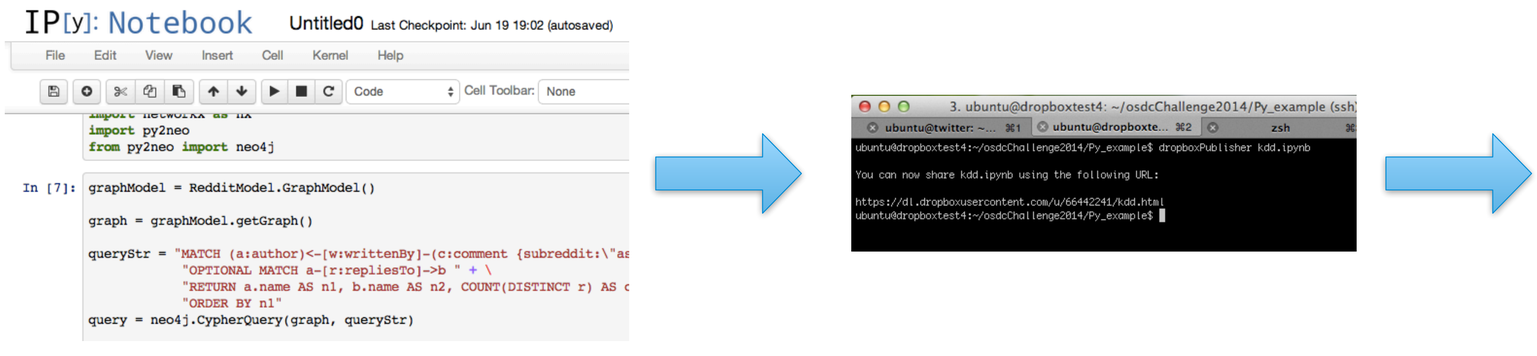
IP[y]: Notebook
Untitled0
Last Checkpoint: Jun 19 19:02 (autosaved)
File Edit View Insert Cell Kernel Help
Code
Cell Toolbar: None
import networkx as nx
import py2neo
from py2neo import neo4j
In [7]:
graphModel = RedditModel.GraphModel()
graph = graphModel.getGraph()
queryStr = "MATCH (a:author)<-[w:writtenBy]-(c:comment {subreddit:\"as
"OPTIONAL MATCH a-[r:repliesTo]->b " + \
"RETURN a.name AS n1, b.name AS n2, COUNT(DISTINCT r) AS c
"ORDER BY n1"
query = neo4j.CypherQuery(graph, queryStr)

3. ubuntu@dropboxtest4: ~/osdcChallenge2014/Py_example (ssh)
ubuntu@twitter: ~... ⌘1
ubuntu@dropboxte... ⌘2
zsh
ubuntu@dropboxtest4:~/osdcChallenge2014/Py_example$ dropboxPublisher kdd.ipynb
You can now share kdd.ipynb using the following URL:
https://dl.dropboxusercontent.com/u/66442241/kdd.html
ubuntu@dropboxtest4:~/osdcChallenge2014/Py_example$


kdd
dl.dropboxusercontent.com/u/66442241/kdd.html
Reader
Google News
In [2]:
import json
# Event files for the Derby and Belmont Stakes
eventFiles = {"Kentucky Derby":"2014-05-03_kentucky_derby", "Belmont Stakes":"2014-06-07_belmont_stakes"}
eventData = {}
eventPath = "../events"
for event in eventFiles.keys():
eventFileName = eventPath + "/" + eventFiles[event] + ".json"
with open(eventFileName, 'r') as f:
eventData[event] = json.load(f)
#print json.dumps(eventData, sort_keys=True, indent=4, separators=(',', ': '))
print "Events:", eventData.keys()
Events: ['Kentucky Derby', 'Belmont Stakes']
In [5]:
from datetime import datetime
import pytz
import time
# Now build a map of windows to events from the input event files
eventTimeMap = {}
for eventName in eventData.keys():
eventList = eventData[eventName]["datasets"][0]["events"]
eventTimeMap[eventName] = {}
for event in eventList:
timeStr = None
if ("EDT" in event["time"]):
timeObj = datetime.strptime(event["time"][:-3], "%Y-%m-%d %H:%M:%S")