Computational Representation of Metabolic Networks

Mark Poolman

June 28, 2018

This is *not* a programming course.

- No assumption of previous programming experience.

- Basic usage of a language as a tool - no technical details.

- Fundamental mathematical concepts as relevant to network analysis.

This is *not* a programming course.

- No assumption of previous programming experience.

- Basic usage of a language as a tool - no technical details.

- Fundamental mathematical concepts as relevant to network analysis.

- Flexibility - define what you want to do.

- Repeatability - apply the actions same actions to many models.

- Reliability - errors are less likely to go unnoticed, code can be analysed.

- Abstract concepts or large data-sets can't always be visualised.

Why Python ?

- Easy to learn.

- Forgiving.

- Flexible.

- Interactive.

- High level - lets you concentrate on the problem, not the computer.

- Wide range of existing software and libraries.

- Free (As in Beer and Freedom).

What's in a program?

- A collection of *data* representing some real-world entity.

- A set of actions that can be performed on that data.

- Some means by which the user can specify which actions to perform.

In Python (and other languages) the data and actions are both defined by *Objects* (aka *types*).

Objects and types

An object is a computational representation of something that exists in the real world.

The type (or class) of an object is defined by its properties.

- Cats:
    - Fur colour,
    - Length of whiskers.

- Proteins:
    - AA sequence,
    - Iso-electric point.

Objects and types

The type of an object defines what it can do, e.g.

- Cats can:
  - Sleep
  - Go miaow

- Proteins can:
  - Precipitate
  - Catalyse a reaction

Objects and types

The type of an object defines what can be done to it, e.g.

- Cats can be:
  - Stroked
  - Chased

- Proteins can be:
  - Crystallised
  - Digested

Objects and types

The type of an object defines its interactions with other objects.

- Cats can:
  - Reproduce with other cats
  - Digest a protein

- Proteins can:
  - Bind to other proteins
  - Poison a cat

The concept of objects that have known properties, can be acted upon and can interact with other objects is central.

Objects are abstract representations of their real-world equivalents (including proteins and cats).

These can be combined to represent data of arbitrary size and structure, including metabolic models and databases.

Types and Classes in Python

Python defines a number of built-in fundamental classes, which can be used to create more complex representations of real-world entities.

The distinction between types and classes in Python is historical, in modern python they are the same thing.

Attributes define the properties of an object and can either be:

Data attributes MyCat.NumberOfWhiskers

OR

Method attributes MyCat.PlayWithString()

Method attributes can be passed additional information:

MyCat.GotoSleep(3600)

Method attributes can *return* information:

FeedNow = MyCat.IsHungry()

The Boolean Class in Python

The simplest of all classes and can take the value of True or False.

FeedNow = MyCat.IsHungry()

FeedNow would be a Boolean value: MyCat is either hungry or it is not.

Numbers come in two main types: integer and float.

They behave *mainly* as expected

But see practical.

Python String Class

Strings are sequences of characters, often used for names and simple descriptions, but could also represent an entire document.

- Create an object called text of type string:

    ```
    >>> text = "My cat plays with string"
    ```

- It has properties, e.g. length:

    ```
    >>> len(text)
    24
    ```

- It can be acted upon, e.g. printed:

    ```
    >>> print text
    My cat plays with string
    ```

- It can interact with other objects:

    ```
    print text + " and mice"
    My cat plays with string and mice
    ```
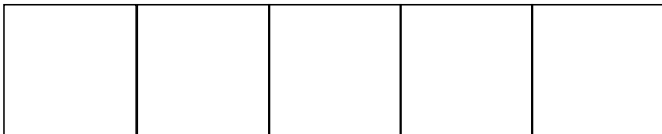
Compound types

Compound types allow arbitrary collections of objects to be held together. The two major compound types are:

- **Lists**: Items are stored in order and are referenced (*indexed*) by an integer.

- **Dictionaries:** Items have no implicit order and can be indexed by a variety of types (commonly strings)
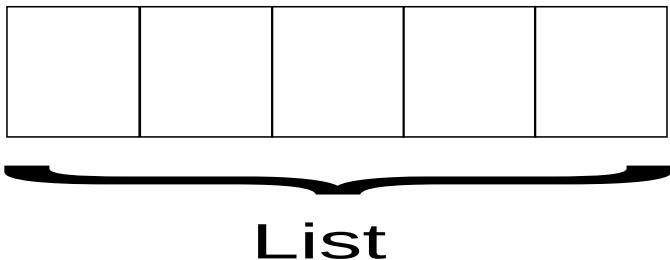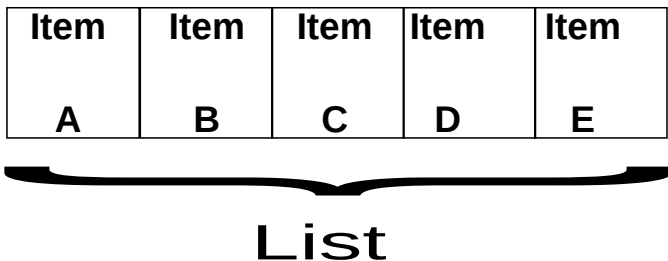
Lists hold collections of objects in order:

Compound types - lists

Lists hold collections of objects in order:

List

Lists hold collections of objects in order:

Item	**Item**	**Item**	**Item**	**Item**
A	**B**	**C**	**D**	**E**

List

Lists hold collections of objects in order:

Index

0	1	2	3	4
Item	**Item**	**Item**	**Item**	**Item**
A	**B**	**C**	**D**	**E**

List

Lists hold collections of objects in order:

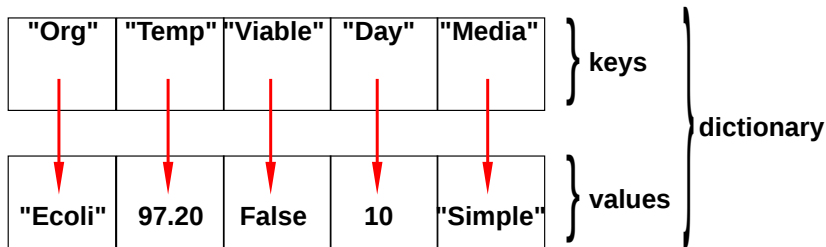| **Index**	**-5**	**-4**	**-3**	**-2**	**-1**
	Item A	**Item B**	**Item C**	**Item D**	**Item E**

List

Compound types - lists

Example:

```
>>> ExampleList = [ "A" , "B" , "C" , "D" , "E" ]
>>> ExampleList [ 0 ]
'A'
>>> ExampleList [ 1 ]
'B'
>>> ExampleList [ 4 ]
'E'
>>> ExampleList[ −1 ]
'E'
>>> ExampleList[ −5 ]
'A'
```

Compound types - dictionaries

Similar in concept to lists, but items held as *key/value* pairs, are not ordered, and key types are not restricted to integer.

Compound types - dictionaries

Creating a dictionary:

```
>>> ExampleDict={"Org":"Ecoli",
    "Temp":97.2,
    "Viable":False,
    "Day":10,
    "Media":"Simple"
}
>>> print ExampleDict
{'Media': 'Simple',
 'Org': 'Ecoli',
 'Viable': False,
 'Temp':97.2,
 'Day': 10
}
```

Compound types - dictionaries

Changing existing values in a dictionary:

```
>>> ExampleDict["Media"] = "Complex"
>>> ExampleDict["Temp"] = 30
>>> ExampleDict["Viable"] = True
>>> print ExampleDict
{'Media': 'Complex',
 'Org': 'Ecoli',
 'Viable': True,
 'Temp': 30,
 'Day': 10
 }
```

Compound types - dictionaries

Adding new key/value pairs to a dictionary:

```
>>> ExampleDict["Recorded_by"] = "Mark"
>>> print ExampleDict
{ 'Media': 'Complex',
  'Org': 'Ecoli',
  'Viable': True,
  'Temp': 30,
  'Recorded_by': 'Mark',
  'Day': 10
  }
```

Functions in Python

Functions behave in the same way as class methods, although they are not an attribute of any particular class.

dir() list the *attributes* of an object.

type() returns the class of an object.

len() returns the length of an object (if that is meaningful)

Functions in Python - Examples

```
>>> L = [1,2,3,4]

>>> dir(L)
['__add__', '__class__', '__contains__', '__delattr__', '__delitem__',
.
.
'append', 'count', 'extend', 'index', 'insert', 'pop',
'remove', 'reverse', 'sort']

>>> type(L)
<type 'list'>

>>> len(L)
4
>>>
```

Modules are used to store pre-written python code for later re-use. They must be *imported* in order to be used:

```
>>> import math
>>> dir(math)
[...,
'pi',...
'sqrt'...]
```

Modules can then be accessed with dot notation:

```
>>> print math.pi
3.14159265359
>>> print math.sqrt(2)
1.41421356237
```

Alternatively selection of items can be imported instead:

```
>>> from math import pi , sin
>>> print sin ( pi /4)
0.707106781187
```

For loops (other loops are available)

We frequently wish to act upon each item in a list in turn. The **for** loop provides a convenient way of doing this.

In general:

```
for Item in MyList :
    # do something
```

Example:

```
>>> for letter in ExampleList :
    print letter
C
B
A
E
D
```

For loops provide a convenient way of scanning across a range of numbers, using, for example the built in **range** function:

```
>>> for x in range(10):
      print x, x**2, x**3
0 0 0
1 1 1
2 4 8
3 9 27
4 16 64
5 25 125
6 36 216
7 49 343
8 64 512
9 81 729
```

We have now covered enough fundamentals to
think about how to use it for modelling.

What do we want to represent and act upon ?

A collection of modules (a **package**) providing the ability to define and analyse models.

Everything revolves around the use of model objects:

```
>>> m = ScrumPy.Model("FileName.spy")
```

Where "FileName.spy" is the name of file describing the model.

The ".spy" extension is conventional and convenient, but not mandatory.

and "m" is the model object. In these talks, "m" will always be used to denote the model.

Model Definition

In ScrumPy, a model is defined by one or more text files, defining:

Comments   Ignored by ScrumPy, but are useful to the human reader.

Directives   Not part of the model *per se*, but specify how the model is to be read.

Reactions   Define the metabolic network.

Initialisations   Define parameter values and initial metabolite concentrations (only in kinetic models)

Model Definition

```
# comment, everything from #
# to the end of the line is ignored

Structural()
# a Directive. Do not do any kinetic processing.

Rubisco:                                    # a reaction name
    x_CO2 + RuBP_ch -> 2 PGA_ch             # stoichiometry
    ~                                       # default kinetic

PGK:
    PGA_ch + ATP_ch <> BPGA_ch + ADP_ch
    ~

G3Pdh:
    BPGA_ch + x_NADPH_ch + x_Proton_ch <>
                        x_NADP_ch + GAP_ch + Pi_ch
    ~
```

Identifiers = Names

Either:
Any sequence of alphanumeric characters and _ (underscore), not starting with a number e.g.
Valid:

```
Fructose6_Phosphate
AlphaAnaline
```

Invalid:

```
2,3−bisphosphoglycerate
TRANS−23−DEHYDROADIPYL−COA
```

Or:
Any quoted (") sequence of characters.

```
"Saturated−Fatty−Acyl−CoA"
"3−oxo−cis−vaccenoyl−ACPs"
```

Accessed as m.sm (internal) and m.smx (external):

	Ru5Pk	Aldo2	TPT_DHAP	Light_react	TKL
RuBP_ch	1/1	0/1	0/1	0/1	0/1
ATP_ch	-1/1	0/1	0/1	1/1	0/1
ADP_ch	1/1	0/1	0/1	-1/1	0/1
GAP_ch	0/1	0/1	0/1	0/1	-1/1
Pi_ch	0/1	0/1	1/1	-1/1	0/1
DHAP_ch	0/1	-1/1	-1/1	0/1	0/1
F6P_ch	0/1	0/1	0/1	0/1	-1/1
E4P_ch	0/1	-1/1	0/1	0/1	1/1
X5P_ch	0/1	0/1	0/1	0/1	1/1
SBP_ch	0/1	1/1	0/1	0/1	0/1
Ru5P_ch	-1/1	0/1	0/1	0/1	0/1

By default values in the stoichiometry matrices are *rational* numbers (ie fractions).

They can are represented as (e.g) 1/2 or mpq(1,2).

This can be changed with the ElType() directive (earlier slide).

For large (genome scale) models it is more common to use *real* numbers
(ElType(**float**))

- Stoichiometry matrices behave as a list of rows:

  ```
  >>> print m.sm[0]
  [mpq(1,1), mpq(0,1), mpq(0,1), mpq(0,1), mpq(0,1), ...]
  ```

- Or as a dictionary of rows:

  ```
  >>> print m.sm["RuBP_ch"]
  [mpq(1,1), mpq(0,1), mpq(0,1), mpq(0,1), mpq(0,1), ...]
  ```

- Individual elements can be accessed as matrix[row,col]:

  ```
  >>> print m.sm[0,0]
  1/1
  >>> print m.sm["RuBP_ch","Ru5Pk"]
  1/1
  ```

The null-space is obtained the matrix.NullSpace() method:

```
>>> k = m.sm.NullSpace()
>>> print k
```

	c_0	c_1	c_2	c_3	c_4
Ru5Pk	0/1	0/1	0/1	0/1	-3/1
Aldo2	0/1	0/1	0/1	0/1	-1/1
TPT_DHAP	0/1	2/1	1/1	1/1	-1/1
Light_react	-1/1	-1/1	0/1	1/1	-9/1
TKL	0/1	0/1	0/1	0/1	-1/1
G3Pdh	0/1	0/1	0/1	1/1	-6/1
PGK	0/1	0/1	0/1	1/1	-6/1
TPI	0/1	1/1	1/1	1/1	-3/1
TKL2	0/1	0/1	0/1	0/1	-1/1

.
.
.

We have covered enough to start the practical.