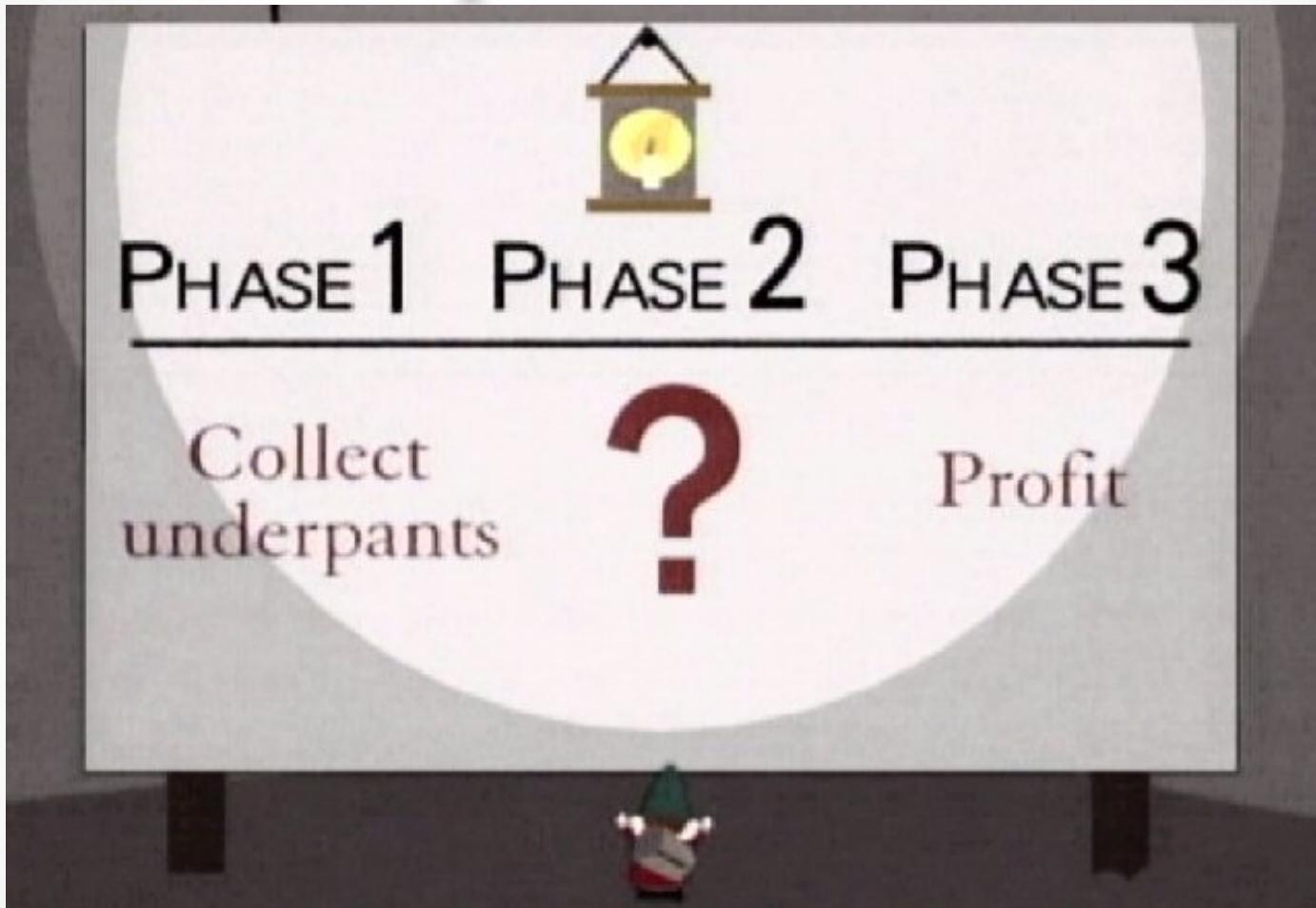


Social Network Analysis – Github Activities

• • •

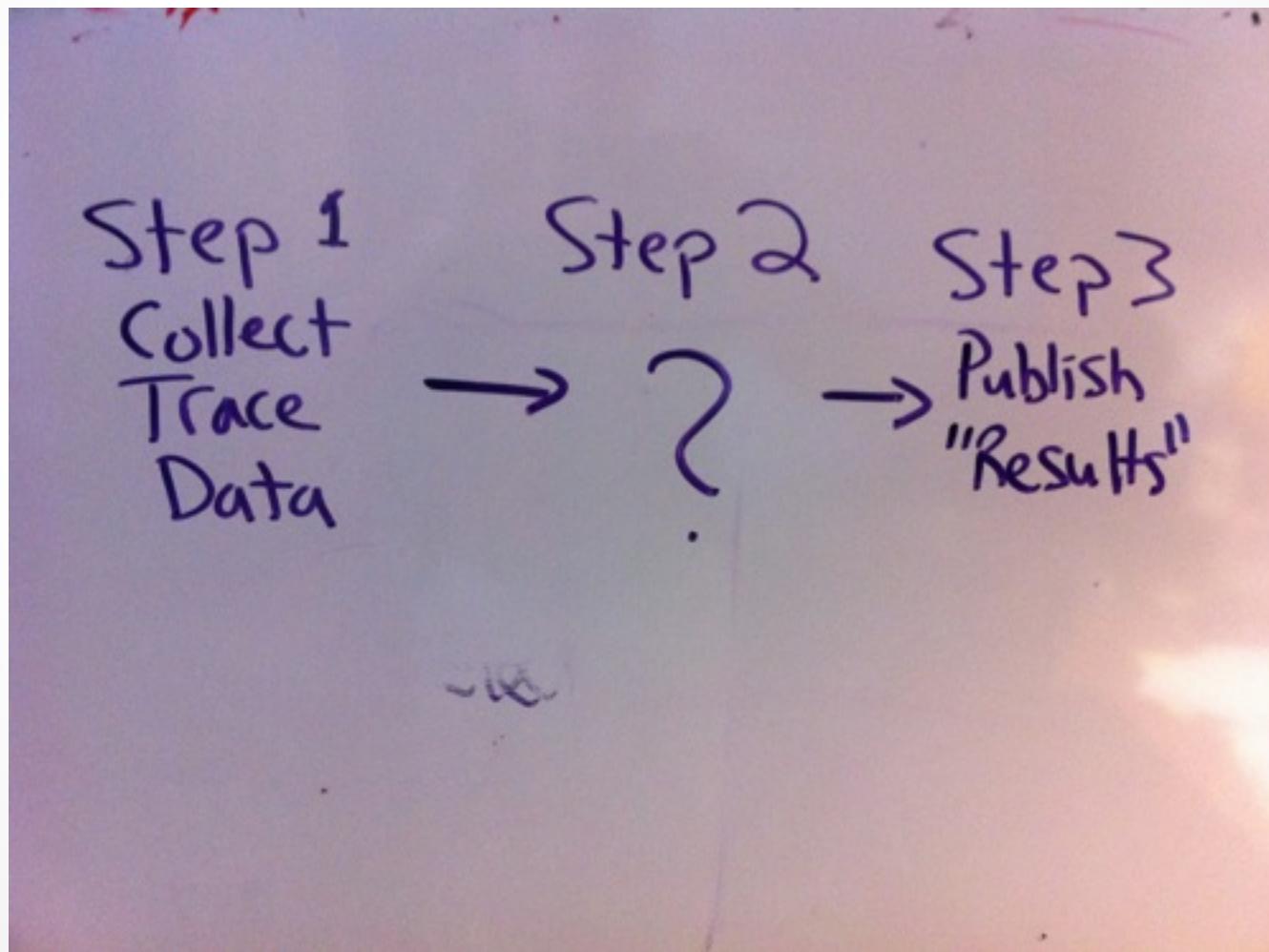
Sean P. Goggins, Ph.D
Associate Professor, University of Missouri

Underpants Gnomes



With much courtesy from the US TV Program “South Park”

Underpants Gnomes

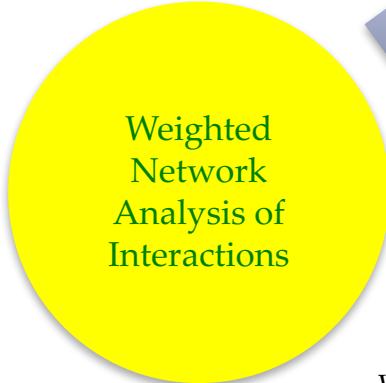


Addressing The Underpants Gnome Postulate

• • •

Group Informatics Described

Identify Key Information Brokers



Discussion Post

- Read
- Response

Classification

- Open Coding
- Axial Coding

Identification of Coordination Events

- Time proximity
- Topical proximity

Weight Connections Based on Time Distance, Grouped By Topic and informed by analysis of time distance between posts.

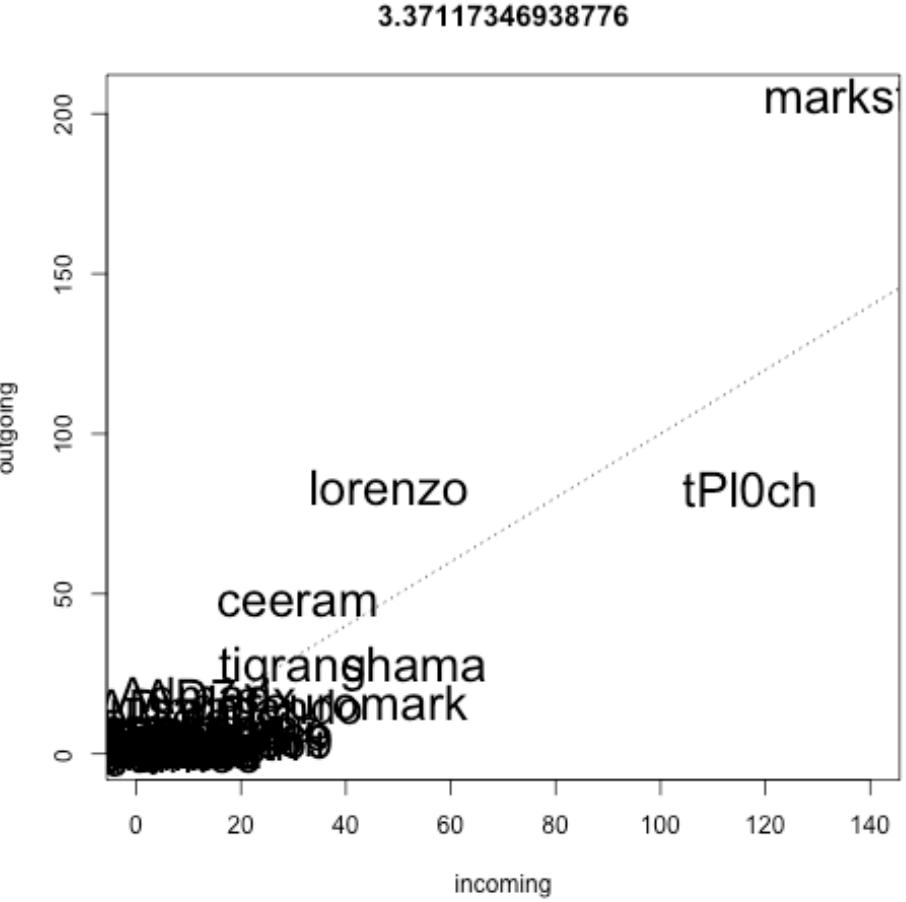
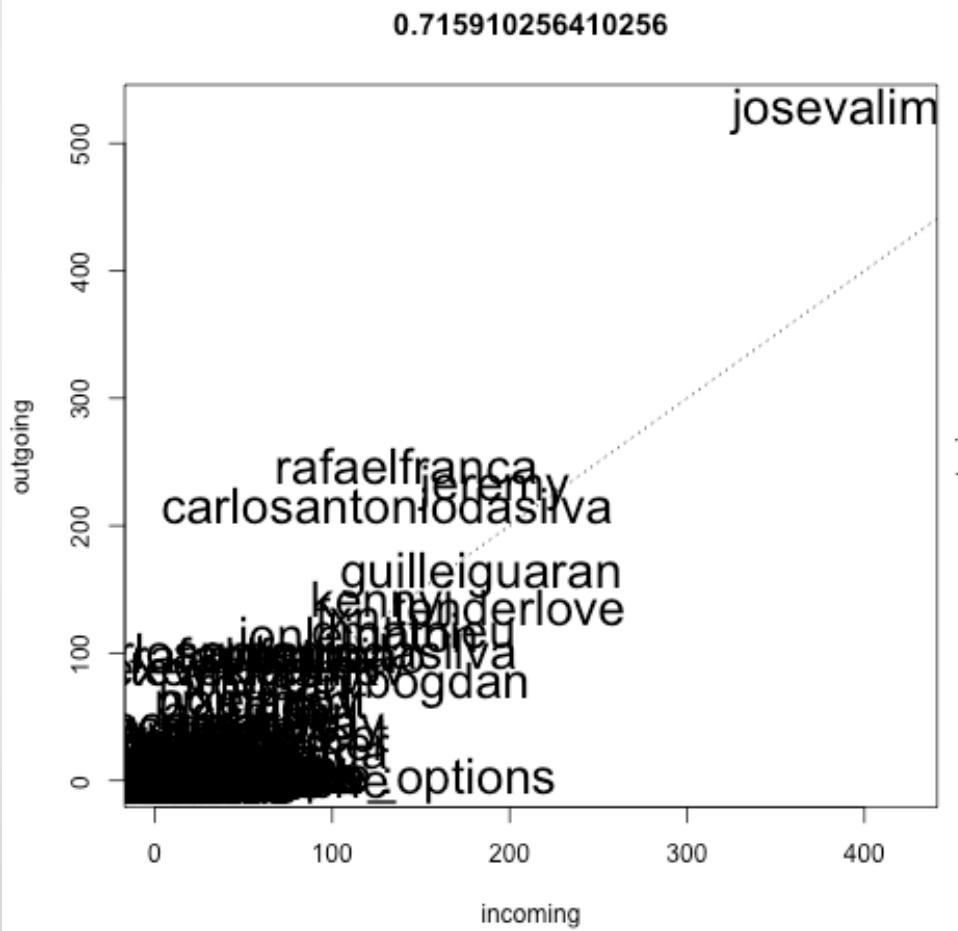
Aggregation of Posts by Topic

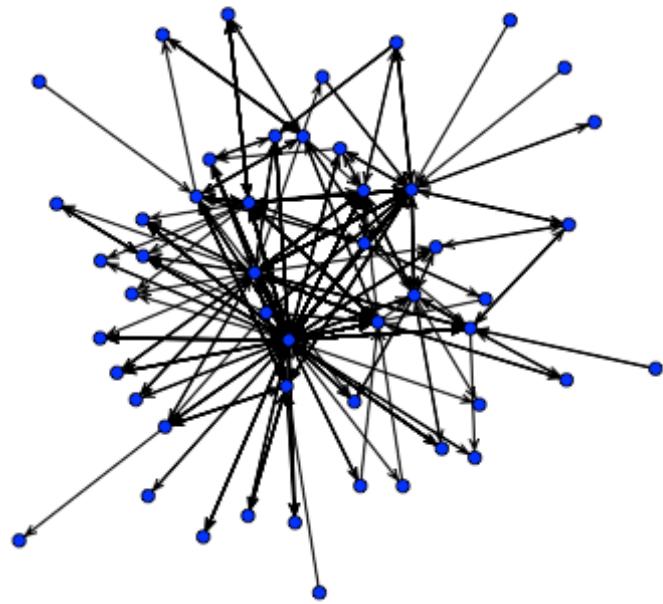
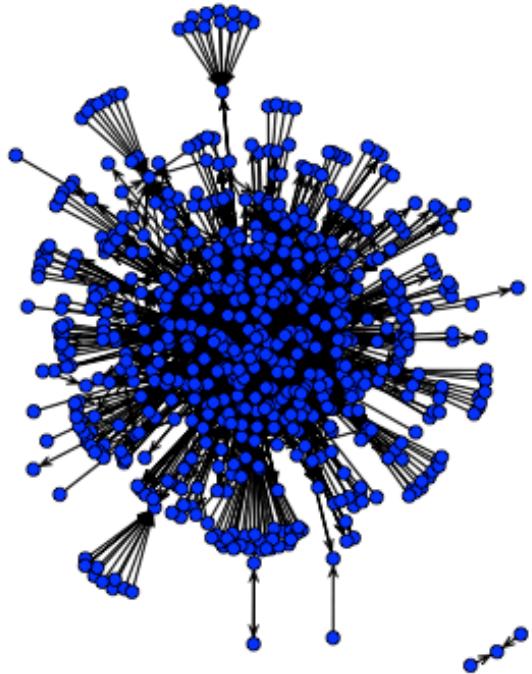
Data: Github

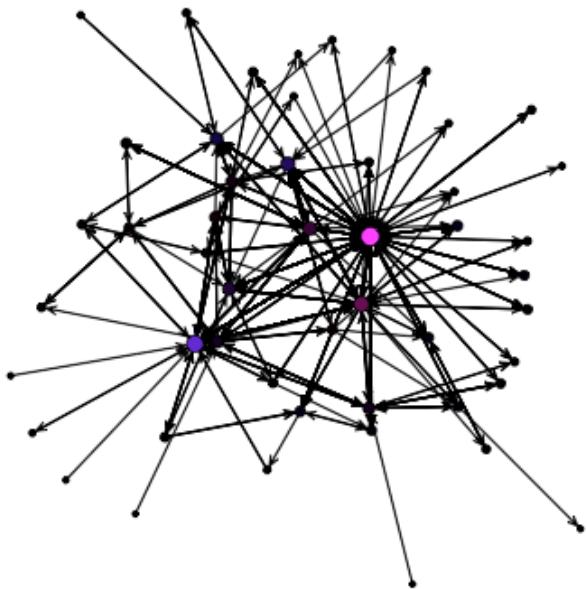
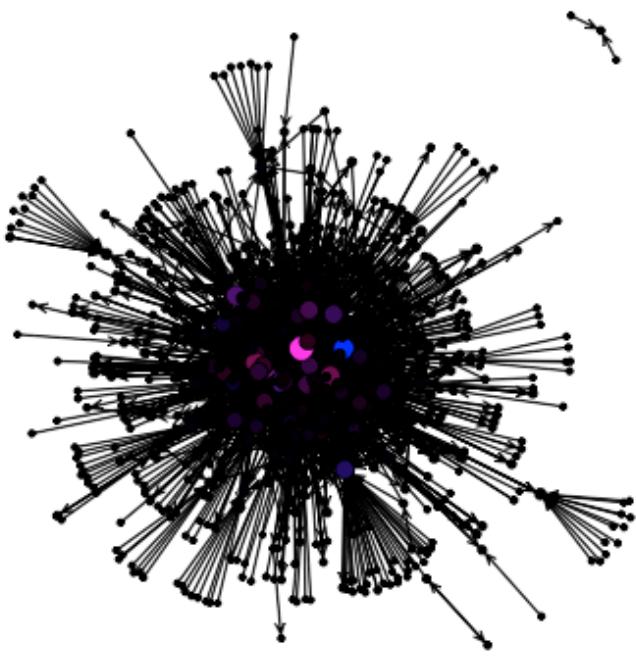
source	target	timeperiodN	STATUS	Merged	Code	codeCode	pullrequestI	type
josevalim	bogdan		0 Closed	1	1		47	comments
bogdan	josevalim		0 Closed	1	1		47	comments
josevalim	bogdan		0 Closed	1	1		47	comments
bogdan	josevalim		0 Closed	1	1		47	comments
vijaydev	bogdan		0 Closed	1	1		47	comments
carlosantonio	bogdan		0 Closed	1	1		47	comments
bogdan	carlosantonio		0 Closed	1	1		47	comments
josevalim	bogdan		0 Closed	1	1		47	comments
bogdan	josevalim		0 Closed	1	1		47	comments
carlosantonio	bogdan		0 Closed	1	1		47	comments
josevalim	bogdan		0 Closed	1	1		48	comments
dasch	amatsuda		0 Closed	0	1		123	comments
spastorino	amatsuda		0 Closed	0	1		123	comments
amatsuda	tenderlove		0 Closed	1	3		124	comments
tenderlove	amatsuda		0 Closed	1	3		124	comments

Github Network Activity

One







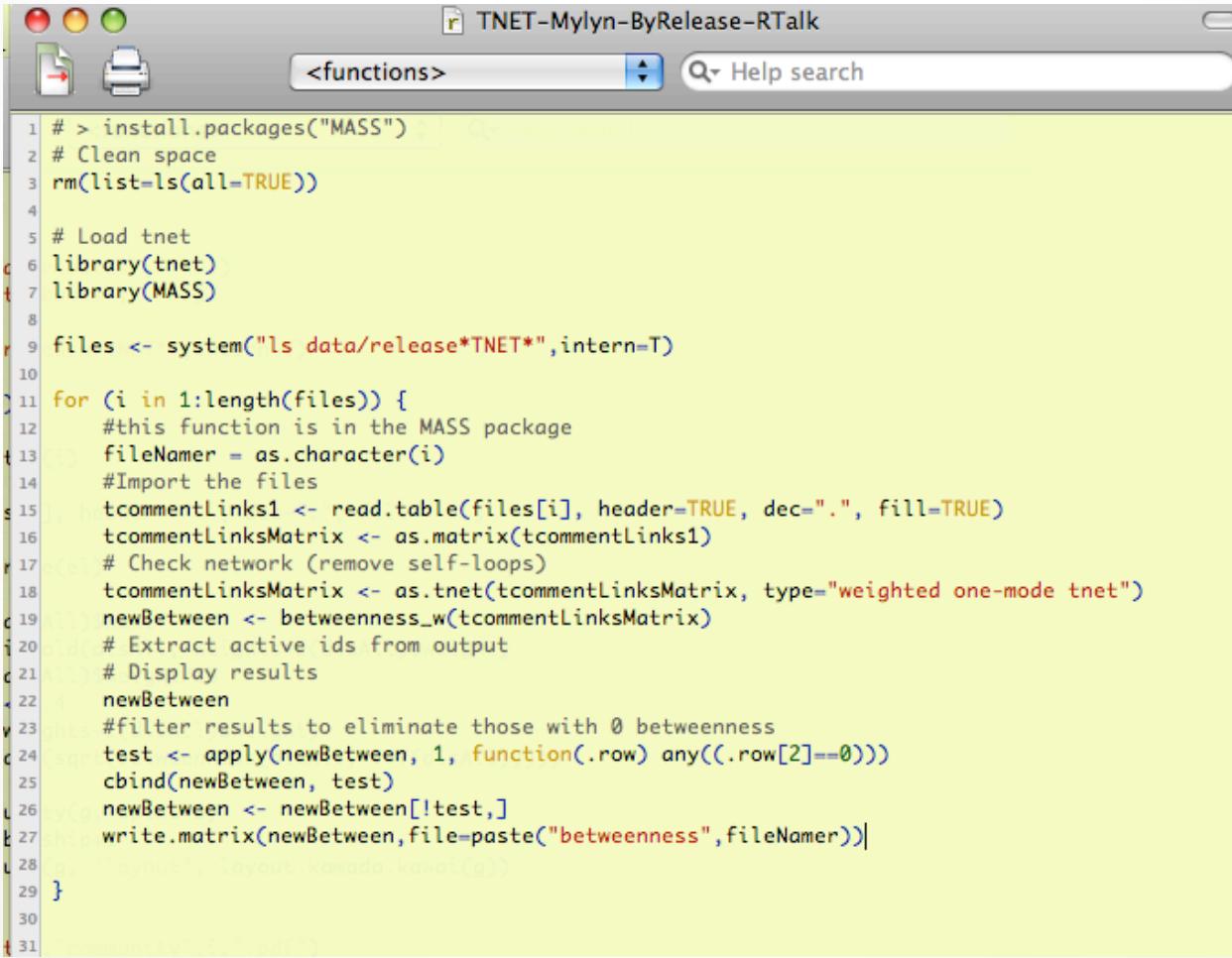
Actual R – Code

- Work through setup
- Scripts are ready to run
- Talk Through Them and Walk around to help

Further Analysis Tools

- Eight Mylyn Releases (Temporal Analysis)
- R Packages Used
 - TNET
 - iGraph
 - Statnet

Weighted Network: TNET



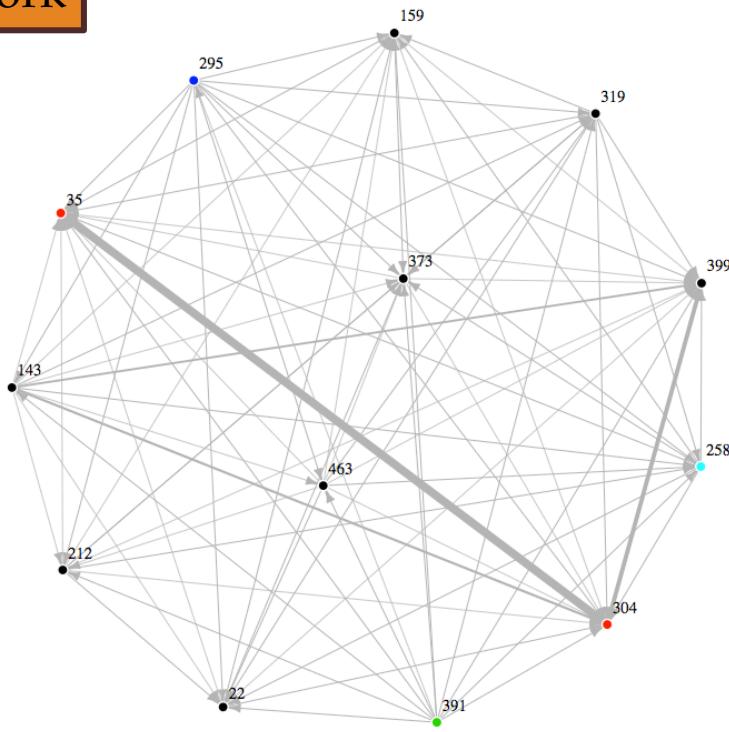
The screenshot shows a software interface titled "TNET-Mylyn-ByRelease-RTalk". The main window is an R script editor with the following code:

```
1 # > install.packages("MASS")
2 # Clean space
3 rm(list=ls(all=TRUE))
4
5 # Load tnet
6 library(tnet)
7 library(MASS)
8
9 files <- system("ls data/release*TNET*",intern=T)
10
11 for (i in 1:length(files)) {
12     #this function is in the MASS package
13     fileNamer = as.character(i)
14     #Import the files
15     tcommentLinks1 <- read.table(files[i], header=TRUE, dec=".",
16     fill=TRUE)
16     tcommentLinksMatrix <- as.matrix(tcommentLinks1)
17     # Check network (remove self-loops)
18     tcommentLinksMatrix <- as.tnet(tcommentLinksMatrix, type="weighted one-mode tnet")
19     newBetween <- betweenness_w(tcommentLinksMatrix)
20     # Extract active ids from output
21     # Display results
22     newBetween
23     #filter results to eliminate those with 0 betweenness
24     test <- apply(newBetween, 1, function(.row) any((.row[2]==0)))
25     cbind(newBetween, test)
26     newBetween <- newBetween[!test,]
27     write.matrix(newBetween,file=paste("betweenness",fileNamer))
28 }
29
30
31
```

The Dense Graph (Work)

- Developers create a dense graph. Not a complete graph, but dense.

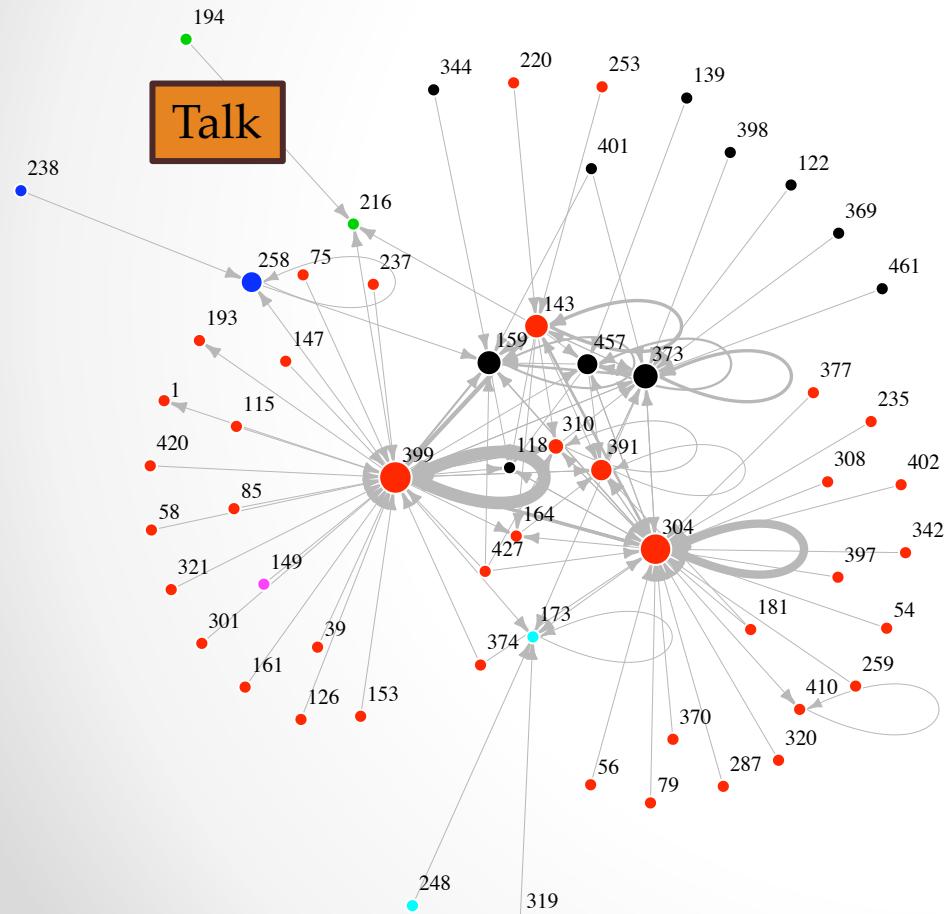
Work



```
1 rm(list=ls(all=TRUE))
2 library(igraph)
3 library(ggplot2)
4 igraph.par("print.vertex.attributes", TRUE)
5 igraph.par("print.edge.attributes", TRUE)
6
7 files <- system("ls mylynCloseness/*", intern=T)
8
9 for (i in 1:length(files)) {
10
11   fileNamer = as.character(i)
12   #Import the files
13   el <- read.table(files[i], header=TRUE, dec=".",
14   #netName = "n"+as.character(i)
15   #print(el)
16
17   #el <- read.table("data/release10TNET.csv", header=TRUE)
18
19   disAll <- graph.data.frame(el)
20
21   V(disAll)$label <- V(disAll)$name
22   layout.fruchterman.reingold(disAll, weights=E(disAll)$weight)
23   E(disAll)$width <- E(disAll)$weight
24   E(disAll)$arrow.size <- 0.4
25   layout.spring(disAll, weights=E(disAll)$weight)
26   V(disAll)$size <- (3+betweenness(disAll, v=V(disAll))))
27   g<-disAll
28   com <- spinglass.community(g, spins=8)
29   V(g)$color <- com$membership+1
30   g <- set.graph.attribute(g, "layout", layout.kamada.kawai(g))
31
32   #output the file
33   filename=paste("output/","closeness",i,".pdf")
34   pdf(filename)
35   #plot.igraph(g)
36   plot(g, vertex.label.dist=.5, vertex.label.cex=.7,
37   vertex.frame.color="white")
38   dev.off()
```

A Sparser Graph (Talk)

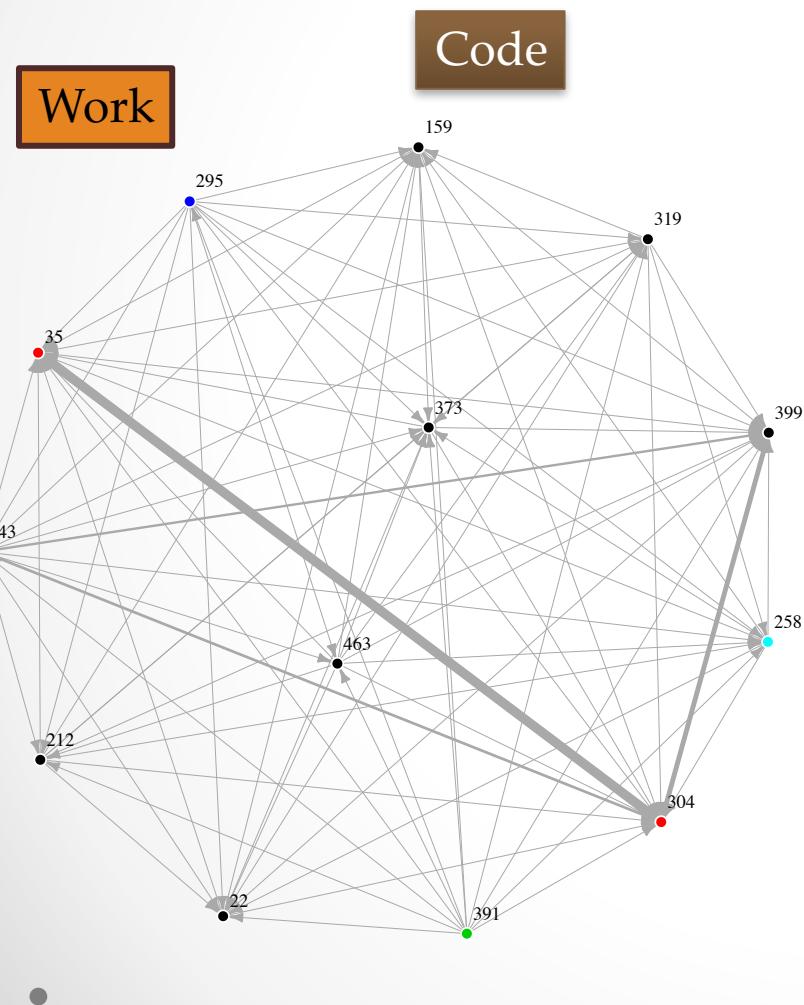
- Commenter's create a sparse graph



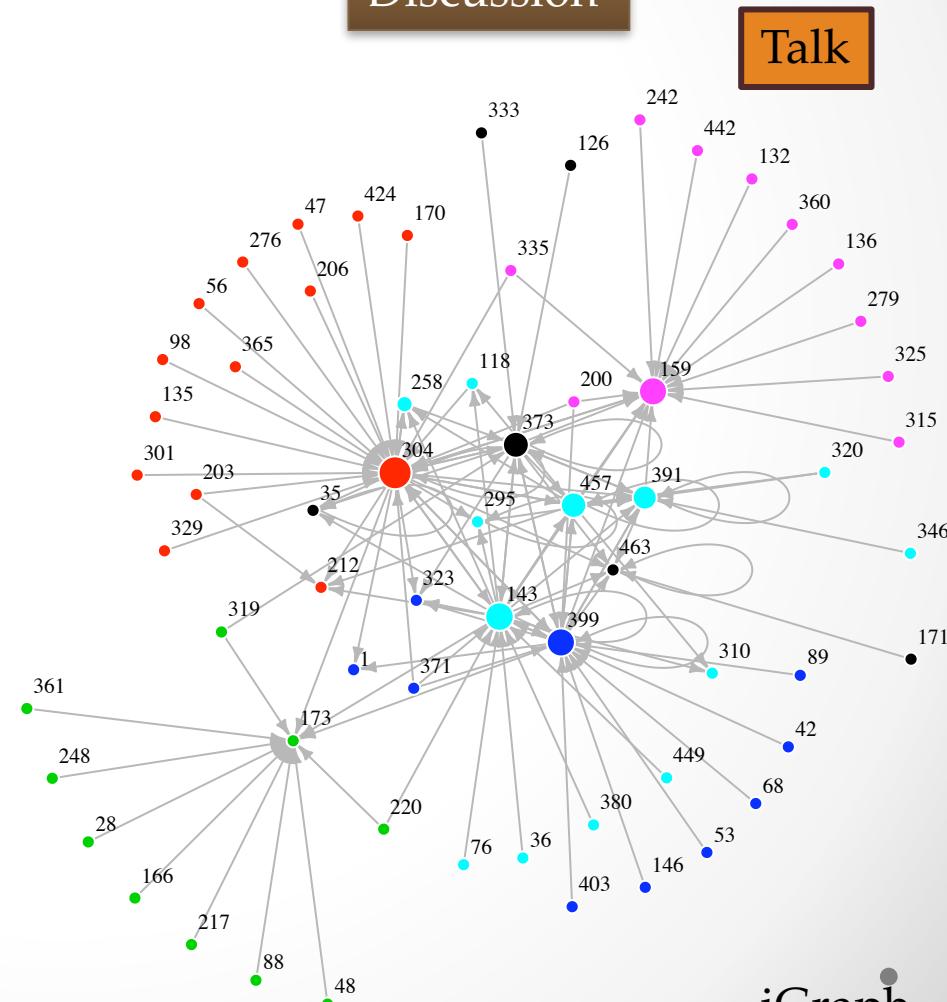
```
1 rm(list=ls(all=TRUE))
2 library(igraph)
3 library(ggplot2)
4 igraph.par("print.vertex.attributes", TRUE) #for release10TNET.csv", header=TRUE)
5 igraph.par("print.edge.attributes", TRUE)
6
7 files <- system("ls data/release*TNET*",intern=T)
8
9 for (i in 1:length(files)) {
10   # Fruchterman-Reingold algorithm
11   # Import the file
12   el <- read.table(files[i], header=TRUE, dec=". ", fill=TRUE)
13
14   disAll <- graph.data.frame(el) #uses community(g, spins=8)
15
16   V(disAll)$label <- V(disAll)$name
17   layout.fruchterman.reingold(disAll, weights=E(disAll)$weight)
18   E(disAll)$width <- E(disAll)$weight/25
19   E(disAll)$arrow.size <- 0.4
20
21   layout.spring(disAll,weights=E(disAll)$weight)
22   V(disAll)$size <- -(3+sqrt(sqrt(betweenness(disAll, v=V(disAll)))))
23   g<-disAll
24   com <- spinglass.community(g, spins=8)
25   V(g)$color <- com$membership+1
26   g <- set.graph.attribute(g, "layout", layout.kamada.kawai(g))
27
28 #output the file
29 filename=paste("output/","community",i,".pdf")
30 pdf(filename)
31 #plot.igraph(g)
32 plot(g, vertex.label.dist=.5, vertex.label.cex=.7, vertex.label.color="black",
33 vertex.frame.color="white")
34 dev.off()
35 }
```

Release One (2.0) Analysis

Release 1



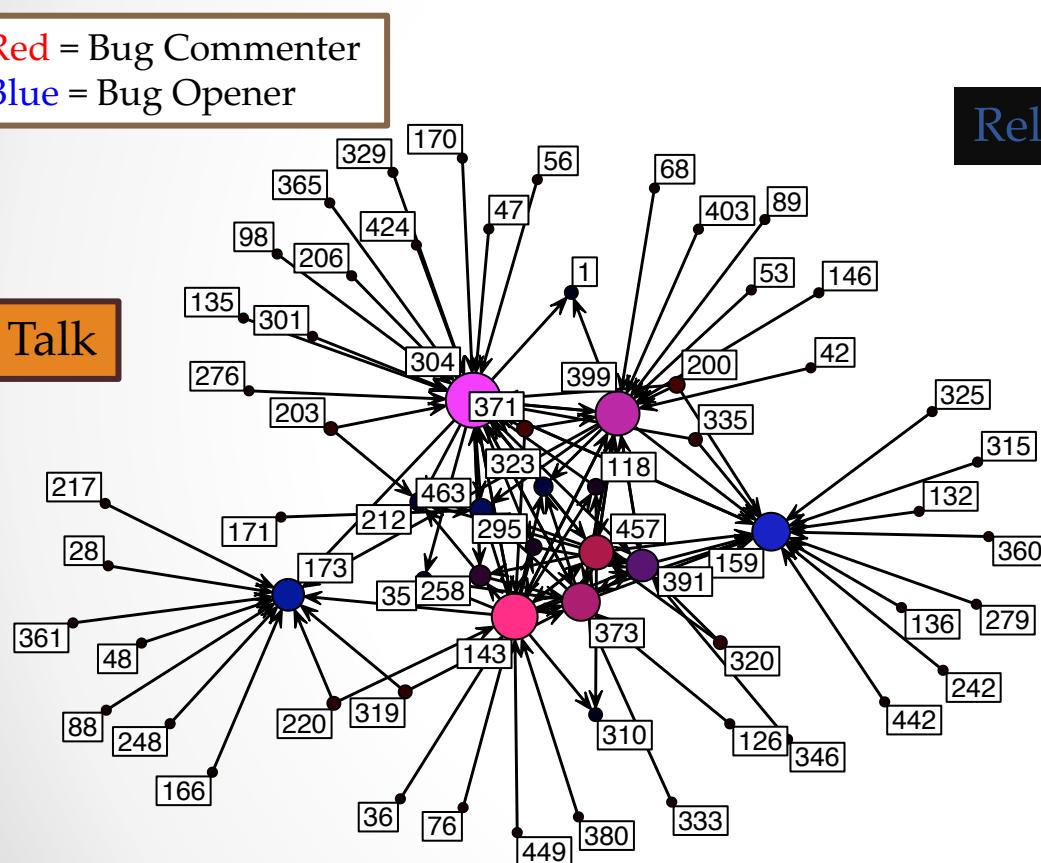
Discussion



STATNET for Discussion

- StatNet

Red = Bug Commenter
Blue = Bug Opener



StatNET

Release 1

```
1 rm(list=ls(all=TRUE))
2
3 library(network)
4 library(statnet)
5 library(MASS)
6
7 #####
8 # Read the Data in
9 #####
10 files <- system("ls data/release*TNET*",intern=T)
11
12 for (i in 1:length(files)) {
13   #this function is in the MASS package
14   fileNamer = as.character(i)
15   #Import the files
16   el <- read.table(files[i], header=TRUE, dec=". ", fill=TRUE)
17   #netName = "n"+as.character(i)
18   print(el)
19   n<-network(el,matrix.type="edgelist")
20   print(n)
21   print(i)
22
23   ## Use centrality scores to size and color the network plot
24   filename=paste("output/", "4graphFourCentralityColorized", i, ".pdf")
25   pdf(filename)
26   gplot(n,vertex.cex=(ideg+odeg)^0.5/2, vertex.sides=50, label.cex=0.8,
27   vertex.col=rgb(odeg/max(odeg), 0, ideg/max (ideg)), label=network.vertex.names(n),
28   displayisolates=FALSE, boxed.labels=TRUE)
29   dev.off()
30 }
```

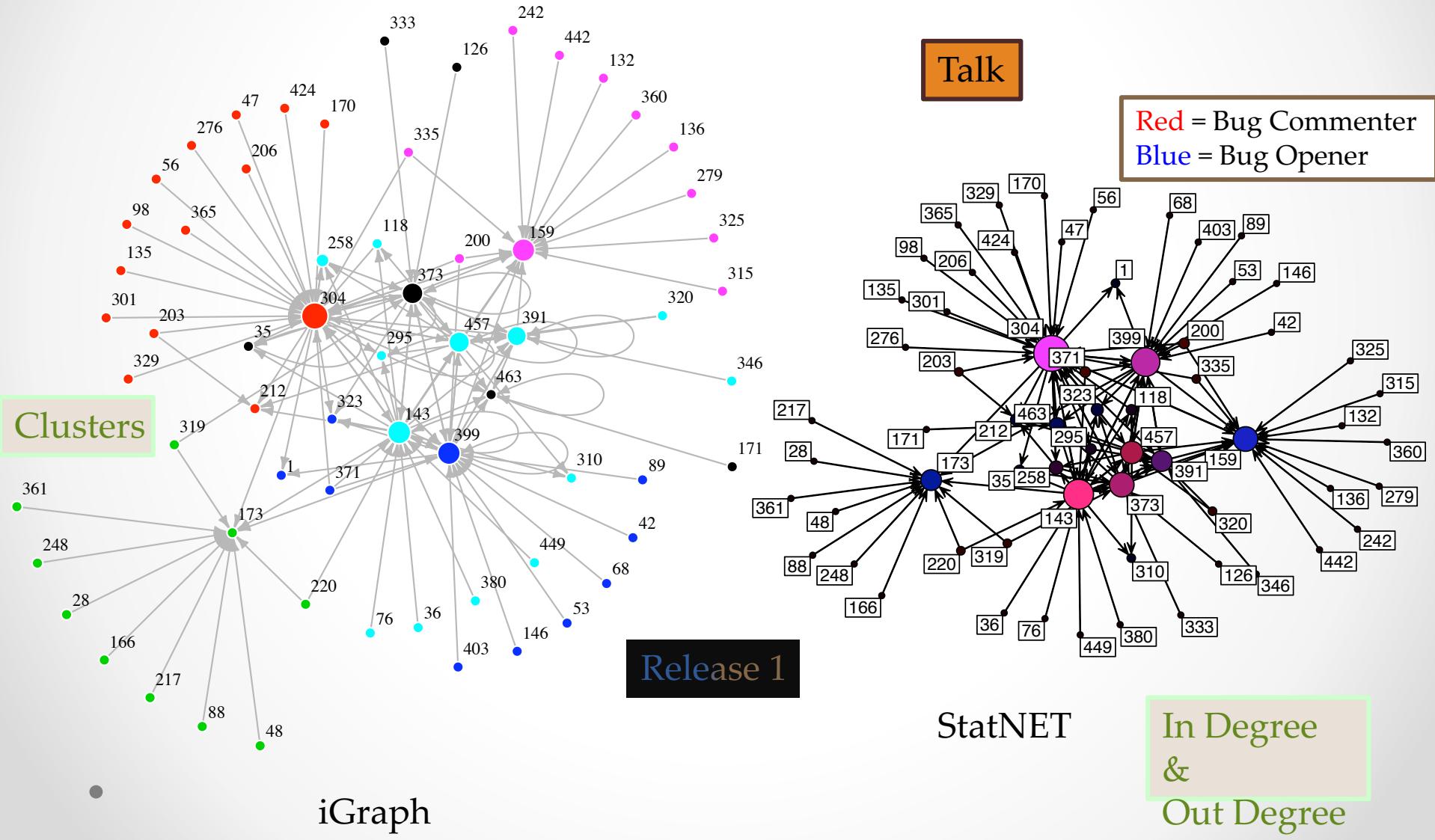
Release One

• • •

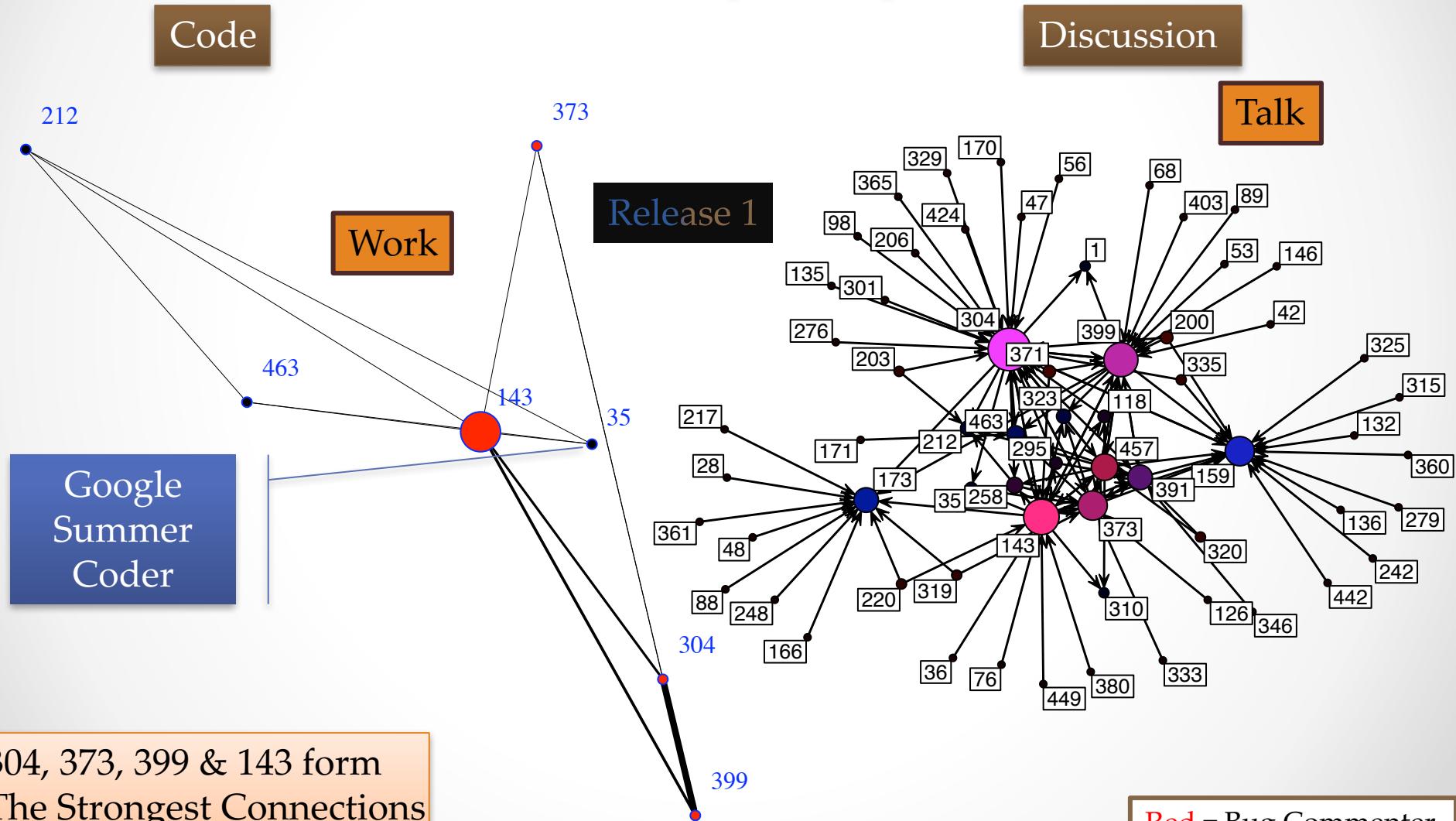
Work & Talk



Release 1 (2.0) iGraph & Statnet



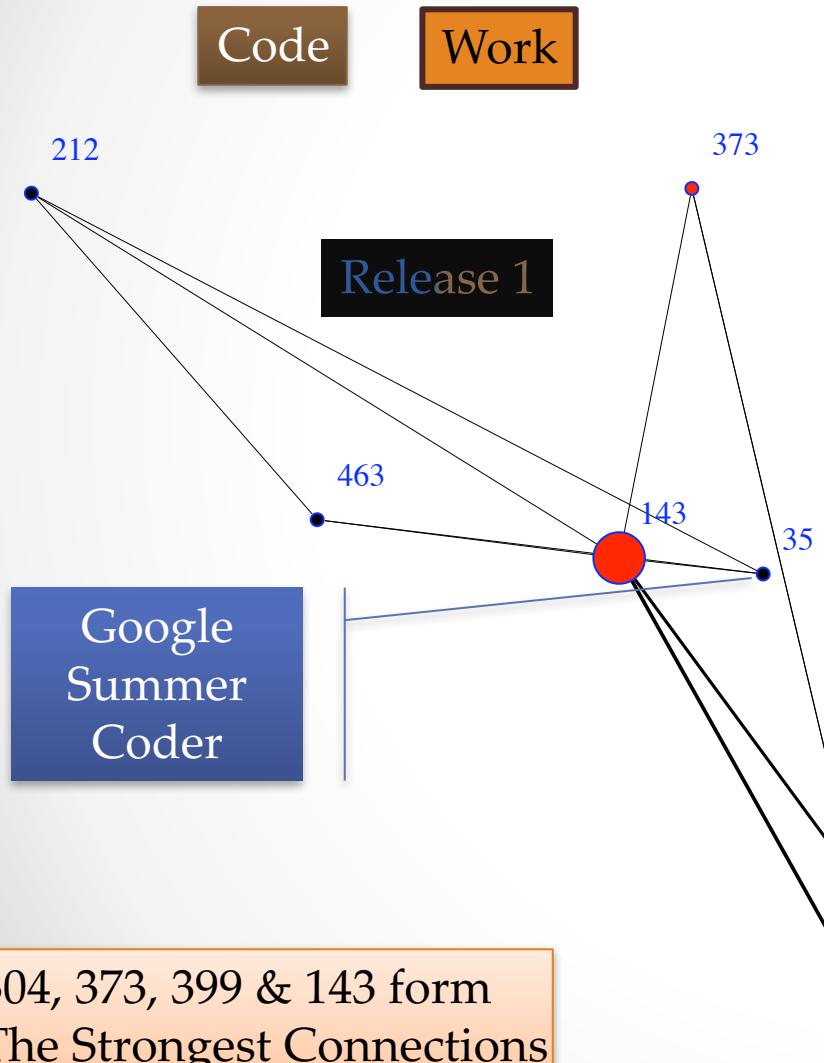
Release One (2.0): Filtered



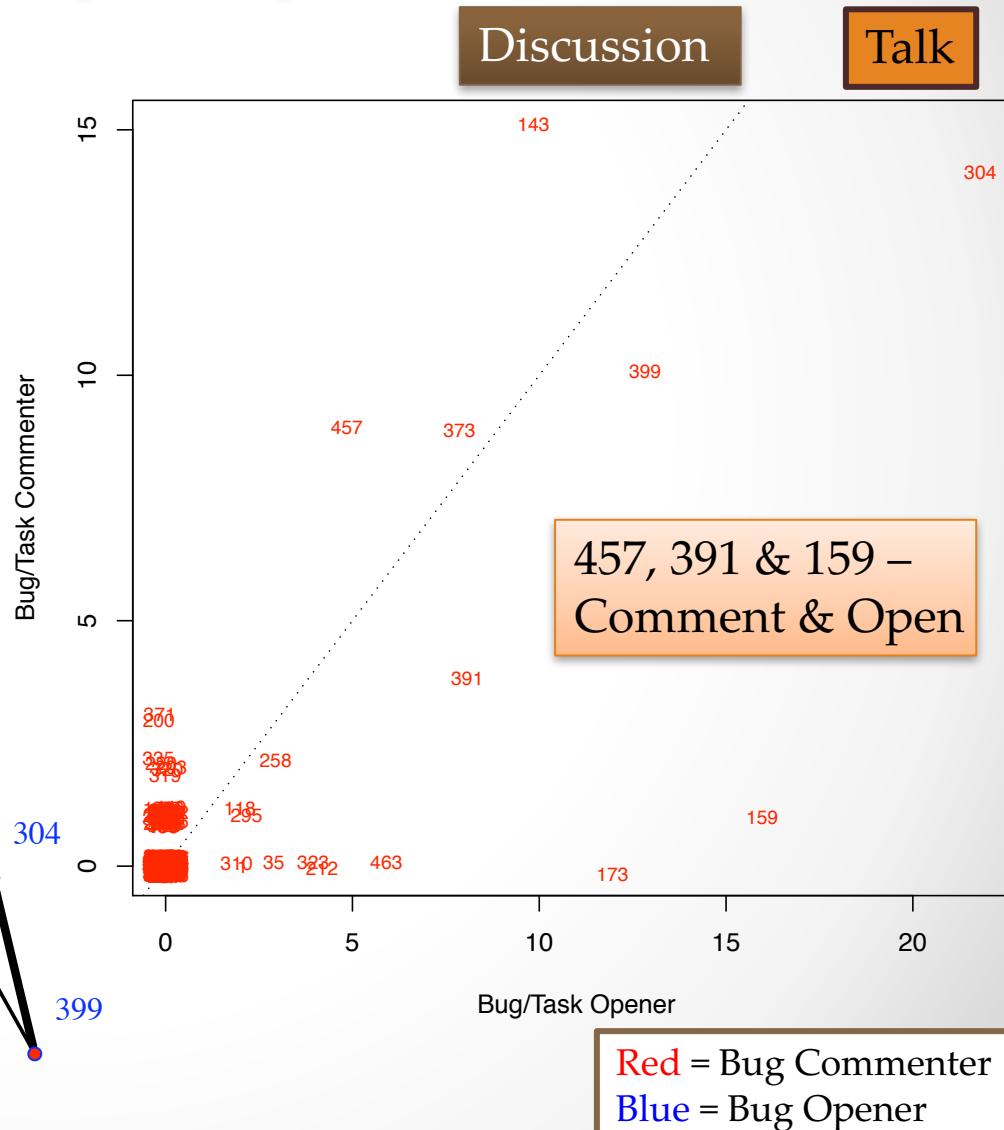
304, 373, 399 & 143 form
The Strongest Connections
In both networks

Red = Bug Commenter
Blue = Bug Opener

Release One (2.0): Filtered



304, 373, 399 & 143 form
The Strongest Connections
In both networks

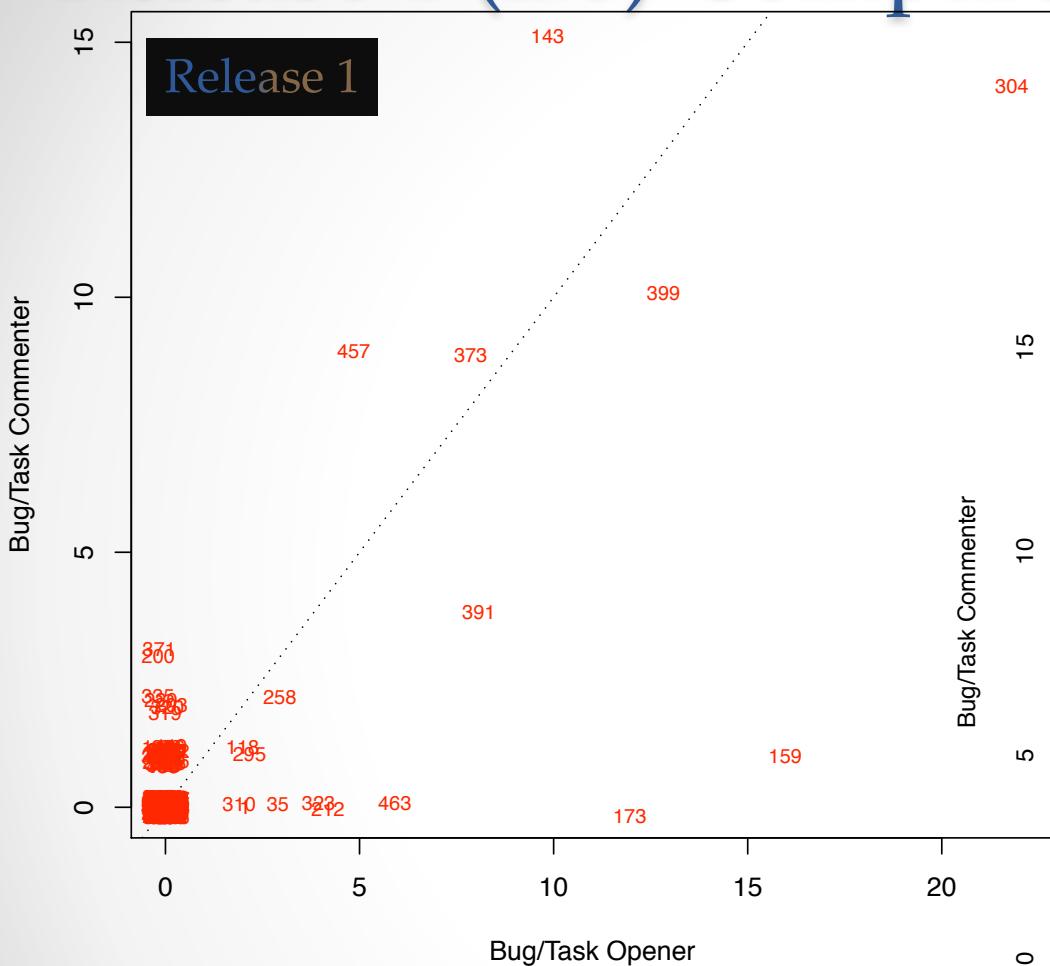


Compare Over Time

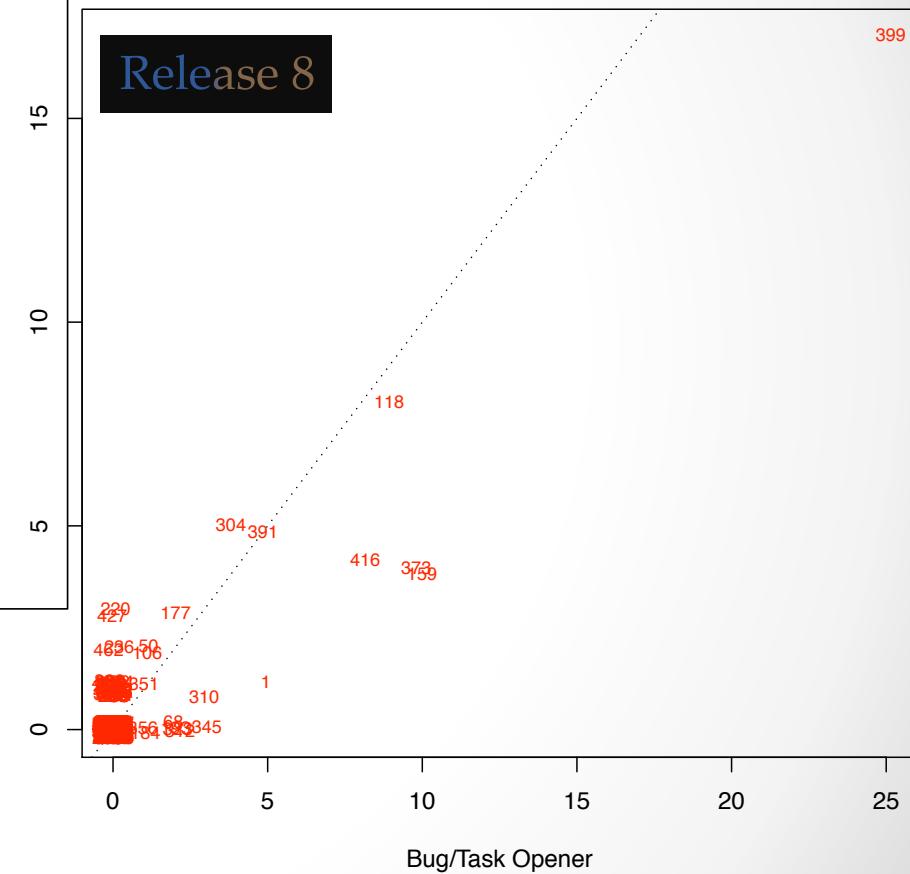
• • •

First & Last Release

Release 1 (2.0) Compared to Release 8 (3.3)

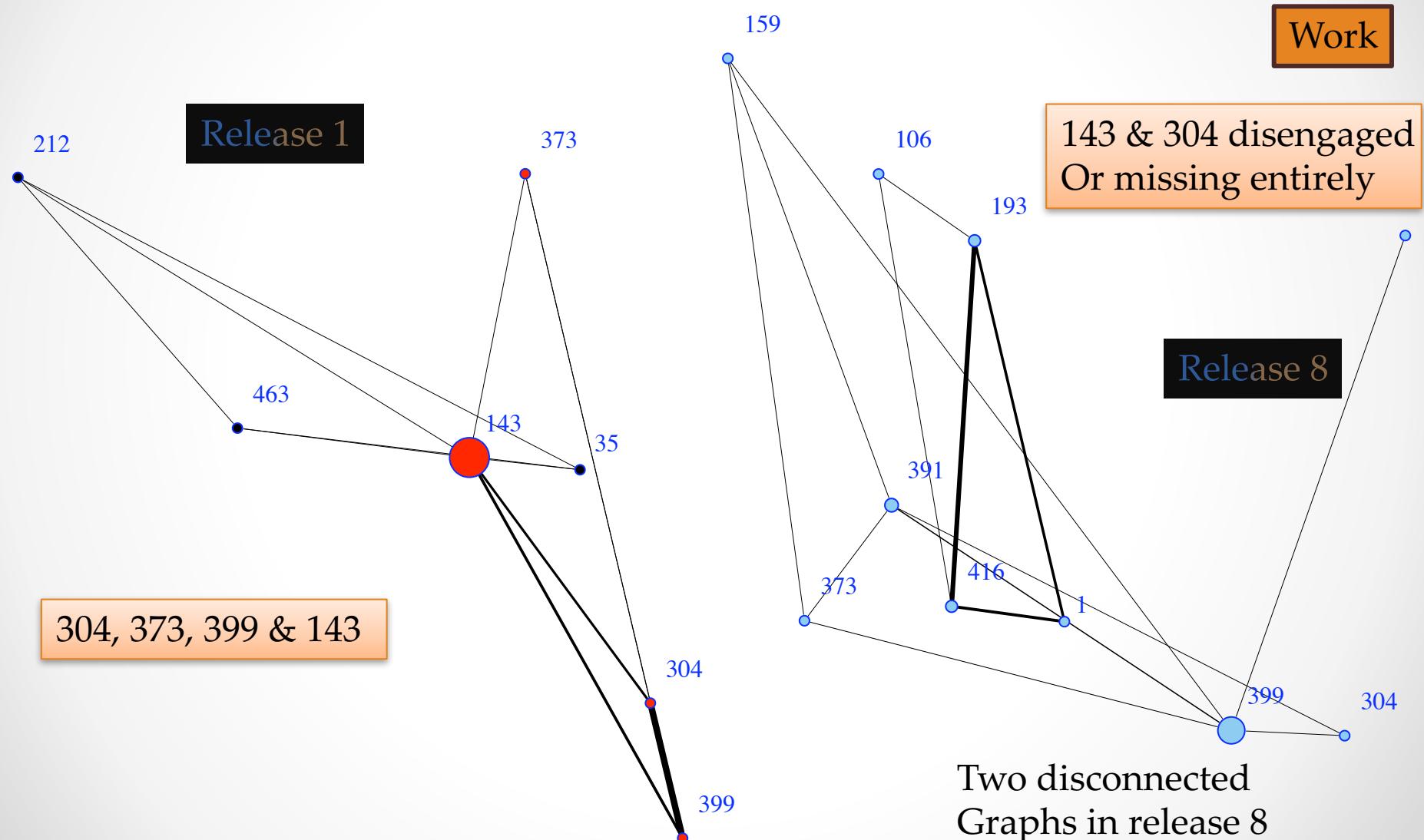


304, 399, 143, 159, 173, 373



399, 118, 304, 159, 391, 416

Release 1 (2.0) Compared to Release 8 (3.3)

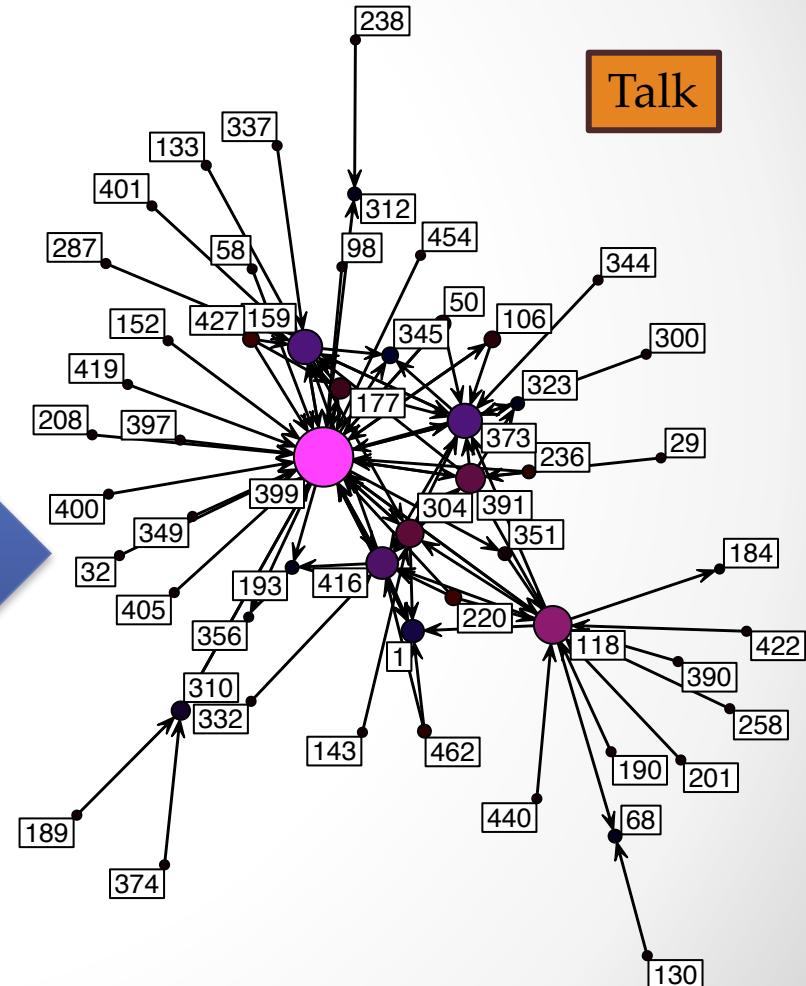
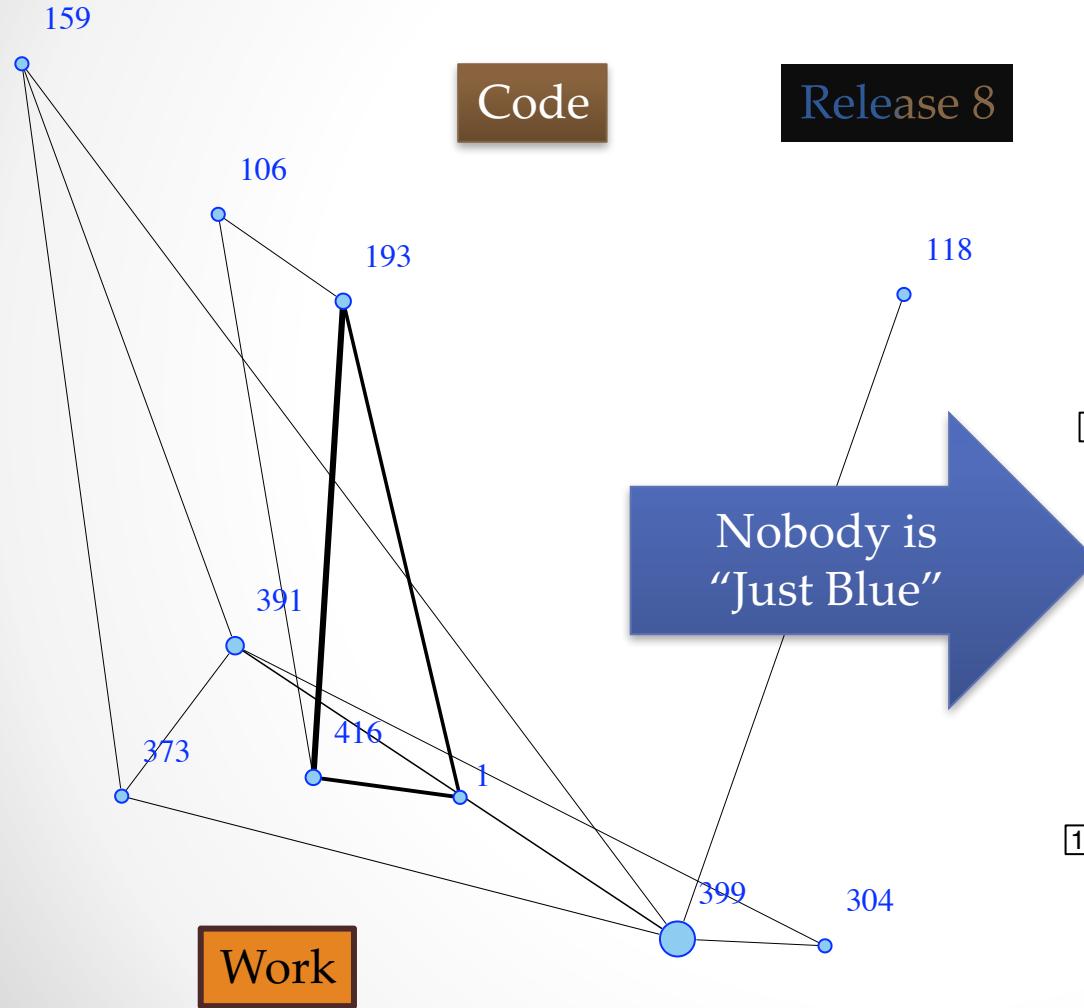


Release Eight

• • •

Work & Talk

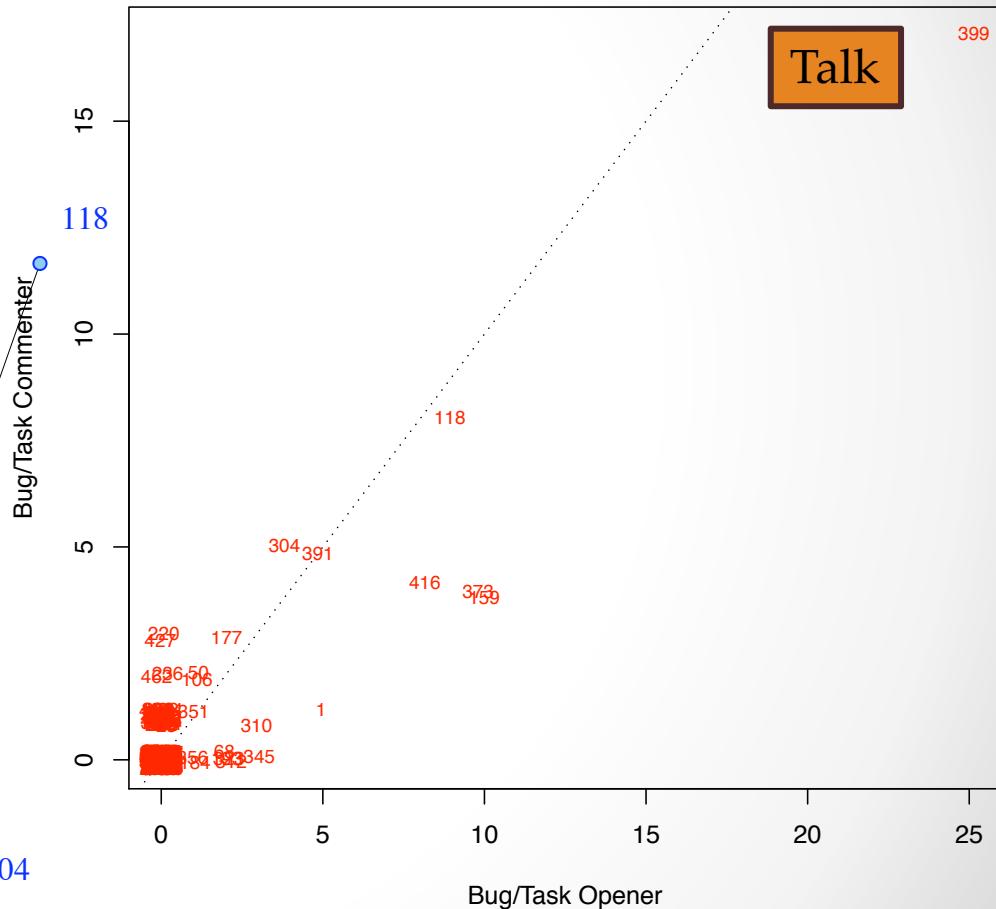
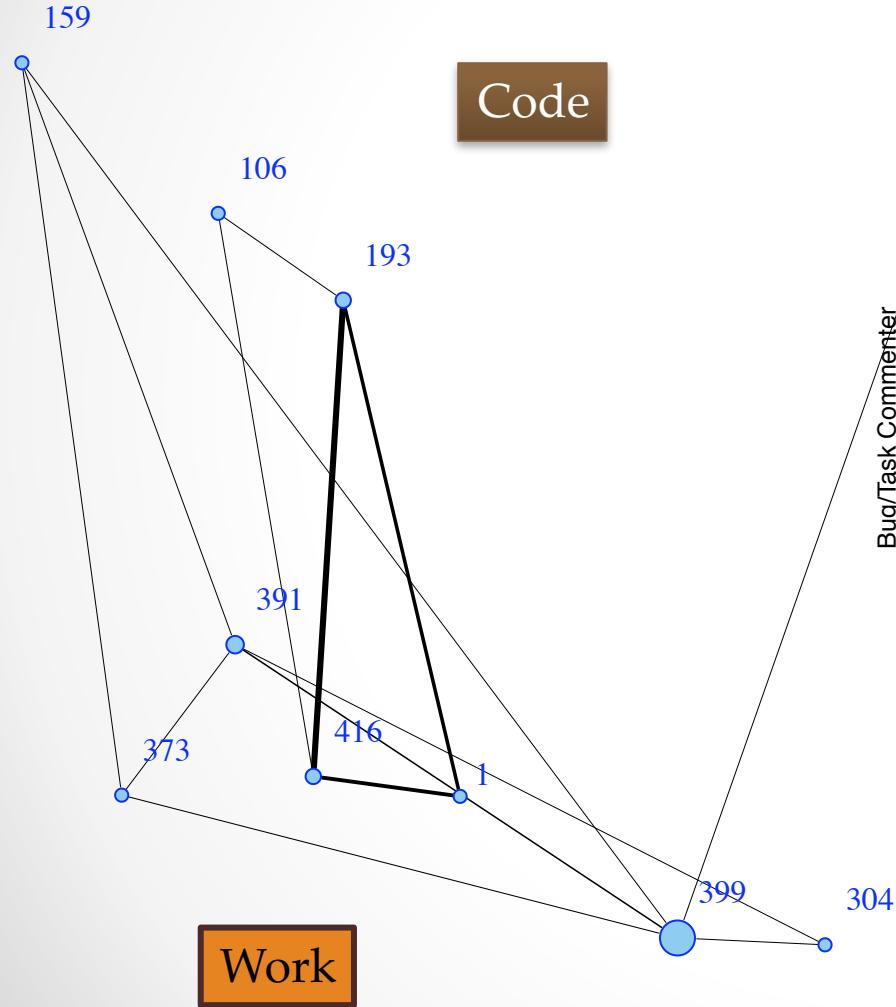
Release 8 (3.3): Filtered



Release 8 (3.3): Filtered

Release 8

Discussion

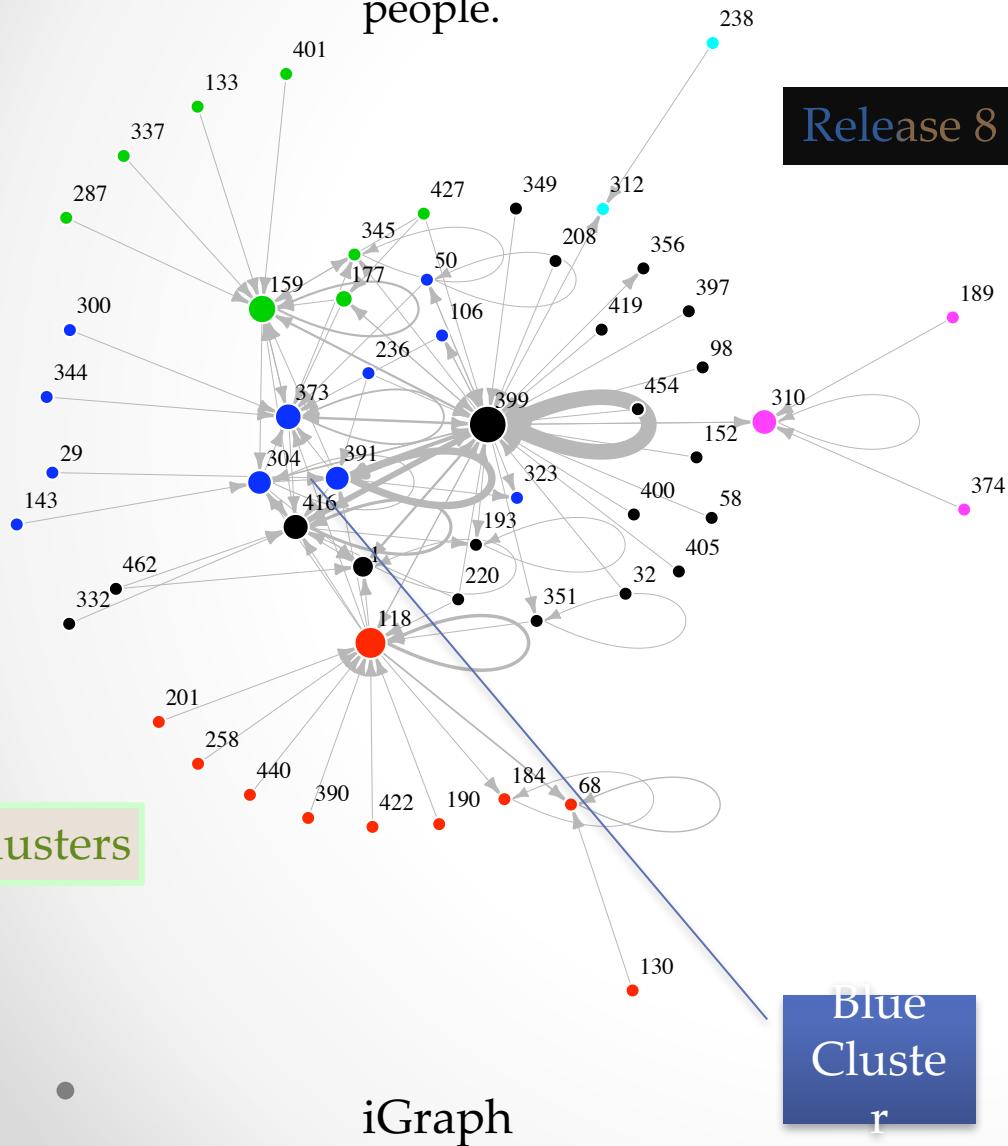


Notice 416 in Talk & Second Coder Graph

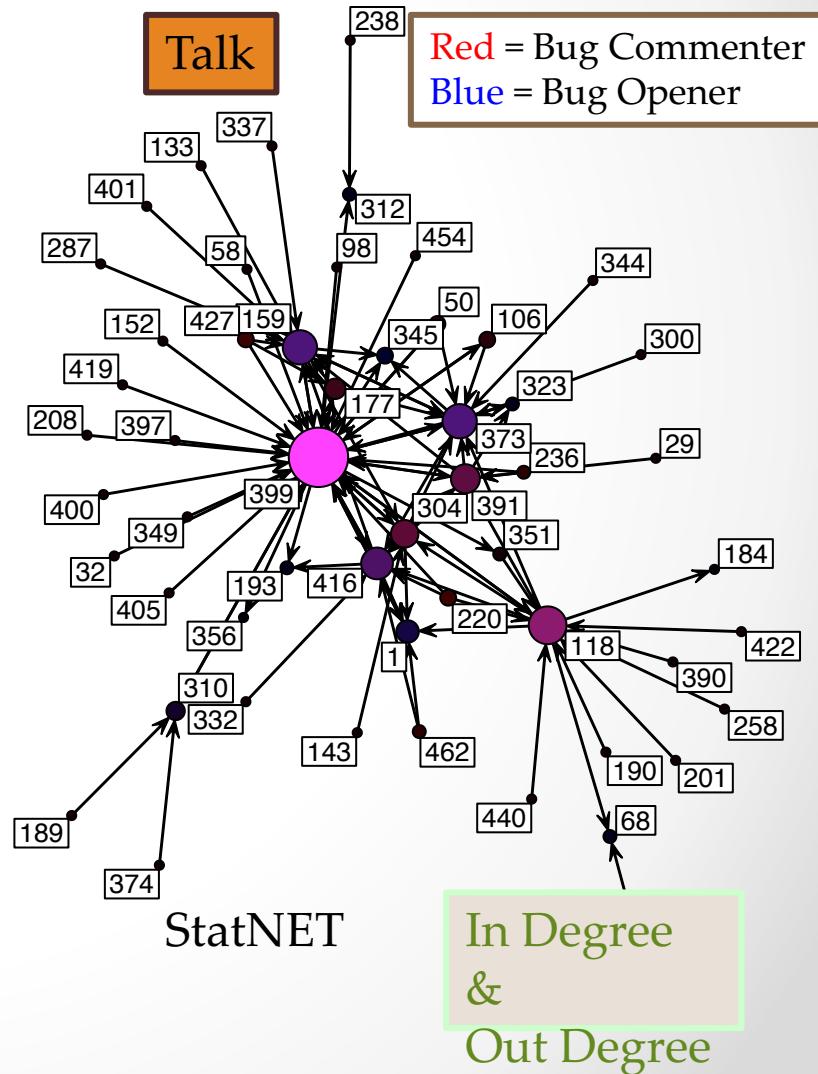
Red = Bug Commenter
Blue = Bug Opener

Release 8 (3.3) iGraph & Statnet

399, 118 & 159 are significant, But play with different clusters of Other people.



Release 8

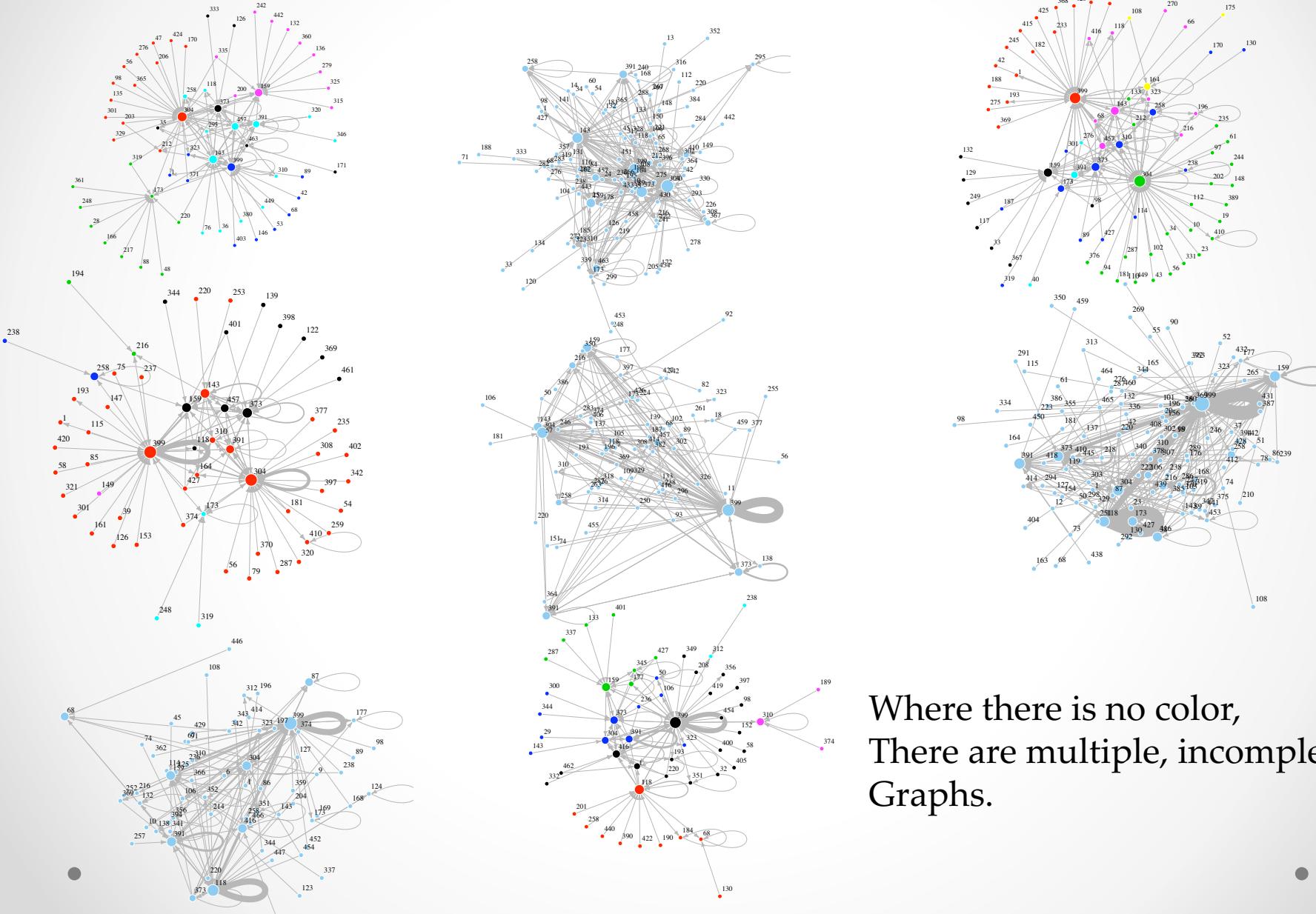


Releases One → Eight

• • •

High Level Views Over Time

Discussion, Releases 1 – 8



Where there is no color,
There are multiple, incomplete
Graphs.

Code, Releases 1 – 8



One Possible explanation:
A few central
People who slowly but
Observably begin to engage
Other contributors in
An open source software
Development project.

Structure evolves
Key Groups Evolve

Github: Longitudinal



Github Statnet Summaries

person	inDegree	outDegree	betweenness	totalDegree	timeperiod	timeperiodN
amatsuda	4	2	66	6	all-fixed.csv	0
Arsen7	1	1	0	2	all-fixed.csv	0
arunagw	1	1	0	2	all-fixed.csv	0
bogdan	8	4	228	12	all-fixed.csv	0
bradediger	1	1	0	2	all-fixed.csv	0
butcher	4	1	24	5	all-fixed.csv	0
carlosantonio	1	2	0	3	all-fixed.csv	0
dasch	2	4	100	6	all-fixed.csv	0
dmathieu	6	4	173	10	all-fixed.csv	0
fxn	0	2	0	2	all-fixed.csv	0
guilleiguaran	9	16	235.666667	25	all-fixed.csv	0
idoru	0	1	0	1	all-fixed.csv	0
isaacsanders	7	8	237.333333	15	all-fixed.csv	0
jasonnoble	0	1	0	1	all-fixed.csv	0
jeremy	1	1	0	2	all-fixed.csv	0

```
inDegree <- rbind()
outDegree <- rbind()
betweenness <- rbind()
totalDegree <- rbind()

fileListA <- data.frame(txt=rep("", N), txt=rep("", N), num=rep(NA, N),
  num=rep(NA, N), num=rep(NA, N), num=rep(NA, N), txt=rep("",N), num=rep(NA, N), # as many cols as you need
  stringsAsFactors=FALSE)
colnames(fileListA)[2] <- 'person'
colnames(fileListA)[3] <- 'inDegree'
colnames(fileListA)[4] <- 'outDegree'
colnames(fileListA)[5] <- 'betweenness'
colnames(fileListA)[6] <- 'totalDegree'
colnames(fileListA)[7] <- 'timeperiod'
colnames(fileListA)[8] <- 'timeperiodNum'

for (j in 1:N)
{
  degree(n)
  names <- network.vertex.names(n)[j]
  idegA <- degree(n, nodes=j, cmode="indegree")
  odegA <- degree(n, nodes=j, cmode="outdegree")
  between <- betweenness(n, nodes=j)
  totDegree <- degree(n, nodes=j)
  timePeriod <- infile
  timePeriodNum <- lfp-1
  fileListA[j,] <- c(N[j],names,idegA,odegA,between,totDegree,timePeriod,timePeriodNum)

}
namerDude <- paste("output/longMatrix",lfp, infile,".csv")
write.csv(file=namerDude, fileListA)
namerTwo <- paste("output/allStuff", infile, ".csv")

if (lfp-1==0){
  write.table(file=namerTwo, fileListA, sep=",", row.names=FALSE, col.names=TRUE)
}
else
  write.table(file=namerTwo, fileListA, sep = ",", append=TRUE,col.names=FALSE, row.names=FALSE)
```

Built with This R Code

Longitudinal Network Libraries

```
| library(network)
| library(statnet)
| library(MASS)
| library(ggplot2)
| library(Hmisc)
| library(reshape)
| library(Matrix)
| library(Scales)
```

GGPLOT2 Code

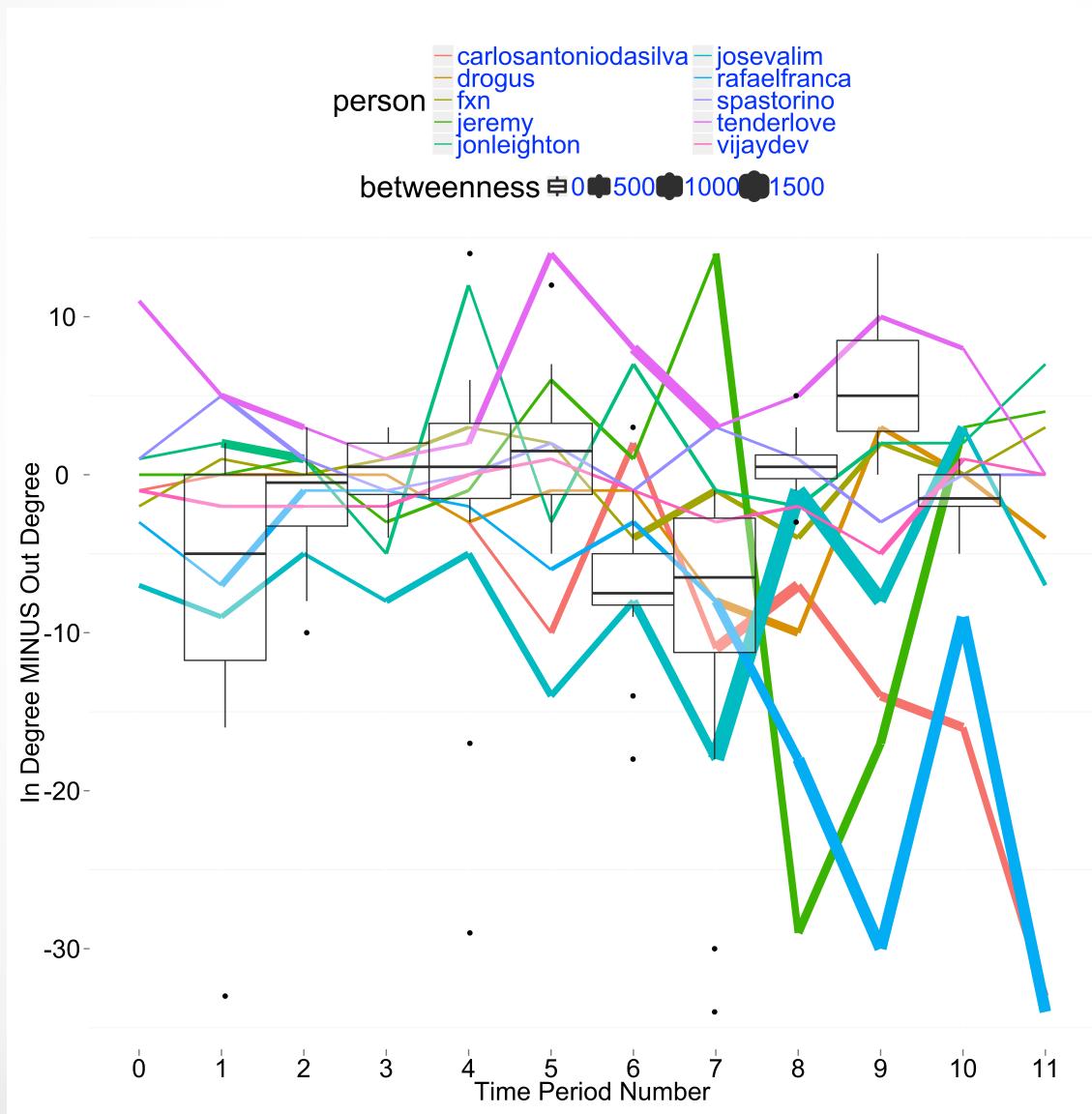
```
pc <- ggplot(gMerged, aes(timeperiodNum, inDegree-outDegree, group=person, size=betweenness))
pc <- pc + geom_line(aes(colour = person)) + guides(colour=guide_legend(nrow=7)) + opts(legend.text =
theme_text(colour="blue", size=20), legend.title=theme_text(size=24), legend.position="top", axis.text.x=theme_text(size=20),
axis.text.y=theme_text(size=20), axis.title.x = theme_text(size=20), axis.title.y = theme_text(size=20, angle=90),
panel.background = theme_rect(fill = "white", colour = NA)) +ylab("In Degree MINUS Out Degree")+xlab("Time Period Number")

ggsave(paste("pubtweak/",filez[kk], subgroup, "MEASURE ROLES - InDegree-OutDegree difference exploratory graph by
time-WITHNAMES + LineThickness.png"), dpi=600, plot=pc, width=12, height=12)
pc <- pc + geom_boxplot(width=.4, alpha=.3)
ggsave(paste("pubtweak/",filez[kk], subgroup, "Boxplot - MEASURE ROLES - InDegree-OutDegree difference exploratory
graph by time-WITHNAMES + LineThickness.png"), dpi=600, plot=pc, width=12, height=12)
##### A plot using points with the same data as the section with betweenness
```

```
pc <- ggplot(gMerged, aes(timeperiodNum, inDegree-outDegree, group=person,
size=betweenness))
```

```
pc <- pc + geom_line(aes(colour = person)) +
guides(colour=guide_legend(nrow=7)) + opts(legend.text =
theme_text(colour="blue", size=20),
legend.title=theme_text(size=24), legend.position="top", axis.text.x=theme_text(size=20),
axis.text.y=theme_text(size=20), axis.title.x = theme_text(size=20), axis.title.y = theme_text(size=20, angle=90),
panel.background = theme_rect(fill = "white", colour = NA)) +ylab("In Degree MINUS Out Degree")+xlab("Time Period Number")
```

Github Longitudinal



GGPLOT2 Code

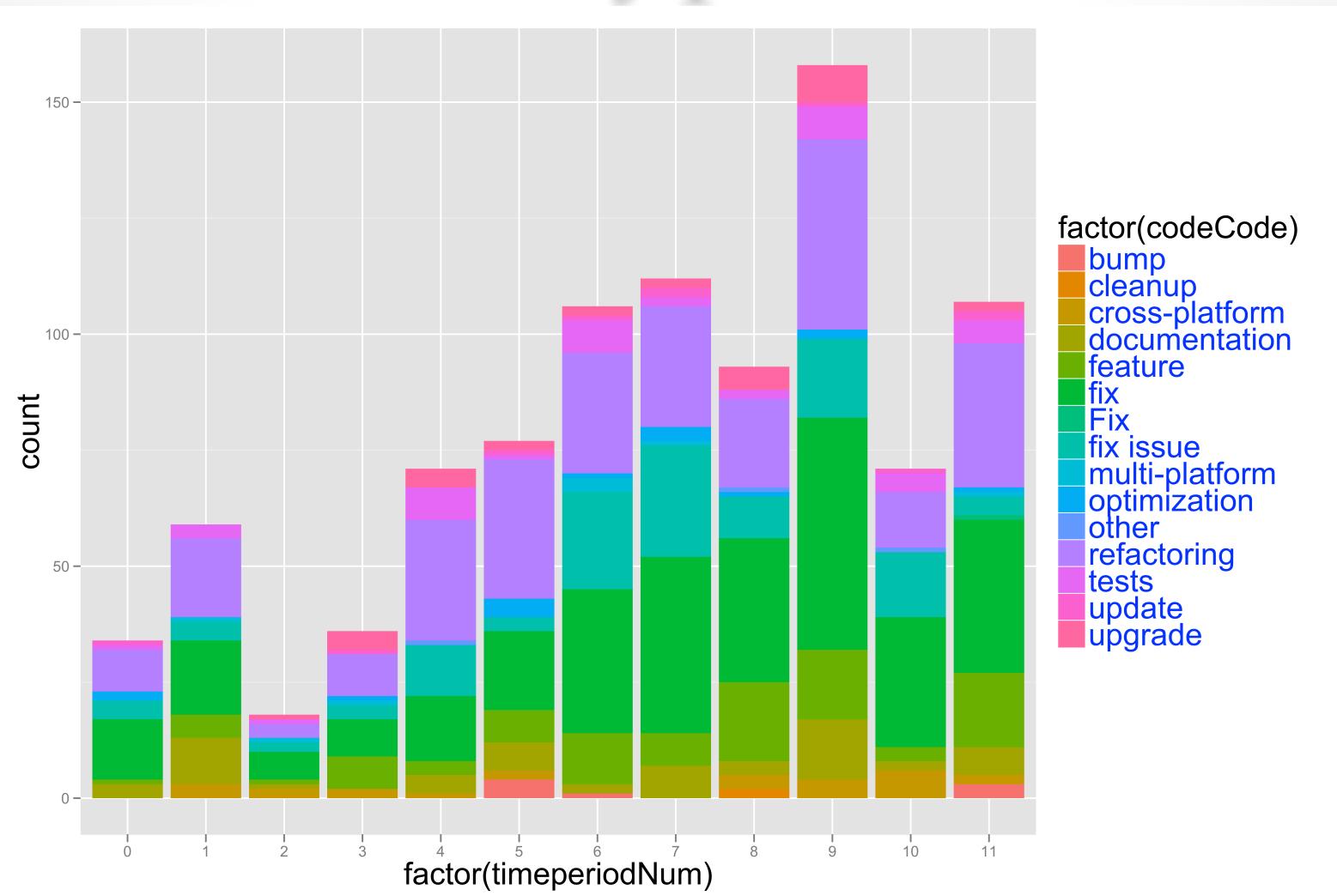
```
pc <- ggplot(gMerged, aes(timeperiodNum, inDegree-outDegree, group=person, size=betweenness))
pc <- pc + geom_line(aes(colour = person)) + guides(colour=guide_legend(nrow=7)) + opts(legend.text =
theme_text(colour="blue", size=20), legend.title=theme_text(size=24), legend.position="top", axis.text.x=theme_text(size=20),
axis.text.y=theme_text(size=20), axis.title.x = theme_text(size=20), axis.title.y = theme_text(size=20, angle=90),
panel.background = theme_rect(fill = "white", colour = NA)) +ylab("In Degree MINUS Out Degree")+xlab("Time Period Number")

ggsave(paste("pubtweak/",filez[kk], subgroup, "MEASURE ROLES - InDegree-OutDegree difference exploratory graph by
time-WITHNAMES + LineThickness.png"), dpi=600, plot=pc, width=12, height=12)
pc <- pc + geom_boxplot(width=.4, alpha=.3)
ggsave(paste("pubtweak/",filez[kk], subgroup, "Boxplot - MEASURE ROLES - InDegree-OutDegree difference exploratory
graph by time-WITHNAMES + LineThickness.png"), dpi=600, plot=pc, width=12, height=12)
##### A plot using points with the same data as the section with betweenness
```

```
pc <- ggplot(gMerged, aes(timeperiodNum, inDegree-outDegree, group=person,
size=betweenness))
```

```
pc <- pc + geom_line(aes(colour = person)) +
guides(colour=guide_legend(nrow=7)) + opts(legend.text =
theme_text(colour="blue", size=20),
legend.title=theme_text(size=24), legend.position="top", axis.text.x=theme_text(size=20),
axis.text.y=theme_text(size=20), axis.title.x = theme_text(size=20), axis.title.y = theme_text(size=20, angle=90),
panel.background = theme_rect(fill = "white", colour = NA)) +ylab("In Degree MINUS Out Degree")+xlab("Time Period Number")
```

GITHUB: Types of Work



Code & Data for Types of Work

```
gg <- ggplot(pullerListTyped, aes(timeperiodNum,  
fill=person))+geom_bar()+facet_grid(facets=codeCode~.)  
ggsave("fix-feature-pullers.png", dpi=600, plot=gg, width=12, height=12)
```

person	codeCode	Merged	PullRequest_Date	timeperiodN
bogdan	fix	Merged	11 months ago	0
bogdan	refactoring	Merged	11 months ago	0
bogdan	refactoring	Merged	11 months ago	0
amatsuda	refactoring	No	11 months ago	0
amatsuda	fix	Merged	11 months ago	0
amatsuda	refactoring	Merged	11 months ago	0
paneq	refactoring	No	11 months ago	0
guilleiguaran	refactoring	Merged	11 months ago	0
guilleiguaran	documentati	Merged	11 months ago	0
guilleiguaran	fix	No	11 months ago	0
guilleiguaran	fix	Merged	11 months ago	0
guilleiguaran	fix	Merged	11 months ago	0