

---

# **LoopRank Documentation**

***Release 0.1***

**Simon Streicher**

October 18, 2014



<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Basic approach . . . . .	3
<b>2</b>	<b>Input data formats</b>	<b>5</b>
2.1	Time series data format . . . . .	5
2.2	Descriptive labels data format . . . . .	5
2.3	Connectivity information . . . . .	5
<b>3</b>	<b>Outputs</b>	<b>7</b>
3.1	Normalised time series data . . . . .	7
3.2	FFT data . . . . .	7
3.3	Band-gap filtered data . . . . .	7
3.4	Time shifted transfer entropy results . . . . .	7
3.5	Full ranking graph . . . . .	7
3.6	Top edges only ranking graph . . . . .	8
3.7	Ordered node importances . . . . .	8
<b>4</b>	<b>Supporting software</b>	<b>9</b>
4.1	TOPCAT - time series and plane plotting from CSV files . . . . .	9
4.2	Cytoscape - graph plotting, layout calculation and visual effects from GML files . . . . .	9
4.3	ViTables . . . . .	9
<b>5</b>	<b>Transfer entropy calculation</b>	<b>11</b>
<b>6</b>	<b>Weight calculation</b>	<b>13</b>
<b>7</b>	<b>Node ranking</b>	<b>15</b>
<b>8</b>	<b>Indices and tables</b>	<b>17</b>



Contents:



---

# Introduction

---

The LoopRank software package provides a set of tools for prioritising base layer control loops from a maintenance perspective on a plant-wide scale.

A mixture of data- and knowledge-based methods are employed.

## 1.1 Basic approach

The basic approach followed follows a two-step process:

- Defining the connectivity between all elements in the process.
- Determining appropriate edge weights between process elements to indicate strength of interaction.
- Using the connectivity to build a digraph which is then analysed according to various graph ranking algorithms.





---

## Input data formats

---

In this section the required and optional data input formats are described.

### 2.1 Time series data format

The input data format for time series data associated with data tags on the plant are comma separated files (CSV) as follows: \*The first row should be a header line, with the first column label being “Time”. \*The first column should contain the time of measurements in UNIX time. \*The rest of the columns should contain the raw data - no need to be normalised, this will be done automatically in the post-processing stages.

### 2.2 Descriptive labels data format

Descriptive labels might be associated with each process data tag. This should be provided in the form of a CSV file with the data tag name in the first column and the description in the second column. The first row should have the labels “Tag name” and “Description” in them. The filename should be “tag\_descriptions.csv”.

### 2.3 Connectivity information

Limiting the connections to certain edges only will be an optional feature. This is provided for the cases where plant topology information is available and considered to be important to include in the analysis.

Please note that adding connectivity information is not always helpful, and in some cases results in poorer analysis of the root cause of the problem as higher-order connections might play an important role in boosting a particular node’s score in the network.



---

## Outputs

---

In this section the outputs made available to the user are described.

### 3.1 Normalised time series data

This is simply a CSV file with the same structure as the time series input file, with the data associated with each tag being normalised. This is useful for qualitatively analysing patterns in the data using tools such as TOPCAT for easy time series plot creation.

### 3.2 FFT data

A CSV file with the first column containing the frequency and the columns after that the normalised magnitude of FFT results for the different tags. Useful for identifying signals with peaks in the same region to decide on which band-gap filters to use for analysis of specific disturbances.

### 3.3 Band-gap filtered data

A CSV file with the same structure as the main time series input data file, except that normalised and band-gap filtered signals are presented.

### 3.4 Time shifted transfer entropy results

A CSV file presenting the directional (difference between forward and backwards) and absolute (forwards only) transfer entropies for each source/destination node pair. A separate file for each source node is created. Useful for investigating the smoothness of transfer entropy with respect to time delays, as well as the different regions that are most likely to be useful for further analysis.

A significance threshold for the zero-offset dataset is available as well.

### 3.5 Full ranking graph

A GML file providing a directed graph with edge weights and node importances. Self-loops are removed by default before ranking is performed. Results for the following scenarios are provided: \*Dummies used in calculation, but

suppressed in the final results. \*No dummies used in calculation. \*Dummies used and still visible.

## 3.6 Top edges only ranking graph

A GML file similar to that for the full ranking graphs, but only including the nodes and edges associated with the specified number of top edge weights.

## 3.7 Ordered node importances

CSV files providing node labels, descriptions and importance scores organised from largest to smallest for the different with results provided in the graph results discussed above.

---

## **Supporting software**

---

In this section the recommended supporting software for analysing the various output files produced is discussed.

### **4.1 TOPCAT - time series and plane plotting from CSV files**

The CSV output files can be imported in TOPCAT and various time series as well as plane plots easily generated and saved in PDF format.

The time series plot format is recommended for all time series data outputs (normalised time series data).

The plane plot format is recommended for all other kinds of plots (transfer entropy vs. time delay, FFT plots, etc.)

### **4.2 Cytoscape - graph plotting, layout calculation and visual effects from GML files**

Cytoscape can be used to produce digraph schematics in specific layouts and with various visual aids from the GML files produces.

### **4.3 ViTables**

ViTables can be used to view and edit H5 data tables.



---

## Transfer entropy calculation

---





---

## Weight calculation

---

Weight calculation is one of the core operations performed on time series data collected from process sensors in order to infer the relative importance of process elements.

The weight calculation methods are called from the `gaincalc` module in the `ranking` library.

---

**Note:** The following is the auto-generated documentation from the `gaincalc` module source:

---



---

## Node ranking

---

The weight calculation methods are called from the `noderank` module in the `ranking` library.

---

**Note:** The following is the auto-generated documentation from the `noderank` module source:

---



---

## Indices and tables

---

- *genindex*
- *modindex*
- *search*