DNM vs Alignment Comparison

Jake Prasad

Tuesday, July 28, 2015

Alignment (Verified/NonVerified):

Our standard measure for alignment uses Verified/Not verified as a power proxy. Using Verified/Not verified, we're able to clearly see that an influential speaker/uninfluential replier pair results in a stronger alignment than an uninfluential speaker/influential replier pair.

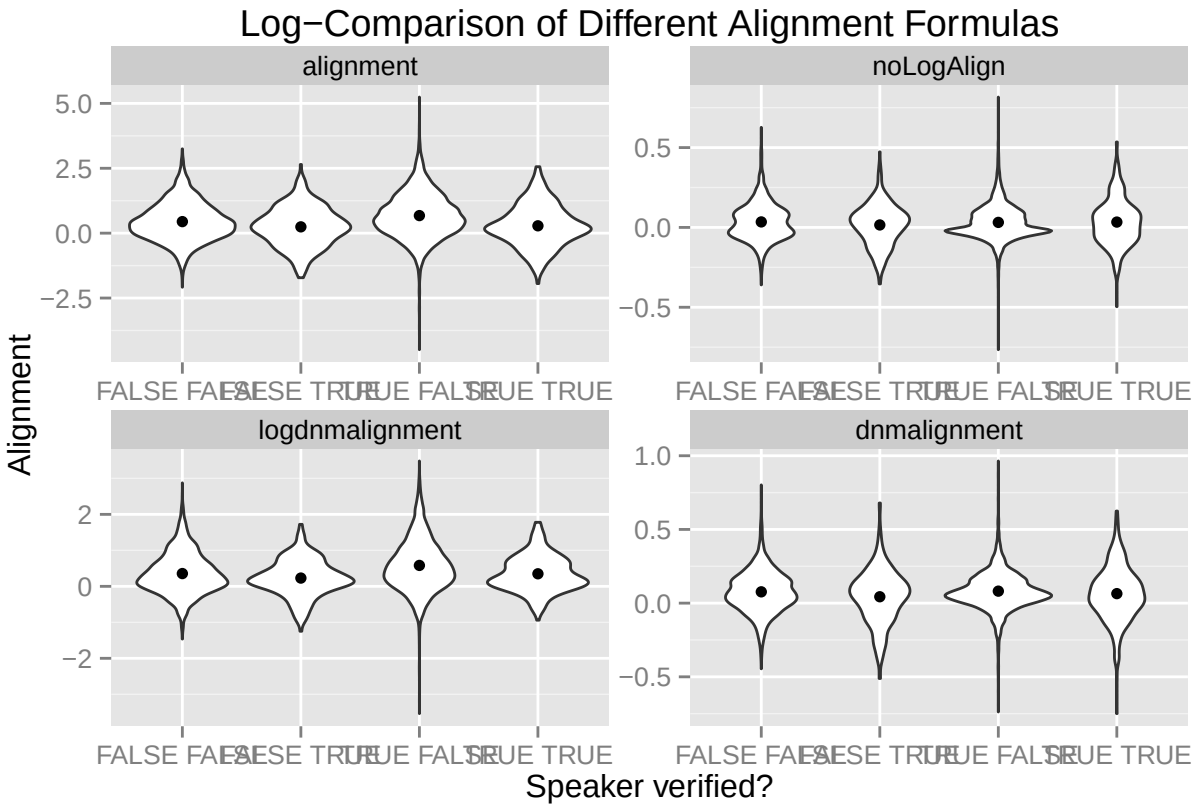
```
subsetted <- subset(df,logdnmalignment!="FALSE"&(ba+nba)>5&(bna+nbna)>5)
subsetted = transform(subsetted,logdnmalignment=as.numeric(logdnmalignment))
```

```
## Warning: NAs introduced by coercion
```

```
d2 <- subsetted %>%
  group_by(verifiedSpeaker,speakerId,replierId) %>%
  summarize(convs=n(), alignment=alignment, vreply=verifiedReplier, dnmalignment=dnmalignment, noLogAlig
  gather(alignmentType,alignmentValue,c(alignment, dnmalignment, logdnmalignment, noLogAlign))

levels(d2$alignmentType)=c("alignment","noLogAlign","logdnmalignment","dnmalignment")

ggplot(d2,aes(x=paste(verifiedSpeaker,vreply),y=alignmentValue)) + geom_violin() + labs(title="Log-Compa
```

```
## Warning: Removed 3309 rows containing non-finite values (stat_ydensity).
```

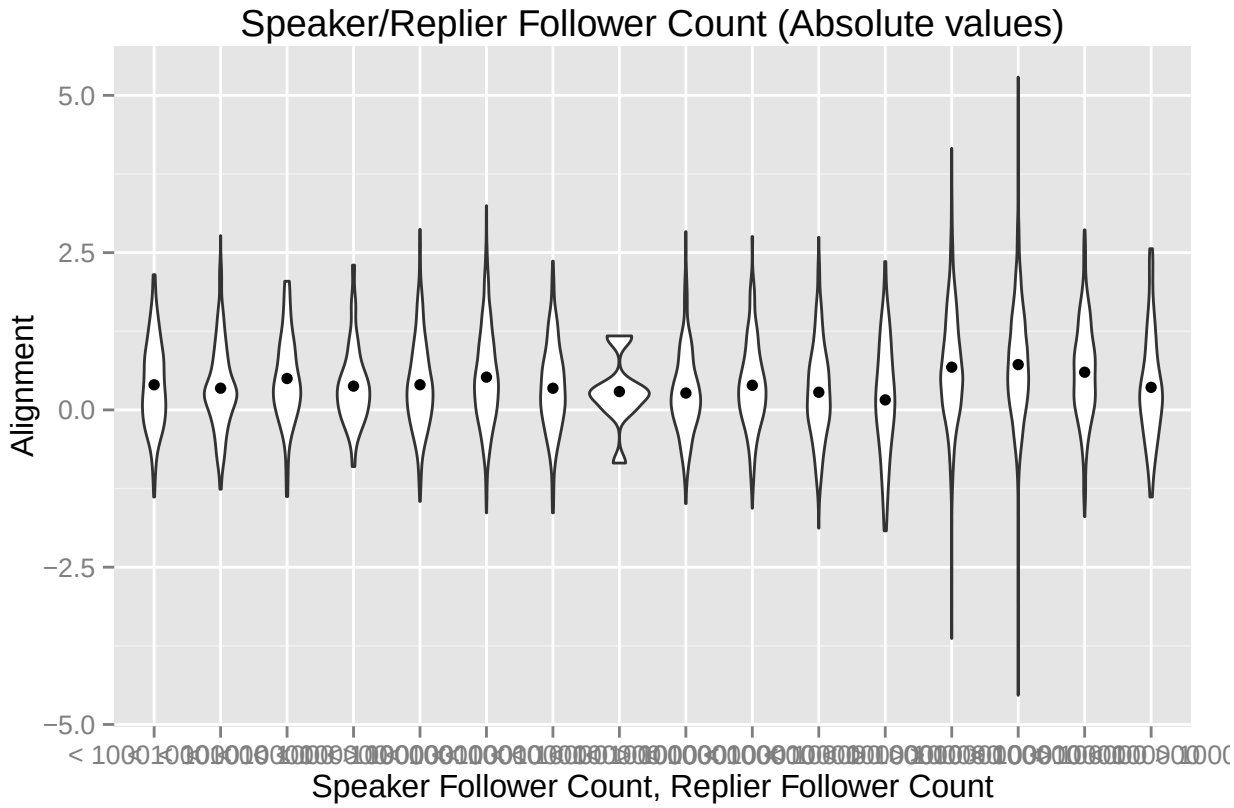
```
## Warning: Removed 3309 rows containing missing values (stat_summary).
```



Follower Bins:

The Echoes Of Power paper used follower count as a power proxy. If follower count isn't a good power proxy it could explain why alignment wasn't found. We bin follower counts for speaker and repliers and plot the results.

```
d2 <- subset(df,logdnmalignment!="FALSE"&(ba+nba)>5&(bna+nbna)>5)
d2$speakerBins <- cut(d2$speakerFollowers, breaks=c(0,1000, 10000, 100000, 100000000), labels=c("< 1000
d2$replierBins <- cut(d2$replierFollowers, breaks=c(0,1000, 10000, 100000, 100000000), labels=c("< 1000
ggplot(d2,aes(x=paste(speakerBins,replierBins),y=alignment)) + geom_violin() + stat_summary(geom="point
```

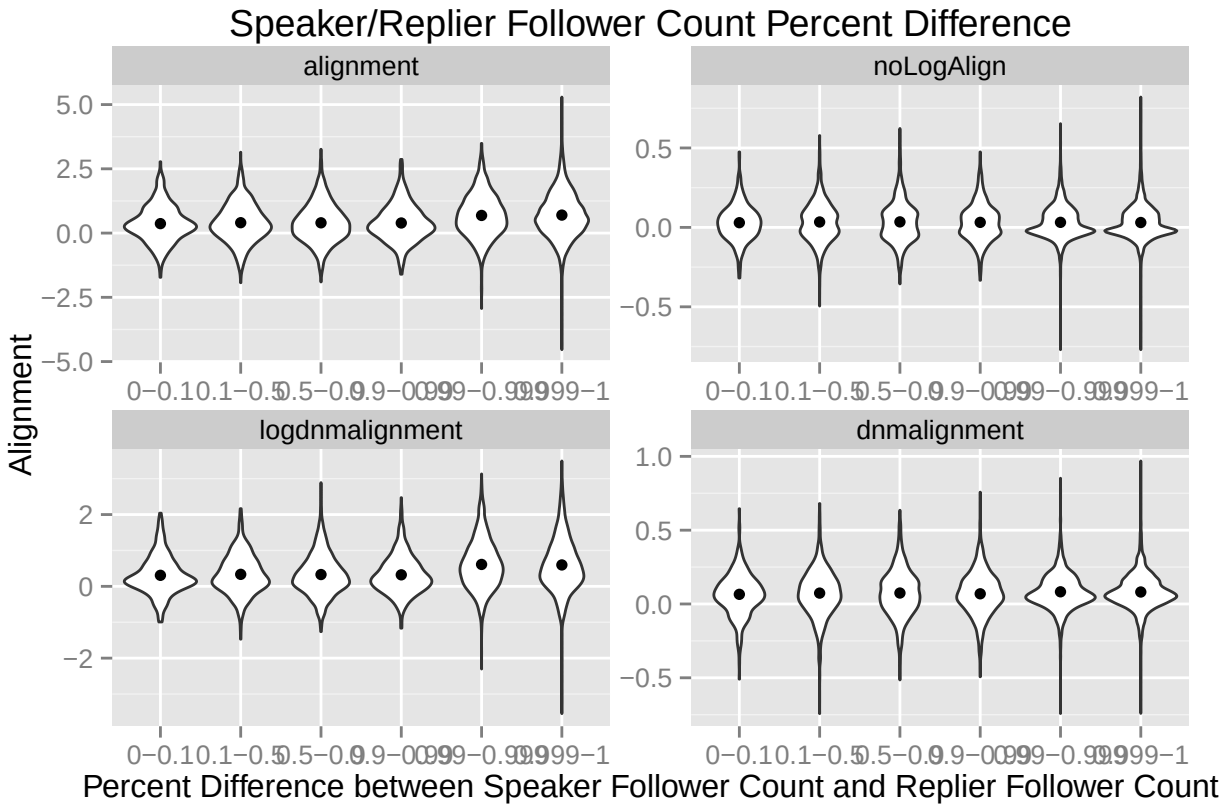


This is an interesting plot. It appears that follower count does have a strong effect on alignment. Let's also try plotting the percent difference between speaker follower count and replier follower count.

```
d2 <- subset(df,logdnmalignment!="FALSE"&(ba+nba)>5&(bna+nbna)>5)
subsetted = transform(subsetted,logdnmalignment=as.numeric(logdnmalignment))
d2 <- subsetted %>%
  group_by(verifiedSpeaker,speakerId,replierId) %>%
  summarize(convs=n(), alignment=alignment, vreply=verifiedReplier, dnmalignment=dnmalignment, noLogAlig
  gather(alignmentType,alignmentValue,c(alignment, dnmalignment, logdnmalignment, noLogAlign))

d2$followerBins <- cut(d2$percentDiff, breaks=c(0,0.1,0.5,0.9, 0.99, 0.999, 1), labels=c("0-0.1","0.1-0

levels(d2$alignmentType)=c("alignment","noLogAlign","logdnmalignment","dnmalignment")

ggplot(d2,aes(x=followerBins,y=alignmentValue)) + geom_violin() + labs(title="Speaker/Replier Follower
```

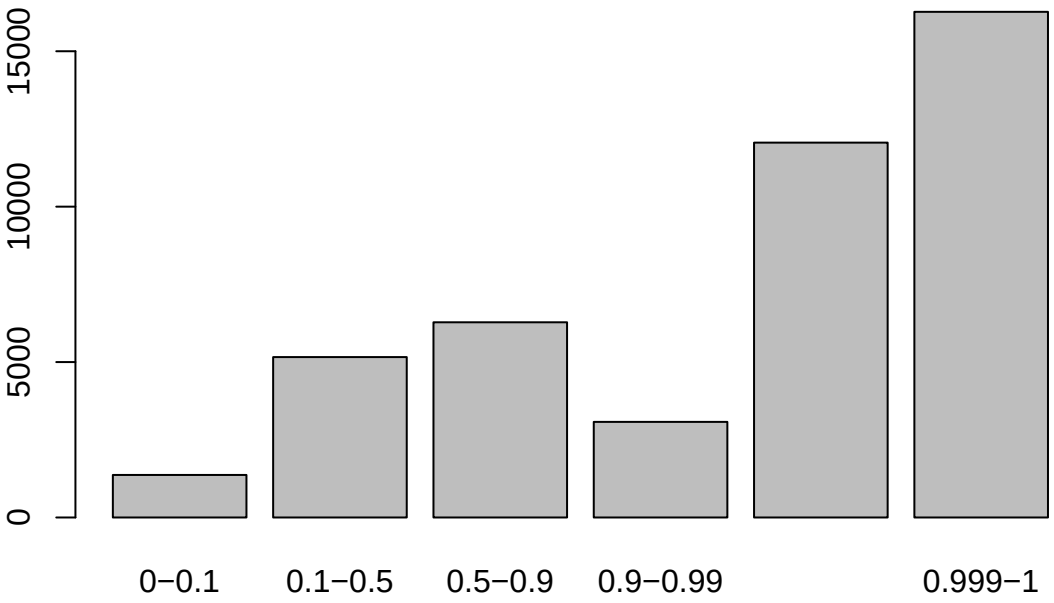
```
## Warning: Removed 3309 rows containing non-finite values (stat_ydensity).
```

```
## Warning: Removed 3309 rows containing missing values (stat_summary).
```



This is unexpected. It appears that we only see power alignment when the speaker/replier follower percent difference >0.99. Could this be because of sparsity?

```
barplot(height=table(d2$followerBins))
```



It seems that there is some sparsity but 0.1-0.5 and 0.5-0.9 have more than 5000 values each, so it may be another problem