Beatrice Garcia

April 11, 2017

Module 9 Homework

The dataset is based off of 100 Nascar tweets. There is an adjacency matrix in the form of a CSV. There is the CSV of the tweets with their ID, created date, the user that Retweeted or sent out the tweet, the raw text, the users the tweet mentioned (with the “@” symbol), and the type of tweet it is (Retweet or original).

Both the tweets and adjacency matrix are stored as a CSV. The nascar\_tweets.csv is actually “|” delimited because there were a lot of commas in the tweets and dates that made the CSV reader read it wrong. The adjacency matrix is as expected.

I collected the data using the free Twitter Rest API with a Python wrapper. Using the wrapper, I retrieved 100 tweets with the term “nascar” or “#nascar” in it as that was one of the trending hashtags of that day. Then, I iterated through all the retrieved tweets and created the CSV with tweet’s ID, user name, created date, raw text, users mentioned, and if it was a Retweet or original content. The raw text was cleaned up a little so that there were no newline characters and deleted all commas so the CSVs are correctly interpreted later on.

Data Details

|  |  |
| --- | --- |
| Response rates/sampling rate | Free Twitter Rest API with targeted search on “nascar”. The Python Twitter Wrapper grabs 100 tweets in one pull. |
| Any theoretical grounding for questions or methods employed | “#nascar” was one of the top five trending hashtags at the time of the pull. |
| A short description of the context | Nascar is a sport involving racing cars. |
| Nature of the respondents | Respondents are affected by at least one social force, homophily. |
| Details about collection intervals | Data was collected once. |
| Temporality of the data | Data was collected on April 9, 2017 around 9pm EST. |
| Known issues that threaten the validity of the data or anything else other social network analysts using these data for teaching or research should be aware of | Data is sparse and randomly collected. |

ERGM model summaries of:

1. Initial restricted model including edges, reciprocity (if undirected), and a triadic term.

==========================

Summary of model fit

==========================

Formula: ga.net ~ edges+triangles

Iterations: 7 out of 20

Monte Carlo MLE Results:

Estimate Std. Error MCMC % p-value

edges -4.7449 0.1059 0 <1e-04 \*\*\*

---

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

Null Deviance: 14473 on 10440 degrees of freedom

Residual Deviance: 1035 on 10439 degrees of freedom

AIC: 1037 BIC: 1044 (Smaller is better.)

1. Final restricted model with significant terms.

==========================

Summary of model fit

==========================

Formula: ga.net ~ edges + triangle

Iterations: 2 out of 20

Monte Carlo MLE Results:

Estimate Std. Error MCMC % p-value

edges -4.7314 0.1086 0 <1e-04 \*\*\*

triangle -Inf 0.0000 0 <1e-04 \*\*\*

---

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

Null Deviance: 14473 on 10440 degrees of freedom

Residual Deviance: NaN on 10438 degrees of freedom

AIC: NaN BIC: NaN (Smaller is better.)

1. Unrestricted model with at least one nodal covariate included

==========================

Summary of model fit

==========================

Formula: ga.net ~ edges + nodematch("type")

Iterations: 7 out of 20

Monte Carlo MLE Results:

Estimate Std. Error MCMC % p-value

edges -4.5555 0.1482 0 <1e-04 \*\*\*

nodematch.type -0.3553 0.2118 0 0.0934 .

---

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

Null Deviance: 14473 on 10440 degrees of freedom

Residual Deviance: 1032 on 10438 degrees of freedom

AIC: 1036 BIC: 1051 (Smaller is better.)

1. Unrestricted model with at least one edge covariate included.

==========================

Summary of model fit

==========================

Formula: ga.net ~ edges + nodematch("text")

Iterations: 7 out of 20

Monte Carlo MLE Results:

Estimate Std. Error MCMC % p-value

edges -4.6821 0.1083 0 <1e-04 \*\*\*

nodematch.text -0.8995 0.5125 0 0.0793 .

---

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

Null Deviance: 14473 on 10440 degrees of freedom

Residual Deviance: 1031 on 10438 degrees of freedom

AIC: 1035 BIC: 1049 (Smaller is better.)