

Computer class

For Lecture 11

Aims

- Sentiment analysis



Predict sentiment from text.

Twitter API key

KEY:

nU02XOBxxuWuvHJiJ42MR6bsW

SECRET:

NFkMtuu9XbekfkbvbW5olzK2QeiVpXCDLM5YWPFYmLYdbHpBXy

Sentiment Analysis

Method

Liu Hu

Vader

Report

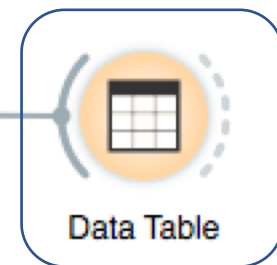
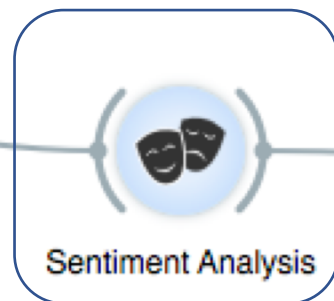
Autocommit is on

	Author	Content True	pos	neg	neu	compound
1	@uz_today	The tradition...	0.000	0.000	1.000	0.000
2	@DataPortal...	Work in full s...	0.000	0.000	1.000	0.000
3	@dylanbeattie	@toddhgard...	0.000	0.091	0.909	-0.273
4	@deanwithay	"Innovation	0.245	0.000	0.755	0.382
5			0.058	0.228	0.714	-0.586
6	@startupcru...	#intel intel C...	0.148	0.000	0.852	0.382
7	@startupcru...	#intel Intel C...	0.140	0.000	0.860	0.382
8	@startupcru...	#intel Intel : ...	0.140	0.000	0.860	0.382
9	@DigitalBan...	China EU pl...	0.357	0.000	0.643	0.612
10	@Nekoterran	Sel		0.000	1.000	0.000
11	@AdamRoge...	Incr		0.000	0.517	0.813
12	@ChrisWilso...	Wa		0.000	1.000	0.000
13	@Buildings_...	@LookersGr...	0.000	0.000	1.000	0.000
14	@MyCloudE...	why Innovati...	0.4		0.8	0.902
15	@MarkMiller1...	The best Cl...	0.2		0.6	0.778
16	@sblackumb...	Human Intuit...	0.2		0.4	0.382
17	@setsquared	Take a listen...	0.12	0.000	0.880	0.382
18	@NairVentur...	How the stu...	0.000	0.000	1.000	0.000
19	@i2CAT	#5G: To go f...	0.000	0.000	1.000	0.000
20	@corcega	#eTconf17 w...	0.178	0.000	0.82	
21	@Stevieboy1...	Maybe The ...	0.405	0.000	0.59	
22	@Gsabiolab	Dundee worl...	0.342	0.000	0.658	0.382
Overall						0.382

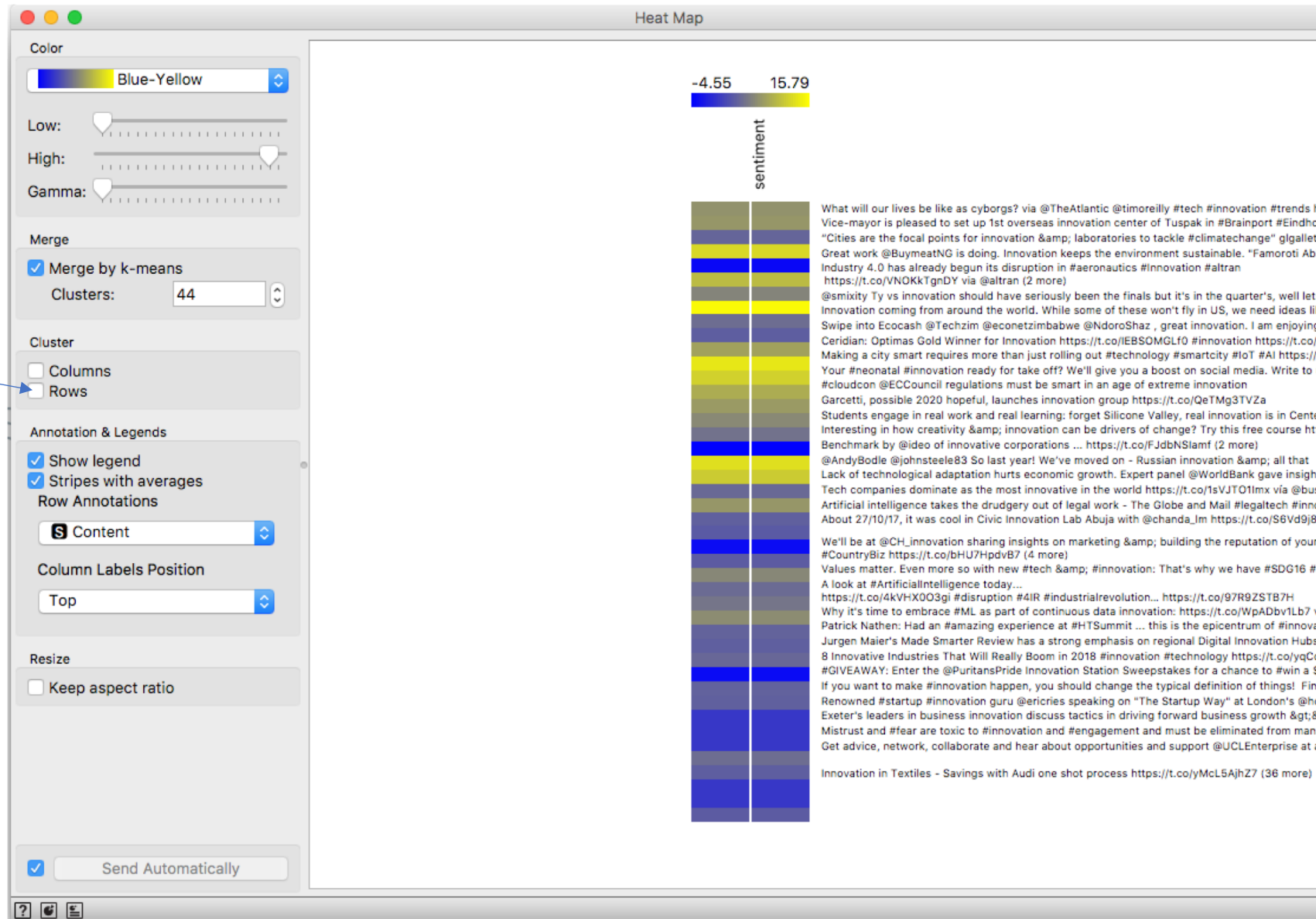
Any Corpus Data

Sentiment analysis widget

Results



Another way of visualizing – connecting to heat maps and applying k-means algorithm



Will group/sort
the result

Coding the data

Raw data can be supplied on general in two formats. As:

Adjacency matrix

	__unit2__	Martin	Peter	Beer	The	Jane	Jason
1	Martin	3	1	2	1	2	1
2	Peter	1	1	1	0	0	0
3	Beer	2	1	2	1	1	0
4	The	1	0	1	1	1	0
5	Jane	2	0	1	1	2	1
6	Jason	1	0	0	0	1	1

Edge List

In	Out	Strength
Martin	Peter	1
Martin	Beer	2
Jane	Jason	1

Creating Adjacency matrix from text

- Splitting the text to segments
- Searching for nodes within each segment
- If nodes are mentioned in the same segment – they have connection

REGEX for exercise

<https://regex101.com/r/Af2Zpd/2>

`^(CHAPTER\s\d+).*?(?=CHAPTER|End of Project)`

<https://regex101.com/r/wT8rE9/9>

`(?<!\.\s)\b[A-Z][a-z]*\b`

1.

Text Field

CHAPTER
Martin is goign to the forest and meeting
Peter, and after they are going for a Beer.

CHAPTER
The Beer was really good and tasty so
Martin decided to continue with it, and met
Jane

CHAPTER
Jane dumped Martin at the bar and went
with Jason for after party.



Send automatically

✓ 1 segment (266 characters) sent to output.

Segment

✓ Advanced settings

Regexes

(s) ^\bCHAPTER\b {type:

Move Up

Move Down

Remove

Clear All

Export List

Import List

Mode:

Split

Regex:

^\bCHAPTER\b

Annotation key:

type

Annotation value:

chapter

☐ Ignore case (i)

☒ Unicode dependent (u)

☒ Multiline (m)

☐ Dot matches all (s)

Add

Segment (1)

✓ Advanced settings

Regexes

(t) \b[A-Z].+?\b {type:

Move Up

Move Down

Remove

Clear All

Export List

Import List

Mode:

Tokenize

Regex:

\b[A-Z].+?\b

Annotation key:

type

Annotation value:

name

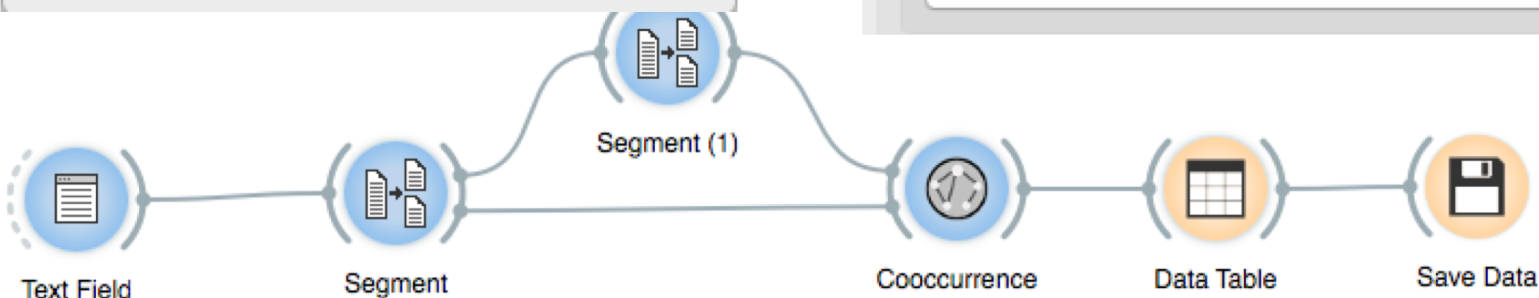
☐ Ignore case (i)

☒ Unicode dependent (u)

☐ Multiline (m)

☐ Dot matches all (s)

Add



2.

Cooccurrence

Units

Segmentation:

Annotation key:

Sequence length:

Intra-sequence delimiter:

Secondary units

☐ Use secondary units

Segmentation:

Annotation key:

Contexts

Mode:

Segmentation:

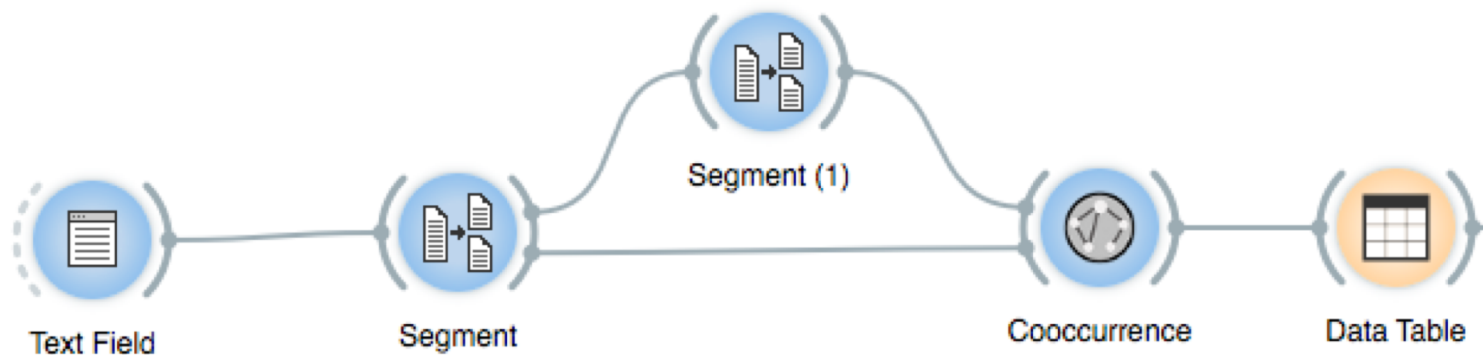
Annotation key:

☐

✔ Table with 34 cooccurrences sent to output.

Data Table

	__unit2__	Martin	Peter	Beer	The	Jane	Jason
1	Martin	3	1	2	1	2	1
2	Peter	1	1	1	0	0	0
3	Beer	2	1	2	1	1	0
4	The	1	0	1	1	1	0
5	Jane	2	0	1	1	2	1
6	Jason	1	0	0	0	1	1



Converting Adjacency matrix to Distance matrix

Orange add-on has no widget for creating network from Adjacency matrix. Therefore we need to create pseudo widget for this purpose.

The screenshot displays the Orange3 data mining software interface. On the left, the 'Data' widget palette is visible, containing various data processing and analysis widgets. A workflow is constructed in the center, starting with a 'Text Field' widget connected to a 'Segment' widget, which then connects to 'Segment (1)'. 'Segment (1)' is connected to a 'Cooccurrence' widget, which outputs to a 'Data Table' widget. The 'Data Table' widget is connected to a 'Python Script' widget. Below the 'Python Script' widget, a 'Network from Distances' widget is connected to a 'Network Explorer' widget. A callout box points to the 'Python Script' widget with the text: 'Paste python code here and click "Run"'. The 'Python Script' widget's interface is shown on the right, with a text area containing the following Python code:

```
import numpy as np
from Orange.misc import DistMatrix

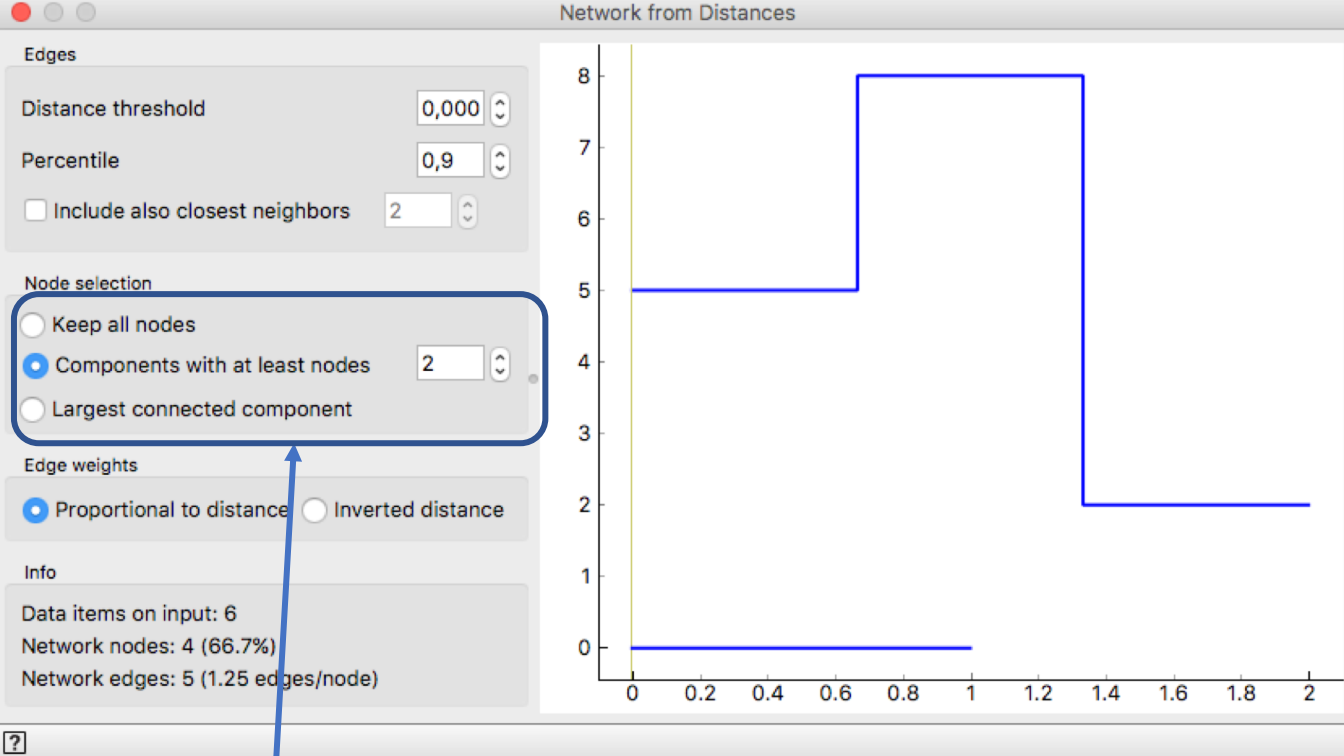
data = np.asarray(in_data) # convert in_data to numpy array
data = data[:, 1:] # strip first column from numeric data (unit2_)
data = data.astype(float) # convert to float, ensures only numbers present

out_object = DistMatrix(data, in_data)
```

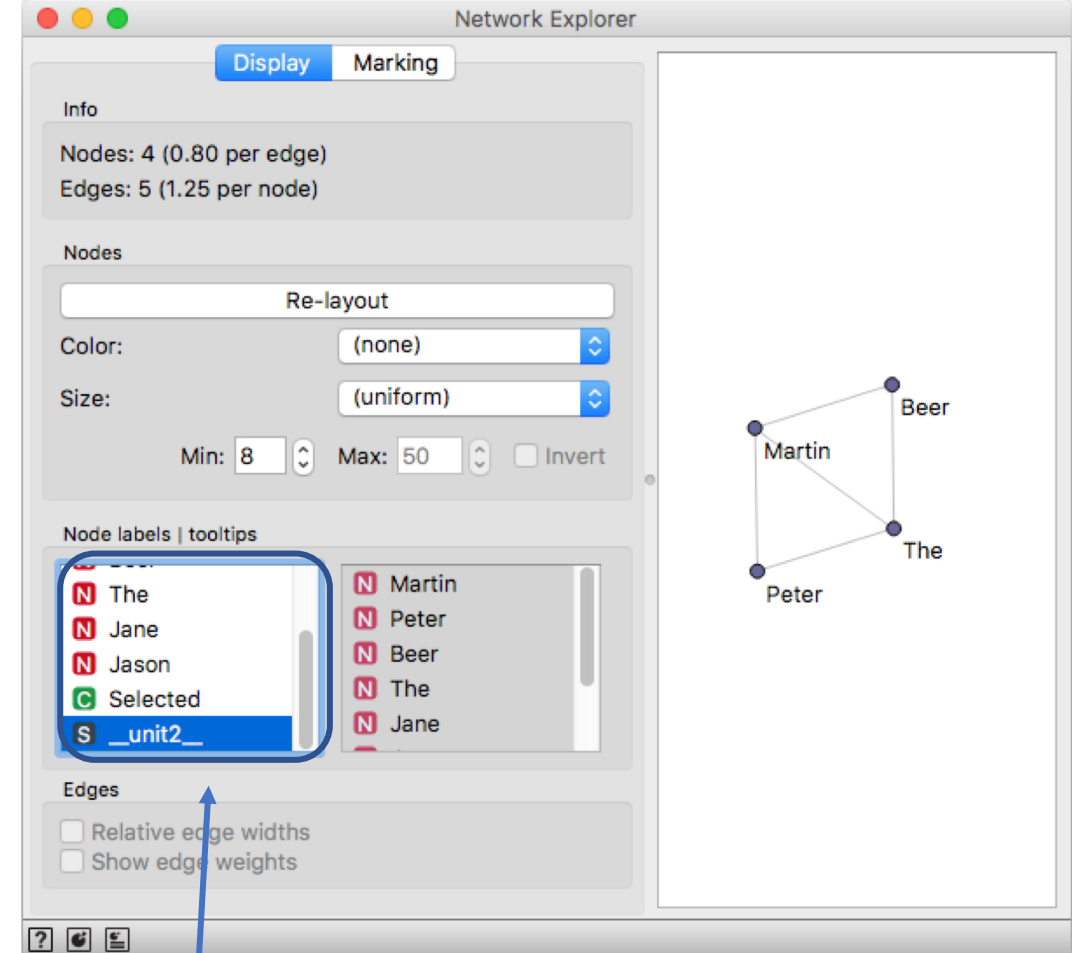
Below the code area, the 'Output variables:' section lists: out_data, out_learner, out_classifier, and out_object. The 'Library' section shows 'Hello world'. The 'Console' section displays the Python version (3.6.1) and the output of the script: 'Hello world'. The 'Run' button is highlighted at the bottom of the 'Python Script' widget.

Python code for Python script

```
import numpy as np
from Orange.misc import DistMatrix
data = np.asarray(in_data) # convert in_data to numpy array
data = data[:, 1:] # strip first column from numeric data (__unit2__)
data = data.astype(float) # convert to float
out_object = DistMatrix(data, in_data)
```

Which connections to leave



Selecting label for nodes