SepMe: 2002 New Visual Separation Measures

Michael Aupetit[*]
Qatar Computing Research Institute, HBKU

Michael Sedlmair[†]
University of Vienna

Abstract

Our goal is to accurately model human class separation judgements in color-coded scatterplots. Towards this goal, we propose a set of 2002 visual separation measures, by systematically combining 17 neighborhood graphs and 14 class purity functions, with different parameterizations. Using a Machine Learning framework, we evaluate these measures based on how well they predict human separation judgements. We found that more than 58% of the 2002 new measures outperform the best state-of-the-art Distance Consistency (DSC) measure. Among the 2002, the best measure is the *average proportion of same-class neighbors among the 0.35-Observable Neighbors of each point of the target class (short GONG 0.35 DIR CPT)*, with a prediction accuracy of 92.9%, which is 11.7% better than DSC. We also discuss alternative, well-performing measures and give guidelines when to use which.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—Theory and methods.

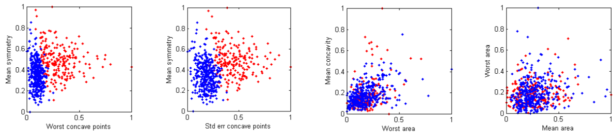
1 Introduction

In visual data analysis a human analyst visually inspects data to identify interesting, yet previously unknown patterns. Given the ever-growing complexity of modern data, however, directly visualizing all the data is usually not possible. It is therefore imperative to support the data analyst in selecting the most relevant data aggregations and graphical representations. The relevance in this case is a matter of human choice and depends on how humans perceive certain patterns in a given representation.

Visual quality measures aim to support this explorative process by pre-selecting visually interesting projections (scatterplots) [21, 27], proposing interesting ways on how to sort axes in parallel coordinates [11, 23], or by guiding the choice of synthetic dimension reduction algorithms before visualizing the data [20]. While many visual quality measures have been proposed [5], studies have shown that these measures are still far from optimal, posing ample opportunities for further improvements [15, 19, 20, 24].

Here, we focus on the specific case of *visual separation measures* in color-coded scatterplots of pre-classified data, which have gained much attention [1, 15, 17, 19–21, 23, 24]. The idea behind visual separation measures is illustrated in Figure 1. In such scatterplots, a human observer has a clear notion of “separated” (a & b) or “not separated” (c & d) classes [15, 20]. Mimicking this human notion of separability is the goal of visual separation measures.

Our goal is to build better visual separation measures, that is, measures that better resemble the human perception [1, 15, 19, 20]. Previous separation measures proposed certain ways of how this visual separability might be expressed directly, for instance, based on distance of points to class centroids, or on the entropy in grid cells [21]. We use a different approach: first, we thoroughly analyze existing separation measures to characterize recurrent features (Section 2). We then systematically generate a large set of 2002 new separation measures by exploring and parameterizing these features (Section 3). Finally, we evaluate and rank these measures following a Machine Learning-based methodology from previous work [19] (Section 4), which allows us to evaluate how well a measure predicts human judgements of visual class separation. In doing so, we not only identified the most promising measures among our set of 2002, but also showed our measures' and approach's superiority over previous work. 1170 of our measures (58.4%) performed better than the best state-of-the-art Distance Consistency (DSC) measure [21]. Furthermore, the best of our measures were better than DSC by 9.7 points (11.7%) of prediction accuracy. This is a substantial improvement over current state-of-the-art measures.



(a) (100, 94) (b) (99, 100) (c) (0, 7) (d) (14, 0)

Figure 1: 2D projections of Wisconsin Cancer data [16] showing two classes as red and blue dots. (a & b) show pairs of dimensions with visibly good class separation and (c & d) with poor separation. Below are the scores of two separation measures (GON, DSC), which are between 100 for best separability, and 0 for worst separability.

In summary, our work makes the following contributions:

- a set of 2002 new visual separation measures;
- a quantitative evaluation of these measures, showing superiority of our measures compared to the current state-of-the-art;
- discussion and guidelines on using alternative measures from the set of 2002;
- a qualitative case study on Wisconsin Cancer data [16], illustrating the differences between the new best measure and the best state-of-the-art one.

2 Background

Not least with the venerable work by Wilkinson and Anand on Scagnostics [27], visual quality measures have become a very active area of visualization research. Bertini et al. [5] provides a good overview over the breadth of research on visual quality measures in general. Here, we focus on a specific type of measures, separation measures, of which many have been defined in the Visualization, Pattern Recognition and Machine Learning communities. Some of these measures were intentionally designed with a specific eye towards human perception (e.g., [17, 21, 23]). Others have been developed with a sole focus on data (e.g., [6, 12, 13]); they could, however, similarly be used for perceptual modeling (although, the expectation is that they would perform inferior).

2.1 Common Features of Separation Measures

Taking a step back, the general idea behind all existing separation measures is to evaluate how “pure” the neighborhoods of a scatterplot's data points are. If neighborhoods include points from many different classes (i.e., colors) they are intermixed, if only from one class then they are pure.

While the general concept is the same, the actual measures differ in terms of how the neighborhood around points is defined. We differentiate between two **types of neighborhoods**:

[*] e-mail: maupetit@qf.org.qa
[†] e-mail: michael.sedlmair@univie.ac.at

- measures with *hard-neighborhoods* look at a specific subset of the data points close to the one under focus;
- *soft-neighborhoods* are based on a weighting of all the data points with respect to their distance to a focus point.

In addition, existing measures differ in terms of how they perform **class-purity evaluation** within these neighborhoods:

- some measures seek to *predict* the class of the focus point based on its neighbors' class: the more accurate this prediction, the higher the class-purity;
- others evaluate the *class-entropy* in these neighborhoods: the lower the entropy, the higher the class-purity;
- yet other measures compare the *within-class and between-class distances*: the larger the between-class and the lower the within-class distances, the higher the class-purity;
- finally, some measures compare the *class density*: the higher one of the class' density, the higher the class-purity.

The local class-purity values are then averaged over *all possible focus points*. A good separation is obtained when the averaged class-purity is high, that is, when the local neighborhood class-purity is high for a large set of focus points. All the measures have been tentatively designed to be monotonic with the human perception of class separation, that is, the higher a measure's value is, the better the perceived class separation (or vice versa).

2.2 Existing Separation Measures

Based on this analysis, we can loosely classify existing measures into 4 different families:

(i) Hard-neighborhood / class-prediction-based: The Distance Consistency (DSC) measure [21] is the proportion of data points x whose nearest class-center-of-mass belongs to the same class as x. **DSC** has been found to be the **best current state-of-the-art** visual separation measure, that is, the one that aligns best with human judgments of class separation [19]. Two other measures follow a similar approach. The Class Separation (CS) measure [17] is the average proportion of the neighbors of each data point x which belong to the same class as x. The neighbors are given considering the Extended Minimum Spanning Tree [17] of all the data points. The Hypothesis Margin (HM) measure [14] is the average of the differences between distances from each data point to its other-class nearest-neighbor and to its same-class nearest-neighbor.

(ii) Hard-neighborhood / class-entropy-based: The Distribution Consistency (DC) [21] looks at how points of different classes mix in an Euclidean ball centered at a pixel z with radius ε (entropy). The final score is the average of the entropy over all pixels. The Histogram Density Measure (HDM) [23] is similar to DC but the average is made over the cells of a square-grid partition of the image space. The class entropy at each cell is computed over the data points within the cell and its 8 adjacent cells in the grid.

(iii) Soft-neighborhood / within-between-class-distances-based: The Average Between (ABTN) and Average Within (AWTN) [15] measures evaluate the between-class separation and within-class homogeneity based on average within-class and between-class distances, respectively. Their ratio (ABW) has been used as a measure as well [19]. The Calinski-Harabasz (CAL) measure [6] quantifies the concentration of the classes around their center-of-mass using squared Euclidean distances. The Dunn's index (DUNN) [12] is the ratio of the minimum between-class distances over the maximum within-class distance. The Gamma (GAM) measure [4] is defined as follows: let ρ^+ be the number of times that a pair of data points that belong to the same class has distance smaller than two data points assigned to different classes, and let ρ^- be the opposite, then the Gamma measure is the ratio $(\rho^+ - \rho^-)/(\rho^+ + \rho^-)$. The Linear Discriminant Analysis (LDA) [13] seeks to find the linear mapping which maximizes the average pairwise distance between class center-of-mass while minimizing the average within-class pairwise distance for all classes. The Silhouette (SIL) measure [18] quantifies the separation as the difference between the average between-class distances and the average within-class distances, normalized by the maximum of these two quantities. Lastly, the Weighted Inter-Intra (WII) measure [22] is the average between-class over average within-class distances weighted by the respective size of the classes.

(iv) Soft-neighborhood / class-density-based: In the Class Density Measure (CDM) [23] the density of each class is estimated at each image pixel z as the inverse distance of z to its K^{th} nearest data point x of this class. The sum over the pixels of the absolute differences between these class-density images gives the CDM.

3 New Separation Measures

As discussed above, all existing measures model class separation in a specific way. Here, we propose a fundamentally different approach, namely, taking a step back and systematical explore how to instantiate the underlying features: (1) neighborhood definitions, and (2) class-purity functions. In a second step, we then use a Machine Learning framework [19] to find which of these instantiations perform best. In our exploration, we specifically focus on hard-neighborhood approaches, such as the ones described in (i) and (ii). Hard-neighborhood approaches provide a clear line between neighborhood and purity evaluation, giving us separate features to vary. Given this clear separation, we can leverage a large set of existing neighborhood or class-purity definitions, as further detailed below.

3.1 Conceptual Overview

Figure 2 gives a global overview over our approach, and shows the measures that we generated by systematically exploring—so far untested—combinations of these features.

In the first step, we define neighborhoods by building different proximity graphs. A proximity graph is a graph whose vertices are the data points and edges connect them depending on their relative positions in the data space. A proximity graph is by itself independent of class labels, and can be directed or not. The Nearest-Neighbor Graph (NNG) is, for instance, directed, and the ε-Ball Graph (EBG) not. It might also need to be parameterized, such as the K-Nearest Neighbor Graph (KNNG), or not, such as the Delaunay Graph (DG).

In the second step, we define two ways to consider the class-purity evaluation within such a graph. For *neighborhood-based class-purity* (Figure 2, top branch), we consider the neighbors of each point as given by the graph, and apply to each of these neighborhoods the class-purity function. Averaging the purity scores of all points will provide the separation measure: the higher the class-purity of each neighborhood, the higher their average, the higher the class separation. This general idea is also used by the state-of-the-art measures DC, HDM, CS, DSC and HM (i, ii). Our separation measures differ in the way the neighborhood is defined and the class-purity is evaluated over each neighborhood. We also explore additional aspects specific to this neighborhood-based class-purity approach. Most importantly, we vary the set of focus points to be used for computing a final separation score. While typically the final separation measure is gained from averaging over *all* points, we also propose to only iterate over the points of a certain *target class*.

In *component-based class-purity* (Figure 2, bottom branch), we cut mixed-edges in order to find the largest sets of neighbors in a given graph, for which the points are at the maximum purity. In other words, we want to find the class-connected-components of the graph [2, 28]. The larger such components or the smaller the number of mixed edges, the higher the class separation. This approach has not been studied for separation measures in scatterplots.

We now provide technical details on the exact instantiations of our features. We used 17 different neighborhood base-graphs (with different parameterizations), and 14 class-purity features. In Section 4, we will systematically evaluate these options and identify which combinations constitute the best separation measures.

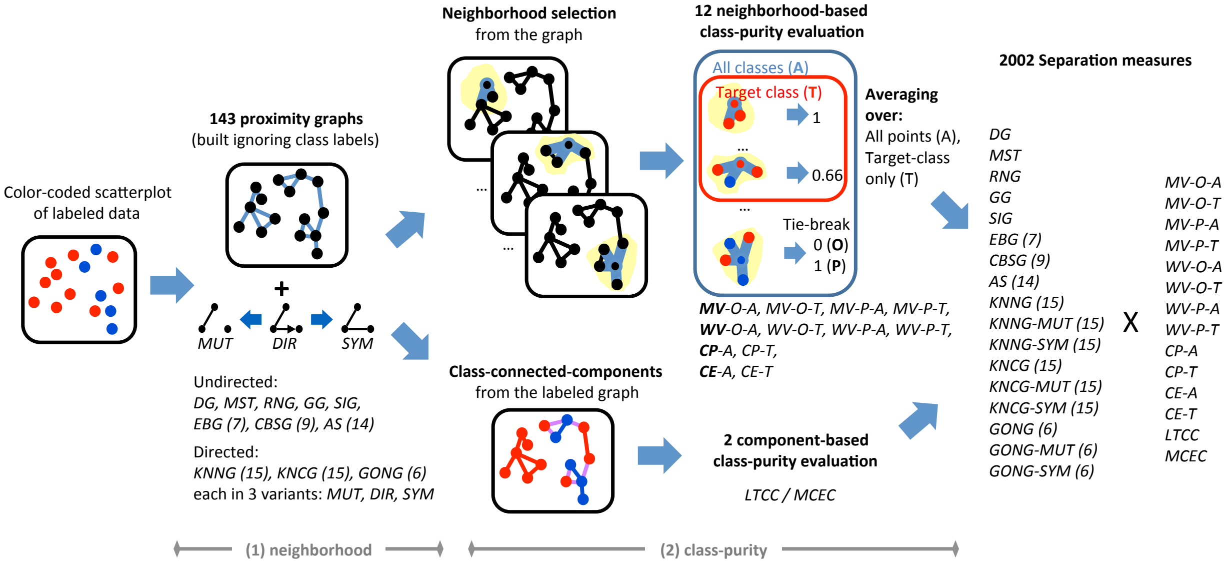


Figure 2: Given a color-coded scatterplot, our global framework (Section 3.1) consists of two major steps. (1) We build a proximity graph of the unlabeled data (Section 3.3). Overall, we explore 143 different graphs from 17 base graphs. Some graphs need to be parameterized (no. of parameters we tested are shown in parentheses). Some graphs are directed, in which case we also look at their mutual (MUT) or symmetrical variants (SYM). (2) We then compute 14 different class-purity functions (Section 3.4). 12 of them are neighborhood-based (top branch), and 2 component-based (bottom branch). For neighborhood-based evaluations, we further differentiate between considering all (A) or target only (T) focus points, and between optimistic (O) and pessimistic (P) tie-breaking rules. This process leads to 2002 new separation measures.

3.2 Formal Notations

We consider color-coded **scatterplots**. A color-coded scatterplot $\mathscr{I}(\mathfrak{s})$ (scatterplot for short in the sequel) is the graphical point-based representation in the discrete pixel space, of a labeled dataset $\mathfrak{s} = \{(x_i, c_i)_{i=1,\dots,n} | x_i \in \mathscr{R}, c_i \in \mathscr{C}_\mathfrak{s}\}$ lying in a 2-dimensional real data space $\mathscr{R} \subset \mathbb{R}^2$. The graphical representation $\mathscr{I}(\mathfrak{s})$ of the n points x_i in $\mathfrak{s}$ is color-coded based on their respective class label $c_i \in \mathscr{C}_\mathfrak{s}$ where $\mathscr{C}_\mathfrak{s} = \{0, \dots, k-1\}$ is a set of k classes.

Without loss of generality, we simplify the following discussion to two-class problems $\mathscr{C}_\mathfrak{s} = \{0, 1\}$. Previous work has shown that class separation can be judged for each class separately [19]. Multi-class problems can then be simply broken down into k two-class problems. We call $c_t = 1$ the target class, that is, the one for which we evaluate the separation from the others, and $c_o = 0$ the union of all other classes. To avoid confusion, we call scatterplot both the abstract dataset $\mathfrak{s}$ and its graphical representation $\mathscr{I}(\mathfrak{s})$ and we distinguish them only when necessary.

3.3 Neighborhood Selection

There is a large literature that deals with defining neighborhoods based on some underlying metric in the data space. These neighborhoods build a basic component of many data analysis tasks. For instance, neighborhoods are used to infer quantities unknown at test points from quantities known at training points (e.g., interpolation, classification), or to define geodesic distances or clusters in the data [8, 25]. These neighborhoods usually come as proximity graphs [7, 26] $G_{\mathfrak{s},g,\theta}$ which span the points x in a scatterplot $\mathfrak{s}$. The existence of edges depends on some function g of the points x ignoring their labels. Edges might be directed or undirected, and might depend on some scale parameter θ, for instance, the maximum distance between a point and its neighbors or the number of nearest neighbors. Any proximity graph $G_{\mathfrak{s},g,\theta}$ determines for each $x_i \in \mathfrak{s}$ a neighborhood $\mathscr{N}_{i,\theta} \subset \mathfrak{s} \backslash x_i$. In graphs with scale parameter θ, we consider parameterizations such that the monotonic increase of the parameter generates a hierarchy of neighborhoods where $\theta_1 \leq \theta_2 \Leftrightarrow G_{\mathfrak{s},g,\theta_1} \subseteq G_{\mathfrak{s},g,\theta_2} \Leftrightarrow \forall i, \mathscr{N}_{i,\theta_1} \subseteq \mathscr{N}_{i,\theta_2}$. We will note $G_{\mathfrak{s},g,\theta} = G_\mathfrak{s}$ and $\mathscr{N}_{i,\theta} = \mathscr{N}_i$ for short in the sequel. We denote $\rho_{i,j} = \|x_i - x_j\|$ the Euclidean distance between x_i and x_j.

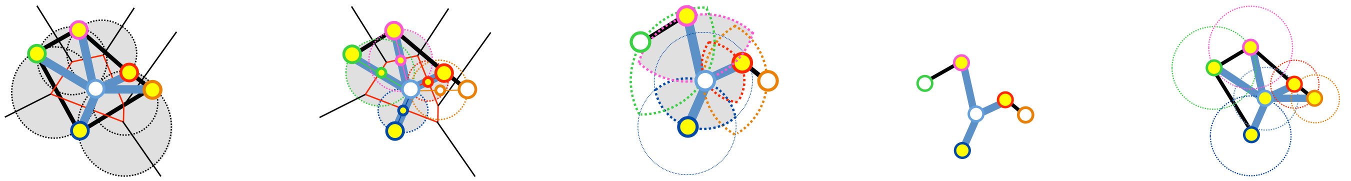
We derive the neighborhood feature from one of the following proximity graphs (all distances considered here are Euclidean and all points refer to points in the scatterplot $\mathfrak{s}$). All these graphs have been studied in [26] except the GONG [3] and the KNCG [9]. Their formal algebraic characterization (g) is beyond the scope of this paper and can be found in the above references. Figures 3(a)-(e) show graphical illustrations of the proximity graphs that do not have scale parameters; Figures 4(a)-(f) show the proximity graphs that have scale parameters. Without loss of generality, we simplify the presentation of graphs to points in general position, that is, no 4 points are concyclic, and no 3 points are aligned.

3.3.1 Base graphs without scale parameter

Fig. 3(a)—The **Delaunay Graph (DG)** [26] connects two points p and q if their Voronoi cells are adjacent (thin red edges), that is, if there exists a point in the plane that has both p and q as nearest neighbors. In other words, circumscribing discs (grey discs) of any edge triangle must be empty (except the vertices of the triangle). Hence, the neighbors of a point tend to surround it.

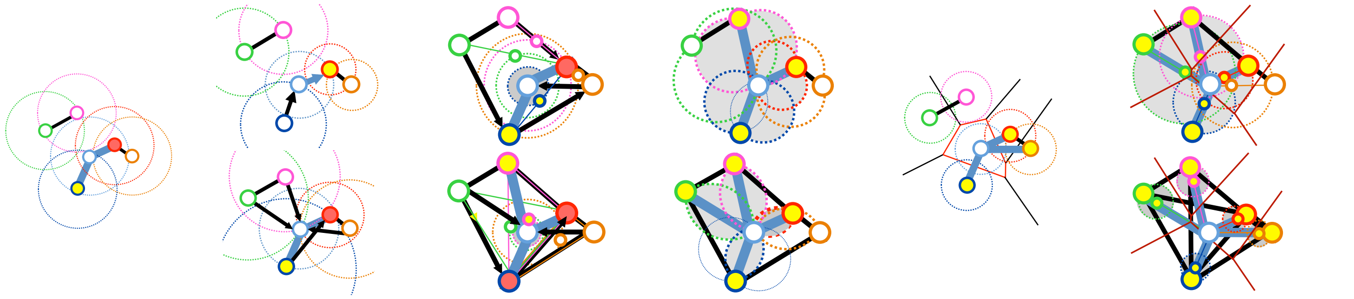
Fig. 3(b)—The **Gabriel Graph (GG)** [26] connects the focus point x_i to another point q if the disc with diameter $x_i q$ (dotted discs) contains no other point (grey discs); The Gabriel Graph is a connected Delaunay subgraph ($GG \subseteq DG$) whose edges cross the shared Voronoi boundary of their endpoints at the center of these discs (small discs). The Gabriel Graph is equivalent to the 0.5-Observable Neighbor Graph (GONG) and to the 0-skeleton graph (CBSG), both discussed later in this section. Similar to DG, neighbors of a point tend to surround it.

Fig. 3(c)—The **Relative Neighbor Graph (RNG)** [26] is a subgraph of the GG ($RNG \subseteq GG$) which connects two points p and q if no other point s has a distance to p and q smaller than the distance between p and q. Conceptually, it is equivalent to say that the intersection of two discs (regions with dotted edges) around a focus point x_i and a neighbor point q with radius $\rho_{x_i,q}$ must be empty (grey areas). Then, x_i and q will be connected (light blue edge). Again, neighbors of a point tend to surround it.



(a) DG (b) GG (c) RNG (d) MST (e) SIG

Figure 3: Base-graphs **without** scale parameter— The points of the scatterplot $\mathfrak{s}$ are represented as *colored circles*. The color is only meant to visually differentiate the points from each other; in particular, color does **not** reflect any class membership. The *light-blue point* is the focus point x_i, connected to its neighbors $\mathcal{N}_i$ through *light-blue edges*. Together *light-blue and black edges* form the proximity graph $G_{\mathfrak{s}}$. *Yellow-filled points* indicates neighbors of x_i. *Thin black and red lines* denote the Voronoi cells of the points. The *colored dotted circles*, as well as other *colored visual marks* show some properties of the graph that directly relate to the points with the same color. *Grey color* indicates empty-region areas which should not contain any point to enable some edge to exist. The actual proximity graphs (a)-(e) are described in Section 3.3.1.



(a) EBG (b) KNNG (c) KNCG (d) CBSG (e) AS (f) GONG

Figure 4: Base graphs **with** scale parameter—The visual encodings are the same as in Figure 4. Additionally *yellow-filled points* indicate neighbors of x_i last entering the neighborhood for the current setting of the scale parameter while *red-filled points* show neighbors also valid for a lower setting of the current scale parameter. Directed edges are indicated as *arrows*. *Small points* in KNCG indicate the center of gravity of the current neighbors plus the candidate one, and in GONG they indicate the intermediary point center of the must-be-empty disc. The actual proximity graphs (a)-(f) are described in Section 3.3.2.

Fig. 3(d)—The **Euclidean Minimum Spanning Tree (MST)** [26] is a subgraph of the RNG ($MST \subseteq RNG$) and is the unique connected subgraph of the Delaunay graph forming a tree and whose total edge length is minimum.

Fig. 3(e)—The **Sphere-of-Influence Graph (SIG)** [26] connects two points p and q if the distance $\rho_{p,q}$ is lower than the sum of the distance to their own nearest neighbors. In other words, if the circles centered at point p and q and passing through their respective nearest neighbor (dotted circles) intersect, then p and q are neighbors. Compared to the EBG (see below), the radius of the ball is "automatically" adapted to the local density of the points.

3.3.2 Base graphs with scale parameter

Fig. 4(a)—The **ε-Ball Graph (EBG)** [26] connects points p and q if their distance $\rho_{p,q} \leq \varepsilon$ (with $\varepsilon > 0$). Any point in the ε-ball (dotted circles) centered at p is neighbor of p.

Fig. 4(b)—The ***K*-Nearest Neighbor Graph (KNNG)** [26] connects each point to its K nearest neighbors (dotted circles passing through the nearest neighbor of each point ($K = 1$ top) or the 2^{nd} nearest neighbor ($K = 2$ bottom)). It generates neighborhoods with equal size $\forall i, |\mathcal{N}_i| = K$ and is similar to the EBG except that the scale is adapted to the local density of the points. The KNNG for $K = 1$ (NNG) is a subgraph of the MST ($NNG \subseteq MST$) and is identical to the GONG for $\gamma = 0$ and to the KNCG for $K = 1$. The KNNG is a directed graph, as shown in Figure 4(b).

Fig. 4(c)—The ***K*-Nearest Center of Gravity Graph (KNCG)** [9] allows generating equal size neighborhoods as the KNNG but with the goal to put each point the closest possible to the center of gravity of its neighbors. Initially, it connects each point x_i to its nearest neighbor for $K = 1$. Then, iteratively at step $K > 1$, it connects each point x_i to the candidate among the $N - K$ remaining points, for which the center of gravity (small circles) of this candidate (yellow-filled point) together with the $K - 1$ x_i's neighbors (red-filled point) is the nearest to x_i (dotted circles). Here the KNCG is shown for $K = 2$ (top) and $K = 3$ (bottom). The KNCG with $K = 1$ is identical to the NNG. The KNCG is a directed graph.

Fig. 4(d)—The **Circle-based β-Skeleton Graph (CBSG)** [26] connects two points p and q if for any other point s the angle formed by the lines joining s to p and q is lower than a threshold angle $\pi(1+\beta)/2$. The parameter β lies in the range of $\in [-1, 1]$: If -1, CBSG is the empty graph; if 0, CBSG is identical to the Gabriel graph; and if 1, then CBSG is the complete graph. Notice that we use a parameterization reverse to the one given in [26], in order to be consistent with the neighborhood hierarchy built on increasing scale parameter values. In Fig. 4(d), the CBSG is shown for $\beta < 0$ (top) and $\beta > 0$ (bottom). The interpretation of the dotted circles and grey area is the same as for the RNG illustration in Fig. 3(c).

Fig. 4(e)—The **α-Shape (AS)** graph [26] is a subgraph of DG ($AS \subseteq DG$) which connects two points if they are neighbors in the DG and their pairwise distance is not greater than $2/\alpha$ ($\alpha > 0$). In other words, points are neighbors if they are Delaunay neighbors and their $(1/\alpha)$-balls (dotted circles) intersect. This allows to select close neighbors of a point (EBG property) which also tend to surround it (DG property).

Fig. 4(f)—The **γ-Observable Neighbor Graph (GONG)** [3] connects each point x_i to a point p if the intermediary point (small circle) $p_i = \gamma p + (1-\gamma)x_i$ has p as its nearest neighbor among $\mathfrak{s} \backslash x_i$. That is, p_i belongs (small yellow-filled circles) to the Voronoi cell (thin red lines) of p over the set $\mathfrak{s} \backslash x_i$, i.e. the ball centered at p_i passing through p (dotted circles) is empty (grey areas). Here, we show the GONG for $\gamma = 0.3$ (top) and $\gamma = 0.8$ (bottom). It is identical to GG for $\gamma = 0.5$, a subgraph of GG for $\gamma \leq 0.5$, and identical

Table 1: Scale parameter settings.

Measure	# param.	Parameters
EBG	7	$\varepsilon \in \{0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5\} \times \Delta_{\mathfrak{s}}$ with $\Delta_{\mathfrak{s}}$ the maximum distance between the points in $\mathfrak{s}$
CBSG	9	$\beta \in \{-0.5, -0.4, \ldots, 0.2, 0.3\}$
AS	14	$\alpha = 1/(\zeta * \Delta_{DG})$ with $\zeta \in \{0.01, 0.02, \ldots, 0.05, 0.1, \ldots, 0.45, 0.5\}$ and Δ_{DG} the length of the longest edge in DG
KNNG	15	$K \in \{1, 2, \ldots, 15\}$
KNCG	15	$K \in \{1, 2, \ldots, 15\}$
GONG	6	$\gamma \in \{0.25, 0.3, 0.35, \ldots, 0.45, 0.5\}$

to NNG for $\gamma = 0$. GONG is a directed graph (not shown here).

As the graph directedness has an impact on the neighborhood $\mathscr{N}_i$ of each point, for the 3 directed (DIR) base graphs $G \in \{\text{GONG}, \text{KNNG}, \text{KNCG}\}$ we consider their mutual (MUT) variant $mut(G)$ deleting one-way edges, and their symmetrical (SYM) variant $sym(G)$ making all one-way edges undirected (see Figure 2).

Thus, we consider 17 different proximity base graphs (MUT and SYM included) plus their parametric variants leading to a total of 143 graphs in our experiments. The set of scale parameters that we explore are shown in Table 1. All these scale parameters are invariant to rescaling of the axes with constant aspect ratio.

3.4 Class-purity Evaluation

Based on these proximity graphs, we define several *separation measures*. We do that by applying different *class-purity functions* to a given neighborhood $\mathscr{N}_i$ given by a specific proximity graph $G_{\mathfrak{s}}$. All our measures are scaled linearly between 0 (no separation) and 100 (separation): the higher the measure, the higher the perceived separation is supposed to be.

As already indicated in Figure 2, we are pursuing two different approaches of class-purity evaluation. On the one hand, we create separation measures based on evaluating class-purity in the local neighborhood of points. We call this approach *neighborhood-based class-purity*; it is conceptually illustrated in Figure 5. On the other hand, we create measures that evaluate how global neighborhood graphs can be broken down into components (sub-graphs) made only of points of one class. We call this approach *component-based class-purity* evaluation, illustrated in Figure 6. After some further notations, we will describe both approaches in more detail.

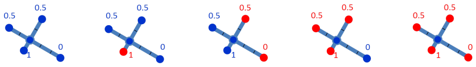
3.4.1 Formal notations

We note $c_i \in \{0,1\}$ the class of x_i and $\mathfrak{s}^1$ the subset of class-1 points in $\mathfrak{s}$ (points of the **target class**). The set of x_i's neighbors with class $v \in \{0,1\}$ is denoted $\mathscr{N}_i^{\,v}$, their proportion as $p_i^v = |\mathscr{N}_i^{\,v}|/|\mathscr{N}_i|$ and the proportion including x_i as $q_i^v = (|\mathscr{N}_i^{\,v}| + \delta_{c_i,v})/(|\mathscr{N}_i| + 1)$. Here, δ is defined as $\delta_{i,j} = 1$ if $i = j$ and 0 otherwise (i.e., Kronecker's *delta*). We call $\mathfrak{m}_C(G)$ the resulting separation measure applying the class-purity evaluation C to the graph G.

3.4.2 Measures based on neighborhood-based class-purity

We first describe four separation measures that are based on neighborhood-based class-purity: *Class-Proportion*, *Class-Entropy*, *Majority-Vote*, and *Weighted-Vote*. All these measures are defined as the average of local purity values over a set S of points in the scatterplot $\mathfrak{s}$. This set of evaluated points can be either *all points* (A), that is $S = \mathfrak{s}$, or only the points of the *target class* (T), that is $S = \mathfrak{s}^1$.

The **Class Proportion (CP)** measure computes the local proportion of the same-class neighbors for each focus point x_i in S. The final score is the average over all these local proportion scores. The local proportion is high in pure-class regions, close to 0.5 in mixed-class regions and low if the focus point x_i is a class-outlier (*e.g.* a blue point in the middle of red points). Formally, it is defined as: $\mathfrak{m}_{CP}(S, G_{\mathfrak{s}}) = \sum_{x_i \in S} p_i^{c_i}/|S|$. If $\mathscr{N}_i = \emptyset$, we set $p_i^{c_i} = 1$. This class-purity function is used in the existing CS [17] and DSC [21] measures.



CP	p_i^1	1	0.75	0.5	0.25	0
CE	q_i^0	0	0.2	0.4	0.6	0.8
	q_i^1	1	0.8	0.6	0.4	0.2
	h_i	0	0.72	0.97	0.97	0.72
MV	z_i	1	1	b	0	0
WV	z_i	1	b	1	0	0

Figure 5: Illustration of the neighborhood-based class-purity evaluation. The focus point x_i has label 1 (blue color). The class 0 is color-coded in red. The weights $w_{ij} \in [0,1]$ are indicated near the x_i's neighbors. For MV and WV, the tie-breaking rule can be either optimistic $b = 1$ (O) or pessimistic $b = 0$ (P). The left case corresponds to pure-class, the right case to class-outlier, and the intermediary cases to mixed-class situations.

The **Class Entropy (CE)** measure averages the local class entropy $h_i = -q_i^0 \log_2 q_i^0 - q_i^1 \log_2 q_i^1$ over all the points in S, depending on the size of the local neighborhood $n_i = |\mathscr{N}_i| + 1$. $\log_2$ accounts for the binary-class setting so the local entropy equals 1 for identical proportions of both classes, and 0 for pure-class situations. It is defined as: $\mathfrak{m}_{CE}(S, G_{\mathfrak{s}}) = \sum_{x_i \in S} n_i h_i / \sum_{x_i \in S} n_i$, and has been used in DC [21] and HDM [23].

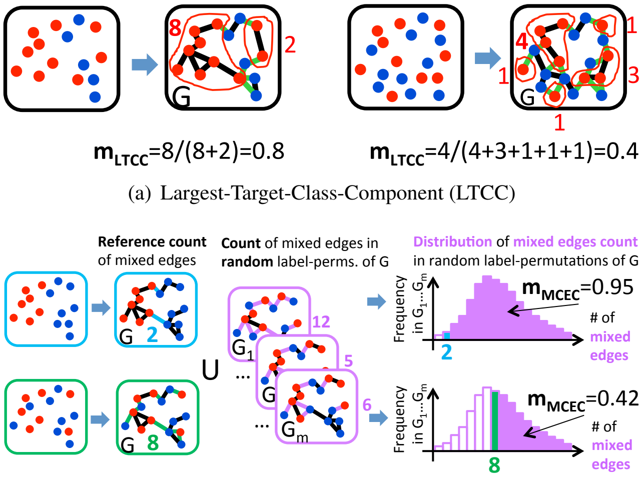
The **Majority-Vote (MV)** measure predicts the label of x_i based on the majority label of its neighbors and average the resulting votes over S. If the prediction is correct (vote = 1) then the local purity is high and if the prediction is false (vote = 0) x_i is a class-outlier possibly catching the attention of the user as a marker of non-separation. Compared to Class Proportion (CP), the vote puts emphasis on pure-class regions (vote = 1) and class-outliers (vote = 0). Mixed-class regions are either classified as pure-class if $p_i^{c_i} > p_i^{1-c_i}$ or as a class-outlier otherwise. The measure is defined as: $\mathfrak{m}_{MV}(S, G_{\mathfrak{s}}) = \sum_{x_i \in S} \delta_{c_i, z_i}/|S|$ where $z_i = 1$ if $p_i^1 > p_i^0$ and $z_i = 0$ if $p_i^1 < p_i^0$. If $p_i^1 = p_i^0$ then we set either $z_i = c_i$ as an *Optimistic* (O) tie-breaking rule, or $z_i = 1 - c_i$ as a *Pessimistic* (P) one. In any case if $\mathscr{N}_i = \emptyset$, then we set $z_i = c_i$.

The **Weighted-Vote (WV)** measure is similar to the Majority-Vote except it accounts for the relative distance to x_i of its neighbors. The closer the neighbors to x_i the higher their weight in the vote: $\forall x_j \in \mathscr{N}_i, w_{i,j} = \frac{\max_{k \in \mathscr{N}_i}(\rho_{i,k}) - \rho_{i,j}}{\max_{k \in \mathscr{N}_i}(\rho_{i,k}) - \min_{k \in \mathscr{N}_i}(\rho_{i,k})}$ is the normalized Euclidean similarity of x_i to its neighbor x_j. $w_{i,j}$ equals 1 for x_j the nearest of the x_i's neighbors $\mathscr{N}_i$ and 0 for the farthest. If $\mathscr{N}_i = \{x_j\}$, $w_{i,j}$ is set to 1. We define the total weight of the neighbors of class $v \in \{0,1\}$ as $W_i^v = \sum_{j \in \mathscr{N}_i^v} w_{i,j}$. The Weighted-Vote measure is then defined over the points in S as: $\mathfrak{m}_{WV}(S, G_{\mathfrak{s}}) = \sum_{x_i \in S} \delta_{c_i, z_i}/|S|$ where $z_i = 1$ if $W_i^1 > W_i^0$ and $z_i = 0$ if $W_i^1 < W_i^0$. The same optimistic and pessimistic tie-breaking rules as for the Majority-Vote apply here in case of $W_i^1 = W_i^0$. If $\mathscr{N}_i = \emptyset$, we set $z_i = c_i$. Both MV and WV have never been used before in separation measures.

Overall, we thus define 12 neighborhood-based separation measures. In addition to the different class-purity functions (MV, WV, CP, CE), the measures depend on the optimistic (O) or pessimistic (P) tie-breaking rule for vote-based measures (MV, WV), and the set of points, all (A) or target class (T), over which the local purity value is averaged (for all four). Hence, we have:

- $\mathfrak{m}_{CPA}$, $\mathfrak{m}_{CPT}$ based on Class Proportion
- $\mathfrak{m}_{CEA}$, $\mathfrak{m}_{CET}$ based on Class Entropy
- $\mathfrak{m}_{MVOA}$, $\mathfrak{m}_{MVPA}$, $\mathfrak{m}_{MVOT}$, $\mathfrak{m}_{MVPT}$ based on Majority-Vote
- $\mathfrak{m}_{WVOA}$, $\mathfrak{m}_{WVPA}$, $\mathfrak{m}_{WVOT}$, $\mathfrak{m}_{WVPT}$ based on Weighted-Vote

Figure 5 illustrates these measures using a simple example.


8
2
G
4
1
3
1
G
1
$\mathbf{m}_{LTCC}$=8/(8+2)=0.8
$\mathbf{m}_{LTCC}$=4/(4+3+1+1+1)=0.4
Reference count of mixed edges
Count of mixed edges in random label-perms. of G
Distribution of mixed edges count in random label-permutations of G
G 2
G 8
U
G_1
12
5
6
...
G
G_m
Frequency in $G_1 \ldots G_m$
$\mathbf{m}_{MCEC}$=0.95
# of mixed edges
2
$\mathbf{m}_{MCEC}$=0.42
# of mixed edges
8

(a) Largest-Target-Class-Component (LTCC)

(b) Mixed-Class-Edges-Cut (MCEC)

Figure 6: Illustration of the component-based class-purity evaluation and associated measures LTCC (top) and MCEC (bottom). **LTCC:** The proportion of points contained in the largest connected component of the target class (here red) is used as a separation measure: the greater the class separation, the lower the number of connected components and the greater their size. **MCEC:** the number of mixed edges (magenta) in the original labeled graph is compared to the distribution of the number of mixed edges in the same graph with randomly permuted labels. The proportion of random counts greater than the original count serves as a separation measure: the lower the number of mixed edges, the greater the class separation.

3.4.3 Measures based on component-based class-purity

We also consider two additional separation measures LTCC and MCEC based on component-based class-purity.

The **Largest-Target-Class-Component (LTCC)** measure is illustrated in Figure 6(a). LTCC is based on the graph $G_{\mathfrak{s}}^*$ which is the graph $G_{\mathfrak{s}}$ for which we delete the edges connecting points with different classes. We compute its class-connected components $CC_{G^*} = \{CC_1, \ldots, CC_\kappa\}$ as proposed in [2]. Here, we use this concept to define a separation measure as the proportion of points in the largest connected component of the target class, to all points of the target class: $\mathfrak{m}_{LTCC}(\mathfrak{s}, G_{\mathfrak{s}}) = \max_i(|CC_i \cap \mathfrak{s}^1|)/|\mathfrak{s}^1|$.

Lastly, the **Mixed-Class-Edges-Cut (MCEC)** is illustrated in Figure 6(b). The MCEC measure is based on the edge-cut statistic which is a measure of class learnability proposed by Zighed et al. [28]. It has been used to distinguish well-separated classes and well-structured chessboard-like class patterns from randomly mixed classes. The measure counts the reference number n_G of mixed-class edges (edges whose end points are of different class) of the proximity graph $G_{\mathfrak{s}}$. Then, it compares this number to its distribution over a set G^{rand} of m identical graphs $G_{\mathfrak{s}}$ with randomly permuted labels of their vertices: $G_1^{rand} \ldots G_m^{rand}$ (Note the number of points of each class remains the same, only the class assignment is changed at random). If n_G is small compared to the distribution values, that means the classes are clustered in a small number of large class-connected-components indicating high separation. Alternatively, n_G should be close to the average of the distribution if the classes are spatially intermixed in the scatterplot. Hence, we define the MCEC measure as the proportion of label-permuted graphs G^{rand} for which the number of mixed-class edges is greater than the reference number n_G. The greater the measure, the greater the separation: $\mathfrak{m}_{MCEC}(\mathfrak{s}, G_{\mathfrak{s}}) = |\{i \in \{1, \ldots, m\} | n_{G_i^{rand}} > n_G\}|/m$.

Finally, we end up with a set of 2002 **separation measures** based on the combination of the 14 distinct class-purity functions applied to each one of the 17 proximity graphs and their parametric variants (143 in total).

4 Evaluation

We now evaluate these measures with four goals in mind: (1) among the 2002, we are interested to identify which of these measures perform best; (2) we are similarly interested in how the newly proposed measures compare to the best current state-of-the-art DSC measure; (3) we want to more closely analyze the impact of the 17×14 different features on the performance of a measure; (4) finally, we want to qualitatively evaluate the performance on a realistic example.

Towards these goals, we performed two evaluations. First, we conducted a large-scale quantitative experiment using a Machine Learning framework for evaluating visual quality measures that we previously proposed [19]. This framework provides a way to quantify how well a measure predicts human judgments assigned beforehand to a large set of data (goals 1–3). We then illustrate the performance of the best new and old measures with a case study on the UCI Wisconsin Breast cancer data [16] (goal 4).

4.1 Quantitative Experiment

After a brief justification and explanation of our methodological choice, we present different results of our experiment.

4.1.1 Methods

To quantitatively evaluate the separation measures, we use an evaluation framework that we recently proposed [19]. We use this evaluation framework for two main reasons: First, traditional methods such as user studies [24], and manual data studies [20] simply do not scale to the sheer number of measures we are interested to test. A more algorithmic approach, as given by this framework, is therefore mandatory. Second, our goal is to evaluate measures as compared to human judgments. The framework we use is grounded in a large set of reliable human judgements and has been thoroughly evaluated, making it the currently best choice for evaluation.

The evaluation framework works as follows: The **best separation measure** $\mathfrak{m}^*$ is determined by evaluating how well measures predict human judgments on yet unseen scatterplots. The prediction accuracy is expressed as the **Area Under the Receiver Operating Characteristic Curve Bootstrapped Average (AUCBA)** [19], which gives a way to quantitatively compare measures among each other. An AUCBA of 50% or less means that a measure is not doing better than random guessing, while 100% indicates perfect alignment of a measure with human judgements. In other words, **the higher the AUCBA, the better the measure.** For our evaluation, we use 768 of the 828 datasets used in [19] containing two-class scatterplots (we had to excluded some that did not work for all measures). All data are publicly available[1]. These data have been carefully cleaned by removing points occluded to human viewers, as well as uncertain human class judgments, and hence can be seen as the most reliable source that is currently available. We use 10000 bootstrap samples to compute the AUCBA score for each measure (bootstrapping allows to generalize to yet unseen datasets). A detailed explanation and justification of the framework is beyond the scope of the paper, but can be found in [19].

4.1.2 Comparison of measures

Figure 7 shows the distribution of the AUCBA scores obtained for the 2002 new separation measures and the best state-of-the-art one, DSC [21]. Overall 1170 (58.4%) of our new measures outperformed DSC, showing the relevance of our global approach.

Figure 8 shows the AUC bootstrap distributions of the best measures given a specific neighborhood or class-purity feature. It shows

[1] http://sepme.cs.univie.ac.at/

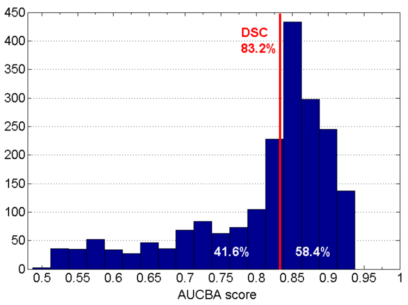


Figure 7: Distribution of AUCBA scores of the 2002 new separation measures and the best state-of-the-art DSC (red line). The closer the AUCBA is to 1, the better the measure; 0.5 equals a random guess. 58.4% of the new measures outperformed DSC.

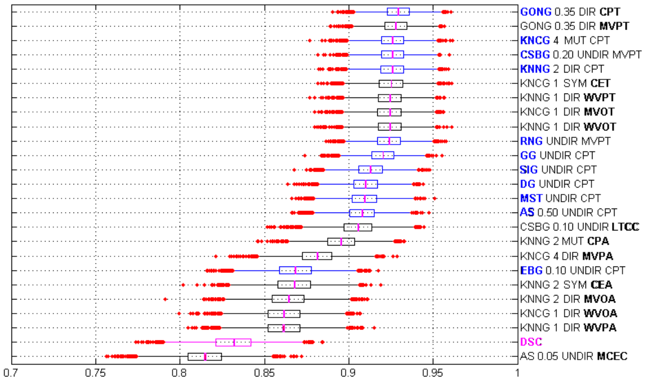


Figure 8: Box plot of the best new measures **given** a specific base graph (11 × bold blue font) and class-purity features (14 × bold black font), sorted by AUCBA. Along with the AUCBA (magenta line in the center of the box), the bootstrapping variance is shown using the default boxplot Matlab function which displays the interquartile range (box), 1.5 times the interquartile range above and below the box (whiskers) covering 99.3% of the data if they are normally distributed, and outliers (red dots). This distribution shows the expected performance on unseen data. The best state-of-the-art measure, DSC, is shown in magenta.

that the **best measure** in terms of AUCBA is the **GONG 0.35 DIR CPT**, that is, *the average Class-Proportion of the 0.35-Observable Neighbors of each point in the Target class*. GONG 0.35 DIR CPT has a 92.9% AUCBA score, which is 9.7 points (11.7%) better than the best state-of-the-art measure, DSC, whose AUCBA equals 83.2%. The remarkable difference between the two bootstrapped AUC distributions underlines the superiority of GONG 0.35 DIR CPT over the state-of-the-art DSC. Similarly good performance is obtained with other combinations of features like the surprisingly simple *average Class-Proportion of the 2-Nearest-Neighbors of each point in the Target class* (**KNNG 2 DIR CPT**). This measure could be used as an alternative to lower the computation complexity: $O(Kn\log(n))$ for KNNG instead of $O(n^2\log(n))$ for GONG.

4.1.3 Effects of features

We evaluate the effect of the different features combined in the separation measures. For each measure based on a specific feature we summarize the distribution of its AUCBA scores across all possible combinations of other independent features and compare their medians based on 95% confidence intervals.

We consider the 11 base graphs (EBG, KNNG, KNCG, CBSG, AS, GONG, MST, RNG, GG, DG, SIG) and the 14 class-purity functions (CPx,CEx,MVxy,WVxy,LTCC,MCEC) as main features.

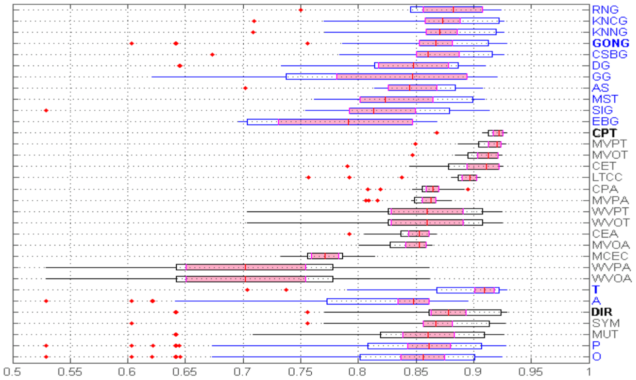


Figure 9: Comparison of the AUCBA distributions (boxplot) across all features for a given fixed one. 95% confidence intervals are shown in pink. Groups of features are separated based on black/blue colors. Features are ranked based on the median value of the AUCBA distribution for each group separately. The features involved in the overall best measure GONG 0.35 DIR CPT are shown in bold font.

As secondary features, we took directedness (DIR,SYM,MUT) for directed neighborhood graphs, and class-focus (A,T) and tie-breaking rules (O,P) for class-purity functions where applicable. We did not consider the scale parameters of the neighborhood graphs as a feature, but simply took the best performing across all possible scale parameter values.

Results for these different groups of features are shown in Figure 9. Two groups are significantly different at a ∼1% significance level if their confidence intervals do not overlap [10] (pink bars in Figure 9). Hence, our significance interpretation in the following is more conservative than the common 5% level.

The RNG is the *neighborhood* with the highest AUCBA median. RNG, KNCG, KNNG, GONG and CBSG have a similar small spread and have all a significantly higher median of the AUCBA score than SIG and EBG. The RNG having the lowest spread could be used as a conservative option to replace the GONG.

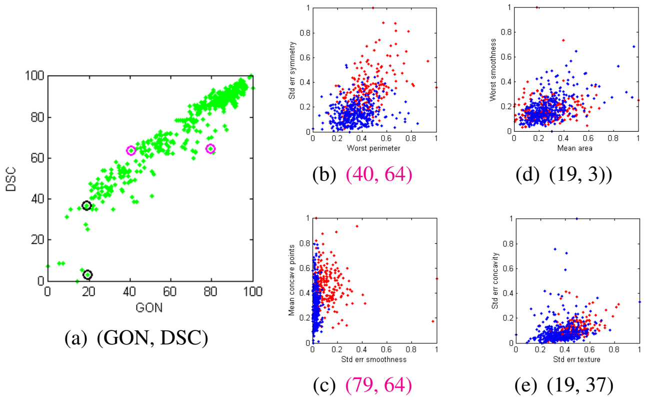
The AUCBA median of the CPT *purity function* is significantly better than all others except MVPT, MVOT and CET, but it has a lower spread than these. Averaging over the target class (T) is significantly better than over all classes (A). There is neither a significant difference between directed (DIR) graphs and their variants (SYM, MUT), nor between Pessimistic (P) and Optimistic (O) tie-breaking rules.

4.1.4 Summary

Our results indicate that the human judgment of class separation in scatterplots seems to be based on averaging over local neighborhoods (RNG, KNCG, KNNG, GONG and CSBG with small scale parameter as seen in the Figure 8), and Class Proportion (CP) or Majority Vote (MV) class-purity functions, focused on the target class only (T). Global statistics (MCEC) or possibly distant neighbors (fixed scale DG, GG, MST, SIG or non-density-adaptive AS and EBG) did not perform as well. These findings can help to study the human visual perception of class separation, and to further improve separation measures.

4.2 Case Study

To qualitatively illustrate our findings, we compare the best new separation measure *GONG 0.35 DIR CPT* (short GON) and the best state-of-the-art measure DSC [21] on scatterplots generated from the UCI Wisconsin Breast Cancer data [16]. This dataset includes 569 instances of normal (357) and tumor (212) cell images, characterized by 30 dimensions. Among the resulting 435 scatterplots



(a) (GON, DSC)

(b) (40, 64)

(d) (19, 3))

(c) (79, 64)

(e) (19, 37)

Figure 10: (a) GON and DSC scores for all pairs of Wisconsin Cancer data scatterplots. Scatterplots with DSC fixed and low (b) and high (c) GON values (magenta circles in (a)). Scatterplots with GON fixed and low (d) and high (e) DSC values (black circles in (a)).

(axis-aligned projections), the goal is to identify those that visually separate the two classes.

We linearly scaled all dimensions between 0 and 1, and computed the DSC and GON scores for each of these scatterplots. We then also linearly scaled these scores between 0 and 100. Figure 1 shows the best and the worst scatterplots with respect to both measures. While there is no strong qualitative difference, we note that the scatterplots are not ranked in the same order. Figure 10(a) shows all pairs of (GON, DSC) scores from all 435 scatterplots. The distribution resides nearby the diagonal and shows, as expected, a strong correlation between DSC and GON scores. We pick the two pairs of scatterplots for which one measure is identical while the other has the largest variation (magenta and black circles in Figure 10(a)). When DSC equals 64 for both scatterplots, GON varies from 40 (b) to 79 (c). The separation looks greater when GON is greater whereas DSC remains oddly fix. When GON equals 19 for both scatterplots, DSC varies from 3 (d) to 37 (e). DSC varies a lot despite the separation looks pretty similar in both views as correctly measured by GON. Both these extreme cases confirm our quantitative results that GON is better than DSC at mimicking human class separation. Inspecting scatterplots in decreasing order of GON scores is likely to be more efficient to discover interesting patterns than using DSC scores.

5 Conclusions and Future Work

We have proposed a broad set of 2002 separation measures for color-coded scatterplots. Through systematic evaluation, we identified *the average Class-Proportion of the 0.35-Observable Neighbors of each points in the Target class* (short *GON*) to be the best separation measure. It predicts human judgment with a 92.9% accuracy (AUC bootstrapped average), and outperforms the best state-of-the-art measure, Distance Consistency, by more than 11.7%.

While 92.9% is a high score, specifically compared to current measures [19], we envision even better measures in the future. For instance, one could consider adapting the parameters of a measure for each scatterplot separately. Also, the evaluation framework that we currently use [19] is based on a binary setting, that is, a class is either labeled separable or not. While it is the best framework available at the moment, this binary setting over-simplifies the rich nature of human separation judgments. Further improving the underlying framework will facilitate more accurate testing and open doors for even better measures. Ideally, such measures would then provide a degree of separation that resembles human perception. Finally, while we have focused on the case of separation measures, we hope that our approaches and insights will inspire researchers to improve other types of visual quality measures as well [5].

Acknowledgements

This work was partly funded by FFG project 845898 (VALID).

References

[1] G. Albuquerque, M. Eisemann, and M. Magnor. Perception-based visual quality measures. In *IEEE VAST*, pages 11–18, 2011.
[2] M. Aupetit and T. Catz. High-dimensional labeled data analysis with topology representing graphs. *Neurocomputing*, 63:139–169, 2005.
[3] M. Aupetit, P. Couturier, and P. Massotte. γ-observable neighbours for vector quantization. *Neural networks*, 15(8):1017–1027, 2002.
[4] F. Baker and L. Hubert. Measuring the power of hierarchical cluster analysis. *J. of the American Statistical Assoc.*, 70(349):31–38, 1975.
[5] E. Bertini, A. Tatu, and D. A. Keim. Quality metrics in high-dimensional data visualization: An overview and systematization. *IEEE TVCG*, 17(12):2203–2212, 2011.
[6] T. Caliński and J. Harabasz. A dendrite method for cluster analysis. *Comm. in Statistics Simulation and Computation*, 3(1):1–27, 1974.
[7] J. Cardinal, S. Collette, and S. Langerman. Empty region graphs. *Computational Geometry*, 42(3):183–195, 2009.
[8] M. . Carreira-Perpin and R. S. Zemel. Proximity graphs for clustering and manifold learning. In *NIPS*, pages 225–232, 2004.
[9] B. B. Chaudhuri. A new definition of neighborhood of a point in multi-dimensional space. *Pattern Recognition Let.*, 17(1):11–17, 1996.
[10] G. Cumming and S. Finch. Inference by eye: Confidence intervals and how to read pictures of data. *American Psych.*, 60(2):170, 2005.
[11] A. Dasgupta and R. Kosara. Pargnostics: Screen-space metrics for parallel coordinates. *IEEE TVCG*, 16(6):1017–1026, 2010.
[12] J. C. Dunn. Well-separated clusters and optimal fuzzy partitions. *J. of Cybernetics*, 4(1):95–104, 1974.
[13] K. Fukunaga. *Introduction to statistical pattern recognition*. Comp. Science and Scientific Comp. Academic Press, 2nd edition, 1990.
[14] R. Gilad-Bachrach, A. Navot, and N. Tishby. Margin based feature selection – theory and algorithms. In *ICML*, pages 43–50, 2004.
[15] J. M. Lewis, M. Ackerman, and V. de Sa. Human cluster evaluation and formal quality measures: A comparative study. In *CogSci*, pages 1870–1875, 2012.
[16] M. Lichman. UCI machine learning repository, 2013.
[17] R. Motta, R. Minghim, A. A. Lopes, and M. C. Oliveira. Graph-based measures to assist user assessment of multidimensional projections. *Neurocomputing*, pages 583–598, 2015.
[18] P. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. of Computational and Applied Mathematics*, 20(1):53–65, 1987.
[19] M. Sedlmair and M. Aupetit. Data-driven evaluation of visual quality measures. *Computer Graphics Forum*, 34(3):201–210, 2015.
[20] M. Sedlmair, A. Tatu, T. M., and T. Munzner. A taxonomy of visual cluster separation factors. *Computer Graphics Forum*, 31(3):1335–1344, 2012.
[21] M. Sips, B. Neubert, J. P. Lewis, and P. Hanrahan. Selecting good views of high-dimensional data using class consistency. *Computer Graphics Forum*, 28(3):831–838, 2009.
[22] A. Strehl. *Relationship-based Clustering and Cluster Ensembles for High-dimensional Data Mining*. PhD thesis, University of Texas at Austin, 2002.
[23] A. Tatu, G. Albuquerque, M. Eisemann, J. Schneidewind, H. Theisel, M. Magnor, and D. Keim. Combining automated analysis and visualization techniques for effective exploration of high-dimensional data. In *IEEE VAST*, pages 59–66, 2009.
[24] A. Tatu, P. Bak, E. Bertini, D. Keim, and J. Schneidewind. Visual quality metrics and human perception: An initial study on 2D projections of large multidimensional data. In *AVI*, pages 49–56, 2010.
[25] G. Toussaint. Geometric proximity graphs for improving nearest neighbor methods in instance-based learning and data mining. *IJCGA*, 15(02):101–150, 2005.
[26] R. Veltkamp. The gamma-neighborhood graph. *CG*, 1:227–246, 1991.
[27] L. Wilkinson and A. Anand. Graph-theoretic scagnostics. *IEEE InfoVis*, pages 157–164, 2005.
[28] D. Zighed, S. Lallich, and F. Muhlenbach. A statistical approach to class separability. *Applied Stochastic Models in Business and Industry*, 21(2):187–197, 2005.