Table of Contents

The **GeneralGeneletModel.py** package (imported as GGM in accompanying scripts) includes a GeneletNetwork class that for initializing, simulating, and compiling sequences for genelet networks. The general genelet model functions at the beginning of **GeneralGeneletModel.py** are helper functions for the GeneletNetwork class. Below are the descriptions for the initialization of GeneletNetwork and how to use its associated functions.

1. Class initialization definition:

CLASS: GeneletNetwork(act_vec=req1,prod_vec=req1,indc_vec=0,blk_vec=inferred,top_mat=req2,Itop_mat=0,genelet_type='KWG')

Initializes the genelet network instance to be simulated

Input definitions:

For act_vec and blk_vec:

List the length of ind_nodes (total nodes)

Numbers represent which orthogonal activators/blockers correspond to which nodes

0 indicates no activator/blocker for a given node

blk_vec is optional if there are no dummy nodes present (i.e. redundant nodes to serve as reporters that do not produce outputs)

If not provided will be inferred from nodes that have coactivators coming to them from the topology matrix

If it is desired to include dummy nodes (redundant reporter nodes that do not produce transcripts) in simulations you must specify blk_vec input

For prod_vec:

List the length of ind_nodes (total nodes)

Numbers represent which RNA repressors and/or RNA coactivators are produced by which nodes

-ve values are repressors
+ve values are coactivators
0 = no production of an inducer RNA

For indc_vec:

List the length of ind_nodes (total nodes)

Numbers represent which inducer RNAs are produced by which nodes

-ve values are inducers that bind repressors (thereby activating)
+ve values are inducers that bind coactivators (thereby BLKing)
0 = no production of an inducer RNA

indc_vec is optional is default to zeros if not provided

For genelet_type:

The type of genelet chemistry used

'KWG' = Kim and Winfree genelet design with free activators and blockers
'STG' = Spatiotemporal genelets with activators directly attached to genelets
This model also allows free activator to be added as a threshold the free activator binds RNA but does not activate the genelets

For top_mat:

A trinary topology matrix that defines how topological nodes connect to one another

A list of lists (ortho_nodes x ortho_nodes) where entries are defined as for prod_vec:

```
  EX:  IFFL network      EX:   TSN network
       TN1  TN2  TN3           TN1  TN2  TN3
  TN1 [ 0    0    0 ]    TN1 [ 0   -1   -1 ]
  TN2 [ 1    0    0 ]    TN2 [-1    0   -1 ]
  TN3 [ 1   -1    0 ]    TN3 [-1   -1    0 ]
```

For Itop_mat:

A trinary topology matrix that defines how topological nodes connect to RNAs in the network

Defines which nodes produce which inducer RNAs
Values are defined as for indc_vec
This is an optional input and will default to 0s if not implemented

There are two ways to initialize the GeneletNetwork

1) input BOTH act_vec and prod_vec (and blk_vec if dummy reporting nodes are present)

```
EX: model = GeneletNetwork(act_vec,prod_vec,indc_vec,blk_vec)
EX: model = GeneletNetwork(act_vec,prod_vec)
EX: model = GeneletNetwork(act_vec,prod_vec,blk_vec=[0,2,2,3])
```

2) input top_mat (is not compatible with defining dummy reporting nodes)

```
EX: model = GeneletNetwork(top_mat=topology_mat)
EX: model = GeneletNetwork(top_mat=topology_mat,Itop_mat=I_mat)
```

Output definitions:

After initializing the GeneletNetwork class the user has access to the following Attributes [where: model = GeneletNetwork( )]

```
model.ortho_nodes     - number of orthogonal genelet nodes
model.ind_nodes       - number of total (individual) genelet nodes
model.act_vec         - activator vector mapping ind_nodes to orthogonal activators
model.prod_vec        - RNA production vector mapping ind_nodes to which RNAs they
                        produce (repressors/coactivators)
model.blk_vec         - blocker vector mapping ind_nodes to orthogonal blockers
model.indc_vec        - RNA inducer production vector mapping ind_nodes to which
                        inducers they produce
model.topology_mat    - topology matrix (an alternative representation of
                        act_vec/prod_vec notation)
model.I_topology_mat  - inducer topology matrix (an alternative representation of
                        act_vec/indc_vec notation)
model.act_mat         - activator connection matrix used in ODEs to map activators to
                        ind_nodes
model.rep_mat         - repressor connection matrix used in ODEs to map repressors to
                        ind_nodes
model.blk_mat         - blocker connection matrix used in ODEs to map blockers to
                        ind_nodes
```

model.ca_mat - coactivator connection matrix used in ODEs to map coactivators to ind_nodes
model.Rprod_mat - repressor production matrix used in ODEs to map which ind_nodes produce which repressors
model.Cprod_mat - coactivator production matrix used in ODEs to map which ind_nodes produce which coactivators
model.Rindc_mat - repressor inducer production matrix used in ODEs to map which ind_nodes produce which repressor inducers
model.Cindc_mat - coactivator inducer production matrix used in ODEs to map which ind_nodes produce which coactivator inducers

2. Function: initial conditions( ) definition:

initial_conditions(dA_tot,G_tot,G_int_vec=[],dB_added=0,rRin=0,rCin=0,rIrin=0,rIcin=0)

Sets the initial conditions for the network. All input concentrations should be in nM

Input definitions:

For dA_tot and G_tot:

numpy arrays the length of ortho_nodes and ind_nodes, respectively

Represent total concentrations (nM) of activators and genelets, respectively

For G_int_vec:

List the length of ind_nodes (total nodes)

Represents the initial state of a node
1 for ON
-1 for BLK
0 for OFF

G_int_vec is optional if there are no dummy nodes present (i.e. redundant nodes that serve as reporters that do not produce RNAs)

If not provided will be inferred from nodes from the topology matrix as follows:

Nodes with no inputs are ON
Nodes with coactivator inputs are BLK
Nodes that are only repressed are ON

If it is desired to include dummy nodes (redundant reporter nodes that do not produce transcripts) in simulations you must specify G_int_vec input

For dB_added:

All BLK genelets are assumed to be annealed with 50% excess dB
Thus for 25 nM of a BLK there will be 32.5 nM total dB and 12.5 nM free dB to as the initial concentration

To specify additional free dB modify the dB_added input (in nM)

EX: model.initial_conditions(dA_tot,G_tot,G_int_vec,dB_added=[0,150,150])

For all other species the default values are 0:

To modify these species input a list of concentrations for any other species by name (in nM):

EX: model.initial_conditions(dA_tot,G_tot,G_int_vec,rCin =[0,150,150])

EX: model.initial_conditions(dA_tot,G_tot,G_int_vec,rRin=[1000,0,0])

Output definitions:

After calling initial_conditions the user will have access to the following attributes:

model.G_int_vec -the list of initial genelet states (these may be determined as a default)
model.ind_cond -the vector holding the initial conditions for each species
model.(species) -all of the individual species concentrations that make up int_cond are also accessible as attributes by name (dB_added, rRin, etc)

3. Function: simulate( ) definition:

simulate(t_vec,iteration,rnase='RnH',leak=0,rate_constants=[],rR=[],dA=[],rC=[],dB=[],rIr=[],rIc=[],dR=[])

Simulates the network with the above initial conditions

Simulate can be called numerous times and specific conditions can be updated by name
Use 'NA' for values that should not be updated for a given species

For updating a species conditions, the list has to be the length of the total number of that species and concentrations in nM

dR is a special option for KWG models where DNA repressors can be added to the reaction to permanently remove activators
dR species are currently not included in the symbolic equations that can be output as they are not really part of the networks

t_vec2 should start at the last timepoint from t_vec1 and go to a later

Input definitions:

For rnase:
Selects the rnase to use in the simulations
'RnH' = RNase H (degrade RNA in RNA:DNA duplex)
'RnA' = RNase A (degrade ssRNA and RNA in RNA:DNA duplex (likely with slower rate than RNase H))
'Both' = RNase H and A

This input is optional and 'RnH' is the default

For leak:
Represents the leak transcription rate of BLK (KWG) or GrR (STG) genelets as a fraction of the ON transcription rate between 0 and 1

Can be entered as a single value applied to all genelets or as a list for different leak rates for each RNA production:
leak = [0.05,0.1,0.075,0.06]
There should be ind_node # of leak inputs as in this model each individual genelet has its own production rate even if two genelets produce the same RNA

For rate_constants:
Use different rate constants than the default values

```
       If genelet_type = 'KWG' then import as a list of single values (1/M-s or 1/s)
       [kpr,kpc,kpi,kd_H,kd_A,kga,kgar,kar,kgb,kgab,kgbc,kbc,kir]

       If genelet_type = 'STG' then import as a list of single values (1/M-s of 1/s)
       [kpr,kpi,kd_H,kd_A,kd_Ag,kd_Hg,kgar,kar,kir]

       Each individual rate can be a list of length ind_nodes for production rates
              Or a list of length ortho_nodes for other rates
              Or a single value which will be assumed to be the same for all nodes

EX: model.simulate(t_vec1,1)
EX: model.simulate(t_vec2,2,rIr=['NA','NA',10000])
EX: model.simulate(t_vec3,3,rR=[1000,'NA','NA'])
EX: model.simulate(t_vec1,1,rnase='RnA')  # use RNase A
EX: model.simulate(t_vec1,1,rnase='both',leak=0.1) # use both RNase A and H
                                                   # 10% leak transcription

EX: model.simulate(t_vec1,1,rnase='both',leak=0.1,rate_constants=rate_list)
                                                  # use both RNase A and H
                                                  # 10% leak transcription for all RNAs
                                                  # user input rate constants

EX: model.simulate(t_vec1,1,leak=[0.05,0.1,0.075]) # different leak rates for 3 RNAs
```

Output definitions:

After running simulate the user has access to the following useful attributes:

```
model.sol    -sol.y contains all of the concentration date and sol.t contains the time
              in seconds

model.(rate_constants)   -all of the individual rate constants can be accessed by name

model.output_concentration    -the dictionary containing all of the concentrations
                               with specific name keys

EX: G1 = model.output_concentration['GdA1'] # for KWG
EX: G1 = model.output_concentration['Gon1'] # for STG
EX: R4 = model.output_concentration['rR4']
```

4. Function: export sym eqs( ) definition:

```
export_sym_eqs(file_name='sym_eqs',path_name='')
-------------------------------------------------------------------------------
Exports the symbolic ODE equations of the model to a text file

Both of the inputs are optional and have default values
```

Input definitions:

For file_name:
       User defined name of the .txt file (do not include .txt in the name)

For path_name:
       User defined path to where the file should be saved (include \ at the end)

       If user does not supply a path the file is saved in current folder

EX: model.export_sym_eqs(file_name='TSN_eqs',path_name='C:\\Desktop\\')

Output definitions:

The output from this is the text file with the symbolic equations in it
with the specified or default filename in the specified or current (default) folder

5. Function: plot topology( ) definition:

```
plot_topology(pos=[],layout='spring',plot_title='',show_rnas=1)
-------------------------------------------------------------------------------
Plots the network topology with initial node states

This considers both genelet nodes and RNA coactivators/repressors as topological nodes
which allows inducer RNA connections to be represented

If this is called right after initializing the network the initial states will be
inferred

If this is called after the initial_conditions function has been called the user
defined initial states will be displayed
```

Input definitions:

For pos:

This is an optional input of each nodes position in the plot

It has to be a dictionary with keys that are the names of the topological nodes and the desired [x,y] position values

For layout:

This is an optional input that selects the algorithm to organize the nodes in the plot

Default is 'spring'
Other options:
'spectral'
'circular'
'shell'

See the networkX package layout options for further details

For plot_title:

This is an optional title for the topology plot

For show_rnas:

This is an option to show the RNA repressors/coactivators in the topology as their own nodes

This is necessary if you want to plot topologies where there are nodes that produce inducers

If this is set to 0 then the RNAs will not be plotted and any inducer nodes will appear free standing

EX: model.plot_topology(layout='shell',plot_title='IFFL network')

Output definitions:

The output from this is the topology plot

model.net_edges gives the user access to all the network edges used in the plot

6. Function: compile network sequences( ) definition:

```
compile_network_sequences(input_file_loc,desired_nodes=[],min_design=0,bth=1,\
save_file_name='genelet_seqs',save_path_name='')
```

Exports genelet sequences for a specified topology to a text file

Algorithm for domain selection for unspecified nodes:

First, nodes that must both be coactivated AND repressed are populated with domains that meet this criteria selected in order of their OCRW score in Supp. Section 12

Second, nodes that do not have ANY coactivator or repressor inputs are Populated

Since these nodes are the likely place a network will be expanded they are:

First populated with domains that can be coactivated AND repressed selected in order of their OCRW score in Supp. Section 12

Second populated with domains that can be coactivated selected in order of their OCRW score in Supp. Section 12

Third populated with domains that can be coactivated selected in order of their OCRW score in Supp. Section 12

Third, nodes that have ONLY coactivator inputs are populated

If min_design = 0 the first available domain that can be coactivated will be selected in order of their OCW score in Supp. Section 12

If min_design = 1 the first available domain that can ONLY be coactivated will be selected in order of their OCW score in Supp. Section 12

If there are not enough min_design domains then the remaining domains will be selected as if min_design = 0

Fourth, nodes that have ONLY coactivator inputs are populated

If min_design = 0 the first available domain that can be repressed will be selected in order of their OCW score in Supp. Section 12

If min_design = 1 the first available domain that can ONLY be repressed will be selected in order of their ORW score in Supp. Section 12

If there are not enough min_design domains then the remaining domains will be selected as if min_design = 0

Input definitions:

For input_file_loc:

The file path to the Excel file holding all of the viable genelet node Sequences (all_genelet_sequences.xlsx)

For desired_nodes:

An optional input that specifies which input domains should be used for each TN in the network

By default, this is empty, algorithm will select its own nodes if not provided

Provide as a list of length ortho_nodes with the names of nodes to be used to compile the sequences (['G1','G5','G8'])

TN1 TN2 TN3

The order of the list of nodes is in the order of the topological nodes as shown above

An error will occur if you select a domain that cannot conduct the regulation required at that position

i.e. if the TN needs to be coactivated and repressed and you select a domain that only represses

If only some nodes are specified and others are to be selected, leave the unspecified positions blank: (['G1',' ','G8'])

TN1 TN2 TN3

The unspecified positions will be filled with other domains using the same algorithm as when no nodes are specified

If there is an unspecified position and min_design = 1 then the unspecified node will attempt to use a min_design domain if possible

For min_design:

An optional input to specified if sequences selected should adopt a minimum design

By default this is 0 so the first nodes that meet the needs of the network will be used as described for the algorithm

If min_design = 1 the algorithm will select nodes with the minimal possible function in the network

For example, if a node is only coactivated in the network then the algorithm will use the first node that can only be coactivated at this position

If there are not enough minimal nodes to fill the network other nodes will be selected as described for the algorithm

For bth:

Optional input that specifies whether or not BTH domains should be included

This is 1 by default which keeps the 8 base 5' blocker toehold (BTH) on all the genelets in the network

This can be set to 0 and then the 8 base 5' blocker toehold will be removed from all genelets that are only repressed in a given network as these nodes do not need a blocker

This makes it possible to order shorter -nt sequences without having to truncate their 3' end (as for the TSN)

These truncated sequences will be denoted with a * in the output file

For save_file_name:

User defined name of the .txt file to be saved (do not include .txt in the name)

If user does not supply a name it will be saved as 'genelet_seqs.txt'

For save_path_name:

User defined path to where the file should be saved (include \ at the end)

If user does not supply a path the file is saved in current folder

EX:
```
model.compile_network_sequences(input_file_loc,save_file_name='TSN_sequences',save_path_name='C:\\Desktop\\')
```

Output definitions:

The output from this is the text file with the genelet sequences in it

Each node is comprised of G-nt / G-t / activator / blocker sequences as needed for the simulated network

All -nt sequences that have the BTH removed will be denoted with a *

Currently, any dummy reporter nodes (redundant nodes that do not produce any RNAs) will not be included as exported sequences

7. TROUBLESHOOTING

Check model.blk_vec if you have not supplied the blk_vec yourself

---

When doing iterative model.simulate calls make sure the t_vec units are correct, if t_vec1 = linspace(0,5,1000)*3600 then t_vec2 = linspace(t_vecl[-1]/3600,10,1000)*3600

---

blk_vec has to be specified as an input for the model to handle redundant dummy reporter nodes that do not produce RNA outputs

```
                G1 G2 G3
  EX: act_vec = [1, 2, 2]
     prod_vec = [2,-2, 0]
```

Here G3 does not possess a new activator and does not connect to anything but could serve as a labeled reporter as in experiments

However, such a node cannot be represented in the simplified topology matrix format which is where the blk_vec is inferred from so things like the blk_vec will be inferred incorrectly from the topology_mat - thus specify blk_vec if you want to include dummy nodes

blk_vec = [0,2,2] (without defining this the model with assume blk_vec = [0,2])

G_int_vec has to be specified as an input for the model to handle redundant dummy reporter nodes that do not produce RNA outputs

```
                G1 G2 G3
  EX: act_vec = [1, 2, 2]
     prod_vec = [2,-2, 0]
```

Here G3 does not possess a new activator and does not connect to anything and could serve as a labeled reporter as in experiments

However, such a node cannot be represented in the simplified topology matrix format which is where the G_int_vec is inferred from so G_int_vec will be inferred incorrectly from the topology_mat - thus specify G_int_vec if you want to include dummy nodes

G_int_vec = [0,-1,-1] (without defining this the model will assume G_int_vec = [0,-1])

The package currently does not populate sequences for any redundant dummy nodes in the system.

When defining a network from the topology matrix, the production vector will always be defined in increasing numeric order.

This could cause confusion when going back and forth between act_vec/prod_vec and top_mat definitions

For example you might define the IFFL as having prod_vec = [3,2,-3,0] which will produce a top_mat = [[0,0,0],[1,0,0],[1,-1,0]] but coverting that top_mat into an act_vec and prod_vec will produce prod_vec = [2,3,-3,0].

Both prod_vec definitions are correct since the first 2 nodes in the IFFL share the same activator but the values of G_tot need to be changed accordingly

---