October 23, 2018

# Social Network Analytics, Empirical Exercise #3
## Due on Wednesday, October 21, 2017 at 12:00pm

### Diffusion of political parties founded after droughts in India

**Setting up diffusion 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 paired 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 diffusion process of political parties entering into districts. A striking feature of this data is that the initial catalyst of political parties entering a region, rainfall, is random, so the analysis is able to isolate and model the diffusion of political activity that is essentially stimulated by chance.

- The file "district_information.csv" contains an table with details about each district, compiled over each of the 14 election periods. Not all districts exist for all 14 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" prefix represent counts for how many new parties of each type contested a Lok Sabha seat in a district—a count of how many new active parties appear in a district that have not been active in this district before
  - "total$_p$$arties''represents the total number of parties contesting seats in a district in a period--the labels for party types overlap, so the total of the new_parties columns may not always sum to total_partie$
  - the column "political_concentration" is a Herfindahl index of concentration in each district, which measures the degree to which a few parties command the majority of the vote share:

$$Herfindahl = \sum_{i}^{n} (market\ share_i)^2$$

- The file "rainfall_information.csv" contains a table with yearly rainfall totals for electoral districts in India. The districts are coded to uniquely match the districts in "district_information.csv" without needing to refer to states. Two columns indicate rainfall measures for each district in each year:
  - the raw level of rainfall, measured in milliliters
  - the Standardized Precipitation Index, a transformation used by meteorological scientists that normalizes the raw rainfall according to a the historical average for the region using a Pearson type III distribution

- The file "border_information.csv" contains an edge list illustrating which districts share borders. Each row represents a border-sharing relationship between two districts, and this relationship is undirected.

1. First, we will set up the relationship between rainfall and political party foundings, and then modify the rainfall measure to generate a statistically independent measure for droughts. This modification will allow us to isolate the effect of economic strain from other underlying features of a region.

   (A) Create a figure, for example, a scatter plot, showing the visual relationship between the level of rainfall in a district in the period leading up to the current election, and the number of political parties that are founded in a region. You can use the raw rainfall measure or the Standardized Precipitation Index. You can consider the level of rainfall for each election period in terms of (1) the sum of the raw rainfall during the interval starting from the year following the previous election up until the year of the current election or (2) the yearly average of the Standardized Precipitation Index during the interval starting from the year following the previous election up until the year of the current election.

(B) Using the election-period level rainfall measures created above, show that the raw level of rainfall, as well as the Standardized Precipitation Index, are not independent from one election period to the next within a district, as well as from neighboring districts from one election period to the next. It is possible to show this relationship by regressing a district's current level of the rainfall variable on (1) its lagged value and (2) the lagged value of its neighbors' rainfall variable. For computing the neighbors' value, you can use an average of each of the surrounding districts' values. Include a control in the regression for the number of years in the election period, and use a fixed effects specification to control for the time-invariant features of a district as well as a control for each election period. This can be accomplished using the plm package, using a model specified in the form of `plm(outcome variable  predictor variables, data, effect = "twoways", model = "within", index = "district")`, where "twoways" "within" provide both sets of fixed effects.

(C) Meteorological scientists consider moderate droughts to occur if the Standardized Precipitation Index falls below -1, and moderate floods to occur if it rises above 1. Create a measure that sums the number of years a district experiences either moderate droughts or floods during the interval starting from the year following the previous election up until the year of the current election. Perform the same test as in (B), using this new transformed measure. This measure will form the basis for the predictors used in the remainder of the regressions in Questions 2-5. Since this is a count outcome that is reported as a discrete number of years, use a regression adopted for data of this form—this can be accomplished with the pglm package, using a model specified in the form of `pglm(outcome variable  predictor variables, data, effect = "twoways", model = "within", index = "district", family = "poisson")`. What differences do you see between the estimates?

2. Next, let's analyze whether there are more new political parties when droughts or floods occur.

   Run a regression predicting the number of new political parties that are formed as a function of the number of years a district experiences droughts or flooding in the interval starting from the year following the previous election up until the year of the current election. The number of new political parties that enter a district is a discrete count outcome, so we should use a regression format adopted for counts, as in (1C). Include a control in the regression for the number of years in the election period, and a control for the time-invariant features of a district, as in Question 1. Are certain kinds of political parties more likely to be formed when a district experiences extreme weather?

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 another district.

   Use a similar regression to Question 2 to show that, even when taking into account a district's own droughts and floods, that district's degree of political foundings will also depend on the number of years its neighboring districts experience years of droughts or flooding in the interval starting from the year following two elections ago, up until the year of the previous election—the election lead-up interval before the current one. Include a control in the regression for the number of years in the current election period, and a control for the time-invariant features of a district.

4. Extreme weather events like droughts or floods can erode the stability of political systems and wear away at the entrenched power bases of large, national-scale parties that have difficulty responding to the needs of affected regions.

   Does experiencing droughts or floods relate to political concentration? Perform a regression, similar to Question 3, predicting the Herfindahl Index of a region as a function of the number of years of droughts or flooding that occur in a district in the interval leading up to the current election, and the number of years of droughts or flooding that occur in its neighboring districts in the interval leading up to the previous election. Include a control in the regression for the number of years in the election period, and a control for the time-invariant features of a district. What does this result illustrate in terms of the concentration or fragmentation of political power in districts affected by extreme weather?

5. Political parties are formed to accomplish a variety of goals. Individual parties can also exist in the context of larger social and cultural trends, especially when regions influence each other as political organizing activity diffuses across regions over time. To understand the diffusion process more, we want to analyze whether the new parties that appear in a district are the same parties that have appeared in neighboring districts in the past, or if it the process of political organization, rather than the content of a specific political party, that is diffusing.

   To analyze this, run two separate regressions predicting the likelihood of (1) new political parties being founded in a district, that have contested an election in a neighboring district in any previous election period, and (2) new political parties being founded in a district that have not contested an election in a neighboring district in any previous election period.

   As in Question 3, estimate these as a function of the number of years of droughts or flooding that occur in a district in the interval leading up to the current election, and the number of years of droughts or flooding that occur in its neighboring districts in the interval leading up to the previous election. Include a control in the regression for the number of years in the election period, and a control for the time-invariant features of a district. What does this illustrate about the diffusion of political organizing?