

DNM vs Alignment Comparison

Jake Prasad

Tuesday, July 28, 2015

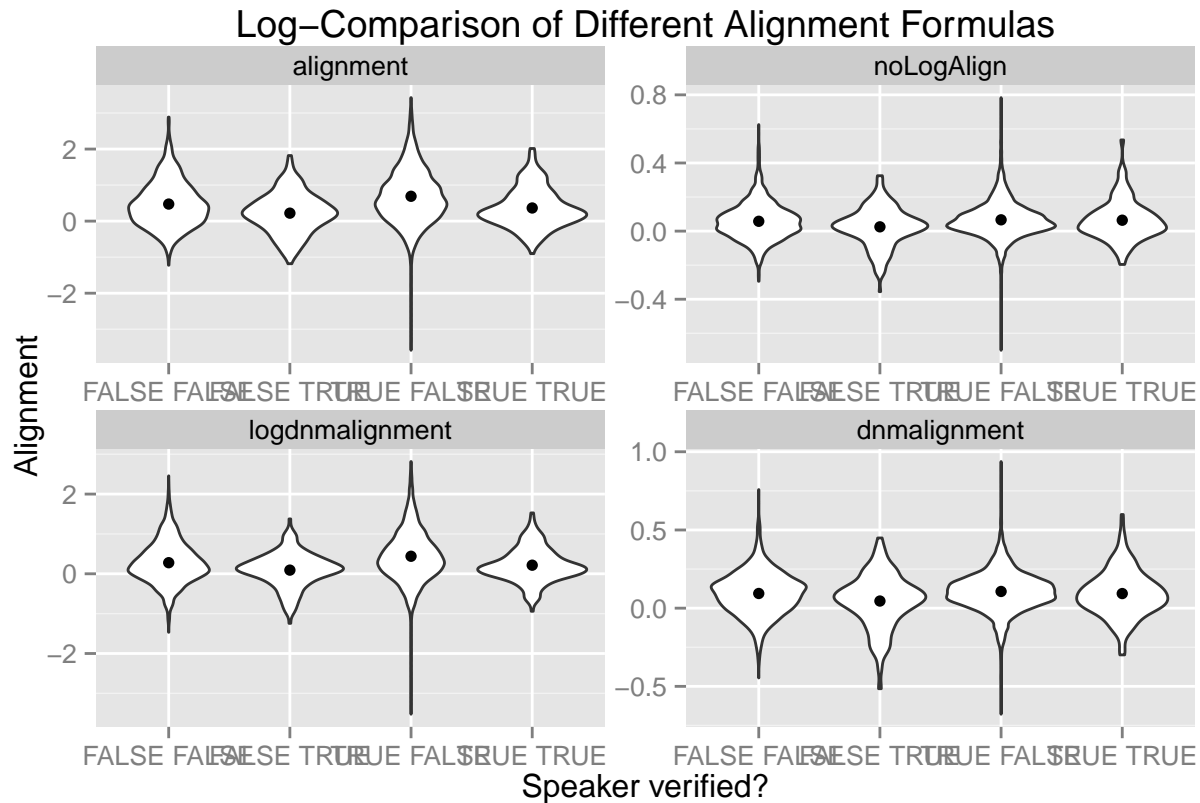
Alignment (Verified/NonVerified):

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.

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

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

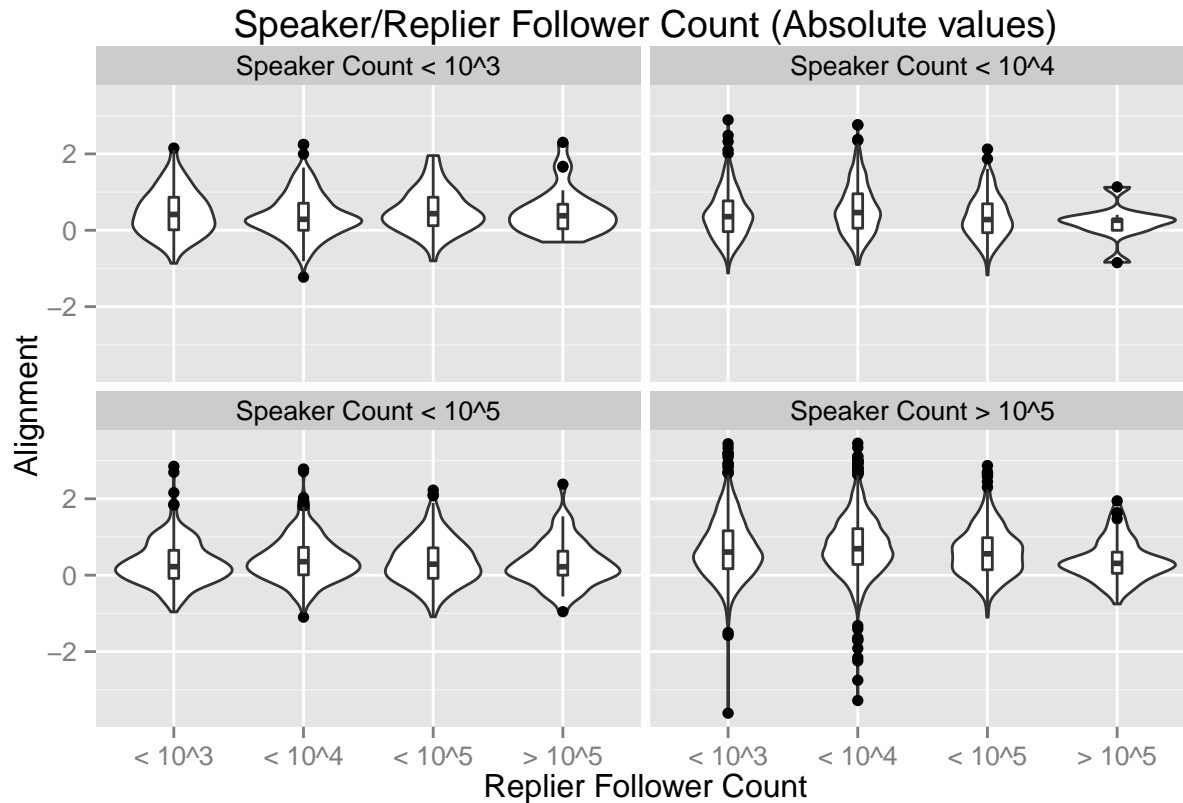
ggplot(d2, aes(x=paste(verifiedSpeaker, vreply), y=alignmentValue)) + geom_violin() + labs(title="Log-Comparison of Different Alignment Formulas")
```



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("Speaker",
d2$replierBins <- cut(d2$replierFollowers, breaks=c(0,1000, 10000, 100000, 100000000), labels=c("< 10^3",
ggplot(d2, aes(x=paste(replierBins), y=alignment)) + geom_violin() + labs(font=10, title="Speaker/Replier
```



This is an interesting plot. It appears that follower count does have an 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)
subsetting = transform(subsetting, logdnmalignment=as.numeric(logdnmalignment))
d2 <- subsetting %>%
  group_by(verifiedSpeaker, speakerId, replierId) %>%
  summarize(convs=n(), alignment=alignment, vreply=verifiedReplier, dnmalignment=dnmalignment, noLogAlign=
  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("<0.1", "<0.5",
levels(d2$alignmentType)=c("alignment", "noLogAlign", "logdnmalignment", "dnmalignment")

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

