# A framework for the evaluation of routing protocols in opportunistic networks

Dimitrios-Georgios Akestoridis[a], Evangelos Papapetrou[a,*]

[a]*Department of Computer Science and Engineering, University of Ioannina, 45110 Ioannina, Greece*

**Abstract**

The evaluation of routing protocols for opportunistic networks can be seen as a multidimensional problem because it involves several performance aspects. To capture these aspects various evaluation metrics are used, such as the number of delivered packets, the delivery delay and the number of transmissions. Unfortunately, in the context of opportunistic networks, these metrics are often highly correlated and usually conflicting. To make things worse, the characteristics of the network affect the importance of each metric as well as the levels of its correlation with other metrics. In this work, we first propose a set of performance evaluation metrics that are normalized with respect to the optimal performance. This approach tackles several of the above-mentioned shortcomings of traditional evaluation metrics. We then formulate the evaluation of routing protocols as a Multiple-Criteria Decision-Making (MCDM) problem where each routing protocol is an alternative and the performance metrics correspond to a set of criteria. We use this formulation to develop an evaluation framework that objectively ranks the performance of opportunistic routing protocols. To this end, we reshape well-known concepts and algorithms from the MCDM field to address the special requirements that are specific to the opportunistic context. We present detailed simulation results of well-known routing protocols in various opportunistic environments and rank their performance according to the proposed framework. In conclusion, no algorithm was able to achieve the best performance in all or the majority of the network topologies that we studied. This demonstrates the diversity of challenges that routing mechanisms face in such networks.

*Keywords:* evaluation framework, multiple-criteria decision-making, routing, opportunistic networks

## 1. Introduction

The evaluation and comparison of routing protocols, especially of those designed for opportunistic networks, is a complex process that involves many performance aspects. Typically, researchers capture different performance features by using a set of performance metrics such as the average ratio of successfully delivered packets, the average end-to-end delay, the average number of transmissions, etc. In general, a detailed assessment requires multiple performance metrics to be evaluated jointly rather than individually. The latter task is not straightforward, mainly because the investigated performance features are often correlated or even conflicting. Moreover, the involved trade-offs are ambiguous because the properties of the intermittent network are typically not known. For example, a small number of transmissions, a key performance indicator for energy efficiency, usually coexists with limited delivery capability. When comparing two protocols, it is not always clear what is an acceptable degradation of delivery capability for achieving increased energy efficiency, even if one has a strong preference for the latter.

In this work we aim at developing a generic and easy to implement method for ranking the performance of a set of protocols in a specific network. The ranking should rely on the *joint evaluation* of all aspects of a protocol's performance and *take into account the performance trade-offs imposed by the network structure*. This evaluation approach should be a valuable tool not only for assessing the performance of existing protocols but also for providing useful insight when designing new ones. Towards achieving our goal, we make the observation that we can formulate the evaluation of opportunistic routing protocols as a *Multiple-Criteria Decision-Making (MCDM)* problem. MCDM methods [1, 2, 3, 4] provide an evaluation of a set of alternatives using a set of criteria. In our approach, *we visualize each routing protocol as an alternative and the performance metrics as the set of criteria*. Unfortunately, legacy MCDM methods cannot address the full extent of the challenges faced in our scenario. This is mainly because of two reasons. The first relates to the normalization of criteria, i.e., the process of adjusting the values of different criteria to a common scale. Indeed, MCDM methods require such a normalization in order to create a common ground for combining the different criteria. We discuss wide-spread normalization methods in Section 3.1. Unfortunately, normalization significantly affects the outcome of the evaluation. In our scenario, an efficient normalization should consider the best performance allowed by the network. Reasonably, the normalization techniques used in MCDM methods are generic and therefore cannot provide the required "network-awareness". The second reason for the limited efficiency of conventional MCDM methods in our scenario pertains to the performance metrics used in the opportunistic routing literature. As discussed in detail in Section 4.1, these

*Corresponding author

*Email addresses:* `dakestor@cse.uoi.gr` (Dimitrios-Georgios Akestoridis), `epap@cse.uoi.gr` (Evangelos Papapetrou)

metrics present a high degree of correlation that significantly limits the information that each metric offers. This correlation is typically non-linear and "network-depended" therefore it cannot be captured by traditional MCDM methods. Moreover, in several cases, the extent of correlation is such that a metric may provide misleading information. For example, a routing protocol may deliver only short-haul packets, i.e. the ones exchanged between non-distant nodes, and therefore exhibit a low average delay. On the other hand, an opponent protocol, delivering short-haul packets with the same delay but also handing over long-haul ones, will present a larger average delay. We further discuss this issue in Sections 4.1 and 4.2.

We address the prior concerns by taking a two-step approach. First, we examine *a new normalization technique through which we redefine traditional performance metrics*. Our aim is to provide a *network-aware* version of those metrics and at the same time reduce their correlation as much as possible. In the second step, we focus on weighting, a key mechanism of MCDM methods. This is a process in which each metric is assigned a number that determines its importance compared to other metrics. We develop a *correlation-aware weighting model* that is suitable for our context. Such weighting models have been studied in the MCDM literature [4] (we review well-known weighting models in Section 3.2) but only examine linear correlation. Instead, we take a more generic approach that responds to both linear and non-linear relationships. Summarizing, our contributions are:

- We formulate the evaluation of opportunistic routing protocols as an MCDM problem where the protocols are the alternatives and the performance metrics are the criteria. Based on this formulation, we develop a framework that ranks each protocol's performance (Section 4).

- We propose a set of modified performance metrics that stem from well-known ones (Section 4.2). The proposed metrics capture the performance of routing protocols with respect to the optimal one in a specific network. Thus, they provide the means for a network-aware assessment when used either in the context of the proposed MCDM framework or as baseline metrics in future evaluation of opportunistic routing protocols.

- We propose VIC (Section 4.3), a weighting method that relies on the variability and the dependency (both linear and non-linear) of criteria in order to determine their relative importance. VIC can be applied to any MCDM problem for the assignment of objective weights and it is also compatible with the use of subjective, i.e. user-defined, ones.

- We perform a detailed performance assessment for a wide range of opportunistic routing protocols using a variety of real-world contact traces that correspond to networks of different scales and structures (Section 5).

In the rest of the paper, we first review opportunistic routing protocols (Section 2.1) and the most popular performance metrics (Section 2.2). In Section 3, we provide background information on traditional MCDM methods. After delineating our contributions (Sections 4-5), we conclude this work in Section 6.

## 2. Background

### 2.1. Routing in Opportunistic Networks

Routing in opportunistic networks follows the store-carry-and-forward approach. That is, a node may store packets for long periods of time and forward/replicate them upon a *contact* with another node. The main challenge in the design of opportunistic routing protocols is to determine whether a packet should be forwarded/replicated to an encountered node with the two extremes being Epidemic Routing [5] and Direct Delivery [6]. The former follows the most aggressive approach, i.e. replicates a packet to every encountered node that is not already a carrier, thus essentially floods the network. On the other hand, Direct Delivery follows a conservative approach, i.e., the source node does not forward/replicate a packet but waits to encounter the destination in order to deliver it. In the case of unlimited resources, Epidemic Routing will deliver every packet that the limited connectivity allows to be delivered. It will do so with the minimum delay but the trade-off is the excessive number of transmissions, i.e., routing overhead. On the contrary, Direct Delivery will produce the minimum routing overhead but often suffers long delivery delays and limited delivery capability.

Most of the proposed routing protocols try to strike a balance between these two extremes. To this end, they employ the concept of *utility*, i.e. a value indicating the fitness of a node for delivering (directly or indirectly) a packet to its destination. The idea is to forward/replicate a packet to a higher utility node. Most algorithms differ in the realization of the utility metric. Various such realizations have been proposed in the literature [7, 8]. For example, the utility of a node may be assembled using the time since its last encounter with the destination [9], its encounter rate with the destination or even a more complex interpretation of a node's history of previous encounters (e.g., the utility used by the PRoPHET algorithm [10]). A utility metric can be classified as either *destination-independent* or *destination-dependent* based on whether the captured information pertains to the destination or not. For instance, the total number of contacts that a node had with any other node is a destination-independent utility metric, while the total number of contacts with a specific destination node is a destination-dependent one [11].

Another design choice for a routing protocol is whether to follow a *single-copy* [12] or a *multi-copy* approach [13]. In single-copy protocols a node forwards the packet to a node with better utility, while in multi-copy ones the packet is replicated therefore multiple copies exist in the network. By spreading copies, multi-copy protocols are more likely to find a faster delivery path than single-copy ones. However, this comes at the expense of routing overhead, i.e. more transmissions. The algorithms that rely on a utility metric to determine when to replicate a packet to an encountered node follow what is known as the "Compare and Replicate" approach. The strategy aims at reducing the routing overhead. Yet, this is still significantly higher compared to the "Compare and Forward" approach, i.e. when a node decides based on a utility to forward, instead of replicating, a packet.

The "Spray" approach takes a more direct path towards controlling the number of a packet's replicas and therefore the routing overhead. According to the strategy, each packet carries in its header a number $R$ that announces the remaining number of replicas that can be created. This number is initialized with a maximum value $L$, i.e. $R = L$, when the packet is generated at the source. Each node carrying more than one copy of a packet (i.e. $R > 1$) may decide to distribute $R' < R$ of those copies to an encountered node. This is done by replicating the packet with the number $R'$ in its header. In this way, there are at maximum $L$ copies of the packet distributed to $L$ carrier nodes. The most representative algorithm in this category is Spray and Wait [14] where a node distributes half of its replicas, i.e. $R' = R/2$, to an encountered node. When $R = 1$, the node waits until it encounters the destination node. Spray and Focus [15] follows a similar approach but when $R = 1$ the node acts as in a single-copy utility-based protocol, i.e. forwards the packet to a higher utility node. Other spray-based protocols emphasize on a more efficient spraying of replicas. More specifically, in Last-Seen-First (LSF) Spraying [16], a node chooses to distribute half of its copies only to an encountered node of higher utility. The latter depends on the time elapsed since the node last encountered the destination. On the other hand, EBR [17] uses the utility metric to determine $R'$, i.e., how many replicas should be forwarded upon an encounter. In particular, the utility value of each node corresponds to an exponentially weighted moving average of its number of encounters. Thus, when two nodes meet, $R'$ depends on the utility values of the two nodes. SimBetTS [18] follows a similar approach but relies on social network analysis in order to determine the utility metric. In the case that there is only one replica of a packet, SimBetTS operates as a single-copy routing protocol.

A different approach towards reducing the routing overhead of multi-copy protocols was introduced by Delegation Forwarding (DF) [19]. Unlike the "Compare and Replicate" approach, where each node replicates a packet to a node with a higher utility value, DF takes into account the history of observed utility values. More specifically, a node will replicate a packet only when it encounters another node whose utility value is greater than the highest utility value that has been observed for that packet so far. Note that DF works with virtually any utility metric. COORD [20] builds upon DF's concept to further reduce redundant replications. To that end, COORD enables nodes to coordinate their replication decisions by exchanging their observations.

Note that MCDM concepts have been used in the context of opportunistic networks, either implicitly [18] or explicitly [21]. The idea is to combine multiple utility metrics in order to enable a node to make the best decision regarding a network function, e.g., forwarding, replication, congestion control, etc. In this paper, we focus on the performance evaluation of routing protocols and formulate this task as an MCDM problem, i.e., we use multiple performance metrics to rank the performance of different routing algorithms.

### 2.2. Performance metrics for Opportunistic Routing Protocols

Routing is a complex process that should meet multiple conflicting objectives. In networks of intermittent connectivity, the trade-offs for achieving these objectives are often unpredictable. Therefore, assessing the performance of an opportunistic routing protocol is a challenging task. Reasonably, this involves the joint evaluation of a set of performance metrics where each metric quantifies a performance feature. Traditionally, the most commonly used metrics are the following:

- *Delivery Ratio (DR)*: The number of successfully delivered packets normalized to the number of generated ones.
- *Average Delay (AD)*: The delay for delivering a packet to its destination. It is calculated as the sum of all delivery delays divided by the number of delivered packets.
- *Overhead Ratio (OR)*: The total number of transmissions normalized to the number of generated packets.[1]

Finally, a less popular performance metric, although one that can provide valuable information, is the *Average Number of Hops (ANH)* [22] required for delivering a packet to its destination.

## 3. Multiple-Criteria Decision-Making

In this section we review the Multiple-Criteria Decision-Making (MCDM) literature. In general, MCDM schemes consist of a *decision-making algorithm* and a *weighting method*.

### 3.1. Decision-Making Methods

In a multiple-criteria evaluation problem there are $n$ alternatives that have been evaluated using $m$ criteria. The evaluation information is summarized in a *normalized decision matrix*:

$$\mathbf{Z} = \begin{array}{c} \\ \mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n \end{array} \begin{array}{c} \begin{array}{cccc} \mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_m \end{array} \\ \begin{pmatrix} z_{1,1} & z_{1,2} & \cdots & z_{1,m} \\ z_{2,1} & z_{2,2} & \cdots & z_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n,1} & z_{n,2} & \cdots & z_{n,m} \end{pmatrix} \end{array}, \tag{1}$$

where $z_{i,j}$ corresponds to the performance value of the alternative $\mathbf{a}_i$ with respect to the criterion $\mathbf{c}_j$ for $i = 1, \ldots, n$ and $j = 1, \ldots, m$. We refer to the criteria that capture a positive performance aspect as *benefit criteria* while those referring to a negative performance feature are known as *cost criteria*. Since the criteria could be expressed in different measurement units it is critical that the decision matrix is normalized so that each element $z_{i,j}$ is dimensionless. A variety of normalization approaches have been proposed in the literature [23]. Table 1 presents the most commonly used normalization methods for the construction of the $\mathbf{Z}$ matrix out of non-normalized performance values $x_{i,j}$ for $i = 1, \ldots, n$ and $j = 1, \ldots m$. Furthermore,

[1] In some studies, the routing overhead may be quantified by the number of transmissions per delivered packet.

Table 1: Most common normalization techniques for decision-making methods.

| | |
|---|---|
| Linear 1 | $z_{i,j} = \frac{x_{i,j}}{x_j^{max}}\,,\ i = 1,\ldots,n$ for benefit criteria<br>$z_{i,j} = \frac{x_j^{min}}{x_{i,j}}\,,\ i = 1,\ldots,n$ for cost criteria |
| Linear 2 | $z_{i,j} = \frac{x_{i,j}-x_j^{min}}{x_j^{max}-x_j^{min}}\,,\ i = 1,\ldots,n$ for benefit criteria<br>$z_{i,j} = \frac{x_j^{max}-x_{i,j}}{x_j^{max}-x_j^{min}}\,,\ i = 1,\ldots,n$ for cost criteria |
| Linear 3 | $z_{i,j} = \frac{x_{i,j}}{\sum_{k=1}^{n} x_{k,j}}\,,\ i = 1,\ldots,n, \quad j = 1,\ldots,m$ |
| Vector | $z_{i,j} = \frac{x_{i,j}}{\sqrt{\sum_{k=1}^{n} x_{k,j}^2}}\,,\ i = 1,\ldots,n, \quad j = 1,\ldots,m$ |
| $x_j^{max} = \max\{\, x_{i,j} \mid i = 1,\ldots,n \,\}$ and $x_j^{min} = \min\{\, x_{i,j} \mid i = 1,\ldots,n \,\}$ | |

in order to be able to determine different levels of importance for various criteria, MCDM methods introduce a *weight vector* $\mathbf{w} = (w_1\ w_2\ \ldots\ w_m)$, where $\sum_{j=1}^{m} w_j = 1$ and $w_j \geq 0$ for $j = 1,\ldots,m$. In this context, $w_j$ corresponds to the relative importance of the $j$th criterion. Using $\mathbf{w}$ it is possible to construct the *weighted normalized decision matrix* $\mathbf{T}$ such that $t_{i,j} = w_j z_{i,j}$ for $i = 1,\ldots,n$ and $j = 1,\ldots,m$. We review weighting methods in the following subsection.

Given a normalized decision matrix $\mathbf{Z}$ and a weight vector $\mathbf{w}$, several decision-making methods have been proposed in order to provide an overall score of each alternative [24, 25]. The simplest and probably the most widely used method is the Simple Additive Weighting (SAW), which is also known as the Weighted Sum Model (WSM) [26]. According to the SAW method, the overall score of each alternative $\mathbf{a}_i$ is determined as:

$$SAW(\mathbf{a}_i) = \sum_{j=1}^{m} w_j z_{i,j}\,, \quad i = 1,\ldots,n\,. \tag{2}$$

This method is applicable only when all criteria are or have been transformed to be either benefit or cost ones. For example, this can be done using the normalization methods that we refer to as “Linear 1” and “Linear 2” in Table 1. Another similar decision-making method is the Multiplicative Exponential Weighting (MEW), which is also known as the Weighted Product Model (WPM) [27, 28]. Their main difference is that MEW relies on multiplication instead of addition. More specifically, the overall score of each alternative $\mathbf{a}_i$ is given by:

$$MEW(\mathbf{a}_i) = \prod_{j=1}^{m} \left(z_{i,j}\right)^{w_j}\,, \quad i = 1,\ldots,n\,. \tag{3}$$

Assuming that all criteria are benefit criteria then the best alternative is the one that yields the highest overall score.

A different approach is followed by the Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) [1, 2]. The key idea here is that the best alternative is the closest to the positive ideal solution and the farthest to the negative ideal solution. The positive ideal solution $\mathbf{a}^+ = \left(t_1^+\ t_2^+\ \ldots\ t_m^+\right)$ consists of the maximum values of all the benefit criteria and the minimum values of all the cost criteria. In other words, $t_j^+$ is the maximum value in the $j$-th column of matrix $\mathbf{T}$ if the $j$-th criterion is a benefit one and the minimum value if the criterion is a cost one. Similarly, the negative ideal solution $\mathbf{a}^- = \left(t_1^-\ t_2^-\ \ldots\ t_m^-\right)$ consists of the minimum values of all the benefit criteria and the maximum values of all the cost criteria. TOPSIS calculates the Euclidean distances from an alternative $\mathbf{a}_i$ to the positive ($\mathbf{a}^+$) and the negative ($\mathbf{a}^-$) ideal solutions. Then, it determines the overall score of $\mathbf{a}_i$ as:

$$TOPSIS(\mathbf{a}_i) = \frac{d_i^-}{d_i^- + d_i^+}\,, \quad i = 1,\ldots,n\,, \tag{4}$$

where $d^+(d^-)$ is the Euclidean distance of $\mathbf{a}_i$ from $\mathbf{a}^+(\mathbf{a}^-)$. The alternative with the highest overall score is considered the best alternative. This is because it has the shortest Euclidean distance from the positive ideal solution and the longest from the negative ideal solution.

A modified version of TOPSIS has also been proposed, which we refer to as mTOPSIS [3]. Instead of calculating the Euclidean distances of each alternative from the positive and negative ideal solutions in matrix $\mathbf{T}$, mTOPSIS uses the weighted Euclidean distances of each alternative from the positive and negative ideal solutions in matrix $\mathbf{Z}$. More specifically, after transforming any cost criteria into benefit ones, mTOPSIS calculates the following weighted Euclidean distances:

$$d_i^{max} = \sqrt{\sum_{j=1}^{m} w_j \left(z_j^{max} - z_{i,j}\right)^2}\,, \quad i = 1,\ldots,n\,, \tag{5}$$

$$d_i^{min} = \sqrt{\sum_{j=1}^{m} w_j \left(z_{i,j} - z_j^{min}\right)^2}\,, \quad i = 1,\ldots,n\,, \tag{6}$$

where $z_j^{max} = \max\{\, z_{i,j} \mid i = 1,\ldots,n \,\}$ and $z_j^{min} = \min\{\, z_{i,j} \mid i = 1,\ldots,n \,\}$. Then, the overall score of each alternative $\mathbf{a}_i$ is given by:

$$mTOPSIS(\mathbf{a}_i) = \frac{d_i^{min}}{d_i^{min} + d_i^{max}}\,, \quad i = 1,\ldots,n\,. \tag{7}$$

### 3.2. Weighting Methods

Determining the weights of each criterion in MCDM is of paramount importance. Numerous weighting methods have been proposed in the literature. These methods can be broadly classified into three categories: subjective, objective, and integrated. Subjective weighting methods rely on the decision maker to determine the importance of each criterion [29, 30, 31, 32, 33, 34] while objective weighting methods determine the importance of each criterion based on the available information in the decision matrix [4, 3]. Integrated weighting methods combine subjective and objective information [35, 36, 37, 38, 39]. In this work, we focus on objective weighting methods in order to provide an unbiased evaluation of routing protocols on contact datasets with vastly different characteristics.

The most simplistic weighting approach is to assume that all the criteria are equally important, i.e., $w_j = \frac{1}{m},\ j = 1,\ldots,m$. This approach is known as the Mean Weights (MW) method [4]. However, advanced objective methods make the observation that criteria exhibiting little variation in the recorded performance values (i.e. the values of a column in matrix **Z**) have little use in differentiating the alternatives [40]. Therefore, focus should be shifted to other criteria. For example, when we want to evaluate a set of routing protocols, we usually consider the fraction of successfully delivered packets as an important performance metric. However, if all the routing protocols deliver about the same number of packets, then this criterion has little influence on the overall evaluation of the routing protocols. If we view criteria as information sources then their importance can be perceived as their contrast intensity [41]. The latter can be quantified by a measure of either entropy or standard deviation. In the Entropy Measure (EM) method [3], the amount of information emitted from each criterion $\mathbf{c}_j$ is measured by:

$$e_j = -\frac{1}{\ln(n)} \sum_{i=1}^{n} z_{i,j} \ln\left(z_{i,j}\right), \quad j = 1, \ldots, m, \tag{8}$$

where the matrix **Z** should be normalized such that $z_{i,j} \in [0, 1]$ and $\sum_{i=1}^{n} z_{i,j} = 1$. The weight of each criterion is given by normalizing its degree of divergence:

$$w_j = \frac{1 - e_j}{\sum\limits_{k=1}^{m} (1 - e_k)}, \quad j = 1, \ldots, m. \tag{9}$$

Similarly, the Standard Deviation (SD) method [4] determines the relative importance of each criterion through the following equation:

$$w_j = \frac{\sigma_j}{\sum\limits_{k=1}^{m} \sigma_k} = \frac{\sqrt{\frac{1}{n}\sum\limits_{i=1}^{n}\left(z_{i,j} - \mu_j\right)^2}}{\sum\limits_{k=1}^{m}\sqrt{\frac{1}{n}\sum\limits_{i=1}^{n}\left(z_{i,k} - \mu_k\right)^2}}, j = 1, \ldots, m, \tag{10}$$

where $\mu_j$ is the arithmetic mean of the values of all alternatives for criterion $\mathbf{c}_j$.

The notion of dependence among criteria was first taken into consideration by the "CRiteria Importance Through Intercriteria Correlation" (CRITIC) method [4]. CRITIC, similar to the SD method, also relies on the standard deviation to quantify the contrast intensity of each criterion. It also uses a symmetric matrix where the element $\rho_{j,k}$ of the matrix is the linear correlation coefficient, also known as the Pearson product-moment correlation coefficient, between criteria $\mathbf{c}_j$ and $\mathbf{c}_k$. The idea is that $\rho_{j,k}$ is a measure of the *conflicting character* of the criteria. CRITIC measures the conflict that each criterion has with the rest of chosen criteria. Then, it combines the conflict of the criterion with its corresponding standard deviation in order to determine its importance:

$$f_j = \sigma_j \sum_{k=1}^{m} \left(1 - \rho_{j,k}\right), \quad j = 1, \ldots, m. \tag{11}$$

In other words, according to CRITIC, the importance of a criterion depends on the amount of variation in the recorded values and the amount of discordance that this criterion has with the other criteria. The objective weights are determined as:

$$w_j = \frac{f_j}{\sum\limits_{k=1}^{m} f_k}, \quad j = 1, \ldots, m. \tag{12}$$

## 4. A Framework for the Evaluation of Opportunistic Routing Protocols

As previously discussed, our observation is that it is possible to formulate the evaluation of opportunistic routing protocols as an MCDM problem where each protocol is an alternative and each performance metric is a criterion. Nevertheless, there are several pitfalls in using the traditional performance metrics as criteria. We address such problems and propose solutions in Sections 4.1 and 4.2. Then, in Section 4.3, we tackle the problem of weighting the importance of criteria.

### *4.1. Pitfalls in using Performance Metrics as Evaluation Criteria*

The first drawback of traditional performance metrics stems from the fact that they exhibit diverse ranges of values. Table 2 reports the simulation results (i.e., the average values of the metrics presented in Section 2.2) for several routing protocols in the Dartmouth dataset.[2] It is clear that some kind of normalization is required in order to perform a multi-criteria assessment of the protocols. Indeed, as we mentioned, decision-making methods include a normalization step when needed. This step produces the normalized decision matrix **Z** in Eq. (1) from a non-normalized matrix **X** where the normalized value $z_{i,j}$ of criterion $\mathbf{c}_j$ for alternative $\mathbf{a}_i$ is produced from $x_{i,j}$ and the observed criteria values for the other alternatives, i.e., $x_{k,j}, \forall k$. Unfortunately, this approach bears a significant disadvantage; clearly the calculation of **Z** depends on the set of chosen alternatives. According to Section 3.1, this, in our context, implies that the score of each routing algorithm depends on the set of the examined algorithms. Regrettably, in many cases this is also true for the rank of each routing algorithm. To illustrate the problem, we used the method entitled as "Linear 1" in Table 1 to normalize the performance of protocols in the example of Table 2. We report two normalizations (Table 3); one considering all six protocols as alternatives and another one with only four protocols (Epidemic and Direct routing are excluded). The two normalizations are entirely different because the set of alternatives has been narrowed down, i.e. the performed normalization depends on the set of alternatives. We argue that an objective performance assessment calls for a more stable normalization method.

Another major challenge in using legacy metrics originates from the fact that they are highly correlated. For example, one can notice in Table 2 the high correlation between most of the metrics with the prominent example being the correlation between the average delay and the overhead ratio (Pearson correlation coefficient $\rho \approx -0.81$). Working with correlated criteria

[2] The details of the simulation setup are presented in Section 5. Results for more protocols and datasets can be found in Section 5 as well as in [42].

Table 2: Average values of the traditional performance metrics for a set of routing protocols on the Dartmouth dataset.

| *Protocol* | *DR* | *AD* (*in hours*) | *OR* |
|---|---|---|---|
| Epidemic | 0.9779 | 27.10 | 560.00 |
| Direct | 0.0453 | 52.22 | 0.05 |
| CnF.PRoPHET | 0.5515 | 77.32 | 8.49 |
| CnR.PRoPHET | 0.9508 | 40.97 | 119.22 |
| DF.PRoPHET | 0.8214 | 59.30 | 14.64 |
| COORD.PRoPHET | 0.7804 | 61.28 | 10.73 |

Table 3: Normalized performance metrics considering all six (four) of the algorithms (Dartmouth dataset, Method: "Linear 1").

| *Protocol* | *Norm. DR* | *Norm. AD* | *Norm. OR* |
|---|---|---|---|
| Epidemic | 1 (-) | 1 (-) | 0.0001 (-) |
| Direct | 0.0463 (-) | 0.5190 (-) | 1 (-) |
| CnF.PRoPHET | 0.5640 (0.5800) | 0.3505 (0.5299) | 0.0059 (1) |
| CnR.PRoPHET | 0.9723 (1) | 0.6615 (1) | 0.0004 (0.0712) |
| DF.PRoPHET | 0.8400 (0.8639) | 0.4570 (0.6909) | 0.0034 (0.5799) |
| COORD.PRoPHET | 0.7980 (0.8208) | 0.4422 (0.6686) | 0.0047 (0.7912) |

is not new. In fact, MCDM algorithms use weighting methods [4] in order to take into account this correlation and perform a more accurate assessment. We discuss the challenges in developing a suitable weighting method in Section 4.3. Here, we examine the problems that cannot be tackled by a weighting method. More specifically, we refer to the fact that some performance metrics may provide erroneous information because of their high degree of correlation. A typical example of such a metric is the average delay due to the statistical bias that a low delivery ratio inflicts to it. To explain this, let us consider the delay comparison of two routing protocols, one with a low ($RP_A$) and another with a high ($RP_B$) delivery ratio. Even if $RP_B$ delivers every packet that $RP_A$ does with the exact same delay, it will probably exhibit a higher average delay. Yet, its performance is not worse. This happens because the higher delivery ratio of $RP_B$ usually implies the ability to reach more distant destinations, i.e. the ones that require longer delays. To illustrate such an example, let us go back to Table 2. Observe that Direct Delivery presents a lower average delay compared to Delegation Forwarding (DF) and COORD when both use the PRoPHET utility (DF.PRoPHET and COORD.PRoPHET respectively in Table 2). Note that in both DF and COORD the source node will always have a replica of the packet. In other words, in the worst case, both protocols match the delay performance of Direct Delivery for the subset of packets delivered by the latter, i.e. they deliver the packets using the exact same single hop paths if not finding shorter-delay multi-hop ones. Consequently, the information provided by the average delay metric is misleading.

Another example of inaccurate information also involves delay and emerges when two protocols deliver the same number of packets but to different destinations. For example, one protocol may deliver a packet with the optimal delay, i.e. the shortest delay allowable by the network. On the other hand, another protocol may deliver a packet to a different destination with a delay that is smaller compared to the first algorithm but is larger than the optimal one. In this case, the average delay for the second protocol will be smaller even though its performance is worse. The obvious reason is that the traditional average delay metric does not take into account the "difficulty" of reaching a destination. A similar phenomenon emerges when interpreting the overhead ratio metric because again the "difficulty" of reaching the destination, i.e. the minimum number of required transmissions, is not taken into account.

As a final note on using metrics as criteria, observe that some of the metrics in Section 2.2 can be considered as benefit while others as cost criteria. This rules out the decision-making methods that require a set of criteria that are all either benefit or cost ones, e.g. SAW and MEW, unless a transformation step that converts cost (benefit) criteria to benefit (cost) ones is included.

### *4.2. Proposed Performance Metrics*

The previous discussion highlights the need for a set of improved performance metrics, suitable for formulating the evaluation of routing protocols as an MCDM problem. To bypass the cons of the normalization step in MCDM algorithms, we opt for inherently normalized metrics. In such a case, a reasonable question that arises is what the normalization basis should be. Since opportunistic networks are characterized by heterogeneous contact rates and unpredictable node mobility that significantly affect the routing process, we argue that a metric should be normalized with respect to the optimal performance in a specific network. To this end, we use two single-copy algorithms that require a priori the entire information about the network's evolution. The first one, which we refer to as $OPT_D$, delivers a packet using the minimum delay path while the hop count is the tie-breaker for equal delay paths. The second algorithm, which we refer to as $OPT_F$, delivers a packet through the minimum hop path with the minimum delay being the tie-breaker. In other words, both algorithms maximize the number of delivered packets while $OPT_D$ minimizes delay and $OPT_F$ minimizes the number of required transmissions. Note that, although non-realistic, both algorithms provide performance limits that can be matched by realistic algorithms. For example, Epidemic routing maximizes the number of delivered packets while minimizing delay, obviously at the expense of transmissions.

To determine the performance of $OPT_D$ and $OPT_F$ it is possible to model an opportunistic network and the contacts between its nodes as a time evolving (or temporal) graph where each edge is associated with a set of labels and each label defines a period of time that the edge is available. Using this modeling, $OPT_D$ corresponds to the algorithm that finds what is known as the "foremost journey", i.e. the path that its arrival time to the destination is the smallest among all other paths. Moreover, $OPT_F$ corresponds to the algorithm that finds the "shortest journeys", i.e., the minimum-hop path, since the minimum number of transmissions can be achieved by following this path. There exist efficient algorithms for identifying both the foremost as well as the shortest journeys in time evolving graphs [43, 44]. Furthermore, in Section 5.1 we present a lightweight simulation-based approach for determining the performance of both $OPT_D$ and $OPT_F$.

After defining $OPT_D$ and $OPT_F$, we are in a position to define a set of normalized performance metrics. We call the first one the *Normalized Delivery Ratio (NDR)* and define it as the total number of successfully delivered packets normalized to the maximum number of packets that can be delivered. Note that the latter number is actually the number of packets delivered by either $OPT_D$ or $OPT_F$. Thus, the *NDR* of a routing protocol *RP* is calculated as:

$$NDR\,(RP) = \frac{|Dlv\,(RP)|}{|Dlv\,(OPT_D)|}\,, \tag{13}$$

where $|Dlv\,(\cdot)|$ denotes the number of delivered packets. *NDR* is particularly useful for scenarios where a large proportion of the generated packets is not delivered due to the network connectivity. Therefore, a low delivery ratio does not necessarily mean that a routing protocol performed poorly. *NDR* is 1 even if the delivery ratio is less than 1, provided that the routing protocol delivered the same packets as $OPT_D$ did.

For assessing the delay profile of a protocol, we use the ratio of the end-to-end delay that $OPT_D$ achieves for a packet to the end-to-end delay of the examined protocol for the same packet. Clearly, this ratio lies between 0 and 1, with 1 meaning that the protocol achieves the minimum delay. Then, we define the *Normalized Delivery Delay (NDD)* Index of a protocol *RP* as:

$$NDD\,(RP) = \frac{\displaystyle\sum_{i \in Dlv(RP)} \frac{D_i\,(OPT_D)}{D_i\,(RP)}}{|Dlv\,(OPT_D)|}\,, \tag{14}$$

where $Dlv\,(\cdot)$ denotes the set of delivered packets and $D_i(\cdot)$ the end-to-end delay for packet $i$. Observe that a high *NDD* value corresponds to a good performance and that $NDD(OPT_D) = 1$. *NDD* tackles the shortcomings of the traditional average delay metric that we discussed previously. One first improvement is that it is now possible to have a more fair comparison of protocols that deliver packets to different destinations since $D_i\,(OPT_D)$ in Eq. (14) quantifies the "difficulty" of reaching the destination of packet $i$. Let us, for example, consider the case of two protocols $RP_A$ and $RP_B$ that deliver one packet each, $k$ and $l$ respectively, to different destinations. Also, assume that the destination of $k$ is a low-delay one while the opposite holds for the destination of $l$, i.e. $D_k(OPT_D) \ll D_l(OPT_D)$. Then with high probability $D_k(RP_A) < D_l(RP_B)$ and $RP_A$ presents a smaller average delay even if $RP_A$ delivers the packet with suboptimal delay, i.e. $D_k(RP_A) > D_k(OPT_D)$, and $RP_B$ delivers its packet with optimal delay. However, using *NDD* secures that $RP_B$ receives the better score. The second important enhancement is that *NDD* minimizes the statistical bias that a low delivery ratio inflicts on the average delay. More specifically, let us assume two protocols $RP_A$ and $RP_B$ where $RP_A$ delivers only a subset of the packets delivered by $RP_B$, i.e. $Dlv\,(RP_A) \subset Dlv\,(RP_B)$, with the exact same delay as $RP_B$ does, i.e. $D_i(RP_A) = D_i(RP_B)$, $\forall\, i \in Dlv\,(RP_A)$. Previously, we explained the anomaly that the average delay for $RP_B$ will probably be higher than that of $RP_A$ even if $RP_B$ delivers the extra packets with the optimal delay, i.e. $D_i(RP_B) = D_i(OPT_D)$, $\forall\, i \in Dlv\,(RP_B) - Dlv\,(RP_A)$. Observe that *NDD* assumes an infinite delay for every packet delivered by $OPT_D$ but not by the examined protocol (due to the normalization based on the number of packets delivered by $OPT_D$). Therefore, in the previous example $RP_A$ will receive a smaller score compared to $RP_B$.

Table 4: Average values of the proposed performance metrics for a set of routing protocols on the Dartmouth dataset.

| *Protocol* | *NDR* | *NDD* | *NRO* |
|---|---|---|---|
| Epidemic | 1.0000 | 1.0000 | 0.0173 |
| Direct | 0.0463 | 0.0222 | 1.0000 |
| CnF.PRoPHET | 0.5640 | 0.2393 | 0.2884 |
| CnR.PRoPHET | 0.9723 | 0.7070 | 0.0919 |
| DF.PRoPHET | 0.8399 | 0.4620 | 0.3050 |
| COORD.PRoPHET | 0.7980 | 0.4298 | 0.3360 |

Regarding the evaluation of routing overhead, we wish to compare the number of transmissions that a routing protocol performed with those performed by $OPT_F$, again in order to quantify the "difficulty" of reaching the destination. We are interested in the transmissions performed for all packets (either delivered or not). This is because a realistic routing protocol does not know if a packet is deliverable or not, therefore it will try to forward (or replicate in the multi-copy case) any generated packet. Based on these principles, we introduce the *Normalized Routing Overhead (NRO)* Index of protocol *RP* as:

$$NRO\,(RP) = \frac{\displaystyle\sum_{i \in Dlv(RP)} \frac{1 + F_i\,(OPT_F)}{1 + F_i\,(RP)} + \sum_{i \in NDlv(RP)} \frac{1}{1 + F_i\,(RP)}}{|Dlv\,(RP) \cup NDlv\,(RP)|}\,, \tag{15}$$

where $NDlv(\cdot)$ denotes the set of packets not delivered and $F_i(\cdot)$ the number of transmissions for packet $i$. Similar to *NDD*, a high *NRO* value indicates a good performance and the highest score corresponds to the optimal algorithm, i.e. $NRO(OPT_F) = 1$. Note that we use two sums in *NRO*; one for delivered packets and one for non-delivered ones. For a delivered packet, a routing protocol will receive the highest score if the optimal number of transmissions $F_i(OPT_F)$ is performed. On the other hand, for non-delivered packets the optimal strategy would be to perform no transmissions. In this case, a protocol will receive a small score when $F_i(RP)$ is large and will only receive the highest score when $F_i(RP) = 0$. Concluding, the *NRO* of a routing protocol is 1 (the highest value) only if it delivered packets with the optimal number of transmissions and did not perform any redundant transmissions for non-delivered packets.

The proposed performance metrics effectively address all the challenges discussed previously. In particular, all of them are bounded between 0 and 1, with 1 indicating the optimal performance in the examined network. Furthermore, their values are monotonically increasing, which means that a higher value indicates a better performance. Therefore all the proposed metrics can be used as benefit criteria, thus enabling the use of decision-making methods that require all criteria to be either benefit or cost ones. In Table 4 we provide the average values of the proposed performance metrics for the same simulation scenario that we examined in Table 2. As one can see, Epidemic Routing achieves the optimal performance value for the *NDR* and the *NDD* metrics but its performance value for the *NRO* is close to 0 because of its excessive number of transmissions. On the other hand, Direct Delivery achieves the optimal performance value for the *NRO* but its performance values for the other metrics are very low compared to the other algorithms. Notice that when using the proposed metrics, both Delegation Forwarding

(DF) and COORD achieve better performance values for the *NDD* metric compared to Direct Delivery, which is reasonable based on a previous discussion. In addition, DF and COORD also achieve better performance values for the *NRO* metric than the single-copy Compare and Forward (CnF) approach. Even though CnF performed less total transmissions, it was also unable to deliver a significant fraction of the deliverable packets, i.e., many of these transmissions were not useful in the sense that they were not used for delivering packets. Finally, note that, contrary to the case of traditional metrics (Table 3), here the values reported in Table 4 are the same regardless of the set of examined algorithms.

As a final note, observe that the previously discussed drawbacks of traditional performance metrics are generic and do not only pertain to the use of those metrics in the context of MCDM-based performance evaluation of opportunistic routing protocols. As a result, we believe that the proposed metrics can be beneficial even outside the context of MCDM-based evaluation, i.e., when used as baseline metrics in traditional performance evaluation methodologies.

### *4.3. Variability and Interdependencies of Criteria*

After constructing a decision matrix with the proposed performance metrics as evaluation criteria, we must determine the relative importance of each criterion. Most of the proposed weighting methods consider the amount of variation in the values of a criterion as an indicator of its importance. However, since this approach examines the importance of a criterion in isolation from the other criteria, several weighting methods that utilize the correlation among criteria have also been proposed. CRITIC [4] was the first weighting method to follow the latter approach. It calculates the importance of each criterion based on the standard deviation of its observed values and then increases this importance according to the amount of conflict that the criterion has with the other criteria. To assess the amount of conflict between two criteria, CRITIC uses their linear correlation coefficient. While we believe that CRITIC is heading in the right direction, there are several pitfalls associated with it. More specifically:

- We argue that the importance of each criterion should be characterized by its *independence* from other criteria. In other words, unlike CRITIC, a highly independent criterion should be more important than criteria that present a high degree of negative correlation.
- If the Pearson correlation between two criteria is equal to 0, this only implies that there is no *linear* association. However, *non-linear* associations are not ruled out. In order to be able to determine when two criteria are independent, another measure of dependence must be used.
- Unlike CRITIC, we argue that the importance of a criterion should be determined as a *fraction* of its variability. For totally independent criteria the fraction should be 1 while for dependent ones the fraction should be reduced based on the degree of correlation between criteria. In other words, we *reduce* the importance of highly dependent criteria instead of increasing the importance of highly independent ones.

The proposed *Variability and Interdependencies of Criteria (VIC)* weighting method addresses the aforementioned issues. In particular, VIC calculates the importance of each criterion as

$$g_j = \frac{\sigma_j}{\sum_{k=1}^{m} \mathcal{R}_{j,k}}, \quad j = 1, \ldots, m, \tag{16}$$

where $\sigma_j$ is the standard deviation of the criterion $\mathbf{c}_j$ and $\mathcal{R}_{j,k}$ is the distance correlation (dCor) [45, 46] between criteria $\mathbf{c}_j$ and $\mathbf{c}_k$. The objective weight of each criterion is then given by:

$$w_j = \frac{g_j}{\sum_{k=1}^{m} g_k}, \quad j = 1, \ldots, m. \tag{17}$$

The main reason we selected the distance correlation as the measure of dependence for the VIC method is because it is equal to 0 if and only if the two criteria are independent. Furthermore, while the Pearson correlation ranges from -1 to 1, the distance correlation ranges from 0 to 1. Therefore, according to Eq. (16), the importance of a totally independent criterion corresponds to its standard deviation. This is because $\mathcal{R}_{j,k} = 0 \ \forall j \neq k$ and $\mathcal{R}_{j,j} = 1$. In the case of a highly dependent criterion $\exists k : \mathcal{R}_{j,k} > 0$ and $\mathcal{R}_{j,j} = 1$, thus the importance reduces since the denominator in Eq. (16) will be greater than 1 regardless of whether the correlation is positive or negative.

Recall that VIC is an objective weighting method. However, it is possible that a protocol designer follows specific guidelines for prioritizing the various performance aspects. For example, a protocol designer may value limited energy consumption, e.g., due to the resulting limited cost, while she/he can tolerate a significant impact on other performance aspects such as the delivery ratio. In such cases, one should resort to a subjective weighting method. Still, we believe that a subjective weight assignment should be combined with VIC because the latter can provide awareness regarding the performance trade-offs in the examined network and at the same time act as a guideline against the pitfalls stemming from criteria dependencies.

#### *4.3.1. Numerical Examples*

We present some numerical examples in order to examine the weights that different objective weighting methods produce in various scenarios. More specifically, we present the weight vectors obtained by the Mean Weights (MW) method, the Standard Deviation (SD) method, the CRITIC method with linear correlation coefficients as it was originally proposed [4], a modified version of CRITIC with distance correlation coefficients (CRITIC.dCor), and the proposed VIC method.

Table 5 depicts a decision matrix with seven alternatives and three highly dependent criteria. It also reports the standard deviations, Pearson correlation coefficients and distance correlation coefficients as well as the weights that each method determines for each criterion. Since the standard deviations of all three criteria are equal in this example, the SD method considers every criterion as equally important, just like the MW method always

Table 5: An example illustrating the determination of weights for highly dependent criteria using different methods.

| Alternatives | Criteria | | |
|---|---|---|---|
| | $\mathbf{c}_1$ | $\mathbf{c}_2$ | $\mathbf{c}_3$ |
| $\mathbf{a}_1$ | 0 | 0 | 1 |
| $\mathbf{a}_2$ | 0.1 | 0.2 | 0.8 |
| $\mathbf{a}_3$ | 0.2 | 0.4 | 0.6 |
| $\mathbf{a}_4$ | 0.3 | 0.7 | 0.3 |
| $\mathbf{a}_5$ | 0.6 | 0.8 | 0.2 |
| $\mathbf{a}_6$ | 0.8 | 0.9 | 0.1 |
| $\mathbf{a}_7$ | 1 | 1 | 0 |

| | $\sigma_j$ | $\rho_{j,k}$ | | | $\mathcal{R}_{j,k}$ | | |
|---|---|---|---|---|---|---|---|
| | | $\mathbf{c}_1$ | $\mathbf{c}_2$ | $\mathbf{c}_3$ | $\mathbf{c}_1$ | $\mathbf{c}_2$ | $\mathbf{c}_3$ |
| $\mathbf{c}_1$ | 0.3493 | 1 | 0.93 | −0.93 | 1 | 0.94 | 0.94 |
| $\mathbf{c}_2$ | 0.3493 | 0.93 | 1 | −1 | 0.94 | 1 | 1 |
| $\mathbf{c}_3$ | 0.3493 | −0.93 | −1 | 1 | 0.94 | 1 | 1 |

| Method | Weights | | |
|---|---|---|---|
| | $w_1$ | $w_2$ | $w_3$ |
| MW | 0.3333 | 0.3333 | 0.3333 |
| SD | 0.3333 | 0.3333 | 0.3333 |
| CRITIC | 0.2500 | 0.2586 | 0.4914 |
| CRITIC.dCor | 0.5000 | 0.2500 | 0.2500 |
| VIC | 0.3382 | 0.3309 | 0.3309 |

does. The CRITIC method considers $\mathbf{c}_3$ as the most important because it has a perfect negative linear correlation with $\mathbf{c}_2$ and a strong negative linear correlation with $\mathbf{c}_1$. However, if we visualize a perfect negative linear correlation between two criteria as a perfect trade-off, their importance should be reduced equally since selecting an alternative with a higher performance value in one criterion will result in a reduction in the other criterion. Similarly, the importance of two criteria that have a perfect positive linear correlation should also be reduced equally since selecting an alternative with a higher performance value in one criterion will also result in a higher performance in the other criterion. Thus, weakly correlated criteria should be considered as the most important since their values are minimally affected by the values of other criteria. In our example, using the CRITIC method with distance correlation instead of the linear one (i.e., Pearson correlation) results in an exaggeration of the importance of $\mathbf{c}_1$ because its importance was increased twice as much compared to the other two criteria. By reducing the importance of each criterion based on its distance correlation with other criteria, VIC produces objective weights that reflect the aforementioned rationale; since in this scenario the criteria are highly correlated, their weights should be almost equal, with $\mathbf{c}_1$ having a slightly higher weight because it is the least correlated with the other criteria.

In the previous example the weights $w_1, w_2$ and $w_3$ that VIC produces would be the same even if we used the absolute values of the Pearson correlation coefficients, i.e., account for linear associations, instead of the distance correlation coefficients that also capture non-linear associations. However, in this case we would not be able to distinguish independent criteria from ones with strictly non-linear associations. To shed light on the situation, let us consider the example in Table 6 where $\mathbf{c}_1$ has a non-linear association with $\mathbf{c}_2$ while $\mathbf{c}_3$ is totally independent from

Table 6: An example illustrating the determination of weights for criteria with non-linear associations using different methods.

| Alter. | Criteria | | | Alter. | Criteria | | |
|---|---|---|---|---|---|---|---|
| | $\mathbf{c}_1$ | $\mathbf{c}_2$ | $\mathbf{c}_3$ | | $\mathbf{c}_1$ | $\mathbf{c}_2$ | $\mathbf{c}_3$ |
| $\mathbf{a}_1$ | 0 | 0 | 0 | $\mathbf{a}_7$ | 0.6 | 1 | 0 |
| $\mathbf{a}_2$ | 0 | 0 | 1 | $\mathbf{a}_8$ | 0.6 | 1 | 1 |
| $\mathbf{a}_3$ | 0.2 | 0.5 | 0 | $\mathbf{a}_9$ | 0.8 | 0.5 | 0 |
| $\mathbf{a}_4$ | 0.2 | 0.5 | 1 | $\mathbf{a}_{10}$ | 0.8 | 0.5 | 1 |
| $\mathbf{a}_5$ | 0.4 | 1 | 0 | $\mathbf{a}_{11}$ | 1 | 0 | 0 |
| $\mathbf{a}_6$ | 0.4 | 1 | 1 | $\mathbf{a}_{12}$ | 1 | 0 | 1 |

| | $\sigma_j$ | $\rho_{j,k}$ | | | $\mathcal{R}_{j,k}$ | | |
|---|---|---|---|---|---|---|---|
| | | $\mathbf{c}_1$ | $\mathbf{c}_2$ | $\mathbf{c}_3$ | $\mathbf{c}_1$ | $\mathbf{c}_2$ | $\mathbf{c}_3$ |
| $\mathbf{c}_1$ | 0.3416 | 1 | 0 | 0 | 1 | 0.519 | 0 |
| $\mathbf{c}_2$ | 0.4082 | 0 | 1 | 0 | 0.519 | 1 | 0 |
| $\mathbf{c}_3$ | 0.5000 | 0 | 0 | 1 | 0 | 0 | 1 |

| Method | Weights | | |
|---|---|---|---|
| | $w_1$ | $w_2$ | $w_3$ |
| MW | 0.3333 | 0.3333 | 0.3333 |
| SD | 0.2733 | 0.3266 | 0.4001 |
| CRITIC | 0.2733 | 0.3266 | 0.4001 |
| CRITIC.dCor | 0.2397 | 0.2865 | 0.4738 |
| VIC | 0.2263 | 0.2705 | 0.5031 |

both $\mathbf{c}_1$ and $\mathbf{c}_2$. Note that the Pearson correlation coefficients are equal to 0 for all combinations of criteria, i.e., $\rho_{j,k} = 0 \; \forall j \neq k$. At the same time, the non-linear association of $\mathbf{c}_1$ and $\mathbf{c}_2$ is only captured by the distance correlation, i.e., $\mathcal{R}_{1,2} = \mathcal{R}_{2,1} \neq 0$. As a result, the weights derived using CRITIC are identical to the weights derived using the SD method. On the other hand, VIC assigns a significantly higher weight to $\mathbf{c}_3$ because its variation is the highest and it is also independent of $\mathbf{c}_1$ and $\mathbf{c}_2$. In Section 5.2 we examine the correlation of the proposed performance metrics on several datasets and we provide the weight vector that each method produces.

### 4.4. Selection of a Decision-Making Method

Up to this point we have addressed the construction of a decision matrix, where each routing protocol corresponds to an alternative and each performance metric corresponds to a benefit criterion, using the normalized performance metrics proposed in Section 4.2. Moreover, the VIC method, proposed in the previous section, tackles the problem of deriving a suitable weight vector that captures the relative importance of each criterion. Provided the decision matrix and the weight vector, we can now use a decision-making method in order to produce a ranking of routing protocols. However, if we try to select the best decision-making method using an MCDM approach then we reach a decision-making paradox [47]. This means that we cannot consider a candidate decision-making method as the best one because we have to use the best decision-making method itself in order to come to that conclusion, i.e., there is an intrinsic circularity problem in making this conclusion. Therefore, our justification for the selection of a decision-making method is based on the way that the scores of each method can be interpreted for our specific problem. Clearly, it is not possible to

Table 7: An example illustrating how different decision-making methods handle trade-offs between criteria.

| Alternatives | Criteria | | SAW | MEW | mTOPSIS |
|---|---|---|---|---|---|
| | $\mathbf{c}_1$ | $\mathbf{c}_2$ | | | |
| $\mathbf{a}_1$ | 0 | 1 | 0.5000 | 0 | 0.5000 |
| $\mathbf{a}_2$ | 0.25 | 0.75 | 0.5000 | 0.4330 | 0.5000 |
| $\mathbf{a}_3$ | 0.5 | 0.5 | 0.5000 | 0.5000 | 0.5000 |
| $\mathbf{a}_4$ | 0.75 | 0.25 | 0.5000 | 0.4330 | 0.5000 |
| $\mathbf{a}_5$ | 1 | 0 | 0.5000 | 0 | 0.5000 |

Table 8: Overall score of routing protocols using different decision-making methods with the proposed normalized metrics and VIC (Dartmouth dataset).

| | SAW | MEW | mTOPSIS |
|---|---|---|---|
| Epidemic | 0.5796 | 0.1762 | 0.5315 |
| Direct | 0.4485 | 0.1446 | 0.4685 |
| CnF.PRoPHET | 0.3685 | 0.3449 | 0.3597 |
| CnR.PRoPHET | 0.5323 | 0.3284 | 0.5086 |
| DF.PRoPHET | 0.5208 | 0.4721 | 0.5031 |
| COORD.PRoPHET | 0.5124 | 0.4754 | 0.4962 |

prove that our selection is the best one since this proof involves the aforementioned circularity problem.

In the context of opportunistic networks, we recommend the use of the MEW method [27, 28] for evaluating the performance of routing protocols. This is mainly because MEW provides a more suitable handling of the trade-offs between different criteria. Let us consider the decision matrix in Table 7 where there are two conflicting criteria that are equally important. Evidently, the SAW method [26] and the modified version of the TOPSIS method [3] produce the same rank for all the alternatives. On the contrary, MEW considers the third alternative as the best one because its performance is more balanced. Since most opportunistic routing protocols target a middle ground between the extreme routing overhead of Epidemic Routing [5] and the unsatisfactory delivery delays of Direct Delivery [6], MEW comes forward as the most suitable for their evaluation because routing protocols with a low performance value for at least one criterion will receive an overall low score.

In general, if the MEW score of a routing protocol is close to 0, this means that it performs poorly in terms of at least one of the performance metrics. In order for a routing protocol to have a MEW score equal to 1, it must perform optimally in terms of all performance metrics with non-zero weights. A high overall score indicates a protocol with a balanced and high efficiency performance, i.e., a protocol that performs efficiently with respect to all important performance metrics.

To further strengthen our reasoning for selecting the MEW method, Table 8 reports the overall performance scores of the routing protocols that we examined in Table 4 using the VIC weighting method with three different decision-making methods: SAW, MEW, and mTOPSIS. Even though Epidemic Routing produces by far the most packet transmissions, SAW and mTOPSIS consider it as the best alternative, followed by the multi-copy Compare and Replicate (CnR) approach. This is because the high performance values of both protocols for the normalized delivery ratio and normalized delivery delay metrics overshadow their very low performance value for the normalized routing overhead metric. Moreover, SAW and mTOPSIS assign a higher overall score to Direct Delivery compared to the single-copy Compare and Forward (CnF) approach, because Direct Delivery achieves the optimal performance value for the normalized routing overhead. This is clearly unreasonable, since Direct Delivery has by far the worst performance in terms of the other two metrics. On the other hand, the approach of MEW is more reasonable since it assigns low overall scores to Epidemic Routing and Direct Delivery because they perform poorly for at least one metric. In the following section, we provide the ranking of an extended set of routing protocols using the MEW decision-making method and the VIC weighting method on datasets with vastly different characteristics, while we also discuss the rankings that other combinations of decision-making and weighting methods produce.

Table 9: Characteristics of the investigated datasets.

| | Reality Mining | INFOCOM 2005 | Lyon | Dartmouth |
|---|---|---|---|---|
| Number of Nodes | 97 | 41 | 242 | 738 |
| Duration (days) | 283 | 3 | 2 | 14 |
| Granularity (secs) | 300 | 120 | 20 | 1 |
| Network Interface | Bluetooth | Bluetooth | RFID | Wi-Fi |

## 5. Case Studies

### 5.1. Simulation Setup

We selected four datasets of varying scale to evaluate the performance of routing protocols for opportunistic networks according to the proposed framework. More specifically, the datasets that we selected are the following: Reality Mining [48, 49], INFOCOM 2005 [50, 51], Lyon [52, 53], and Dartmouth [54, 55, 56]. The Lyon dataset was downloaded from the website of the SocioPatterns collaboration [57], while the rest datasets were downloaded from the website of the CRAWDAD archive [58]. The investigated datasets differ significantly in terms of the number of participants while their duration varies from a few days to several months. Table 9 summarizes their characteristics.

For the simulations we use the Adyton simulator [59]. Adyton is an open source and freely available simulator that is capable of processing contact traces and supports several opportunistic routing protocols. We investigate the performance of the following set of routing algorithms:

- Epidemic Routing (Epidemic) [5].
- Direct Delivery (Direct) [6].
- Compare and Forward (CnF) [12].
- Compare and Replicate (CnR) [13].
- Delegation Forwarding (DF) [19].
- COORD [20].
- Spray and Wait (SnW) [14].
- LSF-Spray and Wait (LSF-SnW) [16].
- Spray and Focus (SnF) [15].
- SimBetTS [18].

- EBR [17].

For the routing protocols that is required to predefine a maximum number of replicas ($L$) (i.e., Spray and Wait, LSF-Spray and Wait, Spray and Focus, SimBetTS, and EBR), we assigned four values: $L = 2$, $L = 4$, $L = 8$, and $L = 16$. In the following, the example notation $X.L = 4$ indicates protocol $X$ with $L = 4$. For the routing protocols that can operate using different utility functions (i.e., Compare and Forward, Compare and Replicate, Delegation Forwarding, and COORD), we used one destination-independent and three destination-dependent utility metrics. In particular, we used the Last Time Seen (LTS) [9], the Destination Encounters (DestEnc) [11], the Encounters (Enc) [11], and the latest version of the PRoPHET utility with the default parameter settings [10]. In the following, the notation $X.Y$ indicates protocol $X$ that uses the $Y$ utility function. For example, *CnR.PRoPHET* is used to refer to the Compare and Replicate protocol when it uses the PRoPHET utility. For multi-copy routing protocols we implemented the VACCINE anti-packet scheme [60] which is used to erase redundant replicas of a packet after its successful delivery.

We were also able to efficiently determine the performance of the two versions of the Optimal Routing algorithm, i.e., $OPT_D$ and $OPT_F$, using the Epidemic Routing algorithm in Adyton. Recall that, if at least a path exists, i.e., a packet is deliverable, $OPT_D$ delivers the packet using the fastest path with the minimum number of hops while $OPT_F$ delivers it using the shortest (in hops) path with the minimum delivery delay. Moreover, bear in mind that Epidemic Routing follows a flooding-based approach, i.e., it delivers multiple packet copies, each one corresponding to a different path. Evidently, the first copy delivered by Epidemic algorithm follows the fastest path. Yet, a minor modification is required to correctly identify minimum-hop paths. More specifically, in Epidemic Routing an intermediate node will typically store the first copy received and reject all subsequent ones. This strategy suppresses possible minimum-hop paths since copies following such paths may arrive later. To tackle the problem, we opt to allow a node to update the hop count of a stored copy if another copy with a smaller hop count is available through a later contact. This modification allows the destination node to receive the minimum-hop copy. At the same time, the fastest arriving copy will contain the correct hop count. With this modified algorithm it is possible to capture the performance of both $OPT_D$ and $OPT_F$. For $OPT_D$, it suffices to record, for each deliverable packet, the delay and the hop count of the first copy delivered to the destination. Note that, since $OPT_D$ and $OPT_F$ are single-copy algorithms, the hop count is equivalent to the number of transmissions. For $OPT_F$, from the multitude of copies received by the destination, we record the delay and the hop count of the copy that traveled the minimum number of hops.

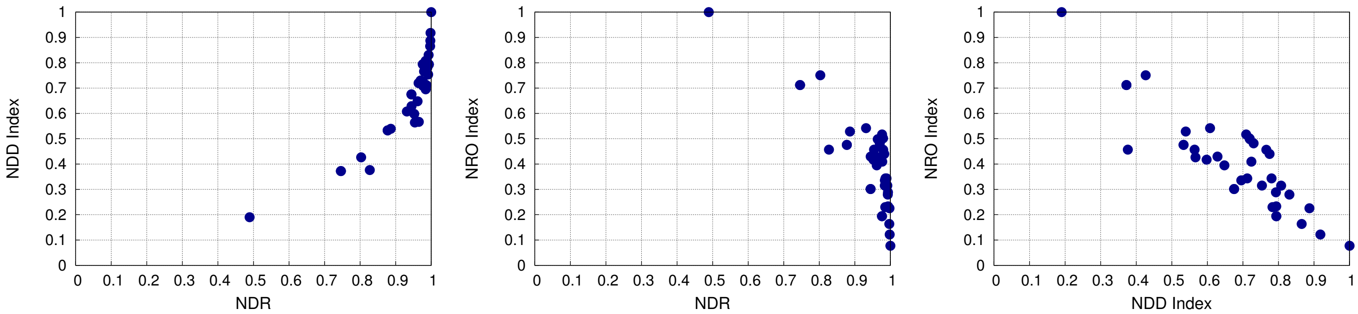
For each simulation, we generated 10000 packets of fixed size with a random pair of source and destination nodes. However, because some nodes are not active for the entire duration of the respective dataset, each node can be the source or the destination only for packets that were generated during its presence in the network. Like most evaluation studies in the literature, we did not consider any resource constraints during our simulations. To avoid statistical bias, the results were collected after a warm-up period and before a cool-down period with the duration of each one being the 20% of the total simulation time. We simulated 25 repetitions of each scenario in order to calculate the average values and 95% confidence intervals of the traditional and the proposed performance metrics.

### 5.2. Simulation Results

#### 5.2.1. Performance Evaluation on the Reality Mining dataset

The Reality Mining dataset [48, 49] consists of contacts between students and faculty members at the MIT. It is one of the most widely used datasets, mainly because of its large number of participants and long duration.

Fig. 1 presents three scatter plots, one for each combination of the three proposed normalized metrics. In each scatter plot, a point represents the performance achieved by one of the routing protocols with respect to two metrics. The three scatter plots allows us to illustrate the associations between the observed normalized performance values. According to Fig. 1(a), the lowest normalized delivery ratio (NDR) that a routing protocol achieved on this dataset is almost equal to 0.5, which corresponds to the performance of Direct Delivery. In other words, half of the deliverable packets could be delivered directly, i.e. with the least possible routing overhead. As we would expect, Direct Delivery also achieves the lowest value for the normalized delivery delay (NDD) index, i.e., high average delivery delay. The rest of the routing protocols are able to deliver a lot more packets with smaller delays (i.e., higher NDD indices). Interestingly, Fig. 1(a) clearly illustrates that the




(a) (b) (c)

Figure 1: Scatter plots of the normalized performance metrics on the Reality Mining dataset.

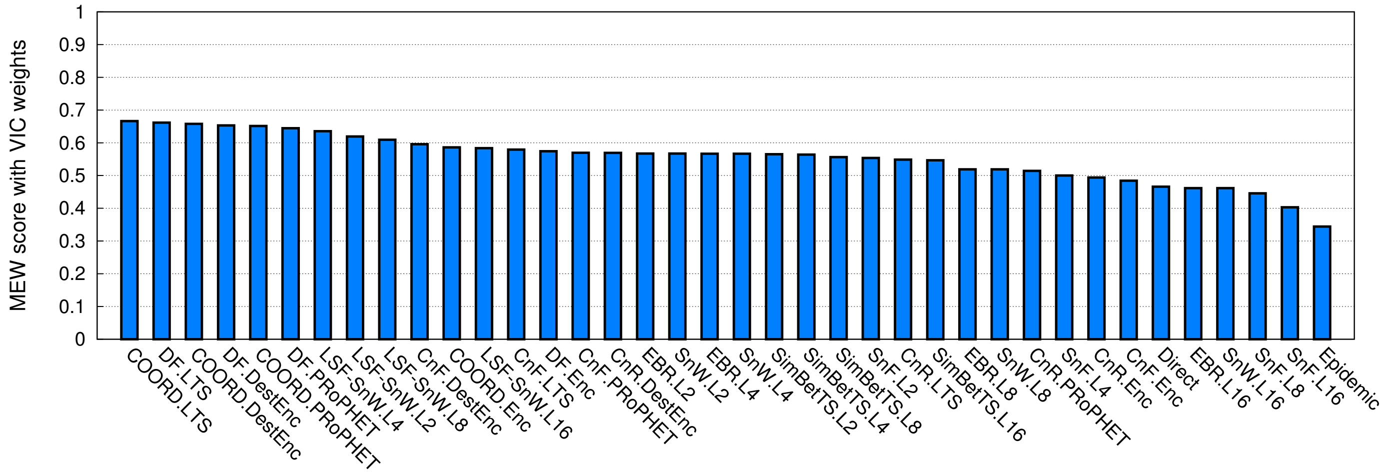



Figure 2: Ranking of the routing protocols using the MEW method with VIC weights on the Reality Mining dataset.

Table 10: Relative importance of each performance metric on the Reality Mining dataset according to different weighting methods.

| | MW | SD | CRITIC | CRITIC.dCor | VIC |
|---|---|---|---|---|---|
| *NDR* | 0.3333 | 0.2197 | 0.1600 | 0.2014 | 0.2173 |
| *NDD* | 0.3333 | 0.3741 | 0.2794 | 0.3102 | 0.3662 |
| *NRO* | 0.3333 | 0.4062 | 0.5606 | 0.4884 | 0.4165 |

association between *NDR* and *NDD* is non-linear. The increase of the delivery capability for all protocols comes at the cost of a significant reduction of their performance in terms of the normalized routing overhead (NRO) index, i.e., increased routing overhead. Fig. 1(b) confirms that most protocols deliver about the same number of packets but with a significantly different routing overhead. This is an indication that in this dataset the routing overhead is of increased importance. Finally, Fig. 1(c) highlights the negative correlation (i.e., the conflicting nature) between the NDD and the NRO indices, which is reasonable since more packet copies typically result in reduced delivery delay because it is more probable that a packet copy will follow the optimal delay path.

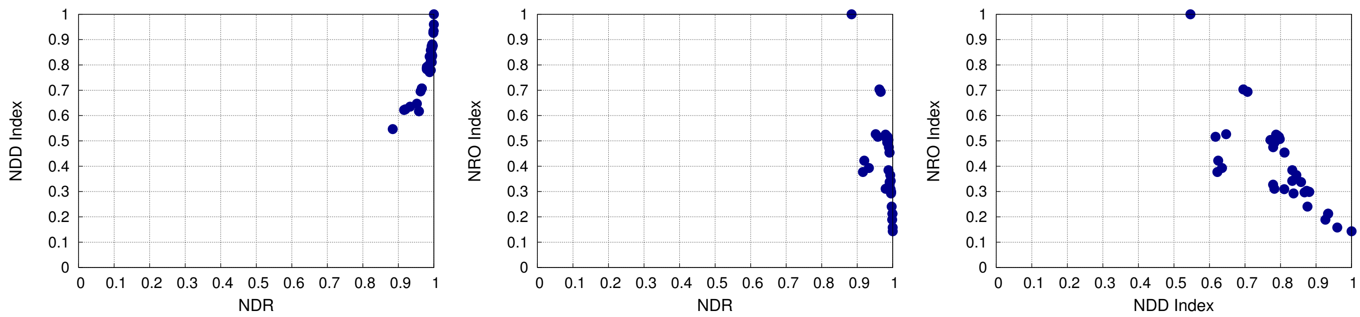
We constructed a decision matrix with the proposed performance metrics as the set of criteria and the routing protocols as the set of alternatives. Based on this matrix, we used different weighting methods to determine the weight of each metric. Table 10 reports the results. All the weighting methods, except for the MW method, consider the NRO index as the most important performance metric in the Reality dataset, followed by the NDD index. CRITIC assigns by far the highest weight value to NRO, because of its strongly conflicting associations with both NDR and NDD. If we use the MEW method with the weight vector of the CRITIC method, LSF-Spray and Wait with $L$=2 receives the highest overall score although it delivers significantly less packets than almost every other routing protocol and it does so with a significantly higher delivery delay. Similarly, the mTOPSIS method with CRITIC weights also assigns the highest score to LSF-Spray and Wait with $L = 2$ while it assigns the second highest score to Direct Delivery. Worse still, if we use the CRITIC weights with the SAW method, which is the suggested decision-making method in [4], the highest SAW score corresponds to the Direct Delivery protocol. Apparently, the reason for all these results is the fact that CRITIC overestimates the importance of the NRO index. On the contrary, the weight vector that results with the VIC method is more reasonable. Indeed, it is close to that of the SD method because NDD and NRO present a more significant variation compared to NDR. However, VIC slightly increases the importance of NRO because its correlation with NDR is smaller than that between the NDD and NDR. When using the VIC weight vector, all three decision-making methods rank COORD.LTS at the top, followed by DF.LTS.

In Fig. 2 we provide, in descending order, the MEW score of each routing protocol when the VIC method is used. COORD and Delegation Forwarding (DF) with the destination-dependent utility metrics (e.g., LTS, DestEnc, PRoPHET) receive the highest rankings in the Reality dataset because they achieve a good trade-off between delivery ratio and routing overhead. When using a destination-independent utility, the routing overhead increases (NRO index decreases) for both algorithms without a parallel increase in the delivery capability or a decrease of delivery delays. This results in lower overall scores. LSF-Spray and Wait with $L = 4$ receives the next best ranking because it performs better, in terms of NRO, than any version of COORD and DF. However, its performance in terms of the other two performance metrics is notably worse. Note that, reasonably, both Direct Delivery and Epidemic receive low scores because they are poor performers in at least one dimension.

### 5.2.2. *Performance Evaluation on the INFOCOM 2005 dataset*

The INFOCOM 2005 dataset [50, 51] is another commonly used dataset for the performance evaluation of opportunistic routing protocols. During the INFOCOM Student Workshop in 2005, 41 devices were distributed to attendees and recorded their contacts.

It is clear from Fig. 3 that this dataset is not challenging in terms of delivering packets. Even Direct Delivery is able to deliver more than 85% of the generated packets and almost 90% of the packets that the optimal algorithm can deliver. Moreover, it also achieves a much higher NDD index (lower delay)




(a) (b) (c)

Figure 3: Scatter plots of the normalized performance metrics on the INFOCOM 2005 dataset.

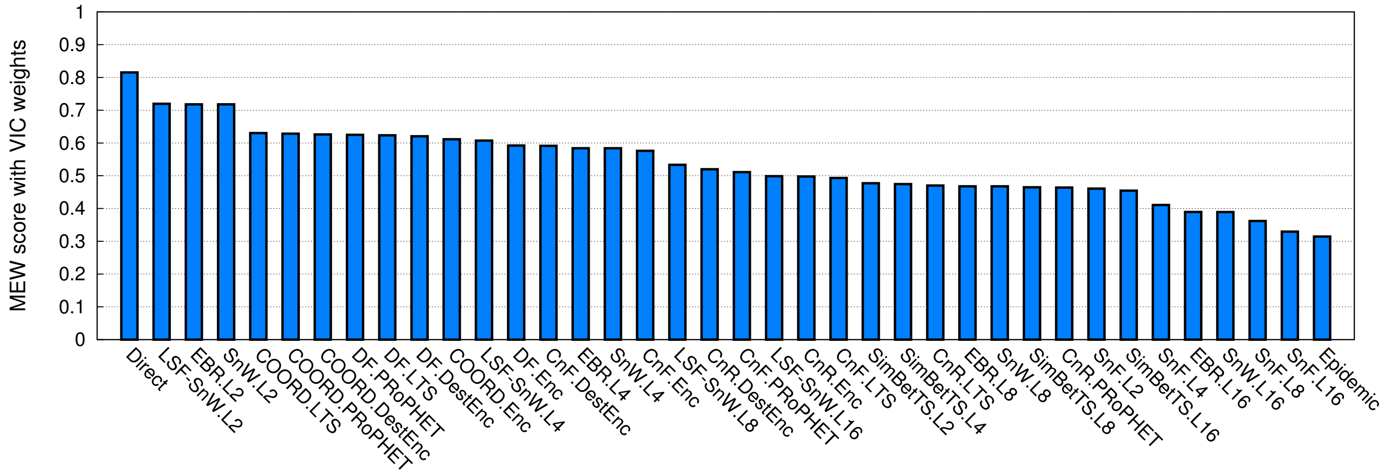



Figure 4: Ranking of the routing protocols using the MEW method with VIC weights on the INFOCOM 2005 dataset.

Table 11: Relative importance of each performance metric on the INFOCOM 2005 dataset according to different weighting methods.

| | MW | SD | CRITIC | CRITIC.dCor | VIC |
|---|---|---|---|---|---|
| *NDR* | 0.3333 | 0.0870 | 0.0557 | 0.0755 | 0.0846 |
| *NDD* | 0.3333 | 0.3420 | 0.2349 | 0.2284 | 0.3207 |
| *NRO* | 0.3333 | 0.5711 | 0.7094 | 0.6962 | 0.5947 |

compared to what it usually achieves on other datasets. Overall, the range of values for all the performance metrics is rather small. However, we can still observe some variation in the performance values of NDD and a greater one in NRO. Based on these observations, it is obvious that, in this scenario, the most important performance metric is NRO. We would therefore expect all the weighting methods, except for MW, to assign a high weight value to NRO and a low weight value to NDR. As seen in Table 11, CRITIC indeed assigns the highest weight value to NRO and the lowest weight value to NDR. However, it seems that CRITIC, again, overestimates the importance of NRO and therefore underestimates the importance of NDD.

Regardless of the weighting method, all the decision-making methods consider Direct Delivery as the best alternative. Fig. 4 presents the ranking of protocols based on the MEW method with the VIC weight vector. Direct Delivery achieves by far the highest overall MEW score. This should not come as a surprise, given that Direct Delivery performs optimally in terms of NRO, which is the most important metric in this scenario. At the same time, it also achieves reasonable performance values for the other two performance metrics (recall that this dataset does not present significant challenges for delivering packets). Observe that the next three routing protocols in the ranking order (LSF-SnW.L2, EBR.L2 and SnW.L2) share some characteristics. They all perform limited replication (only two replicas) and once a node is left with only one replica of a packet it will wait to meet its destination node in order to deliver it. Again, since, in this scenario, delivering packets is rather easy and NRO is the most important metric, these protocols are next in the ranking order due to their very small number of transmissions. Furthermore, every version of COORD is ranked higher than the corresponding version of DF since it performs less transmissions without a noticeable decrease in the other performance metrics. Finally, as we would expect, Epidemic Routing has the lowest overall score.

### 5.2.3. Performance Evaluation on the Lyon dataset

The Lyon dataset [52, 53, 57] consists of close-range contacts between 232 children and 10 teachers over the course of two school days in a primary school in Lyon, France. The fact that these students were from 10 different classes should offer a challenge for most routing protocols.[3] Indeed, as we can see

[3] A visualization of their contacts during the first school day is available at `https://vimeo.com/31490438`.

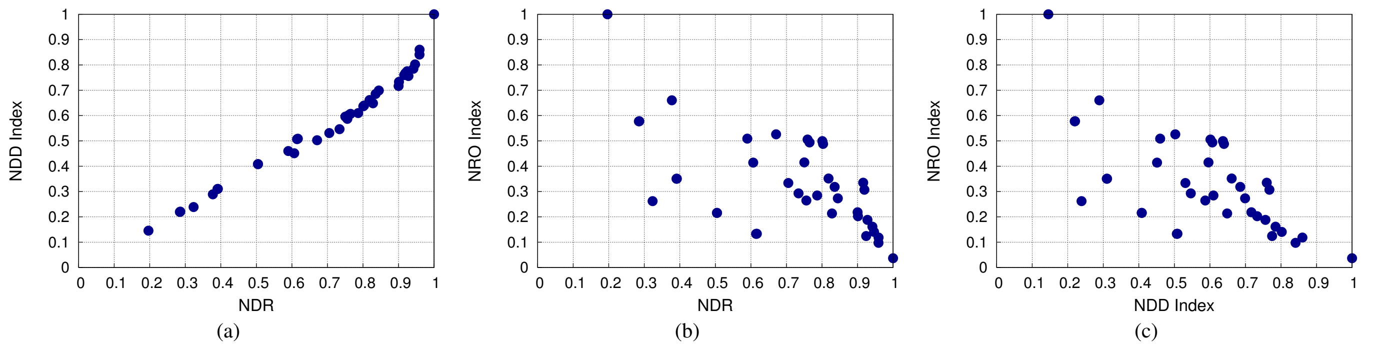



Figure 5: Scatter plots of the normalized performance metrics on the Lyon dataset.

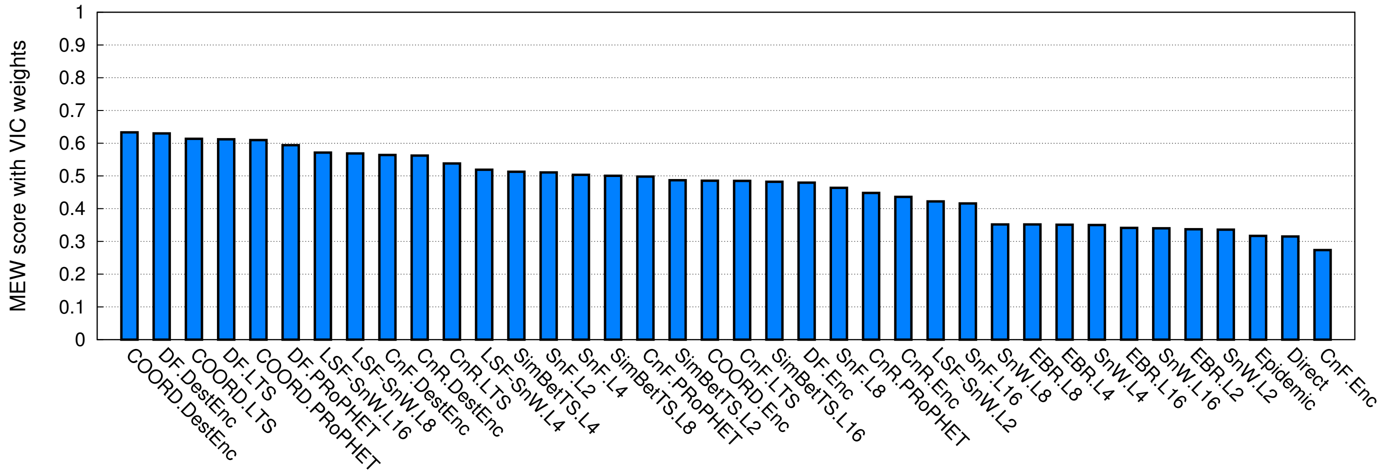



Figure 6: Ranking of the routing protocols using the MEW method with VIC weights on the Lyon dataset.

in Fig. 5, Direct Delivery was only able to deliver about 20% of the packets delivered by the optimal routing algorithm. The large variability of NDR is evident, as well as its high correlation with NDD. On the contrary, NRO depends less on the other two metrics, i.e., routing protocols with similar performance values for NRO may have very different performance values for the other performance metrics.

Table 12 provides the weight vectors produced by different methods. Note that, in contrast to the previous datasets, here different methods assign noticeably different weight values. The SD method considers NDR as the most important criterion, followed by NDD. This is because the SD method does not take into account the interdependences of the criteria. CRITIC increases the importance of NRO, which has the least variation in its values, because it is negatively correlated with the other two criteria. Similar weights are derived from the modified CRITIC method that relies on the distance correlation coefficients of the criteria instead of their linear correlation coefficients. A different assignment of weights is given by the VIC method, which considers NDR and NRO as almost equally important metrics. This is because although NDR has the highest amount of variation, it also has a very strong correlation with NDD. On the other hand, NRO has the lowest amount of variation but, at the same time, it presents a moderate correlation with the other criteria.

Table 12: Relative importance of each performance metric on the Lyon dataset according to different weighting methods.

| | MW | SD | CRITIC | CRITIC.dCor | VIC |
|---|---|---|---|---|---|
| *NDR* | 0.3333 | 0.3631 | 0.2761 | 0.2885 | 0.3441 |
| *NDD* | 0.3333 | 0.3290 | 0.2554 | 0.2456 | 0.3084 |
| *NRO* | 0.3333 | 0.3080 | 0.4686 | 0.4659 | 0.3475 |

The ranking of the routing protocols with the MEW decision-making method and the VIC weighting method is given in Fig. 6. Similarly to the simulation results on the Reality Mining dataset, COORD and DF with destination-dependent utility metrics are the highest ranked. However, on the Lyon dataset, both algorithms were able to deliver noticeably more packets with the DestEnc utility instead of the LTS utility. In fact, COORD and DF were able to deliver even more packets with the PRoPHET utility, but at the expense of significantly lower NRO (higher overhead) so their rank is lower. Observe that the versions of LSF-Spray and Wait with $L = 16$ and $L = 8$ are the next best performing algorithms. Furthermore, Direct Delivery and Epidemic Routing have significantly lower MEW scores compared to the other routing protocols. The reason is because both routing protocols perform poorly in terms of at least one performance metric. If we use SAW or mTOPSIS to calculate the overall scores, Epidemic Routing is ranked significantly higher. In particular, when ranking protocols with the SAW decision-making and the weight vector of VIC, Epi-

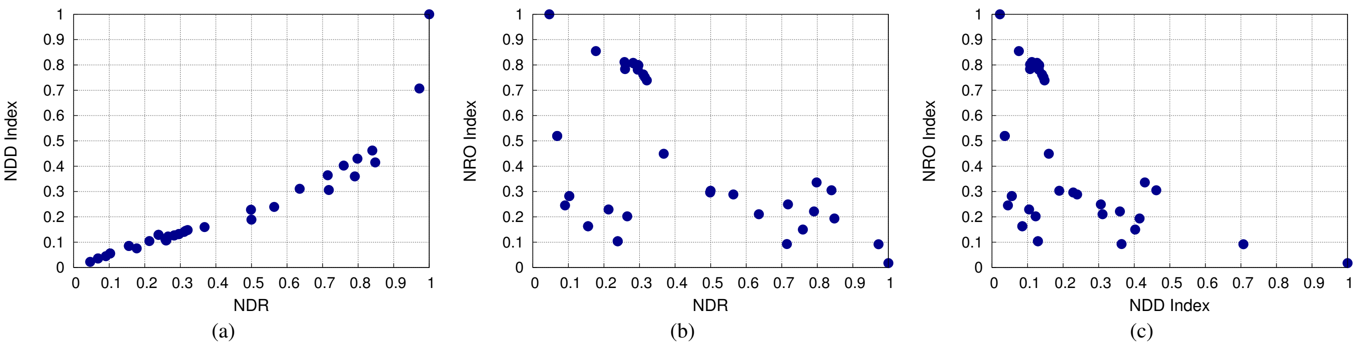



Figure 7: Scatter plots of the normalized performance metrics on the Dartmouth dataset.

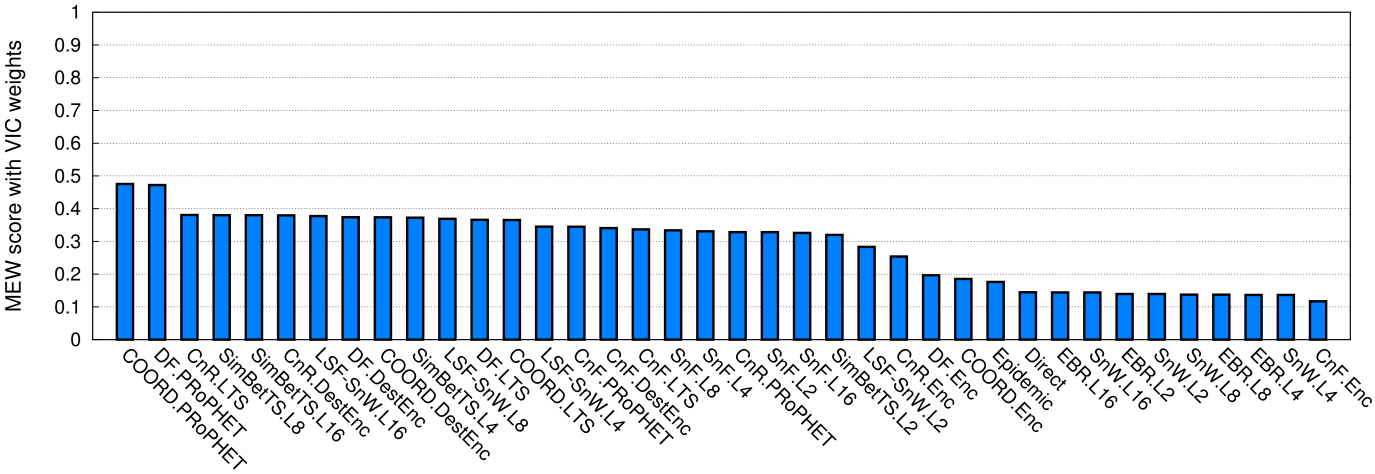



Figure 8: Ranking of the routing protocols using the MEW method with VIC weights on the Dartmouth dataset.

demic Routing is considered the second-best alternative, after COORD.PRoPHET. Worse still, when using the weight vector of the CRITIC method, even Direct Delivery is among the highest-ranked alternatives. The ranking provided by the MEW method with VIC weights seems to be the most reasonable since only CnF.Enc has a lower score than Epidemic Routing and Direct Delivery. This is because CnF.Enc performs poorly in all three performance metrics.

### 5.2.4. *Performance Evaluation on the Dartmouth dataset*

The Dartmouth dataset [54, 55, 56] contains associations of wireless cards with access points at the Dartmouth College campus. We can construct a large-scale opportunistic network if we treat each wireless card as a node and assume that two nodes are in contact when they are associated with the same access point at the same time. The approximated contacts that we extracted are between 738 nodes. From the datasets that we studied, this was the most challenging for the routing protocols as it is evident from Fig. 7. Less than 5% of the deliverable packets could be delivered in a single hop. Even the algorithms that were able to deliver more packets, they achieved that with significantly higher delivery delays compared to the optimal performance. Note that Epidemic Routing delivered every deliverable packet with the optimal end-to-end delay by performing about 560 transmissions for each generated packet.

Notice that there is a set of algorithms with similar performance values for NRO and NDR (Fig. 7(b), group of points in upper left corner). These algorithms relied on destination-dependent utility metrics that required direct contacts with the destination node in order to increase their utility values. However, due to the large-scale of this dataset, utility metrics such as DestEnc rarely consider an encountered node as suitable to carry a packet. As a result, these algorithms perform a very small number of transmissions (high NRO index) and are unable to deliver a large proportion of packets. Interestingly enough, the algorithms that rely on the transitive property of the PRoPHET utility metric perform much better. In particular, the second-highest performance values for NDR and NDD are achieved by CnR.PRoPHET. However, even though this algorithm achieves these performance values with significantly less routing overhead compared to Epidemic Routing (about 119 transmissions for each generated packet) other replication strategies perform a lot less transmissions. For example, COORD.PRoPHET performs about 11 transmissions for each generated packet and delivers about 80% of the deliverable packets.

Table 13: Relative importance of each performance metric on the Dartmouth dataset according to different weighting methods.

| | MW | SD | CRITIC | CRITIC.dCor | VIC |
|---|---|---|---|---|---|
| *NDR* | 0.3333 | 0.3611 | 0.2686 | 0.2687 | 0.3333 |
| *NDD* | 0.3333 | 0.2580 | 0.1939 | 0.1951 | 0.2389 |
| *NRO* | 0.3333 | 0.3809 | 0.5375 | 0.5362 | 0.4278 |

As can be seen in Table 13, all the weighting methods, with the exception of MW, consider NRO as the most important one. This should be expected since NRO presents the highest amount of variation and the least amount of dependence on the other criteria. However, CRITIC once again overestimates the relative importance of NRO. If we use the SAW or the mTOPSIS method with CRITIC weights, Direct Delivery ranks first, which is clearly unreasonable. Even if we use the MEW method with CRITIC weights, the algorithms that we observed to have a high NRO value but low performance values for the other two metrics, are among the highest-ranked algorithms. In Fig. 8 we present the MEW score of each routing protocol with the VIC weight vector. As we can see, COORD and DF with the PRoPHET utility metric have noticeably higher MEW scores than the other algorithms. This is because both algorithms achieve values of NDR and NDD that are among the highest ones. At the same time, they perform a very small number of transmissions compared to every other algorithm that is able to deliver a comparable number of packets. It should be noted that, if we use the SAW or the mTOPSIS method with VIC weights, Epidemic Routing is ranked first. This is because it achieves the optimal performance values for two criteria with high relative importance. This overshadows its very poor performance value for the third criterion, even though this is the most important performance metric in this scenario. On the other hand, as we can see in Fig. 8, the MEW method assigns a low overall score to the Epidemic Routing protocol because of its very poor performance in the most important criterion. Thus, the ranking of the routing protocols that was provided by the MEW method with VIC weights seems to be the most reasonable.

## 6. Discussion and Conclusions

There are some factors that should be taken into consideration in order to apply the proposed framework. First of all, the decision matrix should contain performance values of a representative set of routing protocols. Even if we are interested in the performance comparison of only a small number of routing protocols, the performance values of algorithms such as Epidemic Routing and Direct Delivery should be included. This is because such algorithms provide useful information for the performance limits imposed by the underlying network. For the same reason, the set of alternatives in the decision matrix should consist of only realistic algorithms, since non-realistic ones would affect the relative importance of the criteria that are then used for the evaluation of feasible solutions. Notice that in our analysis, the two versions of Optimal Routing were only used for the normalization of the performance metrics and they were not included in any decision matrix as alternatives.

Another point of consideration is the use of the proposed framework for evaluating protocols with specific performance requirements. In its current form, the framework follows an objective weighting of performance criteria that is based on the performance trade-offs observed in the examined network. Nonetheless, this weighting method corresponds to a type of prioritization that may not fit well with protocols that focus on specific performance goals/requirements. For example, a protocol designer may value a low number of transmissions because of the reduced energy consumption and at the same time be rather unconcerned with delay. In such a case, we should define a subjective weighting function that captures the preference towards specific performance goals, use this function in combination with the VIC weighting method (as discussed in Section 3.2) and reasonably use the framework to rank protocols with the same performance goals. In some cases a performance goal is expressed in the form of a constraint. For example, a protocol may target the best possible performance as long as the number of transmissions does not exceed a certain limit. In those cases, subjective weighting may be used as long as it penalizes protocols that do not meet the constraint. Yet, there is another more straightforward approach that does not involve subjective weighting. The only required modification is to define both $OPT_D$ and $OPT_F$ under the investigated constraint and then determine their performance. Evidently, this process depends on the type of constraint. Apparently, in this case, the framework should be used for comparing protocols operating under the same constraint.

While in this work we used three normalized performance metrics as our criteria for the evaluation of the routing protocols, more criteria could be introduced in the evaluation process. For example, a fairness index could be considered as another criterion.

Finally, we believe that, besides the proposed framework, the modified performance metrics, proposed in Section 4.2, are of a significant value on their own. This is because they are free of the pitfalls witnessed in traditional metrics, therefore they can be more robust when used as baseline criteria even with traditional evaluation methodologies, i.e., outside the context of our framework.

Several conclusions can be drawn from the performance analysis that we conducted with the proposed framework. First of all, no algorithm was able to achieve the best performance on all or most of the datasets that we studied. In small-scale opportunistic networks, Direct Delivery and replication strategies with a small maximum number of replicas were sufficient. In opportunistic networks of larger scale, utility-based replication strategies were typically considered as the best alternatives. In particular, the replication strategies of DF and COORD were often ranked at the top. Furthermore, there was not a single utility metric that performed better than the others on every dataset that we studied. More specifically, in large-scale opportunistic networks, the transitive property of the destination-dependent PRoPHET utility was critical for the performance of the utility-based routing protocols. However, in opportunistic networks of smaller scale, this utility often resulted in a noticeable increase in the number of transmissions compared to other utilities. Concluding, our case studies demonstrated the difficulty of finding a globally successful routing strategy in opportunistic networks.

## References


[1] C.-L. Hwang, K. Yoon, Multiple Attribute Decision Making, Vol. 186 of Lecture Notes in Economics and Mathematical Systems, Springer-Verlag Berlin Heidelberg, 1981.

[2] C.-L. Hwang, Y.-J. Lai, T.-Y. Liu, A new approach for multiple objective decision making, Computers & Operations Research 20 (8) (1993) 889–899. doi:10.1016/0305-0548(93)90109-V.

[3] H. Deng, C.-H. Yeh, R. J. Willis, Inter-company comparison using modified TOPSIS with objective weights, Computers & Operations Research 27 (10) (2000) 963–973. doi:10.1016/S0305-0548(99)00069-6.

[4] D. Diakoulaki, G. Mavrotas, L. Papayannakis, Determining objective weights in multiple criteria problems: The CRITIC method, Computers & Operations Research 22 (7) (1995) 763–770. doi:10.1016/0305-0548(94)00059-H.

[5] A. Vahdat, D. Becker, Epidemic routing for partially connected ad hoc networks, Tech. Rep. CS-2000-06, Duke University (2000).
URL http://issg.cs.duke.edu/epidemic/epidemic.pdf

[6] M. Grossglauser, D. N. C. Tse, Mobility increases the capacity of ad hoc wireless networks, IEEE/ACM Transactions on Networking 10 (4) (2002) 477–486. doi:10.1109/TNET.2002.801403.

[7] T. Spyropoulos, R. N. B. Rais, T. Turletti, K. Obraczka, A. Vasilakos, Routing for disruption tolerant networks: taxonomy and design, Wireless Networks 16 (8) (2010) 2349–2370. doi:10.1007/s11276-010-0276-9.

[8] Y. Zhu, B. Xu, X. Shi, Y. Wang, A survey of social-based routing in delay tolerant networks: Positive and negative social effects, IEEE Communications Surveys & Tutorials 15 (1) (2013) 387–401. doi:10.1109/SURV.2012.032612.00004.

[9] H. Dubois-Ferriere, M. Grossglauser, M. Vetterli, Age matters: Efficient route discovery in mobile ad hoc networks using encounter ages, in: Proceedings of the 4th ACM International Symposium on Mobile Ad Hoc Networking & Computing (MobiHoc), 2003, pp. 257–266. doi:10.1145/778415.778446.

[10] A. Lindgren, A. Doria, E. Davies, S. Grasic, Probabilistic routing protocol for intermittently connected networks, RFC 6693, Internet Research Task Force (IRTF) (2012).
URL http://tools.ietf.org/pdf/rfc6693.pdf

[11] V. Erramilli, A. Chaintreau, M. Crovella, C. Diot, Diversity of forwarding paths in pocket switched networks, in: Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement (IMC), 2007, pp. 161–174. doi:10.1145/1298306.1298330.

[12] T. Spyropoulos, K. Psounis, C. S. Raghavendra, Efficient routing in intermittently connected mobile networks: The single-copy case, IEEE/ACM Transactions on Networking 16 (1) (2008) 63–76. doi:10.1109/TNET.2007.897962.

[13] T. Spyropoulos, K. Psounis, C. S. Raghavendra, Efficient routing in intermittently connected mobile networks: The multiple-copy case, IEEE/ACM Transactions on Networking 16 (1) (2008) 77–90. doi:10.1109/TNET.2007.897964.

[14] T. Spyropoulos, K. Psounis, C. S. Raghavendra, Spray and wait: An efficient routing scheme for intermittently connected mobile networks, in: Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-tolerant Networking (WDTN), 2005, pp. 252–259. doi:10.1145/1080139.1080143.

[15] T. Spyropoulos, K. Psounis, C. S. Raghavendra, Spray and focus: Efficient mobility-assisted routing for heterogeneous and correlated mobility, in: Proceedings of the 5th Annual IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom), 2007, pp. 79–85. doi:10.1109/PERCOMW.2007.108.

[16] T. Spyropoulos, T. Turletti, K. Obraczka, Routing in delay-tolerant networks comprising heterogeneous node populations, IEEE Transactions on Mobile Computing 8 (8) (2009) 1132–1147. doi:10.1109/TMC.2008.172.

[17] S. C. Nelson, M. Bakht, R. Kravets, Encounter-based routing in DTNs, in: Proceedings of the 28th IEEE International Conference on Computer Communications (INFOCOM), 2009, pp. 846–854. doi:10.1109/INFCOM.2009.5061994.

[18] E. M. Daly, M. Haahr, Social network analysis for information flow in disconnected delay-tolerant manets, IEEE Transactions on Mobile Computing 8 (5) (2009) 606–621. doi:10.1109/TMC.2008.161.

[19] V. Erramilli, M. Crovella, A. Chaintreau, C. Diot, Delegation forwarding, in: Proceedings of the 9th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc), 2008, pp. 251–260. doi:10.1145/1374618.1374653.

[20] N. Papanikos, E. Papapetrou, Coordinating replication decisions in multi-copy routing for opportunistic networks, in: Proceedings of the 10th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob), 2014, pp. 8–13. doi:10.1109/WiMOB.2014.6962142.

[21] K. Wei, S. Guo, K. Xu, CACC: A context-aware congestion control approach in smartphone networks, IEEE Communications Magazine 52 (6) (2014) 42–48. doi:10.1109/MCOM.2014.6829943.

[22] S. Grasic, A. Lindgren, An analysis of evaluation practices for DTN routing protocols, in: Proceedings of the 7th ACM International Workshop on Challenged Networks (CHANTS), 2012, pp. 57–64. doi:10.1145/2348616.2348629.

[23] H.-S. Shih, H.-J. Shyur, E. S. Lee, An extension of TOPSIS for group decision making, Mathematical and Computer Modelling 45 (7–8) (2007) 801–813. doi:10.1016/j.mcm.2006.03.023.

[24] S. H. Zanakis, A. Solomon, N. Wishart, S. Dublish, Multi-attribute decision making: A simulation comparison of select methods, European Journal of Operational Research 107 (3) (1998) 507–529. doi:10.1016/S0377-2217(97)00147-1.

[25] E. Triantaphyllou, B. Shu, S. N. Sanchez, T. Ray, Multi-criteria decision making: An operations research approach, in: Encyclopedia of Electrical and Electronics Engineering, Vol. 15, John Wiley & Sons, New York, 1998, pp. 175–186.

[26] P. C. Fishburn, Additive utilities with incomplete product sets: Application to priorities and assignments, Operations Research 15 (3) (1967) 537–542. doi:10.1287/opre.15.3.537.

[27] P. W. Bridgman, Dimensional Analysis, Yale University Press, 1922.

[28] D. W. Miller, M. K. Starr, Executive Decisions and Operations Research, Prentice-Hall, Inc., 1969.

[29] D. von Winterfeldt, W. Edwards, Decision Analysis and Behavioral Research, Cambridge University Press, 1986.

[30] W. Edwards, How to use multiattribute utility measurement for social decisionmaking, IEEE Transactions on Systems, Man and Cybernetics 7 (5) (1977) 326–340. doi:10.1109/TSMC.1977.4309720.

[31] W. Edwards, F. H. Barron, SMARTS and SMARTER: Improved simple methods for multiattribute utility measurement, Organizational Behavior and Human Decision Processes 60 (3) (1994) 306–325. doi:10.1006/obhd.1994.1087.

[32] T. L. Saaty, A scaling method for priorities in hierarchical structures, Journal of Mathematical Psychology 15 (3) (1977) 234–281. doi:10.1016/0022-2496(77)90033-5.

[33] T. L. Saaty, How to make a decision: The analytic hierarchy process, European Journal of Operational Research 48 (1) (1990) 9–26. doi:10.1016/0377-2217(90)90057-I.

[34] K. O. Cogger, P. L. Yu, Eigenweight vectors and least-distance approximation for revealed preference in pairwise weight ratios, Journal of Optimization Theory and Applications 46 (4) (1985) 483–491. doi:10.1007/BF00939153.

[35] J. Ma, Z.-P. Fan, L.-H. Huang, A subjective and objective integrated approach to determine attribute weights, European Journal of Operational Research 112 (2) (1999) 397–404. doi:10.1016/S0377-2217(98)00141-6.

[36] X. Xu, A note on the subjective and objective integrated approach to determine attribute weights, European Journal of Operational Research 156 (2) (2004) 530–532. doi:10.1016/S0377-2217(03)00146-2.

[37] Y.-M. Wang, Y. Luo, Integration of correlations with standard deviations for determining attribute weights in multiple attribute decision making, Mathematical and Computer Modelling 51 (1–2) (2010) 1–12. doi:10.1016/j.mcm.2009.07.016.

[38] A. Jahan, F. Mustapha, S. M. Sapuan, M. Y. Ismail, M. Bahraminasab, A framework for weighting of criteria in ranking stage of material selection process, The International Journal of Advanced Manufacturing Technology 58 (1) (2012) 411–420. doi:10.1007/s00170-011-3366-7.

[39] M. Alemi-Ardakani, A. S. Milani, S. Yannacopoulos, G. Shokouhi, On the effect of subjective, objective and combinative weighting in multiple criteria decision making: A case study on impact optimization of composites, Expert Systems with Applications 46 (2016) 426–438. doi:10.1016/j.eswa.2015.11.003.

[40] M. F. Shipley, A. de Korvin, R. Obid, A decision making model for multi-attribute problems incorporating uncertainty and bias measures, Computers & Operations Research 18 (4) (1991) 335–342. doi:10.1016/0305-0548(91)90094-8.

[41] M. Zeleny, Multiple Criteria Decision Making, McGraw-Hill, New York, 1982.

[42] D.-G. Akestoridis, Evaluation of routing protocols for opportunistic networks with multiple-criteria decision-making methods, Master's thesis, Department of Computer Science and Engineering, University of Ioannina (2016).
URL http://www.cse.uoi.gr/tech_reports/publications/MT-2016-3.pdf

[43] B. B. Xuan, A. Ferreira, A. Jarry, Computing shortest, fastest, and foremost journeys in dynamic networks, International Journal of Foundations of Computer Science 14 (02) (2003) 267–285. doi:10.1142/S0129054103001728.

[44] E. C. Akrida, L. Gasieniec, G. B. Mertzios, P. G. Spirakis, The complexity of optimal design of temporally connected graphs, Theory of Computing Systems 61 (3) (2017) 907–944. doi:10.1007/s00224-017-9757-x.

[45] G. J. Székely, M. L. Rizzo, N. K. Bakirov, Measuring and testing dependence by correlation of distances, The Annals of Statistics 35 (6) (2007) 2769–2794. doi:10.1214/009053607000000505.

[46] G. J. Székely, M. L. Rizzo, Brownian distance covariance, The Annals of Applied Statistics 3 (4) (2009) 1236–1265. doi:10.1214/09-AOAS312.

[47] E. Triantaphyllou, S. H. Mann, An examination of the effectiveness of multi-dimensional decision-making methods: A decision-making paradox, Decision Support Systems 5 (3) (1989) 303–312. doi:10.1016/0167-9236(89)90037-7.

[48] N. Eagle, A. S. Pentland, Reality mining: sensing complex social systems, Personal and Ubiquitous Computing 10 (4) (2006) 255–268. doi:10.1007/s00779-005-0046-3.

[49] N. Eagle, A. S. Pentland, CRAWDAD dataset mit/reality (v. 2005-07-01), Downloaded from http://crawdad.org/mit/reality/20050701 (2005). doi:10.15783/C71S31.

[50] P. Hui, A. Chaintreau, J. Scott, R. Gass, J. Crowcroft, C. Diot, Pocket switched networks and human mobility in conference environments, in: Proceedings of the 2005 ACM SIGCOMM Workshop on Delay-tolerant Networking (WDTN), 2005, pp. 244–251. doi:10.1145/1080139.1080142.

[51] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, A. Chaintreau, CRAWDAD dataset cambridge/haggle (v. 2009-05-29), Downloaded from http://crawdad.org/cambridge/haggle/20090529 (2009). doi:10.15783/C70011.

[52] J. Stehlé, N. Voirin, A. Barrat, C. Cattuto, L. Isella, J.-F. Pinton, M. Quaggiotto, W. Van den Broeck, C. Régis, B. Lina, P. Vanhems, High-resolution measurements of face-to-face contact patterns in a primary school, PLoS ONE 6 (2011) e23176. doi:10.1371/journal.pone.0023176.

[53] V. Gemmetto, A. Barrat, C. Cattuto, Mitigation of infectious disease at school: targeted class closure vs school closure, BMC Infectious Diseases 14 (1) (2014) 1–10. doi:10.1186/s12879-014-0695-9.

[54] D. Kotz, K. Essien, Analysis of a campus-wide wireless network, Wireless Networks 11 (1) (2005) 115–133. doi:10.1007/s11276-004-4750-0.

[55] T. Henderson, D. Kotz, I. Abyzov, The changing usage of a mature campus-wide wireless network, Computer Networks 52 (14) (2008) 2690–2712. doi:10.1016/j.comnet.2008.05.003.

[56] D. Kotz, T. Henderson, I. Abyzov, J. Yeo, CRAWDAD dataset dartmouth/campus (v. 2009-09-09), Downloaded from http://crawdad.org/dartmouth/campus/20090909 (2009). doi:10.15783/C7F59T.

[57] SocioPatterns, Primary school temporal network data, Downloaded from http://www.sociopatterns.org/datasets/primary-school-temporal-network-data/ (2015).

[58] D. Kotz, T. Henderson, C. McDonald, CRAWDAD: A Community Resource for Archiving Wireless Data At Dartmouth.
URL http://crawdad.org

[59] N. Papanikos, D.-G. Akestoridis, E. Papapetrou, CRAWDAD toolset tools/simulate/uoi/adyton (v. 2016-04-21), Downloaded from http://crawdad.org/tools/simulate/uoi/adyton/20160421 (2016). doi:10.15783/C7BG6V.

[60] Z. J. Haas, T. Small, A new networking model for biological applications of ad hoc sensor networks, IEEE/ACM Transactions on Networking 14 (1) (2006) 27–40. doi:10.1109/TNET.2005.863461.