

# Some $k$ -hop based graph metrics and node ranking in wireless sensor networks\*

Csaba Biró, Gábor Kusper

Eszterházy Károly University, Eger, Hungary

[biro.csaba@uni-eszterhazy.hu](mailto:biro.csaba@uni-eszterhazy.hu)

[kusper.gabor@uni-eszterhazy.hu](mailto:kusper.gabor@uni-eszterhazy.hu)

*Submitted: December 14, 2017*

*Accepted: September 13, 2018*

*Published online: February 27, 2019*

## Abstract

Node localization and ranking is an essential issue in wireless sensor networks (WSNs). We model WSNs by communication graphs. In our interpretation a communication graph can be directed, in case of heterogeneous sensor nodes, or undirected, in case of homogeneous sensor nodes, and must be strongly connected. There are many metrics to characterize networks, most of them are either global ones or local ones. The local ones consider only the immediate neighbors of the observed nodes. We are not aware of a metric which considers a subgraph, i.e., which is between global and local ones. So our main goal was to construct metrics that interpret the local properties of the nodes in a wider environment. For example, how dense the environment of the given node, or in which extent it can be relieved within its environment. In this article we introduce several novel  $k$ -hop based density and redundancy metrics: Weighted Communication Graph Density ( $WCGD$ ), Relative Communication Graph Density ( $RCGD$ ), Weighted Relative Communication Graph Density ( $WRCGD$ ), Communication Graph Redundancy ( $CGR$ ), Weighted Communication Graph Redundancy ( $WCGR$ ). We compare them to known graph metrics, and show that they can be used for node ranking.

*Keywords:*  $k$ -hop based graph metrics, communication graph density, com-

---

\*This research was supported by the grant EFOP-3.6.1-16-2016-00001 “Complex improvement of research capacities and services at Eszterházy Károly University”

munication graph redundancy, node ranking

MSC: 68M07 68M17 68M12 68M15 68R10 05C05 05C12 05C40 05C82 90B10

## 1. Introduction

The modeling and analysis of complex networks is an important interdisciplinary field of science. The networks are mathematically related to graph theory. It is known that topology represents the properties of the whole network structure. A topology describes a real network (*with constraints*) and it can be converted to an undirected or directed graph. The common property of topological models is that they are usually calculated based on probabilities [2–4, 11]. The objects of the model can be matched by the vertices of the graph. Edges can be used to describe the relations between the objects. Graph-based modeling can be of two types: ad-hoc or measurement-based. On large wireless networks the traditional measurements based procedures can not be applied efficiently, but  $k$ -hop based approaches can be computed effectively also for large networks.

There are many graph-based metrics for modeling complex networks [9]. Topological metrics commonly used on networks: number of nodes and edges, average degree, degree distribution, connectedness, diameter, number of independent paths. Parameters for measuring the effectiveness of wireless networks: scope and coverage, scalability, expected transmission number, hop count (*number of hops*), power consumption / lifetime.

In graph theory, the density of a graph  $(\mathcal{V}; \mathcal{E})$  can be calculated as  $\frac{|\mathcal{E}|}{|\mathcal{V}|(|\mathcal{V}|-1)}$  [6]. Since the number of edges for a complete directed graph is  $|\mathcal{V}|(|\mathcal{V}|-1)$ , the maximum density is 1. Clearly, the minimum density is 0 (for empty graphs). There are two different approaches [13, 16], but there is no strict distinction between sparse and dense graphs.

### Distance-based metrics

The eccentricity of a node  $u$  is defined as the longest hop count between the node  $u$  and any other node in the graph.

Centralization [8] is a general method for calculating a graph-level centrality score based on some node-level centrality metric. Centrality based metrics are the following ones: degree centrality (*based on degree*), closeness centrality (*based on average distances*), betweenness centrality (*based on geodesics*), eigenvector centrality (*recursive: similar to page rank methods*), eccentricity centrality. In the case of eccentricity centrality, we not use the reciprocal to assure that more central nodes have a higher value of eccentricity.

### Connection-based metrics

The most basic connection-based metrics are the degree of a node, which is the number of edges to other nodes, and the degree distribution. The degree distribu-

tion  $P(k)$  of a graph is then defined to be the fraction of nodes in the network with degree  $k$ . Thus if there are  $n$  nodes in total in a graph and  $n_k$  of them have degree  $k$ , we have  $P(k) = n_k/n$ .

Clustering is a fundamental and important property of networks, just like degree and degree distribution. Clustering coefficient is the measurement that shows the opportunity of a graph to be divided into clusters. Clusters are disjoint subgraphs of the graph. A cluster usually should be a complete subgraph, so in this way it is similar to a clique, but a cluster may consist of one node, on the other hand a clique is a complete subgraph which contains always at least two nodes in case of a communication graph. The clustering coefficient can globally [12, 18] or locally [19] characterize a graph. The global clustering coefficient is based on triplets of nodes. The global clustering coefficient of a network, also known as transitivity  $T$ , which is the ratio of the number of loops of length three and the number of paths of length two.

Let  $u$  be a vertex with  $k$  degree and given by the proportion of  $e$  edges between the  $v$  within its neighborhood  $G$ , then the Local clustering coefficient of  $u$  in  $G$  is given by  $C_u = \frac{2e}{k(k-1)}$ . Thus,  $C_u$  measures the ratio of the number of edges between the neighbors of  $u$  to the total possible number of such edges. The average clustering coefficient is the average of local clustering coefficients.

## Wireless sensor networks

The ad hoc wireless sensor networks (WSN) are used widely (*for example in military to observe environment*). They have the advantage that they consist of sensors with low energy consumption, which can be deployed easily in a cheap way on such areas which are out-of-the-way. These sensors are the nodes of WSN. They are capable to process some limited information and to use wireless communication. A big effort is used to research how to deploy them in an optimal way to keep efficient energy consumption and communication. Although there are many WSN solutions, the deployment of a WSN is still an active research field [1].

One of the important property of an ad hoc wireless network is node density. The dense layout makes the following properties available: high fault tolerance, high-coverage characteristics, but also cause some problems. The interference is high near to dense node areas, and there are a lot of collisions in case of message passing, which requires complicated operations for routing protocols, because of too many possible routes, routing needs lots of resources [5].

The aim of topology control techniques is to reduce the cost of the distributed algorithms interpreted on the network. The graph, which represents a network, has to be thinned because of cost-reduction by techniques like disconnection of nodes, removing links, changing scopes, etc., but the network-quality characteristics (*like scalability, coverage, fault tolerance, etc.*) must not fall below a required level. The overall aim is to create a scalable, fault-tolerant sparse topology, where the degree of the nodes are low, the maximum load is low, energy consumption is low and the paths are short. The following techniques are used to create an optimal topology: reducing the scope of nodes, removing some nodes, introducing a dominating set

of nodes, clustering, and add some new nodes to gain all-all communication [15, 17].

In multi-hop networks one hop is the unit of the path between source and destination. The hop count refers to the number of intermediate nodes through which data must pass between source and destination. Networks can be classified by the number of hops between source nodes, which measures their environment, and a sink node, which collects data. In a single-hop network there is only one (*single*) hop between the source nodes and the sink node. In a multi-hop network a sensor can also transmit data from the source to the sink because there are more than one hop from the source to the sink.

The rest of this article is organized as follows. In Section 2 we give some preliminary definitions, like communication graph. In Section 3 we introduce the new metrics, each of them are  $k$ -hop based. In Section 4 we compare existing metrics and the new ones. In section 5 we show how to use this metrics to rank nodes and Section 6 contains our conclusions.

## 2. Preliminaries

Given a randomly-deployed sensor network with homogeneous or heterogeneous nodes. Also given a mapping which sensor is able to communicate with which sensors directly. Accordingly, by communication graph we mean a weighted directed graph  $\mathcal{D} = (\mathcal{S}; \mathcal{E}_{\mathcal{C}}, \mathcal{W})$ , where  $\mathcal{S}$  is the set of nodes, which represents the sensors,  $\mathcal{E}_{\mathcal{C}} \subseteq \mathcal{S} \times \mathcal{S}$  is the set of edges, and  $\mathcal{W}$  is the set of weights. An edge  $(x_i, x_j) \in \mathcal{E}_{\mathcal{C}}$  represents the possibility of messaging from node  $x_i$  to  $x_j$  in  $\mathcal{D}$ , i.e., the sensor represented by  $x_j$  is in the transmission range of  $x_i$ . The  $\mathcal{W}_{ij}$  denotes the communication cost of the  $(x_i, x_j)$  message. In the case of homogeneous sensors the  $\mathcal{D}$  graph is symmetric, accordingly  $\mathcal{D} = (\mathcal{S}; \mathcal{E}_{\mathcal{C}}, \mathcal{W})$  is equivalent to a simple weighted undirected graph  $\mathcal{G} = (\mathcal{S}; \mathcal{E}_{\mathcal{C}}, \mathcal{W})$ .

In case of an weighted undirected graph  $\mathcal{G} = (\mathcal{S}; \mathcal{E}_{\mathcal{C}}, \mathcal{W})$  we define a clique as a subset of the node set  $\mathcal{Cl} \subseteq \mathcal{S}$ , such that for every two nodes in  $\mathcal{Cl}$ , there exists an edge connecting the two. The weight of a  $\mathcal{Cl}$  is the sum of the weight of their edges.

If our communication graph is directed, we define a clique as a subset of the node set  $\mathcal{Cl} \subseteq \mathcal{S}$ , such that for every two nodes in  $\mathcal{Cl}$ , there exists an edge from the first one to the second one, and from the second one back to the first one. The weight of the  $\mathcal{Cl}$  is defined as above, considering that  $\mathcal{W}_{ij}$  and  $\mathcal{W}_{ji}$  are not necessarily equal. A maximal clique is a clique which is not a proper subset of any other clique. A  $n$ -clique is a clique which contains exactly  $n$  vertices.

In this paper we assume that the communication graph is strongly connected, the cost of communication between each node is constant (*we do not use weights*), and the network consists of homogeneous nodes. To test the new metrics we used our own representation [7] and SAT solver [10].

### 3. $k$ -hop based graph density and redundancy metrics

In this section we present some spanning tree and clique-based graph density metrics. With spanning tree-based metrics, we define graph density, whereas clique-based redundancy metrics mean the degree of relieving in our interpretation. We use the notion of  $k$ -hop environment of a node  $u$ , denoted by  $\mathcal{G}^{[n]}(u)$ , which is a subgraph of graph  $\mathcal{G}$ , which consists  $u$  and the nodes which can be reached from  $u$  from an path, which length is smaller or equal than  $k$ , and which contains edges between these nodes from  $\mathcal{G}$ . We compute local metrics for  $u$  by computing a graph metrics for  $\mathcal{G}^{[n]}(u)$ . The parameter  $k$  should be a relatively small number because otherwise  $\mathcal{G}^{[n]}(u)$  could be the whole graph. The metrics over the  $k$ -hop environment of a node can characterize the node more properly then considering merely the node itself. On the other hand these metrics characterize not only the node but its environment.

Taking into account the constraints mentioned in Section 2, the basic notations are:

- $u$ : the candidate node;
- $k$ : the number of hops;
- $\mathcal{N}$ ,  $\mathcal{V}$ : the number of nodes and edges of graph  $\mathcal{G}$ ;
- $\mathcal{N}^{[k]}(u), \mathcal{V}^{[k]}(u)$ : the number of nodes and edges of graph  $\mathcal{G}^{[k]}(u)$ ;
- $\mathcal{Cl}$ ,  $\mathcal{M}$ : the set of maximum cliques of graph  $\mathcal{G}$  and the cardinality of this set;
- $\mathcal{Cl}^{[k]}(u)$ ,  $\mathcal{M}^{[k]}(u)$ : the set of maximal cliques of graph  $\mathcal{G}^{[n]}(u)$  and the cardinality of this set;
- $\mathcal{T}^{[k]}(u)$ ,  $\mathcal{T}$ : the number of edges of the minimum cost spanning tree of graph  $\mathcal{G}^{[k]}(u)$  and  $\mathcal{G}$ . Note, that in case of a communication graph we have that  $\mathcal{T} = \mathcal{N} - 1$ , regardless whether the graph is directed or undirected;
- $s$ : the spreading factor, which is rather a technical value to enlarge small differences in the metrics, in this article we set  $s = 2.71$ ;
- $cs$ : the clique size, minimum value is 2.

#### 3.1. Spanning tree-based metrics

Spanning tree-based approaches can be found in the wide area of network protocols. For example, a known technique is Time-To-Live ( $TTL$ ). It works as follows, routing methods try to find the best path for forwarding the collected data, the TTL mechanism is used to limit the number of hops to avoid over-overlapping of paths

and to balance the data load on the nodes and the energy consumption [14]. They use also small  $k$  values.

We define graph density of the graphs  $\mathcal{G}$  and  $\mathcal{G}^{[k]}(u)$  as follows:

$$\mathcal{GD} = \frac{\mathcal{V}}{\mathcal{T}}$$

$$\mathcal{GD}^{[k]}(u) = \frac{\mathcal{V}^{[k]}(u)}{\mathcal{T}^{[k]}(u)}.$$

The graph density takes its maximum if the graph is complete. In case of undirected graphs the maximum is:  $\frac{\mathcal{N}(\mathcal{N}-1)}{2(\mathcal{N}-1)} = \frac{\mathcal{N}}{2}$ . In case of directed graphs the maximum is:  $\frac{\mathcal{N}(\mathcal{N}-1)}{\mathcal{N}-1} = \mathcal{N}$ . The graph density takes its maximum if the graph is a tree. In case of undirected graphs the minimum is:  $\frac{\mathcal{N}-1}{\mathcal{N}-1} = 1$ , since the graph is a communication graph, i.e., it is strongly connected. If the graph is directed, then the minimum is:  $\frac{2(\mathcal{N}-1)}{\mathcal{N}-1} = 2$ , because of the same reason.

### Communication and Weighted Communication Graph Density

We define the communication graph density of node  $u$  in its  $k$ -hop environment as follows:

$$\mathcal{CGD}^{[k]}(u) = s^{\frac{\mathcal{V}^{[k]}(u)}{\mathcal{T}^{[k]}(u)}}.$$

The  $\mathcal{CGD}^{[k]}(u)$  can be used also as a local metric for a node, and computed quickly for all nodes and use to rank them.

We define the weighted communication graph density of node  $u$  in its  $k$ -hop environment as follows:

$$\mathcal{WCGD}^{[k]}(u) = s^{\frac{\mathcal{V}^{[k]}(u)}{\mathcal{T}^{[k]}(u)}} \frac{\mathcal{N}^{[k]}(u)}{\mathcal{N}}.$$

The  $\mathcal{WCGD}^{[k]}(u)$  is no longer a purely local metric, but takes into account the number of nodes in the  $k$ -hop environment.

### Relative Communication Graph Density

We define the relative communication graph density of node  $u$  in its  $k$ -hop environment as follows:

$$\mathcal{RCGD}^{[k]}(u) = s^{\frac{\mathcal{CGD}^{[k]}(u)}{\mathcal{CGD}}} = s^{\frac{\mathcal{V}^{[k]}(u)\mathcal{T}}{\mathcal{T}^{[k]}(u)\mathcal{V}}}.$$

It maximizes its value when the  $k$ -hop environment of  $u$ , i.e.,  $\mathcal{G}^{[k]}(u)$  is a complete graph and the rest of the graph is a tree, or consists of several trees.

The minimum is - vice versa - assumes that the  $k$ -hop environment of  $u$  is a tree and the rest of the graph is a complete graph.

If we consider the two extremes, i.e., if the communication graph is a complete graph or if it is a tree, interestingly enough, we get the same relative communication

graph density, which is  $s$ . If the communication graph is a complete graph, then for any  $k \geq 1$  and for any node  $u$  we have that  $\mathcal{G}^{[k]}(u)$  is equal to  $\mathcal{G}$ , so,  $\frac{\nu^{[k]}(u)}{\tau^{[k]}(u)} = \frac{\nu}{\tau}$ , i.e.,  $\frac{\nu^{[k]}(u)\tau}{\tau^{[k]}(u)\nu} = 1$ . On the other hand, if the communication graph is a tree, then its communication graph density is a constant (1 if the graph is undirected, 2 if it is directed) for any  $n$  and  $u$ , so again  $\frac{\nu^{[k]}(u)\tau}{\tau^{[k]}(u)\nu} = 1$ .

We get the same result for the two extreme cases, because this metric shows the relative density of subgraph related to the whole graph. A tree has a very small density, and a complete graph has a very high density, but if we take a subgraph of a tree then it has the same density as the whole, and the same is true for a complete graph. So they have the same relative density.

This metric shows whether the  $k$ -hop environment of a node is more dense as the whole graph, or has the same density, or it is less dense. This means that if

- $\mathcal{RCGD}^{[k]}(u) = s$ , then  $\mathcal{G}^{[k]}(u)$  has the same cgd as  $\mathcal{G}$ ;
- $\mathcal{RCGD}^{[k]}(u) < s$ , then  $\mathcal{G}^{[k]}(u)$  has smaller cgd than  $\mathcal{G}$ ;
- $\mathcal{RCGD}^{[k]}(u) > s$ , then  $\mathcal{G}^{[k]}(u)$  has bigger cgd than  $\mathcal{G}$ ;

where cgd means communication graph density.

Note, that this metric is computed by dividing a local property by a global one.

### Weighted Relative Communication Graph Density

We define the weighted relative communication graph density of node  $u$  in its  $k$ -hop environment as follows:

$$WR\mathcal{CGD}^{[k]}(u) = \mathcal{RCGD}^{[k]}(u) \frac{\mathcal{N}^{[k]}(u)}{\mathcal{N}} = s \frac{\nu^{[k]}(u)\tau}{\tau^{[k]}(u)\nu} \frac{\mathcal{N}^{[k]}(u)}{\mathcal{N}}.$$

Note, that this metric is computed as a multiplication of two numbers, which are both computed by dividing a local property by a global one, so we have  $(local'/global') * (local''/global'')$ .

This metric takes in consideration also how many nodes are in the  $n$ -hop environment of the node  $u$ . A node is more valuable if its  $k$ -hope environment is bigger.

### 3.2. Clique-based metrics

During the work of a WSN the topology of the network may change because some sensors may go wrong, or the transmission range can be less. If a node can be found in a dense (redundant) environment then it may happen more often that communication interference occurs and routing is more resource consuming; on the other hand, the environment itself is more fault tolerant. In a sparse environment routing is easier, communication interference is less frequent, but the environment is less fault tolerant. The aim of topology control techniques is to reduce the cost

of the distributed algorithms interpreted on the network. But the network-quality characteristics (*like scalability, coverage, fault tolerance, etc.*) must not fall below a required level. A clique is a complete subgraph, so they have high communication redundancy, on the other hand they allow high fault tolerance, results in high coverage, etc.

First of all we define the average clique size as follows:

$$\overline{c\mathcal{L}} = \frac{1}{\mathcal{M}} \sum_{i=1}^{\mathcal{M}} |c\mathcal{L}_i|_{>=cs}.$$

The average clique size is maximal, if the graph is complete. Its minimum is  $cs$  if all maximal cliques have the size  $cs$ . It is not defined if there is no clique with size at least  $cs$ . Its maximum is  $\mathcal{N}$  if the communication graph is complete, because then we have only one maximal clique, the graph itself. The clique problem, the problem of finding all maximal size cliques, is a well-known NP-complete problem. It means that it is not feasible to find all maximal cliques in a large graph. So one can not use clique based metrics to guide topology control techniques, except if we work with relatively small graphs, like in the  $k$ -hop environment of a node.

### Clique size-based metrics

So we define the clique size-based communication graph redundancy of node  $u$  within  $k$ -hop environment as follows:

$$\mathcal{CGR}_{sb}^{[k]}(u) = \frac{1}{\mathcal{M}^{[k]}(u)} \sum_{i=1}^{\mathcal{M}^{[k]}(u)} |c\mathcal{L}^{[k]}(u)_i|_{>=cs}$$

It only shows the average clique size within  $k$ -hop environment of node  $u$ , but it ignores the number of nodes within the  $k$ -hop environment.

We define weighted communication graph redundancy of node  $u$  within  $k$ -hop environment as follows:

$$\mathcal{WCGR}_{sb}^{[k]}(u) = \mathcal{CGR}_{sb}^{[k]}(u) \frac{\mathcal{N}^{[k]}(u)}{\mathcal{N}}.$$

This metric uses also the number of nodes. This can be considered to be a local metric, because the computationally intensive tasks (*find cliques*) typically occur in a  $k$ -hop environment.

### Clique value-based metrics

Since a clique of size 4 is more valuable in a graph than 6 in a graph with 100 nodes, we shall take into consideration the number of nodes in the graph, which is denoted by  $\mathcal{N}$ , to compute the value of a clique. We also use the average clique size to normalize this value.

So we define the value of a clique as follows:

$$\mathcal{CL}_V = \frac{|\mathcal{Cl}|_{>=cs}}{\mathcal{N}} s \frac{|\mathcal{Cl}|_{>=cs}}{\overline{\mathcal{CL}}}.$$

We define also the average value of cliques as follows:

$$\overline{\mathcal{CL}_V} = \frac{1}{\mathcal{M}} \sum_{i=1}^{\mathcal{M}} \frac{|\mathcal{Cl}_i|_{>=cs}}{\mathcal{N}} s \frac{|\mathcal{Cl}_i|_{>=cs}}{\overline{\mathcal{CL}}}$$

We define also the average value of cliques within the  $k$ -hop environment, also called clique value-based communication graph redundancy as follows:

$$\mathcal{CGR}_{vb}^{[k]}(u) = \frac{1}{\frac{1}{\mathcal{M}^{[k]}(u)} \sum_{i=1}^{\mathcal{M}^{[k]}(u)} \frac{|\mathcal{Cl}^{[k]}(u)_i|_{>=cs}}{\mathcal{N}^{[k]}(u)} s \frac{|\mathcal{Cl}^{[k]}(u)_i|_{>=cs}}{\overline{\mathcal{CL}^{[k]}(u)}}}.$$

This metric is the pair of  $\mathcal{CL}_V$  in case of  $\mathcal{G}^{[k]}(u)$ . This is a local metric, but this notion does not takes into consideration the number of nodes in the  $k$ -hop environment of  $u$ . Without reciprocal, the peripheral but relievable nodes are ranked in advance.

After considering the number of nodes in the  $k$ -hop and conversion we define weighted clique value-based communication graph redundancy as follows:

$$\begin{aligned} \mathcal{WCGR}_{vb}^{[k]}(u) &= \frac{1}{\mathcal{CGR}_{vb}^{[k]}(u)} \frac{\mathcal{N}^{[k]}(u)}{\mathcal{N}} = \\ &= \frac{1}{\mathcal{M}^{[k]}(u)} \sum_{i=1}^{\mathcal{M}^{[k]}(u)} \frac{\mathcal{N}^{[k]}(u) |\mathcal{Cl}^{[k]}(u)_i|_{>=cs}}{\mathcal{N}^2} s \frac{|\mathcal{Cl}^{[k]}(u)_i|_{>=cs}}{\overline{\mathcal{CL}^{[k]}(u)}} \end{aligned}$$

This metric is the pair of  $\overline{\mathcal{CL}_V}$  in case of  $\mathcal{G}^{[k]}(u)$ .

## 4. Comparisons with other metrics

In this article we considered networks with 200-500 nodes at 15-40% densities. The  $k$  value in each case is less than 3. An important constraint was that the largest  $k$ -hop environment must be smaller than the quarter of a complete graph.

For simulating and analyzing networks we used a self-developed *Python* tool based on *NetworkX*<sup>1</sup>. For computing pairwise correlation of metrics we used *pandas*<sup>2</sup>. Many metrics are only implemented for undirected graphs in *NetworkX*, therefore, the comparisons were done only on undirected graphs.

The results of the correlation analysis are presented in Table 1–3 (*average values of 1000 runs*), shows some interesting phenomena and experience. The abbreviation c. means centrality.

<sup>1</sup><https://networkx.github.io>

<sup>2</sup><https://pandas.pydata.org>

## 4.1. 1-hop based environment

	<i>k</i> -hop based metrics					
	$WCGD^{[k]}$	$RCGD^{[k]}$	$WRCGD^{[k]}$	$WCGR_{sb}^{[k]}$	$CGR_{vb}^{[k]}$	$WCGR_{vb}^{[k]}$
Clustering coeff.	0,06	0,19	0,00	0,37	-0,17	0,14
Eccentricity	-0,07	-0,21	-0,24	-0,15	-0,31	-0,21
Betweenness c.	-0,04	-0,05	0,06	-0,12	0,22	-0,01
Degree c.	<b>0,54</b>	<b>0,87</b>	<b>0,94</b>	<b>0,76</b>	<b>0,84</b>	<b>0,83</b>
Closeness c.	0,05	0,23	0,28	0,17	0,37	0,23
Eigenvector c.	<b>0,67</b>	<b>0,61</b>	<b>0,64</b>	0,49	0,28	<b>0,68</b>

Table 1: Correlations with other metrics, where  $k$  is 1

It can be seen from the Table 1 that within 1-hop environment the defined metrics show their most significant correlation with degree centrality. The correlation is the strongest between  $WRCGD$  and degree centrality, the correlation is over 90%.  $WCGD$ ,  $RCGD$ ,  $WRCGD$  and  $WCGR_{vb}$  metrics are also strongly correlated with the eigenvector centrality.

## 4.2. 2-hop based environment

	<i>k</i> -hop based metrics					
	$WCGD^{[k]}$	$RCGD^{[k]}$	$WRCGD^{[k]}$	$WCGR_{sb}^{[k]}$	$CGR_{vb}^{[k]}$	$WCGR_{vb}^{[k]}$
Clustering coeff.	-0,07	0,04	-0,15	0,08	-0,26	-0,06
Eccentricity	-0,15	-0,24	-0,38	-0,22	-0,45	-0,33
Betweenness c.	0,06	0,03	0,22	0,05	0,32	0,17
Degree c.	<b>0,54</b>	<b>0,78</b>	<b>0,83</b>	<b>0,72</b>	<b>0,67</b>	<b>0,72</b>
Closeness c.	0,08	0,26	0,44	0,24	0,53	0,36
Eigenvector c.	<b>0,87</b>	<b>0,68</b>	<b>0,67</b>	<b>0,57</b>	0,26	<b>0,71</b>

Table 2: Correlations with other metrics, where  $k$  is 2

It can be seen from the Table 2 that the defined metrics within 2-hop environment showed a weaker correlation with the degree centrality and stronger with the eigenvector centrality, since the degree of neighbors of the examined node also affects the density and redundancy of the environment. The strongest correlation with the degree centrality is still shown with  $WRCGD$ , while with the eigenvector centrality correlates best with  $WCGD$ .

## 4.3. 3-hop based environment

The correlations in the 3-hop environment are shown in the Table 3. In general, the correlations with the degree centrality and the eigenvector centrality are no longer significant, the eccentricity and the closeness centrality correlations are reinforced.

In the following, we will analyze in detail the relationship between the new and already known metrics.

	k-hop based metrics					
	$WCGD^{[k]}$	$RCGD^{[k]}$	$WRCGD^{[k]}$	$WCGR_{sb}^{[k]}$	$CGR_{vb}^{[k]}$	$WCGR_{vb}^{[k]}$
Clustering coeff.	0,12	0,03	-0,20	0,03	-0,24	-0,19
Eccentricity	-0,27	-0,18	<b>-0,55</b>	-0,30	<b>-0,56</b>	<b>-0,51</b>
Betweenness c.	0,18	0,07	0,36	0,14	0,38	0,29
Degree c.	0,46	<b>0,55</b>	<b>0,63</b>	<b>0,56</b>	<b>0,54</b>	<b>0,62</b>
Closeness c.	0,38	0,25	<b>0,63</b>	0,34	<b>0,67</b>	<b>0,54</b>
Eigenvector c.	<b>0,72</b>	<b>0,55</b>	<b>0,52</b>	<b>0,52</b>	-0,26	<b>0,59</b>

Table 3: Correlations with other metrics, where  $k$  is 3

### Weighted Communication Graph Density

The metric  $WCGD^{[k]}(u)$  correlates strongly with eigenvector centrality. There is a not too strong but significant correlation with degree centrality also, and there is no relevant correlation with other metrics. If we want to characterize this metric on the basis of the above, then a high  $WCGD^{[k]}(u)$  value node has the following properties (in  $k$ -hop environment, if  $k = 3$ ):

- average probability of high number of direct connections,
- high probability of high degree of neighbors,
- weak probability of central location.

### Relative Communication Graph Density

The metric  $RCGD^{[k]}(u)$  has average correlation with degree centrality and eigenvector centrality, weak but significant contact with closeness centrality, and there is no relevant correlation with other metrics. So a high  $RCGD^{[k]}(u)$  value node has the following properties (in  $k$ -hop environment, if  $k = 3$ ):

- average probability of high number of direct connections,
- average probability of high degree of neighbors.

### Weighted Relative Communication Graph Density

The metric  $WRCGD^{[k]}(u)$  has an average linear correlation with degree centrality, closeness centrality, and eigenvector centrality, and suggests a weak correlation with betweenness centrality, but with eccentricity shows an average but inverse correlation. So a high  $WRCGD^{[k]}(u)$  value node has the following properties (in  $k$ -hop environment, if  $k = 3$ ):

- average probability of high number of direct connections,
- average probability of central location,
- average probability of high degree of neighbors,
- weak probability of high geodesic distance from any other node.

### Weighted Communication Graph Redundancy (*size-based*)

The metric  $\mathcal{WCGR}_{sb}^{[k]}(u)$  has average correlation with degree centrality and eigenvector centrality. It suggests a weak but significant correlation with closeness centrality and inverse correlation with eccentricity. There is no relevant correlation with other metrics. So a high  $\mathcal{WCGR}_{sb}^{[k]}(u)$  value node has the following properties (in  $k$ -hop environment, if  $k = 3$ ):

- average probability of high number of direct connections,
- average probability of high degree of neighbors,
- weak probability of central location.

### Communication Graph Redundancy (*value-based*)

The metric  $\mathcal{CGR}_{vb}^{[k]}(u)$  has an average linear correlation with degree centrality and closeness centrality. It suggests a weak correlation with betweenness centrality. It shows shows an average but inverse correlation with eccentricity. So a high  $\mathcal{CGR}_{vb}^{[k]}(u)$  value node has the following properties (in  $k$ -hop environment, if  $k = 3$ ):

- average probability of high number of direct connections,
- average probability of central location,
- weak probability of great geodesic distance from any other node.

### Weighted Communication Graph Redundancy (*value-based*)

The metric  $\mathcal{WCGR}_{vb}^{[k]}(u)$  has an average linear correlation with degree centrality, closeness centrality, and eigenvector centrality. It suggests a weak correlation with betweenness centrality, but with eccentricity shows an average but inverse correlation. So a high  $\mathcal{WCGR}_{vb}^{[k]}(u)$  value node has the following properties (in  $k$ -hop environment, if  $k = 3$ ):

- average probability of high number of direct connections,
- average probability of high degree of neighbors,
- average probability of central location,
- weak probability of great geodesic distance from any other node.

## 5. Node ranking

In this section we show how to use the different metrics to make node ranking (*top 30 selection*). The generated network (shown in Fig. 1 ) contains 100 randomly deployed and homogeneous sensor nodes (*vertices*) with 926 connections (*edges*). The density is 18.7%, the transmission range is 55  $m$ , the area is  $300m \times 300m$  and the  $k$ -hop number is 3. The communication graph of the exemplary network are shown Fig. 2.

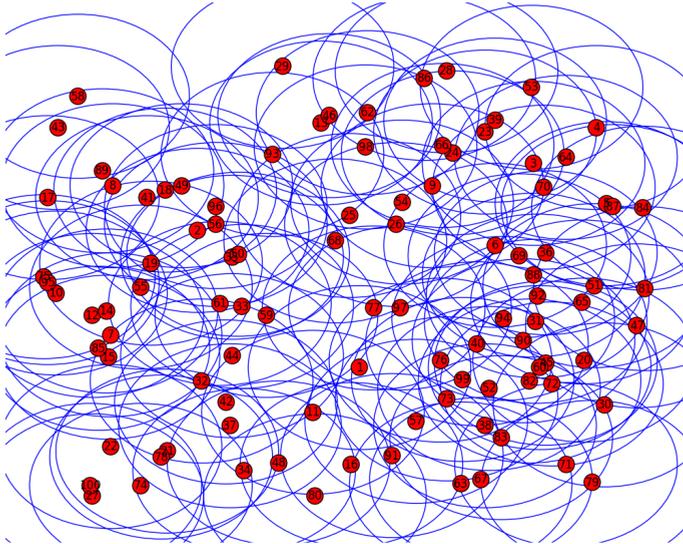


Figure 1: Randomly Deployed Sensor Network

Both the spanning tree and the clique based metrics show the denser environments of the network. Since this network consists of only 100 nodes it does not give a real picture of the metrics, but we can still see the tendencies.

### Spanning tree-based metrics (density)

Figures 3–5 show how to use the spanning tree based metrics to make ranking nodes. The task of all three metrics is to designate the densest environments. Based on the comparisons, the most significant feature of  $WCGD$  is that the neighbors and the neighbors' neighbors have a high degree. Figure 3 shows that the top 3 node and the at least 40% of the selected nodes are centrally located. The metric  $RCGD$  showed no significant correlation with the closeness centrality and eccentricity. In Figure 4 we can see that among the selected nodes there are only few nodes in central position. The metric  $RCGD$  primarily marks the nodes within the densest areas. The metric  $WCGD$  selects those nodes whose density is high within their

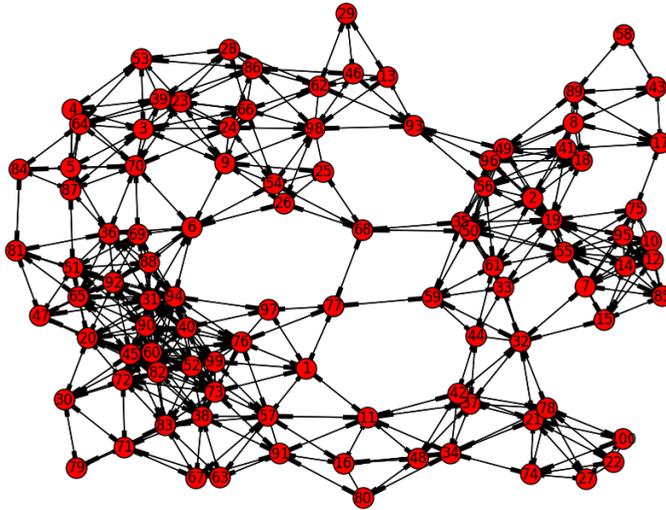


Figure 2: Communication Graph

$k$ -hop environment and centrally located. Figure 5 shows that the top 3 node and the at least 80% of the selected nodes are centrally located.

In these figures we use the following colour codes: top 1 rank node is red, top 2 is green, top 3 is yellow, top 4–10 are blue, top 11–20 are pink, top 21–30 are orange, the rest is cyan.

Figure 3 shows the top 30 ranked nodes based on the *Weighted Communication Graph Density* metric. In 3-hop environments, the highest weighted communication graph density has nodes 77, 68, and 26.

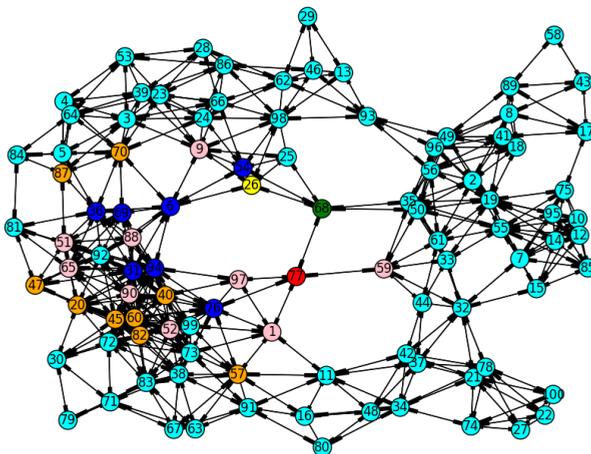


Figure 3: Weighted Communication Graph Density

Figure 4 shows the top 30 ranked nodes based on the *Relative Communication Graph Density* metric. In 3-hop environments, the highest relative communication graph density has nodes 30, 79, and 71.

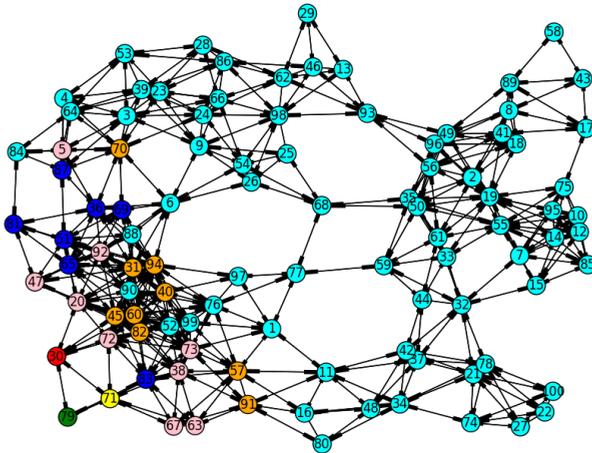


Figure 4: Relative Communication Graph Density

Figure 5 shows the top 30 ranked nodes based on the *Weighted Relative Communication Graph Density* metric. In 3-hop environments, the highest weighted relative communication graph density has nodes 68, 77, and 26.

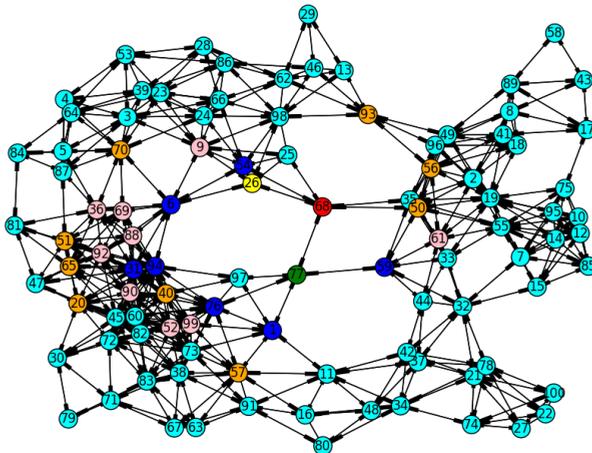


Figure 5: Weighted Relative Communication Graph Density

### Clique-based metrics (degree of relieving)

Figures 6–8 show how can we use the clique based metrics to make ranking nodes. The task of all three metrics is to designate the degree of relieving nodes within their  $k$ -hop environment. Figure 6 shows node ranking created by  $WCGR_{sb}$ . In case of  $WCGR_{sb}$  only the size of cliques in the  $k$ -hop environment of the examined node is relevant.  $WCGR_{sb}$  marks primarily the nodes within the densest areas, just lik  $RCGD$ . Figure 7 shows node ranking created by  $WCGR_{vb}$ . The significant difference between  $WCGR_{vb}$  and  $WCGR_{sb}$  is that  $WCGR_{vb}$  takes into consideration also the degree of neighbors. Figure 8 clearly shows that  $CGR_{vb}$  primarily focuses on centrally located nodes so the top 3 node and the at least 80% of the selected nodes are centrally located.

Figure 6 shows the top 30 ranked nodes based on the *Clique size-based Weighted Communication Graph Redundancy* metric. In 3-hop environments, the most relieved nodes has nodes 79, 30, and 81.

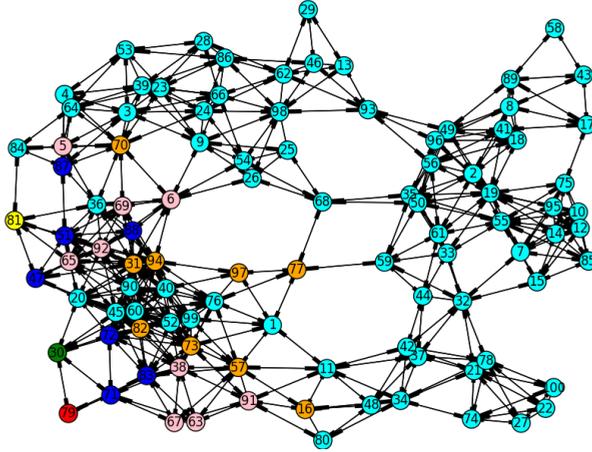


Figure 6: Clique size-based Weighted Communication Graph Redundancy

Figure 7 shows the top 30 ranked nodes based on the *Clique value-based Communication Graph Redundancy* metric. In 3-hop environments, the most relieved nodes has nodes 68, 77, and 26.

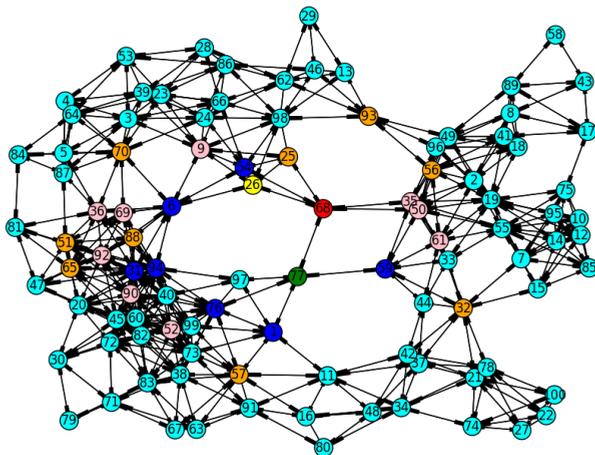


Figure 7: Clique value-based Communication Graph Redundancy

Figure 8 shows the top 30 ranked nodes based on the *Clique value-based Weighted Communication Graph Redundancy* metric. In 3-hop environments, the most relieved nodes has nodes 77, 68, and 26.

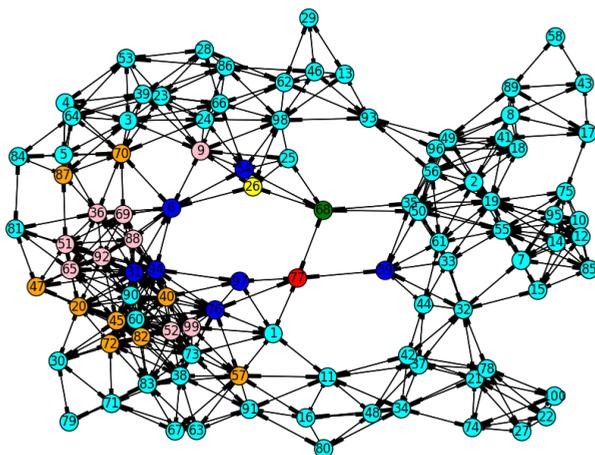


Figure 8: Clique value-based Weighted Communication Graph Redundancy

These figures show that 1-1 densities and redundancy-based metrics similarly rank the nodes. Why? The first reason is that the two concepts are closely related, if the density is high, then the redundancy is high, too. However, if the test is performed with directed graphs, there will be significant differences, because if node  $u$  can send a message to node  $v$ , then  $v$  may not be able to send a message to  $u$ , which means that high density does not mean necessarily also high redundancy.

## 6. Conclusions and Future work

In this paper we introduced several novel  $k$ -hop based density and redundancy metrics. We compared them to well-known graph metrics and we showed how can them be used for node ranking. Our primary goal was to define metrics that are able to rank nodes depending on their immediate environment within the whole network. Based on the results, we think that more sophisticated node ranking can be given using the new metrics. We primarily focused on modelling small heterogeneous networks. Metrics are defined so that they can be used also on networks where communication costs are different (*weighted directed graphs*). Our further goal is to investigate also such networks. An interesting questions is how to use these metrics to increase the efficiency of different (*Tx range-based, hierarchical*) topology control methods and how to use them in different hierarchical topological models (*e.g. clustering, cluster head selection*).

## References

- [1] I. F. AKYILDIZ, W. SU, Y. SANKARASUBRAMANIAM, E. CAYIRCI: *A survey on sensor networks*, IEEE Communications magazine 40.8 (2002), pp. 102–114, DOI: 10.1109/MCOM.2002.1024422.
- [2] R. ALBERT, A.-L. BARABÁSI: *Statistical mechanics of complex networks*, Reviews of modern physics 74.1 (2002), p. 47.
- [3] A.-L. BARABÁSI, R. ALBERT, H. JEONG: *Scale-free characteristics of random networks: the topology of the world-wide web*, Physica A: statistical mechanics and its applications 281.1-4 (2000), pp. 69–77.
- [4] S. BOCCALETTI, V. LATORA, Y. MORENO, M. CHAVEZ, D.-U. HWANG: *Complex networks: Structure and dynamics*, Physics Reports 424.4 (2006), pp. 175–308, ISSN: 0370-1573, DOI: 10.1016/j.physrep.2005.10.009.
- [5] BOLIC, MIODRAG, SIMPLOT-RYL, DAVID STOJMENOVIC, IVAN (EDS.): *RFID Systems - Research Trends and Challenges*, New York: John Wiley & Sons, 2010, p. 576, ISBN: 978-0-470-74602-8, DOI: 10.1002/9780470665251.
- [6] T. F. COLEMAN, J. J. MORÉ: *Estimation of sparse Jacobian matrices and graph coloring blems*, SIAM journal on Numerical Analysis 20.1 (1983), pp. 187–209.
- [7] G. K. CSABA BIRÓ: *Equivalence of Strongly Connected Graphs and Black-and-White 2-SAT Problems*, Miskolc Mathematical Notes 19.2 (2018), pp. 755–768, DOI: 10.18514/MMN.2018.2140.
- [8] L. C. FREEMAN: *Centrality in social networks conceptual clarification*, Social Networks (1978), p. 215, DOI: 10.1.1.227.9549.
- [9] J. M. HERNÁNDEZ, P. VAN MIEGHEM: *Classification of graph metrics*, Delft University of Technology: Mekelweg, The Netherlands (2011), pp. 1–20.
- [10] G. KUSPER, C. BIRÓ, G. B. ISZÁLY: *SAT solving by CSFLOC, the next generation of full-length clause counting algorithms*, in: 2018 IEEE International Conference on Future IoT Technologies (Future IoT), IEEE, 2018, pp. 1–9, DOI: 10.1109/FIOT.2018.8325589.
- [11] A. LESNE: *Complex networks: from graph theory to biology*, Letters in Mathematical Physics 78.3 (2006), pp. 235–262, DOI: 10.1007/s11005-006-0123-1.

- [12] R. D. LUCE, A. D. PERRY: *A method of matrix analysis of group structure*, Psychometrika 14.2 (June 1949), pp. 95–116, ISSN: 1860-0980, DOI: 10.1007/BF02289146, URL: 10.1007/BF02289146.
- [13] J. NEŠETŘIL, P. O. DE MENDEZ: *First order properties on nowhere dense structures*, The Journal of Symbolic Logic 75.3 (2010), pp. 868–887, DOI: 10.2178/js1/1278682204.
- [14] D. REWADKAR, M. P. MADHUKAR: *An adaptive routing algorithm using dynamic TTL for data aggregation in Wireless Sensor Network*, in: Second International Conference on Current Trends In Engineering and Technology-ICCTET 2014, IEEE, 2014, pp. 192–197, DOI: 10.1109/ICCTET.2014.6966286.
- [15] P. SANTI: *Topology Control in Wireless Ad Hoc and Sensor Networks*, ACM Computing Surveys 37.2 (2005), pp. 164–194, ISSN: 0360-0300, DOI: 10.1145/1089733.1089736.
- [16] I. STREINU, L. THERAN: *Sparse hypergraphs and pebble game algorithms*, European Journal of Combinatorics 30.8 (2009), Combinatorial Geometries and Applications: Oriented Matroids and Matroids, pp. 1944–1964, ISSN: 0195-6698, DOI: 10.1016/j.ejc.2008.12.018.
- [17] Y. WANG: *Topology Control for Wireless Sensor Networks*, Springer - Wireless Sensor Networks and Applications 148.2 (2008), pp. 113–147, ISSN: 1860-4862, DOI: 10.1007/978-0-387-49592-7.
- [18] S. WASSERMAN, K. FAUST: *Social network analysis: Methods and applications*, vol. 8, Cambridge university press, 1994.
- [19] D. J. WATTS, S. H. STROGATZ: *Collective dynamics of 'small-world' networks*, Nature 393.6684 (1998), pp. 440–442, ISSN: 00280836, DOI: 10.1038/30918.