**HUBS: WHERE AND WHY**

- Hubness occurs in all intrinsically high-dimensional data, including images and text.
- Hubs emerge as centers of influence within the data and dominate the query result sets.
- They are less relevant to the queries and frequently violate the semantics of the search.
- Hubness is related to distance concentration and other aspects of the dimensionality curse. Different feature representations and different metrics exhibit different degrees of hubness.

**ORPHANS:** Points that are never retrieved in kNN queries.

Most points in high-dim. data are orphans, which leads to an information loss.

**FEATURE ASSESSMENT**

- **Feature Representations**
- **Similarity Measures**
- **k-NN topology**
- **Analyse**
- **Evaluate**

**HUBS: unusually frequent nearest neighbors, similar to many other points.**

Their occurrences are often not informative and act as noise.

**DATA OVERVIEW**

- Main hubs are projected onto a viewing pane by multi-dimensional scaling.
- The average induced label mismatch percentages in kNN sets are used to determine the background landscape, which is smoothed by multiple passes of low-level convolution filters.

**EXAMINE INDIVIDUAL HUBS**

- The Graph View allows for a selection of direct and reverse nearest neighbors of each point, which can be added to the view. Their neighbors can also be selected, etc. This way, users can form a local kNN sub-graph and examine its structure.
- Here is an example of a bad hub from the butterfly image data. The *Artogea rapae* image in the middle is the nearest neighbor to many images from different classes, i.e. different butterfly species. This is not a feature of the image itself, but a consequence of the specific feature representation and metric. We can then change metrics and go back and select the same image and observe how the structure has changed.

**POINT TYPE DISTRIBUTIONS**

- Different classes consist of different types of points, so some are more difficult to handle than others.
- IHE makes it easy to visualize this.

**OBJECT RECOGNITION AND RANKING**

Image Hub Explorer implements several novel approaches to hubness-aware learning and classification, as well as re-ranking. It also implements several standard baselines. These components can be evaluated for each rep./metric choice.

**IMAGE DATA IN THE EXAMPLE: LEEDS BUTTERFLY DATASET**

The data can be accessed at: [http://www.comp.leeds.ac.uk/scs/jwks/dataset/leedsbutterfly/](http://www.comp.leeds.ac.uk/scs/jwks/dataset/leedsbutterfly/)

**CONTACT**

For more information on our work on hubness, visit: [http://aialab.ijs.si/renad_tomasev/](http://aialab.ijs.si/renad_tomasev/)