

## Table of contents

<b>PART I: Defining Hypotheses to Test</b>	1
Hypotheses	1
Hypothesis 6:	2
<b>PART 2: Testing Hypotheses</b>	2
Model 1	2
Model 2	3
Computation of clusters and sentiments to be used in model 3	4
Model 3	5
Model 4	6
Model 5	6
Model 6	7
Model Comparisons	8

Rows: 286561 Columns: 6

-- Column specification -----

Delimiter: "\t"

chr (4): X1, X2, X3, X6

dbl (1): X5

dtm (1): X4

i Use `spec()` to retrieve the full column specification for this data.

i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

Rows: 571927 Columns: 6

-- Column specification -----

Delimiter: "\t"

chr (4): X1, X2, X3, X6

dbl (1): X5

dtm (1): X4

i Use `spec()` to retrieve the full column specification for this data.

i Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

## PART I: Defining Hypotheses to Test

### Hypotheses

#### Hypothesis 1:

The likelihood of a subreddit *i* linking to another subreddit *j* increases if (a) *i* has linked to *j* before and (b) if *j* has linked to *i* before. Measured by: *RSndSnd* and *RRecSnd* (repetition and reciprocity, respectively)  
*Echo chambers should have mutual reinforcement loops among in-group communities, so finding reciprocal linking should show whether this mutual reinforcement is present.*

## Hypotheses 2:

The likelihood of a subreddit *i* linking to subreddit *j* is greater if *j* has previously been linked to frequently by other subreddits. Measured by: NTDegRec *This measures preferential attachment, where nodes are more likely to connect with nodes who are already central to the network. Highly-linked subreddits may become larger centers of opinions, which could form the core of echo chambers*

## Hypothesis 3a & 3b:

3a: The likelihood of a subreddit *i* linking to subreddit *j* is greater if both subreddits are in the same cluster. Measured by: CovInt based on (to be computed) cluster covariate **DID NOT END UP TESTING - would have run for minimum of 4 hours to get output and Stephen got enough similar info**

3b: The likelihood of a subreddit *i* linking to subreddit *j* is greater if *i* and *j* have similar average sentiments. Measured by: CovInt with sentiment similarity *One of the key identifiers of echo chambers is communicating only with others of similar sentiments/opinions. This tells us whether subreddits are more likely to link to others with similar sentiments to help us understand the formation of echo chambers*

## Hypothesis 4:

When an subreddit *i* links to another subreddit *j*, *j* has a greater likelihood of linking back to subreddit *i* in the next event. Measured by: PSAB-BA *Quick back and forth linkage between pairs of subreddits reinforces insularity and short feedback loops, which are important elements of echo chambers*

## Hypothesis 5:

The likelihood of a subreddit *i* linking to a subreddit *k* is greater if *i* links to a third subreddit *j*, and *j* links to *k*. Measured by: PSAB-BY *In echo chambers, subreddits that propagate information/linkage to other subreddits with similar sentiments extend echo chamber narratives when they form information chains*

## Hypothesis 6:

The rate of subreddit *i* linking to other subreddits within its cluster increases over time. Measured by: CovSnd with time covariate *Burst dynamics are often visualized in the formation of echo chambers. A positive trend would suggest engagement going up quickly and intensely, while negative trends point to fading activity*

## PART 2: Testing Hypotheses

### Model 1

# actors (n): 49

# events (m): 2595

```
Prepping edgelist.
Checking/prepping covariates.
Computing preliminary statistics
Fitting model
Obtaining goodness-of-fit statistics
```

Relational Event Model (Ordinal Likelihood)

	Estimate	Std.Err	Z value	Pr(> z )
RRecSnd	9.7452e-01	7.4043e-02	13.161	<2e-16 ***
RSndSnd	3.0427e+00	4.6790e-02	65.029	<2e-16 ***
CovSnd.1	3.6289e-04	8.7381e+01	0.000	1

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
Null deviance: 40290.08 on 2595 degrees of freedom  
Residual deviance: 37113.97 on 2592 degrees of freedom  
Chi-square: 3176.109 on 3 degrees of freedom, asymptotic p-value 0  
AIC: 37119.97 AICC: 37119.98 BIC: 37137.56

*Testing Hypothesis 1: The likelihood of a subreddit i linking to another subreddit j increases if (a) i has linked to j before and (b) if j has linked to i before.*

A strong reciprocity effect here would be when if a subreddit j has recently linked to a subreddit i, i is more likely to link back to j - meaning that the links between i and j are bidirectional. RRecSnd is the statistic we can look at to judge the reciprocity effect: the p-value is far below 0.05, making this a statistically significant finding. The estimate (measurign how much this variable influences the likelihood of an event occurring) is positive and fairly large, showing a strong reciprocity effect. Since this analysis is on the network as a whole, without factoring in clustering, this just tells us that subreddits in general are likely to link back to those that linked to them. To detect echo chambers, we need to find selective reciprocity and determine whether reciprocity is disproportionately concentrated within clusters or distributed evenly across the network, which we'll do with future models in this report.

A strong repetition effect here is when subreddits are more likely to continue linking to other subreddits that they've linked to frequently in the past. The p-value of RSndSnd, the statistic we look at to judge the repetition effect, is far smaller than 0.05, making the results statistically significant. The estimate is positive and very high (3x RRecSnd) which indicates a very strong repetition effect. This tells us that subreddits have very strong interlinking patterns, where subreddits tend to link to others frequently over time.

## Model 2

```
Prepping edgelist.
Checking/prepping covariates.
Computing preliminary statistics
Fitting model
Obtaining goodness-of-fit statistics
```

#### Relational Event Model (Ordinal Likelihood)

```
      Estimate   Std.Err Z value  Pr(>|z|)
NTDegRec 1.1000e+01 4.7779e-01 23.022 < 2.2e-16 ***
RRecSnd  3.9332e-01 8.1770e-02  4.810 1.509e-06 ***
RSndSnd  2.5805e+00 5.2548e-02 49.107 < 2.2e-16 ***
CovSnd.1 6.3228e-04 9.0923e+01  0.000      1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null deviance: 40290.08 on 2595 degrees of freedom
Residual deviance: 36648.16 on 2591 degrees of freedom
Chi-square: 3641.923 on 4 degrees of freedom, asymptotic p-value 0
AIC: 36656.16 AICC: 36656.17 BIC: 36679.6
```

*Testing Hypotheses 2: The likelihood of a subreddit  $i$  linking to subreddit  $j$  is greater if  $j$  has previously been linked to frequently by other subreddits.*

The p-values of all the previous effects are statistically significant. We see positive coefficients for both the reciprocity and repetition effects.

The p-value of NTDegRec is well below 0.05 ( $p < 2.2e-16$ ), indicating the result is statistically significant. The estimate is  $1.1000e+01$  (~11), which is positive and very high. This means that subreddits that have already been frequently linked to are more likely to receive future links — a strong preferential attachment effect. The exponentiated value of this estimate is ~59874.14 which indicates that highly linked (popular) subreddits are far far more likely to receive future links than less linked subreddits. In the context of echo chambers, this suggests that a small number of highly central subreddits may dominate the flow of links, reinforcing central opinion hubs, which is a trademark of echo chamber formation.

### Computation of clusters and sentiments to be used in model 3

```
#| label: sentiment-covar
#| echo: false

# Average sentiment (example uses PROP_21)
sentiment_estimates <- combined_data_top %>%
  group_by(SOURCE_SUBREDDIT) %>%
  summarise(avg_sentiment = mean(PROP_21, na.rm = TRUE), .groups = "drop")

node_sentiment <- sapply(unique_subs, function(sub) {
  idx <- match(sub, sentiment_estimates$SOURCE_SUBREDDIT)
  if (!is.na(idx)) sentiment_estimates$avg_sentiment[idx] else NA
})

## -- Fix: mean-impute any missing sentiment values
mean_sent <- mean(node_sentiment, na.rm = TRUE)
node_sentiment[is.na(node_sentiment)] <- mean_sent
```

```
# Build similarity matrix (negative absolute distance)
sent_sim_matrix <- outer(
  node_sentiment,
  node_sentiment,
  FUN = function(x, y) -abs(x - y)
)
diag(sent_sim_matrix) <- 0 # zero diagonal
```

### Model 3

Prepping edgelist.  
Checking/prepping covariates.  
Computing preliminary statistics  
Fitting model  
Obtaining goodness-of-fit statistics

Relational Event Model (Ordinal Likelihood)

	Estimate	Std.Err	Z value	Pr(> z )
NTDegRec	1.0865e+01	4.7622e-01	22.8141	< 2.2e-16 ***
RRecSnd	3.4036e-01	8.2076e-02	4.1469	3.369e-05 ***
RSndSnd	2.5694e+00	5.2788e-02	48.6739	< 2.2e-16 ***
CovSnd.1	6.3243e-04	8.8939e+01	0.0000	1
CovEvent.1	1.0205e+00	1.4250e-01	7.1611	8.002e-13 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
Null deviance: 40290.08 on 2595 degrees of freedom  
Residual deviance: 36593.05 on 2590 degrees of freedom  
Chi-square: 3697.031 on 5 degrees of freedom, asymptotic p-value 0  
AIC: 36603.05 AICC: 36603.07 BIC: 36632.36

*Testing Hypotheses 3b: The likelihood of a subreddit  $i$  linking to subreddit  $j$  is greater if  $i$  and  $j$  have similar average sentiments. Measured by: CovInt with sentiment similarity.*

The p-value for CovEvent, the indicator of sentiment similarity, is well below 0.05, and therefore is statistically significant. The covariate is strongly positive, and when the estimate is exponentiated, we get  $\sim 2.77$ , meaning that for every one-unit increase in average-sentiment similarity between two subreddits, the instantaneous rate of change at which they link is  $\sim 2.8x$  higher. This confirms our hypothesis 3, telling us that subreddits with more aligned sentiments are more likely to connect. In the context of echo chamber formation, it indicates that there are strong ideological homophily-communication flows along lines of shared opinions, which limits exposure to different view points. The control terms (discussed in previous models) show that links accumulate on past dyads and popular subreddits, which show tight information loops. Overall, this model tells us that Reddit hyperlink traffic is driven by ideological alignments of sentiment (along with the other factors we've previously discussed). When put together, these dynamics (so far) mimic the characteristics of echo chambers where users repeatedly engage with like-minded people without communicating much with others with different sentiments.

## Model 4

```
Prepping edgelist.
Checking/prepping covariates.
Computing preliminary statistics
Fitting model
Obtaining goodness-of-fit statistics
```

Relational Event Model (Ordinal Likelihood)

	Estimate	Std.Err	Z value	Pr(> z )
NTDegRec	1.0902e+01	4.7621e-01	22.8938	< 2.2e-16 ***
RRecSnd	3.0029e-01	8.3706e-02	3.5874	0.0003340 ***
RSndSnd	2.5696e+00	5.2777e-02	48.6875	< 2.2e-16 ***
CovSnd.1	6.3236e-04	8.9915e+01	0.0000	0.9999944
CovEvent.1	1.0222e+00	1.4249e-01	7.1743	7.27e-13 ***
PSAB-BA	1.1569e+00	3.1233e-01	3.7040	0.0002122 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
Null deviance: 40290.08 on 2595 degrees of freedom  
Residual deviance: 36583.05 on 2589 degrees of freedom  
Chi-square: 3707.032 on 6 degrees of freedom, asymptotic p-value 0  
AIC: 36595.05 AICC: 36595.08 BIC: 36630.22

*Testing Hypotheses 4: When an subreddit i links to another subreddit j, j has a greater likelihood of linking back to subreddit i in the next event.*

The p-value of PSAB-BA is far below 0.05, so the results are statistically significant. The estimate is ~1.16, which when exponentiated, yields ~3.18. This means that when j has just linked to i, the likelihood that i immediately links back increases by more than 3x. Hypothesis 4 tells us about the turn taking reciprocity, which is confirmed by the high, positive PSAB-BA. This means subreddits reciprocate quickly, which reinforces dyadic loops. The presence of turn taking can signal stronger echo chambers, since it tells us that subreddits are repeatedly linking to each other in a reinforcing loop instead of reaching out to new communities. Tight feedback loops of information are created as a result, which are an essential signifier of echo chamber formation.

## Model 5

```
Prepping edgelist.
Checking/prepping covariates.
Computing preliminary statistics
Fitting model
Obtaining goodness-of-fit statistics
```

#### Relational Event Model (Ordinal Likelihood)

	Estimate	Std.Err	Z value	Pr(> z )	
NTDegRec	1.1001e+01	4.7621e-01	23.1006	< 2.2e-16	***
RRecSnd	3.0578e-01	8.4229e-02	3.6304	0.0002830	***
RSndSnd	2.5674e+00	5.2717e-02	48.7023	< 2.2e-16	***
CovSnd.1	-6.2675e-05	9.0247e+01	0.0000	0.9999994	
CovEvent.1	1.0277e+00	1.4271e-01	7.2014	5.962e-13	***
PSAB-BA	1.1723e+00	3.1241e-01	3.7525	0.0001751	***
PSAB-BY	9.0007e-01	9.5567e-02	9.4182	< 2.2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Null deviance: 40290.08 on 2595 degrees of freedom

Residual deviance: 36514.5 on 2588 degrees of freedom

Chi-square: 3775.578 on 7 degrees of freedom, asymptotic p-value 0

AIC: 36528.5 AICC: 36528.55 BIC: 36569.53

*Testing Hypotheses 5: The likelihood of a subreddit i linking to a subreddit k is greater if i links to a third subreddit j, and j links to k.*

The p-value of PSAB-BY is statistically significant. PSAB-BY measures the information-chain shift: after subreddit i links to subreddit j, j quickly links to a new subreddit k. The coefficient is positive, and it's exponentiated form is ~2.46. This tells us that the likelihood that j links to a third party increases by ~2.5x in comparison with the baseline. PSAB-BY confirms our hypothesis that once two subreddits are engaged, the second subreddit frequently reaches outward to link to new subreddits. This seems like healthy information diffusion at first, but when paired with the other network dynamics we've observed, it reinforces the echo-chamber processes. We learned that there's ideological homophily (that links are more likely made to other subreddits with similar sentiments) and turn taking reciprocity (dyads bounce content back and forth). With the addition of information chains, this tells us that when the link outward is made, it's likely making a connection with another subreddit that's like-minded, which just extends the echo chamber cluster instead of bridging dissenting communities. These chains of communication accelerate the internal reinforcement of narratives while not assisting with cross-cluster exposure, which is a classic marker of echo chambers.

#### Model 6

Prepping edgelist.

Checking/prepping covariates.

Computing preliminary statistics

Fitting model

Obtaining goodness-of-fit statistics

#### Relational Event Model (Ordinal Likelihood)

	Estimate	Std.Err	Z value	Pr(> z )	
NTDegRec	1.1001e+01	4.7621e-01	23.1004	< 2.2e-16	***
RRecSnd	3.0580e-01	8.4229e-02	3.6306	0.0002828	***

```
RSndSnd      2.5674e+00 5.2717e-02 48.7021 < 2.2e-16 ***
CovSnd.1     6.3235e-04 9.0247e+01  0.0000 0.9999944
CovEvent.1   1.0277e+00 1.4271e-01  7.2015 5.955e-13 ***
PSAB-BA      1.1724e+00 3.1240e-01  3.7527 0.0001749 ***
PSAB-BY      9.0008e-01 9.5567e-02  9.4183 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Null deviance: 40290.08 on 2595 degrees of freedom
Residual deviance: 36514.5 on 2588 degrees of freedom
    Chi-square: 3775.578 on 7 degrees of freedom, asymptotic p-value 0
AIC: 36528.5 AICC: 36528.55 BIC: 36569.53
```

*Testing Hypotheses 6: The rate of subreddit  $i$  linking to other subreddits within its cluster increases over time.*

With hypothesis 6, we wanted to check if burst dynamics could be visualized in the subreddit data. The CovSndTime was intended to tell us whether active subreddits become progressively more likely to link as times go on to answer this hypothesis. However, the CovSndTime effect was not outputted. This could potentially mean that there was an identical value at every event (having zero variance), or that it was perfectly colinear with the other terms. Because of this, H6 cannot be evaluated.

## Model Comparisons

We will compare the BIC of all models that produced successful results (excluding 3a and 6):

Model 1 BIC: 37137.56 Model 2 BIC: 36679.6 Model 3b BIC: 36632.36 Model 4 BIC: 36630.22 Model 5 BIC: 36569.53

The lowest BIC is model 5, as it has a BIC of 36569.53, meaning we prefer model 5. This means that with preferential attachment, reciprocity and repetition, sentiment similarity, turn taking, and information chains all included in the model, we're best able to model and explain the linkage patterns between subreddits in the overall Reddit network.

```
save.image(file = "Echo_Chamber_REM_descriptive.RData")
```