A FLEXIBLE OUTLIER DETECTOR BASED ON A TOPOLOGY GIVEN BY GRAPH COMMUNITIES

A PREPRINT

O. Ramos Terrades, A. Berenguel, D. Gil*
Computer Vision Center - Dept. of Computer Science
Universitat Autònoma de Barcelona
Cerdanyola del Vallés, 08193
oriolrt,aberenguel,debora@cvc.uab.cat

February 19, 2020

ABSTRACT

Outlier, or anomaly, detection is essential for optimal performance of machine learning methods and statistical predictive models. It is not just a technical step in a data cleaning process but a key topic in many fields such as fraudulent document detection, in medical applications and assisted diagnosis systems or detecting security threats. In contrast to population-based methods, neighborhood based local approaches are simple flexible methods that have the potential to perform well in small sample size unbalanced problems. However, a main concern of local approaches is the impact that the computation of each sample neighborhood has on the method performance. Most approaches use a distance in the feature space to define a single neighborhood that requires careful selection of several parameters.

This work presents a local approach based on a local measure of the heterogeneity of sample labels in the feature space considered as a topological manifold. Topology is computed using the communities of a weighted graph codifying mutual nearest neighbors in the feature space. This way, we provide with a set of multiple neighborhoods able to describe the structure of complex spaces without parameter fine tuning. The extensive experiments on real-world data sets show that our approach overall outperforms, both, local and global strategies in multi and single view settings.

***K*eywords** Outlier detectors · graph community detectors · local structure

1 Introduction

Outlier, or anomaly, detection is a major issue in many learning-based algorithms since the presence of outlier data on training data might affect the proper estimation of model parameters. Moreover, outlier detection becomes harder when data changes along time, since it is unclear how to distinguish proper data, coming from a new arising class, from corrupted, or outlier, data. Outlier detection is not just a technical step in a data cleaning process. It is, by itself, a key topic in many fields such as fraudulent document detection, in identity documents or passports, insurances claims and health care fraud; medical applications and assisted diagnosis systems; detection of security threats, etc. In all these tasks, an accurate outlier detection will impact in the economic balance of any company or to properly detect any security threat, for instance.

The outlier concept is fuzzy, it depends on each task and thus each detection method provides its own definition. For instance, Chandola et al. [2009] define an outlier as a sample that "does not conform to expected behavior" and they classify them as *point outliers*, *contextual outliers* and *collective outliers*. Point outliers are samples with abnormal feature values not expected for any of the classes and they correspond to isolates samples either not belonging to any cluster or not following the population distribution. Contextual outliers are samples which are labelled differently

*DGil is a Serra Hunter Fellow

from their neighboring samples, which define their context. Collective outliers are small group of samples sharing unusual features that are clustered together. In Zhao and et al [2018] outliers are split into *attribute outliers* and *class outliers* in the context of multimodal representations. In such representations there are two or more feature spaces (called *views*) for each sample. In this context, attribute outliers correspond to point outliers, while class outliers are similar to contextual outliers in the measure that are samples labelled differently across *views*.

Existing methods for detection of outliers can be categorized into global and local approaches. Global methods are population based and they model the distribution in the feature space of a set of (annotated) samples. Population distribution can be modelled using either parametric global descriptors or unsupervised clustering approaches. These methods are suited to detect point (or attribute) outliers. Local methods are based on a description of the structure of each sample's neighbors in the feature space and, thus, they are better suited to detect, both, attribute and class outliers.

In this work we introduce a local approach, which takes benefit of topological properties of neighboring samples, to detect class and attribute outliers in both single view and multiview data.

2 Related work

2.1 Global methods

Global approaches seek to estimate parametric data distributions and define outliers as points not folowing the estimated distribution. For instance, Yang et al. [2009] use Gaussian mixture models (GMM) for outlier detection in which they define each point as a cluster and the outlierness score is the reciprocal of the point likelihood. In general, data distribution cannot be properly estimated in small sample size sets since outliers become influential points who deviates global approaches from normal population. This leads to lack of reproducibility and drop their potential for outlier detection.

With the advent of big data and deep learning techniques this main drawback of global approaches seems to be mitigated since they can learn complex data structures from big amount of data. A main issue when dealing with big data is labelling enough samples for training and testing deep learning methods, which is especially difficult in such an imbalanced classification task. Chalapathy and Chawla [2019] define four groups of outlier detectors methods: unsupervised, semi-supervised, hybrid and One-Class neural networks. While is clear which methods belong to unsupervised and semi-supervised methods, hybrid methods use deep learning architectures for feature extraction and then use traditional outlier detectors. Finally, One-Class neural network (OC-NN) method is inspired by kernel methods on one-class classification tasks [Chalapathy et al., 2018]. A variant of OC-NN architectures is Deep Support Vector Data Description (Deep-SVDD) [Ruff et al., 2018]. In that work, the authors train a deep neural network to extract common variation factors by mapping close inlier data instances to the center of a hyper-sphere. Generative adversarial active learning (GAAL) networks have also been used for outlier detection [Liu et al., 2018].

There are some real use cases in which there are not enough data for learning end-to-end methods. Although domain adaptation and transfer learning techniques can be used to deal with small datasets, the success of these end-to-end methods relies on their capability to learn specific features for the outlier detection task. However, there are some real use case scenarios, like clinical decision support systems or personalized models, in which feature vector are defined beforehand for the particular task and hence cannot be modified, or changed, to detect outliers.

2.2 Local methods

Local methods are based on a description of each sample's neighborhood usually defined using the Euclidean distance among samples in the feature space. Given that local methods define outliers in terms of such distances, they are distribution free and, thus, better suited for unbalanced small datasets.

Most local approaches, like Ramaswamy et al. [2000] or Angiulli and Pizzuti [2002], define outliers in terms of the distance to the k-th nearest point. LOF is an outlier detector method defined in the context of knowledge discovery in databases that assigns an outlierness score to each sample based on local information [Breunig et al., 2000]. This score is computed in terms of the distance to the k nearest neighbors of each point, meaning a score near to 1 to not be an outlier while higher values provides higher certainty of being it. LOCI [Papadimitriou et al., 2003] also bases on k-th nearest neighbours to define a multi-granularity deviation factor (MDEF) as outlierness measure. The MDEF is the relative deviation of sample's local neighborhood density from the average local neighborhood density, so that a point is an outlier if its MDEF is sufficient large. This way LOCI is effective to detect point outliers and collective outliers, as well. The Isolation Forest, IF, technique [Liu et al., 2012] builds a tree that isolates *attribute* outliers using a binary search. Since an attribute outlier has different values compared to inlier points IF detects them as points such that the length path to reach them is significantly shorter than the mean length to reach any other point.

The selection of the parameters defining neighborhoods is a main bottleneck in local approaches. In particular, the selection of the number k of nearest neighbors is crucial since it greatly affects the performance of methods. Therefore, several strategies for optimal selection of the parameter k have been proposed since the early years of nearest neighborhood approaches.

2.3 Graph based methods

Neighboring relationship can be defined in terms of distances but also in terms of *friendship* relationship on structured data, like graphs. In graph structured data and networks, outliers are also linked to topological variation of subgraphs that broke a repetitive pattern. To detect these local structural singularities, it has also been proposed outlier detectors for graph-based representations.

The early Brito et al. [1997] studied the relationship between connectivity of mutual-knn graph (MKG) and outlier detection, providing a criteria in terms of the graph topology. In particular, they studied the geometric properties of the underlying data points distribution and derived a theoretical criteria to set a value of k ensuring that the connected components of the graph correspond to clusters in the feature space. In that context, outliers were those samples which did not belong to any of this connected components, that is, they correspond to single node connected components. The work in Brito et al. [1997] was later extended in Maier et al. [2007]. There, the authors provide further insights on the mutual k-nn graphs to derive tighter bounds to estimate the optimal k to build a MKG. A main inconvenience for a practical use is that these bounds are still hard to compute with real data having class outliers.

Ning et al. [2018] propose an algorithm to search the optimal k to build the MKG. In that work, the authors introduce the concepts of *stability state* and *appropriate k (apk)* for a MKG and they propose and algorithm to search the optimal k. Finally, the most recent work of Wang et al. [2019] also proposes a variation of MKG to minimize the impact of k. In their approach, they compute multiple local proximity graphs for k sampled uniformly in a range of values. Then, their approach does not rely in finding the *optimal* k of the MKG but in combining the information of all MKGss using a random walk to detect outliers.

Aside fixing the parameter k, another concern about existing local approaches is that outlier scores are defined from the structure of a single neighborhood defined using distances. From a mathematical point of view, this implies that the topological structure of feature spaces is modelled as a norm or metric space [Munkres, 2000]. Although Euclidean spaces admit a topology defined from a metric or norm, these approaches might fail to properly describe more complex spaces (like manifolds [Munkres, 2000]). In this context, topology is a powerful mathematical approach to model the structure of complex manifolds without the assumption of any parametric model for the data.

Methods for the detection of communities in social networks can provide a mean to extract a set of topological neighbours from MKG. In this context, attribute outliers are often non connected, or hardly connected, individuals. Meanwhile, class outliers correspond to community members with a user profile, or interests, far of most of community members and they are often ignored. While detecting attribute outliers is done using graph topological properties, computation of a topology in non-structured spaces given by a discrete set of population samples still remains a challenge for topology. Gil et al. [2017] presented a local method based on neighborhoods given by the communities of the graph built from the samples distances. Despite the promising results in the diagnosis of lung cancer in confocal images, this topology given by graph communities is prone to exclude many points that could not be considered isolated attribute outliers [Mielgo, 2017]. Besides, like other local methods, a main concern is the impact of the parameters used to compute the graph used to detect communities. In case k is too small, communities might exclude points that are not are actually outliers [Mielgo, 2017], while increasing k produces a single community including all points. Thus, the method requires a proper accurate value for the parameter k, which the authors fine-tuned to give optimal results.

Direct analysis of the local structural properties of graphs and networks also allows the detection of outliers and anomalies. The OddBall method is a widely used method for anomaly detection in networks which focus on detecting nodes having topological properties significantly different compared to neighboring nodes [Akoglu et al., 2010]. More recently, Elliott et al. [2019] propose an extension of the NetEMD network method [Wegner et al., 2018] to detect graph anomalies and spectral localization statistics in financial transaction networks. Other methods base on the the Minimum Description Length (MDL) principle. In Noble and Cook [2003] anomalous sub-graphs are detected using variants of MDL. Eberle and Holder [2007] use MDL as well as other probabilistic measures to detect several types of graph anomalies (e.g. unexpected/missing nodes/edges).

2.4 Multiview methods

All methods described above are specifically designed for the detection of attribute outliers in single view problems. Multimodal, or multiview, data can benefit of the different sources of information to detect data outliers. In particular, class outliers are easily spotted when a sample is labelled differently across views. In this context, Gao et al. [2013] detect class and attribute outliers based on the spectral analysis of the combined adjacency matrix. In that paper, those samples that lie in the kernel space of the combined adjacency matrix are identified as outliers. A different approach is the one proposed in Zhao and Fu [2015] and Zhao and et al [2018]. In that works, the authors propose a generalized K-means method that learns cluster label consistencies across views. Samples having different cluster labels are classified as class outliers. A limitation specific to multi-view methods is the combination of information across views, which usually leads to under-detection of abnormalities arising in single views. Finally, another limitation is that being designed for more than view, multi-view methods are prone to perform poorly in single view problems.

2.5 Our Contributions

In this work we present a local approach based on a measure of sample labels diversity in a set of topological neighbourhoods of each sample. We compute this topology extending the graph structural method presented in Gil et al. [2017]. In particular, we refine the initial topology given by community detection methods to include isolated non-outlier points. Sample diversity is computed using probabilistic measures, which summarize the variability of sample labels in the set of topological neighbourhoods samples belong to. These diversity measures provide a normalized *outlierness* representation feature space. This normalized space is independent of the actual sample features and only depends on their topological structure. Finally, a classifier, like support vector machine or random forest, is used in a final step to detect outliers in this representation space (see Figure 1 for further details). We call our method Community-based Outlier Detector, COD.

Comparing to existing outliers detection methods, the proposed approach have the following two main contributions:

- It is based on a topological description of the structure of feature spaces based on the communities of a MKG.
- This description provides an outlierness representation space based on the intrinsic topological structure of outliers, which is independent of the original feature space.
- It is a *training cheap* method. As said above, the source dataset used to train the outlier classifier does not significantly bias the performance of the COD method. This advantage is useful when dealing with small datasets or high dimensionality data, since we can trained a classifier on a big labelled dataset and then use it as an off the shelf classifier.

In summary, we propose a cheap outlier detector that uses a standard classifier but is trained on meaning-full feature space.

3 COD: Outlier Detection based on a Topological Measure of Pattern Diversity

Our method is a local approach based on the communities of a graph encoding the structure of the feature space. Figure 1 sketches the main steps of our method for the two-dimensional single-view space shown in Figure 1.a. The dataset has 2 classes (black dots and red crosses), one class outlier (numbered 6) and one attribute outlier (numbered 10). The 3 main steps of the proposed method applied to the synthetic data are shown in Figure 1.b-d. First, we encode the local structure of samples using the graph representing their mutual k-nearest neighbor. The nodes of the graph are colored using each class color (red and black). Second, we use methods for dynamical analysis of social networks to compute graph communities from an initial set of communities. Figure 1.c shows the initial set of communities on the left and their extension on the right image. These extended communities define in the original feature space a set of neighborhoods of each sample. By definition of class and attribute outlier, isolated nodes not belonging to any community are attribute outliers, while class outliers should belong to communities with an heterogeneous distribution of labels. In order to characterize the latter, we define a local measure of abnormality from two probabilistic measures of each sample heterogeneity computed in its set of neighborhoods. These two measures define a function from the set of samples to the unit square that maps inliers and outliers to different corners of the square. A classifier, C, discriminating between inliers and outliers provides our measure of outlierness, Figure 1.d.

In the next sections, we give details about each of the main steps: graph construction, community detection and definition of the feature space for the classification of inliers and outliers.

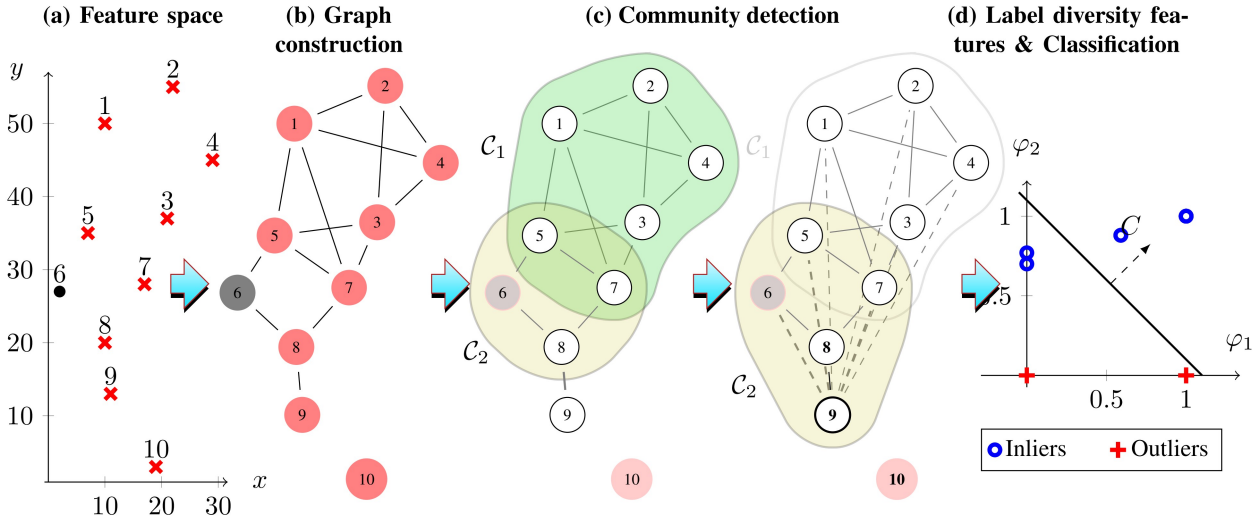

Figure 1: Overview of the method: (a) Two-dimensional feature space with 2 classes (black dots and red crosses). (b) Graph encoding mutual k-nearest neighbors with nodes colored in red and black according to its class. Nodes 6 and 10 correspond, respectively, to the class attribute and attribute outliers. (c) Detection of communities: initial communities, $\mathcal{C}_1, \mathcal{C}_2$, in left graph and their extension in the right graph. (d) Feature space and classifier margin, C, giving the final outlierness measure from the communities.

3.1 Graph Construction

The graph, $G = (D, A)$, is given by the adjacency matrix of the mutual k-nearest neighbor of the set of samples. Let $D := \{(\mathbf{v}^i, \ell_{\mathbf{v}^i}) \mid \mathbf{v}^i = (v_1^i, \ldots, v_n^i) \in \mathbb{R}^n, \ \ell_{\mathbf{v}^i} \in \{1, \ldots, n_l\}\}_{i=1}^N$ be a set of N labelled points in an n-dimensional feature space endowed with a distance, namely d. For any positive integer, k , let $\mathrm{kNN}(\mathbf{v}^i)$ denote the set of $\mathbf{v}^i$ k-nearest neighbors and $\mathrm{mkNN}(V^i)$ the set of $\mathbf{v}^i$ mutual k-nearest neighbors defined as:

$$\mathrm{mkNN}(\mathbf{v}^i) := \{\mathbf{v}^j \text{ such that } \mathbf{v}^j \in \mathrm{kNN}(\mathbf{v}^i) \text{ and } \mathbf{v}^i \in \mathrm{kNN}(\mathbf{v}^j)\} \tag{1}$$

Then, the adjacency matrix, $A = (a(\mathbf{v}^i, \mathbf{v}^j))_{ij} = (a_{ij})_{ij}$, codifying G is defined as:

$$a_{ij} = \begin{cases} \frac{1}{d(\mathbf{v}^i, \mathbf{v}^j)+1} & \text{if} \quad \mathbf{v}^j \in \mathrm{mkNN}(\mathbf{v}^i) \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

for $d(\mathbf{v}^i, \mathbf{v}^j)$ the distance between $\mathbf{v}^j$ and $\mathbf{v}^i$. We note that the graph edges, a_{ij}, are in $[0, 1]$, being close to 0 if mutual neighbors are far from each other, close to 1, otherwise. This way, the number of neighbors for every node, $\mathbf{v}^i$, is related to the sparseness of the point $\mathbf{v}^i$ in the feature space.

3.2 Community Detection

In order to alleviate the impact of the parameters involved in the computation of (2), communities are computed using criteria for dynamic computation of communities [Cazabet and et al, 2010, Kevin et al., 2011, Sekara et al., 2016] to extend an initial set of communities. The initial communities are given by Percolation clusters [Domb et al., 1980] which are defined as maximal unions of adjacent k-cliques. Percolation communities are prone to exclude many points that are not actual attribute outliers [Mielgo, 2017]. In order to add them to the set of initial communities, we extend them following a modification of the community detector proposed in Cazabet and et al [2010]. An isolated node, $\mathbf{w}^j$, is added to a initially detected community, $\mathcal{C}$, if it fulfills that:

$$\mathrm{CS}(\mathcal{C}, \mathbf{w}^j) \geq \delta \, \mathrm{IC}(\mathcal{C}) \tag{3}$$

for $\delta \in [0, 1]$ a tolerance parameter, $\mathrm{IC}(\mathcal{C})$ a measure of the community internal connectivity and $\mathrm{CS}(\mathcal{C}, \mathbf{w}^j)$ a measure of the connectivity between $\mathbf{w}^j$ and the community $\mathcal{C}$. Both measures are computed from a function of the degree of the community nodes as follows.

Let $\mathcal{S}$ be the set composed of all nodes that belong to any of the initially detected communities and $G^{\mathcal{S}}$ and $G^{\mathcal{C}}$ the subgraphs induced by $\mathcal{S}$ and $\mathcal{C}$, respectively. Then, for $\forall \mathbf{v}^i \in \mathcal{C}$ we can define the following function, $\rho_{\mathcal{C}}(\mathbf{v}^i)$, measuring its belongingness to the community:

$$\rho_{\mathcal{C}}(\mathbf{v}^i) := \frac{\deg^{G^{\mathcal{C}}}(\mathbf{v}^i)}{\deg^{G^{\mathcal{S}}}(\mathbf{v}^i)} \tag{4}$$

being $\deg$ the degree function of a node in a graph. The measure of the internal connectivity of $\mathcal{C}$ is defined from $\rho_{\mathcal{C}}(\mathbf{v}^i)$ as:

$$\mathrm{IC}(\mathcal{C}) := \sum_{\mathbf{v}_i \in \mathcal{C}} \rho_{\mathcal{C}}(\mathbf{v}_i) \tag{5}$$

The measure of the connectivity between $\mathbf{w}^j$ and $\mathcal{C}$ is also defined from $\rho_{\mathcal{C}}(\mathbf{v}^i)$ as:

$$\mathrm{CS}(\mathcal{C}, \mathbf{w}^j) := \sum_{\mathbf{v}^i \in \mathcal{C}} \frac{\rho_{\mathcal{C}}(\mathbf{v}^i)}{d(\mathbf{w}^j, \mathbf{v}^i) + 1} \tag{6}$$

since $\mathrm{CS}(\mathcal{C}, \mathbf{w}^j)$ is a weighted average of $\rho_{\mathcal{C}}(\mathbf{v}^i)$ with weights $\frac{1}{d(\mathbf{w}^j, \mathbf{v}^i)+1}$, we have that:

$$\mathrm{CS}(\mathcal{C}, \mathbf{w}^j) \geq \left(\max_{ji} \frac{1}{d(\mathbf{w}^j, \mathbf{v}^i) + 1} \right) \mathrm{IC}(\mathcal{C}) \tag{7}$$

By the above inequality, we could set $\delta = \max_{ji} \frac{1}{d(\mathbf{w}^j, \mathbf{v}^i)+1}$. However, such extreme value could aggregate some attribute outliers to the initial set of communities. In order to avoid such an artifact, we propose to define δ as a percentile of $\frac{1}{d(\mathbf{w}^j, \mathbf{v}^i)+1}$ probabilistic distribution.

3.3 Outlierness Feature Space

Given that communities define a set of neighbors in the feature space, nodes not belonging to any of the extended communities correspond to attribute outliers. Meanwhile, class outliers are expected to belong to communities with either high heterogeneity in nodes label or the majority of nodes with a label different from the class outlier label.

Under the above considerations, we define two quantities, φ_1, φ_2, measuring how homogeneous the labels of the communities a sample belongs to are. For each sample, the function φ_1 quantifies the heterogeneity in community labels, while φ_2 quantifies how many nodes in the community have a label different from the sample label. Both measures are based on the probabilistic distribution of the community node labels and are normalized in $[0, 1]$ in such a way that they define a function φ:

$$\begin{array}{rccc} \varphi: & D & \longrightarrow & [0,1] \times [0,1] \\ & (\mathbf{v}^i, \ell_{\mathbf{v}^i}) & \mapsto & (\varphi_1(\mathbf{v}^i, \ell_{\mathbf{v}^i}), \varphi_2(\mathbf{v}^i, \ell_{\mathbf{v}^i})) \end{array} \tag{8}$$

mapping inliers to $(1, 1)$, attribute outliers to $(0, 0)$ and class outliers to either $(1, 0)$ or $(0, 1)$ depending on whether they belong to one or more communities. In case $\mathbf{v}$ does not belong to any community, we have an attribute outlier and, thus, we set $\varphi(\mathbf{v}) = (0, 0)$. This way, φ defines a 2-dimensional feature space able to discriminate inliers and outliers. The probability of a classifier trained to discriminate between them provides our outlierness score and its output class our outlier detection.

The function φ_1 measuring the heterogeneity in community labels is computed from their entropy as follows. For each sample $\mathbf{v}$, let $\mathcal{S}_{\mathbf{v}}$ denote the set of communities containing $\mathbf{v}$ and $Ent_{\mathcal{C}}$ the entropy of the labels of the nodes in a community $\mathcal{C} \in \mathcal{S}_{\mathbf{v}}$ defined as:

$$Ent_{\mathcal{C}} = -\sum_{i=1}^{n_l} p_i^{\mathcal{C}} \log(p_i^{\mathcal{C}}) \tag{9}$$

for $p_i^{\mathcal{C}}$ the probability in $\mathcal{C}$ of the i-th label. This probability is approximated by the proportion of $\mathcal{C}$ nodes that are labelled i:

$$p_i^{\mathcal{C}} := \frac{|\{\mathbf{w} \in \mathcal{C} \quad | \quad \ell_{\mathbf{w}} = i\}|}{|\{\mathbf{w} \in \mathcal{C}\}|} \tag{10}$$

for $|\cdot|$ denoting the cardinality of a set. Then, φ_1 is defined as:

$$\varphi_1(\mathbf{v}, \ell_{\mathbf{v}}) := \varphi_1(\mathbf{v}) := \frac{1}{|\mathcal{S}_{\mathbf{v}}|} \sum_{\mathcal{C} \in \mathcal{S}_{\mathbf{v}}} \pi(Ent_{\mathcal{C}} \leq T) \tag{11}$$

being $T \in [0, 1/n_\ell]$ a tolerance threshold on the maximum entropy allowed in community labels and $\pi(a) = 1$ if a holds. The score $\varphi_1(\mathbf{v}) \in [0, 1]$ has extreme values $\varphi_1(\mathbf{v}) = 0$ in case all communities in $\mathcal{S}_\mathbf{v}$ have heterogeneous labels and $\varphi_1(\mathbf{v}) = 1$ in case the label within each community is the same for all nodes in the community. We note that, in this last case, by definition of $Ent_\mathcal{C}$, the label must be the same for all communities in $\mathcal{S}_\mathbf{v}$.

The function φ_2 measuring how many nodes in the community have a label equal to the sample label, $\mathbf{v}$, is computed from the probability of $\ell_\mathbf{v}$, as follows. For each $\mathcal{C} \in \mathcal{S}_\mathbf{v}$, let $p^\mathcal{C}_{\ell_\mathbf{v}}$ denote the probability of $\ell_\mathbf{v}$ in $\mathcal{C}$ excluding $\mathbf{v}$:

$$p^\mathcal{C}_{\ell_\mathbf{v}} = \frac{|\{\mathbf{w} \in \mathcal{C} \setminus \mathbf{v} \quad | \quad \ell_\mathbf{w} = \ell_\mathbf{v}\}|}{|\{\mathbf{w} \in \mathcal{C} \setminus \mathbf{v}\}|} \tag{12}$$

Then, φ_2 is defined as:

$$\varphi_2(\mathbf{v}, \ell_\mathbf{v}) := \frac{1}{|\mathcal{S}_\mathbf{v}|} \sum_{\mathcal{C} \in \mathcal{S}_\mathbf{v}} p^\mathcal{C}_{\ell_\mathbf{v}} \tag{13}$$

The measure $\varphi_2(\mathbf{v}, \ell_\mathbf{v}) \in [0, 1]$ and has extreme values $\varphi_2(\mathbf{v}, \ell_\mathbf{v}) = 0$ in case no community in $\mathcal{S}_\mathbf{v}$ is consistent with $\mathbf{v}$ label, $\varphi_2(\mathbf{v}, \ell_\mathbf{v}) = 1$ in case all communities have nodes with labels equal to $\ell_\mathbf{v}$.

Finally, in order to increase the separability across the different types of outliers, we transform φ_1, φ_2 to logarithmic scale by applying the function:

$$f(x) = \frac{1}{1 - \log x} \tag{14}$$

to $x = \varphi_i$, $i = 1, 2$.

In the multi-view case, we model the space as a Cartesian product and, thus, compute a mutual k-nearest neighbor graph for each view. Communities are computed independently for each of these graphs and so are the probabilistic measures. These view-dependent measures are aggregated to define the function that maps samples to the unit square. We compute the two measures for each view and aggregate them to define φ. If φ^j_1, φ^j_2 denote the measures computed for the j-th view, then their minimum values across views defines the function φ as:

$$\begin{array}{rccl} \varphi: & D & \longrightarrow & [0,1] \times [0,1] \\ & (\mathbf{v}^i, \ell_{\mathbf{v}^i}) & \mapsto & (\varphi_1(\mathbf{v}^i, \ell_{\mathbf{v}^i}), \varphi_2(\mathbf{v}^i, \ell_{\mathbf{v}^i})) := (\min_j \varphi^j_1(\mathbf{v}^i, \ell_{\mathbf{v}^i}), \min_j \varphi^j_2(\mathbf{v}^i, \ell_{\mathbf{v}^i})) \end{array} \tag{15}$$

4 Experiments

4.1 Experimental Set-up

The performance of the proposed method has been assessed in UCI [2] datasets altered to have different % of attribute and class outliers in, both, single and multi-view settings. Table 1 reports the main characteristics of the UCI datasets selected for these experiments. The single-view problem was defined directly using UCI data sets features. Following Gao et al. [2013], the multi-view case was defined by splitting UCI features into disjoint sets, each of them defining one view. We considered 2 and 3 views for all sets, with the exception of Iris, which dimensionality only allows splitting features in 2-views.

Datasets	Dimension	Num of Samples	Num of Classes
Iris	4	150	3
BCW	10	699	2
Ionosphere	34	351	2
Letter recognition	16	20,000	26

Table 1: Selected UCI Datasets

We have followed the experimental settings described in Zhao and et al [2018]. In particular, we have considered 3 combinations of percentages in attribute and class outliers: 1) 8% class outlier + 2% attribute outlier, labelled 8-2, 2) 5% class outlier + 5% attribute outlier, labelled 5-2 and 3) 2% class outlier + 8% attribute outlier, labelled 2-8. To simulate attribute outliers, the features of the selected samples were changed by random numbers following a distribution with the highest probability outside the range of the values expected for the original data. In the multi-view case, features were altered in each view. To simulate class outliers, we swapped the labels of points randomly selected from random pairs of classes. In the multi-view case, classes were swapped in views randomly selected.

[2] https://archive.ics.uci.edu/ml/datasets.php

For each outlier configuration, we repeated the experiment 50 times for statistical analysis of results. Our proposal has been compared to state-of-art single and multi view methods methods in terms of Area Under the ROC Curve (AUC). Single-view methods include 4 local and 3 global approaches. The local methods are LOF [Breunig et al., 2000], LOCI [Papadimitriou et al., 2003], KNN [Ramaswamy et al., 2000] and IF [Liu et al., 2012], while the global ones are GMM [Yang et al., 2009], the graph-based APS [Wang et al., 2019] and the deep-learning one SO-GAAL [Liu et al., 2018] model. The multi-view methods are the best performers reported in [Zhao and et al, 2018], DMOD [Zhao and et al, 2018], AP [Alvarez et al., 2013] and MLRA [Li et al., 2015], and the pioneering multi view approach HOAD [Gao et al., 2013].

Our method was computed using the following parameters. The mkNN graph was computed setting $k = 40$. For the extension of the initial communities, the tolerance δ in (3) was computed for the 75% percentile.

Finally, for the computation of the outlierness measure, we use a SVM classifier trained on a sub-set of the MirFlickr dataset [Huiskes and Lew, 2008] altered to have the different types of outliers. To trained it, we use the second last layer of a pre-trained VGG network as feature vector on images with a single annotation. Then, we applied a K-means to identify those annotations that does not highly overlap between them and, to further reduce the size of samples, we select only those samples with minimal overlapping. Then, we follow Zhao and et al [2018] to generate class and attribute outliers.

4.2 Results

Table 2 reports the AUC (average $\pm$ standard deviation) for the results obtained for the proposed COD, LOF, LOCI, IF, GMM, SO-GAAL and APS on Iris, Breast Cancer Winconsin (BCW), Ionosphere and Letter Recognition with different settings [Wang et al., 2019]. Following [Zhao and et al, 2018], the first and second best performers for each configuration are marked in red and blue, respectively. The analysis of the ranges lead to the following observations.

According to the number of times methods are top 2, best performers are COD ($16/24 = 67\%$ times), KNN ($14/24 = 58\%$ times), IF ($4/24 = 17\%$ times) and GMM ($7/24 = 30\%$ times). The remaining methods achieve top ranges in less than 25% of the cases, being the deep-learning method SO-GAAL the worst performer with 0 top ranges. The ranking of top ranges indicates that local methods perform better than global ones, as 3 best ranked methods are local approaches. It is worth noticing that our COD is the only method that has top ranges for all datasets and outlier configurations. Meanwhile, the other top ranked methods achieve their best performance only in some of the datasets. In particular, LOCI did not converge for the Letter dataset.

If we compare ranges between configurations with same quantity of attribute outliers but different quantity of class outliers (2-8 against 0-8, 5-5 against 0-5 and 8-2 against 0-2) we have that, in general, methods are better detecting attribute outliers. In particular, the configuration that has worst ranges is the one with higher number of class-outliers (8-2). Among all methods, our COD is the one that has the lowest drop in performance in the presence of class-outliers.

Table 3 reports the AUC (average $\pm$ standard deviation) for the results obtained for the proposed COD, HOAD, AP, MLRA and DMOD on Iris, Breast Cancer Winconsin (BCW), Ionosphere and Letter Recognition with different settings. As before, the best two performers for each configuration are marked in red and blue. The analysis of the ranges lead to the following observations.

Ranking as before according to the number of top ranges, the best performers by large are COD ($21/21 = 100\%$ times) and DMOD ($17/21 = 81\%$ times). The proposed COD is the best performer for all cases, but three cases (2-View Ionosphere with outlier configuration 2-8 and 3-View Letter with outlier configuration 8-2) that it is the second best. The HOAD method is the worst performer with 0 times having top ranges.

All methods, excluding the proposed COD, perform better in the 2-view case. Our COD is the only method that has similar (top) ranges for, both, 2 and 3 view configurations. Regarding outlier configurations, there is not a clear trend across datasets. For Iris and Breast performance increases with the number of class outliers, while it decreases for Ionosphere and Letter datasets. Although this behaviour holds for both views, the decrease rate seems to be a bit higher for the 3-view case. Therefore, we attribute it to, both, the separability of the original dataset, as well as, the partition of features to simulate the multi-view configuration, which might have selected features having the least discriminating power.

4.3 Discussion

The method has been tested on 4 UCI datasets altered to have different configurations of class and attribute outliers in single and multi view spaces. Datasets have been selected to be representative of the main configurations of classification feature spaces. In particular, they include datasets presenting the most common artifacts dropping performance

Table 2: AUC values (mean ± standard deviation) for the Single View Case

DataSet	Method	2-8	5-5	8-2	0-8	0-5	0-2
Iris	LOF	0.973 ± 0.020	0.958 ± 0.024	0.949 ± 0.039	0.992 ± 0.003	0.978 ± 0.003	0.958 ± 0.024
	LOCI	0.962 ± 0.022	0.888 ± 0.052	0.728 ± 0.058	0.971 ± 0.007	0.966 ± 0.011	0.962 ± 0.006
	KNN	0.974 ± 0.019	0.936 ± 0.042	0.866 ± 0.061	0.990 ± 0.002	0.982 ± 0.004	0.970 ± 0.002
	IF	0.905 ± 0.025	0.855 ± 0.029	0.814 ± 0.036	0.987 ± 0.009	0.975 ± 0.001	0.959 ± 0.000
	APS	0.882 ± 0.047	0.891 ± 0.046	0.882 ± 0.047	0.862 ± 0.07	0.854 ± 0.088	0.910 ± 0.120
	GMM	0.484 ± 0.004	0.484 ± 0.004	0.485 ± 0.012	0.484 ± 0.003	0.484 ± 0.003	0.484 ± 0.002
	SO-GAAL	0.614 ± 0.019	0.605 ± 0.096	0.559 ± 0.087	0.663 ± 0.090	0.631 ± 0.090	0.602 ± 0.134
	COD	0.976 ± 0.045	0.980 ± 0.029	0.979 ± 0.020	0.971 ± 0.052	0.980 ± 0.041	0.989 ± 0.007
BCW	LOF	0.545 ± 0.092	0.528 ± 0.071	0.509 ± 0.032	0.513 ± 0.071	0.525 ± 0.088	0.511 ± 0.044
	LOCI	0.882 ± 0.008	0.735 ± 0.010	0.593 ± 0.013	0.981 ± 0.005	0.972 ± 0.004	0.963 ± 0.002
	KNN	0.889 ± 0.005	0.739 ± 0.008	0.592 ± 0.012	0.991 ± 0.002	0.981 ± 0.002	0.966 ± 0.001
	IF	0.885 ± 0.005	0.733 ± 0.009	0.588 ± 0.013	0.987 ± 0.001	0.973 ± 0.001	0.959 ± 0.000
	APS	0.794 ± 0.029	0.674 ± 0.029	0.580 ± 0.029	0.882 ± 0.033	0.863 ± 0.040	0.895 ± 0.048
	GMM	0.371 ± 0.008	0.472 ± 0.015	0.574 ± 0.015	0.299 ± 0.003	0.299 ± 0.002	0.298 ± 0.002
	SO-GAAL	0.693 ± 0.040	0.597 ± 0.032	0.533 ± 0.023	0.733 ± 0.056	0.709 ± 0.063	0.672 ± 0.084
	COD	0.879 ± 0.039	0.830 ± 0.035	0.747 ± 0.033	0.936 ± 0.042	0.954 ± 0.037	0.978 ± 0.002
Ionosphere	LOF	0.679 ± 0.034	0.647 ± 0.030	0.572 ± 0.019	0.736 ± 0.041	0.797 ± 0.047	0.892 ± 0.059
	LOCI	0.823 ± 0.025	0.722 ± 0.016	0.585 ± 0.020	0.922 ± 0.027	0.956 ± 0.023	0.954 ± 0.018
	KNN	0.874 ± 0.009	0.734 ± 0.011	0.583 ± 0.017	0.987 ± 0.005	0.976 ± 0.002	0.960 ± 0.001
	IF	0.871 ± 0.011	0.734 ± 0.012	0.584 ± 0.019	0.987 ± 0.006	0.974 ± 0.001	0.958 ± 0.010
	APS	0.881 ± 0.025	0.742 ± 0.026	0.579 ± 0.034	0.995 ± 0.002	0.996 ± 0.004	0.996 ± 0.005
	GMM	0.590 ± 0.012	0.697 ± 0.017	0.814 ± 0.019	0.497 ± 0.001	0.497 ± 0.001	0.497 ± 0.001
	SO-GAAL	0.515 ± 0.035	0.503 ± 0.029	0.511 ± 0.025	0.514 ± 0.048	0.509 ± 0.043	0.503 ± 0.055
	COD	0.845 ± 0.023	0.762 ± 0.032	0.686 ± 0.033	0.902 ± 0.007	0.891 ± 0.006	0.890 ± 0.003
Letter Rec.	LOF	0.533 ± 0.009	0.522 ± 0.007	0.508 ± 0.007	0.539 ± 0.011	0.527 ± 0.013	0.504 ± 0.015
	LOCI	±	±	±	±	±	±
	KNN	0.513 ± 0.010	0.500 ± 0.008	0.486 ± 0.006	0.516 ± 0.012	0.510 ± 0.012	0.521 ± 0.014
	IF	0.516 ± 0.013	0.503 ± 0.010	0.494 ± 0.009	0.514 ± 0.013	0.509 ± 0.016	0.534 ± 0.019
	APS	0.492 ± 0.014	0.483 ± 0.010	0.473 ± 0.011	0.503 ± 0.015	0.640 ± 0.015	0.751 ± 0.014
	GMM	0.894 ± 0.010	0.893 ± 0.009	0.901 ± 0.008	0.916 ± 0.011	0.939 ± 0.011	0.968 ± 0.012
	SO-GAAL	0.495 ± 0.021	0.492 ± 0.020	0.487 ± 0.018	0.485 ± 0.024	0.493 ± 0.035	0.492 ± 0.043
	Our COD	0.844 ± 0.013	0.872 ± 0.009	0.906 ± 0.007	0.821 ± 0.017	0.809 ± 0.015	0.795 ± 0.013

Table 3: AUC values (mean ± standard deviation) for the Multi View Case

		2-View Case			3-View Case		
DataSet	Method	2-8	5-5	8-2	2-8	5-5	8-2
Iris	HOAD	0.167 ± 0.057	0.309 ± 0.063	0.430 ± 0.055	−−	−−	−−
	AP	0.326 ± 0.027	0.630 ± 0.021	0.840 ±0.021	−−	−−	−−
	MLRA	0.856 ± 0.063	0.828 ± 0.080	0.826 ± 0.089	−−	−−	−−
	DMOD	0.909 ± 0.044	0.831 ± 0.038	0.799 ± 0.068	−−	−−	−−
	COD	0.975 ± 0.024	0.971 ± 0.023	0.970 ± 0.021	−−	−−	−−
BCW	HOAD	0.555 ± 0.072	0.586 ± 0.061	0.634 ± 0.046	0.538 ±0.027	0.597 ± 0.038	0.643 ± 0.008
	AP	0.293 ± 0.012	0.532 ± 0.024	0.693 ± 0.023	0.190 ± 0.016	0.388 ± 0.012	0.593 ± 0.046
	MLRA	0.745 ± 0.056	0.715 ± 0.022	0.688 ± 0.028	0.614 ± 0.057	0.596 ± 0.032	0.599 ± 0.029
	DMOD	0.824 ± 0.022	0.752 ± 0.019	0.692 ± 0.036	0.657 ± 0.017	0.720 ± 0.013	0.799 ± 0.016
	COD	0.890 ± 0.027	0.935 ± 0.019	0.947 ± 0.013	0.838 ± 0.022	0.897 ± 0.020	0.910 ± 0.014
Ionosphere	HOAD	0.446 ± 0.074	0.442 ± 0.051	0.448 ± 0.041	0.489 ± 0.079	0.477 ± 0.072	0.444 ± 0.065
	AP	0.623 ± 0.033	0.761 ± 0.025	0.822 ± 0.030	0.511 ± 0.027	0.659 ± 0.043	0.758 ± 0.035
	MLRA	0.645± 0.084	0.669 ± 0.028	0.776± 0.037	0.645 ± 0.040	0.663 ± 0.048	0.700 ± 0.045
	DMOD	0.877 ± 0.032	0.801 ± 0.042	0.774 ± 0.049	0.818 ± 0.018	0.787 ± 0.039	0.784 ± 0.037
	COD	0.841 ± 0.024	0.811 ± 0.024	0.780 ± 0.029	0.854 ± 0.019	0.827 ± 0.025	0.791 ± 0.036
Letter Rec.	HOAD	0.536 ± 0.046	0.663 ± 0.057	0.569 ± 0.049	0.193 ± 0.022	0.488 ± 0.111	0.563 ± 0.081
	AP	0.372 ± 0.057	0.550 ± 0.043	0.640 ± 0.051	0.189 ± 0.039	0.340 ± 0.037	0.570 ± 0.63
	MLRA	0.883 ± 0.024	0.817± 0.051	0.786 ± 0.065	0.841 ± 0.055	0.716 ± 0.044	0.640 ± 0.081
	DMOD	0.912 ± 0.029	0.846 ± 0.022	0.762 ± 0.025	0.916 ± 0.031	0.815 ± 0.038	0.664 ± 0.037
	COD	0.926 ± 0.009	0.904 ± 0.011	0.877 ± 0.011	0.843 ± 0.014	0.816 ± 0.016	0.774 ± 0.017

of methods, like small sample size (Ionosphere), large dimensionality (Ionosphere) and large number of classes with some of them being minority groups (Letter Recognition). The method has been compared to several state of the art methods for detection of outliers in single and multi view settings. The analysis of results lead to the following conclusions.

In, both, single and multi view settings, local methods perform better than global ones. In fact, the worst performers in both settings are the global methods SO-GAAL (single view) and HOAD (multi view). Surprisingly, the deep learning approach SO-GAAL is the worst performer in the single view setting without any range in the top two. Although deep learning approaches achieve excellent results in classification problems, they require huge amounts of data to model population distribution. Being based in big data, their capability to model rare cases (like outliers) might be, as our experiments indicate, limited compared to other approaches.

Among all local approaches, the proposed COD outperforms existing methods in, both, single and multi view settings, regardless of the outlier configuration. Unlike most methods, it performs equally well in small size and high dimensionality datasets. It is worth noticing that COD has been applied with the same parameter configuration to all datasets in, both, training and test. Regarding COD training, we used a completely different repository (MirFLick) and no learning transfer was applied. This is a main advantage compared to existing approaches that require fine tuning of parameters and a re-training for new datasets.

Although the choice of parameters (the k for the computation of the mutual k-nearest neighbor graph, in particular) is not critical, the performance of COD significantly increases for low dimensional balanced datasets like Iris. We think that the combination of all information obtained from graphs computed sampling the range of possible values of k could improve the performance of COD in high dimensional unbalanced datasets.

5 Conclusion

This paper presents a local approach for outlier detection based on a outlierness representation space that codifies samples label diversity using a topological description of samples structure in feature space. This representation space is normalized and describes the intrinsic structural properties of feature spaces independently of the particular dataset. It follows that COD needs only to be trained once and can be applied to any data set without any transfer learning.

The reported experiments show that COD outperforms existing methods in, both, single and multi view settings, regardless of the outlier configuration. Although the choice of parameters is not critical, the performance of COD significantly increases for low dimensional balanced datasets like Iris. We think that the description of the feature space topology could be more elaborated and include higher order aspects (like closed paths and cycles). Algebraic topology is a unique mathematical discipline that provides with very low-level descriptions of complex manifolds in high dimensional spaces without the need of either exhaustive training or access to big data. In particular, the groups of persistent homology provide an algebraic description of the structure of point clouds. In order to define intrinsic properties, these groups are computed using a set of neighbourhoods (called filtration) which increases according to a parameter. Parameter values are sampled to obtain such set of neighbourhoods and those properties that are more stable across the filtration constitute the persistent homology. Our current efforts focus on the improvement of COD incorporating the computation of persistent homology to better define space topological properties.

Acknowledgment

This work was supported by Spanish projects RTI2018-095645-B-C21 and RTI2018-095209-B-C21, Generalitat de Catalunya, 2017-SGR-1624 and 2017-SGR-1783 and CERCA-Programme and the ATTRACT project funded by the EC under Grant Agreement 777222. Debora Gil is supported by Serra Hunter Fellow. The Titan X Pascal and Titan V used for this research was donated by the NVIDIA Corporation. DGil is a Serra Hunter Fellow.

References

Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.

H Zhao and et al. Consensus regularized multi-view outlier detection. *IEEE Trans Imag Proc*, 27(1), 2018.

X Yang, L. Jan, and L. D. Pokrajac. Outlier detection with globally optimal exemplar-based gmm. In *SIAM International Conference on Data Mining*, 2009.

Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. *CoRR*, abs/1901.03407, 2019. URL `http://arxiv.org/abs/1901.03407`.

Raghavendra Chalapathy, Aditya Krishna Menon, and Sanjay Chawla. Anomaly detection using one-class neural networks. 02 2018.

Lukas Ruff, Nico Görnitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Robert Vandermeulen, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *International Conference on Machine Learning*, pages 4390–4399, 2018.

Y. Liu, Z. Li, C. Zhou, Y. Jiang, J. Sun, M. Wang, and X. He. Generative adversarial active learning for unsupervised outlier detection. 2018.

S. Ramaswamy, R. Rastogi, and K. Shim. Efficient algorithms for mining outliers from large data sets. *SIGMOD Rec.*, 29(2):427–438, May 2000. ISSN 0163-5808. doi: 10.1145/335191.335437. URL http://doi.acm.org/10.1145/335191.335437.

F. Angiulli and C. Pizzuti. Fast outlier detection in high dimensional spaces. In T. Elomaa, H. Mannila, and H. Toivonen, editors, *Principles of Data Mining and Knowledge Discovery*, pages 15–27, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg. ISBN 978-3-540-45681-0.

Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. Lof: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, SIGMOD '00, pages 93–104, New York, NY, USA, 2000. ACM. ISBN 1-58113-217-4. doi: 10.1145/342009.335388. URL http://doi.acm.org/10.1145/342009.335388.

S. Papadimitriou, H. Kitagawa, P. B. Gibbons, and C. Faloutsos. Loci: fast outlier detection using the local correlation integral. In *Proceedings 19th International Conference on Data Engineering (Cat. No.03CH37405)*, pages 315–326, March 2003. doi: 10.1109/ICDE.2003.1260802.

F. T. Liu, K. M. Ting, and Z.-H. Zhou. Isolation-based anomaly detection. *ACM Trans. Knowl. Discov. Data*, 6(1):3:1–3:39, March 2012. ISSN 1556-4681. doi: 10.1145/2133360.2133363. URL http://doi.acm.org/10.1145/2133360.2133363.

M. R. Brito, E. L. Chávez, A. J. Quiroz, and J. E. Yukich. Connectivity of the mutual k-nearest-neighbor graph in clustering and outlier detection. *Statistics & Probability Letters*, 35(1):33–42, Aug 1997. URL http://www.sciencedirect.com/science/article/B6V1D-3SP2B5D-K/1/d7f9779042ec204c57a711d8f3ba7fd9.

M.s Maier, M.s Hein, and U. von Luxburg. Cluster identification in nearest-neighbor graphs. In M. Hutter, R. A. Servedio, and E. Takimoto, editors, *Algorithmic Learning Theory*, pages 196–210, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. ISBN 978-3-540-75225-7.

J. Ning, L. Chen, C. Zhou, and Y. Wen. Parameter k search strategy in outlier detection. *Pattern Recognition Letters*, 112:56 – 62, 2018. ISSN 0167-8655. doi: https://doi.org/10.1016/j.patrec.2018.06.007. URL http://www.sciencedirect.com/science/article/pii/S0167865518302307.

C. Wang, Z. Liu, H. Gao, and Y. Fu. Applying anomaly pattern score for outlier detection. *IEEE Access*, 7:16008–16020, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2895094.

James Munkres. *General Topology*. Prentice-Hall, 2000.

D. Gil, O. Ramos Terrades, E. Minchole, C. Sanchez, N. Cubero de Frutos, M. Diez-Ferrer, and A. Maria Ortiz, R. ans Rosell. Classification of confocal endomicroscopy patterns for diagnosis of lung cancer. In *CLIP-MICCAI*, pages 151–159. Lecture Notes in Computer Science, 2017.

J Mielgo. *Analysis of Community Detection Algorithms for Image Annotation*, 2017.

L. Akoglu, M. McGlohon, and C. Faloutsos. oddball: Spotting anomalies in weighted graphs. In Mohammed J. Zaki, Jeffrey Xu Yu, B. Ravindran, and Vikram Pudi, editors, *Advances in Knowledge Discovery and Data Mining*, pages 410–421, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. ISBN 978-3-642-13672-6.

A. Elliott, M. Cucuringu, M. Martinez Luaces, P. Reidy, and G. Reinert. Anomaly detection in networks with application to financial transaction networks. *CoRR*, abs/1901.00402, 2019. URL http://arxiv.org/abs/1901.00402.

AE Wegner, Ospina-Forero, RE Gaunt, C Deane, and GD Reinert. Identifying networks with common organizational principles. *Journal of Complex Networks*, 6(6):887–913, 2018.

C. C. Noble and D. J. Cook. Graph-based anomaly detection. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 631–636, New York, NY, USA, 2003. ACM. ISBN 1-58113-737-0. doi: 10.1145/956750.956831. URL http://doi.acm.org/10.1145/956750.956831.

W. Eberle and L. Holder. Discovering structural anomalies in graph-based data. In *Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007)*, pages 393–398, Oct 2007. doi: 10.1109/ICDMW.2007.91.

J. Gao, N. Du, W. Fan, D. Turaga, S. Parthasarathy, and J. Han. *Graph Embedding for Pattern Analysis*, chapter A multi-graph spectral framework for mining multi-source anomalies, pages 205–228. Springer, 2013.

H. Zhao and Y. Fu. Dual-regularized multi-view outlier detection. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 4077–4083. AAAI Press, 2015. ISBN 978-1-57735-738-4. URL http://dl.acm.org/citation.cfm?id=2832747.2832817.

R Cazabet and et al. Detection of overlapping communities in dynamical social networks. *Social Comp*, 2010.

S Kevin, K Mark, and H Alfred. Tracking communities in dynamic social networks. In *Social Computing, Behavioral-Cultural Modeling and Prediction - 4th International Conference*, 2011.

V Sekara, A Stopczynski, and S Lehmann. Fundamental structures of dynamic social networks. *PNAS*, 113 (36): 9977–9982, 2016.

C Domb, E Stoll, and T Schneider. Percolation clusters. *Contemporary Physics*, 21(6):577–592, 1980.

C Wang, Z Liu, H Gao, and Y Fu. Applying anomaly pattern score for outlier detection. *IEEE Access*, 17:16008–16021, 2019.

A. M. Alvarez, M. Yamada, A. Kimura, and T. Iwata. Clustering-based in multi-view data. In *ACM Int. Conf. Conf. Inf. Knowl. Manage*, 2013.

S. Li, M. Shao, and Y. Fu. Multi-view low-rank analysis for outlier detection. In *SIAM Int. Conf. Data Mining*, 2015.

Mark J. Huiskes and Michael S. Lew. The mir flickr retrieval evaluation. In *MIR '08: Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval*. ACM, 2008.

References

Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):15, 2009.

H Zhao and et al. Consensus regularized multi-view outlier detection. *IEEE Trans Imag Proc*, 27(1), 2018.

X Yang, L. Jan, and L. D. Pokrajac. Outlier detection with globally optimal exemplar-based gmm. In *SIAM International Conference on Data Mining*, 2009.

Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. *CoRR*, abs/1901.03407, 2019. URL http://arxiv.org/abs/1901.03407.

Raghavendra Chalapathy, Aditya Krishna Menon, and Sanjay Chawla. Anomaly detection using one-class neural networks. 02 2018.

Lukas Ruff, Nico Görnitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Robert Vandermeulen, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *International Conference on Machine Learning*, pages 4390–4399, 2018.

Y. Liu, Z. Li, C. Zhou, Y. Jiang, J. Sun, M. Wang, and X. He. Generative adversarial active learning for unsupervised outlier detection. 2018.

S. Ramaswamy, R. Rastogi, and K. Shim. Efficient algorithms for mining outliers from large data sets. *SIGMOD Rec.*, 29(2):427–438, May 2000. ISSN 0163-5808. doi: 10.1145/335191.335437. URL http://doi.acm.org/10.1145/335191.335437.

F. Angiulli and C. Pizzuti. Fast outlier detection in high dimensional spaces. In T. Elomaa, H. Mannila, and H. Toivonen, editors, *Principles of Data Mining and Knowledge Discovery*, pages 15–27, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg. ISBN 978-3-540-45681-0.

Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. Lof: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, SIGMOD '00, pages 93–104, New York, NY, USA, 2000. ACM. ISBN 1-58113-217-4. doi: 10.1145/342009.335388. URL http://doi.acm.org/10.1145/342009.335388.

S. Papadimitriou, H. Kitagawa, P. B. Gibbons, and C. Faloutsos. Loci: fast outlier detection using the local correlation integral. In *Proceedings 19th International Conference on Data Engineering (Cat. No.03CH37405)*, pages 315–326, March 2003. doi: 10.1109/ICDE.2003.1260802.

F. T. Liu, K. M. Ting, and Z.-H. Zhou. Isolation-based anomaly detection. *ACM Trans. Knowl. Discov. Data*, 6(1):3:1–3:39, March 2012. ISSN 1556-4681. doi: 10.1145/2133360.2133363. URL http://doi.acm.org/10.1145/2133360.2133363.

M. R. Brito, E. L. Chávez, A. J. Quiroz, and J. E. Yukich. Connectivity of the mutual k-nearest-neighbor graph in clustering and outlier detection. *Statistics & Probability Letters*, 35(1):33–42, Aug 1997. URL http://www.sciencedirect.com/science/article/B6V1D-3SP2B5D-K/1/d7f9779042ec204c57a711d8f3ba7fd9.

M.s Maier, M.s Hein, and U. von Luxburg. Cluster identification in nearest-neighbor graphs. In M. Hutter, R. A. Servedio, and E. Takimoto, editors, *Algorithmic Learning Theory*, pages 196–210, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. ISBN 978-3-540-75225-7.

J. Ning, L. Chen, C. Zhou, and Y. Wen. Parameter k search strategy in outlier detection. *Pattern Recognition Letters*, 112:56 – 62, 2018. ISSN 0167-8655. doi: https://doi.org/10.1016/j.patrec.2018.06.007. URL http://www.sciencedirect.com/science/article/pii/S0167865518302307.

C. Wang, Z. Liu, H. Gao, and Y. Fu. Applying anomaly pattern score for outlier detection. *IEEE Access*, 7:16008–16020, 2019. ISSN 2169-3536. doi: 10.1109/ACCESS.2019.2895094.

James Munkres. *General Topology*. Prentice-Hall, 2000.

D. Gil, O. Ramos Terrades, E. Minchole, C. Sanchez, N. Cubero de Frutos, M. Diez-Ferrer, and A. Maria Ortiz, R. ans Rosell. Classification of confocal endomicroscopy patterns for diagnosis of lung cancer. In *CLIP-MICCAI*, pages 151–159. Lecture Notes in Computer Science, 2017.

J Mielgo. *Analysis of Community Detection Algorithms for Image Annotation*, 2017.

L. Akoglu, M. McGlohon, and C. Faloutsos. oddball: Spotting anomalies in weighted graphs. In Mohammed J. Zaki, Jeffrey Xu Yu, B. Ravindran, and Vikram Pudi, editors, *Advances in Knowledge Discovery and Data Mining*, pages 410–421, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. ISBN 978-3-642-13672-6.

A. Elliott, M. Cucuringu, M. Martinez Luaces, P. Reidy, and G. Reinert. Anomaly detection in networks with application to financial transaction networks. *CoRR*, abs/1901.00402, 2019. URL http://arxiv.org/abs/1901.00402.

AE Wegner, Ospina-Forero, RE Gaunt, C Deane, and GD Reinert. Identifying networks with common organizational principles. *Journal of Complex Networks*, 6(6):887–913, 2018.

C. C. Noble and D. J. Cook. Graph-based anomaly detection. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 631–636, New York, NY, USA, 2003. ACM. ISBN 1-58113-737-0. doi: 10.1145/956750.956831. URL http://doi.acm.org/10.1145/956750.956831.

W. Eberle and L. Holder. Discovering structural anomalies in graph-based data. In *Seventh IEEE International Conference on Data Mining Workshops (ICDMW 2007)*, pages 393–398, Oct 2007. doi: 10.1109/ICDMW.2007.91.

J. Gao, N. Du, W. Fan, D. Turaga, S. Parthasarathy, and J. Han. *Graph Embedding for Pattern Analysis*, chapter A multi-graph spectral framework for mining multi-source anomalies, pages 205–228. Springer, 2013.

H. Zhao and Y. Fu. Dual-regularized multi-view outlier detection. In *Proceedings of the 24th International Conference on Artificial Intelligence*, IJCAI'15, pages 4077–4083. AAAI Press, 2015. ISBN 978-1-57735-738-4. URL http://dl.acm.org/citation.cfm?id=2832747.2832817.

R Cazabet and et al. Detection of overlapping communities in dynamical social networks. *Social Comp*, 2010.

S Kevin, K Mark, and H Alfred. Tracking communities in dynamic social networks. In *Social Computing, Behavioral-Cultural Modeling and Prediction - 4th International Conference*, 2011.

V Sekara, A Stopczynski, and S Lehmann. Fundamental structures of dynamic social networks. *PNAS*, 113 (36): 9977–9982, 2016.

C Domb, E Stoll, and T Schneider. Percolation clusters. *Contemporary Physics*, 21(6):577–592, 1980.

C Wang, Z Liu, H Gao, and Y Fu. Applying anomaly pattern score for outlier detection. *IEEE Access*, 17:16008–16021, 2019.

A. M. Alvarez, M. Yamada, A. Kimura, and T. Iwata. Clustering-based in multi-view data. In *ACM Int. Conf. Conf. Inf. Knowl. Manage*, 2013.

S. Li, M. Shao, and Y. Fu. Multi-view low-rank analysis for outlier detection. In *SIAM Int. Conf. Data Mining*, 2015.

Mark J. Huiskes and Michael S. Lew. The mir flickr retrieval evaluation. In *MIR '08: Proceedings of the 2008 ACM International Conference on Multimedia Information Retrieval*. ACM, 2008.