Cyberbalkanization Online Supplement

XXX

5/11/2017

1 Results of community detection in the Hong Kong Facebook Pages sharing network

Names of the communities were connoted by the first authors according to each community's central political ideology and main topics of the member pages and were then verified by the second author.

Community Name	Top page with the highest betweenness centrality	Number of pages
Pro-activists #1	蘋果日報 (newspaper)	605
Pro-activists #2	健吾 (blogger)	390
Autonomists	聚言時報 Polymer (online media)	377
Conservationists	集雜志 Zine (online media)	69
Blue Ribbon #1	佔中不代表我 (issue based)	60
Teen Magazines	東TOUCH (magazine)	53
Blue Ribbon #2	時聞香港 (online media)	37
Trivia	我係香港人 I am a Hongkonger (issue based)	32
Radical Media	寰雨膠事錄 PlasticNews.wf (online media)	31
Movie and entertainment	Edko Films Ltd. 安樂影片 (movie distributor)	30

2 Robustness of the main conclusion

In this section, we attempted to reanalyse our data to test the robustness of our main conclusion: the cyberbalkanization index is a leading indicator of political polarization index.

We attempt to use other assumptions to verify our findings.

# 2.1 Robustness against interpolation-related assumptions
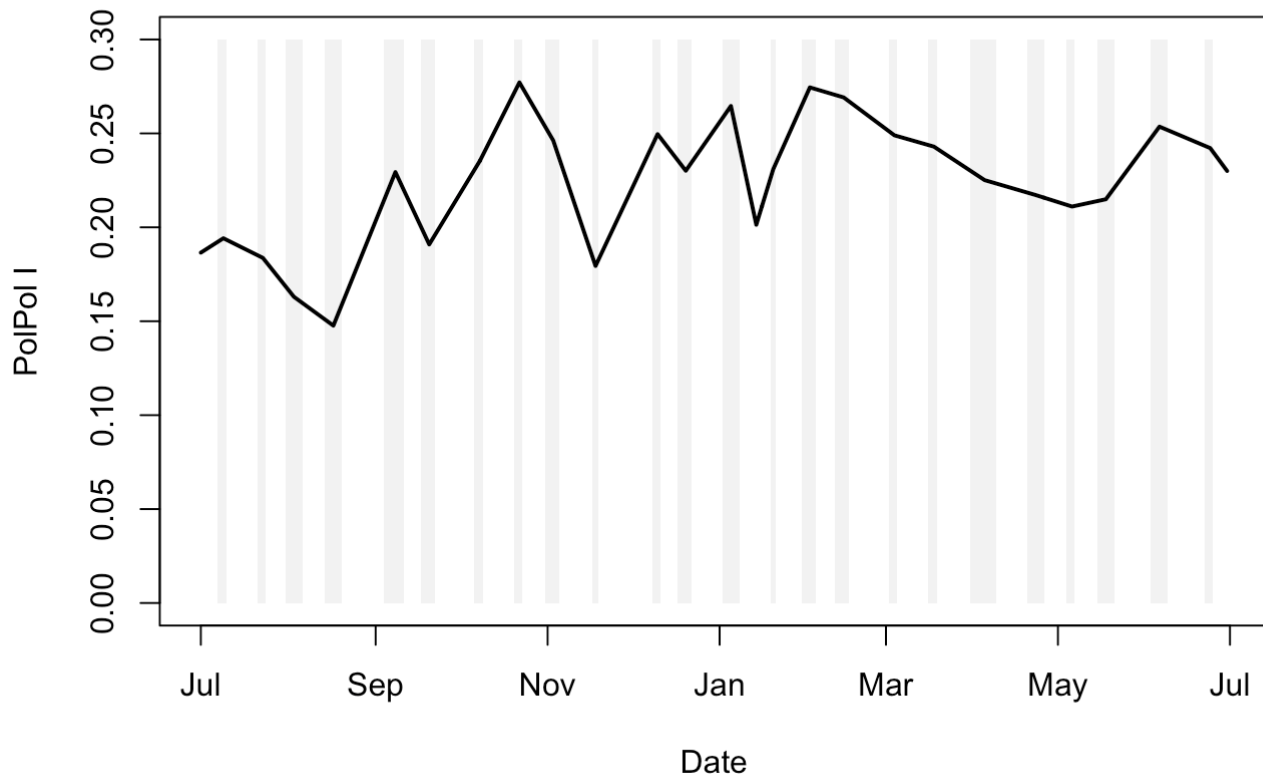
The linear interpolation of Pol Pol I has two assumptions and they are:

1. The telephone polls were conducted over a short period (most of them in 3-days). For each telephone, the midpoint of this short period was used to represent the date of measurement to calculate the PolPol I.
2. The PolPol I dervied from each poll were linearly interpolated to create a time series of daily index.

Therefore, this interpolation assume the poll results represent the public opinion at those midpoints and the PolPol I changing speeds between each mid-dates (slopes) are constant.
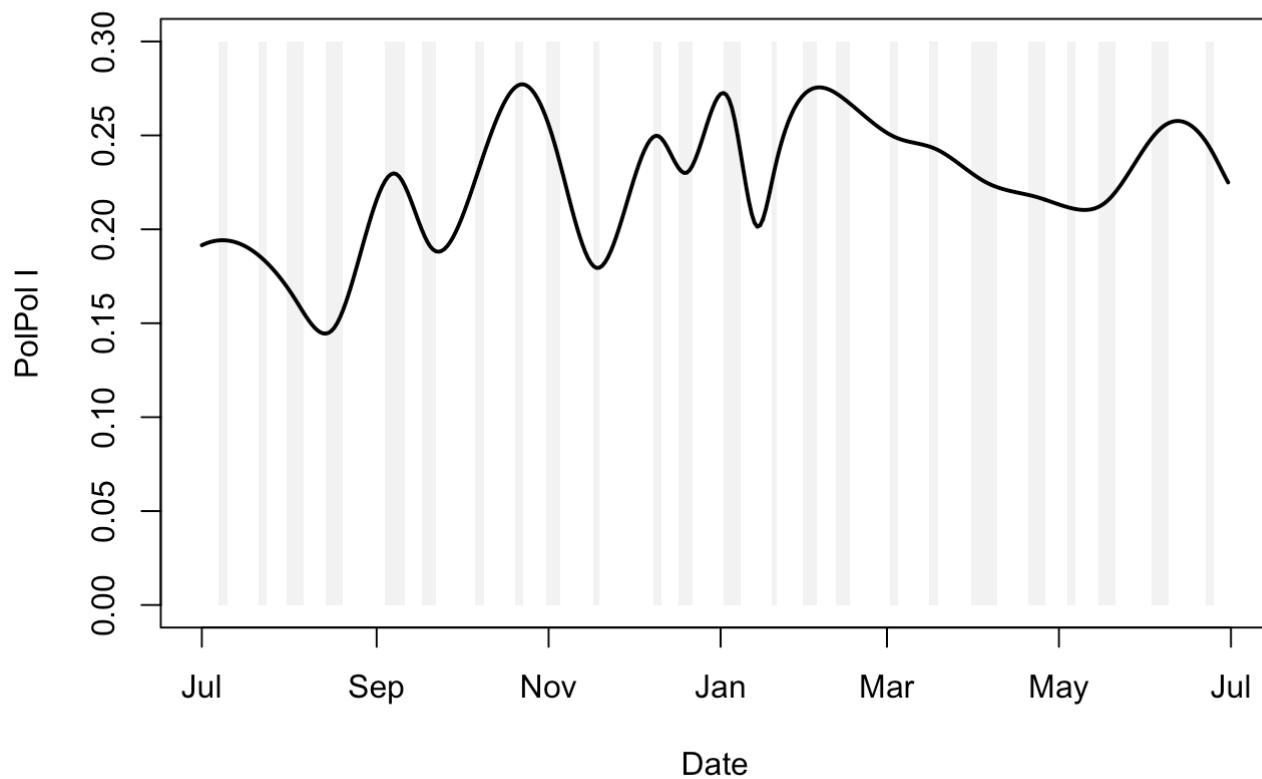
2.1.1 Interpolation between periods
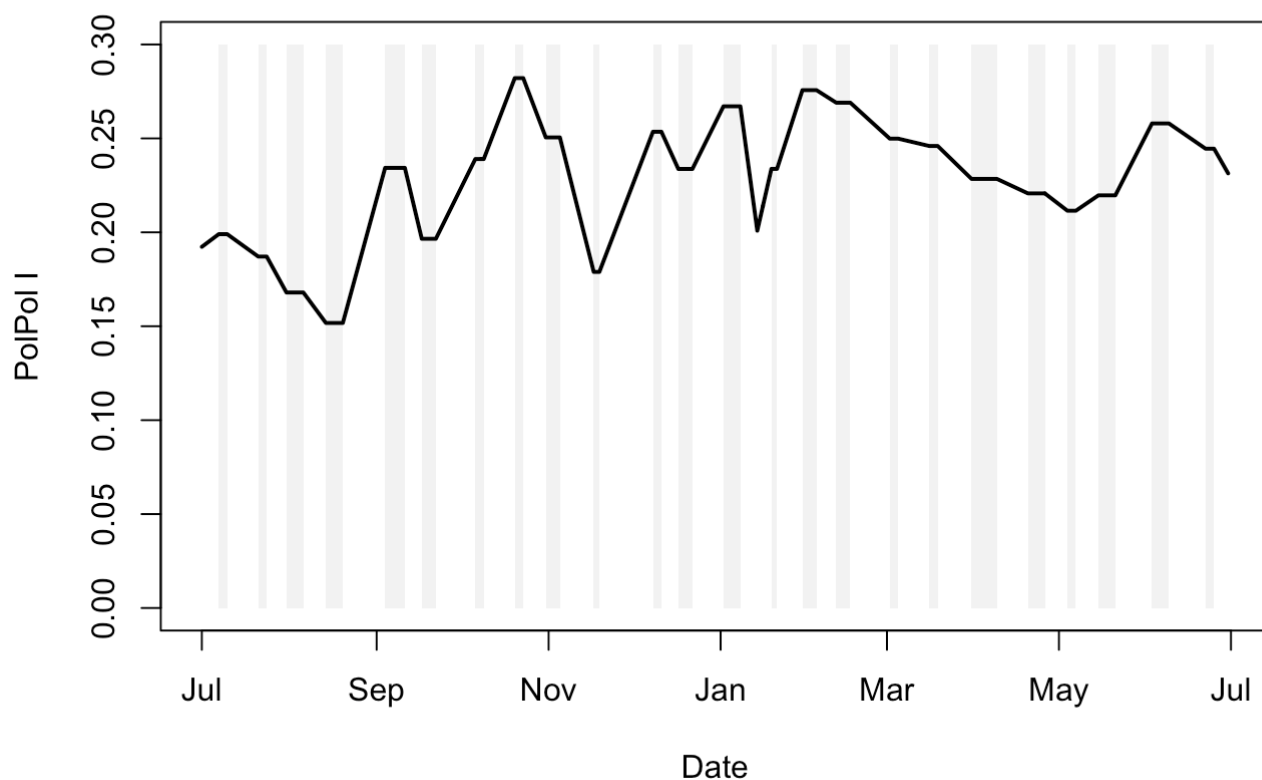
Instead of using midpoints, the entire periods were used to represent the date of measurement. The CCF using this PolPol I time series is:

## Diff, Linear Interpolation between periods
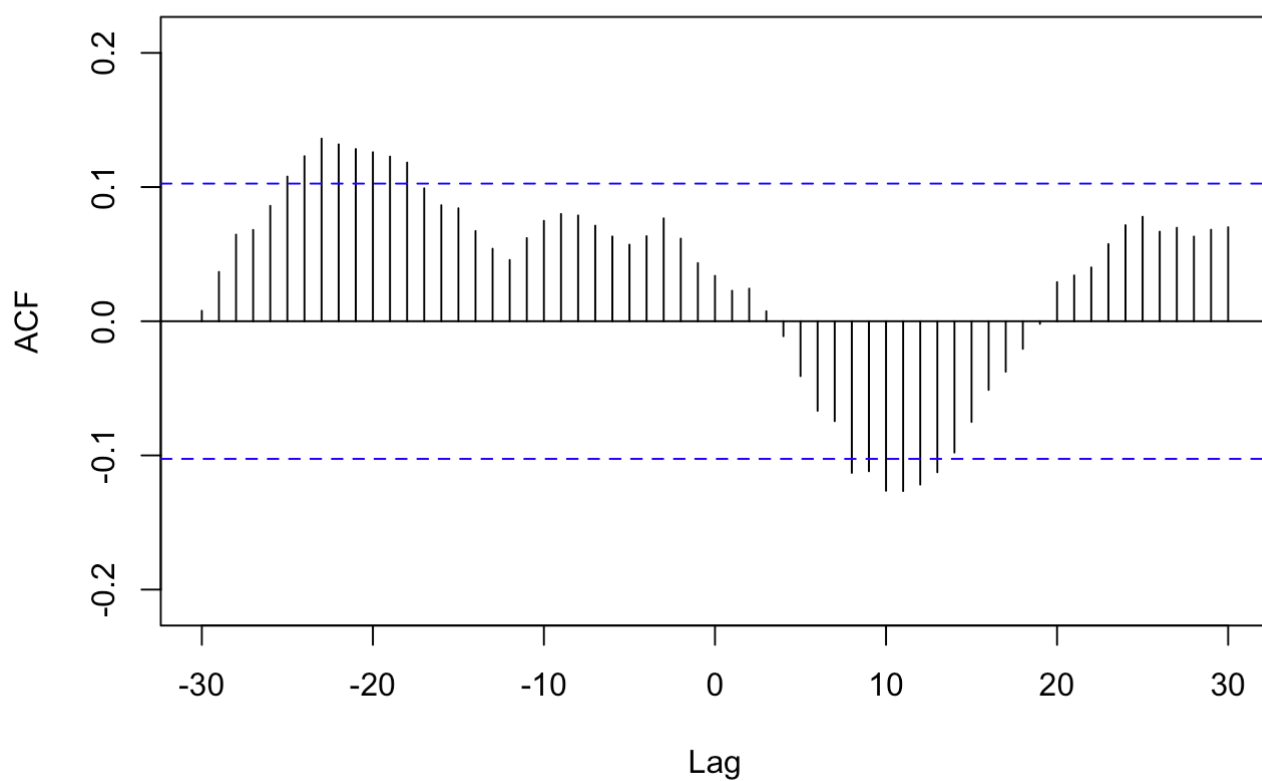
Positive cross-correlations were observed at negative lag units.

2.1.2 Cubic spline interpolation

With spline interpolation, the change in opinion polarization is assumed to be a cubic function, i.e. changes in public opinion are slower around the midpoints. The CCF using this PolPol I time series is:

Diff, Spline Interpolation between midpoints
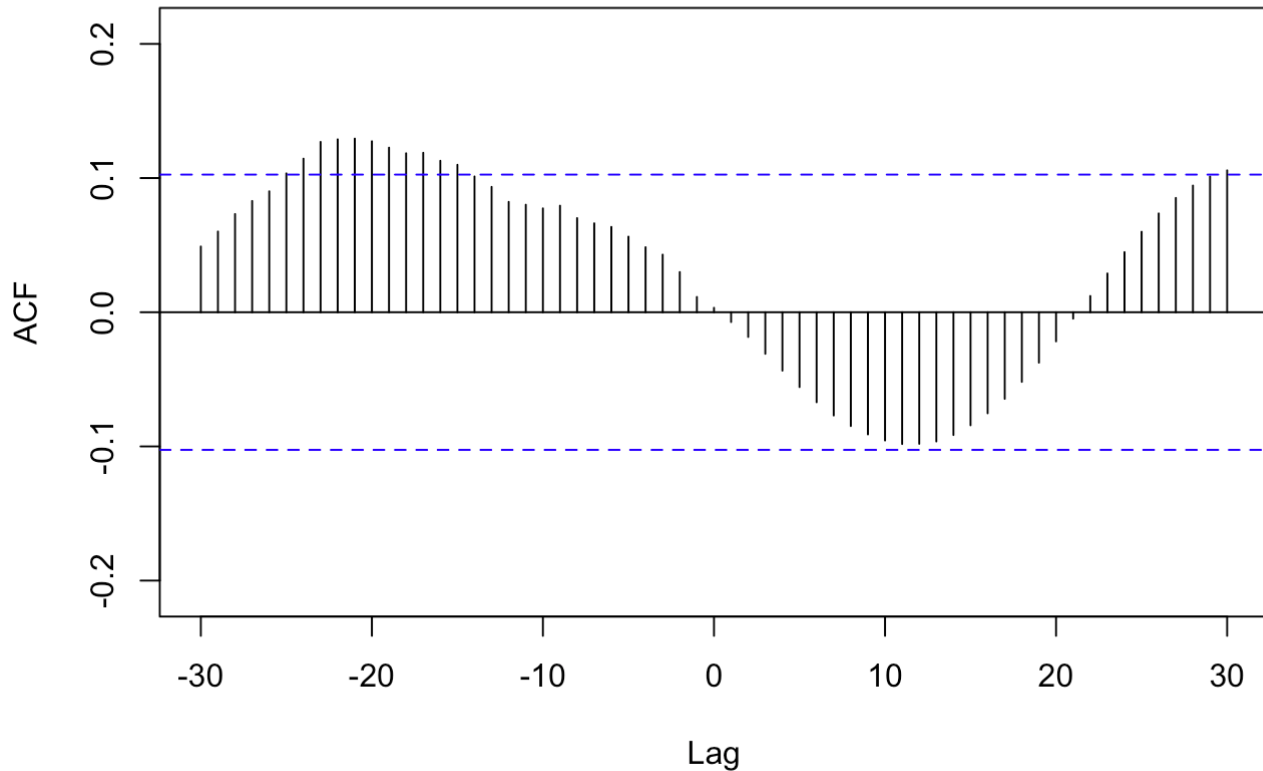
Positive cross-correlations were observed at negative lag units.

# 2.2 Robustness against the intrinsic lag of telephone polls

Because the telephone polls are not conducted daily, the PolPol I time series maybe intrinsically lag behind the online / offline events or comments. In this two robust analysis, we attempt to either assume PolPol I react more quickly or assume the CBI influence slowly.

2.2.1 Last observation carried backward
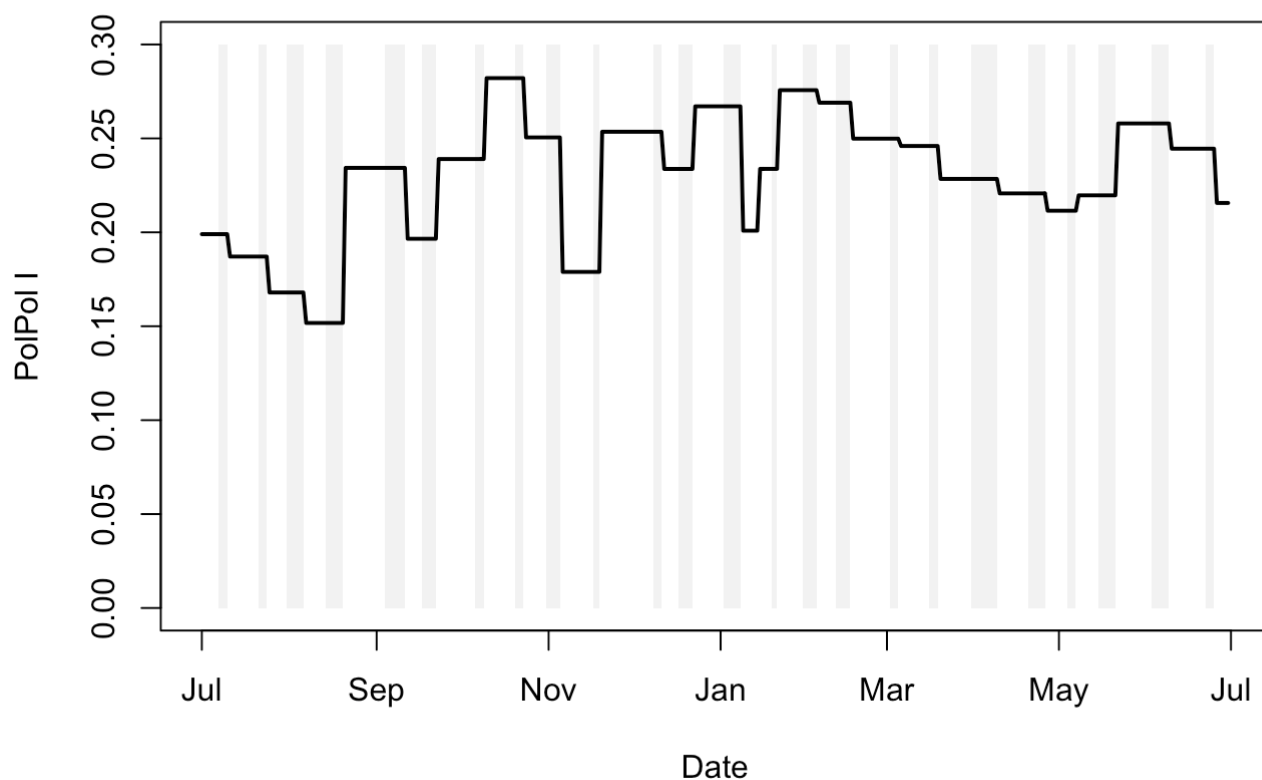
Using the "last observation carried backward" technique, this interpolation assumes the last date of the telephone poll period as the final status of the public opinion. The value for the day after the last date is imputed with the value from the next poll. This interpolation assumes telephone polls react to the public opinion much faster than linear interpolation. The CCF using this PolPol I time series is:

Diff, Last Observation Carried Backward
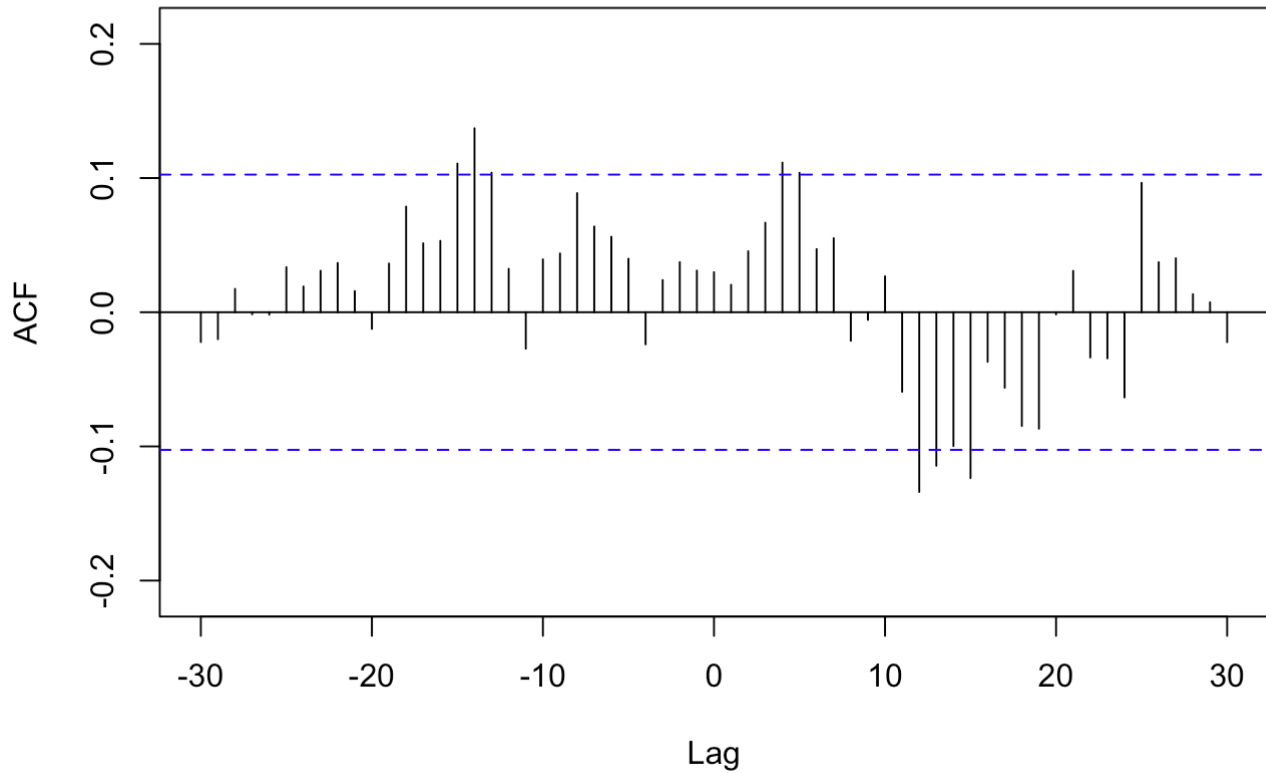
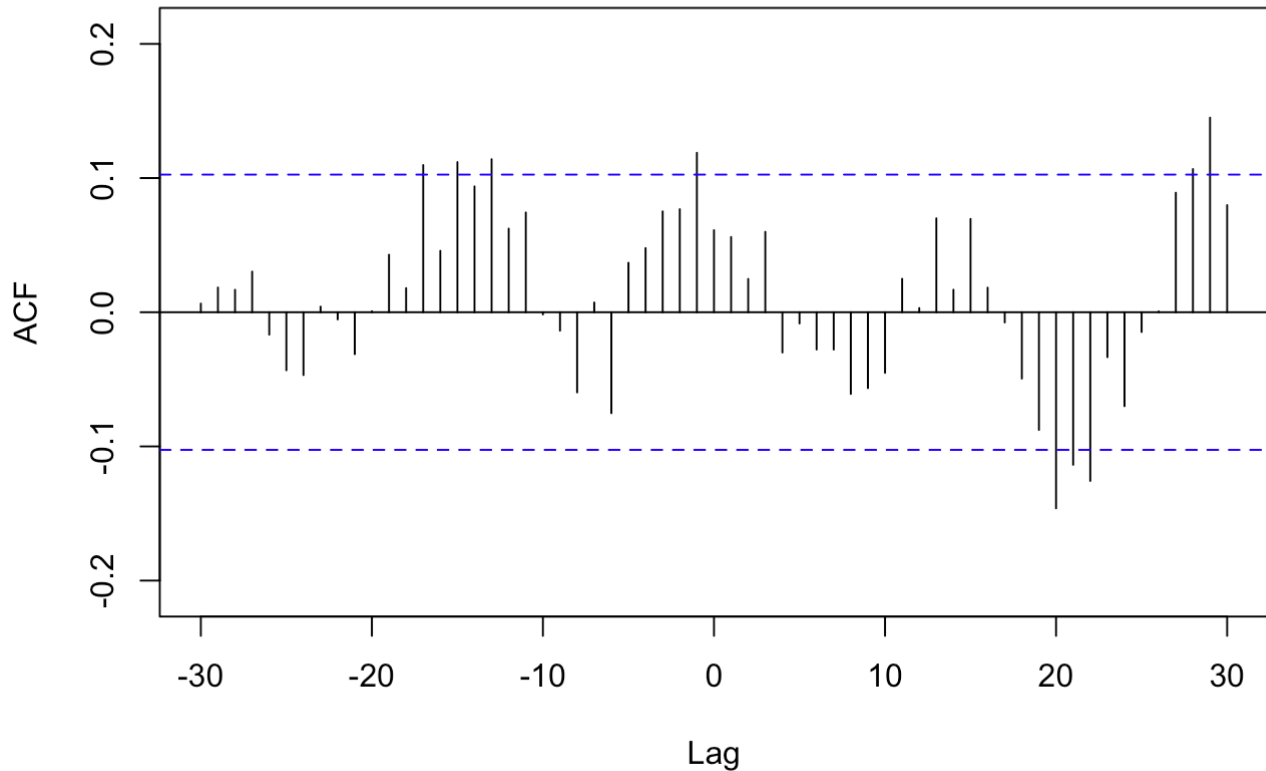Positive cross-correlations were observed at negative lag units.

## 2.2.2 Moving average of Cyberbalkanization index
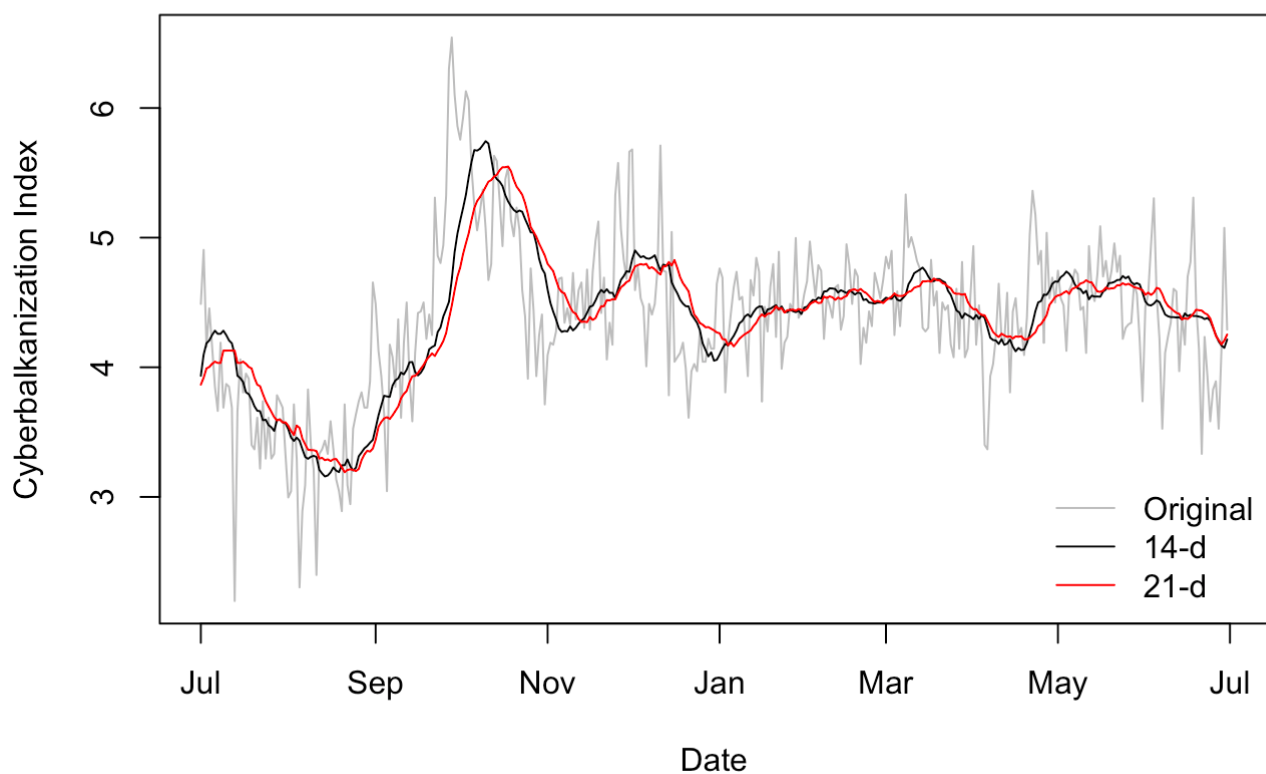
The 14-day and 21-day moving averages of the CBI were calculated. The reason for choosing 14- and 21-day are two-folds: 1) A previous study (Fu & Chan. Cyberpsychol Behav Soc Netw. 2013 Sep;16(9):702-7) suggested that the online sentiment leads the HKUPOP telephone poll by 14 days; 2) The current study found that the CBI leads the PolPol I for over 20 days. The moving average time series assume the influence of CBI is cummulative and therefore reaction slowly to the change in public opinion. The CCFs using the moving average time series are:

Diff, 14-day Moving average of CBI

Diff, 21-day Moving average of CBI
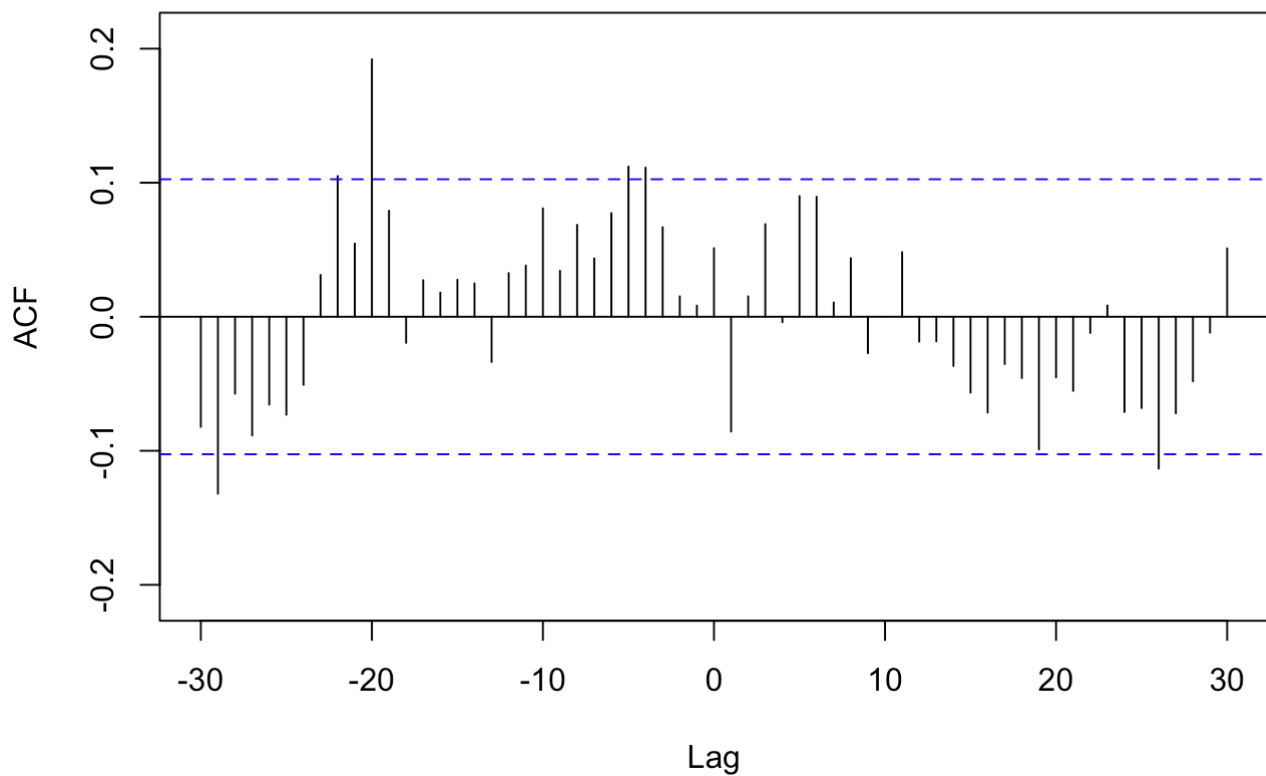
Positive cross-correlations were observed at negative lag units.

3 Model-based approach and forecasting

3.1 Regression

A regression model was constructed with the prewhiten pairs of CBI and PolPol I. As the CCF found a 21-day lead for CBI, the prewhiten CBI was lagged for 21 day in the analysis.

	Dependent variable:
	PolPol I (prewhiten)
CBI (prewhiten, 21-day lag)	9.288***
	(3.151)
Constant	-0.006
	(0.020)
Observations	344
R^2	0.025
Adjusted R^2	0.022
Residual Std. Error	0.378 (df = 342)
F Statistic	8.688*** (df = 1; 342)
Note:	$p<0.1$; ***p<0.05;*** $p<0.01$

The CBI is a very signficant predictor of PolPol I ($p < 0.001$).

3.2 Forecasting

We have created a prediction model of PolPoll using the CBI according to the community structure and 21-day-moving average regression model derived from the data between July 2014 and June 2015. This model is deployed to forecast the PolPoll between July 1, 2015, and August 31, 2015. Based on the results, our simple model are slightly decalibrated in the forecasting period but can still predict the trend of future PolPoll using the CBI with R^2 equals to 0.222, indicating the validity of the model prediction.