Engineering Comparators for Graph Clusterings⋆

Daniel Delling, Marco Gaertler, Robert Görke, and Dorothea Wagner

Faculty of Informatics, Universität Karlsruhe (TH)
{delling,gaertler,rgoerke,wagner}@informatik.uni-karlsruhe.de

Abstract. A promising approach to compare two graph clusterings is based on using measurements for calculating the distance between them. Existing measures either use the structure of clusterings or quality-based aspects with respect to some index evaluating both clusterings. Each approach suffers from conceptional drawbacks. We introduce a new approach combining both aspects and leading to better results for comparing graph clusterings. An experimental evaluation of existing and new measures shows that the significant drawbacks of existing techniques are not only theoretical in nature but manifest frequently on different types of graphs. The evaluation also proves that the results of our new measures are highly coherent with intuition, while avoiding the former weaknesses.

1 Introduction

Finding groups of similar elements in datasets, a technique known as clustering, is an important problem in the analysis and exploration of data. There are numerous applications such as data mining [8], network analysis [1], and biochemistry [16]. While recent research [2,3] focused on measuring the quality of a given clustering of an underlying graph, the problem of *comparing* two graph clusterings becomes more and more important.

There exists a mutual relation between the two concepts quality and distance: One could use a quality index to obtain a distance measure as shown later, while measuring the distance of a given clustering to an "optimal" clustering could yield the quality of the clustering. Current techniques for the comparison of clusterings use only qualitative aspects or transfer existing measures from the field of data mining. Both approaches have certain drawbacks: When comparing clusterings by using qualitative aspects the results are highly dependent on the used quality measure and completely different clusterings may yield the same quality value and are thus indicated as equal. Measures originating from data mining only consider the partition of nodes and ignore the structure of graphs. Due to these conceptional disadvantages, investigated below, the introduction of new measures seems inevitable, using structural and qualitative properties of the clusterings to calculate an appropriate distance. We present a new approach combining structural properties and qualitative aspects. In order to achieve this,

---

we extend data mining measures by adding qualitative features and introduce a new promising measure having its origin in quality measurement. Due to the high complexity of comparing clusterings we focus on the case of static comparison, i.e., the graph is unchanged, but give an outlook on the dynamic case. An experimental evaluation is presented, showing that the drawbacks of data mining measures are not only theoretical in nature but manifest often.

This paper is organized as follows. Section 2 introduces preliminaries and existing measures for comparing (data-)clusterings, including their drawbacks. Two approaches for constructing new measures are presented in Section 3. An evaluation based on artificial data of all presented measures is given in Section 4, while Section 5 shows the applicability of our approach in a real-world scenario. Section 6 concludes this paper.

2   Preliminaries

We assume that $G = (V, E)$ is an undirected, unweighted and connected graph. Let $n := |V|, m := |E|$, and $\mathcal{C} := \{C_1, \ldots, C_p\}$ a partitioning of V. We call $\mathcal{C}$ a clustering and the C_i clusters of the graph. The set of all possible clusterings is $\mathcal{A}(V)$. Let $E(\mathcal{C}) := \{\{u, v\} \in E \mid u, v \in C_i\}$ be the set of *intra-cluster edges* of $\mathcal{C}$ and $\overline{E}(\mathcal{C}) := \{\{u, v\} \in E \mid u \in C_i, v \in C_j, i \neq j\}$ the set of *inter-cluster edges* of $\mathcal{C}$. The cardinalities are indicated by $m(\mathcal{C}) := |E(\mathcal{C})|$ and $\overline{m}(\mathcal{C}) := |\overline{E}(\mathcal{C})|$. We call a graph with disjoint cliques a *clustergraph* and $F_{\mathcal{C}}$, the set of edges to be added or deleted in order to transform a given graph and clustering $\mathcal{C}$ into an according clustergraph, the *cluster editing set* of $\mathcal{C}$. When comparing two clusterings we use $\mathcal{C}$ and $\mathcal{C}'$, with $k := |\mathcal{C}|, l := |\mathcal{C}'|$. With $\deg(C_i) := \sum_{v \in C_i} \deg(v)$ we indicate the sum of all degrees of nodes within a cluster. All presented measures are given in a distance version, normalized to the interval $[0, 1]$. In the following, we give a short overview of existing comparison techniques. Among them are measures based on quality and on comparing the partitions of node-sets, the latter are also called *node-structural*.

Quality-Based Distance. Quality-based measurements can be constructed by comparing the scores of the two clusterings with respect to an arbitrary quality index such as coverage, performance or modularity [1,3]. Note, that a distance measured in such a way is highly dependent on the used index. Furthermore, completely different clusterings can yield the same value. Thus, we neglect purely quality-based distances in the following and focus on measuring the distance based on the structure of the clusterings.

Counting Pairs. In [17] some techniques based on counting pairs are presented. Summarizing, every pair of nodes is categorized based on whether they are in the same (or different) cluster with respect to both clusterings. Four sets are defined: S_{11} (S_{00}) is the set of unordered pairs that are in the same (different) clusters under both clusterings, whereas S_{01} (S_{10}) contains all pairs that are

in the same cluster under $\mathcal{C}$ ($\mathcal{C}'$) and in different under $\mathcal{C}'$ ($\mathcal{C}$). In the following we present two representatives for this class: *Rand* and *adjusted Rand* measure. Rand introduced the distance function $\mathcal{R}$ given in Equation 1 in [12], it suffers from several drawbacks. For example, it is highly dependent on the number of clusters. One attempt to remedy some of these drawbacks, which is known as *adjusted Rand* $\mathcal{AR}$ and given in Equation 1, is to subtract the expected value for clusterings with a hypergeometric distribution of nodes, see [11].

$$\mathcal{R}(\mathcal{C},\mathcal{C}') := 1 - \frac{2(n_{11} + n_{00})}{n(n-1)} \ , \ \ \mathcal{AR}(\mathcal{C},\mathcal{C}') := 1 - \frac{n_{11} - t_3}{\frac{1}{2}(t_1 + t_2) - t_3} \ , \qquad (1)$$

where $t_1 := n_{11} + n_{10}$, $t_2 := n_{11} + n_{01}$, and $t_3 := (2t_1 t_2)/(n(n-1))$ and t_1 (t_2) is the cardinality of all pairs of nodes that are in the same cluster under $\mathcal{C}$ ($\mathcal{C}'$).

Overlaps. Another counting approach is based on the $k \times l$ confusion matrix $CM := (m_{ij})$ whose ij-entry indicates how many elements are in Cluster C_i and C'_j, formally $m_{ij} := |C_i \cap C'_j|$, for $1 \le i \le k$ and $1 \le j \le l$. Several measures are based on the confusion matrix. We restrict ourselves to the measure $\mathcal{NVD}$, introduced by van Dongen in [15], given in Equation 2. Other measures suffer from the obvious disadvantage of asymmetries, thus we exclude them. We use a normalized version to keep the measure to the interval $[0, 1]$.

$$\mathcal{NVD}(\mathcal{C},\mathcal{C}') := 1 - \frac{1}{2n} \sum_{i=1}^{k} \max_j m_{ij} - \frac{1}{2n} \sum_{j=1}^{l} \max_i m_{ij} \qquad (2)$$

One major drawback of $\mathcal{NVD}$ is that the distance between the two trivial clusterings, i.e., $k = 1, l = n$, only yields a value of about 0.5. In addition, this measure suffers from the drawback that only the maximum overlaps contribute, resulting counter-intuitive examples are given in [10].

Information Theory. More promising approaches are based on information theory [4]. Informally, the *entropy* $\mathcal{H}(\mathcal{C})$ of a clustering is the uncertainty of a randomly picked node belonging to a certain cluster. An entropy of a clustering is always positive and is bounded by $\log_2(n)$, see [13]. An extension of entropy is the *mutual information* $\mathcal{I}(\mathcal{C},\mathcal{C}')$. The mutual information of two clusterings is the loss of uncertainty of one clustering if the other is given. With $P(i) := |C_i|/n$ and $P(i,j) := (|C_i \cap C'_j|)/n$, entropy and mutual information are defined as follows.

$$\mathcal{H}(\mathcal{C}) := - \sum_{i=1}^{k} P(i) \log_2 P(i) \ , \ \ \mathcal{I}(\mathcal{C},\mathcal{C}') := \sum_{i=1}^{k} \sum_{j=1}^{l} P(i,j) \log_2 \frac{P(i,j)}{P(i)P(j)} \qquad (3)$$

Note that mutual information is positive and bounded by $\min\{\mathcal{H}(\mathcal{C}), \mathcal{H}(\mathcal{C}')\} \le \log_2(n)$. In the following we present two representatives in this class, namely

one introduced by Fred & Jain [6] and *Variation of Information*, introduced by
Meila [9].

$$\mathcal{FJ}(\mathcal{C},\mathcal{C}') := \begin{cases} 1 - \frac{2\mathcal{I}(\mathcal{C},\mathcal{C}')}{\mathcal{H}(\mathcal{C})+\mathcal{H}(\mathcal{C}')} & \text{, if } \mathcal{H}(\mathcal{C})+\mathcal{H}(\mathcal{C}') \neq 0 \\ 0 & \text{, otherwise} \end{cases} \tag{4}$$

$$\mathcal{VI}(\mathcal{C},\mathcal{C}') := \mathcal{H}(\mathcal{C})+\mathcal{H}(\mathcal{C}') - 2\mathcal{I}(\mathcal{C},\mathcal{C}') \tag{5}$$

The first measure $\mathcal{FJ}$, given in Equation 4, is a normalized version of the mutual
information and stated as a distance function. The case differentiation is used
to deal with the degenerated case of two trivial clusterings, i. e.,$k = l = 1$.

The second measure $\mathcal{VI}$ is motivated by an axiomatic approach and given in
Equation 5. In [10], it is shown that $\mathcal{VI}$ is the only measure fulfilling several ax-
ioms. However, these axioms seem to be inadequate in the special case of graph
clustering. According to these axioms, the movement of a node v from one clus-
ter C_i to another cluster C_j must be equivalent to first splitting v off from C_i and
then merging it with C_j. Figure 1 shows an example regarding this axiom: intu-
itively $d(\mathcal{C},\mathcal{C}'')$ should be greater than $d(\mathcal{C},\mathcal{C}')+d(\mathcal{C}',\mathcal{C}'')$ of which both terms rep-
resent minor changes, but according to the axiom $d(\mathcal{C},\mathcal{C}'') = d(\mathcal{C},\mathcal{C}') + d(\mathcal{C}',\mathcal{C}'')$
must hold. This measure is not normalized and the two possible normalization
factors, which are $1/\log_2(n)$ and $1/\log_2(\max\{k,l\})$, mapping to the intervals
$[0,x], x \leq 1$ and $[0,1]$ respectively, have significant drawbacks. Nevertheless, we
use the $\log_2(n)$ normalized version for comparability with the other measures.

Drawbacks of the Data Mining Approach. All node-structural measures
suffer from the same drawback that they neglect the structure of the graph. The
examples in Figure 2 clarify this circumstance. The figure shows four clusterings
$\mathcal{C}_1,\mathcal{C}_1',\mathcal{C}_2$ and $\mathcal{C}_2'$ on two graphs G_1 and G_2. A measure d not considering the
structure of the graphs fulfills $d(\mathcal{C}_1,\mathcal{C}_1') = d(\mathcal{C}_2,\mathcal{C}_2')$. Intuitively, the distance
$d(\mathcal{C}_1,\mathcal{C}_1')$ has to be greater than $d(\mathcal{C}_2,\mathcal{C}_2')$ since the quality of $\mathcal{C}_1$ is almost equal
to that of $\mathcal{C}_2$, but $\mathcal{C}_1'$ has far lower quality than $\mathcal{C}_2'$. This drawback can become
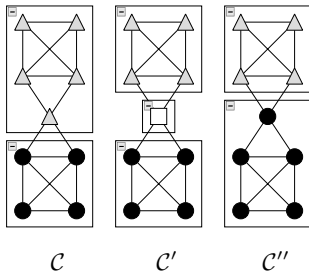arbitrarily grave when the edge set of the graph is allowed to change.

$\mathcal{C}$ $\qquad$ $\mathcal{C}'$ $\qquad$ $\mathcal{C}''$

$\mathcal{C}_1$ $\qquad$ $\mathcal{C}_1'$ $\qquad$ $\mathcal{C}_2$ $\qquad$ $\mathcal{C}_2'$

Fig. 1. The sum of two minor
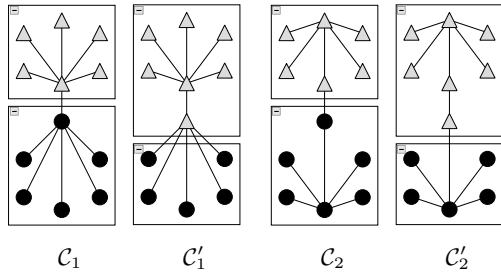changes result in a major one

Fig. 2. Two static comparisons of graph cluster-
ings

3   Engineering Graph-Structural Comparison Measures

In order to remedy some of the disadvantages of node-structural measures, we introduce the concept of *graph-structural* measures. Since they are also based on the underlying graph structure, they can include qualitative aspects for measuring the distance of two clusterings. In the first part, Section 3.1, we extend node-structural measures, while a novel measure is introduced in the second part, Section 3.2.

3.1   Extension of Node-Structural Measures

For consistency, all extended measures should meet the following requirement: If the underlying graph is complete, then both the graph- and node-structural version should yield the same value, since then the graph structure does not provide additional information. A second objective is to adjust the three founding principles—counting pairs, overlaps and information theory—of the existing measures themselves, instead of adjusting each implementation separately.

Counting Local Pairs. Instead of categorizing every pair we only consider those pairs, that are connected by an edge. For $a, b \in \{0, 1\}$ we define $E_{ab} := S_{ab} \cap E$ and $e_{ab} := |E_{ab}|$. It is obvious that $S_{ab} = E_{ab}$ holds for complete graphs. Thus, we obtain the graph-based versions $\mathcal{R}_g$ and $\mathcal{AR}_g$ of the Rand and adjusted Rand measure given in Equation 6:

$$\mathcal{R}_g(\mathcal{C}, \mathcal{C}') := 1 - \frac{e_{11} + e_{00}}{m} \ , \ \ \mathcal{AR}_g(\mathcal{C}, \mathcal{C}') := 1 - \frac{e_{11} - t_3}{\frac{1}{2}(m(\mathcal{C}) + m(\mathcal{C}')) - t_3} \ , \quad (6)$$

where $t_3 := (m(\mathcal{C})m(\mathcal{C}'))/m$. Note, that $m(\mathcal{C}) = e_{11} + e_{10}$ and $m(\mathcal{C}') = e_{11} + e_{01}$, respectively, hold.

Degree-Based Overlaps. Measures based on overlaps can be transformed into graph-structural measures by a slight modification in the definition of the confusion matrix as follows. The ij-th entry of the *degree-based confusion matrix* $CM^d := (m_{ij}^d)$ indicates the sum of the degrees of the nodes that are both in C_i and C', formally $m_{ij}^d := \deg(C_i \cap C'_j)$. Note, that if G is d-regular graph, then the equality $CM = CM^d/d$ holds. In certain cases, this may lead to different normalization factors. The extension of $\mathcal{NVD}$ is given in Equation 7.

$$\mathcal{NVD}_g(\mathcal{C}, \mathcal{C}') := 1 - \frac{1}{4m} \sum_{i=1}^{k} \max_{j} m_{ij}^d - \frac{1}{4m} \sum_{j=1}^{l} \max_{i} m_{ij}^d \qquad (7)$$

The equivalence of the node- and the graph-structural variant of the normalized van Dongen measure for regular graphs follows from $m = dn/2$ and $m_{ij} = m_{ij}^d/d$.

Edge Entropy. The entropy defined in Section 2 solely depends on the node-set, thus we extend it to the edge-set using the following paradigm: Instead of

randomly picking a node from the graph for measuring the uncertainty, we pick the end of an edge randomly. As a consequence, a node with high degree has a greater impact on the distance. The formal definition of *edge entropy* $\mathcal{H}_E$ and *edge mutual information* $\mathcal{I}_E$ is given in Equation 8 and 9.

$$\mathcal{H}_E(\mathcal{C}) := -\sum_{i=1}^{k} P_E(i) \log_2 P_E(i) \ , \tag{8}$$

$$\mathcal{I}_E(\mathcal{C}, \mathcal{C}') := \sum_{i=1}^{k} \sum_{j=1}^{l} P_E(i,j) \log_2 \frac{P_E(i,j)}{P_E(i)P_E(j)} \ , \tag{9}$$

where $P_E(i) := \deg(C_i)/2m$ and $P_E(i,j) := \deg(C_i \cap C'_j)/2m$. Note that for regular graphs, the entropy and the edge entropy coincide. The extensions of $\mathcal{FJ}$ and $\mathcal{VI}$ are given in Equation 10 and 11.

$$\mathcal{FJ}_g(\mathcal{C}, \mathcal{C}') := \begin{cases} 1 - \frac{2\mathcal{I}_E(\mathcal{C},\mathcal{C}')}{\mathcal{H}_E(\mathcal{C})+\mathcal{H}_E(\mathcal{C}')} & \text{, if } \mathcal{H}_E(\mathcal{C}) + \mathcal{H}_E(\mathcal{C}') \neq 0 \\ 0 & \text{, otherwise} \end{cases} \tag{10}$$

$$\mathcal{VI}_g(\mathcal{C}, \mathcal{C}') := \mathcal{H}_E(\mathcal{C}) + \mathcal{H}_E(\mathcal{C}') - 2\mathcal{I}_E(\mathcal{C}, \mathcal{C}') \tag{11}$$

The equivalence of the node- and the graph-structural variant for regular graphs results from the equality of entropy and edge entropy for complete graphs. Meila showed in [10] that $\mathcal{VI} \leq \log_2(n)$ also holds for weighted clusterings. Since the degree of a node can be interpreted as node-weight our $\log_2(n)$-normalization maps to the interval of $[0, 1]$.

3.2   A Novel Approach for Measuring Graph-Structural Distance

Although the extensions introduced in the previous section incorporate the underlying graph structure, they are not suitable for comparing clusterings on different graphs. As a first step to solve this task, we consider the restriction to graphs with the same node-set, but potentially different edge-sets. Motivated by the cluster editing set, we introduce the *editing set difference* defined in Equation 12.

$$\mathcal{ESD}(\mathcal{C}, \mathcal{C}') = \frac{|F_{\mathcal{C}} \cup F_{\mathcal{C}'}| - |F_{\mathcal{C}} \cap F_{\mathcal{C}'}|}{|F_{\mathcal{C}} \cup F_{\mathcal{C}'}|} = 1 - \frac{|F_{\mathcal{C}} \cap F_{\mathcal{C}'}|}{|F_{\mathcal{C}} \cup F_{\mathcal{C}'}|} \tag{12}$$

Small cluster editing sets correspond to significant clusterings. By comparing the two clusterings with a geometric difference, we obtain an indicator for the structural difference of the two clusterings. It easy to see, that in the case of static comparison, $\mathcal{ESD}$ is a metric.

4   Experiments and Evaluation

We evaluate the introduced measures on two setups. The first focuses on structural properties of clusterings, the second concentrates on qualitative aspects:

Initial and Random Clusterings. The tests consist of two comparisons, each including clusterings with the same expected intrinsic structure of the partitions, i. e.,the expected number of clusters and the size of clusters. The first comparison uses one significant clustering and one uniformly random clustering, while the second one uses two uniformly random clusterings.

Local Minimization. The setup consists of two parts, each comparing a reference clustering with a clustering of less significance. The two parts differ in the significance of the reference clustering.

The intuition of the first test is to clarify the drawbacks of the node-structural measures, while the second setup verifies the obtained results. We use the *attractor generator* introduced in [5] which uses geometric properties based on Voronoi Diagrams to generate initial clusterings. The Voronoi cells represent clusters and the maximum Euclidean distance of two nodes being connected is determined by a perturbation parameter. All tests use $n = 1000$ nodes and are repeated until the maximal length of the 0.95-confidence intervals is not larger than 0.1.
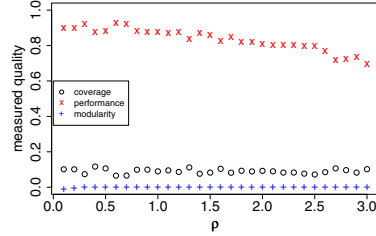
4.1   Initial- and Random Clusterings

The generated clustering is used as a significant clustering. For the random clustering we first pick k uniformly at random between 2 and $\sqrt[3]{n}$ for the number of clusters and assign each node uniformly at random to the k clusters. Figures 3.1 and 3.2 show the measured quality by the indices coverage, performance and modularity [1,3]. The tests consists of two cases. On the one hand, the comparison of the generated and a random clustering (GvR) and on the other hand, the comparison of two random clusterings (RvR). A measure for comparing graph clusterings should differ in the two cases. For GvR, a suitable measure should indicate a decreasing distance with the loss of significance of the reference, while for RvR two interpretations are possible. On the one hand, one could claim that the distance between two random clusterings should be independent of the underlying graphs. On the other hand, the distance should decrease with the loss of significance because two random clusterings on an almost complete graph are closer to each other than on a graph with an existing significant clustering. Another interpretation seems acceptable as well: The distance of a given clustering to a random clustering should always be somehow maximal.

Figure 3 shows the results for the node- and graph-structural measures. By comparing Figure 3.3 and 3.4 it is evident that node-structural measures do not distinguish the two cases. Only Fred & Jain and adjusted Rand reflect the interpretation that the distance to a random clustering is always maximal. However, the situation changes for the graph-structural distance (Figures 3.5 and 3.6). Only Rand and $\mathcal{ESD}$ capture the difference, while the remaining measures show nearly the same behavior as their node-structural counterparts. For GvR, the distance measured by Rand is decreasing with increasing density while for RvR the distance is invariant under the density. Furthermore, the measured distance equals the node-structural measurement for RvR. $\mathcal{ESD}$ has the same behavior for GvR as Rand, whereas RvR reflects the intuition that two random clusterings become more similar with loss of significance. Under the assumption that
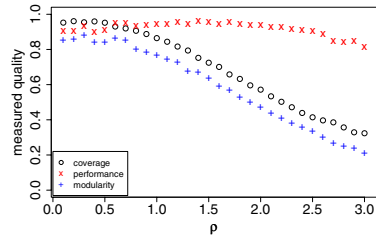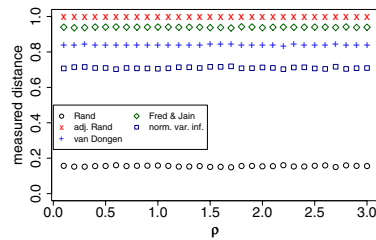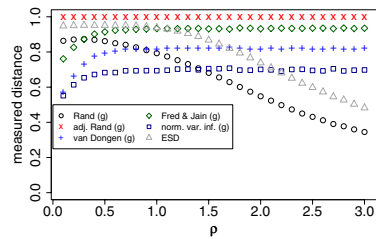
3.1: Quality of initial clustering
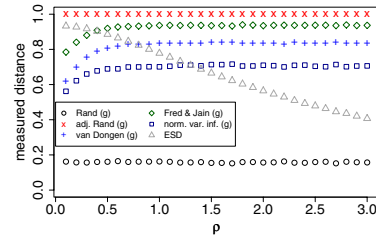
3.2: Quality of random clustering
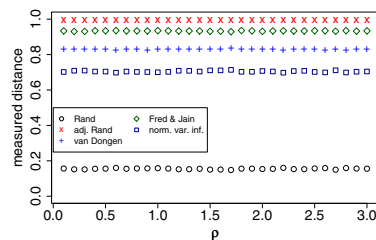
3.3: GvR nodes-structural

3.4: RvR nodes-structural

3.5: GvR graph-structural

3.6: RvR graph-structural

Fig. 3. Results of the initial- and random clustering setup

a comparison to a random clustering should always be interpreted as maximal, adjusted Rand and Fred & Jain can be accepted. Nevertheless, the equivalence of the node- and the graph-structural versions of van Dongen and the normalized Variation of Information is counterintuitive. This partly originates from the fact, that attractors produce graphs that are close to regular for $\varrho > 0.5$. Furthermore, the clusters are equal in size. The strange behavior of Fred & Jain, van Dongen and the variation of Information for very small ϱ stems from the fact that for small ϱ attractors are nearly stargraphs with k centers.

4.2   Local Minimization

Since there are several possible interpretations of graph-structural distance and the structural similarity of the clusterings in Section 4.1 a second test is executed having a precise intuition for graph-structural distance. Again, as a reference
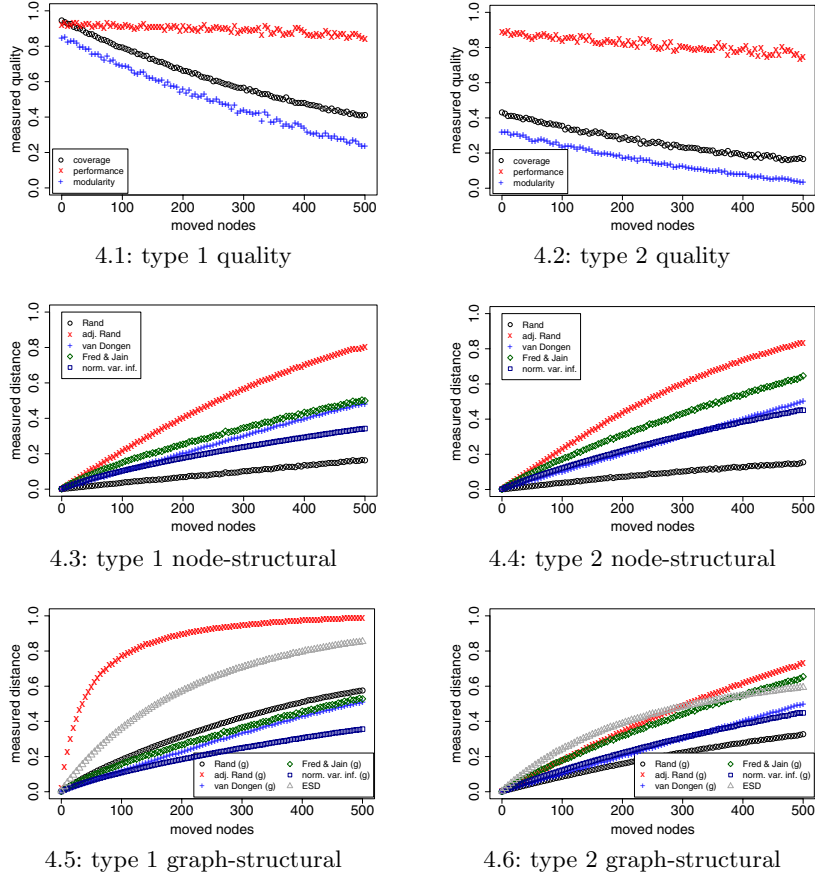
4.1: type 1 quality

4.2: type 2 quality

4.3: type 1 node-structural

4.4: type 2 node-structural

4.5: type 1 graph-structural

4.6: type 2 graph-structural

Fig. 4. Results of the local minimization setup

clustering we use the generated clustering of an attractor graph. The second clustering of less significance is obtained from the reference clustering by locally moving nodes from one cluster to another. Such a shift is executed, if it maximally decreases a given index among all possible shifts. This is done until no decrease of quality can be achieved or the number of moved nodes has reached a maximum value of $M_{\max}$.

In this setup, we use modularity as the index, the density is set to the values $\varrho = 0.5$ (type 1) and $\varrho = 2.5$ (type 2), and $M_{\max}$ increases from 0 to 500 using steps of 5. Figures 4.1 and 4.2 show the measured quality of the locally decreased clusterings on increasing number of moved nodes. Note, that for $M_{\max} = 0$ the reference and the locally decreased clustering coincide. A suitable distance measure should first of all distinguish the two cases. In addition, with increasing $M_{\max}$ the measured distance in type 2 should be smaller than in type 1, since the intuition is that in type 1 a very significant clustering is destroyed while on

type 2 the loss of significance is lower. Figure 4 shows the results for all measures on this specific setup. As shown in Figures 4.3 and 4.4, all node-structural measures hardly distinguish the two cases. This reveals additional disadvantages. Evaluating the graph-structural measures (Figures 4.5 and 4.6), the intuitive behavior of Rand is verified. Furthermore, adjusted Rand and $\mathcal{ESD}$ distinguish both cases very well. The remaining graph-structural measures show the same behavior as their node-structural counterparts. Thus, the failure of van Dongen and the Variation of Information is confirmed. Unlike in Section 4.1 Fred & Jain fails on this setup. The unexpected behavior of the overlap and entropy based measures may be due to—as mentioned in Section 4.1—the fact that for $\varrho = 0.5$ and $\varrho = 2.5$ attractor graphs have a fairly regular structure. As shown in Section 2 the graph-structural versions of overlap- and entropy-based measures equal the node-structural variants for regular graphs.

5   Real-World Scenario

In this section, we discuss a real-world instance in order to illustrate the advantages of graph-structural measures over node-structural ones. As input, we use the e-mail graph (Figure 5) of the Karlsruhe faculty of computer science, introduced in [7]. As a reference clustering, we group by departments. We additionally compute two clusterings by using the greedy modularity approach [3] and the MCL algorithm, introduced in [14]. Table 1 depicts the scores achieved by the quality measures coverage, performance and modularity. With respect to all three quality measures, MCL outperforms the greedy approach and achieves a score close to the reference. Table 2 gives an overview of the measured distances between the abovementioned clusterings. We observe that the MCL-clustering is not as close to the reference than one could expect from the figures in
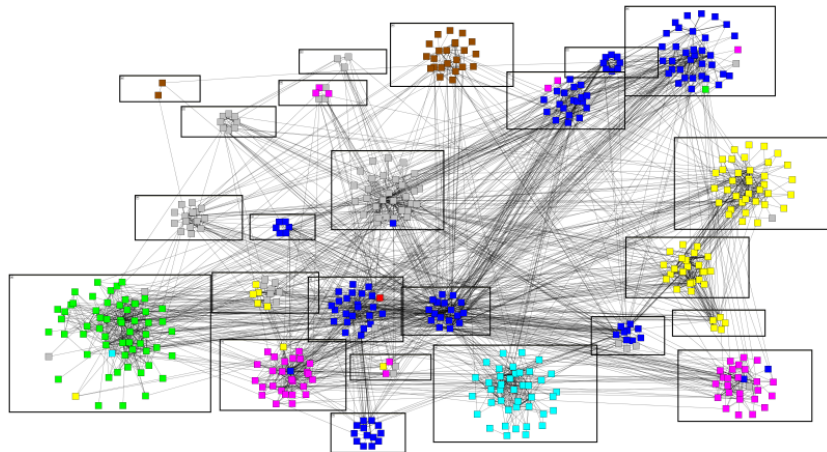

Fig. 5. Karlsruhe e-mail graph. Groups refer to the reference clustering, colors to the clustering obtained by the greedy modularity algorithm.

Table 1. Quality scores achieved by the reference clustering and those computed by the greedy approach and by MCL. The input is the Karlsruhe e-mail graph.

	reference	greedy	MCL
coverage	0.8173	0.8634	0.8182
performance	0.9387	0.8286	0.9238
modularity	0.7423	0.6725	0.7282

Table 2. Measured distances between reference and two computed clusterings. One clustering is obtained by MCL, the other one by the greedy modularity algorithm. The input is the Karlsruhe e-mail graph (cf. Figure 5).

measure type	measure	reference vs. greedy	reference vs. MCL	greedy vs. MCL
quality	modularity difference	0.0697	0.0140	0.0557
node-structural	Rand	0.1233	0.0463	0.1466
	adj. Rand	0.5765	0.3555	0.6549
	van Dongen	0.2676	0.1834	0.3465
	Fred & Jain	0.3137	0.1794	0.3876
	variation of information	0.2425	0.1658	0.2904
graph-structural	Rand	0.1963	0.1305	0.2452
	adj. Rand	0.4689	0.2820	0.5730
	van Dongen	0.2435	0.1714	0.3215
	Fred & Jain	0.2828	0.1623	0.3581
	variation of information	0.2107	0.1427	0.2549
	ESD	0.7325	0.5382	0.7796

Table 1. All graph-structural distance measures indicate a difference of more than 0.1. More interestingly, $\mathcal{ESD}$ yields a lower score than graph-structural adjusted Rand. For artificial data, the contrary is true (cf. Section 4). Most of our graph-structural measures indicate a lower distance between all clusterings than their node-structural versions. As all clusterings score similar quality values, and thus have quite a low distance with respect to quality, the graph-structural measures really incorporate qualitative aspects. Hence, they harmonize better with intuition than the purely node-structural versions. As discussed in Section 2, the node-structural Rand measure yields a very small value due to the high number of small cluster. However, this drawback appears to be remedied by the graph-structural version.

6   Conclusion

The experimental evaluation confirms the drawbacks of node-structural measures while some graph-structural measures, i. e.,$\mathcal{ESD}$, adjusted Rand, and Rand perform more consistently with intuition. Furthermore, this is an indicator for the feasibility of the graph-structural distance in applications such as dynamic graph clustering. More precisely, since graph-structural measures incorporate

both structural and qualitative aspects, they can be used as a foundation for clustering in dynamic scenarios. Summarizing, extensions of node-structural measures are not trivial and need not lead to intuitive results. Furthermore, our presented extensions are only suitable for comparing clusterings on the same graph. In contrast, the editing set distance only requires the same node-set. Thus, this improves the foundation for dynamic graph clusterings. Concluding, this work is a first step towards a unifying comparison framework.

References

1. Brandes, U., Erlebach, T. (eds.): Network Analysis. LNCS, vol. 3418. Springer, Heidelberg (2005)
2. Brandes, U., Gaertler, M., Wagner, D.: Experiments on Graph Clustering Algorithms. In: Di Battista, G., Zwick, U. (eds.) ESA 2003. LNCS, vol. 2832, pp. 568–579. Springer, Heidelberg (2003)
3. Clauset, A., Newman, M., Moore, C.: Finding community structure in very large networks. Physical Review E 70(066111) (2004)
4. Cover, T., Thomas, J.: Elements of Information Theory. John Wiley & Sons, Inc., Chichester (1991)
5. Delling, D., Gaertler, M., Wagner, D.: Generating Significant Graph Clusterings. In: European Conference of Complex Systems (ECCS 2006) (2006)
6. Fred, A., Jain, A.: Robust Data Clustering. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR, pp. 128–136 (2003)
7. Gaertler, M., Görke, R., Wagner, D.: Significance-Driven Graph Clustering. In: Kao, M.-Y., Li, X.-Y. (eds.) AAIM 2007. LNCS, vol. 4508, pp. 11–26. Springer, Heidelberg (2007)
8. Jain, A., Dubes, R.: Algorithms for Clustering Data. Prentice Hall, Englewood Cliffs (1988)
9. Meila, M.: Comparing Clusterings by the Variation of Information. In: 16th Annual Conference of Computational Learning Theory (COLT), pp. 173–187 (2003)
10. Meila, M.: Comparing Clusterings - An Axiomatic View. In: 22nd International Conference on Machine Learning, Bonn, Germany, pp. 577–584 (2005)
11. Morey, R., Agresti, A.: The Measurement of Classification Agreement: An Adjustment to the RAND Statistic for Chance Agreement. Educational and Psychological Measurement 44, 33–37 (1984)
12. Rand, W.: Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association 66, 846–850 (1971)
13. Strehl, A., Ghosh, J.: Cluster ensembles – a knowledge reuse framework for combining multiple partitions. J. Mach. Learn. Res. 3, 583–617 (2003)
14. van Dongen, S.: A cluster algorithm for graphs. Tech. Report INS-R0010 (2000)
15. van Dongen, S.: Performance criteria for graph clustering and markov cluster experiments. Technical Report INS-R0012, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam (May 2000)
16. Vidal, M.: Interactome modeling. FEBS Lett. 579, 1834–1838 (2005)
17. Wagner, S., Wagner, D.: Comparing Clusterings – An Overview. Technical Report 2006-4, Faculty of Informatics, Universität Karlsruhe (TH) (2006)