Descriptions of Functions:

Note: For more information on each function including the arguments that the function needs, read the associated doc string in the utilities.py file or type "help(function_name)" into a jupyter notebook cell and run it.

Utility Functions:

- **view_data_toc**
  Prints the Dataset Table of Contents file. Displays a table describing each available dataset. This is the same table on the main page of the Biological Clocks Class GitLab repository.

- **load_dataset**
  Loads a dataset into a dataframe, using the name of the dataset found in the "Dataset" column of the Dataset Table of Contents.

- **load_results**
  Loads results from periodicity algorithms or LEMpy into a dataframe, using the name of the results file (for pyJTK and pyDL) or directory (or Lomb-Scargle and LEMpy).

- **duplicate_check**
  Checks for duplicates within a dataset dataframe and prints out tips for what to do with the duplicates.

- **remove_duplicates**
  Removes duplicate gene names by one of two methods supplied to the function.
  Method 1: keep only the duplicate with the highest gene expression at any time point.
  Method 2: keep the duplicate with the highest average gene expression.

- **relabel_duplicates**
  Relabels duplicate gene names by appending "dupN" to each duplicate name where N is an integer.

- **intersection**
  Returns a list that consists of the intersection of two supplied lists.

- **uniques**
  Returns the unique genes in each the two supplied lists. Unique genes are those are not in the intersection of the two lists.

- **get_genelist_from_top_n_genes**
  Returns a gene list consisting of the top n genes based off a specified periodicity score and a specified number of genes.

- **get_genelist_from_threshold**
  Returns gene list consisting of the top genes based off a specified periodicity score and a specified numeric threshold on the periodicity score.

- **normalize_data**
  Z-score normalizes a time series dataframe.

Visualization Functions:

- **plot_heatmap**
  Plots gene expression in a heatmap. Genes to plot are based on a user supplied threshold on a periodicity score. Orders the heatmap on the peak gene expression from the first period.
- **plot_heatmap_in_supplied_order**
  Plots genes from a dataset in a heatmap ordered by a supplied order of gene names.
- **plot_linegraphs_from_gene_list**
  Plots genes from a dataset from the supplied gene list in line plots. Can only plot between 1 and 10 genes.
- **plot_linegraphs_from_top_periodicity**
  Plots the top n genes from a dataset in line plots. Top genes are determined based on supplied gene number and the supplied periodicity results.
- **make_network_from_edge_list**
  Makes an interactive network graph from a list of edges in LEM edge specification.
- **make_top_edge_network**
  Makes an interactive network graph from the top n edges from the LEM all scores dataframe.

Periodicity Functions:

Note: For descriptions of individual tools, check the Tools section of the README.

- **run_pyjtk**
  Uses pyJTK to analyze the periodicity of each gene in a time series dataset dataframe. •
run_pydl
  Uses pyDL to analyze the periodicity of each gene in a time series dataset dataframe.
- **run_ls**
  Uses Lomb-Scargle to analyze the periodicity of each gene in a time series dataset dataframe.
- **run_periodicity**
  Uses pyJTK, pyDL, and Lomb-Scargle on a single dataset to analyze and rank the periodicity of a time series dataset. Returns individual results for each periodicity algorithm.

LEM Functions:

Note: For descriptions of individual tools, check the Tools section of the README.

- **run_lem**
  Runs LEMpy on a time series dataset dataframe, specifying what genes are targets, transcriptional repressors, and transcriptional regulators.

Stripey Functions:

- **interpolate_timepoints**
  Interpolates specified timepoints of a timeseries dataframe to handle stripeys.
- **quantile_normalize**
  Quantile normalizes a timeseries dataframe to handle stripeys.