Bo Zhang
01063214

1 D3 graphing

Use D3 to visualize your Twitter followers. Use my twitter account if you do not have >= 50 followers.

Algorithm:

1. Use the script from http://stackoverflow.com/questions/31000178/how-to-get-large-list-of-followers-tweepy to download the information of all followers, and save them to "twitterFollowers.csv".
2. Open "twitterFollowers.csv", read all kinds of the information to different lists.
3. Use the package "gender-detector"(https://pypi.python.org/pypi/gender-detector/0.1.0) to guess the gender and save it to another list.
4. Select 100 from the followers whose gender could be guessed, and mark them. Mark the main account as well.
5. Save all the followers as nodes to "graph.csv".
6. Use the package "tweepy"(http://docs.tweepy.org/en/latest/api.html) to get the friendship within the 100 followers. If anyone of the 2 followers followed the other, save it as an edge to "graph.csv".
7. Open "graph.csv", read all the information and add the edges between the main account and the followers.
8. Use the package "networkx"(https://networkx.readthedocs.io/en/stable/) to save the graph to "graph.json".
9. Use the D3 to visualize the graph.

References:

1. https://bl.ocks.org/puzzler10/4efcb280a23c2f9b824879771ae41592
2. http://www.puzzlr.org/force-directed-graph-using-node-and-link-attributes/
3. http://stackoverflow.com/questions/18164230/add-text-label-to-d3-node-in-force-directed-graph-and-resize-on-hover

Source code:

Listing 1: The content of downloadTwitterFollowers.py

```
import tweepy
import time

#Variables that contains the user credentials to access Twitter API
access_token = "825062339653271552-q2y3e35bUt1pKxdbqZ9leWIcgIT1mvt"
access_token_secret = "GIyuMRZB2xoIFVVJzJBRCt3kFWZCJh36rHY1T125GcVN0"
consumer_key = "EVKHzzDy0B3mtbvN426yrEZOM"
consumer_secret = "jAydpzL5jYnQGcUkxuwG1DeYGYgF6hu3zzlH9Vx1sG0iHbPKPT"

#This handles Twitter authetification
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)

users = tweepy.Cursor(api.followers, screen_name='phonedude_mln').items()
```

```
f = open("twitterFollowers.csv", "w", encoding='utf-8')

#Print the follower# of every follower into the file 1 by 1
index = 1
while True:
    try:
        user = next(users)
        f.write(user.screen_name + '\n')
        f.write(user.name + '\n')
        f.write('{}'.format(user.followers_count) + '\n')
        index = index + 1
    except tweepy.TweepError:
        time.sleep(60)
    except StopIteration:
        break

f.write('phonedude_mln\n')
f.write('Michael L. Nelson\n')
f.write('{}'.format(index-1)+'\n')
f.close()
```

Listing 2: The content of FollowersCheck.py

```
import tweepy
import time
from gender_detector import gender_detector
detector = gender_detector.GenderDetector('us')

f = open("twitterFollowers.csv", "r", encoding='utf-8')
lines = f.readlines()
f.close()

usernames=[]
names=[]
firstNames=[]
followerNumbers=[]
for i in range(int(len(lines)/3)):
    usernames.append(lines[i*3].strip())
    names.append(lines[i*3+1].strip())
    firstNames.append(names[i].split()[0])
    followerNumbers.append(lines[i*3+2].strip())

genders=[]
for firstName in firstNames:
    try:
```

```
        genders.append(detector.guess(firstName))
    except:
        genders.append('unknown')

usernamesCheck=[]
count = 0
for i in range(len(genders)):
    if genders[i]!='unknown' and count<100:
        usernamesCheck.append(1)
        count = count + 1
    else:
        usernamesCheck.append(0)

usernamesCheck[len(genders)-1]=1

f = open("graph.csv", "w", encoding='utf-8')
f.write('nodes:\n')
for i in range(len(usernames)):
    f.write(usernames[i]+'\n'+genders[i]+'\n{}'.format(usernamesCheck[i])+'\n')
f.write('edges:\n')
f.close()

#Variables that contains the user credentials to access Twitter API
access_token = "825062339653271552-q2y3e35bUt1pKxdbqZ9leWIcgIT1mvt"
access_token_secret = "GIyuMRZB2xoIFVVJzJBRCt3kFWZCJh36rHY1T125GcVN0"
consumer_key = "EVKHzzDy0B3mtbvN426yrEZOM"
consumer_secret = "jAydpzL5jYnQGcUkxuwG1DeYGYgF6hu3zzlH9Vx1sG0iHbPKPT"

#This handles Twitter authetification
auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)

friendships=[]
for i in range(len(usernames)-2):
    if usernamesCheck[i]==1:
        for j in range(i+1,len(usernames)-1):
            if usernamesCheck[j]==1:
                print ('{}'.format(i)+'\t{}'.format(j))
                while True:
                    try:
                        friendship = api.show_friendship(source_screen_name=usernames[i],target_sc
                        if friendship[0].following==True or friendship[0].followed_by==True or fri
                            f = open("graph.csv", "a", encoding='utf-8')
                            f.write('{}'.format(i)+'\n{}'.format(j)+'\n')
                            f.close()
```

```
            break
        except Exception as err:
            print (err)
            if str(err)=='Not authorized.':
                break
            else:
                time.sleep(60)
```

Listing 3: The content of graphWrite-json.py

```
import networkx as nx
from networkx.readwrite import json_graph
import json

g = nx.Graph()

f = open("graph.csv", "r", encoding='utf-8')
f.readline()

usernames=[]
genders=[]
usernamesCheck=[]
while True:
    line = f.readline().strip()
    if line!='edges:':
        usernames.append(line)
        line = f.readline().strip()
        genders.append(line)
        line = f.readline().strip()
        usernamesCheck.append(int(line))
    else:
        break

lines=f.readlines()
f.close()

for i in range(len(usernames)):
    g.add_node(i, username=usernames[i], gender=genders[i], usernamesCheck=usernamesCheck[i])

for i in range(len(usernames)-1):
    g.add_edge(len(usernames)-1,i)

for i in range(int(len(lines)/2)):
    g.add_edge(int(lines[i*2].strip()),int(lines[i*2+1].strip()))

data = json_graph.node_link_data(g)
```

```
s = json.dumps(data)

f = open('graph.json', 'w')
f.write(s)
f.close()
```

Listing 4: The content of Q1.html

```
<!DOCTYPE html>
<meta charset="utf-8">
<style>

.links line {
  stroke: #999;
  stroke-opacity: 0.6;
}

.node text {
  pointer-events: none;
  font: 15px sans-serif;
}

.nodes circle {
  stroke: #fff;
  stroke-width: 1.5px;
}

</style>
<svg width="1366" height="768"></svg>
<script src="https://d3js.org/d3.v4.min.js"></script>
<script>

//create somewhere to put the force directed graph
var svg = d3.select("svg"),
  width = +svg.attr("width"),
  height = +svg.attr("height");

d3.json("graph.json", function(graph){
  var simulation = d3.forceSimulation()
    .nodes(graph.nodes);

  simulation
    .force("charge_force", d3.forceManyBody())
    .force("center_force", d3.forceCenter(width / 2, height / 2));

  var node = svg.selectAll(".node")
```

```
      .data(graph.nodes)
      .enter().append("g")
      .attr("class", "node")

  node.append("text")
      .attr("dx", 12)
      .attr("dy", ".35em")
      .text(function(d) { return d.username });

  node.append("title").text(function(d) { return d.username; });

  node.append("circle")
        .attr("r", 5)
        .attr("fill", "red");

  simulation.on("tick", tickActions );

  var link_force =  d3.forceLink(graph.links).distance(100)

  simulation.force("links",link_force)

  var link = svg.append("g")
    .attr("class", "links")
    .selectAll("line")
    .data(graph.links)
    .enter().append("line")
    .attr("stroke-width", 2);

  function tickActions() {
    link
      .attr("x1", function(d) { return d.source.x; })
      .attr("y1", function(d) { return d.source.y; })
      .attr("x2", function(d) { return d.target.x; })
      .attr("y2", function(d) { return d.target.y; });

    node.attr("transform", function(d) { return "translate(" + d.x + "," + d.y + ")"; });
  }
});
</script>
```

Results: https://cdn.rawgit.com/zhangboroy/cs532-s17/b56ae831/assg06_submission/Q1.html

2 Gender homophily in your Twitter graph

Take the Twitter graph you generated in question #1 and test for male-female homophily. For the purposes of this question you can consider the graph as undirected. Use the twitter name and programatically determine if the user is male or female.

Create a table of Twitter users and their likely gender. List any accounts that can't be determined and remove them from the graph. Does your Twitter graph exhibit gender homophily?

Algorithm:

1. Open "graph.csv", read all the information and save it to an output list. Sort the output list with gender and save it to "genderTable.txt".
2. Select the 101 marked account from the output list and save them to another output list.
3. Compute the total edge number and account number of different genders.
4. Save the new output list to "genderTable100.txt".
5. Compute the number of edges between different genders within the 101 accounts.
6. Print the account number of different genders, Randomly assigned cross-gender edge fraction, total edge number, cross-gender edge number and Actual cross-gender edge fraction.
7. Use the result from step 6 in R to run Exact Binomial Test.

Source code:

Listing 5: The content of homophilyTest.py

```
f = open("graph.csv", "r", encoding='utf-8')
f.readline()

usernames=[]
genders=[]
usernamesCheck=[]

while True:
    line = f.readline().strip()
    if line!='edges:':
        usernames.append(line)
        line = f.readline().strip()
        genders.append(line)
        line = f.readline().strip()
        usernamesCheck.append(int(line))
    else:
        break

lines=f.readlines()
f.close()

genderTable=[]
for i in range(len(usernames)):
    row = [genders[i], usernames[i], usernamesCheck[i]]
    genderTable.append(row)
```

```
genderTable.sort()

f = open("genderTable.txt", "w", encoding='utf-8')
f.write('Username')
for i in range(12):
    f.write('␣')
f.write('Gender\tUsername')
for i in range(12):
    f.write('␣')
f.write('Gender\tUsername')
for i in range(12):
    f.write('␣')
f.write('Gender\n')
for i in range(0,int(len(genderTable)/3)*3,3):
    f.write(genderTable[i][1])
    for j in range(20-len(genderTable[i][1])):
        f.write('␣')
    f.write(genderTable[i][0]+'\t'+genderTable[i+1][1])
    for j in range(20-len(genderTable[i+1][1])):
        f.write('␣')
    f.write(genderTable[i+1][0]+'\t'+genderTable[i+2][1])
    for j in range(20-len(genderTable[i+2][1])):
        f.write('␣')
    f.write(genderTable[i+2][0]+'\n')
if i+3<len(genderTable):
    f.write(genderTable[i+3][1])
    for j in range(20-len(genderTable[i+3][1])):
        f.write('␣')
    f.write(genderTable[i+3][0])
if i+4<len(genderTable):
    f.write('\t'+genderTable[i+4][1])
    for j in range(20-len(genderTable[i+4][1])):
        f.write('␣')
    f.write(genderTable[i+4][0])
f.write('\n')
f.close()

genderTable100=[]
for i in range(len(genderTable)):
    if genderTable[i][2]==1:
        genderTable100.append(genderTable[i])

edges = int(len(lines)/2) + (len(genderTable100) - 1)
edgesCross = 0
```

```
f = open("genderTable100.txt", "w", encoding='utf-8')
f.write('Username')
for i in range(12):
    f.write('␣')
f.write('Gender\tUsername')
for i in range(12):
    f.write('␣')
f.write('Gender\tUsername')
for i in range(12):
    f.write('␣')
f.write('Gender\n')
for i in range(0,int(len(genderTable100)/3)*3,3):
    if genderTable100[i][0]=='female':
        edgesCross = edgesCross + 1
    if genderTable100[i+1][0]=='female':
        edgesCross = edgesCross + 1
    if genderTable100[i+2][0]=='female':
        edgesCross = edgesCross + 1
    f.write(genderTable100[i][1])
    for j in range(20-len(genderTable100[i][1])):
        f.write('␣')
    f.write(genderTable100[i][0]+'\t'+genderTable100[i+1][1])
    for j in range(20-len(genderTable100[i+1][1])):
        f.write('␣')
    f.write(genderTable100[i+1][0]+'\t'+genderTable100[i+2][1])
    for j in range(20-len(genderTable100[i+2][1])):
        f.write('␣')
    f.write(genderTable100[i+2][0]+'\n')
if i+3<len(genderTable100):
    if genderTable100[i+3][0]=='female':
        edgesCross = edgesCross + 1
    f.write(genderTable100[i+3][1])
    for j in range(20-len(genderTable100[i+3][1])):
        f.write('␣')
    f.write(genderTable100[i+3][0])
if i+4<len(genderTable100):
    if genderTable100[i+4][0]=='female':
        edgesCross = edgesCross + 1
    f.write('\t'+genderTable100[i+4][1])
    for j in range(20-len(genderTable100[i+4][1])):
        f.write('␣')
    f.write(genderTable100[i+4][0])
f.write('\n')
f.close()

print ('Males:␣{}'.format(len(genderTable100)-edgesCross)+'\tFemales:␣{}'.format(edgesCross))
```

```
print ('Randomly assigned cross-gender edges fraction: {}'.format(2*edgesCross*(len(genderTable100

for i in range(int(len(lines)/2)):
    if genders[int(lines[i*2].strip())]!=genders[int(lines[i*2+1].strip())]:
        edgesCross = edgesCross +1

print ('edges: {}'.format(edges)+'\tedgesCross: {}'.format(edgesCross))
print ('Actual cross-gender edges fraction: {}'.format(edgesCross/edges))
```

Listing 6: The content of Q2.R

```
binom.test(108, 256, p=2*37*64/101^2, alternative='t')
```

Results:

Listing 7: The content of genderTable.txt

Username	Gender	Username	Gender	Username	Gender
1n9r1d	female	5280BigData	female	AliShamim333	female
AlinaDeniau	female	Alisia1092	female	AngelaWoodall	female
AnnaPerricci	female	ArchivingIt	female	BeeneSteph	female
BexAnnalisa	female	CamtheWicked	female	CassPF	female
Cee_Finley	female	GerdaMueller2	female	GetDownODU	female
HBee2015	female	HollyCroft	female	IngeRudomino	female
J_IreneTieman	female	JenServenti	female	Julibabe21	female
K4arkive	female	KarenKvaughan	female	KarolHolu	female
KingsleySteph	female	LizCohee	female	LyndaLSF	female
MarthaBunton	female	MeghanHoyer	female	MellissaBuys	female
Mercedesew211	female	Miel_vds	female	Milena_Dobreva	female
Mittrach_garten	female	MonicaDYOIart	female	MyriamCTraub	female
Naima3704	female	NancyAjarmeh	female	OKMotovlog	female
PruittAL	female	Sandra_Mason	female	SarahBellefleur	female
SciTechProf	female	ShathaJY	female	TeriBerg96	female
TessaFallon	female	abugseye	female	anatbd	female
annika_hinze	female	archivesmatter	female	artlibrariannyc	female
avadigs	female	badgerhartman	female	bethcron	female
brendaberkelaar	female	carole_gagne	female	ccmarshall	female
cklk90	female	continuants	female	crobinj7074	female
dhowell	female	downey_cara	female	elizhoagcarhart	female
emcaulay	female	erikaris	female	es_land	female
gmj2053	female	hhockx	female	hollips	female
hryanski	female	inadlweb	female	infod1va	female
jessicasmith541	female	jessili	female	jessogden	female
johanna_lautner	female	jubrogers	female	k8lin	female
karenhansn	female	karensnet	female	katestarbird	female
katjakra	female	kaylamarie0110	female	kvanmalssen	female
kzwa	female	landeart	female	liblaura	female

lili_czarina	female	lisallynch	female	lissertations	female
margymavery	female	mdiazdigpres	female	meblake	female
mekonkol	female	mgallinger	female	momof3inov	female
nwatzman	female	opba	female	peripatesis	female
pvanhaitsma	female	rebeccaholte	female	rosalielack	female
saraaubry	female	saschel	female	seahorseadmirer	female
shirapeltzman	female	silvertje	female	skhartmann	female
sshreeves	female	stierholz	female	stillinsky	female
taylor_amy	female	technelily	female	toinette1607	female
trinketapp	female	umurthy	female	vneblitt	female
wrap_ed	female	yasmina_anwar	female	74paddycakes	male
AAlasaadi	male	AntonJRasmussen	male	AricAumann	male
BeachAutoBroker	male	BrianDavison	male	BrosephElder	male
CLIRDLF	male	ChuChoResano	male	CrowNaito	male
DanKerchner	male	DataG	male	DavidUnderdown9	male
DevAhmedhosni	male	Dr_Abomohra	male	EJWalters	male
ELittley	male	ErichdotPy	male	Erikfreiser	male
Fernando___Melo	male	FjellHenning	male	Galsondor	male
GardnerCampbell	male	GeorgeRBuchanan	male	GerhardGossen	male
Ghozia	male	GraemeEarl	male	GrahamSeaman	male
Grant_ODU	male	HugoAViana	male	HussamHallak1	male
IgorBrigadir	male	JCrueger	male	JDecourselle	male
Jakdemir	male	JakeOrlowitz	male	JayNuetron	male
JohnChangGWU	male	John_B_Howard	male	JonesWCaleb	male
JoshCowls	male	KennyDiedrich	male	KevinLevrone3	male
LarryWilson1942	male	LuytenBram	male	Manoj_Chandra11	male
MarkGraham	male	MaxKemman	male	MildlyBored	male
Moha_Magdy	male	Mohamma47262302	male	MohdKamal6	male
NKrabben	male	NeilMonday	male	NielsBr	male
OpenMaze	male	PeterOnymos	male	Philipp_Mayr	male
PoorOldMoot	male	RaifordGuins	male	Reloaded2Boot	male
RikerHampton	male	RjLyric	male	RobVesse	male
SalimChemlal	male	Shaaban_Migo	male	SimpleSimon2013	male
StanZheng	male	StanfordLibs	male	SteveMcLaugh	male
Strollerman	male	TEILandmark	male	THEBenLeBrun	male
TaksNz	male	Thaer_samar	male	TimelessFuture	male
TueHLarsen	male	WeskerAlbert777	male	WessamElhefnawy	male
White__Knuckles	male	WolframElvis	male	_matef	male
_mstevenson	male	aag1091	male	aalsum	male
abrennr	male	abziegler	male	acnwala	male
acocciolo	male	aheshamSalem	male	andrewjbtw	male
archivetype	male	arjenpdevries	male	atomotic	male
azaroth42	male	bertramlyons	male	bfluzin	male
bhaslhofer	male	bicho_daniel	male	bindonlane	male
bitarchivist	male	blefurgy	male	bruno_leonard_	male
carltonnorthern	male	cazzerson	male	claussni	male

cyberlaw	male	daniel_rehn	male	dchud	male
digitalfay	male	dineshpaladhi	male	djw0952	male
dmimno	male	docmattweber	male	doppelen	male
doviethung	male	dwhitenist	male	e_mccain	male
edsu	male	ehetzner	male	ekansa	male
elunca	male	ericleasemorgan	male	erikchoi	male
euanc	male	f_nanni	male	felipebravom	male
fmccown	male	furuta	male	futuresma	male
fzaker	male	gauravguptak1	male	gjfowler	male
hariharshankar	male	hdo003	male	helgeho	male
hochstenbach	male	hvdsomp	male	iFromm	male
iamtimmo	male	ianmilligan1	male	involutish	male
itroch	male	j_w_baker	male	jakkbl	male
jasonmarkwebber	male	jefferson_bail	male	jgsmith	male
joc7188	male	johnaberlin	male	junklight	male
justinfbrunelle	male	kberberi	male	keesone	male
kennethec1	male	kevingashley	male	klischka	male
kurtluther	male	kzakza	male	landlibrarian	male
ldmm	male	lintool	male	lmercereau	male
lnsails	male	lorcanD	male	lysander07	male
machawk1	male	maherzog	male	mamund	male
manoj_chandra_k	male	marcinwilkowski	male	mart1nkle1n	male
maturban1	male	mcdonald	male	meterz1	male
minornotez	male	mjgiarlo	male	mlzman	male
mousta	male	mrchyr	male	mwittin	male
mzarro	male	n_audenaert	male	narfman0	male
normeu	male	northgardner	male	nullhandle	male
olyerickson	male	palewire	male	pampel	male
paulwalk	male	peterkz_swe	male	petrknoth	male
phonedude_mln	male	pj_webster	male	pkeane	male
plamen26968325	male	plbogen	male	pmyoung84	male
prwheatley	male	rdhyee	male	risse691	male
rodwittenberg	male	ruebot	male	rundavidrun	male
rwincewicz	male	saadaitomation	male	seamuslawless	male
shajid333	male	smalljones	male	sspranay	male
stevehit	male	subotic	male	subsublibrary	male
tedlawless	male	thabing	male	thefaisalahmad	male
this_phillips	male	tjowens	male	trevormunoz	male
tsuomela	male	twarko	male	txkuhn	male
vector_ctrl	male	vincelebow	male	vphill	male
weblawlib	male	westfood	male	williamjnixon	male
witch_doctor81	male	xabuci	male	ycpdan	male
zeonfernando	male	ziadmatni	male	zittrain	male
6T7VXu11XcTiAbd	unknown	9ulovesu	unknown	ACMSIGIR	unknown
ARLIS_NY	unknown	AT_Webarchive	unknown	AVArchivist	unknown
AlephArchives	unknown	AllThingsGraph	unknown	AndreaGoethals	unknown

AndreeaCorinaM2	unknown	Antenna_Lab	unknown	Arkivum	unknown
ArnoudGoos	unknown	AutonomyIncub8r	unknown	BL_Labs	unknown
BelaGipp	unknown	BergisJules	unknown	BizApproved	unknown
CRajaharsha	unknown	CScoutTech	unknown	ChiefScientist	unknown
ChrisAldrich	unknown	CodioHQ	unknown	CommonCrawl	unknown
CorrenMcCoy	unknown	DLWebBnF	unknown	DLWeb_Seb	unknown
DLib2014	unknown	DanMilanko	unknown	DialUCL	unknown
DigiCultureKCL	unknown	DrAinenwar	unknown	DrDanetteAllen	unknown
DwiDwisyafitri	unknown	EdelOShea83	unknown	Educopia	unknown
EmeraldIKM	unknown	EmeraldLibrary	unknown	FFuqiang	unknown
Faryane	unknown	ForgetITProject	unknown	FredericikB	unknown
GJST	unknown	HCIR_GeneG	unknown	HamptonRFinest	unknown
Hindawi	unknown	HistWebArchives	unknown	History2point0	unknown
ISRJINDIA	unknown	IntelligenceTV	unknown	IuliaCristina12	unknown
JCDLConf	unknown	JS_Smith_71	unknown	JStoddert	unknown
Jack_Stran	unknown	JesseforCouncil	unknown	JoeLemanski	unknown
Karki693	unknown	KhTalha6	unknown	KyleFJackson	unknown
KyleStr	unknown	LulwahMA	unknown	MPanula	unknown
Magda69425548	unknown	Mementoweb	unknown	MeredithA	unknown
Milbala	unknown	MonarchsAimHigh	unknown	NDSA2	unknown
NetLab_dk	unknown	NetPreserve	unknown	NgYewli	unknown
NicolaJBingham	unknown	NorfolkVA_	unknown	ODUAlumni	unknown
ODUAuxServices	unknown	ODUPeninsulaCtr	unknown	ODURugby_	unknown
OTHSEAFOOD	unknown	Oorlogsbronnen	unknown	OpenResearchExe	unknown
PSleeman	unknown	Paisadalo	unknown	PapachriL	unknown
PeerProd	unknown	PolitAdArchive	unknown	Pravacana_Mats	unknown
RJIJDNA	unknown	RSSNewsmaster	unknown	RajnishMallick	unknown
ReedTechCorp	unknown	RossiAtanassova	unknown	SIGIR17	unknown
SPARC_NA	unknown	SoftGid_com	unknown	SouthernKiaGB	unknown
Stesker3D	unknown	TPDL2016	unknown	TagTrees	unknown
TrueHiddenFact	unknown	Tunaz_Islam	unknown	UKWebArchive	unknown
WOSP2014	unknown	WPKefi	unknown	WSUNDSA	unknown
WebART12	unknown	WebArchivists	unknown	WebSciDL	unknown
WehoMan88	unknown	Wikicite	unknown	XYOU	unknown
_akisato	unknown	acquire_UCLA	unknown	agrotke	unknown
ahmedkalka97	unknown	alexwade	unknown	amaranaas	unknown
amlhasbat	unknown	andy_anjoyh	unknown	anjacks0n	unknown
annediaz01	unknown	archiveis	unknown	archiveitorg	unknown
archivelle	unknown	archivportal	unknown	athurman	unknown
awptix	unknown	beatles__beatle	unknown	beyondcitation	unknown
bodhisattvaOS	unknown	brettbobley	unknown	brijmohanrana4	unknown
bumdots4eva	unknown	calpopp	unknown	carolc16	unknown
cbmrccc	unknown	cgknowles	unknown	chglass	unknown
chrisbellekom	unknown	chrisfreeland	unknown	chrislaoscott	unknown
clarellewellyn	unknown	clem_kev	unknown	cni_org	unknown
covenant_ws	unknown	datacartel1	unknown	dbpedia	unknown

degruyter_lib	unknown	desouzatech	unknown	detailmatters	unknown
dhombe3	unknown	digiarchiveteam	unknown	digitopia_nl	unknown
diglib	unknown	documentnow	unknown	dswarm	unknown
dtscnc	unknown	erinengle	unknown	forschungsdaten	unknown
foucaultwelles	unknown	fromADMwithlove	unknown	goodenjm	unknown
grotophorst	unknown	hanysalaheldeen	unknown	iPRES2013	unknown
iPRESconf	unknown	ibnesayeed	unknown	icdm2016	unknown
idokius	unknown	ijdl	unknown	imafuturemedic	unknown
impactzoneco	unknown	internetarchive	unknown	isi2015conf	unknown
jalbertbowdenii	unknown	jcdl2012	unknown	jcdl2013	unknown
jcdl2015	unknown	jcdl2016	unknown	jcdl2017	unknown
jesseajohnston	unknown	jkamps	unknown	joansm1th	unknown
joemcgonegal	unknown	jotschirr	unknown	jschneider	unknown
jsicot	unknown	jteevan	unknown	kboughida	unknown
kingswebit	unknown	klhiggins33	unknown	knmnyn	unknown
kraabus	unknown	kristsi	unknown	kylewilliams87	unknown
ldirks	unknown	lestextesR	unknown	lljohnston	unknown
lmaccork	unknown	lookouthoney	unknown	lwynholds	unknown
maheedhargunnam	unknown	maikhanhlam	unknown	majetisiri	unknown
masroorfreaks	unknown	mbutel	unknown	mdzellerCDAL	unknown
metalinker	unknown	mgome0072	unknown	migcosta	unknown
mixnode	unknown	mmicros1	unknown	mre1920	unknown
mridulish	unknown	mummifyit	unknown	myerscarpenter	unknown
nanopub_org	unknown	nattiyak	unknown	nichworby	unknown
njcomputergroup	unknown	no_identd	unknown	noumenal_woman	unknown
oaisdujour	unknown	oducs	unknown	ohttic	unknown
oreficeplumbing	unknown	oso525gang	unknown	peplluis7	unknown
permacc	unknown	perrycollins	unknown	pigandpublish	unknown
pre11e	unknown	printuu	unknown	ptrourke	unknown
rdayMedia	unknown	robin_ruggaber	unknown	robina_naazli	unknown
robinmkatz	unknown	ruyhliu	unknown	ryanfb	unknown
salamancaschool	unknown	samalanmeister	unknown	samy_tawab	unknown
save4use	unknown	schuurmanna	unknown	shawnmjones	unknown
sheershasamach1	unknown	skurt	unknown	smartfit_atees	unknown
sooraya_sadulla	unknown	sorgentelinda	unknown	space360vr	unknown
squealermusic	unknown	syamiliC	unknown	tamingdata	unknown
theovv	unknown	tmeehleib	unknown	todrobbins	unknown
topgolfvabeach	unknown	tpdl2017	unknown	ttso	unknown
tywalters1	unknown	unknown_6	unknown	unmil	unknown
unstablearchive	unknown	usa8951	unknown	uskudarli	unknown
violetailik	unknown	walkeroh	unknown	webrecorder_io	unknown
weiglemc	unknown	williamjturkel	unknown	woilgoo	unknown
wosp2015	unknown	wospworkshop	unknown	www2013rio	unknown
wwwtxt	unknown	xuxomln	unknown	zhaohuan_wang	unknown
zimeon	unknown	zxie	unknown		

Listing 8: The content of genderTable100.txt

Username	Gender	Username	Gender	Username	Gender
1n9r1d	female	AngelaWoodall	female	BexAnnalisa	female
CassPF	female	Cee_Finley	female	HBee2015	female
HollyCroft	female	KarenKvaughan	female	KingsleySteph	female
MeghanHoyer	female	Miel_vds	female	Mittrach_garten	female
MyriamCTraub	female	OKMotovlog	female	PruittAL	female
ShathaJY	female	annika_hinze	female	artlibrariannyc	female
avadigs	female	brendaberkelaar	female	carole_gagne	female
elizhoagcarhart	female	es_land	female	hollips	female
infod1va	female	jessicasmith541	female	jubrogers	female
katestarbird	female	kzwa	female	lili_czarina	female
lissertations	female	mekonkol	female	nwatzman	female
peripatesis	female	pvanhaitsma	female	rebeccaholte	female
stierholz	female	74paddycakes	male	AntonJRasmussen	male
CLIRDLF	male	CrowNaito	male	DanKerchner	male
DavidUnderdown9	male	DevAhmedhosni	male	Dr_Abomohra	male
ErichdotPy	male	Fernando___Melo	male	GardnerCampbell	male
Grant_ODU	male	HugoAViana	male	HussamHallak1	male
JCrueger	male	JDecourselle	male	JakeOrlowitz	male
JayNuetron	male	John_B_Howard	male	KennyDiedrich	male
KevinLevrone3	male	LarryWilson1942	male	LuytenBram	male
Manoj_Chandra11	male	MarkGraham	male	Mohamma47262302	male
MohdKamal6	male	RikerHampton	male	SalimChemlal	male
Shaaban_Migo	male	TEILandmark	male	TueHLarsen	male
White__Knuckles	male	_mstevenson	male	abziegler	male
andrewjbtw	male	arjenpdevries	male	bicho_daniel	male
bitarchivist	male	bruno_leonard_	male	cazzerson	male
claussni	male	dineshpaladhi	male	doviethung	male
futuresma	male	helgeho	male	johnaberlin	male
landlibrarian	male	lmercereau	male	lysander07	male
manoj_chandra_k	male	marcinwilkowski	male	minornotez	male
mrchyr	male	normeu	male	pampel	male
phonedude_mln	male	pmyoung84	male	risse691	male
ruebot	male	saadaitomation	male	seamuslawless	male
subsublibrary	male	ycpdan	male		

Table 1: Gender Homophily Test Result

Male	Female	Total edges	Cross-gender edges	p-value	95 percent confidence interval
64	37	256	108	0.1882	0.3606438 - 0.4849366

According to these figures, Randomly assigned cross-gender edge fraction is 0.4642682, Actual cross-gender edge fraction is 0.421875 and just 9% off. The Exact Binomial Test p-value is 0.1882, which means the possibility of 81.2% that they are not equal. So the gender homophily exists but slightly.

3 Using D3, create a graph of the Karate club before and after the split.

Have the transition from before/after the split occur on a mouse click. This is a toggle, so the graph will go back and forth beween connected and disconnected.

Algorithm:

1. Copy the last 34 lines of the matrix in "zachary.dat" and save them into "karate.txt".
2. Use the package "igraph"(http://igraph.org/r/) in R to open "karate.txt".
3. Use the edge betweenness community detection algorithm.
4. Cut the merge tree to get 2 communities.
5. Plot the 2 communities.
6. Save the graph in "edgelist" format to "karateGraph.txt" and add the group index of the nodes to the file.
7. Open "karateGraph.txt", for every edge check if the nodes are from different groups and add mark to the edge.
8. Use the package "networkx"(https://networkx.readthedocs.io/en/stable/) to save the graph to "karate.json".
9. Use the D3 to visualize the graph.

Source code:

Listing 9: The content of karate.R

```
library("igraph")
data = as.matrix(read.table("karate.txt"))
karate = graph_from_adjacency_matrix(data, mode = "undirected", weighted = TRUE)
ebc = edge.betweenness.community(karate)
cut = cutat(ebc,2)
colors = rainbow(2)
plot(karate, vertex.color=colors[cut])
cut
E(karate)
write_graph(karate, "karateGraph.txt", format = "edgelist")
write.table("nodes.gender:", "karateGraph.txt", append=TRUE, quote = FALSE, sep ="\n", row.names =
write.table(cut, "karateGraph.txt", append=TRUE, sep ="\n", row.names = FALSE, col.names = FALSE)
```

Listing 10: The content of graphWrite-karate.py

```
import networkx as nx
from networkx.readwrite import json_graph
import json

g = nx.Graph()

f = open("karateGraph.txt", "r", encoding='utf-8')

edge=[]
```

```
while True:
    line = f.readline().strip()
    if line!='nodes.gender:':
        line = line.split()
        row = [int(line[0]),int(line[1])]
        edge.append(row)
    else:
        break

lines=f.readlines()
f.close()

for element in edge:
    if lines[element[0]]==lines[element[1]]:
        g.add_edge(element[0],element[1],broken=0)
    else:
        g.add_edge(element[0],element[1],broken=1)

for i in range(len(lines)):
    g.add_node(i, name=i, group=int(lines[i].strip()))

data = json_graph.node_link_data(g)
s = json.dumps(data)

f = open('karate.json', 'w')
f.write(s)
f.close()
```

Listing 11: The content of karate.html

```
<!DOCTYPE html>
<meta charset="utf-8">
<style>

.links line {
  stroke: #999;
  stroke-opacity: 0.6;
}

.node text {
  pointer-events: none;
  font: 15px sans-serif;
}

.nodes circle {
  stroke: #fff;
```

```
  stroke-width: 1.5px;
}

</style>
<svg width="1366" height="768"></svg>
<script src="https://d3js.org/d3.v4.min.js"></script>
<script>

//create somewhere to put the force directed graph
var svg = d3.select("svg"),
  width = +svg.attr("width"),
  height = +svg.attr("height");

var flag=0;

d3.json("karate.json", function(graph){
  var simulation = d3.forceSimulation()
    .nodes(graph.nodes);

  simulation
    .force("charge_force", d3.forceManyBody())
    .force("center_force", d3.forceCenter(width / 2, height / 2));

  var node = svg.selectAll(".node")
      .data(graph.nodes)
      .enter().append("g")
      .attr("class", "node");

  node.append("text")
      .attr("dx", 12)
      .attr("dy", ".35em")
      .text(function(d) { return d.name });

  node.append("title").text(function(d) { return d.name; });

  var nodeCircle=node.append("circle")
      .attr("r", 5)
      .attr("fill", circleColour);

  function circleColour(d){
      if(flag ==0)
      {
        return "red";
      }
      else if(d.group ==1)
      {
```

```
      return "blue";
    }
    else
    {
      return "yellow";
    }
}

simulation.on("tick", tickActions );

var link_force =  d3.forceLink(graph.links);

simulation.force("links",link_force);

var link = svg.append("g")
  .attr("class", "links")
  .selectAll("line")
  .data(graph.links)
  .enter().append("line")
  .attr("stroke-width", 2)
  .style("display", linkDisplay);

function linkDisplay(d){
    if(flag ==0)
    {
      return "inline";
    }
    else if(d.broken == 0)
    {
      return "inline";
    }
    else
    {
      return "none";
    }
}

svg.on("click",function() {
    flag=1-flag;
    nodeCircle.attr("fill", circleColour);
    link.style("display", linkDisplay);
})

function tickActions() {
  link
    .attr("x1", function(d) { return d.source.x; })
```

```
      .attr("y1", function(d) { return d.source.y; })
      .attr("x2", function(d) { return d.target.x; })
      .attr("y2", function(d) { return d.target.y; });

    node.attr("transform", function(d) { return "translate(" + d.x + "," + d.y + ")"; });
  }
});
</script>
```

Results: https://cdn.rawgit.com/zhangboroy/cs532-s17/01870cb6/assg06_submission/karate.html