Experiments on Comparing 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 graph clusterings is based on using measurements for calculating the distance. Existing measures either use the structure of clusterings or quality–based aspects. Each approach suffers from critical 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 and proves that the results of our new measures are more coherent with intuition.

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 [1], network analysis [2], and biochemistry [3]. While recent research [4,5] 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. For example, dynamic clustering [6] is based on measuring the distance of the original and the updated clustering. For evaluating the usability of a clustering algorithm [4,7] the obtained clustering needs to be compared to a given reference clustering.

At a first glance, the two problems of quality and distance seem to be independent of each other. However, there exists a mutual relation: On the one hand, one can use a quality index to obtain a distance measure as shown later. On the other hand, measuring the distance of a given clustering to an "optimal" clustering could be interpreted as the quality of the clustering.

Current techniques use only qualitative aspects or transfer existing measures from the field of data mining. Both approaches have critical drawbacks: When comparing clusterings by using qualitative aspects the results are highly dependent on the used quality measure and completely different clusterings may have the same significance and are indicated as equal. Measures originating from data mining only consider the partition of nodes and ignore the structure of graphs. In the following, we show that such measures suffer from critical drawbacks even when comparing clusterings on the same graph, which is later called *static comparison*. Due to the fact of these conceptional disadvantages, the introduction of new measures seems inevitable, using structural and qualitative properties of the clusterings to calculate an appropriate distance.

* This work was partially supported by the DFG under grant WA 654/14-3 and EU under grant DELIS (contract no. 001907).

We present a new approach combining structural properties and qualitative aspects resulting in better measures for comparing graph clusterings. 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 clusterings we focus on the case of static comparison. Nevertheless, an outlook on the dynamic case is given. An experimental evaluation is presented, showing that the drawbacks of data mining measures are not only theoretical in nature.

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 of all presented measures is given in Section 4. Section 5 concludes this paper by giving a summary and an outlook on following questions.

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 $\mathbb{P}(V)$. 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 into a 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.

Some existing measures calculate the similarity between clusterings. These measures can be transfered to distance measures. For better understanding, all presented measures are given in a normalized distance version, i. e., equal clusterings yield the value of zero and completely different clusterings have a value of one.

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*.

2.1 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 [2,5]. The comparison can be realized by the absolute difference, for example. Of course, many other functions are also possible. 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 quality–based distances in the following and focus on measuring the distance based on the structure of the clusterings.

2.2 Counting Pairs

In [8] 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 $\mathcal{C}'(\mathcal{C})$. Based on the cardinalities $n_{ab} := |S_{ab}|$, for $a, b \in \{0, 1\}$ several measures are defined. In the following we present two representatives for this class: *Rand* and *adjusted Rand* measure.

Rand has introduced the distance function $\mathcal{R}$ given in Equation (1) in [9]. It is widely spread but suffers from several drawbacks. For example, it is highly dependent on the number of clusters.

$$\mathcal{R}(\mathcal{C}, \mathcal{C}') := 1 - \frac{2(n_{11} + n_{00})}{n(n-1)} = \frac{2(n_{01} + n_{10})}{n(n-1)} \tag{1}$$

One attempt to remedy some of these drawbacks, which is known as *adjusted Rand* and given in Equation 2, is to subtract the expected value for hypergeometric clusterings, see [10].

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

$$\text{where } t_1 := \sum_{i=1}^{k} \binom{|C_i|}{2} = n_{11} + n_{10}, t_2 := \sum_{j=1}^{l} \binom{|C'_j|}{2} = n_{11} + n_{01}, t_3 := \frac{2t_1 t_2}{n(n-1)}$$

Note that $t_1(t_2)$ is the cardinality of all pairs of nodes that are in the same cluster under $\mathcal{C}(\mathcal{C}')$. Although the assumption of hypergeometric clusterings seems to be counterintuitive, the quality measure *modularity* [5], which seems to be a quite promising approach for quality measurements of clusterings, is also built on a strong – although different – assumption.

2.3 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 $1 \leq i \leq k$ and $1 \leq j \leq l$. Several measures are based on the confusion matrix. We restrict ourselves to the measure $\mathcal{NVD}$, introduced by van Dongen in [11], given in Equation (3). 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 $[0, 1]$.

$$\mathcal{NVD}(\mathcal{C}, \mathcal{C}') := 1 - \frac{1}{2n} \sum_{i=1}^{k} \max_j m_{ij} - \frac{1}{2n} \sum_{j=1}^{l} \max_i m_{ij} \tag{3}$$

One potential drawback of $\mathcal{NVD}$ is that the distance between the two trivial clusterings, i. e., $k = 1, l = n$, does not have not a maximum value.

2.4 Information Theory

More promising approaches are based on information theory [12]. Entropy in the context of clustering is defined by:

$$\mathcal{H}(\mathcal{C}) := -\sum_{i=1}^{k} P(i)\log_2 P(i) \quad \text{where} \quad P(i) := \frac{|C_i|}{n}$$

Informally, the entropy 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 of two clusterings $\mathcal{C}, \mathcal{C}'$:

$$\mathcal{I}(\mathcal{C},\mathcal{C}') := \sum_{i=1}^{k}\sum_{j=1}^{l} P(i,j)\log_2 \frac{P(i,j)}{P(i)P(j)} \quad \text{where} \quad P(i,j) := \frac{|C_i \cap C'_j|}{n}$$

Informally, the mutual information of two clusterings is the loss of uncertainty of one clustering if the other is given. Thus, mutual information is positive and bounded by $\min\{\mathcal{H}(\mathcal{C}), \mathcal{H}(\mathcal{C}')\} \leq \log_2(n)$. In the following we present two representatives in this class, namely one introduced by Fred & Jain [14] and *Variation of Information*, introduced by Meila [15].

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

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

The second measure *Variation of Information*, introduced by Meila in [16], is motivated by an axiomatic approach and given in Equation (5).

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

It is shown that $\mathcal{VI}$ is the only measure fulfilling several axioms. 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 cluster C_i to another cluster C_j is equivalent to first splitting v from C_i and then merging it with C_j. In addition, this measure is not normalized and the two possible normalization factors, which are $1/\log_2(n)$ and $1/\log_2(\max\{k,l\})$ respectively, have significant drawbacks. Nevertheless, we use the $\log_2(n)$ normalized version for better comparability with the other measures.

2.5 Drawbacks of the Data Mining Approach

All node–structural measures suffer from the same drawback that they neglect the structure of the underlying graph. Even comparing clusterings on the same graph can lead to very counterintuitive situations. Examples in Figure 1 clarifies this circumstance.

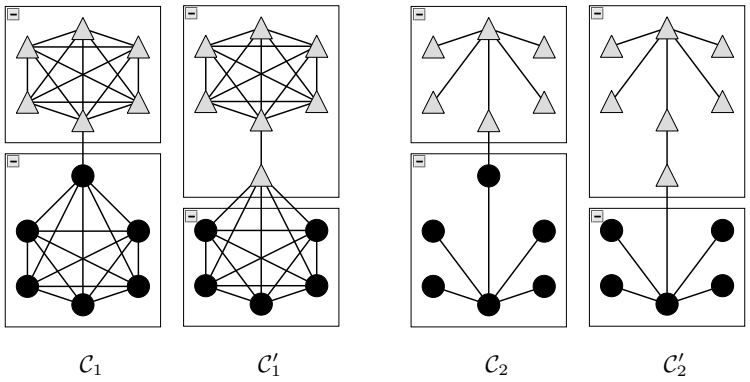


Fig. 1. two static comparisons of graph clusterings

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 underlying graphs fulfills $d(\mathcal{C}_1, \mathcal{C}_1') = d(\mathcal{C}_2, \mathcal{C}_2')$. Although, the distance $d(\mathcal{C}_1, \mathcal{C}_1')$ has to be greater than $d(\mathcal{C}_2, \mathcal{C}_2')$.

Furthermore, when comparing clusterings on different underlying graphs, worse situations can arise. For example, when comparing the same clustering on an arbitrary graph and its complement graph, all node–structural measures yield a distance of zero.

3 Engineering Graph–Structural Comparison Measures

In order to remedy some of the disadvantages of node–structural measures presented in Section 2, 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 a 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

Here, we present extensions of the node–structural measures given in Section 2. 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. A second objective is to adjust the three founding principles – counting pairs, overlaps and information theory – of the existing measures instead of adjusting each formula separately.

Counting Local Pairs Instead of categorizing every pair of the graph we only consider those pairs, that are connected by an edge in the graph. 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 because every pair of nodes is connected.

Thus, we obtain the graph–based versions of the Rand and adjusted Rand measure given in Equation (6) and (7):

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

$$\mathcal{AR}_g(\mathcal{C},\mathcal{C}') := 1 - \frac{e_{11}-t_3}{\frac{1}{2}(m(\mathcal{C})+m(\mathcal{C}'))-t_3} \text{ where } t_3 := \frac{m(\mathcal{C})m(\mathcal{C}')}{m} \tag{7}$$

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. The ij–th entry of the *degree–based confusion matrix* $CM^d := (m^d_{ij})$ indicates the sum of the degrees of the nodes in C_i and C', formally $m^d_{ij} := \deg(C_i \cap C'_j)$. Note, that if G is d–regular graph, then the equality $CM = CM^d/d$ holds.

The extensions of overlap based measures are obtained by substituting the confusion matrix by the degree–based confusion matrix. In certain cases, this may lead to different normalization factors. The extension of $\mathcal{NVD}$ is given in Equation (8).

$$\mathcal{NVD}_g(\mathcal{C},\mathcal{C}') := 1 - \frac{1}{4m}\sum_{i=1}^{k}\max_j m^d_{ij} - \frac{1}{4m}\sum_{j=1}^{l}\max_i m^d_{ij} \tag{8}$$

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 $m_{ij} = m^d_{ij}/d$.

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 picking a node randomly from the graph for measuring the uncertainty, we pick the end of an edge randomly. As a consequence, a node with high degree has an greater impact on the distance than a node with smaller degree. The formal definition is given in Equation (9) and (10).

$$\mathcal{H}_E(\mathcal{C}) := -\sum_{i=1}^{k} P_E(i)\log_2 P_E(i) \tag{9}$$

$$\mathcal{I}_E(\mathcal{C},\mathcal{C}') := \sum_{i=1}^{k}\sum_{j=1}^{l} P_E(i,j)\log_2\frac{P_E(i,j)}{P_E(i)P_E(j)} \ , \tag{10}$$

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 extension of $\mathcal{FJ}$ and $\mathcal{VI}$ are given in Equation (11) and (12)

$$\mathcal{FJ}_g(\mathcal{C},\mathcal{C}') := \begin{cases} 1 - \frac{2\mathcal{I}_E(\mathcal{C},\mathcal{C}')}{\mathcal{H}_E(\mathcal{C})+\mathcal{H}_E(\mathcal{C}')} & \text{, if } \mathcal{H}_E(\mathcal{C})+\mathcal{H}_E(\mathcal{C}') \neq 0 \\ 0 & \text{, otherwise} \end{cases} \tag{11}$$

$$\mathcal{VI}_g(\mathcal{C},\mathcal{C}') := \mathcal{H}_E(\mathcal{C}) + \mathcal{H}_E(\mathcal{C}') - 2\mathcal{I}_E(\mathcal{C},\mathcal{C}') \tag{12}$$

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. According to the node–structural version of the Variation of Information we use a $\log_2(n)$ normalized version. Meila showed in [16] that $\mathcal{VI} \leq \log_2(n)$ also holds for weighted clusterings.

3.2 A Novel Approach for Measuring Graph–Structural Distance

Although the extensions introduced in the previous section incorporates 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 (13).

$$\mathcal{ESD}(\mathcal{C}, \mathcal{C}') = \frac{|F_{\mathcal{C}} \cup F_{\mathcal{C}'}| - |F_{\mathcal{C}} \cap F_{\mathcal{C}'}|}{|F_{\mathcal{C}} \cup F_{\mathcal{C}'}|} = 1 - \frac{|F_{\mathcal{C}} \cap F_{\mathcal{C}'}|}{|F_{\mathcal{C}} \cup F_{\mathcal{C}'}|} \tag{13}$$

The size of a cluster editing set indicates the significance of a clustering, a smaller value indicates a significant clustering, a high value corresponds to an unrelated clustering. 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 setup focuses on structural properties of the clusterings, while the second setup concentrates on qualitative aspects. More precisely, they are built in the following way:

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 an *attractor generator* to create a graph with an initial clustering based on the following idea: k cluster centers, called *attractor* nodes, are placed uniformly at random in the plane and the corresponding clusters are filled up with $n - k$ additional nodes, called *satellite* nodes. We implemented this approach by first placing k attractor nodes on a discrete 2–dimensional grid such that the distance between each pairs is greater than $\alpha_{\min}$. Afterwards we insert for each empty grid position a satellite node with probability

$1/d$, where d is the euclidean distance to the closest attractor node a_j. If a node u is inserted in such a way, it is assigned to the cluster represented a_j and the edge $\{u, a_j\}$ is inserted. As a final step, all pairs of nodes having a distance less than the threshold $\alpha_{\max}$ are connected. To ensure global connectivity the attractor nodes are connected pairwise. In the following, we consider the ratio $\alpha_{\max}/\alpha_{\min}$, denoted by the density f. Furthermore, we pick k uniformly at random between 2 and $\sqrt[3]{n}$ at each execution.

All tests use $n = 1000$ nodes and are repeated at least 50 times 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. Figure 2 shows the measured quality by the indices coverage, performance and inter–cluster conductance. The tests consists of two

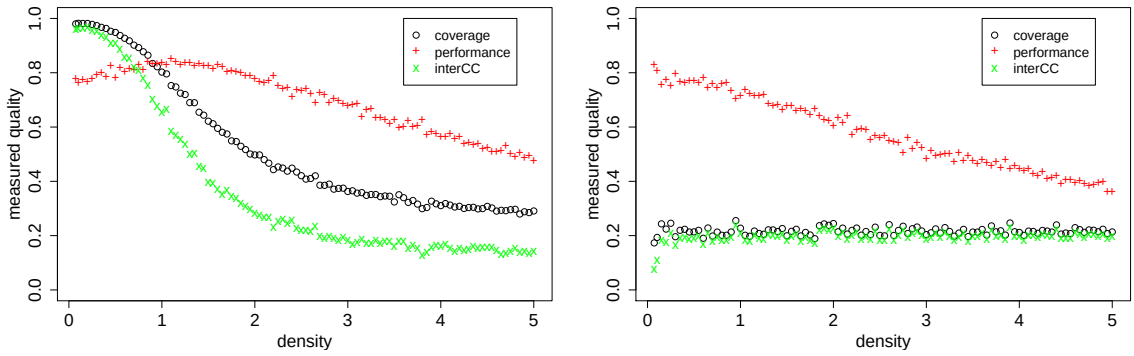
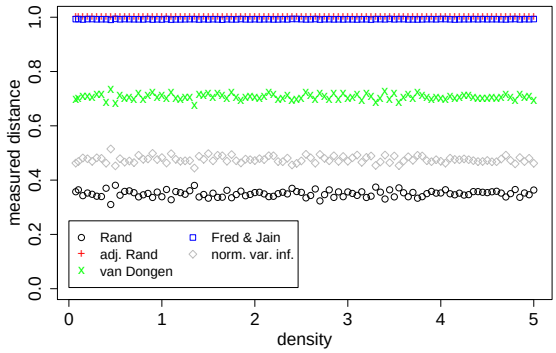


Fig. 2. Measured quality of the initial (left) and random (right) clustering with increasing density f. For random clusterings, the value of coverage and inter–cluster conductance almost coincides.

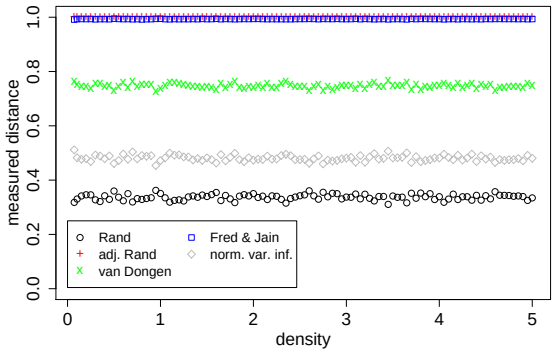
cases. On the one hand, the comparison of the generated and a random clustering (GvR) and on the other hand, the comparison of two random clusterings (RvR).

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, one could claim that 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.

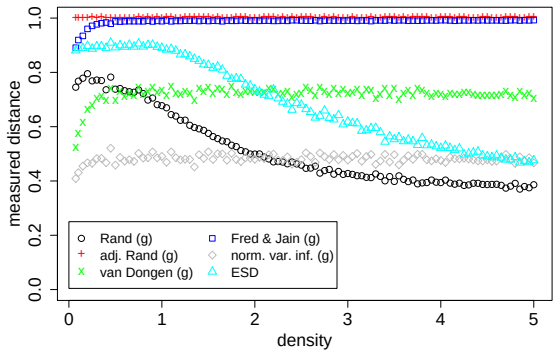
Figure 3 shows the result for the node– and graph–structural measures. By comparing Figure 3.1 and 3.2 it is evident that node–structural measures do not distinguish the two



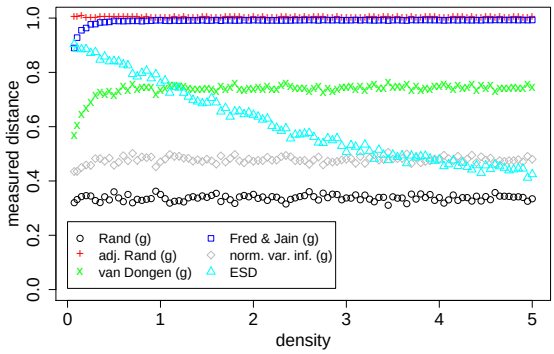
3.1: GvR nodes–structural



3.2: RvR nodes–structural



3.3: GvR graph–structural



3.4: RvR graph–structural

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

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 3.3 and 3.4). 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 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 are fairly regular for $f > 0.5$. Furthermore, the cluster are equal in size.

The strange behavior of Fred & Jain, van Dongen and the variation of Information for very small f is originated from the fact that for small f 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 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 the mean of coverage, performance, and inter–cluster conductance as index, the density is set to the values $f = 1$ (type 1) and $f = 3$ (type 2), and $M_{\max}$ increases from 0 to 500 using steps of 5.

Figure 4 shows the measured quality of the locally decreased clusterings with 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.

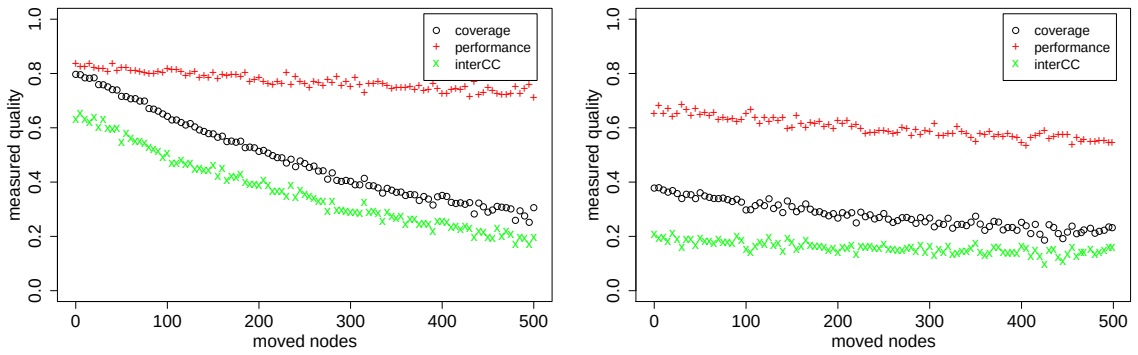
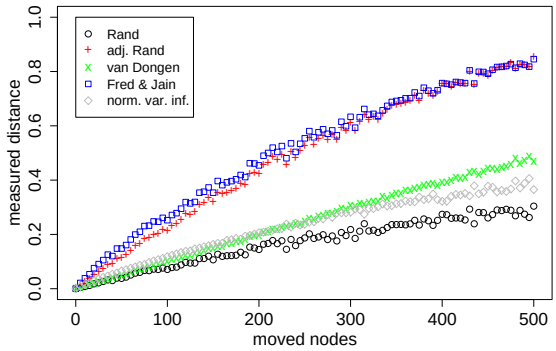


Fig. 4. Measured quality of the locally decreased clustering on attractors with $f = 1$ (left) and $f = 3$ (right).

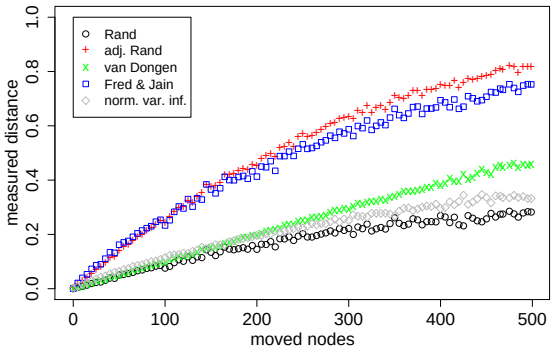
In addition, with increasing $M_{\max}$ the measured distance in type 2 should be smaller than in type 1, since the intuition is that in type 1 a very significant clustering is destroyed while on type 2 the loss of significance is lower.

Figure 5 shows the results for all measures on this specific setup. As shown in Figures 5.1 and 5.2, all node–structural measures hardly distinguish the two cases. This reveals additional disadvantages. Evaluating the graph–structural measures (Figures 5.3 and 5.4), 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.

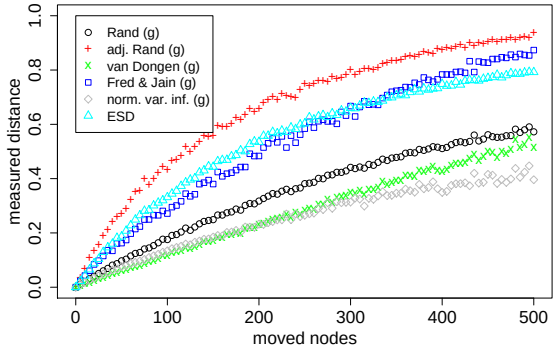
The unexpected behavior of the overlap and entropy based measures may be due to – as mentioned in Section 4.1 – the fact that for $f = 1$ and $f = 3$ attractor graphs have a



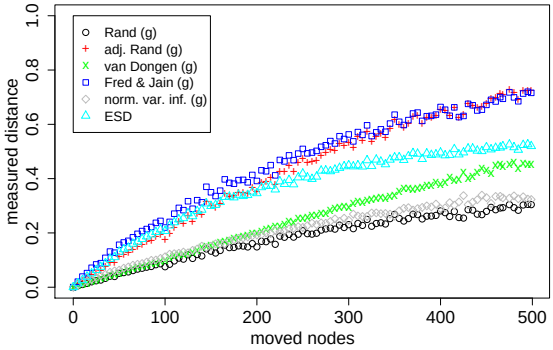
5.1: type 1 node–structural



5.2: type 2 node–structural



5.3: type 1 graph–structural



5.4: type 2 graph–structural

Fig. 5. Results of the local minimization setup

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 Conclusion

The experimental evaluation confirms the critical 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 incooperate structural and qualitative aspects, they can be used as a foundation for dynamic graph clustering.

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. Such a framework should yield a appropriate comparison measurement for every set of requested intuitions.

References

1. Jain, A.K., Dubes, R.C.: Algorithms for Clustering Data. Prentice Hall (1988)
2. Brandes, U., Erlebach, T., eds.: Network Analysis: Methodological Foundations. Volume 3418 of Lecture Notes in Computer Science. Springer-Verlag (2005)
3. Vidal, M.: Interactome modeling. FEBS Lett. **579** (2005) 1834–1838
4. Brandes, U., Gaertler, M., Wagner, D.: Experiments on graph clustering algorithms. In: Proceedings of the 11th Annual European Symposium on Algorithms (ESA'03). Volume 2832 of Lecture Notes in Computer Science. (2003) 568 – 579
5. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Physical Review E **70** (2004)
6. Gaertler, M., Görke, R., Wagner, D., Wagner, S.: How to Cluster Evolving Graphs. In: Proceedings of the European Conference of Complex Systems ECCS '06. (2006) to appear
7. van Dongen, S.: A cluster algorithm for graphs. Technical Report INS-R0010, National Research Institute for Mathematics and Computer Science in the Netherlands, Amsterdam, May (2000)
8. Wagner, S., Wagner, D.: Comparing Clusterings – An Overview. Technical Report 2006-4, ITI Wagner, Faculty of Informatics, Universität Karlsruhe (TH) (2006)
9. Rand, W.: Objective criteria for the evaluation of clustering methods. Journal of the American Statistical Association **66** (1971) 846–850
10. Morey, R., Agresti, A.: The measurement of classification agreement: An adjustment to the rand statistic for chance agreement. Educational and Psychological Measurement **44** (1984) 33–37
11. 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)
12. Shannon, C.: A mathematical theory of communication. Bell System Tech. J. **27** (1948) 379–423, 623–656
13. Strehl, A., Ghosh, J.: Cluster ensembles — a knowledge reuse framework for combining multiple partitions. J. Mach. Learn. Res. **3** (2003) 583–617
14. Fred, A., Jain, A.: Robust data clustering. In: In Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition,CVPR (3). (2003) 128–136
15. Meila, M.: Comparing clusterings by the variation of information. In Schölkopf, B., Warmuth, M., eds.: COLT. Volume 2777 of Lecture Notes in Computer Science., Springer (2003) 173–187
16. Meila, M.: Comparing clusterings - an axiomatic view. In Proceedings of the 22nd International Conference on Machine Learning, Bonn, Germany (2005)