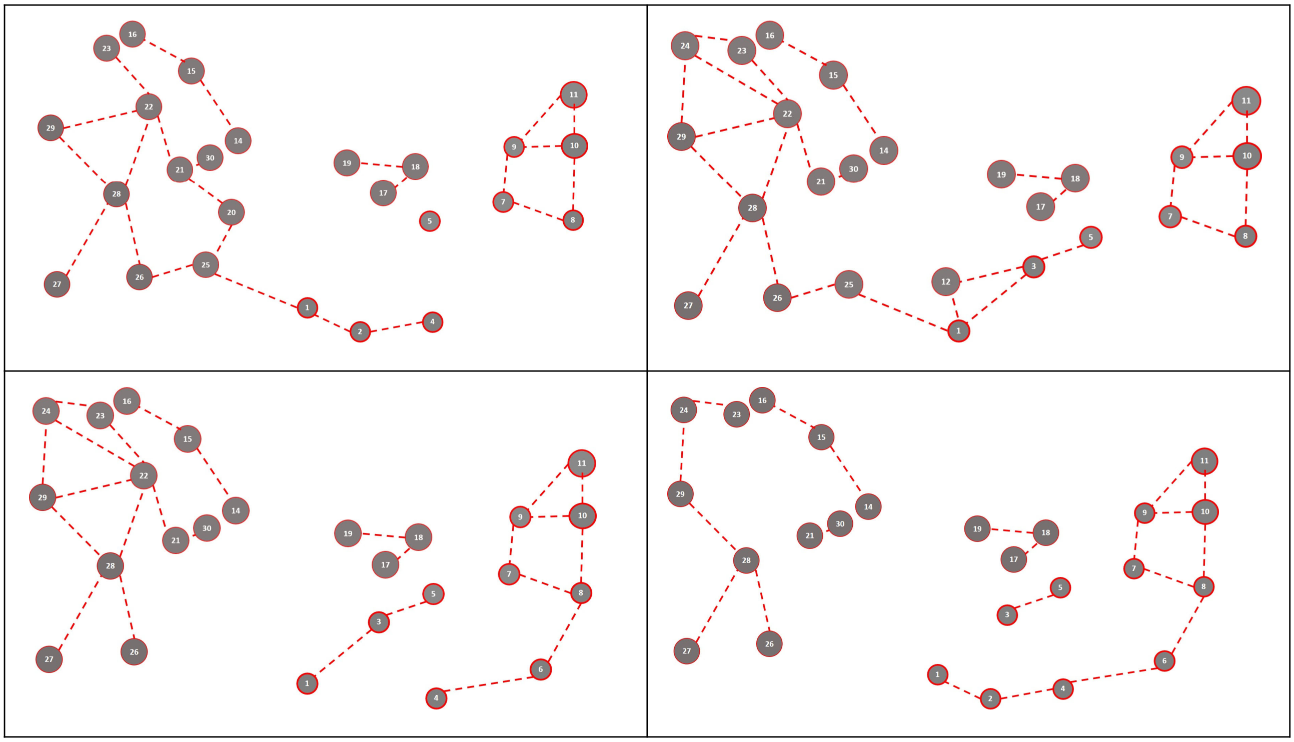

Fig. 5: Network fragmentation after the five most relevant nodes were removed. Top left: *Closeness*. Top right: *Betweenness*. Bottom left: *Information Index*. Bottom right: *Status Index*.

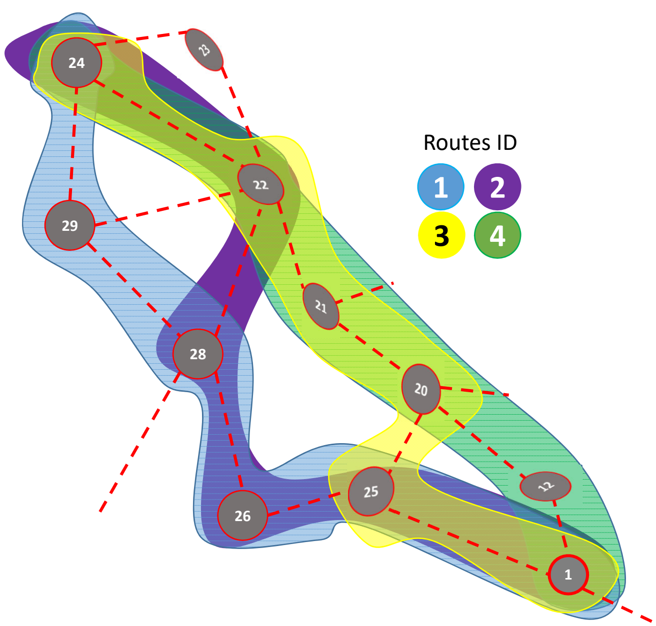

Fig. 6: Determining the most reliable route using a significance metric.

To illustrate, suppose we wish to establish the most reliable route connecting Node 1 and 24 in Fig. 1. As we have seen in Section IV, there are four potential routes having the same hop count (ref. to Fig. 6). Which of them is the most reliable route? If we rank the intermediate nodes according to their significance, we can answer this question objectively. For comparison, we employ the *Information Index*, the modified *Status Index*, and the *Significance Index* to rank the nodes. Table V lists the most vulnerable intermediate nodes in each route along with their indices. The bottom row identifies the route(s) containing the least vulnerable intermediate node. Thus, the *Information Index* identifies **R1** and **R2** as the most reliable routes whereas the modified *Status Index* and the *Significance Index* identify **R4** and **R3** as the most reliable routes. As can be seen, Nodes 26 and 29, which make up **R1** and **R2**, are leaf (outer) nodes and it is appropriate that the modified *Status Index* and the *Significance Index* exclude them.

E. Duty-Cycle

Another advantage of ranking nodes is the possibility of specifying their duty-cycle according to their significance. A duty cycle specifies the portion of time a node is active [1], [28] and is expressed as:

$$D = \frac{t_a}{t_a + t_s} \times 100\% \tag{20}$$

where D is the duty cycle in percent, t_a is the active duration, t_s is the sleep duration, and $T = t_a + t_s$ is the period. If, for example, one of the leaf nodes has a duty-cycle $D_l = 10\,\%$, the duty-cycle of all the remaining nodes can be expressed in terms of their relative significance:

$$D_i = D_l \times \frac{r_i}{r_l} \tag{21}$$

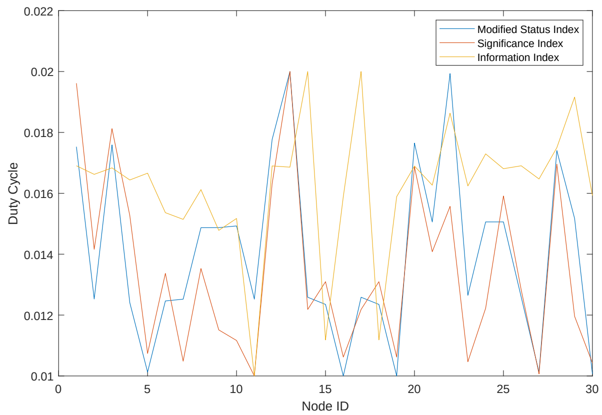

Fig. 7: Defining duty cycle using the *Status Index*, the modified *Status Index*, and the *Information Index*.

where r_i and r_l refer to the ranks of node i and the leaf node, respectively. Fig. 7 compares the duty cycle of nodes we computed using the *Information Index*, the modified *Status Index*, and the *Significance Index*. A closer examination of the duty cycles of the individual nodes reveals that the *Information Index* is not suitable for this task. There is no apparent correlation between the duty cycle it assigns to the nodes and their actual significance in the network. The modified *Status Index* and the *Significance Index* agreed in their assignment most of the time, however, the latter moderates the appropriation of active time to nodes which are far away from the base station even though they are well-connected. Examples are: Nodes 8, 9, 22. On the other hand, critical nodes near the base station are assigned relatively longer active durations. Examples are Nodes 6 and 25.

VIII. Scalability and Application

It is legitimate to ask at this point (1) whether the approach scales and (2) what additional applications it has. The answer to the first question is affirmative as long as the adjacency matrix (which is positive-definite) is invertible. The computational complexity (i.e., the time complexity) associated with solving Equation 15 is in the neighbourhood of $\mathcal{O}^{2.3}$ [29], on account of the inverse matrix operation.

In order to demonstrate the scalability of our approach, consider Fig. 8, where we have a hierarchical-topology wireless sensor network consisting of a base station, three cluster heads, and 125 child nodes. The topology forms a polar coordinate. For this network, the original *Status Index* breaks not only on account of p but also of m (ref. to Equation 9), which becomes inordinately large and dwarfs the contribution of all the other terms. Interestingly, for a well-structured topology like this, the *Modified Status Index* and the *Significance Index* rank the nodes in a similar fashion. This is because, the position of each node and its contribution to the connectivity of the entire network is implicitly encoded in the topology itself, without the need to explicate it with Equation 17. Fig. 9 displays the significance indices. Interestingly, both approaches have accurately reflected the unique role border nodes (such as nodes: 7, 35, 48, 60, 69, and 96) play in maintaining connectivity.

IX. Conclusion and Future Work

In this paper we extended Katz's *Status Index* to measure the significance of nodes in a wireless sensor network. We interpreted significance as the contribution of individual nodes to the connectivity or reliability of the network. Originally proposed to rank the status of people in a social circle, the *Status Index* assigns a quantitative value to each node in a network according to the strength of its links with its peers, taking both direct and all possible indirect links into account. However, it does not take into consideration the physical placement of nodes in the network and the role they play in maintaining unbroken multi-hop links. As these aspects are essential for wireless sensor networks (wherein the flow of traffic is predominately from child nodes to a single base station), we introduced a new normalisation term to contextualise the *Status Index*.

Our approach assumes that the network's topology does not change or changes slowly over time. This assumption justifies the cost associated with the construction and evaluation of the adjacency matrix. When the network topology is dynamic, this cost becomes appreciable. Nevertheless, this cost is not unique to our approach. Indeed, most of the proposed approaches, besides relying on some form of adjacency matrices, take into consideration additional metrics to increase the expression power of their model, which have to be updated whenever the topology of the underlying networks and the roles played by the nodes change.

From the way the *Significance Index* ranks the nodes, it is apparent that the second term of the right-hand side in Equation 18 may become dominant in some network topologies. In which case, a fine-tuning should be made, so that each term – i.e., a node's relationship with its peers, a node's placement in the network, and the presence of alternative relay nodes in the hop chain – gets its proper significance. This is particularly the case when the network's size is appreciably large (in which case, h_{max} becomes disproportionately large, thereby forcing the equation to overemphasize the placement of nodes in the network).

In future, our focus will be on exploring additional application areas for the *Status Index*. For example, in Figure 8 we computed the *Significance Index* of the nodes given the structure of the network. An interesting question to address will be: Given the *Significance Index* of the nodes and a flat network, is it possible to (1) dynamically cluster the network, (2) identify (or designate) Nodes **BS**, **c1**, **c2**, **c3**, as cluster heads, and (3) associate child nodes with these nodes?

References

[1] W. Dargie and C. Poellabauer, *Fundamentals of wireless sensor networks: theory and practice*. John Wiley & Sons, 2010.

[2] A. F. Molisch, K. Balakrishnan, C.-C. Chong, S. Emami, A. Fort, J. Karedal, J. Kunisch, H. Schantz, U. Schuster, and K. Siwiak, "Ieee 802.15. 4a channel model-final report," *IEEE P802*, vol. 15, no. 04, p. 0662, 2004.

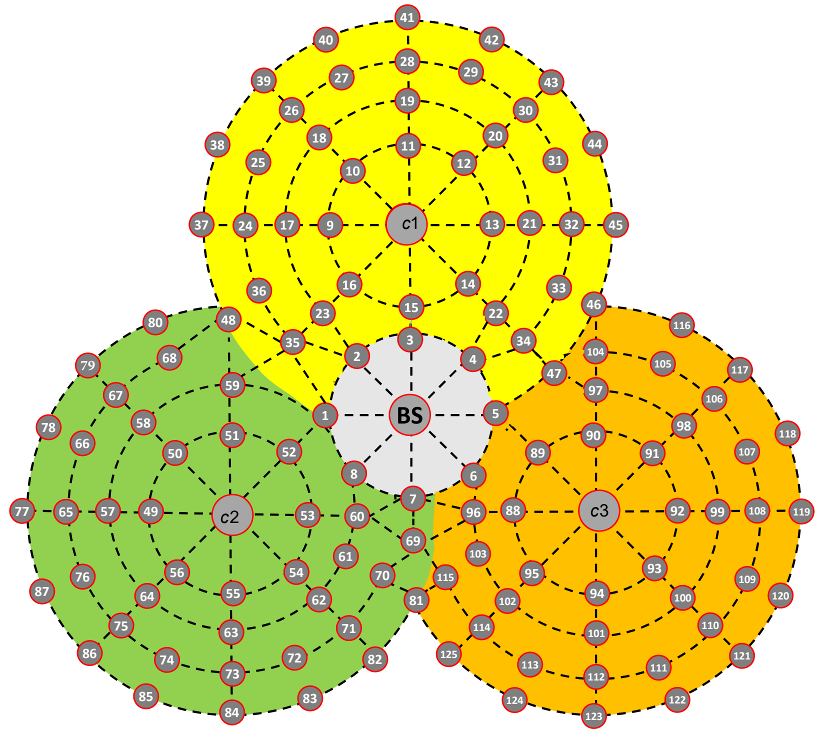

Fig. 8: A cluster-based wireless sensor network deployment.

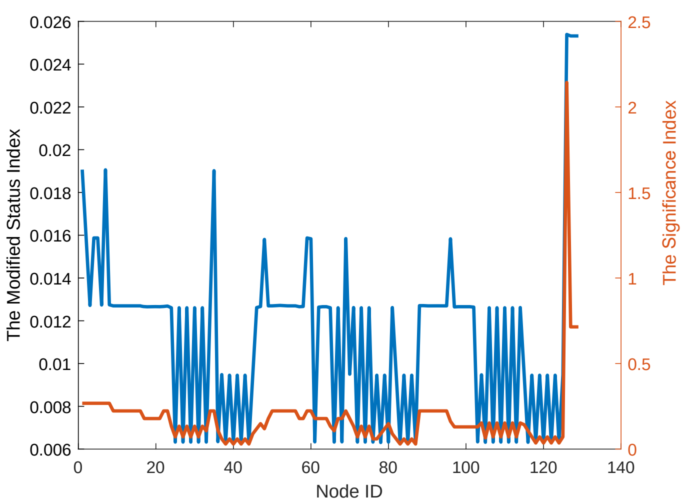

Fig. 9: Ranking the nodes in Fig. 8 according to their significance using the *Modified Status Index* and the *Significance Index*.

[3] I. Kuzminykh, A. Snihurov, and A. Carlsson, "Testing of communication range in zigbee technology," in *2017 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM)*, pp. 133–136, 2017.

[4] C. Park and T. S. Rappaport, "Short-range wireless communications for next-generation networks: Uwb, 60 ghz millimeter-wave wpan, and zigbee," *IEEE Wireless Communications*, vol. 14, no. 4, pp. 70–78, 2007.

[5] D. Dina and K. L. Downie, "Wireless communication network sensor information for control of industrial equipment in harsh environments," Feb. 23 2016. US Patent 9,270,520.

[6] Y. Qiu, S. Li, X. Xu, and Z. Li, "Talk more listen less: Energy-efficient neighbor discovery in wireless sensor networks," in *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pp. 1–9, IEEE, 2016.

[7] Y. Sun, O. Gurewitz, and D. B. Johnson, "Ri-mac: a receiver-initiated asynchronous duty cycle mac protocol for dynamic traffic loads in wireless sensor networks," in *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pp. 1–14, 2008.

[8] L. Katz, "A new status index derived from sociometric analysis," *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.

[9] K. Kandhway and J. Kuri, "Using node centrality and optimal control to maximize information diffusion in social networks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 47, no. 7, pp. 1099–1110, 2016.

[10] F. Dablander and M. Hinne, "Node centrality measures are a poor substitute for causal inference," *Scientific reports*, vol. 9, no. 1, pp. 1–13, 2019.

[11] T. Nie, Z. Guo, K. Zhao, and Z.-M. Lu, "Using mapping entropy to identify node centrality in complex networks," *Physica A: Statistical Mechanics and its Applications*, vol. 453, pp. 290–297, 2016.

[12] R. Ghanbari, M. Jalili, and X. Yu, "Correlation of cascade failures and centrality measures in complex networks," *Future generation computer systems*, vol. 83, pp. 390–400, 2018.

[13] L. Freeman, "The development of social network analysis," *A Study in the Sociology of Science*, vol. 1, 2004.

[14] T. Opsahl, F. Agneessens, and J. Skvoretz, "Node centrality in weighted networks: Generalizing degree and shortest paths," *Social networks*, vol. 32, no. 3, pp. 245–251, 2010.

[15] H. Kim and R. Anderson, "Temporal node centrality in complex networks," *Physical Review E*, vol. 85, no. 2, p. 026107, 2012.

[16] A. M. M. González, B. Dalsgaard, and J. M. Olesen, "Centrality measures and the importance of generalist species in pollination networks," *Ecological complexity*, vol. 7, no. 1, pp. 36–43, 2010.

[17] K. Stephenson and M. Zelen, "Rethinking centrality: Methods and examples," *Social networks*, vol. 11, no. 1, pp. 1–37, 1989.

[18] A. Abbasi, "h-type hybrid centrality measures for weighted networks," *Scientometrics*, vol. 96, no. 2, pp. 633–640, 2013.

[19] D. Taylor, S. A. Myers, A. Clauset, M. A. Porter, and P. J. Mucha, "Eigenvector-based centrality measures for temporal networks," *Multiscale Modeling & Simulation*, vol. 15, no. 1, pp. 537–574, 2017.

[20] W. Dargie and J. Wen, "A simple clustering strategy for wireless sensor networks," *IEEE Sensors Letters*, 2020.

[21] M. Dong, K. Ota, and A. Liu, "Rmer: Reliable and energy-efficient data collection for large-scale wireless sensor networks," *IEEE Internet of Things Journal*, vol. 3, no. 4, pp. 511–519, 2016.

[22] F. Bouabdallah, N. Bouabdallah, and R. Boutaba, “Efficient reporting node selection-based mac protocol for wireless sensor networks,” *Wireless networks*, vol. 19, no. 3, pp. 373–391, 2013.

[23] F. Dobslaw, T. Zhang, and M. Gidlund, “End-to-end reliability-aware scheduling for wireless sensor networks,” *IEEE Transactions on Industrial Informatics*, vol. 12, pp. 758–767, April 2016.

[24] I. Silva, L. A. Guedes, P. Portugal, and F. Vasques, “Reliability and availability evaluation of wireless sensor networks for industrial applications,” *Sensors*, vol. 12, no. 1, pp. 806–838, 2012.

[25] E. Ruijters and M. Stoelinga, “Fault tree analysis: A survey of the state-of-the-art in modeling, analysis and tools,” *Computer science review*, vol. 15, pp. 29–62, 2015.

[26] C. A. Ericson, “Fault tree analysis,” in *System Safety Conference, Orlando, Florida*, vol. 1, pp. 1–9, 1999.

[27] W. Dargie, “A quantitative measure of reliability for wireless sensor networks,” *IEEE Sensors Letters*, vol. 3, pp. 1–4, Aug 2019.

[28] W. Dargie, “Dynamic power management in wireless sensor networks: State-of-the-art,” *IEEE Sensors Journal*, vol. 12, no. 5, pp. 1518–1528, 2011.

[29] N. Karmarkar, “A new polynomial-time algorithm for linear programming,” in *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pp. 302–311, ACM, 1984.