

```

In [21]: import networkx as nx
import matplotlib.pyplot as plt
import numpy as np
import urllib

In [5]: fill=urllib.request.urlretrieve("http://www.stats.ox.ac.uk/~reinert/dtc/
polblogs.gml")

In [6]: polblog=networkx.Graph(networkx.read_gml(fill[0]))

In [7]: polblog.remove_edges_from(polblog.selfloop_edges())

In [8]: graph = polblog

In [148]: Nodes=list(graph.nodes())
len(Nodes)

Out[148]: 1490

In [177]: #induced random sampling
rs = list(np.random.choice(Nodes, 300))

In [178]: H = graph.subgraph(rs)

In [180]: giant_comp = [node for (node, degree) in H.degree() if degree<3]

In [181]: #Need to create a new graph object to "unfreeze graph"
frozen_graph = nx.freeze(H)
H_new = nx.Graph(frozen_graph)
nx.is_frozen(H_new)

Out[181]: False

In [182]: #remove giant component
H_new.remove_nodes_from(giant_comp)

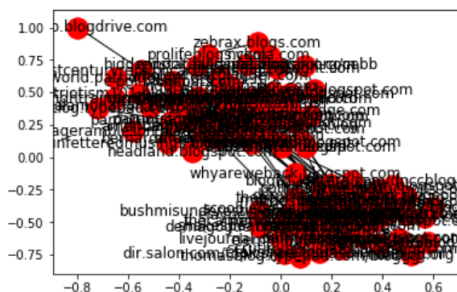
In [292]: H_new.degree()

Out[292]: DegreeView({'pardonmyenglish.com': 6, 'commonsenserunswild.typepad.co
m': 17, 'politicalwire.com': 18, 'democrats.org/blog': 13, 'patriotism1
01.blogspot.com': 3, 'thomasgalvin.blogspot.com': 16, 'everythingiknowi
swrong.com': 7, 'mdcbowen.org/cobb': 4, 'gmscorner.blogspot.com': 5, 'h
igherpieproductions.com': 7, 'alamonation.blogspot.com': 5, 'bushmisund
erestimated.blogspot.com': 3, 'txfx.net': 7, 'fairshot.org': 8, 'scruti
nyhooligans.blogspot.com': 6, 'alvintostig.typepad.com': 5, 'msnbc.msn.
com/id/3449870': 17, 'victorhanson.com': 3, 'jeffblanco.blogdrive.com':
1, 'americanfuture.typepad.com': 4, 'theagitator.blogspot.com': 3, 'the
donovan.com': 7, 'littlegreenfootballs.com/weblog': 27, 'powerageramble
son.blogspot.com': 2, 'hiddennook.blogspot.com': 3, 'prestoncaldwellblo
g.blogspot.com': 3, 'nykola.com': 3, 'bensworld.patriotforum.org': 3,
'takebackthemedias.com': 5, 'electionprojection.com': 11, 'robschumache
r.blogspot.com': 13, 'dailywarnews.blogspot.com': 8, 'martinipundit.co
m': 16, 'ndnblog.org': 3, 'atrios.blogspot.com': 44, 'elemming2.blogspo
t.com': 5, 'scoobiedavis.blogspot.com': 10, 'c-pol.blogspot.com': 3, 's
treetlog.typepad.com': 3, 'professorbainbridge.com': 9, 'blog01.kinter
a.com/dnccblog': 7, 'scrappleface.com': 12, 'jadedreality.blogspot.co
m': 4, 'needlenose.com': 5, 'bernardmoon.blogspot.com': 4, 'ospolitics.
org': 10, 'thomasblog.typepad.com/bloggg': 3, 'headland.blogspot.com':
4, 'twentyfirstcenturyrepublican.blogspot.com': 4, 'nomayo.mu.nu': 3,
'alphapatriot.com': 20, 'unfetteredmusings.blogspot.com': 3, 'fromthero
ots.org': 9, 'wholewheatblogger.com': 3, 'wizbangblog.com': 21, 'zebra
x.blogs.com': 3, 'allspinzon.blogspot.com': 4, 'righttrainbow.com': 3,
'thepoorman.net': 16, 'sgtstryker.com': 5, 'nomoremister.blogspot.com':
12, 'stcfoundation.blogspot.com': 5, 'crankybastard.blogspot.com': 4,
'elayneriggs.blogspot.com': 15, 'quante.blogspot.com': 4, 'evangelicalo
utpost.com': 19, 'lt-smash.us': 13, 'prolifeblogs.com': 5, 'bartcop.co
m': 7, 'politicalmusings.net': 8, 'oliverwillis.com': 24, 'mydd.com': 2
0, 'marcbrazeau.blogspot.com': 5, 'demagogue.blogspot.com': 5, 'rogerls
imon.com': 11, 'livejournal.com/users/jmhm': 4, 'therightcoast.blogspo
t.com': 5, 'rightword.blogspot.com': 6, 'whyareweback.blogspot.com': 7,
'mathewgross.com/blog': 7, 'mahablog.com': 14, 'dir.salon.com/topics/jo
e_conason/index.html': 4, 'pudentilla.blogspot.com': 7, 'blogsofwar.co
m': 19, 'lefterer.com': 6, 'dneiwert.blogspot.com': 21, 'thecore4.com':
4, 'shakespeareassister.blogspot.com': 5, 'diplomadic.blogspot.com': 9,
'donaldsensing.com': 8, 'buzzmachine.com': 19, 'freedomofthought.com':
3, 'thelleftcoaster.com': 16, 'etherealgirl.blogspot.com': 2, 'redstatea
lliance.blogspot.com': 3, 'anncoulter.org': 9, 'markarkleiman.com': 12,
'thecarpetbaggerreport.com': 6, 'patriotsforbush.com': 6})

In [186]: nx.write_graphml(H_new, "subgraph.graphml")

```

```
In [184]: networkx.draw_networkx(H_new)
```



```
In [250]: urllib.request.urlretrieve("http://www.stats.ox.ac.uk/~reinert/dtc/commu
nityDetection.py", 'communityDetection.py')
```

```
Out[250]: ('communityDetection.py', <http.client.HTTPMessage at 0xa1983a6a0>)
```

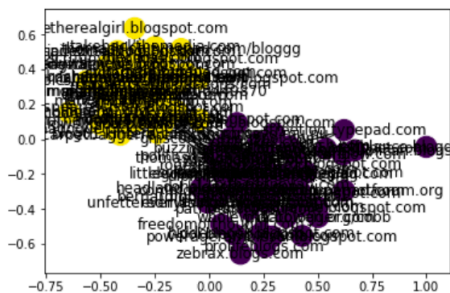
```
In [252]: import communityDetection as comDet
```

```
In [279]: H_Bet = comDet.betweenComDetect(H_new)
```

```
print(comDet.modularity(H_new,H_Bet))
```

```
0.46840129641549394
```

```
In [256]: nx.draw_networkx(H_new,nodelist=H_new.nodes(),node_color=[H_Bet[x] for x
in H_new.nodes()])
```



```
In [333]: import pandas as pd
site = [node for (node, val) in H_new.degree()]
degrees = [val for (node, val) in H_new.degree()]
```

```
In [334]: cluster = pd.DataFrame.from_dict(H_Bet, orient='index', columns=['cluster'])
degree = pd.DataFrame.from_dict(dict(zip(site,degrees)), orient='index',
columns=['degree'])
```

```
In [338]: df = cluster.join(degree)
```

```
In [339]: df.groupby(['cluster']).mean()
```

```
Out[339]:
```

	degree
cluster	
0	7.660377
1	9.391304