

# Social Network Analytics, Empirical Exercise #5

## Due on Sunday, December 10, 2017 at 11:59pm

### Parliamentary elections and droughts in India

**Background on data:** This exercise analyzes electoral data from each Lok Sabha—the lower house of Parliament in India—election from 1951 until 1999. These data are compiled at the district level. The electoral data is combined with meteorological data that details each district’s monthly level of rainfall. We can use the occurrence of extreme weather events in a district as a proxy for economic disruption. We will use the data to answer the question of whether abnormal levels of rainfall during the time leading up to an election cause more political parties to enter into a district during that election. We will also use the data to find out whether there is a geographical diffusion process of political parties entering into districts. One nice feature of this data is that the initial catalyst of political parties entering a region, rainfall, is random, so the regression is able to isolate and model political activity that is essentially stimulated by chance.

1. The file “district\_information.csv” contains a table with many details about each district, compiled over each of the 12 election periods. Not all districts exist for all 12 election periods (some districts are created later on, and some districts dissolve). The data contain columns with information about the formation and geographical growth of political parties:
  - columns with a “new\_parties” indicator are a count for how many new parties contested a Lok Sabha seat in a district—i.e., how many new active parties that have not been active in this district before
  - columns with a “first\_foundings” indicator are a count for how many *brand* new parties contested a Lok Sabha seat in a district—i.e., how many new active parties that have not been active in *any* district before, and are appearing in the data during this election period for the first time
  - columns with a “franchise\_foundings” indicator are a count for how many *franchise* new parties contested a Lok Sabha seat in a district—i.e., how many new active parties that have been active in a *different* district before, but not this particular district
  - “voting\_participation” is the percentage of registered voters that turned out to cast a vote

There is also information about the meteorological conditions of each district, compiled over the election period:

- raw measure of rain fallen in a district over the election—this is a continuous measure of weather
- standardized measure rain fallen in a district, standardized within a district, compared to its historical average rainfall—this is a continuous measure of weather
- whether districts experienced droughts, floods, or a combination of the two during the election period. Droughts and floods are defined as periods of low or high rainfall during an election period that are more extreme than a district’s historical tendencies for this type of weather. So, this is a within-district comparison as well—this is a binary measure of weather

First, let’s make sure that our droughts and floods are actually random from one election period to the next. Let’s also make sure that a neighbor experiencing a drought or flood in the previous period isn’t related to a district experiencing a drought or flood in the current election period. Run a regression illustrating that experiencing a drought or flood in the previous period is not associated with experiencing a drought or flood in the current period, and that a neighbor experiencing a drought or flood in the previous period is also not associated with experiencing a drought or flood in the current period. Perform the same test for standardized rainfall. What does this suggest?

2. Next, let's figure out if there are more political parties when there are mild droughts or floods. Since the new political parties that enter a district is a count outcome, we should use a regression format adopted for counts. The data are overdispersed, so use a negative binomial model (check the MASS library). Include a district-level fixed effect, i.e., control for the district in the regression. The district-level fixed effect controls for any underlying difference between districts that might be correlated with its political or economic conditions. What if there are moderate droughts or floods? Also use a regression to illustrate if there are more or less new political parties when standardized rainfall is high, and also include a district-level fixed effect.
3. Now that we have established the baseline effect, we can look at how political activity stimulated by droughts or floods in one district might affect political activity in an other district. Use a regression to show that, even when taking into account a district's own droughts and floods, that district will have more political parties enter when its neighboring districts experience droughts or floods in the previous election period. Focus the regression on moderate droughts or floods, and include a district-level fixed effect. Does a similar pattern exist for standardized rainfall as well?
4. What kinds of parties tend to get formed as a result of a district experiencing droughts or floods, and its neighbors experiencing droughts or floods in the previous period? Use regressions to illustrate whether droughts and floods tend to produce brand new parties that have never appeared in the political system before (`first_foundings`) or franchise parties that have contested elections in other districts in previous election periods (`franchise_foundings`). Why might this be the case? Focus the regression on moderate droughts or floods, and include a district-level fixed effect.

**Extra challenge problem:** It is possible to ask whether the diffusion of political parties entering may hop several borders at one time. Extend the regression model from Question 4 to show whether the droughts or floods experienced two districts away in the previous election period have any influence on the franchise parties that enter into a district in the current election period. What about the droughts or floods that appear two districts away, two election periods ago? What do these results suggest about the diffusion process of political parties?