

UNIVERSIDAD NACIONAL DEL LITORAL

Facultad de Ingeniería y Ciencias Hídricas



**Manual de Usuario del entorno web para
minería de redes de regulación entre genes en
Bioinformática**

Autor:
Eduardo Oliva

Revisó:
Dr. Mariano Rubiolo

Santa Fe, Agosto de 2017

Índice.

1. Descripción.....	3
2. GRNNMiner.	3
3. Pantalla Principal.....	3
4. Data.....	4
5. Simulate Data.	5
6. Label.....	6
7. Parameters y Rules	7

1. Descripción

El presente documento contiene información básica para poder utilizar la herramienta web para minería de redes de regulación entre genes, basada en el conjunto de algoritmos GRNNMiner.

2. GRNNMiner.

El modelo GRNNminer fue desarrollado en el Instituto de Investigación en Señales, Sistemas e Inteligencia Computacional sinc (i), de doble dependencia UNL- CONICET, ubicado en la Facultad de Ingeniería y Ciencias Hídricas de la Universidad Nacional del Litoral. El mismo utiliza modelos neuronales para encontrar la GRN subyacente a los datos, modelando la posible relación existente entre cada par de perfiles de expresión de genes existentes en el conjunto de datos.

El objetivo de obtener las relaciones existentes entre los genes de un conjunto de datos, es poder determinar el patrón de activaciones e inhibiciones entre los genes que componen la GRN resultante.

3. Pantalla Principal.

Como se puede observar en la Figura 1, el entorno web cuenta con diferentes funcionalidades y servicios que requieren cierto conocimiento técnico, debido a que se necesitan settear de manera correcta diversas variables, parámetros, entre otros, para lograr de esta manera el correcto funcionamiento del conjunto de algoritmos. En las siguientes secciones se realizará el desarrollo de las mismas.

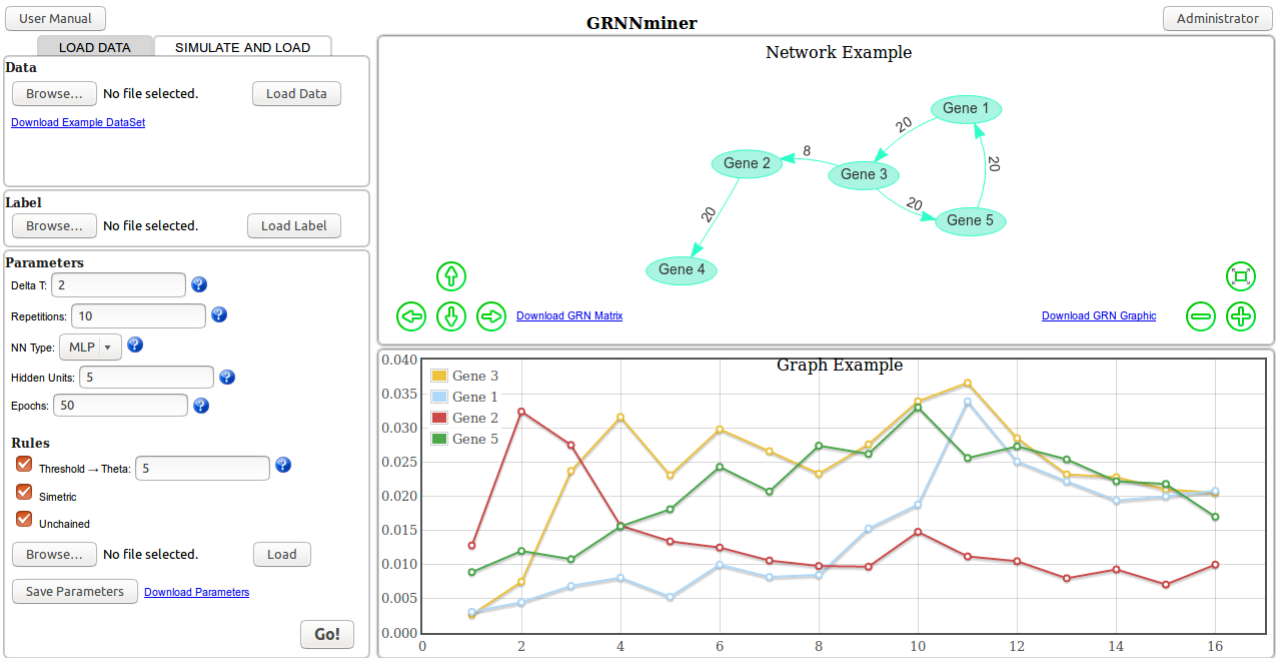


Figura 1: Pantalla Principal.

4. Data.

Al momento de comenzar a utilizar la herramienta web es necesario contar con un conjunto de datos específicos. Si no se cuenta con los mismos, se brinda la posibilidad de generarlos (detallado en el punto 5. Simulate Data).

En la Figura 2 podemos ver donde se debe cargar los datos que poseemos.

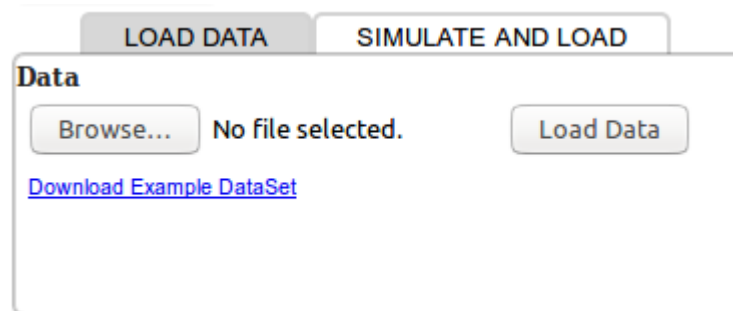


Figura 2: Carga de datos.

Como puede observarse, se permite la carga de un archivo en formato "Comma separated value" con extensión ".csv" que contenga información del comportamiento de los genes a analizar en un determinado intervalo de tiempo. Es decir, el archivo debe contener los instantes de tiempo para cada uno de los genes. Cabe aclarar, que los datos pueden ser

datos reales (obtenidos de experimentos biológicos) o datos artificiales (previamente simulados).

Además se puede ver la presencia de un link “Download Example DataSet” el cual inicia la descarga de un archivo de ejemplo con el formato correcto que debe poseer el archivo de datos.

Luego, para comenzar con la carga del mismo se debe presionar el botón “Browse”, seleccionar el archivo y presionar el botón “Load Data” de esta manera el entorno informará si se cargó el archivo de manera correcta o existió alguna falla.

Ejemplo de Archivo de datos:

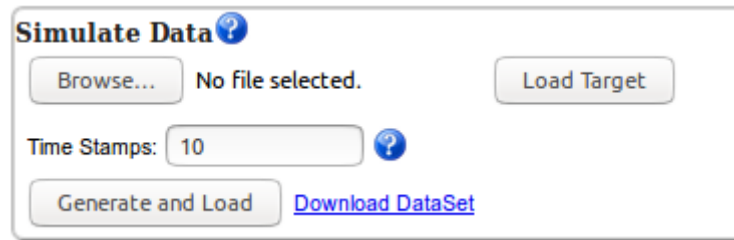
```
0.0031000000000000,0.0128000000000000,0.0027000000000000,0.0087000000000000,0.0089000000000000
0.0045000000000000,0.0324000000000000,0.0075000000000000,0.0221000000000000,0.0120000000000000
0.0069000000000000,0.0275000000000000,0.0237000000000000,0.0240000000000000,0.0108000000000000
0.0081000000000000,0.0157000000000000,0.0316000000000000,0.0078000000000000,0.0156000000000000
0.0053000000000000,0.0134000000000000,0.0231000000000000,0.0061000000000000,0.0181000000000000
0.0100000000000000,0.0125000000000000,0.0298000000000000,0.0085000000000000,0.0243000000000000
0.0082000000000000,0.0106000000000000,0.0266000000000000,0.0093000000000000,0.0207000000000000
0.0085000000000000,0.0098000000000000,0.0233000000000000,0.0076000000000000,0.0274000000000000
0.0153000000000000,0.0097000000000000,0.0276000000000000,0.0080000000000000,0.0262000000000000
0.0188000000000000,0.0148000000000000,0.0339000000000000,0.0096000000000000,0.0330000000000000
0.0339000000000000,0.0112000000000000,0.0366000000000000,0.0119000000000000,0.0256000000000000
0.0251000000000000,0.0105000000000000,0.0285000000000000,0.0084000000000000,0.0273000000000000
0.0222000000000000,0.0080000000000000,0.0232000000000000,0.0067000000000000,0.0254000000000000
0.0194000000000000,0.0093000000000000,0.0228000000000000,0.0059000000000000,0.0222000000000000
0.0200000000000000,0.0071000000000000,0.0210000000000000,0.0048000000000000,0.0218000000000000
0.0208000000000000,0.0100000000000000,0.0205000000000000,0.0084000000000000,0.0170000000000000
```

Figura 3: Ejemplo archivo de datos.

5. Simulate Data.

Analizando la Figura 4, se brinda la posibilidad de generar ese archivo de datos de expresión de genes (que es necesario cargar en el sub-área Data). Es por eso que si no se cuenta con el mismo, solo se tiene que cargar el archivo Target y setear la cantidad de instantes de tiempo que se desea que posea ese conjunto de genes que se va a generar y presionar el botón “Generate and Load”. Para simular los datos se utiliza un algoritmo generador denominado GeneSim.

Para ello el archivo Target debe contener una matriz que represente la relación entre los genes. Dicha matriz debe ser cuadrada y del tamaño de la cantidad de genes que se poseen. El formato del archivo a ser cargado debe ser tipo “.csv”, similar al del ejemplo en Figura 5:



Simulate Data ?

Browse... No file selected. Load Target

Time Stamps: 10 ?

Generate and Load [Download DataSet](#)

Figura 4: Simulate Data.

```
0,1,0,0,0
0,0,0,1,0
1,0,0,1,1
0,1,0,0,0
1,0,0,0,0
```

Figura 5: Ejemplo de archivo de target.

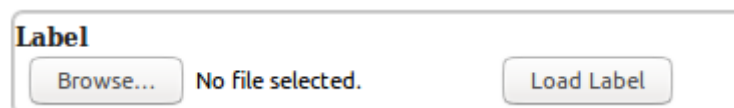
Siguiendo el mismo esquema que en la sub-área anterior, los pasos a seguir para cargar el archivo Target son los mismos. Dicho archivo solo debe ser cargado si no se posee un conjunto de datos, ya que el mismo va a ser utilizado para luego generar los datos. Además también con una ayuda en la cual se visualiza el formato correcto que debe poseer un archivo de datos.

Además de poder generar ese archivo que contiene el comportamiento de los genes a lo largo de los distintos instantes de tiempo, se pueden descargar los mismos, presionando el link “Download DataSet”, de esta manera comienza la descarga de los datos en un archivo en formato .csv.

Cabe aclarar que si no se carga el archivo Target, no se podrán simular los datos.

6. Label

En la Figura 6, se puede ver que es posible cargar un archivo que contenga etiquetas con los nombres de los genes, el mismo debe ser un archivo en formato .csv y para la carga se deben seguir los mismos pasos que en las sub-áreas anteriores.



Label

Browse... No file selected. Load Label

Figura 6: Label.

Es muy importante saber que es opcional la carga de etiquetas, ya que si no se carga un archivo que contenga las mismas, el entorno etiqueta los genes de manera automática bajo la nomenclatura “Gen X”. Archivo a modo de ejemplo:

CBF1, GAL4, SWI5, GAL80, ASH1

Figura 7: Ejemplo de archivo de label.

7. Parameters y Rules

Se dispone de varios campos que deben ser establecidos para poder utilizar el entorno. Por defecto, el mismo ya posee valores, pero éstos pueden ser modificados a necesidad, dependiendo de la utilidad y el ensayo que se desea realizar (Figura 8).

The screenshot shows a configuration window titled "Parameters" and "Rules". Under "Parameters", there are input fields for "Delta T:" (value 2), "Repetitions:" (value 10), "NN Type:" (dropdown menu showing "MLP"), "Hidden Units:" (value 5), and "Epochs:" (value 50). Each field has a blue question mark icon to its right. Under "Rules", there are three checked checkboxes: "Threshold → Theta:" (value 5), "Simetric", and "Unchained". Below these, there is a "Browse..." button, a text field showing "No file selected.", and a "Load" button. At the bottom, there are "Save Parameters" and "Download Parameters" buttons, and a "Go!" button in the bottom right corner.

Figura 8: Parameters y Rules.

En la figura 8 se visualizan los diferentes campos a setear:

Parameters:

- Delta T: Ventana temporal a ser considerada.
- Repetitions: Número de repeticiones de cada experimento configurado en el entorno.
- NN Type: Tipo de red neuronal, que al momento son MLP (MultiLayer Perceptron) y ELM (Extreme learning machines).
- Hidden Units: Porcentaje de entradas consideradas sobre el número de neuronas ocultas de la red neuronal.

- Epochs: Número de épocas usadas para entrenar la red neuronal, cuando la misma es del tipo MLP.

Rules (sirven para resolver algunas relaciones dudosas entre distintas series de tiempo):

- Threshold: Las conexiones débiles se eliminan estableciendo un umbral (theta) por encima del cual sólo permanecen las relaciones más fuertes.
- Theta : Umbral para reglas de Minería.
- Simetric: Determina la dirección verdadera de la regulación entre dos genes.
- Unchained: Define el sentido correcto de la regulación de un gen a otro, cuando hay más de un gen involucrado en una cadena o un árbol.

Como se puede visualizar en la figura 8, esta sub-área posee un campo (Browse) que permite la carga de un archivo en formato .txt, el cual contenga los parámetros y reglas en el orden en que se encuentran en la sub área. Preferentemente se recomienda que dicho archivo haya sido creado por el entorno en alguna simulación precedente, evitando de esta manera cualquier tipo de error. Una vez seleccionado el archivo se debe presionar el botón “Load” para realizar la carga de los parámetros y reglas.

Si no se cuenta con un archivo de parámetros y reglas, los mismos deben ser establecidos manualmente. Una vez fijados estos valores se debe proceder a presionar el botón “Save Parameters”, de esta manera se guardan los parámetros y las reglas (presionar el botón “Save Parameters” solamente si NO se cargó un archivo con los mismos previamente.)

Además se brinda la posibilidad de poder exportar los parámetros y reglas fijadas manualmente en un archivo de formato .txt, para ellos es necesario utilizar el link “Download Parameters”.

El archivo de parámetros y reglas, deberá tener el siguiente formato:

```
DeltaT: 2
Repetitions: 10
HiddenUnits: 5
Epochs: 50
NeuralNetwork: MLP
Threshold: on
Theta: 5
Simetric: on
Unchained: on
```

Figura 9: Ejemplo de archivo de parámetros exportados.

Finalmente, en dicha sub-área se encuentra un botón denominado “Go!” cuando se presiona desenlaza los algoritmos de GRNNminer, calculando de esta manera la GRN en base a los archivos previamente cargados y los parámetros establecidos.

Por otra parte, en el área gráfica, se cuenta con dos links que permiten la descarga de los resultados obtenidos, ya que sea de manera matemática como de manera gráfica.

El link denominado “Download GRN Matrix” brinda la posibilidad de exportar la red obtenida en un formato de archivo “.csv”. Por otro lado el link “Donwload GRN Graphic” permite descargar una imagen de la GRN en formato “.png”. En las figuras 10 y 11 se pueden observar los resultados exportados.

```
0,0,20,0,0
0,0,0,20,0
0,8,0,0,20
0,0,0,0,0
20,0,0,0,0
```

Figura 10: Ejemplo de archivo de resultados (GRN).

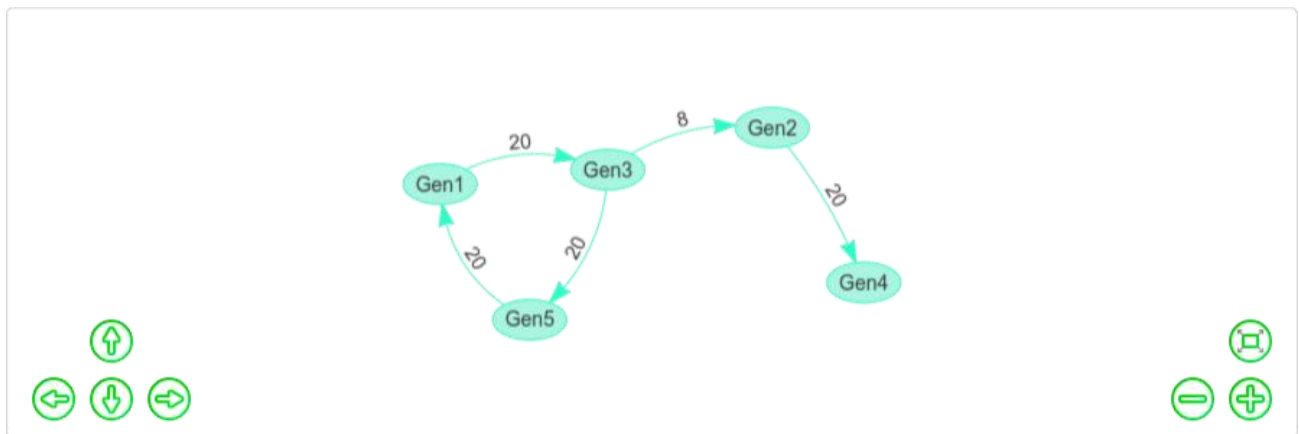


Figura 11: Ejemplo de archivo de resultados en forma gráfica (GRN).