

# p1 to bonus 1

## Summaries

Run time is appended to the title of the plot.

We convert all graphs to simple ones by looking only at upper triangular matrix of the adjacency, and make it symmetric.

All networks with node size greater than 5000, we use a uniformly random sampled set (size 5000, or 40000 when graph is sparse) of nodes to approximate clustering, path, component, degree correlation and degree cluster correlation, but not degree and eigen.

For eigen distribution, we only plot top 20 largest eigenvalues. And the eigen gap defined in class is computationally infeasible for even small graph. Thus we used the difference between largest and second largest eigen value as eigen gap.

For BA model, we procedurally generate preferential attachment starting from a fully connected graph. And we don't deal with multiple edges, and allow self loops.

For bonus 1, we modify the growth factor  $m$  to be growing logarithmically to the size of the network.

## Findings

Many degree distribution has a curvy tail on the upper left, indicating not as many low degree nodes as BA model suggests, this might be relevant to the latest paper on why log normal might fit real network better than power law.

BA model approximates degree distribution well, modified BA model creates some outliers for degree distribution, since some nodes have way more added connections than others, thus forming mega hubs.

For clustering, higher degree nodes tend to have lower clustering. BA model cannot approximate clustering well.

For path, BA model approximates the shape of the plot well, but not average length well.

Component analysis is highly biased by subsampling and no good insights can come out.

Top 20 eigen values normally fall into distinct groups, at least 2.

Most degree correlation coefficient is positive.

And degree clustering correlation is negative, as observed in clustering plot.

To improve the results, using random walk sampling or forest fire sampling might be better than uniform node sampling.

## Plots

### collaboration

```
In [1]: datas = [
    ('collaboration', 'datasets/collaboration_in.npz'),
    ('citation', 'datasets/citation_in.npz'),
    ('actor', 'datasets/actor_in.npz'),
    ('email', 'datasets/email_in.npz'),
    ('internet', 'datasets/internet_in.npz'),
    ('metabolic', 'datasets/metabolic_in.npz'),
    ('phonecalls', 'datasets/phonecalls_in.npz'),
    ('powergrid', 'datasets/powergrid_in.npz'),
    ('protein', 'datasets/protein_in.npz'),
    ('www', 'datasets/www_in.npz')
]
metrics = [
    'degree',
    'clustering',
    'path',
    'component',
    'eigen',
    'degreecorr',
    'degreecluster',
]
```

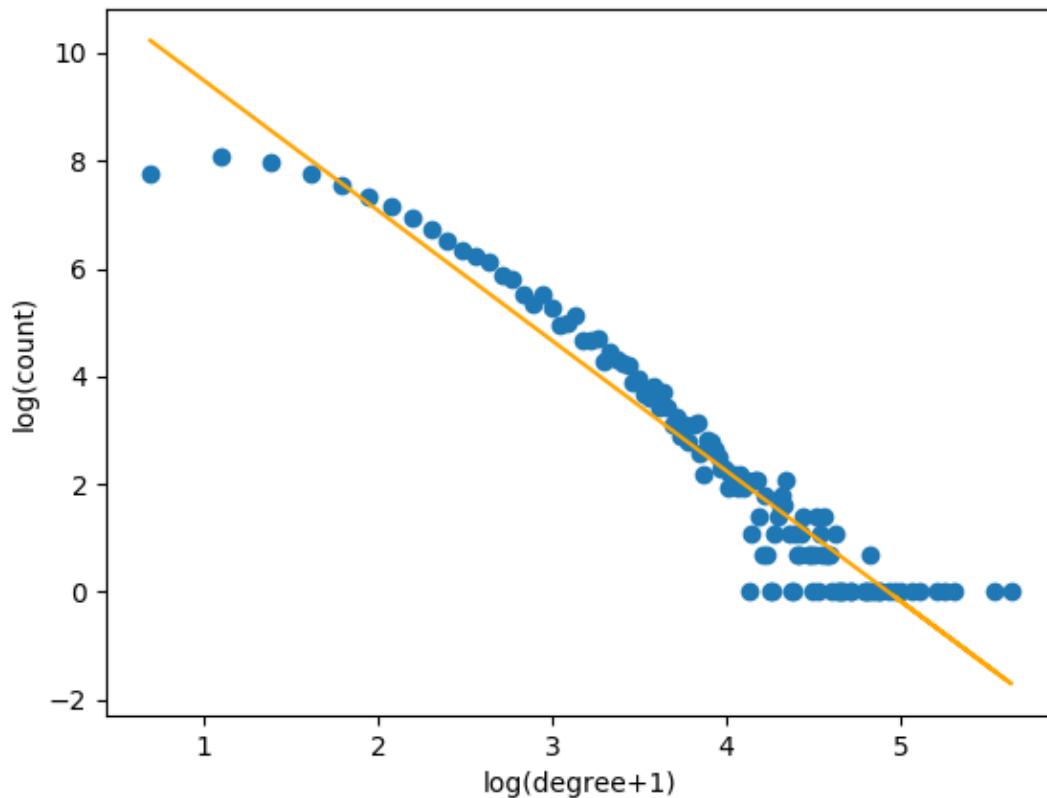
```
In [7]: for metric in metrics:
    print('# {}'.format(metric))
    for data in datas:
        data = data[0]
        print('## {}'.format(data))
        for mod in ['', 'ba_', 'mod_ba_']:
            if mod == '':
                print('### {}'.format('original network'))
            elif mod == 'ba_':
                print('### {}'.format('BA network'))
            else:
                print('### {}'.format('modified BA network'))
    print('![title](results/{}_{}.png)'.format(mod + data, metric))
    print('\n')
```

## degree

### collaboration

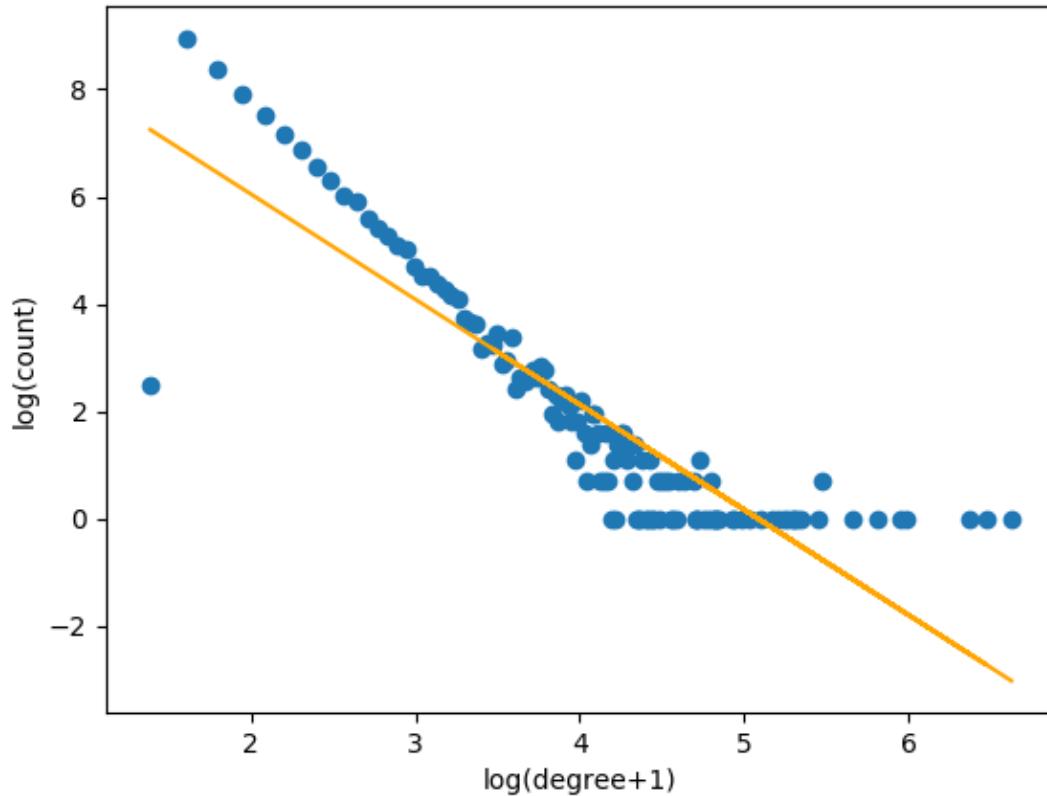
#### original network

degree distribution plot, slope: -2.42, bias: 11.91,  
time: 0.0081



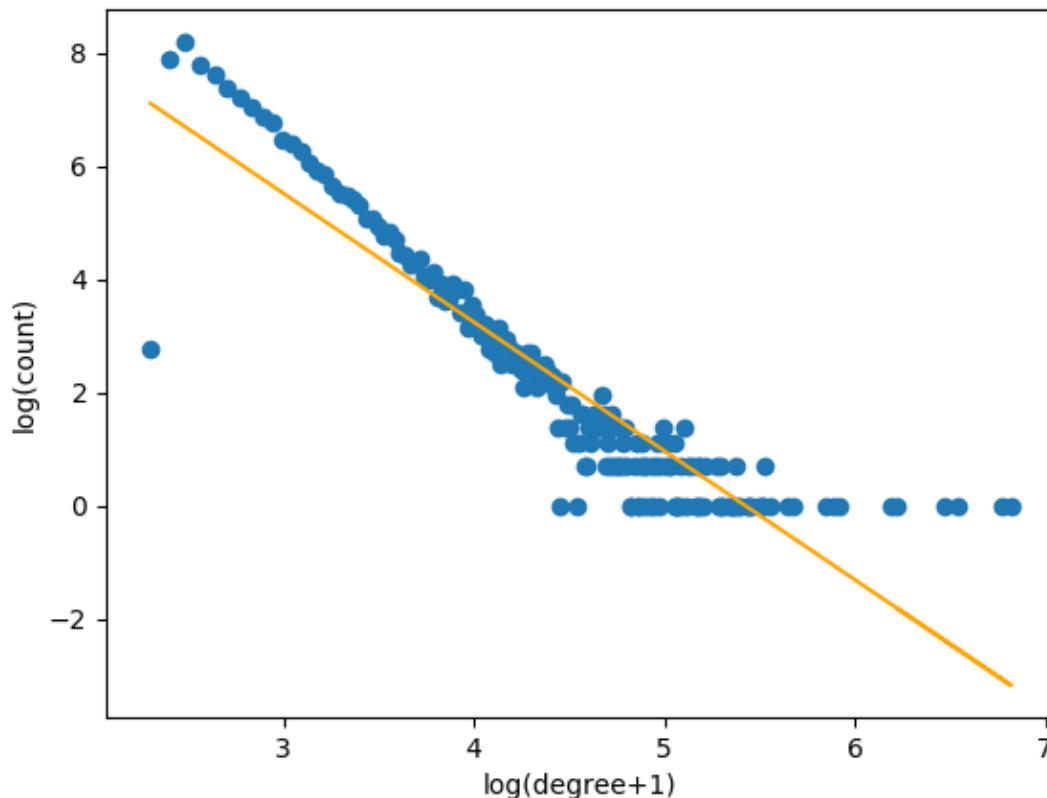
#### BA network

degree distribution plot, slope: -1.96, bias: 9.96,  
time: 0.0071



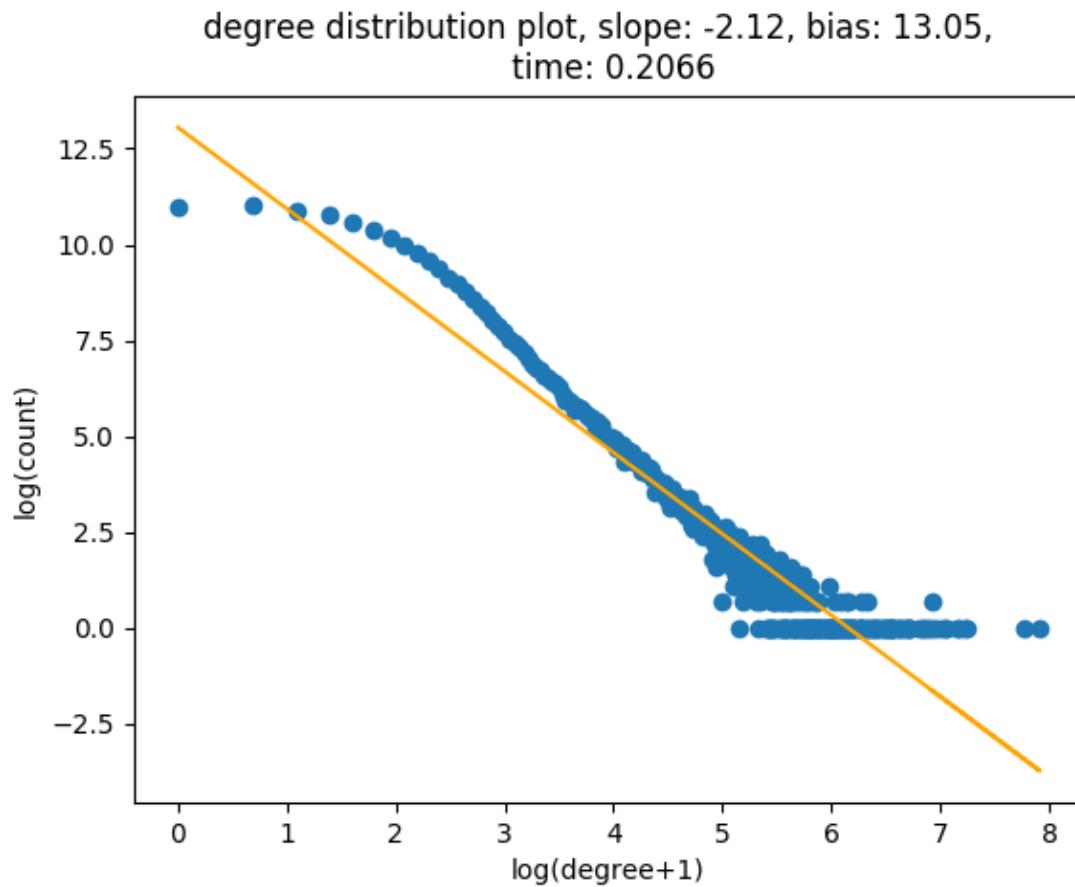
### modified BA network

degree distribution plot, slope: -2.27, bias: 12.34,  
time: 0.0077



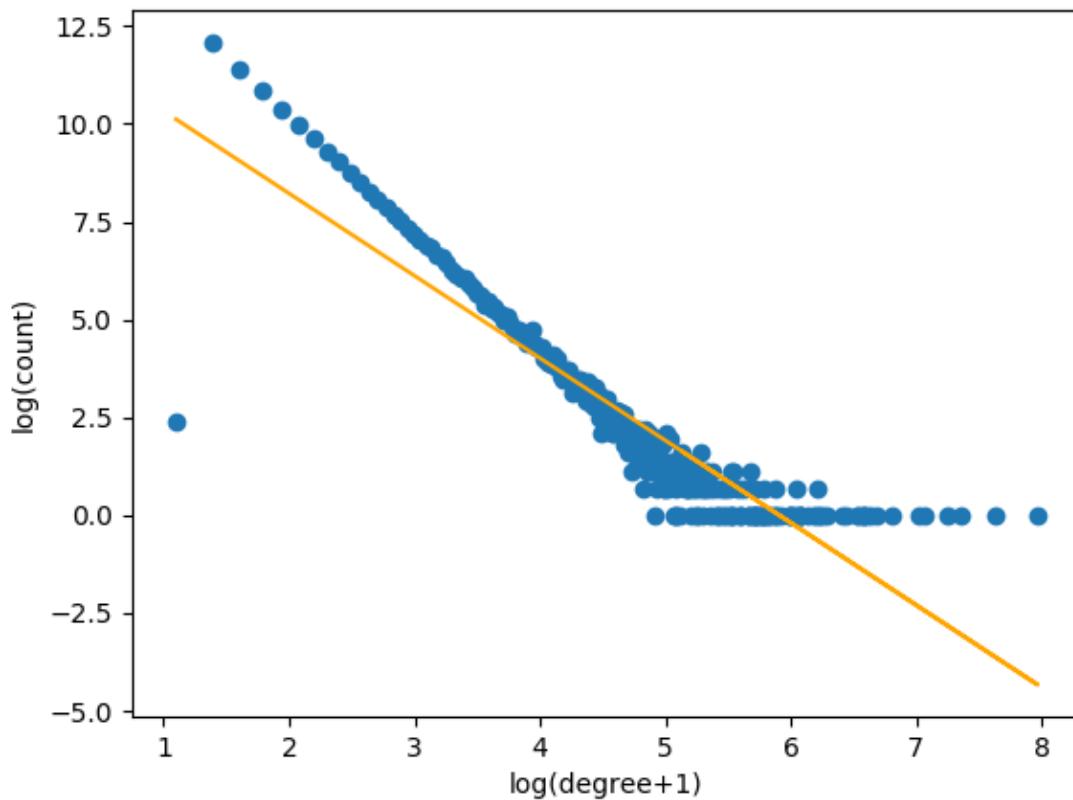
## citation

### original network



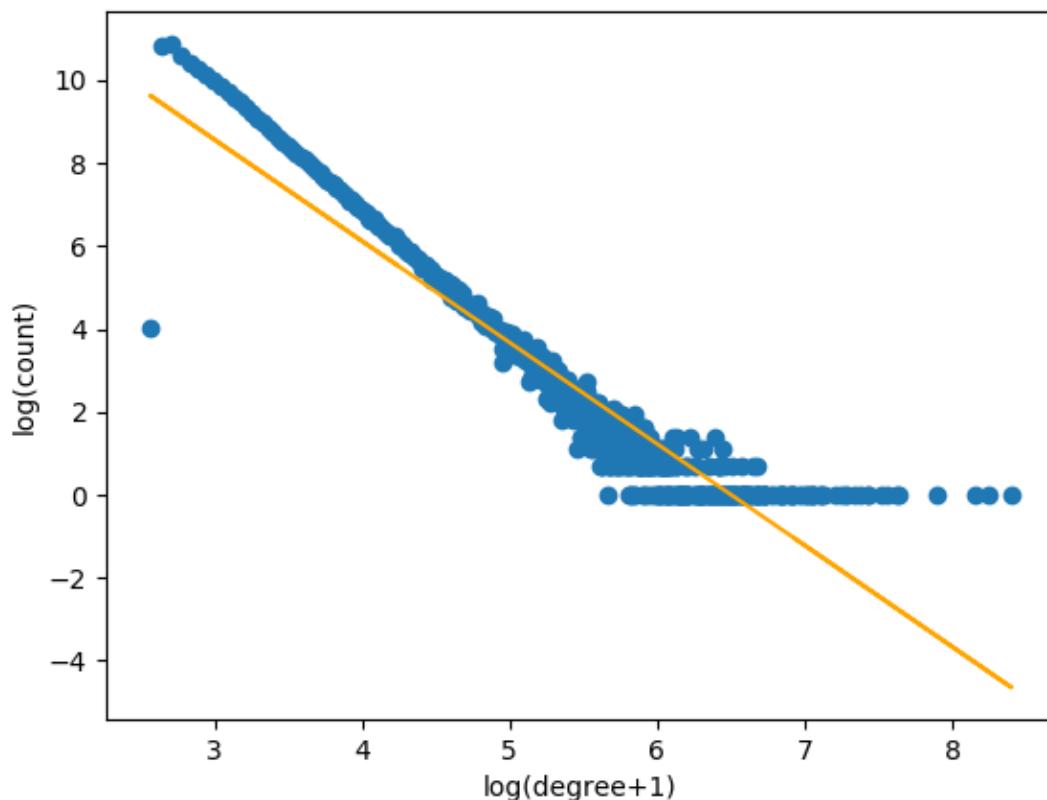
### BA network

degree distribution plot, slope: -2.1, bias: 12.43,  
time: 0.1076



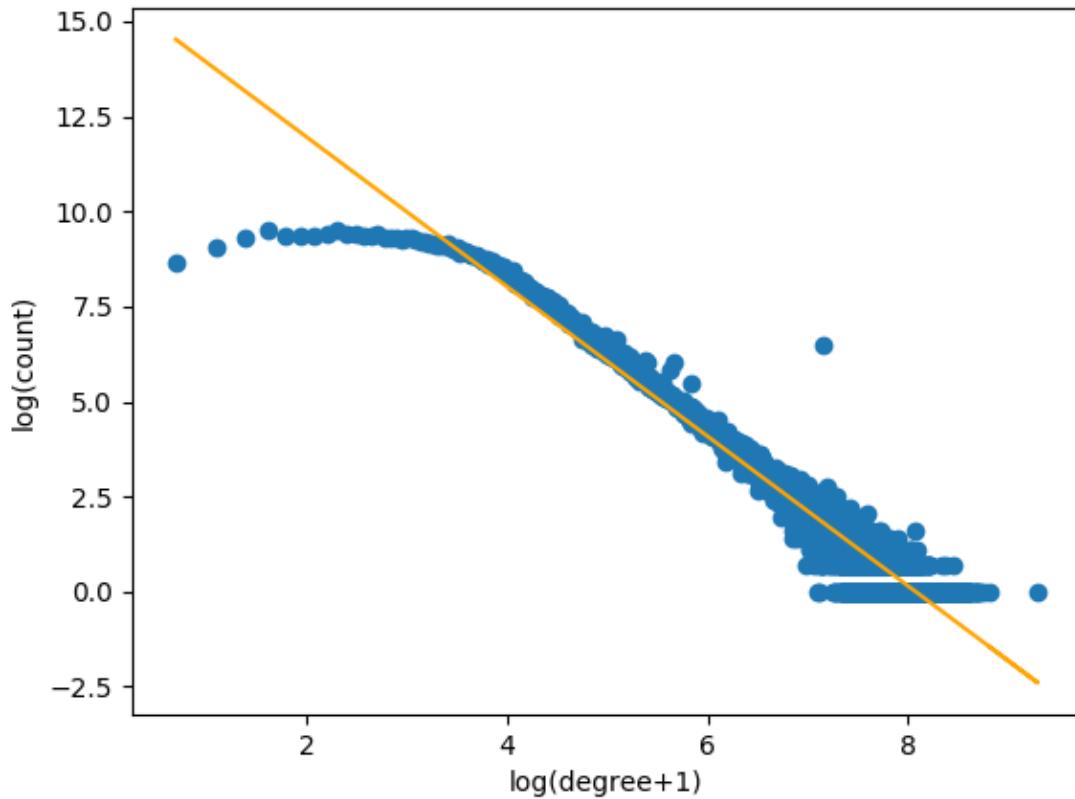
### modified BA network

degree distribution plot, slope: -2.45, bias: 15.9,  
time: 0.1111

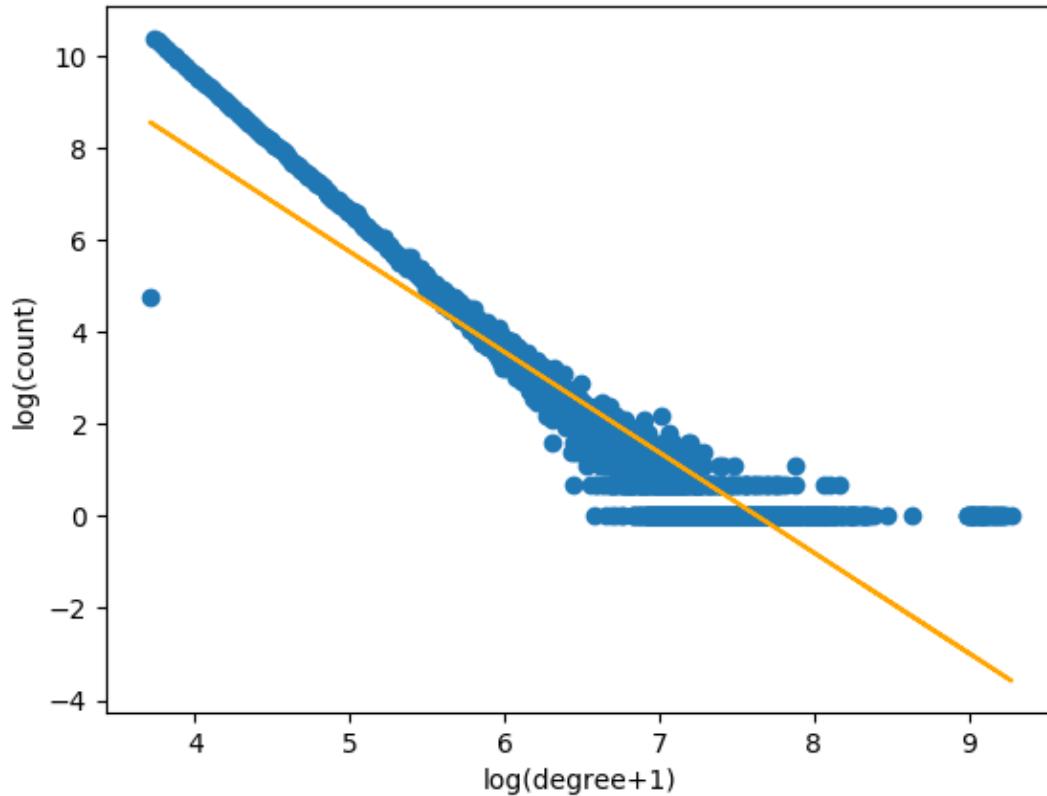


**actor****original network**

degree distribution plot, slope: -1.97, bias: 15.89,  
time: 0.2328

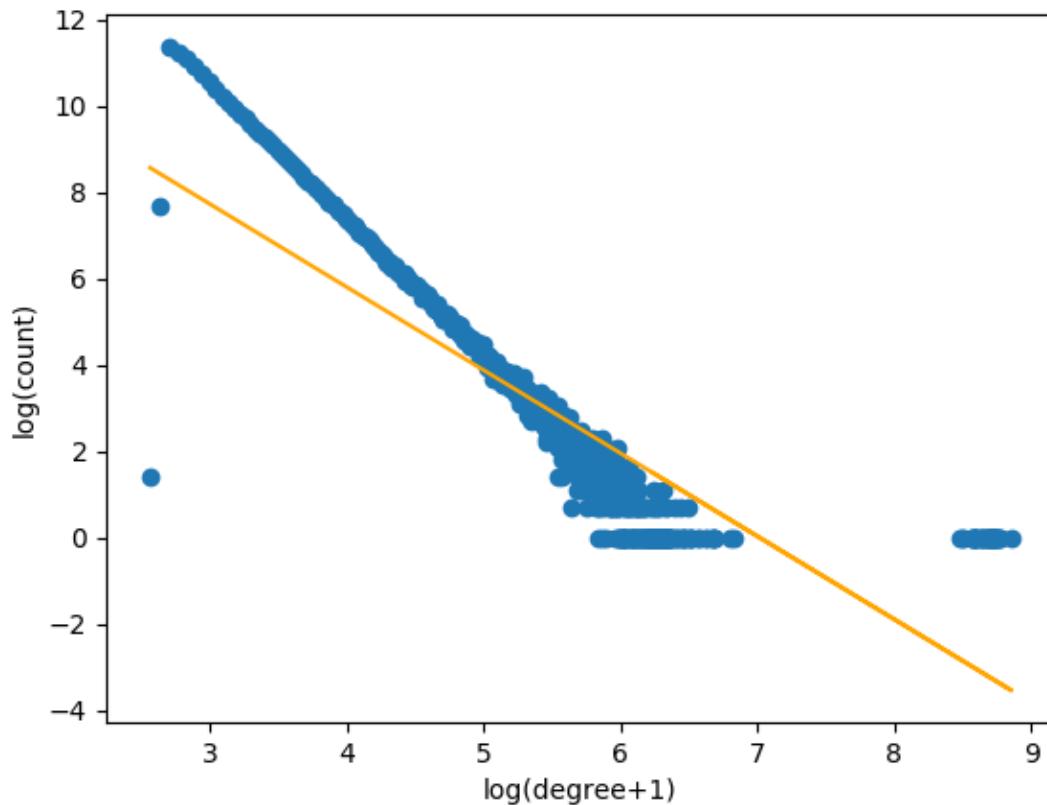
**BA network**

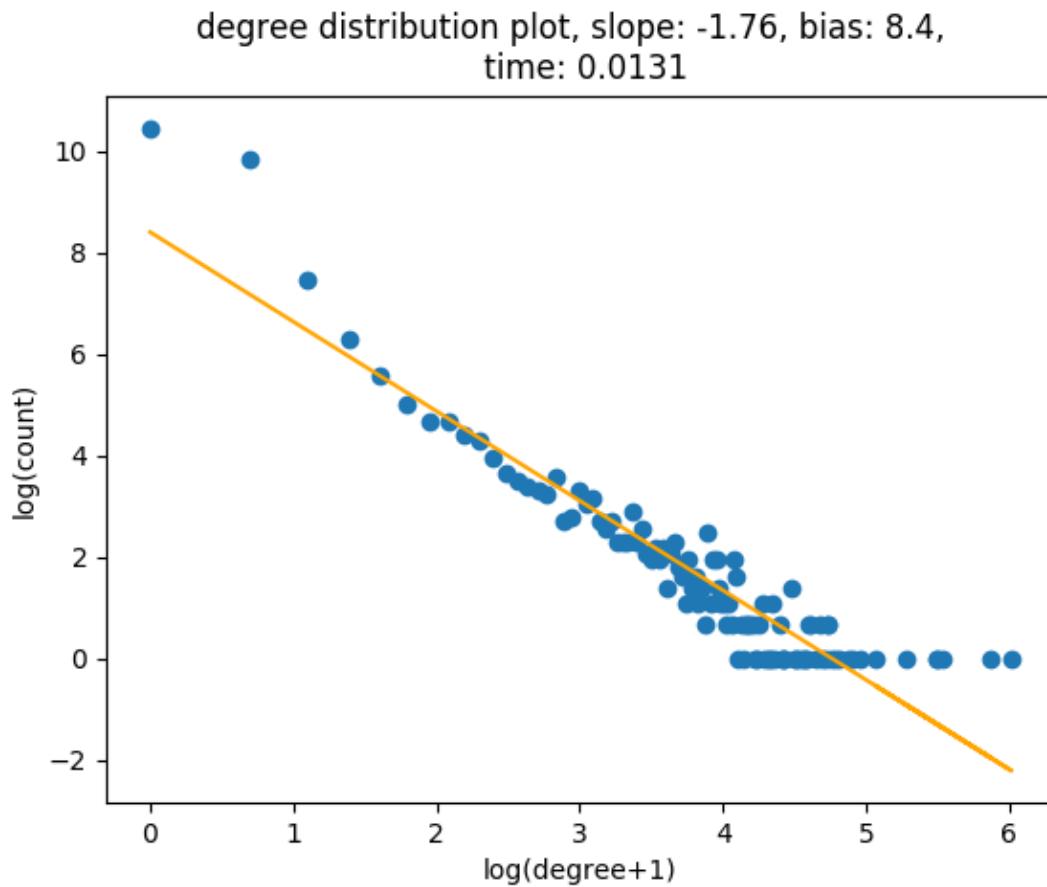
degree distribution plot, slope: -2.18, bias: 16.66,  
time: 0.2135



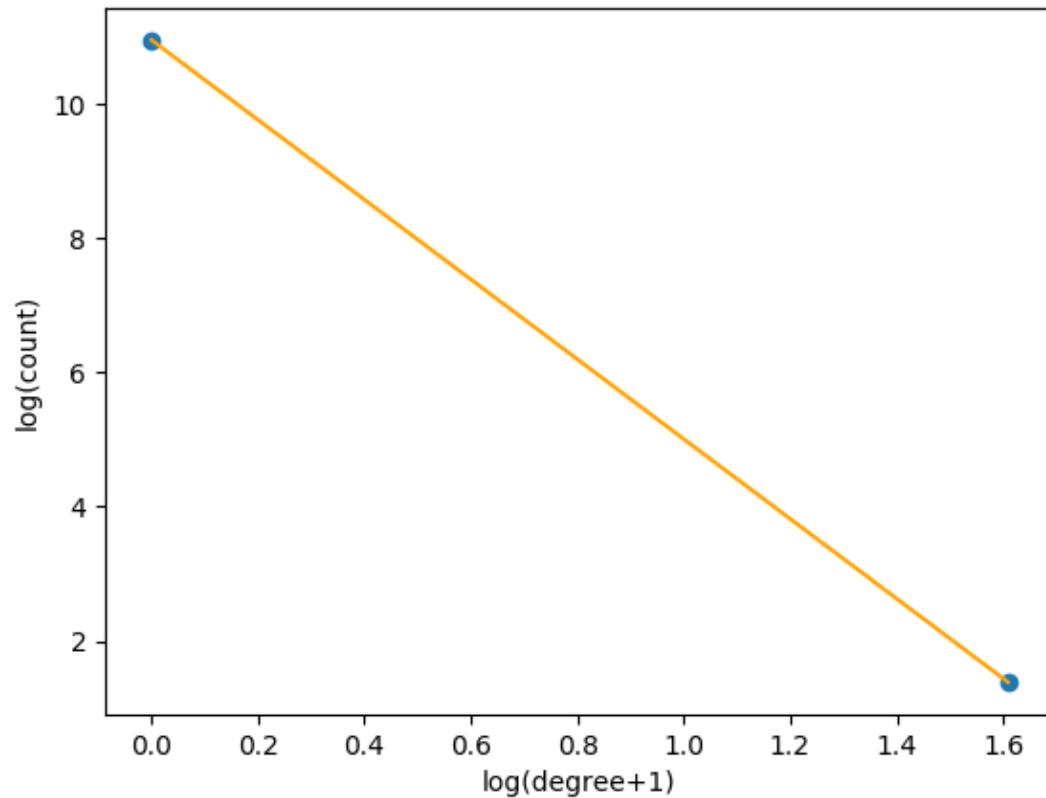
### modified BA network

degree distribution plot, slope: -1.92, bias: 13.5,  
time: 0.1721



**email****original network****BA network**

degree distribution plot, slope: -5.94, bias: 10.95,  
time: 0.0113



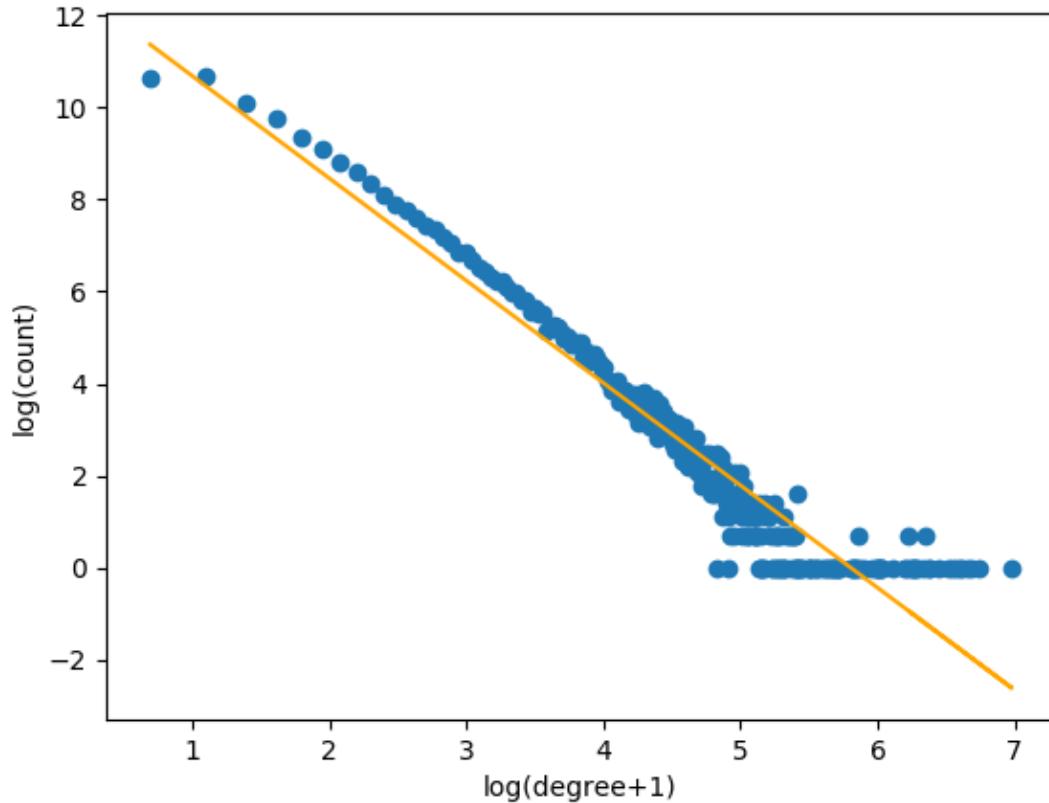
### modified BA network



### internet

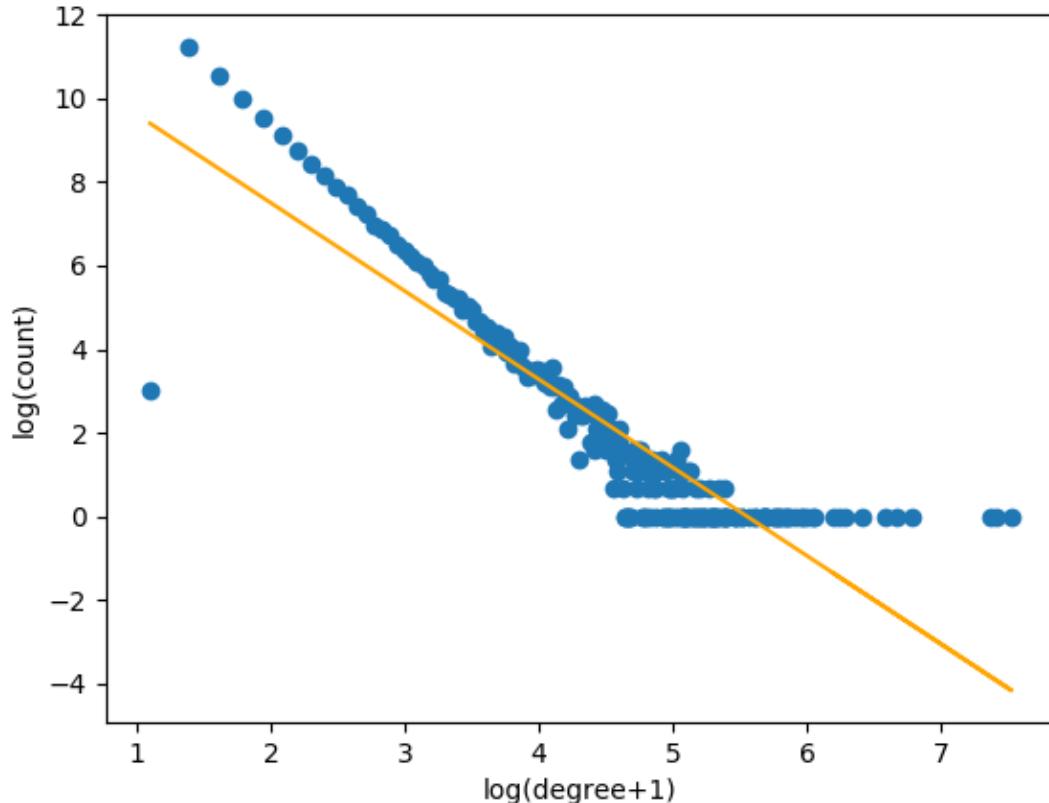
### original network

degree distribution plot, slope: -2.22, bias: 12.9,  
time: 0.0868

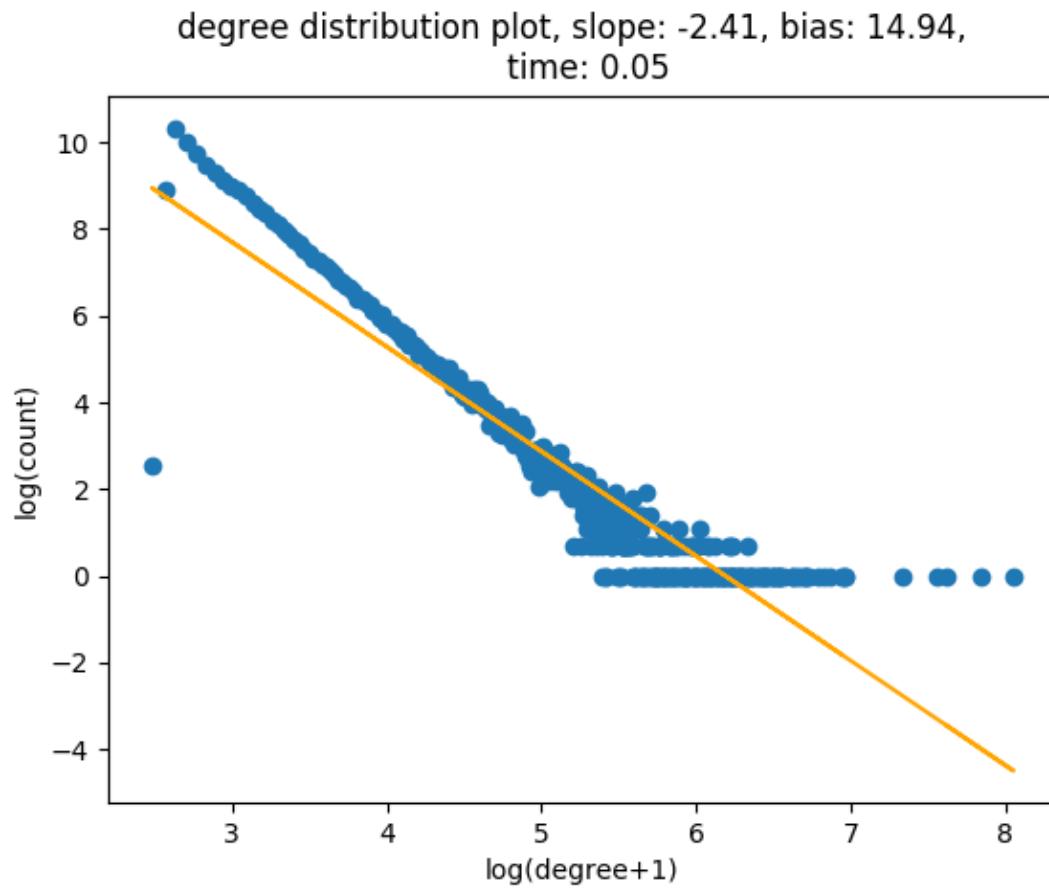


## BA network

degree distribution plot, slope: -2.11, bias: 11.72,  
time: 0.0978



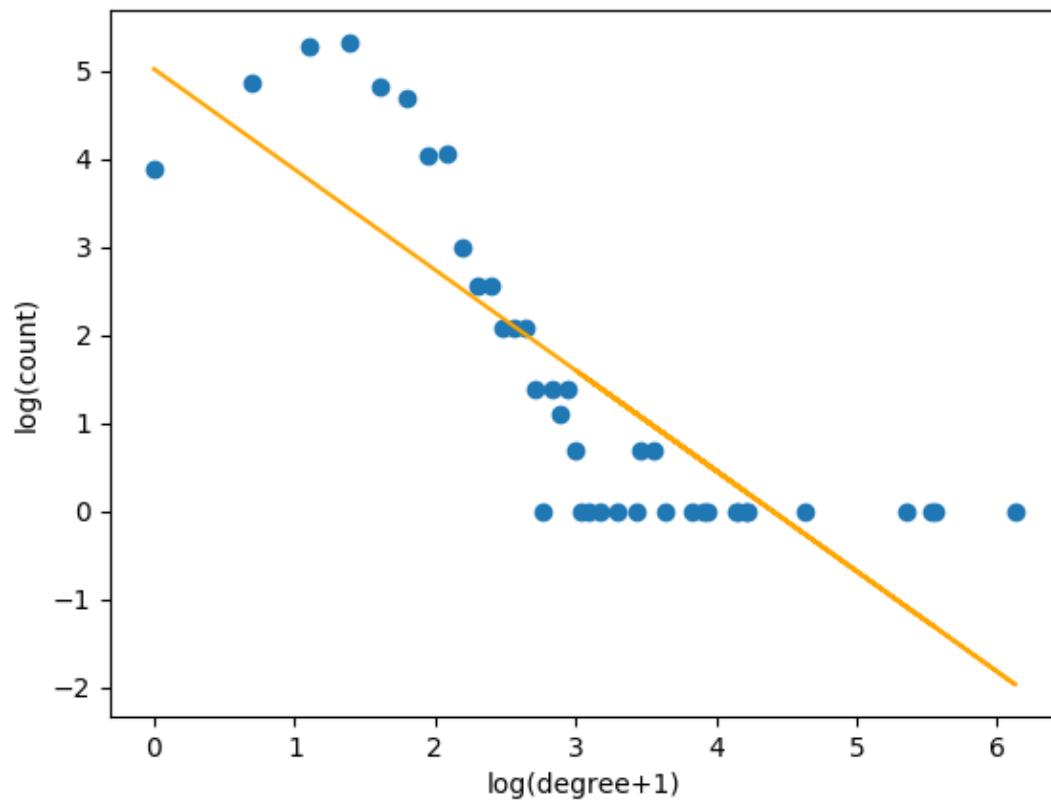
## modified BA network



## metabolic

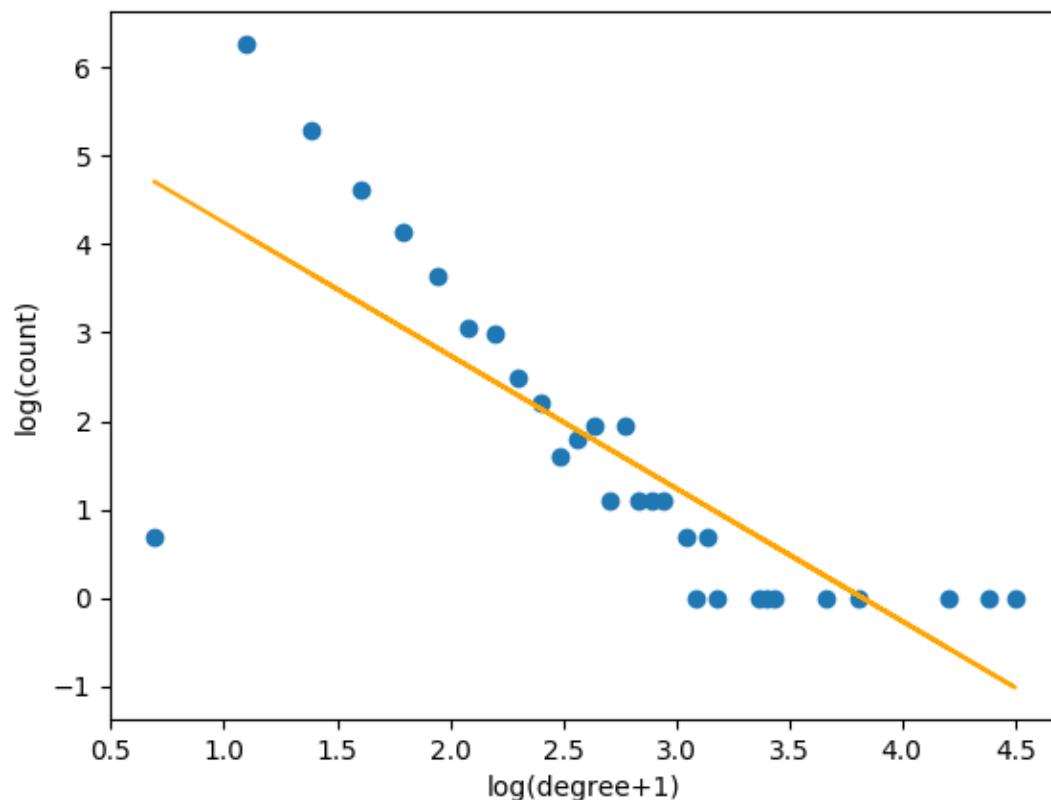
### original network

degree distribution plot, slope: -1.14, bias: 5.02,  
time: 0.0016

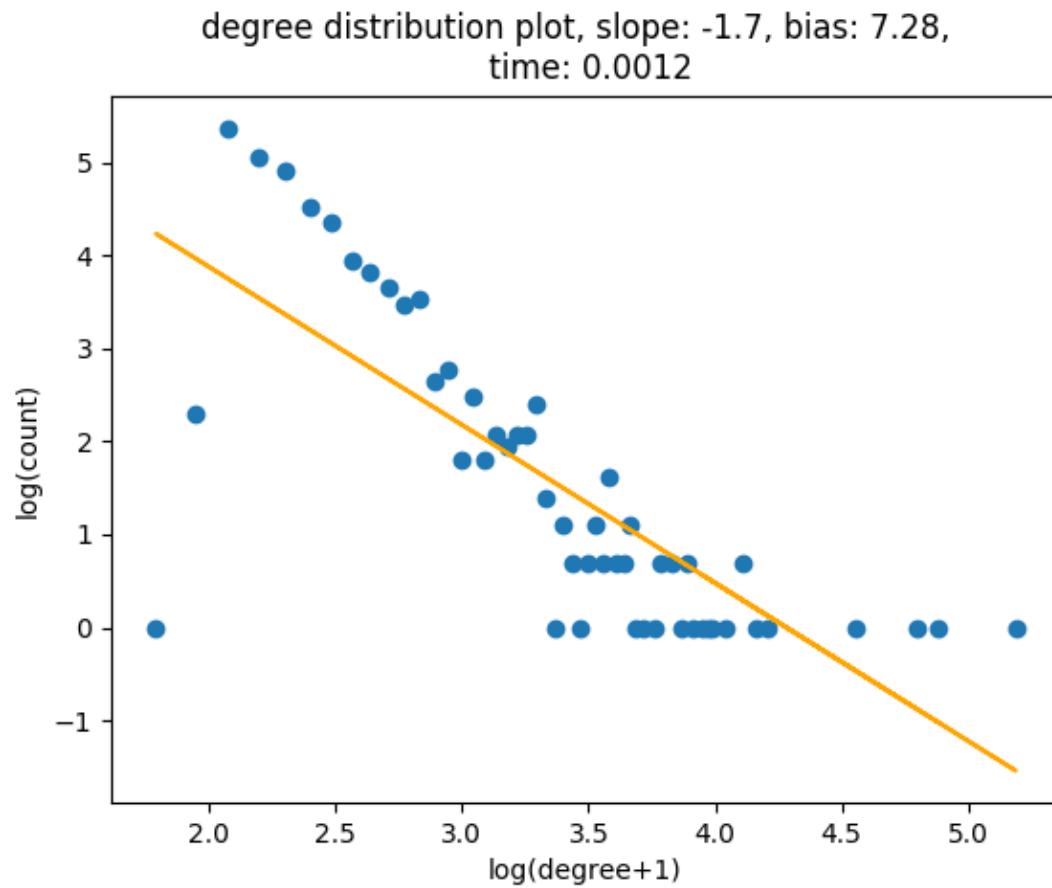


## BA network

degree distribution plot, slope: -1.5, bias: 5.74,  
time: 0.0011



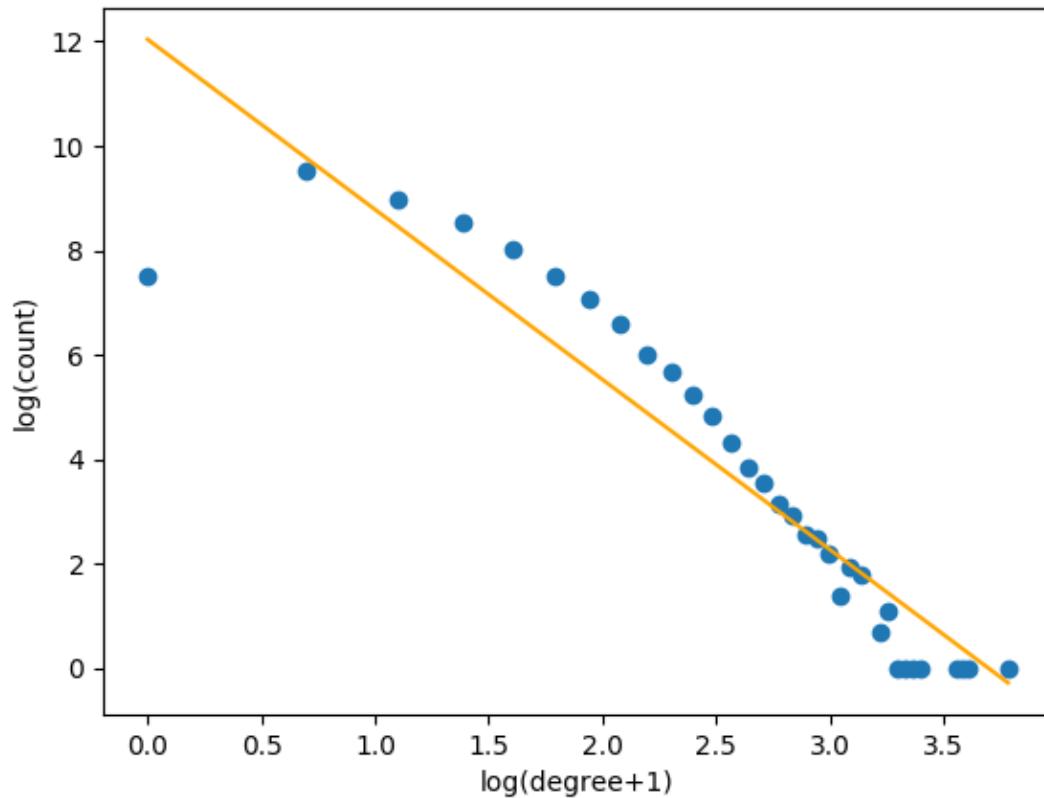
## modified BA network



## phonecalls

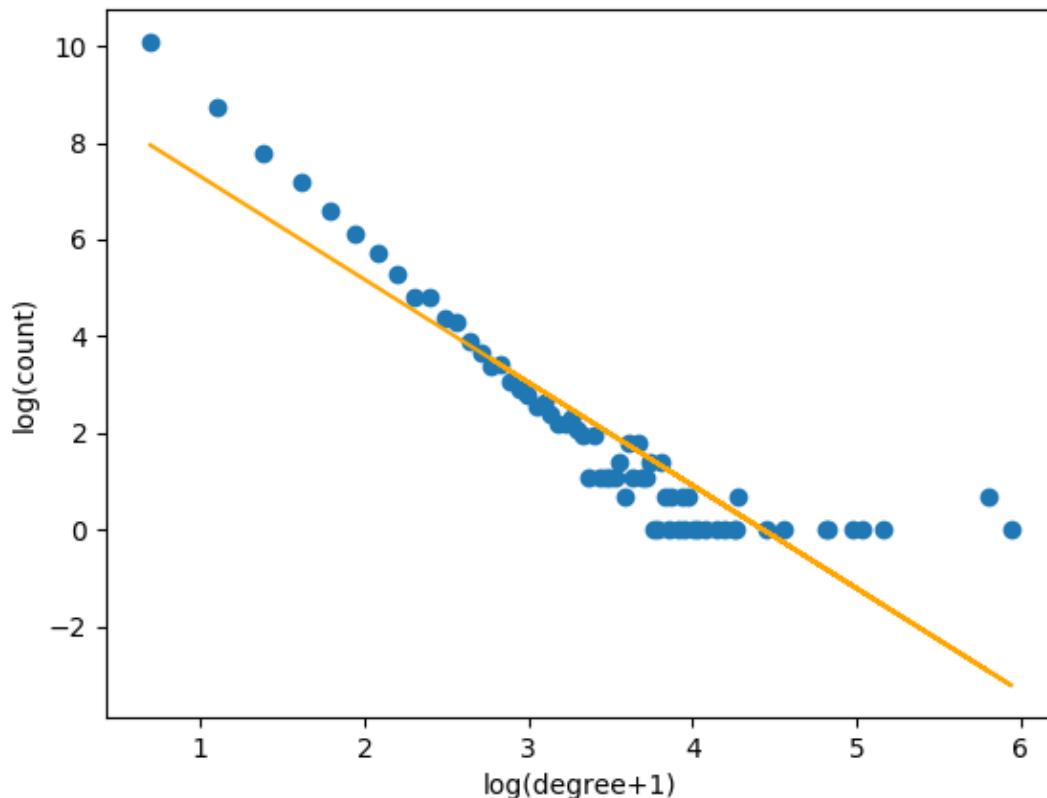
### original network

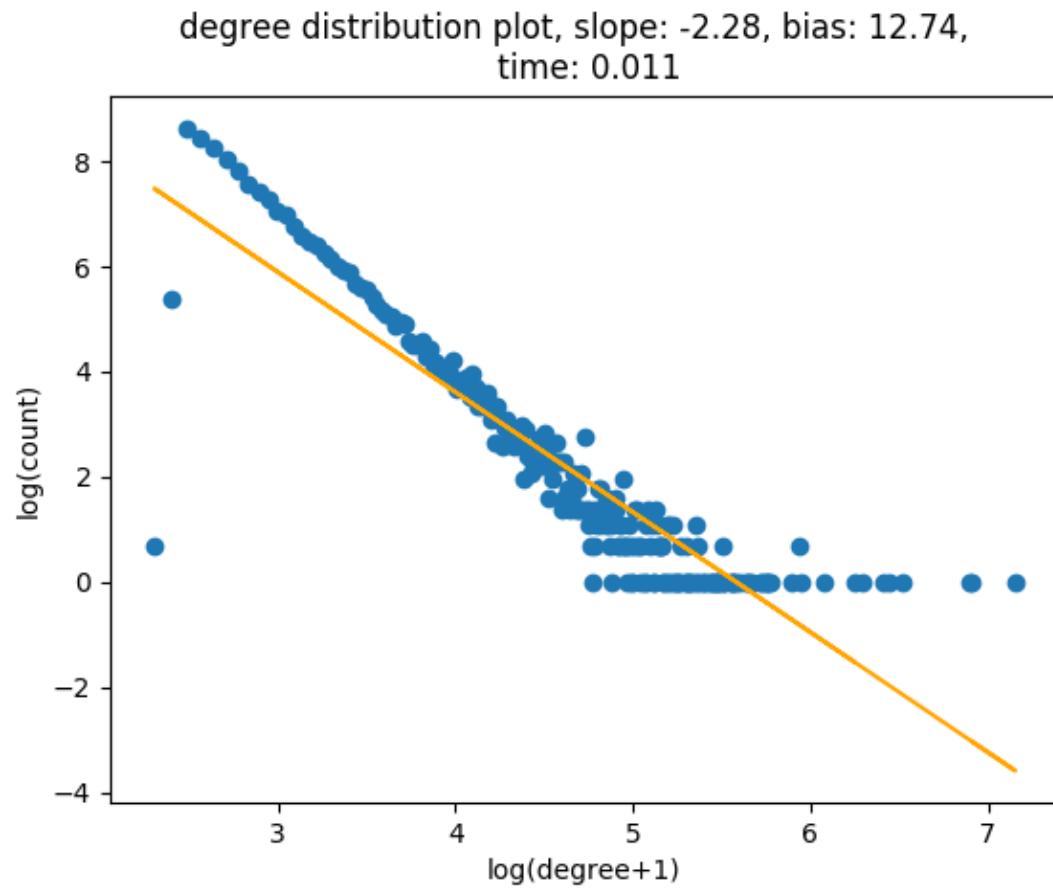
degree distribution plot, slope: -3.25, bias: 12.04,  
time: 0.0487



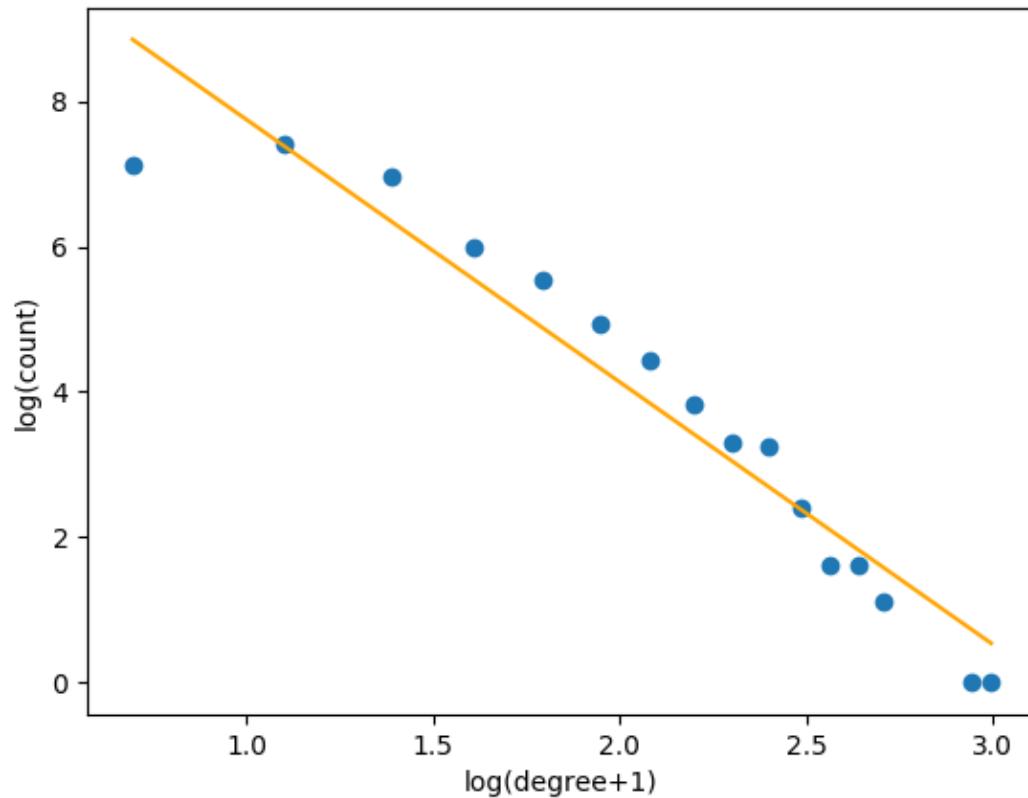
## BA network

degree distribution plot, slope: -2.13, bias: 9.43,  
time: 0.0126



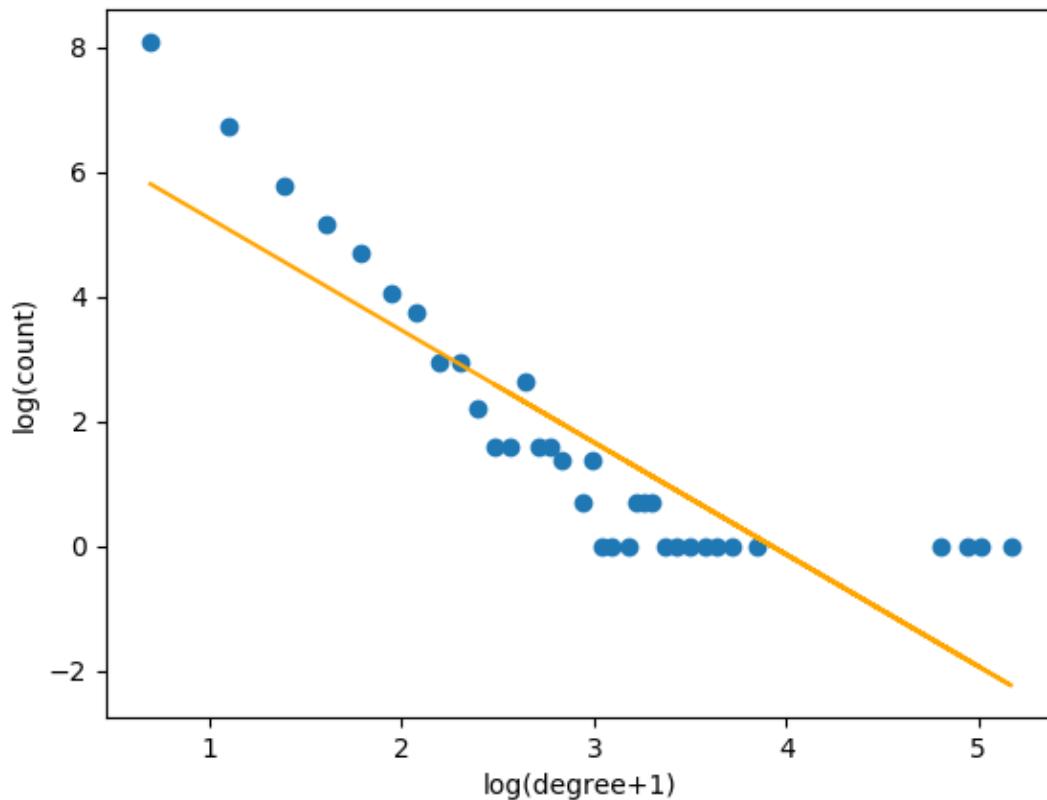
**modified BA network****powergrid****original network**

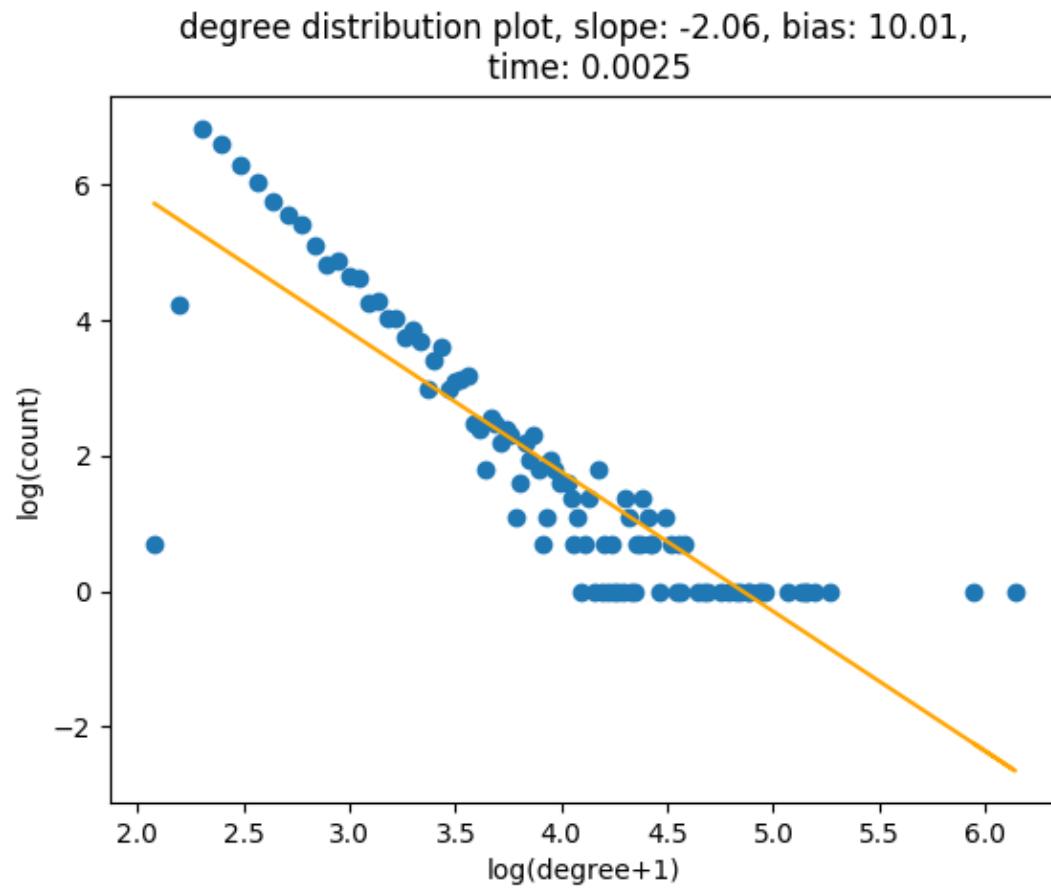
degree distribution plot, slope: -3.61, bias: 11.35,  
time: 0.0019



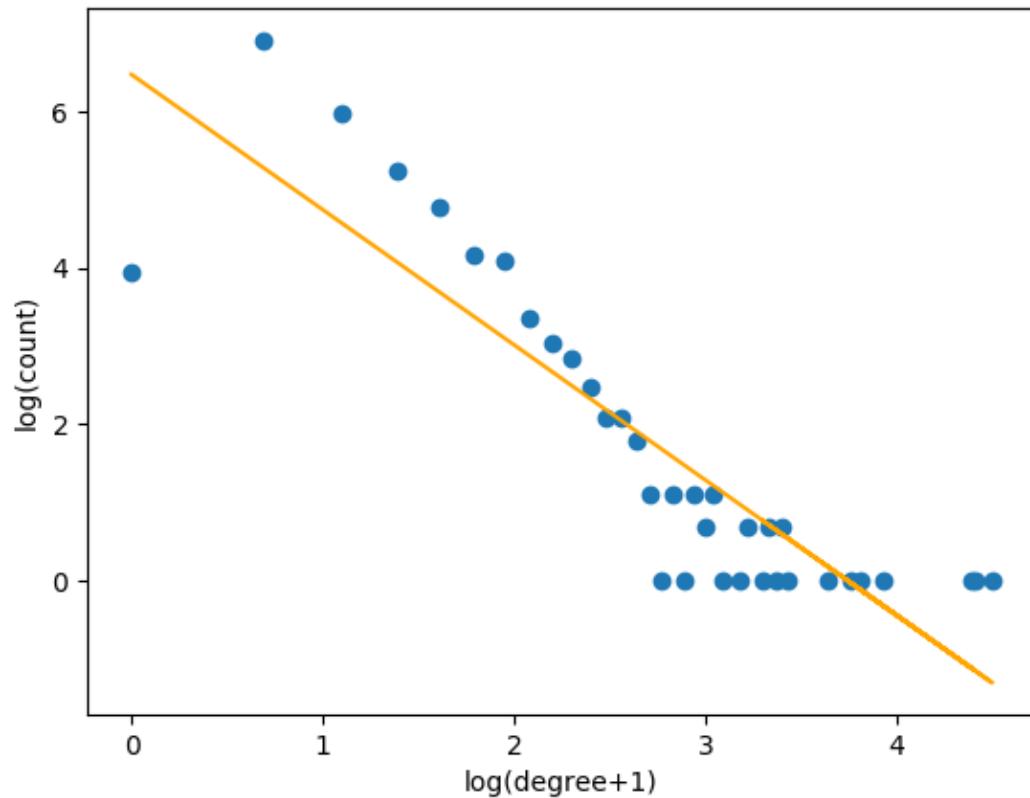
## BA network

degree distribution plot, slope: -1.8, bias: 7.06,  
time: 0.002



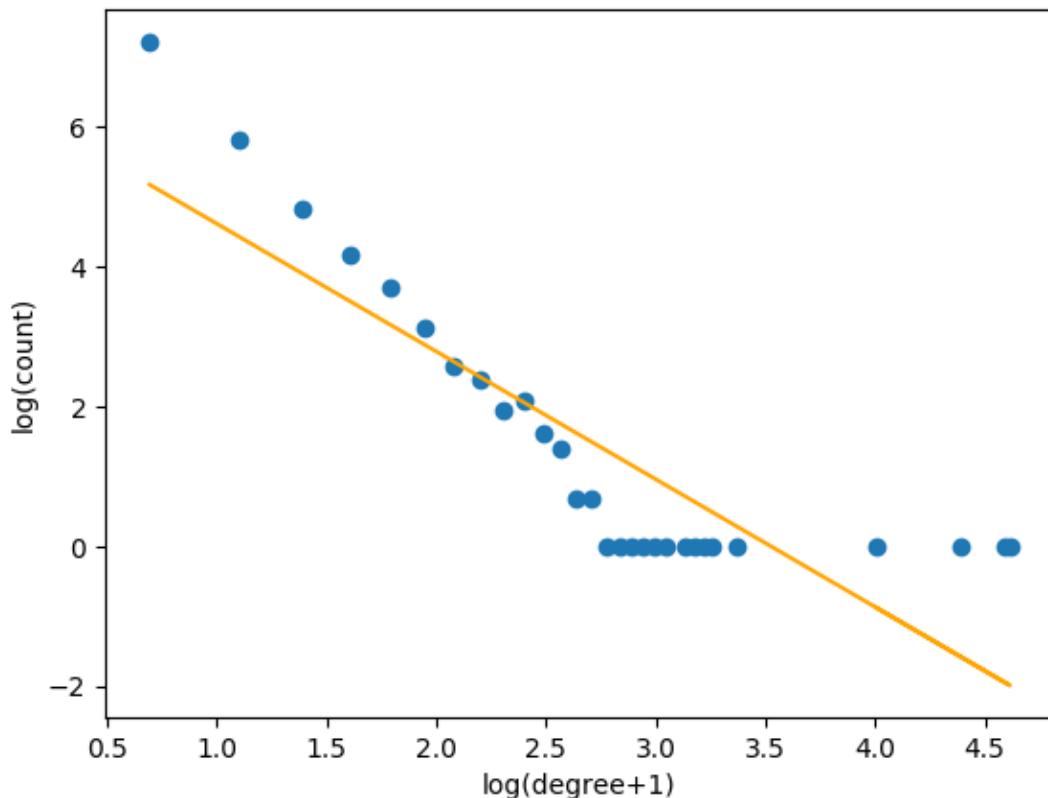
**modified BA network****protein****original network**

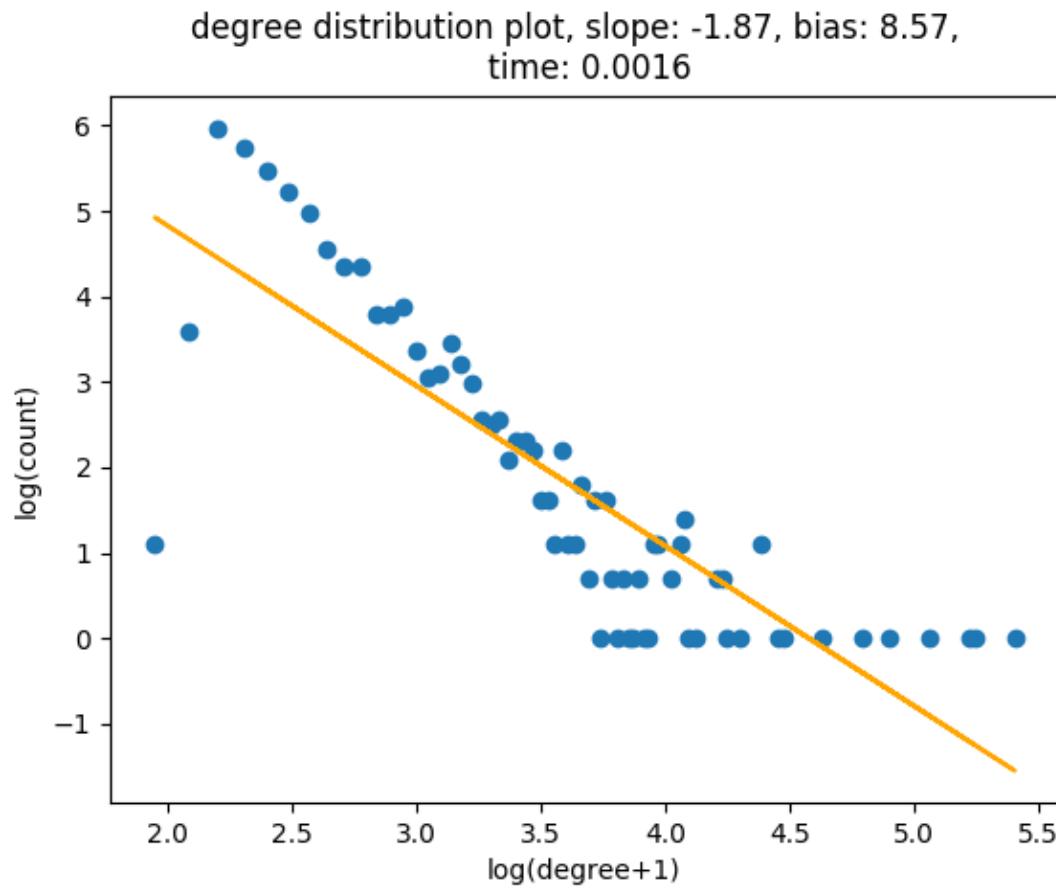
degree distribution plot, slope: -1.73, bias: 6.47,  
time: 0.0148



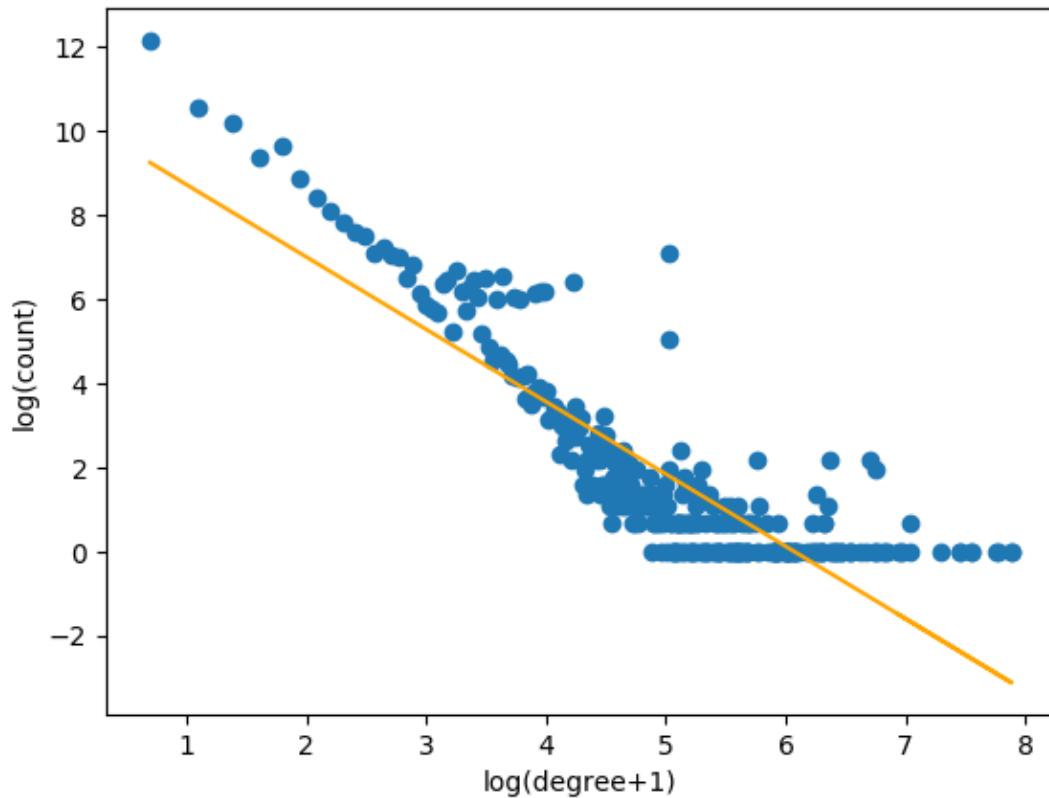
## BA network

degree distribution plot, slope: -1.82, bias: 6.44,  
time: 0.0012



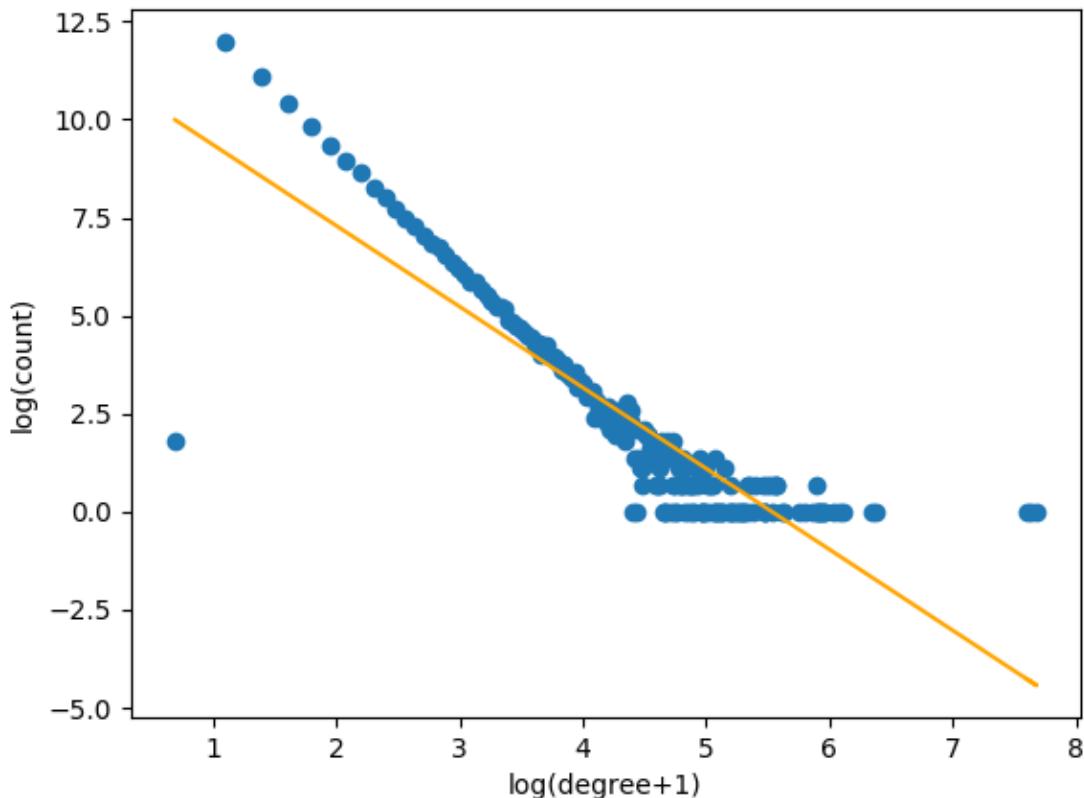
**modified BA network****WWW****original network**

degree distribution plot, slope: -1.72, bias: 10.44,  
time: 0.2047

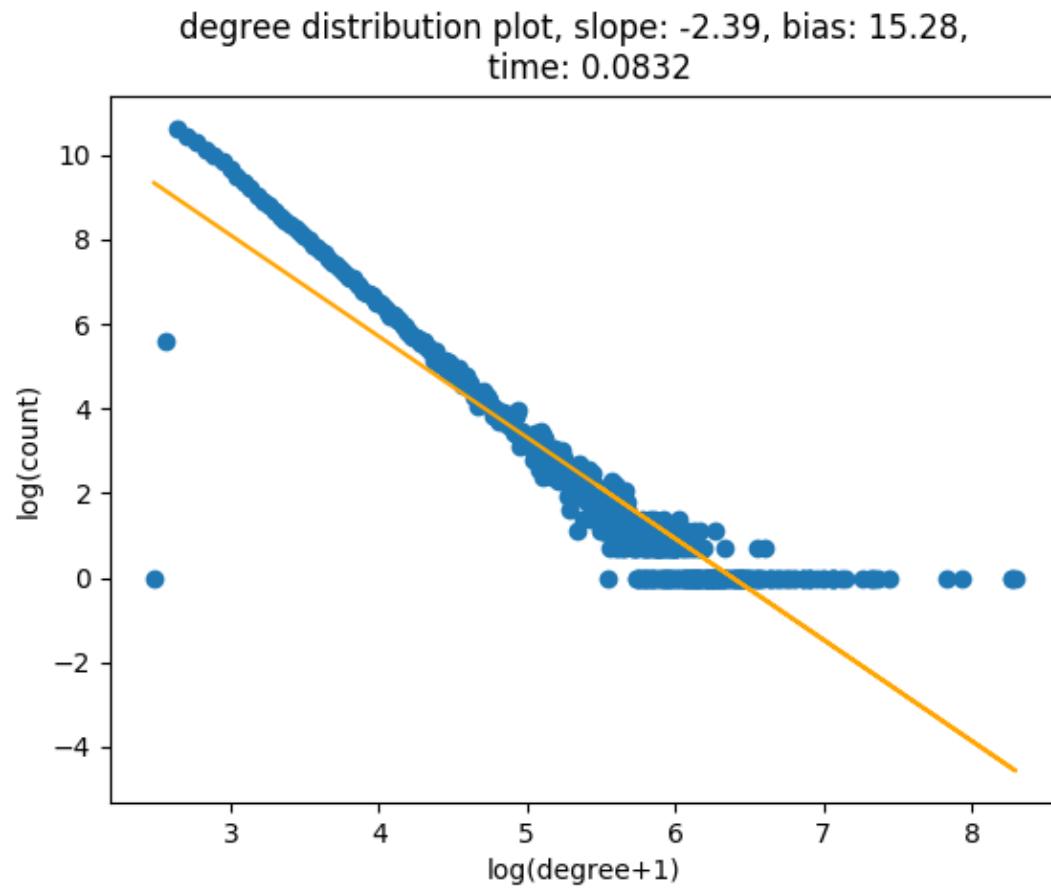


## BA network

degree distribution plot, slope: -2.06, bias: 11.42,  
time: 0.0774



## modified BA network

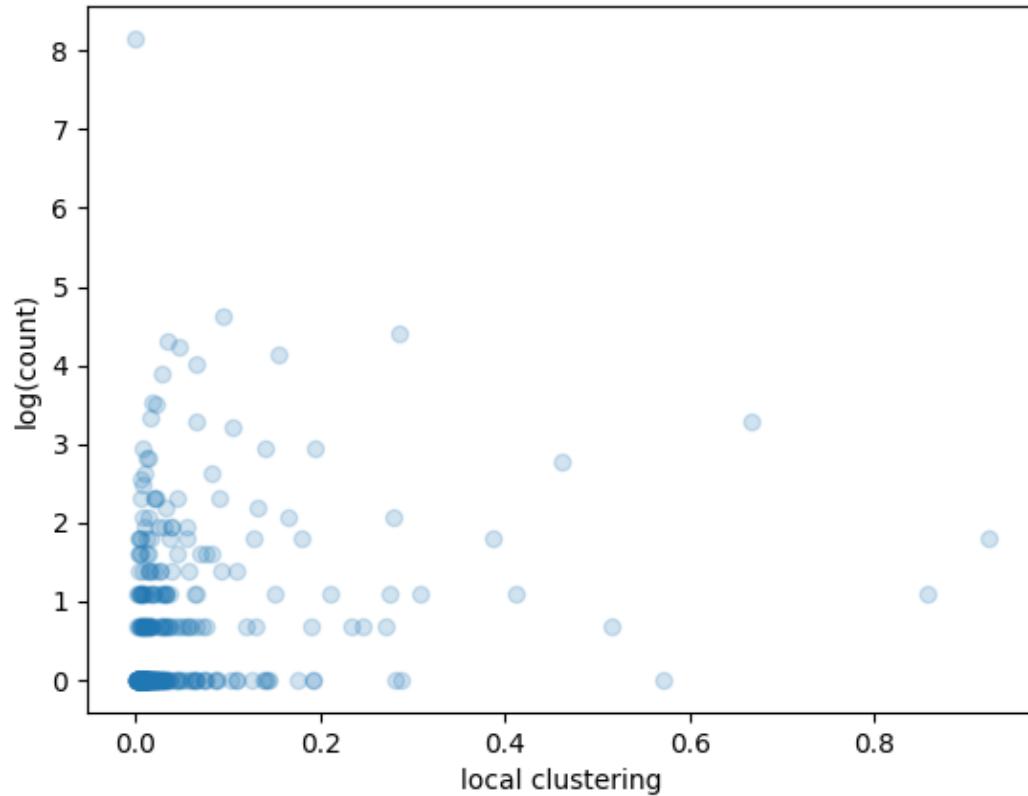


**clustering**

**collaboration**

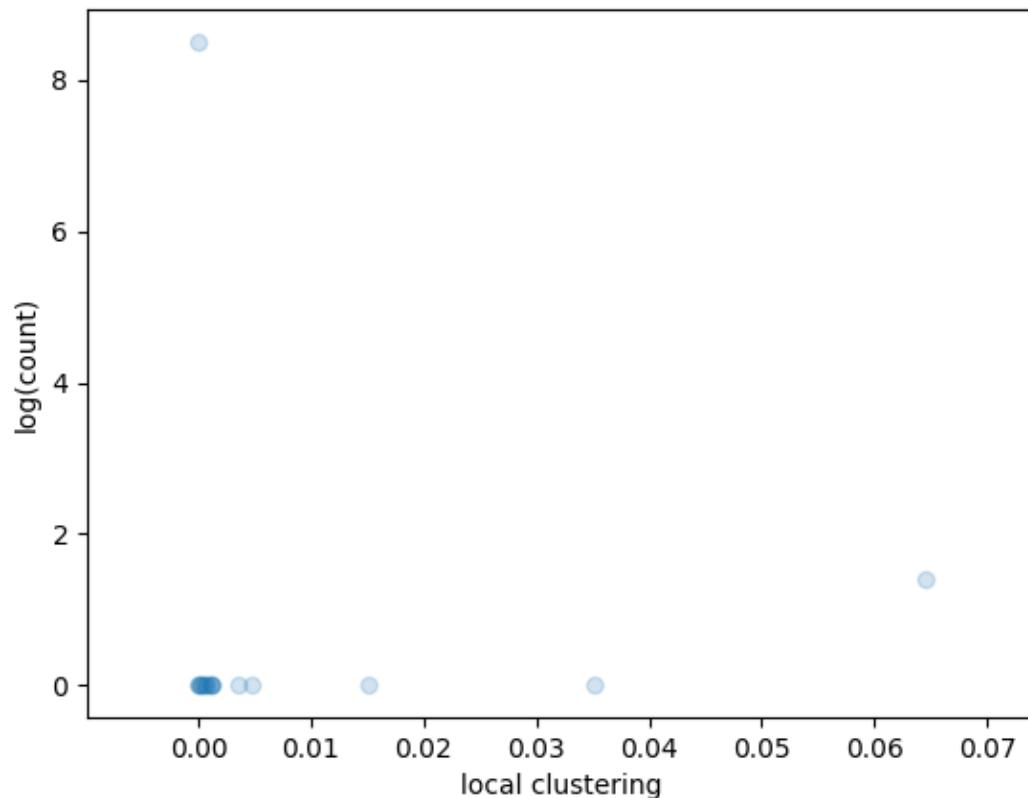
**original network**

local clustering distribution plot, local clustering mean: 0.03,  
time: 0.0194

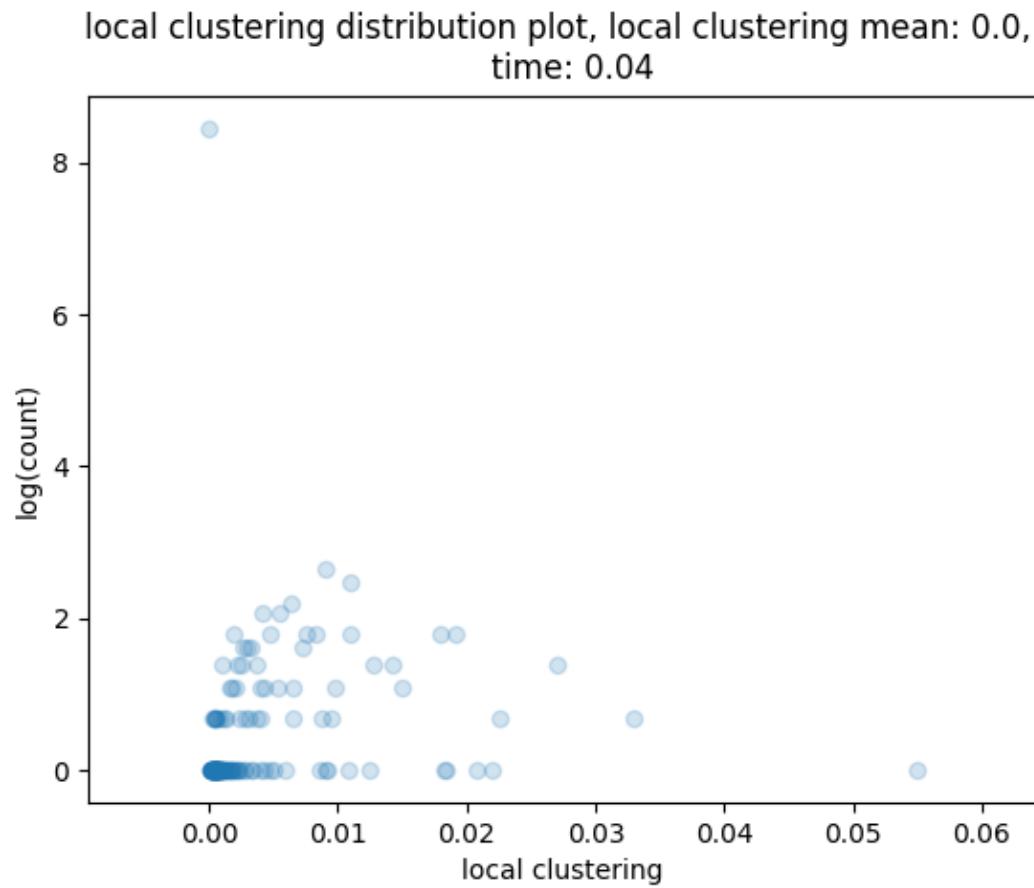


## BA network

local clustering distribution plot, local clustering mean: 0.0,  
time: 0.0171



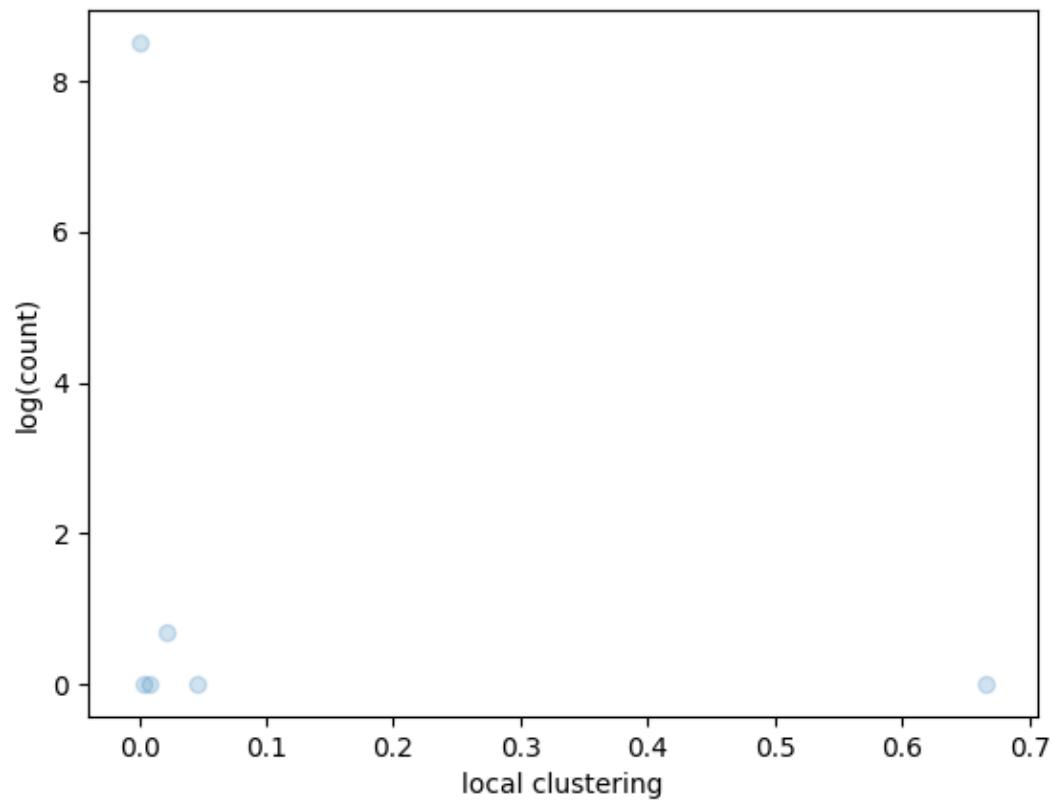
## modified BA network



## citation

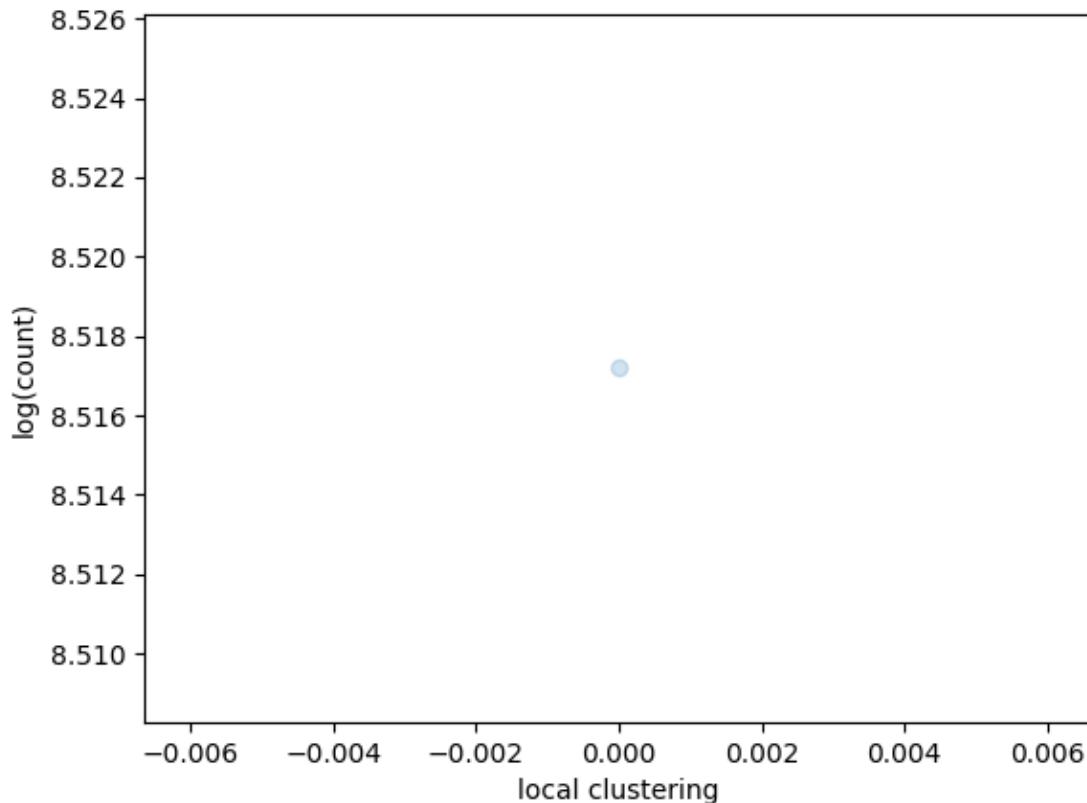
### original network

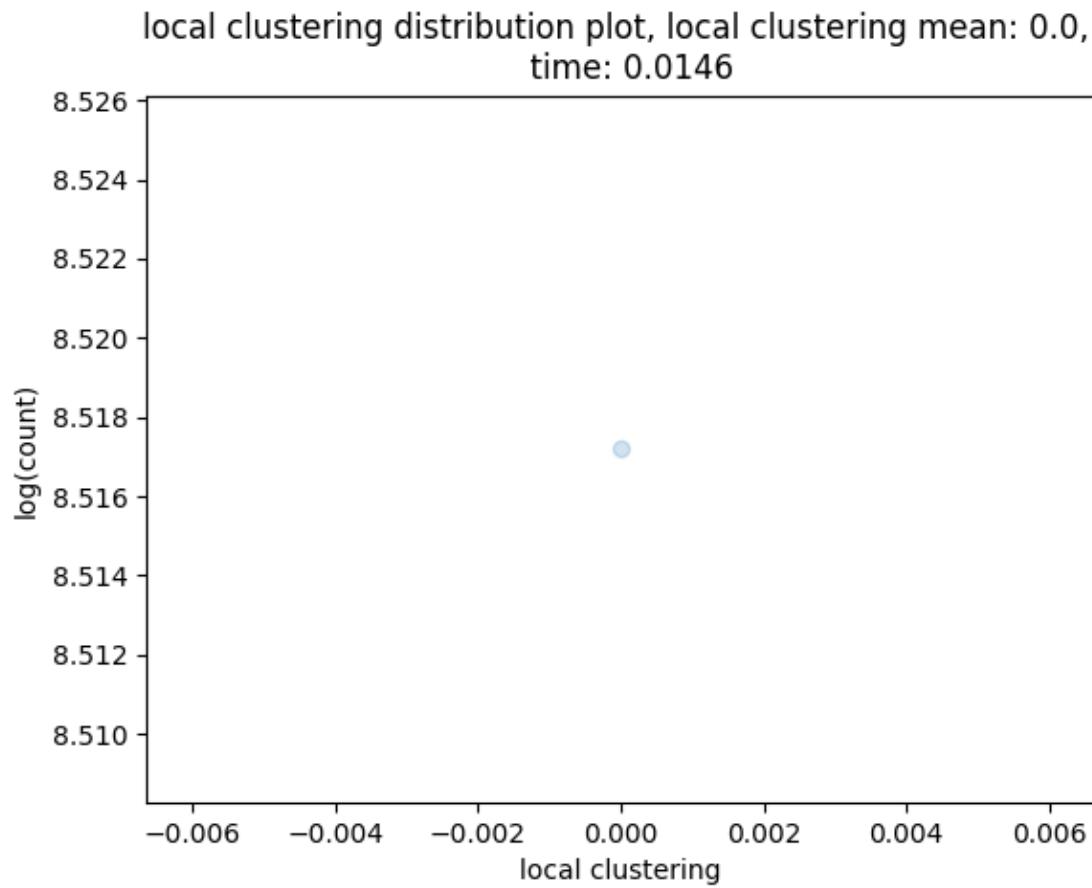
local clustering distribution plot, local clustering mean: 0.0,  
time: 0.0156



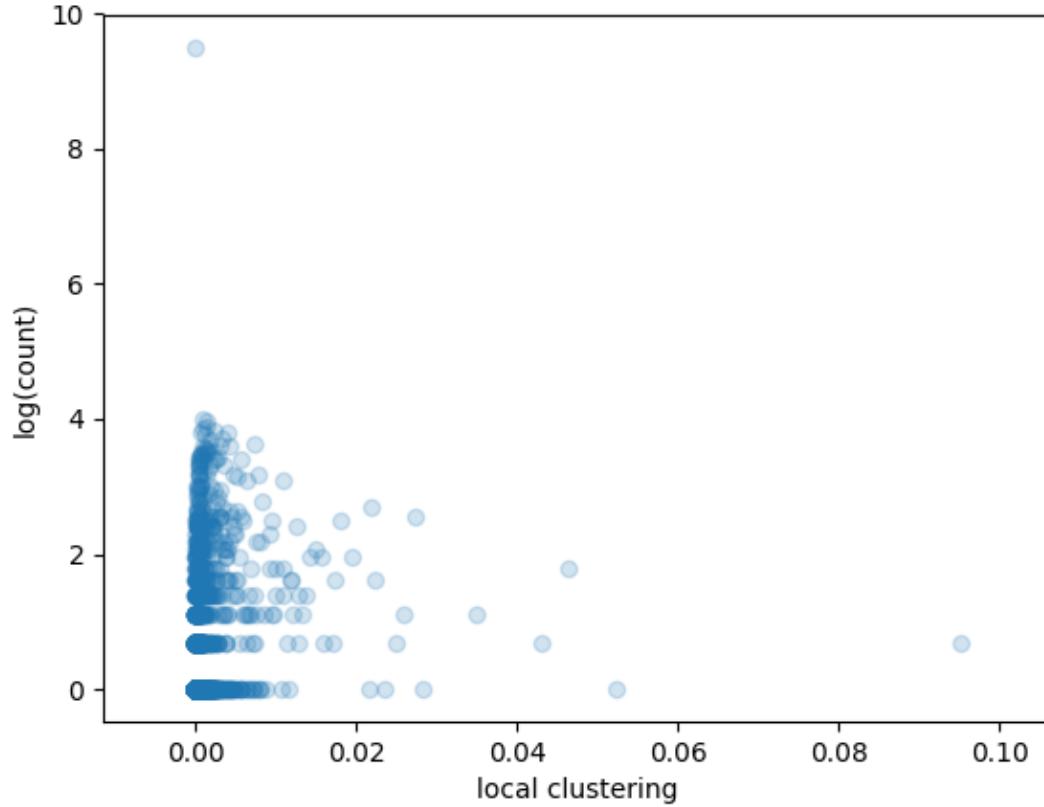
## BA network

local clustering distribution plot, local clustering mean: 0.0,  
time: 0.0153



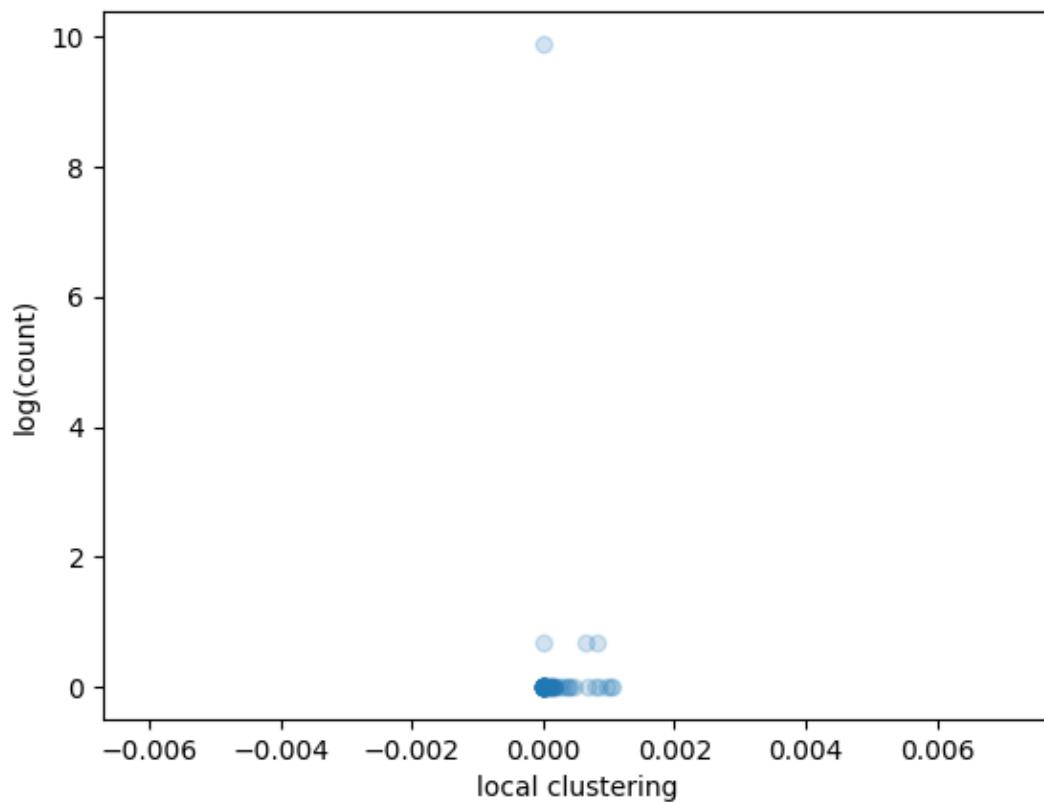
**modified BA network****actor****original network**

local clustering distribution plot, local clustering mean: 0.0,  
time: 0.1375



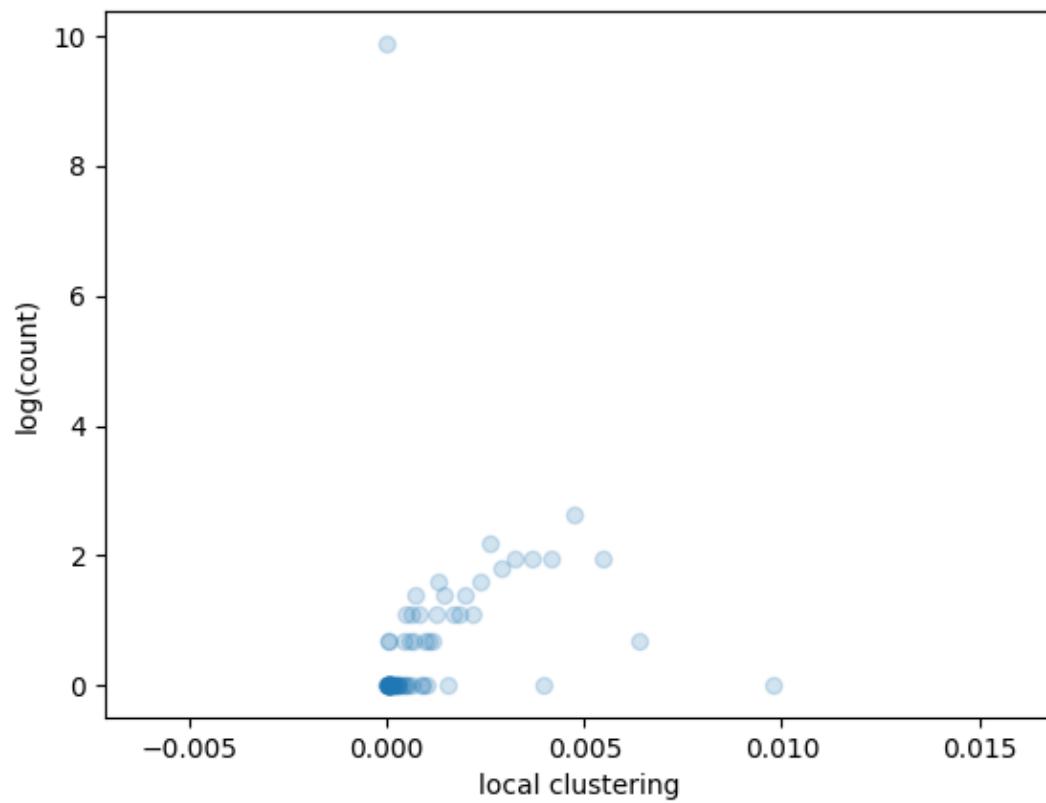
## BA network

local clustering distribution plot, local clustering mean: 0.0,  
time: 0.055

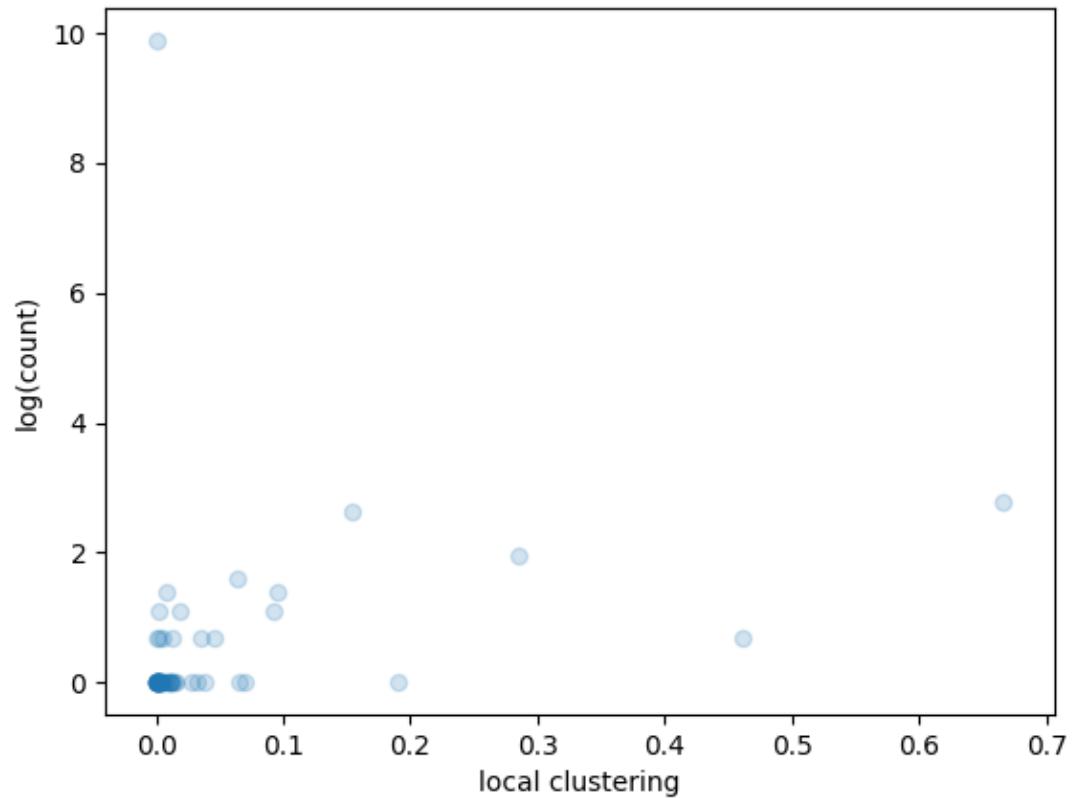


**modified BA network**

local clustering distribution plot, local clustering mean: 0.0,  
time: 0.022

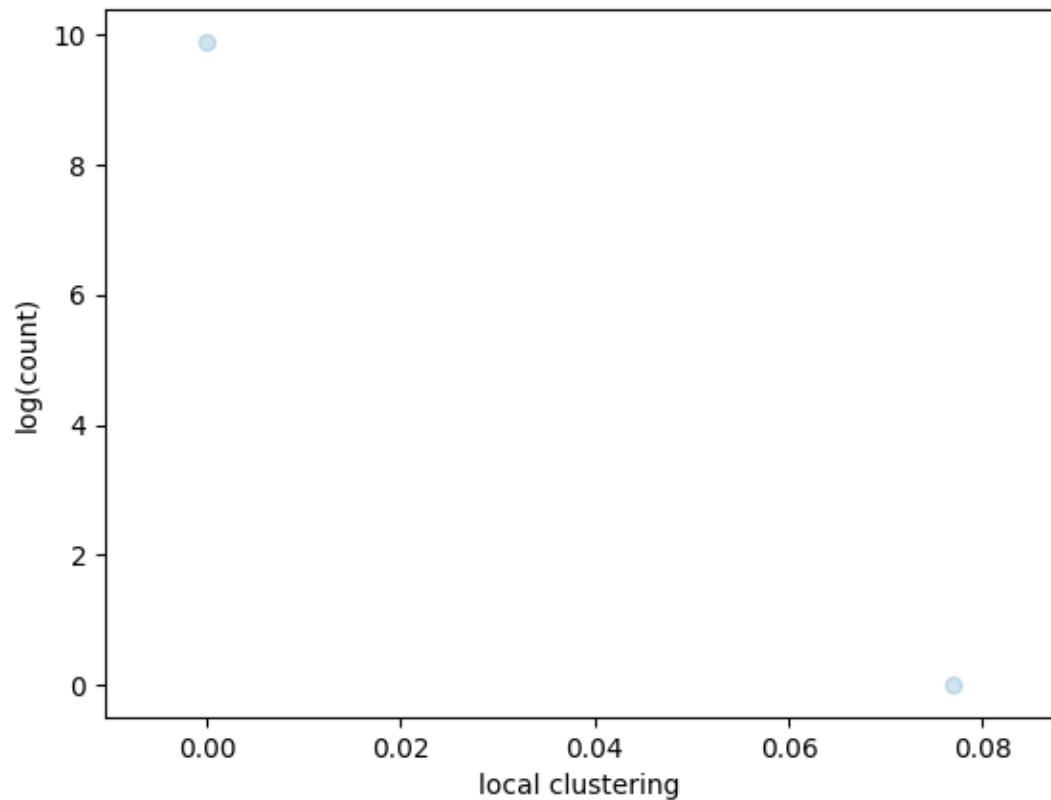
**email****original network**

local clustering distribution plot, local clustering mean: 0.0,  
time: 0.0215



## BA network

local clustering distribution plot, local clustering mean: 0.0,  
time: 0.0168



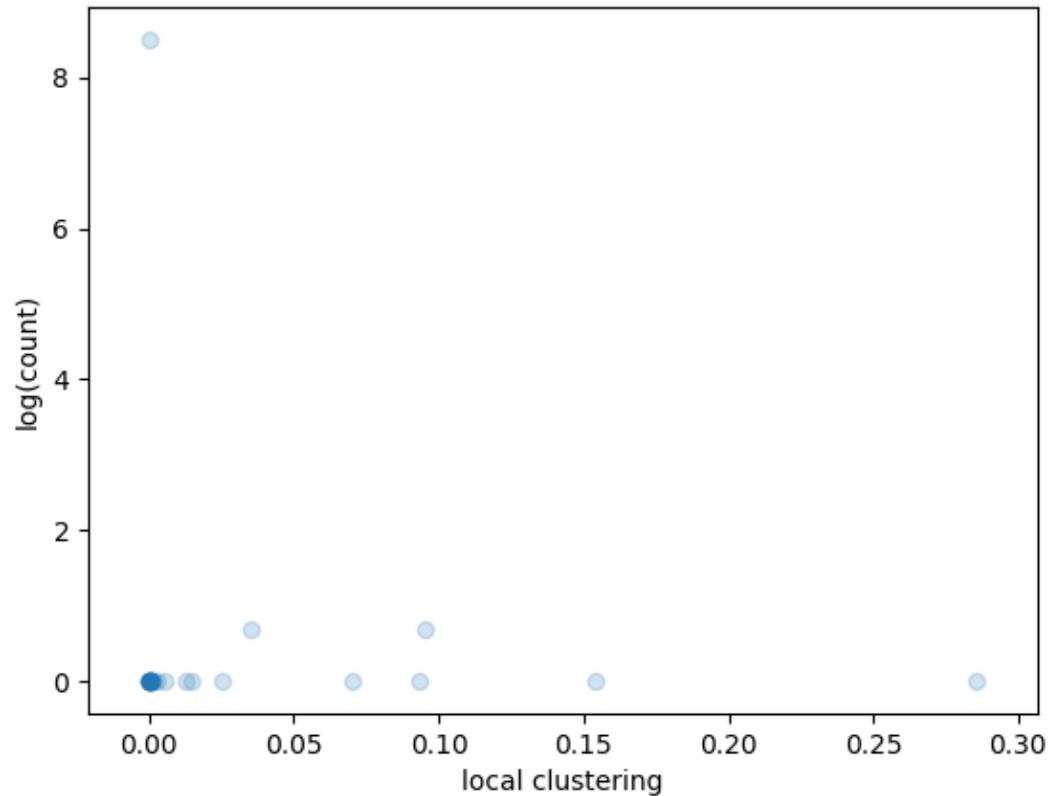
## modified BA network



## internet

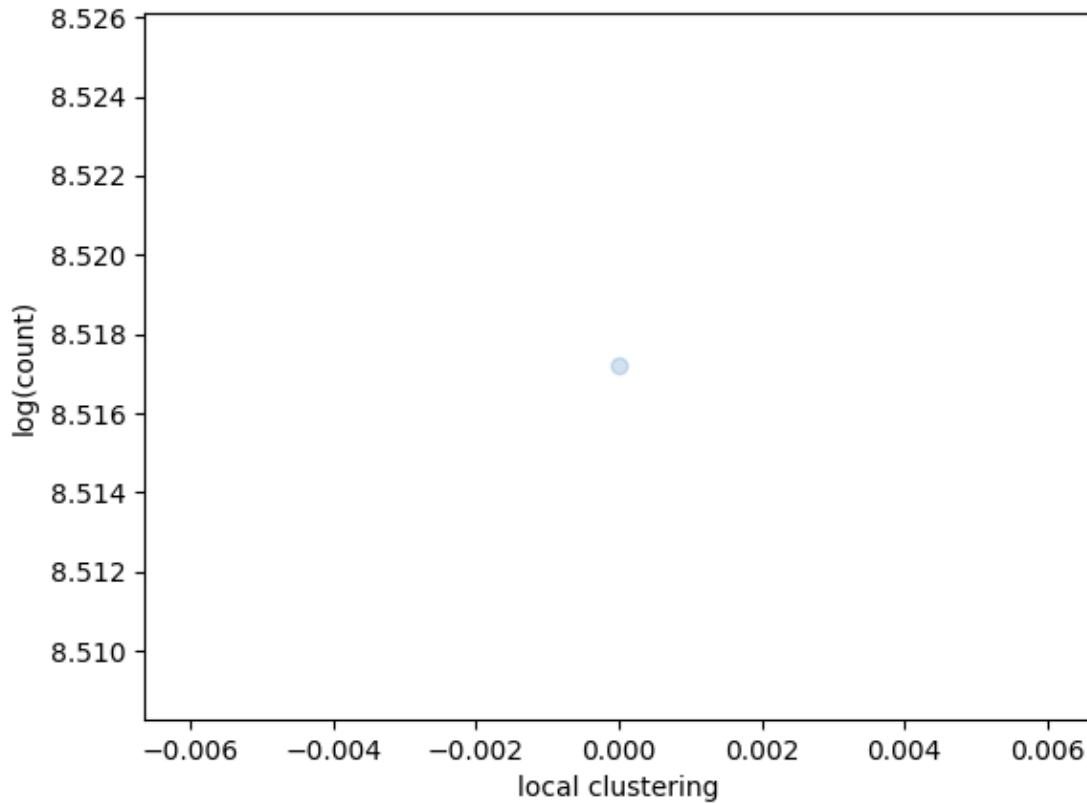
## original network

local clustering distribution plot, local clustering mean: 0.0,  
time: 0.0249



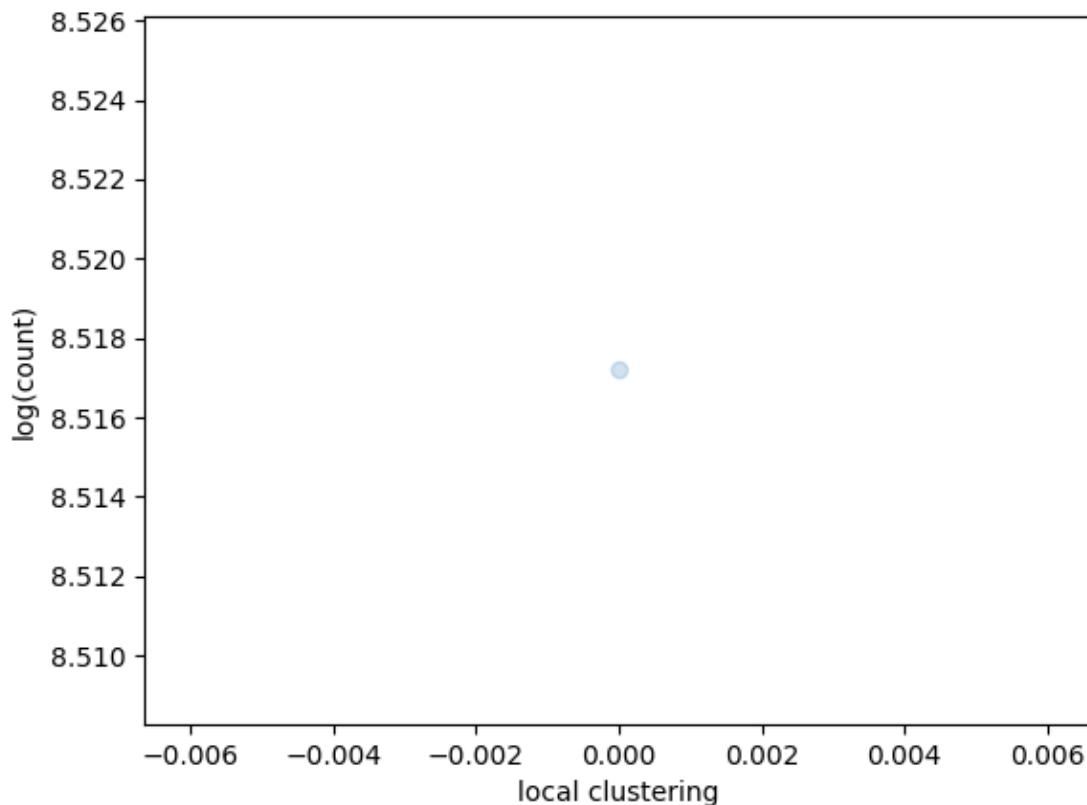
## BA network

local clustering distribution plot, local clustering mean: 0.0,  
time: 0.0569



## modified BA network

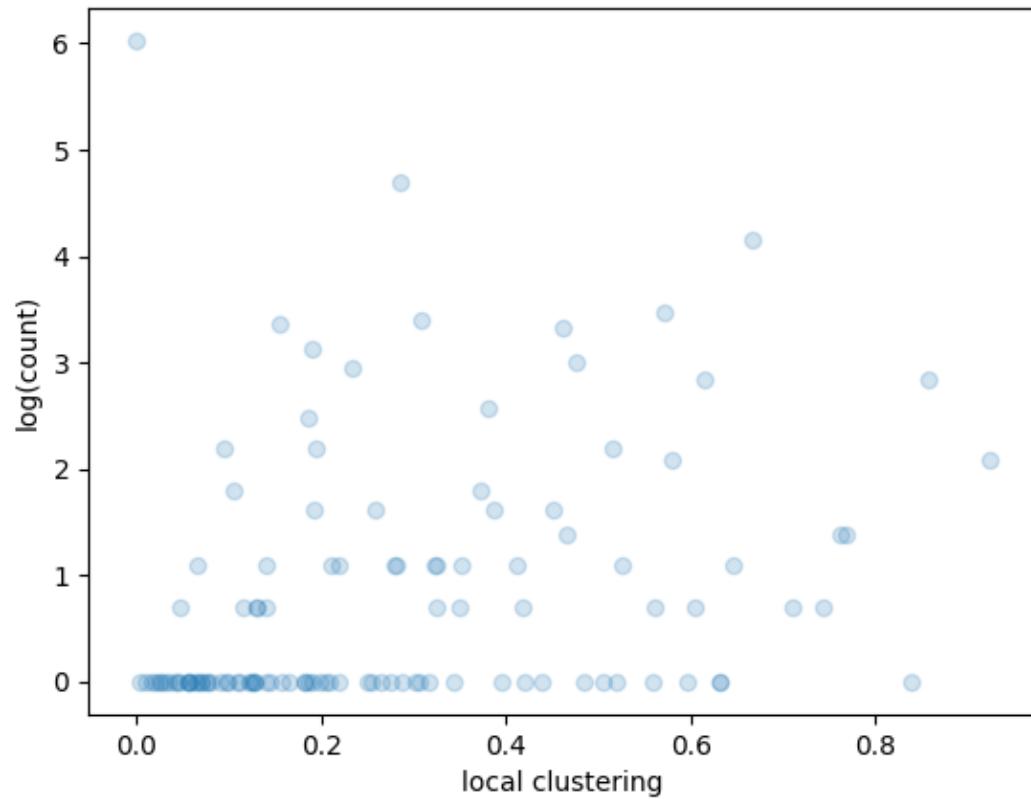
local clustering distribution plot, local clustering mean: 0.0,  
time: 0.018



# metabolic

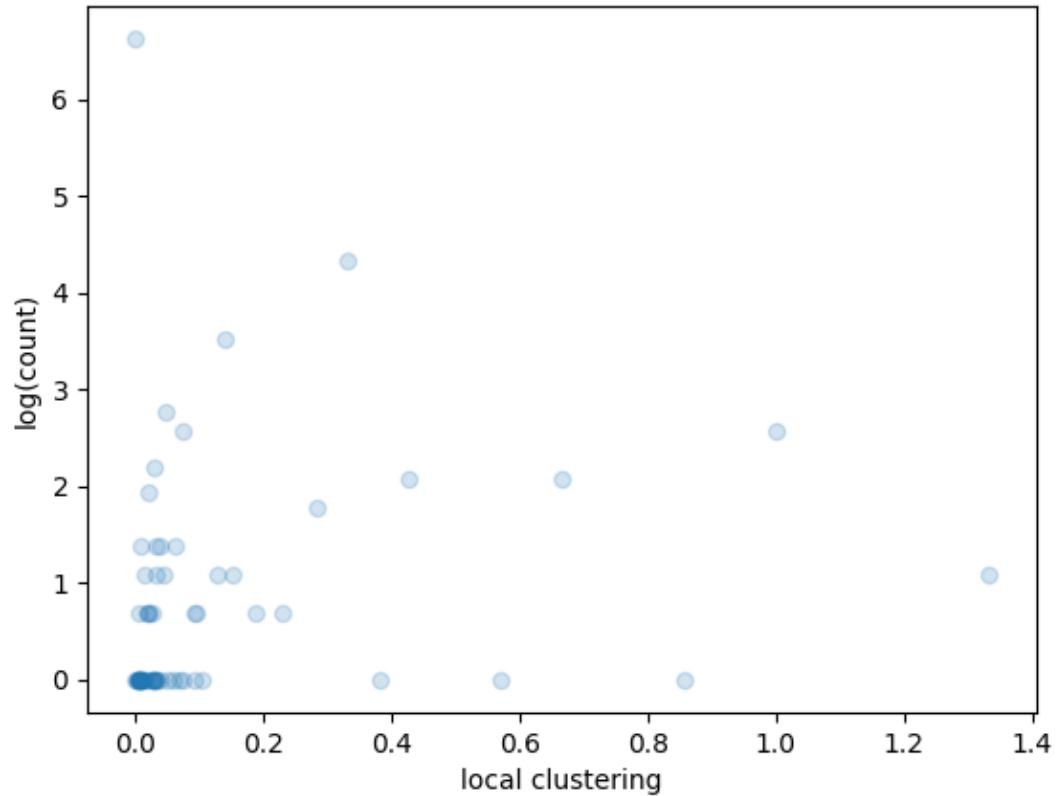
## original network

local clustering distribution plot, local clustering mean: 0.23,  
time: 0.0529



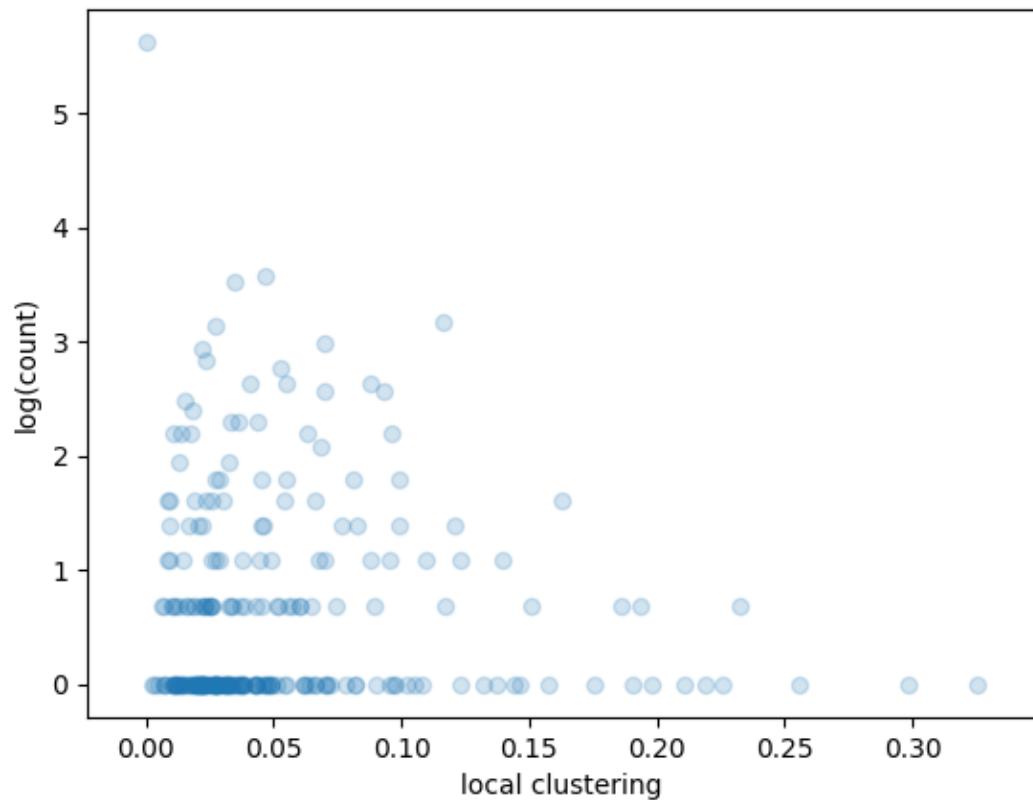
## BA network

local clustering distribution plot, local clustering mean: 0.06,  
time: 0.0312



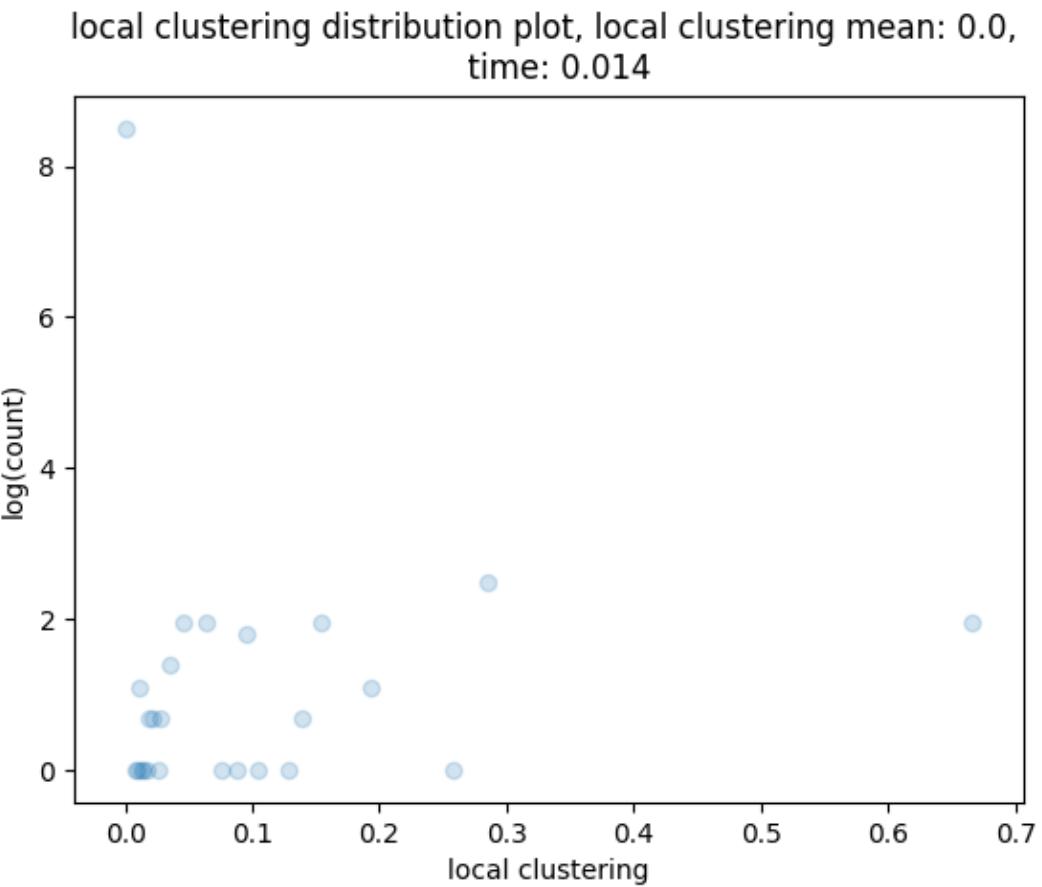
## modified BA network

local clustering distribution plot, local clustering mean: 0.04,  
time: 0.0774



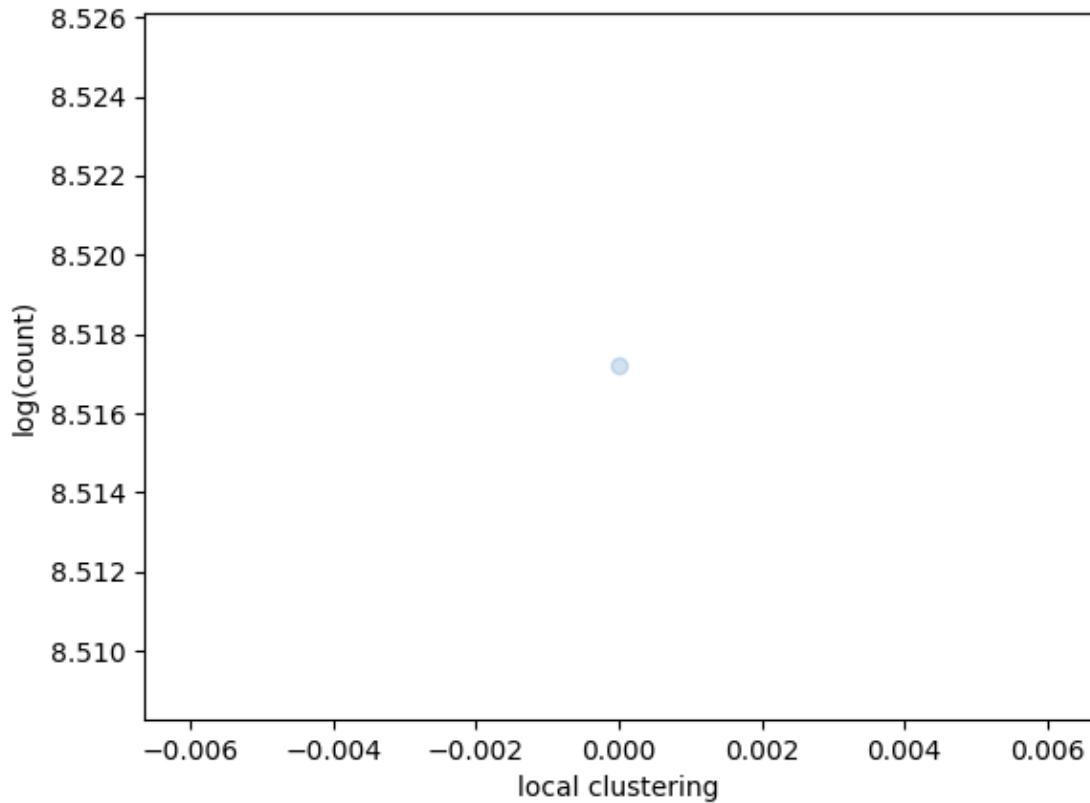
# phonecalls

## original network



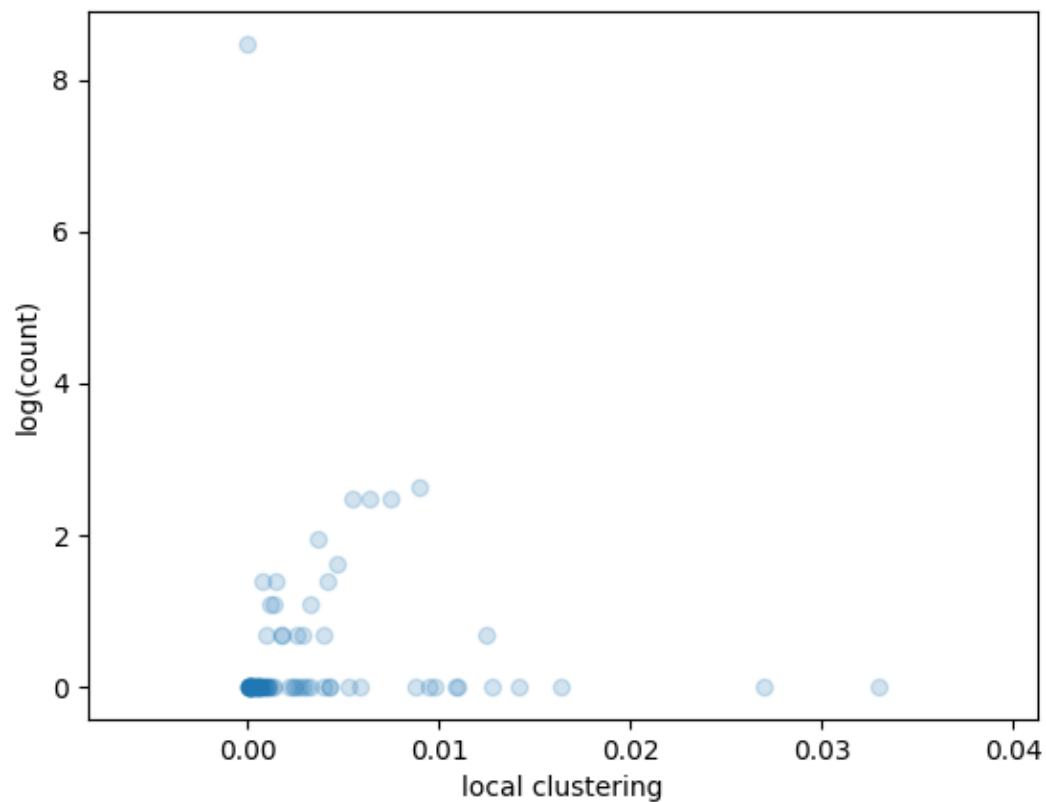
## BA network

local clustering distribution plot, local clustering mean: 0.0,  
time: 0.0132



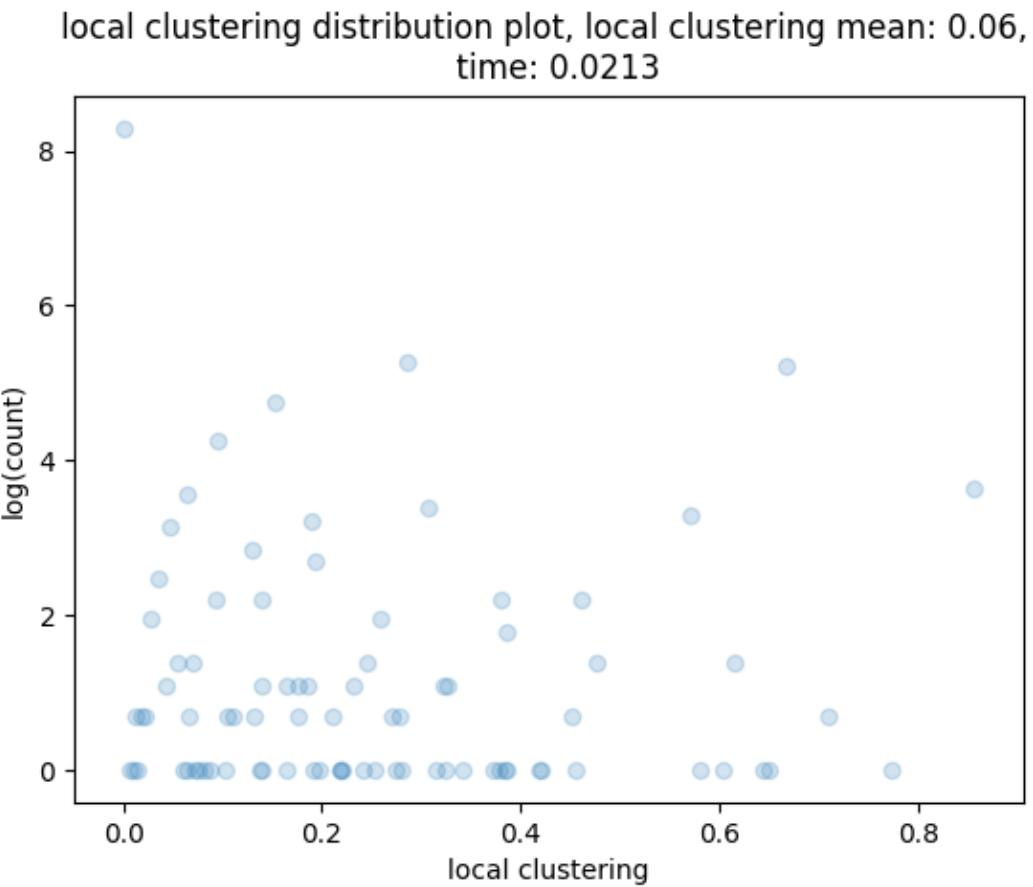
## modified BA network

local clustering distribution plot, local clustering mean: 0.0,  
time: 0.0374



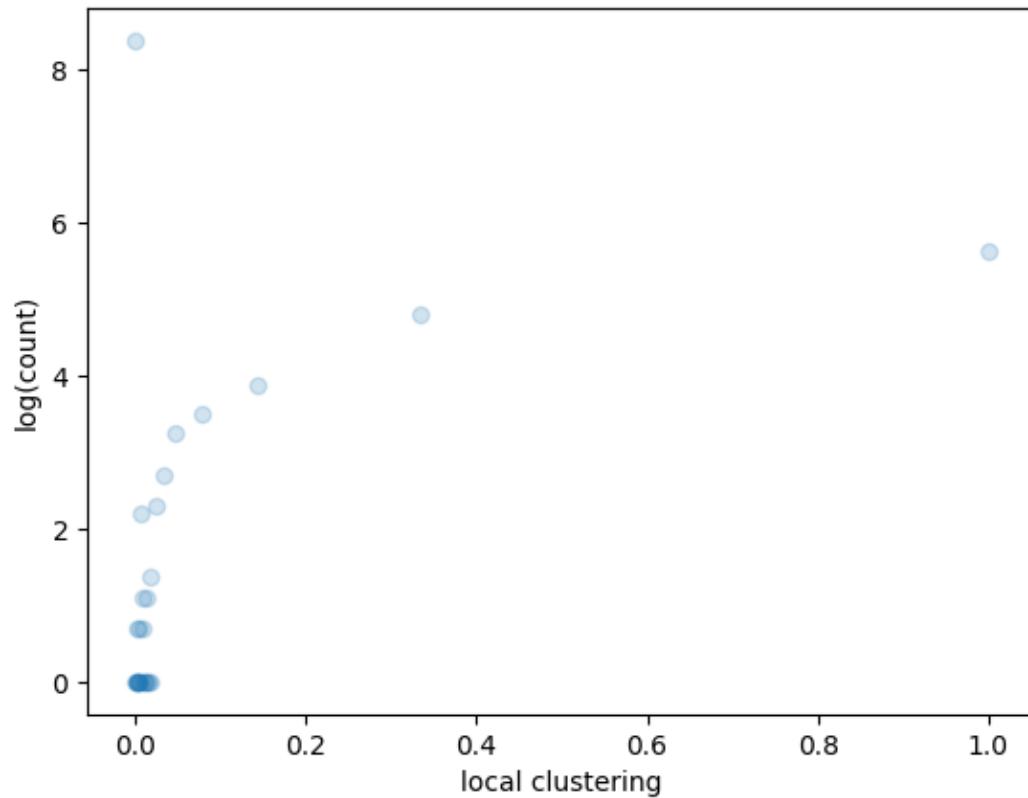
# powergrid

## original network



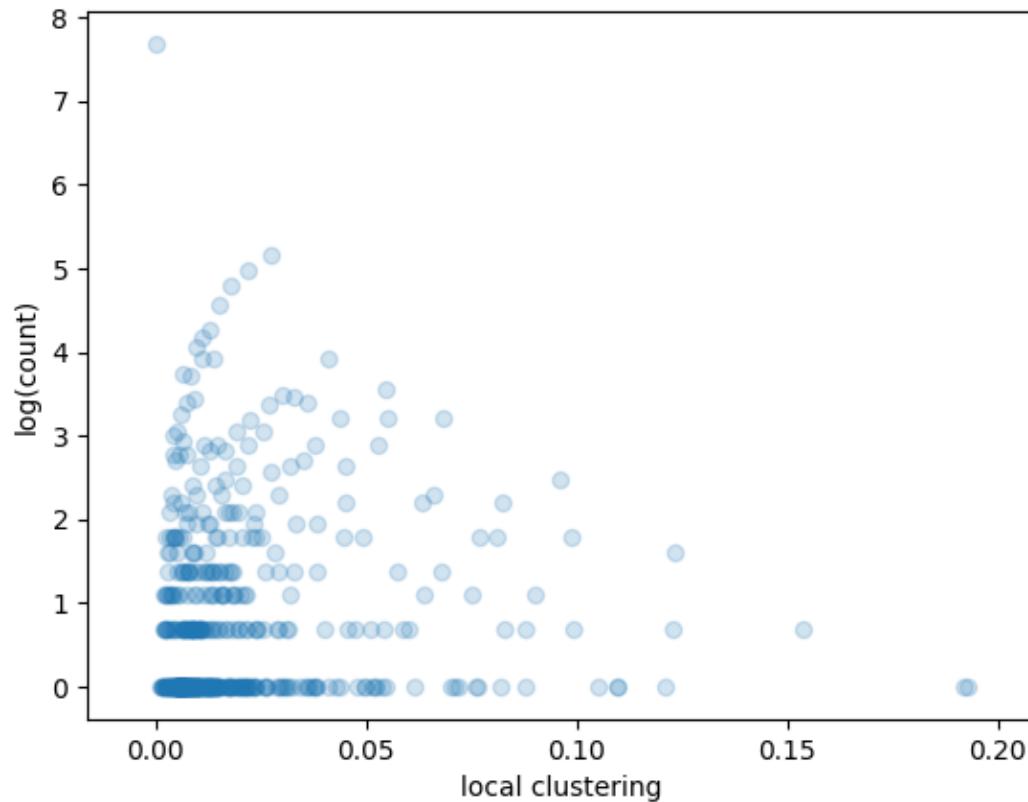
## BA network

local clustering distribution plot, local clustering mean: 0.07,  
time: 0.024



## modified BA network

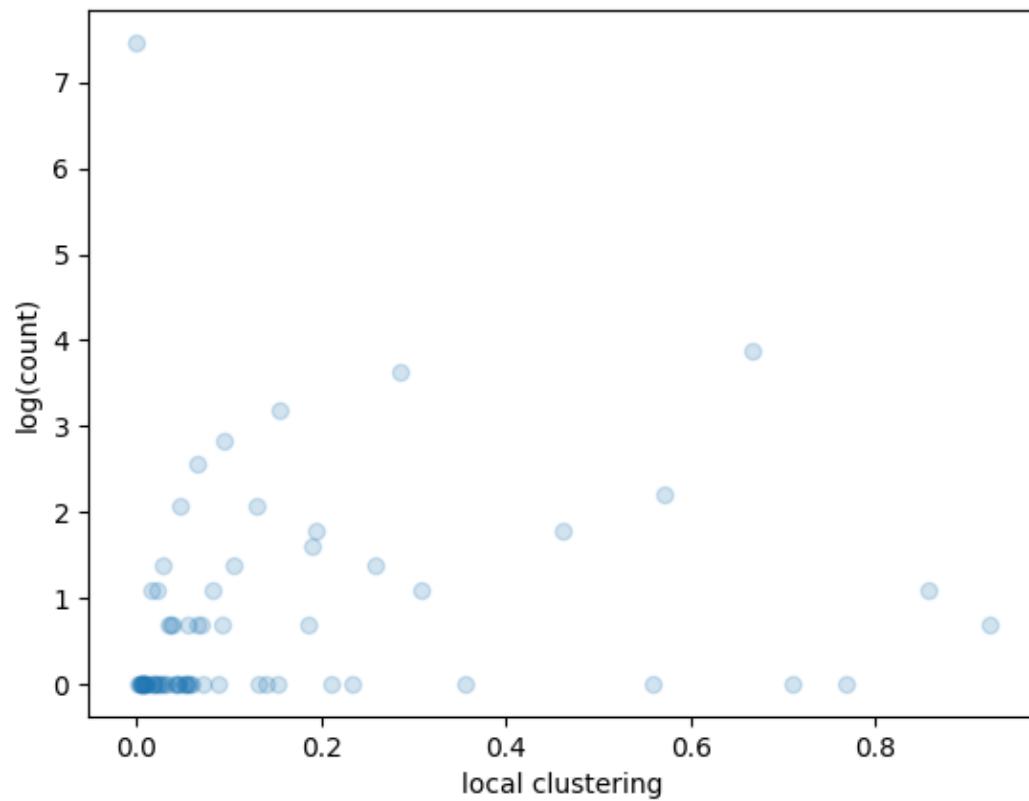
local clustering distribution plot, local clustering mean: 0.01,  
time: 0.7875



# protein

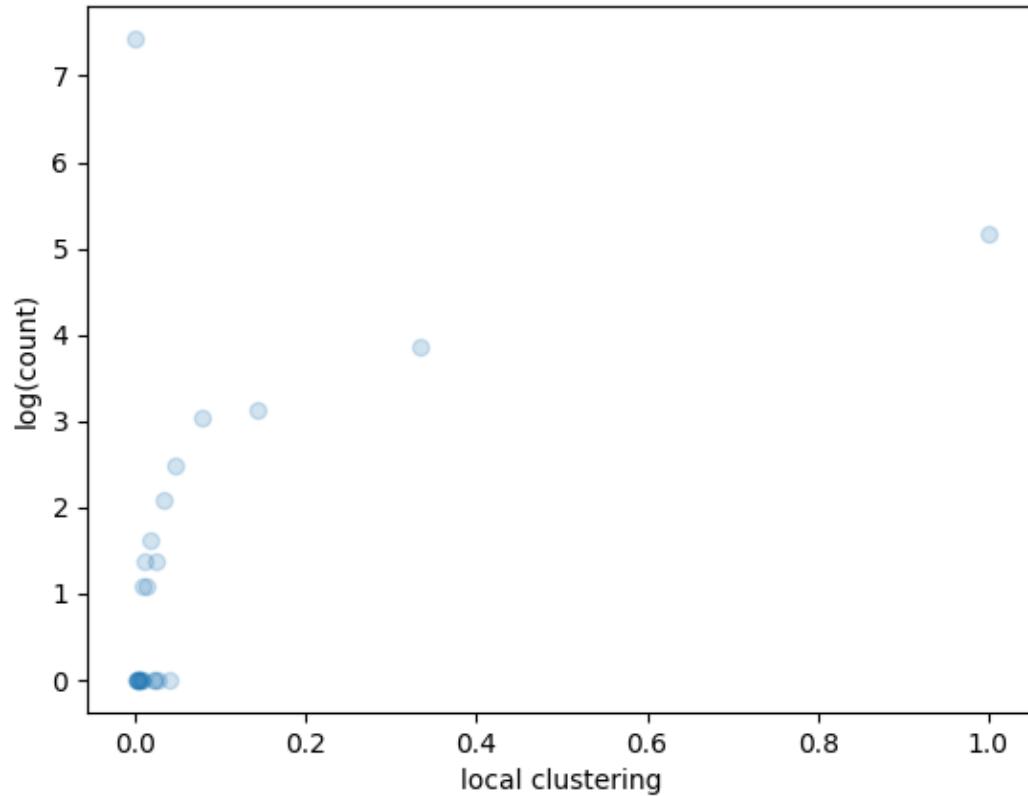
## original network

local clustering distribution plot, local clustering mean: 0.04,  
time: 0.082



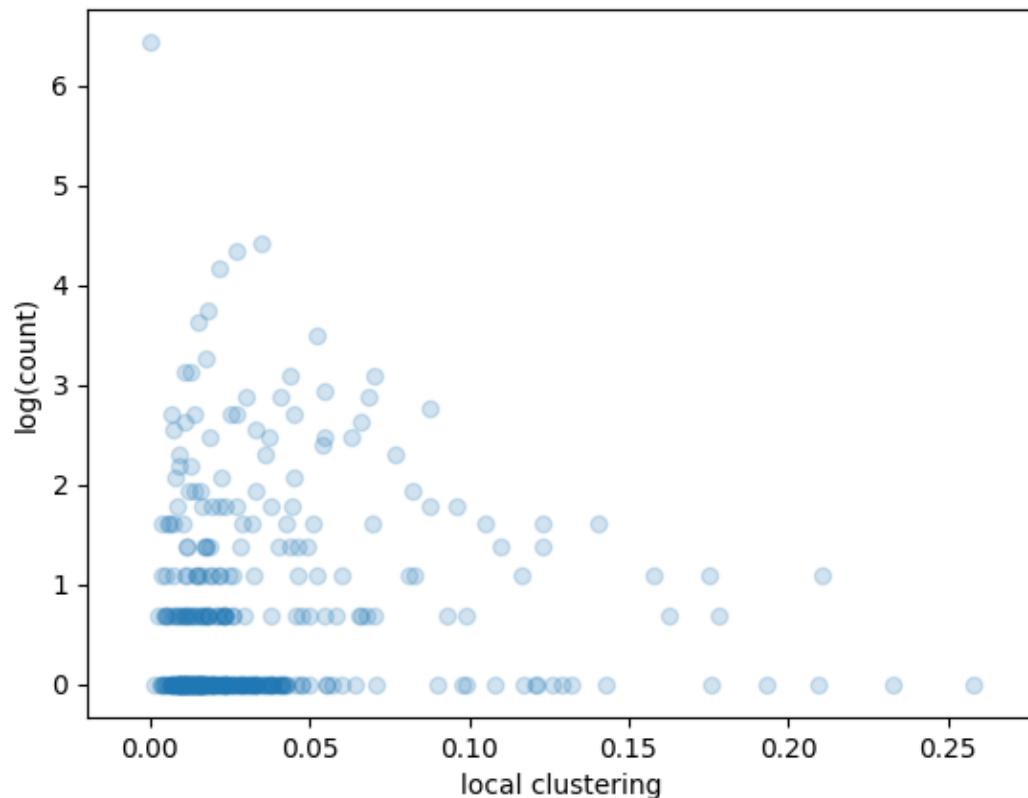
## BA network

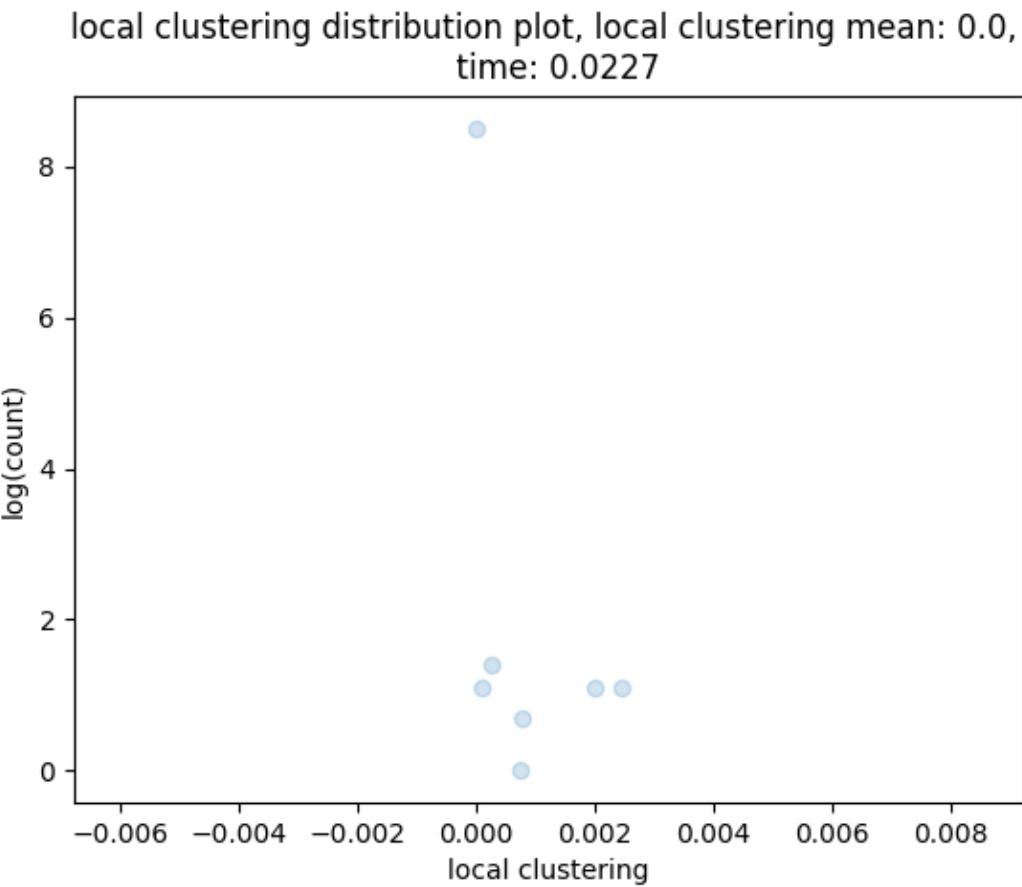
local clustering distribution plot, local clustering mean: 0.1,  
time: 0.0165



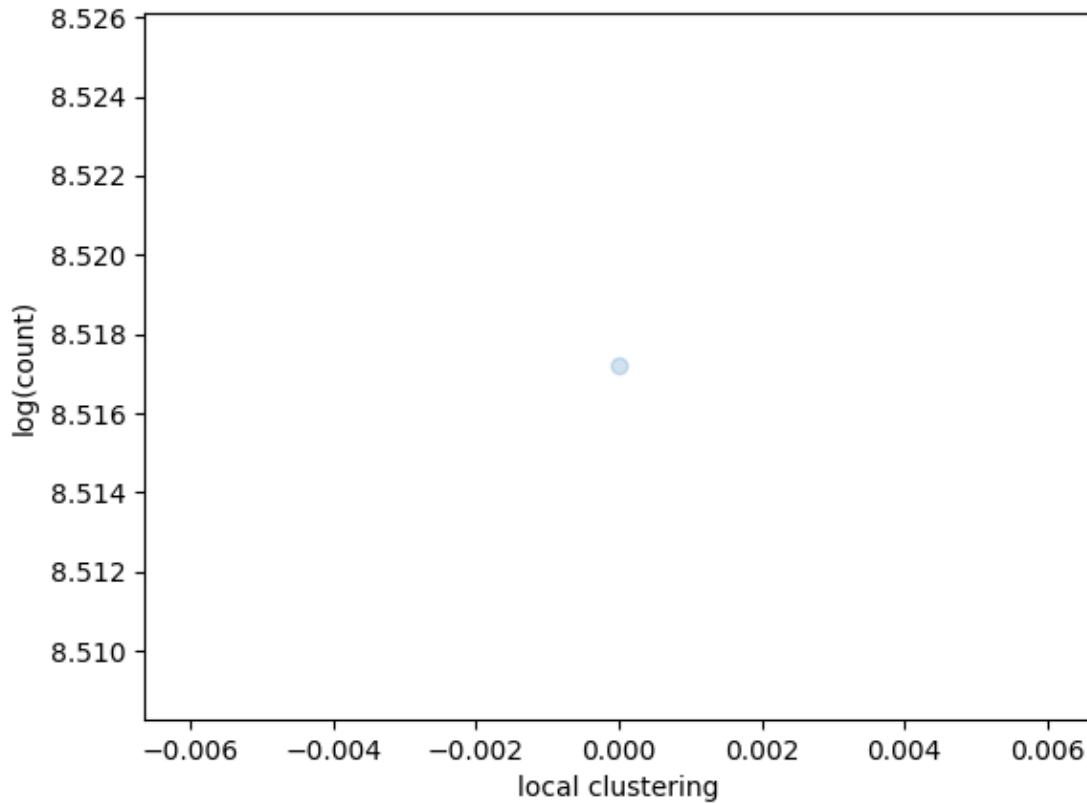
## modified BA network

local clustering distribution plot, local clustering mean: 0.02,  
time: 0.2683



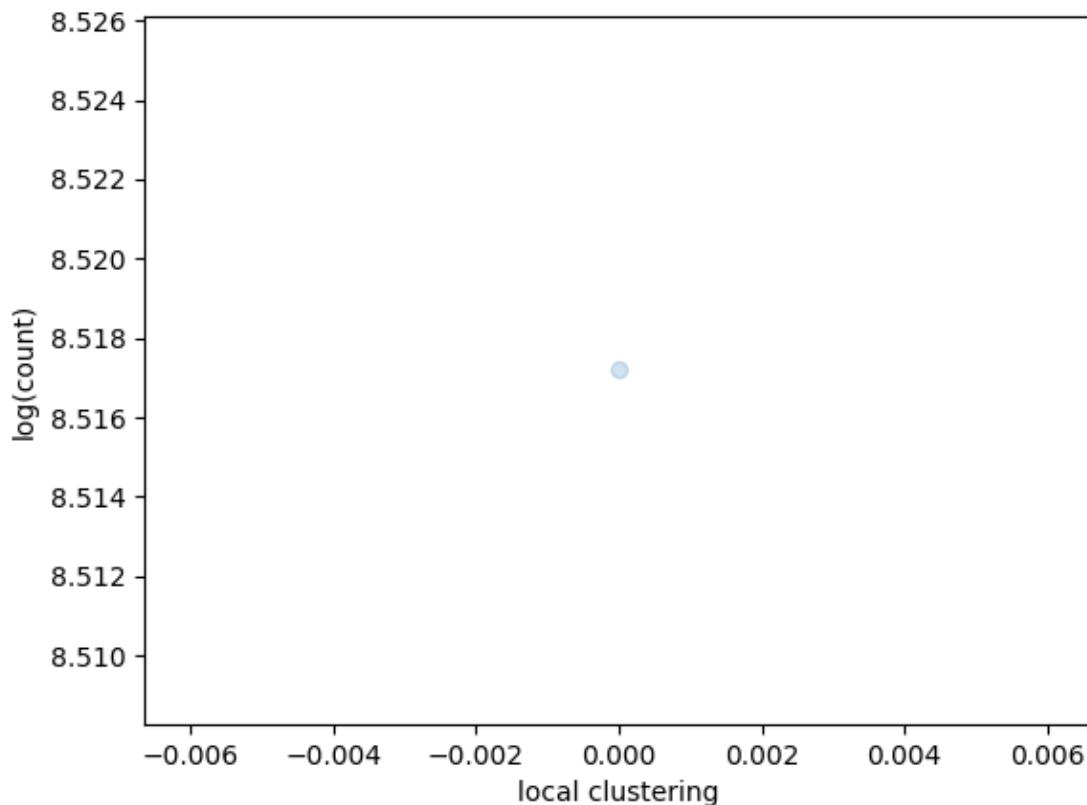
**WWW****original network****BA network**

local clustering distribution plot, local clustering mean: 0.0,  
time: 0.0151



## modified BA network

local clustering distribution plot, local clustering mean: 0.0,  
time: 0.0154

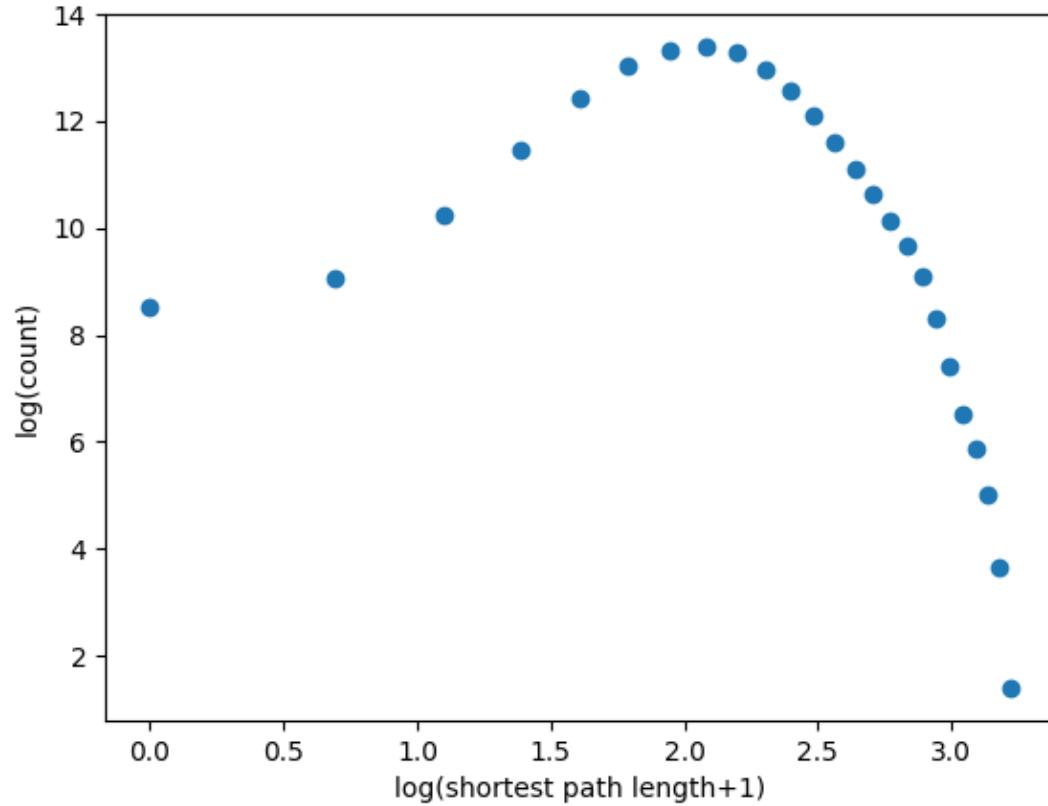


**path**

**collaboration**

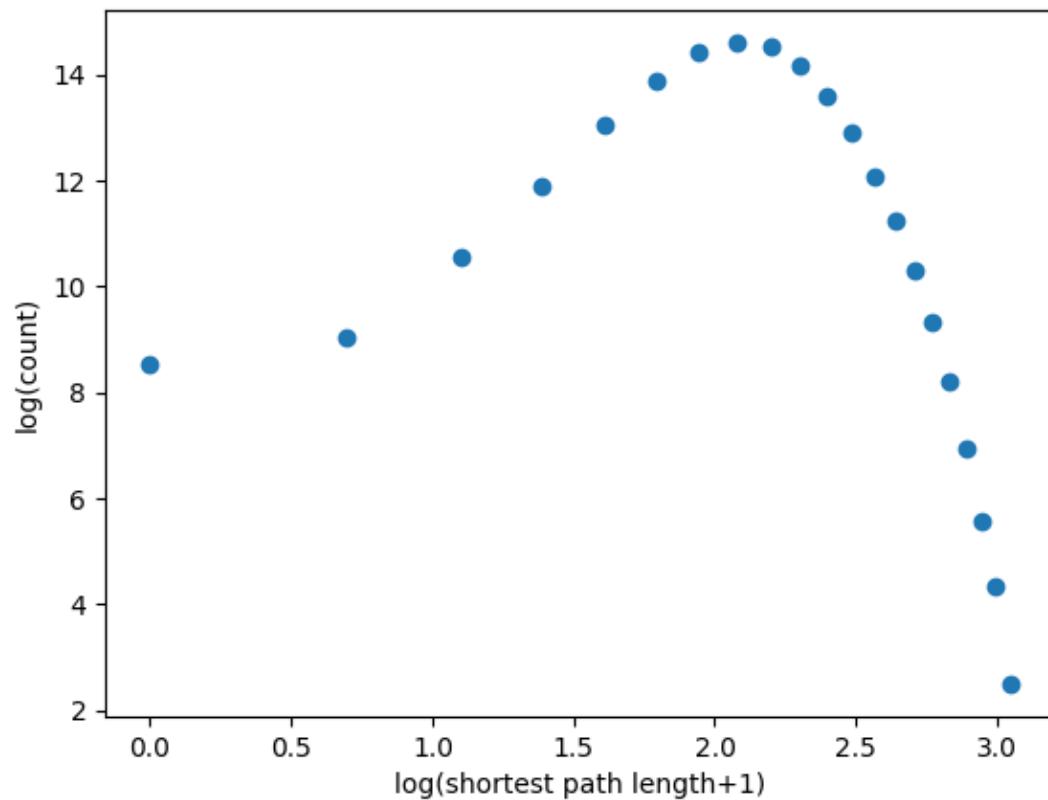
**original network**

all pair shortest path length distribution plot, shortest path length mean: 7.4!  
time : 3.6104



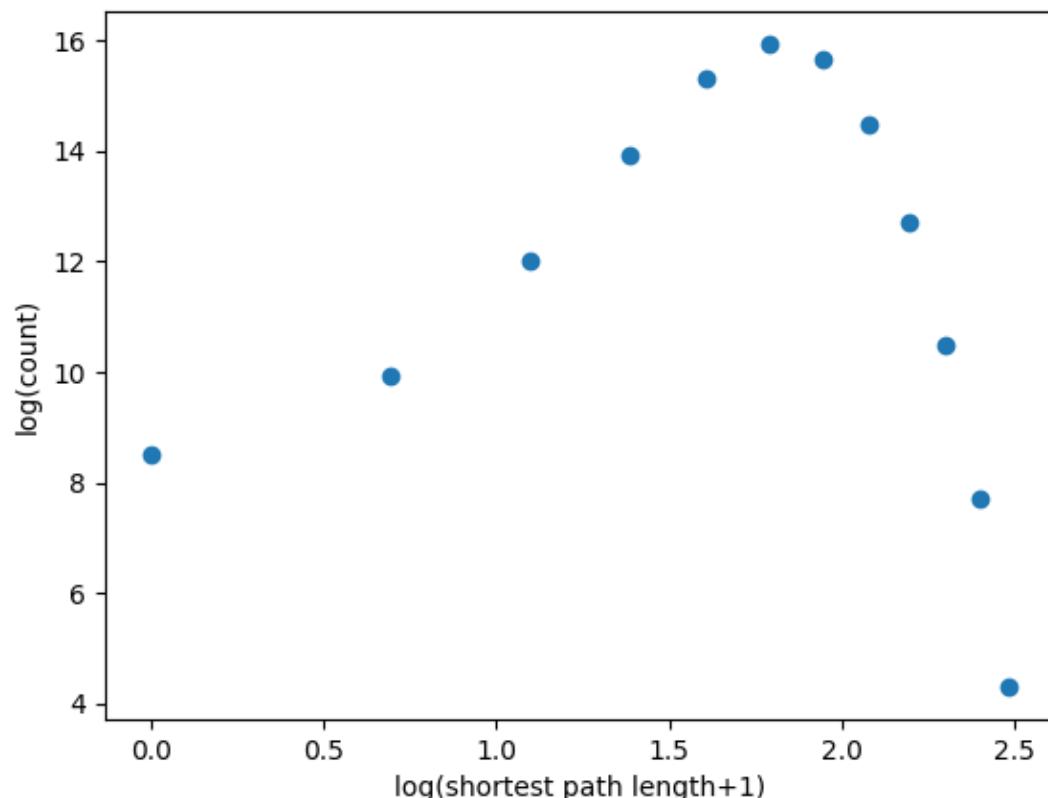
**BA network**

all pair shortest path length distribution plot, shortest path length mean: 7.4  
time : 9.4146



### modified BA network

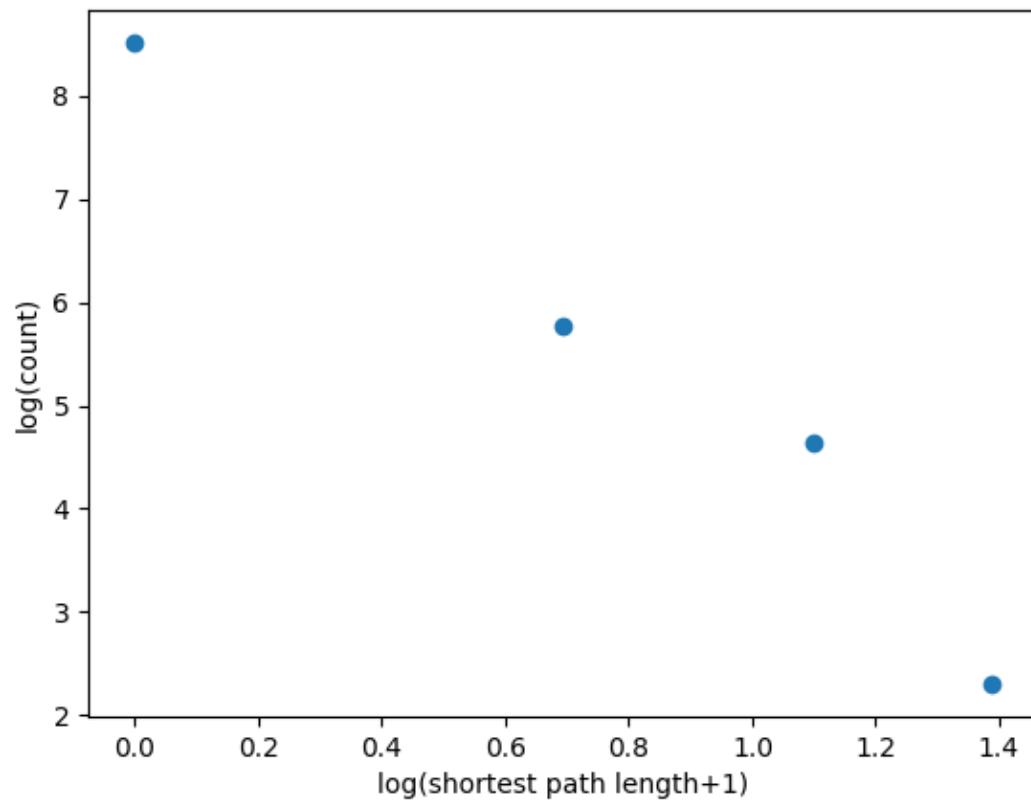
all pair shortest path length distribution plot, shortest path length mean: 5.1  
time : 30.0717



## citation

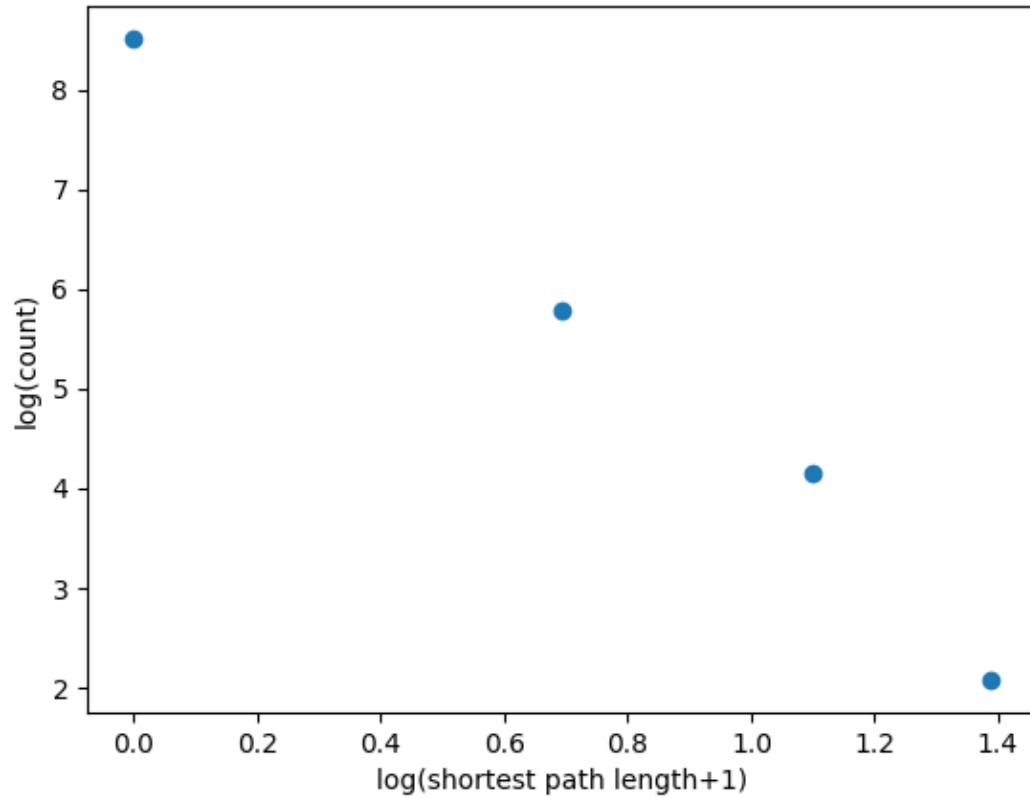
### original network

all pair shortest path length distribution plot, shortest path length mean: 0.1  
time : 0.0144



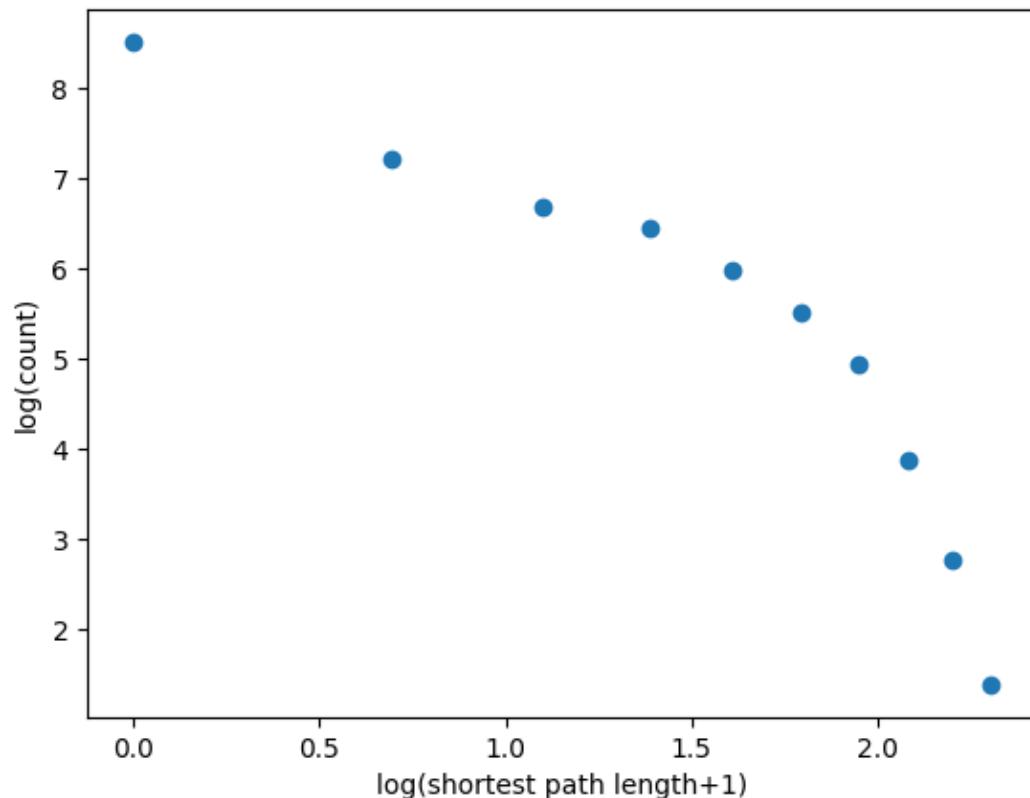
### BA network

all pair shortest path length distribution plot, shortest path length mean: 0.0!  
time : 0.0218



### modified BA network

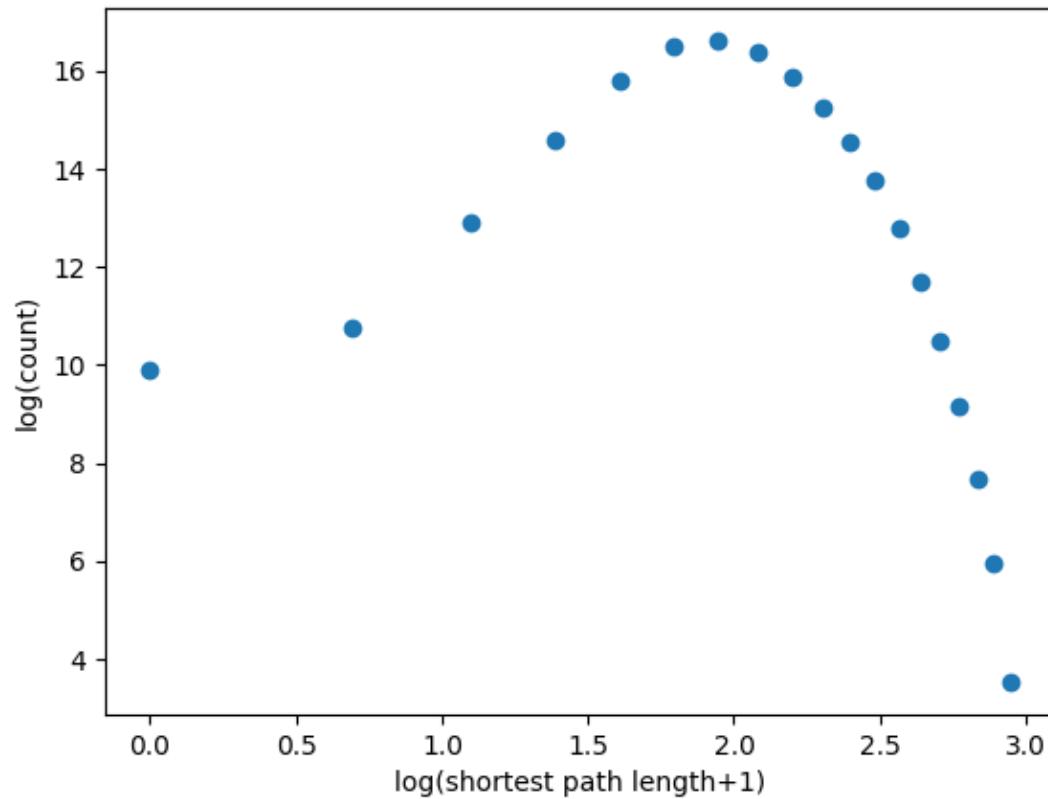
all pair shortest path length distribution plot, shortest path length mean: 1.0!  
time : 0.0194



## actor

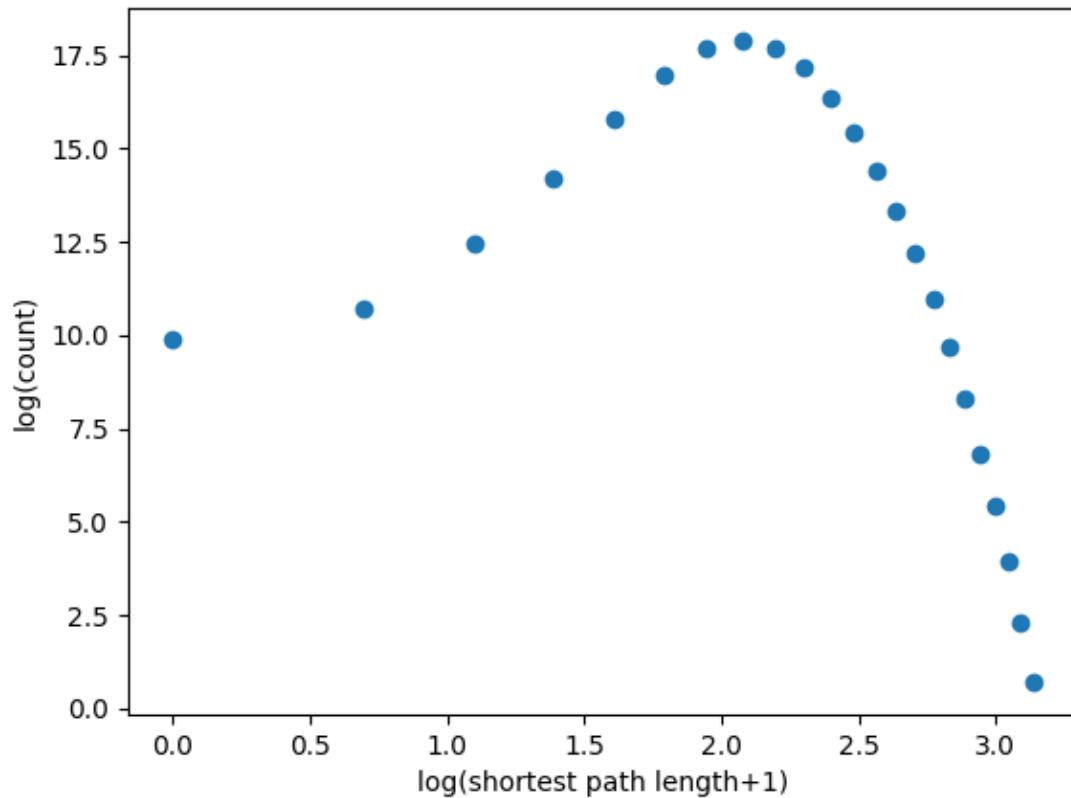
### original network

all pair shortest path length distribution plot, shortest path length mean: 6.29  
time : 82.6954



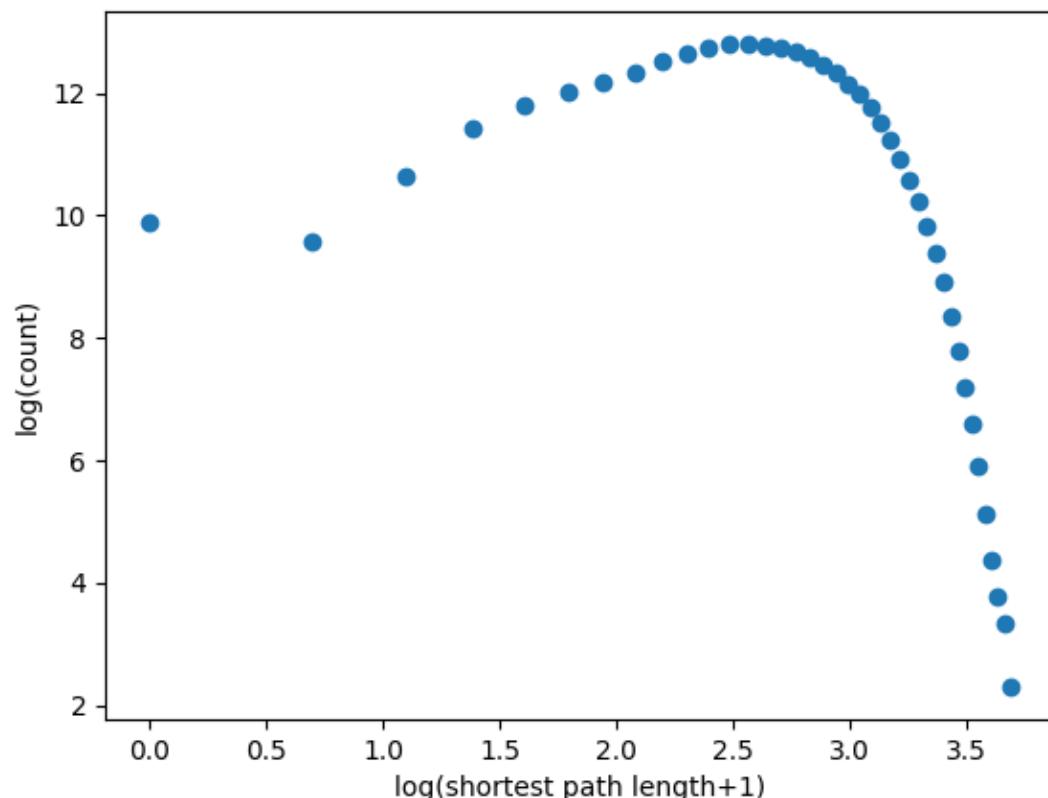
### BA network

all pair shortest path length distribution plot, shortest path length mean: 7.21  
time : 246.4448



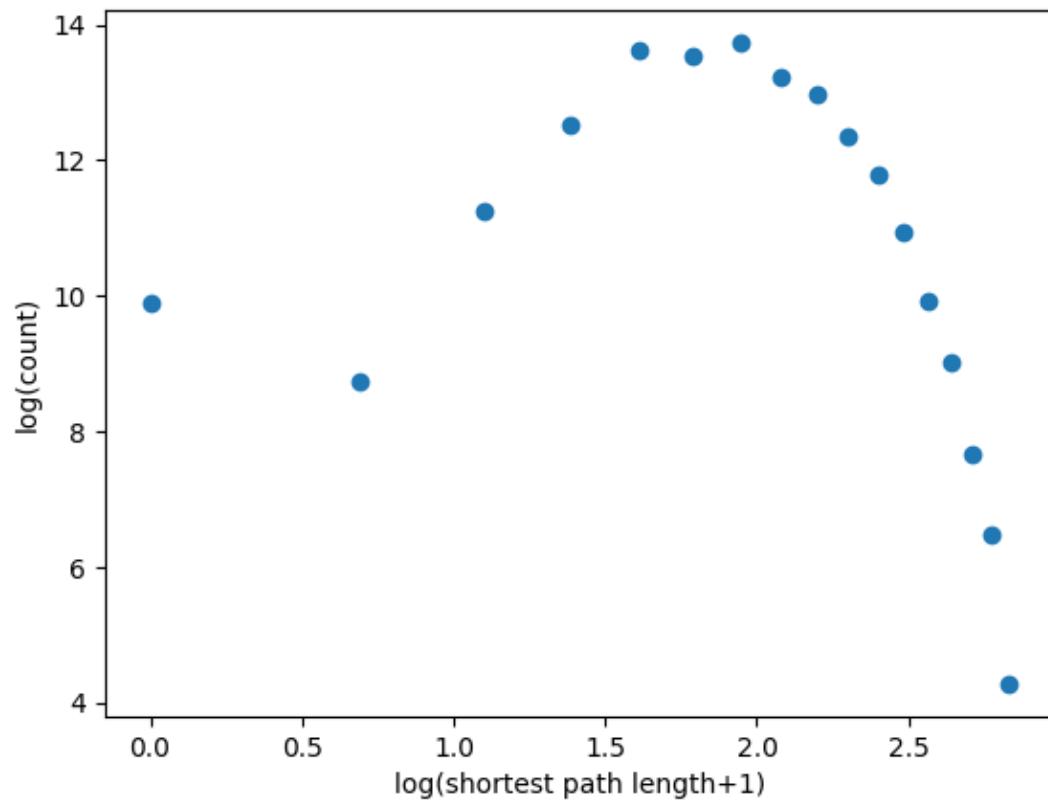
### modified BA network

all pair shortest path length distribution plot, shortest path length mean: 12.8  
time : 4.2592

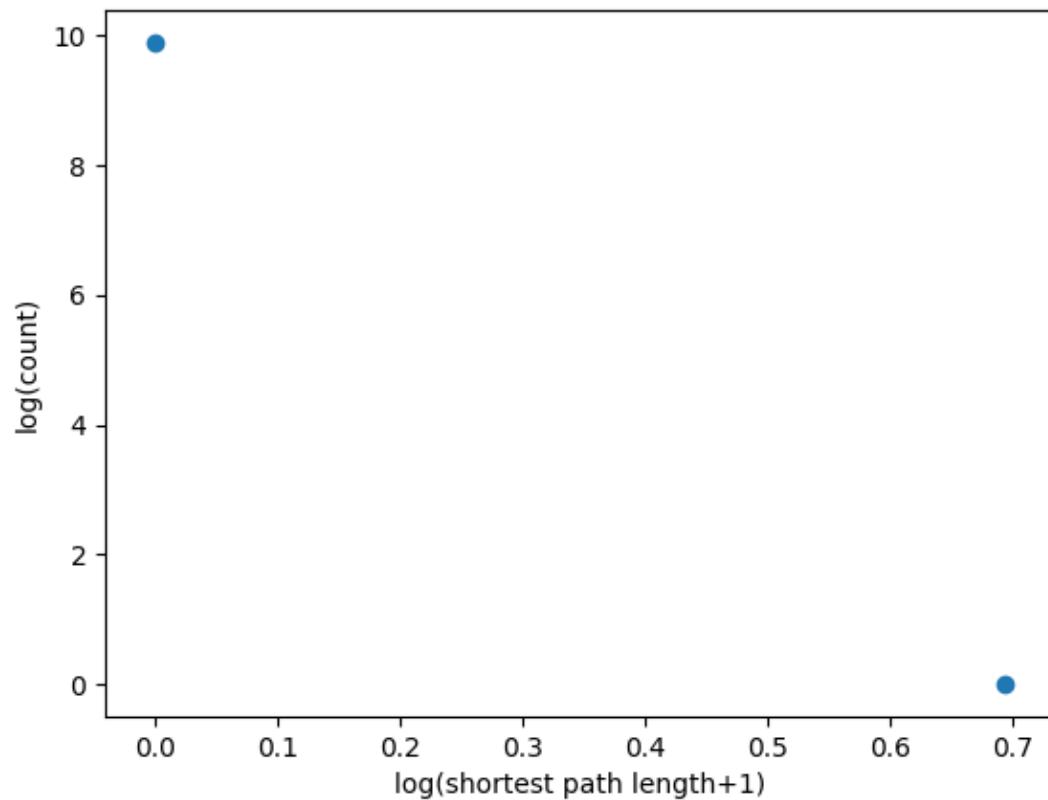


**email****original network**

all pair shortest path length distribution plot, shortest path length mean: 5.89  
time : 3.3552

**BA network**

all pair shortest path length distribution plot, shortest path length mean: 0.0  
time : 0.1157



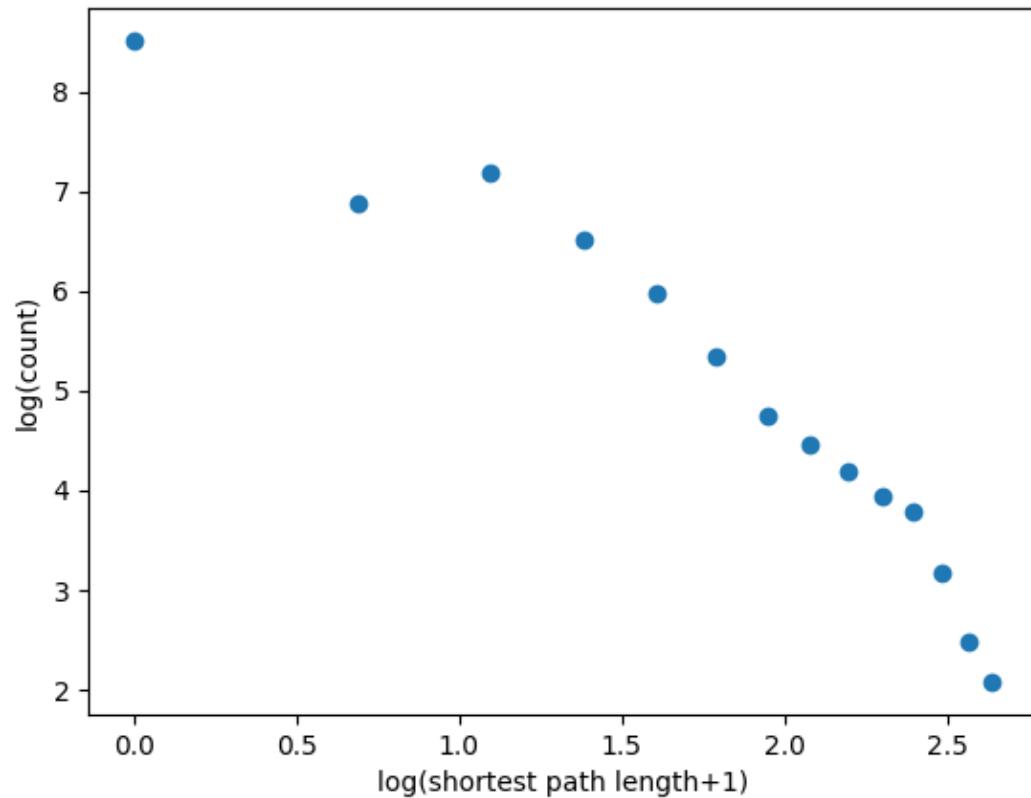
### modified BA network



### internet

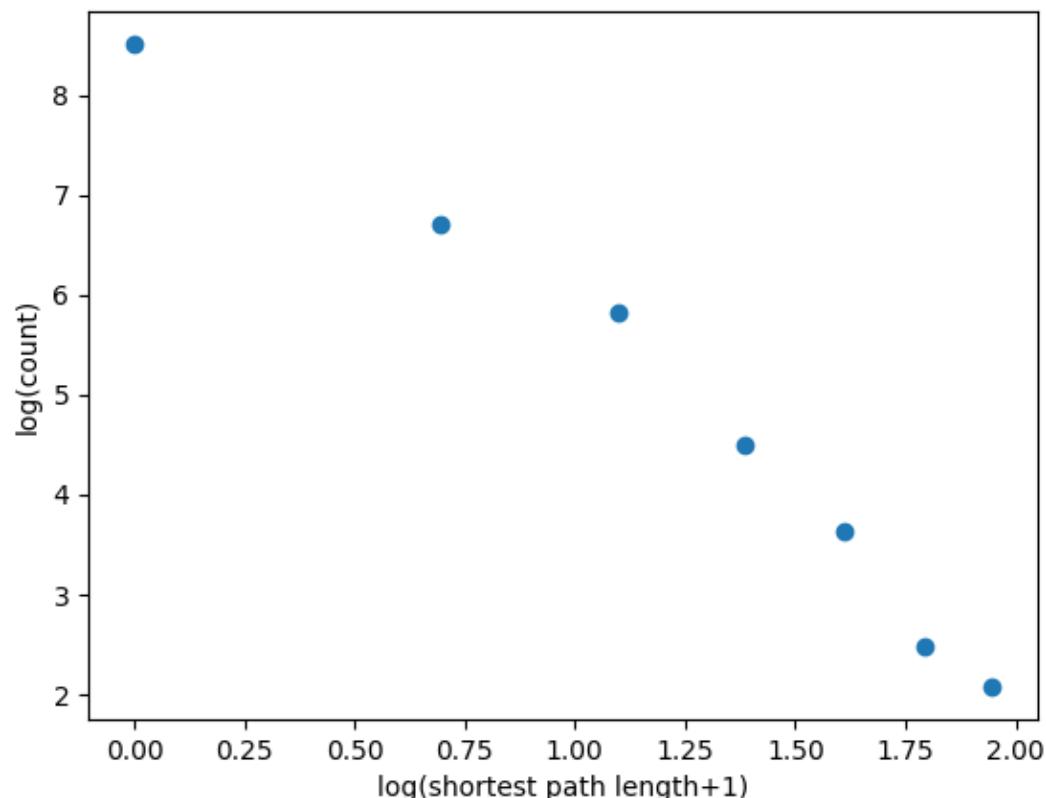
### original network

all pair shortest path length distribution plot, shortest path length mean: 1.28  
time : 0.0188



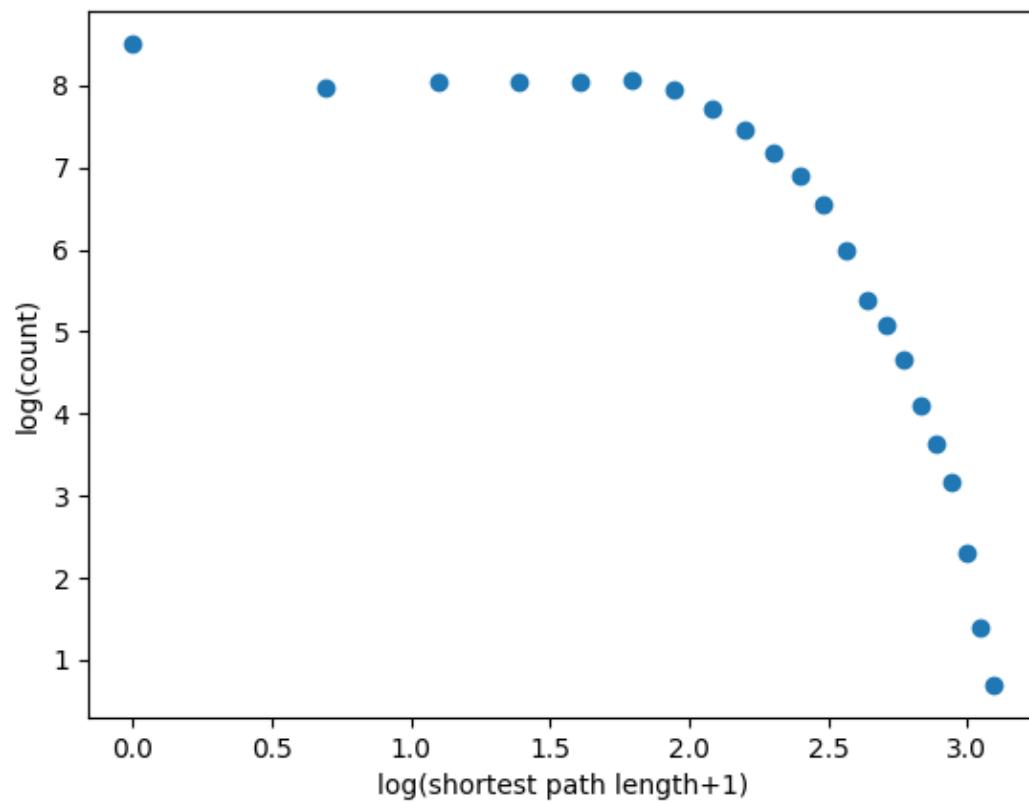
## BA network

all pair shortest path length distribution plot, shortest path length mean: 0.31  
time : 0.0293



## modified BA network

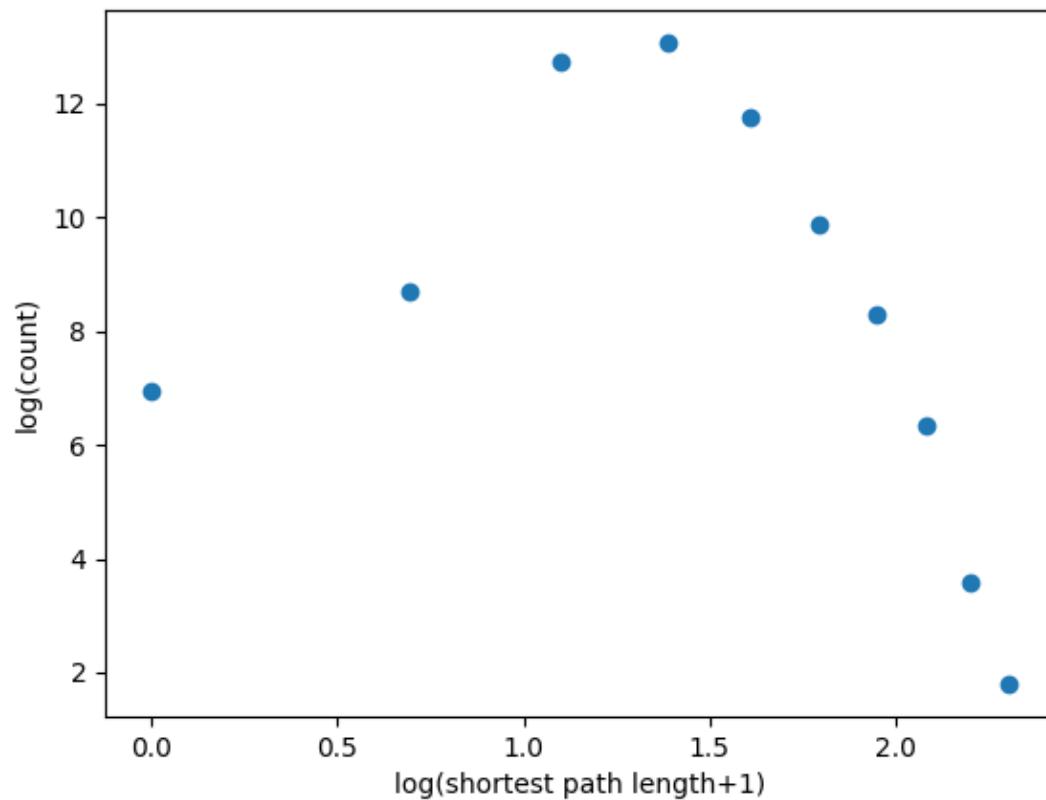
all pair shortest path length distribution plot, shortest path length mean: 4.3  
time : 0.0566



## metabolic

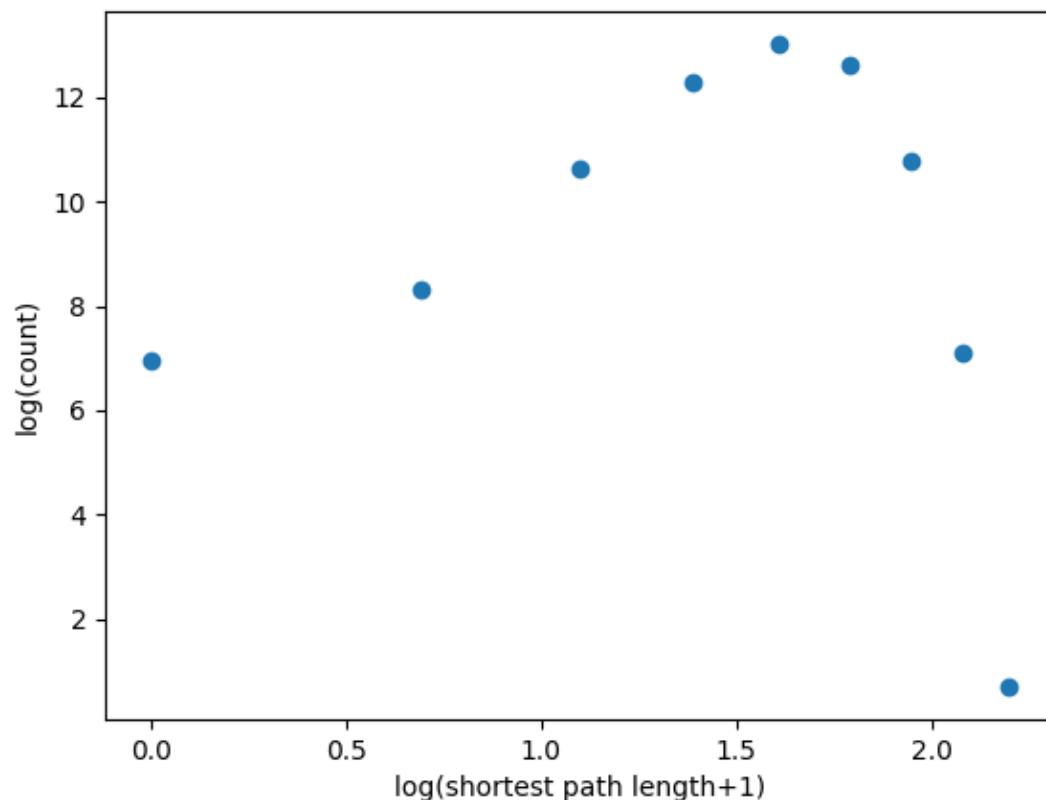
### original network

all pair shortest path length distribution plot, shortest path length mean: 2.81  
time : 1.0824



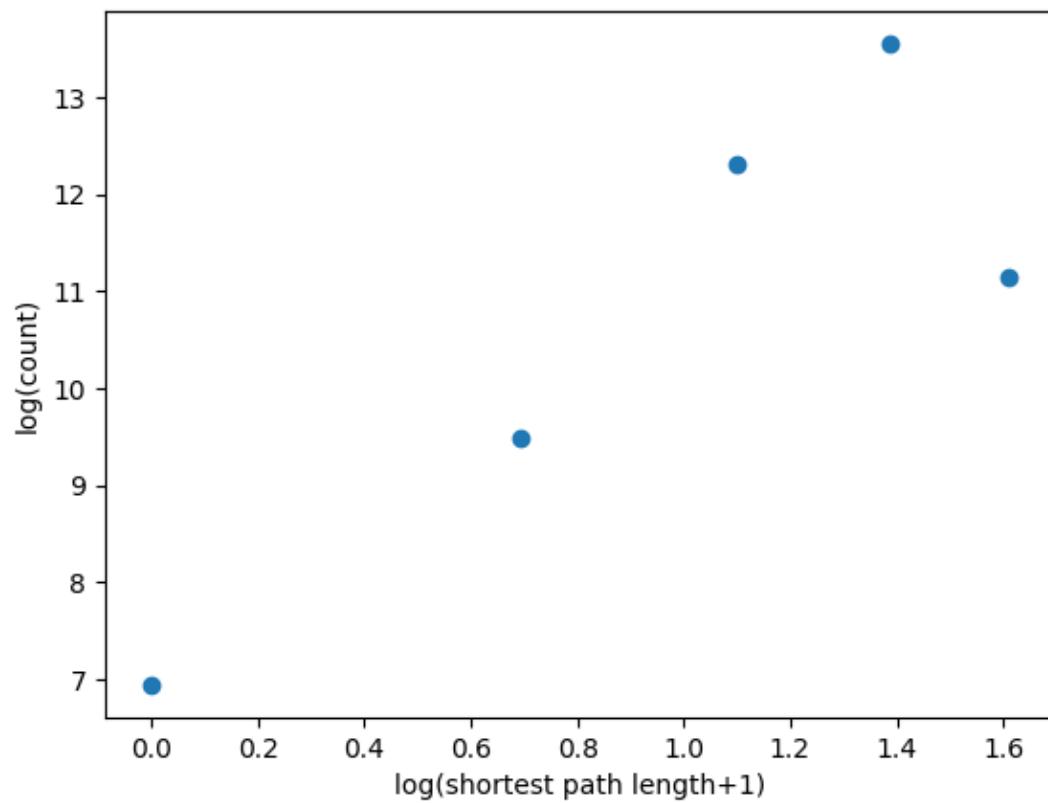
## BA network

all pair shortest path length distribution plot, shortest path length mean: 4.01  
time : 1.0575



## modified BA network

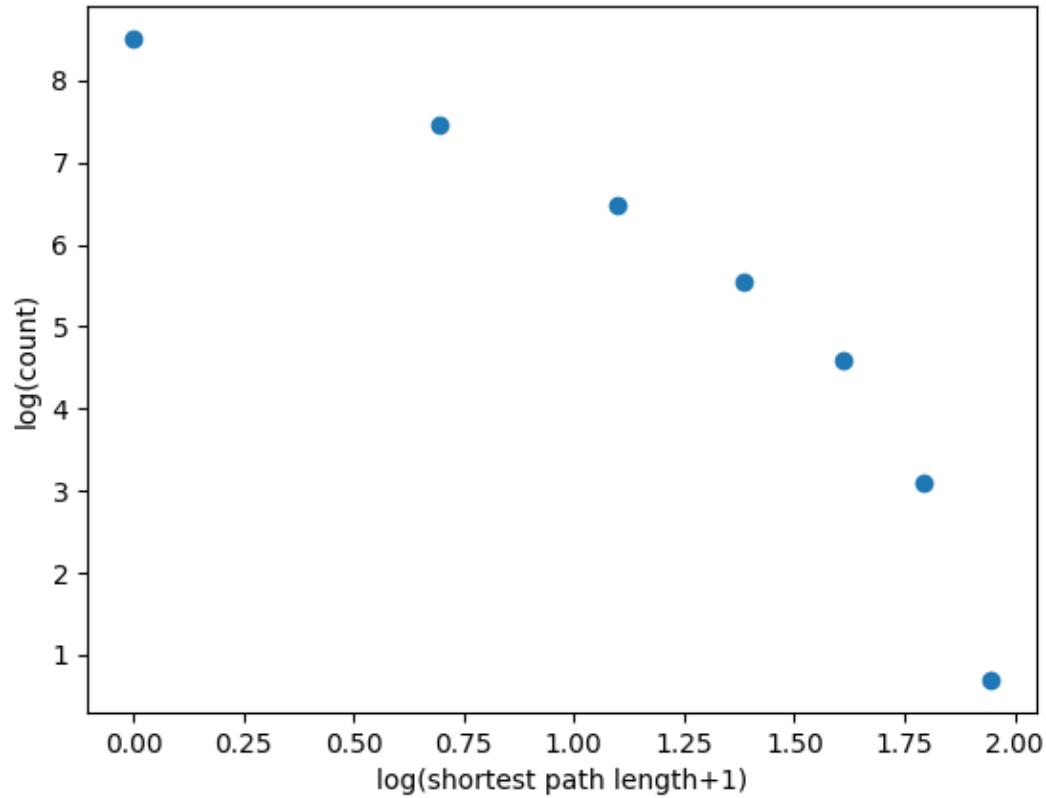
all pair shortest path length distribution plot, shortest path length mean: 2.81  
time : 1.8511



## phonecalls

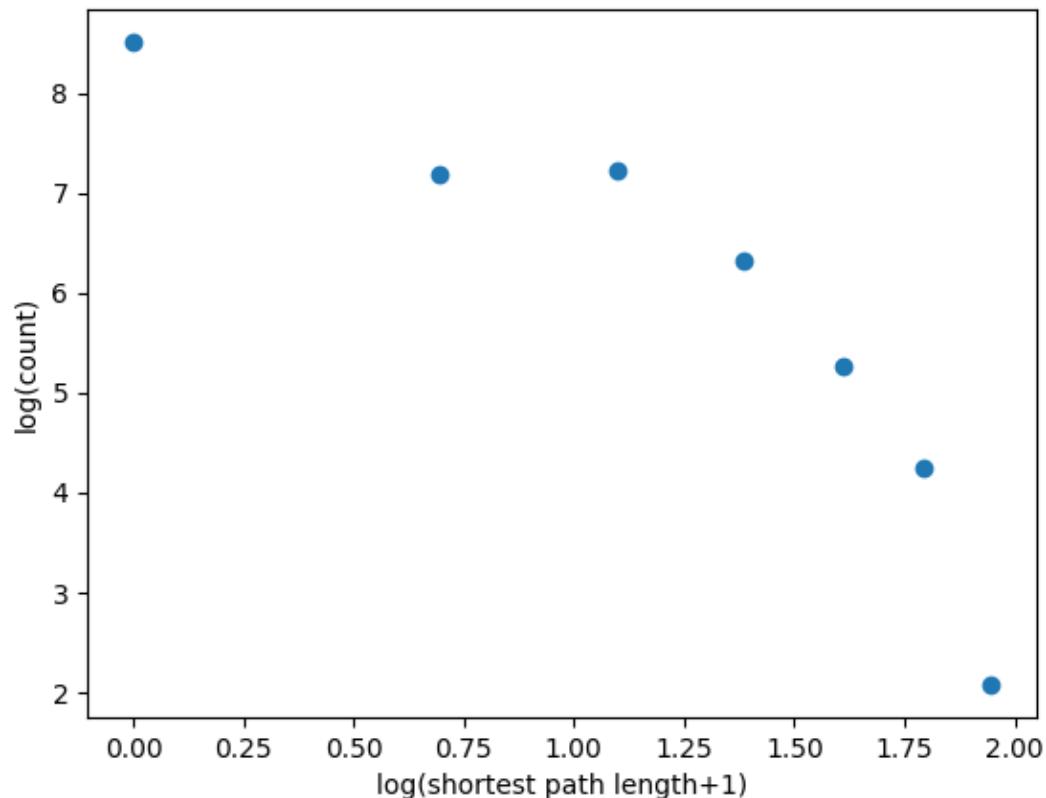
### original network

all pair shortest path length distribution plot, shortest path length mean: 0.51  
time : 0.0374



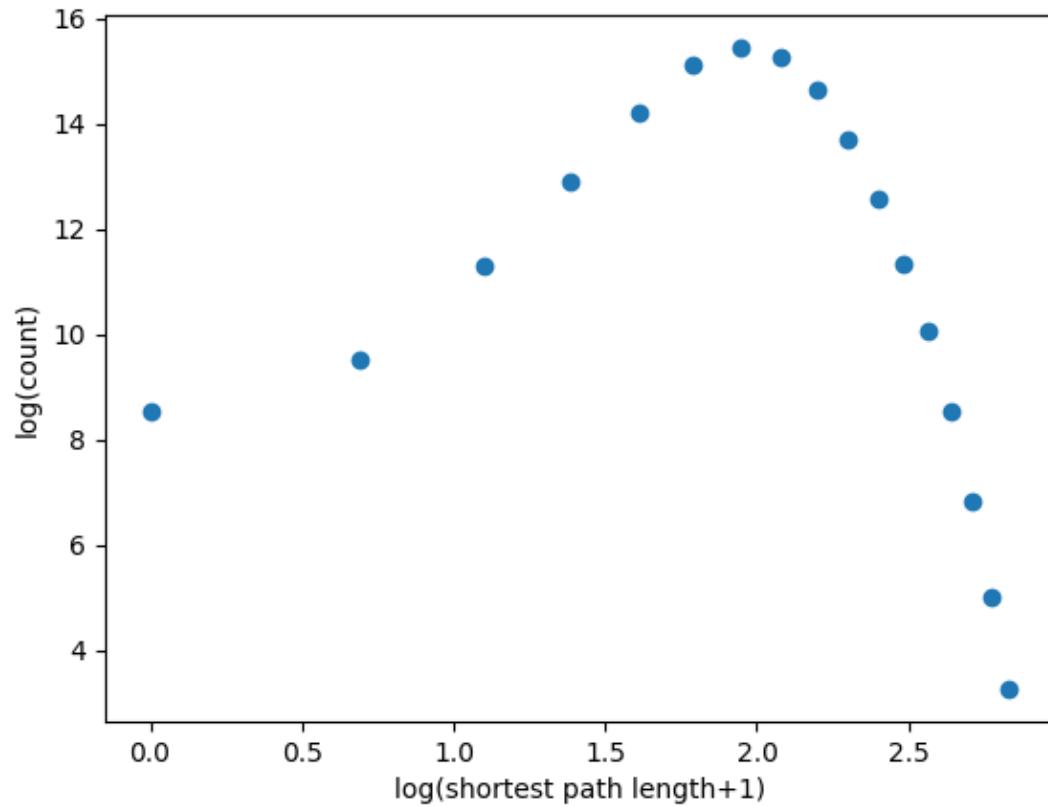
## BA network

all pair shortest path length distribution plot, shortest path length mean: 0.81  
time : 0.0345



## modified BA network

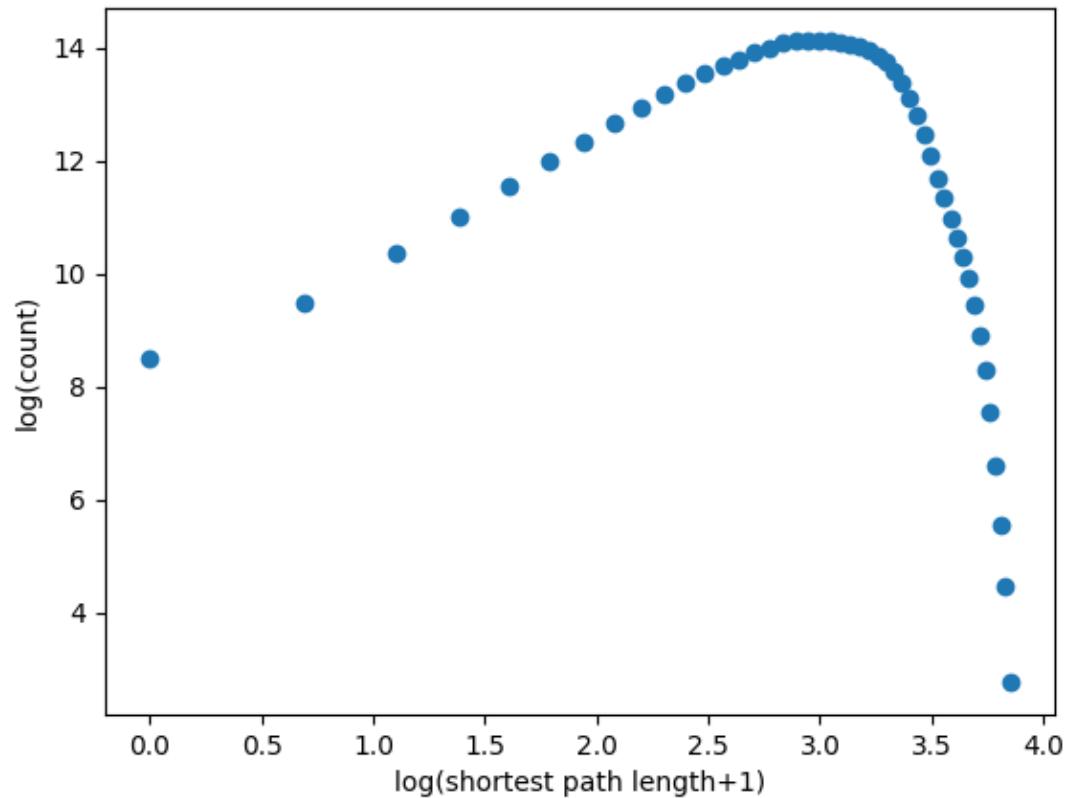
all pair shortest path length distribution plot, shortest path length mean: 6.27  
time : 27.8906



## powergrid

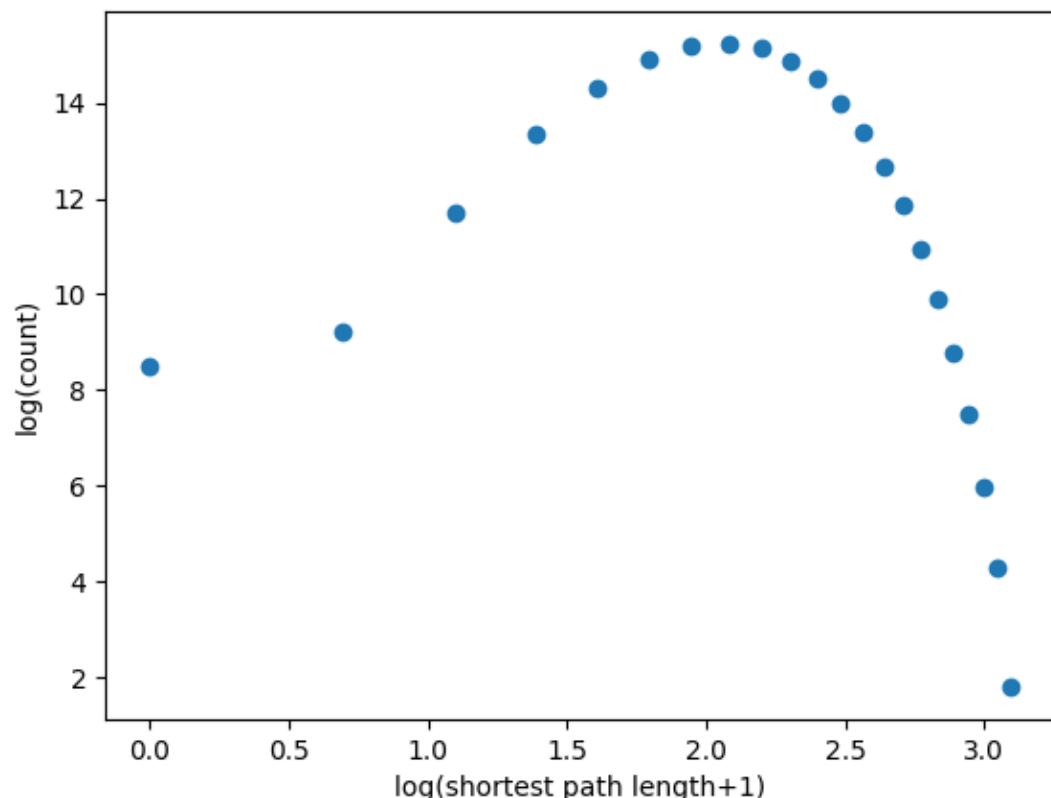
### original network

all pair shortest path length distribution plot, shortest path length mean: 18.9  
time : 22.7659



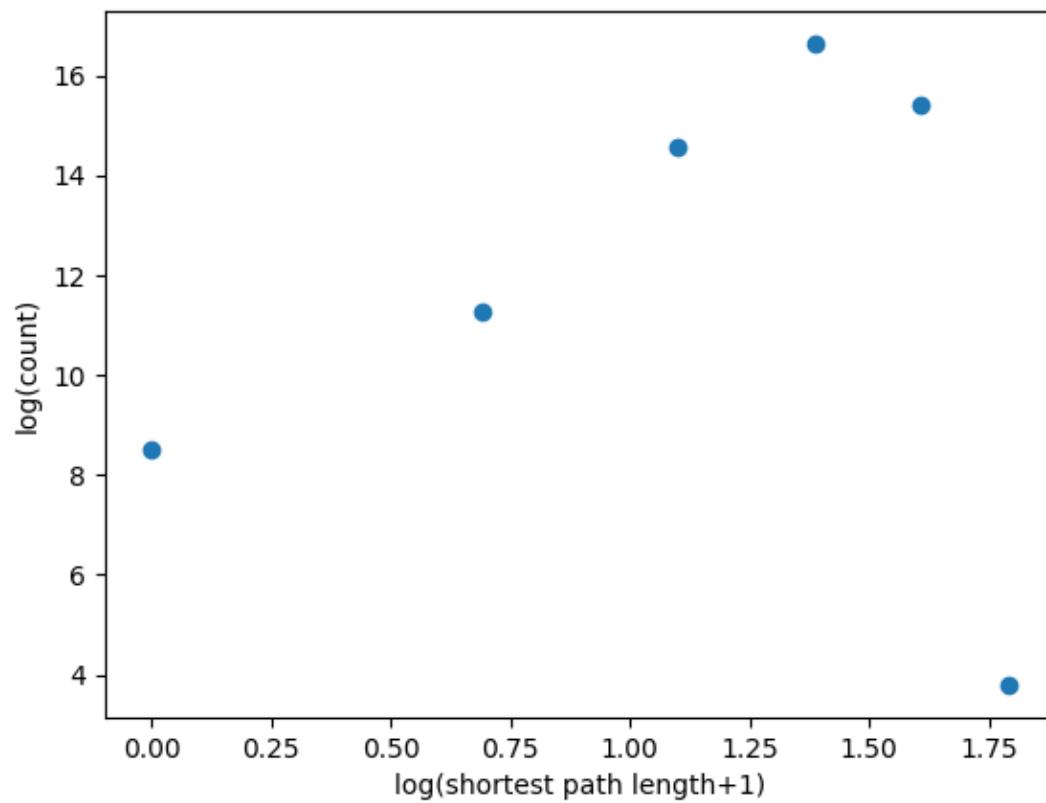
## BA network

all pair shortest path length distribution plot, shortest path length mean: 7.3  
time : 24.1837



## modified BA network

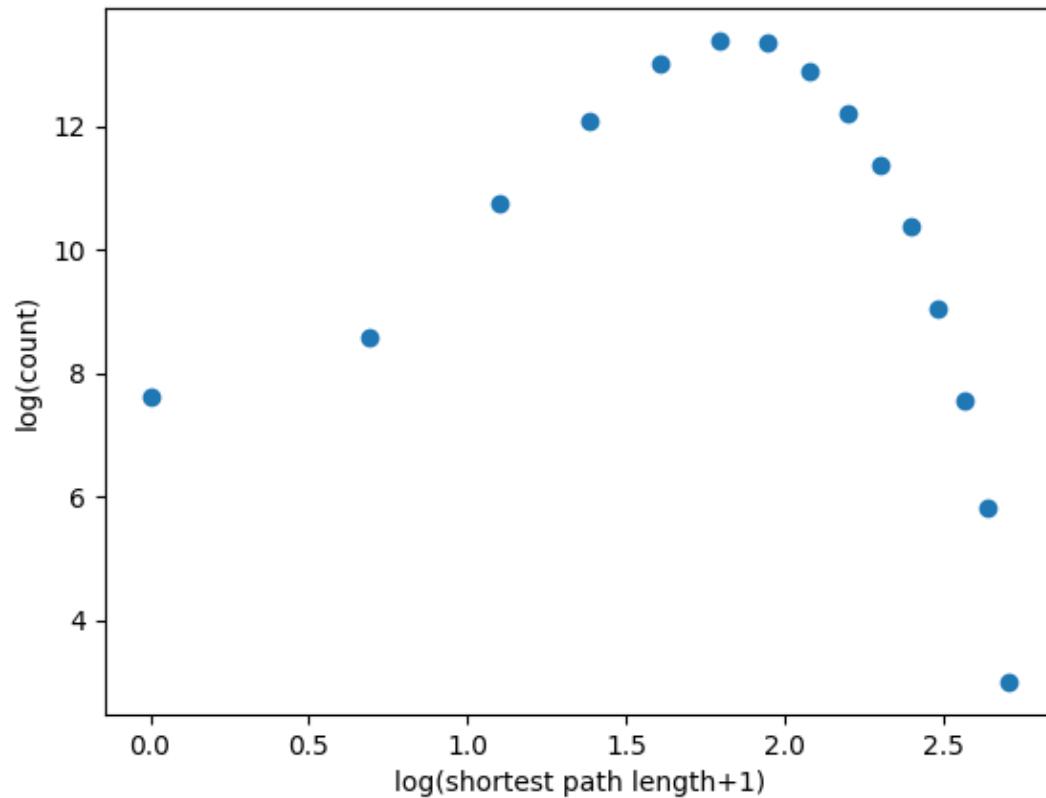
all pair shortest path length distribution plot, shortest path length mean: 3.1:  
time : 59.3067



## protein

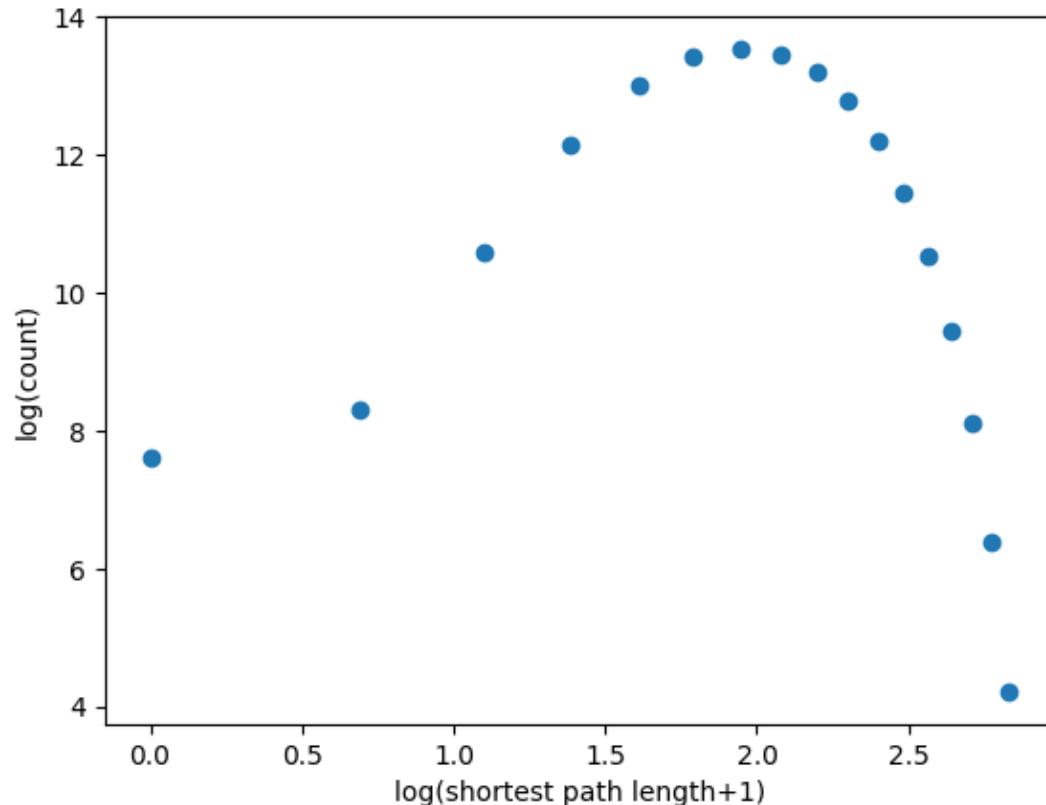
## original network

all pair shortest path length distribution plot, shortest path length mean: 5.61  
time : 3.3496



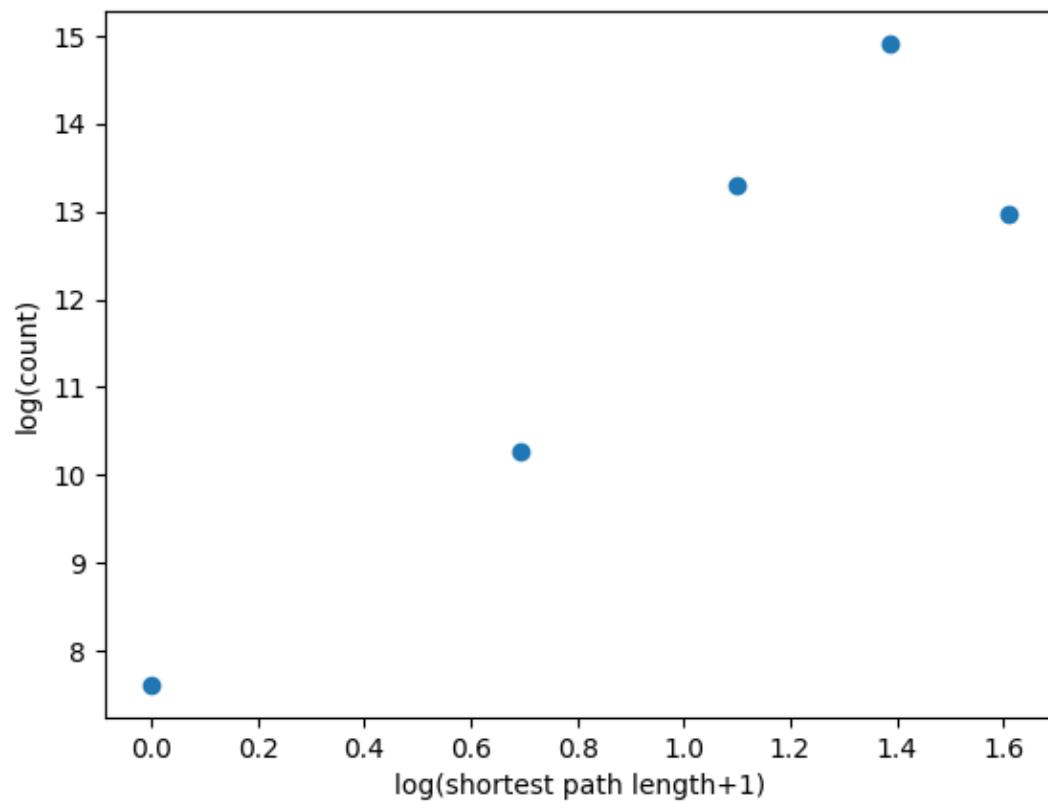
## BA network

all pair shortest path length distribution plot, shortest path length mean: 6.54  
time : 3.8661

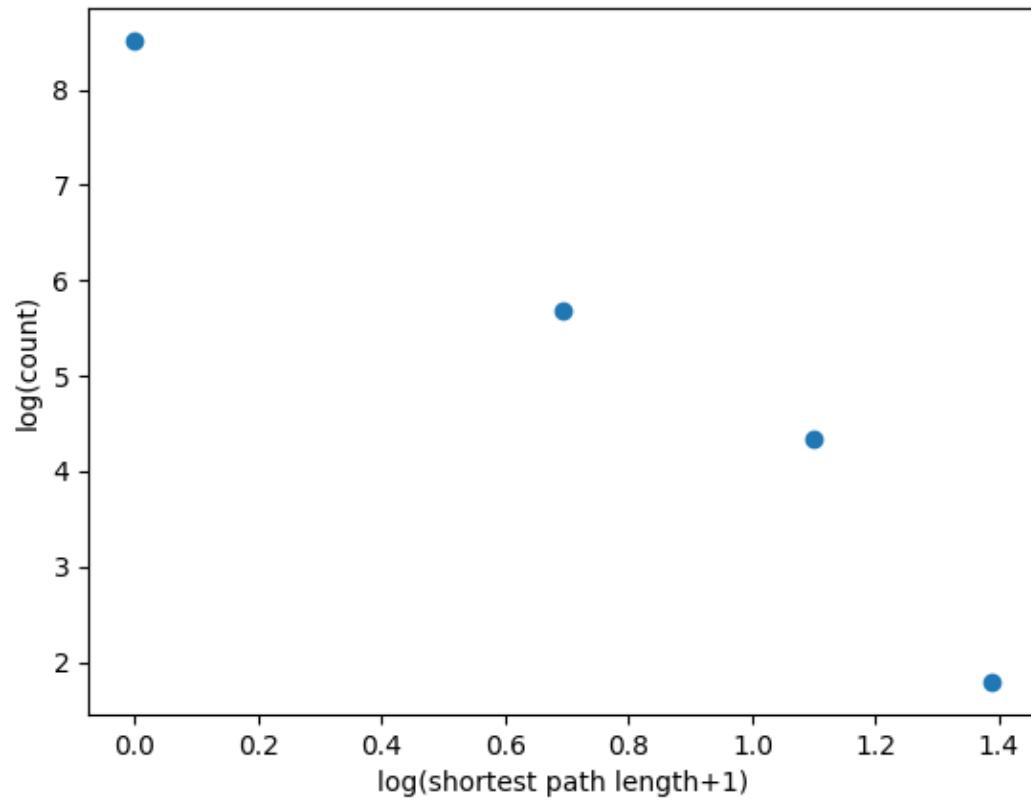


**modified BA network**

all pair shortest path length distribution plot, shortest path length mean: 2.94  
time : 12.6525

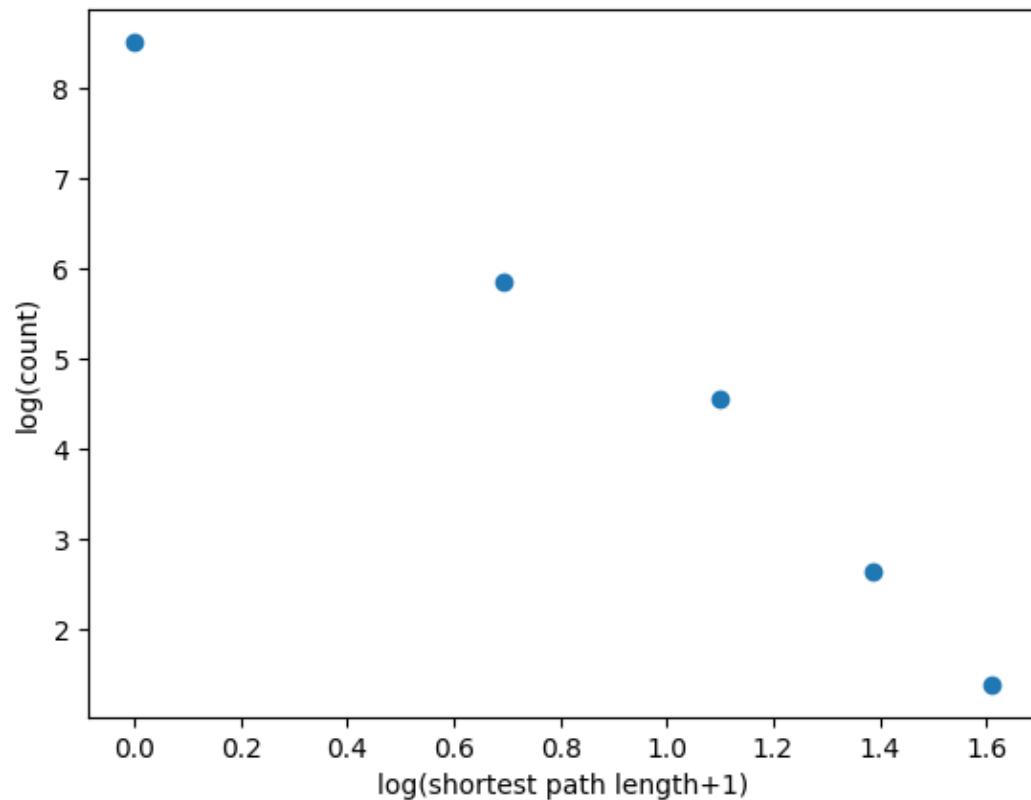
**WWW****original network**

all pair shortest path length distribution plot, shortest path length mean: 0.0!  
time : 0.013



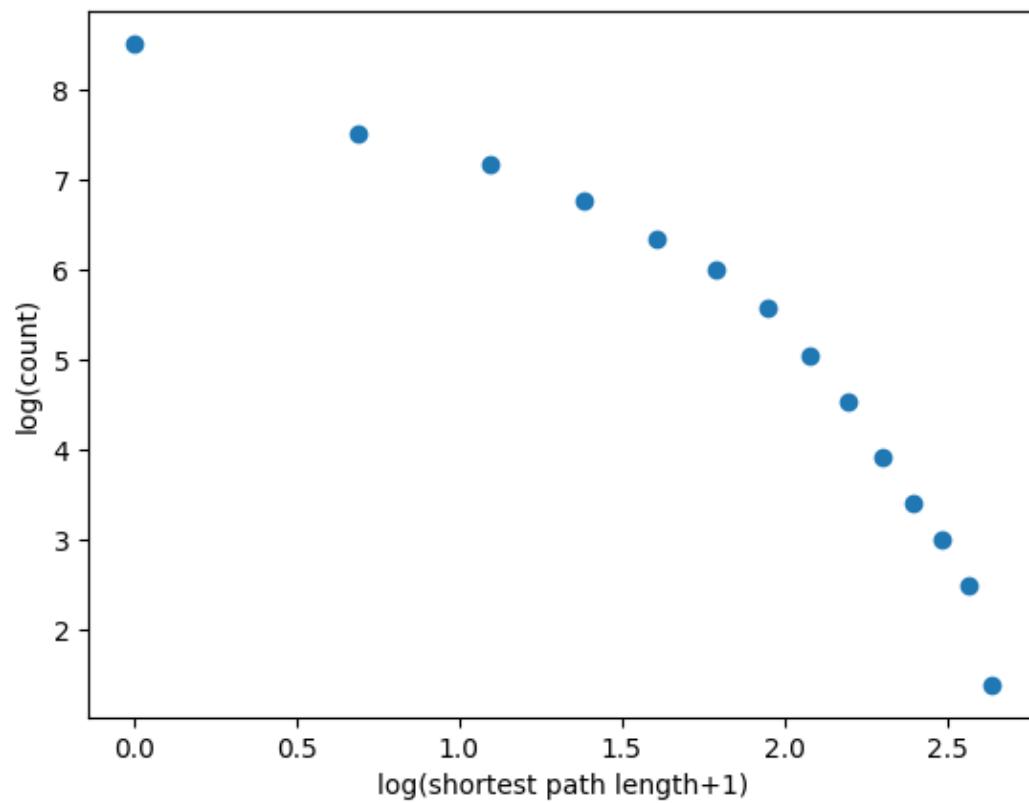
## BA network

all pair shortest path length distribution plot, shortest path length mean: 0.1!  
time : 0.0123



## modified BA network

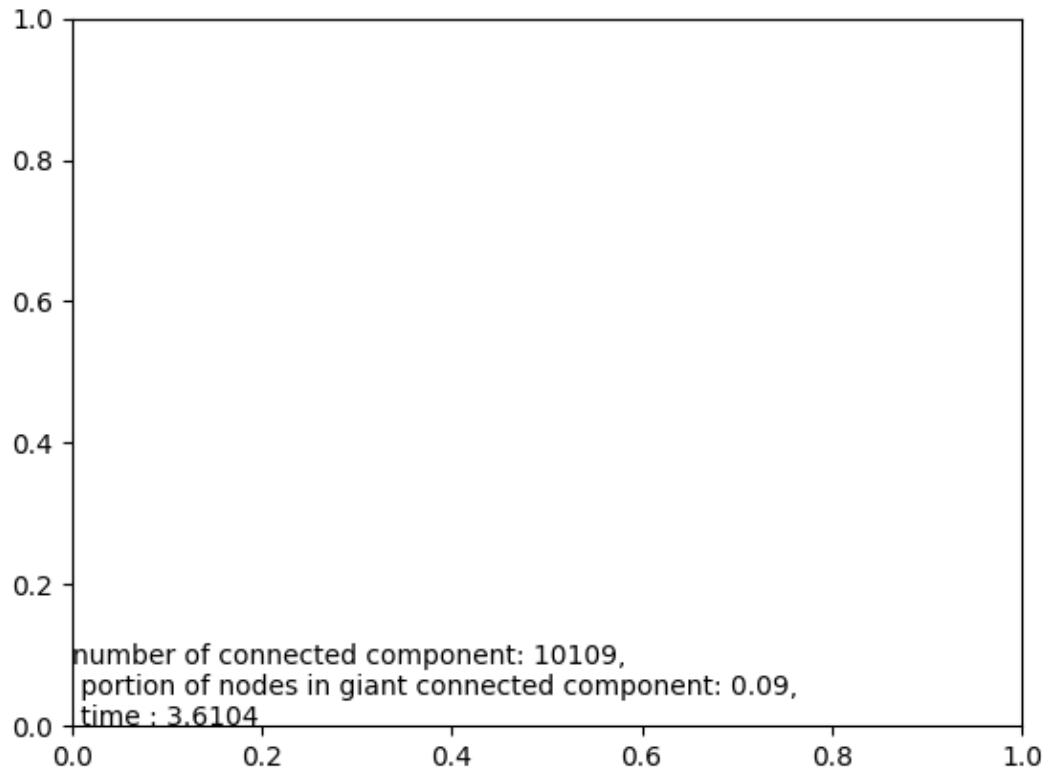
all pair shortest path length distribution plot, shortest path length mean: 1.5  
time : 0.0228



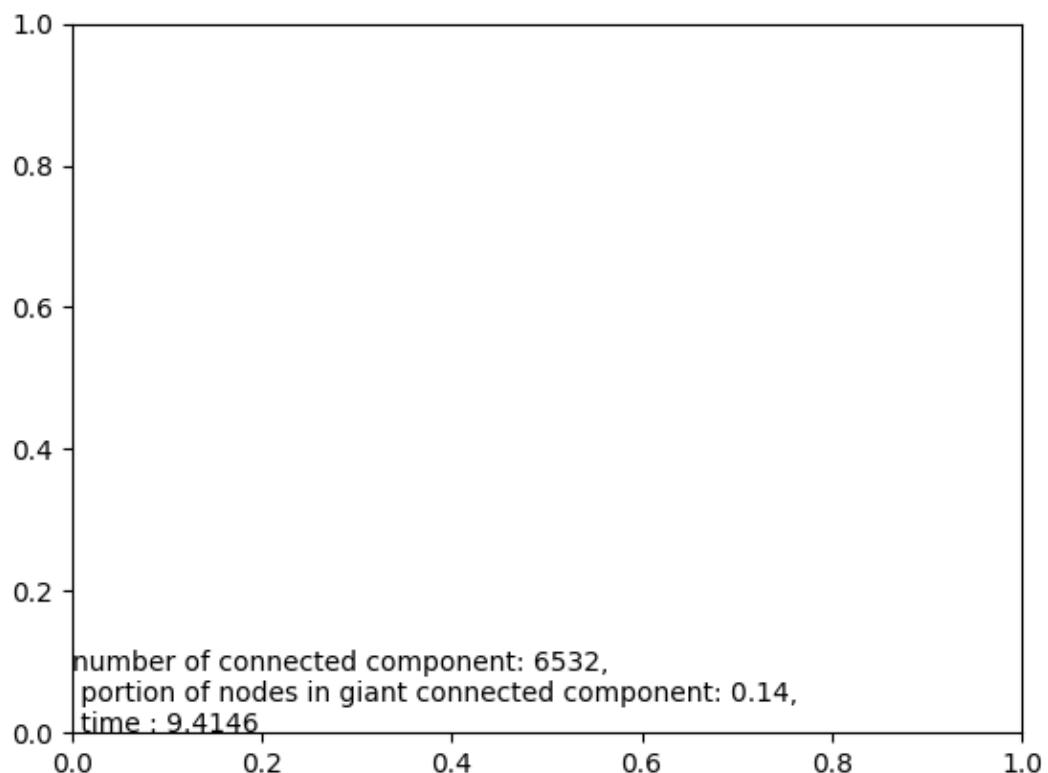
## component

### collaboration

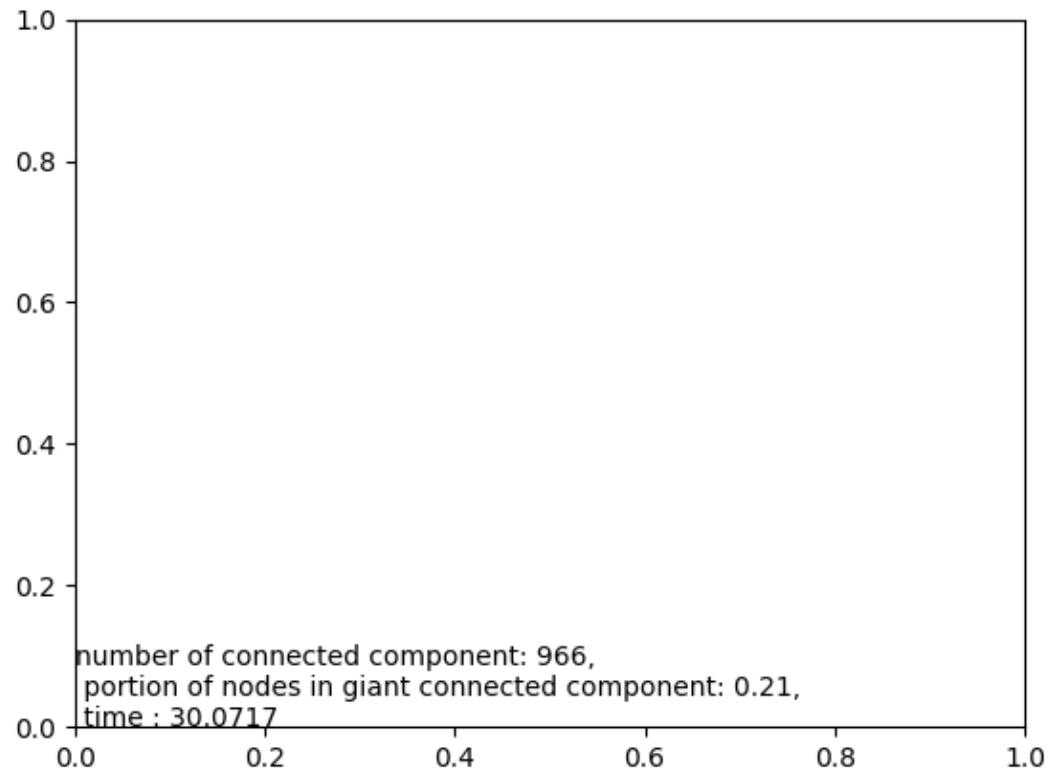
### original network



## BA network

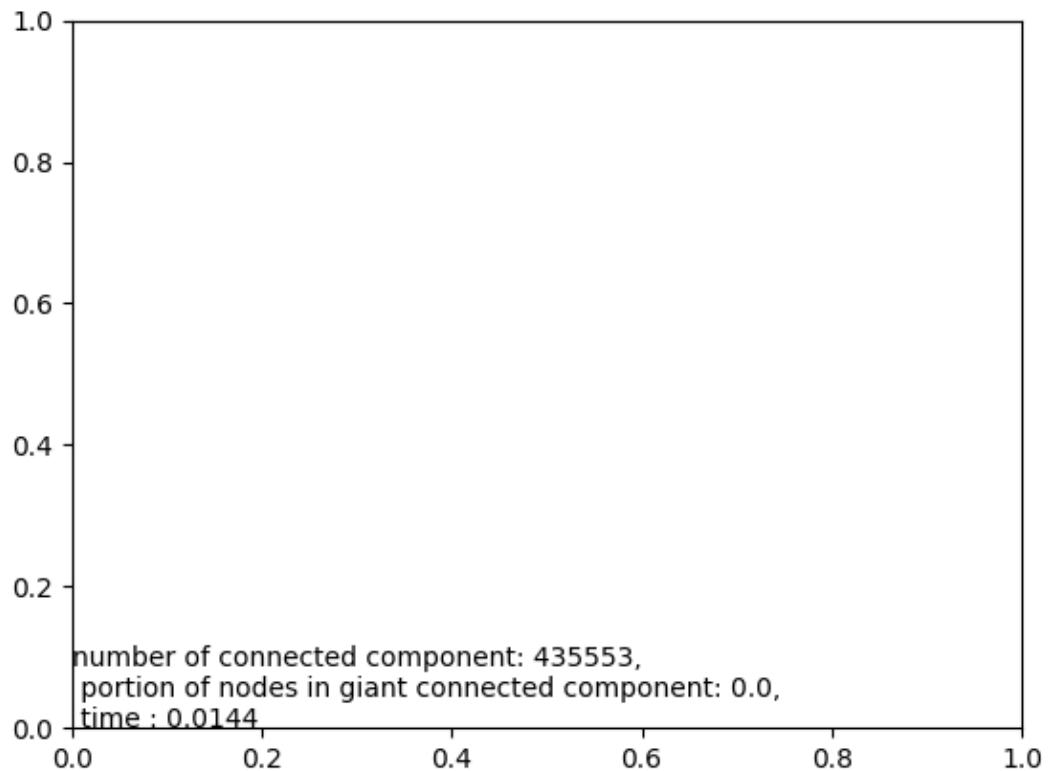


## modified BA network

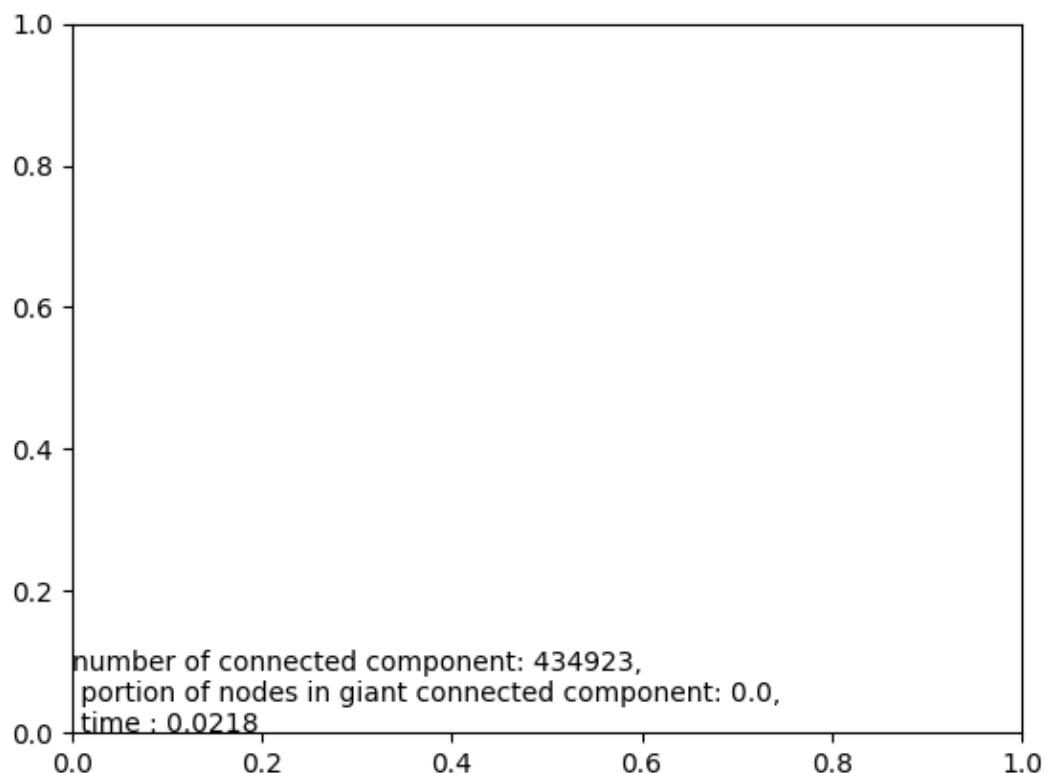


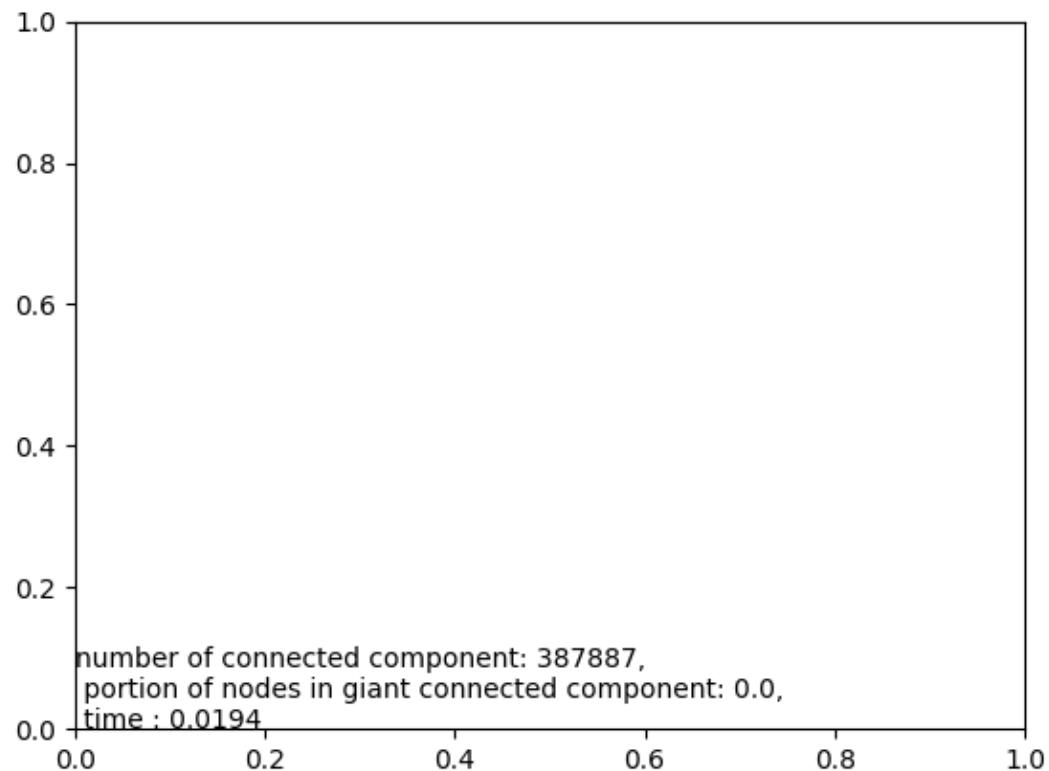
## citation

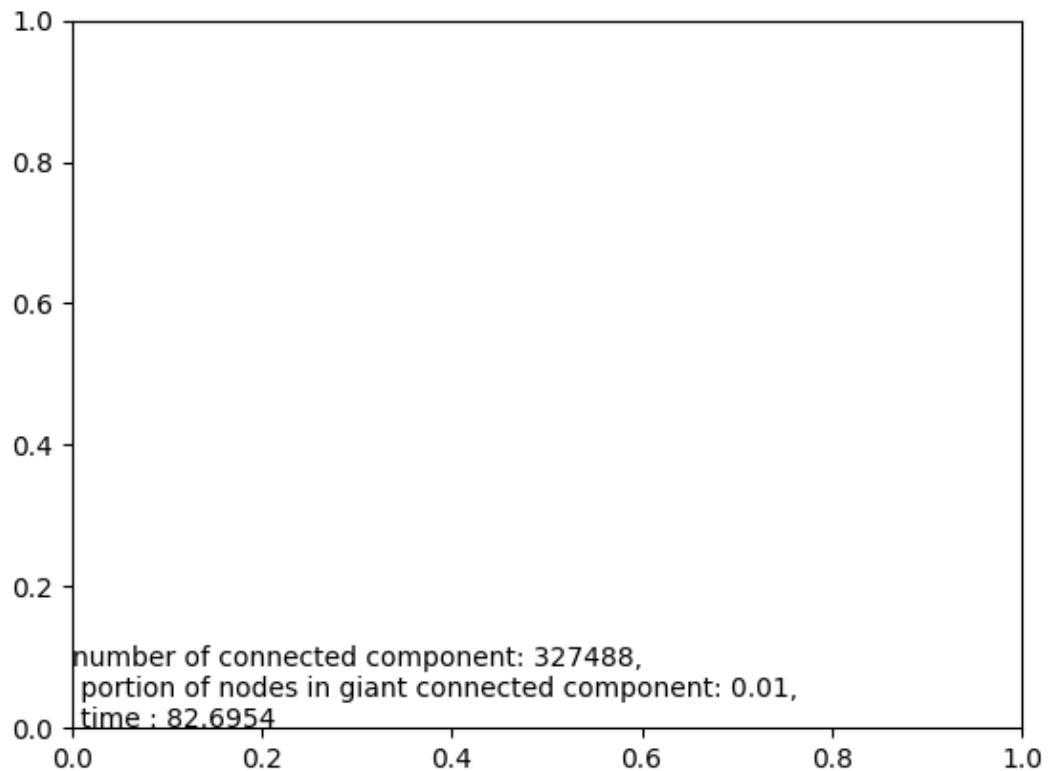
### original network



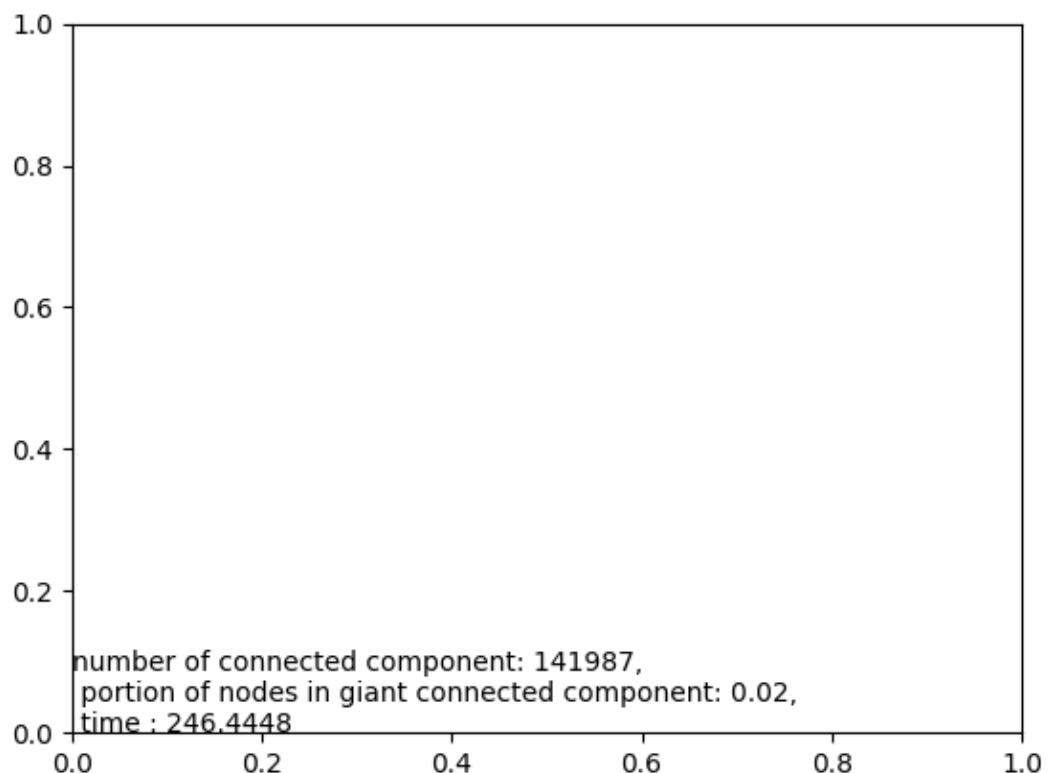
## BA network



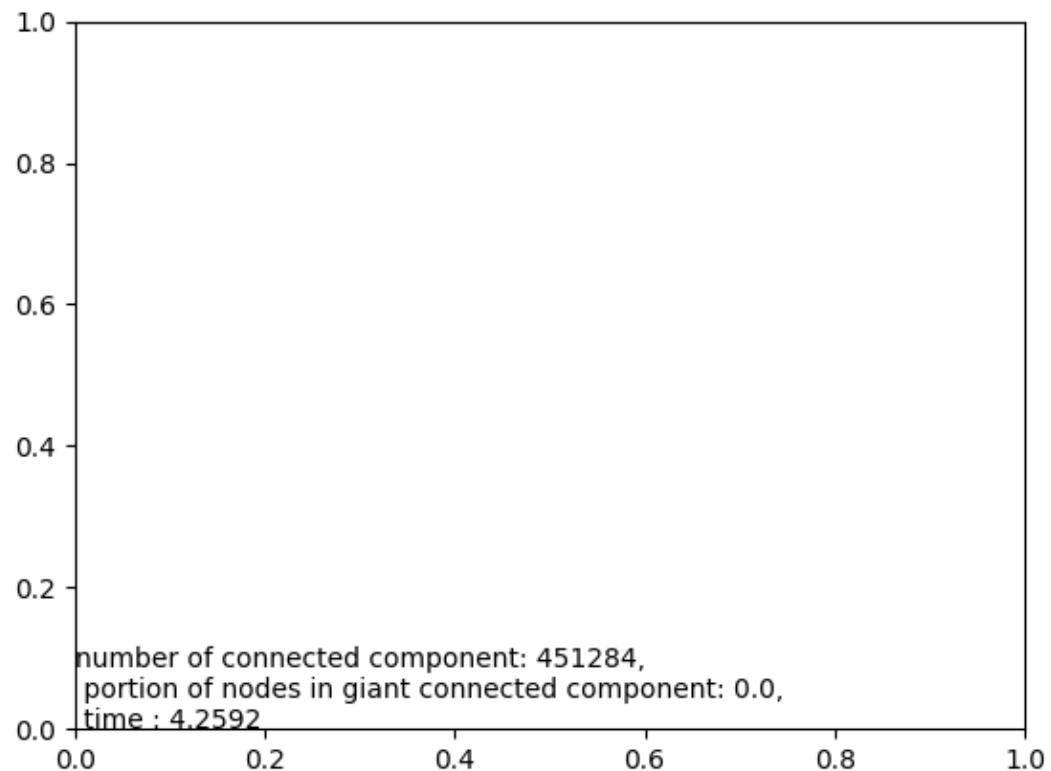
**modified BA network****actor****original network**



## BA network

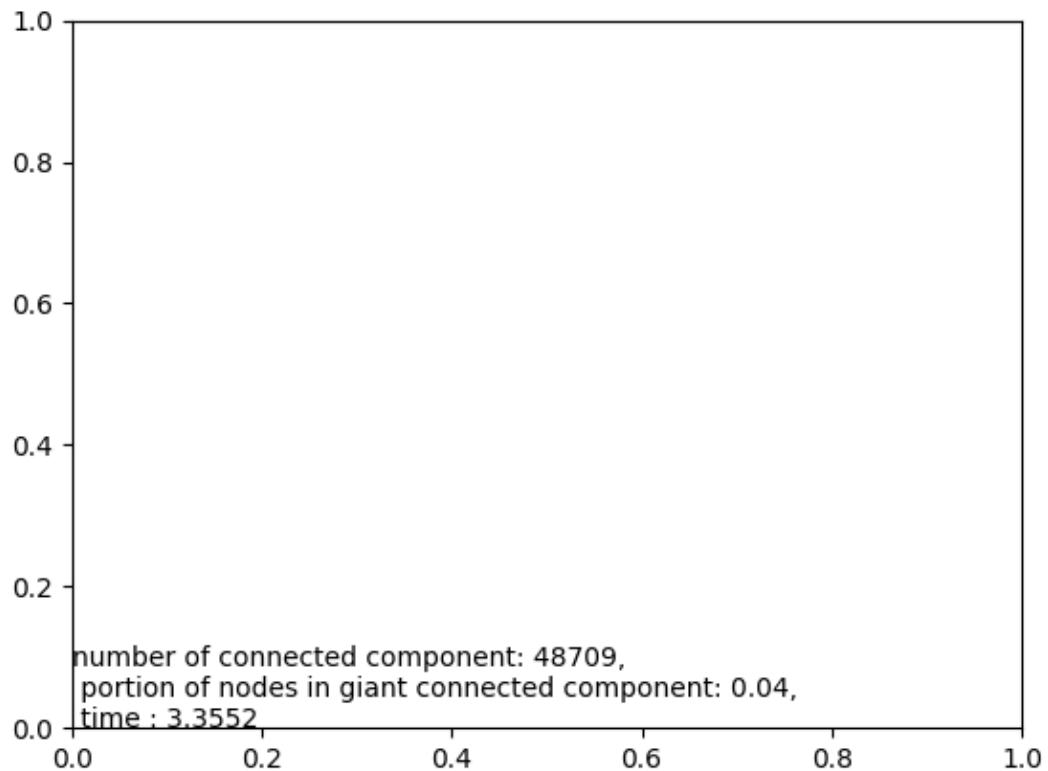


## modified BA network

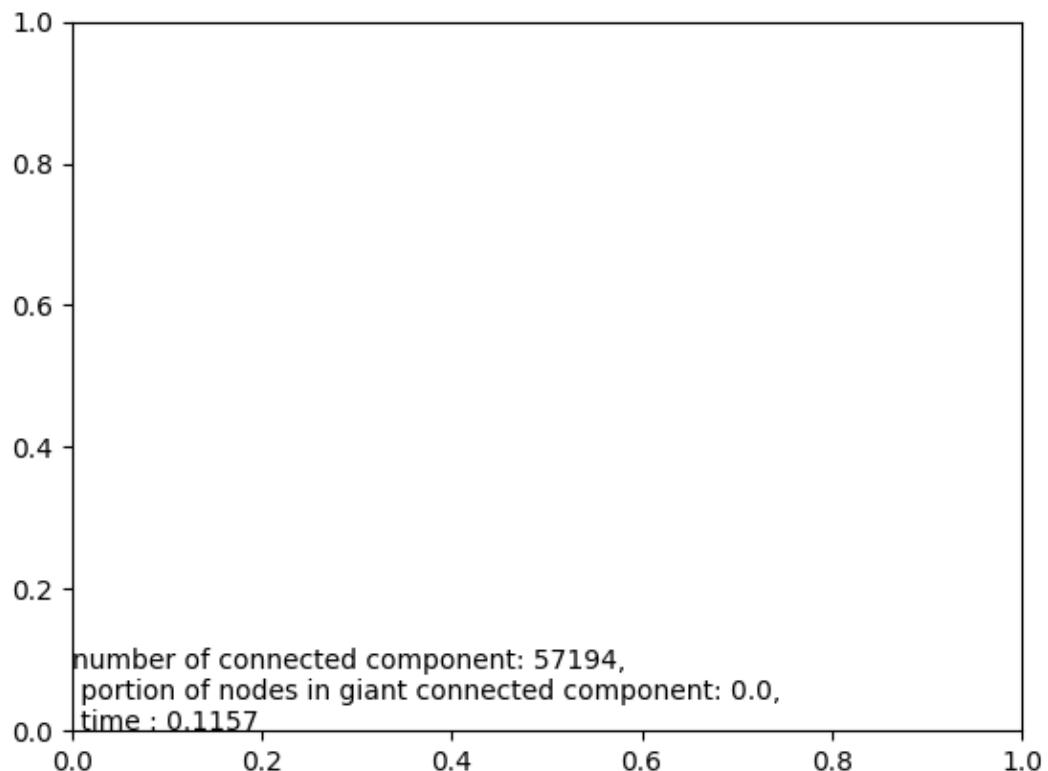


## email

### original network



## BA network

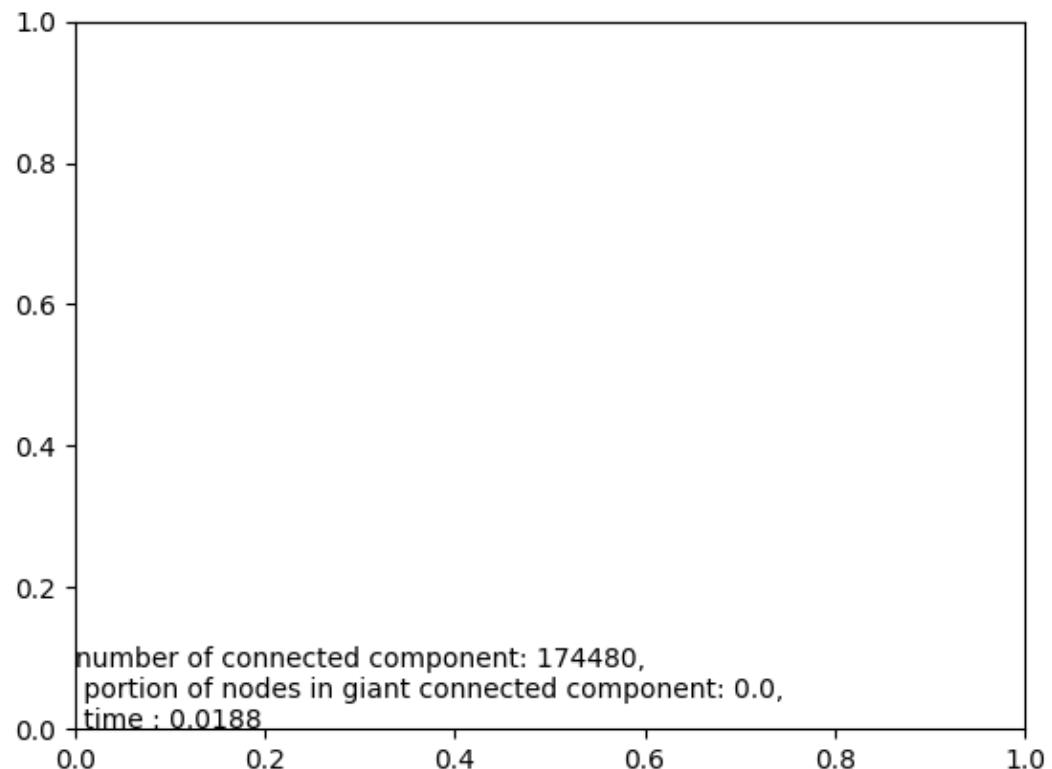


## modified BA network

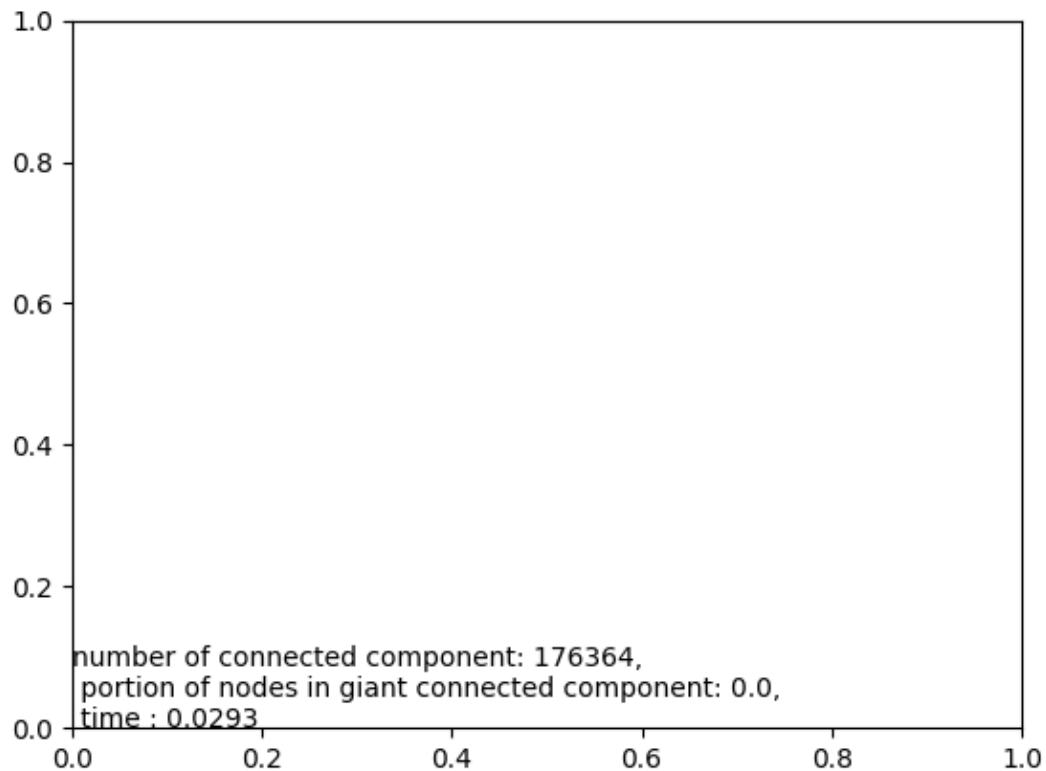


**internet**

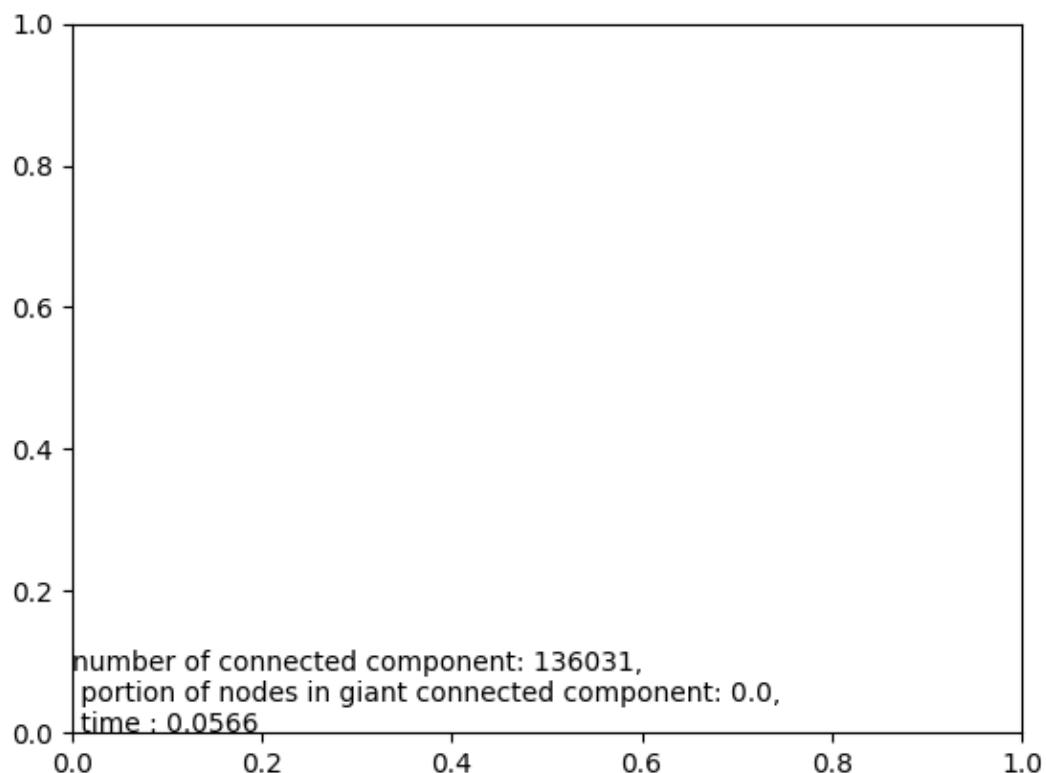
**original network**



**BA network**

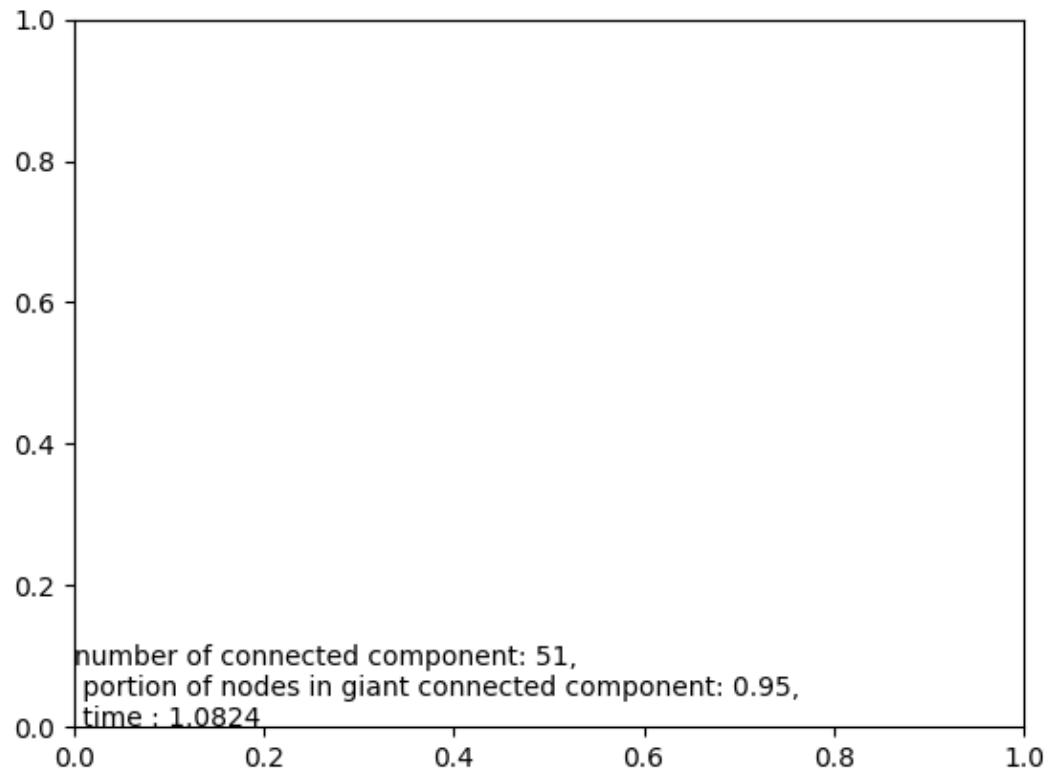


## modified BA network

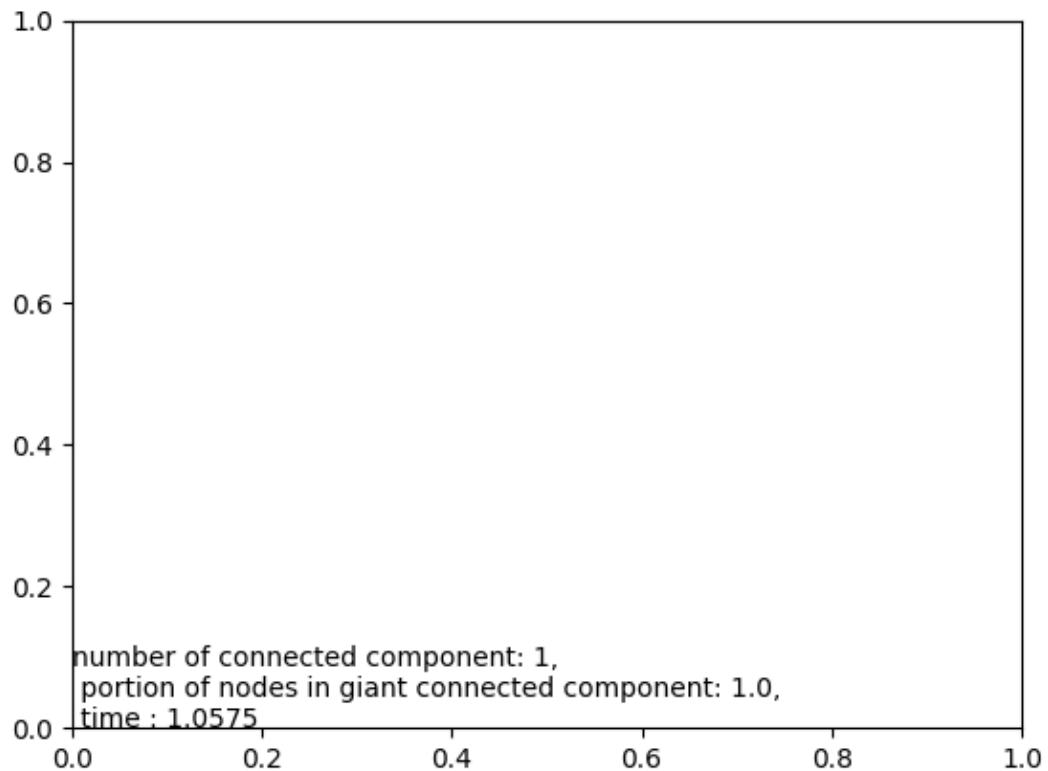


## metabolic

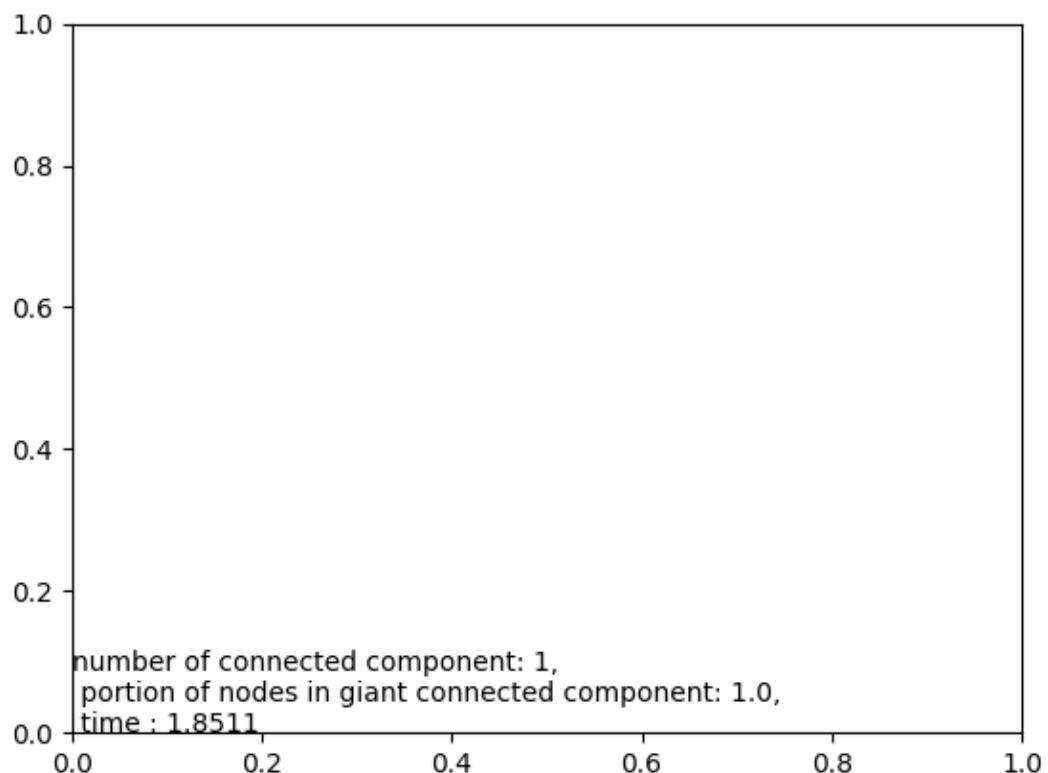
### original network



### BA network

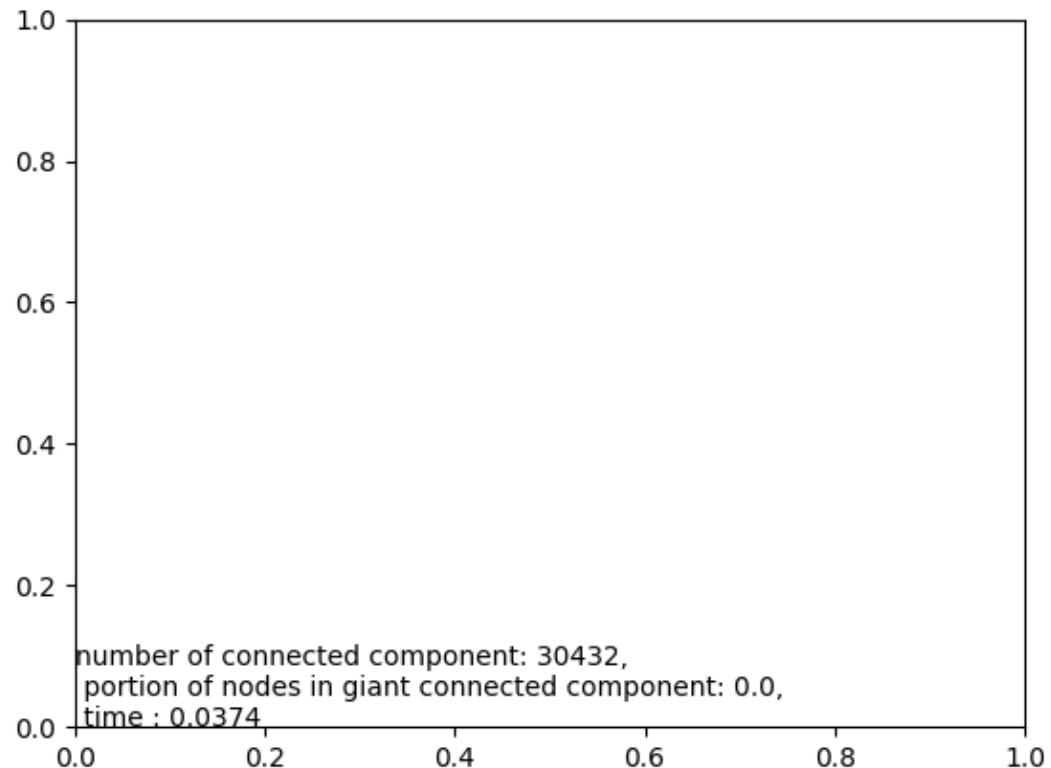


## modified BA network

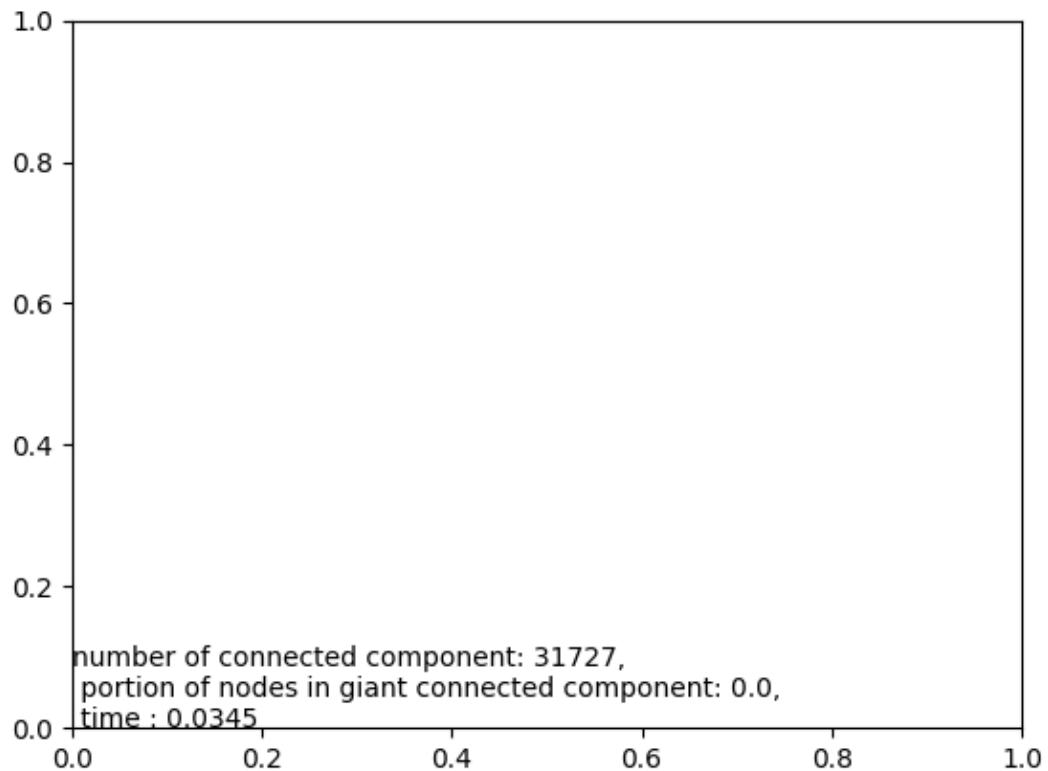


# phonecalls

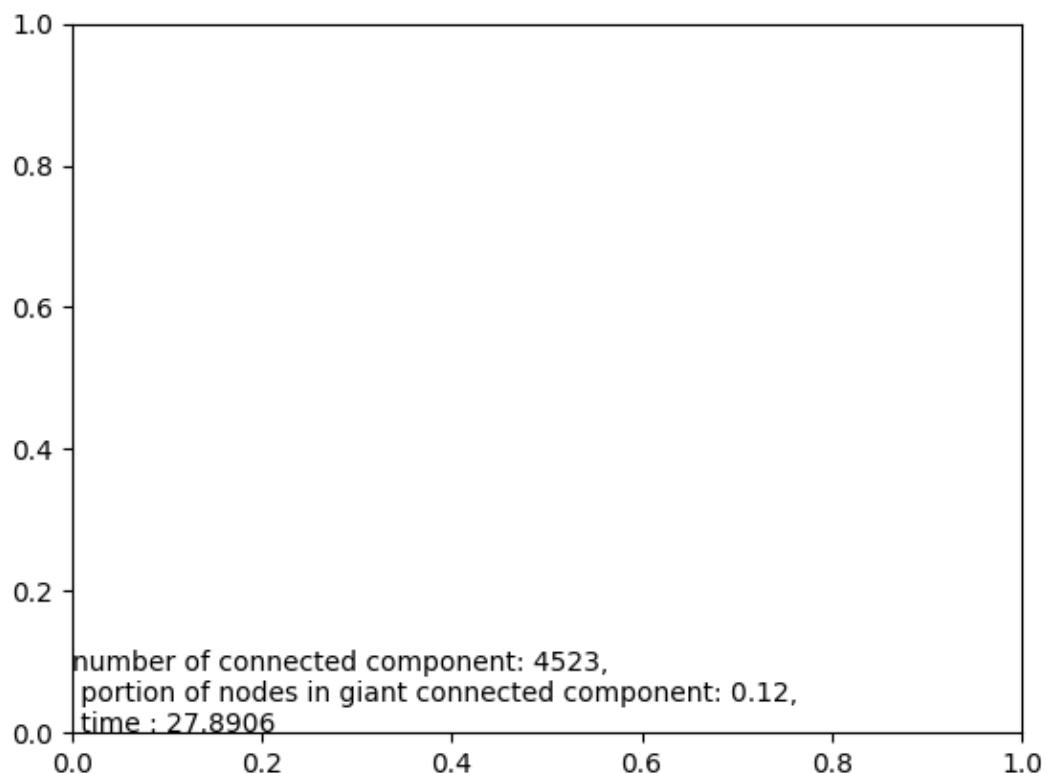
## original network



## BA network

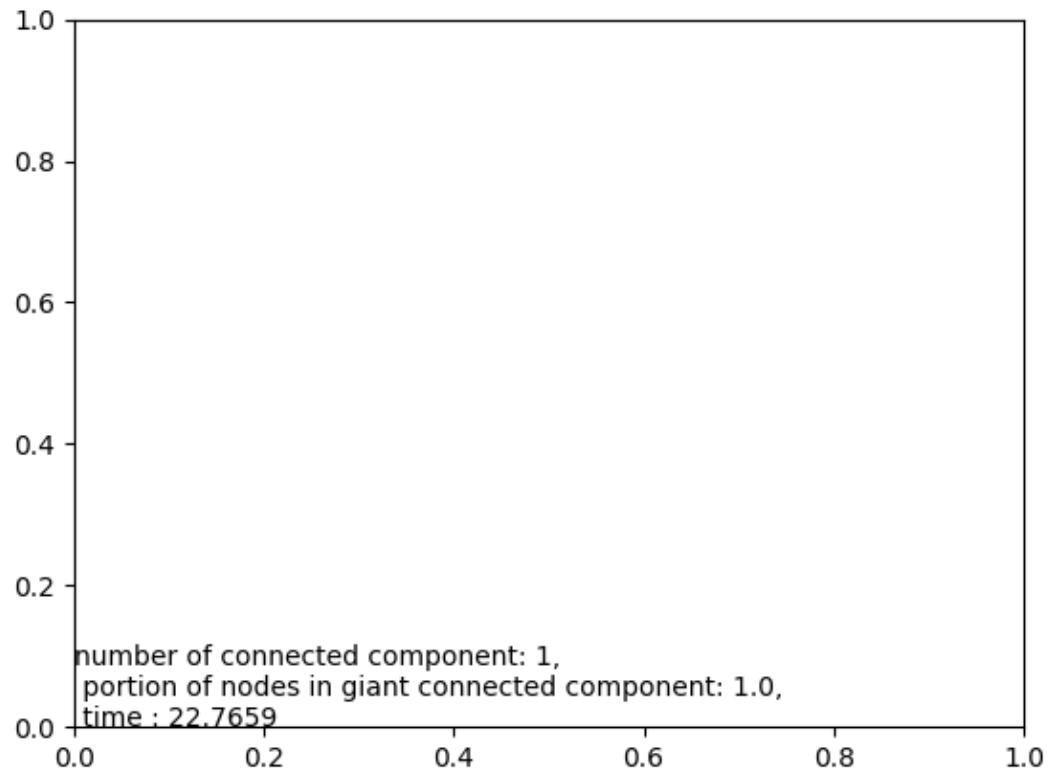


## modified BA network

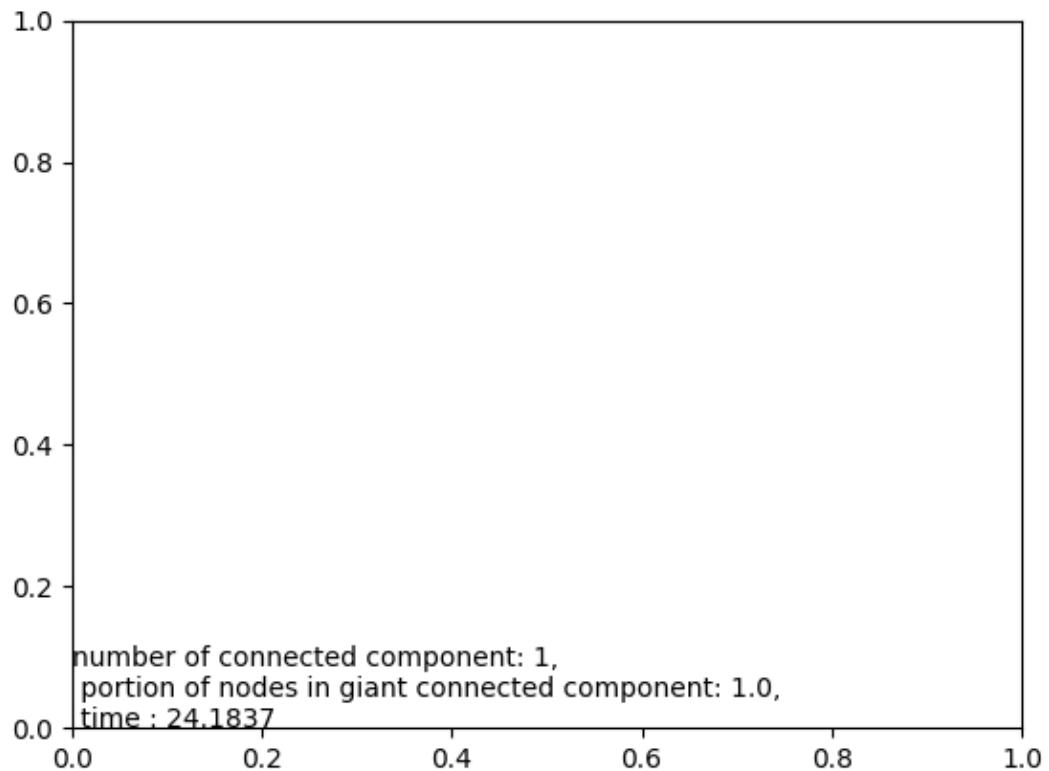


# powergrid

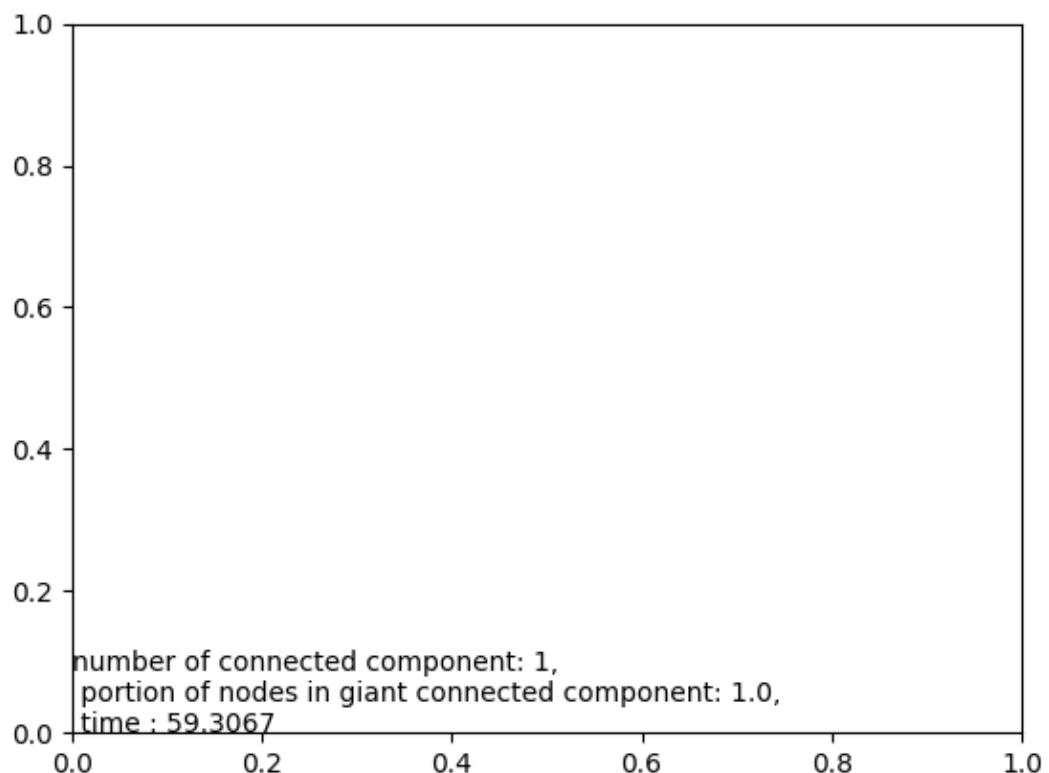
## original network



## BA network

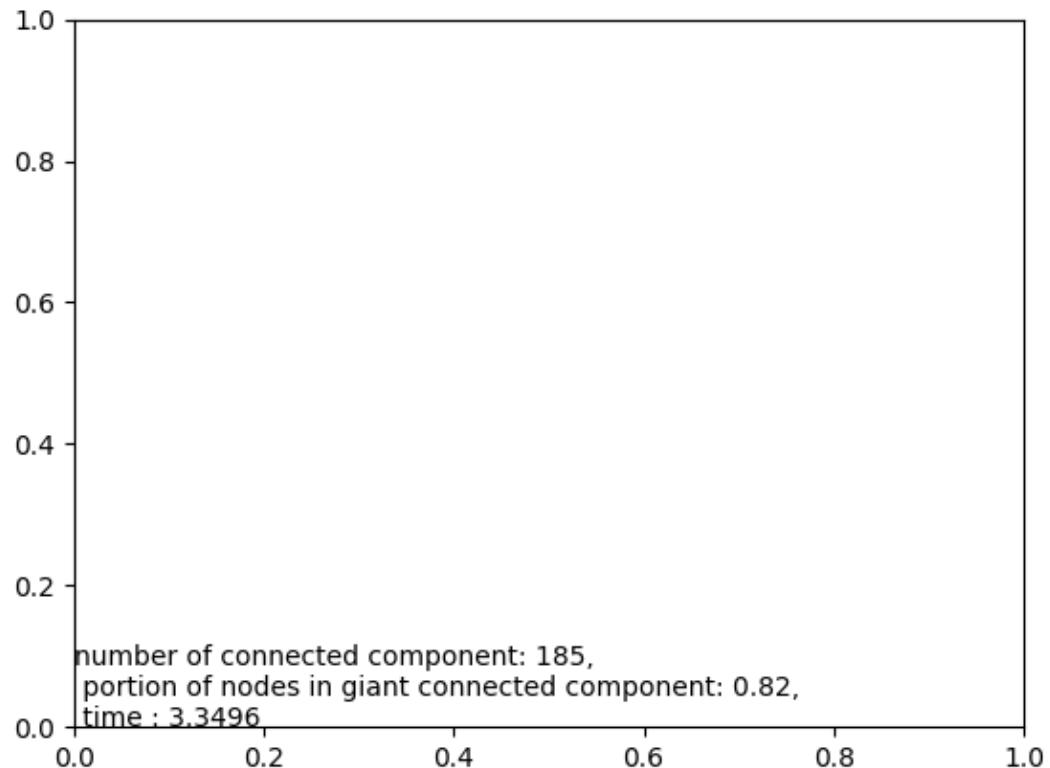


## modified BA network

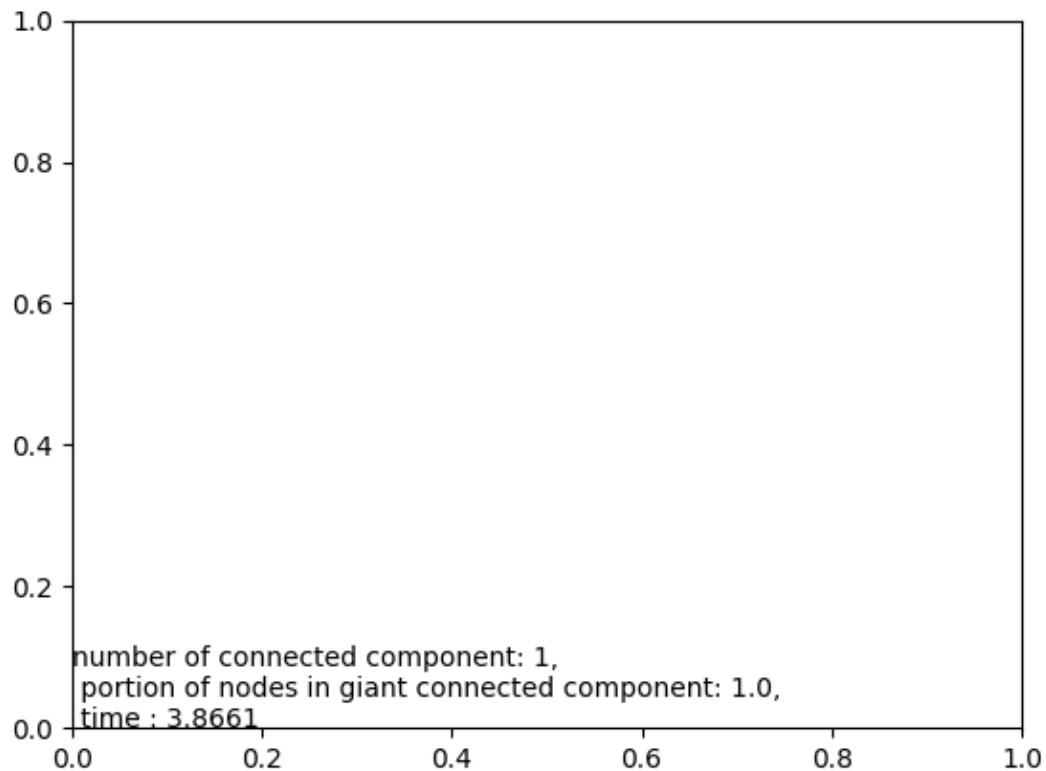


# protein

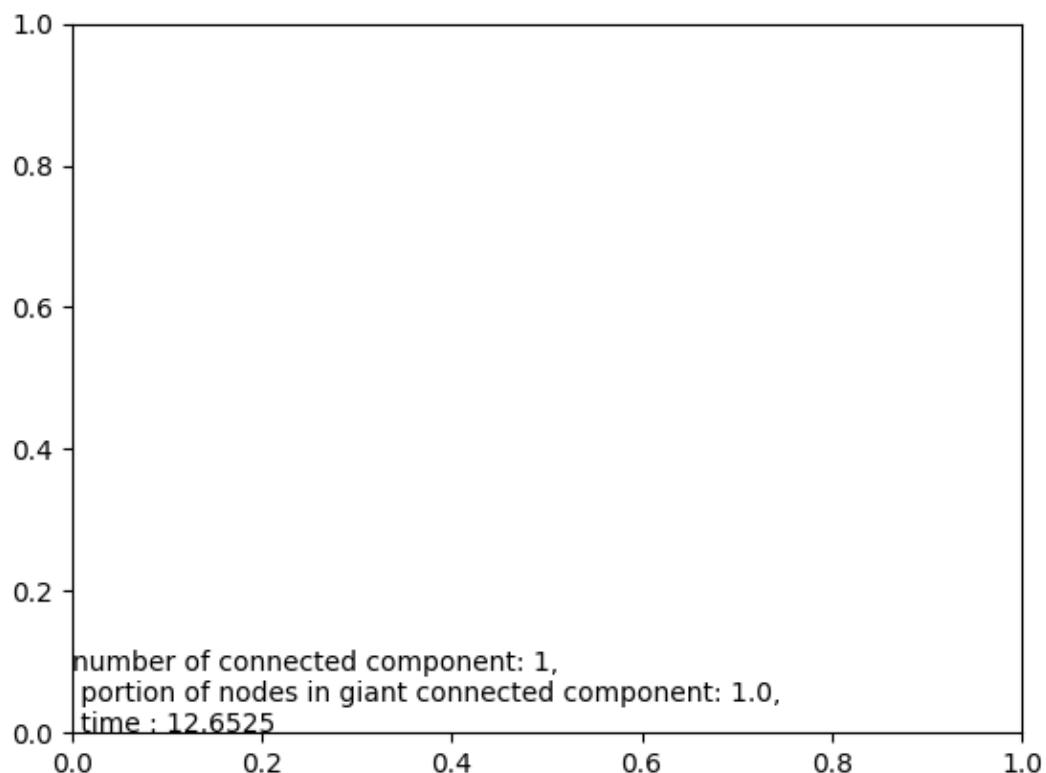
## original network

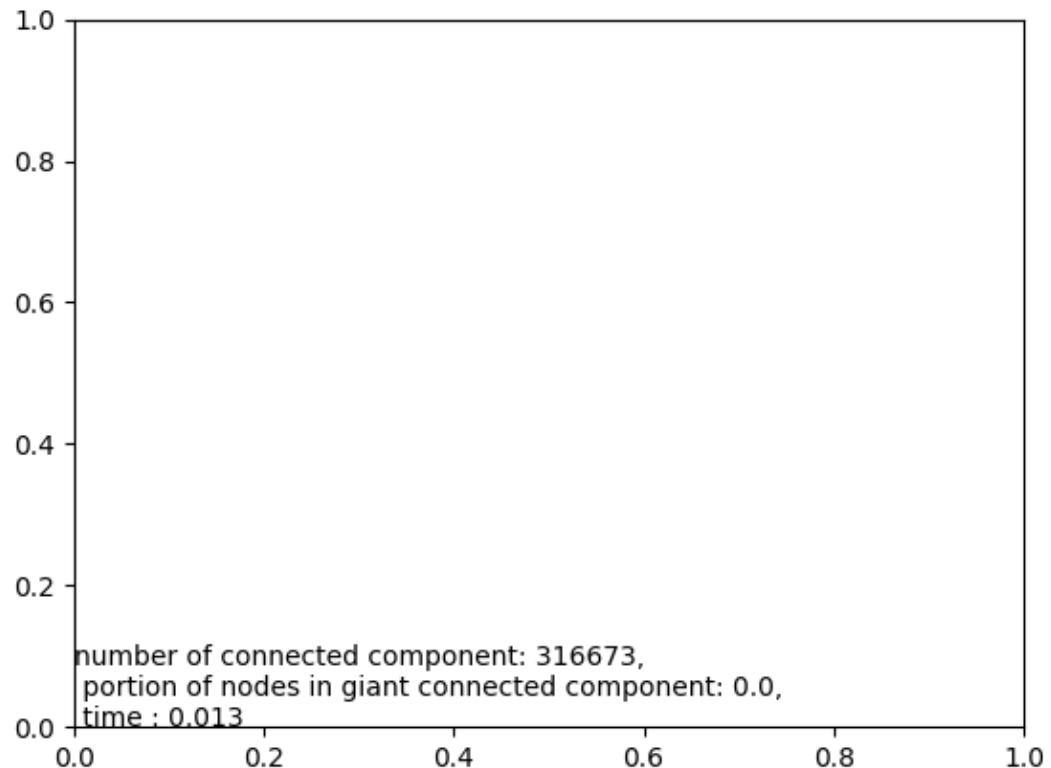


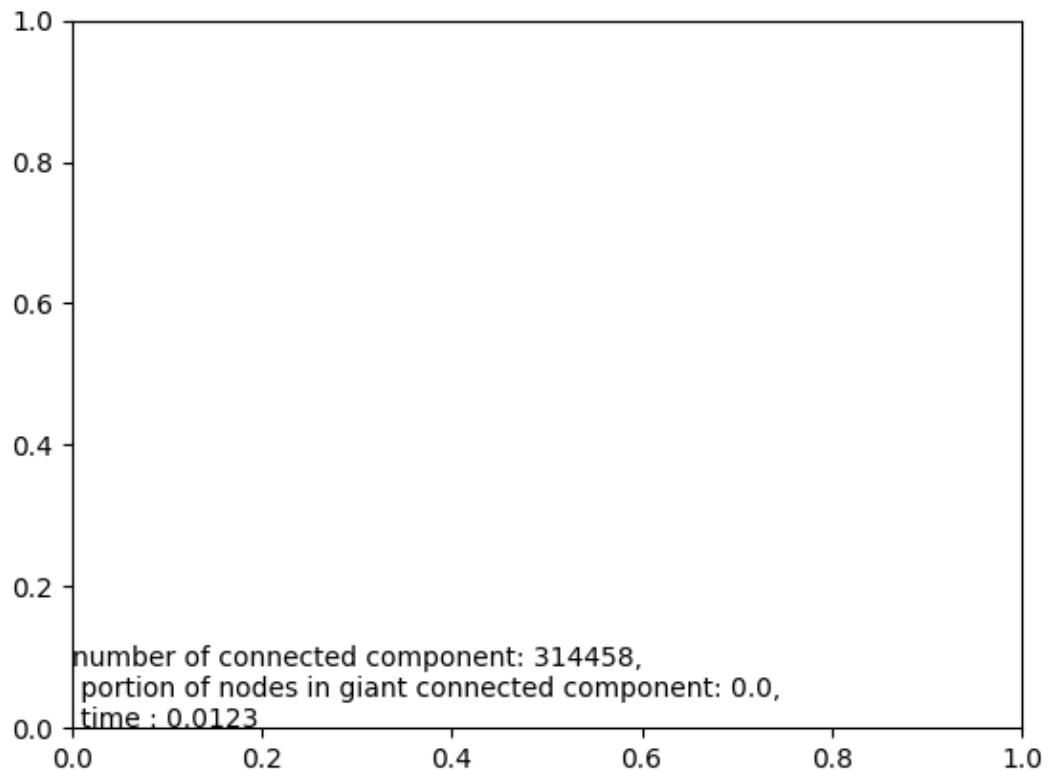
## BA network



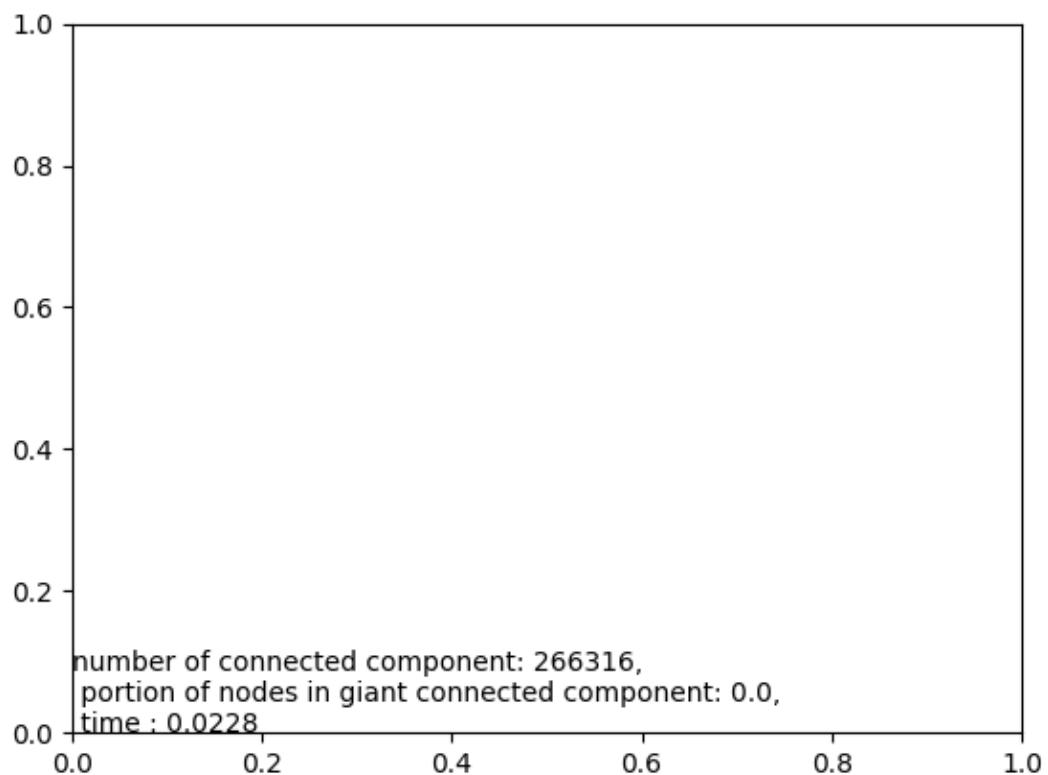
## modified BA network



**WWW****original network****BA network**

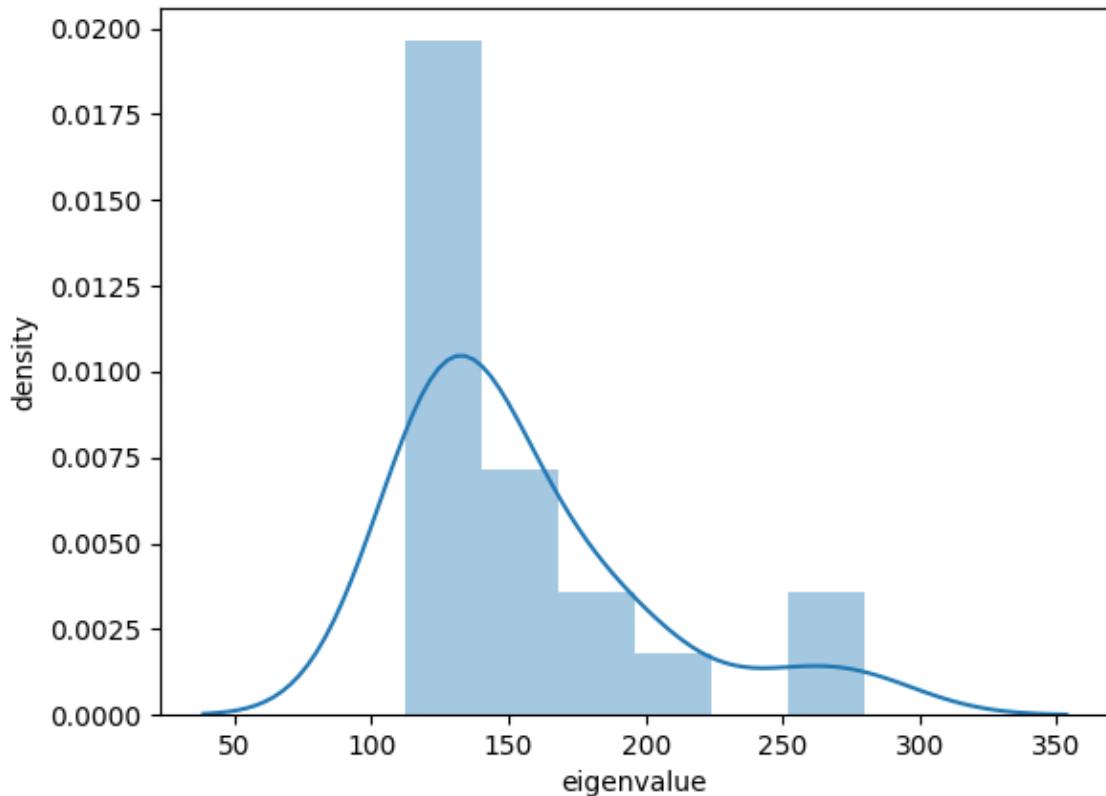


## modified BA network

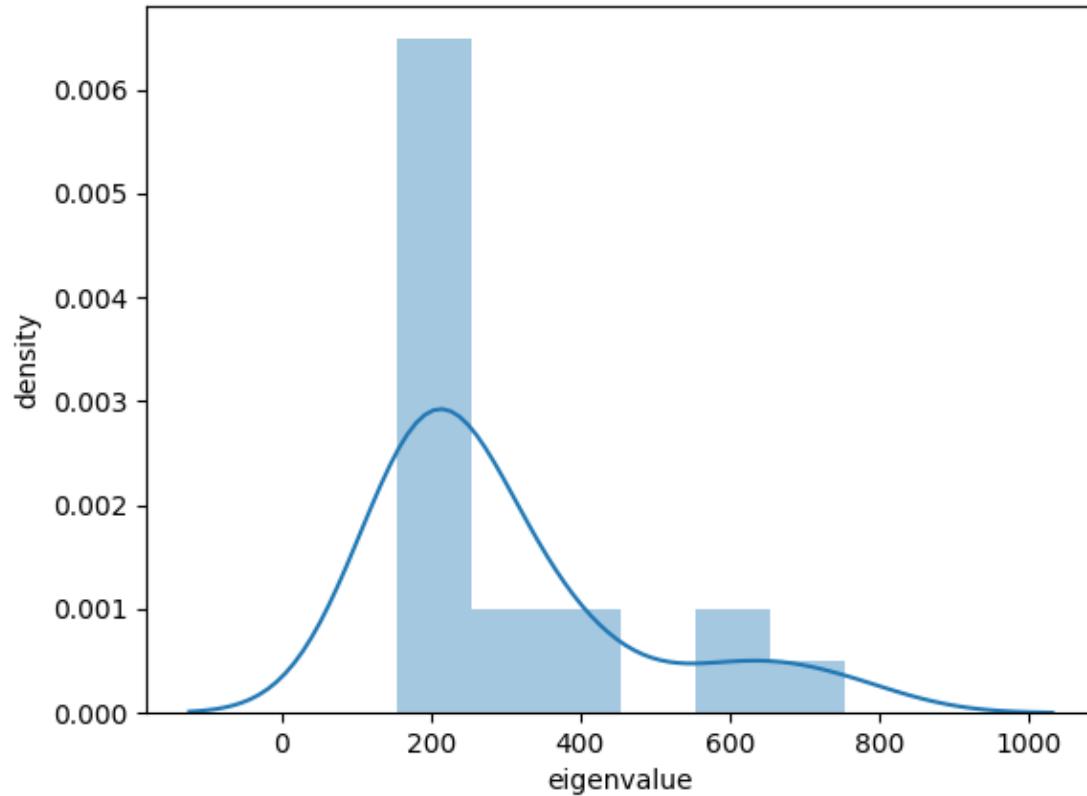


**eigen****collaboration****original network**

top 20 laplacian eigenvalue distribution, eigengap: 27.02,  
time: 0.1326

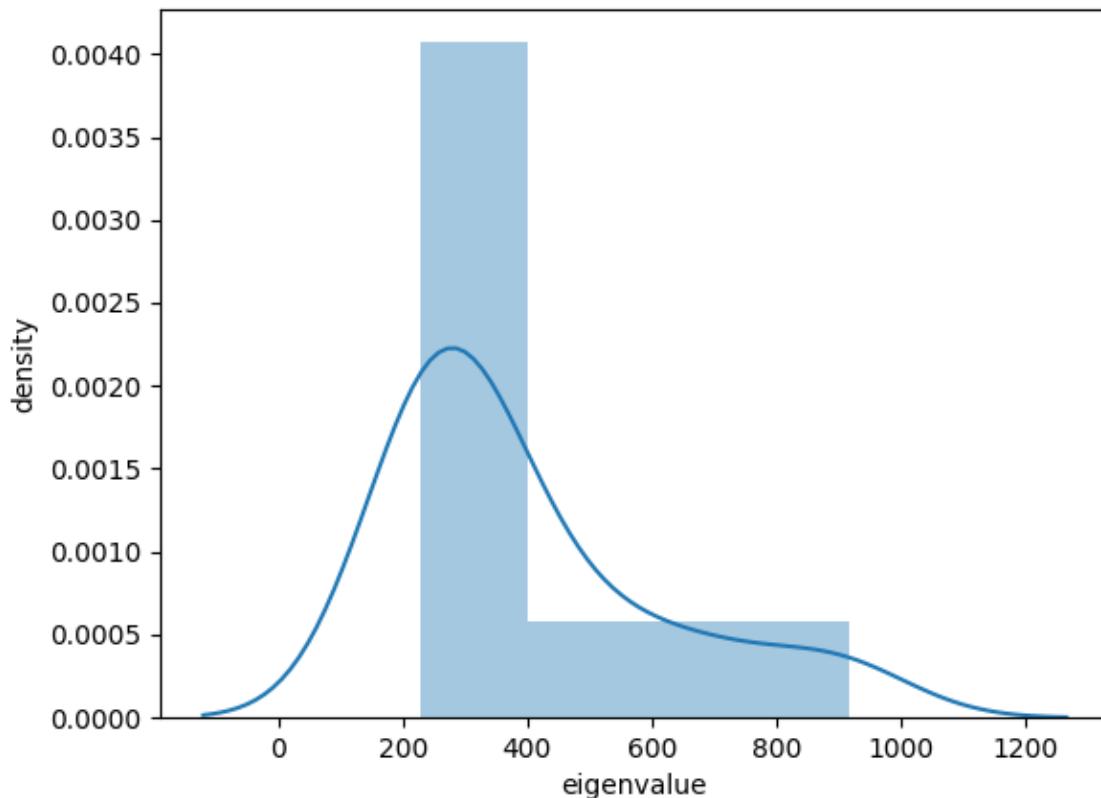
**BA network**

top 20 laplacian eigenvalue distribution, eigengap: 109.0,  
time: 0.1138



### modified BA network

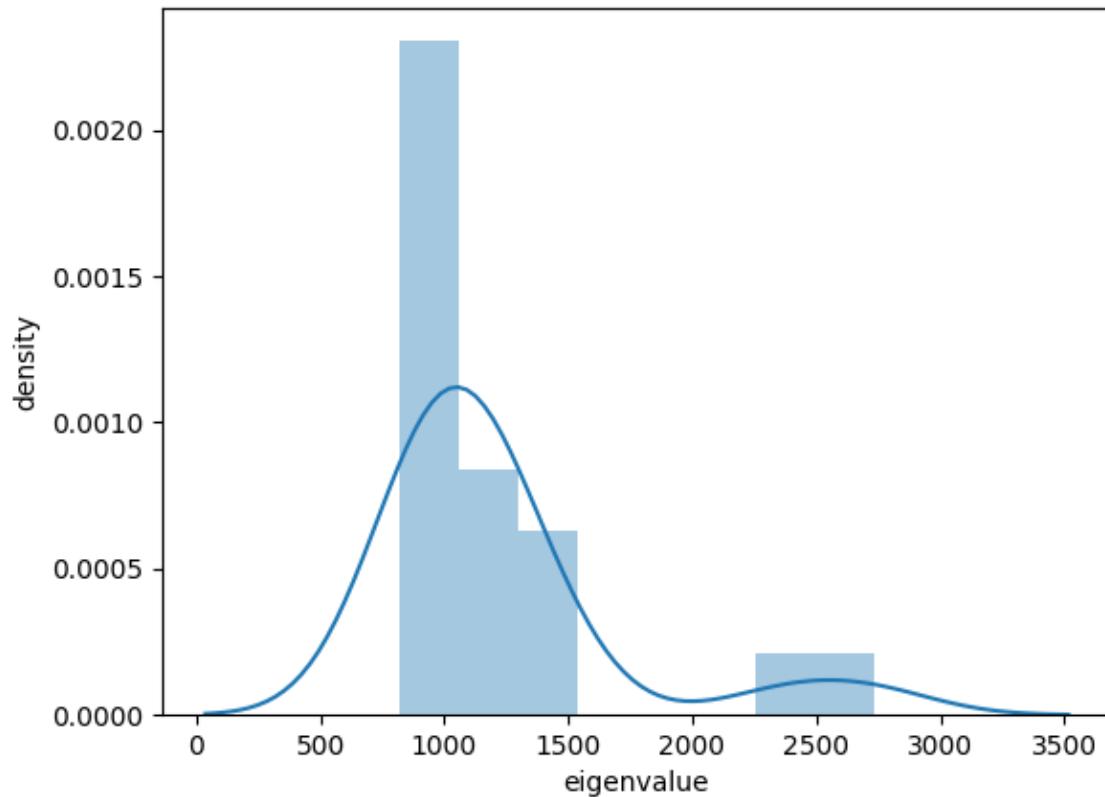
top 20 laplacian eigenvalue distribution, eigengap: 44.04,  
time: 0.5127



## citation

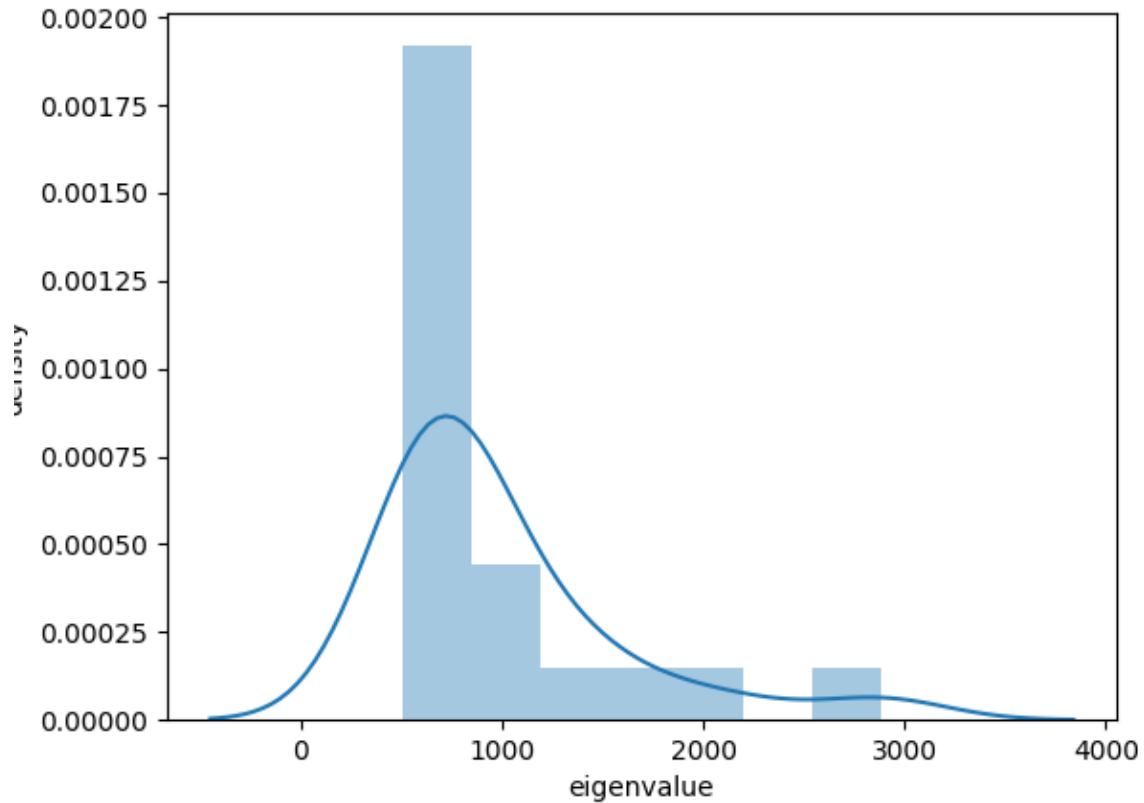
### original network

top 20 laplacian eigenvalue distribution, eigengap: 357.0,  
time: 6.6235



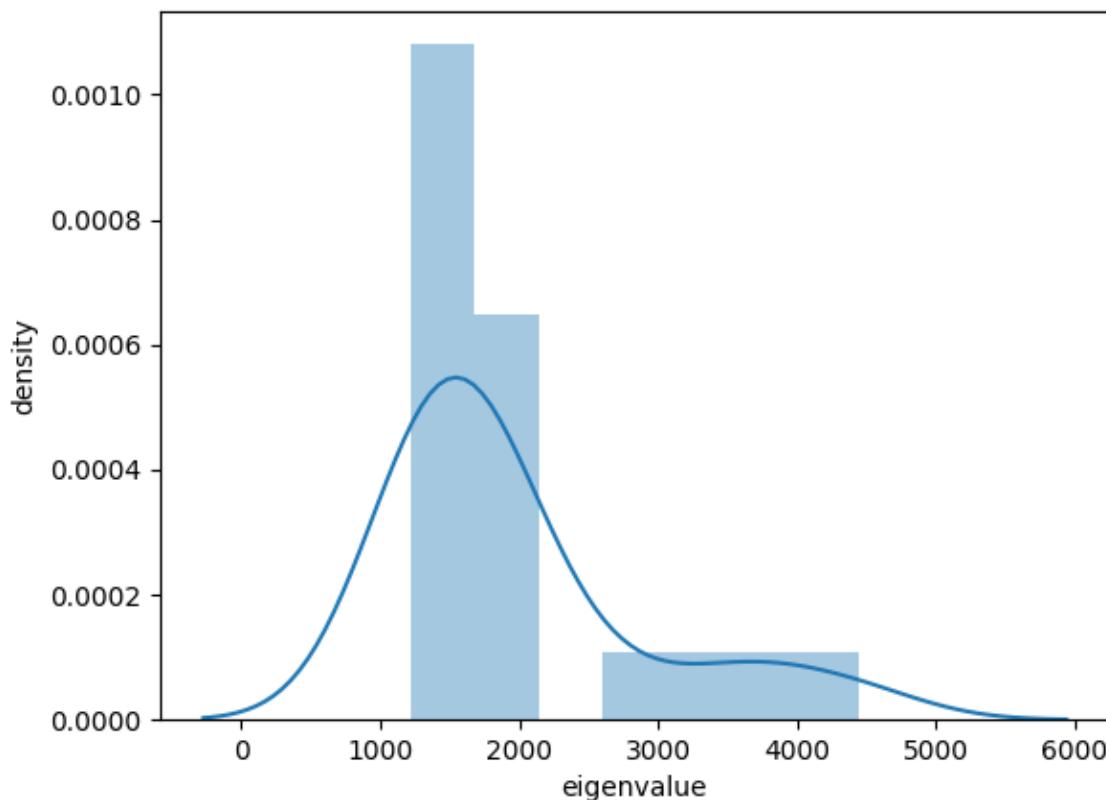
### BA network

top 20 laplacian eigenvalue distribution, eigengap: 809.0,  
time: 6.9878



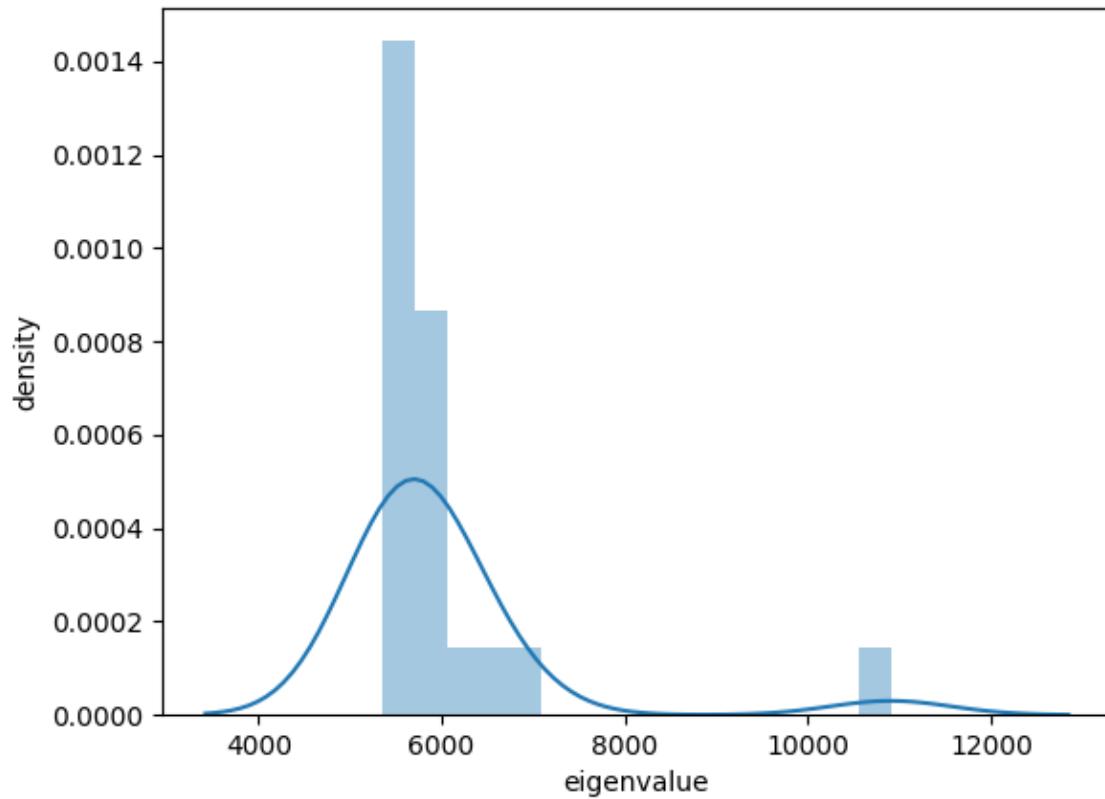
### modified BA network

top 20 laplacian eigenvalue distribution, eigengap: 636.0,  
time: 11.895

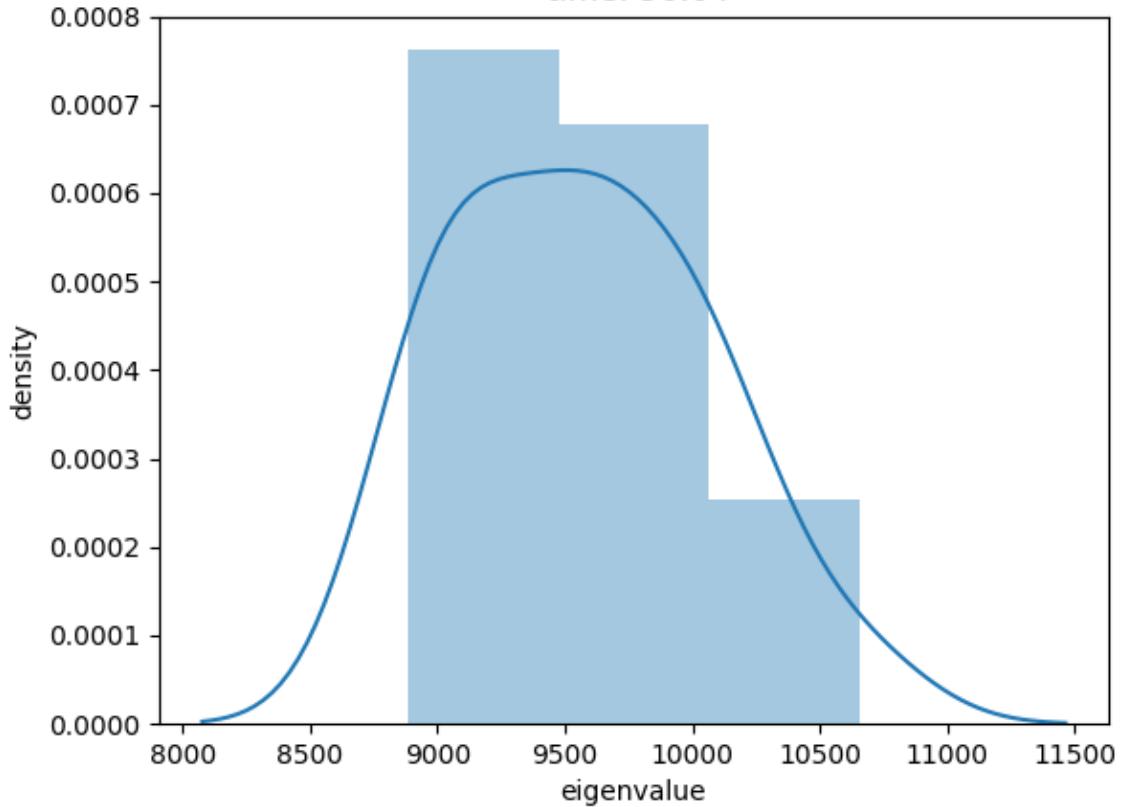


**actor****original network**

top 20 laplacian eigenvalue distribution, eigengap: 4136.95,  
time: 64.0479

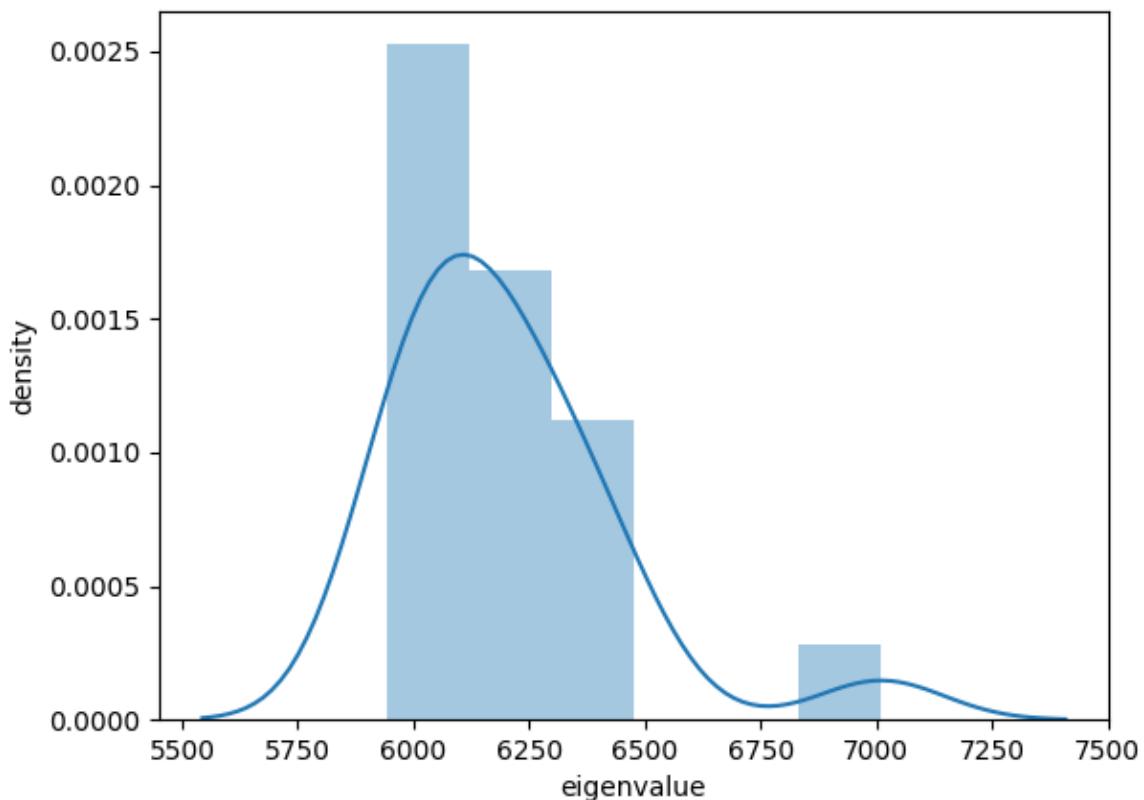
**BA network**

top 20 laplacian eigenvalue distribution, eigengap: 407.99,  
time: 56.64



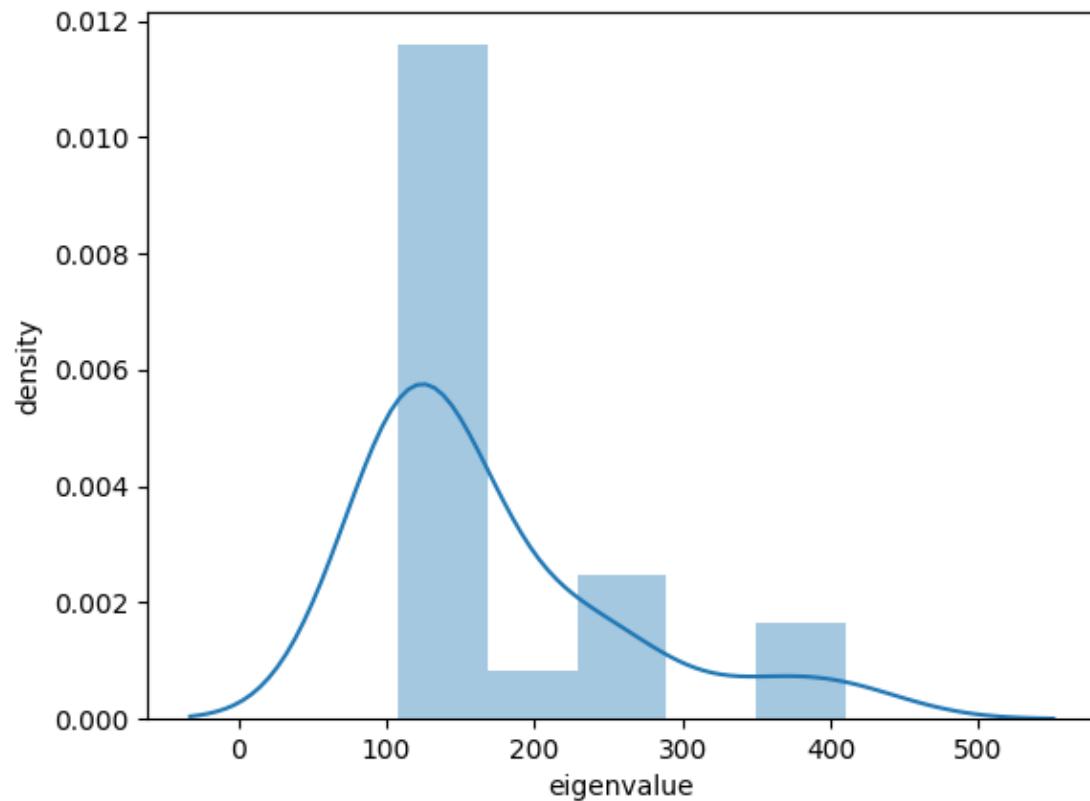
## modified BA network

top 20 laplacian eigenvalue distribution, eigengap: 558.9,  
time: 18.797

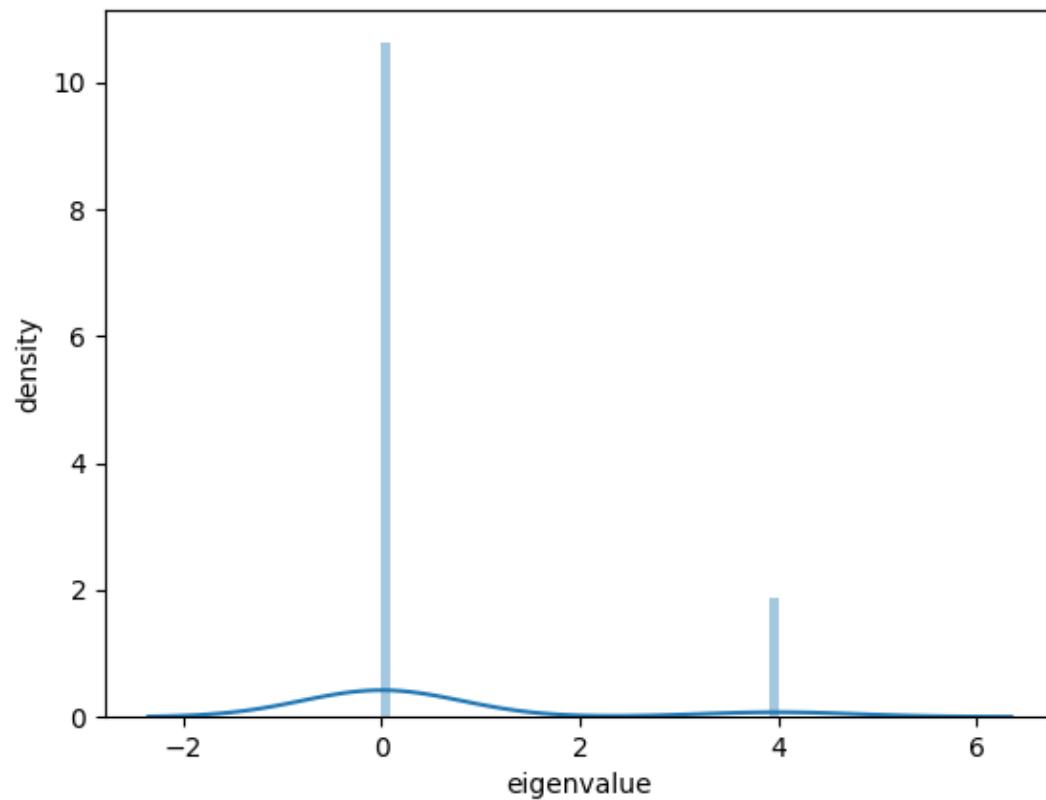


**email****original network**

top 20 laplacian eigenvalue distribution, eigengap: 55.05,  
time: 0.2037

**BA network**

top 20 laplacian eigenvalue distribution, eigengap: 0.0,  
time: 0.0562



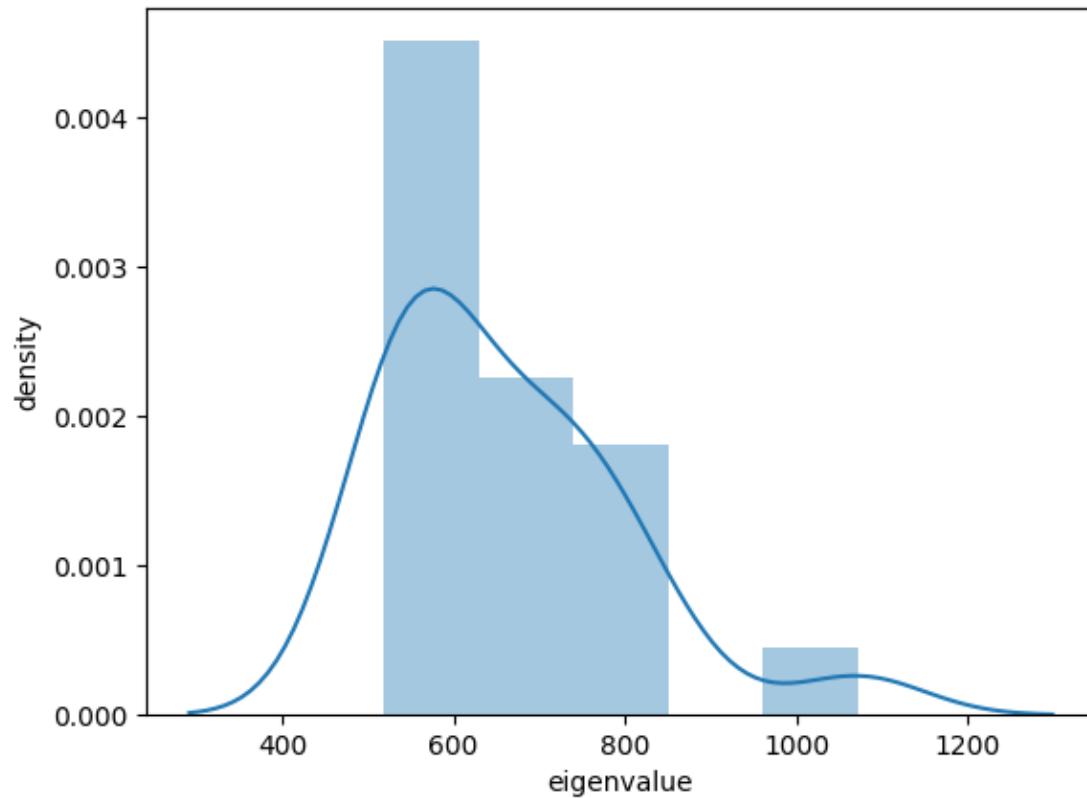
### modified BA network



### internet

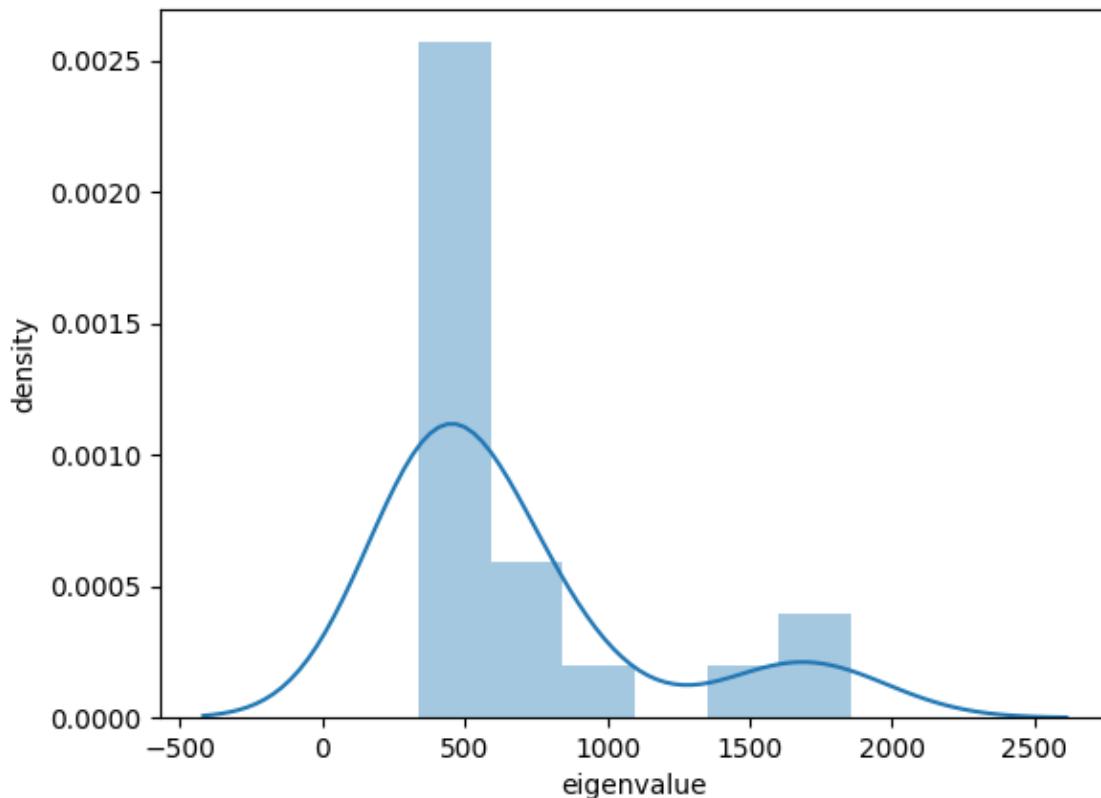
### original network

top 20 laplacian eigenvalue distribution, eigengap: 229.92,  
time: 1.9075



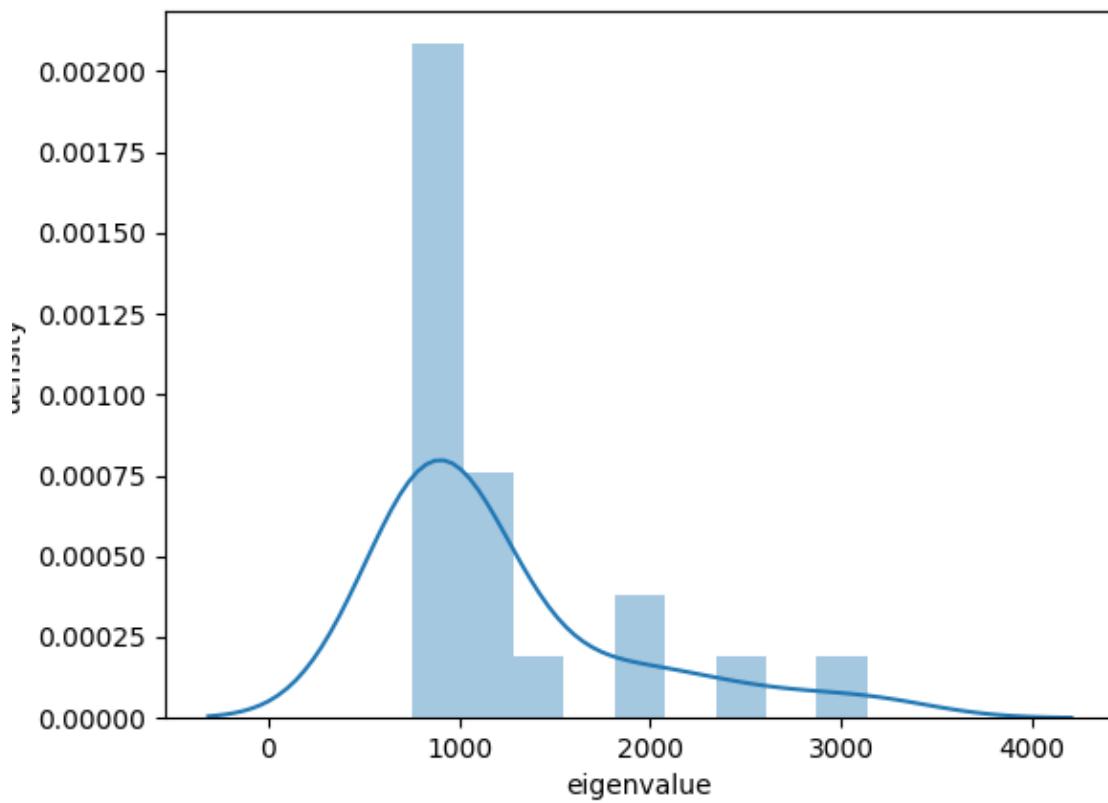
## BA network

top 20 laplacian eigenvalue distribution, eigengap: 210.0,  
time: 2.9974



## modified BA network

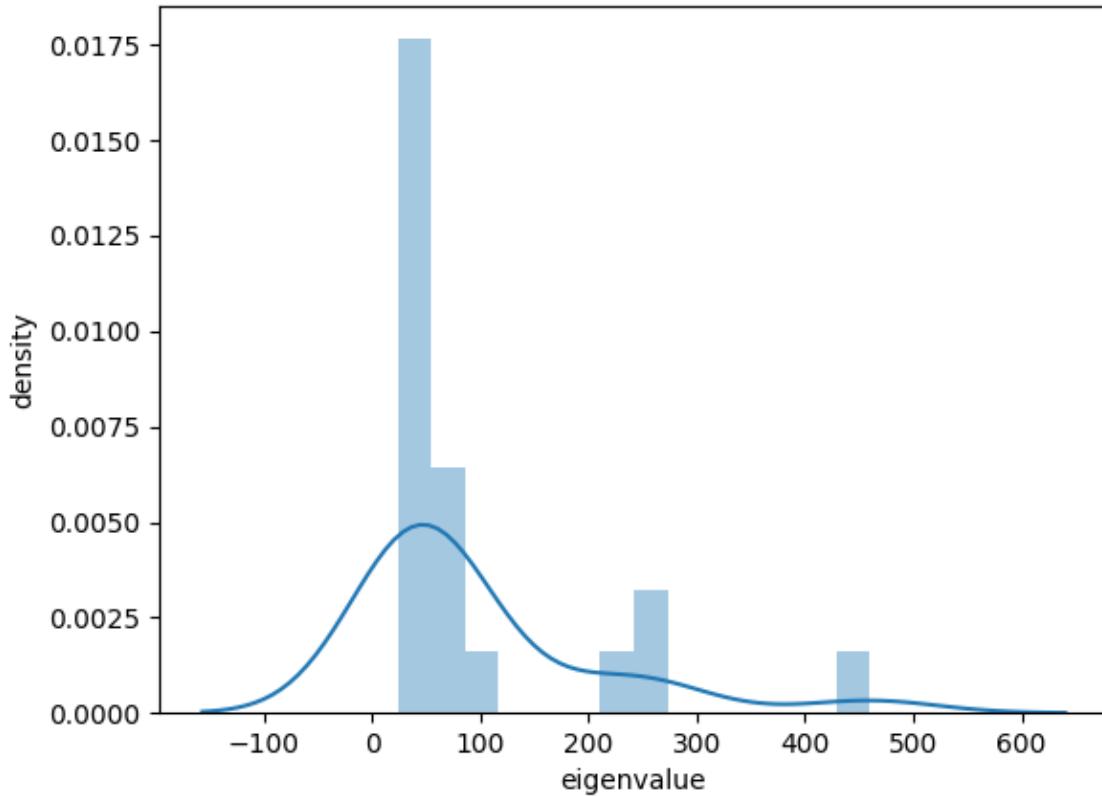
top 20 laplacian eigenvalue distribution, eigengap: 606.0,  
time: 5.2321



## metabolic

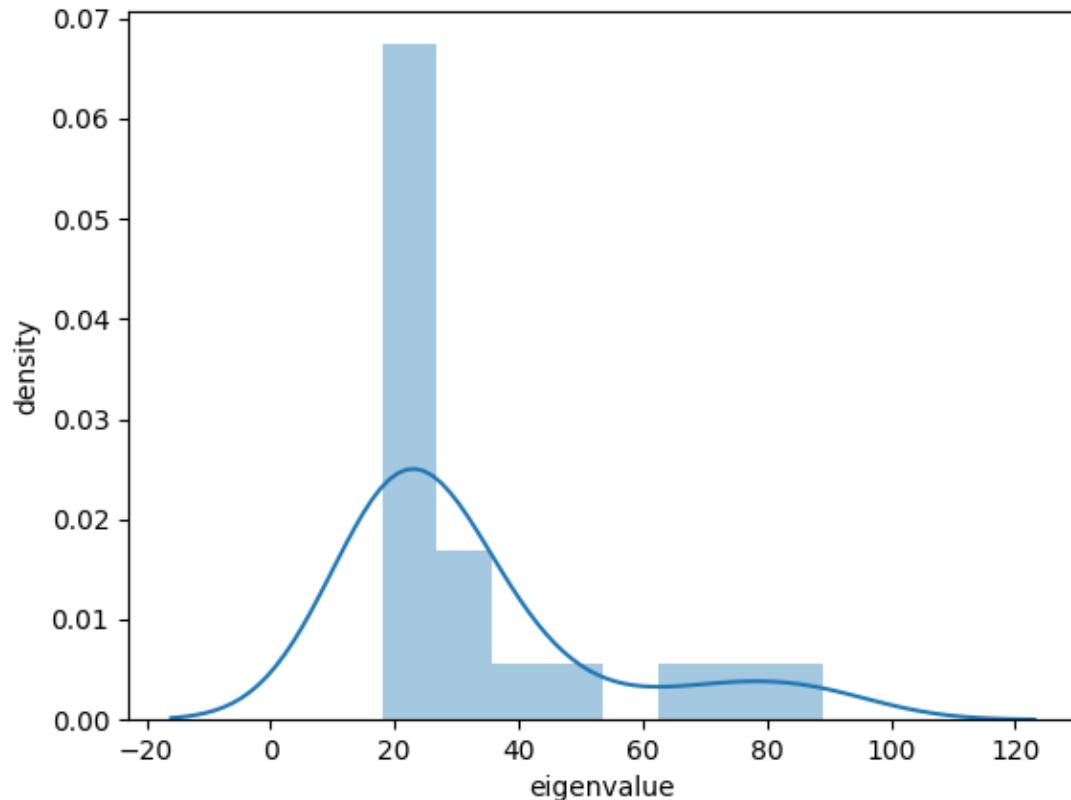
### original network

top 20 laplacian eigenvalue distribution, eigengap: 201.94,  
time: 0.0354



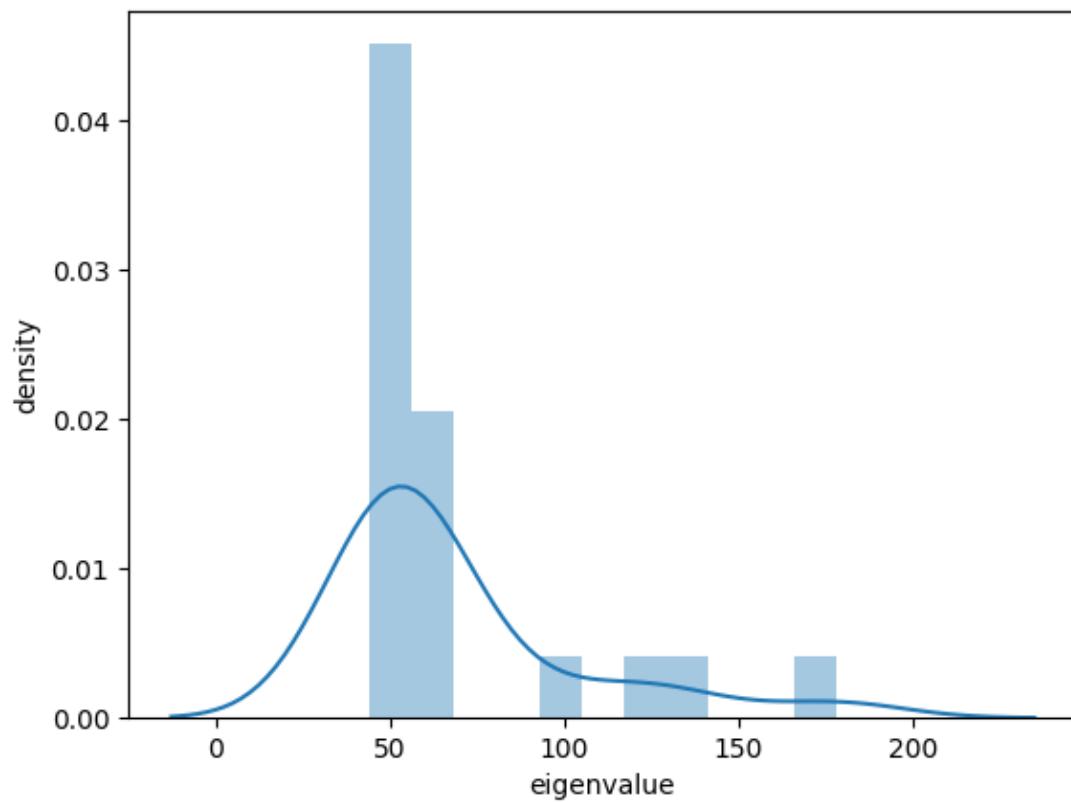
## BA network

top 20 laplacian eigenvalue distribution, eigengap: 10.13,  
time: 0.1833



## modified BA network

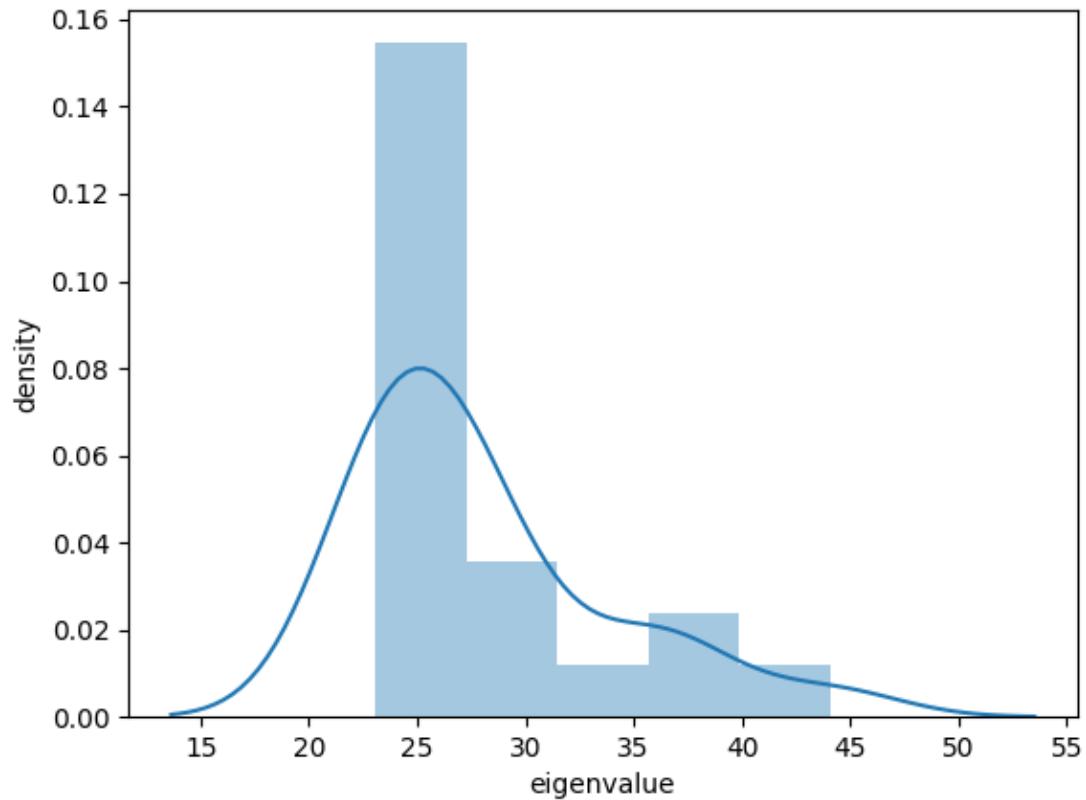
top 20 laplacian eigenvalue distribution, eigengap: 46.92,  
time: 0.0599



## phonecalls

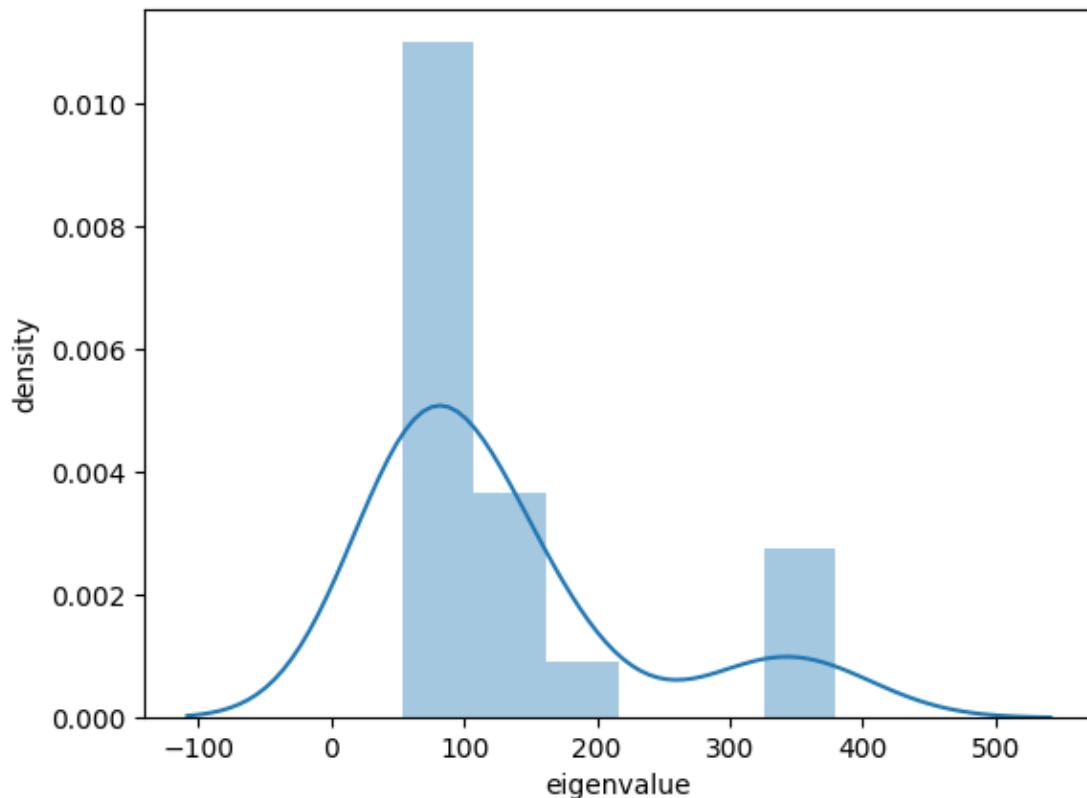
### original network

top 20 laplacian eigenvalue distribution, eigengap: 6.99,  
time: 0.2658



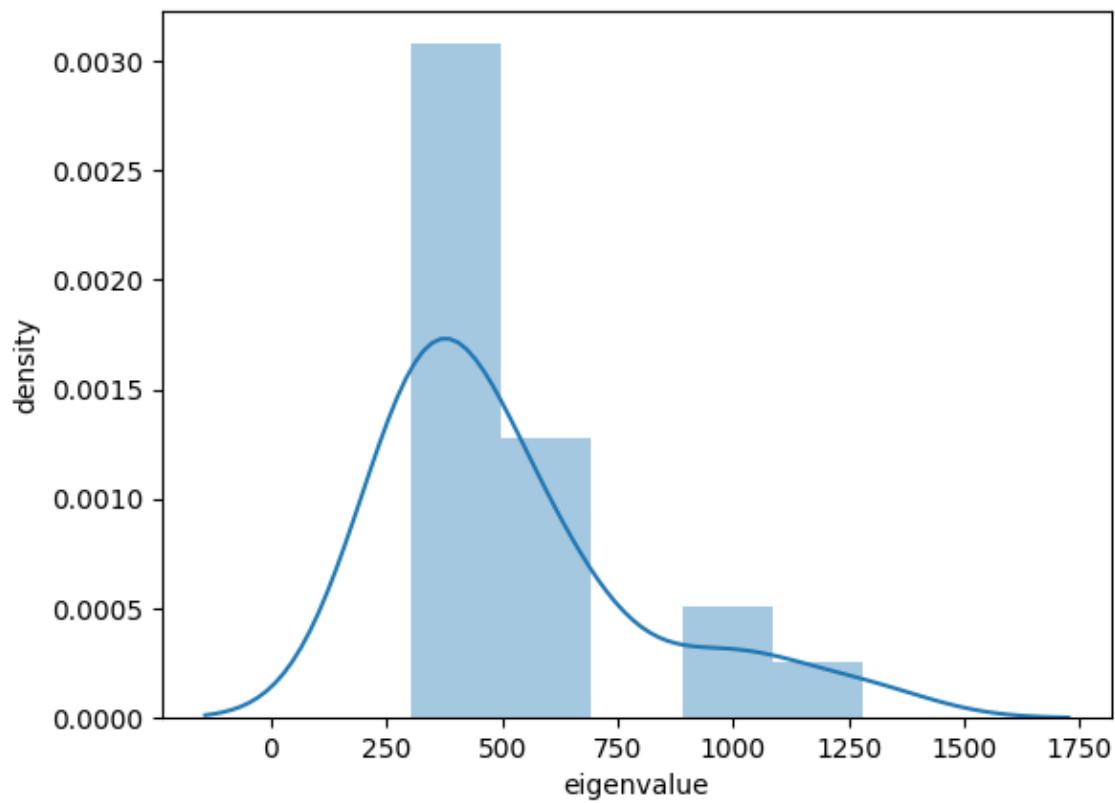
## BA network

top 20 laplacian eigenvalue distribution, eigengap: 50.04,  
time: 0.2908



## modified BA network

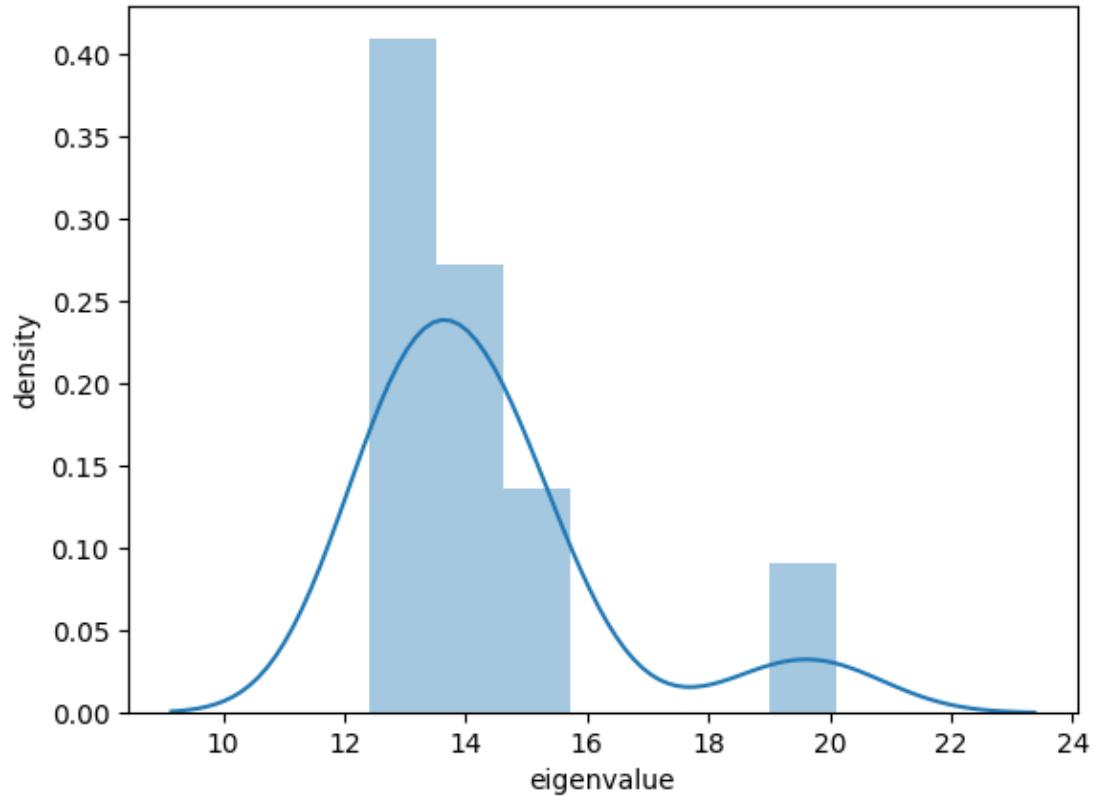
top 20 laplacian eigenvalue distribution, eigengap: 275.92,  
time: 0.6519



## powergrid

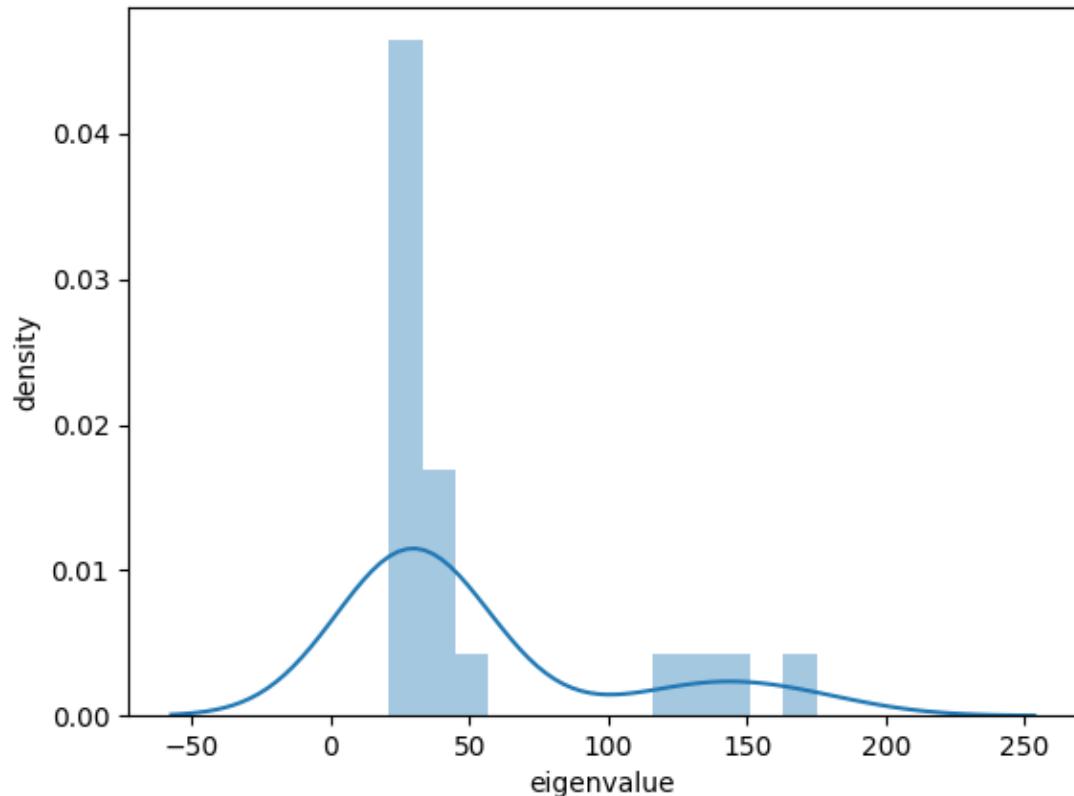
### original network

top 20 laplacian eigenvalue distribution, eigengap: 0.99,  
time: 0.2642



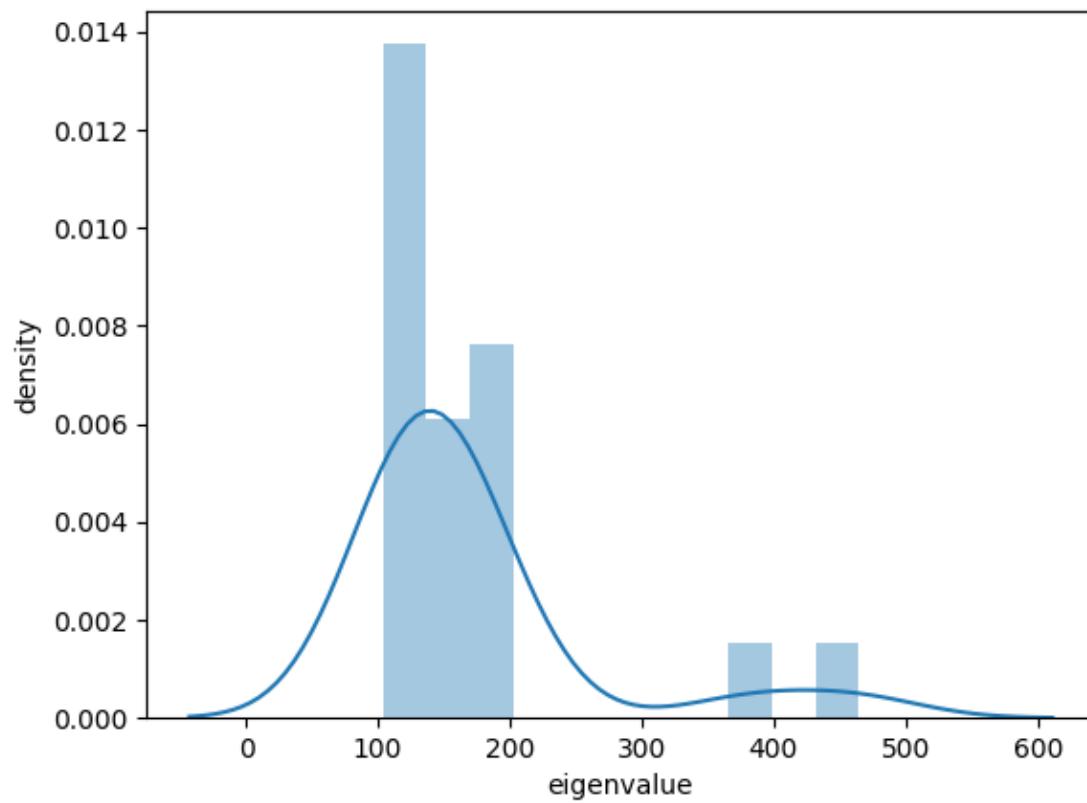
## BA network

top 20 laplacian eigenvalue distribution, eigengap: 25.98,  
time: 0.0748

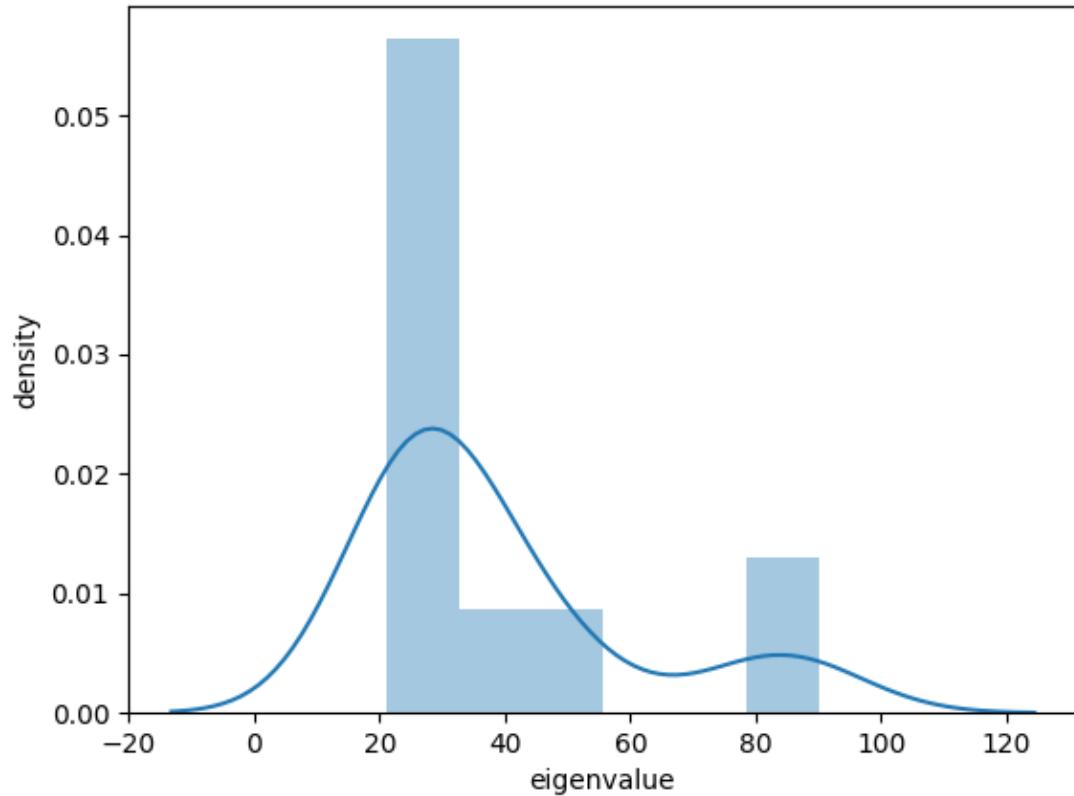


**modified BA network**

top 20 laplacian eigenvalue distribution, eigengap: 82.0,  
time: 0.0793

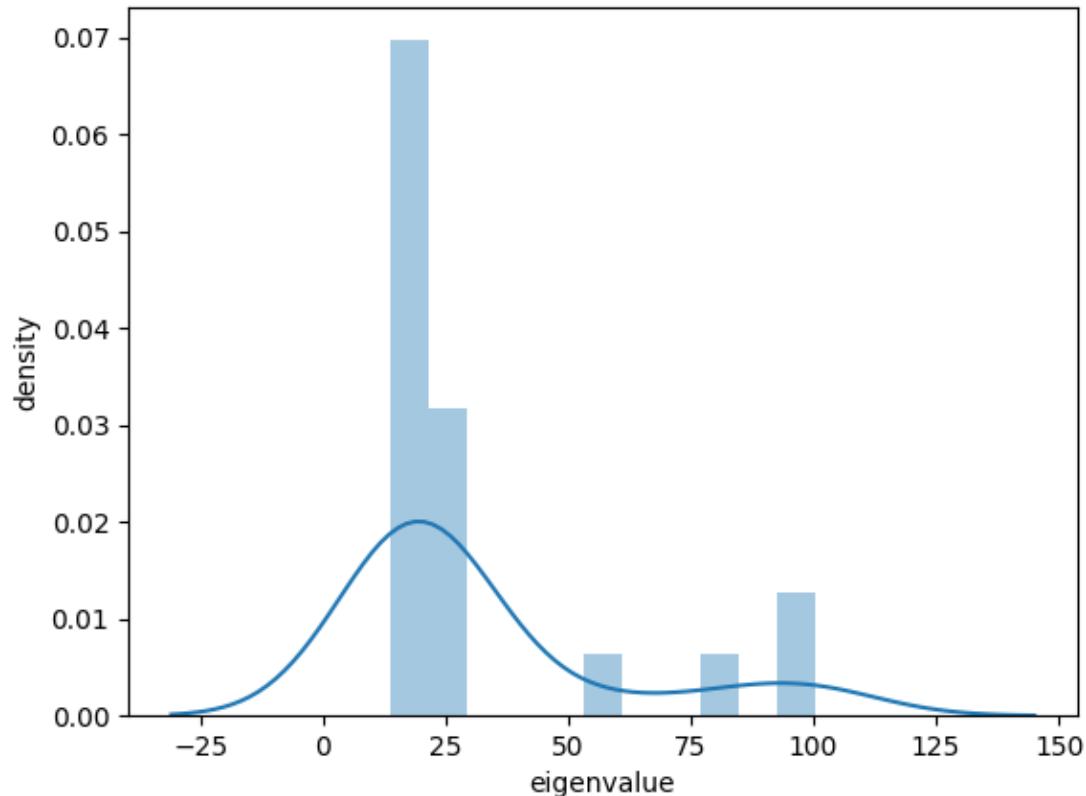
**protein****original network**

top 20 laplacian eigenvalue distribution, eigengap: 8.12,  
time: 0.1159



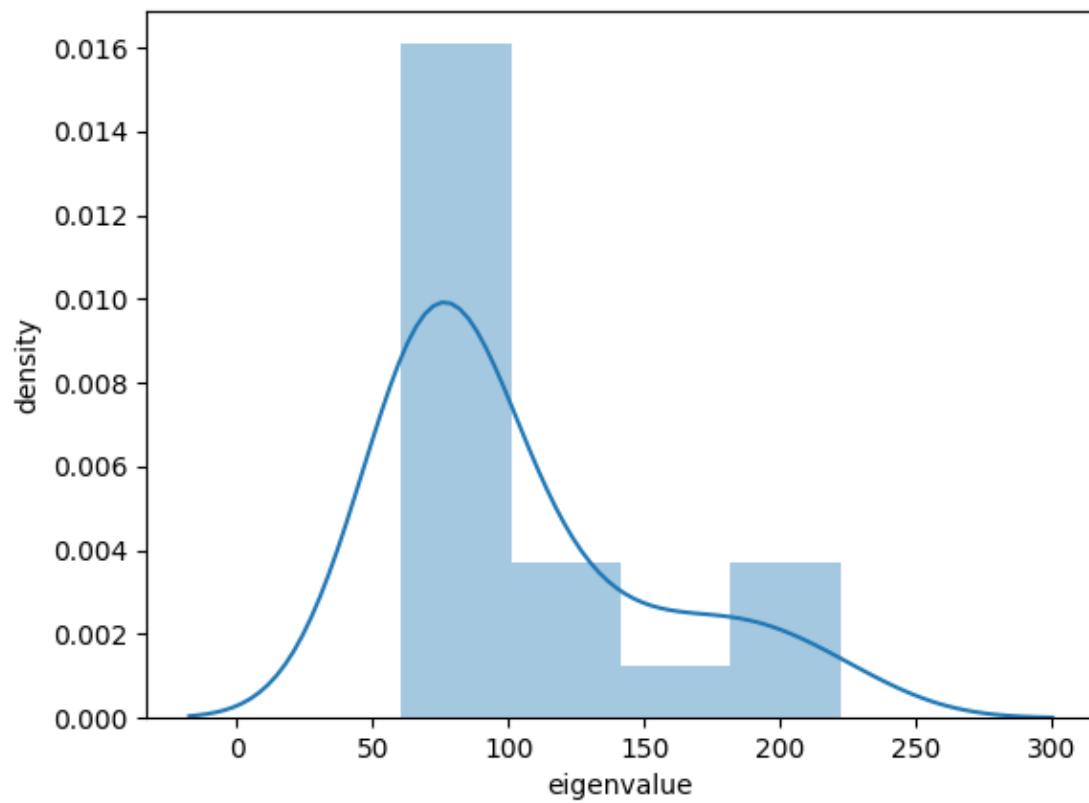
## BA network

top 20 laplacian eigenvalue distribution, eigengap: 2.71,  
time: 0.0249



## modified BA network

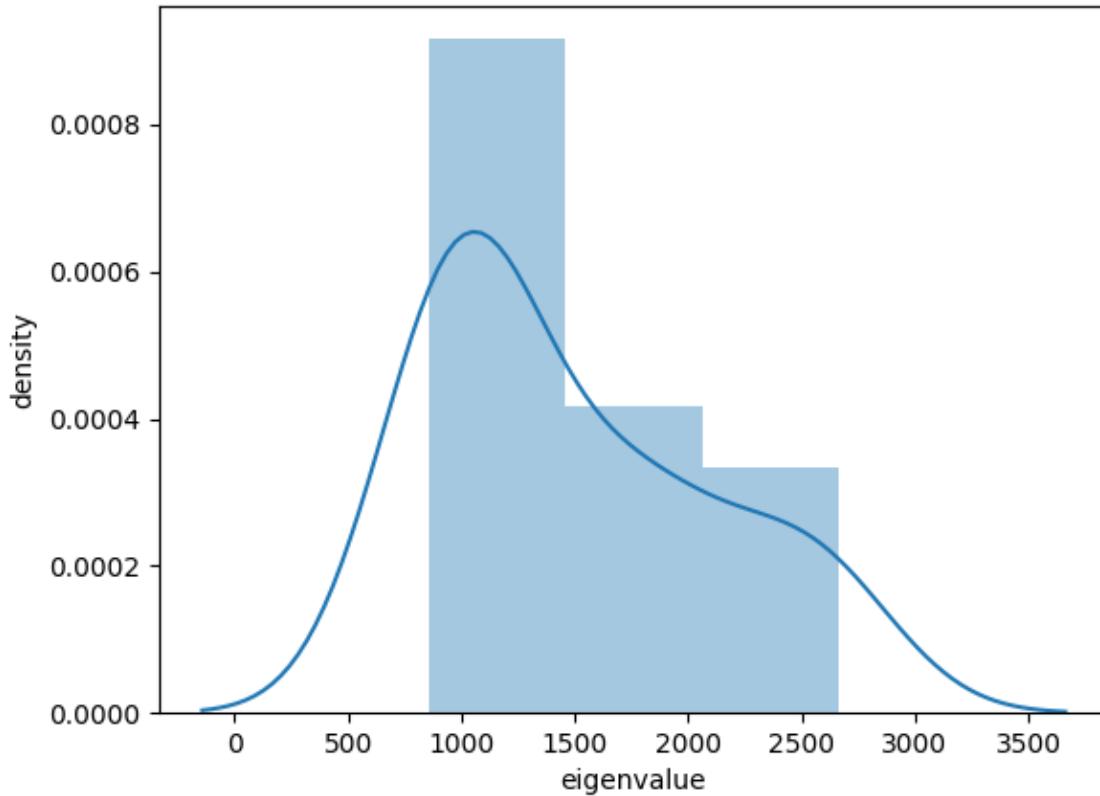
top 20 laplacian eigenvalue distribution, eigengap: 32.85,  
time: 0.1088



## WWW

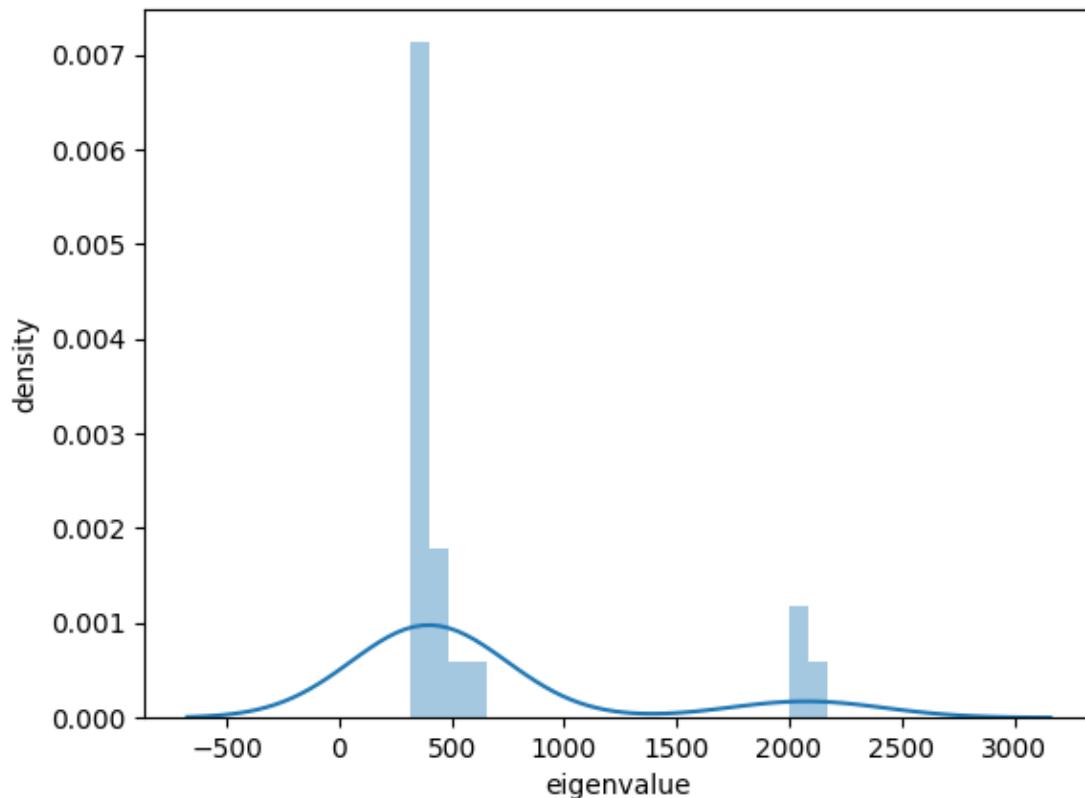
## original network

top 20 laplacian eigenvalue distribution, eigengap: 13.01,  
time: 3.2348



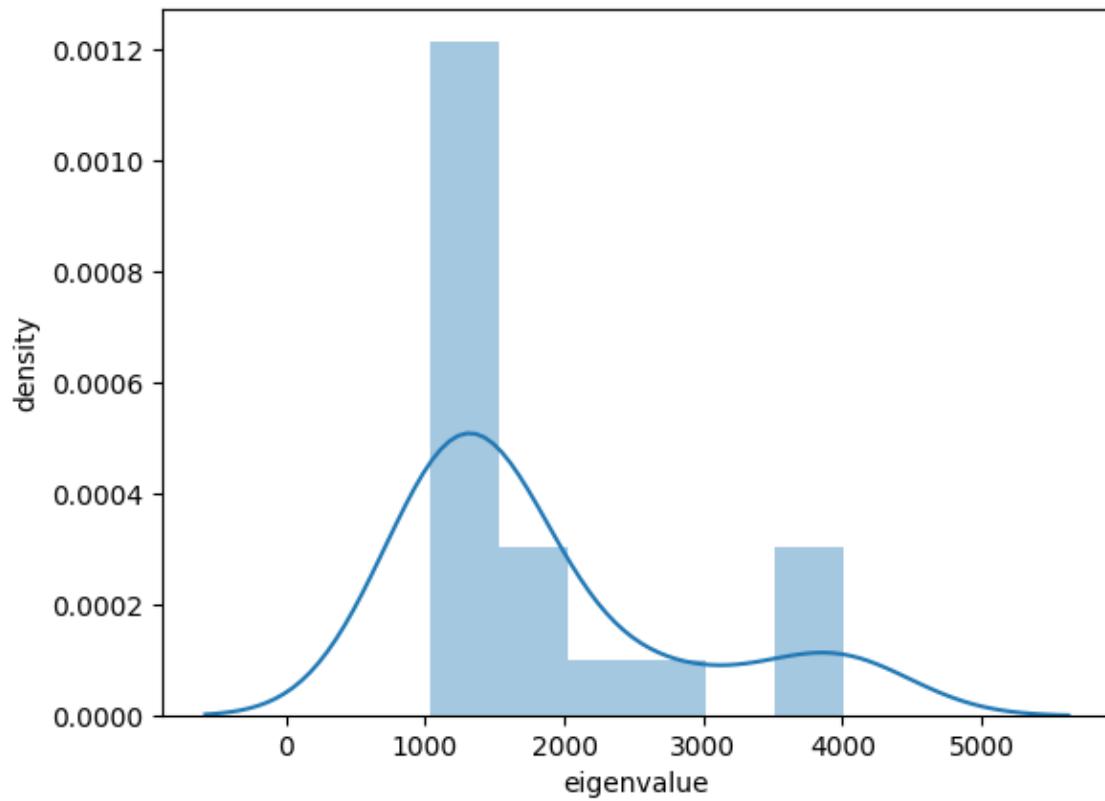
## BA network

top 20 laplacian eigenvalue distribution, eigengap: 113.0,  
time: 3.1193



## modified BA network

top 20 laplacian eigenvalue distribution, eigengap: 95.96,  
time: 8.363

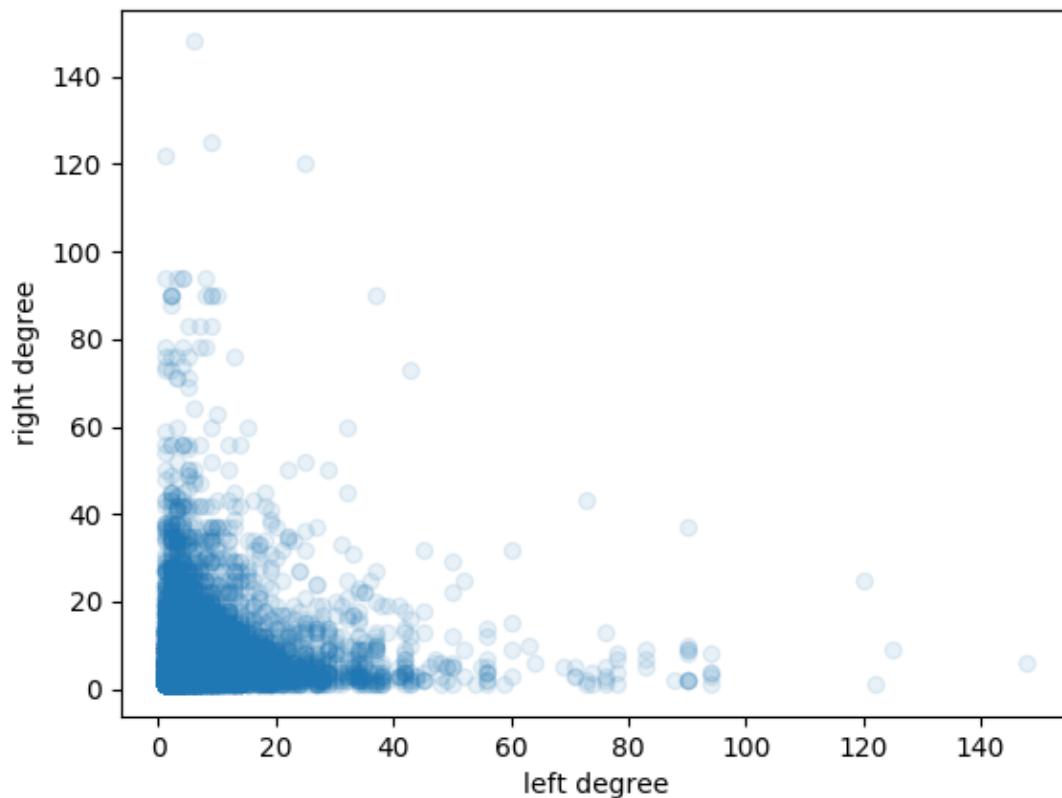


**degreecorr**

**collaboration**

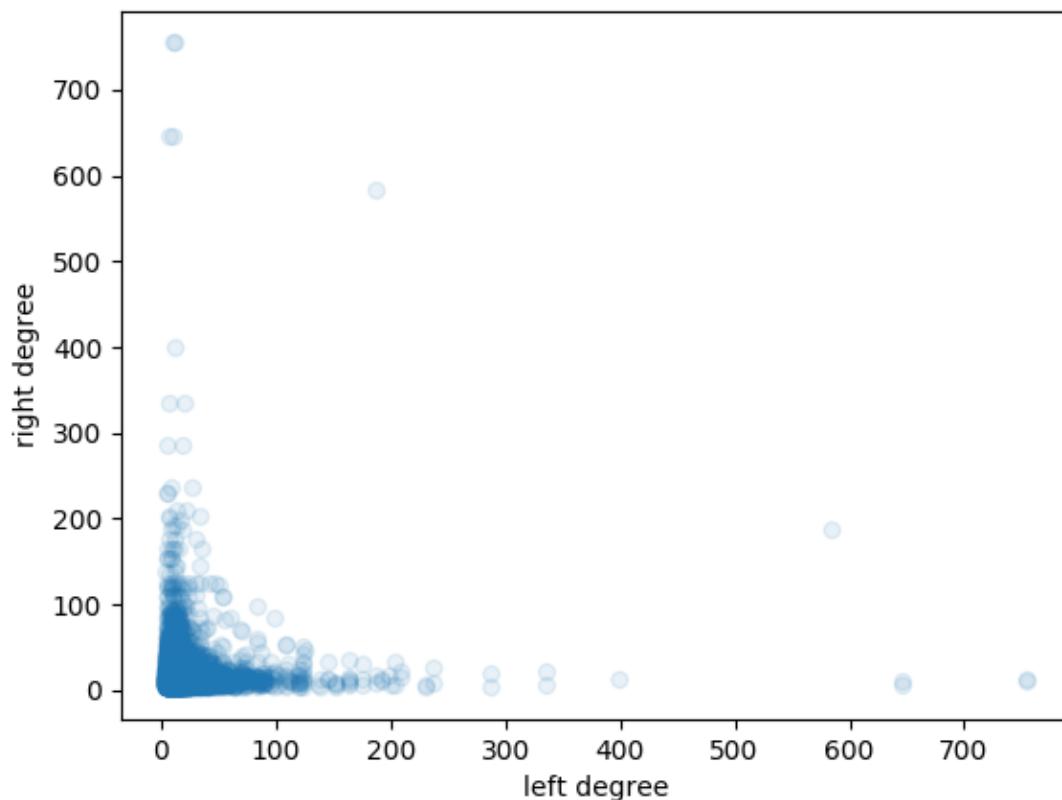
**original network**

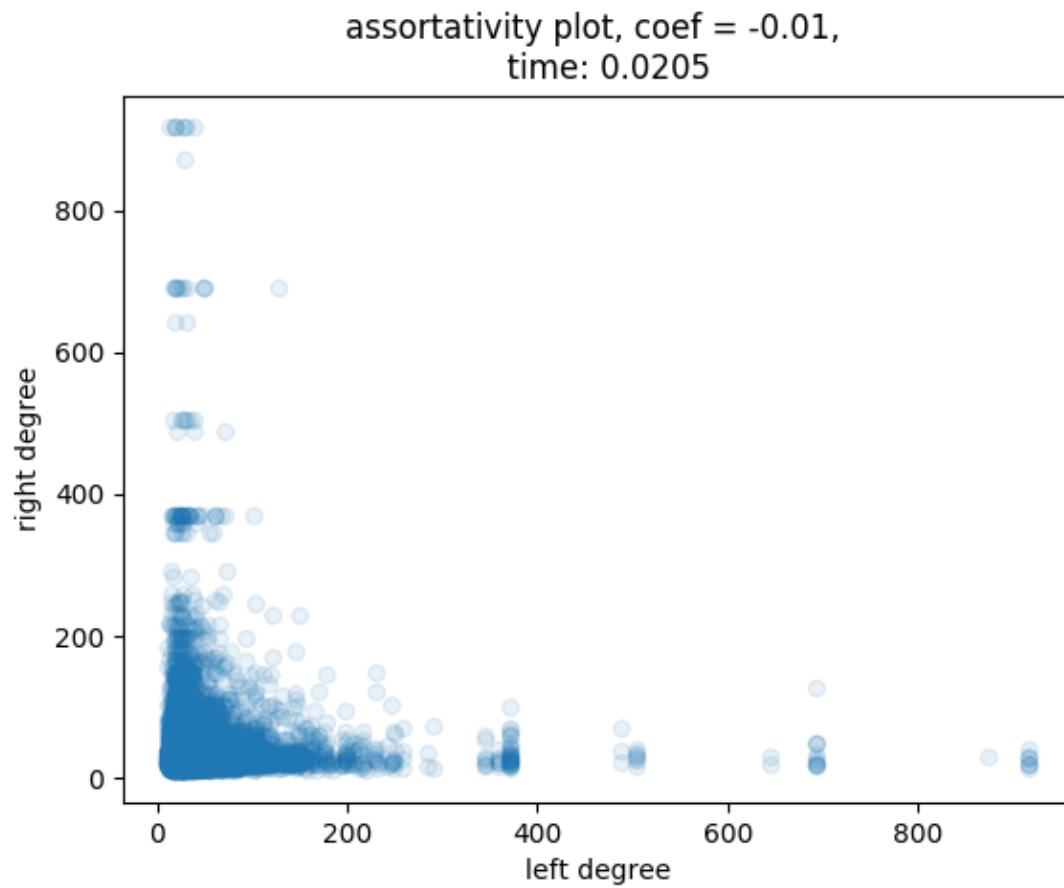
assortativity plot, coef = 0.0,  
time: 0.0092



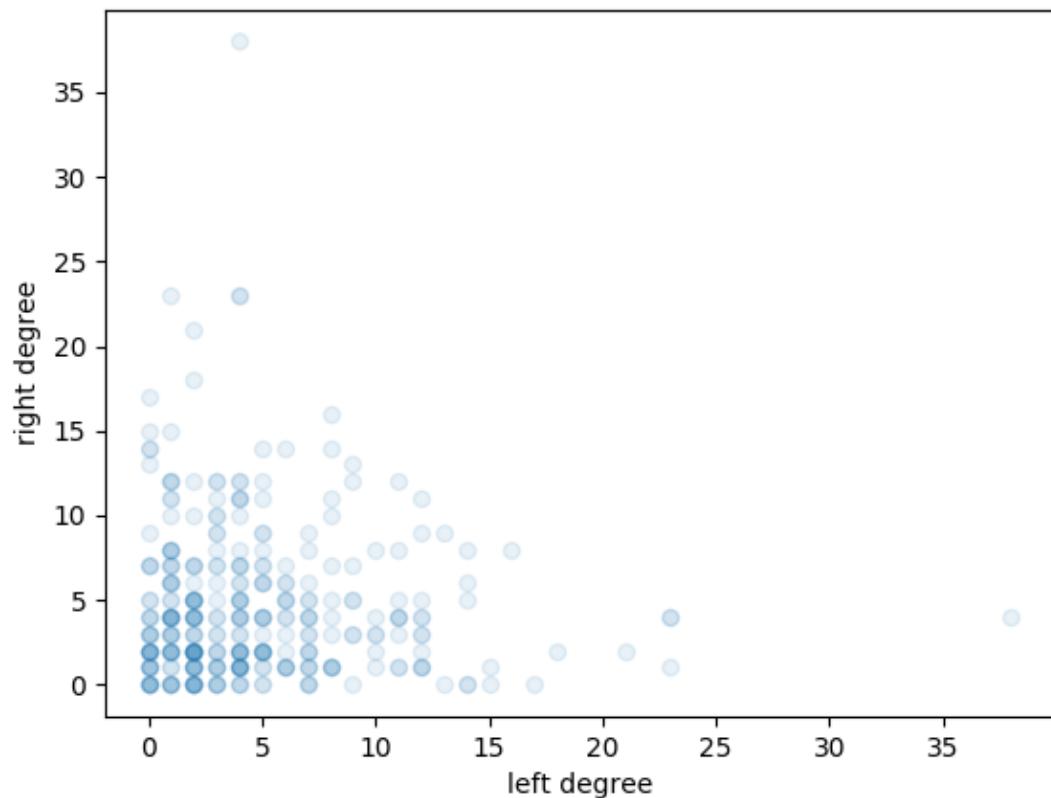
## BA network

assortativity plot, coef = 0.03,  
time: 0.0087



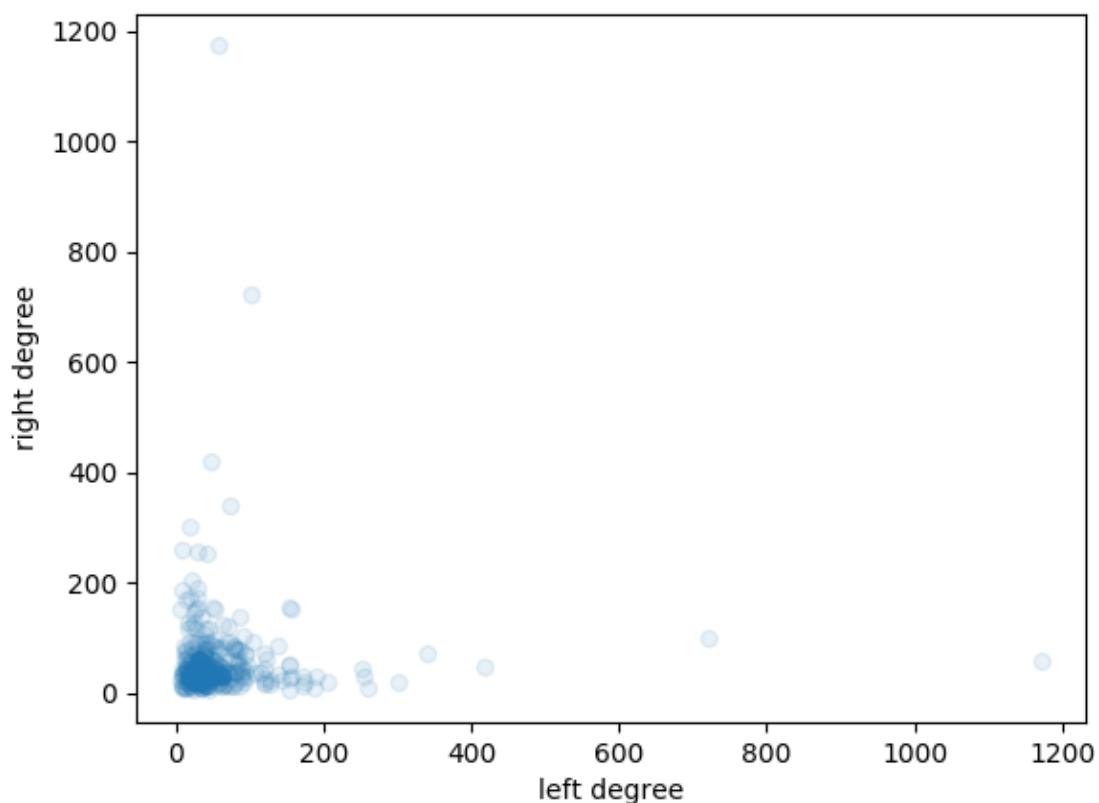
**modified BA network****citation****original network**

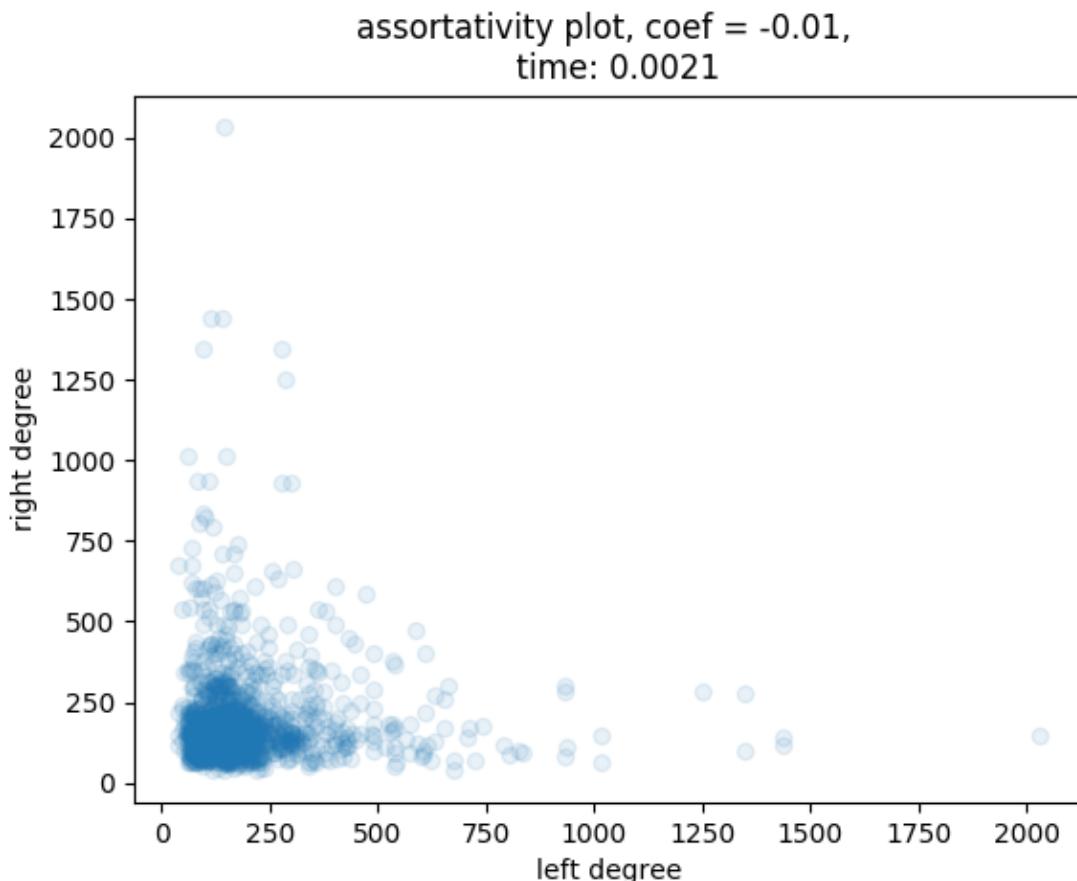
assortativity plot, coef = -0.06,  
time: 0.0006



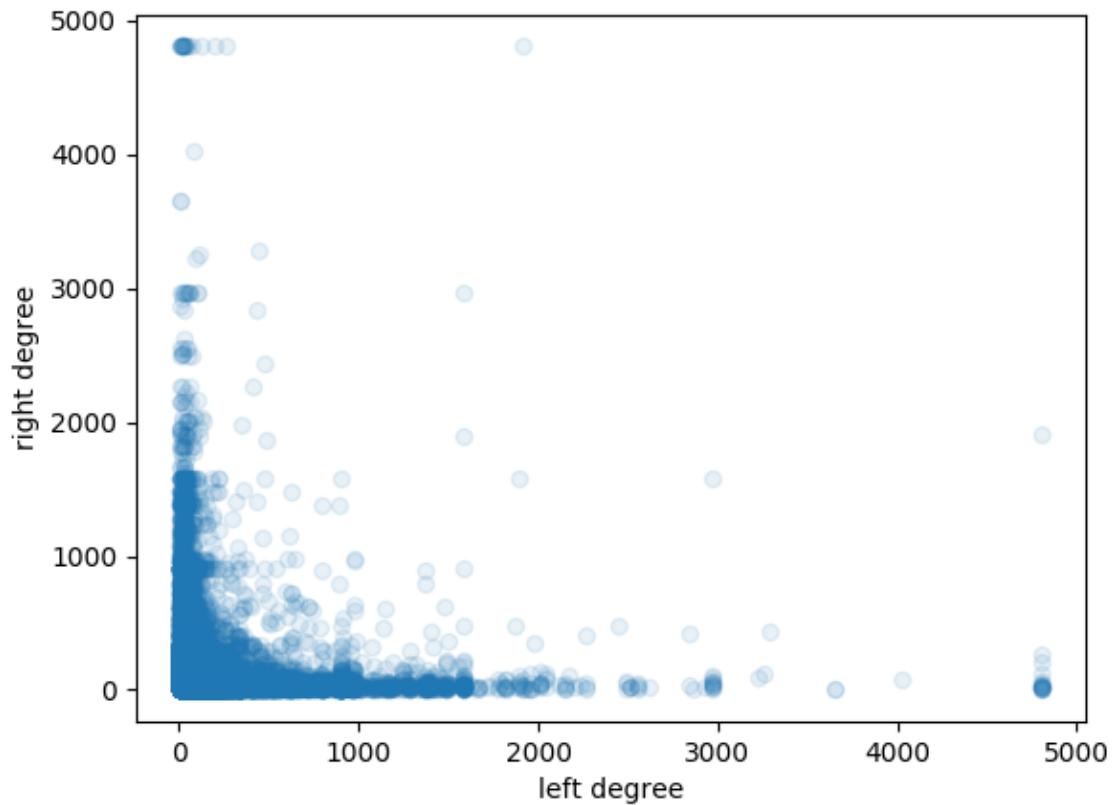
## BA network

assortativity plot, coef = 0.0,  
time: 0.0008



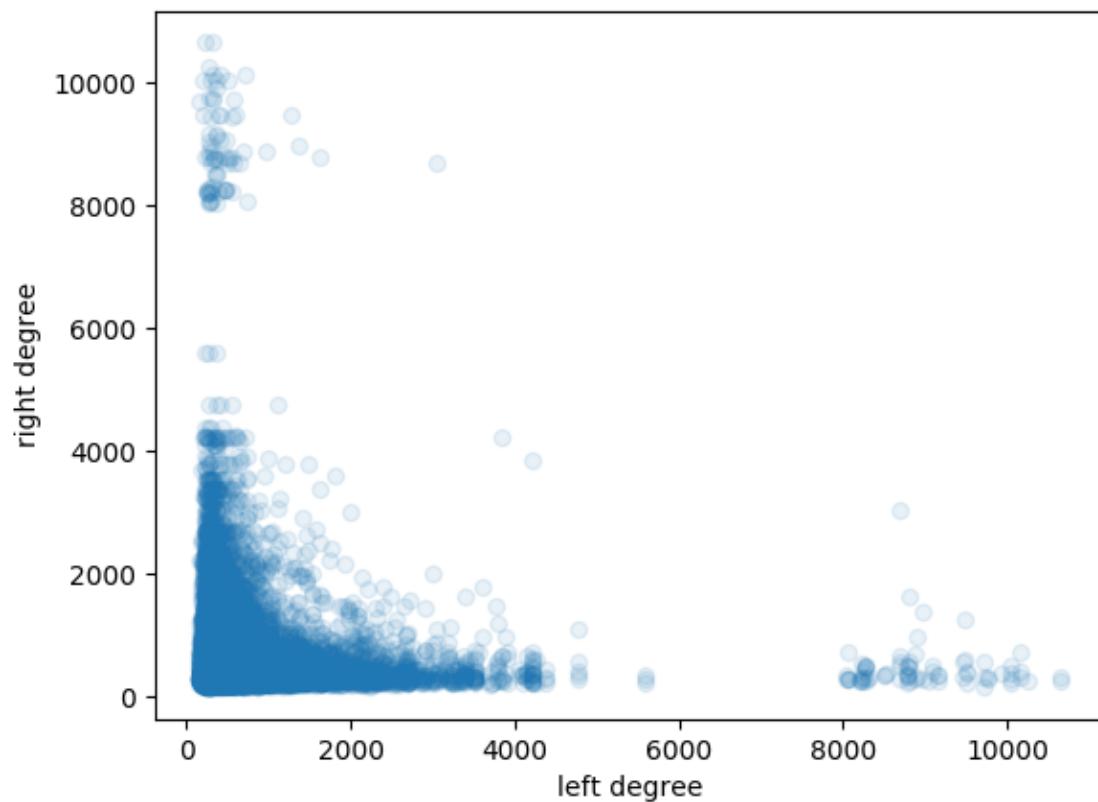
**modified BA network****actor****original network**

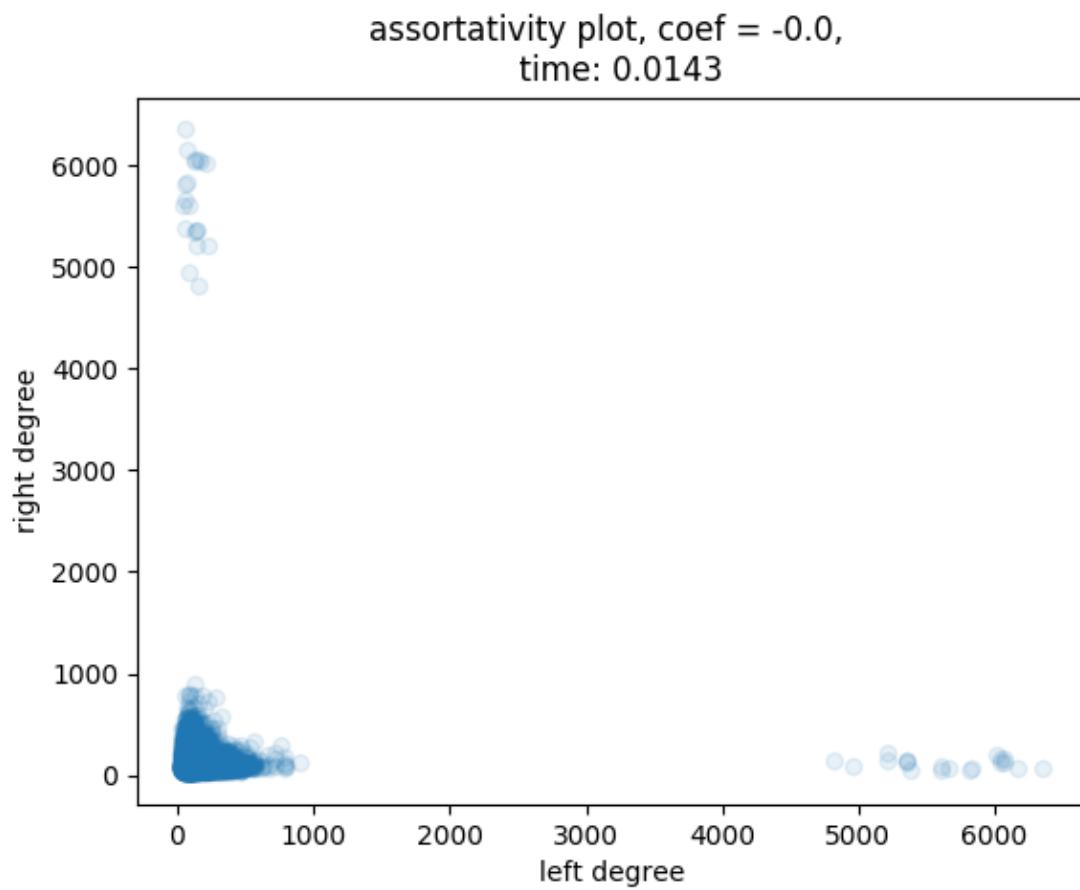
assortativity plot, coef = 0.01,  
time: 0.0496



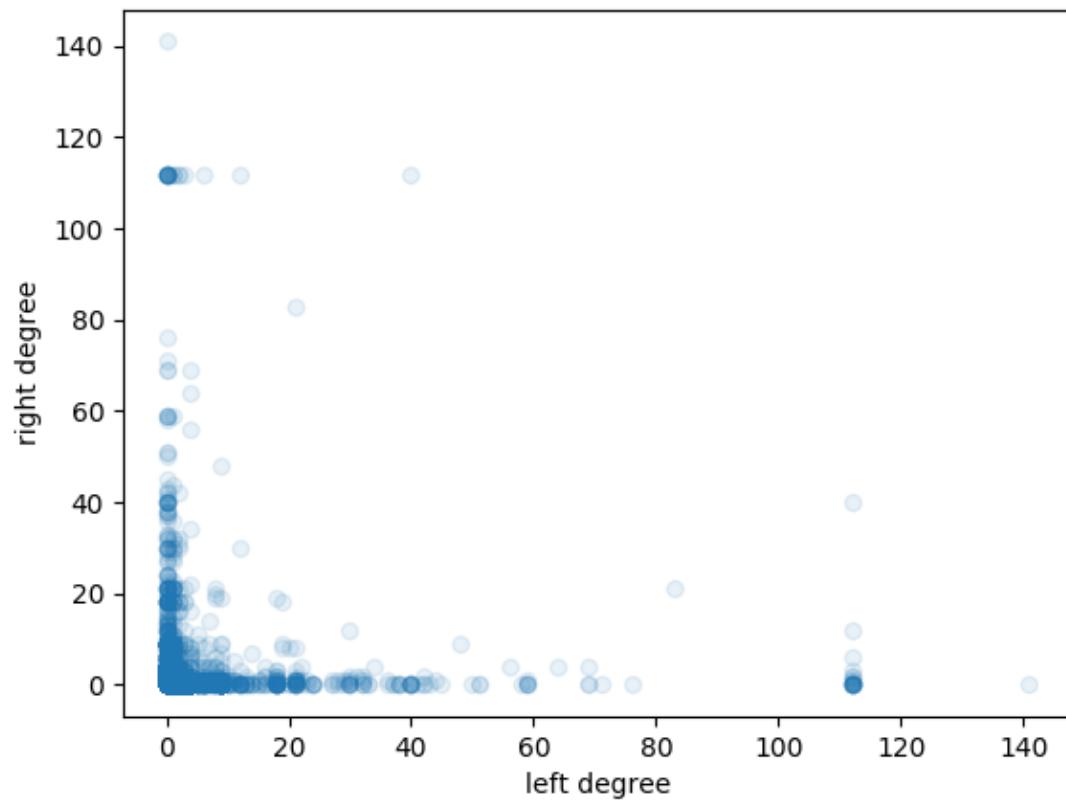
## BA network

assortativity plot, coef = -0.0,  
time: 0.0443



**modified BA network****email****original network**

assortativity plot, coef = 0.02,  
time: 0.0057



## BA network



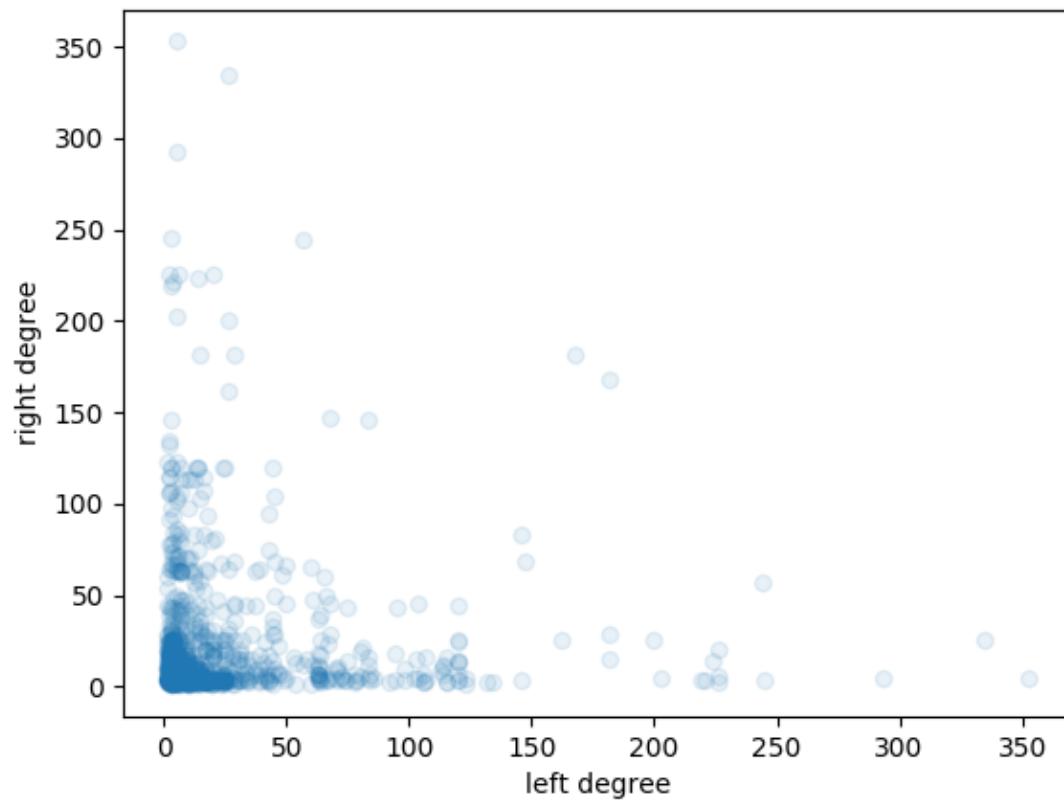
## modified BA network



## internet

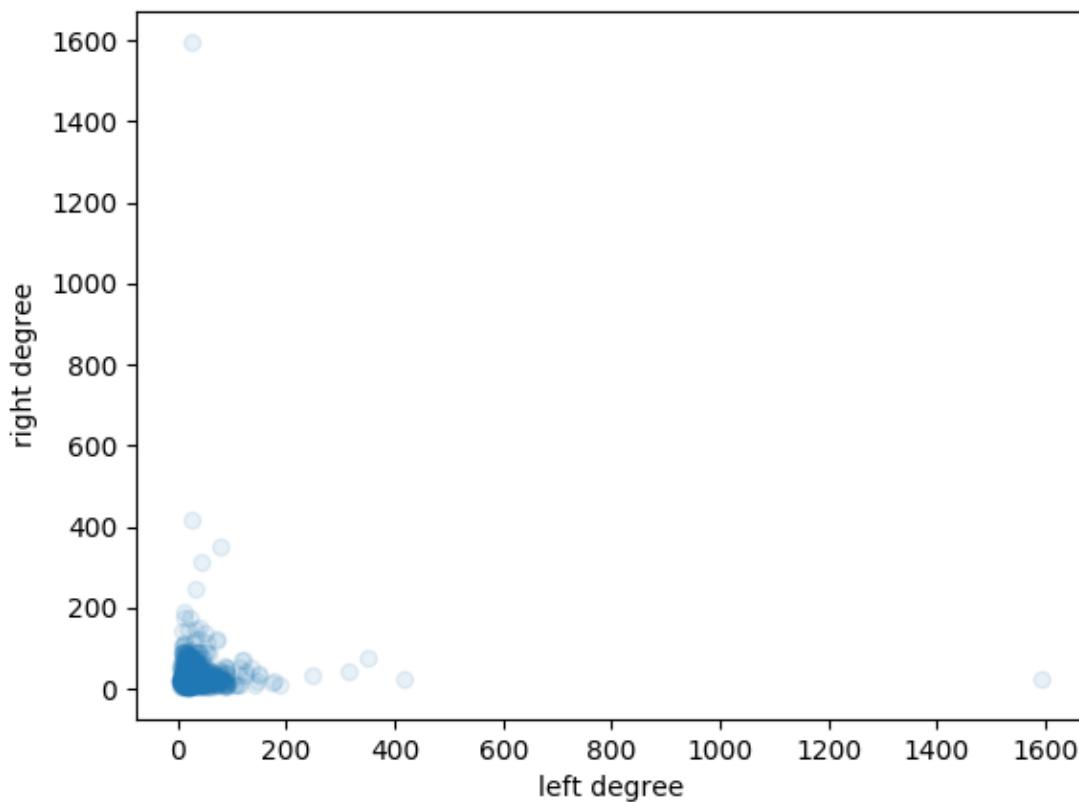
## original network

assortativity plot, coef = -0.01,  
time: 0.0019

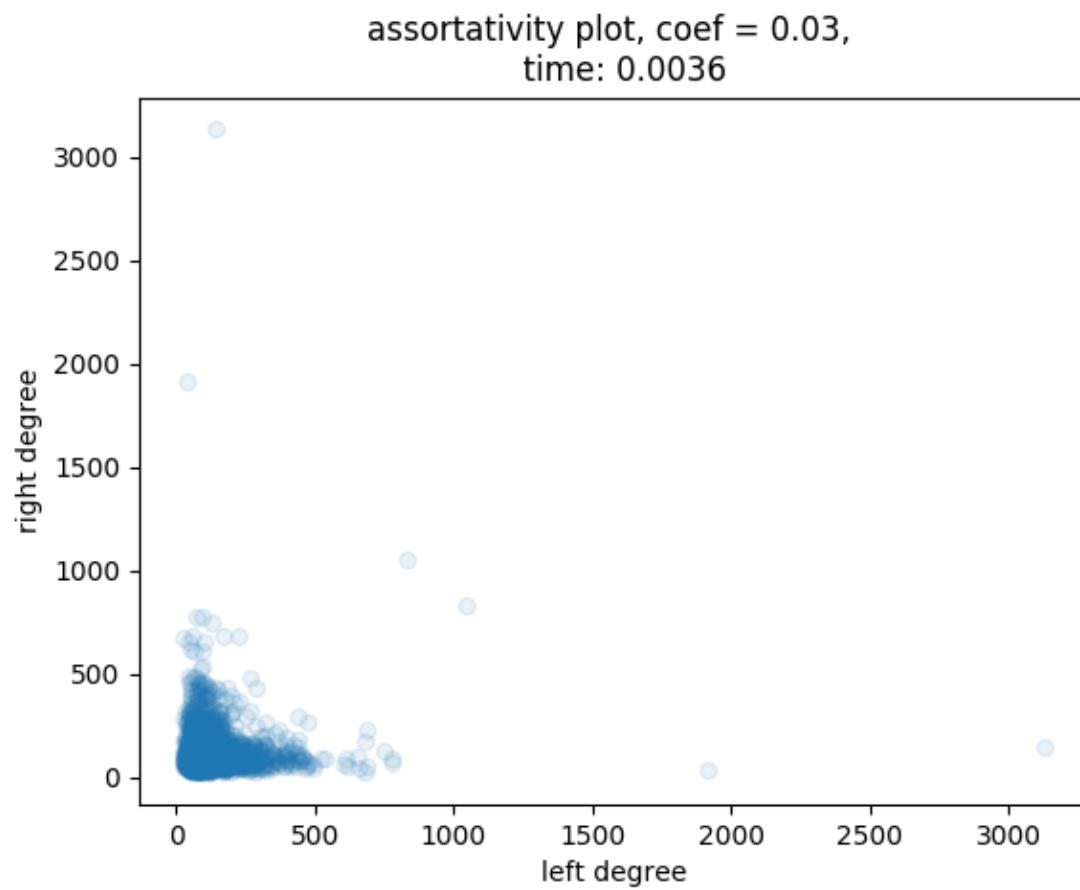


## BA network

assortativity plot, coef = -0.01,  
time: 0.0014



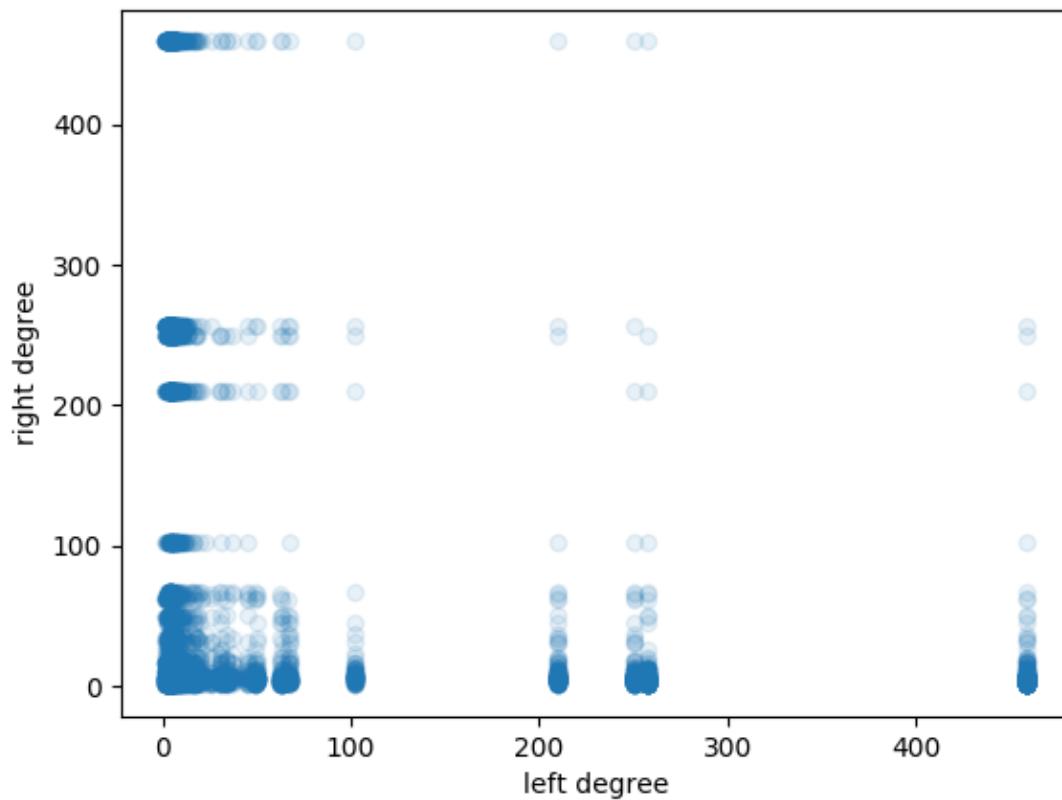
## modified BA network



## metabolic

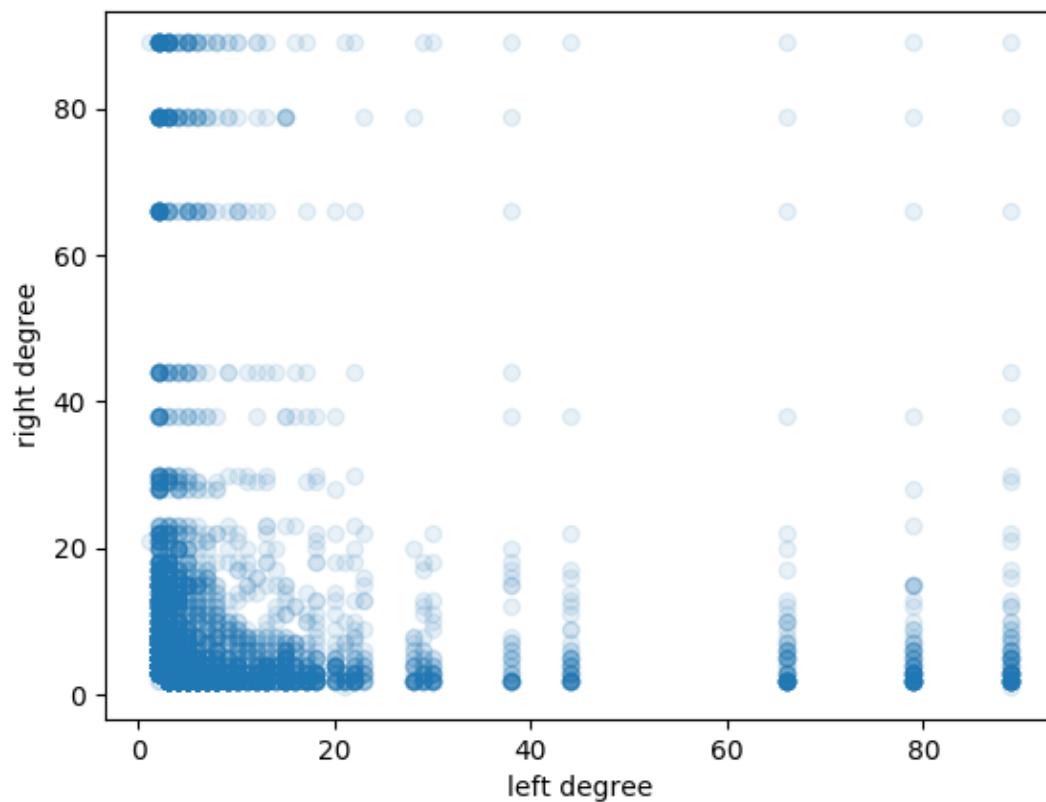
### original network

assortativity plot, coef = -0.26,  
time: 0.006

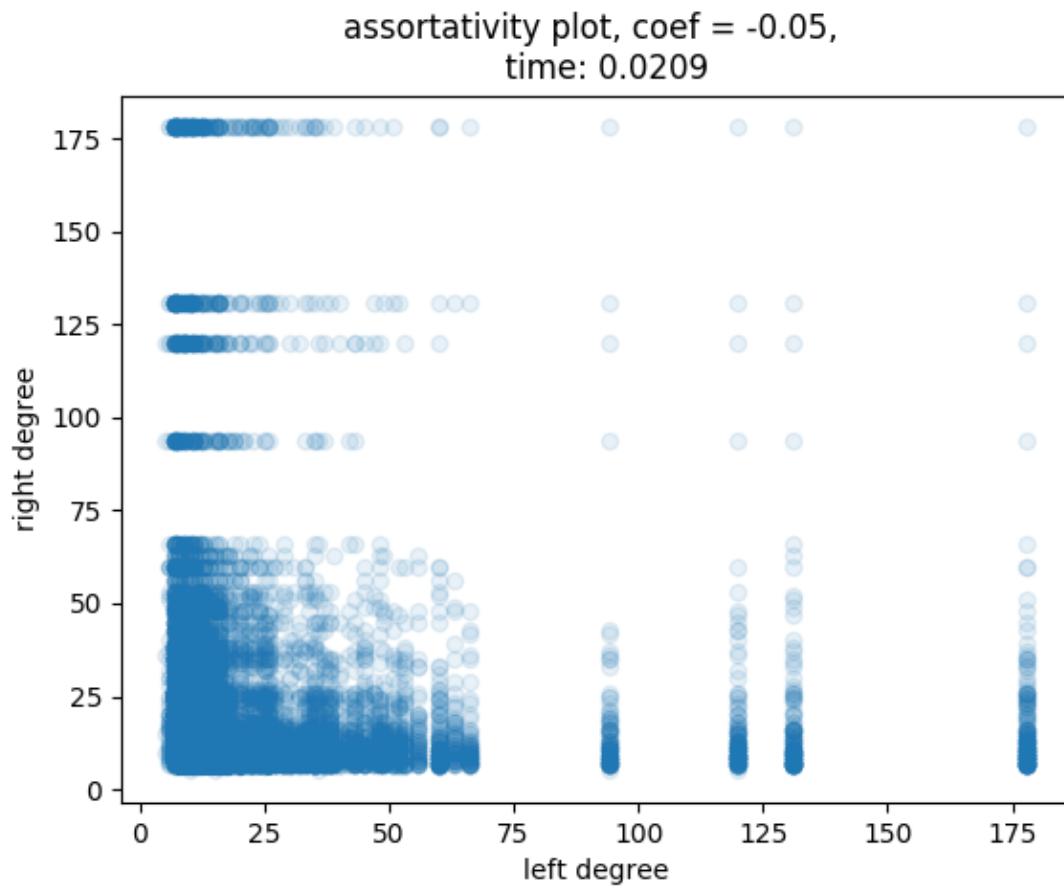


## BA network

assortativity plot, coef = -0.07,  
time: 0.0042



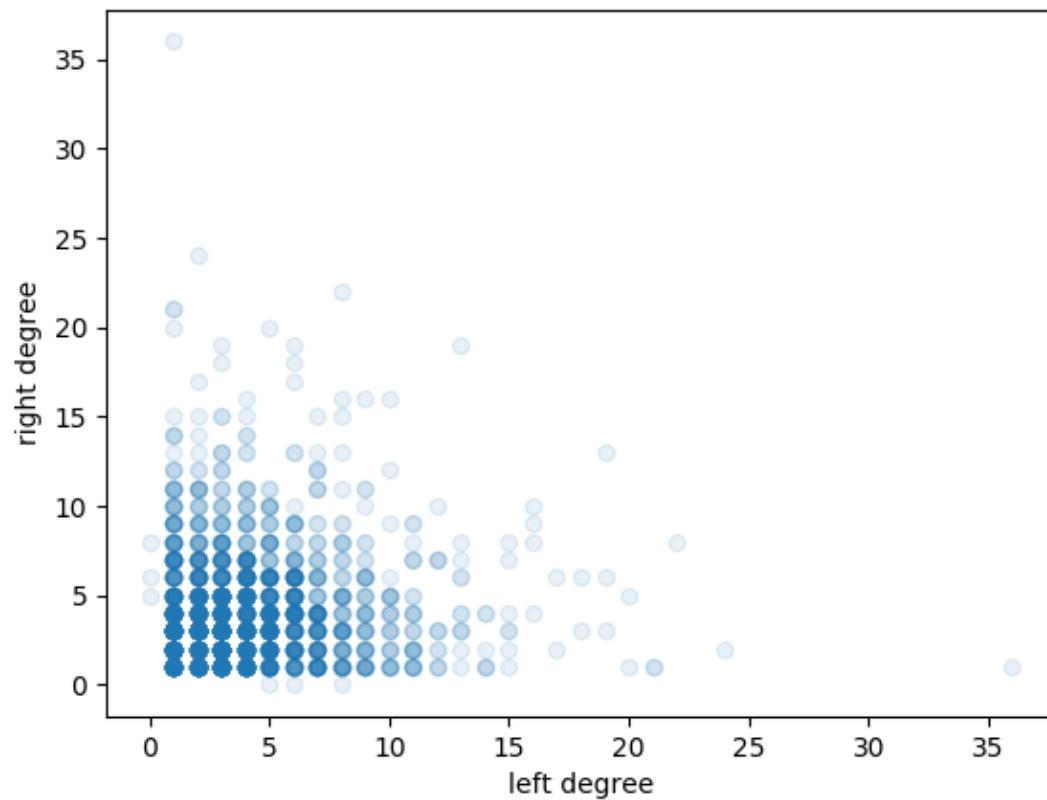
## modified BA network



## phonecalls

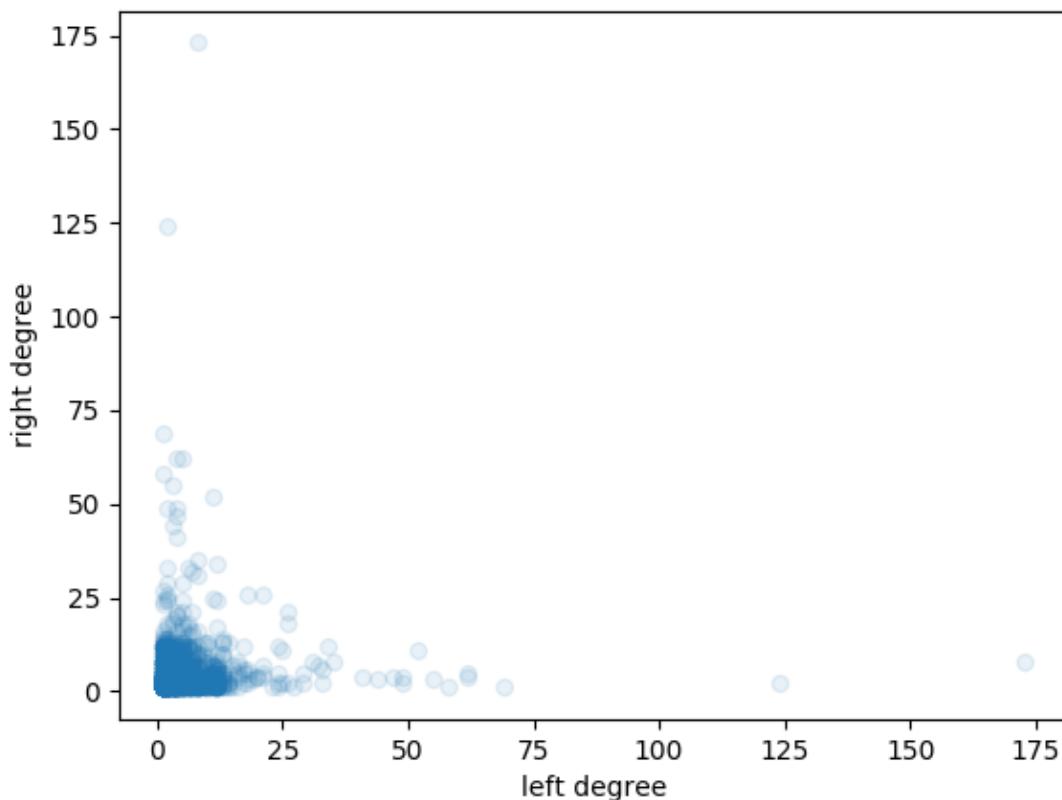
### original network

assortativity plot, coef = -0.01,  
time: 0.0038

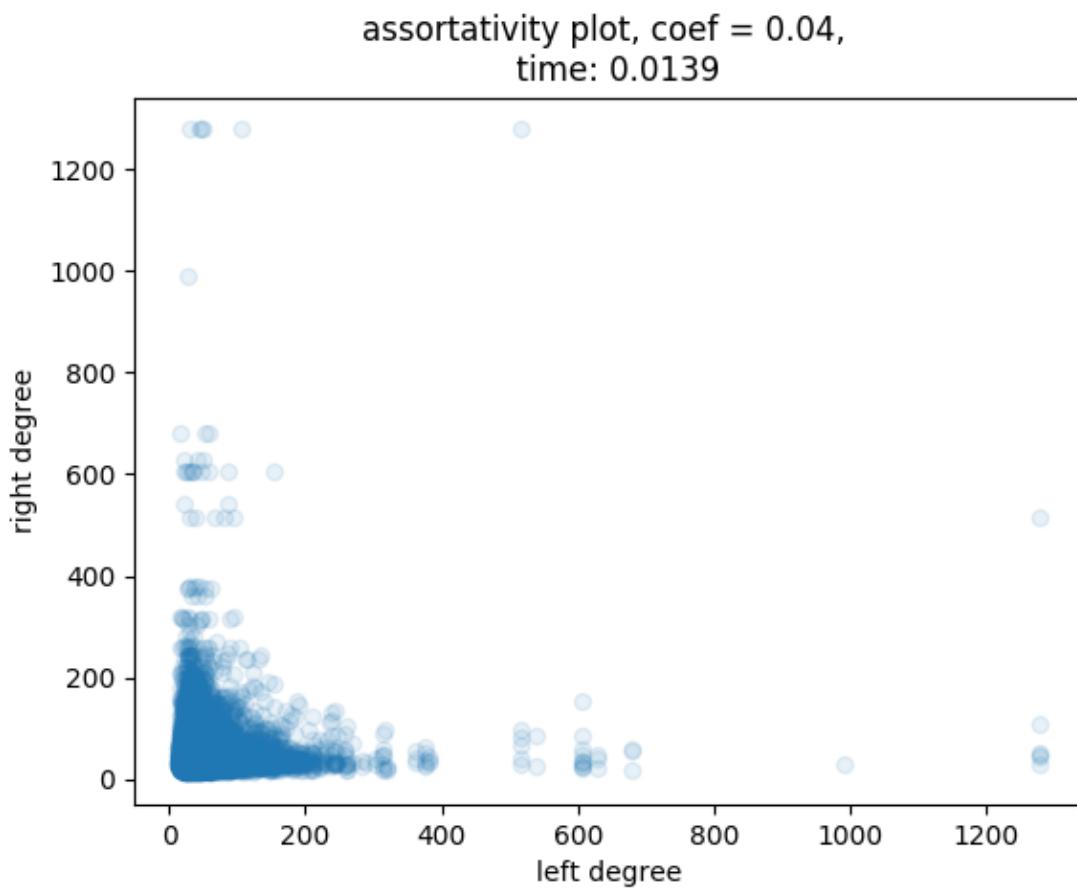


## BA network

assortativity plot, coef = 0.02,  
time: 0.0026



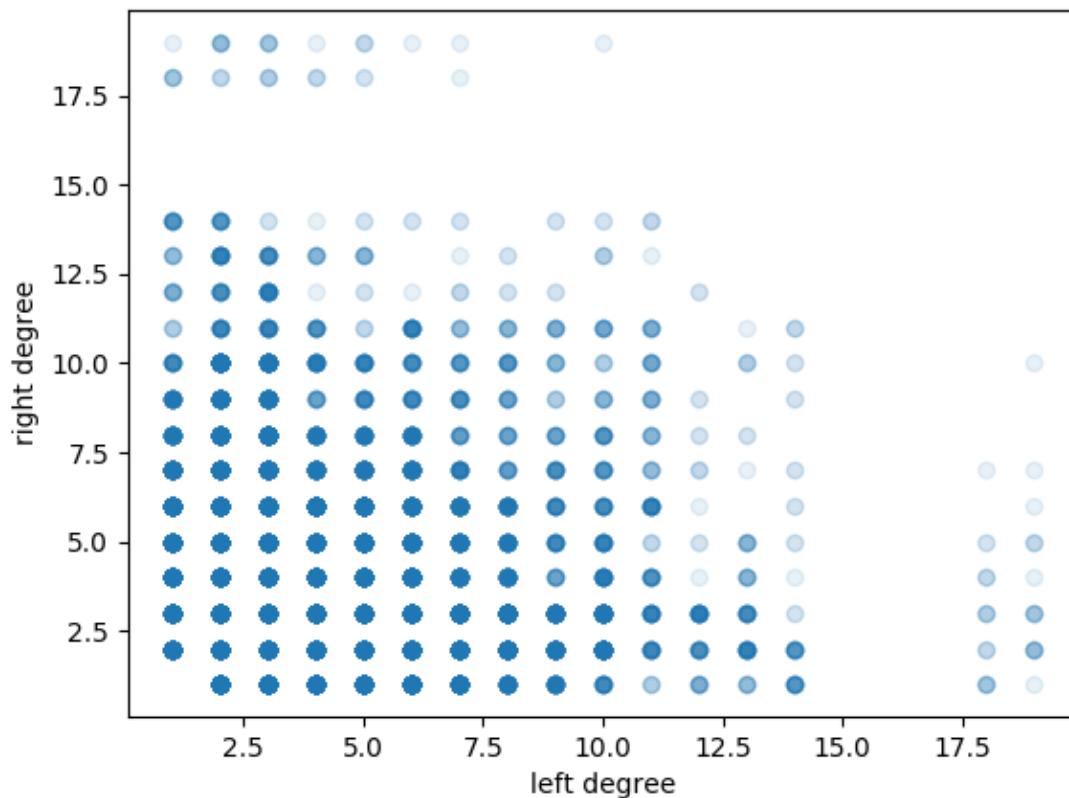
## modified BA network



## powergrid

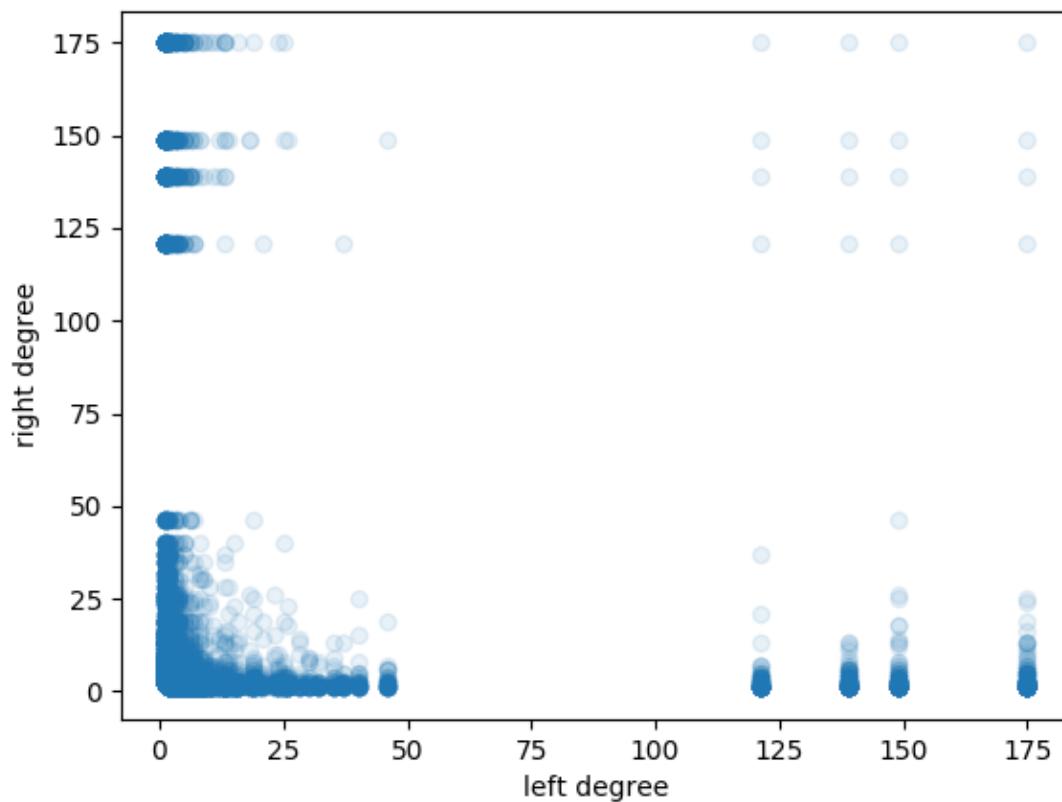
### original network

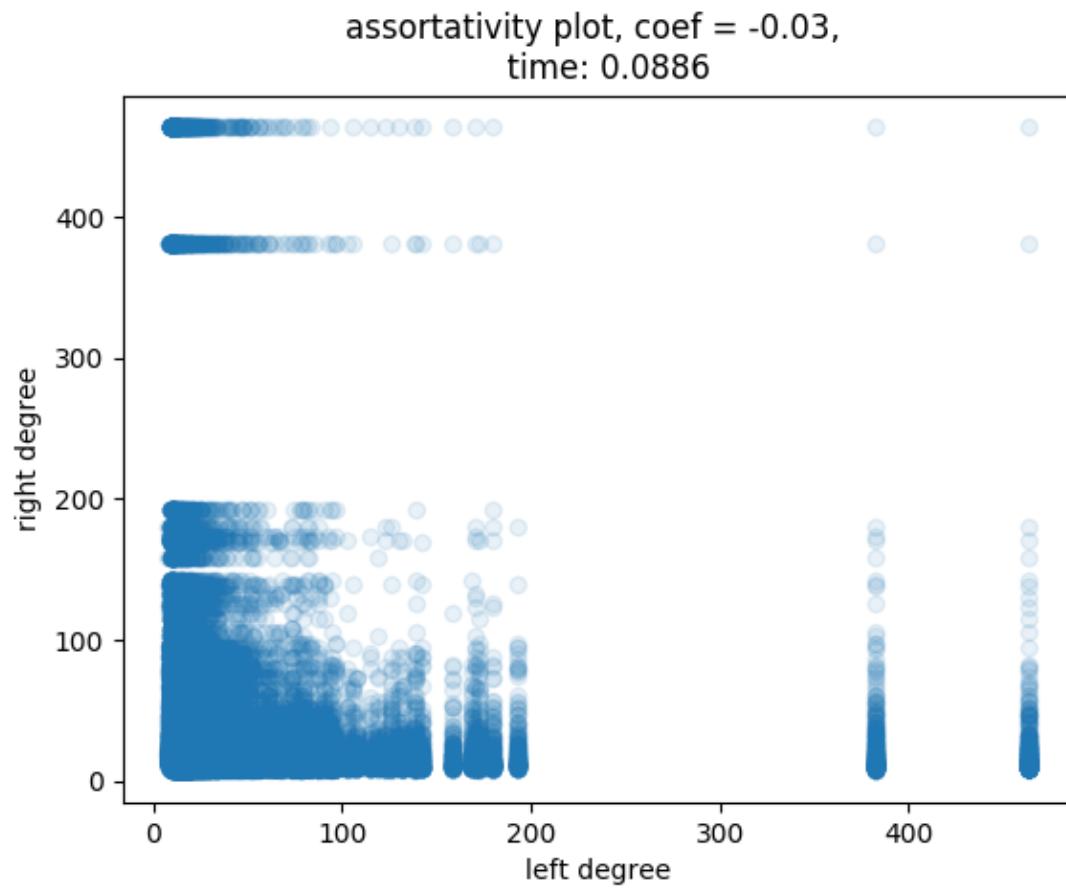
assortativity plot, coef = 0.0,  
time: 0.0238



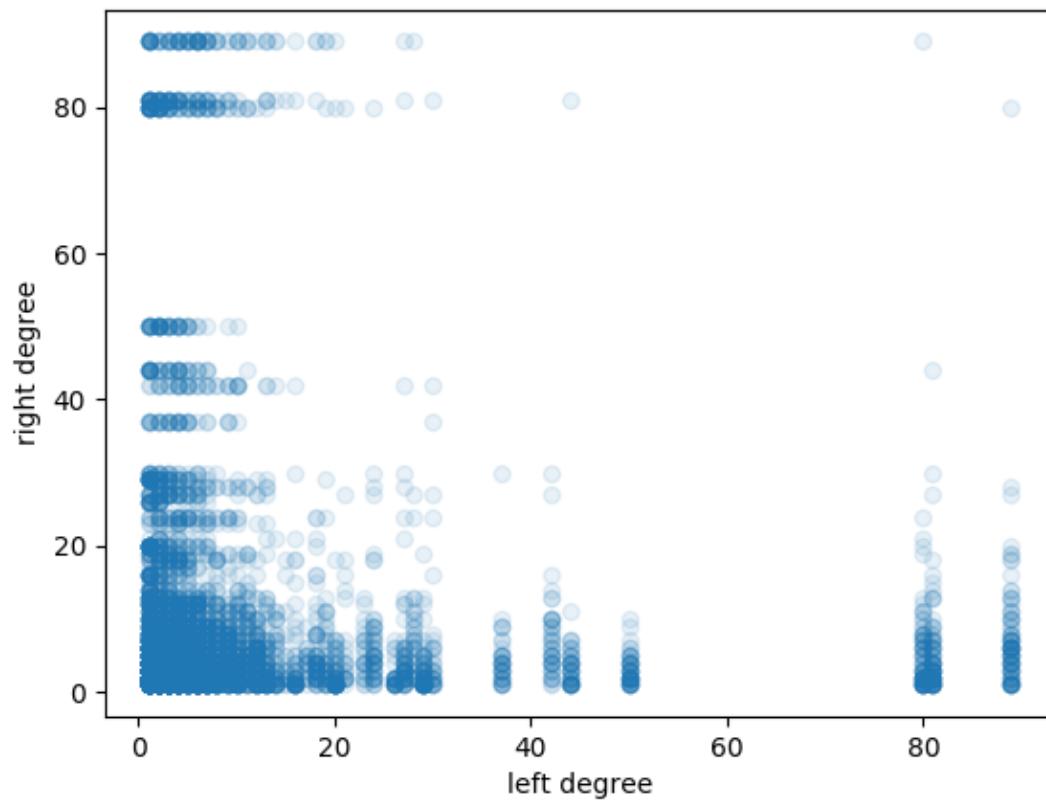
## BA network

assortativity plot, coef = -0.07,  
time: 0.0103



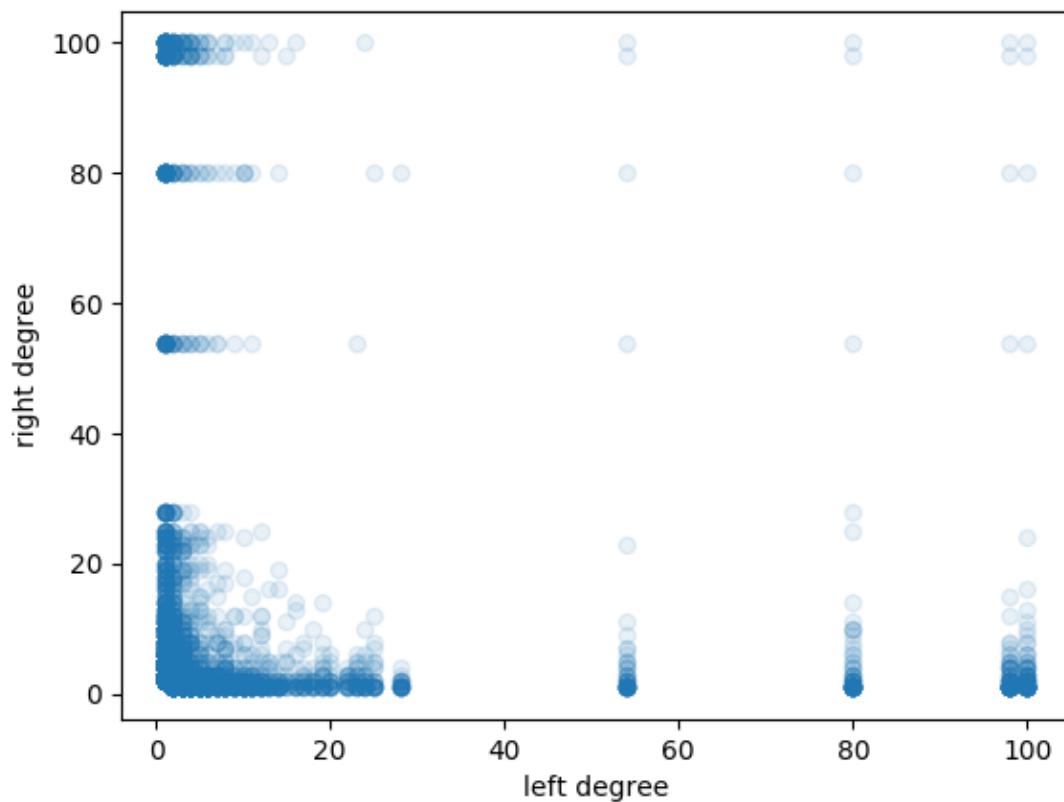
**modified BA network****protein****original network**

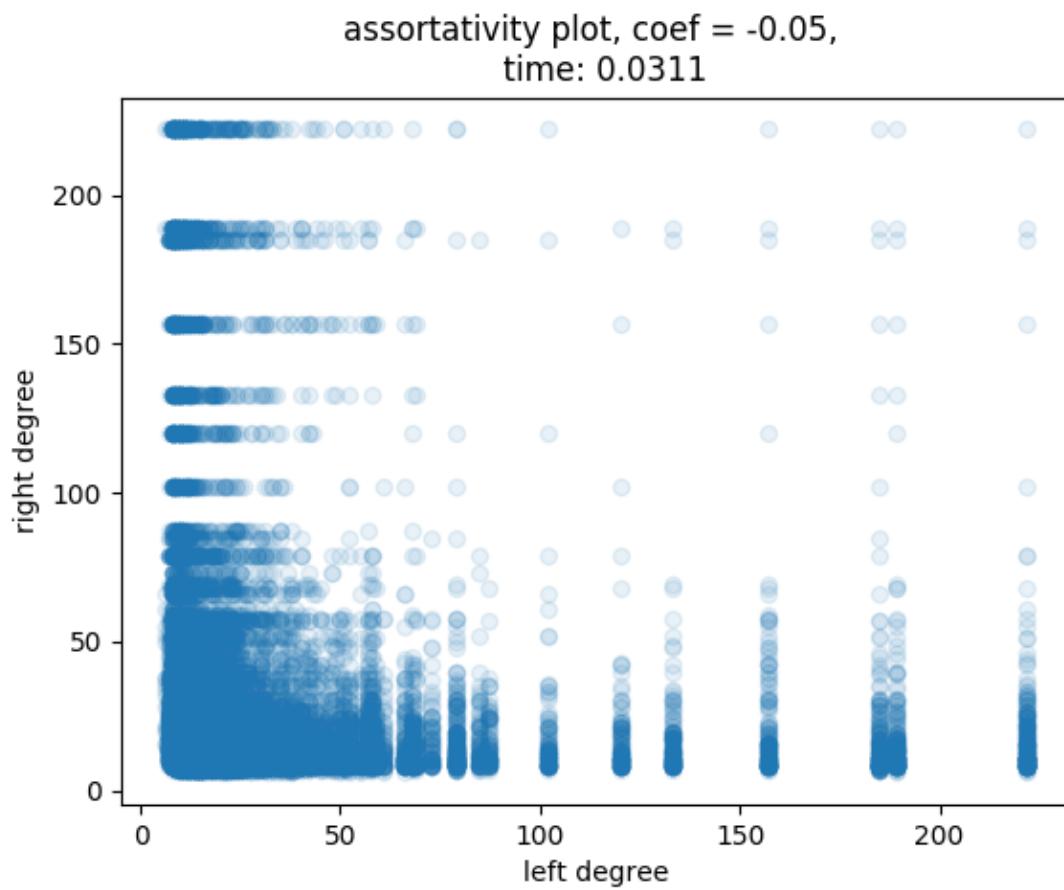
assortativity plot, coef = -0.08,  
time: 0.0079



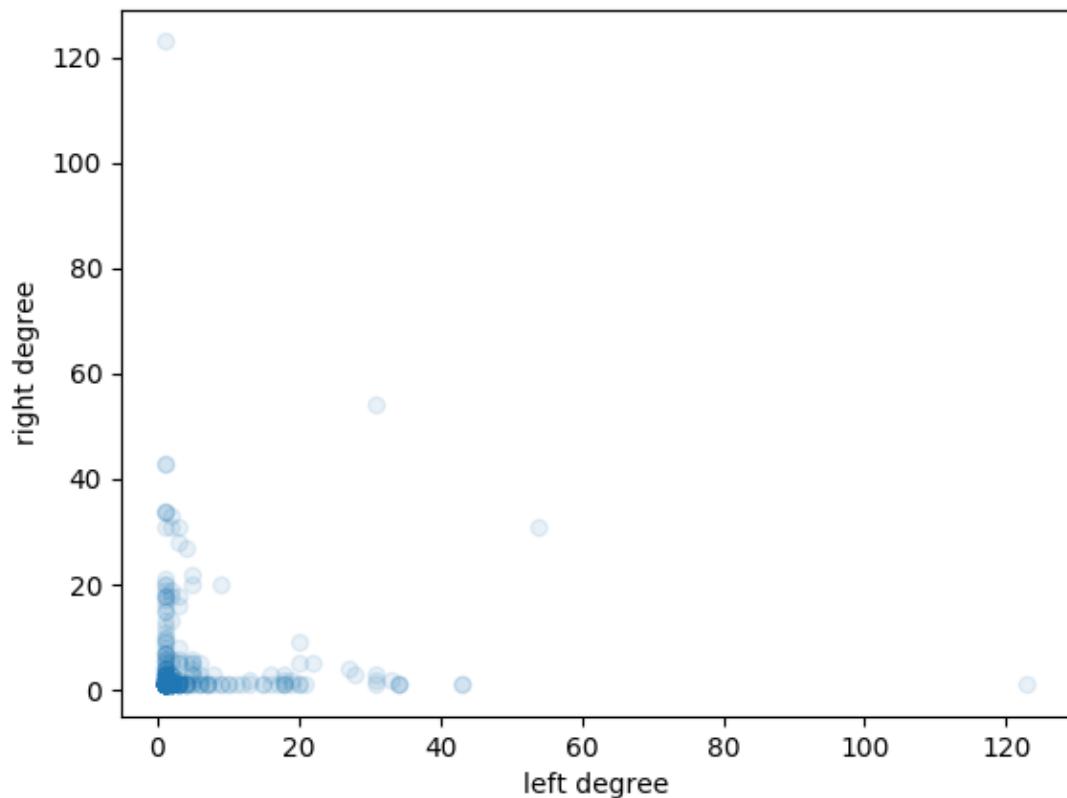
## BA network

assortativity plot, coef = -0.09,  
time: 0.0044



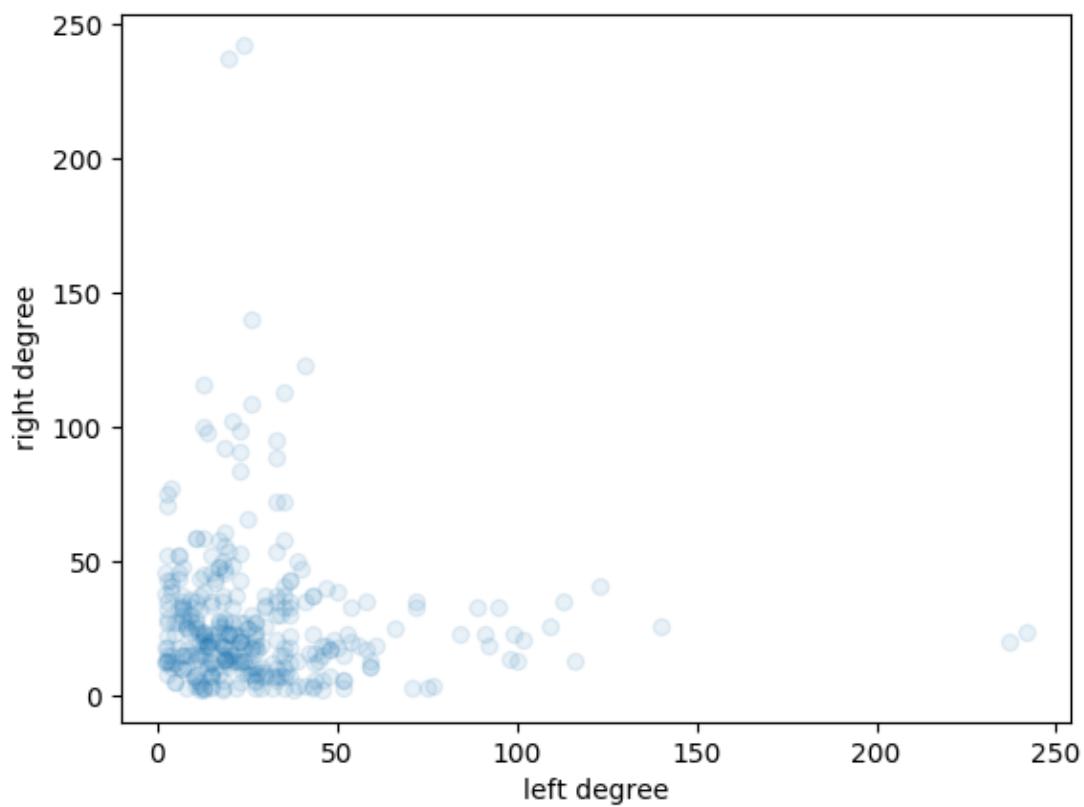
**modified BA network****WWW****original network**

assortativity plot, coef = -0.0,  
time: 0.0008



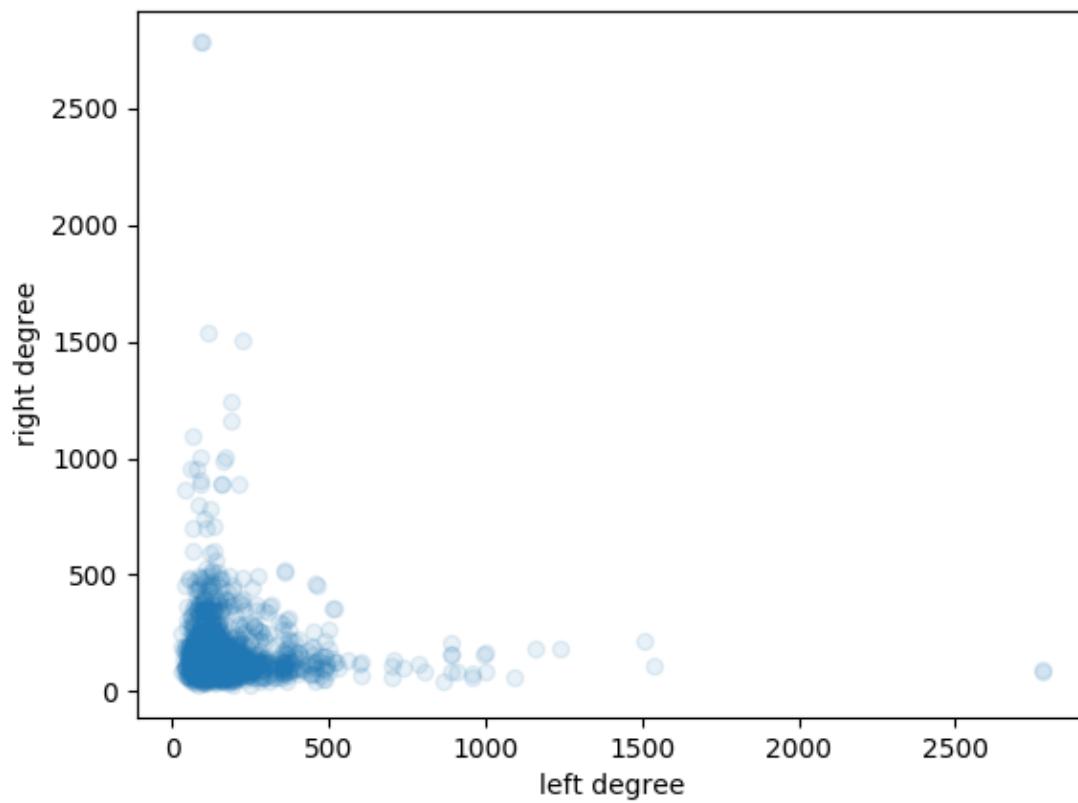
## BA network

assortativity plot, coef = -0.06,  
time: 0.0009



## modified BA network

assortativity plot, coef = -0.05,  
time: 0.0026

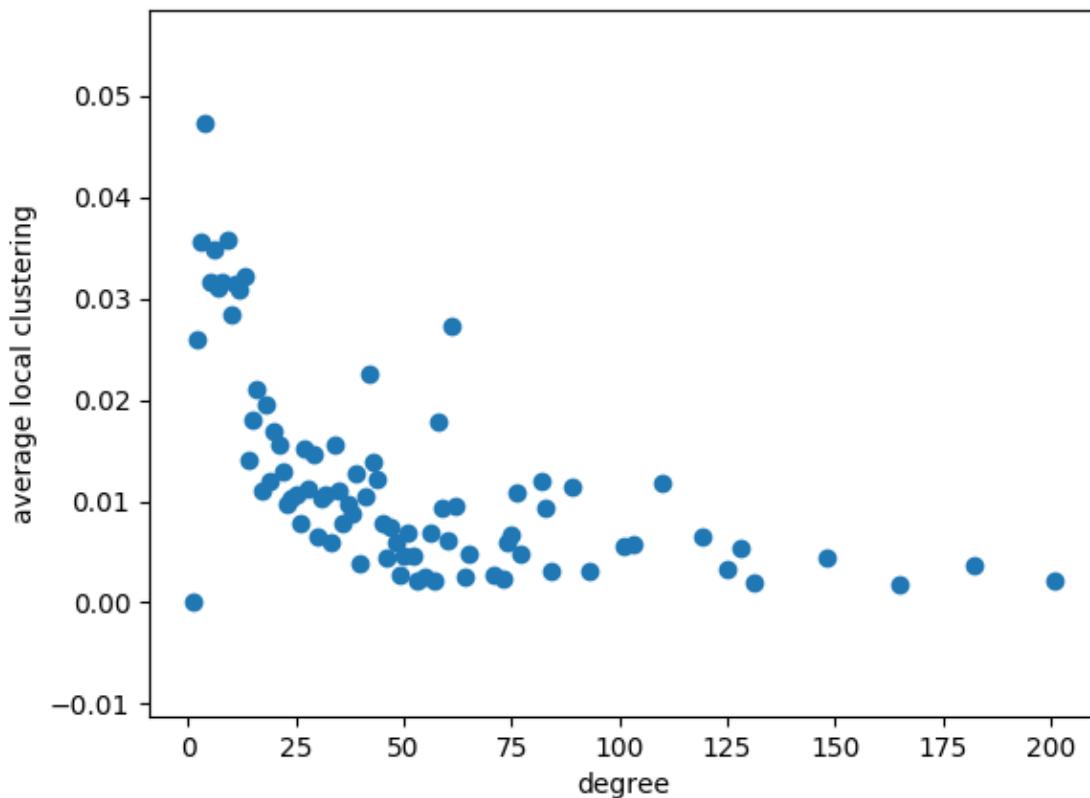


## degreecluster

### collaboration

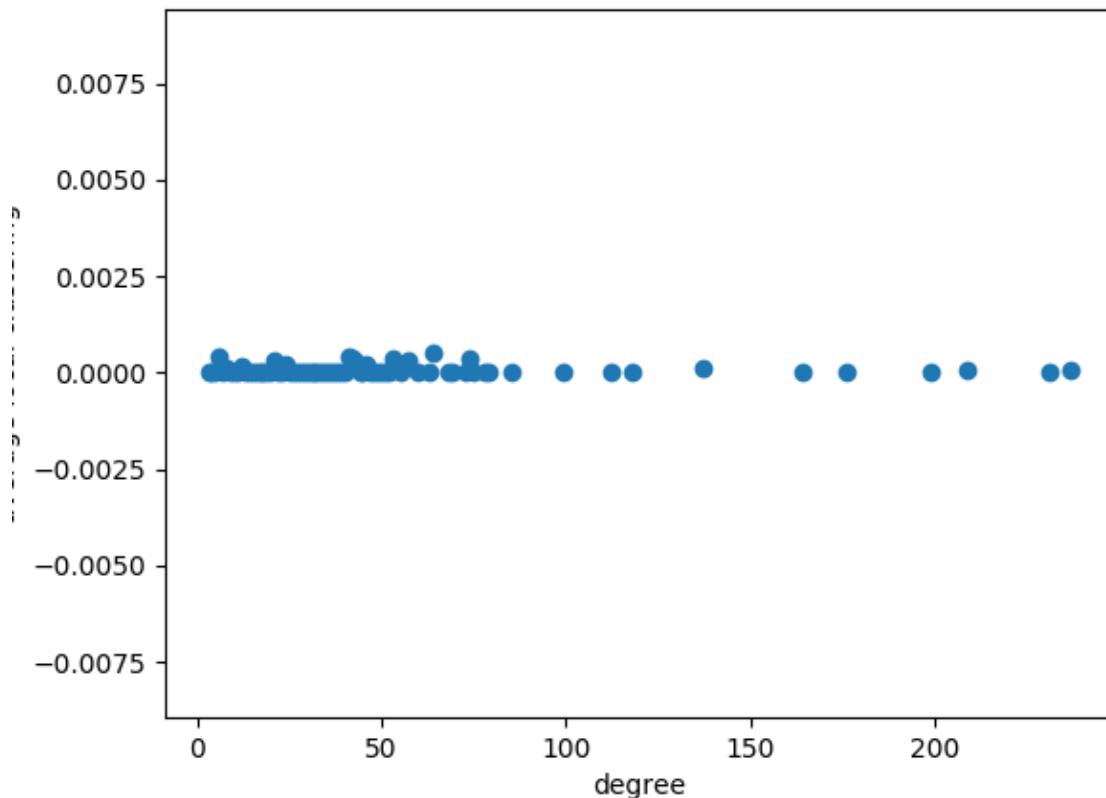
### original network

degree clustering plot,  
time: 0.0042

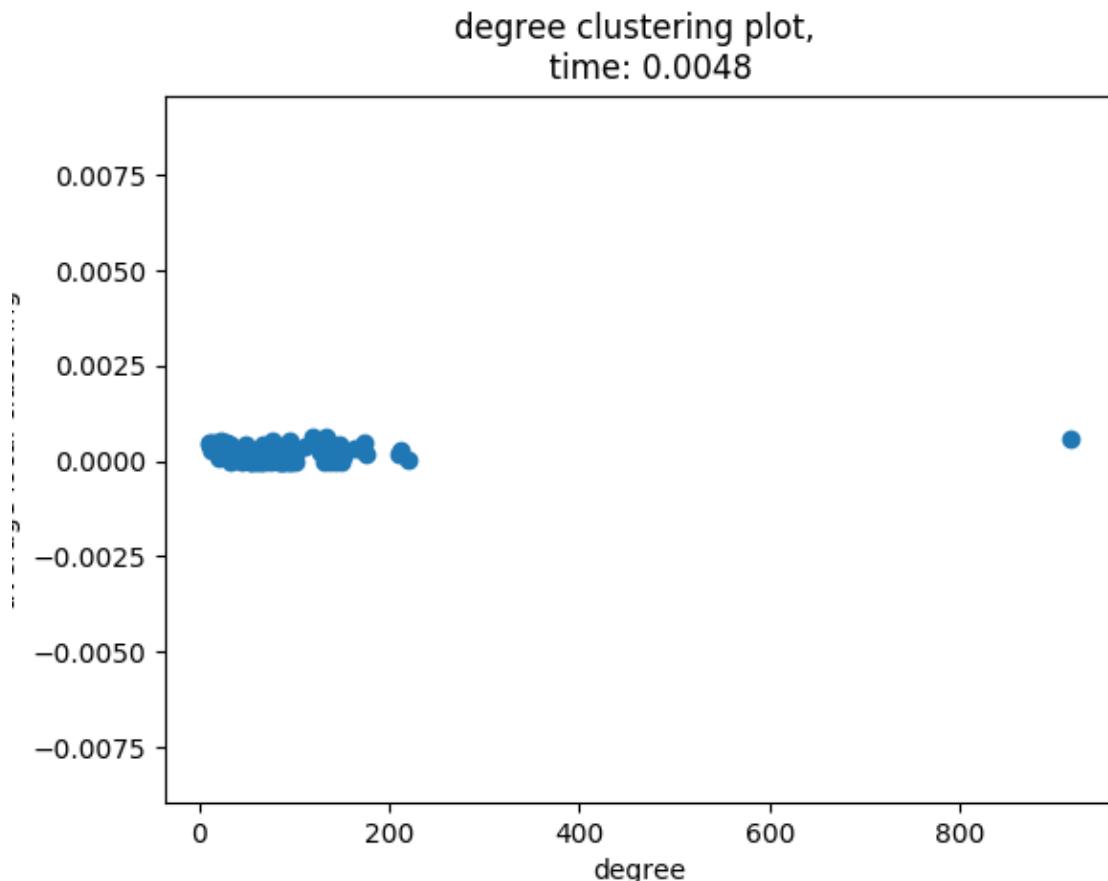


## BA network

degree clustering plot,  
time: 0.0041

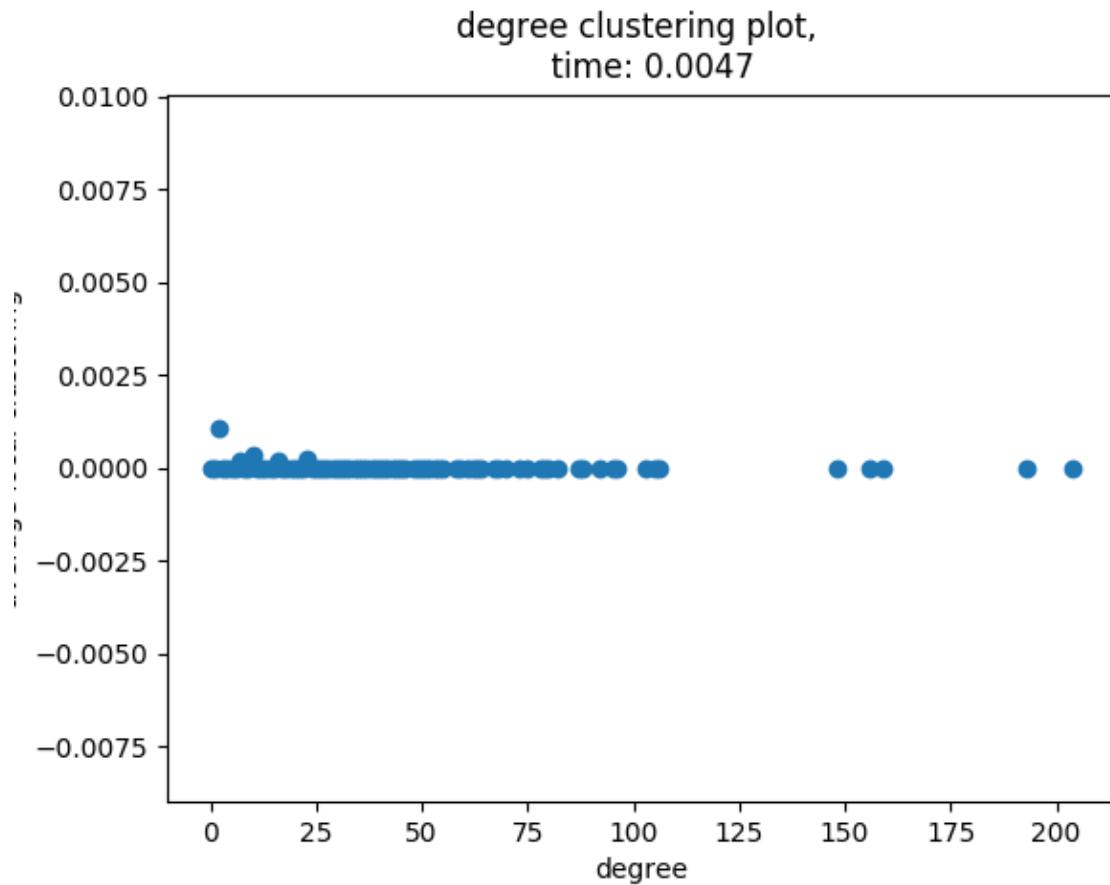


## modified BA network

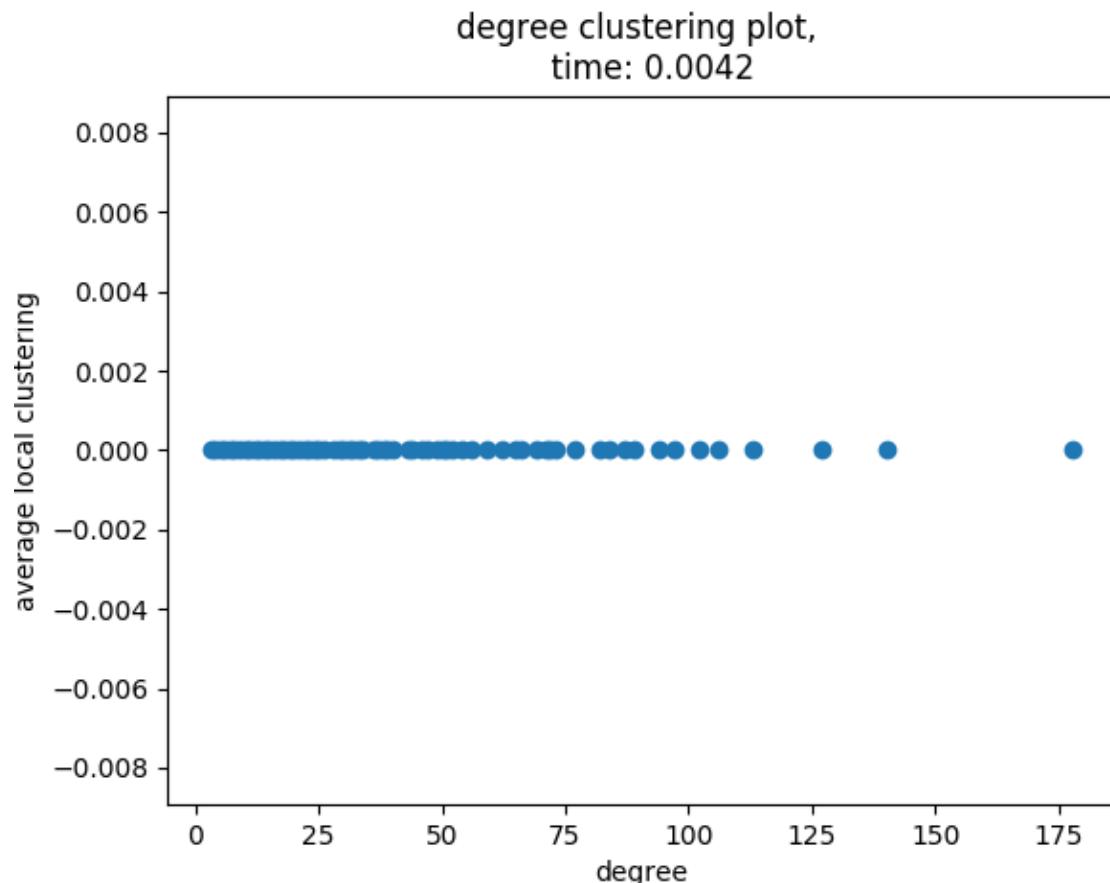


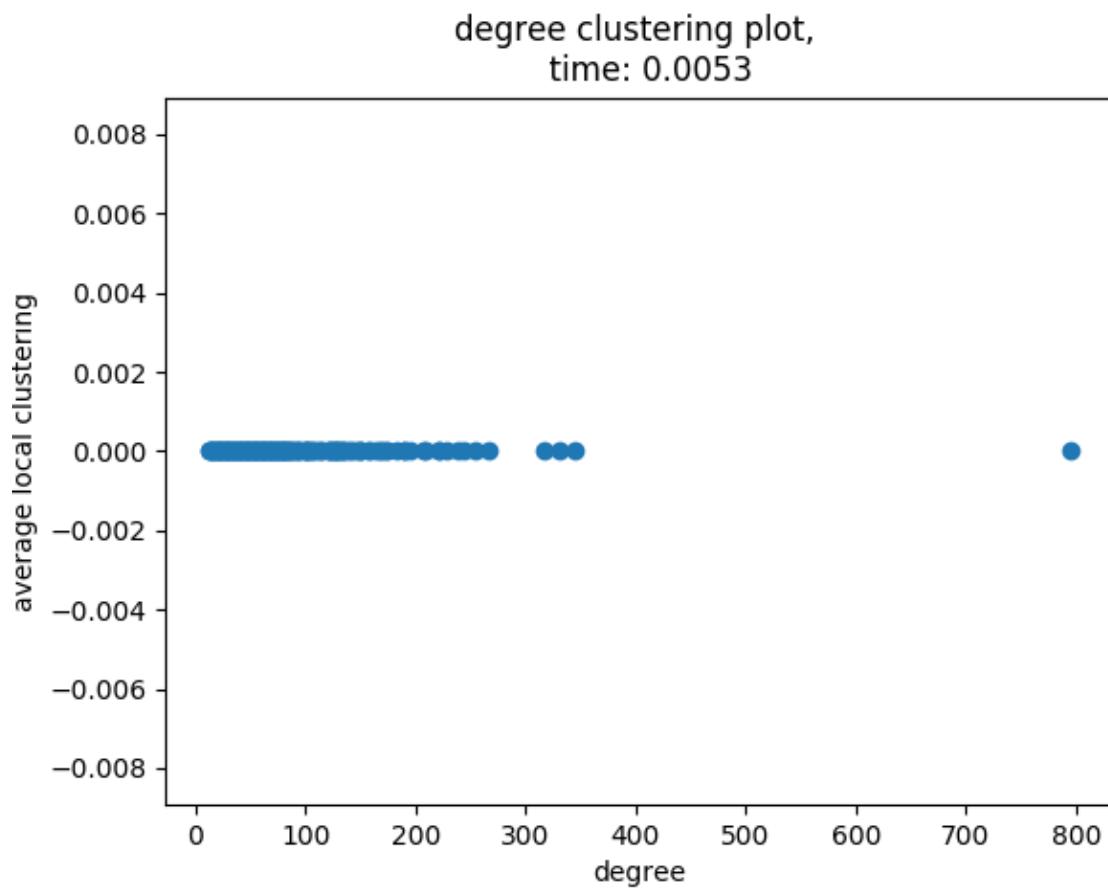
## citation

### original network

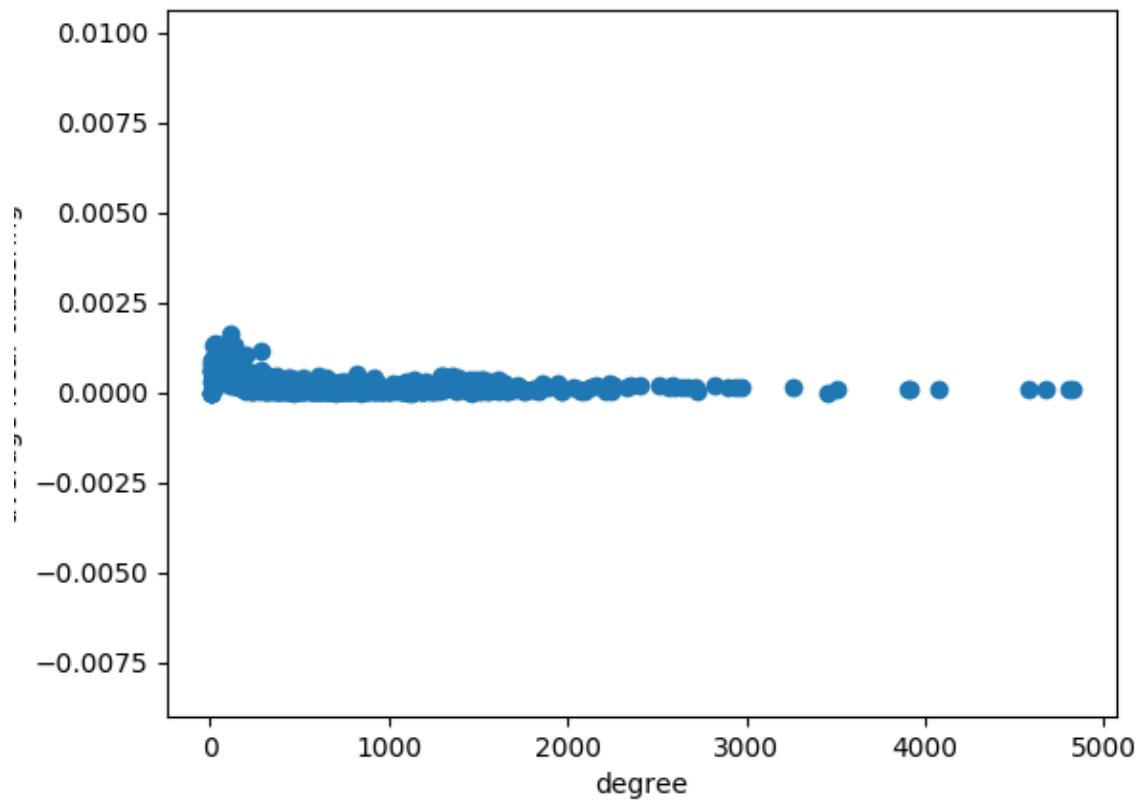


## BA network



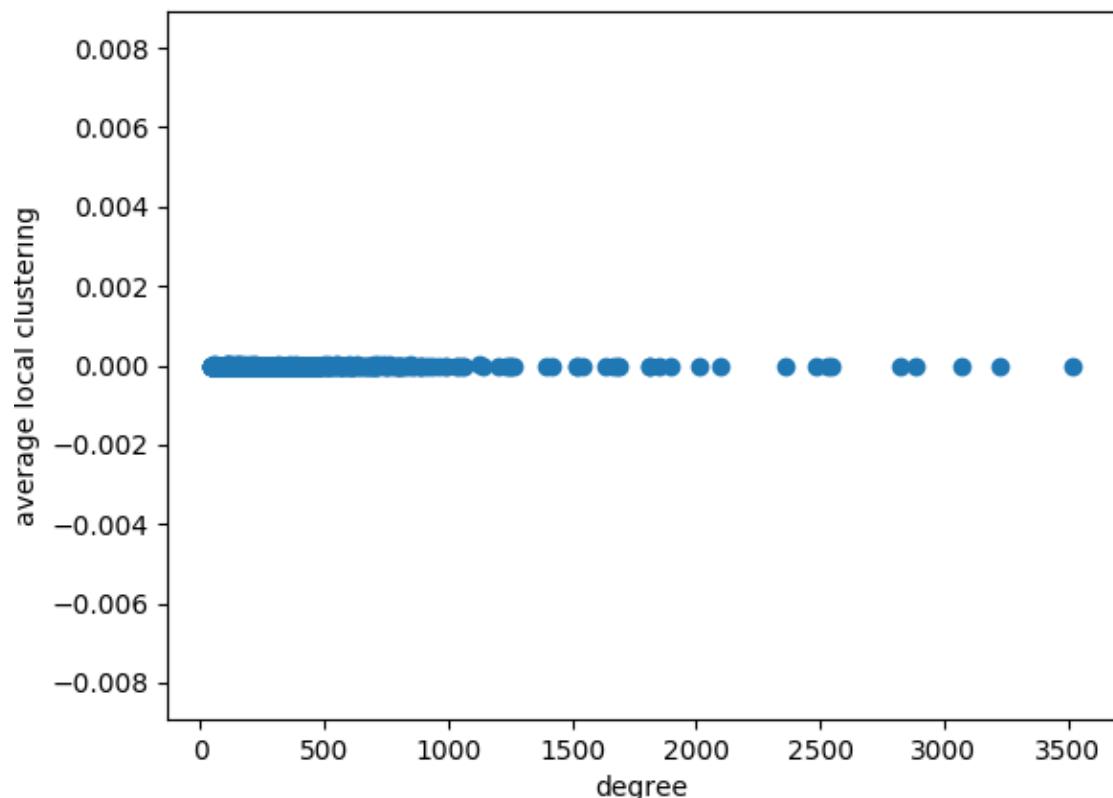
**modified BA network****actor****original network**

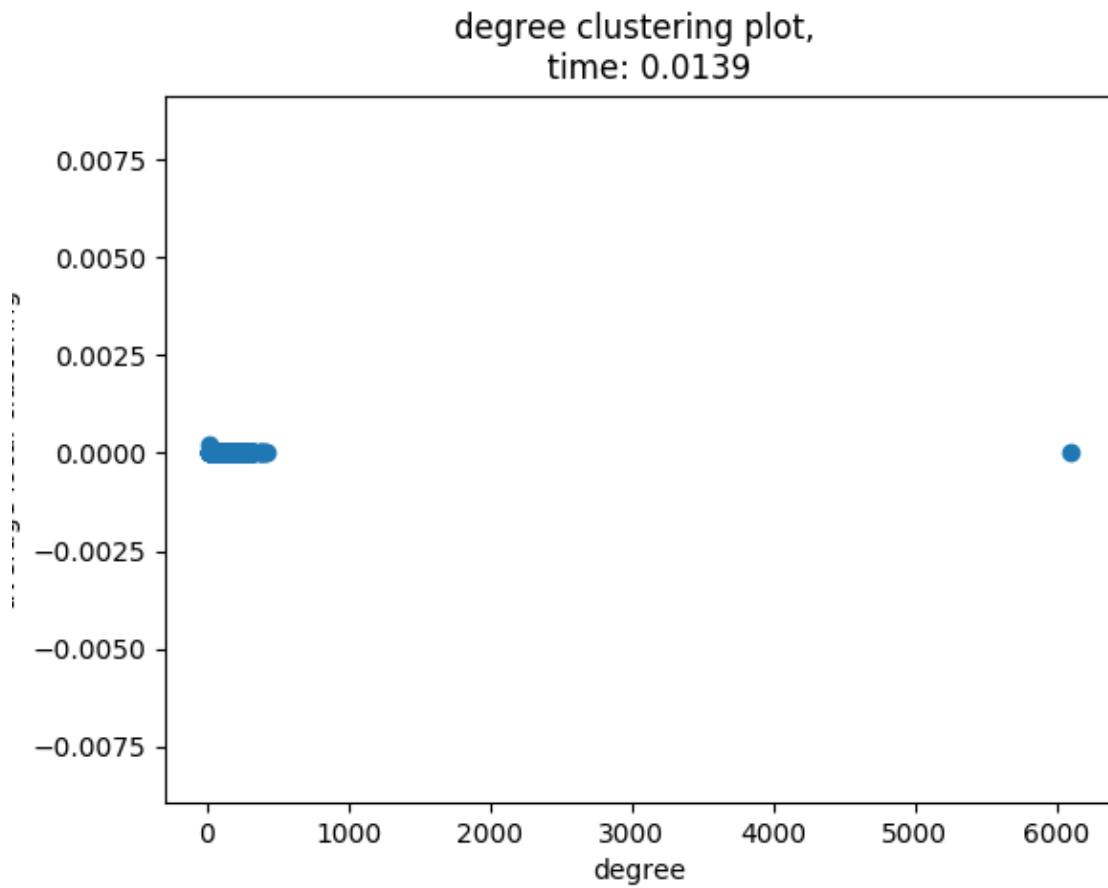
degree clustering plot,  
time: 0.0254

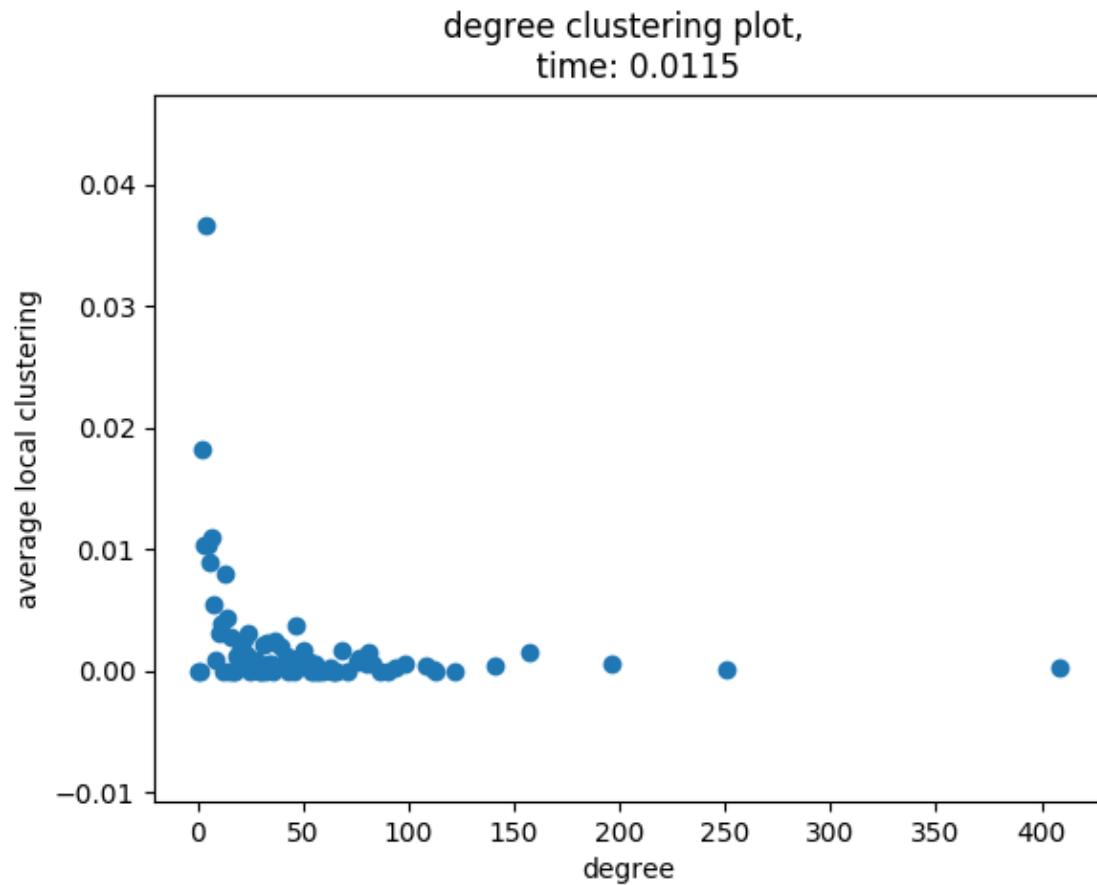


## BA network

degree clustering plot,  
time: 0.0183



**modified BA network****email****original network**



**BA network**

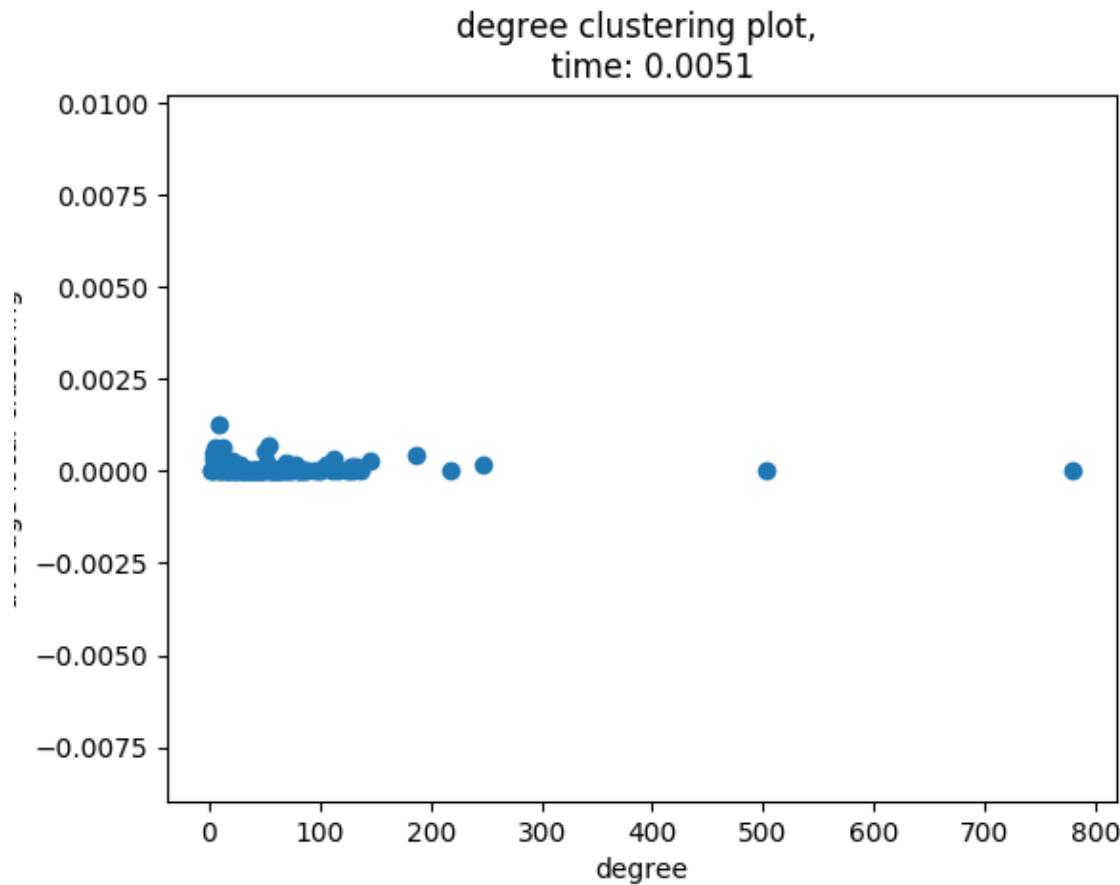


**modified BA network**

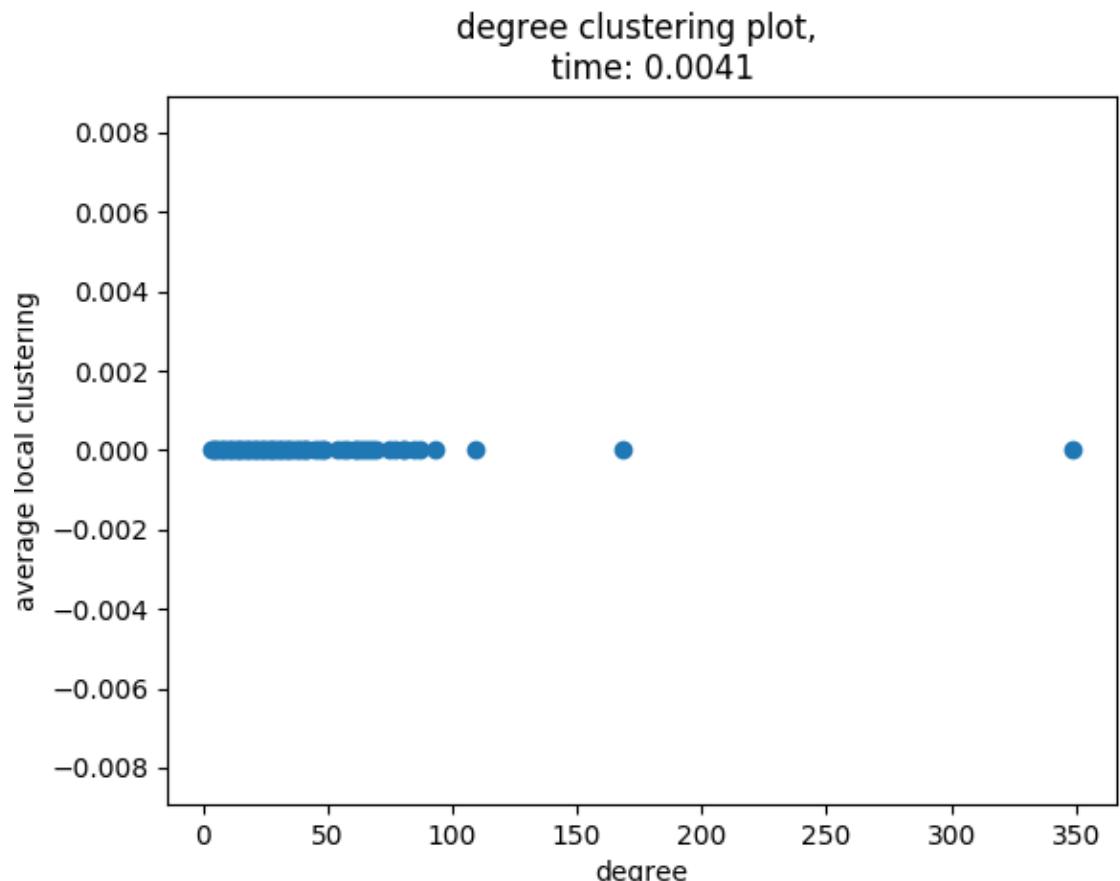


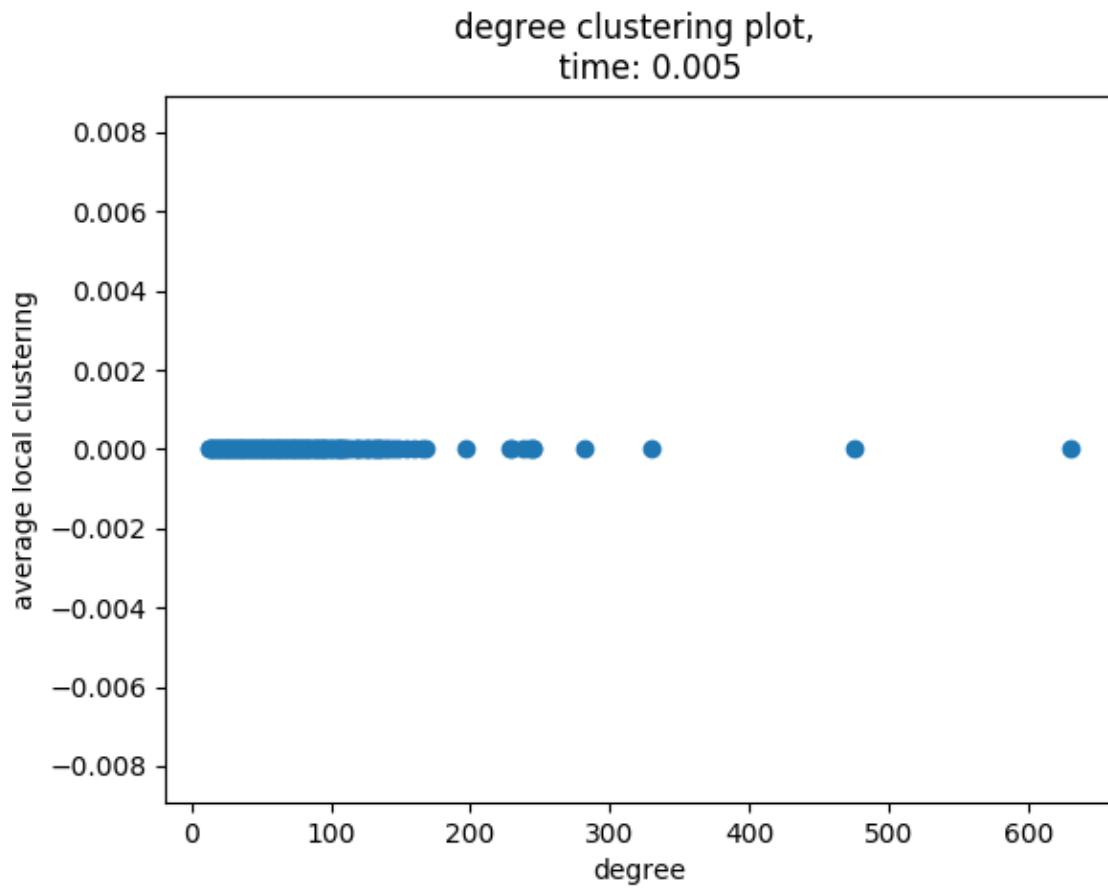
**internet**

**original network**

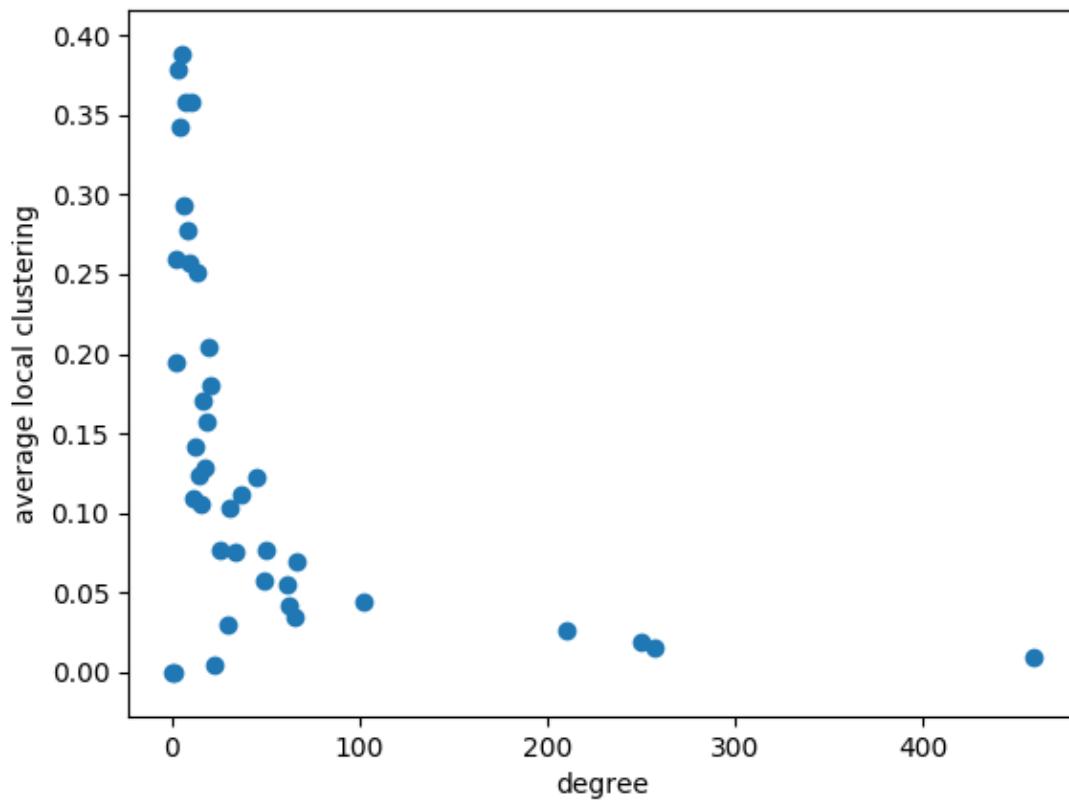


## BA network



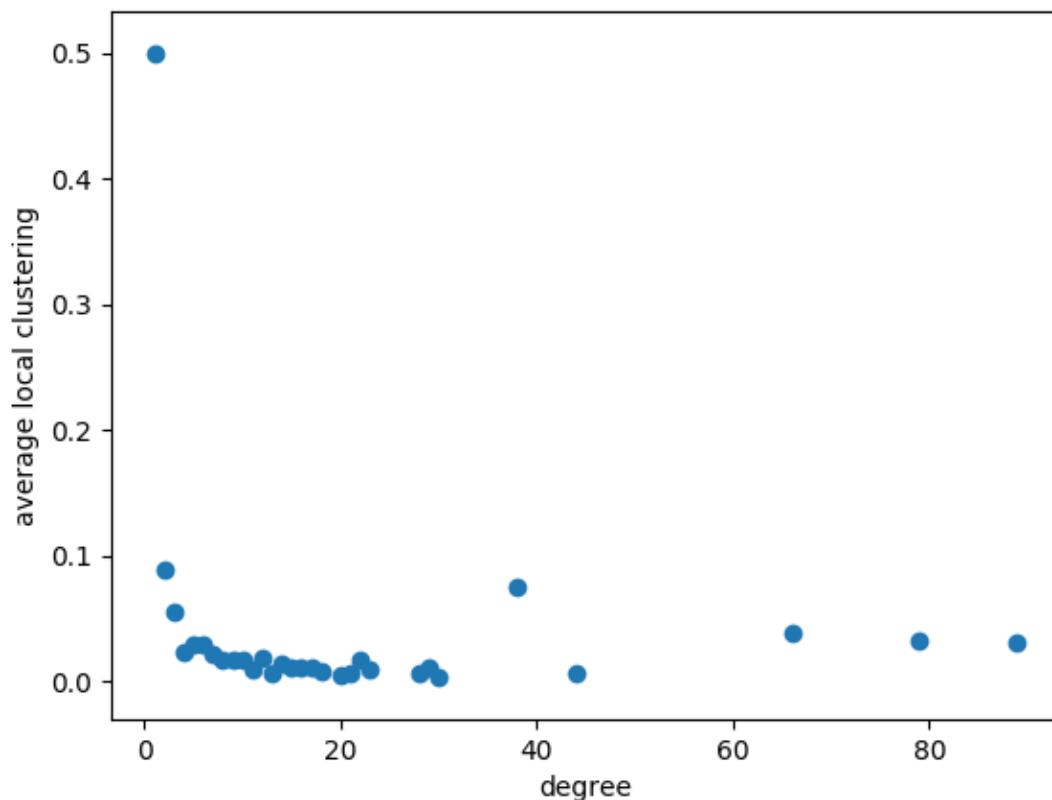
**modified BA network****metabolic****original network**

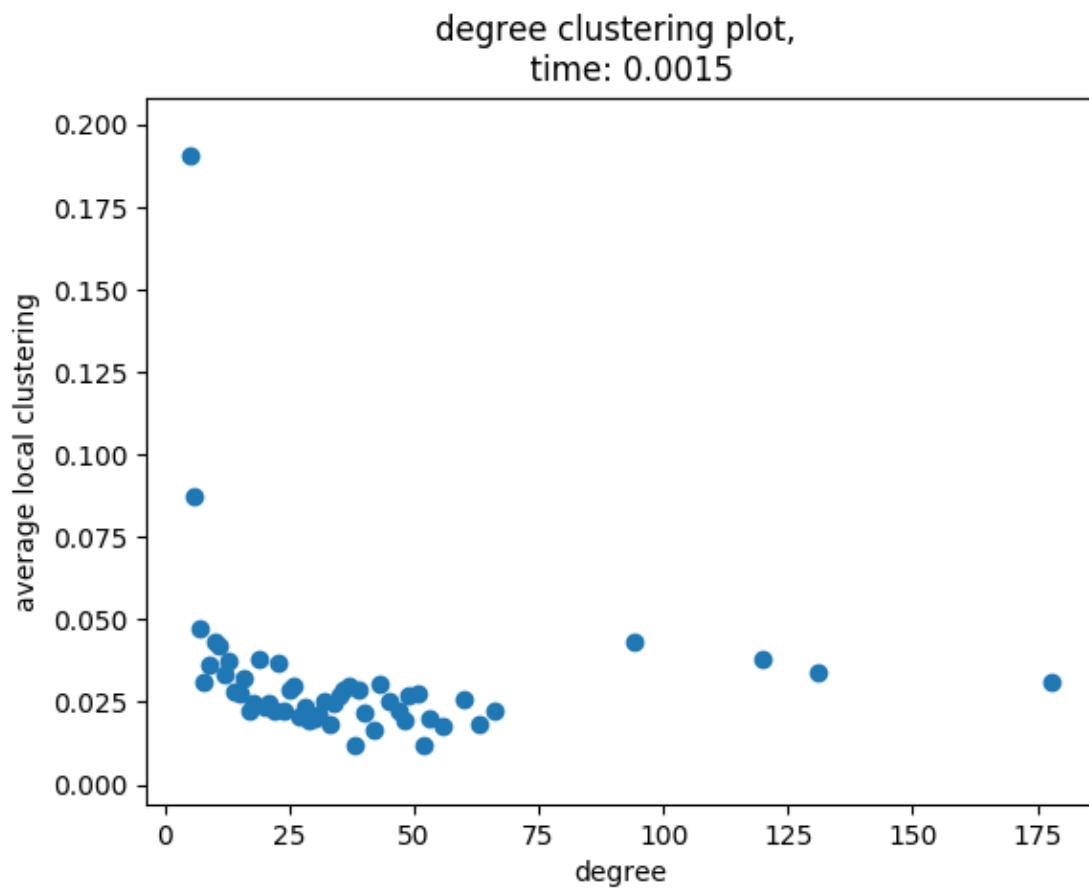
degree clustering plot,  
time: 0.0021



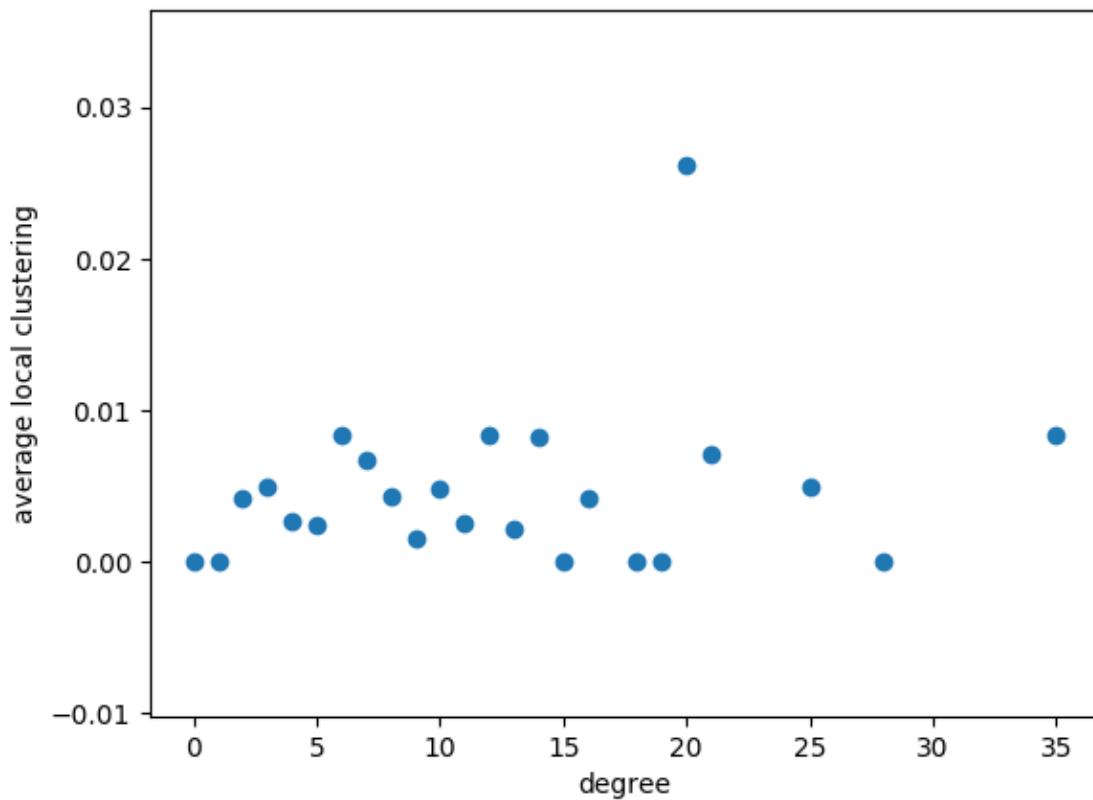
## BA network

degree clustering plot,  
time: 0.0013



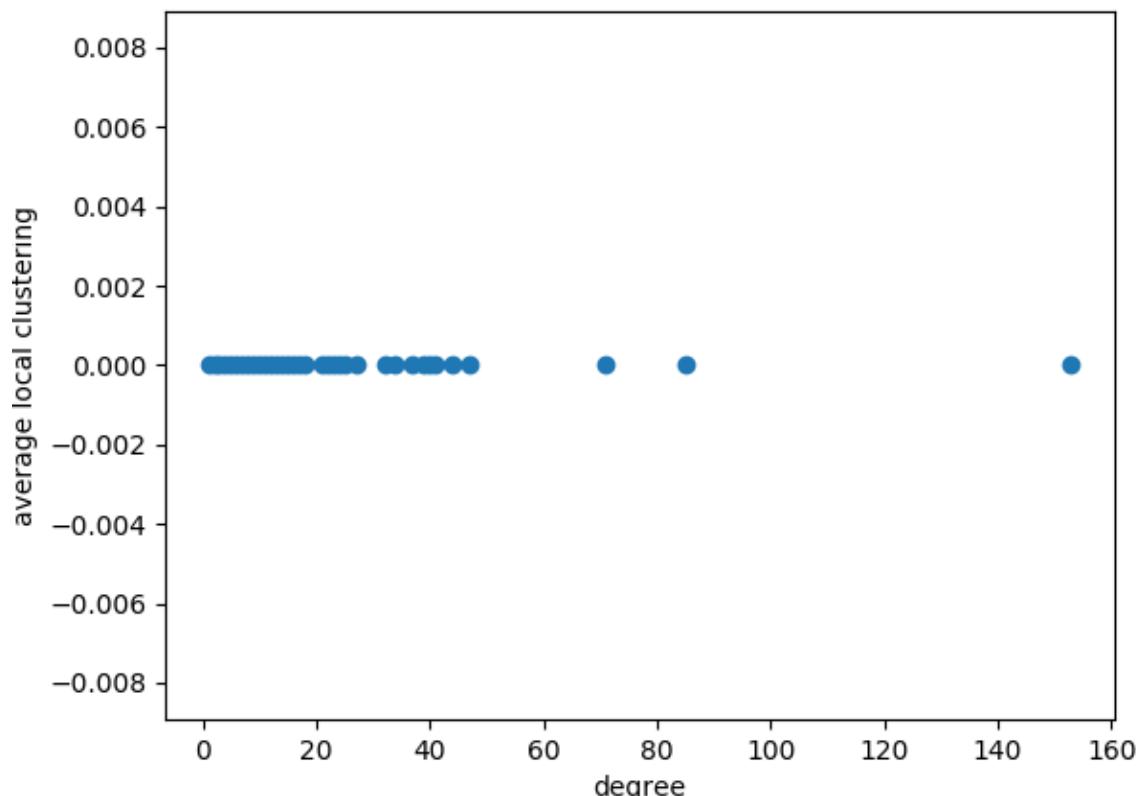
**modified BA network****phonecalls****original network**

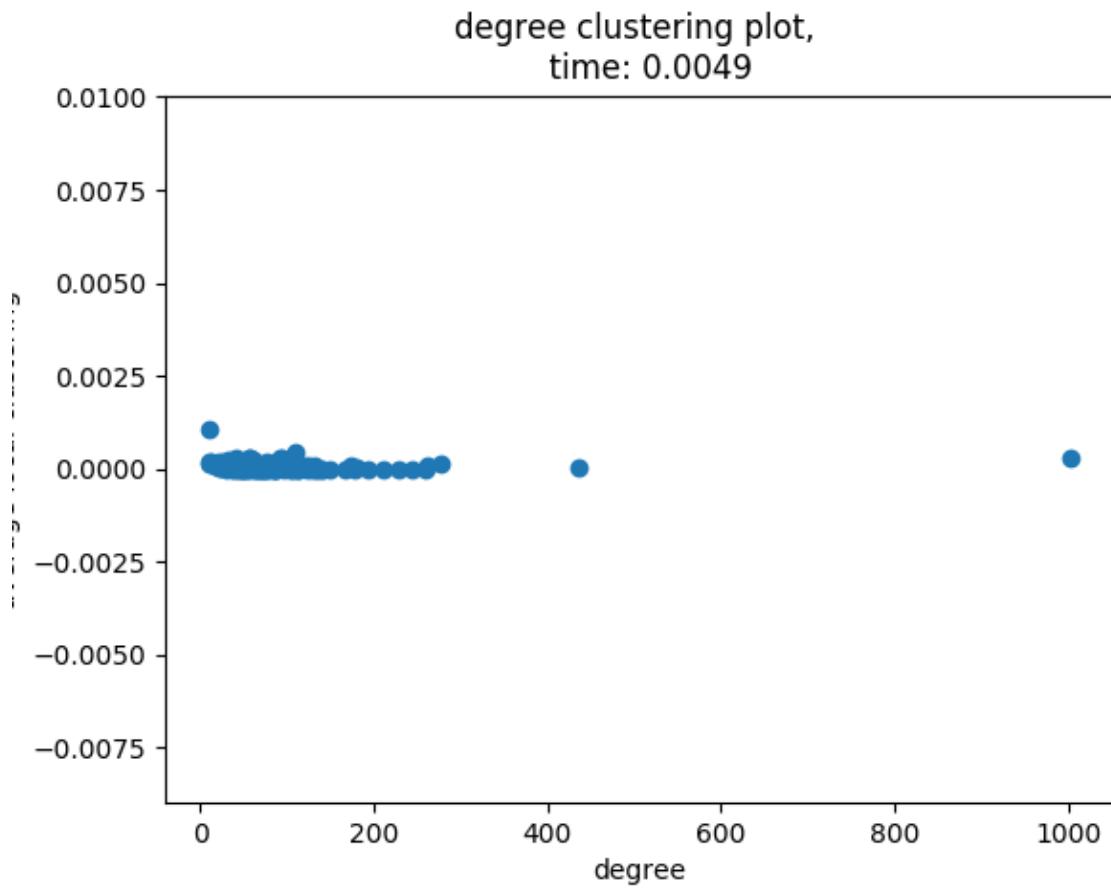
degree clustering plot,  
time: 0.0038



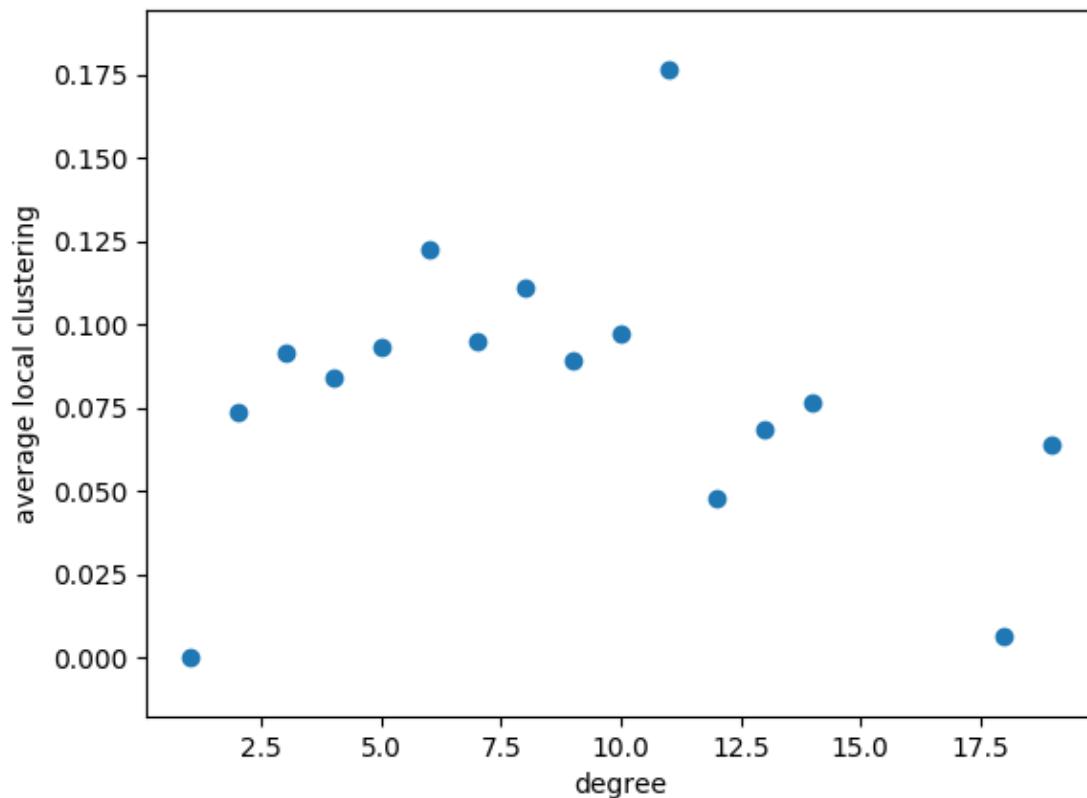
## BA network

degree clustering plot,  
time: 0.0053



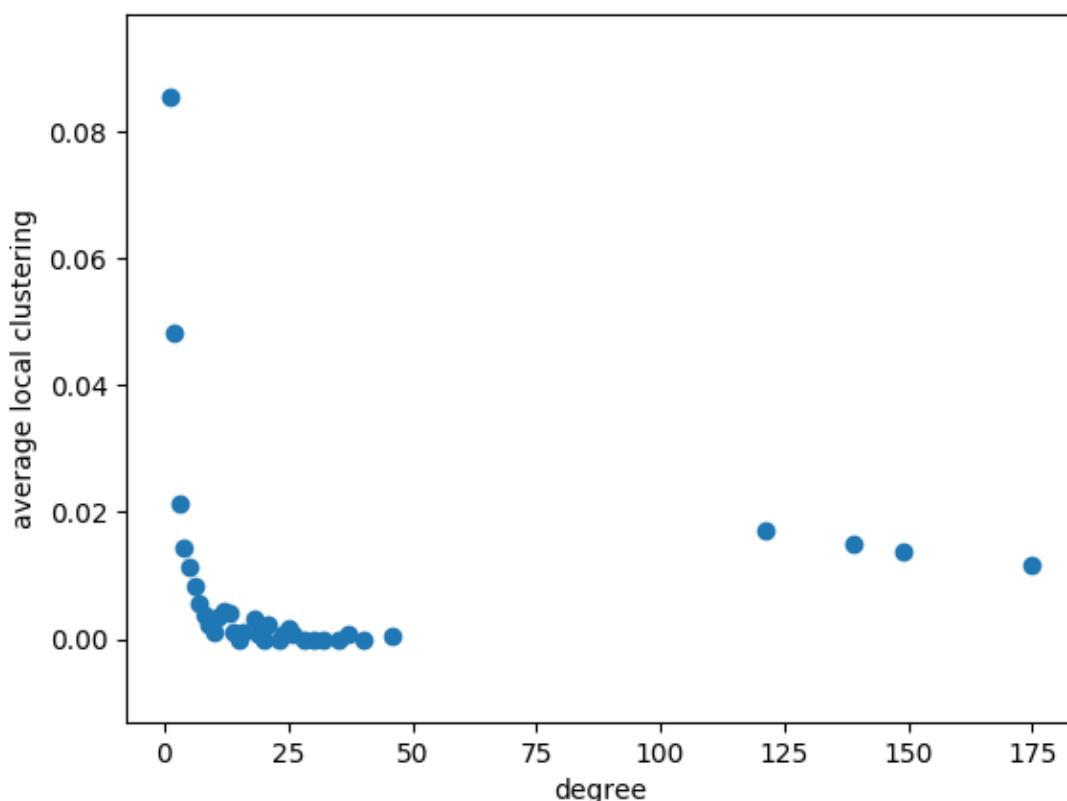
**modified BA network****powergrid****original network**

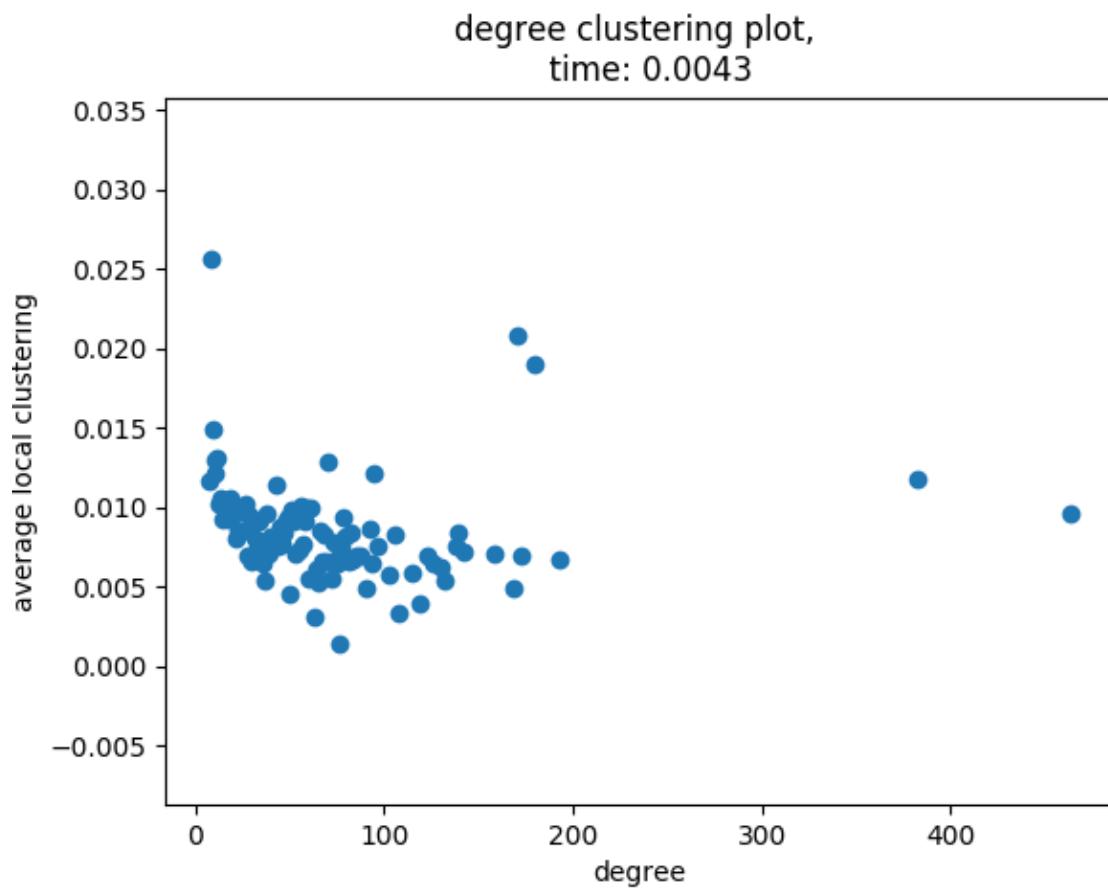
degree clustering plot,  
time: 0.0035



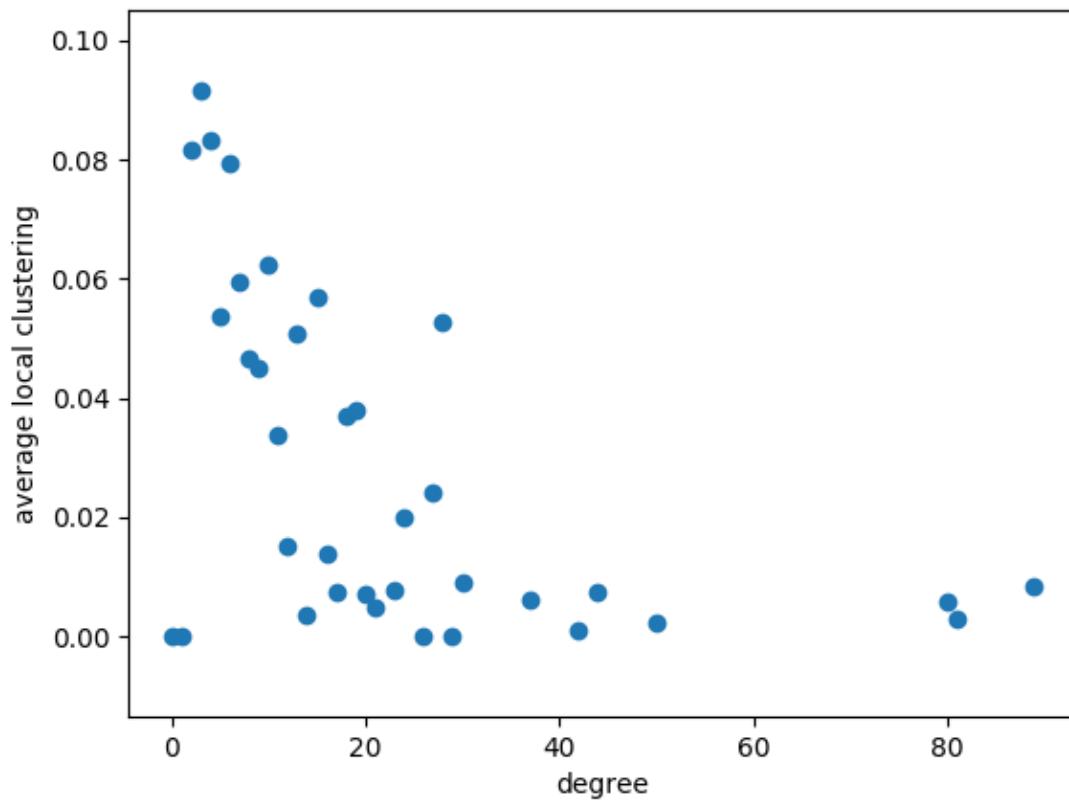
## BA network

degree clustering plot,  
time: 0.0072



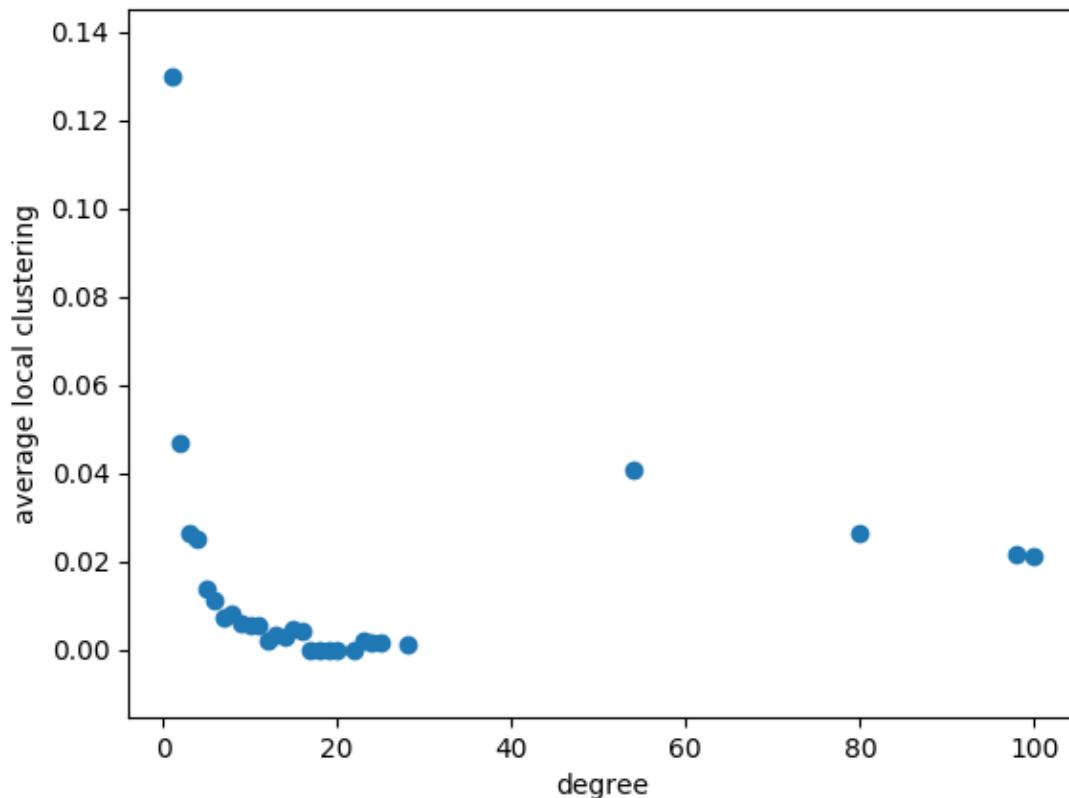
**modified BA network****protein****original network**

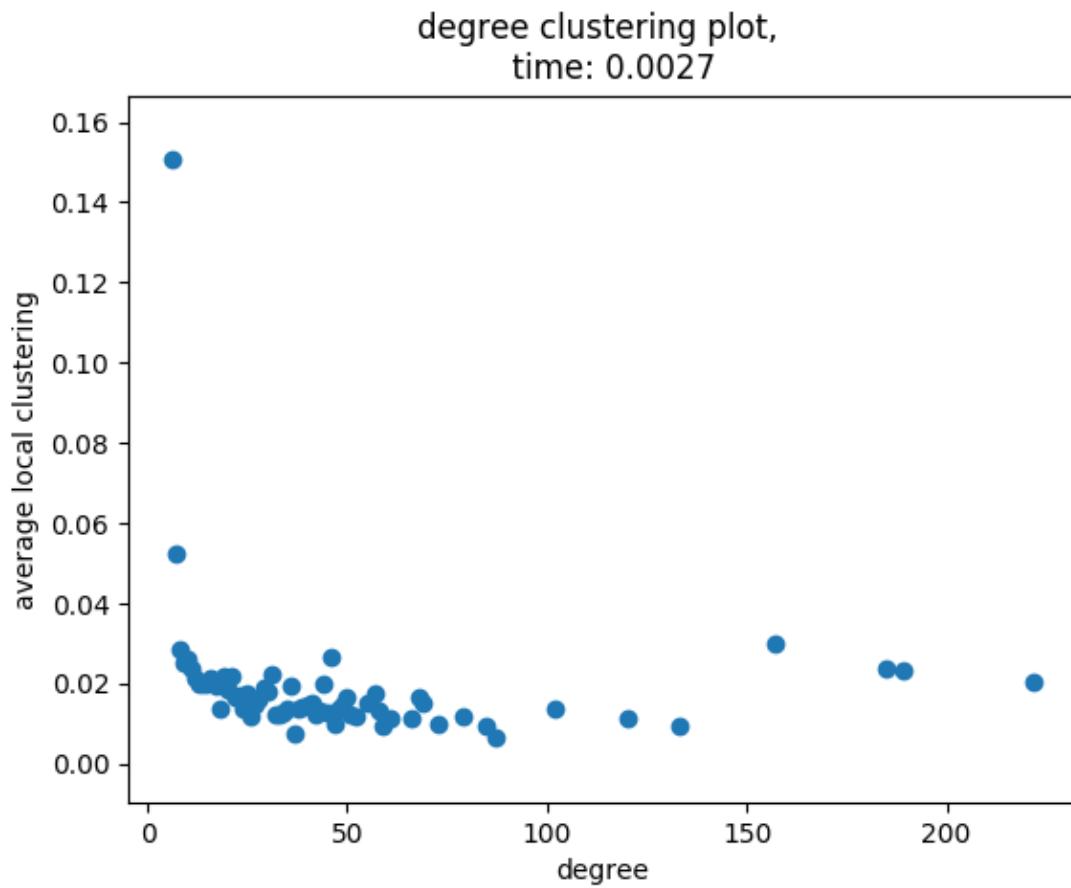
degree clustering plot,  
time: 0.0022



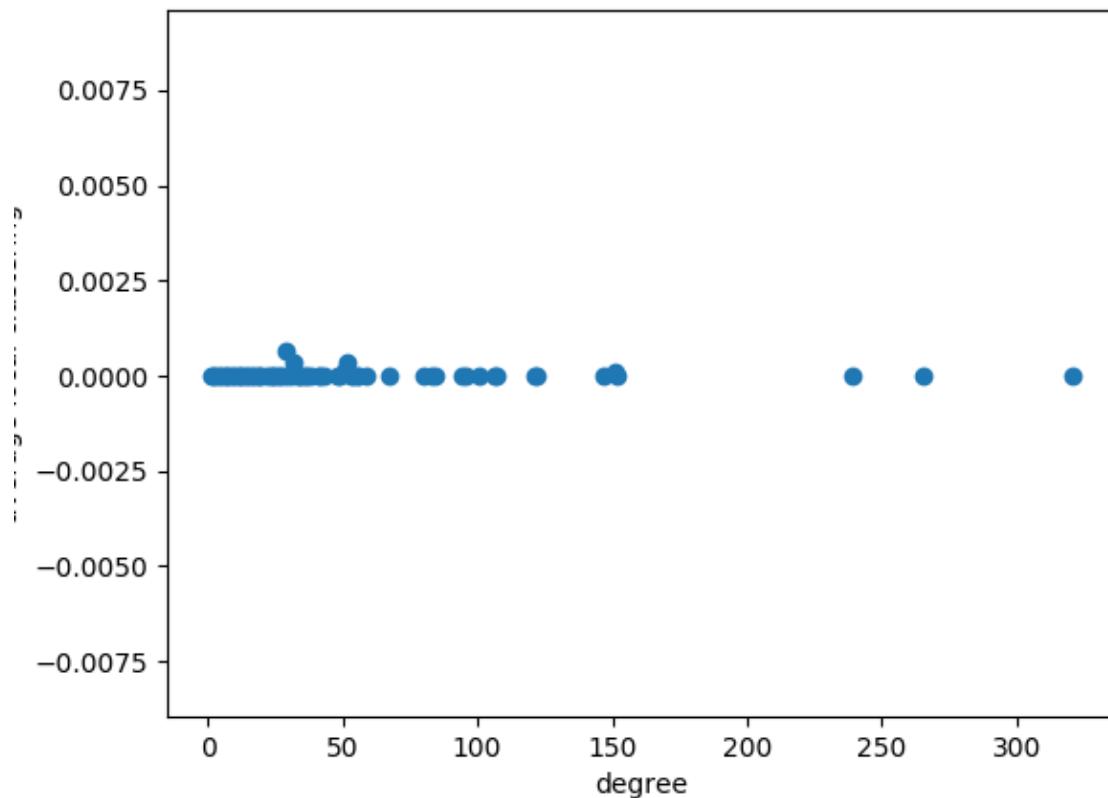
## BA network

degree clustering plot,  
time: 0.0028



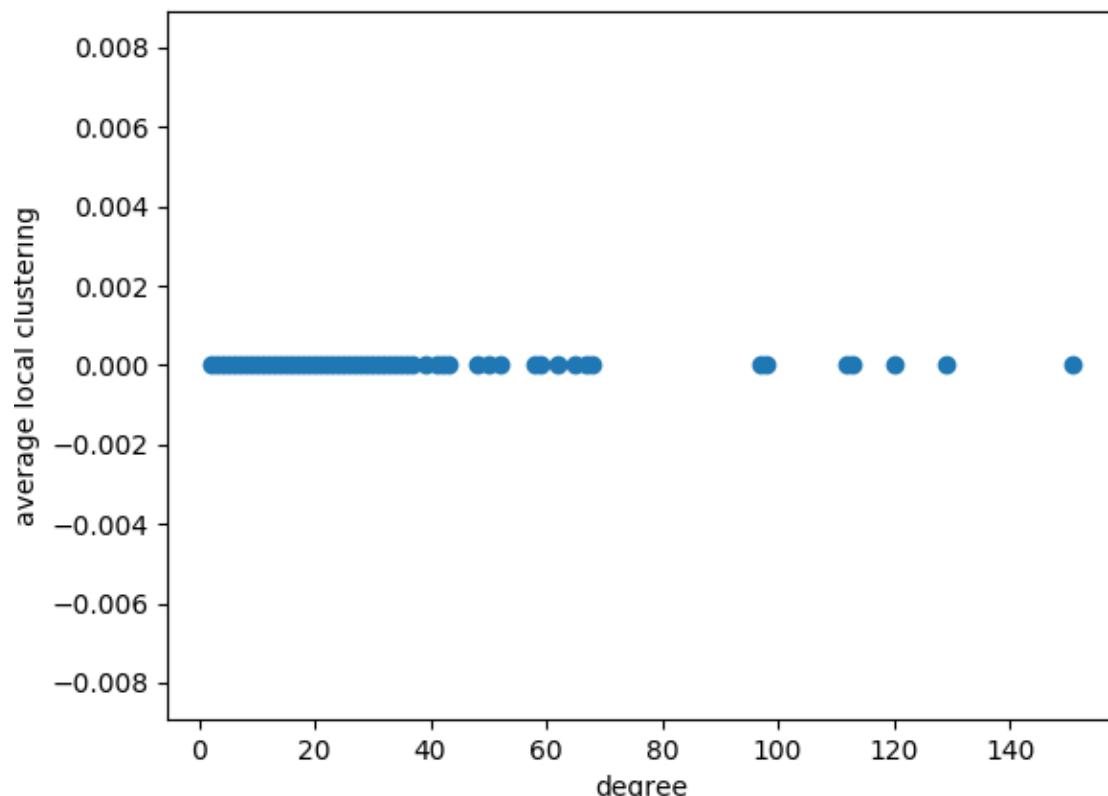
**modified BA network****WWW****original network**

degree clustering plot,  
time: 0.0114

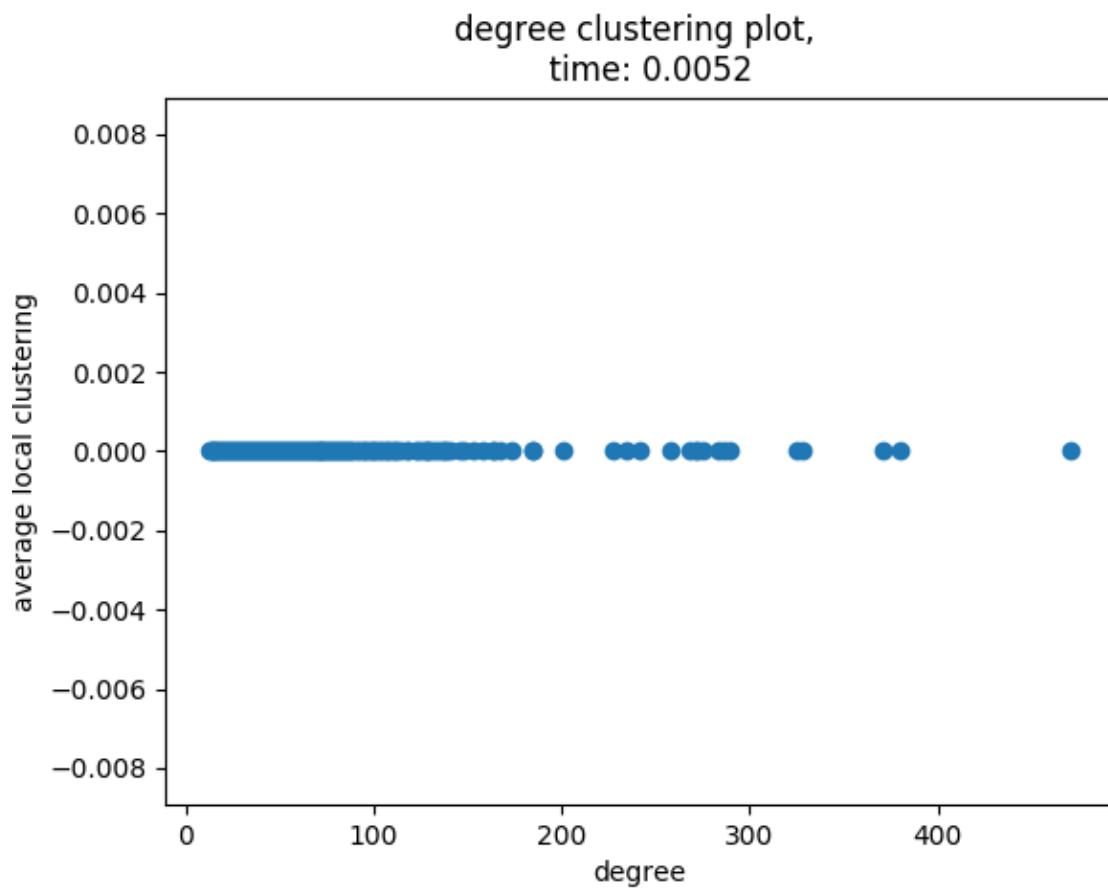


## BA network

degree clustering plot,  
time: 0.0072



## modified BA network



In [ ]: